Wed. Jun 6, 2018

Room A

Oral presentation | General Session | [General Session] 2. Machine Learning

[2A1] [General Session] 2. Machine Learning

座長:竹内 孝（NTTコミュニケーション科学基礎研究所）
9:00 AM - 10:40 AM Room A (4F Emerald Hall)

[2A1-01] Dual Convolutional Neural Network for Graph of Graphs Link Prediction

○Shonosuke Harada[1], Hirotaka Akita[1], Masashi Tsubaki[2], Yukino Baba[1], Ichigaku Takigawa[3], Yoshihiro Yamanishi[4], Hisashi Kashima[1,5] （1. Kyoto University, 2. National Institute of Advanced Industrial Science and Technology, 3. Hokkaido University, 4. Kyushu University, 5. RIKEN Center for Advanced Intelligence Project）

9:00 AM - 9:20 AM

[2A1-02] Predicting Eruptions of Sakurajima by Stacked Recurrent Neural Network

○Tsuyoshi Murata[1], Hiep Le[1], Masato Iguchi[2] （1. Tokyo Institute of Technology, 2. Kyoto University）

9:20 AM - 9:40 AM

[2A1-03] Anomaly Manufacturing Product Detection using Unregularized Anomaly Score on Deep Generative Models

Ryosuke Tachibana[1], Takashi Matsubara[1], ○Kuniaki Uehara[1] （1. Graduate School of System Informatics, Kobe Universiry）

9:40 AM - 10:00 AM

[2A1-04] Pseudo-feature generation from feature map in deep learning for imbalanced data multi-class image classification

○Tomohiko Konno[1], Hideaki Fujii[1], Michiaki Iwazume[1] （1. AI Science R&D Promotion Center in National Institute of Information and Communication Technology ）

10:00 AM - 10:20 AM

[2A1-05] Route Generation of Construction Machine by Deep Reinforcement Learning

○Shunya Tanabe[1], Zeyuan Sun[1], Masayuki Nakatani[1], Yutaka Uchimura[1] （1. Shibaura Institute of Technology）

10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 2. Machine Learning

[2A2] [General Session] 2. Machine Learning

座長:石畠 正和（NTTコミュニケーション科学基礎研究所）
1:20 PM - 3:00 PM Room A (4F Emerald Hall)

[2A2-01] Counting garbage bags from garbage collecting video using deep learning

○Kazuhiro Mikami[1], Makoto Kawano[1], Yin Chen[1], Jin Nakazawa[1] （1. Keio University）

1:20 PM - 1:40 PM

[2A2-02] Deep few-shot learning with pseudo example optimization

○Akisato Kimura[1,2], Zoubin Ghahramani[2], Koh Takeuchi[1], Tomoharu Iwata[1], Naonori Ueda[1] （1. Nippon Telegraph and Telephone Corporation, 2. University of Cambridge）

1:40 PM - 2:00 PM

[2A2-03] Verification of the harvesting date prediction method using deep learning

○Kengo Miyoshi[1], Yukikazu Murakami[1] （1. National Institute of technology, Kagawa College）

2:00 PM - 2:20 PM

[2A2-04] Visualizing the Behavior of the Inner Layers of Convolutional Neural Networks by Layer-wise Relevance Propagation

Hirotaka Sakai[1], ○Yoshitaka Kameya[1], Takahiro Sota[2], Hiroaki Arie[2] （1. Meijo University, 2. DENSO CORPORATION）

2:20 PM - 2:40 PM

[2A2-05] Solving Four Arithmetic Operation Problems and Word Algebra Problems by Neural Programmer-Interpreters

○Shota Katsumata[1], Katsumi Inoue[1,2] （1. Tokyo Institute of Technology, 2. National Institute of Informatics）

2:40 PM - 3:00 PM

Oral presentation | General Session | [General Session] 2. Machine Learning

[2A4] [General Session] 2. Machine Learning

座長:椿 真史（産業技術総合研究所）
5:20 PM - 7:00 PM Room A (4F Emerald Hall)

[2A4-01] Collecting whole sky images and classification of cloud genera and conditions

○Yu Morikawa[2], Haru Nakanishi[2], Naoki Inamura[1], Nobuaki Kondo[1], Hiroki Obuchi[3], Teruo Ohsawa[5], Takashi Matsubara[4], Kuniaki Uehara[4] （1. BANYAN PARTNERS Inc., 2. Kobe Digital Labo Inc., 3. SKY Perfect JSAT Corporation, 4. Graduate School of

©The Japanese Society for Artificial Intelligence

System Informatics, Kobe University, 5. Graduate School of Maritime Sciences, Kobe University)
5:20 PM - 5:40 PM

[2A4-02] Detecting community structure in layered neural networks for diagram recognition
○Chihiro Watanabe[1], Kaoru Hiramatsu[1], Kunio Kashino[1] (1. NTT Communication Science Laboratories)
5:40 PM - 6:00 PM

[2A4-03] A study on traffic flow prediction method using probe information in transportation system
○Keisuke Ohno[1], Eichi Takaya[1], Hiroshi Matsumoto[2], Tetsuo Morita[2], Satoshi Kurihara[1] (1. Graduate School of Science and Technology, Keio University, 2. Sumitomo Electric Industries, Ltd)
6:00 PM - 6:20 PM

[2A4-04] Analysis of Machine-Friendly Knowledge Representation toward More Efficient Knowledge Tracing
○Hiromi Nakagawa[1], Yusuke Iwasawa[1], Kaoru Nasuno[1], Yutaka Matsuo[1] (1. The University of Tokyo)
6:20 PM - 6:40 PM

[2A4-05] Highlighting Non-contributing Pixels for Visual Explanation of CNNs
○Koichi Ikeno[1], Satoshi Hara[1], Takashi Washio[1] (1. The Institute of Scientic and Industrial Research, Osaka University)
6:40 PM - 7:00 PM

Oral presentation | General Session | [General Session] 11. Robot / Real World

[2A3] [General Session] 11. Robot / Real World

座長:稲邑 哲也（国立情報学研究所）
3:20 PM - 5:00 PM Room A (4F Emerald Hall)

[2A3-01] Bidirectional Translation between Robot Actions and Linguistic Descriptions by Aligned Representations
○Tatsuro Yamada[1,2], Hiroyuki Matsunaga[1], Tetsuya Ogata[1] (1. Waseda University, 2. Research Fellow of Japan Society for the Promotion of Science)
3:20 PM - 3:40 PM

[2A3-02] Trials Reduction Method for Reinforcement Learning in Trajectory Discovery
○Yusuke Kato[1,2], Tomoaki Nakamura[3], Takayuki Nagai[3], Natsuki Yamanobe[1], Kazuyuki Nagata[1], Jun Ozawa[1] (1. Advanced Industrial Science and Technology, 2. Panasonic Corporation, 3. The University of Electro-Communications)
3:40 PM - 4:00 PM

[2A3-03] An Approach to Robot Control using Deep Reinforcement Learning and Discretization by Words
○Sayuri Hashimoto[1], Akira Kaneko[1], Ichiro Kobayashi[1] (1. Ochanomizu University)
4:00 PM - 4:20 PM

[2A3-04] Understanding Ambiguous Instructions Using Generative Adversarial Nets for Object Disposal Tasks
○Komei Sugiura[1], Aly Magassouba[1], Hisashi Kawai[1] (1. NICT)
4:20 PM - 4:40 PM

[2A3-05] Tool-use Model Considering Selecting Tool by Deep Learning
Namiko Saito[1], Kitae Kim[1], ○Dai Ba Nguyen[1], Shingo Murata[1], Tetsuya Ogata[2], Shigeki Sugano[1] (1. Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University, 2. Department of Intermedia Art and Science, School of Fundamental Science and Engineering, Waseda University)
4:40 PM - 5:00 PM

Room B

Oral presentation | General Session | [General Session] 1. Basis / Theory

[2B1] [General Session] 1. Basis / Theory

座長:小林 亮太（国立情報学研究所）
9:00 AM - 10:40 AM Room B (4F Moon Light)

[2B1-01] Analysis of the Dictionary Bases obtained by Applying Sparse Coding to Brain Activity and the Semantic Representation based on Distributed Representation
○Chiaki Kawase[1], Ichiro Kobayashi[1], Shinji Nishimoto[2], Satoshi Nishida[2], Hideki Asoh[3] (1. Ochanomizu University, 2. National Institute of Information and Communications Technology, 3. National Institute of Advanced Industrial and Technology)
9:00 AM - 9:20 AM

[2B1-02] Identification of EEG features during motor imagery by heuristic BCI with L-FTM, after learning of EEG features during exercise.

©The Japanese Society for Artificial Intelligence

○Teruo Oda[1], Suguru N. Kudoh[1] （1. Kwansei gakuin university）

9:20 AM - 9:40 AM

[2B1-03] Creating Artificial Brain Activity Data Based on Mutual Data Transformation between Different Subjects

○Kaei Cho[1], Ichiro Kobayashi[1], Shinji Nishimoto[2], Satoshi Nishida[2], Hideki Asoh[3] （1. Ochanomizu University, 2. National Institute of Information and Communications Technology, 3. National Institute of Advanced Industrial Science and Technology）

9:40 AM - 10:00 AM

[2B1-04] Reconsidering the computational model for the auditory nerve: unsupervised learning or task-based optimization?

○Hiroki Terashima[1], Shigeto Furukawa[1] （1. NTT Communication Science Laboratories）

10:00 AM - 10:20 AM

[2B1-05] Top-down and bottom-up classification between areas in mouse cerebral cortex to connect machine learning modules on connectomes

○Taku Hayami[1], So Negishi[1], Rintaro Komori[1], Haruo Mizutani[1,2], Hiroshi Yamakawa[1,2] （1. Dwango Artificial Intelligence Lab., 2. Whole Brain Architecture Initiative）

10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 1. Basis / Theory

[2B4] [General Session] 1. Basis / Theory

座長:佐々木 耀一（NEC）

5:20 PM - 7:00 PM Room B (4F Moon Light)

[2B4-01] Human body as a set of rigid units: finding correspondence between points of two sparse point-sets under distance-preserving transformation

○Torii Takuma[1], Shohei Hidaka[1] （1. Japan Advanced Institute of Science and Technology）

5:20 PM - 5:40 PM

[2B4-02] An attempt of feature extraction of neuronal activity patterns in living neuronal network, by multi-layered artificial neuronal network

○Hiromichi Sakuta[1], Suguru N. Kudoh[1] （1. Kwansei Gakuin University）

5:40 PM - 6:00 PM

[2B4-03] Structural Description and Behavior Analysis of Neural Substrate Subserving Saccadic Eye Movement with Qualitative Neuron Model

○Yoshimasa Tawatsuji[1], Tatsunori Matsui[1] （1. Waseda University）

6:00 PM - 6:20 PM

[2B4-04] Search Space Reduction in Automatically Solving Elementary Number Theory Problems in National Center Test

○Shinya Inuzuka[1], Takuya Matsuzaki[1], Satoshi Sato[1] （1. Nagoya University）

6:20 PM - 6:40 PM

[2B4-05] A Method to Estimate the Original Form of Kaomoji Using Neural Network with Character Embedding Function

○Rei Okumura[1], Noriyuki Okumura[1] （1. National Institute of Technology, Akashi College）

6:40 PM - 7:00 PM

Oral presentation | Organized Session | [Organized Session] OS-19

[2B2-OS-19a] [Organized Session] OS-19

1:20 PM - 2:40 PM Room B (4F Moon Light)

[2B2-OS-19a-01] How the learning environment in which high-quality opportunities exist for collaborative problem solving skills is linked to fostering ability

○Hiroyuki Masukawa[1], Ikuo Endo[2] （1. University of The Sacred Heart, 2. Shizuoka University）

1:20 PM - 1:40 PM

[2B2-OS-19a-02] Toward Understanding of Embodied Knowledge in Rehabilitation

○Kentaro Kodama[1], Kazuhiro Yasuda[2], Ryosaku Makino[2] （1. Kanagawa University, 2. Waseda University）

1:40 PM - 2:00 PM

[2B2-OS-19a-03] Designing Themes of Regional Revitalization on the basis of On-Site Inquiry by a Method of letting Subjectivity and Objectivity Coexist.

○Haruyuki Fujii[1], Ken-ichi Shinozaki[2] （1. Tokyo Institute of Technology, 2. Nihon University）

2:00 PM - 2:20 PM

[2B2-OS-19a-04] Analysis of the Handling Technique on a Rehabilitation Therapist

○Ikko Miyamoto[1,2], Takayuki Shiose[3], Masaaki Sakagami[3] （1. Ritsurin Hospital, 2.

©The Japanese Society for Artificial Intelligence

Waseda University, 3. Kyoto University)
2:20 PM - 2:40 PM

Oral presentation | Organized Session | [Organized Session] OS-19

[2B3-OS-19b] [Organized Session] OS-19

3:20 PM - 4:40 PM Room B (4F Moon Light)

[2B3-OS-19b-01] Stay mobile and carry on
○Fumitoshi Kato[1] (1. Keio University)
3:20 PM - 3:40 PM

[2B3-OS-19b-02] The Story of the REMAINS
○Kikuna Kuwayama[1], Minami Koseki[2], Masaki Suwa[1] (1. Keio University, 2. Keio Research Institute at SFC)
3:40 PM - 4:00 PM

[2B3-OS-19b-03] Consideration for succession of knowledge, technology and skill
○Yasuyoshi Taruta[1,2], Satoshi Yanagihara[2], Yukihiro Iguchi[1,2], Koichi Kitamura[1], Masashi Tezuka[1], Yuya Kouda[1] (1. Japan Atomic Energy Agency, 2. Fukui Univ.)
4:00 PM - 4:20 PM

[2B3-OS-19b-04] A Proposal of the Constructed Exhibition
○Minayo Okumura[1], Tomoharu Hoshina[2], Ken-ichi Kimura[3] (1. Graduate School of Systems Information Science, Future University Hakodate, 2. Hakodate City Museum, 3. Future University Hakodate)
4:20 PM - 4:40 PM

Room C

Oral presentation | General Session | [General Session] 6. Web Mining

[2C2] [General Session] 6. Web Mining

座長:柳井 孝介(日立製作所 研究開発グループ)
1:20 PM - 2:40 PM Room C (4F Orchid)

[2C2-01] Proposal of "Social Porn" Hypothesis and Pre-analysis towards its Observation
○Takeshi Sakaki[1], Fujio Toriumi[2] (1. Hottolink, Inc., 2. The University of Tokyo)
1:20 PM - 1:40 PM

[2C2-02] Communication based on strong preference in Twitter
○kimitaka asatani[1], yasuko kawahata[2], Fujio Toriumi[1], Ichiro Sakata[1] (1. The University of Tokyo, 2. Gunma University)
1:40 PM - 2:00 PM

[2C2-03] Finding Emerging Named Entities from Social Media Streams
○Satoshi Akasaki[1], Naoki Yoshinaga[2], Masashi Toyoda[2] (1. The University of Tokyo, 2. Institute of Industrial Science, The University of Tokyo)
2:00 PM - 2:20 PM

[2C2-04] Effects of Community Structure of a Social Network on the Cascade Sizes of Tweets on Social Media
○Sho Tsugawa[1] (1. University of Tsukuba)
2:20 PM - 2:40 PM

Oral presentation | General Session | [General Session] 9. NLP / IR

[2C1] [General Session] 9. NLP / IR

座長:宮西 大樹(国際電気通信基礎技術研究所)
9:00 AM - 10:40 AM Room C (4F Orchid)

[2C1-01] A Study of Document Summarization for Knowledge Presentation from Stored Information of Business Activities
○Yohko Konno[1], Takuya Nakamura[1], Masayuki Yoshida[2], Hidenori Kawamura[3] (1. Chowa Giken corporation, 2. HAJ Corporation, 3. Hokkaido University)
9:00 AM - 9:20 AM

[2C1-02] Solving the opinion summarization problem in English in the "Can a Robot Get into the University of Tokyo?" project
○Ryuichiro Higashinaka[1], Hiroaki Sugiyama[1], Hiromi Narimatsu[1], Hideki Isozaki[2], Genichiro Kikui[2], Kohji Dohsaka[3], Hirotoshi Taira[4], Tomoya Kita[4], Yasuhiro Minami[5], Takeru Kazama[5], Junji Yamato[6] (1. NTT Corporation, 2. Okayama Prefectural University, 3. Akita Prefectural University, 4. Osaka Institute of Technology, 5. The University of Electro-Communications, 6. Kogakuin University)
9:20 AM - 9:40 AM

[2C1-03] Sentiment Classification with Gated CNN using Spatial Pyramid Pooling
○Makoto Okada[1], Hirokazu Yanagimoto[1], Kiyota Hashimoto[2] (1. Osaka Prefecture University, 2. Prince of Songkla University, phuket campus)
9:40 AM - 10:00 AM

[2C1-04] An Application of Recursive Neural Tensor Network to Sentiment Analysis of Japanese Sentences

©The Japanese Society for Artificial Intelligence

○Ryuichi Akai[1], Masayasu Atsumi[1] （1. Soka University）

10:00 AM - 10:20 AM

[2C1-05] A study on document classification focusing on the output side weight on Word2Vec

○Shuto Uchida[1], Tomohiro Yoshikawa[1], Felix Jimenez[1], Takeshi Furuhashi[1] （1. Nagoya university）

10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 12. HI / Education Aid

[2C4] [General Session] 12. HI / Education Aid

座長:稲葉 通将（広島市立大学）

5:20 PM - 6:40 PM Room C (4F Orchid)

[2C4-01] Performance Evaluation of Automatic Gesture Generation System Using Bi-Directional LSTM on Humanoid Robot

○Kodai Hiyori[1], Kenji Araki[1], Dai Hasegawa[2], Satoshi Yoshio[1] （1. Graduate School of Information Science and Technology, Hokkaido University, 2. Tokyo University of Technology (Currently: Hokkai Gakuen University)）

5:20 PM - 5:40 PM

[2C4-02] Contextual Dependent Gaze Aversion in a Communication Robot

○Yukiko Kawaguchi[1], Fumihide Tanaka[2] （1. College of Engineering Systems, School of Science and Engineering, University of Tsukuba, 2. Faculty of Engineering, Information and Systems, University of Tsukuba）

5:40 PM - 6:00 PM

[2C4-03] Dialogue Selection Algorithm Based on the Amount of User’ s Emotional Expression in a Circulate-Learning Robot Connecting Elderly People

○Tsubasa Kasai[1], Fumihide Tanaka[1] （1. University of Tsukuba）

6:00 PM - 6:20 PM

[2C4-04] Input Interface for Robot System Using Mixed Reality that Encourages Retrospect of Driving

○Yuya Aikawa[1], Masayoshi Kanoh[1], Felix Jimenez[2], Mitsuhiro Hayase[3], Takahiro Tanaka[2], Hitoshi Kanamori[2] （1. Chukyo University, 2. Nagoya University, 3. Toyohashi Sozo University）

6:20 PM - 6:40 PM

Oral presentation | Organized Session | [Organized Session] OS-17

[2C3-OS-17] [Organized Session] OS-17

3:20 PM - 5:00 PM Room C (4F Orchid)

[2C3-OS-17-02] Preference Prediction using User/Property Features and Floorplan Images

○Naoki Kato[1], Toshihiko Yamasaki[1], Kiyoharu Aizawa[1], Takemi Ohama[2] （1. The University of Tokyo, 2. ietty Co., Ltd.）

3:40 PM - 4:00 PM

[2C3-OS-17-03] Extraction of User Latent Preference for Real Estate Website

○Kenshin Ikegami[1], Hironori Ito[2], Shimpei Nomura[1] （1. Recruit Sumai Company Ltd., 2. Graduate School of Informatics, Kyoto University）

4:00 PM - 4:20 PM

[2C3-OS-17-04] Multivariate analysis of the transition probability of an occupation of rental rooms by using the housing information website data

○Hayafumi Watanabe[2], Yu Ichifuji[3], Masato Suzuki[4], Satoshi Yamashita[1] （1. The institute of statistical mathematics, 2. Research Organization of Information and Systems, 3. Nagasaki University, 4. UD Asset Valuation Co., Ltd.）

4:20 PM - 4:40 PM

[2C3-OS-17-05] Classification of Real Estate Images using Bottleneck Features

○Suguru Toyohara[1], Youiti Kado[2], Toshihiko Yamasaki[3], Susumu Fujimori[1], Ikuo Tahara[1] （1. Tokyo University of Science, 2. At Home Co., Ltd., 3. The University of Tokyo）

4:40 PM - 5:00 PM

Room D

Oral presentation | General Session | [General Session] 13. AI Application

[2D1] [General Session] 13. AI Application

座長:尾形 哲也（早稲田大学）

9:00 AM - 10:40 AM Room D (4F Cattleya)

[2D1-01] Parallel Implementation of Probability Occupancy Map Generation for Multiple Human Tracking Based on Graph Optimization

○Yuri Nishikawa[1,2], Hitoshi Sato[1], Jun Ozawa[1] （1. National Institute of Advanced Industrial Science and

©The Japanese Society for Artificial Intelligence

Technology (AIST), 2. Panasonic Corporation)
9:00 AM - 9:20 AM

[2D1-02] Towards Accurate Multiple Human Tracking Using Scalable Distributed Deep Learning
○Hitoshi Sato[1], Yuri Nishikawa[1,2], Jun Ozawa[1] (1. National Institute of Advanced Industrial Science and Technology (AIST), 2. Panasonic)
9:20 AM - 9:40 AM

[2D1-03] Performance enhancement of multi-human tracking based on K-Shortest Paths by data reduction
○Nozomi Hata[1], Yuri Nishikawa[3,4], Takashi Nakayama[1], Jun Ozawa[3], Katsuki Fujisawa[2,3] (1. Faculty of Mathematics, Kyushu University, 2. Institute of Mathematics for Industry, Kyushu University, 3. National Institute of Advanced Industrial Science and Technology, 4. Panasonic Corporation)
9:40 AM - 10:00 AM

[2D1-04] An action recognition system using a highly abstracted infrared array sensor
○Shogo Murakami[1], Ikuko Eguchi Yairi[1] (1. Graduate School of Science and Technology, Sophia University)
10:00 AM - 10:20 AM

[2D1-05] Detection of a Painter School with Deep Learning Method: A Case Study of Scenes from the Tale of Genji
○Takuya Kato[1], Mariko Inamoto[2], Akihiko Konagaya[1] (1. Tokyo Institute of Technology, 2. Keisen University)
10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 13. AI Application

[2D4] [General Session] 13. AI Application

座長:森川 幸治（パナソニック株式会社）
5:20 PM - 7:00 PM Room D (4F Cattleya)

[2D4-01] Utopia of 81 Squares
○Naoyuki Soma[1] (1. The University of Tokyo)
5:20 PM - 5:40 PM

[2D4-02] Evaluation of Hybrid Reward Architecture on various learning policies and environments
○Yutaro Fujimura[1], Tomoyuki Kaneko[1] (1. The University of Tokyo)
5:40 PM - 6:00 PM

[2D4-03] Identification of the Expression of the Move of Go by Classifier Learning
○Natsumi Mori[1], Takehito Utsuro[1] (1. University of Tsukuba)
6:00 PM - 6:20 PM

[2D4-04] Development of a Jam Session System with Hatsune Miku That Learns User's Favorite Scat by Reinforcement Learning
○Joji Tsuruda[1], Natsuki Oka[1], Kazuaki Tanaka[1] (1. Kyoto Institute of Technology)
6:20 PM - 6:40 PM

[2D4-05] Automatic Music Composition System Based on Genetic Programming and Surrogate Model with Deep Learning
○Hironori Yamamoto[1], Taku Hasegawa[2], Mori Naoki[2], Keinosuke Matsumoto[2] (1. Osaka Prefecture University, 2. Graduate School of Engineering, Osaka Prefecture University)
6:40 PM - 7:00 PM

Oral presentation | Organized Session | [Organized Session] OS-21

[2D2-OS-21a] [Organized Session] OS-21

1:20 PM - 3:00 PM Room D (4F Cattleya)

[2D2-OS-21a-01] Real-time Forecasting of Big Time-series Data
○Yasuko Matsubara[1] (1. Kumamoto University)
1:20 PM - 2:00 PM

[2D2-OS-21a-02] Anomalous popularity growth in social tagging ecosystems
○Yasuhiro Hashimoto[1], Mizuki Oka[2], Takashi Ikegami[1] (1. Univ. of Tokyo, 2. Univ. of Tsukuba)
2:00 PM - 2:20 PM

[2D2-OS-21a-03] Deep Learning Approach for Life Emerged from High-dimensional Data Recognition
○Wataru Noguchi[1], Hiroyuki Iizuka[1], Masahito Yamamoto[1] (1. Hokkaido University)
2:20 PM - 2:40 PM

[2D2-OS-21a-04] Designing a Communication Space containing Agency using Writing Record Information
○Dominique Chen[1], Hiroki Kojima[2], Mizuki Oka[3], Takashi Ikegami[2] (1. Faculty of Letters, Arts and Science, Waseda University, 2. Graduate School of Arts and Sciences,

©The Japanese Society for Artificial Intelligence

University of Tokyo, 3. Graduate School of Systems and Information Engineering, University of Tsukuba）
2:40 PM - 3:00 PM

Oral presentation | Organized Session | [Organized Session] OS-21

[2D3-OS-21b] [Organized Session] OS-21

3:20 PM - 4:40 PM Room D (4F Cattleya)

[2D3-OS-21b-01] Bursting Behavior of Social Insects through Agent-based Simulation
○Doi Itsuki[1], Takashi Ikegami[1] （1. the University of Tokyo）
3:20 PM - 3:40 PM

[2D3-OS-21b-02] Generation and Selection Mechanism on Social Tagging Systems
○Yoshimasa Nishikawa[1], Mizuki Oka[1], Yasuhiro Hashimoto[2], Takashi Ikegami[2] （1. University of Tsukuba, 2. The University of Tokyo）
3:40 PM - 4:00 PM

[2D3-OS-21b-03] Modelling and experimental approaches to continuous dynamics of social interactions
○Reiji Suzuki[1], Takaya Arita[1] （1. Nagoya University）
4:00 PM - 4:20 PM

[2D3-OS-21b-04] Autonomous Regulation of Self and Non-Self in Embodied Neural Networks.
○Atsushi Masumori[1], Norihiro Maruyama[1], Lana Sinapayen[1], Takashi Ikegami[1] （1. The University of Tokyo）
4:20 PM - 4:40 PM

Room E

Oral presentation | General Session | [General Session] 9. NLP / IR

[2E2] [General Session] 9. NLP / IR

座長:田中 貴秋（NTT コミュニケーション科学基礎研究所）
1:20 PM - 2:40 PM Room E (4F Queen)

[2E2-01] A Framework of Generating a Narrative Using Metaphor Mapping and Post-editing
○Suguru Matsuyoshi[1], Akira Utsumi[1] （1. The University of Electro-Communications）
1:20 PM - 1:40 PM

[2E2-02] A Generator of Various Styles of Utterances Using Utterance Expression Dictionary
○Ryo Kimura[1], Kazuko Natsume[1], Satoshi Sato[1], Takuya Matsuzaki[1] （1. Nagoya University）
1:40 PM - 2:00 PM

[2E2-03] Headline Generation of Yahoo! Chiebukuro Based on Learning to Rank
○Tatsuru Higurashi[1], Hayato Kobayashi[1,2], Kazuma Murao[1], Masuyama Takeshi[1] （1. Yahoo Japan Corporation, 2. RIKEN AIP）
2:00 PM - 2:20 PM

[2E2-04] TULIP: Generation of Screenplay-style Novel (SS) Using Three-step LSTM Trained on Web Novels
○Atsushi Suzuki[1] （1. The University of Tokyo）
2:20 PM - 2:40 PM

Oral presentation | NFC Session | [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

[2E3-NFC-1a] [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

座長:成田 雅彦（産業技術大学院大学）
3:20 PM - 5:00 PM Room E (4F Queen)

[2E3-NFC-1a-01] Robot based Active Sensing System and "Integrated Platform for Cloud-based Robot Services Project"
○Masahiko Narita[1], Nobuto Matsuhira[2], Yosuke Tsuchiya[1], Yuka Kato[3], Yoshihiko Murakawa[4] （1. Advanced Institute of Industrial Technology, 2. Shibaura Institute of Technology, 3. Tokyo Woman's Christian University, 4. FUJITSU LABORATORIES LTD）
3:20 PM - 3:40 PM

[2E3-NFC-1a-02] Behavior information acquisition using wireless tag in conventions
○Tokuro Matsuo[1], Hidekazu Yanagimoto[2], Masahiko Narita[1], Naoki Fukuta[3], Katsuhide Fujita[4], Kiyota Hashimoto[5], Hidekazu Iwamoto[6] （1. Advanced Institute of Industrial Technology, 2. Osaka Prefecture University, 3. Shizuoka University, 4. Tokyo University of Agriculture and Technology, 5. Prince of Songkla University, 6. Josai International University）
3:40 PM - 4:00 PM

[2E3-NFC-1a-03] Development on a Large Display Type Tele-presence Robot
○Tatsuno Kyoichi[1], Hideharu Oobitsu[2], Katsumi Yoshino[2] （1. Meijo University, 2.

©The Japanese Society for Artificial Intelligence

Shimane Institute For Industrial Technology)
4:00 PM - 4:20 PM

[2E3-NFC-1a-04] An analysis of user's utterance to reduce inappropriate response of chat robot
○Shodai Aoyagi[1], Kazuki Hirata[1], Tomoya Takatani[2], Eri Sato-Shimokawara[1], Toru Yamaguchi[1] (1. Tokyo Metropolitan University, 2. Toyota Motor Corporation)
4:20 PM - 4:40 PM

[2E3-NFC-1a-05] Development of a partner robot aiming at reducing the psychological burden of child-care environment
○Yosuke Tsuchiya[1], Jun Uchiyama[2], Kazunari Hirakoso[2], Daichi Takeshima[2], Rika Sekita[2], Jinnan Zhang[2], Yoshio Kondo[2], Masahiko Narita[2] (1. Tokyo Online University, 2. Advanced Institute of Industrial Technology)
4:40 PM - 5:00 PM

Oral presentation | NFC Session | [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

[2E4-NFC-1b] [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

座長:成田 雅彦（産業技術大学院大学）
5:20 PM - 7:00 PM Room E (4F Queen)

[2E4-NFC-1b-01] Predicting a Pedestrian Trajectory Using Sequence to Sequence Learning for Mobile Robot Navigation
Natsuki Sakata[1], Yuka Kinoshita[1], ○Yuka Kato[1] (1. Tokyo Woman's Christian University)
5:20 PM - 5:40 PM

[2E4-NFC-1b-02] Classification of Utterances Using Temporal Relation in Cloud-based Dialogue System
○Kazuki Hirata[1], Eri Shimokawara[1], Tomoya Takatani[2], Toru Yamaguchi[1] (1. Tokyo Metropolitan University, 2. Toyota Motor Corporation)
5:40 PM - 6:00 PM

[2E4-NFC-1b-03] Structure of RT system to realize various robot services
○Nobuto Matsuhira[1], Takamasa Ikeda[1], Mirai Shimoyama[1], Satoshi Okano[1], Masahiko Narita[2], Yosuke Tsuchiya[2], Yuka Kato[3] (1. Shibaura Institute of Technology, 2. Advanced Institute of Industrial Technology, 3. Tokyo Woman's Christian University)
6:00 PM - 6:20 PM

[2E4-NFC-1b-a] (Panel discussion) The next direction of "Integrated Platform for Cloud-based Robot Services Project"
Masahiko Narita, Tokuro Matsuo, Nobuto Matsuhira, Tatsuno Kyoichi, Toru Yamaguchi, Yuka Kato
6:20 PM - 7:00 PM

Room F

Oral presentation | General Session | [General Session] 4. Knowledge Utilization and Sharing

[2F4] [General Session] 4. Knowledge Utilization and Sharing

座長:森田 武史（慶應義塾大学）
5:20 PM - 7:00 PM Room F (4F Garreria)

[2F4-01] Ontology-based Knowledge Systemization for Interpreting the Mechanism of Liver Toxicity
○Yuki Yamagata[1], Yoshinobu Igarashi[1], Noriyuki Nakatsu[1], Katsuhisa Horimoto[2], Kazuhiko Fukui[2], Yoshihiro Uesawa[3], Hiroshi Yamada[1] (1. National Institutes of Biomedical Innovation, Health and Nutrition, 2. National Institute of Advanced Industrial Science and Technology, 3. Meiji Pharmaceutical University)
5:20 PM - 5:40 PM

[2F4-02] Reorganizing Hierarchical Category Structure of Japanese Wikipedia for Knowledge Engineers
○Takanori Nakagawa[1], Masaharu Yoshioka[1,2] (1. Hokkaido Unviersity, 2. RIKEN AIP)
5:40 PM - 6:00 PM

[2F4-03] Towards Implementation of Integrated Knowledge Infrastructure to Create Knowledge Graphs
○Ryutaro Ichise[1,2], Natthawut Kertkeidkachorn[2], Lihua Zhao[2] (1. National Institute of Informatics, 2. National Institute of Advanced Industrial Science and Technology)
6:00 PM - 6:20 PM

[2F4-04] Training Data on Japanese Archery Motion Recognition for Self-training Support
○Takahiro Morozumi[1], Tadachika Morozumi[1], Toramatsu Shintani[1] (1. Department of Computer

©The Japanese Society for Artificial Intelligence

Science, Graduate School of Engineering, Nagoya Institute of Technology)
6:20 PM - 6:40 PM

[2F4-05] A Model for Knowledge Representation Based on the Concepts of Things and Events
○Takahiro Yamada[1] (1. Japan Aerospace Exploration Agency)
6:40 PM - 7:00 PM

Oral presentation | General Session | [General Session] 12. HI / Education Aid

[2F1] [General Session] 12. HI / Education Aid

座長:松原 行宏(広島市立大学大学院)
9:00 AM - 10:40 AM Room F (4F Garreria)

[2F1-01] Measuring Beginner Friendliness of Chinese Web Pages Explaining Academic Concepts using HTML Structures
○BINGCAI HAN[1], Takahide Kasuga[1], Hayato Shiokawa[1], Kota Kawaguchi[1], Takehito Utsuro[1], Yasuhide Kawada[2] (1. University of Tsukuba, 2. Logworks Co., Ltd.)
9:00 AM - 9:20 AM

[2F1-02] Estimation of understanding of words by lifelog analysis for vocabulary visualizing
○Kazuma Nozawa[1], Noriki Mochizuki[1], Gakuto Masuyama[2], Sousuke Nakamura[3] (1. Department of Science and Engineering, Hosei University Graduate School, 2. Faculty of Science and Technology, Meijo University, 3. Faculty of Science and Technology, Hosei University)
9:20 AM - 9:40 AM

[2F1-03] Implementing a Weekly Report Management System Based on Activity Records for Reuse of Past Reports
○Rikiya Ando[1], Tadachika Ozono[1], Toramatsu Shintani[1] (1. Nagoya Institute of Technology)
9:40 AM - 10:00 AM

[2F1-04] Effectiveness of Partial Fading of Text on Reading Comprehension
○Jumpei Kobayashi[1,2], Toshio Kawashima[2] (1. Dai Nippon Printing Co., Ltd., 2. Future University Hakodate)
10:00 AM - 10:20 AM

[2F1-05] Proposal of fashion coordinate system based on color harmony
○Takumi Saito[1], Mizuki Oka[1], Kazuhiko Kato[1] (1. University of Tsukuba, Graduate School of Systems and Information Engineering)
10:20 AM - 10:40 AM

Oral presentation | Organized Session | [Organized Session] OS-4

[2F2-OS-4a] [Organized Session] OS-4

1:20 PM - 3:00 PM Room F (4F Garreria)

[2F2-OS-4a-01] Well-being and Haptics
○Junji Watanabe[1] (1. NTT Communication Science Laboratories)
1:20 PM - 1:40 PM

[2F2-OS-4a-02] Emotional responses are changed bi-directionally through face-to-face interaction
○Aiko Murata[1] (1. Waseda University)
1:40 PM - 2:00 PM

[2F2-OS-4a-03] Measurement trials of crowd’ s well-being using visible light communication
○Takashi G Sato[1], Yoshifumi Shiraki[1], Takehiro Moriya[1] (1. Nippon Telegraph and Telephone)
2:00 PM - 2:20 PM

[2F2-OS-4a-04] Analysis of the Relationship between Sleep Quality and Stress with a focus on Circadian Rhythm for Well-being
○Ryo Takano[1], Satoshi Hasegawa[1], Yuta Umenai[1], Takato Tatsumi[1], Keiki Takadama[1], Toru Shimuta[2], Toru Yabe[2], Hideo Matsumoto[2] (1. The University of Electro-Communications, 2. Murata Manufacturing Company, Ltd.)
2:20 PM - 2:40 PM

[2F2-OS-4a-05] Investigation of appropriate fNIRS feature to evaluate cognitive load
○Kota Suzuki[1], Tatsuya Suzuki[1], Sotaro Shimada[1], Atsumichi Tachibana[2], Yumie Ono[1] (1. Meiji University, 2. Dokkyo Medical University)
2:40 PM - 3:00 PM

Oral presentation | Organized Session | [Organized Session] OS-4

[2F3-OS-4b] [Organized Session] OS-4

3:20 PM - 5:00 PM Room F (4F Garreria)

[2F3-OS-4b-01] Direction to quantifying well-being: Visualizing mindfulness based on functional brain network
○Satoru Hiwa[1], Tomoyuki Hiroyasu[1] (1.

©The Japanese Society for Artificial Intelligence

Doshisha University)
3:20 PM - 3:40 PM

[2F3-OS-4b-02] Estimation of low rank matrix between two correlation matrices based on clustering
○Kensuke Tanioka[1], Satoru Hiwa[2], Tomoyuki Hiroyasu[2], Hiroshi Yadohisa[2] (1. Wakayama Medical University, 2. Doshisha University)
3:40 PM - 4:00 PM

[2F3-OS-4b-03] Meditation Changes the Time Required for Attention Shifting from Mind-Wandering
○Issaku Kawashima[1,2], Toru Takahashi[1,3], Masahiro Fujino[3,4], Hiroaki Kumano[2] (1. Graduate School of Human Sciences, Waseda Univ., 2. Faculty of Human Sciences, Waseda Univ., 3. Japan Society for the Promotion of Science, 4. Graduate School of Education, Kyoto Univ.)
4:00 PM - 4:20 PM

[2F3-OS-4b-04] A study on Detection of Symptoms of Dementia Based on Peak Latency of P300.
○Koki Miwa[1], Tomohiro Yoshikawa[1], Takeshi Furuhashi[1], Minoru Hoshiyama[2], Taeko Makino[3], Madoka Yanagawa[4], Yusuke Suzuki[5], Hiroyuki Umegaki[6], Masahumi Kuzuya[7] (1. Dept. of Information and Communication Eng., Graduate School of Eng., Nagoya University, 2. Brain &Mind Research Ceneter, Nagoya University, 3. School of Rehabilitation and Care, Seijo University, 4. Dept. of Geriatrics, Nagoya University Hospital, 5. Center for Community Liaison and Patient Consultations, Nagoya University Hospital, 6. Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University, 7. Institute of Innovation for Future Society / Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University)
4:20 PM - 4:40 PM

[2F3-OS-4b-05] Improvement in Chronic Stress Level Recognition by Using Both Full-term and Short-term Measurements of Physiological Features
○Yoshiki Nakashima[1], Masanori Tsujikawa[1], Onishi Yoshifumi [1] (1. NEC Corporation)
4:40 PM - 5:00 PM

Room G

Oral presentation | General Session | [General Session] 11. Robot / Real World

[2G1] [General Session] 11. Robot / Real World
座長:長井 隆行（電通大）
9:00 AM - 10:20 AM Room G (5F Ruby Hall Hiten)

[2G1-01] Comparison of Methods for Localizing A Target with UAV
○Ryosuke Matsuo[1], Koya Kishi[1], Koichi Hori[1], Takehisa Yairi[1] (1. The University of Tokyo)
9:00 AM - 9:20 AM

[2G1-02] Difference evaluation of daily-life activities in physical and immersive VR environments
○Yoshiaki Mizuchi[1], Tetsunari Inamura[1,2] (1. National Institute of Informatics, 2. SOKENDAI (The Graduate University for Advanced Studies))
9:20 AM - 9:40 AM

[2G1-03] Generation of presentation motion using sequence to sequence model
○Akihito Shimazu[1], Chie Hieida[1], Takayuki Nagai[1], Tomoaki Nakamura[1], Yuki Takeda[2], Takenori Hara[2], Osamu Nakagawa[2], Tsuyoshi Maeda[2] (1. The University of Electro-Communications, 2. Dai Nippon Printing Co., Ltd.)
9:40 AM - 10:00 AM

[2G1-04] Validation of Effective Active Learning Method in Semantic Learning of Context-dependent Motions
○Tatsuya Sakato[1], Tetsunari Inamura[1,2] (1. National Institute of Informatics, 2. SOKENDAI(The Graduate University for Advanced Studies))
10:00 AM - 10:20 AM

Oral presentation | General Session | [General Session] 11. Robot / Real World

[2G4] [General Session] 11. Robot / Real World
座長:大澤 博隆（筑波大学）
5:20 PM - 6:40 PM Room G (5F Ruby Hall Hiten)

[2G4-01] Emotion Estimation from Robot Movements Using Weighted Random Forest and Pairwise Classification
○Hideki Kato[1], Shohei Kato[1,2] (1. Dept. of Computer

©The Japanese Society for Artificial Intelligence

Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, 2. Frontier Research Institute for Information Science, Nagoya Institute of Technology)
5:20 PM - 5:40 PM

[2G4-02] Emotional Differentiation based on Decision-Making Process

○Chie Hieida[1], Takato Horii[1], Takayuki Nagai[1] (1. The University of Electro-Communications)
5:40 PM - 6:00 PM

[2G4-03] Effect of Sympathy expression method by Face and Body of Educational-Support Robot in Long Term Experiment

○Yuhei Tanizaki[1], Felix Jimenez[1], Tomohiro Yoshikawa[1], Takeshi Furuhashi[1], Masayoshi Kanoh[2] (1. Nagoya University, 2. Chukyo University)
6:00 PM - 6:20 PM

[2G4-04] Segmenting Time Series Data Using GP-HSMM with Nonparametric Bayesian Model

○Masatoshi Nagano[1], Tomoaki Nakamura[1], Takayuki Nagai[1], Daichi Mochihashi[2], Ichiro Kobayashi[3], Masahide Kaneko[1] (1. The University of Electro-Communications, 2. Institute of Statistical Mathematics, 3. Ochanomizu University)
6:20 PM - 6:40 PM

Oral presentation | Organized Session | [Organized Session] OS-10

[2G2-OS-10b] [Organized Session] OS-10

1:20 PM - 2:40 PM Room G (5F Ruby Hall Hiten)

[2G2-OS-10b-01] Establishment of Core Vocabulary based on Knowledge Graph for Interoperability between Agricultural Data

○Sungmin Joo[1], Hideaki Takeda[1,2], Akane Takezaki[3], Tomokazu Yoshida[3] (1. National Institute of Informatics, 2. The Graduate University for Advanced Studies, 3. National Agriculture and Food Research Organization)
1:20 PM - 1:40 PM

[2G2-OS-10b-02] Fruits Growth Information Extracting Application by using Deep Learning Network

○Kazuki Kobayashi[1], Toshiki Nakamura[1] (1. Shinshu University)
1:40 PM - 2:00 PM

[2G2-OS-10b-03] Development of Crop Vocabulary to facilitate interoperability among agricultural IT systems

○Akane Takezaki[1], Sungmin Joo[2], Hideaki Takeda[2], Tomokazu Yoshida[1] (1. National Agriculture and Food Research Organization, 2. National Institute of Informatics)
2:00 PM - 2:20 PM

[2G2-OS-10b-04] Agricultural Field Monitoring for Multi-Lens Camera TOMBO

○Fumiya Shimobayashi[1], Kazuki Kobayashi[1], Keiichiro Kagawa[2], Wei Guo[3], Masayuki Hirafuji[3], Jun Tanida[4] (1. Shinshu Univ., 2. Shizuoka Univ., 3. Univ. of Tokyo, 4. Osaka Univ.)
2:20 PM - 2:40 PM

Oral presentation | Organized Session | [Organized Session] OS-10

[2G3-OS-10c] [Organized Session] OS-10

3:20 PM - 4:40 PM Room G (5F Ruby Hall Hiten)

[2G3-OS-10c-01] Features values for recognition of swine's head direction from video.

○Takuya Sato[1], Misaki Mito[1], Takuji Kawagishi[1], Koichi Mizutani[1,2], Keiichi Zempo[1,2], Naoto Wakatsuki[1,2], Nobuhiro Takemae[3], Takehiko Saito[3] (1. Graduate School of Systems and Information Engineering, University of Tsukuba, 2. Faculty of Engineering, Information and Systems, University of Tsukuba, 3. National Institute of Animal Health, NARO)
3:20 PM - 3:40 PM

[2G3-OS-10c-02] Development of Image Sensing Method to Automatically Obtain Soybean Growth Condition

○Kazuki Omura[1], So Yahata[1], Seiichi Ozawa[1], Takenao Ohkawa[1], Noriyuki Murakami[2], Hiroyuki Tsuji[2] (1. Kobe University, 2. Hokkaido National Agricultural Research Center)
3:40 PM - 4:00 PM

[2G3-OS-10c-03] Detection of Cattle States Relevant to Calving from Video

○Yusuke Okimoto[1], Kazuma Sugawara[1], Susumu Saito[1], Teppei Nakano[1,2], Makoto Akabane[1,2], Tetsunori Kobayashi[1], Tetsuji

©The Japanese Society for Artificial Intelligence

Ogawa[1] (1. Waseda University, 2. Intelligence Framework Lab)
4:00 PM - 4:20 PM

[2G3-OS-10c-04] A Study on Controlling Body Water Content of Tomato Using Image Processing Technology
○Shuto Namba[1], Junpei Tsuji[1], Masato Noto[1] (1. Kanagawa University)
4:20 PM - 4:40 PM

Room H

Oral presentation | General Session | [General Session] 3. Data Mining

[2H2] [General Session] 3. Data Mining
座長:檜山 敦（東京大学）
1:20 PM - 3:00 PM Room H (10F Sky Hall)

[2H2-01] Pattern Recognition for Tennis Tactics using Hidden Markov Model from Rally Series
○Natsuki Miyahara[1], Taro Tezuka[2], Yasushi Nakauchi[3] (1. Graduate School of Library Information and Media Studies, University of Tsukuba, 2. Faculty of Library Information and Media Science, University of Tsukuba, 3. Faculty of Engineering, Information and Systems, University of Tsukuba)
1:20 PM - 1:40 PM

[2H2-02] New indicator for centrality measurement in passing network analysis of football
○Masatoshi Kanbata[1], Ryohei Orihara[1], Yuichi Sei[1], Yasuyuki Tahara[1], Akihiko Ohsuga[1] (1. The University of Electro-Communications)
1:40 PM - 2:00 PM

[2H2-03] Estimation of Cognitive Processes of Shield Machine Operator Based on Eye Tracking
○Nao Fujimoto[1], Junya Morita[1], Toru Okubo[1], Nobuhiko Obayashi[1], Takayasu Shirai[1] (1. KAJIMA CORPORATION)
2:00 PM - 2:20 PM

[2H2-04] Extracting Long-term Patterns from Electrocardiogram Waveform using Deep Learning
○Noriyasu Omata[1], Yoshihiro Nakamura[2], Susumu Shirayama[1] (1. School of Engineering, The University of Tokyo, 2. Faculty of Engineering, The University of Tokyo)
2:20 PM - 2:40 PM

[2H2-05] Learning sense of locality?
daisuke moriwaki[1], ○isshu munemasa[2], toshikazu fukami[1] (1. CyberAgent, Inc., 2. Meiji University)
2:40 PM - 3:00 PM

Oral presentation | General Session | [General Session] 4. Knowledge Utilization and Sharing

[2H1] [General Session] 4. Knowledge Utilization and Sharing
座長:川村 隆浩（科学技術振興機構）
9:00 AM - 10:40 AM Room H (10F Sky Hall)

[2H1-01] Implementation of adaptive displaying function for action decomposition trees based on various viewpoints
○Masaaki Imazono[1], Yoshinobu Kitamura[1] (1. Ritsumeikan University)
9:00 AM - 9:20 AM

[2H1-02] Knowledge Sharing System in an Organization Using Extended Prediction Market
○Seiyu Yamaguchi[1], Hajime Mizuyama[1], Tomomi Nonaka[1] (1. Aoyama Gakuin University)
9:20 AM - 9:40 AM

[2H1-03] Toward an Elderly Care Ontology for Knowledge Sharing among Care Workers
○Yoshiki Fukushima[1], Satoshi Nishimura[2], Kenichiro Fukuda[2], Takuichi Nishimura[2], Yoshinobu Kitamura[1] (1. Graduate School of Information Science and Engineering, Ritsumeikan University, 2. Artificial Intelligence Research Center, National Institute for Advanced Industrial Science and Technology)
9:40 AM - 10:00 AM

[2H1-04] Constructing knowledge models and an ontology for adjusting specifications of industrial machinery
○Taisei Ido[1], Yoshinobu Kitamura[1] (1. Ritsumeikan University)
10:00 AM - 10:20 AM

[2H1-05] An Testing Environment to Accumulate Test Cases for Execution Order Variation
○Yusuke Aoyama[1], Takeru Kuroiwa[1], Noriyuki Kushiro[1] (1. Kyushu Institute of Technology)
10:20 AM - 10:40 AM

Oral presentation | NFC Session | [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)

[2H3-NFC-4a] [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)
座長:谷田 泰郎（シナジーマーケティング）

©The Japanese Society for Artificial Intelligence

3:20 PM - 5:00 PM Room H (10F Sky Hall)

[2H3-NFC-4a-01] Issues on Translating Information of Food Culture between Different Languages
○Masami Suzuki[1] (1. KDDI Research Inc.)
3:20 PM - 3:40 PM

[2H3-NFC-4a-02] Analysis of free description answer data on "super aged society"
○Yukiko Saito[1], Kotomi Takamuku[1], Yasuo Tanida[1] (1. Synergy Marketing, Inc.)
3:40 PM - 4:00 PM

[2H3-NFC-4a-03] About ‘Creative Genome Project’ and Relevant Intelligence Communication Consortiam
○Atsushi Sasaki[1,2] (1. AOI TYO Holdings Inc., 2. Relevant Intelligence Communication Consortiam)
4:00 PM - 4:20 PM

[2H3-NFC-4a-04] ‘Two-way communication’ and ‘Mind design’
○Yasuo Tanida[1], Kotomi takamuku[1] (1. Synergy Marketing Inc.)
4:20 PM - 4:40 PM

[2H3-NFC-4a-05] Text Analysis and Structuring of Persons in Kabuki
○Takashi Ogata[1] (1. Iwate prefectural university)
4:40 PM - 5:00 PM

Oral presentation | NFC Session | [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)

[2H4-NFC-4b] [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)

座長:谷田 泰郎（シナジーマーケティング）
5:20 PM - 6:40 PM Room H (10F Sky Hall)

[2H4-NFC-4b-01] A Latent Class Analysis using Association between Consumer Values and Behaviors.
○Yoshihide Nishio[1] (1. Synergy Marketing, Inc.)
5:20 PM - 5:40 PM

[2H4-NFC-4b-02] A Study of the Values and the Revisit Intention of Inbound Tourists in Japan
○Hiromi Kamata[1] (1. Hitotsubashi University)
5:40 PM - 6:00 PM

[2H4-NFC-4b-03] Inbound tourism Marketing Model Green AI Academy (GAIA)
○Izawa Yasushi[1] (1. EcoTourLab EcoGuideCafe)
6:00 PM - 6:20 PM

[2H4-NFC-4b-04] Stability and similarity of clusters under reduced response data
Marisciel Litong-Palima[1], Kristoffer Jon Albers[2], ○Fumiko Kano Gluckstad[1] (1. Copenhagen Business School, 2. Technical University of Denmark)
6:20 PM - 6:40 PM

Room J

Oral presentation | General Session | [General Session] 13. AI Application

[2J1] [General Session] 13. AI Application

座長:星野 崇宏（慶応大）
9:00 AM - 10:40 AM Room J (2F Royal Garden B)

[2J1-01] Effect of Increasing of Horizontal ShareHolding by Index Funds to Competitions and Market Prices -- Investigation by Agent-Based Model --
○Takanobu Mizuta[1] (1. SPARX Asset Management Co.,Ltd.)
9:00 AM - 9:20 AM

[2J1-02] Indexation of sentiment of Bank of Japan using text of Financial report and macroeconomic index
○Kyoto Yono[1], Kiyoshi Izumi[1], Hiroki Sakaji[1] (1. Graduate School of Engineering,The University of Tokyo)
9:20 AM - 9:40 AM

[2J1-03] How Does the Rule for Investment Diversification Affect Market Fluidity, Efficiency, and Investor's Performance?
○shunya maruyama[1], takanobu mizuta[2], isao yagi[1] (1. Kanagawa Institute of Technology, 2. SPARX Asset Management CO.,Ltd.)
9:40 AM - 10:00 AM

[2J1-04] Analyzing Mergers and Acquisitions (M&A) in Japan Using AI methods
○Bohua Shao[1], Kimitaka Asatani[1], Ichiro Sakata[1] (1. School of Engineering, the University of Tokyo)
10:00 AM - 10:20 AM

[2J1-05] Consideration on substitution by artificial

©The Japanese Society for Artificial Intelligence

intelligence and adaptation of organization structure

○Shunei NORIKUMO[1] （1. Osaka University of Commerce）

10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 13. AI Application

[2J2] [General Session] 13. AI Application

座長:水田 孝信（ スパークス・アセット・マネジメント株式会社）

1:20 PM - 2:40 PM Room J (2F Royal Garden B)

[2J2-01] Improvement of Prediction Accuracy in Predicting Market Trends by Newspaper Article Analysis Using Deep Learning

○Kazuki Matsumoto[1], Tohgoroh Matsui[1] （1. Chubu University）

1:20 PM - 1:40 PM

[2J2-02] A Novel Approach of Purchase Prediction Considering the Purchase on Competing Company by Using ID-POS Combined with Geofencing Location Data

○Junichiro NIIMI[1], Takahiro HOSHINO[2,1] （1. RIKEN Center for Advanced Intelligence Project, 2. Faculty of Economics, Keio University）

1:40 PM - 2:00 PM

[2J2-03] Classification by Time-Series Gradient Boosting Tree

○Kei Nakagawa[1,2], Mitsuyoshi Imamura[1,3], Kenichi Yoshida[2] （1. Nomura Asset Management Co.,Ltd., 2. University of Tsukuba Graduate School of Business Sciences, 3. University of Tsukuba Graduate School of Systems and Information Engineering）

2:00 PM - 2:20 PM

[2J2-04] Construction of Strategy using Binary Decision Diagram in Stock Investment using Technical Indicators

○Ohki Kato[1], Hazime Anada[1] （1. Tokyo City University）

2:20 PM - 2:40 PM

Oral presentation | General Session | [General Session] 13. AI Application

[2J3] [General Session] 13. AI Application

座長:小澤 順（産業技術総合研究所）

3:20 PM - 5:00 PM Room J (2F Royal Garden B)

[2J3-01] Objective evaluation for the cystoscopic diagnosis of bladder cancer using transfer learning

○Yutaro Hoshino[1,2], Hidenori Sakanashi[2], Masahiro Murakawa[2], Nagatsugu Yamanouchi[1], Hirokazu Nosato[1,2] （1. Graduate School of Toho University, 2. Artificial Intelligence Research Center, National Institute of Adanced Industrial Science and Technology）

3:20 PM - 3:40 PM

[2J3-02] Estimation of Depressive Tendency Based on Lifestyle and Constitution

○Shiori Yamaguchi[1], Hiroki Tanaka[1], Hayato Maki[1], Shigehiko Kanaya [1,2], Nobutaka Suzuki[3], Satoshi Nakamura [1,2] （1. Nara Institute of Science and Technology, Graduate School of Science and Technology, 2. Nara Institute of Science and Technology, Data Science Center, 3. Kanazawa University, Graduate School of Medical Sciences）

3:40 PM - 4:00 PM

[2J3-03] Study on Measurement of Pelvic Angle from Point Cloud

○Emiri Sakiyama[1], Atsushi Hiyama[2,3], Sohei Wakisaka[3], Atsushi Izumihara[3], Masahiko Inami[3] （1. The Faculty of Engineering, The University of Tokyo, 2. Center for Advanced Intelligence Project, RIKEN, 3. Reseach Center for Advanced Science and Technology, The University of Tokyo）

4:00 PM - 4:20 PM

[2J3-04] Denoising autoencoder-based modification method for RRI data with premature ventricular contraction

○Shota Miyatani[1], Koichi Fujiwara[1], Manabu Kano[1] （1. Kyoto University）

4:20 PM - 4:40 PM

[2J3-05] Consideration of relations sit-to-stand movement and strength of lower-limbs

○Takahiro Hiyama[1], Yoshikuni Sato[1], Jun Ozawa[2] （1. Panasonic, 2. Advanced Industrial and Technology）

4:40 PM - 5:00 PM

Oral presentation | General Session | [General Session] 13. AI Application

[2J4] [General Session] 13. AI Application

座長:小澤 順（産業技術総合研究所）

5:20 PM - 6:40 PM Room J (2F Royal Garden B)

[2J4-01] Improvement of CT image segmentation by Deep Residual 3D U-Nets and 3D-CNNs

○Keita Ninomiya[1], Yoshinobu Furuyama[2], Joji Ota[2], Hiroki Suyari[1] （1. Chiba University, 2. Department of Radiology, Chiba University Hospital）

©The Japanese Society for Artificial Intelligence

5:20 PM - 5:40 PM

[2J4-02] Feature selection and over-adaptation prevention in neural networks using the auxiliary weight method

○Akira NODA[1] (1. Technology Research Laboratory, Shimadzu Corp.)

5:40 PM - 6:00 PM

[2J4-03] Mental Disorder Diagnosis using fMRI Images by Structured Deep Generative Model

○Tetsuo Tashiro[1], Takashi Matsubara[1], Kuniaki Uehara[1] (1. Kobe University)

6:00 PM - 6:20 PM

[2J4-04] Segmentation of anatomical structures in chest X-ray pictures using U-Net

○Kenji Kondo[1,2], Jun Ozawa[1], Masaki Kiyono[2,3], Shinichi Fujimoto[3], Masato Tanaka[3], Toshiki Adachi[3], Harumi Ito[3], Hirohiko Kimura[3] (1. Advanced Industrial Science and Technology, 2. Panasonic Corporation, 3. University of Fukui)

6:20 PM - 6:40 PM

Room K

Oral presentation | General Session | [General Session] 12. HI / Education Aid

[2K2] [General Session] 12. HI / Education Aid

座長:瀬田 和久（大阪府立大学）

1:20 PM - 3:00 PM Room K (3F Azisai Mokuren)

[2K2-01] Pulley Learning Support with Feedback based on Haptic and Pseudo-Haptic

○Takanori Kono[1], Yukihiro Matsubara[1], Masaru Okamoto[1] (1. Graduate School of Information Sciences, Hiroshima City University)

1:20 PM - 1:40 PM

[2K2-02] Virtual Antique Folding Screen with Optical Effects of Sacred Light

○Toshiyuki Tsuno[1], Mariko Inamoto[2], Akihiko Konagaya[1] (1. Tokyo Institute of Technology, 2. Keisen University)

1:40 PM - 2:00 PM

[2K2-03] Analysis of Gaze Behaviors in Virtual Environments for Cooperative Pattern Formation

○Norifumi Watanabe[1], Kota Itoda[2] (1. Advanced Institute of Industrial Technology, 2. Keio University)

2:00 PM - 2:20 PM

[2K2-04] Development and Evaluation of a AR Drawing System suitable for real time gripping movement by using Kinect

○Hiroki Inatome[1], Masato Soga[1] (1. Wakayama University)

2:20 PM - 2:40 PM

[2K2-05] An Effect of User's Degree of Intimacy on Agent that Expresses Requital of a Favor

○Mitsuhiro Hayase[1], Taro Nakamura[2], Masayoshi Kanoh[2] (1. Toyohashi Sozo University, 2. Chukyo University)

2:40 PM - 3:00 PM

Oral presentation | General Session | [General Session] 13. AI Application

[2K1] [General Session] 13. AI Application

座長:高橋 大志（慶応大 ）

9:00 AM - 10:40 AM Room K (3F Azisai Mokuren)

[2K1-01] The customyzation of neural machine translation with user's dictionary

○Shota Yamauchi[1], Yohei Murakami[2] ,Takao Nakaguchi[3], Toru Ishida[4] (1. Kyoto University, 2.Unit of Design, Kyoto University, 3.Department of Web Business Technology, Kyoto College of Graduate Studies for Informatics, 4.Department of Social Informatics, Kyoto University)

9:00 AM - 9:20 AM

[2K1-02] Dialog Summarization using Phrase Structure Analysis and Text Clustering

○Ayako Hoshino[1], Itaru Hosomi[1] (1. NEC Corp.)

9:20 AM - 9:40 AM

[2K1-03] Language generation system for sightseeing guidance using neural language model

Kazuya Ikuta[1], Seitaro Shinagawa[1], ○Koichiro Yoshino[1,3], Yu Suzuki[1,2], Satoshi Nakamura[1,2] (1. Graduate School of Information Science, Nara Institute of Science and Technology, 2. Data science center, Nara Institute of Science and Technology, 3. PRESTO, Japan Science and Technology Agency)

9:40 AM - 10:00 AM

[2K1-04] Analysis of News and Stock Returns in Korean Markets Using High Frequency Data: Exploring Conditions of Classification Algorithm

Sungjae Yoon[1], ○Aiko Suge[1], Hiroshi Takahashi[1] (1. Keio University)

10:00 AM - 10:20 AM

[2K1-05] A Consideration of Estimating Spoiler Pages in Comics

○Yoshiki Maki[1], Yuji Shiratori[1], Kenta Sato[1], Satoshi

©The Japanese Society for Artificial Intelligence

Nakamura[1] （1. Meiji University）
10:20 AM - 10:40 AM

Oral presentation | NFC Session | [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

[2K3-NFC-3a] [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

座長:西村 拓一（産業技術総合研究所）
3:20 PM - 5:00 PM Room K (3F Azisai Mokuren)

[2K3-NFC-3a-01] Consideration of application cases of structured manual and its utilization

〇Satoshi Nishimura[1], Ken Fukuda[1], Takuichi Nishimura[1] （1. National Institute of Advanced Industrial Science and Technology）
3:20 PM - 3:40 PM

[2K3-NFC-3a-02] Knowledge Sharing and Acquisition among Teachers using Guitar Rendition Ontology

〇Nami Iino[2,3], Satoshi Nishimura[2], Takuichi Nishimura[2], Ken Fukuda[2], Hideaki Takeda[1,3] （1. National Institute of Informatics, 2. National Institute of Advanced Institute Science and Technology , 3. Graduate University for Advanced Studies）
3:40 PM - 4:00 PM

[2K3-NFC-3a-03] Development of "Tailor-made Brain Training" program suitable for environment

〇Chiaki Oshiyama[1], Hiroyasu Miwa[2], Chihiro Sutou[3], Eiji Shimizu[3], Takuichi Nishimura[2] （1. Osaka University, 2. National Institute of Advanced Industrial Science and Technology, 3. Chiba University）
4:00 PM - 4:20 PM

[2K3-NFC-3a-04] Design of Activity Program for expression :project for expression on caring mind

〇maiko kobayakawa[1], takeshi sunaga[2], sunao maruyama[2], tomoki hirano[1], kousuke yamada[4], takuichi nishimura[3], kentaro watanabe[3], sachiko fujimitsu[5] （1. Tokyo University of the arts, 2. Tokyo University of the Arts, 3. National Institute of Advanced Industrial Science and Technology, 4. Keio University , 5. Saga University hospital）
4:20 PM - 4:40 PM

[2K3-NFC-3a-05] Treatment strategy of Septicemia based on DPC data analysis by using pLSA method

〇Kazuya Yamashita[1], Yuichiro Sakamoto[2], Eiichi Sakurai[1,2], Yoichi Motomura[1,2] （1. National Institute of Advanced Industrial Science and Technology Artificial Intelligence Research Center, 2. Saga University）
4:40 PM - 5:00 PM

Oral presentation | NFC Session | [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

[2K4-NFC-3b] [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

座長:西村 拓一（産業技術総合研究所）
5:20 PM - 7:00 PM Room K (3F Azisai Mokuren)

[2K4-NFC-3b-01] Analysis and utilization of the action history of the event participants

〇Tomoyuki Owada[1], Kazuya Yamashita[1], Tomotaka Omae[1], Yoichi Motomura[1] （1. National Institute of Advanced Inductrial Science and Technology）
5:20 PM - 5:40 PM

[2K4-NFC-3b-02] Reflecting COTO Database Research: toward Implementation in the Society

〇Kentaro Watanabe[1], Hiroyasu Miwa[1], Satoshi Nishimura[1], Ken Fukuda[1], Takuichi Nishimura[1] （1. National Institute of Advanced Industrial Science and Technology）
5:40 PM - 6:00 PM

[2K4-NFC-3b-03] Product and Activity Design Support based on Event Database

〇Takuichi Nishimura[1], Kentarou Watanabe[1], Ken Fukuda[1], Satoshi Nishimura[1], Yoichi Motomura[1] （1. Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)）
6:00 PM - 6:20 PM

[2K4-NFC-3b-a] (Panel discussion) Data and Knowledge Construction, Sharing and Utilization in Real World

Satoshi Nishimura[1], Takuichi Nishimura[1], Ken

©The Japanese Society for Artificial Intelligence

Fukuda[1], Kentaro Watanabe[1], Kristina Jokinen[1], Yasuyuki Yoshida[1], Nami Iino[1], Lihua Zhao[1] （1. National Institute of Advanced Industrial Science and Technology）
6:20 PM - 7:00 PM

Room L

Oral presentation | General Session | [General Session] 9. NLP / IR

[2L1] [General Session] 9. NLP / IR

座長:柳瀬 利彦（株式会社 日立製作所）
9:00 AM - 10:40 AM Room L (3F Sapphire Hall Asuka)

[2L1-01] Location identification using local context and relevant documents
〇Ryo Seki[1], Takashi Inui[1] （1. University of Tsukuba）
9:00 AM - 9:20 AM

[2L1-02] Convolutional Neural Network for Event Knowledge Extraction
〇Kai Ishikawa[1], Hiroya Takamura[2], Manabu Okumura[2] （1. NEC Corporation, 2. Tokyo Institute of Technology）
9:20 AM - 9:40 AM

[2L1-03] Search for New Application of Invention based on Extraction of Causality from Patents
〇Takahisa Ota[1], Takuya Minami[1], Yusuke Yamazaki[1], Yoshishige Okuno[1], Chinatsu Tanabe[1], Hiroyuki Sakai[2], Hiroki Sakaji[3] （1. SHOWA DENKO K.K., 2. Seikei Univ., 3. The Univ. of Tokyo）
9:40 AM - 10:00 AM

[2L1-04] Mathematical Expression Retrieval in PDF documents from Web using Mathematical Terms as Queries
〇Kuniko Yamada[1], Harumi Murakami[1] （1. Osaka City University）
10:00 AM - 10:20 AM

[2L1-05] Politician classification based on political probrems extracted from Twitter Data
〇Mayumi Nomura[1], Tomonobu Ozaki[1] （1. College of Humanities and Sciences）
10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 9. NLP / IR

[2L4] [General Session] 9. NLP / IR

座長:角森 唯子（NTTドコモ）
5:20 PM - 7:00 PM Room L (3F Sapphire Hall Asuka)

[2L4-01] Measuring Beginner Friendliness of Web Pages explaining Academic Concepts by Integrating Heterogeneous Features
〇Shiokawa Hayato[1], Takahide Kasuga[1], Kota Kawaguchi[1], Bingcai Han[1], Takehito Utsuro[1], Yasuhide Kawada[2] （1. University of Tsukuba, 2. Logworks Co., Ltd.）
5:20 PM - 5:40 PM

[2L4-02] The Poc on automatic detection of erroneous telop using AI technology
〇Makoto Nakano[1] （1. Nippon Television Network Corporation）
5:40 PM - 6:00 PM

[2L4-03] A Simple Reproducible Judgment Method for Lexical Simplification based on Word Difficulty
〇Riku Takada[1], Tomoyuki Kajiwara[2], Noriyuki Okumura[1] （1. National Institute of Technology, Akashi College, 2. Osaka University）
6:00 PM - 6:20 PM

[2L4-04] Country Adaptation in Neural Machine Transliteration of Person Names
〇Dawoon Jung[1], Satoshi Sato[1] （1. Nagoya University）
6:20 PM - 6:40 PM

[2L4-05] Scoring support system using sequence alignment
〇Kohei Takai[1], Kengo Taketani[1], Yasukuni Mori[2], Hiroki Suyari[2] （1. Department of Informatics and Imaging Systems, Faculty of Engineering, Chiba University, 2. Graduate School of Engineering, Chiba University）
6:40 PM - 7:00 PM

Oral presentation | Organized Session | [Organized Session] OS-6

[2L2-OS-6a] [Organized Session] OS-6

1:20 PM - 3:00 PM Room L (3F Sapphire Hall Asuka)

[2L2-OS-6a-01] Emerging Interactions
〇Ichiro Tsuda[1] （1. Chubu University）
1:20 PM - 1:40 PM

[2L2-OS-6a-02] Spatial Concept Formation-based Semantic Mapping
〇Yuki Katsumata[1], Akira Taniguchi[1], Yoshinobu Hagiwara[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）
1:40 PM - 2:00 PM

[2L2-OS-6a-03] Learning actions using Generative Adversarial Imitation Learning added task achievement reward

©The Japanese Society for Artificial Intelligence

○Akira Kinose[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）
2:00 PM - 2:20 PM

[2L2-OS-6a-04] Analysis of Action Decision Algorithm Using Episodic Memory and Value Association

○Shintaro Kayanuma[1], Sana Kawazoe[1], Yuna Tsutsumi[1], Masahiro Miyata[2], Takashi Omori[1] （1. College of Engineering, Tamagawa University, 2. Graduate School of Engineering, Tamagawa University）
2:20 PM - 2:40 PM

[2L2-OS-6a-05] Reward learning through interaction between self-organization of a spiking neural network and a bodily behavior

○Yuji Kawai[1], Tomohiro Takimoto[1], Jihoon Park[1], Minoru Asada[1] （1. Graduate School of Engineering, Osaka University）
2:40 PM - 3:00 PM

Oral presentation | Organized Session | [Organized Session] OS-6

[2L3-OS-6b] [Organized Session] OS-6

3:20 PM - 5:00 PM Room L (3F Sapphire Hall Asuka)

[2L3-OS-6b-01] Symbol Emergence in Robotics towards Adaptive Autonomous AI Architecture

○Tadahiro Taniguchi[1] （1. Ritsumeikan University）
3:20 PM - 3:40 PM

[2L3-OS-6b-02] Validation of symbolic inference based on a value driven associative memory

○Masahiro Miyata[1], Takashi Omori[2] （1. Graduate School of Engineering, Tamagawa University, 2. College of Engineering, Tamagawa University）
3:40 PM - 4:00 PM

[2L3-OS-6b-03] Active SLAM and Spatial Concepts Formation for a Mobile Robot in Home Environment

○Yoshiki Tabuchi[1], Akira Taniguchi[1], Yoshinobu Hagiwara[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）
4:00 PM - 4:20 PM

[2L3-OS-6b-04] Stories as Living Representation

○Taisuke Akimoto[1] （1. Kyushu Institute of Technology）
4:20 PM - 4:40 PM

[2L3-OS-6b-05] A Capability Map for Guiding AGI Developments

Masahiko Osawa[2], ○Takashi Omori[1], Koichi Takahashi[3], Naoya Arakawa[4], Naoyuki Sakai[4], Michihiko Ueno[4], Michita Imai[2], Hiroshi Yamakawa[4] （1. Tamagawa University, 2. Keio University, 3. RIKEN, 4. Dwango Artificial Intelligence Labo.）
4:40 PM - 5:00 PM

Room M

Oral presentation | General Session | [General Session] 13. AI Application

[2M2] [General Session] 13. AI Application

座長:大西 貴士（NEC/産総研）
1:20 PM - 2:40 PM Room M (2F Amethyst Hall Hoo)

[2M2-01] Utilizing deception information for dialog management of doctor-patient conversations.

○Tung The Nguyen[1], Koichiro Yoshino[1], Sakriani Sakti[1], Satoshi Nakamura[1] （1. Graduate School of Information Science, Nara Institute of Science and Technology）
1:20 PM - 1:40 PM

[2M2-02] Cost reduction using stochastic programming in supply chain management

○TAKESHI KONNO[1], Ken Kobayashi[1], Shuji Awai[2], Go Yamamoto[2] （1. FUJITSU LABORATORIES LTD., 2. FUJITSU SOCIAL SCIENCE LABORATORY LTD.）
1:40 PM - 2:00 PM

[2M2-03] AI utilization for membrane structure design and manufacturing

○Kazuo Yokobori[1], Tomo Miura[1] （1. Taiyo Kogyo Corporation）
2:00 PM - 2:20 PM

[2M2-04] Optimization for the Migration Units in Information Systems based on CRUD Relations with Semi-Supervised Graph Neural Networks

○Kugatsu Sadamitsu[1] （1. Future Corporation）
2:20 PM - 2:40 PM

Oral presentation | General Session | [General Session] 13. AI Application

[2M3] [General Session] 13. AI Application

座長:荒井 幸代（千葉大学）
3:20 PM - 5:00 PM Room M (2F Amethyst Hall Hoo)

[2M3-01] Drone Flying Area Estimation Method based on Deep Learning II

○Masatoshi Hamanaka[1] （1. RIKEN）

©The Japanese Society for Artificial Intelligence

3:20 PM - 3:40 PM

[2M3-02] Toward Building the Evaluation Dataset of Non-Signal Crossing Images

○Shinichiro Goto[1], Yuki Murata[1], Masayasu Atsumi[1] （1. Graduate school of Eng., Soka University）

3:40 PM - 4:00 PM

[2M3-03] Large Scale Driver Identification Using Driving Data

○Daiki Tanaka[1,2], Yukino Baba[1,2], Hisashi Kashima[1,2], Tomoya Saito[3,2], Yuta Okubo[4,2] （1. Kyoto University, 2. RIKEN Center for AIP, 3. SOMPO RISK MANAGEMENT &HEALTH CARE, 4. Sompo Japan Nipponkoa Insurance Inc.）

4:00 PM - 4:20 PM

[2M3-04] Application of individual behavior model in machine learning based cafeteria visitor prediction model and its effectiveness analysis

○Kiyotaka Matsue[1], Hideki Noda[1], Koichi Kondo[1] （1. Toshiba Infrastructure Systems &Solutions Corporation）

4:20 PM - 4:40 PM

[2M3-05] A Preliminary Implementation of Followee Recommendation Mechanism with Self-evaluation Maintenance Model on Mastodon

○Takuma Matsushita[1], Naoki Fukuta[1] （1. Shizuoka University）

4:40 PM - 5:00 PM

Room N

Oral presentation | General Session | [General Session] 10. Vision / Speech

[2N1] [General Session] 10. Vision / Speech

座長:辻川 剛範（NEC）

9:00 AM - 10:20 AM Room N (2F Sakurajima)

[2N1-01] Expressive Speech Synthesis through modeling the variety of expressions by Variational Autoencoder

○Kei Akuzawa[1], Iwasawa Yusuke[1], Matsuo Yutaka[1] （1. University of Tokyo）

9:00 AM - 9:20 AM

[2N1-02] Response Timing Estimation Using Dialogue Situations for Spoken Dialogue Systems

○Genki Akai[1], Ryu Takeda[1], Kazunori Komatani[1] （1. The Institute of Scientific and Industrial Research, Osaka University）

9:20 AM - 9:40 AM

[2N1-03] Incremental estimation of user's turn-taking state

○Shinya Fujie[1,2], Katsuya Yokoyama[2], Tetsunori Kobayashi[2] （1. Chiba Institute of Technology, 2. Waseda University）

9:40 AM - 10:00 AM

[2N1-04] Semantic-Usage Classification of Mimetic Words using Sound Symbolism

○Natsumi Takeuchi[1], Daiki Urata[1], Tsuyoshi Nakamura[1], Masayoshi Kanoh[2], Koji Yamada[3] （1. Nagoya Institute of Technology, 2. Chukyo University, 3. Institute of Advanced Media Arts and Sciences）

10:00 AM - 10:20 AM

Room O

Oral presentation | General Session | [General Session] 3. Data Mining

[2O1] [General Session] 3. Data Mining

座長:原 聡（大阪大学）

9:00 AM - 10:20 AM Room O (2F Kaimon)

[2O1-01] Modeling Stream Data with Fluctuated Periodicity and Periodicity Analysis Using Decision

○Koji Ohba[1], Mitsuki Kimura[1], Kazuya Watanabe[1] （1. Honda R& D Co.,Ltd）

9:00 AM - 9:20 AM

[2O1-02] Marketing data analysis by hierarchical pattern mining based on swarm intelligence

○Kazuaki Tsuboi[1], Satoshi Suga[1], Hiroaki Kanemitsu[2], Keiji Sakamoto[3], Satoshi Kurihara[1] （1. The University of Electro-communications, 2. {Trial Company, Inc., 3. Trial Holdings, Inc.）

9:20 AM - 9:40 AM

[2O1-03] Graph-based Visualization of Timeseries Information in Multiple Texts

○Hironari Kawada[1], Mina Akaishi[2], Hiroshi Hosobe[2] （1. Graduate School of Computer and Information Sciences, Hosei University, 2. Faculty of Computer and Information Sciences, Hosei University）

9:40 AM - 10:00 AM

[2O1-04] Tourist Transition Model among Sightseeing Spot based on Trajectory Data

○Takeshi Watabe[1], Hidekazu Kasahara[2], Masaaki Iiyama[2], Michihiko Minoh[2] （1. Undergraduate School of Informatics and Mathematical Science, Kyoto University, 2. Academic Center for Computing and

©The Japanese Society for Artificial Intelligence

Media Studies, Kyoto University)
10:00 AM - 10:20 AM

Oral presentation | General Session | [General Session] 3. Data Mining

[2O4] [General Session] 3. Data Mining

座長:大知 正直（東京大学）
5:20 PM - 7:00 PM Room O (2F Kaimon)

[2O4-01] Sightseeing Tweet Classification using Convolutional Neural Networks
○Shuichi Hashida[1], Keiichi Tamura[1], Tatsuhiro Sakai[1] (1. Hiroshima city university)
5:20 PM - 5:40 PM

[2O4-02] User constrained independent topic analysis
○KENTA YAMAMOTO[1], TAKAHIRO NISHIGAKI[1], TAKASHI ONODA[1] (1. AOYAMA GAKUIN UNIVERSITY)
5:40 PM - 6:00 PM

[2O4-03] Discussion on topic change detection method in Incremental Independent Topic Analysis
○Takahiro Nishigaki[1], Takashi Onoda[1] (1. Aoyama Gakuin University)
6:00 PM - 6:20 PM

[2O4-04] Extraction of Causal Knowledge from Annual Securities Report
○Fumihito Sato[1], Hiroaki Sakuma[1], Shunya Kodera[1], Yoshinori Tanaka[1], Hiroki Sakaji[2], Kiyoshi Izumi[2] (1. Nikko Research Center, Inc., 2. Graduate School of Engineering, The University of Tokyo)
6:20 PM - 6:40 PM

[2O4-05] Nursing Literature Review Matrix System based on Machine Learning
○Takanori Yamashita[1], Rieko Izukura[1], Naoki Nakashima[1], Sachio Hirokawa[1] (1. Kyushu University)
6:40 PM - 7:00 PM

Oral presentation | Organized Session | [Organized Session] OS-24

[2O2-OS-24a] [Organized Session] OS-24

1:20 PM - 3:00 PM Room O (2F Kaimon)

[2O2-OS-24a-01] Analysis using a Triangular Pyramid Game Simulating Restaurant Service
○Tomomi Nonaka[1], Nobutada Fujii[2], Takeshi Shimmura[1] (1. Ritsumeikan University, 2. Kobe University)
1:20 PM - 1:40 PM

[2O2-OS-24a-02] Finding cultural differences on food using dish descriptions from social media: a comparative study of culinary perceptions and preferences between Japanese and Americans
○Yukie Nakamoto[1], Masanao Ochi[1], Takeshi Sakaki[2], Junichiro Mori[3], Ichiro Sakata[1] (1. Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 2. Hotto Link Inc., 3. Graduate School of Information Science and Technology, The University of Tokyo)
1:40 PM - 2:00 PM

[2O2-OS-24a-03] An analysis of sizzle words based on co-occurrence networks
Shingo Kasai[1], ○Fumiaki Saitoh[1], Syohei Ishizu[1] (1. Aoyama Gakuin University)
2:00 PM - 2:20 PM

[2O2-OS-24a-04] Analysis of waiters know-how using a game
○Fumihide Nishimura[1], Emi Kinoshita[1], Tomomi Nonaka[1], Hajime Mizuyama[1] (1. aoyamagakuindaigaku)
2:20 PM - 2:40 PM

[2O2-OS-24a-05] Approach to improvement of restaurant management using machine language
○Tomohiro Hoshino[1], Takashi Tanizaki[1], Takeshi Shimmura[2], Takeshi Takenaka[3] (1. Graduate School of Kindai University, 2. Ritsumeikan University, 3. National Institute of Advanced Industrial Science and Technology)
2:40 PM - 3:00 PM

Oral presentation | Organized Session | [Organized Session] OS-24

[2O3-OS-24b] [Organized Session] OS-24

3:20 PM - 4:40 PM Room O (2F Kaimon)

[2O3-OS-24b-01] Estimation of Unwritten Cooking Utensils in Recipe Texts
Kento Nagai[1], ○Takafumi Ohsugi[1], Mitsunori Matsushita[1] (1. Kansai University)
3:20 PM - 3:40 PM

[2O3-OS-24b-02] How can we apply AI to a study of gastronomy?
○Hiroshi Izawa[1], Ryosuke Yamanishi[1] (1. Ritsumeikan University)
3:40 PM - 4:00 PM

©The Japanese Society for Artificial Intelligence

[2O3-OS-24b-03] Improving Labor Productivity for Japanese Cuisine Restaurant Using Simulator

○Takeshi Shimmura[1], Nobutada Fujii[2], Tomomi Nonaka[3] （1. Ritsumeikan University, 2. Graduate School of System Informatics, Kobe University, 3. Department of Industrial and System Engineering, Aoyama Gakuin University）

4:00 PM - 4:20 PM

[2O3-OS-24b-04] Community Value Co-creation Using Home Cooking-plate Network Systems

○Nobutada Fujii[1], Tomomi Nonaka[2], Akio Shimomura[3], Jun Yamadera[4] （1. Kobe University, 2. Ritsumeikan University, 3. Shimomura Shikkiten Co., Ltd., 4. Eyes, JAPAN Co., Ltd.）

4:20 PM - 4:40 PM

Room P

Oral presentation | General Session | [General Session] 2. Machine Learning

[2P3] [General Session] 2. Machine Learning

座長:木村 圭吾（NEC）

3:20 PM - 5:00 PM Room P (4F Emerald Lobby)

[2P3-01] Feasibility Study on PUC for Measurement Noise Reduction

○Takeshi Yoshida[1], Takashi Washio[1], Takahito Ooshiro[1], Masateru Taniguchi[1] （1. The Institute of Scientific and Industrial Research Osaka University）

3:20 PM - 3:40 PM

[2P3-02] Bayesian network classifier learned by conditional marginal likelihood

○Shouta Sugahara[1], Masaki Uto[1], Maomi Ueno[1] （1. The University of Electro-Communications）

3:40 PM - 4:00 PM

[2P3-03] Influence of Deterministic Decision Rules in Fair Logistic Regression

○Toshihiro Kamishima[1], Shotaro Akaho[1], Hideki Asoh[1], Jun Sakuma[2,3] （1. National Institute of Advanced Industrial Science and Technology (AIST), 2. University of Tsukuba, 3. RIKEN Center for Advanced Intelligence Project）

4:00 PM - 4:20 PM

[2P3-04] Data Fusion Method with Kernel Matching for Estimated Canonical Variables

○Masaki Mitsuhiro[1,2], Takahiro Hoshino[3,4] （1. Nikkei Research Inc., 2. Graduate School of Economics, Keio University, 3. Department of Economics, Keio University, 4. RIKEN Center for Advanced Intelligence Projec）

4:20 PM - 4:40 PM

[2P3-05] Multi-label Logistic Regression Using Relative Density Ratio

○Masaaki Okabe[1], Jun Tsuchida[2], Hiroshi Yadohisa[3] （1. Graduate School of Cluture and Infomation Science, Doshisha University, 2. Faculty of Engineering, Tokyo University of Science, 3. Faculty of Culture and Information Science, Doshisha University）

4:40 PM - 5:00 PM

Oral presentation | General Session | [General Session] 7. Agent

[2P1] [General Session] 7. Agent

座長:田中 友紀子（NEC）

9:00 AM - 10:40 AM Room P (4F Emerald Lobby)

[2P1-01] Building Cooperative Relationship by Continuous Observation of Interaction between Two Agents

○Junya Karasaki[1], Yoshimasa Ohmoto[1], Toyoaki Nishida[1] （1. Graduate School of Informatics, Kyoto University）

9:00 AM - 9:20 AM

[2P1-02] Detecting Suspects by Observing Behavioral Changes of Surrounding Pedestrians

○Reina Saitou[1], Ei-Ichi Osawa[1] （1. Future University Hakodate）

9:20 AM - 9:40 AM

[2P1-03] Parallelization of evolution of reinforcement learning agents using GPGPU

○Yoshiki Senga[1], Kouichi Moriyama[1], Atsuko Mutoh[1], Tohgoroh Matsui[2], Inuzuka Nobuhiro[1] （1. Nagoya Institute of Technology, 2. Chubu University）

9:40 AM - 10:00 AM

[2P1-04] Hedonic games considering agents sociality

○Kazunori Ota[1], Nathanael Barrot[2], Yuko Sakurai[3], Makoto Yokoo[1,2] （1. Graduate School of Information Science and Electrical Engineering, Kyushu University, 2. RIKEN, Center for Advanced Intelligence Project AIP, 3. AIST Artificial Intelligence Research Center International Symposium）

10:00 AM - 10:20 AM

©The Japanese Society for Artificial Intelligence

[2P1-05] Intelligent agents supporting to estimation of cyber threat intelligence

○Hidetoshi Kawaguchi[1], Yuichi Ishihara[1] （1. NIPPON TELEGRAPH AND TELEPHONE CORPORATION）

10:20 AM - 10:40 AM

Oral presentation | General Session | [General Session] 7. Agent

[2P2] [General Session] 7. Agent

座長:落合 友四郎（大妻女子大学）

1:20 PM - 3:00 PM Room P (4F Emerald Lobby)

[2P2-01] Accelerate Deep Q Network by weighting experiences

○Kazuhiro Murakami[1], Koichi Moriyama[1], Atsuko Mutoh[1], Tohgoroh Matsui[2], Nobuhiro Inuzuka[1] （1. Nagoya Institute of Technology, 2. Chubu University）

1:20 PM - 1:40 PM

[2P2-02] Effect of Asset Price Fluctuation and Inter-bank Lending and Borrowing Network on Chain Bankruptcy of Bank

○Ryo Hamawaki[1], Kiyoshi Izumi[1], Hiroki Sakaji[1], Hiroto Yonenoh[1] （1. The University of Tokyo）

1:40 PM - 2:00 PM

[2P2-03] Trade-off between search time and responsiveness of MCTS in a ghting video game

○Wataru Kamegaki[1], Koichi Moriyama[1], Atsuko Mutoh[1], Nobuhiro Inuzuka[1] （1. Nagoya Institute of Technology）

2:00 PM - 2:20 PM

[2P2-04] 3D L-Systems Path-finding

○Djuned Fernando Djusdek[1], Yoshiteru Ishida[1] （1. Toyohashi University of Technology）

2:20 PM - 2:40 PM

[2P2-05] Effects Analysis of CAR Regulations on Financial Markets using Artificial Market Simulations

○Masanori HIRANO[1], Hiroto YONENOH[2], Kiyoshi IZUMI[2] （1. Faculty of Engineering, The University of Tokyo, 2. School of Engineering, The University of Tokyo）

2:40 PM - 3:00 PM

Room Z

Oral presentation | General Session | [General Session] 5. Web Intelligence

[2Z2] [General Session] 5. Web Intelligence

座長:西村 悟史（産業技術総合研究所）

1:20 PM - 2:40 PM Room Z (3F Matsu Take)

[2Z2-01] Synthetic Accessibility Assessment Using Auxiliary Responses

○Shun Ito[1], Yukino Baba[1], Tetsu Isomura[2], Hisashi Kashima[1,3] （1. Graduate school of Informatics, Kyoto University, 2. Mitsubishi Chemical Holdings Corporation, 3. RIKEN Center for Advanced Intelligence Project）

1:20 PM - 1:40 PM

[2Z2-02] Crowd Neural Network

○Yusuke Sakata[1], Yukino Baba[1], Kashima Hisashi[1,2] （1. Kyoto University Department of Intelligence Science and Technology, 2. RIKEN Center for AIP）

1:40 PM - 2:00 PM

[2Z2-03] LOD Surfer API: A web API for LOD exploration based on class-class relationships

○Atsuko Yamaguchi[1], Norio Kobayashi[2], Hiroshi Masuya[2], Yasunori Yamamoto[1], Kouji Kozaki[3] （1. Research Organization of Information and Systems, 2. RIKEN, 3. Osaka University）

2:00 PM - 2:20 PM

[2Z2-04] A Preliminary Approach on a Cooperative Query Editing System using Dynamic Ontology Mappings

○Takuya Adachi[1], Naoki Fukuta[2] （1. Graduate School of Integrated Science and Technology, Shizuoka University, 2. College of Informatics, Academic Institute, Shizuoka University）

2:20 PM - 2:40 PM

Oral presentation | General Session | [General Session] 5. Web Intelligence

[2Z3] [General Session] 5. Web Intelligence

座長:大向 一輝（国立情報学研究所）

3:20 PM - 4:40 PM Room Z (3F Matsu Take)

[2Z3-01] Quantification of Diverse Personal Attributes in Tweets

○take yo[1], Ayahito Saji[1], Kazutoshi Sasahara[1] （1. nagoya university）

3:20 PM - 3:40 PM

[2Z3-02] Analysis of Minor Women's Online Risk

○Yuichi Hirano[1], Fujio Toriumi[1], Masanori Takano[2], Kazuya Wada[2], Ichiro Fukuda[2] （1. The University of Tokyo, 2. CyberAgent, Inc.）

3:40 PM - 4:00 PM

[2Z3-03] Preliminary Investigation for Analysis of Browsing News Pages on Smartphone App and Wikipedia

○Mitsuo Yoshida[1], Yoshifumi Seki[2] （1. Toyohashi University of Technology, 2. Gunosy Inc.）

©The Japanese Society for Artificial Intelligence

4:00 PM - 4:20 PM

[2Z3-04] Comparison of Classification Methods for Ascii Art

○Kazuyuki Matsumoto[1], Akira Fujisawa[2], Minoru Yoshida[1], Kenji Kita[1] （1. Tokushima University, 2. Aomori University）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

Oral presentation | General Session | [General Session] 2. Machine Learning

[2A1] [General Session] 2. Machine Learning

座長:竹内 孝（NTTコミュニケーション科学基礎研究所）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room A (4F Emerald Hall)

[2A1-01] Dual Convolutional Neural Network for Graph of Graphs Link Prediction

○Shonosuke Harada[1], Hirotaka Akita[1], Masashi Tsubaki[2], Yukino Baba[1], Ichigaku Takigawa[3], Yoshihiro Yamanishi[4], Hisashi Kashima[1,5] （1. Kyoto University, 2. National Institute of Advanced Industrial Science and Technology, 3. Hokkaido University, 4. Kyushu University, 5. RIKEN Center for Advanced Intelligence Project）

9:00 AM - 9:20 AM

[2A1-02] Predicting Eruptions of Sakurajima by Stacked Recurrent Neural Network

○Tsuyoshi Murata[1], Hiep Le[1], Masato Iguchi[2] （1. Tokyo Institute of Technology, 2. Kyoto University）

9:20 AM - 9:40 AM

[2A1-03] Anomaly Manufacturing Product Detection using Unregularized Anomaly Score on Deep Generative Models

Ryosuke Tachibana[1], Takashi Matsubara[1], ○Kuniaki Uehara[1] （1. Graduate School of System Informatics, Kobe Universiry）

9:40 AM - 10:00 AM

[2A1-04] Pseudo-feature generation from feature map in deep learning for imbalanced data multi-class image classification

○Tomohiko Konno[1], Hideaki Fujii[1], Michiaki Iwazume[1] （1. AI Science R&D Promotion Center in National Institute of Information and Communication Technology ）

10:00 AM - 10:20 AM

[2A1-05] Route Generation of Construction Machine by Deep Reinforcement Learning

○Shunya Tanabe[1], Zeyuan Sun[1], Masayuki Nakatani[1], Yutaka Uchimura[1] （1. Shibaura Institute of Technology）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Graph of Graphsに対する二重畳み込みニューラルネットワーク

Dual Convolutional Neural Network for Graph of Graphs Link Prediction

原田 将之介 *1 Shonosuke Harada
秋田 大空 *1 Hirotaka Akita
椿 真史 *2 Masashi Tsubaki
馬場 雪乃 *3 Yukino Baba
瀧川 一学 *4 Ichigaku Takigawa
山西 芳裕 *5 Yoshihiro Yamanishi
鹿島 久嗣 *1*6 Hisashi Kasima

*1京都大学 Kyoto University
*2産業技術総合研究所 National Institute of Advanced Industrial Science and Technology
*3筑波大学 University of Tsukuba
*4北海道大学 Hokkaido University
*5九州大学 Kyushu University
*6理化学研究所 革新知能統合研究センター RIKEN Center for Advanced Intelligence Project

A graph is a general and powerful data representation for complex real-world phenomena ranging from chemical compounds to social networks; however, effective feature extraction from graphs is not a trivial task, and much work has been done in the field of machine learning and data mining. In this paper, we propose a dual convolution approach that extracts node representations of a *graph of graphs (GoG)* consisting of an external graph and internal graphs, by combining the external and internal graph structures in an end-to-end manner. Experiments on various types of link prediction tasks demonstrate the effectiveness of the proposed method.

1. はじめに

グラフは、化合物やソーシャルネットワークのような現実世界の複雑な現象をモデル化する一般的かつ強力なデータ表現である。しかしながら、データ解析においてそのようなグラフ構造の扱いは自明ではない。というのも、ほとんどの機械学習手法はデータは予め固定長の特徴ベクトルで表現されうるという前提の上に成り立っているからである。従って、グラフ構造データの分析は多くの研究者の関心を集めている課題のひとつである。

グラフ構造は大きく 2 種類に分けることができる。データ内部の構造を表現する内部グラフ及びデータ間の関係性を表現する外部グラフである。内部グラフの例として化合物や Web ページ、外部グラフの例としてソーシャルネットワークや Web ページのリンク関係を表すネットワーク等がそれぞれ挙げられる。本研究では、両者の特徴を持つ図 1 のような graph of graphs (GoG) と呼ばれるグラフに対する適切な特徴表現の獲得を目的とする。GoG の例として、化合物相互作用ネットワークや Web ページリンクネットワークが挙げられる。

単純な一つのグラフの集合より複雑で一般的なグラフ構造を持つ GoG に対して、これまで異なる文脈で別々に研究が行われてきた内部、外部グラフの両者の特徴を共に考慮する手法を提案し、3 つの実 GoG データセットを使ったリンク予測の実験において、二重畳み込み法がベースライン手法と比較して低次元特徴ベクトルを用いた場合でも、ベースライン手法より優れていることを示す。

2. 問題設定

本研究において、GoG とは、グラフの各ノード自体が、ある別のグラフ構造を持つグラフ構造であると定義する。本論文では元になるグラフを外部グラフ、外部グラフのノードを構成するグラフを内部グラフと呼び、外部グラフにおけるノード間のリンク予測を行い、得られた特徴ベクトルの適切さを評価する。一般的に、GoG は内部グラフと外部グラフの 2 層だけでなく、何層にでもなりうるが、本研究では簡単のため 2 層の GoG のみを扱う。なお、本研究のアプローチは何層の GoG に対しても容易に応用可能である。

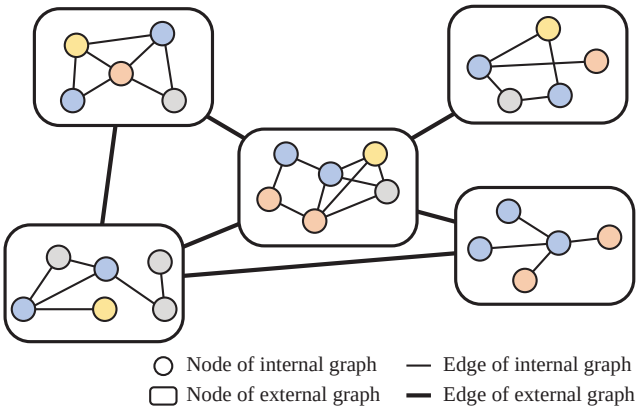


図 1: GoG の例

2 層の GoG は、$\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{E})$ として表されるグラフである。$\mathcal{V}$ は外部グラフのノードの集合、$\mathcal{A}$ は外部グラフの隣接リスト、$\mathcal{E}$ は外部グラフのノード間のリンクの集合を表す。外部グラフ $\mathcal{G}$ の i 番目のノードは内部グラフ $G_i \in \mathcal{V}$ をもつ。簡単のために、以降 $G_i = G = (V, A, E)$ と標記する。ここで、V は内部グラフのノードの集合、A は内部グラフの隣接リスト、E は内部グラフのノード間のエッジの集合である。簡単のために、本研究では E の種類については考慮しない。

本問題の入出力をまとめると以下のようになる：

入力： $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{E})$

出力： $G_i, G_j \in \mathcal{V}$ 間にリンクが存在する確率

連絡先: 原田 将之介, sh1108@ml.ist.i.kyoto-u.ac.jp

3. 関連研究

本研究は、これまで異なる背景で別々に研究が行われてきた内部グラフ及び外部グラフの統合を試みているため、これまでのグラフ学習の研究をそれら 2 種類のグラフの観点から述べる。

内部的なグラフ構造を持ったデータに対する分析は、データマイニング分野に端を発する。その根底にある考え方は、グラフの特徴は局所的な部分構造にあるとするものであり、そのような局所部分構造を見つけるような拡張が成された頻出パターンマイニング手法 [Inokuchi 00, Yan 02] が提案された。明示的にグラフの部分構造の特徴を見つけるグラフマイニングとは対照的に、カーネル法は暗黙的に特徴を抽出する。初期のグラフカーネル法 [Kashima 03, Gärtner 03] は、パスを特徴ベクトルとして用いた。昨今のグラフカーネルの最先端の手法 [Shervashidze 11] はより複雑な部分構造の形式を用いている。

ニューラルネットワークもまたグラフに対する柔軟な特徴抽出に成功している。Scarse らによるもの [Scarselli 09] に続いて、様々なグラフ畳み込みモデルが提案されてきた。例えば、Duvenaud らによるもの [Duvenaud 15] や Niepert らによるもの [Niepert 16] が挙げられる。

外部グラフ構造を持ったデータに対する分析は、しばしばリンクマイニングと呼ばれる [Getoor 05]。リンクマイニングの典型的なタスクには、リンク予測の他にランキング、クラスタリング、ノード分類等がある。近年、グラフ内でのノードの近接性を維持するノード埋め込みによる研究 [Socher 13, Yang 15] がますます行われている。特に Yang ら [Yang 15] の考え方は、評価実験におけるベースラインとして用いている。

グラフ構造を持つデータに対する研究は非常に幅広いため、ここで全ての範囲を網羅することは困難である。重要な点は、これまで内部グラフと外部グラフは別々に研究が行われてきていたという点である。

4. 二重畳み込みニューラルネットワーク

GoG の特徴表現を学習するために、本研究ではニューラルネットワークを用いて内部グラフ及び外部グラフを考慮した一貫学習を行う。

4.1 内部グラフに対する畳み込み

G の k 番目のノード v_k の特徴ベクトルを $\mathbf{v}_k \in \mathbb{R}^d$ とする。d は特徴ベクトルの次元を表し、モデルのハイパーパラメータの一つである。各 $\mathbf{v}_k$ の初期値 $\mathbf{v}_k^{(0)}$ はノードの種類 (例えば、水素や酸素) に応じてランダムに初期化され、その後、外部グラフに対するリンク予測の際の誤差逆伝搬で更新される。

グラフ $G = (V, A)$ におけるノードの特徴ベクトルを $\mathbf{v}_k$ として、畳み込みステップ $t = 1, 2, \ldots, T$ における $\mathbf{v}_k$ を $\mathbf{v}_k^{(t)}$ と表し、各畳み込みステップにおいて以下のニューラルネットワークを用いて特徴を更新する。

$$\mathbf{v}_k^{(t+1)} = f_G \left(\mathbf{W}\mathbf{v}_k^{(t)} + \sum_{m \in A_k} \mathbf{M}\mathbf{v}_m^{(t)} \right) \tag{1}$$

ここで f_G は例えば ReLU のような非線形関数、A_k は k 番目の隣接リストの索引、$\mathbf{W} \in \mathbb{R}^{d \times d}$、$\mathbf{M} \in \mathbb{R}^{d \times d}$ は学習される重み行列を示す。内部グラフに対する畳み込みについての関数の詳細な設定は、次節で説明する外部グラフに対する畳み込みの設定と共に、 5.2 節で改めて述べる。

最終的に $G = (V, A)$ における全てのノードの特徴を以下のように全ての畳み込みステップを通して足し合わせることで、内部グラフの特徴表現 $\mathbf{g}^{(T)} \in \mathbb{R}^d$ を得る。

$$\mathbf{g}^{(T)} = \sum_{k \in V} g \left(\sum_{t=0}^{T} \mathbf{v}_k^{(t)} \right) \tag{2}$$

ここで、T は内部グラフに対する畳み込みを行う回数、g は非線形関数である。続いて、i 番目の内部グラフの特徴表現 $\mathbf{g}^{(T)}$ を以降では改めて $\mathbf{g}_i^{(0)}$ と表記する。右肩の 0 は次節で説明する外部グラフに対する畳み込みの初期特徴ベクトルであることを表している。

4.2 外部グラフに対する畳み込み

上述した内部グラフの特徴表現 $\mathbf{g}_i^{(0)}$ を用いて以下に示すような外部グラフに対する畳み込みを行う。

$$\mathbf{g}_i^{(\ell+1)} = f_{\mathcal{G}} \left(\mathbf{U}\mathbf{g}_i^{(\ell)} + \sum_{n \in \mathcal{A}_i} \mathbf{V}\mathbf{g}_n^{(\ell)} \right) \tag{3}$$

ここで、$f_{\mathcal{G}}$ は非線形関数、$\mathcal{A}_i$ は i 番目の隣接リストの索引、$\mathbf{U} \in \mathbb{R}^{d \times d}$ 及び $\mathbf{V} \in \mathbb{R}^{d \times d}$ は学習される行列パラメータを表す。そして、以下のように全ての畳み込みステップを考慮して新たな内部グラフの特徴表現を得る。

$$\mathbf{g}_i^{(L)} = g \left(\sum_{\ell=0}^{L} \mathbf{g}_i^{(\ell)} \right) \tag{4}$$

ここで L は外部グラフに対する畳み込みを行う回数、g は非線形関数を表す。また $\mathbf{g}_i^{(L)}$ は内部グラフ G_i の外部グラフに対する畳み込みの畳み込みステップ L における特徴表現を表す。本手法の目的は外部グラフ $\mathcal{G}$ の特徴表現を獲得することではなく、内部グラフ G_i の適切な特徴表現を獲得することであることに注意されたい。本研究では $\mathbf{g}_i^{(L)}$ を内部グラフ G_i の二重畳み込み特徴表現と呼ぶ。これは、次節で説明する外部グラフにおけるリンク予測の際に入力される特徴ベクトルとなる。

4.3 リンク予測

2 つの内部グラフ G_i 及び G_j 間のリンク予測は、それらの二重畳み込み特徴表現 $\mathbf{g}_i^{(L)}$ 及び $\mathbf{g}_j^{(L)}$ を用いて行われる。以下のように入力として、2 つの特徴ベクトル $\mathbf{g}_i^{(L)}$ 及び $\mathbf{g}_i^{(L)}$ を受け取り、ニューラルネットワークを用いてリンクの有無の度合いを出力する関数 h によって 2 次元ベクトル $\boldsymbol{y} \in \mathbb{R}^2$ を出力する。

$$\mathbf{y} = h \left(\mathbf{g}_i^{(L)}, \mathbf{g}_j^{(L)} \right) \tag{5}$$

さらに、softmax 層によってリンクの確率を以下のように与える。

$$p_t = \frac{\exp(y_t)}{\sum_k \exp(y_k)}$$

ここで $t \in \{0, 1\}$ は 2 値 (すなわちリンクが存在しているか否か) のラベル、p_t は t の確率をそれぞれ表している。関数 h について詳細は 5.2 節で述べるが、入力される i 及び j について、h によって対称性は維持されていることに注意されたい。

5. 評価実験

5.1 データセット

GoG に対する内部グラフ及び外部グラフを考慮する一貫学習の有用性を確認するため、リンク予測を行う。リンク予測実験で用いた 3 つの実データセットについて説明する。

5.1.1 標的共有ネットワーク

標的共有ネットワークはタンパク質と化合物の相互作用についてのデータセット [Takigawa 11] である。元々のグラフは 2 部グラフであるが、この実験では、このデータセットでは同じタンパク質と作用する薬同士を繋いだ薬剤のみのネットワークに変換して用いる。化合物に含まれる原子数が 64 以下の化合物 3,918 を用いた。化合物のペア ${}_{3918}C_2 = 7{,}673{,}403$ のうち、リンクがあるものは 35,562 であり、残りは負例として扱う。上述のように、薬剤のみのネットワークに変換しているが、本研究では便宜上このデータを標的共有ネットワークと呼ぶ。

5.1.2 薬剤相互作用ネットワーク

薬剤化合物のネットワークで、2 つの化合物間で、同時に服用すると相互作用や、人体に悪影響のある反応が見られる場合にその 2 つの化合物の間にエッジが存在する。化合物に含まれる原子数が 64 以下の化合物 1,993 を用いる。薬剤ペア ${}_{1993}C_2 = 1{,}985{,}028$ のうち、リンクがあるものは 186,555 であり、残りは負例として扱う。

5.1.3 代謝化合物ネットワーク

代謝化合物のネットワークであり、2 つの化合物が代謝経路上の酵素反応において、基質及び生成物の対である場合にそれらの間にリンクを張る [Kotera 14]。化合物に含まれる原子数が 64 以下の化合物 5,920 を用いる。化合物ペア ${}_{5920}C_2 = 17{,}520{,}240$ のうち、リンクがあるものは 5,041、リンクがないものは 220,096 であり、残りは正解ラベルが与えられておらず、今回の実験では扱わない。

5.2 提案手法の設定の詳細

2 つの畳み込みネットワーク内で用いる非線形関数には選択の幅がある。本実験では以下のような選択を行う。

式 (1) における関数 f_G として ReLU を用いる。また行列パラメータ $\mathbf{W}$ 及び $\mathbf{M}$ を次数 (すなわち $|A_k|$ 及び $|A_m|$) とその畳み込みステップに応じて初期化する。重み行列パラメータをそれぞれの次数について初期化するという考え方は、Duvenaud ら [Duvenaud 15] の考えに基づいている。式 (2) 及び式 (4) における関数 g として softmax 関数を用いる。式 (1) 及び式 (3) における関数 f_G として softmax 関数を用い、行列パラメータ $\mathbf{U}$ 及び $\mathbf{V}$ を畳み込みステップに応じて初期化する。$\mathbf{W}$ 及び $\mathbf{M}$ とは異なり、$\mathbf{U}$ 及び $\mathbf{V}$ については本実験で扱ったデータセットでは、外部グラフが巨大で各ノードが非常に大きな次数を持っているために次数を考慮しない。また今回の実験では $\mathbf{U}$=$\mathbf{V}$ としている。 4.3 節における関数 h として以下のような関数を用いた。関数 h に対して入力される $\mathbf{g}_i^{(L)}$ 及び $\mathbf{g}_j^{(L)}$ は $(\mathbf{g}_i^{(L)} + \mathbf{g}_j^{(L)}) \oplus (\mathbf{g}_i^{(L)} \odot \mathbf{g}_j^{(L)})$ と処理された後、ニューラルネットワークの入力として扱われる。ここで、$\oplus$ は 2 つのベクトルの連結、$\odot$ はアダマール積を表す。ニューラルネットワークの構造は $(128, 64, 2)$ であり、全ての非線形活性化関数は ReLU を用いる。

5.3 比較手法

比較手法として、内部グラフに対する畳み込みのみを行う手法 (internal)、Yang らによるもの [Yang 15] の踏襲で外部グラフにおけるノードの特徴ベクトルを外部グラフにおけるリンク関係から学習する外部グラフへの埋め込み (external)、化合物の部分構造に対してインデックスを割り当て特徴表現を獲得する Morgan fingerprint [Morgan 65] (morgan) を用いる。今回の実験では Morgan fingerprint は 2048 次元、それ以外は 64 次元の特徴ベクトルを用いた。これは Morgan fingerprint を小さな次元で適用すると衝突により、性能が大きく低下するためである。

5.4 実験結果

訓練データの大きさを変えながら評価したものを図 2a、図 2b、図 2c に示す。横軸が訓練データ（リンク）のサイズ、縦軸は AUC である。なお、訓練データは正例、負例の割合をそれぞれ 50%とした。

5.4.1 標的共有ネットワーク

標的共有ネットワークでは、外部グラフの埋め込み手法の予測性能は低い。これは、ネットワークが疎であるために単純にノードの組み合わせのみを考えるだけでは予測が難しいためだと考えられる。一方で、Hashed Morgan fingerprint は訓練データが小さい場合は提案手法を上回る性能を発揮しており、fingerprint に基づく手法の頑健性を示している。しかし、訓練データのサイズが大きくなると、提案手法が Hashed Morgan fingerprint を上回る性能を示している。これは外部グラフに対する畳み込みによって提案手法の予測性能が大きく向上したためであると考えられる。また内部グラフに対する畳み込みのみの手法を提案手法が上回っており、二重畳み込みが有効に働いていることが分かる。

5.4.2 薬剤相互作用

薬剤相互作用では、ネットワークが密であることから外部グラフのみで特徴学習を行う外部グラフの埋め込み手法が優れた性能を示しているが、訓練データサイズが小さくなるにつれて、その性能を大きく低下させている。一方で、訓練データサイズが大きい時は提案手法と外部グラフの埋め込み手法はほぼ同様の性能を示しており、また訓練データサイズが小さい時であっても、内部グラフ及び外部グラフ両方の情報を取り入れているため大幅な性能の低下は免れており、頑健性を示している。

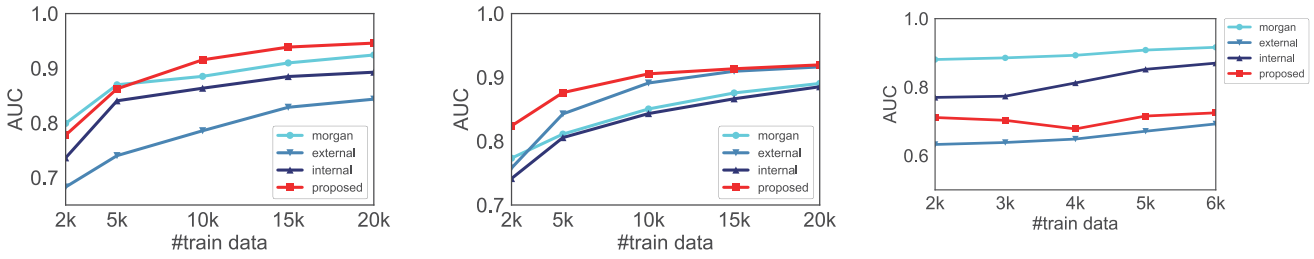
5.4.3 代謝化合物

代謝化合物ネットワークは本研究の中で最も疎なデータセットであり、提案手法が効果を発揮していないだけでなく逆に悪化させてしまう結果を示している。これは、外部グラフが非常に疎であるために、情報を取り入れることが困難になり有効な特徴ベクトルを抽出できていないためと考えられる。同様に外部グラフのスパース性によって外部グラフの埋め込み手法も大きく性能を低下させていることが分かる。一方でベースライン手法である Hashed Morgan fingerprint と内部グラフに対する畳み込みは外部グラフのスパース性の影響を受けない特徴ベクトルを用いているためか、優れた性能を示している。

6. おわりに

本研究では、これまであまり研究対象としては扱われてこなかった GoG に対して内部グラフ及び外部グラフを共に考慮し、一貫で特徴表現を学習する二重畳み込みニューラルネットワークを提案し、特徴表現学習の問題について取り組んだ。3 つの実データに対するリンク予測実験で提案手法の有用性を示した。提案手法である二重畳み込みによるアプローチは、標的共有ネットワーク及び薬剤相互作用ネットワークという一定のリンク密度があるネットワークで高い予測性能を示し、特徴ベクトルの次元がベースライン手法の 1 つである Hashed Morgan fingerprint に比べて低次元であっても同等以上の性能を示した。また提案手法が、外部グラフが非常に疎であるとき、情報を十分に取り入れることが困難になるため性能が大きく低下するということを確認した。

将来の展望として、ノード分類やノードクラスタリング等のタスクについても応用する所存である。本実験では、外部グラフに対する畳み込みの際用いる訓練データから得られる完全



(a) 標的共有ネットワーク　(b) 薬剤相互作用　(c) 代謝化合物

図 2: 各データセットにおける訓練データの変化に対する AUC の推移

な隣接行列について、経験上性能が向上したため 2 割の隣接関係を無視した隣接行列を用いている。隣接行列の運用法による影響については今後詳しく調べていく必要があると考える。上記も含めたハイパーパラメータの調査や、比較手法として新たな比較手法の採用を検討する。

参考文献

[Duvenaud 15] Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R. P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints, in *Advances in Neural Information Processing Systems (NIPS)* (2015)

[Gärtner 03] Gärtner, T., Flach, P., and Wrobel, S.: On graph kernels: hardness results and efficient alternatives, in *Learning Theory and Kernel Machines*, pp. 129–143 (2003)

[Getoor 05] Getoor, L. and Diehl, C. P.: Link mining: a survey, *SIGKDD Explorations*, Vol. 7, No. 2, pp. 3–12 (2005)

[Inokuchi 00] Inokuchi, A., Washio, T., and Motoda, H.: An apriori-based algorithm for mining frequent substructures from graph data, in *Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)* (2000)

[Kashima 03] Kashima, H., Tsuda, K., and Inokuchi, A.: Marginalized kernels between labeled graphs, in *Proceedings of the 20th International Conference on Machine Learning (ICML)* (2003)

[Kotera 14] Kotera, M., Tabei, Y., Yamanishi, Y., Muto, A., Moriya, Y., Tokimatsu, T., and Goto, S.: Metabolome-scale prediction of intermediate compounds in multistep metabolic pathways with a recursive supervised approach, *Bioinformatics*, Vol. 30, No. 12, pp. i165–i174 (2014)

[Morgan 65] Morgan, H. L.: The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service, *Journal of Chemical Documentation*, Vol. 5, No. 2, pp. 107–113 (1965)

[Niepert 16] Niepert, M., Ahmed, M., and Kutzkov, K.: Learning convolutional neural networks for graphs, in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 2014–2023 (2016)

[Scarselli 09] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G.: The graph neural network model, *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp. 61–80 (2009)

[Shervashidze 11] Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M.: Weisfeiler-Lehman graph kernels, *Journal of Machine Learning Research*, Vol. 12, No. Sep, pp. 2539–2561 (2011)

[Socher 13] Socher, R., Chen, D., Manning, C. D., and Ng, A.: Reasoning with neural tensor networks for knowledge base completion, in *Advances in Neural Information Processing Systems (NIPS)* (2013)

[Takigawa 11] Takigawa, I., Tsuda, K., and Mamitsuka, H.: Mining significant substructure pairs for interpreting polypharmacology in drug-target network, *PLoS ONE*, Vol. 6, No. 2, p. e16999 (2011)

[Yan 02] Yan, X. and Han, J.: gSpan: Graph-based substructure pattern mining, in *Proceedings of the Second IEEE International Conference on Data Mining (ICDM)* (2002)

[Yang 15] Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L.: Embedding entities and relations for learning and inference in knowledge bases, in *Proceedings of the Third International Conference on Learning Representations (ICLR)* (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Stacked Recurrent Neural Networkによる桜島噴火予測

Predicting Eruptions of Sakurajima by Stacked Recurrent Neural Network

村田 剛志[*1] ヒープ レ[*1] 井口 正人[*2]
Tsuyoshi Murata Hiep V. Le Masato Iguchi

[*1]東京工業大学情報理工学院情報工学系
Department of Computer Science, School of Computing, Tokyo Institute of Technology

[*2]京都大学防災研究所火山活動研究センター
Sakurajima Volcano Research Center, Disaster Prevention Research Institute, Kyoto University

Volcanic eruptions sometimes cause severe damage to many people. This paper explains our attempts for predicting volcanic eruptions from time series sensor data obtained from volcanic monitoring systems (strainmeters) located in Sakurajima. Given the time series data of strainmeters for 100 minutes, our goal is to predict future status of the volcano which is either *explosive* or *not explosive* for the 60 minutes immediately after the 100 minutes. We use stacked recurrent neural network for this task, and our method achieves 66.1% F-score on average. We also propose a four-stage warning system that classifies time series sensor data into the following categories: "Non-eruption", "May-eruption", "Warning" and "Critial". The percentage of *explosive* cases in "Critial" category is 51.9%.

1. はじめに

日本は世界有数の火山国であり、世界の活火山の約 7%が存在する。火山の噴火は時として甚大な人的・物的被害を及ぼすものであり、2014 年の御嶽山や 2018 年の草津白根山での火山災害は記憶に新しい。火山活動を正確に把握して必要な対応を取ることは、専門家のみならず周辺地域の多くの人々にとっての重大な関心事である。

火山から観測される時系列データは、火山活動の分析において非常に重要である [Sparks 03]。火山から観測される時系列データとしては、地形の歪みや、地震の振動の大きさやエネルギーなどが挙げられる。機械学習を用いた火山活動の分析の例として、Malfante らの研究がある [Malfante 18]。これはペルーのウビナス火山の 6 年分のデータを対象とし、バリエーションの大きい火山性地震を SVM を用いてタイプ毎に分類するものである。著者らは、日本における有数の活火山である鹿児島県の桜島に設置された装置から観測される時系列データに対して、畳み込みニューラルネットワーク (CNN) を用いて分析した結果について報告した [村田 17][日経 18]。これも観測した 100 分間で爆発的噴火が起こるか否かを「分類」するタスクについてであった。

本稿では、与えられた 100 分間の時系列データから、その直後の 60 分間に爆発的噴火が起こるか否かを「予測」するタスクについて述べる。本稿における火山噴火予測の概要を図 1 に示す。時系列データにおける時間変化を検出するために Stacked 2-Layer LSTM を用いて実験を行った結果、噴火・非噴火の 2 クラス平均の F 値による精度で 66.1%であった。また、与えられた時系列データを "Non-eruption", "May-eruption", "Warning", "Critial" の 4 つに分類する警告システムを構築したところ、"Critial" に分類された時系列データで爆発的噴火が起こったものの割合は 51.9%であった。

連絡先: 東京工業大学情報理工学院情報工学系 村田剛志
〒 152-8552 東京都目黒区大岡山 2-12-1 W8-59
E-mail: murata@c.titech.ac.jp

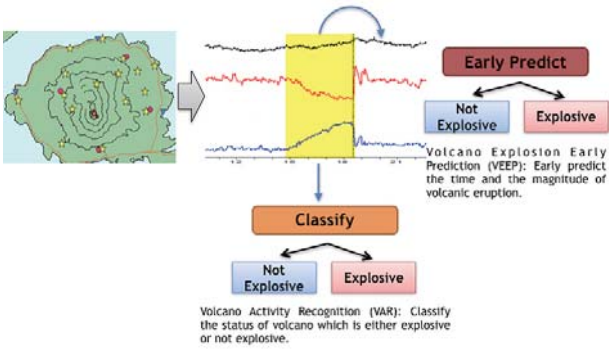


図 1: 火山活動分類 (下) と火山噴火予測 (右)

2. データセット

本稿では、時系列データとして、国土交通省の九州地方整備局の火山活動データを用いる。これは日本で最も活発な火山の一つである桜島の噴火に関するデータである。桜島は時期によっては毎週のように噴火しており、対象データとして適している。

火山から観測されるデータとしては、地震計による最大振幅や振幅二乗和 (エネルギー)、傾斜計による火口の上昇や降下、火山ガスの量や成分などもあるが、本稿では伸縮計によって観測される時系列データに注目する。伸縮計は火口方向・直交方向の地面の伸び縮みを観測するものである (図 2)。

観測データは数値として毎分記録される。データは全体で 2009 年から 2016 年までの 8 年分あり、日本の火山の時系列データとしては最大級である。ただし、大規模な台風の接近や他地域での地震発生などによるノイズや、観測装置の点検期間中におけるデータ欠損などが起こり得る。

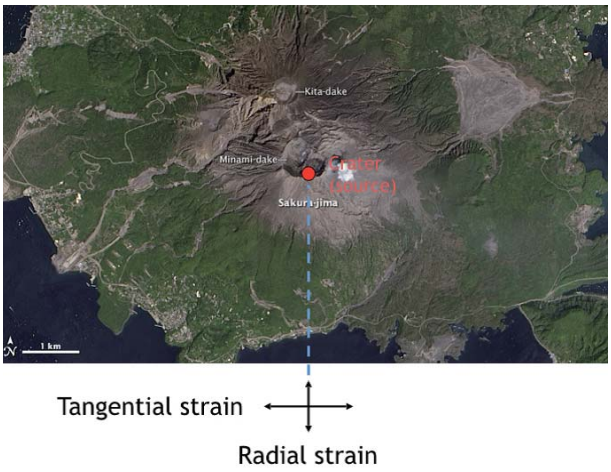


図 2: 伸縮計の火口方向・直交方向

3. モデルアーキテクチャ

本稿においては、火山噴火予測問題を時系列データのパターン認識問題としてとらえる。100 分間の時系列データを入力とし、その直後の 60 分間において爆発的噴火が起こるか否かを予測する。空振による音が 10 パスカル以上のものを爆発的噴火と定義する。

3.1 Recurrent Neural Network

Recurrent Neural Network (RNN) はデータ内の時間的な依存関係を、層内ユニット間の有向の結合によって学習するニューラルネットワークである。時刻 t の隠れ状態 h は以下のように計算される。

$$h_t = \tanh(Wx_t + Uh_{t-1})$$

ここで tanh は双曲線正接関数、W と U はそれぞれ入力に対する重み行列と recurrent weight matrix である。予測は最終隠れ状態 h_t を用いて行う。

$$y_t = \sigma(W_p h_t)$$

3.2 Long Short-Term Memory

一般に RNN では、誤差逆伝播法における誤差信号が隠れ層を経るごとに小さくなってしまう勾配消失問題があることが指摘されている。Long short-term memory (LSTM) [Hochreiter 97] はこの問題を解決するための改良法の一つである。LSTM は状態変化においてゲートを導入することによって勾配消失問題に対処している。各時刻において、LSTM は隠れ状態 h とメモリーセル状態 c を保持して、状態更新と出力を以下のようにコントロールする。

$$\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1}) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1}) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1}) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1}) \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}$$

ここで f_t、i_t、o_t はそれぞれ忘却ゲート、入力ゲート、出力ゲートである。σ はシグモイド関数、$\odot$ は要素毎の積を表している。W_f, W_i, W_o, W_c と U_f, U_i, U_o, U_c は、各ゲートに付随した入力重み行列と recurrent weight matrix であり、学習によってパラメータ値を得る。メモリーセル状態 c は、忘却ゲート f_t、入力ゲート i_t および 1 ステップ前のセル状態によってコントロールされる。このセル状態は全ての時刻の状態を集めることができ、それによって時系列データの長期的な依存関係を把握することができる。LSTM の出力はメモリーセル状態 c と出力ゲート o_t から計算される。

LSTM の隠れ状態 h を別の LSTM への入力とすることによって、より深い構造である Stacked LSTM を作ることができる。1 つ前の層 $l-1$ から作られる層 l の隠れ状態 h は以下で表される。

$$h_t^l = lstm(h_t^{l-1}, h_{t-1}^l)$$

ここで $lstm$ は LSTM のゲートの計算、h_t^{l-1} は層 $l-1$ の隠れ状態で次の LSTM への入力となるもの、h_{t-1}^l は時刻 $t-1$ における層 l の隠れ状態である。

3.3 提案モデル

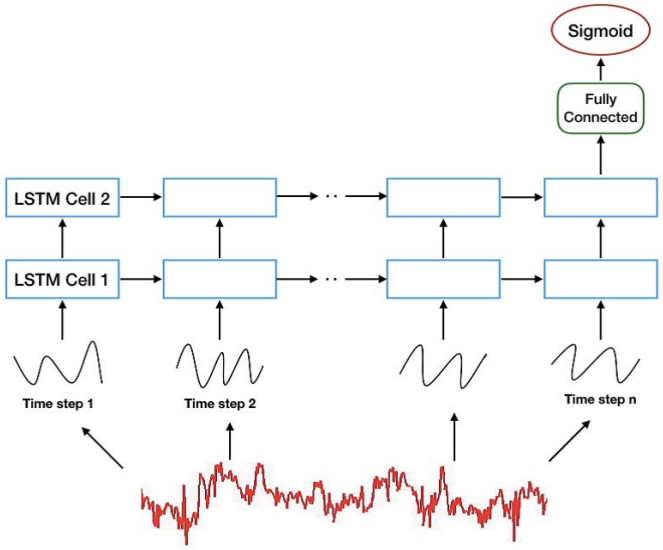


図 3: Stacked 2-Layer LSTM

火山噴火予測のための Stacked 2-Layer LSTM を提案する。提案モデルの全体像を図 3 に示す。まず、入力の時系列データを複数の時間ステップに分割する。各時間ステップは、短時間での動きを把握するためのもので、例えば入力が 100 分の時系列データとしたとき、それを 5 分の時間ステップ 20 個に分割する。この時間ステップの数はハイパーパラメータである。次に、各時間ステップを LSTM セルの入力として与え、動的な振舞いを学習する。LSTM セルは (横方向に) 結合しており、入力時系列データの開始から終了までの時間依存の特徴量を学習する。これを 2 段にした Stacked LSTM にすることで、より複雑な特徴を学習することが可能になる。2 段目の LSTM の最終隠れ状態から全結合層とシグモイド関数を経て最終的な出力が得られ、その値をもとに予測を行う。

3.4 モデルの訓練

モデルの訓練においては、重みつき二値交差エントロピー損失関数を最小化している。正の重みを増やすことによって、火山噴火予測において避けるべき False Negative(見逃し) に

対処している。

$$L = \sum_{i=1}^{batch\_size} y_i \log(y'_i) \times weight + (1 - y_i) \log(1 - y'_i)$$

ここで y は目標値、y' は予測値である。1 以上の値のパラメータ $weight$ を損失関数に導入し、目標値が 1 であるのに予測値が 0 に近い値である場合により多くの罰則を与えるようにした。このような損失関数を最適化することによって、False Negative の事例数を減らすことができると考えられる。本稿の実験では $weight$ の値は 4 とした。

LSTM セルのサイズは 128 に固定して実験を行っている。モデルの訓練は Adam[Kingma 14] を用いて行い、初期学習率は $1e-3$ とした。ハイパーパラメータのチューニングにはランダムサーチを用いた [Bergstra 12]。

4. 実験結果

本稿の実験では伸縮計の直交方向の時系列データを用いた。もちろん、他の時系列データを用いてさらに精度を向上させることも可能であると考えられる。以下で述べる実験結果は、全て 2015 年のデータを用いたものである。

4.1 火山噴火予測

上述の Stacked 2-Layer LSTM を用い、入力として与えた 100 分間の時系列データに対する出力が閾値以上であるものを噴火、そうでないものを非噴火と予測する。閾値を 0.5 として実験した結果の混同行列を表 1 に示す。また噴火・非噴火のそれぞれの場合の F 値は以下の通りである。

表 1: Stacked 2-Layer LSTM での火山噴火予測の混同行列 (閾値 0.5)

	非噴火 (予測)	噴火 (予測)
非噴火 (正解)	42,708	5,033
噴火 (正解)	1,667	2,184

- 非噴火の場合の F 値 $F_{not\_explosive} = 92.7\%$
- 噴火の場合の F 値 $F_{explosive} = 39.5\%$
- 両者の平均の F 値 $F_{avg} = 66.1\%$

対象とする時系列データにおいては、噴火の事例数に比べて、非噴火の事例数が圧倒的に多い。単純に精度を求めると、全ての入力に対して無条件に非噴火と出力したとしても 90%以上の高精度となってしまうことから、上記のように噴火・非噴火の両方の場合の F 値を求めることで評価を行っている。火山噴火予測においては噴火の場合の F 値が重要である。本稿で述べた Stacked 2-Layer LSTM は、噴火の場合の F 値 ($F_{explosive}$) が 40% 近くに達している。

著者らは、多層 CNN(畳み込みニューラルネットワーク) を用いて同様の火山噴火予測の実験を行っている (図 4)[村田 18]。その実験結果を表 2 に示す。また噴火・非噴火のそれぞれの場合の F 値は以下の通りである。

- 非噴火の場合の F 値 $F_{not\_explosive} = 93.4\%$
- 噴火の場合の F 値 $F_{explosive} = 37.6\%$

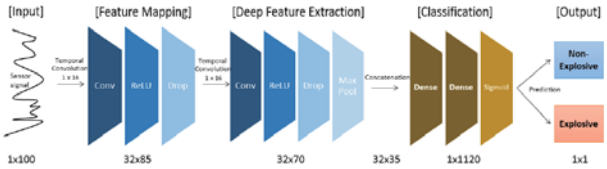


図 4: 多層 CNN(畳み込みニューラルネットワーク) での火山噴火予測

表 2: 多層 CNN での火山噴火予測の混同行列 (閾値 0.5)

	非噴火 (予測)	噴火 (予測)
非噴火 (正解)	43,598	4,143
噴火 (正解)	2,000	1,851

- 両者の平均の F 値 $F_{avg} = 65.5\%$

表 1 と表 2 を比較すると、多層 CNN での火山噴火予測と比べて、Stacked 2-Layer LSTM での火山噴火予測の F_{avg} の方が高精度であり、また False Negative の事例数も少ないことがわかる。

また火山噴火予測において、60 分間のどのタイミングでの火山噴火予測が容易か (困難か) を調べた。60 分間の時系列データを 10 分ずつ 6 つの間隔に分けて、6 つの間隔のどこで噴火が起こったかを、予測に成功した例と失敗した例のそれぞれについて調べた。

- 予測成功: [521, 458, 383, 329, 267, 226]
- 予測失敗: [202, 238, 277, 301, 321, 328]

上記の数字は、10 分ずつ 6 つの間隔のそれぞれでの予測成功と予測失敗の事例数を表している。この結果から、観測した 100 分の直後の 10 分間の火山噴火予測は比較的高精度であるのに対し、観測した 100 分から時間が経つに従って精度が低下していることがわかる。

また、本稿での火山噴火予測に成功した例と失敗した例それぞれの噴火の強度についても調べた。ここでは伸縮計での収縮量を噴火の強度とみなしている。

- 予測成功: 中央値:18.2, 平均値 20.0, 標準偏差 11.4
- 予測失敗: 中央値:9.9, 平均値 11.8, 標準偏差 7.7

上記の数字は、予測で成功した例と失敗した例それぞれの噴火の強度の中央値、平均値、標準偏差を表している。この結果から、本稿の手法は、比較的強度の大きい噴火の予測に成功しており、予測に失敗した噴火の強度は比較的小さいことがわかる。

4.2 4 段階警告システム

前節では噴火か非噴火かの 2 クラス分類を行ったが、Stacked 2-Layer LSTM の出力値の大きさと、爆発的噴火の可能性との間には相関があると考えられる。本節では 3 つの閾値によって時系列データを 4 つのカテゴリに分類し、それぞれのカテゴリにおける爆発的噴火の割合を調べた。閾値としては便宜的に 0.2、0.5、0.8 とした。

- Non-eruption: Stacked 2-Layer LSTM の出力値が 0 以上 0.2 未満

- May-eruption: Stacked 2-Layer LSTM の出力値が 0.2 以上 0.5 未満
- Warning: Stacked 2-Layer LSTM の出力値が 0.5 以上 0.8 未満
- Critical: Stacked 2-Layer LSTM の出力値が 0.8 以上

表 3: 4 段階警告システムの実験結果 (2015 年データ)

	Non-eruption	May-eruption	Warning	Critical
合計	27,967	16,408	6,623	594
非噴火	27,509	15,199	4,747	286
噴火	458	1,209	1,876	308
精度	98.4%	92.6%	28.3%	51.9%

実験結果を表 3 に示す。表 3 における精度は、Non-eruption と May-eruption のカテゴリにおいては非噴火の割合、Warning と Critical のカテゴリにおいては噴火の割合を表している。当然ながらこれらの値は先の 3 つの閾値によって変わってくる。注目すべきなのは、Critical のカテゴリに分類された時系列データにおいては噴火するものの割合が 51.9%になっていることである。このカテゴリに分類される時系列データの事例数はやや少ないが、これらについては爆発的噴火の可能性が十分高く、このような 4 段階警告システムによる火山噴火予測は検討に値すると考えられる。

Kamo ら [Kamo 89] は、桜島の火口の上昇・降下を計測する傾斜計の 5 分間のデータをもとに、あらかじめ定めたルールに基づいて 5 段階に自動分類する手法を提案し実験を行っている。この手法において、最も危険度が高いカテゴリにおける予測精度は 45.76% であった。30 年近く前の研究であり、また用いている時系列データの観測装置も伸縮計と傾斜計とで異なっているが、本稿での 4 段階警告システムによる火山噴火予測はこれよりも高精度である。

5. おわりに

本稿では、桜島周辺で観測される時系列データをもとに、Stacked 2-Layer LSTM を用いて爆発的噴火を予測する手法および実験結果について述べた。実験の結果、多層 CNN を用いたものや関連研究 [Kamo 89] よりも高精度であることが示された。

本稿での火山噴火予測の具体的な応用例としては、以下のものが考えられる。

工事従事者の安全確保 [Kamo 89] をベースにした噴火予測手法は、現在でも桜島で使われている。桜島では火口から 2km までが 1955 年から立ち入り規制されているが、2km 付近での砂防工事があり、その工事従事者の携帯電話に火山噴火予測のメールが届くようになっている。[Kamo 89] では傾斜計のデータだけを使っているが、現在実装されているものは傾斜計と伸縮計の 2 成分 (火口方向・直交方向) から圧力源の体積変化を見積もり、体積変化を指標としている。工事従事者の安全確保のためには、高精度の火山噴火予測が必要である。

降灰予報 気象庁は、蔵王山、草津白根山、桜島、口永良部島、諏訪之瀬島、霧島山 (御鉢)、霧島山 (新燃岳)、霧島山 (えびの高原 (硫黄山) 周辺) の降灰予報 (定時) を行っている (`http://www.data.jma.go.jp/svd/vois/data/kouhai/jishin/ashfall.html`)。これは火山周辺の風向をもとに、噴火が発生したらという条件つきでの降灰予報になっているが、精度の高い火山噴火予測と組み合わせることによって、周辺住民に対してさらに有益な予報にすることが期待できる。

謝辞

火山データをご提供いただいた国土交通省九州地方整備局大隅河川国道事務所に感謝致します。

参考文献

[Bergstra 12] James B., Bengio Y. “Random Search for Hyper-Parameter optimization.” Journal of Machine Learning Research, Vol.13, pp.281-305 (2012).

[Hochreiter 97] Hochreiter, S., Schmidhuber, J., “Long Short-Term Memory”, Neural Computation, Vol.9, No.8, pp.1735-1780 (1997).

[Kamo 89] Kamo K., Ishihara K. “A Preliminary Experiment on Automated Judgement of the Stages of Eruptive Activity Using Tiltmeter Records at Sakurajima, Japan”, in Latter J. H. (eds.) Volcanic Hazards. IAVCEI Proceedings in Volcanology, Vol.1, pp.585-598, Springer (1989).

[Kingma 14] Kingma, D. P., Jimmy B. “Adam: A Method for Stochastic Optimization.”, The Third International Conference for Learning Representations (ICLR 2015), arXiv:1412.6980 (2014).

[Malfante 18] Malfante M., Mura M. D., Metaxian J. -P., Mars J. I., Macedo O., Inza A., “Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives”, IEEE Signal Processing Magazine, Vol.35, Issue.2, pp.20-30 (2018).

[Sparks 03] Sparks, R. S. J, “Forecasting Volcanic Eruptions”, Earth and Planetary Science Letters, Vol.210, No.1-2, pp.1-15 (2003).

[日経 18] 日本経済新聞, “東工大・京大、AI で火山活動分析 噴火予測めざす”, 2018 年 1 月 15 日 第 9 面 (2018).

[村田 17] 村田 剛志, ヒープ レ, 井口 正人, “深層学習による火山活動認識”, 第 112 回人工知能学会知識ベースシステム研究会, SIG-KBS-B508, pp.38-45 (2017).

[村田 18] 村田 剛志, “深層学習による桜島噴火予測”, 桜島研究課題研究集会, 2018 年 1 月 6 日 (2018). `http://www.svo.dpri.kyoto-u.ac.jp/svo/topic-news/meeting20180105/`

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層生成モデルによる非正則化異常度を用いた工業製品の異常検知

Anomaly Manufacturing Product Detection
using Unregularized Anomaly Score on Deep Generative Models

立花亮介 *1　松原崇 *1　上原邦昭 *1
Ryosuke Tachibana　Takashi Matsubara　Kuniaki Uehara

*1神戸大学 大学院 システム情報学研究科
Graduate School of System Informatics, Kobe University

One of the most common needs in manufacturing plants is rejecting products not coincident with the standards as anomalies. Manufacturing companies usually employ numerous inspectors for anomaly detection and it takes a high cost. Accurate and automatic anomaly detection reduces inspection cost and improves product reliability. In unsupervised anomaly detection, a probabilistic model detects test samples with lower likelihoods as anomalies. Recently, a probabilistic model called deep generative model (DGM) has been proposed for end-to-end modeling of natural images and already achieved a certain success. However, anomaly detection of machine components with complicated structures is still challenging because they produce a wide variety of the normal image patches with lower likelihoods. For overcoming this difficulty, we propose unregularized score for the DGM. As its name implies, the unregularized score is the anomaly score of the DGM without the regularization terms. The unregularized score is robust to the inherent complexity of a sample and has a smaller risk of rejecting a sample appearing less frequently but being coincident with the standards. Experimental results of anomaly detection on the real manufacturing datasets show better performance of the unregularized score compared to existing approaches.

1. はじめに

工業製品の製造現場において最も重要であることの一つは，製造された工業製品が期待される仕様を満たしていないときに，その製品を不良品であるとみなして取り除くことである．現状では異常な製品の発見や除去は人手で行うことが一般的であり，企業は異常製品の除去に対して高い人的コストを支払っている．このような状況において，正確かつ自動的な異常検知によって製品の信頼性を高めると同時に，製品の点検にかかるコストを削減することが求められている．

教師なし学習における異常検知では，確率モデルによってデータの負の対数尤度を求め，その値を異常度としてこれが高いものを異常とみなす．従来では，データから主成分分析（Principal Components Analysis; PCA）などで特徴抽出を行った後，得られた特徴を用いて混合ガウスモデル（Gaussian Mixture Model; GMM）などの確率モデルを学習する [Saligrama 12]．最近では，画像から直接尤度を求めることが出来る深層生成モデル（Deep Generative Model; DGM）と呼ばれる確率モデルが提案されており，異常検知において一定の成果を達成している [Ribeiro 17]．しかしながら，工業製品はその構造が複雑かつ多様であるため製品を撮影した際に画像内における各部分の出現頻度が異なり，尤度が必ずしも異常度に対応せず，正しく評価できないという問題がある．

このような問題の解決のため，本論文では深層生成モデルにおいて，非正則化異常度を用いた異常検知手法を提案する．非正則化異常度という名前が示すように，これは深層生成モデルにおける異常度，つまり負の対数数度から正則化項を取り除いたものである．非正則化異常度はデータが潜在的に含有する複雑さに対して堅牢であり，画像内における各部分の出現頻度に依存することなく評価を行える．提案手法の有効性を検証するために，工業製品の画像データに対して本手法を適用させ，異常検知性能においてその有効性を既存の手法と比較する．

連絡先: 松原崇，神戸大学大学院システム情報学研究科，
matsubara@phoenix.kobe-u.ac.jp

2. 非正則化異常度

確率的モデルである VAE や GMM における従来の異常度である負の対数尤度は，いくつかの意味をもつ項に分割できる．提案手法である非正則化異常度は，その項の一つを取り出し新たな異常度として再定義したものである．ここでは，VAE 及び GMM の異常検知における従来の異常度から，提案手法である非正則化異常度の導出を行う．

VAE はパラメータ θ で表現される生成モデル $p_\theta(x)$ に対してモデルエビデンスの下限（ELBO）を最大化することで学習を行うモデルである．VAE は入力 x を潜在変数 z に変換する符号化器及び潜在変数 z から入力の再構成を行う復号化器から構成される．変分事後分布 $q_\phi(z|x)$ 及び条件付き確率 $p_\theta(x|z)$ は，共に分散共分散行列が対角行列となっている多変量正規分布としてモデル化される．符号化器は入力 x を受け取った後，変分事後分布 $q_\phi(z|x)$ における平均ベクトル μ_z と標準偏差ベクトル σ_z の二つの要素の出力を行う．復号化器は，条件付き確率 $p_\theta(x|z)$ における平均ベクトル μ_x と標準偏差ベクトル σ_x の二つの要素の出力を行う．変分事後分布 $q_\phi(z|x)$ のモンテカルロサンプリングを行う代わりに潜在変数 z の最大事後確率推定 μ_z が異常検知において使用される．よって，負の ELBO$\mathcal{L}_{VAE}(x)$ は以下の式 (1) によって表される．

$$\begin{aligned}\mathcal{L}_{VAE}(x) &= D_{KL}(q_\phi(z|x)||p(z)) - \log p_\theta(x|\mu_z) \\ &= D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x).\end{aligned} \tag{1}$$

ただし，$D_{VAE}(x)$，$A_{VAE}(x)$，$M_{VAE}(x)$ はそれぞれ

$$\begin{aligned}D_{VAE}(x) &= \sum_{j=1}^{N_z} \frac{1}{2}(-\log \sigma_{z_j}^2 - 1 + \sigma_{z_j}^2 + \mu_{z_j}^2), \\ A_{VAE}(x) &= \sum_{i=1}^{N_x} \frac{1}{2} \log 2\pi\sigma_{x_i}^2 \Big|_{z=\mu_z}, \\ M_{VAE}(x) &= \sum_{i=1}^{N_x} \frac{1}{2} \frac{(\mu_{x_i} - x_i)^2}{\sigma_{x_i}^2} \Big|_{z=\mu_z},\end{aligned} \tag{2}$$

である．また，i と j はそれぞれデータ x と潜在変数 z の要素の番号を示す．ニューラルネットワークの見地からみると，$A_{VAE}(x) + M_{VAE}(x)$ は再構成誤差を表しており，また $D_{VAE}(x)$ は正則化項を表している．$A_{VAE}(x)$ は正規化定数の対数を取ったものに相当し，これは正規分布 $p_\theta(x|z)$ の確率密度関数の積分を 1 にするものである．$M_{VAE}(x)$ は式の形としてはマハラノビス距離，あるいはユークリッド距離を正規化したものに近いことがわかる．

ここで，導出された $\mathcal{L}_{VAE,M}(x) = M_{VAE}(x)$ を，従来の負の ELBO$\mathcal{L}_{VAE}(x) = D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x)$ の代わりとして，VAE による異常検知における非正則化異常度とする．非正則化異常度という名前は，このように従来の異常度から正規化定数の対数 $A_{VAE}(x)$ 及び正則化項 $D_{VAE}(x)$ を取り除いていることに由来する．

GMM において，データ x が属する隠れクラス z は最大事後確率推定によって決定することができる．このときのデータ x の負の対数尤度は以下の式 (3) で表される．

$$-\log w_k p(x|z=k) = D_{GMM}(x) + A_{GMM}(x) + M_{GMM}(x). \tag{3}$$

ただし，$D_{GMM}(x)$，$A_{GMM}(x)$，$M_{GMM}(x)$ はそれぞれ

$$\begin{aligned} D_{GMM}(x) &= -\log w_k, \\ A_{GMM}(x) &= \frac{1}{2}\log(2\pi)^{N_x}|\Sigma_k|, \\ M_{GMM}(x) &= \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k), \end{aligned} \tag{4}$$

である．但し，$k = \arg\max_z p(z|x)$ である．そして，導出された $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ を，従来の負の対数尤度 $-\log\sum_{k=1}^{N_z} w_k p(x|z=k)$ の代わりとして，GMM による異常検知おける非正則化異常度とする．ここでもし GMM における隠れクラスの数 N_z を 1 とすると，非正則化異常度 $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ は明らかにマハラノビス距離 $(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)$ を 2 で割ったものと等しくなる．

3. 非正則化異常度の解釈

GMM における $D_{GMM}(x)$, $A_{GMM}(x)$, and $M_{GMM}(x)$ と，VAE における $D_{VAE}(x)$, $A_{VAE}(x)$, and $M_{VAE}(x)$ は，それぞれの式の意味において対応していると考えることができる．

工業製品は平らな表面，曲がった部分，ネジの穴など，様々な種類の要素から構成されている．このような場合，図 1 に示すように違う種類の要素グループに属する正常なデータ同士の違いの大きさが，同じ種類の要素のグループに属する正常なデータと異常なデータの違いの大きさよりも大きくなってしまうと考えられる．GMM によって，隠れクラスに各要素グループからの正常なデータが属すことになる．そして，異常なデータは特定の異常クラスとして分離されてそれぞれ属するのではなく，多くのデータが属しているのと同じ隠れクラスに属すことになる．このような仮定のもとでは，混合係数 w_k は対応しているグループ z にデータが属す頻度を表していることになり，また分散共分散行列 Σ_k は対応しているグループ z からのデータの分布を表していることになる．よって，混合係数の負の対数 $D_{GMM}(x)$ と正規化定数の対数 $A_{GMM}(x)$ は，異常を上手く捉えることができないと考えられる．一方で，マハラノビス距離を 2 で割ったものに等しい $M_{GMM}(x)$ については，分散共分散行列 Σ_k によって正規化が行われたグループ z の分布の中心からの距離を表現しているため，グループ z 内部における異常度として機能することが期待できる．

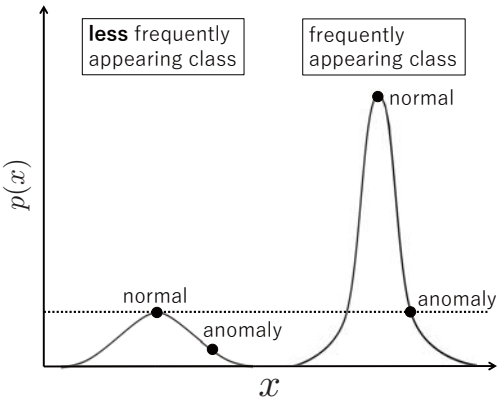


図 1: 工業製品画像における尤度の直感的な図解

VAE を PCA のように教師なし特徴抽出手法として利用すると，潜在変数 z は抽出された特徴として取り扱うことができ，正則化項 $D_{VAE}(x)$ は異常度として機能することが期待できる．しかしながら，VAE は一般的に $N_z < N_x$ であり入力次元よりも出力次元の方が次元数が低い．これは，高次元のデータ空間 $\mathcal{X}$ におけるデータ x を，低次元の潜在空間 $\mathcal{Z}$ へと写像しているということである．入力したデータ x を再構成するためには，VAE は与えられたデータにおいて最も顕著に現れる特徴を学習し，逆に頻度の少ない特徴は無視しなければならない．つまり，異常データ x_a が微小な異常を含んでいたとしても，それが他の部分において同じ種類の要素から取った正常データ x と似通っている場合は，潜在空間 $\mathcal{Z}$ における異常データ x_a の写像 z は正常データ x の写像の近くに分布することになる．このことから，正則化項 $D_{VAE}(x)$ は，混合係数の負の対数 $D_{GMM}(x)$ と同じように異常を検知することに関して有効であるとは言えない．

VAE がデータ x を正確に再構成することが出来る，つまり平均ベクトル μ_x を x に限りなく近づけることが出来るのであれば，標準偏差ベクトル σ_x はほぼ 0 に近い値となるため，これは正規化定数の対数 $A_{VAE}(x)$ を最小化することになる．しかしながら，VAE は最適な平均ベクトル μ_x を推定することが出来ないため，正規化距離 $M_{VAE}(x)$ の値が発散することを防ぐために，標準偏差ベクトル σ_x の値は大きくなくてはならなくなる．VAE では，再構成の不確かさに関して標準偏差 σ_x が調整されることにより，$A_{VAE}(x)$ と $M_{VAE}(x)$ の二つの基準のバランスが取れるように学習が行われる．このことは，複雑な要素から取られた正常データ x に対して，正規化定数の対数 $A_{VAE}(x)$ が大きな値をとるということを意味する．言い換えれば，標準偏差 σ_{x_i} が潜在的にデータ x の不確かさや本質的な複雑さを表現していることになり，またデータ x が属するグループの不確かさや本質的な複雑さを表現していることにもなる．これは，GMM における分散共分散行列 Σ_k と類似する点である．つまり，正規化定数の対数 $A_{VAE}(x)$ は異常検知では有効であるとは言えない．

ここでの結論としては，正規化距離である $M_{VAE}(x)$ のみが直接的に再構成誤差と関係しているということである．この距離 $M_{VAE}(x)$ は標準偏差 σ_{x_i} によって正規化されている．つまり，正規化距離 $M_{VAE}(x)$ はデータ x の不確かさや本質的な複雑さに対して堅牢であり，これは異常検知を行うにあたって有用な値である．以降の章では，VAE による異常検知において，正規化距離 $M_{VAE}(x)$ を非正則化異常度として実験を行い，異常検知における手法の有効性を検証する．

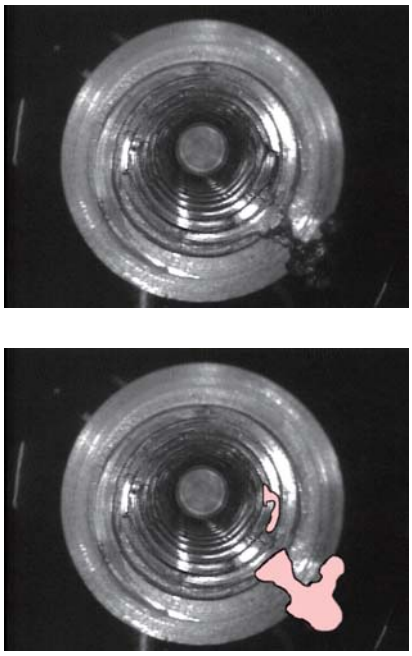


図 2: 異常データと異常箇所の ground truth.

4. 評価実験と考察

4.1 データセット

ネジ穴データセット（The screw dataset）は，ネジの穴を撮影したモノクロ画像のデータセットであり，画像サイズは 640 × 480 である．各画像の中央部にはネジ穴が写っており，その周囲は製品の平らで黒い表面に取り囲まれている．さらに外側にはカメラの囲いが写っている．ネジ穴データにおける異常及び ground truth を図 2 に示す．このような異常による品質の悪化は，製品の強度を低下させ，また破片は故障の原因となる．画像から読み取れるように，このデータセットにおいては，ネジ穴，平らな表面，ネジ穴の底面，などの複数の種類のクラスタがあることがわかる．ネジ穴データセットは 12,406 枚の訓練データと，995 枚のテストデータから構成されている．テストデータの 995 枚の内，888 枚は正常データ，107 枚は異常データである．テストデータの正常，異常のラベル付けに関しては，アイシン・エィ・ダブリュ株式会社の専門家によって分類が行われており，これは手法の性能評価に用いられた．訓練データにはラベルが付いておらず，また訓練データはパラメータの調整に用いられた．訓練データも正常データと異常データが混在しており，各データの正確な割合は確認されていないが，テストデータと同程度の割合であるということを前提として実験を行った．

4.2 モデル構成

提案手法を検証にするにあたって，VAE， 自己符号化器（Autoencoders; AE）[Hinton 06]，GMM の 3 つのモデルを使用した．ここでは，それぞれのモデル構造及び学習方法について述べる．非正則化異常度の検証を行うにあたって，VAE による学習を行った．VAE の実装には PyTorch v0.1.12 を用いた．VAE の学習を行う際は，画像の一部をランダムに 96×96 のサイズで切り取り，このパッチを訓練データとした．テストを行う際は，同サイズの画像を 16 ピクセルの間隔で切り取り，このパッチをテストデータとした．このパッチ全てに関して異常度を算出し，少なくとも 1 枚が閾値を超えている場合，テストデータは異常であると判断した．非正則化異常度 $\mathcal{L}_{VAE,M} = M_{VAE}(x)$ の評価のため，従来の異常度である $\mathcal{L}_{VAE} = D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x)$ を使用した異常検知も同様に行い，比較した．VAE の符号化器は 5 層の畳み込み層と 1 層の全結合層で構成した．n 番目の畳み込み層はカーネルサイズ 4×4，ストライド 2 であり，$N_c \times 2^{n-1}$ チャネルの特徴マップを出力し，出力にはさらに Batch Normalization [Ioffe 15] と活性化関数として ReLU [Nair 10] を適用した．全結合層からは平均ベクトル μ_z と標準偏差ベクトル σ_z を出力させた．この出力層は $2 \times N_z$ ユニットで構成されており，活性化関数として恒等関数を使用した．N_z ユニットは平均ベクトル μ_z として使用し，それ以外は $\log \sigma_z^2$ の獲得のために使用した．復号化器の構造は，符号化器の構造と対になっており，符号化器と逆の構造となるようにした．復号化器の出力は，平均ベクトル μ_x と標準偏差ベクトル σ_x とした．VAE のパラメータ学習にあたっては勾配降下法を用いるが，この際の最適化アルゴリズムとしては Adam を用いた．Adam のパラメータは $\alpha = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ とした．また荷重減衰を 0.9999 とした．畳込み層の開始チャネル数 N_c は $N_c \in \{16, 32, 64\}$ から探索し，潜在空間 $\mathcal{Z}$ における次元数 N_z は $\{2, 5, 10, 20, 50, 100\}$ から探索を行った．上記で述べた条件以外に関しては，VAE の元論文や，先行研究 [Ribeiro 17] に従った．

VAE の比較として，AE の評価も行った．AE は VAE の確定的な簡易モデルとしてみなすことができ，異常検知手法としても使用されている [Zhou 17]．AE は符号化器と復号化器から構成されており，符号化器による出力は潜在変数 z の点推定，復号化器による出力は入力データ x の再構成 $\tilde{x}$ である．AE の目的関数は平均二乗誤差であり，これを異常度 $\mathcal{L}_{AE} = \frac{1}{N_x} \sum_{i=1}^{N_x} (x_i - \tilde{x}_i)^2$ として使用する．AE はモンテカルロサンプリングを必要としない．また，正則化項 $D_{VAE}(x)$ を持たず，標準偏差 σ_{x_i} は定数となる．AE の構造は VAE の構造と同一にしており，データセットの利用方法や勾配降下法における条件も同じである．畳込み層の開始チャネル数 N_c は $N_c \in \{16, 32, 64\}$ から探索し，潜在空間 $\mathcal{Z}$ における次元数 N_z は $\{2, 5, 10, 20, 50, 100\}$ から探索を行い，1 番優れたモデルの選択を行った．

GMM については，scikit-learn v0.19.1 における基本的なモデルを使用した．訓練データとして，1 枚あたり画像サイズ 96×96 のパッチを 100 枚ランダムにサンプルし，PCA を用いて各パッチから N_h 次元の特徴を抽出した．隠れクラスが各々の分散共分散行列を持つ（full covariance）GMM に対して，EM アルゴリズムを使用して抽出した特徴の学習を行う．PCA のコンポーネント数 N_h は $N_h \in \{2, 5, 10, 20, 50, 100, 200, 500\}$ から探索し，GMM の隠れクラスの数 N_z は $N_z \in \{2, 5, 10, 20, 50, 100\}$ から探索した．GMM の異常度として，負の対数尤度 $\mathcal{L}_{GMM}$ と非正則化異常度 $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ を用いて検証を行った．

4.3 異常検知の精度

評価指標として Receiver Operating Characteristic（ROC）曲線の下部面積（ROC-AUC）を用いた．ROC-AUC は最大値が 1.0 であり，値が 1.0 に近いほど異常検知性能が高いことを意味する．ROC 曲線及び ROC-AUC の結果を図 3 に示す．図 3 から，VAE における非正則化異常度の ROC-AUC が他の手法よりも高くなっていることがわかる．

次に，異常度の可視化を行う．異常度とは正常であるほど値が低く，正常とは違う状態であるほど値が高くなる．つまり，異常度を可視化することによって，その度合から異常箇所を発見することができるということになる．これを検証するために，異常箇所が予めわかっているデータに対して VAE におけ

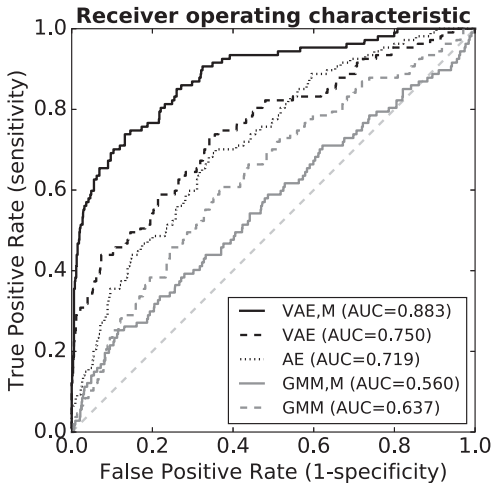


図 3: 各手法における ROC 曲線．括弧内は ROC-AUC の値を表す．

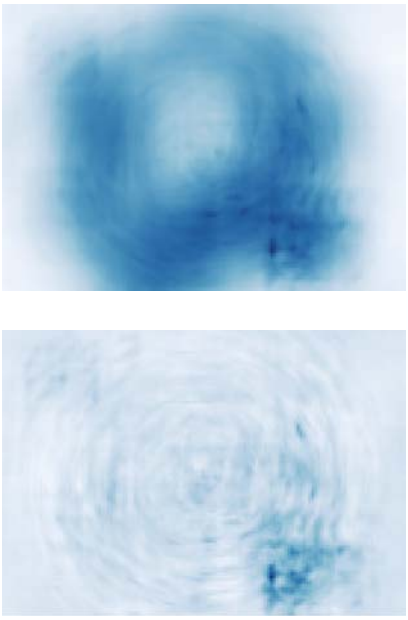


図 4: （上図）従来の異常度 $\mathcal{L}_{VAE}(x)$ のヒートマップ．(下図) 非正則化異常度 $\mathcal{L}_{VAE,M}$ のヒートマップ．

る非正則化異常度 $\mathcal{L}_{VAE,M}$ を求め，その可視化を行うことで，実際の異常箇所との比較を行う．使用するモデルのパラメータは，ROC-AUC の 1 番高いものを採用する．図 2 に示した異常データについて，VAE における従来の異常度 $\mathcal{L}_{VAE}(x)$ を用いたヒートマップ及び図非正則化異常度を用いたヒートマップを図 4 に示す．異常度ヒートマップは，画像を 96×96 のサイズに 4 ピクセルの間隔で切り取り，それぞれの異常度を並べて可視化したもので，色が明るいほど異常度が低いことを示している．異常のない部分の異常度について注目すると，従来の VAE の異常度 $\mathcal{L}_{VAE}(x)$ はネジ穴の周囲の平らな黒い表面及びネジ穴の底面において小さな値を取っているが，ネジ穴の溝では正常にも関わらず大きな値を取っている．これはネジ穴の溝から得られたパッチは平らな部分から得られたパッチと比べて，元の画像における相対的な位置の違いやネジ穴の溝の開始位置によって多様なものになり，また撮影方向や光の当たり具合によって変化するからであると考えられる．よって，ネジ穴の溝から得られたパッチの尤度は低くなる．つまり，異常度が異常それ自体よりも，パッチ自体の複雑さに依存してしまっているということになる．一方，正規化距離 $M_{VAE}(x)$ はパッチが元画像において正常な部分から取られたものであれば，場所によらず値が一定して低くなっている．誤差の絶対値 $|x - \mu_x|$ はネジ穴の溝において大きいが，正規化距離 $M_{VAE}(x)$ は標準偏差 σ_x によって正規化されており，この標準偏差 σ_x は場所による複雑さを表しているため，結果的に複雑さに対して堅牢となることで異常そのものを評価出来るようになると考えられる．このことから，正規化距離 $M_{VAE}(x)$ は異常箇所においてのみ大きな値を取っている．以上の理由により，非正則化異常度 $\mathcal{L}_{VAE,M} = M_{VAE}(x)$ は複雑な構造をもつ工業製品の異常検知において，性能の向上を達成する．

5. 結論

本論文では，VAE による異常検知において，非正則化異常度を使用することを提案した．非正則化異常度とは，その名前が指し示すように生成モデルの目的関数において正則化項を取り除いたものである．VAE において非正則化異常度を異常検知に使用することで，複雑な構造を持つ工業製品の画像に対して堅牢となり，従来の異常度よりも高い異常検知性能を実現することが出来た．また，異常度の可視化を行うことにより，異常箇所の特定を行えることを示した．

本研究は科研費 (16K12487) とアイシン・エィ・ダブリュ株式会社，株式会社エィ・ダブリュ・エンジニアリングの支援，総務省 SCOPE(受付番号 172107101) の委託を受けて行われた．

参考文献

[Hinton 06] Hinton, G. E. and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol. 313, No. 5786, pp. 504–507 (2006)

[Ioffe 15] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in *International Conference on Machine Learning (ICML)* (2015)

[Nair 10] Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, in *International Conference on Machine Learning (ICML)*, pp. 807–814 (2010)

[Ribeiro 17] Ribeiro, M., Lazzaretti, A. E., and Lopes, H. S.: A study of deep convolutional auto-encoders for anomaly detection in videos, *Pattern Recognition Letters* (2017)

[Saligrama 12] Saligrama, V. and Chen, Z.: Video anomaly detection based on local statistical aggregates, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2112–2119 (2012)

[Zhou 17] Zhou, C. and Paffenroth, R. C.: Anomaly Detection with Robust Deep Autoencoders, *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 665–674 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習抽出特徴量から生成した擬似特徴量を用いた不均衡データ多クラス画像分類

Pseudo-feature generation from feature map in deep learning for imbalanced data multi-class image classification

紺野 友彦　藤井 秀明　岩爪 道昭
Tomohiko Konno　Hideaki Fujii　Michiaki Iwazume

国立研究開発法人　情報通信研究機構　知能科学融合研究開発推進センター
AI Science R&D Promotion Center in National Institute of Information and Communications Technology

If some classes of the data have an only small number of samples, the accuracies of the classes become too low. It is well known as an imbalanced data problem. We often encounter imbalanced data in reality. In a sense, all the wild data are imbalanced.

In this paper, we make pseudo-feature from feature map in lower layers of deep neural networks, and we augment the data of minor classes to improve the imbalanced-data problem. We compare our proposed method with existing ones in imbalanced data multi-class image classification problems.

1. はじめに

データの中に少量しかないクラスがあると、そのクラスの識別率が極端に低下してしまう場合がある。そのようなデータは不均衡データと呼ばれており、現実のデータ分析では頻繁に遭遇する。不均衡データ問題が起きる理由は直感的には以下の通りである。例えば、がん検診で負例が 99%、正例が 1%であったとしよう。学習機は全データに対して負例であると言うだけで 99%の精度で予測できてしまい、そのような学習をしてしまう傾向がある。不均衡データ問題の解決方法としては例えば以下のような方法が知られている 。

1. アップサンプリング: マイナークラス（少量しかデータの無いクラス）のデータを重複してサンプリングして他のクラスのデータ量に合わせる方法
2. ダウンサンプリング:　マイナークラスのデータ量に合わせて他のクラスのサンプル数も少なく取る方法
3. コスト関数の重みの変更: クラスのサンプル数に合わせてコスト関数の重みを変更する方法
4. 人工的なマイナークラスデータの生成 (参考 SMOTE(Synthetic Minority Over-sampling Technique)[Chawla 02]）
5. Bagging: bootstrap sampling を行い、アンサンブルで決定する方法。

連絡先: 紺野 友彦 tomohiko@nict.go.jp

実際には特定の方法のみを用いるのではなく、ダウンサンプリング+Bagging など複数の方法を組み合わせて不均衡データ問題に取り組むことが多い。([Wallace 11] など参照)

ダウンサンプリングを行うとマイナークラスの識別率が高くなる。しかしながら、マイナークラスに合わせて他クラスの利用データも減らすことになるため学習データの総量は少なくなってしまう。結果として識別問題全体の予想精度が低下してしまう。

我々は以下に示すように、深層学習の途中層で得られた特徴量を基に人工的な擬似特徴量を生成し、マイナークラスのデータ数を増強する方法を提案する。この方法による擬似特徴量生成は不均衡データ問題のみに限られる手法ではないと期待される。

2. 提案手法

提案手法の概要を図 1 に示した。手法は次の手続きからなる。

Step1: (図 1 左) 全データを用いて Neural Network 1 で学習を行う。

Step2: (図 1 中) 学習済み Neural Network 1 の途中層から訓練データの特徴量分布を取り出す。

Step3: （図 1 中）取り出した特徴量分布をガウス分布と仮定し、それに従う擬似特徴量を生成する。(2.1 で解説)

Step4: （図 1 中）生成された擬似特徴量と実訓練データ（の抽出特徴量）を合わせて Neural Network 2 で学習を行う。

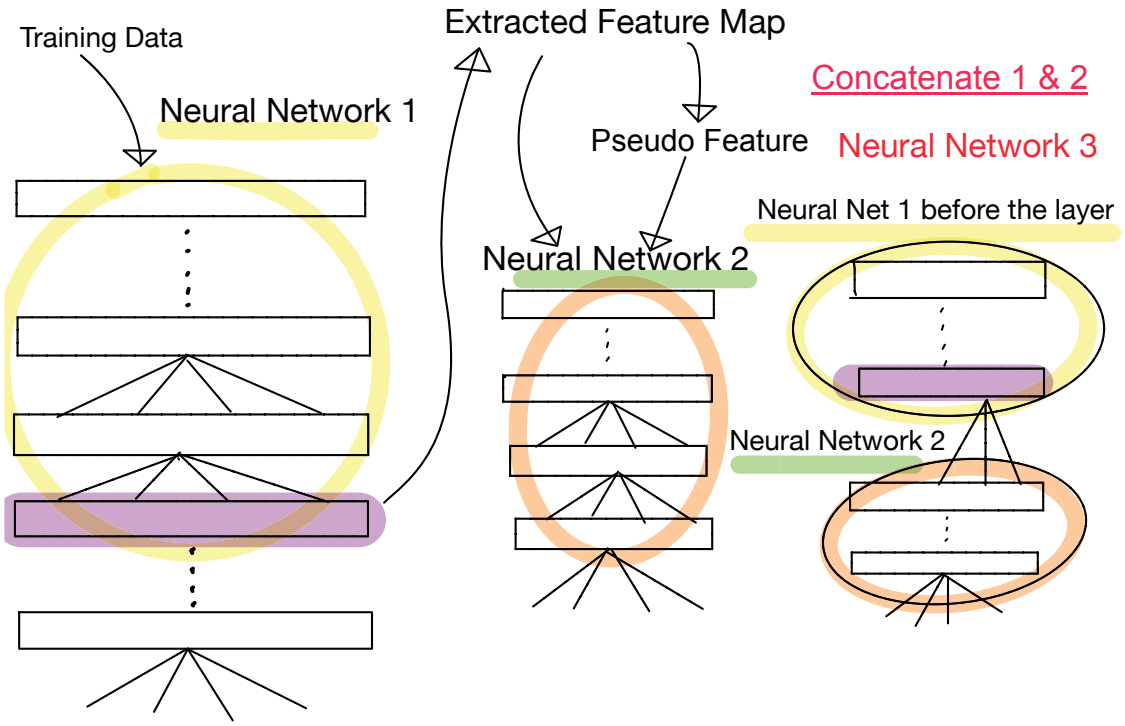


図 1: 提案手法の概要

Step5: (図 1 右) 学習済み Neural Network 1 の上部（特徴量を抽出した層以上の部分）と学習済み Neural Network 2 を結合して Neural Network 3 を作る。

Step6: (図 1 右) Neural Network 3 を用いて推論を行う。

今回は不均衡データの分析であるので、擬似特徴量で増強するのはマイナークラスである。

2.1 擬似特徴量の生成方法

擬似特徴量の作り方を説明する。クラス毎にそれぞれ別の擬似特徴量を生成する。特徴量分布がガウス分布に従うと仮定して 2 つの方法で擬似特徴量を生成した。特徴量分布の次元を D とする。1 つは特徴量が D 次元それぞれで独立なガウス分布に従うと仮定して生成した擬似特徴量分布である。つまり、$\mu_{i,c}, \sigma^2_{i,c}$ をそれぞれクラス c の特徴量分布の i 次元目の平均と分散として、クラス c の擬似特徴量の i 次元目の値 $x_{i,c}$ は以下の確率で次元ごとに独立なガウス分布に従うと仮定して得たものである。

$$\Pr(x_{i,c}) = \mathcal{N}(\mu_{i,c}, \sigma^2_{i,c}) \quad \text{for } i = 1, 2, \cdots, D \quad (1)$$

2 つ目は特徴量が多変量ガウス分布に従うと仮定して得られた擬似特徴量である。つまり $\vec{\mu_c}, \Sigma_c$ をそれぞれクラス c の D 次元特徴量分布の平均ベクトルと分散共分散行列とし、クラス c の擬似特徴量の分布 $\vec{x_c}$ は以下の多変量ガウス分布に従うと仮定して得られたものである。

$$\Pr(\vec{x_c}) = \mathcal{N}_D(\vec{\mu_c}, \Sigma_c) \quad (2)$$

図 2 で我々の提案した擬似特徴量が識別決定境界を押し出す直感的な様子をスケッチする。(図 2 左) 不均衡データの状況ではデータ数が過小なクラスは本来の境界（破線）よりも内側にサンプルが収まってしまう。識別決定境界（黒実線）を押せれば良い。（図 2 右) 我々の擬似特徴量生成は実データが作る境界（赤実線）を超える事が出来る。SMOTE ではこの境界を超えることはできない。この効果によって識別決定境界をより押し出すことが出来る。

3. 実験

結果は表 1 と表 2 にまとめた。現時点で、我々は入手の容易な Cifar10 [Krizhevsky 09] データセットを用いて人工的に不均衡データの状況を作り実験を行なった。Cifar10 は 10 クラスの画像を持ち訓練データ 5 万、テストデータ 1 万から構成される。クラス毎にそれぞれ 5000 枚、1000 枚の画像がある。値は試行の平均値である。

表 1: 全体の識別率 (Accuracy)

#Minor	Baseline	Down-Sampling	Proposed: Multivariate Normal	Proposed: Independent Normal
4	0.711	0.518	0.699	0.697
5	0.662	0.513	0.660	0.655
6	0.608	0.525	0.603	0.600
7	0.585	0.521	0.589	0.583
8	0.558	0.517	0.549	0.556
9	0.511	0.522	0.501	0.532

表 2: 識別率が最低のクラスの識別率 (Min of class-accuracy)

#Minor	Baseline	Down-Sampling	Proposed: Multivariate Normal	Proposed: Independent Normal
4	0.344	0.356	0.479	0.444
5	0.188	0.327	0.351	0.314
6	0.227	0.349	0.323	0.296
7	0.207	0.368	0.291	0.260
8	0.250	0.333	0.298	0.280
9	0.219	0.350	0.268	0.296

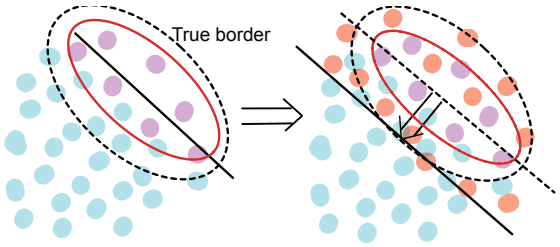


図 2: 直感的なスケッチ　青: メジャークラス　赤: マイナークラス　オレンジ: 擬似的に生成されたデータ

表中の#Minor とはデータの少ないクラスの数である。例えば、#Minor が 4 であるとは学習時には 10 種類の画像クラスのうち 4 クラスが 500 枚しかなく、残り 6 クラスは 5000 枚ある状況である。テスト時には不均衡データではなく 1 万枚のテストデータをそのまま用いた。表中の Baseline とは何もせずに全てのデータを用いて Neural Network 1 を使って学習したものである。Down-Sampling とは Neural Network 1 でダウンサンプリングで学習したものである。Proposed: Multivariate Normal とは特徴量分布から多次元ガウス分布を用いて擬似特徴量を生成した提案手法の結果である。Proposed: Independent Normal とは特徴量分布のそれぞれの次元で独立なガウス分布を用いて擬似特徴量を生成した提案手法の結果である。

3.1　実験結果のまとめ

結果は表 3 にまとめた。マイナークラス数が 4 以上の場合、我々の提案手法は全体の識別率 (Accuracy) は全データを利用した場合 (Baseline) と同じ程度に良く、識別率が最低なクラスの識別率 (Min of class-accuracy) もダウンサンプリングと同じ程度に良い識別率を得ることができた。つまり識別率 (Accucary) も識別率が最低なクラスの識別率 (Min of class accuracy) も共に良い [*1]。

表 3: 結果のまとめ (Summary)

	全体の識別率 (Accuracy)	最低識別率 (Min of Acc.)
Baseline	◯	×
Down-Sampling	×	◯
Proposed	◯	≈ ◯

4.　おわりに

深層学習の途中層から抽出した特徴量の分布から擬似的な特徴量分布を人工的に生成し、それを用いてマイナークラスのデータを増強した。それによって実験では全体の精度も最低精度も共に良い分類を行うことができた。我々が人工的に生成したデータは特徴量分布の境界を超えるものである。

*1　現時点の実験ではマイナークラス数が 3 以下の場合は我々の提案手法はそれほど良い結果を収めなかった。例えばマイナークラスが 1 クラスで残りが 9 クラスがメジャーの場合、これは転移学習のように残り 9 クラスで十分に学習してしまい、マイナーな 1 クラスのサンプル数が少なくてもマイナークラスの識別能力が高くなってしまうからでは無いかと推測している。

参考文献

[Chawla 02] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P.: SMOTE: synthetic minority over-sampling technique, *Journal of artificial intelligence research*, Vol. 16, pp. 321–357 (2002)

[Krizhevsky 09] Krizhevsky, A. and Hinton, G.: Learning multiple layers of features from tiny images (2009)

[Wallace 11] Wallace, B. C., Small, K., Brodley, C. E., and Trikalinos, T. A.: Class imbalance, redux, in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pp. 754–763IEEE (2011)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層強化学習による施工機械の経路生成

Route Generation of Construction Machine by Deep Reinforcement Learning

田邉峻也 Shunya Tanabe　孫澤源 Zeyuan Sun　中谷優之 Masayuki Nakatani　内村裕 Yutaka Uchimura

芝浦工業大学
Shibaura institute of technology

Recently, research on artificial intelligence has been progressing in various fields. In some of the Atari 2600games, the AI player has scored higher than the skilled human players by using deep reinforcement learning techniques. In this paper, autonomous ground leveling work by a bulldozer is targeted, which are expected to optimize action of the bulldozer. In previous work, we implemented deep Q learning method by giving images of simulator as the input data for the network. However, when learning the image using the convolution layer as input using deep reinforcement learning, it requires large computational cost for learning process. This research aims to reduce the computational cost by giving smaller order of input data. This paper describes the comparison results in different order of input data. Transition of the learning sequence is also evaluated.

1. 緒言

近年，自動車の自動運転，囲碁でプロ棋士に勝利する [David 16] など様々な分野で人工知能の研究が進んでいる．Atari2600 のゲーム [Bellemare 13] をルールの教示無しに画面の状態とスコアのみからコンピュータが繰り返し学習を行うことで，人間よりもスコアを上回った結果が報告されている [Mnih 15]．同論文では Deep Q Network（DQN）と呼ばれる強化学習の手法が用いられている．DQN は，環境の状況からエージェントが採るべき行動を選択し，報酬によってより良い行動の選択をする Q 学習法 [Riedmiller 09] と，画像処理の分野において大きな成果 [Krizhevsky 12] を上げている深層学習を組み合わせた手法である．さらに発展させた手法として Double Q learning[Hasselt 15] や Dueling Double DQN[Ziyu 16] など様々な手法によって人工知能の性能は向上している．

このような人工知能の応用分野として，本研究では建設作業の自動化を対象として行っている。建設作業のなかでも不整地の整地作業は，切り土，盛り土の状況に応じた運転経路の判断が必要なため，最終的な整地形状（出来型）に到達するための逐次の運転動作は，囲碁における大局観に通じるものがある。よって本研究では，不整地をブルドーザのような施工機械で整地する作業の自律化を想定し，DQN の手法を用いることによりブルドーザが自律的に最適な運転経路を生成することを目指す．本研究では，不整地をブルドーザのような施工機械で整地する作業の自律化を想定し，DQN の手法を用いることによりブルドーザが自律的に最適な運転経路を生成することを目指す．しかし，DQN による学習においては，ニューラルネットワークの入力，層数，各層のニューロン数，学習率など様々なハイパーパラメータの調整が必要であり，正しい値を設定しないと学習時間が長くなることや，学習がうまく行かないことがあるためハイパーパラメータ調整は重要である．ハイパーパラメータの調整には試行錯誤を重ねていく必要があるが，畳み込み層を利用した画像を入力として学習する場合，マルチコアの CPU や GPU を使っても学習の過程に非常に長い時間を要することが課題である．これまでに筆者らは，DQN を整地シミュレーションに適用して経路計画する研究を行ったが [孫 17, 中谷 17]，ネットワークへの入力が画面データのため，入力サイズが大きく計算負荷が高くなることが課題であった。そこで本研究では入力サイズを小さくすることで，ニューラルネットワークの規模を縮減し，ニューラルネットワークにおける計算コストを減らすことで学習の高速化を図ることを目的とする．

連絡先: 内村 裕，芝浦工業大学，東京都江東区豊洲 3-7-5，uchimura@shibaura-it.ac.jp

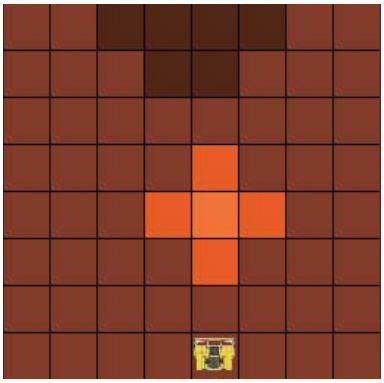


図 1: シミュレーション概要

具体的には，ネットワークへの入力を画像データの代わりに，ブルドーザや土砂の座標データもしくは，ブルドーザと土砂との相対座標とし，入力のサイズを小さくした場合の学習を行い，経路の最適性と高速化について検証する．

2. 問題設定

土砂の整地作業を単純化してシミュレートするために図 1 に示すように整地エリアを 8 × 8 のグリッド上に設定した．エージェントであるブルドーザで土砂山を移動し，図 2 のように地面の凹凸を無くし整地を行うことを目的とする．本シミュレーションでは図 3 のように土砂の高さを 6 段階で表現し，明度が高いと土砂が高く，暗い色ほど土砂が低いことを表している．ブルドーザは上下左右の 4 方向に動き，土砂を押すことがで

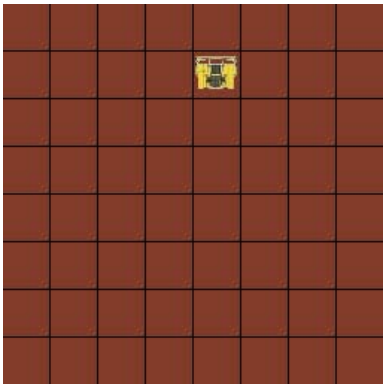


図 2: 整地終了時

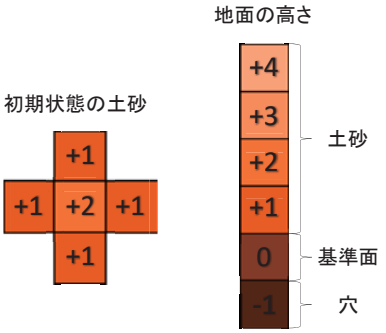


図 3: 土砂と穴の概要

きるが，引くことはできない．基準となる地面の高さを 0 としたとき，高さ −1 の地点の穴に高さ 1 以上の土砂を移動させることで整地を行う．また，図 6 のように，各エピソードにおいて 6 パターンの初期位置をランダムに与えた．報酬の設定は，土砂の高さを −1 の地点に高さ 1 以上の土砂を近づけて移動させることで +1，土砂を 1 エリア整地することで+10，すべてのエリアを整地することで+50 とした．すべての穴が整地されるか，ブルドーザの行動回数が 100 回に達した場合に 1 エピソードの終了とし，エピソードを繰り返し行うことで Q ネットワークの学習を行った．

ネットワークに与える入力の違いによる学習結果の比較のために以下の 2 つの異なる入力データを用意した．

- ケース 1 : ブルドーザの，土砂山・穴の座標値を入力とした場合
- ケース 2 : ブルドーザの周囲のセルの地面の高さを入力とした場合

ケース 2 では入力をシミュレータ画面（8 × 8）より大きなサイズ（15 × 15）のセルとしているが画面外のセルの値は 0 とした。ケース 1 とケース 2 のネットワークの各パラメータを表 1 に示す．中間層に 3 層の全結合層を用いており，各層のニューロン数は 1000，250，50 とした。また，ブルドーザの前後左右方向の移動の 4 つのアクションを出力とした．

学習過程では，割引率 γ を 0.99 に設定した．ϵ-greedy 法によるランダム率は，学習開始時では 100%とし，その後，学習回数が増えるに従って徐々に減らし，学習回数が 100 万回以降は 10%とした．また，学習データを保存するメモリサイズ

表 1: ケース 1 とケース 2 のパラメータ

パラメータ	ケース 1	ケース 2
入力サイズ	38	15 × 15
出力数	4	4
中間層	3 層	3 層
中間層の種類	全結合層	全結合層
ミニバッチ数	50	50
学習率	1.0×10^{-5}	1.0×10^{-5}
割引率	0.99	0.99

表 2: 文献 [孫 17] のネットワークのパラメータ

パラメータ	文献 [孫 17]
入力サイズ	40 × 40 × 4
出力数	4
中間層	7 層（畳み込み 3 層，プーリング 3 層，全結合 1 層）
ミニバッチ数	50
学習率	1.0×10^{-5}
割引率	0.99

は 10 万エピソード分のリングバッファとした．本研究では，Double Q learning を適用し，教師信号を出力するニューラルネットワークはランダムに初期化すると同時に，1 万回の学習ごとに学習済みのニューラルネットワークで上書きした．シミュレーションの 1 エピソードの最大行動回数は 100 回に設定した．また学習回数が 10 万回間隔で学習したネットワークを保存し，ランダム率を 0%としたシミュレーションを行い，行動回数と整地した穴の数を記録した．なお，シミュレータの作成には Pygame を使用し，Double Q learning の実装においては，TensorFlow をフレームワークとして用いた。また，GPU として GeForce GTX 1070（NVIDIA 製）を使用した．

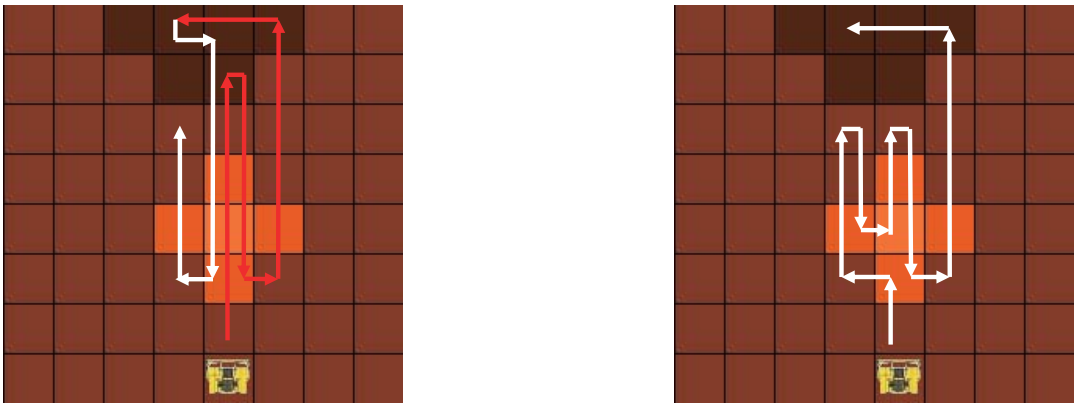
さらに，ニューラルネットワークに与える入力サイズによる計算コストの比較のために，文献 [孫 17] における画像データを入力とした場合の学習も行った．文献 [孫 17] では，1 グリッドを 5 × 5 ピクセルで構成した 8 × 8 のグリッドのシミュレータの画面（40 × 40 ピクセル，8bit 階調）のピクセルデータを 1 画面として，過去 3 画面を含めて 4 画面分を入力としている．同ネットワークに使用したパラメータを表 2 に示す．中間層には畳み込み層が 3 層，プーリング層が 3 層，全結合層が 1 層の合計 7 層をニューラルネットワークに与えられる．しかし，ニューラルネットワークの規模が大きいため，学習時間が長くなるのが問題点である．

3. シミュレーションによる学習結果と比較

3.1 入力データの相違による学習結果

入力層に与える入力データとしてをブルドーザの，土砂，穴の座標値とした場合（ケース 1）と，ブルドーザの周囲のセル（15 × 15）の地面高さを入力とした場合（ケース 2）のシミュレーションによる学習結果を図 5 に示す．グラフ横軸が学習回数，縦軸が各初期位置の穴をすべて整地するまでの行動回数の平均を示している．本結果に示すようにケース 1 ではブルドーザの行動回数が減少せず，学習回数 1,300 万回の試行を行っても穴を全て埋めるには至らず，学習による経路生成が不調に終わった．

ケース 2 では，学習を重ねるごとにブルドーザが全ての穴



(ケース 2)　　(文献 [孫 17])

図 4: 生成経路の比較

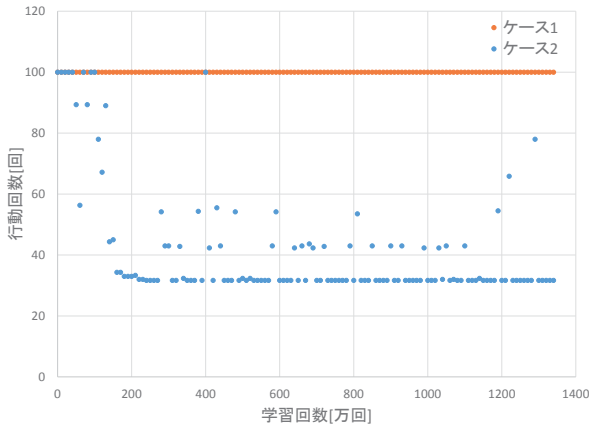


図 5: ケース 1 とケース 2 の学習結果の比較

を整地するまでの行動回数が減少していることがわかる．また穴や土砂の初期位置の異なるパターンにおいても整地を完了する経路の最適化することができた．また，学習回数が 240 万回のときに保存したニューラルネットワークを用いてランダム率 0% でブルドーザをシミュレーション上で動かしたときの 6 パターンの各初期位置の平均行動回数が 31.6 回となった．その後，学習回数 1,300 万回まで学習を行ったが，ブルドーザは 31.6 回より行動回数が少なくなることはなかった。

検証結果が示すように，ケース 2 は学習が進むにつれ整地を完了するまでの行動回数を減少したが，ケース 1 は最大行動回数の 100 回以内に全ての穴の整地を行うことができなかった．また，学習回数を増やしてもすべての穴を整地するまでのブルドーザの行動回数は減少することはなかった．このとき，エピソード開始時は，ブルドーザは上方への行動を数回選択し，いくつかの穴を整地したが，その後全く整地に至る動作を行わなかった．この結果から，ケース 1 で与えられた入力データでは行動価値関数の関数近似が上手く行われていないと考えられる。ケース 1 はケース 2 と比べて入力サイズが少ないことから情報量が少ないこと，入力データが座標データのスカラーの数値のみであり，ケース 2 のような 2 次元配列としての相関がないため学習がうまく進まなかったと推測できる。

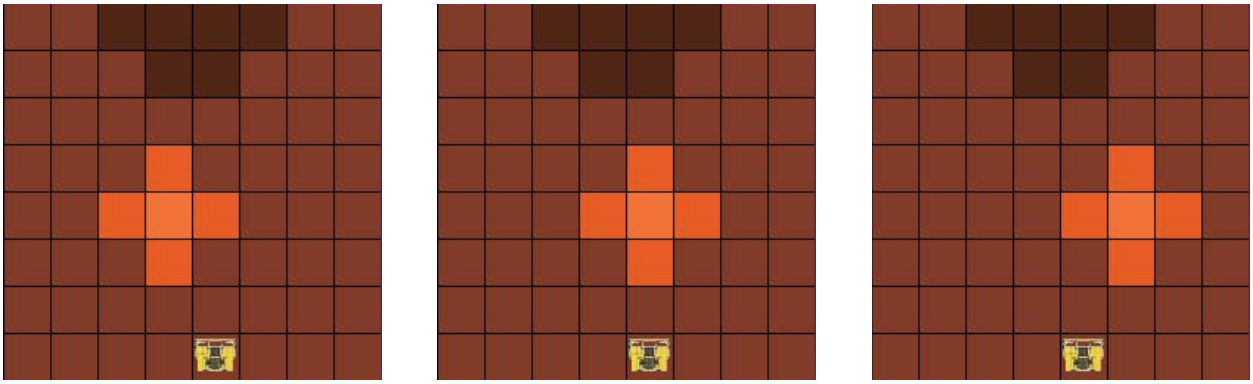
表 3: ケース 2 と文献 [孫 17] の比較

パラメータ	学習時間 [s]	最小行動回数 [回]
ケース 2	5320	31.7
文献 [孫 17]	12300	29.0

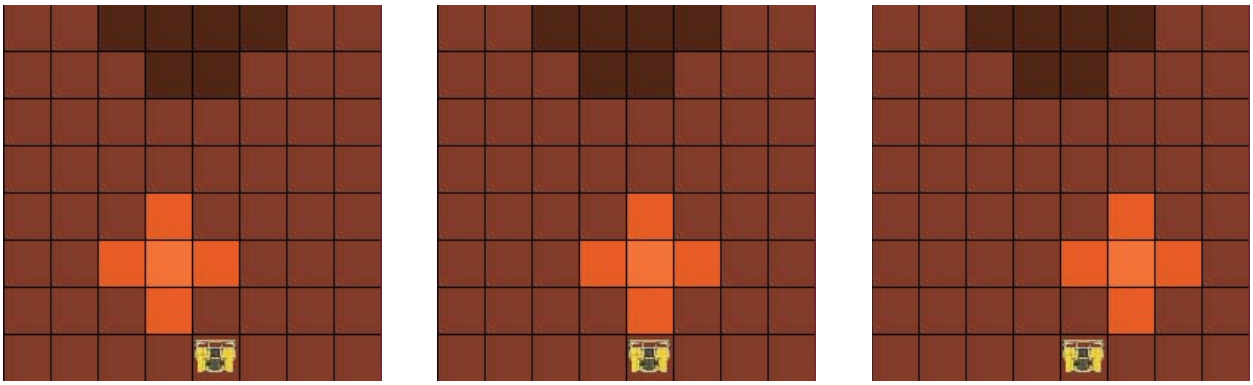
3.2　ケース 2 と文献 [孫 17] の比較

ブルドーザの周囲のセルの地面高さを入力としたケース 2 と，文献 [孫 17] の学習結果の比較を図 7 に示す．縦軸に初期位置の異なる穴をすべて整地するまでのエージェントの行動回数の平均を示し，横軸に学習時間を示している．本結果が示すように，文献 [孫 17] が収束するまでの時間が約 5 時間かかるに対して，ケース 2 とした場合では約 2 時間で収束することがわかる．また，ケース 2 と文献 [孫 17] を学習回数 100 万回にかかる時間とすべての穴を埋めるまでのブルドーザの平均行動回数を比較したものを表 3 に示す．これらの結果が示すように，ケース 2 では文献 [孫 17] と比べて入力サイズを少なくすることでニューラルネットワークの規模が小さくなり，学習時間が約 2 分の 1 に短縮できたことがわかる．しかし，文献 [孫 17] のブルドーザの最小行動回数が 29.0 回に対して，ケース 2 では 31.7 回と行動回数が多くなっていた．また，同じ初期位置のブルドーザの経路をケース 2 と文献 [孫 17] で比較したものを図 4 に示す．矢印はすべての穴を整地するまでのブルドーザの軌跡であり，ブルドーザが 1 マス移動することで行動回数が増えていく．この初期位置の場合ではブルドーザの行動回数は，ケース 2 としたとき 28 回，文献 [孫 17] では 22 回となった．ケース 2 を入力として学習された経路では，最初に 4 つの土砂を運んで穴の整地を行っているが，最後の 2 つの土砂を 1 つずつ運ぶことにより，ブルドーザの行動回数が増えていることがわかる．文献 [孫 17] の経路は，最初に 1 つずつ土砂を運んでいるが，これ以上行動回数を減らすことができない最も効率の良い経路であった．

本研究で適用した手法 (Double DQN) では，現在の状態においてエージェントが最も価値の高い行動を選択するが，人間と違い現在の状態から未来の状態を予測することができない問題点があり，文献 [孫 17] に比べても効率が悪い経路を選択している．文献 [孫 17] と違い，入力数が少ないことや，ニューラルネットワークの規模，畳み込み層の有無などが影響して局所解に陥ってしまった結果，最短経路に到達できなかったと考えられる．



(a) (b) (c)



(d) (e) (f)

図 6: 初期位置

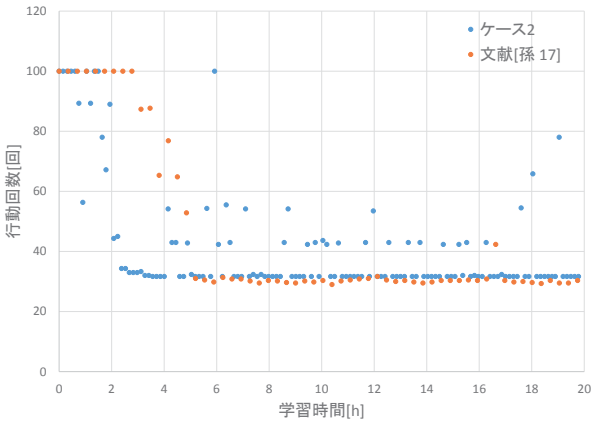


図 7: ケース 2 と文献 [孫 17] の学習結果の比較

4. 結言

本研究では，ブルドーザによる整地作業の自律化を目的に，ニューラルネットワークの入力データとして，ブルドーザや土砂の座標データ，もしくはブルドーザ周囲のセルの状態を与えて学習結果の比較を行った．ブルドーザ周囲のセルの状態を与えることで入力サイズを小さくすることで，ネットワークの規模も縮減し，学習の高速化を実現した．今後は経路をより最適化することを目指すと同時に，より現実に近い問題設定における学習の高速化を図る予定である．

参考文献

[David 16] David, S., Aja, Huang. ”Mastering the game of Go with deep neural networks and tree search” Nature 529 484/489, (2016)

[Bellemare 13] Bellemare, M. G. Naddaf, Y. Veness, J. Bowling, M. :”The arcade learning environment: An evaluation platform for general agents” Journal of Artificial Intelligence Research 47 253-279 (2013)

[Mnih 15] Mnih, V., Kavukcuoglu, K. and Silver, D. ”Human level control through deep reinforcement learning” Nature 518 pp.529-533, 2015.

[Riedmiller 09] Riedmiller, M. Gabel, T. Hafner, R. :”Reinforcement learning for robot soccer” Auton Robot (2009) 27: 55-73 DOI:10.1007/s10514-009-9120-4 (2009)

[Krizhevsky 12] Krizhevsky, A. Sutskever, I. Hinton, G. :”ImageNet classification with deep convolutional neural networks” Adv. Neural Inf. Process. Syst.25, 1106-1114 (2012)

[Hasselt 15] Hasselt, H., Guez, A. and Silver, D.”Deep Reinforcement Learning with Double Q-learning” arXiv:1509.06461v3 cs.LG (2015)

[Ziyu 16] Ziyu, W., Tom, S., Matteo, H., Hado, V,H., Marc, L., Nando, F. ”Dueling Network Architectures for Deep Reinforcement Learning” arXiv:1511.06581v3 cs.LG (2016)

[孫 17] 孫澤源, 中谷優之, 内村 裕, ”深層強化学習による整地作業の自律制御”, ロボット学会学術講演会論文集, 3I2-04, (2017)

[中谷 17] 中谷優之, 孫澤源, 内村 裕, ”深層強化学習による最適経路学習”, 第 60 回自動制御連合講演会, SuD1-4, (2017)

Oral presentation | General Session | [General Session] 2. Machine Learning

[2A2] [General Session] 2. Machine Learning

座長:石畠 正和（NTTコミュニケーション科学基礎研究所）

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room A (4F Emerald Hall)

[2A2-01] Counting garbage bags from garbage collecting video using deep learning

○Kazuhiro Mikami[1], Makoto Kawano[1], Yin Chen[1], Jin Nakazawa[1] （1. Keio University）

1:20 PM - 1:40 PM

[2A2-02] Deep few-shot learning with pseudo example optimization

○Akisato Kimura[1,2], Zoubin Ghahramani[2], Koh Takeuchi[1], Tomoharu Iwata[1], Naonori Ueda[1] （1. Nippon Telegraph and Telephone Corporation, 2. University of Cambridge）

1:40 PM - 2:00 PM

[2A2-03] Verification of the harvesting date prediction method using deep learning

○Kengo Miyoshi[1], Yukikazu Murakami[1] （1. National Institute of technology, Kagawa College）

2:00 PM - 2:20 PM

[2A2-04] Visualizing the Behavior of the Inner Layers of Convolutional Neural Networks by Layer-wise Relevance Propagation

Hirotaka Sakai[1], ○Yoshitaka Kameya[1], Takahiro Sota[2], Hiroaki Arie[2] （1. Meijo University, 2. DENSO CORPORATION）

2:20 PM - 2:40 PM

[2A2-05] Solving Four Arithmetic Operation Problems and Word Algebra Problems by Neural Programmer-Interpreters

○Shota Katsumata[1], Katsumi Inoue[1,2] （1. Tokyo Institute of Technology, 2. National Institute of Informatics）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習を用いたゴミ収集動画からゴミ袋個数の測定

Counting garbage bags from garbage collecting video using deep learning

三上量弘 *1 河野 慎 *1 陳 寅 *1 中澤 仁 *1
Kazuhiro Mikami Makoto Kawano Yin Chen Jin Nakazawa

*1慶應義塾大学
Keio University

Reducing garbage generated in our daily life is a key aspect to fulfill a sustainable human society. The lack of an effective method to obtaining information regarding the collected garbage, such as where, when and what kinds of waste are collected, is one of the major obstacles to promoting the garbage reduction in cities. In this research, we propose a system to automatically count the number of garbage bags, by using convolutional neural networks (CNN)-based technology to process the videos of garbage collecting process taken by a camera mounted at the back of garbage trucks. The proposed system is evaluated using the realistic garbage collecting videos of Fujisawa city. The experiment result shows that the proposed method achieves a high recall score of 0.90.

1. はじめに

持続可能な社会を実現するために，ゴミ減量化運動が日本各地で行われている．ゴミを減量するためには，人々に自分がどれだけゴミを出したのかや削減したのかなどを知らせることによって，ゴミ減量への意識を高める必要がある．しかし，現在得られるゴミの量のデータ粒度は図 1 のように市や区単位といったものであり，人々のゴミ減量への意識を高めるにはさらに微細粒度なゴミの量のデータが必要となる．これらのゴミ量のデータを人で集めることは可能であるが，コストの観点から現実的ではない．したがって，自動的にそれらのデータを収集可能なシステムが必要である．

そこで，深層学習による物体検出と清掃車に取り付けられたドライブレコーダーに着目した．清掃車の後部にはドライブレコーダーが取り付けられ，清掃員らがゴミを清掃車のトランクに投入する様子が全て撮影されている．これらの動画からゴミ袋の種類と数を検出できれば，清掃車が街の中でゴミを収集しながら，人間を介せずに，収集されたゴミ袋の種類と数を検出することができる．清掃車にセンサを取り付けた様々な研究が行なっている [中澤 17, Kawano 17]．図 2 のように清掃車は網羅的に市内を走行している [Chen 17]． これらの清掃車に提案するシステムを実装することによって，微細粒度のゴミの収集量を自動的に取得することができると考えられる．

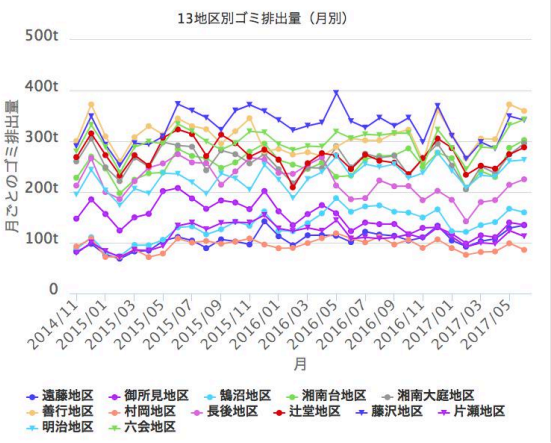


図 1: 藤沢市各地区のゴミ排出量の推移 [筑波]

連絡先: 三上量弘，慶應義塾大学，quantan@ht.sfc.keio.ac.jp

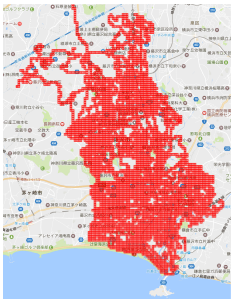


図 2: 慶應大学が開発した清掃車を用いたセンサーネットワークのカバレッジ

本研究では，清掃車に取り付けられたドライブレコーダーで撮影された動画と GPS によって取得された位置情報を用いて，地域の微細粒度なゴミの量を自動的にリアルタイムセンシングするシステムを提案する．今回は，畳み込みニューラルネットワーク (CNN) を用いて動画のフレームにおいて目標のゴミ袋を検出，トラッキングし，清掃車に投げ込まれたかを判定するアルゴリズムによりゴミ袋を数える手法を実装した．また可燃ゴミを例としてデータセットを作成し，CNN で可燃ゴミを検出できるよう学習した．実際に藤沢市を走行する清掃車後部に取り付けられたドライブレコーダーで撮影された動画を用いて実験を行い，手法の精度について評価を行う．

2. 微細粒度な地域のゴミ収拾量測定システム

2.1 システム構成

システムの全体図を図 3 に示す．清掃車後部に設置されたカメラでゴミが投げ込まれる動画を撮影し，動画から GPU が搭載された小型コンピューターでゴミ袋の種類と個数を検出して，各種類のゴミの個数と位置情報をデータサーバに 3G ネットワークで送信する．これにより微細粒度な地域毎のゴミ量を自動的に取得することが可能になる．また，動画そのものを送る必要がないので，市民の個人情報の安全性の観点や，通信コストや維持コストにおいても優れている．

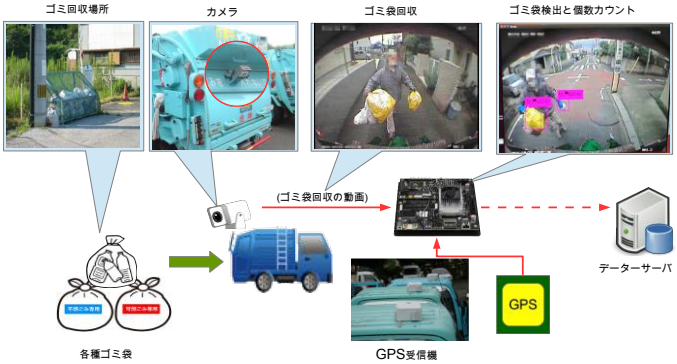


図 3: システムの全体図

2.2 ゴミの個数カウント手法

動画に写る物体のカウントには，物体検出，トラッキング，カウントの判定といったプロセスが必要となる．ゴミ袋の個数を清掃車のトランクに投げ込まれた個数と考えると，清掃車に取り付けられた動画を用いてゴミ袋の個数をカウントするには以下の 3 つのプロセスが必要である．

1. 動画の各フレームに映るゴミ袋を検出
2. フレーム間において，検出されたゴミ袋を同定
3. ゴミ袋が清掃車に投げ込まれたかを判定

提案手法では，物体検出可能な畳み込みニューラルネットワーク (CNN) である Single shot multibox detector（SSD）を用いて動画のフレームにおいて目標のゴミ袋を検出，Intersection over Union（IoU）によるトラッキングアルゴリズムを用いて検出したゴミ袋をフレーム間で同定，清掃車に投げ込まれたかを判断するアルゴリズムによりゴミ袋の個数を数える．手法のフローチャートを図 4 に示す．

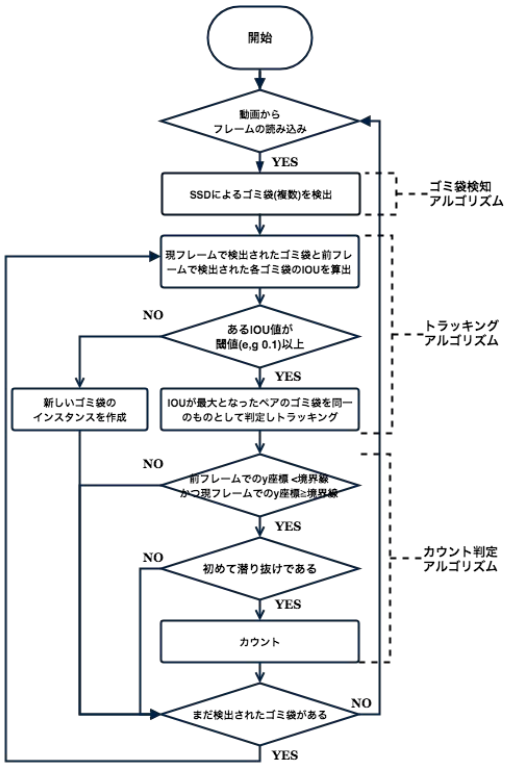


図 4: ゴミの個数カウントのフローチャート

2.2.1 ゴミ袋の検出

ゴミ袋の検出には物体検出可能な CNN である Single shot multibox detector（SSD）を用いる [Liu 16]．SSD はベースとなる CNN に VGG-16[Simonyan 14] を使用し，VGG-16 の fully-connected 層の代わりに新たに 6 層の畳み込み層を加え，end-to-end で物体検出をすることが可能である．物体検出可能な CNN は SSD の他にも多くあるが，その中でも SSD は検出の精度と速度において優れている．SSD は入力画像が異なる SSD-300 と SSD-500 の 2 つのタイプがある．リアルタイム性を考慮し，検出精度は劣るが速度が高い SSD-300 を今回は採用した．また SSD よりも精度と速度の両方において優れている YOLOv2[Redmon 16] が存在するが，画像に大きく写る物体の検出において SSD は検出精度が優れている．この性質が図 6 のように動画の手前で行われるカウントの判定において効果的と考え，ゴミ袋検出に SSD を採用した．

2.2.2 トラッキングアルゴリズム

動画に映る物体のトラッキングアルゴリズムは， カラーヒストリグラムを利用する Mean shift[Fukunaga 75] がある．しかし，SSD によって出力されるのはゴミ袋の画像における領域情報である．そのため，トラッキングアルゴリズムには領域情報を使用する IoU を用いた．図 5 のように IoU は 2 つの領域を和と積で割ったものであり，重なっている割合を表す値である．SSD によって検出された現在のフレームのゴミ袋と 1 つ前のフレームのゴミ袋全ての IoU を算出し，閾値以上かつ IoU が最大となったゴミ袋のペアを同一のものとして，トラッキングする．閾値を越えなかった場合はそのゴミ袋を新しいゴミ袋と判定する．ドライブレコーダーで撮影された動画は 8pfs とフレームレートが低く，閾値を高く設定してしまうとトラッキングすることが難しかったため，今回のアルゴリズムでは閾値を 0.1 とした．

$$IOU = \frac{A \cap B}{A \cup B} =$$

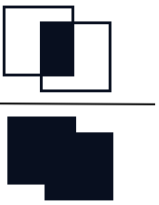


図 5: Intersection over Union

2.2.3 カウント判定アルゴリズム

清掃車後部に設置されたドライブレコーダーで撮影された動画は図 6 のような角度で撮影されているため，ゴミ袋は画面上方から下方へとフレーム間を移動していくように映っている．この特性を利用し，カウントは図 6 のように SSD で出力されたゴミ袋のボックスの中にある赤い点が，前フレームでは青いライン越えておらずかつ次フレームでは越えていたときにカウントされる．下方から青いラインを超えてもカウントされない．また，トラッキングが成功している限り，1 度カウントされたゴミ袋は条件判定により 2 度カウントされない．赤い点はボックスの幅の 1/2，高さの 6/7 に位置する点を表している．中心点をカウント判定とするとゴミがカウントされずそのまま画面外へ行ってしまい，カウントされないことが多かったため，カウント判定するゴミ袋の位置座標をこのような値に設定した．



図 6: （上）: 実際に清掃車後部に取り付けられたドライブレコーダーで撮影された動画のフレーム画像.（下）: ゴミ袋がフレーム間において上方から下方へ移動しトラッキング，カウントされている．

3. 実験

実際に藤沢市を走行する清掃車の後部に取り付けられたカメラで撮影された動画を用いて，提案手法の精度を評価した．

3.1 データセット

学習用のデータセットを作成するため，藤沢市を走行する異なる 2 台の清掃車の後部に取り付けられたドライブレコーダーで可燃ゴミが収集される場面が撮影された動画を用いた．今回可燃ゴミを学習対象としたのは，可燃ゴミが回収されるゴミの量の 76.2%[?] を占め，可燃ゴミを検出できるメリットが大きいと考えたためである．動画は 8fps， 640 × 480 ピクセルで撮影されている．その動画をフレーム化し，その中からゴミ袋が映る画像のみを選別し，画像内に映るゴミ袋にアノテーションを施した．訓練データセットは 2016 年 10 月 28 日に撮影された動画から 1372 枚，2017 年 5 月 11 日に撮影されたものから 259 枚を作成した．テストデータには 2017 年 5 月 11 日に撮影された動画から 539 枚作成した．

3.2 SSD の学習

学習では可燃ゴミ袋を分類するように学習した．ベースとなる VGG-16 の重みには ILSVRC CLS-LOC で事前に学習されたものを使用した．学習したのち，ゴミをどれだけ正確に識別できるかを示す Average Precision を PASCAL VOC の基準を用いてテストデータで評価したところ 0.53 と算出された．

3.3 評価手法

評価は F 値を用いた．TP をゴミを正確にカウントできた数，FN をゴミを見逃してカウントできなかった数，FP をゴミでないものをゴミと判断してカウントした数として，式 1 のようにそれぞれ Precision，Recall，F 値を求めた．

表 1: ゴミ袋カウントの混同行列

		実際の結果	
		True	False
アルゴリズムによる判定	True	TP	FP
	False	FN	TN

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F - 値 = \frac{2Precision * Recall}{Precision + Recall} \quad (3)$$

3.4 実験結果

実験結果を表 2 に示す．Precision が 0.42 とあまり高くない一方で，Recall は 0.90 と高い値である．これはゴミ袋を見逃してカウントできないことが多い一方，ゴミ袋ではないものを誤ってゴミ袋としてカウントすることが少ないことを示している．

表 2: 指標と精度

指標	ゴミの個数
TP	184
FP	255
FN	21

指標	精度
Precision	0.42
Recall	0.90
F-measure	0.57

4. 考察

実験結果に基づく考察を以下に示す．ゴミ袋でないものを誤ってゴミ袋としてカウントする例として，図 7 のように道路の白線のかすれや清掃員の帽子をゴミと誤識別してしまうことが多く見られた．この問題点は，新しく帽子のクラスを作ることによって帽子を検出できるようにし，帽子が通過してもカウントしないようにすることで解決できると考える．

また，ゴミ袋を見逃してしまいカウントできない原因として，動画のフレームレートが 8FPS と低いことが考えられる．フレームレートが低いため，清掃車にゴミ袋を投げ入れる速度が高いと IOU の値が低くなり，閾値を超えることができずトラッキングできない事例が多く見られた．図 8 が示すように，画面に映るゴミ袋の大きさも関係があり，小さいと IOU の値もまた小さくなりやすく，トラッキングができないことも多く見られた．また図 9 の左側のように，ゴミ袋が画面横方向に離れているとカウントできる一方で，右側のように画面横方向に近接し直列的にゴミ袋があると，ゴミ袋が画面上方から下方へ移動するという収集作業の特性のため，前フレームにあったゴミ袋が次のフレームで後ろにあるゴミ袋との IOU 値が高くなり同定されてしまい，カウントされないことが多く見られた．これらの問題点の解決法として，動画のフレームレートを高める方法がある．フレームレートを高めることによって，フレーム間でのゴミの領域情報の類似性が向上し，IOU の値が高まりトラッキングをより正確に行うことができると考える．



図 7: （左）: 掠れている白線をゴミ袋と誤認している.（右）: 帽子を誤認識し，誤ってカウントしてしまっている．



図 8: （左）: ゴミ袋が大きく写っているためトラッキングが成功して正しくカウントできている.（右）: 2 つのゴミ袋が画面横方向に近いためトラッキングが失敗しカウントできていない.

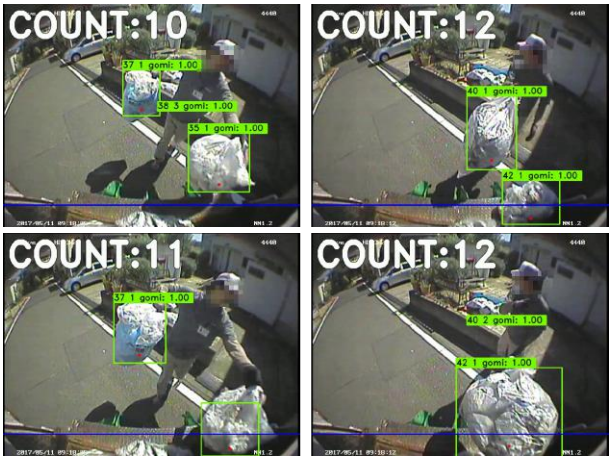


図 9: （左）: 2 つのゴミ袋が画面横方向に離れているため正しくトラッキングできている.（右）: 2 つのゴミ袋が画面トラッキングができていない.

5. 結論と今後

本研究では，地域毎のゴミの量を自動的に推定するシステムを考案した，また，物体検出可能な CNN と清掃車後部に設置されたカメラで撮影された動画を用いて，ゴミ袋の個数を自動的にカウントする手法を実装した．実際に藤沢市を走行する清掃車後部に取り付けられたカメラで撮影された動画を用いて実験を行い，提案手法の精度を評価した．実験の結果，Precision が 0.42 とあまり高くない一方で，Recall は 0.90 と高い数値を示し，提案手法がゴミ袋ではないものをゴミ袋としてカウントすることが少ないことを明らかにした．また，提案手法の精度を高めるためには動画のフレームレートを高める必要があることがわかった．

今後の研究の発展として，SSD のアーキテクチャーを軽量化しゴミ袋を検出する速度を高めることがある．清掃車に用いられる予定の GPU が搭載された小型コンピューターで SSD によるゴミ袋の検出を試してみたところ，1 フレームあたり約 0.3 秒の処理時間を要した．8fps の動画を処理するとなると，動画 1 秒間あたり約 2.4 秒の処理時間が必要となる．本手法を実際の清掃車に搭載しリアルタイムでゴミ袋の個数をカウントすることはできなくはないが，不便を生じさせる．そのため，SSD のアーキテクチャーの軽量化が必要となる．今回 SSD のベースネットワークに用いた VGG-16 は SSD の処理時間の 80%が費やされている．したがって，ベースネットワークに VGG-16 ではなく，より軽量なものを用いることによってリアルタイム化ができると考える．また，軽量化することによって，処理速度が高まる一方検出精度は低くなるが，ゴミ袋でないものを誤ってカウントすることが少なかったという今回の実験結果から，その影響は低いと考えられる．

謝辞

本研究の一部は NTT ソフトウェアイノベーションセンタに支援頂いた．

参考文献

[Chen 17] Chen, Y., Yonezawa, T., and Nakazawa, J.: Towards Vehicle-Assisted Adaptive Wireless GeoMesh Network for Smarter Cities, in *VEHICULAR 2017: The Sixth International Conference on Advances in Vehicular Systems, Technologies and Applications* (2017)

[Fukunaga 75] Fukunaga, K. and Hostetler, L. D.: The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp. 32–40 (1975)

[Kawano 17] Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T., and Nakazawa, J.: Road Marking Blur Detection with Drive Recorder, pp. 4010–4015 (2017)

[Liu 16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C.: Ssd: Single shot multibox detector, in *European conference on computer vision*, pp. 21–37Springer (2016)

[Redmon 16] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *arXiv preprint arXiv:1612.08242* (2016)

[Simonyan 14] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *ArXiv e-prints* (2014)

[筑波] 筑波大学：ゴミ減量 G1 グランプリ, http://bigdata.kde.cs.tsukuba.ac.jp/ g1gp/app/index.html/

[中澤 17] 中澤 仁, 陳 寅, 米澤 拓郎, 大越 匡, 徳田 英幸：ユニバーサルセンサネットワークと清掃車を活用した藤沢市のスマート化, デジタルプラクティス, Vol. 8, No. 3, pp. 244–252 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

疑似訓練サンプル最適化による ニューラルネットワークの少数ショット学習

Deep few-shot learning with pseudo example optimization

木村 昭悟 *1*2 Akisato Kimura
Zoubin Ghahramani*2 Zoubin Ghahramani
竹内 孝 *1 Koh Takeuchi
岩田 具治 *1 Tomoharu Iwata
上田 修功 *1 Naonori Ueda

*1日本電信電話 (株) コミュニケーション科学基礎研究所
Communication Science Laboratories, NTT Corporation

*2ケンブリッジ大学 工学部
Department of Engineering, University of Cambridge

This paper proposes a novel method for training neural networks with a limited amount of training data. Our approach is based on knowledge distillation that transfers knowledge from a deep reference neural network to a shallow target one. The proposed method employs this idea to mimic predictions of reference models more robust against overfitting than the target neural network. Different from almost all the previous work for knowledge distillation that requires a large amount of labeled training data, the proposed method requires only a small amount of training data. Instead, we introduce pseudo training data that is optimized as a part of model parameters.

1. はじめに

画像認識に関連する様々なタスクの state-of-the-art は，数多くの層を持つ深いニューラルネットワーク (neural networks; NN) モデルと，大規模教師付データセットによって支えられている．しかし，解くべき問題に合わせて大規模教師付データセットを都度準備することは非常に多くの人的労力が必要となると共に，医療応用などいくつかの実世界の問題においては，そもそもデータそのものを大量に獲得すること自体が困難である場合も珍しくない．少量の訓練データで NN を学習すると，訓練データに過剰に適合する過学習を引き起こし，タスクの性能が大幅に低下する．NN モデルの過学習を避ける方策として，大量の教師なしデータを追加で用いる半教師付学習 [Goodfellow 14] や，大量の教師付データから学習したモデルを追加で利用する転移学習 [Caruana 94, Koch 15] が広く用いられている．しかし，いずれの方策においても，大量の教師なしデータや学習済モデルなど何らかの追加データを準備する必要があり，少数の訓練データのみから NN を適切に学習する方法は，依然として重要な研究課題の一つとなっている．

一方，SVM やガウス過程回帰など，従来から用いられている機械学習モデルのいくつかは，ニューラルネットワークと比較して過学習を引き起こしにくいことが知られている．本研究では，この点に着目した新しいニューラルネットワークの学習方法を提案する．図 1 に提案手法の概要を示す．提案手法では，半教師付学習や転移学習とは異なり，所与の少量訓練データ以外のデータ資源は仮定しない．提案手法では，まず NN よりも過学習に強いとされる NN 以外の予測器を所与の訓練データで学習し，この予測器を参照モデルとして，目的の NN モデルを知識蒸留 (knowledge distillation) [Buciluǎ 06, Hinton 15] により学習する．すなわち，参照モデルの出力と目標 NN モデルの出力が同様となることを目指して，目標モデルを学習する．従来の知識蒸留手法では，参照モデル・目標モデルの双方の学習において共通の大規模訓練データを用いることが一般的であり，訓練データが少量の場合には目標モデルを十分に学習することができない．提案手法では，この問題を解決するために，少数の訓練データから生成した大量の疑似訓練データ (pseudo training data) を導入すると共に，この疑似訓練データを目標モデルの学習の過程で学習損失がより大きくなるように追加更新を行う．これにより，現在の目標モデルで十分に学習が行われていない領域を探し出し，その領域における学習を促進させることを目指す．さらに，提案手法では，参照モデルによる予測の不確かさに基づいて疑似訓練サンプルの重みを調整する fidelity weighting [Dehghani 17] を導入する．これにより，参照モデルの予測が確からしい領域のみを目標モデルに転移でき，目標モデルの予測性能のさらなる向上が期待できる．

連絡先: 木村昭悟，日本電信電話（株），神奈川県厚木市森の里若宮 3-1，kimura.akisato@lab.ntt.co.jp

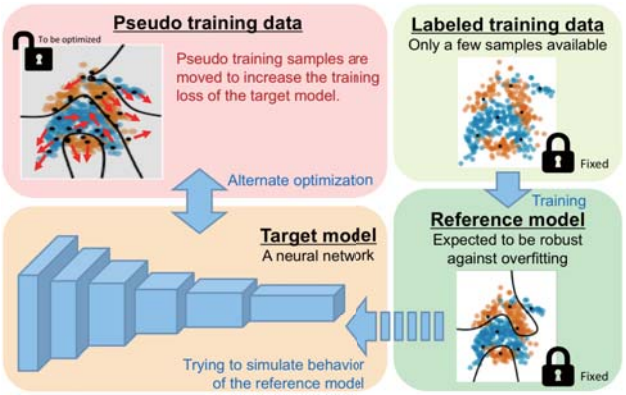


図 1: 提案手法の概略

2. モデル最適化

2.1 知識蒸留

知識蒸留 [Buciluǎ 06, Hinton 15] は，訓練済の参照モデル $g(\boldsymbol{x};\boldsymbol{\theta}_g)$ が持つ予測性能を，目標モデル $f(\boldsymbol{x};\boldsymbol{\theta}_f)$ に転移させるための枠組である．ここで，$\boldsymbol{x}$ は入力サンプル，$\boldsymbol{\theta}_g$ と $\boldsymbol{\theta}_f$ はそれぞれ参照モデル・目標モデルのモデルパラメータである．以降では簡単のため，特に必要な場合を除いてモデルパラメー

タを省略する．多くの知識蒸留手法では，参照モデルとして層が多いもしくは幅が広い NN モデルを仮定し，層が少ないもしくは幅の狭い目標 NN モデルに参照モデルと同等程度の予測性能を持たせることを目的としている．その際に用いられる訓練損失の最も一般的な定式化 [Hinton 15] を以下に示す．以降では，この訓練損失を蒸留損失 (distillation loss) と呼ぶ．

$$L_{\mathrm{dis}}(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}}) = \frac{\lambda_1}{N_{\mathrm{L}}} \sum_{n=1}^{N_{\mathrm{L}}} D_1(\boldsymbol{y}_n^{\mathrm{L}}, f(\boldsymbol{x}_n^{\mathrm{L}})) + \frac{\lambda_2}{N_{\mathrm{L}}} \sum_{n=1}^{N_{\mathrm{L}}} D_2(g(\boldsymbol{x}_n^{\mathrm{L}}), f(\boldsymbol{x}_n^{\mathrm{L}})), \quad (1)$$

ここで，$\boldsymbol{X}^{\mathrm{L}} = \{\boldsymbol{x}_1^{\mathrm{L}}, \ldots, \boldsymbol{x}_{N_{\mathrm{L}}}^{\mathrm{L}}\}$ は教師付訓練サンプルの集合，$\boldsymbol{Y}^{\mathrm{L}} = \{\boldsymbol{y}_1^{\mathrm{L}}, \ldots, \boldsymbol{y}_{N_{\mathrm{L}}}^{\mathrm{L}}\}$ はそれに対応する教師情報の集合，$D_1(\boldsymbol{y}, \hat{\boldsymbol{y}})$ は教師情報 $\boldsymbol{y}$ とモデル予測 $\hat{\boldsymbol{y}}$ との乖離を評価する損失関数，$D_2(\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2)$ は 2 つの異なるモデル予測 $\hat{\boldsymbol{y}}_1$ と $\hat{\boldsymbol{y}}_2$ との乖離を評価する損失関数である．知識蒸留は分類問題に適用する，すなわち教師情報 $\boldsymbol{y}$ を one-hot vector とするのが一般的であるが，教師情報や損失関数を適切に設定することで回帰問題など他の問題にも適用可能である．

2.2 提案手法の損失関数

式 (1) に示した蒸留損失を含め，従来の知識蒸留手法の多くは，教師付訓練サンプル $(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}})$ を用いて目標モデル $f(\cdot)$ を学習する．しかし，少量の教師付訓練データのみでモデル学習を行うと過学習を引き起こす．ここで，蒸留損失の式 (1) を確認すると，右辺第 2 項は教師情報 $\boldsymbol{Y}^{\mathrm{L}}$ がなくても計算できることに気がつく．提案手法では，この点を考慮して，新しい訓練損失 L_{imi} を定義する．この損失では，少数の教師付訓練サンプル $(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}})$ に加えて，新たに疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}} = \{\boldsymbol{x}_1^{\mathrm{P}}, \ldots, \boldsymbol{x}_{N_{\mathrm{P}}}^{\mathrm{P}}\}$ を導入し，この疑似訓練サンプルから蒸留損失の第 2 項に相当する項を計算する．以降では，この訓練損失を模倣損失 (imitation loss) と呼ぶ．

$$L_{\mathrm{imi}}(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}}, \boldsymbol{X}^{\mathrm{P}}) = \frac{\lambda_1}{N_{\mathrm{L}}} \sum_{n=1}^{N_{\mathrm{L}}} D_1(\boldsymbol{y}_n^{\mathrm{L}}, f(\boldsymbol{x}_n^{\mathrm{L}})) + \frac{\lambda_2}{N_{\mathrm{P}}} \sum_{n=1}^{N_{\mathrm{P}}} D_2(g(\boldsymbol{x}_n^{\mathrm{P}}), f(\boldsymbol{x}_n^{\mathrm{P}})). \quad (2)$$

2.3 疑似訓練サンプルの重み付け

先の模倣損失では，いずれの疑似訓練サンプルも同様に重要であることを仮定しているが，実際には目標モデルの学習に悪影響を与えるサンプルも含まれている．例えば，参照モデルの予測の不確実性 (uncertainty) が大きい場合，その参照モデルの予測を目標モデル学習の教師情報として用いることは望ましくない．そこで提案手法では，fidelity weighting [Dehghani 17] の考え方に基づく疑似訓練サンプルの適応的重み付けを導入する．Fidelity weighting では参照モデルの予測の不確実性に応じてモデル更新の学習率を調整しているが，本論文では模倣損失における疑似訓練サンプルの重みを調整する．

$$L_{\mathrm{imi}}(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}}, \boldsymbol{X}^{\mathrm{P}}) = \frac{\lambda_1}{N_{\mathrm{L}}} \sum_{n=1}^{N_{\mathrm{L}}} D_1(\boldsymbol{y}_n^{\mathrm{L}}, f(\boldsymbol{x}_n^{\mathrm{L}})) + \frac{1}{N_{\mathrm{P}}} \sum_{n=1}^{N_{\mathrm{P}}} \lambda_2(g, \boldsymbol{x}_n^{\mathrm{P}}) D_2(g(\boldsymbol{x}_n^{\mathrm{P}}), f(\boldsymbol{x}_n^{\mathrm{P}})). \quad (3)$$

ここで，$\lambda_2(g, \boldsymbol{x}_n^{\mathrm{P}})$ は疑似訓練サンプル $\boldsymbol{x}_n^{\mathrm{P}}$ の重みであり，この疑似訓練サンプルを入力とした参照モデルの予測 $g(\boldsymbol{x}_n^{\mathrm{P}})$ の不確実性 $\sigma_g(\boldsymbol{x}_n^{\mathrm{P}})$ から以下のように計算される．

$$\lambda_2(g, \boldsymbol{x}_n^{\mathrm{P}}) = \hat{\lambda}_2 \exp(-\overline{\lambda}_2 \sigma_g(\boldsymbol{x}_n^{\mathrm{P}}) / \overline{\sigma}_g), \quad (4)$$

ここで，$\hat{\lambda}_2$ は重みの上限値，$\overline{\sigma}_2$ は不確実性の平均値，$\overline{\lambda}_2$ は不確実性が平均値であったときの重みを示すパラメータである．上記の重みを計算するためには，参照モデルから予測の不確実性を求める必要がある．そこで本論文では，fidelity weighting と同様，ガウス過程 (Gaussian process; GP) 分類器を参照モデルとして採用する．

3. 疑似訓練サンプル最適化

ここまで，提案手法で用いる訓練損失について説明してきたが，高い予測性能を持つ目標モデルを学習するためには，疑似訓練サンプルの配置が非常に重要となる．本節では，疑似訓練サンプルの配置方法について述べる．

疑似訓練サンプルの考え方は，ガウス過程のスケーラブルな推論を目的とする 誘導点法 (inducing point method) [Snelson 06] から着想を得ている．誘導点は学習の過程で更新される疑似訓練サンプルであり，モデルパラメータと同様に，目的関数である経験尤度の下限 (evidence lower bound; ELBO) を最大化するように更新される．一方で，本手法では，目標モデルのパラメータは前述の模倣損失を小さくするように更新するが，疑似訓練サンプルは逆に模倣損失を大きくするように更新する．この戦略により，疑似訓練サンプルを現在の目標モデルの学習が十分に進んでいない領域へ移動させ，目標モデルの学習をより進めることを目指す．以下では，疑似訓練サンプルの更新方法について，具体的に説明する．

提案する疑似訓練サンプルの更新方法は，敵対的サンプルを用いた学習 [Szegedy 15] から着想を得ている．敵対的サンプルは，訓練サンプルに注意深く設計されたわずかな摂動を与えることで，元の訓練サンプルから得られるモデル予測とは大きく異なる予測を与えるサンプルのことを指す．最も基本的な敵対的なサンプルの生成手法として，最急勾配符号法 (fast gradient sign method; FGSM) [Goodfellow 15] が知られている．この方法では，所与の教師付訓練サンプル $(\boldsymbol{x}^{\mathrm{L}}, \boldsymbol{y}^{\mathrm{L}})$ から以下の式で敵対的サンプル $\boldsymbol{x}_{\mathrm{AT}}(\boldsymbol{x}^{\mathrm{L}}, \boldsymbol{y}^{\mathrm{L}})$ を計算する．

$$\boldsymbol{x}_{\mathrm{AT}}(\boldsymbol{x}^{\mathrm{L}}, \boldsymbol{y}^{\mathrm{L}}) = \boldsymbol{x}^{\mathrm{L}} + \epsilon \ \mathrm{sign}\{\nabla_{\boldsymbol{x}^{\mathrm{L}}} D_1(\boldsymbol{y}^{\mathrm{L}}, f(\boldsymbol{x}^{\mathrm{L}}))\}, \quad (5)$$

ここで $\nabla_{\boldsymbol{x}}$ は $\boldsymbol{x}$ についての偏微分，$\epsilon \geq 0$ は定数である．

上記の式 (5) は，定数 ϵ を学習率とみなし，勾配符号項 $\mathrm{sign}\{\nabla_{\boldsymbol{x}^{\mathrm{L}}} D_1(\boldsymbol{y}^{\mathrm{L}}, f(\boldsymbol{x}^{\mathrm{L}}))\}$ を単なる勾配 $\nabla_{\boldsymbol{x}^{\mathrm{L}}} D_1(\boldsymbol{y}^{\mathrm{L}}, f(\boldsymbol{x}^{\mathrm{L}}))$ に置き換えると，確率的勾配降下法 (stochastic gradient descent method; SGD) によるモデルパラメータの更新則に極めて類似していることに気がつく．ただし，この更新は目的関数 $D_1(\boldsymbol{y}^{\mathrm{L}}, f(\boldsymbol{x}^{\mathrm{L}}))$ を増加させる方向に進むことに注意する．教師付訓練サンプル $\boldsymbol{x}^{\mathrm{L}}$ に代えて疑似訓練サンプル $\boldsymbol{x}^{\mathrm{P}}$ を，教師情報 $\boldsymbol{y}^{\mathrm{L}}$ に代えて参照モデルの出力 $g(\boldsymbol{x}^{\mathrm{L}})$ をそれぞれ用いると，以下の式によって疑似訓練サンプル $\boldsymbol{x}^{\mathrm{P}}$ を更新できる．

$$\boldsymbol{x}_{\mathrm{imi}}(\boldsymbol{x}^{\mathrm{P}}) = \boldsymbol{x}^{\mathrm{P}} + \epsilon \nabla_{\boldsymbol{x}^{\mathrm{P}}} D_2(g(\boldsymbol{x}^{\mathrm{P}}), f(\boldsymbol{x}^{\mathrm{P}})). \quad (6)$$

上記は最急勾配符号法及び確率的勾配降下法を基礎とする更新式であるが，深層学習で発展した各種の確率的最適化手法や，対象サンプルが画像である場合にはより自然な敵対的サンプルを生成する方法 [Zhao 17] などを適用することも可能である．

目標モデルのパラメータと疑似訓練サンプルとを交互に更新する学習手順は，敵対的生成モデル (generative adversarianl networks; GAN) の生成器と識別器との交互更新と方針が似ており，このような敵対的なモデル学習は一般的に不安定になりやすいことが知られている [Salimans 16]．そこで，本論文では，疑似訓練サンプルを逐次更新するのではなく，実際のモデル学習に用いるサンプルと更新するサンプルとを分離して，適切なタイミングで更新用サンプルを学習用サンプルに統合する方法を採用する．疑似訓練サンプルの更新を含む，提案手法の具体的なアルゴリズムは，以下の通りとなる．

1. （少数の）教師付訓練サンプル $(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}})$ を用いて参照モデル $g(\cdot)$ を学習する．
2. 初期疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(0)}$ 及びそのコピー $\boldsymbol{X}^{\mathrm{P}(1)}$ を教師付訓練サンプル $\boldsymbol{X}^{\mathrm{L}}$ から作成，ステップを表現するインデックスを $t=0$ とする．
3. 教師付訓練サンプル $(\boldsymbol{X}^{\mathrm{L}}, \boldsymbol{Y}^{\mathrm{L}})$，現在の疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(t)}$ 及び参照モデル $g(\cdot)$ を用いて，目標モデル $f(\cdot)$ のモデルパラメータ $\boldsymbol{\theta}_f$ を更新（式 (3)）．
4. 参照モデル $g(\cdot)$ 及び現在の目標モデル $f(\cdot)$ を用いて，次の疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(t+1)}$ を更新（式 (6)）．
5. 3-4 を所定のエポック数だけ繰り返す．
6. 次の疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(t+1)}$ のコピー $\boldsymbol{X}^{\mathrm{P}(t+2)}$ を作成，現在の疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(t)}$ を次の疑似訓練サンプル $\boldsymbol{X}^{\mathrm{P}(t+1)}$ に統合 $\boldsymbol{X}^{\mathrm{P}(t+1)} \leftarrow \boldsymbol{X}^{\mathrm{P}(t)} \cup \boldsymbol{X}^{\mathrm{P}(t+1)}$，インデックスをインクリメント $t \leftarrow t+1$．
7. 3-6 を所定のステップ数だけ繰り返す．

4. 実験

4.1 定性的評価

提案手法の挙動を確認するために，まず最初に 2 次元 2 値分類データセットである Banana dataset [*1] を用いた実験を行った．総計 400 個の 2 次元サンプルを事前に固定したランダム行列により 100 次元空間に線形変換を行い，各クラスから 5 個ずつサンプルを選択し，この 10 個のサンプルを教師付訓練サンプルとして採用した．図 2 (a) に，教師付訓練サンプル（橙・青×印，色がクラスに対応）及びその他の（未使用の）サンプル（橙・青点）の配置を示す．放射基底関数 (radial basis function; RBF) カーネルを用いたガウス過程分類器を参照モデルとして採用し，変分ガウス近似 [Opper 09] を用いてモデル学習及びカーネルパラメータ最適化を行った．目標モデルとして，7 層の全結合層からなる NN を採用し，中間層のユニット数はすべて 1,000 とした．最適化手法として，モデルパラメータの更新には Nesterov 加速勾配を用いた Adam (Nadam) を，疑似訓練サンプルの更新には Adam をそれぞれ用い，初期学習率をそれぞれ 0.001 及び 0.05 に設定した．NN 学習のバッチサイズは 100，エポック数は 200 とした．

まず，最も極端な例として，図 2 (b) に示すように，特徴空間をグリッド上に敷き詰めるように疑似訓練サンプルを配置した場合を検証した．この設定では，疑似訓練サンプルを最適化せず，予測の不確実性に基づく疑似訓練サンプルの重み付けも行わなかった．損失関数として，D_2 には KL ダイバージェンスを用い，D_1 は使用しなかった．すなわち，この設定では，教師付訓練サンプルを目標モデルの学習に用いなかった．図 2 (c)(d) に参照モデルと目標モデルの識別境界をそれぞれ示す．この図から，提案手法を用いることで，極限的には参照モデルの挙動をほぼ完全に目標モデルに転移することができることがわかる．

*1 https://github.com/GPflow/GPflow/tree/master/doc/source/notebooks/data

表 1: MNIST（上）及び fashion MNIST（下）に対する分類性能．“Distill”, “opt” 及び “fidelity” はそれぞれ，模倣損失の利用，疑似訓練サンプル最適化，予測の不確実性を考慮した疑似訓練サンプルの重み付けに対応．

Methods / #labeled	10	20	50	100	200
NN plain	37.9	46.0	66.0	78.3	86.7
GP plain	39.9	51.6	64.6	73.2	80.0
Distill	43.5	51.2	67.7	78.1	86.1
Distill, opt	44.1	53.5	70.0	79.3	**86.7**
Distill, opt, fidelity	**44.1**	**53.9**	**70.4**	**80.0**	86.4

Methods / #labeled	10	20	50	100	200
NN plain	39.3	47.9	58.3	64.9	71.3
GP plain	44.6	52.4	59.9	65.7	71.4
Distill	43.6	50.9	60.0	67.3	**72.5**
Distill, opt	41.2	49.6	60.1	67.3	72.2
Distill, opt, fidelity	**44.8**	**52.7**	**62.1**	**68.0**	**72.5**

続いて，より現実的な例として，教師付訓練サンプルを用いて 250 個の初期疑似訓練サンプルを生成し，提案手法のアルゴリズムを用いて最終的に 1,000 個まで疑似訓練サンプルを増やす場合を検証した．すなわち，目標モデル学習は 4 ステップに別れ，各ステップではそれぞれ 250, 500, 750 及び 1000 個の疑似訓練サンプルを用いて 50 エポックの学習を行った．初期疑似訓練サンプルは，半分を教師付訓練サンプルの線形内挿により，残り半分を教師付訓練サンプルに小さなガウスノイズを加えることにより，それぞれ生成した．図 2 (e) に初期疑似訓練サンプルの配置を示す．損失関数として，D_1 にはヒンジロス，D_2 には KL ダイバージェンスを用い，両者の重み λ_1 及び $\overline{\lambda}_2$ はいずれも 1.0 とした．図 2 (f)(g)(h) に学習済目標モデルの識別境界及び最終的な疑似訓練サンプルの配置を示す．(f)(g)(h) はそれぞれ，疑似訓練サンプルのみを学習に用いた場合，加えて教師付訓練サンプルを用いた場合，さらに予測の不確実性に基づく重み付けを用いた場合を示している．いずれの場合においても，疑似訓練サンプルが特徴空間を覆い尽くしていなくても，参照モデルの挙動をある程度目標モデルに転移できていることがわかる．また，図 2 (g)(h) から，教師付訓練サンプルを学習に含めることで，参照モデルの識別境界をできる限り維持しつつ，教師付訓練サンプルの予測をより良くできていることがわかる（中央下部の橙×印に着目）．

4.2 定量的評価

次に，標準的ベンチマークである MNIST [Lecun 98] 及び fashion MNIST [Xiao 17] を用いて，提案手法の分類性能を定量的に評価した．参照モデルとして再びガウス過程分類器を用い，参照モデルに関する設定はすべて前節と同じものを用いた．目標モデルとして，3 層の畳み込み NN を用い，各畳み込み層は カーネルサイズを 3×3，ストライドを 2（奇数層）もしくは 1（偶数層），第 l 層のチャネル数を $16\times(\mathrm{floor}(l/2)+1)$ と設定した．また，各畳み込み層の後に batch normalization, pReLU 活性化層 及び dropout (0.5) を導入すると共に，最終畳み込み層の後に 2 層の全結合層（batch normalization, pReLU 及び dropout を第 1 層の後に追加）を配置した．最適化手法は前節と同様であり，初期学習率をいずれも 0.02 に設定した．各データセットに含まれる 60,000 個の訓練サンプルの中から各クラス 1 から 20 個のサンプルをランダムに 20 通り選択し，これを教師付訓練サンプルとして用いた．また，教師付訓練サンプルの線形内挿により 1,250 個の初期疑似訓練サンプルを生成し，提案手法のアルゴリズムにより最終的に 10,000 個まで疑似訓練サンプルを増やした．すなわち，目標モデル学習は 8 ステップで構成され，各ステップで 25 エポッ

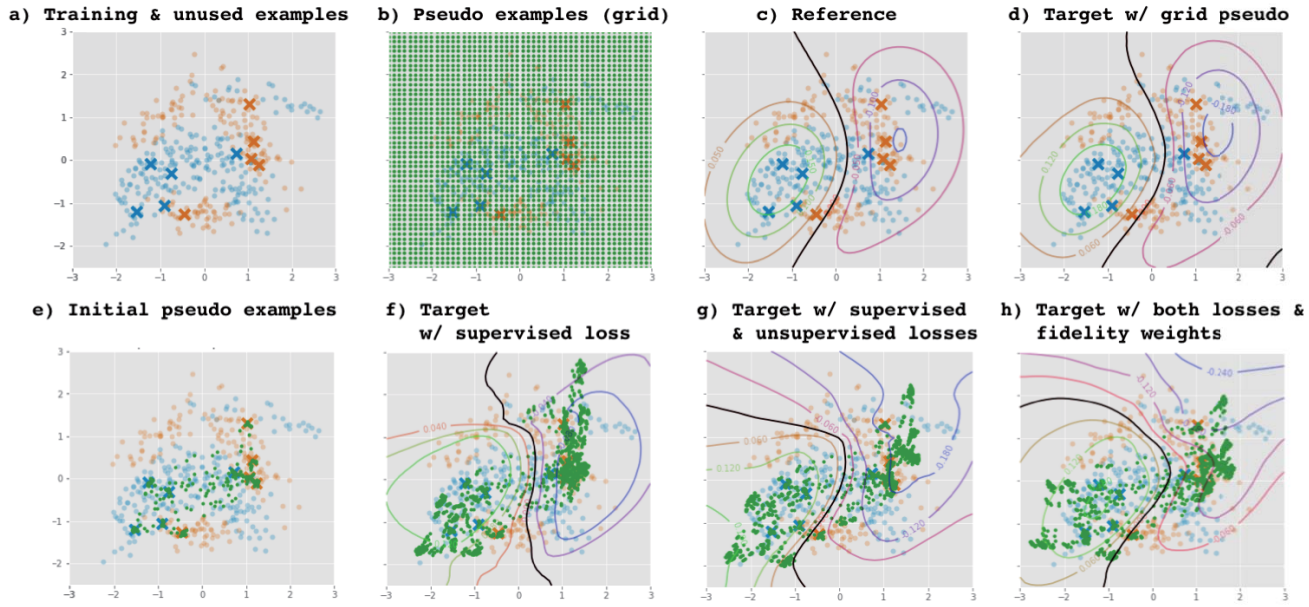


図 2: Banana データセットを用いた際の提案手法の予測結果の可視化．×印：教師付訓練サンプル，橙・青点：データセット内の他のサンプル（学習未使用），緑点：疑似訓練サンプル，黒線：分類境界．

クの学習を行った．損失関数として，D_1 にはヒンジロス，D_2 には KL ダイバージェンスを用い，重み λ_1 は 1.0 に，$\overline{\lambda}_2$ は教師付訓練サンプルの増加に伴って 100.0 から 1.0 まで徐々に減らした．評価は，各データセットに含まれる 10,000 個のテストサンプルを用いて，20 通りの教師付訓練サンプルの選択に対する分類正解率の平均値で評価した．その他の実験設定は，前節と同様である．この実験では，以下の 5 手法を比較する．(1) 教師付訓練サンプルのみで学習した NN 分類器，(2) 参照モデルであるガウス過程分類器，(3) 固定した疑似訓練サンプル 10,000 個と模倣損失を用いて学習した目標モデル，(4) 模倣損失と疑似訓練サンプルの更新を用いて学習した目標モデル，(5) 4 に加えて予測の不確実性を考慮した疑似訓練サンプルの重み付けを用いて学習した目標モデル．

表 1 に実験結果を示す．この結果から，提案手法 (3)-(5) のいずれもが，単純な NN 学習を上回るもしくはほぼ同等の分類性能を発揮できたことがわかる．また，提案手法は，参照モデルの挙動を目標モデルに転移する戦略を採用しているにもかかわらず，ほぼすべての場合において参照モデルの分類性能を同等もしくは上回る分類性能を示したことがわかる．さらに，MNIST データセットでは，教師付訓練サンプルの数が増えると，教師付訓練サンプルのみで学習した NN 分類器が，参照モデルであるガウス過程分類器よりも優れた分類性能を持つ．この設定は，参照モデルの挙動を模擬しようとする提案手法にとって望ましくないが，このような状況であっても，提案手法はさらに優れた分類性能を示した．提案手法の個別要素について確認すると，模倣損失の導入のみでもある程度の分類性能を担保できるが，疑似訓練データの最適化と予測の不確実性に基づく重み付けを導入することでさらに分類性能を向上できる．

5. Conclusion

本論文では，少量の訓練データのみから NN を学習するための手法として，疑似訓練サンプルの導入及びその最適化に基づく手法を提案した．提案手法の基本的な枠組は非常に一般的であり，予測の不確実性に基づく疑似訓練サンプルの重み付けを除けば，どのようなタイプの参照モデル・目標モデルにも適用が可能である．例えば，従来の知識蒸留と同様に，深い NN モデルを参照モデル，浅い NN モデルを目標モデルとすることも可能であり，その逆の場合にも適用可能である．さらには，SVM や random forests など，別の種類の予測器を参照モデルとして利用することも可能である．疑似訓練サンプルの最適化についても，最新の最適化手法や敵対的サンプル生成手法を導入することで，さらなる性能向上の可能性が見込める．

参考文献

[Buciluǎ 06] Buciluǎ et al.: Model Compression, in *Proc. KDD.* (2006)

[Caruana 94] Caruana et al.: Learning Many Related Tasks at the Same Time with Backpropagation, in *Proc NIPS* (1994).

[Dehghani 17] Dehghani et al.: Fidelity-Weighted Learning, *ArXiv* (2017)

[Goodfellow 14] Goodfellow et al.: Generative Adversarial Nets, in *Proc. NIPS.* (2014)

[Goodfellow 15] Goodfellow et al.: Explaining and Harnessing Adversarial Examples, in *Proc. ICLR.* (2015)

[Hinton 15] Hinton et al.: Distilling the Knowledge in a Neural Network, in *NIPS Workshop.* (2015)

[Koch 15] Koch et al.: Siamese Neural Networks for One-shot Image Recognition, in *ICML Workshop.* (2015)

[Lecun 98] Lecun et al.: Gradient-based learning applied to document recognition, *Proc. IEEE.* (1998)

[Opper 09] Opper et al.: The Variational Gaussian Approximation Revisited, *Neural Computation.* (2009)

[Salimans 16] Salimans et al.: Improved Techniques for Training GANs, in *Proc. NIPS.* (2016)

[Snelson 06] Snelson et al.: Sparse Gaussian Processes using Pseudo-inputs, in *Proc. NIPS.* (2006)

[Szegedy 15] Szegedy et al.: Going Deeper with Convolutions, in *Proc. CVPR.* (2015)

[Xiao 17] Xiao et al.: Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, *ArXiv* (2017)

[Zhao 17] Zhao et al.: Generating Natural Adversarial Examples, *ArXiv* (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習を用いた収穫日予測手法の検証

Verification of the harvesting date prediction method using deep learning

三好健悟[*1]　村上幸一[*1]
Kengo Miyoshi　Yukikazu Murakami

[*1]独立行政法人 国立高等専門学校機構 香川高等専門学校
National Institute of Technology, Kagawa College

Contract farming has a managerial advantage that farmers can directly negotiate prices with business partners. And it is necessary to predict harvesting date of agricultural crops precisely for contract farming. Conventionally, this precondition has been solved only by experience rule of experienced farmers. However, for new farmers who do not have experiential rules or for large-scale field managers, contract farming is difficult management method. In order to solve this problem, we proposed automatic harvesting date prediction method using statistical model by deep learning. We confirmed that deep learning models exceed the accuracy of non-deep learning models.

1. はじめに

農業の経営手法の一種である契約栽培は, 営農者が取引先との価格交渉を直接行えるという利点がある. しかし契約栽培では多くの場合において農作物の収穫日や収量を正確に予測することが前提条件であり, これには技術的あるいは技能的な予測手法が必要となる.

従来この問題は熟練営農者の経験則によって解決されてきた. しかし新しく農産業に参加する新規営農者にとっては経験則に依る手法は意味を持たない. また熟練営農者にとっても異なる品種, 圃場, 作業日程で複数の作物を管理する場合, それらすべての収穫日予測を行うことは困難である. 本研究グループではこれまで，深層学習の一種であるディープニューラルネットワークによる自動的な収穫日予測手法について提案してきた. 本論では提案手法の予測精度について検証する.

2. 理論

本研究では収穫日予測を行う回帰出力モデルとしてディープニューラルネットワーク [1](以下,DNN) を主に用いる. DNN は人間などが持つ情報処理システムの中枢である脳神経系を模擬した数理的モデルであるニューラルネットワーク [1] (以下,NN) を深層化したものであり, その構造は図 1 に示される.

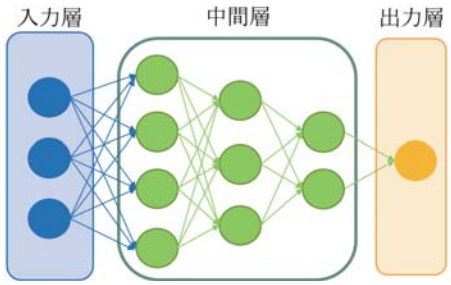


図 1: ディープニューラルネットワーク

連絡先: 村上幸一,（独）国立高等専門学校機構　香川高等専門学校，香川県高松市勅使町 355，087-869-3917，087-869-3909，murakami@t.kagawa-nct.ac.jp

DNN は誤差逆伝播法によって各層の結合重みが更新され, 学習を進める. しかし DNN の深層化は勾配消失問題を引き起こす可能性を持つ. 問題によりノード間の結合重みの更新量が小さくなりすぎると学習が停滞する.

3. 要素技術

(1) MXNet
MXNet [2] は種々のホスト言語にサポートされているマシンラーニング (以下,ML) フレームワークの一つである. 提案手法はこれを R で用い, DNN を設計する. MXNet は他の ML フレームワークと比べて計算機リソースによるスケーリングの効率やホスト言語の多さなどで優れる.

(2) ベイズ最適化とランダムサーチ
本研究では DNN のハイパーパラメータ探索を行う作業工程にベイズ最適化とランダムサーチを用いる.

4. 予測手法の設計

DNN の設計にあたり, 以下の手順を踏んで予測モデルを構築した.

1. 収集した計 35 サイクルのレタス栽培データをデータセットとして用いた. 各サイクルの収穫日を出力パラメータ, またそれ以外の各作業日程と気象データを入力パラメータとして設定した. 各パラメータは正規化した.

2. K-fold cross validation 法を用いてデータセットを訓練データとテストデータに分割した. 分割サイズは 10 分割とした. モデルの精度は 10 回分の出力における MAE を算出し, それらを平均した平均 MAE で評価した. MAE は (1) 式により求められる. ここでは，真の値は出力された予測値，N は予測値の数である.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (i = 1,2,\ldots,N) \qquad (1)$$

3. MXNet を用いてデータセットを元に回帰出力を行う NN, DNN を設定した.

5. 予測手法の設計

収穫日予測手法として， DNN を用いた深層学習で高精度な結果が得られるかを以下の手順で実験した．

(a) 中間層が 1 層の NN と中間層が 2 ～ 5 層の DNN について，ベイズ最適化とランダムサーチによるハイパーパラメータの決定を行う．

(b) 上記の 5 種類の NN, DNN を用いてレタス栽培データセットの回帰出力モデルを構築する．

(c) 学習回数を 10 ～ 1500 回の範囲において 10 回刻みで変化させ，学習回数ごとに出力される平均 MAE を記録する．

(d) (a) の工程と (c) の工程を通して記録された 平均 MAE の中で最も値が小さな学習回数のモデルを，最適な NN, DNN として評価する．

6. 実験結果

実験結果を表 1 に示す．深層化した DNN の中で最良な平均 MAE を持つモデルは前節 (a) でベストチューニングが得られた中間層が 4 層の場合であった．図 2 に前節 (c) でベストチューニングが得られた 3 層 DNN の平均 MAE-学習回数の関係を示す．また図 3 に 3 層 DNN の MAE が最小になった学習回数 980 回時点での各テストデータに対応する MAE を示す．図 3 においてその標準偏差は 1.774 であった．図 4 に，今回の実験で得られた中間層の構造を示す．実験では 3 層 DNN のうち 2 層目のノードが少なくなる特徴的なネットワークの獲得を確認した．

表 1: 実験結果

モデル	MAE [day]
1 層 NN	2.346
2 層 DNN	2.264
3 層 DNN	2.050
4 層 DNN	2.013
5 層 DNN	2.099

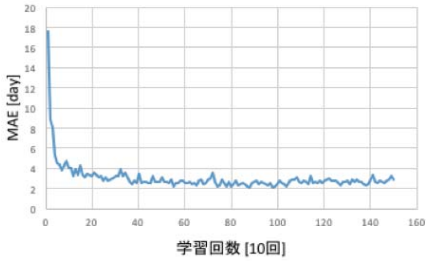


図 2: 3 層 DNN における平均 MAE-学習回数の関係

7. 考察

表 1 より, 中間層が 2 ～ 5 層の DNN による平均 MAE は全てが中間層数 1 の NN の平均 MAE より優れた結果を得た．

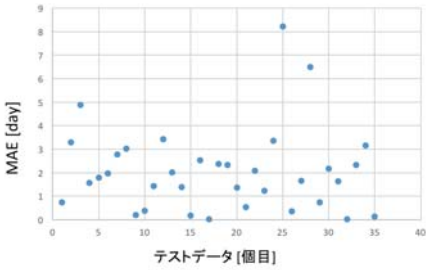


図 3: 3 層 DNN における学習回数 980 回時の各テストデータ MAE

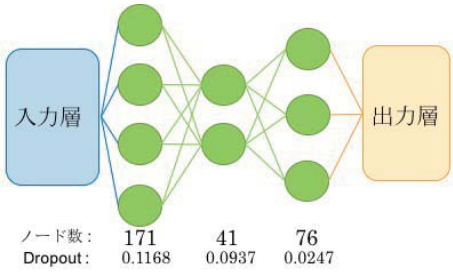


図 4: 3 層 DNN における特徴的な中間層構造

このことは提案した DNN の設計手法が局所最適解や勾配消失の問題を堅実に回避することができることを示す．また 4 層 DNN による MAE は 5 層 DNN による 平均 MAE より優れていた．このことは 5 層 DNN のチューニングが不十分であることを示す．この問題は, 中間層数の増加による, 学習時の計算コストの増大に起因すると考えられる．計算コストの増大はハイパーパラメータ探索に多大な時間を要し, 時間的制約により最適解を見つける可能性を低める．これを解消するためには, 本実験より多くの計算機リソースの使用や実験の効率化が必要である．

8. 結論

本研究では，深層学習を用いた収穫日予測手法としての DNN モデルと, 深層化されない NN モデルの予測精度を比較し, 深層学習を用いた収穫日予測手法がより優れた精度を持つことを確認した．今後の課題として，学習を多数の高性能な計算機リソース上で分散実行させ, 予測精度がより高いモデルを探索することがあげられる．

参考文献

[1] 麻生 英樹・安田 宗樹・前田 新一・岡野原 大輔・岡谷 貴之・久保 陽太郎・ボレガラ ダヌシカ (2015)『深層学習-Deep Learning-』, 近代科学社

[2] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang. ”MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems”. arXiv preprint arXiv:1512.01274, 2017.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

LRP 法による畳み込みニューラルネットワークにおける中間層の振る舞いの可視化

Visualizing the Behavior of the Inner Layers of Convolutional Neural Networks by Layer-wise Relevance Propagation

酒井博貴 *1 Hirotaka Sakai　亀谷由隆 *1 Yoshitaka Kameya　曽田尚宏 *2 Takahiro Sota　有江浩明 *2 Hiroaki Arie

*1名城大学 理工学部 情報工学科
Department of Information Engineering, Meijo University

*2株式会社デンソー 先端研究 4 部
Advanced Research and Innovation Div. 4, DENSO CORPORATION

In recent years, complex machine learning models like deep neural networks play a central role in many real applications, due to their high predictive performance. Interpreting machine learning models is then considered to be important since practitioners constantly need clues for improvement on such complex models, whose behavior is not directly visible to human. In this paper, we focus on the inner workings of convolutional neural networks, visualize them by a method called layer-wise relevance propagation, and report several findings from the visualization.

1. はじめに

深層学習技術の進展により，画像分類等のタスクでは人の能力を超える予測・認識結果を得られるようになり，現実場面でも深層学習が取り入れられ始めている．その一方で，予測・認識結果を解釈・説明することが困難であることが問題となってきており，予測・認識結果の説明責任が問われる医療等のミッションクリティカルな分野への応用の妨げとなっている．予測・認識結果の解釈や説明が可能になれば，学習モデルの改善，責任所在の特定，学習モデルからの知見の獲得が可能になり，より多くの分野への応用が容易になると考えられる．

人工知能分野ではこれまでもニューラルネットワーク（neural network, 以下 NN）等の学習モデルや出力結果を解釈・説明する手法が提案されてきたが [Hotho 03, Duch 04, Yuan 11], 上述の背景によりその進展が著しくなっており，特に畳み込み NN (convolutional NN, 以下 CNN）に対する解釈・説明手法は数多く提案されている [Grün 16, Selvaraju 17]．本研究では，画像分類タスクにおける CNN の中間層の振る舞いに着目し，入力画像内で中間層の出力に強く関連する箇所の可視化を試みる．そのため，学習モデルの入出力のみを解釈・説明する手法 [Ribeiro 16, Lundberg 17] ではなく，CNN の構造に沿って出力への関連度を逆伝播する LRP (layer-wise relevance propagation) 法 [Bach 15] を利用する．

2. LRP 法

d 次元入力ベクトル x に対して値 $f(x)$ を出力する学習モデルがあったときに，LRP 法では各入力要素 x_i $(1 \le i \le d)$ が $f(x)$ に関連する度合い（関連度, relevance）R_i を求める．関連度 R_i は入力ベクトル x 中の重要箇所を示しており，特に x の各要素が画像のピクセルに対応する場合は R_i の大きさに基づいてヒートマップを作成することで画像中の重要箇所をハイライトできる．

LRP 法は NN に限らず利用可能な手法であるが [Bach 15], ここでは図 1 の全結合 NN に対して LRP 法が動作する様子を説明する．NN における通常の予測では，図 1 (a) のように，第 l 層の素子 i の活性値 x_i と第 $l+1$ 層の素子 j への重み w_{ij} から $z_{ij} = x_i w_{ij}$ を計算する．そして第 $l+1$ 層の素子 j の活性値は $x_j = \sigma(\sum_i z_{ij} + b_j)$ として計算される．ここで，b_j はバイアス項，σ は（非線形の）活性化関数である．一方，LRP 法では入力 x と出力 $f(x)$ の対が与えられたときに，図 1 (b) のように関連度を出力層から入力層に向かって逆伝播させる．例えば，simple-LRP と呼ばれる簡単な伝播規則では

$$R_i^{(l)} = \sum_j R_{i\leftarrow j}^{(l,l+1)}, \quad R_{i\leftarrow j}^{(l,l+1)} = \frac{z_{ij}}{z_j} R_j^{(l+1)} \tag{1}$$

のように第 l 層の関連度 $R_i^{(l)}$ が第 $l+1$ 層の素子 j 間の関連度 $R_j^{(l+1)}$ に基づき計算される．ただし，$z_j = \sum_i z_{ij}$ である．入力要素 x_i の関連度 R_i は $R_i^{(1)}$ と定義する．更に，LRP 法では ϵ-LRP（式 2）や $\alpha\beta$-LRP（式 3）という伝播規則も導入されている．

$$R_{i\leftarrow j}^{(l,l+1)} = \frac{z_{ij}}{z_j + \epsilon \cdot \mathrm{sign}(z_j)} R_j^{(l+1)} \tag{2}$$

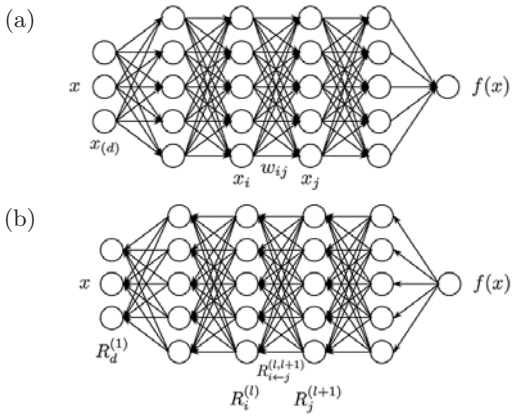


図 1: (a) 通常の予測における活性値の順伝播，(b) LRP 法における関連度の逆伝播（[Bach 15] より引用）

連絡先: 亀谷由隆，名城大学理工学部情報工学科，468-8502 愛知県名古屋市天白区塩釜口 1-501，ykameya@meijo-u.ac.jp

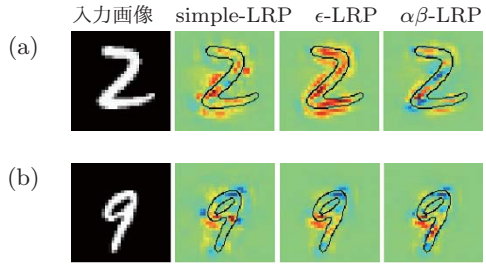


図 2: (a) 入力画像「2」と予測出力 2（正解）を与えた場合，(b) 入力画像「9」と仮想出力 4（誤り）を与えた場合

$$R_{i\leftarrow j}^{(l,l+1)} = \left(\alpha\cdot\frac{z_{ij}^{+}}{z_{j}^{+}} + \beta\cdot\frac{z_{ij}^{-}}{z_{j}^{-}}\right) R_{j}^{(l+1)} \qquad (3)$$

ϵ-LRP は z_j が小さな場合に式 1 が数値的に不安定になるのを避けることができる．また，$\alpha\beta$-LRP において，$z_{ij}^{+} = \max\{z_{ij}, 0\}$, $z_{ij}^{-} = \min\{z_{ij}, 0\}$, $z_{j}^{+} = \sum_i z_{ij}^{+}$, $z_{j}^{-} = \sum_i z_{ij}^{-}$ であり，出力との不整合を意味する負の関連度を多く伝搬するよう α, β を設定できる．ただし，$\alpha+\beta = 1$ という制約を満たさなければならない．なお，CNN においてはプーリング層や入力層について特別な伝播規則を用いるが [Samek 17, Montavon 18]，ここでは詳細を省略する．これらの伝播規則は各層の間で関連度の総和を保存する（すなわち $\sum_i R_i = \cdots = \sum_i R_i^{(l)} = \sum_j R_j^{(l+1)} = \cdots = f(x)$）ように定義されており，出力 $f(x)$ は入力要素の関連度 R_i の和として近似的に分解される．この意味で LRP 法は，LIME 法 [Ribeiro 16] と同様に，NN には手を加えずに近似モデルで特定入力に対する出力を事後解釈 (post-hoc interpretation) [Lipton 16] する手法であると言え，予測モデルと解釈・説明モデルを同時に学習する手法 [Lei 16, Al-Shedivat 17] とは対照的である．

CNN に対する LRP 法の実行例として，MNIST データセット中の手書き文字画像を入力として与えた場合を図 2 に示す．図 2 (a) では，入力画像「2」と正解の予測出力 2 を LRP 法に与えたときに得られる関連度のヒートマップが伝播規則ごとに示されている ($\epsilon = 1$, $\alpha = 3$, $\beta = 2$)．本論文で示すヒートマップでは赤色が関連度が強い箇所，青色が関連度の低い箇所を表す．図 2 (a) を見ると「2」という数字を認識するのに重要な箇所が関連度が高くなっている．理論的裏付けがあるのは $\alpha = 1, \beta = 0$ の $\alpha\beta$-LRP であるが [*1]，著者らが試した範囲では ϵ-LRP が全般的に見やすいヒートマップを出力したため，以降の実行結果では $\epsilon = 1$ の ϵ-LRP で得られたものを示す．また，LRP 法では実際の予測とは異なる仮想的な出力を正解として与えることもできる．図 2 (b) は入力画像「9」に対して誤りの仮想出力 4 を与えた場合の関連度のヒートマップである．これを見ると，出力 4 が正解として与えられたこと整合しないと思われる箇所が青色で示されている．

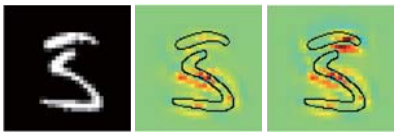


図 3: CNN が正しく判断できなかった場合

*1 [Montavon 17] に一部の伝播規則で得られる関連度が出力 $f(x)$ の一次テイラー展開の結果であることが示され，[Montavon 18] に伝播規則の使い分けのガイドラインが記述されている．

表 1: MNIST 用 CNN の構成

層番号	種類	入力次元 H	入力次元 W	入力次元 D	出力次元 H	出力次元 W	出力次元 D	カーネル	ストライド
0	入力	32	32	1	32	32	1	—	—
1	畳み込み	32	32	1	28	28	10	5×5	1×1
2	プーリング	28	28	10	14	14	10	2×2	2×2
3	畳み込み	14	14	10	10	10	25	5×5	1×1
4	プーリング	10	10	25	5	5	25	2×2	2×2
5	畳み込み	5	5	25	2	2	100	4×4	1×1
6	プーリング	2	2	100	1	1	100	2×2	2×2
7	畳み込み	1	1	100	1	1	10	1×1	1×1
8	Flatten	1	1	10		10		—	—

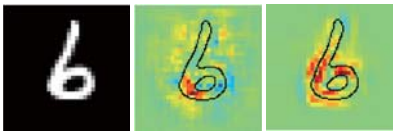


図 4: 入力画像「6」と関連度のヒートマップ

この性質を使うと予測に失敗した理由が分かることがある．例えば図 3 の入力画像は人間にも「3」か「5」の判断が難しいが，CNN は 3 と予測した．その予測出力 3 に対するヒートマップは左のように得られ，確かに青色の部分が少ない．一方，5 を正解出力として与えると右のヒートマップが得られる．上部の横線が下と繋がっていない箇所，上部の横線が丸みを帯びている箇所が青く，ここが問題だと判断されたことが分かる．

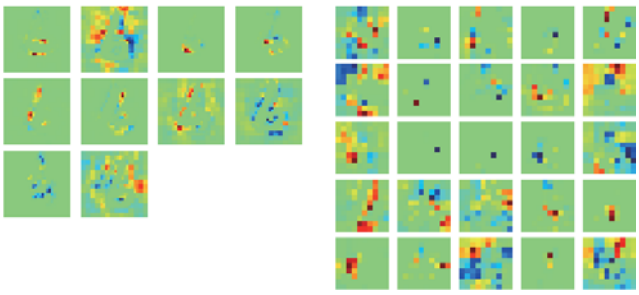
3. 中間層の振る舞いの可視化

前節の LRP 法の実行例では入出力の対を与えて入力と出力の対応関係を可視化した．更に本研究では，CNN における中間層の振る舞いを可視化するために，入力と出力の対応関係に限らない形で LRP 法を利用する．対象としたデータセットは MNIST と CIFAR-10 である．

3.1 MNIST データセットを用いた実験

MNIST データセットに対して LeNet-5 を参考に構成された表 1 の CNN を動かす．畳み込み層では活性化関数に ReLU を用い，プーリング層では平均値プーリングを行っている．Flatten 層では直前の 3 次元情報を 1 次元に落とす．実験では LRP 法の研究グループによる Python 独自実装版 [*2] を用いた．

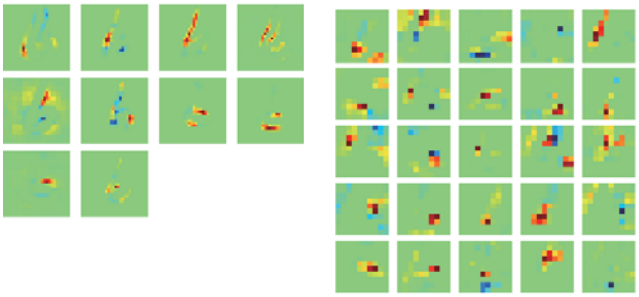
この実験では表 1 の CNN の中間層における特徴マップを関連度でヒートマップ化して入力の特徴がどのように抽出さ



(a) 1–2 層間　(b) 3–4 層間

図 5: 低精度モデルに画像「6」を入力したときの特徴マップ

*2 https://github.com/sebastian-lapuschkin/lrp_toolbox



(a) 1–2 層間　(b) 3–4 層間

図 6: 高精度モデルに画像「6」を入力したときの特徴マップ

れているか観察する．用いる関連度は第 2 層（プーリング層）から第 1 層（畳み込み層）へ渡されるもの，第 4 層（プーリング層）から第 3 層（畳み込み層）へ渡されるものである．また学習の効果を見るために，十分に学習を行った高い予測精度（正答率）99.23 %の CNN と学習回数を減少させた低い予測精度 41.94 %の CNN の 2 種類で比較する．

図 4 は入力画像「6」と出力 6 を与えたときの入力層でのヒートマップ（左：低精度，右：高精度）である．これを見ると高精度モデルの方がより数字「6」の特徴を捉えていると言える．また，同じ入力画像を与えたときの低精度モデルと高精度モデルの特徴マップは各々図 5, 6 のようにヒートマップ化される．まず低精度モデルの特徴マップでは，正解出力を与えたにも関わらず青色の箇所が多く，数字「6」の特徴がうまく捉えられていないように見える（図 5）．一方，高精度モデルの入力側に近い特徴マップ（図 6 (a)）では入力のエッジを捉えており，入力から離れた特徴マップ（図 6 (b)）では数字「6」の特徴を更に細かく捉えている．このように特徴マップを観察することで，学習がうまく行われたかどうかをある程度判断できるのではないかと考えられる．

3.2 CIFAR-10 データセットを用いた実験

CIFAR-10 データセットに対して表 1 の構成の CNN を動かす．予測精度は 69.92 %である．畳み込み層では活性化関数に ReLU を用い，プーリング層では最大値プーリングを行っている．第 9 層と第 10 層の全結合層では活性化関数として ReLU を用いるが，第 11 層では恒等関数を用いる．また，学習時には第 3 層と第 5 層の畳み込み層でドロップアウトを各々 0.9, 0.8 と設定した．実験では LRP 法の研究グループによる TensorFlow 実装版 [*3] を用いた．

この実験では，CNN の全結合層・出力層は使わずに，畳み込み層およびプーリング層によって特徴がうまく抽出できているかどうかを確認する．具体的には，第 8 層（プーリング層）と第 9 層（全結合層）の間で伝播される関連度を人為的に変更してそこから入力層に向かって関連度を逆伝播する．関連度の変更においては第 8 層中の 300 素子から無作為に素子 i $(0 \leq i \leq 299)$ を 1 つ選択し，素子 i の関連度を 1 に，他の 299 素子の関連度を 0 にする．この操作を行うと入力層で得られるヒートマップは素子 i と強く関連する入力要素をハイライトすることになる．また入力画像は，テスト用画像 10,000 枚の中から無作為に選び，元画像そのまま，左右反転した画像，色相をランダムに変化させた画像の 3 種類を与える．

図 7 は馬の画像に対して，関連度を 1 に変更した素子と入力画像の加工方法を変えたときの入力層におけるヒートマップ

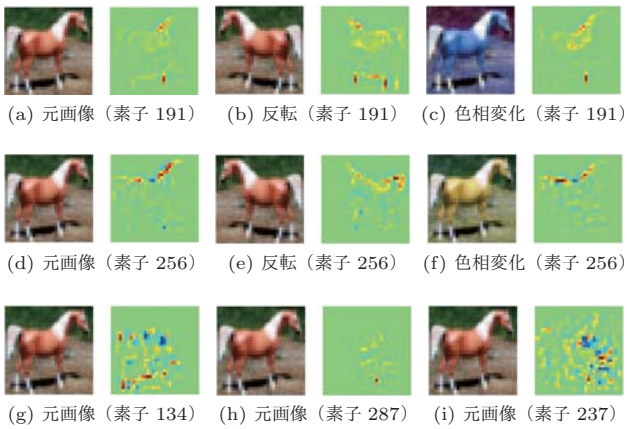
*3 https://github.com/VigneshSrinivasan10/interprettensor

表 2: CIFAR-10 用 CNN の構成

層番号	種類	入力次元 H	入力次元 W	入力次元 D	出力次元 H	出力次元 W	出力次元 D	カーネル	ストライド
0	入力	32	32	3	32	32	3	—	—
1	畳み込み	32	32	3	30	30	300	3×3	1×1
2	プーリング	30	30	300	15	15	300	2×2	2×2
3	畳み込み	15	15	300	14	14	300	2×2	1×1
4	プーリング	14	14	300	7	7	300	2×2	2×2
5	畳み込み	7	7	300	5	5	300	3×3	1×1
6	プーリング	5	5	300	3	3	300	2×2	2×2
7	畳み込み	3	3	300	2	2	300	2×2	1×1
8	プーリング	2	2	300	1	1	300	2×2	2×2
9	全結合	1	1	300		300		—	—
10	全結合		300			100		—	—
11	全結合		100			10		—	—
12	Softmax		10			10		—	—

である．191 番目の素子の関連度を 1 に変更した場合，(b)(c) のように入力画像を加工しても (a) と変わらず馬の胴体・首と足の関連度が高いことから，この素子は胴体と足を特徴として捉えている素子であると考えられる．左右反転しても入力画像に対応して胴体や足の部分が捉えられており，使用している CNN は左右反転に対応できるよう学習できていることが分かる．また，256 番目の素子の関連度を 1 に変更した場合，(d)~(f) のように馬の背中の関連度が高くなっていることから，背中を特徴として捉えている素子であると分かる．一方で，(g)~(i) のようにヒートマップが全体的に乱れた素子や，ヒートマップ中に高い関連度を示す箇所が見られずに入力画像のどこを特徴として捉えているか不明な素子も存在した．

図 8 は，同じクラスの異なる入力画像に対して，特定の素子の関連度を 1 に変更したときのヒートマップである．まず自動車クラスの画像に対するヒートマップ (a)~(c) において，(a) 素子 101 では自動車のタイヤに対して高い関連度は見られないが，(c) 素子 252 ではタイヤに該当する箇所の関連度が高いことが分かる．また，素子 252 と素子 295 はどちらもタイヤ部分を捉えていることが多い．同様に馬クラスの画像に対するヒートマップ (d)~(f) において，素子 191 と 256 は図 7 でも注目した素子である．また，図 8 の入力画像の中には図 7 と異なり，足や胴体が画像中にない画像も含めている点に注意する．図 7 の説明において素子 256 は馬の背中を特徴として捉えていると解釈していたが，他の画像では足部分の関連度が



(a) 元画像（素子 191）　(b) 反転（素子 191）　(c) 色相変化（素子 191）

(d) 元画像（素子 256）　(e) 反転（素子 256）　(f) 色相変化（素子 256）

(g) 元画像（素子 134）　(h) 元画像（素子 287）　(i) 元画像（素子 237）

図 7: 画像加工時の特定素子から逆伝播した関連度の違い



(a) 自動車，素子 101 (b) 自動車，素子 252 (c) 自動車，素子 295

(d) 馬，素子 84 (e) 馬，素子 191 (f) 馬，素子 256

図 8: 同じクラスにおける同じ素子の振舞い

高い．一方，素子 191 はやはり馬の胴体と足を特徴として捉えている．興味深いのは素子 191 が胴体や足がない馬の画像においては顔の輪郭を捉えている点である．NN の素子と入力における特徴の間には多対多の（複雑な）関係が成り立っているものと予想される．

最後に，図 9 は，素子 10 の関連度の値を 1 に変更したときの猫と自動車のクラスから選択した入力画像とヒートマップである．この場合，対象（猫や自動車）が存在する箇所への関連度は低く，対象以外の部分に関連度が見られる．これらのことから，素子 10 は猫や自動車の特徴を捉えていない素子なのではないかと考えられる．他のクラスに興味がない場合，このような素子は削除して計算コストを削減することも考えられる．

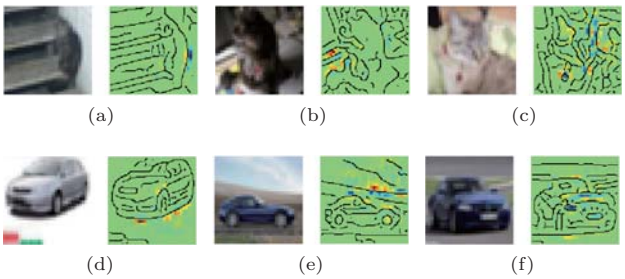
4. おわりに

LRP 法による CNN の中間層の振舞いをヒートマップとして可視化して分析した．中間層の素子が入力のどの部分を特徴として捉えているのか一定の傾向を把握できた．その一方で，CIFAR-10 については実装上の理由により 70 %弱の予測精度を持つ CNN でしか実験を行うことができなかった．現時点では予測精度 90 %を大きく超える CNN が存在しており，改めて検証実験を行う必要がある．また，学習モデルの解釈・説明ができたとしても，そこから学習モデルの改善へはまだギャップがあり [*4]，引き続き他のデータセットでも実験を行い，知見を増やしたい．また，LRP 法では得られたヒートマップの定量的な評価手法が提案されており [Samek 17, Montavon 18]，本実験で行った主観的評価とこれらの評価手法の対応関係を確認する必要がある．

参考文献

[Al-Shedivat 17] Al-Shedivat, M., Dubey, A., and Xing, E. P.: The Intriguing Properties of Model Explanations, in *Proc. of NIPS-17 Symp. on Interpretable Machine Learning* (2017)

*4 [Zeiler 14] では CNN の可視化の結果からストライド幅等を調整して予測精度向上につなげられたとの報告がある．



(a) (b) (c)

(d) (e) (f)

図 9: 素子 10 の関連度を 1 としたときのヒートマップ

[Bach 15] Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W.: On Pixel-wise Explanations for Non-linear Classifier Decisions by Layer-wise Relevance Propagation, *PLOS ONE*, Vol. 10, No. 7 (2015)

[Duch 04] Duch, W., Setiono, R., and Zurada, J. M.: Computational Intelligence Methods for Rule-based Data Understanding, *Proc. of the IEEE*, Vol. 92, No. 5, pp. 771–805 (2004)

[Grün 16] Grün, F., Rupprech, C., Navab, N., and Tombari, F.: A Taxonomy and Library for Visualizing Learned Features in Convolutional Neural Networks, in *Proc. of ICML-16 Workshop on Visualization for Deep Learning* (2016)

[Hotho 03] Hotho, A., Staab, S., and Stumme, G.: Explaining Text Clustering Results using Semantic Structures, in *Proc. of PKDD-03*, pp. 217–228 (2003)

[Lei 16] Lei, T., Barzilay, R., and Jaakkola, T.: Rationalizing Neural Predictions, in *Proc. of EMNLP-16*, pp. 107–117 (2016)

[Lipton 16] Lipton, Z. C.: The Mythos of Model Interpretability, in *Proc. of ICML-16 Workshop on Human Interpretability in Machine Learning* (2016)

[Lundberg 17] Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in *Proc. of NIPS-17* (2017)

[Montavon 17] Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Müller, K.-R.: Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition, *Pattern Recognition*, Vol. 65, pp. 211–222 (2017)

[Montavon 18] Montavon, G., Samek, W., and Müller, K.-R.: Methods for Interpreting and Understanding Deep Neural Networks, *Digital Signal Processing*, Vol. 73, pp. 1–15 (2018)

[Ribeiro 16] Ribeiro, M. T., Singh, S., and Guestrin, C.: Why Should I Trust You? — Explaining the Predictions of Any Classifier, in *Proc. of KDD-16*, pp. 1135–1144 (2016)

[Samek 17] Samek, W., Binder, A., Montavon, G., Lapuschkin, S., and Müller, K.-R.: Evaluating the Visualization of What a Deep Neural Network Has Learned, *IEEE Trans. on Neural Networks and Learning Systems*, Vol. 28, No. 11, pp. 2660–2673 (2017)

[Selvaraju 17] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, in *Proc. of ICCV-17*, pp. 618–626 (2017)

[Yuan 11] Yuan, C., Lim, H., and Lu, T.-C.: Most Relevant Explanation in Bayesian Networks, *J. of Artificial Intelligence Research*, Vol. 42, pp. 309–352 (2011)

[Zeiler 14] Zeiler, M. D. and Fergus, R.: Visualizing and Understanding Convolutional Networks, in *Proc. of ECCV-14*, pp. 818–833 (2014)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Neural Programmer-Interpreters の拡張と四則演算を用いる文章問題の解答

Solving Four Arithmetic Operation Problems and Word Algebra Problems by Neural Programmer-Interpreters

勝俣 翔太 *1　井上 克巳 *1*2
Shota Katsumata　Katsumi Inoue

*1東京工業大学 情報理工学院
Tokyo Institute of Technology, School of Computing

*2国立情報学研究所
National Institute of Informatics

In this paper we extend the arithmetic operations of the Neural Programmer-Interpreters (NPI). NPI is a recurrent and compositional neural network that learns to represent and execute programs. First, we enable NPI to execute not only the addition that NPI was originally possible but also the other three arithmetic operations, i.e. subtraction, multiplication and division. Then, we extend NPI by allowing it to share subprograms between tasks for improving learning efficiency. Next, we solve word algebra problems for elementary school-level mathematics which can be solved by using four arithmetic operations. For this purpose, we develop a converter that converts word algebra problems into mathematical expressions. This neural network is based on the Sequence-to-Sequence model with the attention mechanism. Using this neural network and NPI, we solve the data sets of word algebra problems and show that the accuracy of our method is better than the other existing methods.

1. はじめに

Programming by example (PBE) は実行したいタスクの入出力例からコンピューターが自動的にタスクを達成するためのプログラムを推論する技術である [9]．Neural Programmer-Interpreters (NPI) とは Reed らによって提案されたプログラムの実行を学習できる PBE の一手法であり、ニューラルネットワークを用いる [10]．学習後の NPI は実行すべきタスクと環境を情報として受け取り，そのタスクを達成するためのプログラムシーケンスを出力する．NPI は再帰構造とニューラルネットワークを用いることにより、タスクの長さに対する強力な汎化性能を持ち、高い精度でプログラムを出力できる．

ニューラルネットワークを使ってコンピューターにプログラムを理解させる研究は近年大きな成果を見せている．Graves らは NTM を開発した [1]．これはチューリングマシンの読み書きなどの動作を微分できる行列計算として定義し、外部メモリと LSTM コントローラを用いて、簡単なプログラムの実行を学習する技術である．Shu らはテキストを変換するプログラムを変換前と変換後のテキストのペアから導出する NPBE を提案している [3]．類似する手法として、Parisotto らは NSPS を提案している [5]．

本研究ではまず、NPI が可能であった加算の筆算に加えて、減算、乗算、除算のそれぞれの筆算を NPI ができるようにタスクを定義する．その際、タスク間でサブプログラムを共有可能にさせる拡張を行い、学習効率を向上させる．これは [10] で今後の課題とされていたことのうちの 1 つである．次に、四則演算を使って解くことのできる、小学生レベルの算数の文章問題を解くことを考える．このために、文章問題を数式表現へと変換する変換器を作成する．この変換器と NPI を用いて文章問題のデータセットを解き、他の既存手法と比較する．本研究では他の既存手法と違い、答えを求めるとともに NPI のプログラムシーケンスも得られる．すなわち、自然言語表現からプログラムを出力することができる．

連絡先: 勝俣 翔太，東京工業大学，shota@il.c.titech.ac.jp

2. NPI

NPI の構造を簡潔に表した図を 1 に示す．NPI は環境からの情報、現在のプログラム ID、その引数を入力として受け取り、プログラムの終了フラグ、次のプログラム ID、その引数をステップごとに出力する．プログラム ID は整数であり、ID のリストはユーザーによって与えられ、引数は整数 3 つのタプルである．環境からの情報は環境によって可変だが、本研究では環境に存在する 4 つのポインターが何を指しているかを LSTM-core に与える．学習を行う際はこれら入出力のペアを訓練データとし、テスト時には各ステップにおける出力を次ステップの入力として用いる．NPI はこの入力と出力のステップを繰り返し行うことで出力系列を生成し、NPI 全体の出力とする．図 2 は我々が定義した減算タスクを NPI が実行している様子を表している．NPI はプログラムの実行手順をステップごとに区切って学習することで、出力すべきプログラム長に対する強力な汎化性能を持つ．Cai らはこの汎化性能を完全にする改良を行なっており、本研究ではこの手法に基づいて拡張を行う [8]．

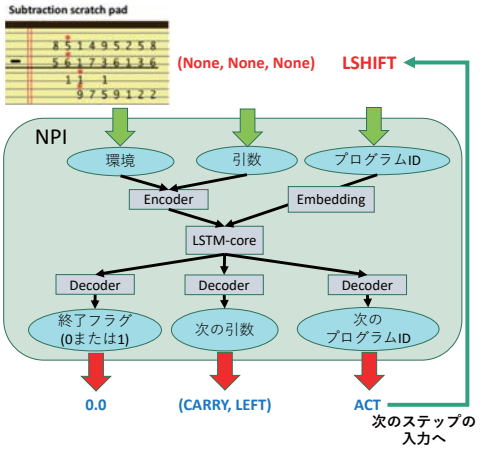


図 1: NPI の構造

3. NPI の拡張

3.1 モデル

NPI を四則演算に対応させるための拡張を行った．その際、環境の種類とエンコーダーを統一することでサブプログラムをタスク間で共通することを可能としている．環境には 4 つのポインターが存在し、NPI は各ポインターを移動させたり、ポインターの位置に任意の数字を書くことができる．また、そのポインターを通じて NPI に環境を観測させる．図 2 は拡張した NPI の減算における動作の様子を示している．

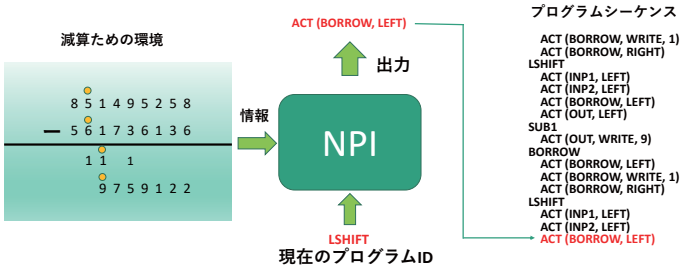


図 2: NPI の減算実行時の様子

3.2 タスクの定義について

NPI を四則演算に対応させるためには、環境とプログラムリストを定義する必要がある．加算については [10] と同じものであるが、減算については、桁上げに代わって桁借りを組み込む．これらを行うかどうかも NPI は学習することができ、桁上げの必要がない場合に 0 を桁上げしたりすることはない．以下では乗算と除算について説明する．

3.2.1 乗算

環境は図 3 の通りである．乗算タスクでは通常の筆算と同じように，大きく分けて 2 つのステップで処理を行う．まず、図 3 の左のように，最初に 1 桁同士の乗算部分を行う．次に，図 3 の右のように前ステップで出力された数字を合計していく．

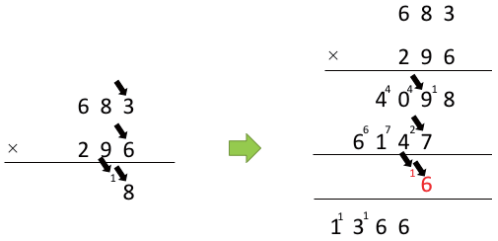


図 3: 乗算プログラム実行時の環境 (一部簡略化している)

3.3 除算

本研究では，乗算と減算を繰り返し実行することによって除算を実行する．その様子を図 4 に表す．図の上部では乗算，および減算の様子を表している．初めに，商を出力するべき欄の最左端に 9 を立てる．その後、出力した数字、割る数、割られる数とで乗算と減算を行う．その結果が負であれば乗算と減算の結果を消去し、次は 9 から 1 を引いて 8 を立てる，というようにして商を探索していく．図の下部では、立てた商が正しかった場合に次の桁へと移る処理の様子を表している．

3.4 サブプログラムの共有

本研究ではサブプログラムの共有によって学習効率を向上させている。図 5 は NPI が実際に出力する加算と減算のプログラムシーケンスの一部である。図中にある 2 つの LSHIFT は全く同じ動作をするプログラムである。このようなプログラムに同一のプログラム ID を与え、学習状況を共有させた。このような全てのプログラムに対して共有を行なった。

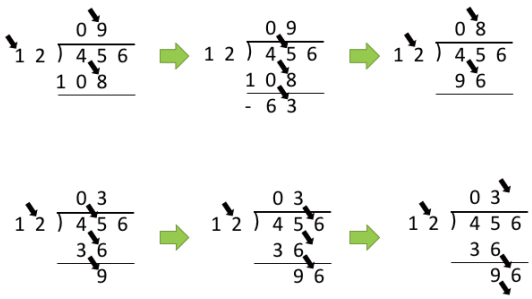


図 4: 除算プログラム実行時の環境 (一部簡略化している)

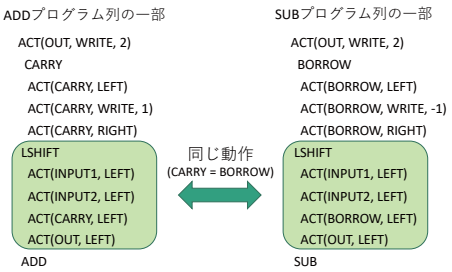


図 5: 加算と減算のプログラムシーケンスの一部

4. 文章問題への対応

4.1 概要

ここで扱う文章問題とは “Luke was trying to expand his game collection. He bought 4 games from a friend and bought 5 more at a garage sale. How many good games did he end up with?” のような，四則演算のみを 1 回または 2 回使った数式表現に変換できるものを指す．これをアテンションメカニズム [4] を組み込んだ Sequence-to-Sequence モデル [7] を基にした変換器を使って数式表現へと変換し (上の例であれば「4 + 5」)，それを NPI を使って解き，プログラムシーケンスを出力することを目的とする．一連の流れを図 6 に、変換器の図を 7 に示す．

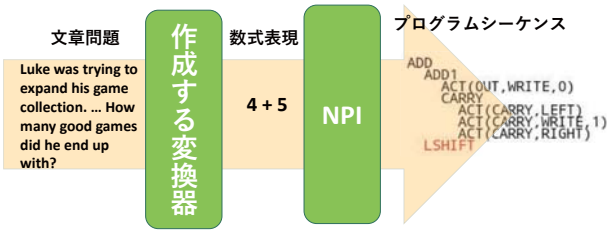


図 6: 文章問題からプログラムシーケンスまでの流れ

4.1.1 エンコーダー

エンコーダーは，アテンションレイヤーとデコーダーに、入力された文章を固定長ベクトルにして渡す．ここでは、双方向 LSTM (BiLSTM) を使用している．まず，入力単語列

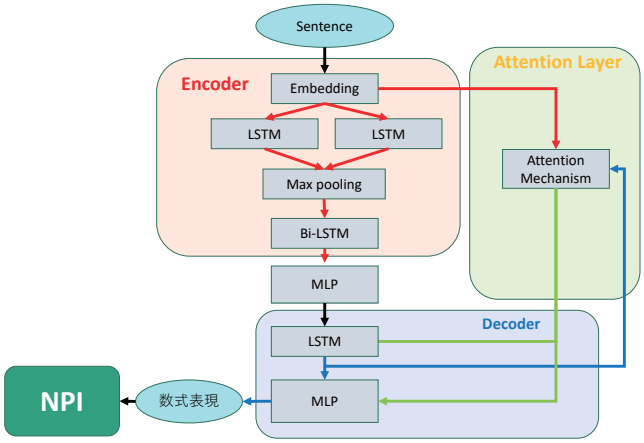


図 7: 変換器の構造

$w_1, w_2, ..., w_L$ (単語 ID) それぞれに対して学習可能な行列を使って固定長ベクトル $e(i_k)$ へと変換する．次に，それらを最初からと最後から、別々の LSTM を使って，最初からの文脈表現 $l(i_k)$ と，最後からの文脈表現 $r(i_k)$ へと式 (1) のように変換する．ここで f_l および f_r は LSTM の計算を表している．得られた $l(i_k)$ と $r(i_k)$ を一つの固定長ベクトル c_k とするため，式 (2) のようにして c_k を計算する．ここで，W_e, b_e はパラメーターであり，[;] はベクトル同士の結合を表している．最後に Bi-LSTM を使って c_k をまとめてエンコードは終了する．

$$l(i_k) = f_l(l(i_{k-1}), e(i_k)), \quad r(i_k) = f_r(r(i_{k+1}), e(i_k)) \quad (1)$$

$$c_k = tanh(W_e[max(l(i_k), r(i_k)); e(i_k)] + b_e) \quad (2)$$

4.1.2 アテンションレイヤー

アテンションレイヤーの役割は，デコーダーが入力のどこに注目すればいいかという情報をデコーダーに渡すことである．はじめに，エンコーダーから受け取った情報 c_k とデコーダーの前回の出力 h_{t-1} を式 (3) のようにして変換し，a_k^{key} と a_t^{query} を得る．それらを結合し，e_{kt} を式 (4) のように計算する．ここで，$f_{key}, f_{query}, f_{score}$ は全て，全結合層の処理を表している．最後に e_{kt} を正規化し，α_{kt} を重みとして計算した c_k の合計，att_t を式 (5) のようにして計算する．

$$a_k^{key} = f_{key}(c_k), \ a_t^{query} = f_{query}(h_{t-1}) \quad (3)$$

$$e_{kt} = f_{score}(a_k^{key}, a_t^{query}) \quad (4)$$

$$\alpha_{kt} = \frac{exp(e_{kt})}{\sum_{i=1}^{k} exp(e_{it})}, \ att_t = \sum_{i=1}^{k} \alpha_{it} c_i \quad (5)$$

4.1.3 デコーダー

デコーダーは最終的な出力系列の計算を行う．まず，アテンションレイヤーにデコーダーの前回の出力を渡し，アテンションレイヤーからの出力 att_t を受け取る．デコーダーの前回の出力 (最初の出力はエンコーダーからの出力を使う) とアテンションレイヤーからの出力をデコーダー側の LSTM に渡し，現在の出力を計算する．それと，2 層の全結合層から構成される分類器によって最終的な出力を得る．これを決められた回数繰り返し，出力系列を得る．本論文では 3 回出力している．

5. 実験

この章では前章で述べた二つのモデルに対して行った実験の説明をする．実験環境として，CPU は Intel Core i7 4790K 4.00GHz，GPU は NVIDIA GeForce GTX 950，メモリは 16GB で OS は Windows8.1 を使用した．

5.1 NPI の拡張

5.1.1 データセット

[8] によって再帰化されたタスクを行う NPI は，入出力の長さに対して完全な汎化性能を持つ．よって各ステップにおいて，起こりうる全通りのポインターの値とその出力のペアを学習データとして用いる．

5.1.2 結果

上記の学習データを学習した NPI に対して評価を行う．テストデータは各 4 つのタスクについて，入力の桁数 20 までのそれぞれ 16 個，合計 1280 個のランダムな数値のペアから作った問題を使う．入力の桁数を大きく増やしてしまうとメモリ不足になってしまい，学習ができなかった．これらのテストデータ総数に対して，正しくプログラムシーケンスを出力できた割合を NPI の評価とする．その結果を表 1 に示す．

Method	加算	減算	乗算	除算
Reed ら 2016 [10]	97%	−	−	−
Cai ら 2017 [8]	100%	−	−	−
提案手法	100%	100%	100%	100%

表 1: NPI の四則演算のテスト精度

5.1.3 サブプログラム共有の効果

表 2 にサブプログラム共有を行なった場合と行なっていない場合の比較を示す。訓練データ数は NPI の各ステップにおける入力と出力のペアの数である．学習は精度が 100%に達するか 1500epoch を超えるまで続ける．

共有しない場合では訓練データ数やプログラムリストの長さが約 2 倍必要であった．また、共有しない場合では学習を続けても精度が 100%に到達しなかった．共有した場合では 700 〜800epoch の間に精度が 100%に到達することが多かった．

共有の有無	必要訓練データ数	プログラム数	精度
共有する	1,259,211 個	27 個	100%
共有しない	2,664,721 個	53 個	99.8%

表 2: サブプログラム共有の有無による性能の比較

5.2 文章題への対応

5.2.1 準備

データセットは Roy らと同じ 3 つを使用した [12].

データセット名	AI2	IL	CC
問題数	395	562	600
演算子の数	1 個または 2 個	1 個	2 個
演算子の種類	加算・減算	全て	全て
解答に関係のない数字	あり	なし	なし

表 3: データセットの特徴

比較対象となる既存手法について紹介する．Roy らによって提案された Solving General Arithmetic Word Problems (SGAWP)[12] は文章から，数字の計算順を示す配列と，計算式を表現する構文木を生成する技術である．GENERE[6]

は Bouchard らによって提案された学習データ生成モデルのパラメーターを自動的に獲得する手法である．[6] では，このGENERE と seq2seq を使って文章問題のデータセットを解いている．∂4 は Bošnjak らによって提案された Forth 言語を理解するモデルである [11]．[11] では加算，バブルソートに加えて，文章問題のデータセットを ∂4 を使って解いた．また，∂4 は AI2 データセットのような関係のない数字に対する処理を行なっておらず，計算に使う演算子も 2 個で固定されている．

5.2.2 結果

他の既存手法と比較した結果を表 4 に示す．[12] に従い，k-分割交差検証で評価する．AI2 は 3 分割，IL は 5 分割で行う．CC は [12] に従い、学習データ 300 個、バリデーションデータ 100 個、テストデータ 200 個で評価を表している．提案手法は AI2，IL において他の手法よりも精度で上回っている．しかし，CC に関しては GENERE が一番精度が良かった．

手法	AI2	IL	CC
数式表現のみ			
SGAWP [12]	82.4%	75.4%	55.5%
GENERE [6]	77.9%	56.7%	**98.5%**
seq2seq [6]	72.5%	53.7%	95.0%
解答まで			
∂4 [11]	−	−	96.0%
NPI プログラムシーケンス + 解答			
提案手法	**83.3%**	**85.1%**	97.4%

表 4: 既存手法と提案手法を比較した表

6. 考察

NPI の実験において，サブプログラムを共有して学習することは、今まで学習したことの一部を新しい学習に応用するという人間が実際に行うことをコンピューターに可能にさせているとも捉えられる．また、本研究では end-to-end 学習を行なっていない．しかし、これも人間と同様で、四則演算の計算を覚えてから文章問題に取り組むということをコンピューターに行なわせていると考えられる．コンピューターに段階的な学習をさせることの有用性は [13] でも述べられている．[13] では、あるタスクを学習させる際により簡単な訓練データから学習させることで精度が向上することを示している．

本研究ではアテンションメカニズムによって通常の seq2seq よりも性能を向上させている．例えば “Sam’s dog had puppies and 8 had spots . He gave 2 to his friends . He now has 6 puppies . How many puppies did he have to start with ?” という問題に対しては “has”や “gave” に注目していることがわかった．このような動詞から演算子の種類を判断することは非常に良いことであり、文章問題の読解にアテンションメカニズムが貢献していることがわかる．アテンションメカニズムの有効性を検証するために、提案手法のモデルからアテンションメカニズムを除いたモデルを作成し，性能の比較を行なった．その結果の表を表 5 に示す．表からわかるようにアテンションメカニズムは全てのケースにおいて良い結果を得られている．

アテンションメカニズムの存在	AI2	IL	CC
アテンションメカニズムあり	83.3%	85.1%	97.4%
アテンションメカニズムなし	71.0%	78.9%	95.9%

表 5: アテンションメカニズムの有無による性能の比較

7. まとめと今後の課題

本論文では NPI の四則演算への対応と [10] で課題として挙げられていたサブプログラム共有のための拡張を提案し、それぞれの拡張がうまく機能していることを示した．また、NPI を利用した小学生レベルの算数の文章問題のデータセットを解く手法を提案した．そして提案手法が他の先行研究よりも優れていること、アテンションメカニズムが今回使用した文章問題に有効であることを示した．

NPI の除算は商を立てる際の効率が悪い．NPI がニューラルネットワークを使用しているという利点を活かし、商を賢く立てられるようにするべきである．NPI は出力すべきプログラムの長さに対しては汎用性があるものの未知のタスクに対応できない．[2] は [1] 同様、外部メモリとコントローラーを組み合わせたモデルであり、様々なタスクに取り組んでいる．このような算術演算以外のタスクに対して NPI を適用し、プログラムを出力させるということも今後の課題としたい．

参考文献

[1] Alex Graves, Greg Wayne and Ivo Danihelka. Neural Turing Machine. *arXiv preprint arXiv:1410.5401*, 2014.

[2] Alex Graves, Greg Wayne, Malcolm Reynolds et al. Hybrid computing using a neural network with dynamic external memory. Nature, 538(7626):471-476, 2016.

[3] Chengxun Shu and Hongyu Zhang. Neural Programming by Example. In *Proc. AAAI*, pp.1539-1545, 2017.

[4] Dzmitry Bahdanaue, Kyunghyyun Cho and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv:1409.0473*, 2014.

[5] Emilio Parisotto, Adbel-rahman Mohamed, Rishabh Singh, Lihong Li, Dengyoung Zhou and Pushmeet Kohli. Neuro-Symbolic Program Synthesis. In *arXiv preprint arXiv:1611.01855*, 2016.

[6] Guillaume Bouchard, P Saito Stenetorp and Sebastian Riedel. Learning to Generate Textual Data. In *Proc. EMNLP*, pp.1608-1616, 2016.

[7] Ilya Sutskever, Oriol Vinyals and Quoc V. Le. Sequence to Sequence Learning with Neural Networks. In *Proc. NIPS*, pp.3104-3112, 2014.

[8] Jonathon Cai, Richard Shin and Dawn Song. Making Neural Programming Architectures Generalize via Recursion. In *arXiv preprint arXiv:1704.06611*, 2017.

[9] Lieberman Henry. Your wish is my command: Programming by example. Morgan Kaufmann. 2001.

[10] Reed Scott and Nando De Freitas. Neural programmer-interpreters. *arXiv preprint arXiv:1511.06279*, 2015.

[11] Sebastian Riedel, Matko Bošnjak and Tim Rocktäschel. Programming with a Differentiable Forth Interpreter. In *arXiv preprint arXiv:1605.06640*, 2017.

[12] Subhro Roy and Dan Roth. Solving General Arithmetic Word Problems. In *Proc. EMNLP*, pp.1743-1752, 2015.

[13] Yoshua Bengio, Jérôme Louradour, Ronan Collobert and Jason Weston. Curriculum Learning. In *Proc. ICML*, pp.41-48, 2009.

[2A4] [General Session] 2. Machine Learning

座長:椿 真史（産業技術総合研究所）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room A (4F Emerald Hall)

[2A4-01] Collecting whole sky images and classification of cloud genera and conditions

○Yu Morikawa[2], Haru Nakanishi[2], Naoki Inamura[1], Nobuaki Kondo[1], Hiroki Obuchi[3], Teruo Ohsawa[5], Takashi Matsubara[4], Kuniaki Uehara[4] （1. BANYAN PARTNERS Inc., 2. Kobe Digital Labo Inc., 3. SKY Perfect JSAT Corporation, 4. Graduate School of System Informatics, Kobe University, 5. Graduate School of Maritime Sciences, Kobe University）

5:20 PM - 5:40 PM

[2A4-02] Detecting community structure in layered neural networks for diagram recognition

○Chihiro Watanabe[1], Kaoru Hiramatsu[1], Kunio Kashino[1] （1. NTT Communication Science Laboratories）

5:40 PM - 6:00 PM

[2A4-03] A study on traffic flow prediction method using probe information in transportation system

○Keisuke Ohno[1], Eichi Takaya[1], Hiroshi Matsumoto[2], Tetsuo Morita[2], Satoshi Kurihara[1] （1. Graduate School of Science and Technology, Keio University, 2. Sumitomo Electric Industries, Ltd）

6:00 PM - 6:20 PM

[2A4-04] Analysis of Machine-Friendly Knowledge Representation toward More Efficient Knowledge Tracing

○Hiromi Nakagawa[1], Yusuke Iwasawa[1], Kaoru Nasuno[1], Yutaka Matsuo[1] （1. The University of Tokyo）

6:20 PM - 6:40 PM

[2A4-05] Highlighting Non-contributing Pixels for Visual Explanation of CNNs

○Koichi Ikeno[1], Satoshi Hara[1], Takashi Washio[1] （1. The Institute of Scientic and Industrial Research, Osaka University）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

全天球画像のデータ収集と 雲形と状態判定

Collecting whole sky images and classification of cloud genera and conditions

森川 優 *1 Yu MORIKAWA
中西 波瑠 *1 Haru NAKANISHI
稲村 直樹 *5 Naoki INAMURA
近藤 伸明 *5 Nobuaki KONDO
小渕 浩希 *4 Hiroki OBUCHI
大澤 輝夫 *2 Teruo OHSAWA
松原 崇 *3 Takashi MATSUBARA
上原 邦昭 *3 Kuniaki UEHARA

*1株式会社神戸デジタル・ラボ Kobe Digital Labo Inc
*2神戸大学大学院海事科学研究科 Graduate School of Maritime Sciences, Kobe University
*3神戸大学大学院 システム情報学研究科 Graduate School of System Informatics, Kobe University
*4スカパー JSAT 株式会社 SKY Perfect JSAT Corporation
*5バニヤン・パートナーズ株式会社 BANYAN PARTNERS Inc

Maritime meteorological observation is critical for a safe voyage, and general ships are required in Japan to report the observations to parties concerned. Since it is difficult to recognize the meteorological conditions for non-experts, the demand of automatic recognition arises. Many studies have tackled the classification of cloud genera and the regression of cloud cover. However, less attention has been paid for cloud conditions. Thus, we developed a machine learning system for classification of cloud conditions. We first developed a dedicated equipment for photographing whole sky images and collected data samples. Then, we tagged cloud genera and conditions in each cloud layer (high, middle, and low). Using the dataset, we built a deep convolutional neural network to classify the cloud genera and conditions via fine-tuning ResNet50. The network achieved accuracies higher than 0.9 for both cloud genera and conditions.

1. はじめに

現在、一般船舶では 3 時間毎に海上の気象観測を行い、「船舶気象観測表」を気象庁へ報告することが気象業務法で定められている。[5] 海上気象観測の報告はこれまで電報にて実施されていたが、気象測器の開発や通信インフラの発達によって自動測定およびデータ通信が可能となり、気象観測や報告の自動化が進められている。しかし、雲の観測だけは従来の気象測器での測定が困難であり、乗組員による目視で観測されているが正確性や人的負荷の問題が存在する。

このように雲の観測はまだ研究の途にあるが、雲の自動観測を行うにあたって解決しなければならない 3 つの課題が存在している。まず第 1 に、船上に設置して長期の観測に耐えるデバイスの開発が必要である。第 2 に、収集された画像を自動的に分析して雲形・状態や雲量を測定する手法の開発が必要である。また、近年の畳み込みニューラルネットワーク (CNN) の発達を見るに深層学習の利用が有望であるが、雲の観測ではまだ研究の域を出ていない。そもそも学習に用いることができる雲形のデータセットは存在しているが、日本の海上気象観測で求められている、雲の状態と雲量に関するラベル付けされたデータセットは存在していない。これが第 3 の課題である。本論文ではこれら 3 つの課題にアプローチする。

先行研究において、全天球画像における雲形に関しては高い識別率で識別できている [1, 2, 3]。しかし、状態に関しては先行事例がなく、また実際の現場では下層・中層・上層における雲形と状態を報告しなければならないため、先行事例において作成された学習器 [1, 2, 3] を今回のシステムに適応することはできない。このため、雲形と状態を同時に判定する学習器が必要である。

本研究では、オープンデータとして存在しない雲形のみではなく、状態までラベル付けされたデータセットの構築を行い、雲形と状態の学習器を CNN によって構築するアプローチを取る。

連絡先: 稲村 直樹，バニヤン・パートナーズ株式会社，〒 650-0031 神戸市中央区東町 116-2 オールドブライトビル 7F，078-335-5955，078-335-5431，inamura@banyanpartners.co.jp

2. 全天球画像のデータ収集

データセットを開発するにあたって、撮影デバイスが存在していないことから、撮影デバイスのプロトタイプ開発を行っている。図 1 は撮影デバイス本体の写真で、図 2 は実際に撮影デバイスを設置した船舶、図 3 は撮影デバイスを防水ケースに収納し船上に設置した際の写真である。



図 1: デバイス稼働実験時



図 2: 神戸大学大学院海事科学研究科所属・深江丸



図 3: 撮影デバイス設置時

当初は、デバイスとしてコダック アクションカメラ PIXPRO SP360 4K（以下アクションカメラ）単体で撮影を行う予定であった。しかし、アクションカメラ単体ではバッテリーの容量が一日の撮影を行うには足りず、その上、撮影時の熱により 30 分程度で撮影不可能になってしまう問題点が存在する。

そのため、アクションカメラにデスクトップ PC 用の水冷キットを接合することによって冷却し、バッテリーに 12 万 mAh の物を用意し接続をしている。また、デバイスの撮影タイミングやデバイスの稼働監視、カメラの再起動装置等の制御を Raspberry Pi 3 Model B で行っている。結果として、図 1 のような大きなデバイスとなってしまっている。図 2 の深江丸に搭載することを考慮すると、潮風や風雨による防水加工されていない部分へのダメージが大きいことは明白である。このため、図 3 のように防水加工されたケースを改造し、デバイスを収納して撮影を行っている。定期的な保守等は必要ではあるが、安定した撮影を行うことに成功している。

全天球画像収集のため、撮影デバイスを用いて構築した自動撮影システムを、神戸大学大学院海事科学研究科所属の深江丸に搭載して 1 分毎の撮影を行っている。1 分毎の写真だけでは写真の際の部分の雲や、重なった部分の雲が判定できないため、10 分毎に 1 度判定することによりラベル付けを行っている。また、夜間は人間の目では判定できないため、ラベル付けを行っていない。

ラベルとしては、世界気象機関 (WMO) により定められた 10 基本形と国際式天気記号 [5] における状態のラベル付けを行っている。雲の層ごとに雲形と状態のラベル付けを行ったため、一つの画像に対してラベルは 6 種類存在している。表 1 は層ごとに出現する雲形と状態のラベルの範囲（0 から 9 まで）を示している。

表 1: 雲形と状態

層	雲形	状態
下層	積雲	0〜9
	層積雲	
	積乱雲	
	層雲	
中層	高積雲	0〜9
	乱層雲	
	高層雲	
上層	巻雲	0〜9
	巻積雲	
	巻層雲	

今回は深江丸において撮影できなかった、もしくは枚数が少なすぎる雲形と状態が存在したため、学習の際はこれらを除外し表 2 と表 3 の雲形と状態を対象にして学習を行っている。

表 2 と表 3 を見ると、雲形と状態の枚数が揃っていない状況にある。この理由としては、撮影した範囲が限られていることや、季節による雲の出現の違いが理由としてあげられる。撮影開始の時期が秋頃からだったため、夏場によく現れる積乱雲や乱層雲を撮影できなかったことや、上層の雲が出現しやすい海域での撮影が出来ていないことにより、枚数に偏りが出てしまっている。

表 2: 雲形

層	雲形	図	枚数
下層	積雲	図 4a	388 枚
	層積雲	図 4b	120 枚
	層雲	図 4c	67 枚
中層	高積雲	図 5a	257 枚
	高層雲	図 5b	336 枚
上層	巻雲	図 6a	118 枚
	巻積雲	図 6b	91 枚
	巻層雲	図 6c	73 枚

表 3: 状態

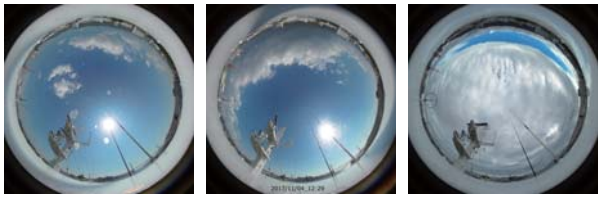
層	状態	図	枚数
下層	0		640 枚
	1	図 4d	332 枚
	4	図 4e	116 枚
	7	図 4f	123 枚
中層	0		257 枚
	1	図 5c	134 枚
	2	図 5d	219 枚
	3	図 5e	113 枚
	7	図 5f	89 枚
上層	0		933 枚
	1	図 6d	86 枚
	2	図 6e	24 枚
	7	図 6f	30 枚
	8	図 6g	24 枚
	9	図 6h	91 枚

雲形に「なし」がない理由としては、先に状態を識別してから雲形を識別する方法をとったため、雲形の「なし」を学習させていない。このような方法を取った理由としては、2 回判定することにより整合が取れない場合（例：下層に雲が存在するにも関わらず雲形はなしと識別）を回避するためにある。結果として、計 6 つの学習器を作成し、識別システムの実装を試みている。

図 4〜図 6 では実際に撮影し、ラベル付けされた雲の画像の例を示している。雲ごとに十分な差異があることが明らかであるが、図 5c と図 5d のように人間の目では区別がつかない雲も存在している。また 3 層分の雲を判別することから別の層の雲が重なり合う事例が発生するため、ラベル付けの作業はさらに困難である。このように、雲自体の判別が観測未経験者では困難である上、雲の層ごとの高さの判別はさらに困難であるため、専門に教育された人材が必要である。

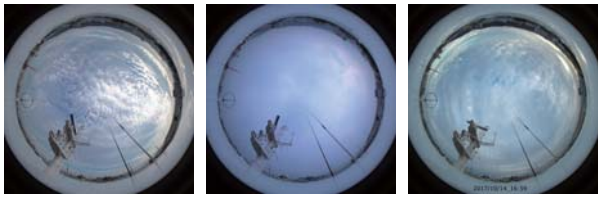


(a) 積雲　(b) 層積雲　(c) 層雲



(d) 下層状態 1　(e) 下層状態 2　(f) 下層状態 7

図 4: 下層雲形・状態

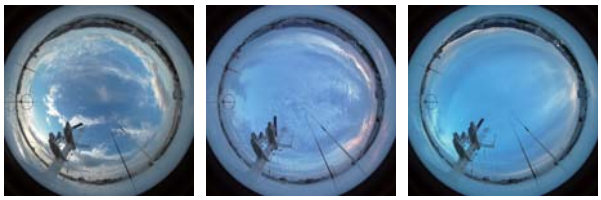


(a) 高積雲　(b) 高層雲　(c) 中層状態 1

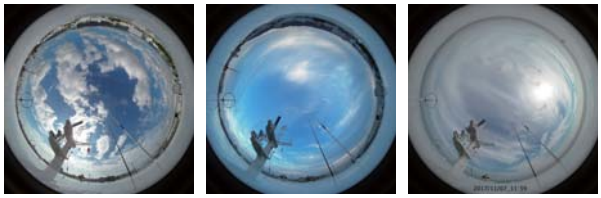


(d) 中層状態 2　(e) 中層状態 3　(f) 中層状態 7

図 5: 中層雲形・状態



(a) 巻雲　(b) 巻積雲　(c) 巻層雲



(d) 上層状態 1　(e) 上層状態 2　(f) 上層状態 7



(g) 上層状態 8　(h) 上層状態 9

図 6: 上層雲形・状態

3. 学習環境と実験結果

Deeplearning のライブラリは TensorFlow-gpu(1.4.1)[4] と Keras(2.0.6) を用いている。また、既存の学習モデルである ResNet50 を用いて finetuning を行っている。49 層目まで固定し、全結合層のみをデータに対応した形に変形し、学習を行っている。学習時は 5-fold の cross-validation により accuracy を計算している。学習率は 0.0002 から始めて、200 epoch 時に 0.0001 になるよう、指数関数にしたがって減少させながら学習を行っている。勾配法は確率的勾配降下法を用いて momentum を 0.9 に設定している。さらに、Keras の Data augmentation の機能を使って訓練データの変形を行っている。

- データ拡張は横・縦方向に幅に対して 0〜0.1 の間で変形
- 反時計回りシアー変換を 0〜0.2 π の間で変形
- 画像中心に対してズームを 0.6〜1.4 の間で変形
- 画像を-180〜180 の間でランダムに回転
- RGB の値を-0.2〜0.2 の間でランダムに変動
- 水平・垂直方向にランダムで反転

以下の図 7 と図 8 は 1 epoch ごとのテストデータに対する精度である。

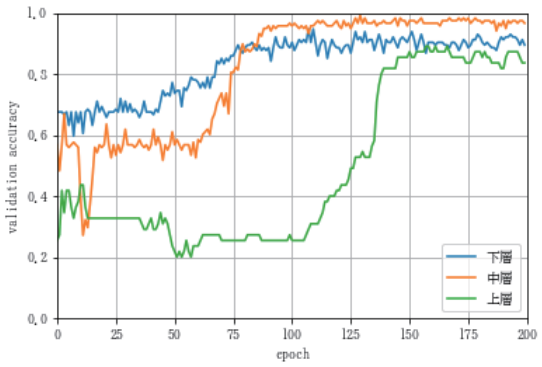


図 7: 雲形

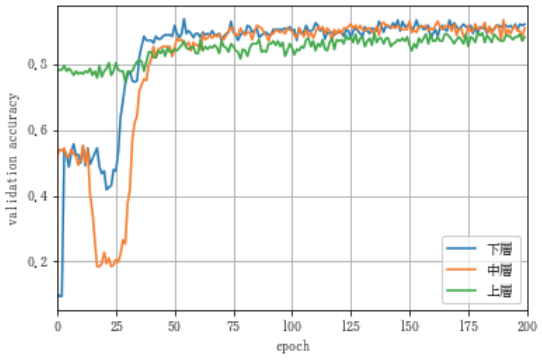
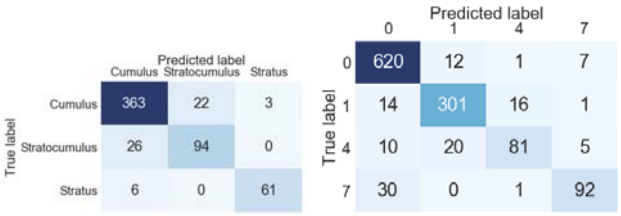


図 8: 状態

図 7 と図 8 では、雲形・雲の状態共に上層以外は 0.9 を超える高い accuracy を示している。まだすべてのラベルの全天球画像が集まっていない問題はあるが、先行研究において雲形は高い精度であることを考慮すると、全ラベルの場合は若干

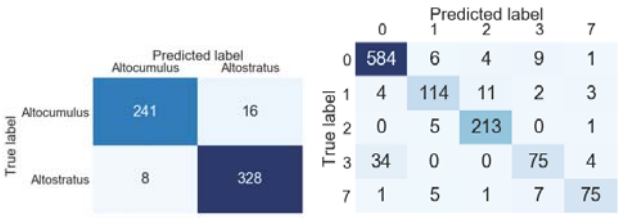
分類しにくい程度で収まる可能性を示している。この事実は、今後、訓練データが完全な形で収集できれば、チューニング次第では、通常航海においても実用に足る学習器の生成が可能であることも示している。

実運用では未知のデータを扱うことから、図 9〜図 11 では過学習の度合いを示す指標である、validation loss が最も低い場合におけるテストデータに対する accuracy の confusion-matrix を示している。下層・中層の精度は上層よりも高い結果となった。表 2 や表 3 を見ると分かるが、各データの偏りが大きく、特に上層は枚数も少ない上に大きなデータの偏りが生じていることが大きな原因であると考えられる。上層の枚数が少ない理由としては、下層と中層の雲に遮られることで収集画像に写り込まないことである。このことから、今後も収集できる上層の枚数は大幅に増えない可能性があるため、収集量を増やす方法が必要となる。



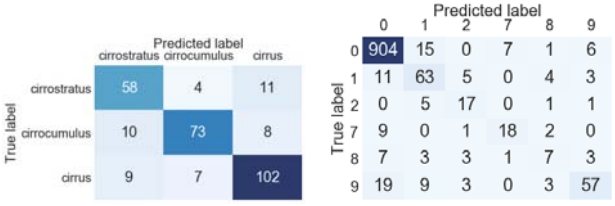
(a) 下層雲形　　(b) 下層状態

図 9: validation loss の最も低い学習器における下層の validation accuracy



(a) 中層雲形　　(b) 中層状態

図 10: validation loss の最も低い学習器における中層の validation accuracy



(a) 上層雲形　　(b) 上層状態

図 11: validation loss の最も低い学習器における上層の validation accuracy

4. 結論と今後の課題

本論文では、ラベルは不完全ではあったが、収集した雲の全天球画像データにラベル付けを行って構築したデータセットにて、雲形と状態の両方の識別において高い精度を出すことに成功した。これにより現状のデータ収集方法と学習方法に問題がないことを示すことができた。また、先行研究では取り組まれていなかった雲の状態においても識別が可能であることを示すことができた。しかし、学習精度の向上や一般船舶にて実運用を行うために解決しなければならない課題が多数存在する。

まず第 1 に、データ収集のための撮影は深江丸と沖縄県伊是名島の 2 箇所で実施しているが、網羅的なデータセット構築のために観測地点を増やす必要がある。また、データ収集量の増大に合わせて、効率的なラベル付け作業の実行方法も同時に検討が必要である。

第 2 に、全天球画像の特性上、1 枚の画像だけでは雲の高さの判別が困難である。気象測器として雲底高度計があるが測定距離の限界と直上地点のみという制限がある。正確なラベル付けおよび学習精度向上のために、全天球画像の中でどの範囲にどの高さの雲が存在するかを判別する方法の検討が必要である。

第 3 に、乗組員が雲の観測をする際には前回の観測結果を参考に時系列に測定を行なっている。今回の学習では、1 枚の画像だけで学習を行っているが、前後の画像も併せて学習することで、雲の流れによる雲の高度や雲形・状態の変化を検証する必要がある。

第 4 に、本研究では全天球画像をそのまま利用しているが、画像のセンター部分やエッジ部分などをマスクしたり、目視観測により近い矩形画像に変換して学習させた場合の全天球画像との精度比較を検証する必要がある。

5. 謝辞

田中謙次氏にコーディングの作業の実地や学習方法について討論を行って頂いた。兵庫県立大学大学院応用情報科学研究科申吉浩教授に GPU マシンの貸出や学習方法について指摘を頂いた。深江丸での観測の実地は、矢野吉治船長、青山克己一等航海士をはじめとする乗組員の皆様にご協力頂いた。なお、本研究は、国土交通省の「先進船舶・造船技術研究開発費補助事業（先進船舶技術研究開発）」の支援対象事業における「海上気象観測の自動観測・自動送信システムの開発 *1」の一環として行われたものである。

参考文献

[1] Jinrui Gan, Weitao Lu, “Cloud Type Classification of Total-Sky Images UsingDuplex Norm-Bounded Sparse Coding” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017.

[2] A. Kazantzidis, P. Tzoumanikas, “Cloud detection and classification with the use of whole-skyground-based images” Atmospheric Research, vol.113 pp.80–88, 2012

[3] A. Heinle, A. Macke, “Automatic cloud classification of whole sky images”, Atmos. Meas. Tech., vol.3, p557–567, 2010

[4] Martn Abadi et al.,“TensorFlow: A System for Large-Scale Machine Learning”, 2016

[5] “第 5 章　雲の観測”. 船舶気象観測指針, 改定第 7 版 (2013), 気 象 庁,　https://marine.kishou.go.jp/brochure/guide-jp.pdf,(参照 2018-03-06)

*1　株式会社商船三井、古野電気株式会社、スカパー JSAT 株式会社を主体とする産学官連携の共同研究

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

図形認識のための多層ニューラルネットにおける大局構造の抽出

Detecting community structure in layered neural networks for diagram recognition

渡邊千紘 *1 Chihiro Watanabe　平松薫 *1 Kaoru Hiramatsu　柏野邦夫 *1 Kunio Kashino

*1NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

Layered neural networks (LNNs) have realized high recognition performance for various real datasets, however, it is difficult for human beings to understand their training results. Conventionally, we have proposed network analysis methods for extracting simplified structure of a trained LNN, by detecting communities of units based on the similarity of connection patterns. In this work, we propose a new method for representing the community structure in a LNN, by using connection weights between pairs of communities. By experiment using the dataset of diagram recognition, we show that our new method provides clues for interpreting the roles of each community in a LNN, in terms of which community in input-side adjacent layer is the most important for it in prediction.

1. はじめに

多層ニューラルネットは高次元空間に存在する様々な実データにおいて，高い予測性能を実現している．しかし，その内部表現は非常に多くのパラメータが非線形かつ階層的に組み合わせられた形で構成され，人が理解することは難しい．我々はこれまで，データから学習された多層ニューラルネットに対し，ネットワーク解析を適用することにより，各層において隣接する層のユニットと似た結合パターンを持つユニットのグループ (コミュニティ) を推定する方法を提案してきた [渡邊 17a, Watanabe 18, Watanabe 17b, Watanabe 17c]．これにより，学習結果のネットワーク構造を単純化し，大局的に捉えた表現を獲得することが可能となり，この結果から，ニューラルネットの各部分が推論において果たす役割を推察することができるようになった．これらの研究においては，ニューラルネットから抽出された各コミュニティ間に存在する複数の結合を，閾値処理に基づいて 1 本の結合束として表現することにより，構造の単純化を行っていたが，コミュニティ間がどの程度強く結びついているかを知る方法は存在していなかった．

本研究では，多層ニューラルネットから抽出されたコミュニティ構造に対し，新たな結合束の定義法と可視化手法を適用することにより，各コミュニティが入力側の層におけるどのコミュニティと最も強く結びついているかを知ることを可能にした．また，実際に図形認識のデータセットを用いて学習した多層ニューラルネットに対し，提案法を適用することにより，各コミュニティの役割について考察を行った．

2. ニューラルネットのコミュニティ構造抽出

確率的最急降下法に基づき，データから多層ニューラルネットの学習を行う．この際，LASSO[Ishikawa 90, Tibshirani 94] と呼ばれる手法を適用することにより，冗長な結合重みの値をゼロに近づけ，疎なネットワークを獲得することができる．詳細なニューラルネットの学習方法は，既存研究 [渡邊 17a] と同じである．ただし，実験では，学習を安定させるために，各反復において学習データを確率的に選ぶのではなく，各クラスの学習データを 1 つずつ順に選ぶこととした．

連絡先: watanabe.chihiro@lab.ntt.co.jp

データから学習された多層ニューラルネットにおいて，結合重みの符号を区別した上で，似た結合パターンを持つユニットを分類するコミュニティ抽出法として，既存研究 [渡邊 17a] が存在する．これは，確率モデルに基づくコミュニティ抽出法 [Newman 07] を，多層ニューラルネット向けに拡張した手法であり，本研究でもこの手法 [渡邊 17a] を用いてニューラルネットの各層におけるコミュニティ構造を推定する．

多層ニューラルネットの各層において，隣接する層との間の結合関係を 4 つの隣接行列 $A^+ = \{A^+_{i,k}\}$, $A^- = \{A^-_{i,k}\}$, $B^+ = \{B^+_{k,j}\}$, $B^- = \{B^-_{k,j}\}$ で定義する．ただし，$k = 1, \cdots, l$. ここで，$A^+(A^-)$ の各要素 $A^+_{i,k}(A^-_{i,k})$ は，入力側の隣接層における i 番目のユニットと，各層における k 番目のユニットとの間にある結合重みが ξ 以上 ($-\xi$ 以下) ならば 1, そうでなければ 0 として定義する．同様に，隣接行列 $B^+(B^-)$ を，各ユニットと，出力側の隣接の層における各ユニットとの間の結合重みから定義する．

各層で，コミュニティcにユニットが属する確率を $\pi = \{\pi_c\}$, コミュニティcのユニットに入力される正 (負) の結合の結合元が入力側の隣接層における i 番目のユニットである確率を $\tau^+ = \{\tau^+_{c,i}\}(\tau^- = \{\tau^-_{c,i}\})$, コミュニティ$c$のユニットから出力される正 (負) の結合の結合先が出力側の隣接層における j 番目のユニットである確率を $\tau'^+ = \{\tau'^+_{c,j}\}(\tau'^- = \{\tau'^-_{c,j}\})$ とおく．

隣接行列 A^+, A^-, B^+, B^- の尤度を最大化するパラメータ $\pi, \tau^+, \tau^-, \tau'^+, \tau'^-$ は，以下の更新式を繰り返し計算することにより，局所解として求めることができる (EM 法)．ここで，$q_{k,c}$ は k 番目のユニットがコミュニティcに属する確率を表し，k 番目のユニットが属するコミュニティは，EM 法の最終反復において $q_{k,c}$ を最大化するコミュニティcとして定義する．

$$q_{k,c} = \Big\{\pi_c \Big[\prod_i \Big(\tau^+_{c,i}\Big)^{A^+_{i,k}} \Big(\tau^-_{c,i}\Big)^{A^-_{i,k}}\Big] \Big[\prod_j \Big(\tau'^+_{c,j}\Big)^{B^+_{k,j}} \Big(\tau'^-_{c,j}\Big)^{B^-_{k,j}}\Big]\Big\} \Big/ \Big\{\sum_s \pi_s \Big[\prod_i \Big(\tau^+_{s,i}\Big)^{A^+_{i,k}} \Big(\tau^-_{s,i}\Big)^{A^-_{i,k}}\Big] \Big[\prod_j \Big(\tau'^+_{s,j}\Big)^{B^+_{k,j}} \Big(\tau'^-_{s,j}\Big)^{B^-_{k,j}}\Big]\Big\}, \quad \pi_c = \frac{\sum_k q_{k,c}}{l},$$

$$\tau^{+}_{c,i} = \frac{\sum_k q_{k,c} A^{+}_{i,k}}{\sum_{k,i} q_{k,c} A^{+}_{i,k}}, \quad \tau^{-}_{c,i} = \frac{\sum_k q_{k,c} A^{-}_{i,k}}{\sum_{k,i} q_{k,c} A^{-}_{i,k}},$$
$$\tau'^{+}_{c,j} = \frac{\sum_k q_{k,c} B^{+}_{k,j}}{\sum_{k,j} q_{k,c} B^{+}_{k,j}}, \quad \tau'^{-}_{c,j} = \frac{\sum_k q_{k,c} B^{-}_{k,j}}{\sum_{k,j} q_{k,c} B^{-}_{k,j}}.$$

3. 結合重みの値に基づく大局構造の定義

多層ニューラルネットから推定された各コミュニティ間に存在する複数の結合を，以下に示す方法に基づいて 1 本の結合束として表現する．まず，入力側の層のコミュニティc と出力側の層のコミュニティc' に対し，その間にある全ての結合について重みの絶対値の平均 $r_{c,c'}$ を結合束の重みとし，重み $r_{c,c'}$ に比例した太さの線で結合束を描画するものとする．ただし，結合束の太さは全ての層において最小値, 最大値が同じになるように正規化した．次に，結合束を疎にするために，各層間において，r_{c_1,c_2} が $\min_{c'} \max_c r_{c,c'}$ 未満の場合はコミュニティc_1, c_2 間の結合束を削除するものとし，それ以外の場合はコミュニティc_1, c_2 間の結合束があるものとする．また，実験では，さらに視認性を向上するために，各出力側のコミュニティc' に対し，$r_{c,c'}$ の値を降順に並べたとき，1 番目 (2 番目) のコミュニティc からの結合束を青色 (黄土色) で描画した．

4. 実験

実際にデータセットから学習されたニューラルネットに対し，提案法を適用することにより，ニューラルネットの大局構造を可視化し，各コミュニティが推論において果たす役割について考察を行う．本実験では，入力画像と，その中に写っている図形の種類 (クラス) が組になった図形認識のためのデータセットを生成し，ニューラルネットの学習に用いた．このデータセットは 10 クラスの図形が描かれた 15×15 画素の画像からなり，各クラス 1000 個ずつの学習データを含む．各画像は，それぞれ決められた平均と分散を持つ分布からランダムに生成された点を直線で結ぶことにより二値画像を生成した後，全ての画素に対し平均 0, 標準偏差 0.1 の正規分布から生成した雑音を加えることにより生成した．ここで，各図形において生成する点の x 座標，y 座標における平均値と，直線で結ぶ点の組み合わせを以下のように設定した．ただし，各画像における x 座標, y 座標の最大値を 1, 最小値を 0 とおいた．

- 長方形 (Rectangle)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.2, 0.8)$, 点 c: $(0.8, 0.8)$, 点 d: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (c, d), (d, a) とした．
- ハート型 (Heart): $\cdots$ 点 a: $(0.1, 0.5)$, 点 b: $(0.3, 0.8)$, 点 c: $(0.5, 0.6)$, 点 d: $(0.7, 0.8)$, 点 e: $(0.9, 0.5)$, 点 f: $(0.5, 0.2)$ とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (c, d), (d, e) (e, f), (f, a) とした．
- 逆三角形 (Triangle)$\cdots$ 点 a: $(0.5, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.2, 0.8)$ とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (c, a) とした．
- バツ型 (Cross)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.2, 0.8)$, 点 d: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを (a, b), (c, d) とした．
- 斜め線 (Line)$\cdots$ 点 a: $(0.2, 0.8)$, 点 b: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを (a, b) とした．
- ダイヤ型 (Diamond)$\cdots$ 点 a: $(0.5, 0.9)$, 点 b: $(0.9, 0.5)$, 点 c: $(0.5, 0.1)$, 点 d: $(0.1, 0.5)$ とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (c, d), (d, a) とした．
- 左向き矢印 (Arrow)$\cdots$ 点 a: $(0.4, 0.9)$, 点 b: $(0.1, 0.5)$, 点 c: $(0.4, 0.1)$, 点 d: $(0.9, 0.5)$ とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (b, d) とした．
- リボン型 (Ribbon)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.8, 0.2)$, 点 d: $(0.2, 0.8)$) とし，直線で結ぶ点の組み合わせを (a, b), (b, c), (c, d), (d, a) とした．
- 顔型 (Face)$\cdots$ 点 a: $(0.3, 0.8)$, 点 b: $(0.3, 0.6)$, 点 c: $(0.7, 0.8)$, 点 d: $(0.7, 0.6)$, 点 e: $(0.2, 0.3)$, 点 f: $(0.8, 0.3)$ とし，直線で結ぶ点の組み合わせを (a, b), (c, d), (e, f) とした．
- 2 本の水平線 (Two lines)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.2)$, 点 c: $(0.2, 0.8)$, 点 d: $(0.8, 0.8)$ とし，直線で結ぶ点の組み合わせを (a, b), (c, d) とした．

全ての図形について，上記の平均値を持ち標準偏差が 0.07 の正規分布から各点の座標を生成した．図 1 に学習データにおける各画像の例を，また図 3 に学習データにおける各画像の平均値を示した．

上記のデータセットを用いて多層ニューラルネットを学習し，コミュニティ抽出を行った結果を図 4 に示した．また，コミュニティ間に存在する複数の結合の重みから，結合束の定義を行うことにより，大局構造を可視化した結果を図 5 に示した．また，入力層における各コミュニティに属する画素を図 6 に示した．

ここで，入力データは最大値が 1, 最小値が -1, 出力データは最大値が 0.99, 最小値が 0.01 になるように正規化して用いた．また，ニューラルネットの学習における反復数を各学習データごとに 100 回ずつとし，t 回目の反復での学習率 $\eta(t)$ を $\eta(t) = 0.7 \times 10^6/(10^6 + 5 \times t)$ とし，LASSO のハイパーパラメータ λ を 5×10^{-7} とした．コミュニティ抽出においては，ニューラルネットの結合重みの閾値 ξ を 0.3 とし，各層のコミュニティ数を 10 個とし，反復数を 1000 反復の EM アルゴリズムを 300 回繰り返し，最終反復における対数尤度の値が最大となる回の結果を用いた．

図 2 に，学習の各反復における図形の認識率の変化を示した．

5. 考察

図 5，図 6 より，図形認識を行うニューラルネットの構造について，以下のような考察が得られる．

最大重みの結合束を出力層から入力層に向かって辿ることにより，リボン型, 2 本の水平線は，主に Com 4 に含まれる画素の情報を用いて認識されるものと推察される．Com 4 に含まれる画素の部分を見ると，リボン型は白であり，2 本の水平線は黒であることが特徴となっている．また，長方形, ハート型, 逆三角形, 左向き矢印, バツ型, ダイヤ型, 顔型, 斜め線は，主に Com 2 に含まれる画素の情報を用いて認識されるものと推察される．この中でも，長方形，ハート型，逆三角形，左向き矢印，バツ型は，出力層の 1 つ手前の隠れ層における同じコミュニティ (図 6 の右から 4 番目のコミュニティ) に対して最大重みの結合束でつながっているため，このコミュニティに含まれるユニットが，これらの図形の識別に用いられていると考えられる．さらに，これらの図形と斜め線は，出力層の 2

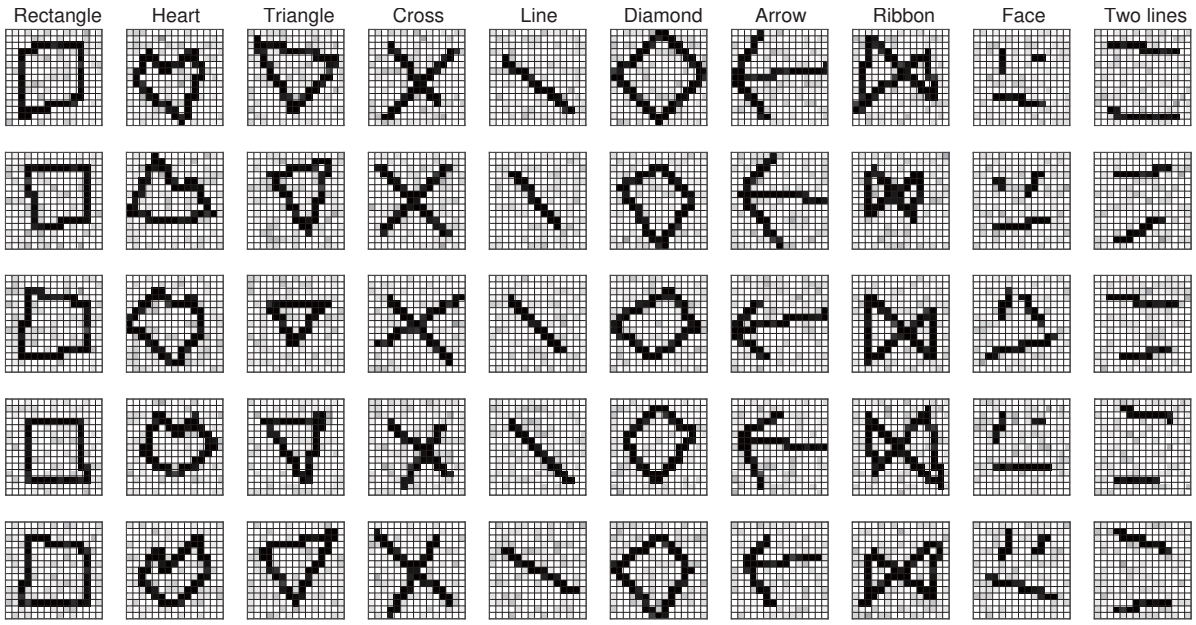


図 1: 学習データとして用いた図形画像の例.

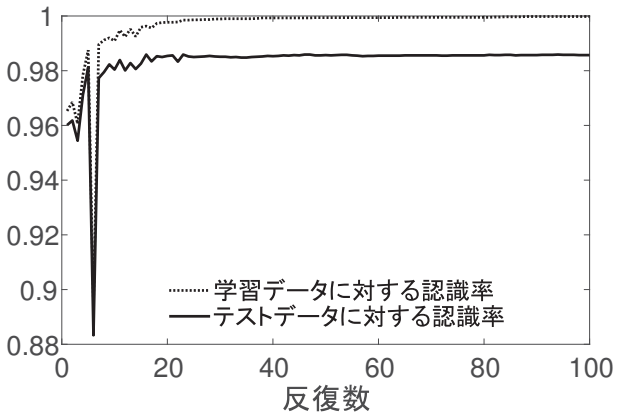


図 2: ニューラルネットの学習における認識率の変化.

つ手前の隠れ層における同じコミュニティ(図 6 の一番右のコミュニティ) に対して最大重みの結合束でつながっており，このコミュニティに含まれるユニットが，これらの図形と斜め線とを識別しているものと考えられる．入力層の Com 9 は入力画像の端の部分にあたり，隣接層への結合束を持たないことから，図形を認識する際には比較的重視されていないことが読み取れる．

6. 結論

本研究では，データから学習されたニューラルネットに対し，新たな大局構造の抽出法を適用することにより，各層におけるユニットのコミュニティが入力側のどのコミュニティと強く結びついているかを知ることを可能にした．また，実際にデータセットを用いて実験を行い，ニューラルネットにおける各部分が図形認識において果たす役割の考察を行った．ニューラルネットから抽出された各コミュニティが果たす役割について，更なる詳細な理解を可能とすることや，その結果をニューラルネットの構成の改善に応用することは，今後の課題である．

参考文献

[渡邊 17a] 渡邊千紘，平松薫，柏野邦夫．多層ニューラルネットにおける正負の結合重みに基づく大局構造抽出．情報科学技術フォーラム (FIT2017), 2017 年.

[Watanabe 18] Watanabe. C and Hiramatsu. K and Kashino. K. Modular Representation of Layered Neural Networks. Neural Networks. Vol. 97. 2018. pp. 62–73.

[Watanabe 17b] Watanabe. C and Hiramatsu. K and Kashino. K. Recursive Extraction of Modular Structure from Layered Neural Networks Using Variational Bayes Method. Proceedings of Discovery Science 2017, Lecture Notes in Computer Science. Vol. 10558. 2017. pp. 207–222.

[Watanabe 17c] Watanabe. C and Hiramatsu. K and Kashino. K. Modular Representation of Autoencoder Networks. Proceedings of 2017 IEEE Symposium on Deep Learning, 2017 IEEE Symposium Series on Computational Intelligence. 2017.

[Ishikawa 90] Ishikawa. M. A Structural Connectionist Learning Algorithm with Forgetting. Journal of Japanese Society for Artificial Intelligence. Vol. 5. 1990. pp. 595–603.

[Tibshirani 94] Tibshirani. R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B. Vol. 58. No. 1. 1994. pp. 267–288.

[Newman 07] Newman. M and Leicht. E. Mixture models and exploratory analysis in networks. Proceedings of the National Academy of Sciences. Vol. 104. No. 23. 2007. pp. 9564–9569.

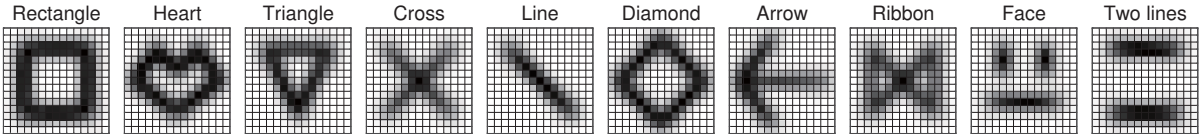


図 3: 学習データとして用いた図形画像の平均値.

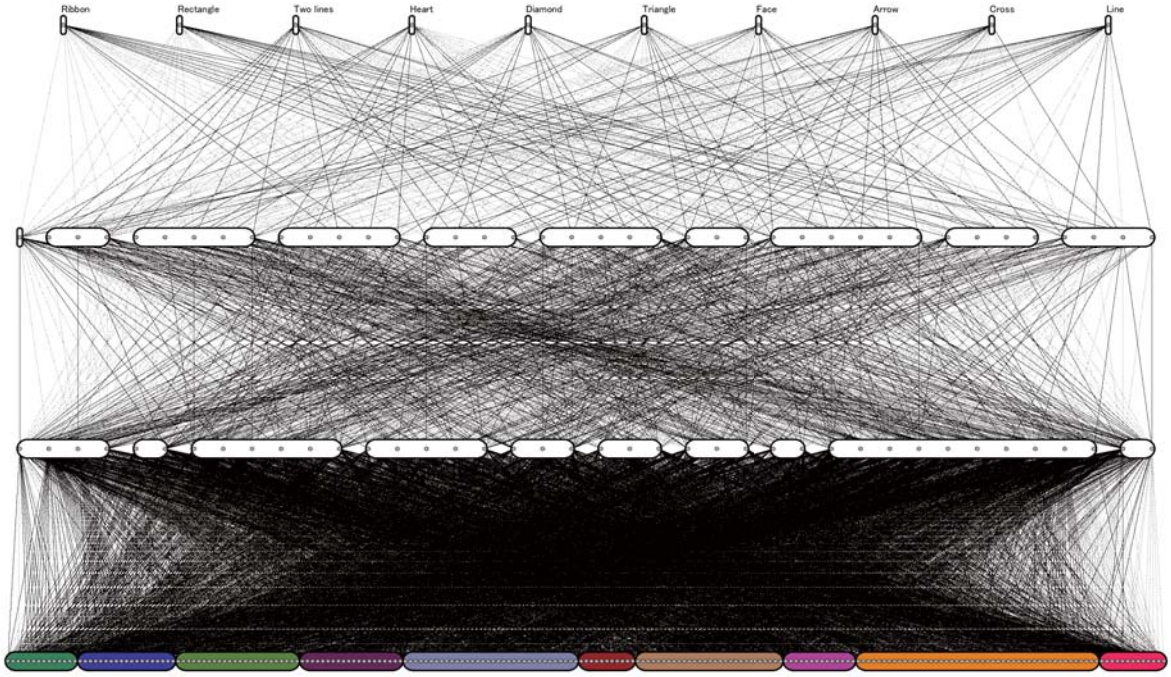


図 4: ニューラルネットから抽出されたコミュニティ構造．入力層の各ユニットは，入力データの画像における 1 画素に対応する.

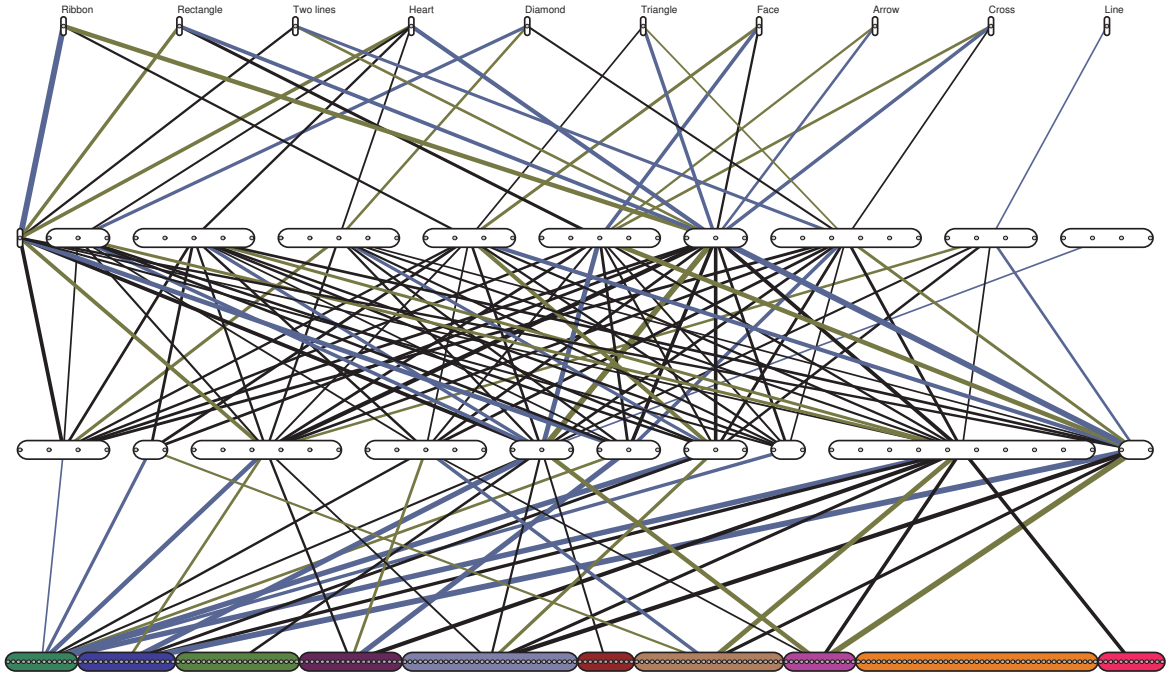


図 5: 各コミュニティ間に存在する複数の結合をまとめて 1 本の結合束として表現した大局構造.

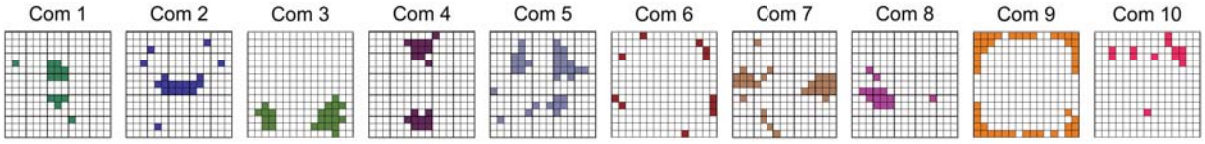


図 6: 入力層の各コミュニティに含まれるユニットと，入力データの画像における画素の対応関係.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

道路交通システムにおけるプローブ情報を用いた交通流予測手法の検討

A study on traffic flow prediction method using probe information in transportation system

大野啓介 *1 Keisuke OHNO　高屋英知 *1 Eichi TAKAYA　松本洋 *2 Hiroshi MATSUMOTO　森田哲郎 *2 Tesuo MORITA　栗原聡 *1 Satoshi KURIHARA

*1慶應義塾大学 大学院理工学研究科 Graduate School of Science and Technology, Keio University

*2住友電気工業株式会社 Sumitomo Electric Industries, Ltd

With the remarkable development of Intelligent Transportation System in recent years, it is possible to easily collect traffic information and various information of the vehicle. Probe information provides more extensive traffic information in addition to the observation information. In this paper, we consider the traffic flow prediction method on urban road using probe information. Accurate and real-time traffic information is indispensable for the deployment of high-performance intelligent transportation systems. Traffic flow is complicated, but by deep learning that can acquire feature quantities automatically, it is possible to express the characteristics of the traffic flow without the prior knowledge such as the characteristics of the site, and it is expected to improve the prediction accuracy. Therefore, in this research, we consider a traffic flow prediction model using deep learning. Also, we compared it with other traffic flow prediction method.

1. はじめに

近年の ITS(Intelligent Transport Systems:高度道路交通システム) の著しい発展により，道路交通情報や車両の様々な情報を容易に収集することが可能になっている．交通流を把握・予測するために収集される情報には，路上に設置されたセンサにより得られる観測情報と道路上を走行する車両から収集されるプローブ情報がある．プローブ情報によって，センサが設置された地点の観測情報より広範囲の道路交通情報が得られる．本研究では，プローブ情報を用いた交通流予測手法を検討する．高性能な高度道路交通システムの配備には正確でリアルタイムな道路交通情報が欠かせない．この情報によって，道路利用者がより良い経路判断をし，交通渋滞の緩和や排気ガスによる大気汚染の削減などにつながる可能性を秘めている．交通流予測の目的はこのような道路交通情報を提供するためである．これまでに，数多くの交通流予測モデルが提案されてきたが，扱う交通データの現場の特性を考慮したモデルがほとんどであり，この特性などの事前知識が必要である．近年，機械学習手法の一種である深層学習が学術的及び産業的に多くの関心を集めており [1]，画像認識や自然言語処理など様々なタスクにおいて深層学習の適用により精度が向上している [2][3][4][5]．交通流は複雑であるが，特徴量を自動的に獲得できる深層学習によって，現場の特性といった事前知識を把握していなくても交通流の特徴を表すことができ，交通流予測においても精度が向上すると期待される．

そこで，本研究では，プローブ情報を用いた深層学習による交通流予測手法を検討する．また，深層学習による交通流予測手法の妥当性を検証するために，従来の機械学習手法との比較を行う．

2. 関連研究

これまで提案されてきた予測手法では，扱うデータセットが観測情報かプローブ情報かに大きく分けられる．

連絡先: 慶應義塾大学 大学院理工学研究科 大野啓介
〒 223-8522 神奈川県横浜市港北区日吉３－１４－１
E-mail: k.o.lib65@gmail.com

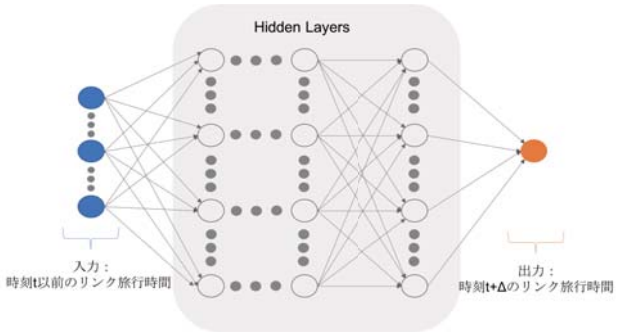


図 1: DNN の構成

観測情報を用いた研究として，Yi ら [6] は深層学習である Stacked Auto Encoder(SAE) を用いて交通流の予測を行い，Support Vector Machine(SVM) などの従来の機械学習手法によるモデルより精度が上がることを示した．しかし，交通流の指標として車の台数のみを扱っており，正確な交通流を把握できない．本研究では，正確な交通流を把握するために，交差点間の通行にかかる時間であるリンク旅行時間を用いる．

プローブ情報を用いた研究として，Zheng ら [7] は隠れ層 1 層からなる Neural Network(NN) を用いてリンク旅行時間を予測している．この Zheng らの研究では，実験に使用するデータセットとして，シミュレーションにより得られたデータと 1 台のプローブカーより得られた情報を用いているが，現実的なデータセットではない．本研究ではこの点を考慮し，1 年間分のプローブ情報から得られるリンク旅行時間を用いた実験を行う．

3. 交通流予測

本研究での交通流予測手法では，図 1 のような Deep Neural Network(DNN) を用いる．入力をプローブ情報から得られる時刻 t 以前のリンク旅行時間とし，教師データである時刻 $t+\Delta$ のリンク旅行時間が出力となるように学習を行う．入力は予測

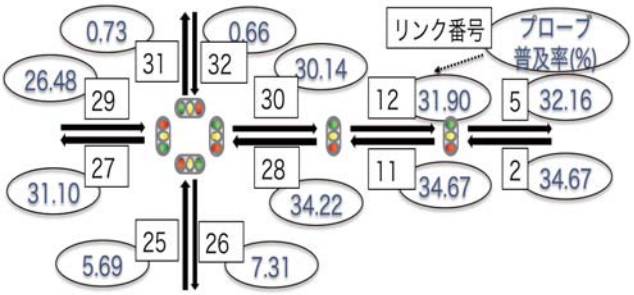


図 2: 道路リンクの一例

対象リンクを含めた時刻 t 以前の複数リンク，出力は予測対象リンクのみのリンク旅行時間とする．

4. 実験

前節で述べた予測手法でリンク旅行時間の予測を行う．また，その予測の妥当性を示すために重回帰分析及びサポートベクター回帰 (SVR) の予測精度と比較を行う．

4.1 使用するデータセット

使用するデータは，図 2 に示すような，岡山県の国道 2 号線付近 33 リンク分の 2016 年 6 月 1 日から 2017 年 5 月 31 日までの 1 年間のプローブ情報より得られたリンク旅行時間であり，リンク旅行時間は 5 分おきに平均されている．データの無い箇所は補完されており，リンク毎のプローブ情報の普及率は図 2 の楕円で囲われた数字で示しており，補完割合は約 70%から 99%である．

4.2 実験設定

本研究では 20 分後のリンク旅行時間の予測を行うので，DNN の出力は時刻 $t+4$ の予測対象リンクのリンク旅行時間とする．データセットの 2016 年 6 月から 9ヶ月間を学習データ，残りの 3ヶ月間を評価データとする．入力には予測対象リンクを含めた複数のリンク旅行時間とする．入力に含める複数リンクのパターンとして，予測対象リンクとその隣接リンクの場合，さらにその複数リンクに隣接しているリンクの場合，といったように全リンク含まれるまで段階的に増やしていく．各パラメータの値は，入力に含めるリンク数を 1 から 33，過去時刻数を 1 から 6，隠れ層のユニットを [10,20,...,100]，隠れ層数 1 から 3 の範囲とする．

4.3 評価指標

評価指標として平方根平均二乗誤差 (RMSE) を用いる．RMSE は以下の式 1 で表され，

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} \left(|f_i - \hat{f}_i| \right)^2 \right]^{\frac{1}{2}} \quad (1)$$

f_i は教師データ (実測値)，$\hat{f}_i$ は出力データ (予測値) である．

4.4 実験結果及び考察

比較的リンク旅行時間が大きい Link31 を予測対象リンクとした．上で述べた全パラメータに対して実験を行なった結果，最も RMSE が低かった組み合わせは表 1 のようになり，RMSE は学習データに対して 54.3229，評価データに対して 58.4035 となった．この表 1 のパラメータ構成で学習されたモデルを用いてリンク旅行時間を予測した結果を実測値と比較したものが図 3 である．横軸が時刻，縦軸がリンク旅行時間であり，青線が実測値，緑線が予測値を表している．このプロットの期間は 2017 年 3 月 1 日から 3 月 4 日までである．この図 3 から，大まかな周期は捉えられていることがわかる．

表 1: 最適なパラメータ

パラメータ	値
リンク数	33
過去時刻数	6
入力次元数	198
隠れ層のユニット	[80,50,40]
隠れ層数	3

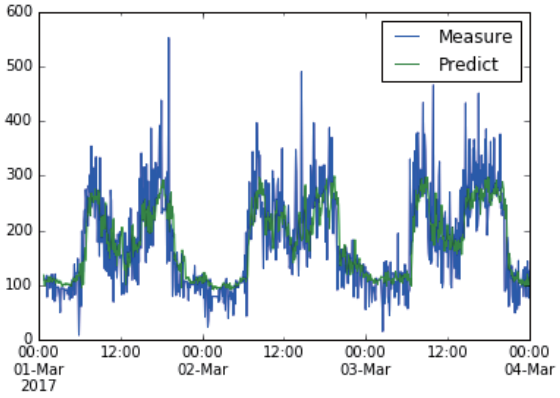


図 3: 実測値と予測値の比較 (Link31) 期間：2017 年 3 月 1 日～2017 年 3 月 4 日

また，出力が Link31 のリンク旅行時間である DNN に対して，入力に含めるリンク数及び過去時刻数を変化させることにより，RMSE にどのような影響を与えるかを調べた．図 4 は，横軸が入力に含めるリンク数の変化で，縦軸が RMSE を表している．入力に含める過去時刻数毎に，リンク数の変化による RMSE の値がプロットされている．この図 4 より，リンク数を増やすことによる RMSE の変化がほとんど見られないことがわかる．図 5 は，横軸が入力に含める過去時刻数の変化で，縦軸が RMSE を表している．入力に含めるリンク数毎に，過去時刻数の変化による RMSE の値がプロットされている．この図 5 から，どのリンク数においても過去時刻数を増やすことで，RMSE の値が減る傾向にあることが分かる．よって，DNN の入力に含める情報として，リンク数を増やすことよりも過去時刻数を増やすことの方が予測精度に大きく影響を与えることが読み取れる．

深層学習による予測手法の妥当性を示すために，重回帰分析及び SVR との比較を行なった結果が表 2 である．また，予測対象リンクとして，リンク旅行時間が比較的大きい Link5 及び Link30 の場合も含めて比較を行なった．各手法を比較した結果，どのリンクにおいても DNN を用いた予測の RMSE が一番小さい値となったが，大差ない結果となっている．本研究での予測手法の妥当性は示されたが，その他手法と比べて大きな精度の向上はみられなかった．これは，本研究で扱ったプローブ情報の補完割合が多いことが関係しており，データセットの補完割合が多いことでデータの多様性がなくなり，非線形な関数の近似能力に優れた深層学習の本来の能力が十分に発揮されなかったからであると考えられる．

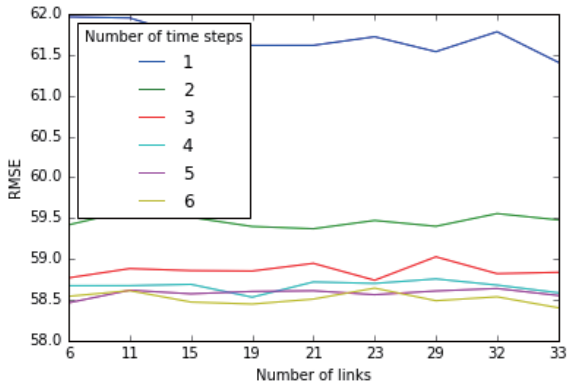


図 4: 過去時刻数毎の，リンク数を変化させた時の RMSE の変化

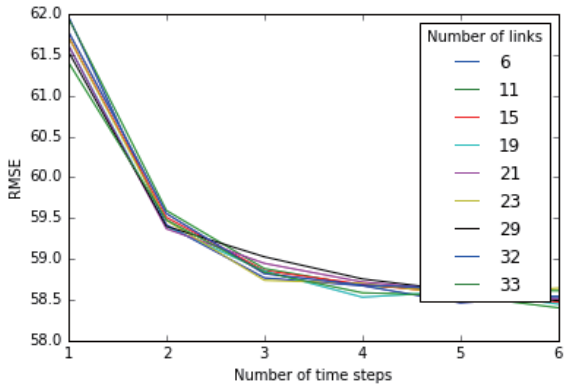


図 5: リンク数毎の，過去時刻数を変化させた時の RMSE の変化

5. まとめと今後の展望

本研究では，プローブデータを用いた深層学習によってリンク旅行時間を予測し，その検証を行った．対象リンクの時刻 $t+4$ のリンク旅行時間を予測するために，予測先リンクとその周辺リンクの時刻 t 以前のリンク旅行時間を用いて予測した．リンク旅行時間の予測をするにあたって，入力に用いる周辺リンクや考慮すべき時刻の範囲を検証した．その結果，入力として用いる周辺リンクの範囲を広げるよりも，過去の時刻を考慮する範囲を広げる方が精度向上に寄与することが示された．

深層学習による予測結果の妥当性を確認するため，重回帰分析および SVR による予測結果と比較を行った．その結果，深層学習による予測結果の妥当性は示され，僅かに深層学習の方が予測精度が優れていることが示されたが，大差無い結果となった．これは，扱ったデータセットの補完割合が高いことでデータの多様性がなくなったからだと考えられる．

また，本研究では 20 分後のリンク旅行時間の予測を行なったが，現在，5 分後，10 分後，15 分後と予測期間を変えることで結果が変わる可能性も考えられるので，予測期間を変えた実験を行うことも必要である．さらに，入力に含める過去時刻数を増やすことで予測精度に影響を与えることが示されたので，過去時刻数をさらに増やした場合の実験も行う必要がある．

表 2: 各手法による予測結果の比較

	予測先リンク	DNN	重回帰分析	SVR
RMSE	Link5	37.64	38.10	37.80
	Link30	65.42	65.96	66.12
	Link31	58.40	59.60	58.71

謝辞

本研究で使用したデータセットは，住友電気工業株式会社様よりご提供いただきました．心より感謝いたします．

参考文献

[1] Yoshua Bengio, et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, Vol. 2, No. 1, pp. 1–127, 2009.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[3] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pp. 160–167. ACM, 2008.

[4] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, Vol. 313, No. 5786, pp. 504–507, 2006.

[5] Brody Huval, Adam Coates, and Andrew Ng. Deep learning for class-generic object detection. *arXiv preprint arXiv:1312.6885*, 2013.

[6] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, pp. 865–873, 2015.

[7] Fangfang Zheng and Henk Van Zuylen. Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, Vol. 31, pp. 145–157, 2013.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

効率的な知識獲得予測に向けた知識の表現方法に関する分析

Analysis of Machine-Friendly Knowledge Representation toward More Efficient Knowledge Tracing

中川 大海 *1 Hiromi Nakagawa　岩澤 有祐 *1 Yusuke Iwasawa　那須野 薫 *1 Kaoru Nasuno　松尾 豊 *1 Yutaka Matsuo

*1東京大学工学系研究科
The University of Tokyo

Recent advancements in computer-assisted learning systems have increased research in the area of *knowledge tracing*. Knowledge tracing is a task to estimate student proficiency based on their past interaction with the learning systems, and it is reported that leveraging neural networks enables efficient estimation. However, such a development of neural network-based knowledge tracing methods suggests the necessity to review the definition of "knowledge", which previously has been designed by human experts and treated as given. In this context, recently a method to automatically learn efficient knowledge representation from student exercise logs has been proposed, and it is becoming important to designing more machine-friendly knowledge representation, which enables efficient performance of knowledge tracing. In this paper, we analyze the properties of knowledge representation learned to maximize the performance of knowledge tracing, and investigate the important factors for machines to efficiently perform knowledge tracing. Using a math open dataset, we empirically validated that the learned representation has more information-efficient structure than the existing representation, and that they have mutually complementing character to improve the performance of knowledge tracing. These results provide useful insights for reviewing the definition of knowledge, which previously has been treated as given, and designing machine-friendly knowledge representation, which could help improve the learning experience of students in more diverse environments.

1. はじめに

近年，オンライン教育サービスの発展に伴い，知識獲得予測 (knowledge tracing) [Corbett 94] の研究が活発化している．知識獲得予測は，生徒の学習行動ログに基づいて，各生徒が知識をどの程度習熟しているかを推定するもので，個人の習熟状況に即した教材推薦などに活用することで，生徒の学習効率を向上させることが可能である．歴史的に様々な手法が提案されてきたが，[Piech 15] が提案した Deep Knowledge Tracing(DKT) という手法は，深層学習の Recurrent Neural Networks(RNN) [Hochreiter 97] を活用することで，従来の手法が必要としていた，人間による複雑な特徴量の設計を必要とせずに，モデルが自動的に知識間の潜在的な関係性を学習できることが報告されている．

このような深層学習による知識獲得予測の発展は，従来は人手で設計されてきた「知識」自体の見直しを要請するものでもある．例えば，DKT を含む従来の知識獲得予測の手法は，人間の専門家によって事前に定義された，演習問題を解く上で必要となる知識を表す「知識タグ」の存在を前提としていたが，[Nakagawa 18] はこの知識タグ自体も生徒の回答ログから学習することを提案し，このタグを用いることで既存の知識タグと比較して知識獲得予測の精度が向上することを報告している．すなわち，従来人手で設計されてきた知識は，人間にとって解釈しやすく，取り扱いやすいものではあるものの，機械で知識獲得予測を行う上では必ずしも最適な構造ではない可能性が示唆されており，今後，知識獲得予測に基づいた学習内容の個人最適化が一般的になっていく中で，機械にとってより予測しやすい知識表現を設計することの重要性が高まっている．

本研究では，[Nakagawa 18] の研究を発展させ，知識獲得の予測性を最大化するように回答ログから学習された知識タグの性質を分析することで，機械にとって予測しやすい知識構造についての分析を行う．まず，[Nakagawa 18] で提案されたタグ学習モデルによって，DKT における予測性を最大化するようなタグ表現を学習する．次に，学習された知識タグ（以下，学習タグ）と既存の知識タグ（以下，既存タグ）が，それぞれ DKT の予測性能に与える影響を計測することで，両者の関係性を分析し，また各タグの構造を定量的に比較することで，両者の性質の違いを検証する．実験の結果，学習タグは，既存タグと組み合わせることでより高い予測性能を発揮し，かつ既存タグと相互に補完し合う性質を有することが確認され，また，既存タグに比べて生徒の問題回答の情報を効率的に表現する構造を獲得していることが検証された．これらの結果を踏まえ，既存タグと学習タグそれぞれが有する補完的な性質について考察し，機械にとって予測しやすい知識表現を設計する上での要件について議論した．

本論文の貢献は以下の 2 点である．

- 生徒の問題回答ログから学習された知識タグ表現は，DKT による知識獲得予測を行う上で有効な表現を獲得しており，また，既存の知識タグと相互に補完し合う性質を持つことを示した．
- 学習されたタグ表現の構造や，既存の知識タグと補完し合う性質について定量的に分析することで，知識獲得の予測性を向上させる知識タグの設計に関する洞察を得た．

2. Deep Knowledge Tracing (DKT)

DKT の構造は伝統的な RNN の構造に基づいており，時刻 t における入力を $\mathbf{x}_t$，隠れ層を $\mathbf{h}_t$，出力を $\mathbf{y}_t$ とすると，その関係性は以下の式で表される．

$$\begin{aligned} \mathbf{h}_t &= f(\mathbf{x}_t, \mathbf{h}_{t-1}) \\ \mathbf{y}_t &= g(\mathbf{h}_t) \end{aligned}$$

連絡先: 中川大海, 東京大学工学系研究科技術経営戦略学専攻, nakagawa@weblab.t.u-tokyo.ac.jp

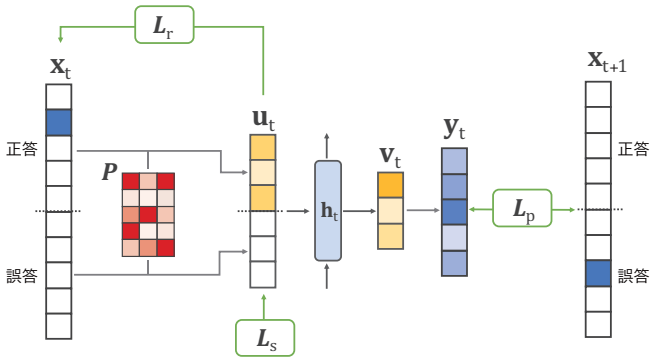


図 1: タグ学習モデルの構造

モデルは関数 f と g によって定義されており，LSTM [Hochreiter 97] や GRU[Cho 14] など様々な活性化関数を利用できる．入力 $\mathbf{x}_t$ は，全部で N 個ある既存の知識タグが紐付いた問題のうち，正答したタグに対応する要素のみ 1 で他は 0 の長さ N のベクトルと，誤答したタグに対応する要素のみ 1 で他は 0 の長さ N のベクトルを結合した，長さ $2N$ の二値のベクトルとして表現される．

最適化は式 (1) で表される回答正誤予測に関する損失関数 L_p を最小化することにより行われる．

$$L_p = \sum_t l(\mathbf{y}_t^T \tilde{\delta}(\mathbf{q}_{t+1}), \mathbf{a}_{t+1}) \tag{1}$$

ここで $\mathbf{y}_t$ は時刻 t の出力である問題正答予測ベクトルであり，$\tilde{\delta}(\mathbf{q}_{t+1})$ は実際の時刻 $t+1$ の問題回答ベクトルであり，$\mathbf{a}_{t+1}$ は時刻 $t+1$ に対応する問題で正答したか否か（1 か 0）のベクトルであり，l は交差エントロピー誤差関数である．

学習済みの DKT のモデルからは，知識獲得の時系列性や知識間の関係性を考慮した，包括的な知識間の関係性を表すネットワークを抽出することができる．問題 i と j のすべての有向ペア (i, j) のうち，問題 i が出現した後に残りの問題系列の中で問題 j が出現する系列数が問題 i が出現する問題系列数全体の V% 以上であるものに対して，式 (2) の影響度 J_{ij} を割り当てる．

$$J_{ij} = \frac{y(j|i)}{\sum_k y(j|k)} \tag{2}$$

ここで $y(j|i)$ は，ある生徒が最初に問題 i に正答した場合に，次の時刻に問題 j に正答する確率である．

3. タグ学習モデル

3.1 モデルの構造

一般的な DKT では，生徒の問題に関する回答は，専門家が定義したルールに基づいて，知識タグに関する回答に変換されてモデルに入力される．この処理は，問題の数が M，知識タグの数が N である場合に，問題回答を表現する長さ M の二値ベクトルに，問題とタグの対応を表す $M \times N$ の大きさの二値の行列 $\mathbf{P}$ をかけることで，タグに対する回答を表現する長さ N の二値ベクトルを得る処理と定式化できる．ここで，$\mathbf{P}$ は，各行が問題を，各列がタグを表し，各問題に紐づくタグに対応する要素が 1 でそれ以外が 0 の行列である．

従来の DKT ではこの行列が暗黙的に所与のものとされてきたが，タグ学習モデルでは，この写像行列を生徒の問題回答ログから学習する．モデルの構造を図 1 に示す．タグ学習モデルでは，問題に関する $2M$ 次元の回答ベクトルをそのまま入力 $\mathbf{x}_t$ とした上で，問題に対する M 次元の正答予測 $\mathbf{y}_t$ を出力する．また，問題空間を低次元のタグ空間に変換する写像行列 $\mathbf{P}$ を学習するために，N' 次元のタグの存在を仮定し，入力 $\mathbf{x}_t$ と隠れ層の $\mathbf{h}_t$ の間に $2N$' 次元の $\mathbf{u}_t$，および $\mathbf{h}_t$ と出力層 $\mathbf{y}_t$ の間に N' 次元の $\mathbf{v}_t$ という中間層を設ける．タグの次元数 N' は任意に設定することが可能であるが，生徒の学習過程を個別の問題よりも抽象的な粒度で捉えるという目的を踏まえると，問題の次元 M よりも小さい次元に設定することが適切である．特に本研究では，既存の知識タグとの比較を行うという背景から，比較を公平に行い差分を明確に検証するために，$N' = N$ として既存の知識タグと同じ次元数に設定する．以上を踏まえて，モデルは以下の式で定義される．

$$\begin{aligned}
\mathbf{x}_t &= [\mathbf{x}_{t_pos}, \mathbf{x}_{t_neg}] \\
\mathbf{u}_t &= [\mathbf{P}\mathbf{x}_{t_pos}, \mathbf{P}\mathbf{x}_{t_neg}] \\
\mathbf{h}_t &= \varphi(\mathbf{u}_t, \mathbf{h}_{t-1}, \theta) \\
\mathbf{v}_t &= \sigma(\mathbf{W}_{hv}\mathbf{h}_t + \mathbf{b}_v) \\
\mathbf{y}_t &= \sigma(\mathbf{W}_{vy}\mathbf{v}_t + \mathbf{b}_y)
\end{aligned}$$

ここで，$\mathbf{x}_{t\_pos}$, $\mathbf{x}_{t\_neg}$ はそれぞれ $\mathbf{x}_t$ の前半と後半の長さ M のベクトルで，問題の正答・誤答を表現しており，$\mathbf{W}_{hv}$，$\mathbf{W}_{vy}$ は重み行列を指し，$\mathbf{b}_v$，$\mathbf{b}_y$ はバイアス項を指し，φ は活性化関数を表す．なお，θ は任意のパラメータを指し，用いる活性化関数によって異なる．

3.2 目的関数

タグ学習モデルでは，より良質な写像行列 $\mathbf{P}$ を学習するために，式 (1) の既存の損失関数 L_p に加え，2 つの損失関数を導入する．

1 つは式 (3) で表される「再構成誤差 (L_r)」である．これは，タグ空間の各要素が，問題の背後に潜む知識間の依存関係などを適切に学習できていれば，生徒の問題に対する正答はそれに対応するタグ空間の状態から予測できるはずであるという前提を反映することを目的としたものであり，以下の式で定義される．

$$\begin{aligned}
g(\mathbf{u}_t) &= \sigma(\mathbf{W}_{vy}\mathbf{u}_t + \mathbf{b}_y) \\
L_r &= \sum_t l(\mathbf{x}_t, g(\mathbf{u}_t))
\end{aligned} \tag{3}$$

もう 1 つは式 (4) で表される「スパース正則化項 (L_s)」である．タグ学習モデルで学習される $\mathbf{P}$ の要素はそのままでは連続値であるため，学習後に一定の条件に基づき $\mathbf{P}$ の各値を 0 または 1 に離散化することでタグを生成するが，初めから $\mathbf{P}$ が離散表現に近いほうが離散化する際に情報量の損失が少ない．よってこのスパース正則化項を導入することにより，$\mathbf{P}$ を事前に 0 または 1 に近づくように学習させる．

$$L_s = \sum_t (0.5 - |\mathbf{u}_t - 0.5|)^2 \tag{4}$$

最終的に，モデル全体の損失関数 L は以下の式 (5) によって定められ，この損失関数を最小化するようにモデルが最適化される．

$$L = \alpha L_p + \beta L_r + \gamma L_s \tag{5}$$

ここで，α，β, γ は任意の非負の実数である．

表 1: データセットの統計量

生徒数	問題数	既存タグ数	ログ数
1,132	3,439	193	592,407

表 2: 各モデルにおける実験結果

入力→予測	AUC	フロー階層性	GRC	σ
既存→既存	0.810	0.72	0.70	9,701
学習→学習	**0.823**	**0.88**	**0.87**	**3,674**
学習→既存	0.816	-	-	-
既存+学習→既存	0.826	-	-	-

4. 実験

4.1 実験設定

実験には，KDDCup の Educational Data Mining Challenge で使用された「Bridge to Algebra 2006-2007」[Stamper 10] という，小学校から高校レベルの算数・数学に関するオンライン教育サービス上の生徒の問題回答ログからなるデータセットを用いた．Bridge to Algebra 2006-2007 は，「Calculate area of overlap」や「Identify GCF」など，専門家によって設計された知識タグ (以下，既存タグ) の情報を有するため，タグ学習モデルによって学習したタグ（以下，学習タグ）との比較検証が可能である．まず，[Nakagawa 18] と同様の条件に基づき，実験に用いるデータを抽出した．データセットの統計量を表 1 に示す．

次に，生徒の問題に対する回答ログにタグ学習モデルを適用した．RNN の部分には GRNN を用い，隠れ層のユニット数は 400 とした．ドロップアウトは $\mathbf{u}_t$ から $\mathbf{h}_t$ の方向にドロップアウト率 0.5，$\mathbf{h}_t$ から $\mathbf{y}_t$ の方向にドロップアウト率 0.2 で適用した．最適化は，式 (5) の L を目的関数として Adam[Kingma 14] によって最小化した．評価指標は AUC を採用した．訓練：検証：テスト $= 8:1:1$ となるようにユーザを分け，訓練ユーザのデータでモデルを構築し，検証ユーザのデータでハイパーパラメータを調整し，検証ユーザのデータで精度が最も高かったモデルを採用した．

最後に，タグ学習モデルによって学習された写像行列 $\mathbf{P}$ を以下の式に基づいて離散化し，タグ表現を獲得した．

$$\mathbf{P'}_{i,j} = \begin{cases} 1 & \text{if } \mathbf{P}_{i,j} = \max(\mathbf{P}_i) \text{ or } \mathbf{P}_{i,j} \geq \theta \\ 0 & \text{else} \end{cases} \quad (6)$$

ここで，i, j は $\mathbf{P}$ の行と列に対応する．θ はハイパーパラメータであり，学習されたタグを DKT に適用した際に検証ユーザのデータで最も高い精度を発揮する値を探索した．

4.2 実験結果

まず，[Nakagawa 18] と同様に，各問題を既存タグに変換して DKT に適用した場合（以下，「既存→既存」）と，学習タグに変換して DKT に適用した場合（以下，「学習→学習」）の精度を比較した結果を表 2 に示す．精度の高いものを太字で示した．タグ学習モデルのハイパーパラメータは，$\alpha = 1.0, \beta = 0.5, \gamma = 1.0, \theta = 0.7$ とした．実験の結果，既存タグを用いる場合と同等以上の精度で知識獲得予測を行うことのできる学習タグを抽出できた．

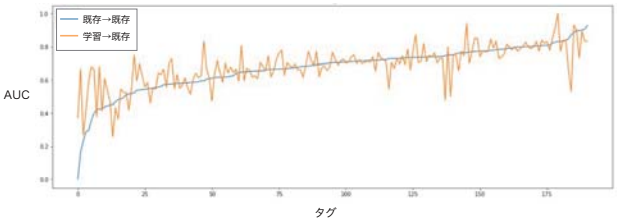


図 2: 各モデルにおける各タグの予測精度

次に，学習タグが予測精度向上に与える寄与や，学習タグと既存タグとの関係性を検証するために，モデルの出力は問題を既存タグに変換したベクトルとした上で，モデルの入力でも問題を既存タグに変換したベクトルを用いた場合（「既存→既存」），モデルの入力では問題を学習タグに変換したベクトルを用いた場合（「学習→既存」），モデルの入力に，問題を既存タグに変換したベクトルと学習タグに変換したベクトルを結合したものを用いた場合（「既存+学習→既存」）の精度を比較した．結果を表 2 に示す．まず，「学習→既存」の精度が「既存→既存」よりも高いことから，学習タグが，既存タグに基づいた回答行動を予測するという，従来のタスクを解く上でもより効率的な表現となっていることが示された．また，「既存+学習→既存」の精度が最も高かったことから，既存タグと学習タグが相互に補完的な性質を有することが示唆された．この補完性について検証するために，「既存→既存」と「学習→既存」のモデルにおいて，モデルが各既存タグに対して出力する正答予測確率を元に，各タグについて個別に AUC を算出し，図 2 に示した．青の線が「既存→既存」を，オレンジの線が「学習→既存」を表しており，各点を 1 つのタグとしてプロットし，「既存→既存」における AUC スコアのの昇順に左から並べた．図より，「既存→既存」と「学習→既存」それぞれで，予測精度の高いタグと低いタグが存在することがわかり，こうした性質を補完し合うことにより「既存+学習→既存」において最も高い精度を発揮できたと考えられる．

さらに，[Piech 15] に基づき，「既存→既存」と「学習→学習」の各モデルで知識間影響ネットワークを作成し，その階層性を計測した．階層性の指標としてフロー階層性 [Luo 11] と GRC[Mones 12] を採用した．結果を表 2 に示す．値が大きいものを太字で示した．表より，学習タグのネットワークは既存タグのネットワークに比べて高い階層性を有することがわかった．このことから，学習タグが，教育学的に「手続き的知識」として階層性が高いといわれている [Anderson 82, Anderson 90] 数学の学問的な特性を反映していることがわかる．また，階層的にカリキュラム設計が生徒の学習効率の向上を促すことも報告されており [Block 71, Cohen 79]，そのような効率的な構造を反映した階層的な構造になっている可能性も考えられる．

最後に，既存タグと学習タグについて，各タグが回答ログに出現する回数の分布を比較した．DKT においては，入出力の各要素はタグに該当するため，各タグが回答ログに出現する回数，すなわちモデルに入力される回数は，DKT の学習の過程で各ユニットに関する重みの学習されやすさに直接影響する．例えば，頻繁に回答されるタグに対応するユニットは十分に学習が進む一方，あまり回答されないタグに対応するユニットは十分に学習できない可能性が高い．既存タグと学習タグにおいて，各タグが回答ログに出現する回数の分布の標準偏差 (σ) を表 2 に示す．値の小さいものを太字で示した．表より，既存タグはタグ間で回答頻度の分散が大きい一方，学習タグは分

散が小さいことがわかる．これは，学習タグはニューラルネットワークの最適化の過程で学習されたことで，入出力層の各ユニットに相当する各タグが効率よく情報を有することができるように，情報が均等に分布するよう学習された一方，既存タグは，学問の伝統的な背景や，人間にとっての取扱やすさなどを考慮して設計されており，回答の情報を均等化することは考慮されていないためだと考えられる．

5. 考察

実験結果より，DKT において既存タグよりも高い予測性能を発揮した学習タグは，既存タグに比べて情報が効率よく分配され，高い階層性を有するものであることが示され，深層学習を用いた知識獲得予測の予測性を向上させる上で重要な素性が示唆された．一方，図 2 の分析からわかるように，全てのタグの予測において学習タグが既存タグを上回っているわけではなく,「既存→既存」「学習→既存」に比べ「既存+学習→既存」が最も高い精度を発揮したことからも，既存タグと学習タグにはそれぞれ異なる点で優れた性質を持ち，相互に補完し合うことが可能なものであることが示された．これは，人間の解釈しやすさなどに基づいて既存タグを設計される過程で失われた，個別の問題回答が有する豊富な情報を，学習タグが効果的に学習できた一方，学習タグが問題回答ログのみからでは十分に学習できなかった，複雑な知識の特性や教育環境などの前提を，人間の設計した既存タグは暗黙的に反映できていた可能性が考えられる．より多様なデータセットで同様の実験を行い，かつ既存タグと学習タグでそれぞれ予測が優れている点を詳細に比較することで，既存タグと学習タグそれぞれが反映できている情報，反映できていない情報などに関してより詳細な分析が可能になり，知識の設計を考える上でより深い洞察を得られると考えられ，今後の研究課題である．

6. 結論

本研究では，近年の深層学習による知識獲得予測の発展に伴い生じた，従来は人手で設計されてきた「知識」自体の見直しの必要性に着目し，知識獲得の予測性を最大化するように回答ログから学習された知識タグの性質を分析することで，機械にとって予測しやすい知識構造についての分析を行った．実験の結果，学習された知識タグは，既存の知識タグと組み合わせることでより高い予測性能を発揮し，かつ既存タグと相互に補完し合う性質を有することが確認され，また，既存タグに比べて生徒の問題回答の情報を効率的に表現できる構造を獲得していることが検証された．これらの結果は，従来人手で設計されてきた知識は，人間にとって解釈しやすく，取り扱いやすいものではあるものの，機械で知識獲得予測を行う上では必ずしも最適な構造ではないという可能性を示唆するものであることに加え，機械にとって予測しやすい知識を設計する上で，有益な示唆を与えるものである．より多様な教科やデータセットにおいて同様の検証を行うことで，今後さらに普及が進むと考えられる知識獲得予測に基づく学習内容の個人最適化を，さらに効果的なものにしていく上で重要な「知識」を適切に定義する上での要件が明らかにされると考えられる．本研究が，そうした新たな「知識」の解明につながると信じている．

謝辞

本研究は JSPS 科研費 JP25700032, JP15H05327, JP16H06562 の助成を受けたものです．

参考文献

[Anderson 82] John R Anderson: Acquisition of cognitive skill, Psychological review 89, 4 (1982), 369.

[Anderson 90] John R Anderson: The adaptive character of thought, Psychology Press (1990).

[Block 71] James H Block and Peter W Airasian: Mastery learning: Theory and practice. Holt Rinehart & Winston (1971).

[Cho 14] Kyunghyun Cho, Bart Van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).

[Cohen 79] S Alan Cohen and Joan S Hyman. Learning for Mastery: Ten Conclusions after 15 Years and 3,000 Schools. Educational Leadership 37, 2 (1979), 1049.

[Corbett 94] Albert T Corbett and John R Anderson: Knowledge tracing: Modeling the acquisition of procedural knowledge, User modeling and user-adapted interaction 4, 4 (1994), 253278.

[Hochreiter 97] Sepp Hochreiter and Jurgen Schmidhuber: Long short-term memory, Neural computation 9, 8 (1997), 17351780.

[Kingma 14] Kingma, Diederik and Ba, Jimmy: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[Luo 11] Jianxi Luo and Christopher L Magee: Detecting evolving patterns of self-organizing networks by flow hierarchy measurement, Complexity 16, 6 (2011), 5361.

[Mones 12] Enys Mones, Lilla Vicsek, and Tamas Vicsek: Hierarchy measure for complex networks, PloS one 7, 3 (2012), e33799.

[Piech 15] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein: Deep Knowledge Tracing, In Advances in Neural Information Processing Systems. 505513 (2015).

[Stamper 10] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R Koedinger. Bridge to Algebra 2006-2007, Development data set from KDD Cup 2010 Educational Data Mining Challenge, http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp. (2010).

[Nakagawa 18] 中川 大海, 那須野 薫, 岩澤 有祐, 上野山 勝也, and 松尾 豊: Deep Knowledge Tracing の拡張による擬似知識タグの生成, 人工知能学会論文誌 33, 3 (2018).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

入力データ非貢献部抽出によるCNNの出力の説明

Highlighting Non-contributing Pixels for Visual Explanation of CNNs

池野 光一 *1　原 聡 *1　鷲尾 隆 *1
Kouichi Ikeno　Satoshi Hara　Takashi Washio

*1大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

Explaining the output of Convolutional Neural Networks (CNNs) is a challenging topic. A typical explanation is to identify which pixels are contributing to the output of CNN. In this paper, we propose a new approach for explaining the output of CNNs by finding pixels that are *not* contributing to the output. To highlight non-contirbuting pixels, we propose optimizing a noise level so that additive noise to the input image does not change the CNN output. The experimental results on MNIST show that the proposed method can idntify non-contributing pixels adequately.

1. 概要

近年の機械学習技術の研究の成果では，特に認識・予測の精度の高さに注目が集まっている．しかし，実応用の場面においては，単に精度だけでなくあわせてモデルの説明性が必要とされることがある．例えば医療現場における病気の診断では，単に患者が特定の疾患に罹患しているかを高い精度で判断するだけでは不十分である．医療サービス提供側には，判断結果だけでなくなぜそのような疾患であると判断したのか，その理由を患者に説明する責任が発生する．このような機械学習の実応用における要請を背景に，機械学習モデルの説明性のニーズが高まっている [1].

本研究では，画像認識のための深層学習モデルである Convolutional Neural Network (CNN)[2] を対象に，その説明性を高める方法を提案する．CNN の説明性を高める方法として，モデルが入力画像のどの部分を根拠に認識を行っているかを特定してハイライトする方法が研究されている [3, 4]．このような説明法を用いることで，モデルの認識根拠を人間が視覚的に検証できるようになる．

提案法では，モデル説明のための新しいハイライト法として「モデル出力に寄与しない入力画像の非貢献部をハイライトする」方法を提案する．従来の手法では，モデル出力と関連が強い入力画像領域を貢献部としてハイライトする．これに対し，本研究では非貢献部に着目することで，これら従来の手法とは逆のアプローチを取る．提案法は従来法のようにモデル線形化のような粗い近似を必要としないという点において，モデルのより本質的な情報に基づいたハイライトを生成できると考えている．

2. 既存手法

本節では,CNN の判断根拠の説明のための代表的なハイライト法を紹介する．

2.1 感度マップ

感度マップでは，入力画像中のピクセルのうち, その値の微小変化が出力に大きく影響するピクセルを重要なピクセルだと考えハイライトする手法である．これは説明対象の CNN モデルを線形近似し，入力画像のピクセルごとに出力の勾配を調べることに相当する．

今，入力画像を x とし，p をピクセルのインデックスとする．これを全 C クラスのうちのある 1 つのクラスに分類するようなモデルを考える．入力 x を受けそれぞれのクラス $c \in C$ に対しスコア $S_c(x)$ が計算され，最終的にこのスコアが最も大きいクラス c に分類される．これは以下の式で表される．

$$class(x) = \mathrm{argmax}_{c \in C} S_c(x)$$

これに対し, 感度マップは以下のようにスコアの勾配として定義できる．

$$(M_c(x))_p = \frac{\partial S_c(x)}{\partial x_p}$$

ただし，$(M_c(x))_p$ は $M_c(x)$ の第 p 成分である．$(M_c(x))_p$ は x_p の微小変化に対してスコア $S_c(x)$ がどれほど変化するかを表したものであるから，$(M_c(x))_p$ の値が大きいピクセル x_p がモデルの出力に大きく寄与していると考えられる．図 1 に感度マップの例を示す．

2.2 SmoothGrad [3]

SmoothGrad は，感度マップを改良した手法である．図 1 に見られるように，感度マップでは認識対象でない背景部分もハイライトされている．SmoothGrad は平均化によりこのような背景のノイズを低減する．

元入力 x にガウシアンノイズ $\mathcal{N}(0, \sigma^2)$ を乗せたものを n 個用意し，それぞれを新たな入力として感度マップ $M_{c,k}(x)(k = 1, ..., n)$ を生成する．それらの平均 $\hat{M}_c(x)$ をとり，モデルの説明とする．

$$\hat{M}_c(x) = \frac{1}{n} \sum_{k=1}^{n} M_c(x + \epsilon_k), \quad \epsilon_k \sim \mathcal{N}(0, \sigma^2)$$

このように,x に微小なノイズを加えてもモデルの判断はほとんど変わらないが，異なるノイズが乗ることで感度マップによってハイライトされる部分はそれぞれ少しづつ異なる．SmoothGrad では, それらの平均を取ることで判断の際に頻繁に重視される部分が浮かび上がり，ノイズに反応してしまった部分はハイライトされづらくなる．

連絡先: 池野 光一，大阪大学 産業科学研究所，
〒 567-0047 大阪府茨木市美穂ヶ丘 8-1
k1keno@ar.sanken.osaka-u.ac.jp

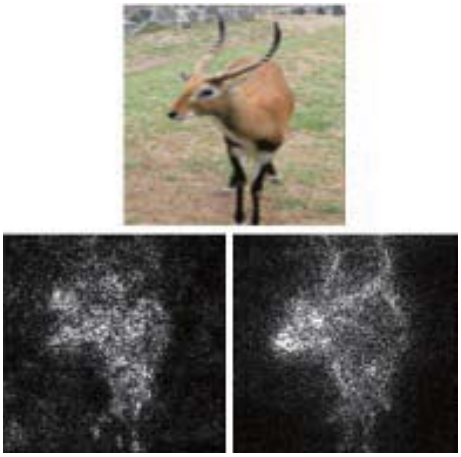


図 1: 上が入力画像，左下が感度マップ，右下が SmoothGrad である．認識対象のぼやけが改善されている．(図は [3] の Fig.3 より引用)

2.3 deep Taylor Decomposition [4]

deep Taylor Decomposition は, 各入力ピクセルの貢献量を出力から計算する手法である．そのために,deep Taylor Decomposition では各層の各ユニットでの貢献量を考える．なお, この際に各層の貢献量の総和は等しいとする以下のような保存則を仮定する．いま, 第 a 層とその次の第 $a+1$ 層について、各層のユニットのインデックスをそれぞれ i, j, 各ユニットの貢献量をそれぞれ $R_i^{(a)}, R_j^{(a+1)}$ とする．このとき, 保存則とはそれぞれの層の貢献量の総和が等しいことを指す：$\sum_i R_i^{(a)} = \sum_j R_j^{(a+1)}$．deep Taylor Decomposition では, 保存則と一次のテイラー展開に基づいて各層の貢献量を計算する．一次のテイラー展開より,$R_j^{(a+1)}(\{\tilde{x}_i\}) = 0$ となる $\{\tilde{x}_i\}$ を用いて,

$$\begin{aligned}\sum_j R_j^{(a+1)} &= \left(\left.\frac{\partial\left(\sum_j R_j^{(a+1)}\right)}{\partial\{x_i\}}\right|_{\{\tilde{x}_i\}}\right)^T \cdot (\{x_i\} - \{\tilde{x}_i\}) + \epsilon \\ &= \sum_i \sum_j \left.\frac{\partial R_j^{(a+1)}}{\partial x_i}\right|_{\{\tilde{x}_i\}} \cdot (x_i - \tilde{x}_i) + \epsilon\end{aligned}$$

と表せる．ϵ_j は十分に小さいものとして無視できるとすると,

$$\begin{aligned}\sum_i R_i^{(a)} &= \sum_j R_j^{(a+1)} \\ &= \sum_i \sum_j \left.\frac{\partial R_j^{(a+1)}}{\partial x_i}\right|_{\{\tilde{x}_i\}} \cdot (x_i - \tilde{x}_i) \\ \therefore R_i^{(a)} &= \sum_j \left.\frac{\partial R_j^{(a+1)}}{\partial x_i}\right|_{\{\tilde{x}_i\}} \cdot (x_i - \tilde{x}_i)\end{aligned}$$

という更新則が導かれる．この貢献量の逆伝搬則を出力層から入力層まで順次適用することで, 入力画像のピクセルごとの貢献量 $\{R_p\}$ を求める．図 2 に deep Taylor Decomposition の一例を示す.

3. 提案法

本節では，モデル説明のための新しいハイライト法として「モデル出力に寄与しない入力画像の非貢献部をハイライトする」方法を提案する．提案法は感度マップや SmoothGrad, deep Taylor Decomposition のようにモデル線形化のような粗い近似を使わないので，モデルのより本質的な情報に基づいたハイライトを生成できると期待できる．

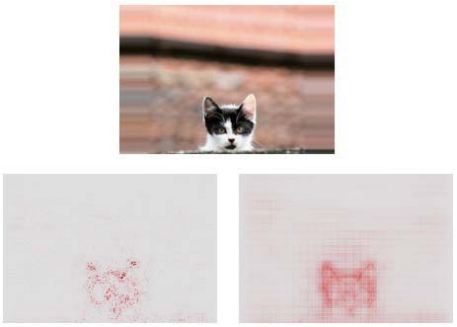


図 2: 上が入力画像，左下が感度マップ，右下が deep Taylor Decomposition である．認識対象のぼやけが改善されている．(図は [3] の Fig.3 より引用)

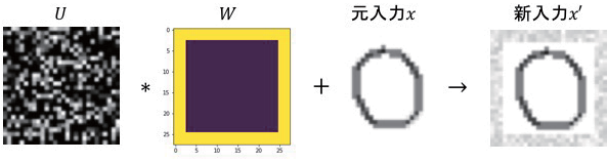


図 3: ノイズを付加する入力側の処理．黄色の部分に大きな値が，紫色の部分に大きな値がある．

3.1 基本アイディア

一般に，画像には認識対象の他に，背景部分など認識とは無関係なものも写っている．このような画像から深層学習モデルが対象を認識して分類するにあたり，認識対象は出力に貢献する貢献部だと言える．一方，認識対象以外の背景部分などは出力に貢献しない非貢献部だと言える．

提案法では，非貢献部をハイライトするため，図 3 のように説明したい入力にノイズを付加した新たな入力を考える．新たな入力 x' に対するモデルの出力 y' は，説明したい入力 x の本来の出力 y から変動するが，この変動はノイズのかかる領域が非貢献部であるか否かに大きく影響を受けると推測できる．例えば，非貢献部だけにノイズが付加されたときは認識対象自体に変化はないため，モデルの出力はほぼ変動しないと考えられる．逆に，認識対象にノイズが付加されると出力は大きく変動すると考えられる．これらの考察に基づいて，本研究では出力がほぼ変動しないノイズを見つけることで，入力画像 x の非貢献部を特定する．

提案法では，視覚的にわかりやすい説明を得るために，説明したいモデルの内部ネットワークに前処理を施す．そのために，出力の変動を考えるにあたって，x を CNN に入力したときに”特徴が見られない”として，あるしきい値より小さな値が伝搬していた部分に注目する．このような”特徴が見られない”部分は出力に寄与しないので，説明のためには不要である．そこで，この部分には必ず 0 が伝搬するようにする．本研究では, この処理を行うユニットをマスクと名付ける.

3.2 学習

提案法では，出力が変動しないようなノイズを，ノイズの大きさパラメータを求めることによって非貢献部をハイライト

する．入力に付加するノイズは以下のように生成する：U を $[-1,1]$ 上の一様分布に従う縦横サイズ L の乱数行列，w を縦横サイズ L のノイズの大きさを決めるパラメータ行列とし，ノイズを $U*w$ により生成する．つまり，新しい入力 x' はピクセル単位で $x'_{ij}=U_{ij}w_{ij}+x_{ij}$ である．このとき，w_{ij} の値が大きくても出力 y' が変動しない，つまりピクセル x_{ij} に大きなノイズを乗せても出力に影響がない場合には，x_{ij} は出力に貢献していないと判断できる．

提案法では，最適なノイズパラメータ w を求めるために以下の目的関数 $E(w)$ を定義し，その最小化により最適な w を求める．

$$E(w)=\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\max(|\,y_k-{y'}_k\,|-a,0)-\frac{1}{L^2}\sum_{i,j=0}^{L-1}w_{ij}$$

$$(0\leq w_{ij}\leq 1)$$

提案法では最適化アルゴリズムとして Adam[5] を使う．ここで，N は最適化の各ステップで生成する乱数行列 U の数，K は分類クラス数，また，y_k, y'_k はそれぞれ出力 y, y' の k 番目の要素である．右辺第 1 項は出力の変動に対する制約で，変動が許容幅 a を超えた時に 0 より大きな値をとる．このとき a を y_k の大小関係が変わらない程度の値に設定することで，認識対象の特徴を損なわないように w を学習させることができる．右辺第 2 項は多くの w_{ij} ができるだけ大きな値をとるように学習を進め，w をマップ化した際に鮮明なイメージとする．最後の制約は，ノイズの大きさを現実的な値に制限するものである．

4. 実験

本節では，提案法の有効性を確認するための実験とその結果について述べる．

4.1 実験概要

本研究では，図 4 のような CNN を説明対象とし，手書き数字画像データ MNIST[6] を入力画像として実験を行った．このモデルは入力画像を 0〜9 の 10 クラスに分類するもので $(K=10)$，また，説明対象の入力画像のいずれもが正しく分類された．前処理として，図 5 のように，マスクのしきい値は，0.2(第 1 畳み込み層)，0(第 2 畳み込み層・全結合層) とした．学習については，$N=100$ で反復回数を 15k 回とし，Adam の学習率は 1×10^{-4} を用いた．

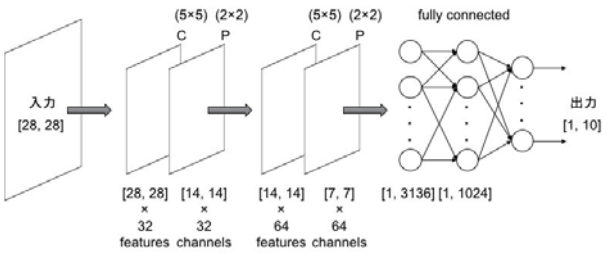


図 4: 説明対象の 3 層 CNN．(図は [3] の Fig.3 より引用)

4.2 実験結果

図 6 のように, 提案法により数字の背景部分が非貢献部としてハイライトされた．逆に，数字及びその近接部は出力に貢献する部分と判断された．これは，モデルが数字とその近接部の

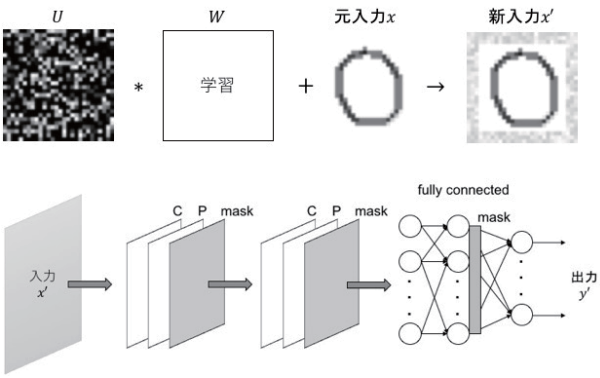


図 5: ノイズパラメータ w を学習するためのネットワークの構造図

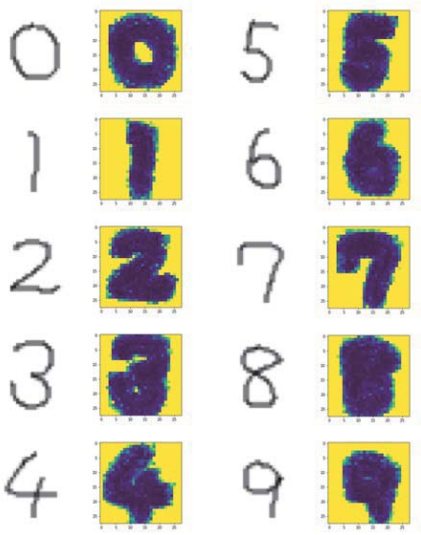


図 6: 左が入力画像，右が提案法による出力の説明である．

両方を用いて数字の輪郭を捉えているためだと考えられ, 直感に即した結果だと言える．

5. まとめ

本研究では，CNN の出力根拠を示すために入力の非貢献部をハイライトする手法を提案した．また，MNIST データを用いた実験により，提案法の有効性を検証した．結果，CNN が適切に数字の輪郭を認識した上で出力を決定していることが確認できた．

参考文献

[1] Riccardo Guidotti, Anna Monreale, Franco Turini, Dino Pedreschi, Fosca Giannotti. A Survey Of Methods For Explaining Black Box Models. arXiv:1802.01933v2. (2018)

[2] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1986(11), Pages 2278-2324. (1998)

[3] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, Martin Wattenberg. SmoothGrad :removing noise by adding noise. arXiv:1706.03825. (2017)

[4] Grégoire Montavon, Sebastian Bach, Alexander Binder, Wojciech Samek, Klaus-Robert Müller. Explaining NonLinear Classification Decisions with Deep Taylor Decomposition. Pattern Recognition Volume 65, Pages 211-222. (2017)

[5] D.P.Kingma and J.Ba.Adam A method for stochastic optimization. arXiv:1412.6980. (2014)

[6] Yann LeCun, Corinna Cortes, Christopher J.C. Burges. THE MNIST DATABASE of handwritten digits. http://yann.lecun.com/exdb/mnist/

[2A3] [General Session] 11. Robot / Real World

座長:稲邑 哲也（国立情報学研究所）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room A (4F Emerald Hall)

[2A3-01] Bidirectional Translation between Robot Actions and Linguistic Descriptions by Aligned Representations

○Tatsuro Yamada[1,2], Hiroyuki Matsunaga[1], Tetsuya Ogata[1] （1. Waseda University, 2. Research Fellow of Japan Society for the Promotion of Science）

3:20 PM - 3:40 PM

[2A3-02] Trials Reduction Method for Reinforcement Learning in Trajectory Discovery

○Yusuke Kato[1,2], Tomoaki Nakamura[3], Takayuki Nagai[3], Natsuki Yamanobe[1], Kazuyuki Nagata[1], Jun Ozawa[1] （1. Advanced Industrial Science and Technology, 2. Panasonic Corporation, 3. The University of Electro-Communications）

3:40 PM - 4:00 PM

[2A3-03] An Approach to Robot Control using Deep Reinforcement Learning and Discretization by Words

○Sayuri Hashimoto[1], Akira Kaneko[1], Ichiro Kobayashi[1] （1. Ochanomizu University）

4:00 PM - 4:20 PM

[2A3-04] Understanding Ambiguous Instructions Using Generative Adversarial Nets for Object Disposal Tasks

○Komei Sugiura[1], Aly Magassouba[1], Hisashi Kawai[1] （1. NICT）

4:20 PM - 4:40 PM

[2A3-05] Tool-use Model Considering Selecting Tool by Deep Learning

Namiko Saito[1], Kitae Kim[1], ○Dai Ba Nguyen[1], Shingo Murata[1], Tetsuya Ogata[2], Shigeki Sugano[1] （1. Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University, 2. Department of Intermedia Art and Science, School of Fundamental Science and Engineering, Waseda University）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

共有表現の学習によるロボット動作と指示説明文の双方向変換

Bidirectional Translation between Robot Actions and Linguistic Descriptions by Aligned Representations

山田竜郎 *1*2 Tatsuro Yamada　松永寛之 *1 Hiroyuki Matsunaga　尾形哲也 *1 Tetsuya Ogata

*1早稲田大学　理工学術院 Faculty of Science and Engineering, Waseda University

*2日本学術振興会　特別研究員 Research Fellow of Japan Society for the Promotion of Science

We propose a novel deep learning framework to bidirectionally translate between robot actions and their linguistic descriptions. The model consists of two recurrent autoencoders, one of which is employed to make vector representations of robot actions and the other is for descriptions. The learning algorithm produces representations shared between actions and their descriptions by creating an additional loss function in which the representation of an action and that of its description become close to each other in the latent vector space. Across the shared representation, the trained model can produce a linguistic description given a robot action. The model is also able to generate an appropriate action by receiving a linguistic instruction, conditioned on the current visual input. Visualization of the shared representations shows that the robot actions are represented in a semantically compositional way in the vector space by being learned jointly with their descriptions.

1. はじめに

人間と協調して働くロボットは，言語的にコミュニケーションをとりつつ働くことが求められる．一方では，人間の言語指示に応えて正しい動作を生成する能力が必要である．また他方で，ロボット自身の生成した動作を文章で説明する能力も，挙動の解釈可能性の観点から鑑みて必要である．

近年，シーケンス間の対応を学習するシステムとしてエンコーダ–デコーダモデル（あるいは Sequence-to-sequence モデル）が注目されている．翻訳 [Sutskever 14]，音声認識，言語指示からロボット動作への変換 [Yamada 17]，ロボット動作から説明文への変換 [Heinrich 14] など多岐に用いられるが，これらはすべて一方向変換のみを実現する．本研究は，並列する二つのエンコーダ–デコーダモデルを用意し，その中間表現を共有するように学習を行うというアプローチにより，ロボット動作と指示説明文を双方向に変換可能なモデルを提案する．

2. 提案手法

2.1 モデル概要

本研究は，ロボットが (1) 言語指示に応えて動作を生成する能力と，(2) 自身の動作からその説明文を生成する能力の，双方を対データから学習するニューラルネットワークモデルおよび学習アルゴリズムを提案する．図 1 に示すように，本モデルは 2 つの Recurrent Autoencoder (RAE) [Srivastava 15] からなる．一方が単語シーケンス，すなわち文章を，もう一方がロボットの動作シーケンスを扱う．RAE は，エンコーダ–デコーダモデルの一種であり，出力シーケンスが入力シーケンスと同じになるように，すなわち恒等写像を目的として学習を行う．学習により RAE は，シーケンスを表す固定長ベクトル表現を，そのシーケンスを再生成可能な形で獲得する．

今回の提案モデルでは，各 RAE における復元誤差に加え，互いに対となる文章と動作の表現同士が近くなるように拘束する損失関数を設ける．これら二つの学習により，上記の二つの機能が実現することが期待される．(1) の機能は，指示文を言語 RAE がエンコードし，その表現を動作 RAE がデコードすることで実現する．(2) は反対に，動作 RAE のエンコーダ，言語 RAE のデコーダと順に順伝播計算を行うことで実現する．

連絡先: 尾形哲也，早稲田大学理工学術院，ogata@waseda.jp

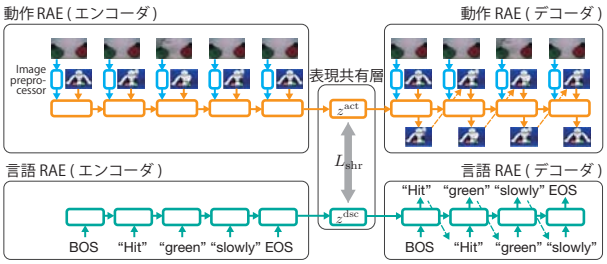


図 1: 双方向変換モデルの概要．

2.2 言語 RAE

本節では言語 RAE (図 1 下部) の詳細を述べる．エンコーダ RNN は式 (1), (2) によって，単語シーケンス $(x_1, x_2, ..., x_T)$ を固定長ベクトル z に埋め込む．

$$h_t^{\text{enc}} = \text{EncCell}(x_t, h_{t-1}^{\text{enc}}) \quad (1 \le t \le T), \tag{1}$$

$$z = W^{\text{enc}} h_T^{\text{enc}} + b^{\text{enc}}. \tag{2}$$

EncCell は，GRU や LSTM [Gers 00] などの任意の学習可能な再帰セルである．h_t^{enc} は時刻 t におけるエンコーダのセル状態で，h_0^{enc} は零ベクトルとする．W^{enc} と b^{enc} は学習可能な重みとバイアスである．エンコーディング後，デコーダ RNN は式 (3)-(5) によって表現 z を $(y_1, y_2, ..., y_{T-1})$ として展開する．

$$h_0^{\text{dec}} = W^{\text{dec}} z + b^{\text{dec}}, \tag{3}$$

$$h_s^{\text{dec}} = \text{DecCell}(y_{s-1}, h_{s-1}^{\text{dec}}) \quad (1 \le s \le T-1), \tag{4}$$

$$y_s = f(W^{\text{out}} h_s^{\text{dec}} + b^{\text{out}}) \quad (1 \le s \le T-1). \tag{5}$$

DecCell は EncCell 同様，任意の再帰セルであり，h_s^{dec} は時刻 s におけるデコーダのセル状態である．W^{out} と b^{out} は，出力層の重みとバイアスである．本研究では，単語シーケンスを 1-hot ベクトルの系列として表現するため，活性化関数 f とし

て softmax 関数を選ぶ．ここで，1-hot ベクトルは単語数と同じ次元数のベクトルであり，単語に対応する要素のみが 1 をとり，他は 0 となる表現形式のことである．y_0 には，文章開始シンボルを表す 1-hot ベクトルを与える．学習は入力と出力の交差エントロピー最小化として行われる．

$$L_{\mathrm{dsc}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\Big(-\sum_{w=1}^{W} x_{t+1}(w)\log y_t(w)\Big). \tag{6}$$

W は語彙数である．

2.3 ロボット動作 RAE

次にロボット動作 RAE の挙動について説明する．動作シーケンスはロボットの関節角度 $(j_1, j_2, ..., j_{T'})$ とそれに同行する視覚情報 $(v_1, v_2, ..., v_{T'})$ からなり，動作 RAE はこれらを連結した $((j_1; v_1), (j_2; v_2), ..., (j_{T'}; v_{T'}))$ を固定長ベクトルに埋め込む．モデルはほぼ言語 RAE と同様であるが，デコーダの挙動が一部異なる．式 (7), (8) が示すように，デコーダが予測するのは関節角度だけであり，視覚情報は外部的に与えられる．

$$h_s^{\mathrm{dec}} = \mathrm{DecCell}(v_s, y_{s-1}, h_{s-1}^{\mathrm{dec}}) \quad (1 \le s \le S'), \tag{7}$$

$$y_s = f(W^{\mathrm{out}} h_s^{\mathrm{dec}} + b^{\mathrm{out}}) \quad (1 \le s \le S'). \tag{8}$$

v_s が時刻 s における視覚入力であり，y_s は時刻 $s+1$ における関節角度 j_{s+1} を予測する．y_0 には初期姿勢 j_1 を与える．行動用 RAE の誤差関数としては，二乗誤差を使用する．

$$L_{\mathrm{act}} = \frac{1}{T-1}\sum_{t=1}^{T-1} ||j_{t+1} - y_t||_2^2. \tag{9}$$

このようなモデルとしたのは，本研究が，ロボットが生成するべき行動の状況依存性を扱うことを目的としているからである．以下，具体例を用いて説明する．例えば，“赤いベルを叩く”という文章に対応する動作シーケンスは，赤いベルがどこにあるかによって変化する．次節で見るように，これらの多様でありうる動作シーケンス群の表現は，すべて “赤いベルを叩く”という文章の表現と拘束される．その結果，これらのシーケンス群はそれぞれ全く異なるシーケンスとしてではなく，同一の意味を持ったシーケンスとして互いに近接した空間に埋め込まれることになる．それゆえ，動作 RAE のデコーダは，この意味的に埋め込まれた表現と，視覚入力により与えられる現在の状況から，正しい行動を生成することを求められる．

2.4 表現の共有

対となる文章と動作のシーケンスの表現を互いに近づけるために，二つの RAE の復元誤差に加え，式 (10) で表される損失関数を設ける．ここで，動作シーケンスの表現のバッチを $\{z_i^{\mathrm{act}} | 1 \le i \le N\}$，対応する文章の表現のバッチを $\{z_i^{\mathrm{dsc}} | 1 \le i \le N\}$ と表記する．N はバッチサイズである．

$$L_{\mathrm{shr}} = \sum_{i}^{N} \psi(z_i^{\mathrm{act}}, z_i^{\mathrm{dsc}}) + \sum_{i}^{N}\sum_{j\neq i} max\{0, \Delta + \psi(z_i^{\mathrm{act}}, z_i^{\mathrm{dsc}}) - \psi(z_i^{\mathrm{act}}, z_j^{\mathrm{dsc}})\} \tag{10}$$

関数 ψ は二変数間のユークリッド距離である．式 (10) の右辺第 1 項は，対となる表現を互いに近づけることを意味する．第 2 項は対でないシーケンス間の距離を互いに遠ざけることを意味する，ただし，対となる表現同士の距離より Δ 以上遠く離れているとき，この力は働かない（この条件を与えない場合，距離が無限大に発散してしまうことが，予備実験で確認された）．



動作名	動作の内容
PUSH-L-SLOW	左の物体をゆっくり押す
PUSH-L-FAST	左の物体を素早く押す
PUSH-R-SLOW	右の物体をゆっくり押す
PUSH-R-FAST	右の物体を素早く押す
PULL-L-SLOW	左の物体をゆっくり引く
PULL-L-FAST	左の物体を素早く引く
PULL-R-SLOW	右の物体をゆっくり引く
PULL-R-FAST	右の物体を素早く引く
SLIDE-L-SLOW	左の物体をゆっくり右に動かす
SLIDE-L-FAST	左の物体を素早く右に動か
SLIDE-R-SLOW	右の物体をゆっくり左に動かす
SLIDE-R-FAST	右の物体を素早く左に動かす

図 2: （左）実験環境と使用したロボット．（右）動作一覧．

2.5 学習方法

学習はランダムバッチを用いた勾配法によって行われる．全体の損失関数は式 (11) によって表される．

$$L_{\mathrm{all}} = \alpha L_{\mathrm{dsc}} + \beta L_{\mathrm{act}} + \gamma L_{\mathrm{shr}}. \tag{11}$$

α, β, γ は各損失関数の影響度合いを調整するハイパーパラメータである．全ての学習パラメータにおける勾配は，Backpropagation Through Time [Rumelhart 86] により計算される．

3. 実験デザイン

3.1 タスク概要

提案手法の有効性を評価するために，ロボット実験を行なった．実験用ロボットとして，小型のヒューマノイドロボット NAO を用いた．ロボットの前には色付きの立方体オブジェクトが 2 つ決められた位置に置かれる (図 2)．オブジェクトの色は赤，緑，黄のいずれかであり，2 つの色は必ず異なるようにするため，可能なオブジェクト配置は ${}_3P_2 = 6$ 通りである．各配置で，ロボットは図 2 に示す 12 通りの動作を行うことができる．それぞれの動作に対応する文章は，オブジェクトの配置によって異なるように設定した．具体的に説明すると，文章は動詞，目的語，副詞の 3 単語からなり，動詞と副詞はオブジェクト配置に依存せず動作のみによって決まるが，目的語部分には，動作の対象となったオブジェクトの色を示す語が入る．例えば左に赤，右に緑と置かれた場合における PUSH-L-SLOWLY の動作に対応する文章は，“push red slowly”である．すなわち可能な文章は，(push, pull, slide)*(red, green, yellow)*(slowly, fast) の 18 通りとなる．

3.2 データ

ロボットの動作軌道はあらかじめコンピュータ上で作成した．FAST 動作群は約 26 ステップ，SLOWLY 動作群は約 39 ステップとなるように動作を設計した．可能な 6 通りのオブジェクト配置で，12 パターンの動作を実際にロボットに生成させ，その際のロボットの腕の関節角度 10 次元，カメラ画像 (H: 120, W: 160, RGB) を収集した．6 × 12=72 パターンの可能な状況それぞれを，6 回収集した．画像は，別に用意した Autoencoder によってあらかじめ 10 次元のベクトルとして圧縮し，これをモデルの視覚入力として扱う．

対応する文章は，1-hot ベクトルの系列として表現した．文章の最初と最後には，開始終端シンボルを挿入した．

3.3 学習

72 パターンの可能な状況のうち，18 パターンを除外して学習を行った．言語 RAE のエンコーダとデコーダはそれぞれ 100 ノード 1 層の LSTM，動作 RAE のエンコーダとデコーダはそれぞれ 100 ノード 2 層の LSTM であり，共有層は 100 次元，損失関数の混合率である α, β, γ はすべて 1.0 とした．オプティマイザを Adam [Kingma 15] (学習率 0.001)，バッチサイズを 50 として，20,000 回の学習を行った．

4. 結果

4.1 説明文の生成

学習後の動作 RAE のエンコーダに動作シーケンスを入力して，その表現を元に言語 RAE のデコーダを計算することで，正しい文章が生成されるかを評価した．言語 RAE の出力層の活性化関数は softmax であるが，もっとも高い値をとった要素に対応する単語を，各時刻におけるモデルの出力と見なした．モデルは，既学習の 54 パターンと未学習の 18 パターンすべてにおいて正しい文章を生成することができた．

4.2 ロボット行動の生成

言語 RAE のエンコーダに文章を入力して，その表現を元に動作 RAE のデコーダを計算することで正しい動作が生成されるか，ロボットを使用して評価した．目的語の示すオブジェクトを，動詞が指示する方向に 3cm 以上動かすことができた場合成功と見なした．また，30 ステップ以内に初期姿勢近傍に戻った場合は FAST 動作，それ以上かかった場合は SLOWLY 動作を生成したと見なした．この条件で，既学習状況については 54 パターン中 36 パターン，未学習状況については 18 パターン中 12 パターンの動作生成に成功した．

失敗したケースについても，明らかに指示された動作と違う軌道を生成していたわけではなく，軌道はほぼ正しいが，わずかにずれたことでオブジェクトの正しい面に触れることができず，オブジェクトが指示された方向に動かなかったというケースがほとんどであった．そこで時系列の類似度を測る指標である Dynamic Time Warping [Senin 08] によって，失敗した際の軌道と，元の 12 パターンの動作の軌道との類似度を計算したところ，全てのケースにおいて，生成された軌道は正解の動作の軌道との類似度がもっとも高いということが分かった．

4.3 共有表現の解析

学習後の共有空間におけるシーケンスの表現を主成分分析によって可視化した．図 3 左は，18 本の文章の表現である．動詞，目的語，副詞のそれぞれの品詞について，体系的にエンコードされていることがわかる．一方，図 3 右は，72 パターンの動作シーケンスの表現であるが，文章の表現と正しく対応づけられていることがわかる．ここで注意されたいのは，オブジェクト配置によって，同じ動作でも違う文章に対応づけられ，反対に，異なる動作でも同じ文章に対応づけられていることである．すなわち，動作シーケンスは，視覚情報と総合することで意味的な行動の表現としてエンコードされた．

5. まとめと展望

本研究ではロボット動作とそれに対応する指示説明文を双方向に変換可能なエンコーダ–デコーダ型の RNN モデルを構築し，その有効性を評価するためロボット実験を行なった．提案モデルは与えられたタスクにおいて，動作と指示説明文の双方向の変換に成功した．また共有空間の可視化により，動作シーケンスが，言語表現と意味的に結び付けられていることが明らかとなった．今後は，より複雑なタスク下でのモデル評価や，聴覚・触覚情報も統合したモデルへの拡張を行う予定である．

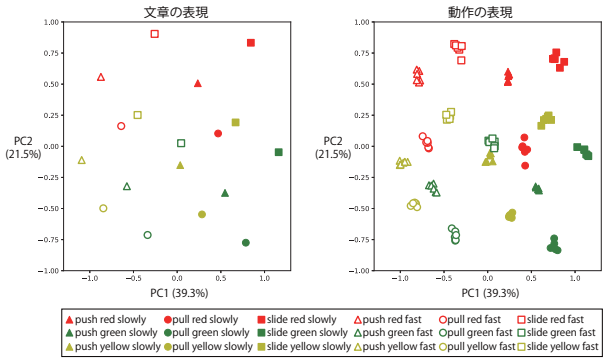


図 3: 主成分分析による文章と動作の表現の可視化．

謝辞

本研究は，科研費特別研究員奨励費 (17J10580)，JST, CREST (JPMJCR15E3)，文部科学省博士課程教育リーディングプログラム「実体情報学博士プログラム」の助成を受けた．

参考文献

[Gers 00] Gers, F. A., and Schmidhuber, J.: Recurrent nets that time and count. In Neural Networks, *Proceedings of the IEEE-INNS-ENNS Int. Joint Conf. on Neural Networks (IJCNN2000)*, Vol. 3, pp. 189-194 (2000).

[Heinrich 14] Heinrich, S. and Wermter, S.: Interactive Language Understanding with Multiple Timescale Recurrent Neural Networks, *Artificial Neural Networks and Machine Learning — ICANN 2014*, LNCS Vol. 8681, pp. 193-200 (2014).

[Kingma 15] Kingma, D., and Ba, J.: Adam: a method for stochastic optimization, in *International Conference on Learning Representations (ICLR2015)*, (2015).

[Rumelhart 86] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT Press), pp. 318-362 (1986).

[Senin 08] Senin, P.: Dynamic time warping algorithm review." *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA* 855, pp. 1-23 (2008).

[Srivastava 15] Srivastava, N., Mansimov, E., and Salakhudinov, R.: Unsupervised Learning of Video Representations using LSTMs, *International conference on machine learning*, (2015).

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, V. Q.: Sequence to Sequence Learning with Neural Networks, in *Neural Information Processing Systems 2014 (NIPS2014)* (2014).

[Yamada 17] Yamada, T., Murata, S., Arie, H., and Ogata, T.: Representation Learning of Logic Words by an RNN: From Word Sequences to Robot Actions, *Frontiers in Neurorobotics*, Vol. 11, No. 70, pp. 1-18 (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

軌道学習における試行回数削減のための強化学習手法

Trials Reduction Method for Reinforcement Learning in Trajectory Discovery

嘉藤 佑亮[*1,2] Yusuke Kato　中村 友昭[*3] Tomoaki Nakamura　長井 隆行[*3] Takayuki Nagai　山野辺 夏樹[*1] Takayuki Nagai　永田 和之[*1] Kazuyuki Nagata　小澤 順[*1] Jun Ozawa

[*1] 産業技術総合研究所 Advanced Industrial Science and Technology

[*2] パナソニック株式会社 Panasonic Corporation

[*3] 電気通信大学 The University of Electro-Communications

In recent years, there are many researches of deep reinforcement learning to realize autonomous motion of robots. In deep reinforcement learning, a large number of trials such as thousands of times or more are required to realize sufficient performance as a learning result. However, learning in a real environment often requires assistance by people, so it is difficult to do thousands of trials. In this research, we create a learning database from efficient reinforcement learning that utilizes knowledge about tasks given by people in advance, and realize learning with a relatively small number of trials by performing mini batch learning using that database. We apply our proposed method to learning of picking task in the logistics warehouse and show the usefulness of our proposed method by comparing the results with other methods.

1. はじめに

近年，ロボットの自律的学習の実現を目指して深層強化学習の研究が多くなされている[Finn 2016] [Gu 2017] [Levine 2016]. 深層強化学習の手法では一般的に学習を完了させるために多くの学習データが必要とされる[Levine 2016] [Pinto 2016]. 特に，学習のための行動を評価した報酬は各行動ではなく，一連の行動の結果として与えられるため，学習初期の段階ではどの行動が良い結果につながるかが分からない．そのため試行錯誤の軌道制御を繰り返す必要があり，1 つの学習に何千回，何万回といった試行回数が必要となる．しかしながら，実環境での学習，特にロボットを用いたマニピュレーションタスクではロボットの周囲環境の条件を整えるためにユーザの補助が必要となる場合が多く，何千回という回数の施行を行うことは難しい．

そこで，本研究では，ロボットが把持した物体をとある場所から別の場所へ移動させるための軌道の学習において，比較的少ない試行回数で学習を完了するための手法を提案する．人間があらかじめ軌道探索について方針を与えることでロボットを用いた効率的な探索を行い，そのときの状態遷移の情報についてデータベースを構築する．そして，構築したデータベースを用いてミニバッチ学習を行うことで，比較的少ない回数での学習を完了させる．

本手法の有効性を確認するために，倉庫環境における物品のピッキング作業を通して評価実験を実施した．その結果，100 回という比較的少ない試行回数で 95.6%の成功率を実現した．

2. 強化学習の試行回数削減手法

2.1 実環境での強化学習における課題

本研究ではロボットが把持した対象物体をとある地点から別の場所へ移動させるようなマニピュレーションを行うタスクについて考える．マニピュレーションのための軌道制御を図1左に示すようなロボットの手先位置の移動によって定義する．図 1 では，

連絡先：嘉藤 佑亮，産業技術総合研究所 知能システム研究部門，〒305-8560 茨城県つくば市梅園 1-1-1 中央第 1 つくば本部・情報技術共同研究棟，TEL:029-861-3537，E-mail:kato.yusuke@aist.go.jp

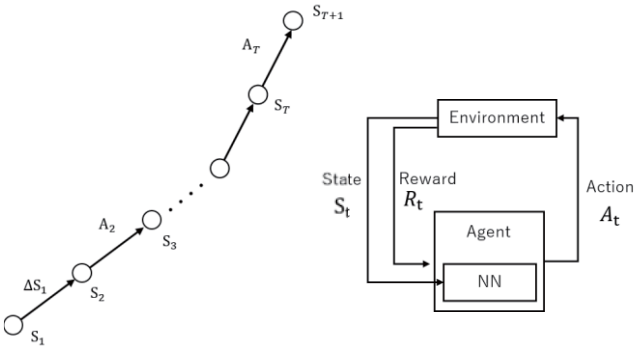


図 1 軌道制御（左）とシステム構成（右）

ある時刻tにおける状態として，ロボットの手先位置姿勢$P_t = (x_t, y_t, z_t,\ \varphi_t, \theta_t, \psi_t)$と手先にかかる力$F_t = (Fx_t, Fy_t, Fz_t)$とトルク$M_t = (Mx_t, My_t, Mz_t)$からなる状態$S_t = (P_t, F_t, M_t)$と定義し，次にロボットの手先の行動（移動量）を$A_t$で表現している．とある状態で選択した行動に対して，環境から報酬R_tを得ることで，もっとも高い報酬を得られるよう学習するシステムは図 1 右のように構築できる．

ロボットが把持した物体をとある地点から別の場所へ移動させるための軌道について考えるとき，軌道を構成する行動のひとつひとつに対して報酬を定義することは難しく，一連の行動の結果として報酬を与えることとなる．そのため学習初期においては，良い報酬を獲得できる軌道を発見するまでに多く試行を必要とする．また，実環境でロボットを用いて強化学習を実施した場合，ロボットが稼働する時間や動作後に対象物を初期位置へ戻すための時間が必要となるため，一般的に TV ゲーム等の学習で行われている何千回といった試行を行うことはあまり現実的ではない．

そのため，実環境ではなくシミュレーション上で環境を再現し，強化学習を行うシステムも提案されている[Mahler 2016]. しかしながら，対象物の変形，摩擦，干渉等の厳密な環境モデルを構築することが困難なため，使用できる環境には制限がある．

2.2 強化学習データベースを用いたミニバッチ学習

本論文では，実環境でロボットを稼働させながらも比較的少ない試行回数で軌道学習を実現するための学習システムを提案する．所定の施行回数で強化学習を実環境でロボットを用い

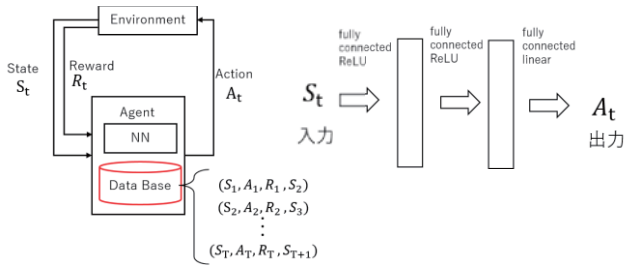


図 2 学習システム構成

て行い，そのときの状態，行動，報酬の情報を強化学習データベースに蓄積する．その後，先の学習で得た強化学習データベースを用いてミニバッチ学習を行い，その結果を最終的な学習結果として利用する．図 2 に本提案システムの概要を示す．

従来の Deep Q Network[Mnih 2013] [Mnih 2015]のような学習では少ない試行回数ではニューラルネットワークの学習回数も少なくなってしまうため，学習の結果が十分に安定しなかった．そこで本論文では，実環境でロボットを使用した強化学習時の状態，行動，報酬をすべて強化学習データベースにいったん蓄積し，その後，蓄積したデータベースを用いてミニバッチ学習を行うことでニューラルネットワークを十分に学習させる．実環境での強化学習から蓄積するデータベースのデータ構造を表1に示す．ここでは 1 回の施行を既定のT回の行動，または途中で行動不能となったときと定義する．総試行回数をN回の内，n回目の試行のt回目の行動での行動前の状態を$S_{n,t}$，選択した行動を$A_{n,t}$，行動後の状態を$S_{n,t+1}$，行動した結果の報酬$R_{n,t}$で表す．これら$(S_{n,t}, A_{n,t}, R_n, S_{n,t+1})$を 1 個としてデータベースに蓄積していく．蓄積するデータ数は行動不能で動作を途中で中断することがあるため，最小N個，最大$N \times T$個となる．N回の施行終了後，蓄積されたデータを用いて改めてミニバッチ法によりニューラルネットワークを学習させる．

2.3 人の知識に基づいた探索

従来手法では学習初期にランダムな探索を行い，学習を進めいくため，少ない試行回数での学習がどれだけうまく進むかはそのランダムな探索にどれだけ有効な情報を得られるかに大きく左右されていた．そこで本手法ではユーザが軌道に関する知識を提供することで探索に一定の方向性を持たせ，効率的な探索を行う．

表 1 データ

Num	*Step*	*State*	*Action*	*Reward*	*Next State*
1	1	$S_{1,1}$	$A_{1,1}$	$R_{1,1}$	$S_{1,2}$
	2	$S_{1,2}$	$A_{1,2}$	$R_{1,2}$	$S_{1,3}$
	…	…	…	…	…
	T	$S_{1,T}$	$A_{1,T}$	$R_{1,T}$	$S_{1,T+1}$
2	1	$S_{2,1}$	$A_{2,1}$	$R_{2,1}$	$S_{2,2}$
	2	$S_{2,2}$	$A_{2,2}$	$R_{2,2}$	$S_{2,3}$
	…	…	…	…	…
	6	$S_{2,6}$	$A_{2,6}$	$R_{2,6}$	$S_{2,7}$
3	1	$S_{3,1}$	$A_{3,1}$	$R_{3,1}$	$S_{3,2}$
	2	$S_{3,2}$	$A_{3,2}$	$R_{3,2}$	$S_{3,3}$
	…	…	…	…	…
…	…	…	…	…	…
N	1	$S_{N,1}$	$A_{N,1}$	$R_{N,1}$	$S_{N,2}$
	2	$S_{N,2}$	$A_{N,2}$	$R_{N,2}$	$S_{N,3}$
	…	…	…	…	…
	T	$S_{N,T}$	$A_{N,T}$	$R_{N,T}$	$S_{N,T+1}$

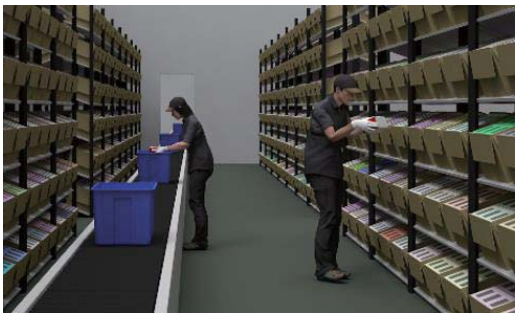


図 3 物流倉庫でのピッキング作業

従来の強化学習では，タスクに関する事前知識などは与えず，取りうる行動をランダムに探索を行っていた．しかしながら，本研究で対象とする把持した物体を移動させるタスクでは移動先の地点，または領域が存在するため，大まかな移動方向えお決めることができる．また，把持した物体を移動させる場合，移動の軌道としては急激な変化は少ない方が望ましい．このことから本手法では，探索のための方針として以下の 2 つを与える．

(1)初期の移動方向の教示

(2)急激な移動方向の変化の制約の教示

(1)に関して，移動のための軌道学習を行う場合，タスクの解となる全体の詳細な軌道は分からなくとも軌道の方向性は分かっていることが多い．そこでその方向を基準に重点的に探索することで，完全ランダムに探索を行うよりも効率よく質の良いデータを集める．

(2)に関して，把持した物体を移動させるための軌道制御においては，安定した運搬を行うために軌道はある程度滑らかな軌道の方が良く，極端な移動が極力避けるべきである．そこで，ランダムに探索を行うときの取りうる行動の選択確率に差を付け，現在の移動方向と同じ方向が選ばれやすくなるよう設計する．

3. 実験

この章では第 2 章で提案した手法を倉庫環境におけるピッキングタスクに適応し，学習させた例を示す．

3.1 倉庫環境におけるピッキング作業

本提案手法を用いて動作の学習を行うことに適した環境の一つとして物流倉庫環境を紹介する．物流倉庫内では小売店舗向けのピッキング作業が行われている．作業者は各小売店舗からの発注に応じた個数の商品をピッキングし，輸送用コンテナへ移動させる作業を行う(図 3)．物流倉庫内で扱われる商品はビニールやフィルム等で包装され（図 4a），箱の中に隙間なく整列して詰められた状態で棚に置かれている(図 4b)．我々は図 4b のように箱の中に隙間なく整列して詰められている状態を充

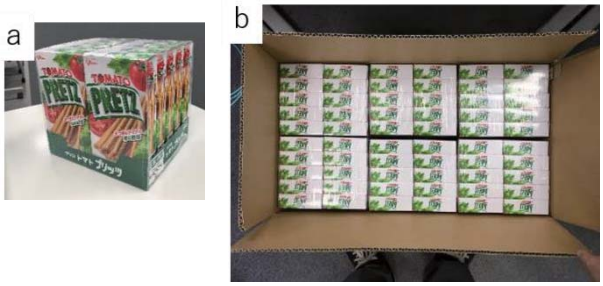


図 4 物流倉庫内の商品の一例

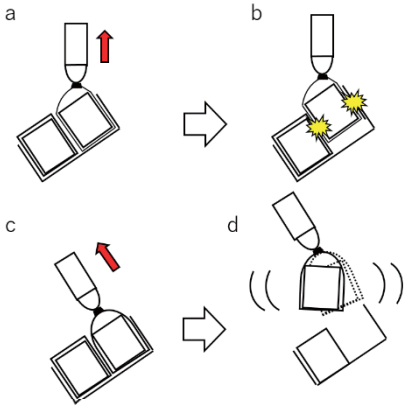


図 5 物流倉庫環境での課題

填状態と呼んでいる. また, 商品の置かれている棚の棚板には取り出しやすさを考慮して傾斜がつけられていることが多い.

以上のような環境である物流倉庫内でロボットシステムが自律的にピッキング作業を行うことを考える. 物流倉庫内の環境で吸着によって商品を持ち上げて取り出すときに課題が 2 つある. 1 つは, 持ち上げる方向である. 対象物を吸着によって把持し, 地面と垂直方向に持ち上げようとしたところ, 物体同士の隙間が小さいため, 傾斜をつけて置かれているこの環境下では周囲の物体と衝突し, 持ち上げることができなかった(図 5a-b). そのため, ピッキング動作の初めは周囲の物体と干渉しない方向に持ち上げる必要がある.

もう一つはピッキング動作の軌道である. 棚についている傾斜に対して垂直方向に引っ張ると, 周囲と干渉することなく持ち上げることができる. しかし, 箱から対象物の全体が取り出されたときに対象物が回転し大きく揺れてしまい吸着が外れ落とすことがある(図 5c-d).

これらのことから物流倉庫内のピッキング作業では単純な方法によるピッキングは成功せず, 対象物を落とさないように取り出せる軌道を見つける必要がある.

また, 物流倉庫内のピッキング動作はビニールの伸びや物体同士の摩擦などに影響を受ける. そのためシミュレーション上で動作学習をさせることは難しく, 実環境で行う必要がある.

3.2 実験システムの構成

図 6a のように倉庫環境を模した実験環境を用意した. 棚板に 20 度の傾斜がついた棚に充填状態にある商品が置き, その正面に手先に吸着機構を備えたマニピュレーションロボットを設置している. 対象物体の 1 つあたりの大きさは 17[cm]×14[cm]×16[cm], 重さ 915[g]であり, 1 つの段ボール箱の中に 12 個詰められている. ロボットは KUKA 社の 7 軸多関節ロボットアーム LBR iiwa14 R820 を使用した. 図 5b にロボットの手先部分を示す. 手先には真空吸着式の吸着機構と力覚センサが備え付けられている. 力覚センサは wacoh-tech 社の静電容量型 6 軸力覚センサ DynPick を使用しており, 3 方向の力成分と 3 方向のモーメント成分の 6 軸を同時に検出することが可能である. 以上のロボットシステムを使用し, 棚に置かれた充填状態の箱から対象物体を 1 つ吸着し取り出す軌道を学習させる.

3.3 学習動作

学習動作は, 物体の取り出し軌跡であり, 2.2 節で紹介したような各状態における一連の行動として定義する. 今回の実験では 1 回の行動での移動量は 2.0[cm]と固定し, 各状態における移動方向の変化を学習させる. なお, 手先の姿勢は移動方向に合わせて変化させるものとする.

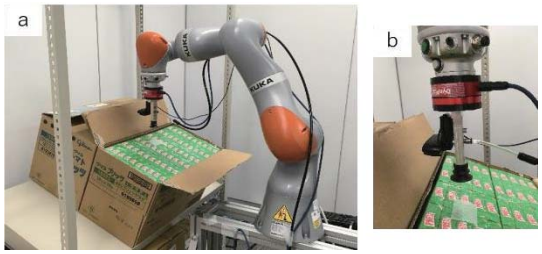


図 6 実験システム

まずロボットはユーザからあらかじめ与えられた手先位置で図 5c のように対象物に対して垂直となる姿勢で吸着を行う. この位置姿勢をロボットの手先の初期の位置姿勢とする. この初期の位置姿勢から吸着したまま 2[cm]単位の移動を繰り返し行い, 12 回移動する, または 12 回の移動を終えるまでに対象物体から吸着が外れたとき動作を終了する. 終了後, 実験者によって対象物体を元の位置に戻す. ここまで一連の流れを 1 試行として扱う.

また, 本実験環境ではロボットの手先移動はx-z 平面上の移動のみを行い, y 軸方向への移動は行わない. 同様に手先姿勢についても y 軸まわりの回転だけを行い, x,z 軸の周りの回転は行わない.

3.4 軌道学習のための制御プログラム

動作学習のための制御プログラムをオープンソフトウェアである深層学習ライブラリ chainerrl を利用し, Deep Q Network（DQN）の手法を基本として実装した.

2.3 節, 3.1 節で説明したタスクに関する知識を本実験における人の知識として実装に組み込む. 本実験ではロボットが取りうる行動を, 手先を2[cm]移動するときの移動方向の変化量として表す. 変化量は-10[deg]から 10[deg]までの 5[deg]刻みの値とし, 変化量が 0[deg]のとき, 現在の移動方向のまま移動することになる. また, 探索の手法として Linear Decay ε-greedy 法があるが, 本実験ではこれに変更を加えたものを探索の手法として使用する. 通常の Linear Decay ε-greedy の手法では行動ごとに線形的に減少するε の値を用いて, ε の確率でランダムな行動（探索）, 1－εの確率で最も Q 値の高い行動を選択するが, 本実験ではε の確率での探索時にランダムに行動選択をするのではなく, 移動方向の変化を 0[deg]とする行動が他の行動に比べて選択されやすくなるように設計する. 今回の実装では, 探索時に移動方向の変化を 0 度とする行動が選ばれる確率を 50%とし, 残りの 4 つの行動はそれぞれ 12.5%の確率で選ばれるように設計した.

学習に使用するロボットの手先位置などの関係を示した図を図 7 に示す. 学習に際して, ニューラルネットワークの入力となる状態として対象物を吸着した位置（初期位置）からの現在のロボットの手先の位置へのベクトルと手先姿勢, 力覚センサから得られた手先に掛かる力とモーメントを用いた. ただし, 本実験では y 軸方向への移動と x,z 軸周りの回転は行わないため, y 座標, y 方向の力, x,z 軸周りのモーメントは入力として使用しない. ネットワークはチャンネル数 30 の中間層 2 層からなるものを使用する. 出力として, 手先の現在の移動方向の変化量を-10[deg]から 10[deg]までを 5[deg]刻みの 5 つの選択肢に対する Q 値を出力する. 図 8 にネットワークの構造を示す.

以上のような設定を用いて強化学習による動作学習を進め, 2.2 節で紹介したように状態と行動, 報酬をデータベースに蓄積していく. この実験では 100 試行分のデータを蓄積させた. そして収集したデータを用いて改めてミニバッチ学習を用いてニュ

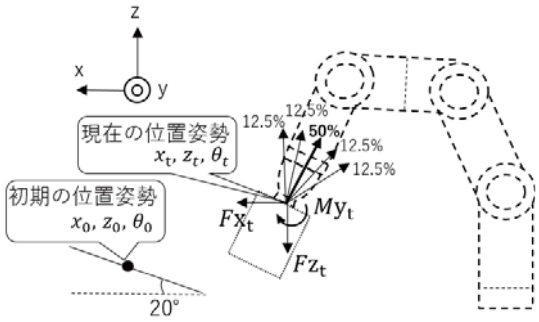


図 7 実験システムの概略図

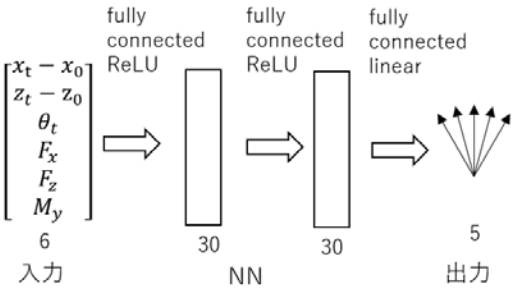


図 8 ネットワーク構造

ーラルネットワークの学習を 100epoch 行い，その結果を最終的な学習結果として使用する．このときの学習に使用するネットワークの構造も図 8 と同じものを使用する．

4. 評価

ここでは 3. 4 節で紹介した学習を行った結果を用いて，取り出し動作を実施し，本提案手法の有用性を評価する．

評価のために学習データベース(DB)や人の知識の有無で下記の 4 条件を用意し，ぞれぞれの条件での学習結果を用いた物流倉庫内の取り出し動作の成功率を比較する．

(a) 強化学習 DB と人の知識，どちらも使用しない学習
(b) 強化学習 DB を使用した学習
(c) 人の知識を使用した学習
(d) 強化学習 DB と人の知識を利用した学習（提案手法）

試行回数を 100 として各条件で学習を 3 回ずつ行い，それぞれの学習結果を用いて 30 回ずつの取り出し動作を行った．

各条件での取り出し動作の成功率を表 2 に示す．本提案手法は 100 回という比較的少ない試行回数の強化学習において全体で 95.6%と高い成功率を示し，学習ごとのばらつきも小さくなった．強化学習 DB を使用していない条件 a,c は，成功率に大きく差が出ており，全体的な成功率も 50.0%，62.2%となった．条件 b は条件 a,c に比べると学習ごとのばらつきは小さくなっているが，全体的な成功率は 87.8%と条件 d よりも低い結果となった．

今回の実験では 1 種類の物体のみを対象として学習を行った．そのため，重さや大きさが変わったときに，今回の学習結果が利用できるかは確認できてない．学習時に入力として対象物体の情報を使って無次元化すれば，ある程度の汎用性は出せると予想できる．しかしながら，物体同士の隙間がより狭いなど環境が大きく異なる場合は，求められる軌道も変わってくるため，その環境に適した学習が必要になると考えられる．

また，今回の学習の特徴として，学習の中に人の知識を組み込んだが，人の知識は画一的なものではなく，その人の熟練度によって内容に差が生じる．そのため，知識の質の良し悪しが学習結果に影響を及ぼすことも考えられる．

表 2 学習結果の比較

学習＼条件	a	b	c	d
1	50.0	73.3	56.7	100.0
2	0.0	100.0	93.3	100.0
3	100.0	90.0	36.7	86.7
全体	50.0	87.8	62.2	95.6

5. おわりに

本稿ではロボットが把持した物体をとある場所から別の場所へ移動させるための軌道の学習において，比較的少ない試行回数で学習を完了するための手法を紹介した．所定回数の強化学習を行い，その時の状態，行動，報酬をデータベースに蓄積する．そして蓄積したデータベースを用いて後に再度ミニバッチの手法で学習させる．この手法により比較的少ない施行回数で安定した学習を実現する．また，データの蓄積の際に人から与えられた教示に基づいて学習を進めることで，効率よく有用なデータを収集する．評価のため本提案手法を用いて学習した結果を，データベースや人の知識を使わないで学習させたものと比較した．その結果データベースや人の知識を使わないで学習させた結果が 50.0%，62.2%，87.8%だったの対し，提案手法は 95.6%の成功率を示した．

参考文献

[Finn 2016] Finn, Chelsea, et al. "Deep spatial autoencoders for visuomotor learning." Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016.

[Gu 2017] Gu, Shixiang, et al. "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates." Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, 2017.

[Levine 2016] Levine, Sergey, et al. "End-to-end training of deep visuomotor policies." The Journal of Machine Learning Research 17.1, 1334-1373, 2016.

[Levine 2016] Levine, Sergey, et al. "Learning hand-eye coordination for robotic grasping with large-scale data collection." International Symposium on Experimental Robotics. Springer, Cham, 2016.

[Mahler 2016] Mahler, Jeffrey, et al. "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards." Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016.

[Mnih 2013] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 , 2013.

[Mnih 2015] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 , 2015.

[Pinto 2016] Pinto, Lerrel, and Abhinav Gupta. "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours." Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層強化学習と言葉による離散化を用いたロボット制御への取組み

An Approach to Robot Control using Deep Reinforcement Learning and Discretization by Words

橋本　さゆり *1　　金子　晃 *2　　小林　一郎 *3
Sayuri Hashimoto　　Akira Kaneko　　Ichiro Kobayashi

*1お茶の水女子大学人間文化創成科学研究科理学専攻情報科学コース
Ochanomizu University

*2*3お茶の水女子大学基幹研究院自然科学系
Ochanomizu University

In recent years, the necessity for robots working in society has been spreading as the aging society has come. To easily be able to communicate with robots, it is expected that they can understand natural language and learn how to behave spontaneously through the interaction with humans. In this study, we aim to ground the meaning of natural language onto their behaviors by using reinforcement learning. In particular, we have proposed an efficient method to learn robot's motion with deep reinforcement learning by descritizing a robot's motion into a hierarchical structure consisting of basic motion elements which can be represented by words.

1. はじめに

近年，ロボットの普及が広がりつつある．今後ロボットが家庭に入ってくることが予想される．その際に人が行なった動作を見てロボットが自発的に学習することや，その学習した動作の意味を理解して人とコミュニケーションをとることが考えられる．そこで本研究では，ロボットに動作を学習させ，その学習した動作と言葉のグラウンディングを取ることでロボットに動作の意味を理解させることを試みる．具体的には，物理シミュレータ MuJoCo の上で動くロボットアーム (以下簡単のため ロボットと略称する) を用い, 以下の図 1 のように右側の白い円柱とタブがある状態から, 左側の状態にするためにはロボットはどのような動作を行なわなければならないかを学習させる．

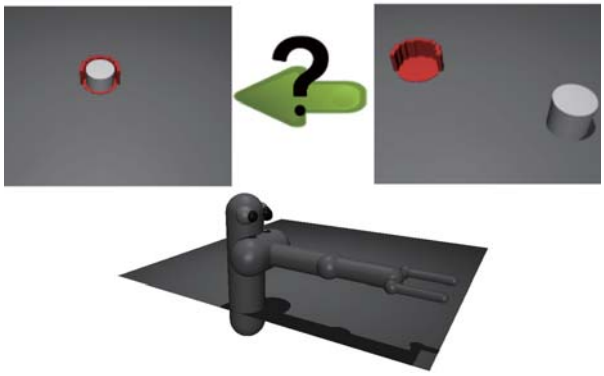


図 1: 動作の概念

本研究では，ある動作はいくつかの動作が組み合わさってできていると考える．例えば，この具体例のようにロボットが白い円柱を赤いタブにはめるためにはまずロボットが腕を移動させ，白い円柱を掴み，次に掴んだ円柱を運んで，押すという 4 つの動作の構成要素に分解される．さらに掴む，運ぶ，押すという動作は下腕を上げる，下げる，指を開く，閉じるといったロボットにとっての基本的な動作の単位から成ると考える．

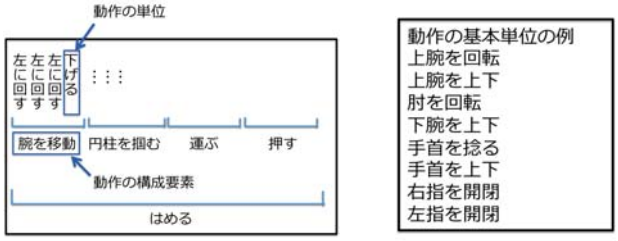


図 2: 動作の構成

2. 深層強化学習

ロボットに動作を獲得させるには強化学習という技術が用いられる．この技術は探索と評価を繰り返して行なうことで最適な行動を学習するものである．全体の枠組みとしては，学習を行なうエージェント（本研究ではロボット）がまず環境から状態を観測し，その観測結果を元に行動を行ない，その行動に対する報酬を貰うことで評価を行なう．エージェントは累積報酬の最大化を目指し，行動を行なっていく．一方で深層学習は，高次元データを扱い，高精度な結果を出すことができる．深層強化学習はこの二つの技術を組み合わせたもので，強化学習のアルゴリズムに深層学習による関数近似を組み込むことで強化学習における高次元で複雑な問題に対しても対処できるようになっている．以下に代表的な３つの深層強化学習の手法について説明する．

2.1 Deep Q-Network

Deep Q-Network（DQN）[2] は Atari のゲームを学習させて様々な Atari のゲームで人間以上の高得点を出したことで有名になった．DQN の元となっているのは強化学習における Q 学習というものである．Q 学習とは各状態において、可能な行動の中で最も行動価値関数の値が高い行動をとるように学習

連絡先: 橋本さゆり，お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース小林研究室，〒 112-8610 東京都文京区大塚 2-1-1，g1320530@is.ocha.ac.jp

を行なう方法であり，行動価値関数の最大値で行動を決定していくアルゴリズムである．元来 Q 学習では Q-table を用いるのが一般的であるが， Q-table では高次元データに対して次元の呪いを起こしてしまうなどの問題点があった．DQN ではこの Q-table の代わりに状態を入力とし，それぞれの行動に対する Q 値を出力とするネットワークを生成し関数近似を行なっている．

2.2 DDPG

DQN では，行動が離散的な場合しか扱うことができなかった．Deep Deterministic Policy Gradients（DDPG）[3] は行動が連続なものに対して使うことができる．DDPG の元となっているのは強化学習における Actor-critic という手法である．Actor-critic とは価値関数とは独立に，方策を表現する構造を別に持っており，Actor に基づき行動選択，Critic が報酬を得て誤差を計算するというアルゴリズムである．Actor-critic は特にロボットなどの強化学習に使用されることが多い．これらを深層化したものが DDPG である．

2.3 A3C

Asynchronous Advantage Actor-critic method（A3C）では，DDPG と同様に強化学習における Actor-critic という手法が元になっている．A3C で最も特徴的なのは GPU を用いず，CPU のスレッドを使って複数のエージェントを学習させることである．それぞれのスレッドで別々のエージェントが学習を行ない勾配更新を非同期に行なうことで，より少ない時間で高性能な学習をすることができるという利点を持つ．

我々は最初に DQN を用いて実験を開始し，ついで DDPG を用いてみた．しかし，DQN はパラメータ調整が難しく学習が困難であった．また DDPG は行動が連続値の場合にしか使用できず，学習に多大な時間がかかることから我々のロボットシミュレータが動く PC 環境では学習を完成させるには至らなかった．よって本研究では最終的に A3C を用いることとした．

3. 動きの辞書による動作の離散化

本研究では実験の際に動作を以下のような辞書にして学習を行なっている．

名称	方向	ID	方向	ID	動きの単位
上腕を回転	左	A	右	B	0.05×60
上腕を上下	下	C	上	D	0.05×60
肘を回転	右	E	左	F	0.05×60
下腕を上下	下	G	上	H	0.005×60
手首を捻る	右	I	左	J	0.05×60
手首を上下	下	K	上	L	0.05×60
右指を開閉	左	M	右	N	0.05×60
左指を開閉	左	O	右	P	0.05×60

この表にある動きの単位とは，ロボットの各関節の最小単位の動作 0.05 を 60 回繰り返したものを指している．これによってロボットの連続な動作を離散化し

1 学習時間の短縮
2 言語との対応付けの容易化

という効用を期待した．

4. 実験

動作を離散化した辞書と深層強化学習を用いて, ロボットに「はめる」という動作を学習させる. 具体的に学習させた内容は, 既に述べた通りロボットが白い円柱を掴み, 赤いタブまで持っていくという課題である. 本実験では ChainerRL 中の example として実装されていたものを我々の目的に合わせ，修正して使用した．

4.1 使用したロボット

本研究では,MuJoCo[*1] [4] という商用物理シミュレータを使用した. またロボットアームは GitHub 上で 公開されていた Pusher [*2] を改変し 2 本の指を付け加えた ‘Picker‘を使用している. このロボットモデルは全部で 8 自由度を持ち, 構成は図 3 のようになっている．

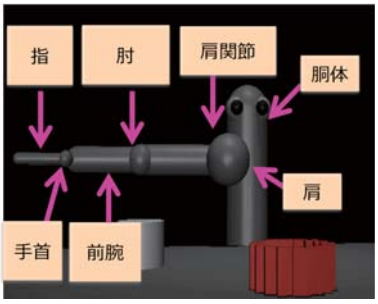


図 3: ロボットの構成図

4.2 実験設定

上記のロボットを OpenAIGym[*3] というフリーの強化学習環境に組み込んで, ChainerRL[*4] の上で深層強化学習を行なった. 上述した A3C を用い，1 エピソードを 300 回の試行とし, 全 80,000,000 エピソード学習させる実験を行なった．

円柱を掴んではめるといった動作はいくつかの動作の構成要素から成り立つ．本研究では，それぞれの構成要素に対してその動作を実現するための行動価値関数を設定し，動作の系列に従ってそれらの行動価値関数を重み係数を持つヘビサイド関数を掛けて累加していくことにより，最終的な動作達成のための単独な行動価値関数を構築している．
具体的には

1 白い円柱に腕を近づける
2 白い円柱を掴む
3 白い円柱を赤いタブまで持っていく

というような 3 つの段階に動作を細分化し，その各々の条件をクリアするまでその後の行動の分が報酬に寄与しないように行動価値関数を設計した．

4.3 実験結果

以下に実験結果を示す．それぞれ取り上げたエピソード番号は報酬の値がそれまでで最も高かったものを取り上げている．

- 100,132 エピソード学習時（図 4 参照）

100,132 エピソード学習時のモデルは以下のようになった．初期の状態から上腕を y 軸方向に回転させた状態で，少し腕を下げる動作をしていた．その後白い円柱に腕全体を持っていく動作を行なっていた．掴むという動作をしないまま円柱を赤い

*1 http://www.mujoco.org/
*2 https://github.com/openai/gym/pull/557
*3 https://github.com/openai/gym
*4 https://github.com/chainer/chainerrl

タブに持って行ってしまい,「掴む」という動作の学習は上手くできていなかった.

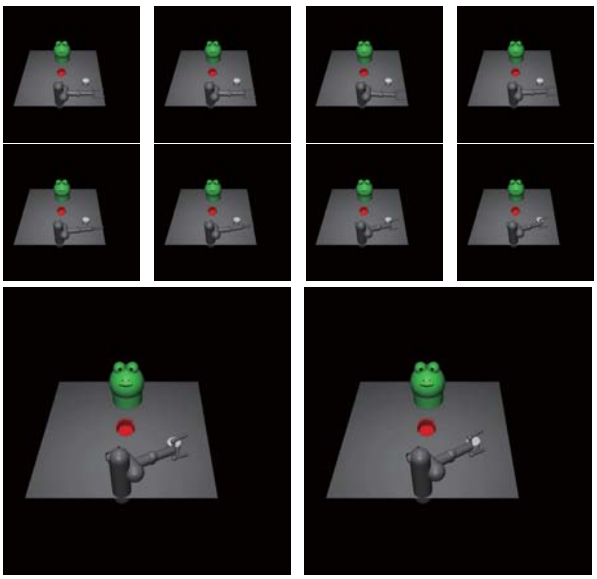


図 4: 100,132 エピソード学習時

- 200,327 エピソード学習時（図 5 参照）

200,327 エピソード学習時のモデルは以下のようになった．初期の状態から腕を平行に移動させ，白い円柱の上まで持っていく様子が確認された．しかし，その後腕を下に降ろすことがなく，円柱を掴むことはできなかった．

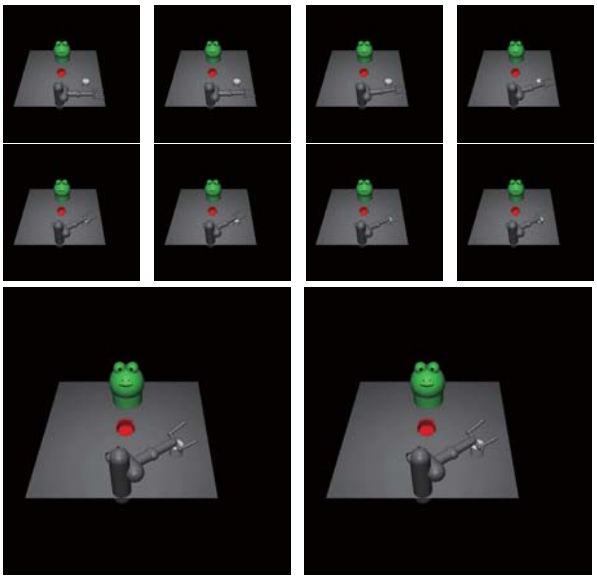


図 5: 200,327 エピソード学習時

- 300,570 エピソード学習時（図 6 参照）

300,570 エピソード学習時のモデルは以下のようになった．初期の状態からすぐに腕を平行に移動させ，白い円柱の上まで持っていき，その状態から腕を下ろすことを確認した．しかし，腕を下ろして掴んだ後に白い円柱を動かすような動作は見られなかった．

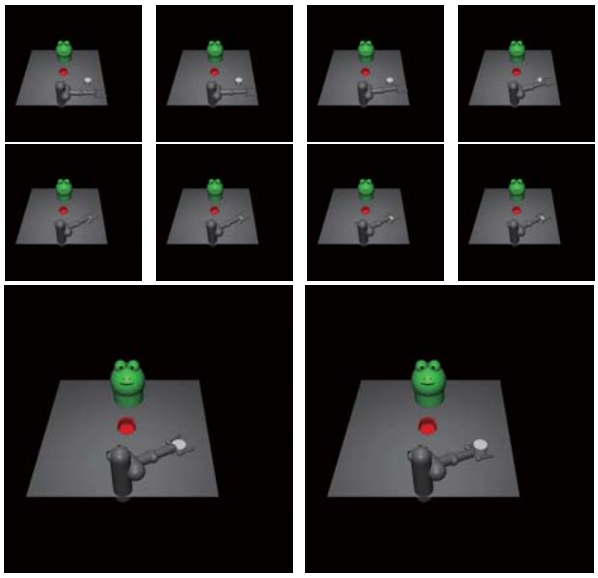


図 6: 300,570 エピソード学習時

- 1,100,031 エピソード学習時（図 7 参照）

1,100,031 エピソード学習時のモデルは以下のようになった．初期の状態からすぐに腕を平行に移動させ，白い円柱の上まで持っていく様子が確認された．その状態から平行に下に腕を下ろし，掴んでいる様子が確認できた．掴んだ後に白い円柱に赤いタブを近づける動作も見られた．しかし，指が赤いタブに当たってしまい，しっかりとはめることはできなかった．

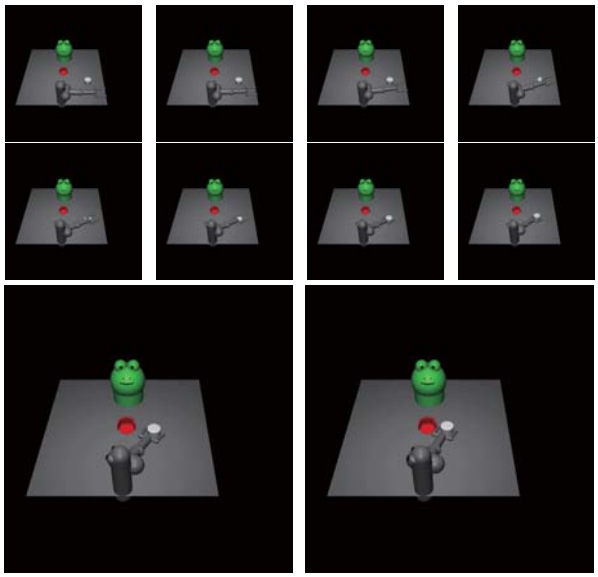


図 7: 1,100,031 エピソード学習時

- 4,600,083 エピソード学習時（図 8 参照）

4,600,083 エピソード学習時のモデルは以下のようになった．初期の状態から腕を滑らかに平行に持っていき，腕を下に下ろして掴んだ後に白い円柱を赤いタブに近づける動作を確認できた．赤いタブに最も近づけていたが，一時的に円柱をはめた後にそこから取り出そうとするような動作を行なっていた．

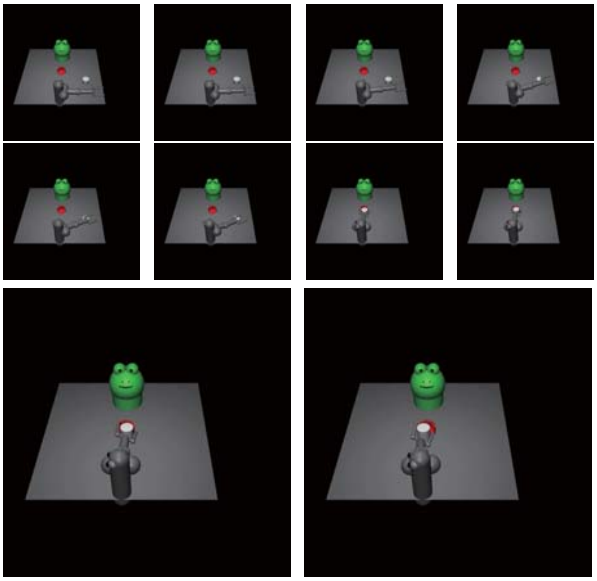


図 8: 4,600,083 エピソード学習時

4.4 考察

白い円柱を掴んで赤いタブにはめるまでの動作を一連の動作として学習させた．「はめる」という動作は腕を移動し，白い円柱を掴んで，赤いタブまで持っていくという動作を組み合わせることで構成されているが，この動作を一連の動作として学習させるために，行動価値関数を動作の構成要素に従って重み係数を掛けて累加していくという工夫を行なった．はめるという動作は人間にとっては簡単なように感じるが，ロボットにとっては複雑な動作であることがわかった．

5. まとめ

本研究では，動作を離散化した辞書と深層強化学習を用いてロボットに「はめる」という動作を学習させることを試みた．動作を離散化し辞書化することにより，学習時間の短縮と言語との対応付けの容易化を試みた．具体的には「はめる」という動作を学習させた．この動作は複雑であることから報酬を白い円柱に腕を近づける，白い円柱を掴む，白い円柱を赤いタブまで持っていく，という 3 段階に分けて考え，それらの動作に対する行動価値関数を重み付き係数を掛けて累加させることによって得られる単独の行動価値関数を用いて学習させた．今後の展望として，今回は白い円柱の初期位置を一定にして学習を行なっているが，今後は白い円柱の初期位置をランダムに与えて学習させたいと考えている．また，今回細分化して学習させた動作を別の動作の学習の際に使用するなど転移学習に応用できるかどうか試みたい．

参考文献

[1] Volodymyr Mnih,Adrià Puigdoménech Badia,Mehdi Mirza, Alex Graves,Tim Harley,Timothy P. Lillicrap,David Silver,Koray Kavukcuoglu,“Asynchronous methods for deep reinforcement learning”, PMLR 48:1928-1937, 2016.

[2] Volodymyr Mnih,Koray Kavukcuoglu,David Silver,Daan Wierstra, Alex Graves,Ioannis Antonoglou and Martin Riedmiller,“Playing Atari with Deep Reinforcement Learning”,In NIPS Deep Learning Workshop. 2013.

[3] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, NicolasHeess, Tom Erez, Yuval Tassa, David Silver and Daan Wierstra,“Continuous control with deep reinforcement learning”, ICLR2016.

[4] Emanuel Todorov, Tom Erez and Yuval Tassa, “MuJoCo: A physics engine for model-based control” IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

生活支援ロボットにおける Generative Adversarial Nets を用いた曖昧な指示の理解

Understanding Ambiguous Instructions Using Generative Adversarial Nets for Object Disposal Tasks

杉浦 孔明　マガスーバ アリー　河井 恒
Komei Sugiura　Aly Magassouba　Hisashi Kawai

情報通信研究機構
National Institute of Information and Communications Technology

This paper focuses on a multimodal language understanding method for “Carry and Place” tasks with domestic service robots. We address the case of ambiguous instructions, that is when the target area is not specified. For instance “Put away the milk and cereal.” is a natural instruction where there is ambiguity on the target area, considering daily life environments. Conventionally, this instruction can be disambiguated from a dialogue system, but at the cost of time and cumbersomeness. Instead, we propose a multimodal approach, where the instructions are disambiguated from the robot state and environment context. We develop MultiModal Classifier Generative Adversarial Network (MMC-GAN) to predict the likelihood of the different target areas considering the robot physical limitation and the target clutter. Our approach, MMC-GAN, significantly improves accuracy compared to baseline methods using instructions only or simple deep neural networks.

1. はじめに

生活支援ロボットによる曖昧な音声命令の理解は，多くの関連課題を有する挑戦的な分野である [杉浦 16]．本研究では，特に Carry and Place タスクにおける曖昧な命令文の理解を扱う．Carry and Place タスクとは，ユーザに指示された物体を適切な場所に移動させるタスクを指すものとする．例として，日常環境で「ペットボトルを片付けておいて」という音声命令をロボットが実行するタスクを考える．人間同士の場合はこのような省略された命令文で通じる場合が多いが，ロボットが行動を開始するために十分な情報を含んでいない．一方，十分な情報を含む命令文は，不自然であることが多い．例えば,「現在把持中のペットボトルをキッチンの棚の３段目の右側に片付けておいて」のような発話をユーザが行うことは考えにくい．

上記の問題への単純なアプローチとして，スロット値がすべて確定するまで聞き返す戦略が考えられる [Kollar 13, Gemignani 15]．実際に，世界最大の生活支援ロボットのベンチマークテストであるロボカップ＠ホーム [Iocchi 15] においても，このアプローチが支配的である．しかしながら，このアプローチでは「どのペットボトルですか？」「キッチンのどの棚ですか？」「棚の何番目の段ですか？」など多くの確認発話が生成されるため，動作実行するまでに必要な時間が長く，不便である．

本研究では，上記のタスクに対し，マルチモーダル情報を用いて省略内容の推定を行う．具体的には，曖昧な命令を入力として，Place 動作の対象領域の尤度を予測する手法を提案する．そのために，[Sugiura 17] で提案された Latent Classifier Generative Adversarial Nets (LAC-GAN) を拡張し，マルチモーダル情報入力を可能とする．

GAN [Goodfellow 14] に関する先行事例としては，Conditional GAN [Mirza 14] や InfoGAN [Chen 16] などが挙げられる．また，AC-GAN [Odena 16] のように，カテゴリを出力に利用した GAN も提案されている．これらの手法は，画像や文の生成などに適用され，品質の良い疑似サンプルの生成が報告されている．さらに，近年，GAN を分類問題に適用した研究として，文献 [Springenberg 15, Shen 17] が挙げられる．ロボティクスにおいては教師データ収集が高コストであるため，Generative Adversarial Nets (GAN) によるデータ拡張と分類を同時に可能な提案手法が重要性を持ち得る．この事実に着想を得て提案された手法が，LAC-GAN [Sugiura 17] である．LAC-GAN と提案手法 MMC-GAN（MultiModal Classifier GAN）の差異は，マルチモーダル情報を扱う Extractor を持つ点である．

本研究の独自性は以下である．

- GAN によるデータ拡張と分類を同時に行う手法 MMC-GAN を提案する．
- Carry and Place タスクにおける曖昧な命令理解に MMC-GAN を適用する．

2. 問題設定

本研究の対象タスクは,「Cary and Place」タスクにおける曖昧な命令（「お茶を片付けて」など）を理解することである．この命令に対する出力として，適切な対象領域を予測する．本研究では，以下の入出力を想定する．

- 入力： 命令文，状態を表す文，（環境中を巡回して得た）対象領域のカメラ画像
- 出力： オブジェクトを片付ける対象領域の尤もらしさ．

ユーザインタフェースとしてタブレット端末を想定し，対象領域（数十箇所以上）は画像でユーザに提示されるものとする．全対象領域を一度にユーザが視認することはできないため，正確な尤度スコアに応じて対象領域をランキングすることが利便性の向上（ユーザタップ数の減少）につながる．

本研究では，様々な実験条件に対して複数回の実験を行うことを主眼とするため，実環境・実ロボットでの実験ではなく，データセット上の分類問題としての統計的評価を論じることを優先する．一方，提案手法をロボットへ移植する際に，学習済みのモデルが実環境とミスマッチになることは避けなければならない．そのため，ハードウェアとしては標準化された生活支援ロボットと同等のデバイスを用いる．実際に，本研究で構築

連絡先: 杉浦孔明，京都府相楽郡精華町光台 3-5

したモデルはそのままロボットに移植可能である．ロボットを用いた動画に関しては，以下の URL[*1] を参照されたい．

生活支援ロボットのハードウェアとして，標準化されたロボットプラットフォームを前提とする．本研究では，HSR（Human Support Robot）[寺田 12] をプラットフォームとして想定する．世界最大の生活支援ロボットのベンチマークテスト RoboCup@Home において，HSR は標準化されている．

3. 提案手法: Multimodal Classifier GAN

3.1 LAC-GAN

LAC-GAN [Sugiura 17] は，Extractor E，Generator G，Discriminator D，の 3 つの部分から構成される．LAC-GAN の構造を図 1 に示す．

いま，学習サンプルが，$(\boldsymbol{x}_{raw}, y^*)$ の組で与えられるものとする．ここに，$\boldsymbol{x}_{raw} \in \mathbb{R}^N$ は E による抽出前の特徴量，y^* は対応するカテゴリのラベルである．y^* は d_y 次元の one-hot ベクトルであるものとする．

E の目的は，特徴抽出前の入力 $\boldsymbol{x}_{raw}$ から，分類に適した特徴量 $\boldsymbol{x}_{real}$ を抽出することである．GAN を用いる目的が $\boldsymbol{x}_{raw}$ に似た $\boldsymbol{x}_{fake}$（画像等）の生成であれば，$\boldsymbol{x}_{raw}$ を D への入力とすることに合理性がある．一方，GAN を用いて分類器を構成する場合，$\boldsymbol{x}_{raw}$ をそのまま用いるより，分類に適した特徴量 $\boldsymbol{x}_{real}$ を用いたほうが有利である．

E の学習は，以下の交差エントロピー J_C を最小化することで行う．

$$J_C = -\sum_n \sum_j y^*_{nj} \log p_D(y_{nj}) \tag{1}$$

ここに，$p(y_{nj})$ は E の出力層の値である．

G の入力は，カテゴリ c および $\boldsymbol{z} \in \mathbb{R}^{d_z}$ である．c および $\boldsymbol{z}$ は，それぞれカテゴリカル分布および標準正規分布からサンプルされるものとする．紙幅の都合上，$\boldsymbol{z} = \{\boldsymbol{z}, c\}$ と表記する．G の出力は，$\boldsymbol{x}_{fake} = G(\boldsymbol{z})$ である．

D の入力源 S は $\{real|fake\}$ のいずれかから選択される．そのときの D の入力をそれぞれ $\boldsymbol{x}_{fake}$，$\boldsymbol{x}_{real}$ と書くこととする．$D(\boldsymbol{x})$ は，入力源 S が $real$ である確率の推定値 $p(S)$ を出力する．すなわち，

$$D(\boldsymbol{x}) = p(S = real|\boldsymbol{x})$$

である．

G および D の学習では，以下のコスト関数 $J^{(D)}$ および $J^{(G)}$ をそれぞれ最小化する．

$$J^{(D)} = J_S + \lambda J_C$$

$$J^{(G)} = -J_S$$

$$J_S = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}_{real}} \log D(\boldsymbol{x}_{real}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log(1 - D(G(\boldsymbol{z})))$$

ここに，λ は重みである．パラメータ学習では，D と G の学習を交互に行う．まず，D の学習を行い，その後 D のパラメータを固定して G の学習を行う．

3.2 Multimodal Classifier GAN

以下では，提案手法 MMC-GAN について説明する．LAC-GAN では，E は Multi-Layer Perceptron 型の DNN であった．一方，画像を含むマルチモーダル入力に対しては，この構造が最適とはいえない．よって，本研究では，E を拡張した Multimodal Classifier GAN (MMC-GAN) を提案する．MMC-GAN における E の構造を図 2 に示す．

*1 https://youtu.be/_YQuziz4eGY

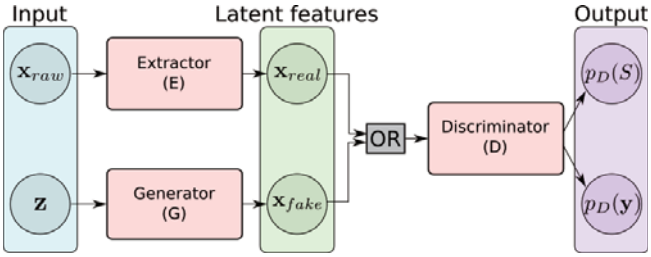


図 1: LAC-GAN および MMC-GAN の構成の概要



図 3: Samples of the 1282 depth images (right) $\mathbf{v}_d$ and their corresponding RGB images (left). Only the depth data is used.

MMC-GAN の Extractor E では，以下を入力とする．

$$\mathbf{x}_{raw} = \{\mathbf{w}_{inst}, \mathbf{w}_c, \mathbf{v}_d, \mathbf{v}_{meta}\}$$

ここに，$\mathbf{w}_{inst}, \mathbf{w}_c, \mathbf{v}_d, \mathbf{v}_{meta}$ はそれぞれ，命令文，コンテキストの自然言語表現，画像，画像メタ情報である．$\mathbf{w}_{inst}, \mathbf{w}_c$ の例を以下に挙げる．

$$\begin{cases} \mathbf{w}_{inst} = \text{“Move the towel to the shelf.”} \\ \mathbf{w}_c = \text{“The robot is holding a towel.”} \end{cases}$$

$\mathbf{v}_d$ としては Depth 画像を用いるが，RGB 画像や RGB-D 画像を用いてもよい．ハードウェアの制限があるため，Place 動作を行う場合，候補領域の適切性が床からの高さに依存することは自明である．よって，画像を取得した時点で得られる候補領域の高さ，カメラの高さおよび角度を $\boldsymbol{w}_c$ として入力に加える．

Extractor E は，上述の入力を潜在表現に変換する．まず，$\mathbf{w}_{inst}, \mathbf{w}_c$ を固定長の分散表現で表すために，PV-DM [Le 14] を用いる．これにより，文の長さに関わらず，固定長の分散表現 $\boldsymbol{x}_{inst}$ および $\boldsymbol{x}_c$ が得られる．一方，$\mathbf{v}_d, \mathbf{v}_{meta}$ を潜在表現に変換するため，図 2 に示す Convoluational Neural Network (CNN) を用いる．CNN の学習におけるコスト関数は，上述の J_C を用いる．中間の隠れ層の出力を $\boldsymbol{x}_v$ とする．

最終的に，E の出力として以下が得られる．

$$\mathbf{x}_{real} = \{\mathbf{x}_{instr}, \mathbf{x}_c, \mathbf{x}_v\}.$$

ここに，$\mathbf{x}_{real}$ は各潜在表現を連結したベクトルであり，D に入力される．

4. Carry and Place マルチモーダルデータセットの構築

Carry and Place タスクにおける命令理解の標準的なマルチモーダルデータセットは我々の知る限り存在しないため，実験で用いるデータセットを構築した．本節では，我々が構築した「Carry and Place マルチモーダルデータセット」について説明する．構築手順を以下に示す．

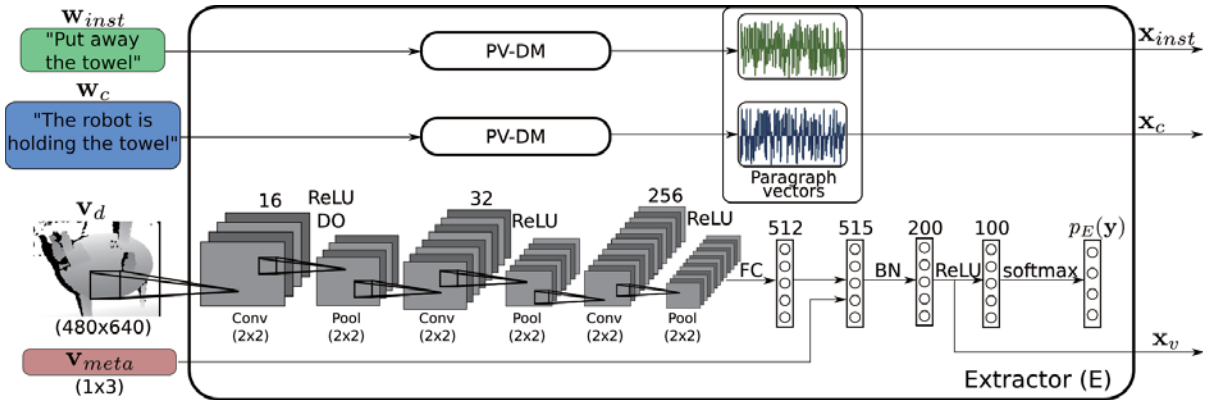


図 2: MMC-GAN における Extractor E の構造

表 1: 対象領域のバリエーション

#	家具の種類	対象領域数	画像数
テーブル	4	18	321
引き出し	4	10	336
机	4	8	425
棚	2	5	200
$\sum$	**14**	**37**	**1282**

1 日用品を対象領域にランダムに配置する（図3 参照).

2 物体検出位置にロボットカメラを配置し，RGB-D 情報を取得する.

3 ランダムに命令文を生成する.

4 ランダムに状態を表す文を生成する.

5 命令文と状態を表す文が，実世界状況において成り立たなければ，ステップ 3 に戻る．成り立てば，6 に進む.

6 ラベラーが正解ラベルを付与する.

本研究では，標準化されたロボットでの動作を想定するため，ハードウェアとして生活支援ロボット HSR を想定する．よって，画像の取得には，HSR と同様，Asus Xtion Pro カメラを用いた．入力として，RGB-D 画像を用いることも可能であるが，本研究では必要最小限の非言語情報で動作を確認するために，Depth 画像のみを用いた．対象領域を表1 にまとめる．1 つの家具に対して複数の対象領域が存在すること（棚の 1 段目，2 段目など），各画像では日用品の配置が異なること，に注意されたい．なお，引き出しは開放されているものとした.

各サンプルは以下の基準により，ラベル付けを行った.

($A1$) 対象領域は十分なスペースがあるとともに，ロボットアームの経路に障害が少なく，自律制御によるタスク達成が簡単である.

($A2$) 対象領域には障害物が一部存在するためロボットアームの経路計画が失敗する可能性があるものの，自律制御によるタスク達成が十分見込まれる.

($A3$) 対象領域には限られたスペースしかなく，ロボットアームが衝突する可能性が高いため，自律制御によるタスク達成の見込みは低い.

($A4$) 対象領域には障害物が多く，タスク達成は不可能に近い.

データセットをランダムにシャッフルし，学習セット (Train)，検証セット (Valid)，テストセット (Test) に分割した．データセットのサイズを表2 に示す.

表 2: 実験で用いたデータセットの概要

#	Train	Valid	Test	$\sum$
$A1$	158	29	25	**212**
$A2$	359	34	39	**432**
$A3$	350	26	22	**398**
$A4$	203	17	20	**240**
$\sum$	**1070** (*83* %)	**106** (*8.5* %)	**106** (*8.5* %)	**1282**

5. 実験

5.1 設定

パラメータ設定を表3 に示す．$\mathbf{x}_v$，w_{inst}，w_c，の次元数は各 200 次元とした．ゆえに，$\mathbf{x}_{real}$ は 600 次元である．提案手法における G として，2 種類の構造（GAN，CGAN）について検討した．それぞれの隠れ層のノード数は同等であるが，CGAN では入力にカテゴリカル分布からサンプルされる 4 次元の乱数が追加されている.

5.2 結果

定性的結果を図4 に示す．上段および中段は正しい予測結果であり，下段は誤りの例である．左上の記号は正解ラベルおよび予測結果を示す．例えば，「A1/A2」はラベルが「A1」であり，予測結果が「A2」であったことを示す.

次に，提案手法とベースライン手法（通常の DNN）を比較評価した．定量的結果を表4 に示す．一般に，DNN の精度比較では，エポックごとにモデルパラメータが更新される．よって，テストセットの最大値を比較しても，未知のデータに対する精度を表すものにならない．ゆえに，本実験では，標準的な手順に従い，検証セットの精度が最大値を示したモデルを各手法の最良モデルとした．表4 に示す結果は，初期値をランダムに変更して行った 10 回の実験の平均および標準偏差を示す．ただし，*を付した結果は，一部の結果が収束しなかったため，収束した結果の平均および標準偏差を示した.

表において，(I)，(I + C)，(I + C + V) は，命令文のみ，命令文および状態，命令文・状況・画像，をそれぞれ入力特徴とした場合である．表より，(I) および (I + C) に比べ，(I + C + V)

表 4: ベースライン手法と提案手法 MMC-GAN の比較結果

	[%]	入力特徴					
	GAN	**I**		**I + C**		**I + C + V**	
手法	type	Valid	Test	Valid	Test	Valid	Test
ベースライン	-	35.3 ± 2.0	28.6 ± 3.3	34.4 ± 1.5	26.7 ± 6.1	77.0 ± 1.2	76.6 ± 1.2
提案手法	GAN	27.1 ± 2.1	24.8* ± 6.3	28.0 ± 2.9	27.5* ± 8.7	84.2 ± 1.0	79.1 ± 1.6
提案手法	CGAN	28.5 ± 2.5	28.8* ± 7.9	28.1 ± 2.7	27.6* ± 8.2	83.7 ± 0.4	**81.0** ± **2.5**

表 3: MMC-GAN のパラメータ設定

最適化手法	Adam (学習率 0.00005, $\beta_1 = 0.5, \beta_2 = 0.9$), λ=0.2
バッチサイズ	64 (E), 50 (G and D)
GAN ノード数	100, 100, 100, 100
CGAN ノード数	100, 100, 100, 100

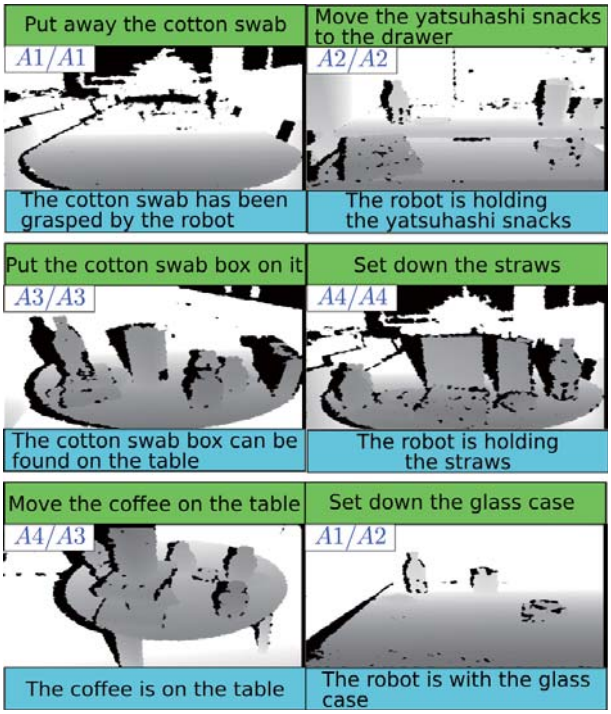


図 4: 対象領域の分類に関する定性的結果．上段および中段は正しい予測結果であり，下段は誤りの例である．

が高い精度であることから，本タスクは言語情報のみの特徴量では不十分であることを示している．また，(I) と (I + C) の間には統計的に有意な差は存在しなかった．

公平な比較を行うために，ベースラインとして構造が同等な DNN を用いた．ベースラインと提案手法を比較すると，提案手法の方が良好な結果であった．このことは，MMC-GAN によるデータ拡張がテストセット精度向上に寄与したことを示唆している．さらに，MMC-GAN に関して G の構造を比較すると，クラス情報を入力とした CGAN の方が高い精度であった．

6. おわりに

高齢化社会のなかで体の不自由な人々を物理的に支援可能な生産年齢人口は十分ではなく，彼らの家族が離職を余儀なくされるケースさえある．この社会課題の解決策として，生活支援ロボットは体の不自由な人々の生活の質を向上させるだけでなく，その家族を時間的拘束から解放するとともに，社会全体の生産性向上につながり得る．本論文では，生活支援ロボットのための命令理解手法を提案した．提案手法 MMC-GAN を Carry-and-Place タスクにおける命令に適用し，ベースラインを上回る精度を得た．

謝辞

本研究の一部は，JST CREST の助成を受けて実施されたものである．

参考文献

[Chen 16] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P.: InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets, in *Advances in Neural Information Processing Systems*, pp. 2172–2180 (2016)

[Gemignani 15] Gemignani, G., Veloso, M., and Nardi, D.: Language-based Sensing Descriptors for Robot Object Grounding, in *Robot Soccer World Cup*, pp. 3–15 (2015)

[Goodfellow 14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative Adversarial Nets, in *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)

[Iocchi 15] Iocchi, L., Holz, D., Solar, Ruiz-del J., Sugiura, K., and Zant, van der T.: RoboCup@Home: Analysis and Results of Evolving Competitions for Domestic and Service Robots, *Artificial Intelligence*, Vol. 229, pp. 258–281 (2015)

[Kollar 13] Kollar, T., Perera, V., Nardi, D., and Veloso, M.: Learning Environmental Knowledge from Task-based Human-Robot Dialog, in *IEEE ICRA*, pp. 4304–4309 (2013)

[Le 14] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, in *Proc. ICML*, pp. 1188–1196 (2014)

[Mirza 14] Mirza, M. and Osindero, S.: Conditional Generative Adversarial Nets, arXiv preprint arXiv:1411.1784 (2014)

[Odena 16] Odena, A., Olah, C., and Shlens, J.: Conditional Image Synthesis with Auxiliary Classifier GANs, arXiv preprint arXiv:1610.09585 (2016)

[Shen 17] Shen, P., Lu, X., Li, S., and Kawai, H.: Conditional Generative Adversarial Nets Classifier for Spoken Language Identification, Proc. Interspeech (2017)

[Springenberg 15] Springenberg, J. T.: Unsupervised and Semi-Supervised Learning with Categorical Generative Adversarial Networks, arXiv preprint arXiv:1511.06390 (2015)

[Sugiura 17] Sugiura, K. and Kawai, H.: Grounded Language Understanding for Manipulation Instructions Using GAN-Based Classification, *IEEE ASRU* (2017)

[杉浦 16] 杉浦 孔明：ロボットによる大規模言語学習に向けて-実世界知識の利活用とクラウドロボティクス基盤の構築-, 計測と制御, Vol. 55, No. 10, pp. 884–889 (2016)

[寺田 12] 寺田 耕志, 橋本 国松, 齋藤 史倫, 池田 幸一, 山本 貴史：生活支援ロボット HSR の開発と実証評価, 第 13 回システムインテグレーション部門講演会資料集, pp. 2A1–2 (2012)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習による道具選択を考慮したロボット道具使用モデル

Tool-use Model Considering Selecting Tool by Deep Learning

斎藤 菜美子 *1 Namiko Saito　金 杞泰 *1 Kitae Kim　グエン バ ダイ *1 Ba Dai Nguyen　村田 真悟 *1 Shingo Murata　尾形 哲也 *2 Tetsuya Ogata　菅野 重樹 *1 Shigeki Sugano

*1早稲田大学 創造理工学部 総合機械工学科
Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University

*2早稲田大学 基幹理工学部 表現工学科
Department of Intermedia Art and Science, School of Fundamental Science and Engineering, Waseda University

We propose a tool-use model that makes robots choose and use tools to carry out tasks. In these days, researches on the tool-use by robots have been done aiming at robots that are useful in daily life. However, conventional researches have two problems, (1)experimenters need to label tools or environment, (2)it is impossible to perform a series of operation from tool selection to task execution. In this research, we propose a model which does not require labeling, and can use tools in consideration of selection. In order to construct a model which can solve the two problems, we let a robot select a tool, hold it and perform the task, and have a series of experiences. Then, train the sensory-motor data that acquired during the experience and task command with deep learning. At last, to evaluate the model, we confirmed ability of motion generation in the untrained situation.

1. はじめに

ロボットが人間社会で活躍するためには，多様なタスクをこなせなければならない．人間は身体の限界を超えた様々なタスクを行うため，適切な道具を選択し使用している．同様にロボットも道具を使用できれば，生活空間で役立つことが予想されるため，ロボットによる道具使用の研究が行われてきた．しかし従来研究は，操作対象物体や道具，動作を事前にラベリングしている [Stoytchev 05]，または道具の把持と選択を考慮していない [Nishide 12][Mar 15] ものであり，(1) ラベル付けしていない物体や道具の使用と動作の実行に限界がある，(2) 道具選択からタスク遂行にかけての一連の操作が困難である，という二つの問題がある．本研究ではこれらの問題を解決するため，ラベル付けを必要としない，道具の選択を考慮した道具使用モデルの提案をする．

モデル実現のため，ロボットに道具を選択し，把持しタスクを遂行させる一連の経験をさせ，その間の運動情報と感覚情報を取得する．また，目標タスクをロボットに指定するタスクコマンドを用意する．運動情報，感覚情報，タスクコマンドをディープニューラルネットワークに入力し，この３つの関係性を学習させることにより，(1) と (2) の問題を解決する．そして，未学習の道具選択状況をロボットに提示し，入力された感覚情報とタスクコマンドから適切な動作生成が可能であるか確認する．

2. 提案手法

2.1 道具選択と道具使用経験によるデータ取得

道具の選択が求められる状況は大きく分けて二つある．一つ目は，機能が異なる複数の道具の中から，遂行するタスクに適した機能を持つ道具を選択する場合である．二つ目は，同じ機能を持つ異なる大きさの道具の中から，適切な大きさの道具を選択する場合である．そのため，この二つの場合を考慮可能な実験設定をする必要がある．まず，道具の機能について，棒型と熊手型という機能の異なる道具を用意することで，一つ目の選択状況の経験を可能とする．続いて道具の大きさについて，棒型，熊手型共に複数の長さの道具を用意し，操作対象物体についても道具の長さに合わせて複数の位置に配置することで，二つ目の選択状況の経験を可能とする．なお，棒型道具は物体を横方向へスライドさせる機能，熊手型道具は手前へ物体を引き寄せる機能を持つものとする．そして，ロボットに道具選択から把持，道具を用いた物体の移動まで一連のタスクを経験させる．その間，運動情報であるロボットアームの関節角度と把持グリッパーを開く角度及び，感覚情報であるロボットの視野画像を取得する．そして，遂行する目標のタスクに合わせてタスクコマンドを設定する．運動情報，感覚情報，タスクコマンドの三者の関係を，ロボットに経験させ，学習させることにより，状況を考慮した適切な道具選択と動作生成を可能にさせる．

2.2 道具選択を考慮した道具使用モデルの学習

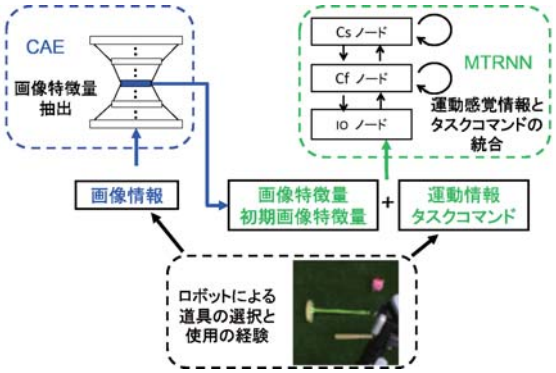


図 1: 道具使用モデル

本研究の提案モデルは図 1 に示すように，画像特徴量を抽出する Convolutional Auto-Encode (CAE) [Masci 11] と運動感覚情報を統合学習する Multiple Timescale Recurrent Neural

連絡先: 斎藤菜美子，早稲田大学創造理工学部総合機械工学科，n_saito@sugano.mech.waseda.ac.jp

Network (MTRNN) [Yamashita 08] から構成される．

道具使用経験の際，取得する画像は次元が大きく計算コストが高い．そのため，次元削減及び画像情報の特徴を抽出する事ができる CAE を利用する．CAE は砂時計型の構造を持っており，入力データがニューロン数の最も少ない中間層を経由した後，再び元の次元数に出力され，この出力が入力データを復元するよう学習される．これにより，中間層ニューロンから画像の特徴量を低次元で取り出すことができ，物体と道具にラベル付けを行う必要が無くなる．続いて，画像特徴量と経験時に取得した運動情報，タスクコマンドを MTRNN により統合学習する．MTRNN は現在の状態から次の状態を予測可能な学習器である RNN の一種であり，IO 層と，ニューロンの発火速度の異なる Cf 層，Cs 層から構成される．この構造を持つことで，データの急激な変化と時系列の順序が学習でき，データのダイナミクスを学習することができる．なお，画像特徴量に関しては，現在の情報だけでなく初期の画像情報についても毎回入力する．これにより，物体や道具がアームの影になり見えない時でも，用いている道具や物体の情報を常に入力することができるようになる．MTRNN により，身体情報，視覚情報，タスクコマンドの関係性が学習でき，目標のタスクに適した道具選択が可能となる．

3. 評価実験

3.1 実験設定

ヒューマノイドロボット NEXTAGE を用い，道具選択と動作生成の実験を行う．道具は熊手型と棒型それぞれ 2 種類の長さの計 4 種を用意し，一試行でそのうち 2 つの道具を使用する．熊手型は物体を引き寄せる，棒型は物体を横にスライドさせるタスクに用いられる．2 つの道具はロボットの目の前に置かれた台の上に配置し，初期位置は，大きく離れた (15cm) 2 か所の基準位置から，上下にも 1.5cm ずつ平行移動させた計 6 か所を設定する．このとき，配置する 2 つの道具のうち，選択しない方は基準位置に置く．この組み合わせにより，動作を 72(4(道具数) × 3(選択しない方の道具の組み合わせ数) × 2(基準位置の数) × 3(平行移動する位置の数)) セット設定する．なお，物体の位置に関しては，選択する道具の長さに合わせて，近い位置と遠い位置 2 か所を設定した．72 セットのうち 48 セットを用いてモデルの学習を行い，24 セットを用いてモデルの汎化性能を確認する．最後に，オンラインで取り込んだ画像から動作生成を行い，未学習位置及び未学習道具への道具選択と動作生成能力を評価する．

3.2 実験結果

MTRNN に画像特徴量，タスクコマンドを入力し，ロボットアームの関節角度とグリッパーの開閉角度を予測させることで，モデルの汎化性能を評価した．評価に使用するデータは (1) 配置する道具の組み合わせが未経験のデータが 12 セット，(2) 未学習位置にある道具を選択するデータが 12 セットである．目標関節角度を生成し実際に動作させた結果，(1) のデータは 83.3%，(2) のデータでは 33.3%の精度でタスク遂行に成功し，物体を目標通り移動させることができた．失敗したデータに関しても，物体の移動はかなわなかったものの，目標角度と似たデータが生成でき，運動自体は正しいものが行えたと言える．この結果から，モデルは未学習データに対して汎化性能を持つことが確認できた．続いて，未学習の道具と位置に関してオンライン動作生成を行った結果，道具の選択及び目標角度の生成に成功し，目標のタスクを遂行できた．動作生成結果の例を図 2 に示す (動作は図の左から右の順で行われる)．この結果から，このモデルにより，ロボットは動作と物体，道具のラベル付け無しに道具の選択及び動作生成が可能であることが確認された．



図 2: 未学習の状況における動作生成結果

4. まとめ

本研究では，道具の選択を考慮した道具使用モデルを提案した．異なる機能を有する道具間，同じ機能を有する大きさの異なる道具間において，状況に応じた適切な道具選択が考慮可能なタスクを設定し，ロボットに経験させ，その間の運動感覚情報とタスクコマンドを学習させた．評価実験として，未学習道具と，経験した道具の未学習位置の感覚情報を入力し，動作生成をさせた．その結果，モデルの汎化性能により，未学習の状況でも道具の選択と動作生成が可能であることを確認した．

5. 謝辞

本研究成果の一部は，文部科学省科研費基盤研究 (S) (No. 25220005),JST，CREST（No: JPMJCR15E3）の助成を受けたものです．本研究成果の一部は，早稲田大学理工研プロジェクト研究「自然と共生する知能情報機械系に関する基盤研究」の一環として行われたものです．ここに謝意を表します．

参考文献

[Stoytchev 05] A.Stoytchev: Behavior-grounded representation of tool affordances, International Conference on Robot and Automation - ICRA, pp.3060-3065, 2005.

[Nishide 12] S.Nishide, J.Tani, T.Takahashi, H.G.Okuno, T.Ogata: Tool-body Assimilation of Humanoid Robot Using a Neurodynamical System, IEEE Transactions on Autonomous Mental Development, vol.4, no.2, pp.139-149, 2012.

[Mar 15] T.Mar, V.Tikhanoff, G.Metta and L.Natale: Self-supervised learning of grasp dependent tool affordances on the iCub Humanoid robot, in IEEE International Conference on Robotics and Automation, pp.3200-3206, 2015.

[Masci 11] J. Masci, U.Meier, D.Ciresan, and J. Schmidhuber: Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction, LNCS 6791, pp.52-59, 2011.

[Yamashita 08] Y. Yamashita and J. Tani: Emergence of Functional Hierarchy in a Multiple Timescale Recurrent Neural Network: A Humanoid Robot Experiment, PLoS Computational Biology, Vol. 4, No.11, 2008.

[2B1] [General Session] 1. Basis / Theory

座長:小林 亮太（国立情報学研究所）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room B (4F Moon Light)

[2B1-01] Analysis of the Dictionary Bases obtained by Applying Sparse Coding to Brain Activity and the Semantic Representation based on Distributed Representation

○Chiaki Kawase[1], Ichiro Kobayashi[1], Shinji Nishimoto[2], Satoshi Nishida[2], Hideki Asoh[3] （1. Ochanomizu University, 2. National Institute of Information and Communications Technology, 3. National Institute of Advanced Industrial and Technology）

9:00 AM - 9:20 AM

[2B1-02] Identification of EEG features during motor imagery by heuristic BCI with L-FTM, after learning of EEG features during exercise.

○Teruo Oda[1], Suguru N. Kudoh[1] （1. Kwansei gakuin university）

9:20 AM - 9:40 AM

[2B1-03] Creating Artificial Brain Activity Data Based on Mutual Data Transformation between Different Subjects

○Kaei Cho[1], Ichiro Kobayashi[1], Shinji Nishimoto[2], Satoshi Nishida[2], Hideki Asoh[3] （1. Ochanomizu University, 2. National Institute of Information and Communications Technology, 3. National Institute of Advanced Industrial Science and Technology）

9:40 AM - 10:00 AM

[2B1-04] Reconsidering the computational model for the auditory nerve: unsupervised learning or task-based optimization?

○Hiroki Terashima[1], Shigeto Furukawa[1] （1. NTT Communication Science Laboratories）

10:00 AM - 10:20 AM

[2B1-05] Top-down and bottom-up classification between areas in mouse cerebral cortex to connect machine learning modules on connectomes

○Taku Hayami[1], So Negishi[1], Rintaro Komori[1], Haruo Mizutani[1,2], Hiroshi Yamakawa[1,2] （1. Dwango Artificial Intelligence Lab., 2. Whole Brain Architecture Initiative）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

脳活動と分散表現による意味表象へのスパースコーディング適用により獲得された辞書基底の分析

Analysis of the Dictionary Bases obtained by Applying Sparse Coding to Brain Activity and the Semantic Representation based on Distributed Representation

川瀬千晶 *1 Chiaki Kawase　小林一郎 *1 Ichiro Kobayashi　西本伸志 *2 Shinji Nishimoto　西田知史 *2 Satoshi Nishida　麻生英樹 *3 Hideki Asoh

*1お茶の水女子大学 Ochanomizu University　*2情報通信研究機構 脳情報通信融合研究センター National Institute of Information and Communications Technology

*3産業技術総合研究所 人工知能研究センター National Institute of Advanced Industrial and Technology

It is known that primary visual cortex uses a sparse code to efficiently represent natural scenes. Based on the fact, we build up a hypothesis that the same phenomenon happens at the higher cognitive function, here we focus on semantic representation, in the cerebral cortex. To proof the hypothesis, we applied sparse coding to the brain activity while the subject is watching the video and the semantic representation of the video. By means of this method, we obtained the dictionary matrix consisting of bases which represent the corresponding relation between the brain activity and the semantic representation, and we analyzed the characteristics of each basis.

1. はじめに

近年，動画像などを視聴した際の脳の活動パターンから人がどのような意味カテゴリを想起しているかを調査する研究が盛んになってきており，多くの新しい知見が得られている．Huth ら [5] は，動画像中に現れる物体や動作を類義語体系である WordNet の語彙で表現し，動画像の刺激（WordNet 語彙 [8]）と脳神経活動との関係について調査し，脳の皮質における意味のマップを作成した．Stansbury ら [3] は，潜在的意味解析手法 LDA[9] を用いて，静止画に対して付与された語彙からシーンに対するラベル付けを教師なし学習で行い，その結果と静止画に対する脳神経活動の関係を結びつけ，カテゴリに対する脳の意味解釈の活動領域を明確にするとともにモデルを構築した．Cukur ら [2] は，動画像中の物体に注意を払い認識する際に，どのように認識の意味形態が変化しているかを脳活動データから推定している．このように統計的な言語モデルは脳活動における感覚や文脈の情報に基づく表象表現を説明するのに適したモデルであることが指摘されてきたが，さらに近年，西本，西田らは，Mikolov ら [11] によって提唱された word2vec を構築する際に採用された Skip-gram と呼ばれる言語モデルが潜在意味解析手法等のこれまでの言語モデルに較べて，より適していることを同じ実験設定の下で確認し，日本語 Wikipedia をコーパスとし，Skip-gram を利用することで得られる日本語の語彙の分散意味表現と血中酸素飽和度で計測される脳神経活動の間に相関関係が存在することを示している [7]．本研究では，脳の活動に対応する word2vec による表現を意味表象と呼ぶ．また，動画視聴時のヒトの脳活動と，その動画を説明する文との対応関係をスパースコーディングにより分析し，それぞれの基底がどの様な機能を表現しているかについて調査する．

2. 脳活動と意味表象の辞書学習

まず，fMRI を用いて計測した脳活動データをサンプルごとに計測した各ボクセルの観測値を入れて行列化し，これを脳活動行列とする．また，その画像説明文もサンプルごとに出現する単語（名詞，動詞，形容詞）の分散意味表現の和のベクトルからなる行列を作り，これを意味表象行列とする．これら 2 つの行列を縦に結合させ，脳活動と意味表象の結合行列を作成する．この結合行列に対し辞書学習を行い辞書と係数に分解する（図 1）．これにより，辞書行列には，脳活動の特徴と意味表象の特徴が 1 列になった基底が作られ，係数行列は脳活動と意味表象において共通の基底になる．このようにして作られた辞書を構成する基底について分析を行う．

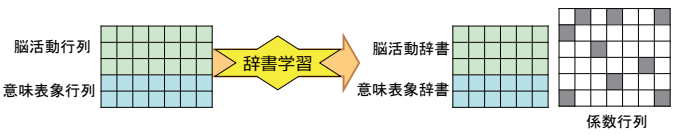


図 1: 脳活動データと意味表象の対応を保持する辞書学習

3. 実験

3.1 データ

使用するデータは，動画視聴時の脳活動データと動画説明文である [10]．このデータセットを 3600 サンプルを使用する．脳活動データは，一人の被験者に動画像を見せ，fMRI を用いてその時の脳神経活動を 2 秒で 1 サンプル記録したものである．脳活動の観測領域は 100×100×32 ボクセルであり，そのうち大脳皮質部分が 30662 ボクセルある．脳活動データの辞書学習をする際に，データ数 3600 サンプルよりもデータの次元を少なくしなくてはいけないため，30662 ボクセルのうち，先行研究 [7] で予測精度が 0.36 以上の 1404 ボクセルを抽出した．動画説明文は被験者に見せた動画像から 1 秒ごとに抽出した静止画に対し，アノテータ 60 人のうちランダムに抽出された 5 人が静止画を見て書いた説明文を使用した．説明文はその静止画を見て想起したことを書いてもらったものである．この脳活動データと画像説明文のデータを 2 秒分ずつ対応づける．

連絡先: 川瀬千晶，お茶の水女子大学，g1220516@is.ocha.ac.jp

3.2 実験設定

脳活動行列の辞書学習アルゴリズムには Lasso-LARS と LARS を用い，辞書の基底数は辞書学習を行う際のデータ数，データの次元数と基底数の制約条件を考慮し，2500 に設定した．

4. 基底の分布

辞書学習で作られた脳活動基底と意味表象基底の分布をそれぞれ可視化した．図 2 は，これらの基底をそれぞれ主成分分析し，寄与率の高い 3 次元を抽出し，その 3 次元空間上で表示した．元の脳活動基底の次元数は 1404，意味表象基底の次元数は 1000 である．主成分分析の結果，脳活動基底の主成分の寄与率は第一主成分が約 0.244，第二主成分が約 0.068，第三主成分が約 0.039 となり，これらの累積寄与率は約 0.350 となった．意味表象基底の主成分の寄与率は第一主成分が約 0.095，第二主成分が約 0.063，第三主成分が約 0.041 となり，これらの累積寄与率は約 0.199 となった．

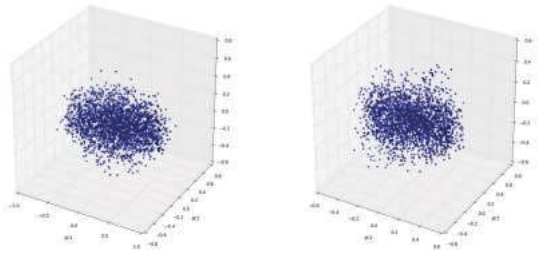


図 2: 基底の分布可視化（3 次元）

そこで，一方の基底を k-means 法によりクラスタ数 k(=30) を決め，クラスタリングを行いクラスタごとに色分けをし，それと対応するもう一方の基底に同じ色をつけ，それを主成分分析による 3 次元空間で表示した．k-means 法の距離には，脳活動基底にはユークリッド距離を，意味表象基底には cos 類似度を用いた．脳活動基底 2500 個を 30 個のクラスタにクラスタリングし，それと対応する意味表象基底に同じ色をつけ，そのうちクラスタ 2 組ずつに含まれる基底の例を 3 つ表示した．左側に脳活動基底，右側に意味表象基底を示す（図 3）．多くの意味表象の基底は，脳活動のクラスタごとにまとまって現れた．また，意味表象基底を 30 個のクラスタにクラスタリングし，それと対応する脳活動基底に同じ色をつけ，そのうちクラスタ 2 組ずつに含まれる基底の例を 3 つ表示した．左側に脳活動基底，右側に意味表象基底を示す（図 4）．多くの脳活動の基底は，意味表象のクラスタごとにまとまって現れた．これらの図から，類似した脳活動基底と対応する意味表象基底は類似している場合が多いと考えられ，また，その逆も同様に言えると考えられる．

5. 基底の分析

5.1 実験方法

5.1.1 意味表象基底の表す意味

辞書学習により得た辞書の意味表象辞書の部分において各意味表象基底が表す意味を分析する．意味表象基底の持つ意味は単語で表せるものではないが，単語での近似を行うことで意味を理解する目安とする．各意味表象基底と本研究で word2vec 空間を学習するのに用いた word2vec コーパスに含まれる全単語との cos 類似度を計り最も小さい単語から上位 3 単語とその cos 類似度を出力し，これらの単語を各意味表象基底の表す意味の近似単語とした．

5.1.2 同じ基底を用いて復元されるサンプル同士の関連性

特定の基底を大きな重みで使って復元されるサンプル同士に関連が見られるかを検証する．これは，特定の基底を大きな重みで使って復元されるサンプル上位 5 つに対応する刺激画像同士に関連があるかを見ることで行う．基底には画像を見たときの脳活動と意味表象の特徴が表されると考えられるので，人が見て近い印象を抱く画像同士が出力されることが予想される．特定の基底に対応する係数ベクトルの非ゼロ要素の中で最も重みの大きいサンプル上位 5 つに対応する刺激画像とその重みを出力する．動画 1 サンプルは 2 秒間であり，その中央である 1 秒目の静止画像を出力した．

5.1.3 基底と ROI との関係

ヒトの大脳皮質には，ROI（Region Of Interest）という部位に分けられる部分がある．そのうち特定のカテゴリに反応する 2 つの ROI と基底との関係について分析を行った．1 つは FFA（fusiform face area）という顔に強い反応を示す領域，もう 1 つは PPA（parahippocampal place area）という風景や場所に強い反応を示す領域とした．本研究で実験対象とした脳活動データ 1404 ボクセルのうち FFA に属するボクセル数は 57 であり，PPA に属するボクセル数は 14 であった．これら 2 つの ROI が大きく寄与する基底に対して，近似単語と，その基底を用いて復元されるサンプルに対する刺激画像を出力し比較を行い，そのカテゴリに違いが見られるかを検証する．これらの ROI が大きく寄与している基底として，脳活動行列の対象 ROI に属しているボクセルに対する値を全て足した値が大きい基底上位 4 つを抽出し，これらの基底について実験を行った．

5.2 実験設定

基底は前から順に番号を付与する．まず，ランダムな基底の代わりに前から順に 4 つの基底に対して実験を行った（表 1）．次に，FFA が大きく寄与している基底上位 4 つ（基底 1585，基底 745，基底 563，基底 719）について実験を行った（表 2）．また，PPA が大きく寄与している基底上位 4 つ（基底 1436，基底 1364，基底 2400，基底 590）について実験を行った（表 3）．

5.3 実験結果

表 1，2，3 には，各基底番号，その基底の意味表象部分の近似単語上位 3 つとその cos 類似度，出力された刺激画像と画像の下にその重みを 1 位から 5 位まで左から順に示す．

5.4 考察

基底 0〜4 に対して実験を行ったところ，似たようなイメージの刺激画像同士が得られた．ここでは 4 例しか示していないが，各基底における実験結果についての考察を述べる．基底 0 においては，動物，特に動物の群れが映っている画像が多く出力された．基底 1 においては，1 位，2 位のサンプルの重みがそれ以下の重みと比べて高くなっており，これらの画像には男女が映っている．基底 2 においては，1 位，2 位，3 位の刺激画像は同じ動画内から抽出された画像となっており非常によく似ている．また，5 位の画像にもこれらの画像と類似した模様のある生物が映っている．基底 3 においては，1 位，3 位，5 位の画像には文字が見られた．2 位，4 位の画像には文字は映っていないが，その前後のサンプルと対応する画像を見てみたところ，文字が映っている画像であることを確認した．これは，動画 1 サンプルは 2 秒間であり，画像説明文はその中の

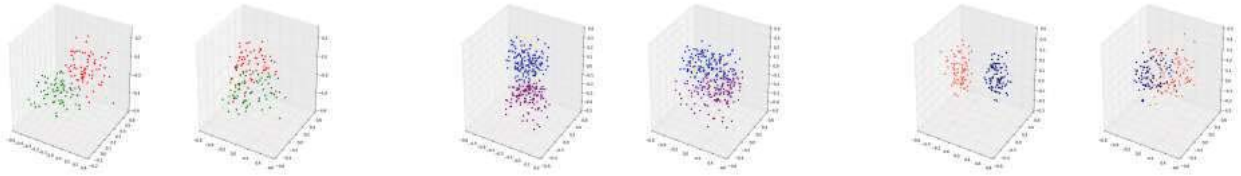


図 3: 脳活動基底のクラスタ（左）と対応する意味表象基底（右）

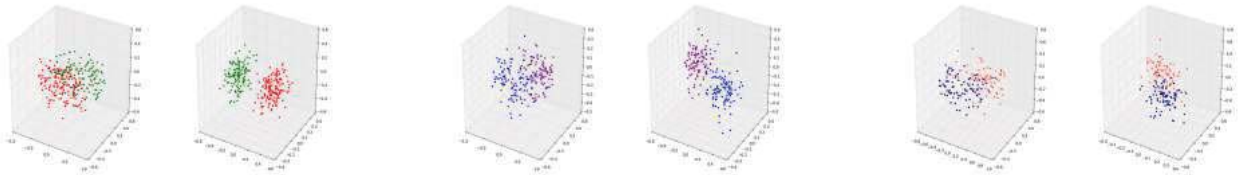


図 4: 意味表象基底のクラスタ（右）と対応する脳活動基底（左）

表 1: 特定基底 (ランダム) が大きく使われるサンプル

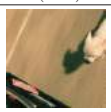
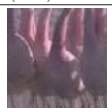
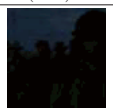
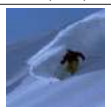
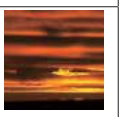
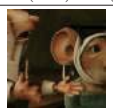
近似単語：和室 (0.23)，洋室 (0.22)，ミュージアムショップ (0.22)					
基底 0	 18.7	 15.3	 14.3	 12.8	 11.9
近似単語：粘板岩 (0.34)，吹き上げる (0.34)，等高線 (0.33)					
基底 1	 34.5	 24.4	 4.40	 3.57	 2.84
近似単語：ウツボ (0.60)，ウミヘビ (0.59)，アメフラシ (0.59)					
基底 2	 41.3	 17.1	 12.3	 6.31	 4.07
近似単語：文字 (0.40)，タイポグラフィ (0.35)，シンタックス (0.34)					
基底 3	 45.5	 7.45	 6.46	 4.77	 4.52

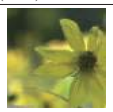
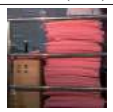
表 2: 特定基底 (FFA が強い) が大きく使われるサンプル

近似単語：男性 (0.36)，雇い主 (0.31)，チカチーロ (0.29)					
基底 1585	 46.7	 7.92	 7.12	 5.62	 3.90
近似単語：スーツ (0.41)，目元 (0.41)，いでたち (0.40)					
基底 745	 41.0	 12.8	 5.64	 5.29	 3.65
近似単語：パーティー (0.36)，女性 (0.35)，会食 (0.34)					
基底 563	 133	 109	 108	 96.8	 96.5
近似単語：包帯 (0.47)，服 (0.45)，ネズミ (0.43)					
基底 719	 49.4	 10.1	 5.98	5.40	3.20

0.5 秒目と 1.5 秒目の静止画像に対するものを合わせているのに対し，その中の 1 秒目の静止画像を出力したことが原因になっている可能性がある．このような例は他の基底においても見られたことから，刺激画像の抽出方法には改善の余地があると考える．他の基底について実験を行っても，同じ基底を大きな重みで使って復元されるサンプル同士は類似している例が多く見られた．このことから，辞書学習により意味表象の特徴を捉えた基底が得られ，これらの意味表象基底はさまざまな意味表象を表現するのに効率の良さに役立っていると考えられる．また，今回は辞書学習の条件の下，訓練データ数とデータの次元数を考慮した上で基底数を 2500 と設定したが，基底数を大きくするほど細かい特徴を捉える基底が得られやすく，基底数を小さくするほど大まかな特徴を捉える基底が得られ易くなると考えられ，最適な基底数を見つけることが重要だと考える．

また，意味表象基底の近似単語について考察を行うと，近似単語と刺激画像とに関連がある例（基底 2，3）とない例（基底 0，1）が見られた．これは，基底 2，3 においては近似単語との cos 類似度が高いのに対し，基底 0，1 においては低いことが関係していると考えられる．このことから，意味表象基底にはその意味を単語で近似しやすいものと近似しにくいものがあると考えられる．FFA と PPA についての実験結果からは，意味表象基底の近似単語を見ると，比較的に FFA の寄与が大きい基底は人を連想させるような単語が多く，PPA の寄与が大きい基底は風景や場所を連想させるような単語が多いことを観測した．出力された刺激画像においても FFA と PPA のカテゴリの違いが多少見られたものの，特に FFA の寄与率が 3 番目に大きい基底 563 に対して出力された刺激画像はどれも風景を表しており予想と反する結果となっており，この点に関

表 3: 特定基底 (PPA が強い) が大きく使われるサンプル

近似単語：地層 (0.54)，堆積岩 (0.52)，火山岩 (0.51)					
基底 1436	 48.6	 8.48	 7.61	 4.79	 4.22
近似単語：道路 (0.32)，街路 (0.30)，街並 (0.29)					
基底 1364	 48.3	 7.69	 3.82	 3.48	 3.47
近似単語：工場 (0.40)，アウトソーシング (0.31)，テクノロジーズ (0.31)					
基底 2400	 47.0	 18.6	 16.5	 14.1	 9.80
近似単語：並木道 (0.49)，小道 (0.46)，木漏れ日 (0.45)					
基底 590	 48.5	 8.50	 5.20	 3.74	 3.60

してはさらなる検証が必要である．

6. おわりに

本研究では，ヒトの動画視聴時の脳活動データと意味表象を辞書学習することにより，これらの特徴を捉えた基底を作ることに成功した．提案手法で得られた基底の分布を見ると，脳活動基底が類似していればそれと対応する意味表象基底も類似している場合が多いと考えられ，その逆も同様に考えられる．また，各意味表象基底の表現する意味を単語で近似した．また，同じ基底を大きな重みで使い復元されるサンプルには関連性が見られることを確認した．このことから，スパースコーディングにより得られた意味表象基底は，さまざまな意味表象を表現するのに効率の良いような特徴を捉えていると考えられる．さらに，2 種類の ROI の寄与が大きい基底での実験を行い比較をしたところ，これらのカテゴリに多少の違いが見られたがさらなる検証の余地が残されていることを観測した．今後はさらに実験を進め，脳活動と意味表象の対応関係についてより深い考察を進めたい．

参考文献

[1] Olshausen BA,Field DJ,“Sparse coding of sensory inputs”,Current Opinion Neurobiology 2004, 14:481-487.

[2] Tolga Cukur, Nishimoto S, Alexander G Huth and Jack L Gallant,“Attention during natural vision warps semantic representation across the human brain”,Nature Neuroscience, Volume 194, January 2013, pp.240-252, 2013.

[3] Stansbury DE1, Naselaris T, Gallant JL ,“Natural scene statistics account for the representation of scene categories in human visual cortex”,Neuron79(5):1025-34. j.neuron,2013.

[4] Francisco Pereira, Matthew Botvinicka, Greg Detre,“Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments”,Volume 5, Article 72, 2011.

[5] Huth AG, Nishimoto S, Vu AT, Gallant JL,“A continuous semantic space describes the representation of thousands of object and action categories across the human brain”,Neuron76(6):pp.1210-1224,2012.

[6] T. Horikawa, M. Tamaki, Y. Miyawaki, Y. Kamitani,“Neural Decoding of Visual Imagery During Sleep”,SCIENCE VOL 340,2013.

[7] Nishida S, Huth AG, Gallant JL, Nishimoto S,“Word statistics in large-scale texts explain the human cortical semantic representation of objects, actions, and impressions”,Society for Neuroscience Annual Meeting,2015.

[8] George A. Miller,“WordNet: A Lexical Database for English”,Communications of the ACM Vol. 38, No. 11: pp.39-41,1995.

[9] David M. Blei, Andrew Y. Ng, and Michael I. Jorfan,“Latent Dirichlet Allocation”,Journal of Machine Learning Research 3, pp.993-1022, 2013.

[10] Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL,“Reconstructing visual experiences from brain activity evoked by natural movies”,Current Biology 21(19):pp.1641-1646,2011.

[11] T.Mikolov, I.Sutskever, K.Chen, G.Corrado and J.Dean, “Distributed Representations of Words and Phrases and their Compositionality ”,Advances in Neural Information Processing Systems 26, pp. 3111-3119,2013.

[12] Jeffrey Pennington, Richard Socher and Christopher D. Manning,“Glove: Global Vectors for Word”, Conference on Empirical Methods in Natural Language Processing (EMNLP 2014),2014.

[13] Quoc Le, Tomas Mikolov, “Distributed Representations of Sentences and Documents”,Google Inc, 1600 Amphitheatre Parkway, Mountain View, CA 94043,2014.

[14] William E. Vinje and Jack L. Gallant,“Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision”,Science 18, Vol. 287, Issue 5456, pp.1273-1276,2000.

[15] Kawase C, Kobayashi I, Nishimoto S, Nishida S, Asoh H,“Semantic representation in the cerebral cortex with sparse coding”,2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

L-FTMを用いた探索型BCIの，運動時発現脳波の学習による運動イメージ脳波特徴識別

Identification of EEG features during motor imagery by heuristic BCI with L-FTM, after learning of EEG features during exercise.

小田 輝王 [1*] 工藤 卓 [1,CA]
Teruo Oda Suguru N. Kudoh

[1] 関西学院大学
Kwansei Gakuin University

It is known that EEG drastically changes depending on external influences and health condition of an experimental participant. BCI based on the focused features of EEG signal, such as frequency band or measurement sites is suitable only for the users with reproducible, major EEG features evoked by a certain cognitive task. To avoid this problem, we developed a BCI using Learning-type-Fuzzy-Template-Matching (L-FTM) method. In addition, we implemented pruning procedure that deletes unsuitable fuzzy rules with high compatibility degree to both of task and non-task status. We confirmed that was BCI system detected the EEG features of a participant during imaging movement task, after learning the EEG features accompanied by motor task.

1. はじめに

神経細胞の集合体である脳は，一つ一つの神経細胞から起こる活動電位を回路の出力として統合し，脊髄や筋を介して動作を実現している．また，脳は運動・知覚などの行動だけでなく，感情・理性などの精神活動の源である．脳に存在する多数の神経細胞で発現した活動電位が重なって生じた集合電位が頭蓋表皮に伝搬したものを記録したものが脳波（ElectroEncephaloGram, EEG）である [Berger 29]．脳波を応用した研究として近年注目されているのは非侵襲型のブレイン-コンピュータ・インターフェース（Brain-Computer Interface, BCI）である．これは頭皮上に電極を設置して脳波計測を行い，その活動パターンから人の意図を検出して機器を操作するもので，特に脳波 BCI は非侵襲であり，感染症のリスクの心配がないため，安全で日常的に使用できる [櫻井 07]．脳波を用いた BCI 研究では，事象関連電位（Event-Related Potentials, ERP）や定常状態視覚誘発電位（Steady State Visual Evoked Potentials, SSVEP）などのようにあらかじめ決まった脳波特徴に着目し，特定の周波数帯域と計測部位に限定して信号を検出するものがほとんどである [Faller 10, Sakurada 13, Kaufmann 14, Daly 18]．着目した脳波特徴から，これを誘発する認知的活動が検出されたときに，これに従って機器の制御を行うのが従来の BCI であった．しかし，EEG は外部の影響によって大きく変化することが知られており，実験参加者の健康状態や集中力によって変動するため，再現性に問題がある [市川 08, Krauledat 08]．本研究では，再現性の問題を克服するために，認知的活動とそれに伴う脳波をあらかじめターゲットにする方法を取らない BCI を提案する．我々は，これまでにあらかじめ検出する対象となる脳波特徴を限定するかわりに，短時間の学習を行い，ある特定の認知的タスクを実行するときに特異的に出現する脳波パターンを自動抽出する探索型 BCI を提案してきた [山口 12]．さらに探索型 BCI として，学習型簡略ファジィ推論を応用した学習型ファジィテンプレートマッチング（Learning Fuzzy Template Matching, L-FTM）法を実装した．また，認知的タスクを行っている状態（タスク状態）と行っていない状態（ノンタスク状態）の両方に出現する脳波特徴に対して適合度が高いテンプレート（ルール）を削減し，タスク又はノンタスクのどちらか一方の状態に適合度が高いテンプレートを残す「プルーニング」を実装した．本研究では，開発した探索型 BCI を用いて，学習時に運動を実際に行い，ファジィルールで構成された探索空間から運動をイメージした状態（タスク状態）で出現する脳波特徴パターンに対応するルールを学習によって自動的に選定し，運動イメージによる BCI を実現した．

連絡先: 工藤卓，関西学院大学　理工学部　人間システム工学科，神経知能工学研究室，三田市学園 2-1，080-2401-4782，snkudoh@kwansei.ac.jp

2. 提案手法

2.1 学習型-ファジィテンプレートマッチング

本研究で用いた学習型ファジィテンプレートマッチング（Learning Fuzzy Template Matching, L-FTM）は，構成としては学習型簡略ファジィ推論と同様である [菅野 88]．通常のファジィテンプレートマッチングではメンバシップ関数の頂点の位置を用いてテンプレートが構成されているのに対し，L-FTM では入力値に対して"高い"，"低い"のファジィラベルを用いて前件部のルールをテンプレートとして構成する．これにより，様々な入力に対してルール（テンプレート）を構成することが出来る．例えば，"α 波周波数帯の O1 計測部位の脳波"，"β 波周波数帯の C3 計測部位の脳波"，"上腕の筋電"のように異なる特性をもった入力に対して柔軟にテンプレートを構成することが可能である．また入力に対して，はっきりとした閾値を設定する必要がなく，生成された入力パターンの各ラベルごとに設定されたメンバシップ関数からメンバシップ値を算出し，各ルールの適合度を求めた．本研究で開発した L-FTM では測定部位数とラベル数によって構成されたルールから有効なルール（テンプレート）を学習によって網羅的に探索するが，実時間での外部機器制御を目標としているため，計算負荷による遅延は可能な限り抑える必要がある．そこで，後件部はマムダニの推論の結論部であるファジィ集合ではなく，実数値を用いた簡略ファジィ推論とした．通常のファジィ推論では重心計算を行うため，時間がかかるのに対し，簡略ファジィ推論は後件部値を実数値として表現することで計算量を軽減することができる．具体的には，後件部値の実数値を，各ルール（テンプレート）の適合度を重みとした荷重平均して出力値とする [Takagi 93]．有効なルールを学習によって選別する際の探索空間となるベースのルールセットは脳波の計測部位と周波数帯の組み合わせで自動生成し，後件部値が目標値となる教師信号に近似するよう学習によって設定した．

2.2 出力値の計算

脳波の入力を得るために Biosemi 社の ActiveTwo システムを用いて脳波信号を計測した．計測部位ごとに高速フーリエ変換（Fast Fourier Transform, FFT）を行い，選択した周波数帯域の範囲の振幅パワーの面積値を求めて入力値とした．面積値を計算する各周波数帯域は δ 波（1∼3 Hz），θ 波（4∼7 Hz），α 波（8∼13 Hz），β 波（14∼50 Hz）とし，タスクに応じて，使用する計測部位と周波数帯域を設定した．各テンプレートの適合度の計算は，i 番目のルールのファジィラベルに対応した j 番目の入力に対するメンバシップ関数に脳波パワーの面積値を入力し，メンバシップ関数の出力 Gj を掛け合わせたものを適合度 μ_i と定義した（式 1）．μ_i は各入力脳波パターンと各テンプレートとの適合度を表す．出力値は μ_i を重みとした各ルールの後件部値 Z_i の荷重平均とした（式 2）．

$$\mu_i = \prod_{j=0}^{n} G_j \tag{1}$$

$$Z = \frac{\sum_{i=1}^{n}(\mu_i * Z_i)}{\sum_{i=1}^{n}\mu_i} \tag{2}$$

2.3 後件部値の学習

後件部値 Z_i の学習は，最急降下法を用いて，出力値 Z が目標値となる教師信号 T に近似するように学習する（式 3）．

$$Z_i = Z_i' + \rho * \mu_i * (T - Z) \tag{3}$$

ただし，Z_i' は更新前の後件部値であり，Z_i とは異なる値である．本研究では，実験参加者に特定のタスクを指示して，そのタスク時に出現する脳波パターンを L-FTM に入力し，この脳波に対する出力目標値を定めて教師信号とした．本研究では脳波の識別が目的であり，タスク実行時の脳波を区別できるように，適当な数値を教師信号として設定した．ρ は学習係数であり，経験的に 0.9 と設定した．

2.4 プルーニング

自動生成した脳波特徴パターンの探索空間であるルールにおいては，タスク時とノンタスク時のどちらにおいても適合度が高くなるルールが存在している可能性がある．両方に適合度が高いルールが残っている場合，2 状態の識別精度を減少させてしまう．この問題を回避するために，両状態で適合度が高いルールを削除し，タスク状態・ノンタスク状態のどちらか一方に適合度が高いルールのみ保持するプルーニングを，開発した BCI システムに実装した．はじめに各ルールのタスク時・ノンタスク時の適合度の合計値から最大値 $\mathrm{Ob}_{(i)}$ を算出した（式 4）．

$$Ob_{(i)} = max\{\sum_{t=1}^{te} Ot_{(i,t)}, \sum_{t=1}^{te} On_{(i,t)}\} \tag{4}$$

t は時刻，te は最終時刻，$Ot_{(i,t)}$ は i 番目のルール $R_{(i)}$ のタスク時に発現した脳波の時刻 t の適合度，$On_{(i,t)}$ はルール $R_{(i)}$ のノンタスク時に発現した脳波に対する適合度である．次に，タスク時・ノンタスク時の各適合度の合計値の差分を求め，その絶対値 $Os_{(i)}$ を算出した（式 5）．

$$Os_{(i)} = |\sum Ot_{(i)} - \sum On_{(i)}| \tag{5}$$

$Os_{(i)}$ を $Ob_{(i)}$ で正規化し，その値が設定した閾値 th より小さい場合，その i 番目のルールは両状態において適合度が高いルールと判別され，削除される．差分が大きい場合，タスク時・ノンタスク時どちらか一方に適合度が高いルールと判断し，ルールは保持される（式 6）．

$$Pruning(R_{(i)}) \begin{cases} retain, if\left(\frac{Os_{(i)}}{Ob_{(i)}} \geq th\right) \\ delete, if\left(\frac{Os_{(i)}}{Ob_{(i)}} < th\right) \end{cases} \tag{6}$$

保持されたルールセットを用いて，保存済みの学習データの入力値により再度学習を行い，後件部値を再設定する．探索空間であるルールのプルーニングを行うことにより，広範な探索空間から限定した新しいテンプレートセットが構成される．新しいテンプレートセットを用いることで，2 状態の識別精度を高くすることが可能である．

3. 実験手法

3.1 実験概略

全ての実験は関西学院大学「人を対象とする医学系研究」倫理規定に従い，倫理委員会の承認を経て行った．実験には，健常な男性 3 名と女性 2 名（21∼24 歳）を実験参加者としてそれぞれ同日に 5 回行った．実験参加者には脳波計と接続された電極をセットしたヘッドキャップを被り，通常の室内で椅子に座って安静にしてもらい，実験を行った．今回選択した計測部位は国際式 10-20 法に基づき，F4，F3，C3，C4，P3，P4 を選択し，各周波数領域は α 波と β 波に設定した．この組み合わせの 6×2 の 12 入力として，各入力に対してファジィラベルは High，Low の 2 種の組み合わせで 2^{12} の 4096 テンプレートルールが生成される．実験参加者には学習時に腕の上げ下げを 1 秒間隔で繰り返し行ってもらい，その後，タスク時にはこの運動をイメージしてもらった．また，別タスクとして右足を用いて背屈運動を 1 秒間隔で繰り返し行ってもらい，その後タスク時にはこの足の運動をイメージすることを繰り返し行ってもらった．このタスクについて，2 つの運動イメージの識別を検証した．メンバシップ関数の定義域を決定するために運動動作 30 秒・安静状態 30 秒の計 60 秒の初期計測を行った．後件部値の初期値は全て 0 として，運動イメージ時（タスク状態）の学習では教師信号を 5，学習係数を 0.9 とし，安静状態（ノンタスク状態）を教師信号を 0，学習係数を 0.9 としてそれぞれの脳波特徴パターンを学習し，プルーニングを行った．各パラメータはタスク状態とノンタスク状態の 2 状態の識別が可能な値を経験的に設定した．学習後に，30 秒ずつ交互にタスク状態・ノンタスク状態を行うことを 1 セットとして，2 セット行ってもらい，2 状態の識別が可能か検証した．

3.2 結果・考察

腕の運動イメージをタスクとした実験において，各セットの出力の平均値は，実験参加者ごとに，タスク状態で出力値が高く，ノンタスク状態では低くなり，この 2 状態が識別可能であることが確認できた（表 1）．学習時にプルーニングの閾値の値によって保持される有効なルールの数が変動し，タスク・ノンタスク状態の出力値に影響が出るため，閾値の設定を考慮する必要がある．また，タスク状態だけでなく，ノンタスク状

態においても，出力値が高い場合があるのは，入力値である脳波が常に変動して不安定であるため，ノンタスク状態においても偶発的にタスク状態におけるルール（テンプレートパターン）に近い脳波パターンが発現して適合度が高くなった可能性が考えられる．各実験参加者の出力平均値は，おおむね運動イメージを行った場合（タスク状態）の方が大きく，実験参加者 A と E は 2 回のタスク状態とノンタスク状態で有意差があった（それぞれ $p<0.01$， 両側 t 検定）．また，他の実験参加者では有意差が 2 回のタスクの内一方にしかないもの（$p<0.05$，両側 t 検定），両方有意差がないものが存在していた．いずれも傾向として 2 状態の識別が可能であることから，タスク状態における脳波特徴を検出したと考えられる（図 1)．足の運動イメージをタスクとした実験において，各実験参加者の出力平均値は，実験参加者 D と E では 2 回のタスク状態とノンタスク状態で有意差があり，他の実験参加者は 2 回のタスクの内一方だけに有意差があるものとないものがあった（それぞれ $p<0.01$， $p<0.05$，両側 t 検定，表 2)．実験参加者 C においては，足の運動イメージを用いた場合，2 状態の識別が難しいが，腕の運動イメージでは識別可能である．このような場合，足の運動イメージではなく，腕の運動イメージをタスクとすれば機器の制御ができる（図 2)．このように人によってより特徴的な脳波が発現しやすいタスクを選択して機器制御が行えることは本 BCI システムの特徴の一つである．また，5 人の実験参加者の腕と足のノンタスク時の出力を 1 としたとき，それぞれのタスク時の出力値は，2.39±0.63（平均 ± 標準誤差），2.27±0.60（平均 ± 標準誤差）であった（図 3)．腕の運動イメージと足の運動イメージにおいてタスク時の出力はノンタスク時と比して約 2 倍となり，従来識別が難しいと言われている足の運動イメージでも腕の運動イメージと同様にタスク時とノンタスク時の出力値に十分な差があり，運動イメージをタスクとした場合に識別することが可能であった．2 つの実験において，学習時に実際に動作を行い，学習後は運動イメージをタスクとして行った場合に，タスク識別が可能であったことから，この実験タスクにおいては，実際に動作を行った脳波と運動イメージの脳波は類似しており，イメージを行うだけで BCI の動作を発現することが可能であった．実験参加者によって出力値は変動するが，開発した BCI システムを用いることによって，脳波特徴をあらかじめ限定することなく，学習時に実際に動作を行ってその時に発現する脳波を学習することで，運動意思を検出する BCI を実現したと言える．特に全く同じ計測系，プログラムを用いて，足と腕の異なる運動状態を学習してそれぞれ識別が出来たことは，本 BCI システムならではの性能であると考えている．

	一回目		二回目	
実験参加者	タスク	ノンタスク	タスク	ノンタスク
A	1.79±0.22	0.67±0.15	1.86±0.17	0.91±0.26
B	0.25±0.03	0.14±0.02	0.25±0.03	0.19±0.02
C	0.65±0.11	0.37±0.10	0.60±0.11	0.26±0.07
D	0.53±0.06	0.37±0.06	0.55±0.06	0.45±0.07
E	3.00±0.51	0.61±0.31	2.84±0.49	0.60±0.25

表 1: 腕の運動イメージ時の出力値平均 ± 標準誤差．

	一回目		二回目	
実験参加者	タスク	ノンタスク	タスク	ノンタスク
A	0.51±0.03	0.30±0.04	0.39±0.05	0.31±0.03
B	0.05±0.01	$0.02\pm0.31\times10^{-2}$	0.05±0.01	$0.03\pm0.27\times10^{-2}$
C	$0.62\times10^{-2}\pm0.14\times10^{-2}$	$0.57\times10^{-2}\pm0.12\times10^{-2}$	$0.80\times10^{-2}\pm0.23\times10^{-2}$	$0.56\times10^{-2}\pm0.11\times10^{-2}$
D	5.60±0.42	1.98±0.40	4.60±0.62	2.62±0.49
E	2.03±0.40	0.53±0.15	1.60±0.35	0.26±0.04

表 2: 足の運動イメージ時の出力値平均 ± 標準誤差．

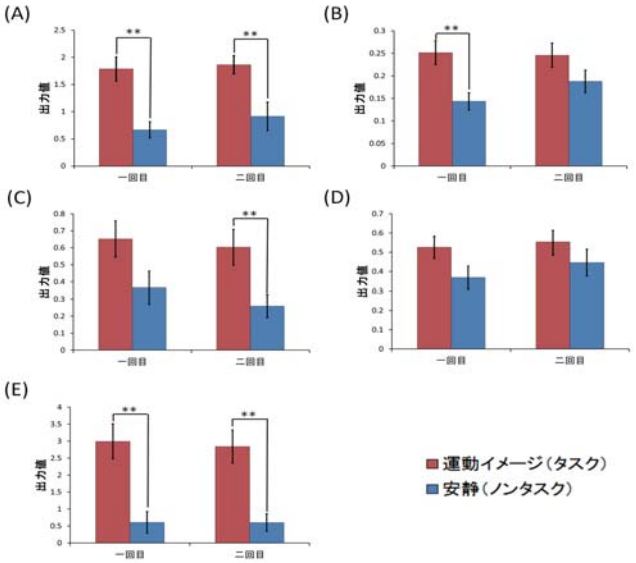


図 1: 腕の運動イメージタスクに対する出力平均値．A)~E) はそれぞれ単一の実験参加者に対応するを示す．エラーバーは標準誤差を示す．

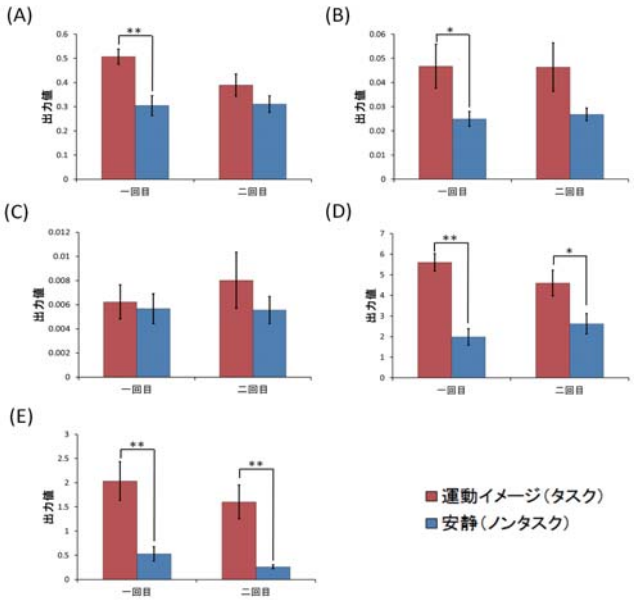


図 2: 足の運動イメージタスクに対する出力平均値．A)~E) はそれぞれ単一の実験参加者に対応するを示す．エラーバーは標準誤差を示す．

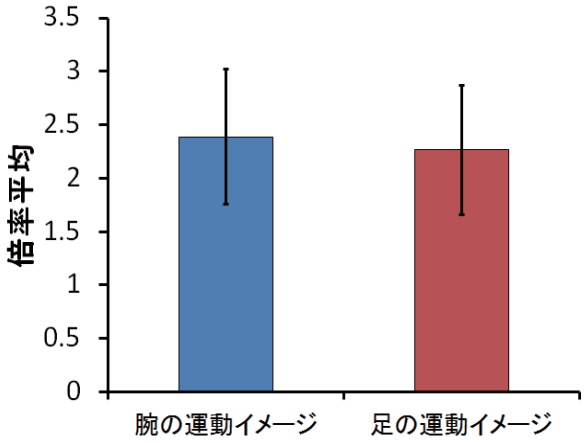


図 3: ノンタスク時の出力を 1 としたときのタスク時の出力の倍率平均．エラーバーは標準誤差を示す．

4. 結論

本研究では，事前情報による特定の計測部位・周波数帯域の脳波特徴パターンを対象とせずに学習によって出現脳波パターンに適合するルールを自動抽出する L-FTM を用いた探索型 BCI を開発した．ファジィラベルによるテンプレートで表現したパターンテンプレートを構成し，これを前件部としたファジィルールから運動動作を行った時に出現する脳波特徴を自動抽出した．運動状態と安静状態のどちらか一方のみに適合度が高くなるルール以外のルールを削除するプルーニングを実装し，学習時と同じ運動をイメージすることで，それに伴う脳波が検出されて，BCI の出力値が上昇することを確認した．本システムは，全く同じ計測系，プログラムを用いて，足と腕の異なる運動状態を学習してそれぞれ識別可能であり，使用者の脳の特性に合わせた柔軟な制御が可能な点が特徴である．

参考文献

[Berger 29] Berger, H.: Über das elektrenkephalogramm des menschen, *European Archives of Psychiatry and Clinical Neuroscience*, Vol. 87, No. 1, pp. 527–570 (1929)

[Daly 18] Daly, I., Blanchard, C., and Holmes, N. P.: Cortical excitability correlates with the event-related desynchronization during brain–computer interface control, *Journal of neural engineering*, Vol. 15, No. 2, p. 026022 (2018)

[Faller 10] Faller, J., Müller-Putz, G., Schmalstieg, D., and Pfurtscheller, G.: An application framework for controlling an avatar in a desktop-based virtual environment via a software SSVEP brain–computer interface, *Presence: teleoperators and virtual environments*, Vol. 19, No. 1, pp. 25–34 (2010)

[Kaufmann 14] Kaufmann, T., Herweg, A., and Kübler, A.: Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials, *Journal of neuroengineering and rehabilitation*, Vol. 11, No. 1, p. 1 (2014)

[Krauledat 08] Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K.-R.: Towards zero training for brain-computer interfacing, *PloS one*, Vol. 3, No. 8, p. e2967 (2008)

[Sakurada 13] Sakurada, T., Kawase, T., Takano, K., Komatsu, T., and Kansaku, K.: A BMI-based occupational therapy assist suit: asynchronous control by SSVEP, *Frontiers in neuroscience*, Vol. 7, p. 172 (2013)

[Takagi 93] Takagi, T. and Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control, in *Readings in Fuzzy Sets for Intelligent Systems*, pp. 387–403, Elsevier (1993)

[山口 12] 山口洋平, 工藤卓:脳波特性変化の自動探索によるブレイン-コンピューター・インタフェース (BCI) の提案, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 24, No. 5, pp. 979–987 (2012)

[市川 08] 市川忠彦：新版 脳波の旅への誘い, 第 2 版 (星和書店) (2008)

[菅野 88] 菅野道夫：ファジィ制御, 日刊工業新聞社 (1988)

[櫻井 07] 櫻井芳雄, 八木透, 小池康晴, 鈴木隆文 : ブレイン-マシン・インタフェース最前線, ブレイン-マシン・インタフェース最前線 (2007)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

異被験者間の脳活動データ相互変換による擬似データ作成

Creating Artificial Brain Activity Data Based on Mutual Data Transformation between Different Subjects

張 嘉瑩 *1 Kaei Cho　小林 一郎 *1 Ichiro Kobayashi　西本 伸志 *2 Shinji Nishimoto　西田 知史 *2 Satoshi Nishida　麻生 英樹 *3 Hideki Asoh

*1お茶の水女子大学 Ochanomizu University　*2情報通信研究機構 脳情報通信融合研究センター National Institute of Information and Communications Technology

*3産業技術総合研究所 人工知能研究センター National Institute of Advanced Industrial Science and Technology

Quantitative analysis of human brain activity has been actively studied in brain and neuroscience. In the analysis, there has recently been getting many opportunities that deep leaning is applied to deal with brain activity data observed with fMRI. Whereas, it costs expensive to collect such data and this always causes the lack of enough data to train a model in deep learning framework. With this background, we aim to artificially increase the number of such data and apply them to a caption generation task with deep learning for raising the accuracy of the model.

1. はじめに

近年，脳神経科学分野において，脳神経活動を定量的に理解する研究が盛んに行われている [Cukur 13, Huth 12, Nishimoto 11]．また，近年の深層学習の成果を取り入れて脳活動データを解読する研究も報告されている．松尾ら [松尾 17] は画像刺激を受けた際の脳活動データを入力としてその人が想起している意味情報，すなわち画像によって想起された事象を深層学習を用いて自然言語文で説明する手法を提案している．そのような深層学習を用いて脳活動データを取り扱う際，学習のために大量のデータが必要である．一方，functional Magnetic Resonance Imaging (fMRI) を用いた脳活動データの収集コストが高いことや，脳のサイズに個人差があるためにたとえ同じ課題の実験であったとしても同じ被験者の脳活動データしか使用できないことにより，機械学習のために使用できるデータ数が不足するといった問題が生じる．

これらの背景を踏まえて，本研究では，異なる被験者の脳活動データ同士の対応関係を学習し一人の被験者のデータに統一することで，データを擬似的に増やし，それを実際に深層学習に使用できるようにすることを目指す．

2. 提案手法

異なる被験者同士の脳活動データを一人の被験者のデータに統一する際に，異なる被験者から取得される脳活動データそれぞれをマルチモーダルデータとみなし，マルチモーダル情報を相互に変換する手法を適用して，ある被験者の脳活動データを他の被験者の脳活動データから擬似的に作り出すことを考える．そして，擬似的に増やしたデータに深層学習を用いた際にモデルの精度を向上させることが出来るかを検証する．

2.1 BiDNN

マルチモーダル情報の対応関係を学習できるものとして，Ngiam らによる Multimodal Autoencoders [Ngiam 11] がある．また，Vukotic らは 2 つの対称的なネットワークを用いてその対称性を考慮した結合荷重の入れ換えを行う Bidirectional Deep Neural Network (BiDNN) を提案している．BiDNN は，Multimodal Autoencoders より精度が良いと報告されている [Vukotic 16] ことから本研究では BiDNN を用いることにする．

BiDNN では，一つ目のモダリティを入力，二つ目のモダリティを出力とするネットワークと，二つ目のモダリティを入力，一つ目のモダリティを出力とする先ほどとは別のネットワークの二つを図 1 のように別々に学習させる．損失関数は平均二乗誤差とする．その際，それぞれのネットワークの中間層を図 1 のように重みを転置して共有させることにより，マルチモーダルの対応関係を取ることができる．

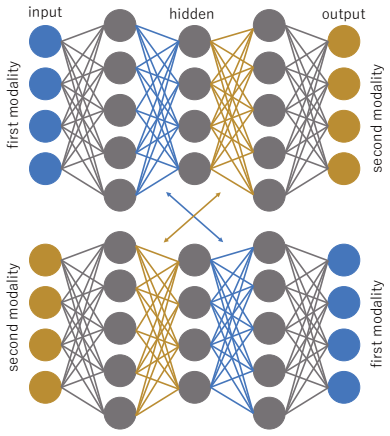


図 1: BiDNN のモデル

2.2 提案手法の概要

図 2 に提案手法の概要を示す．具体的には，動画像タスクを視聴している際の脳活動データを使用するのだが，そのデータを動画像タスク 1 のみを視聴した被験者 A と動画像タスク 1 と 2 を視聴した被験者 B のものと仮定して，以下の手続きを行う．

連絡先: 張嘉瑩，お茶の水女子大学大学 理学部　情報科学科 小林研究室，〒 112-8610 東京都文京区大塚 2-1-1，03-5978-5708，g1420526@is.ocha.ac.jp

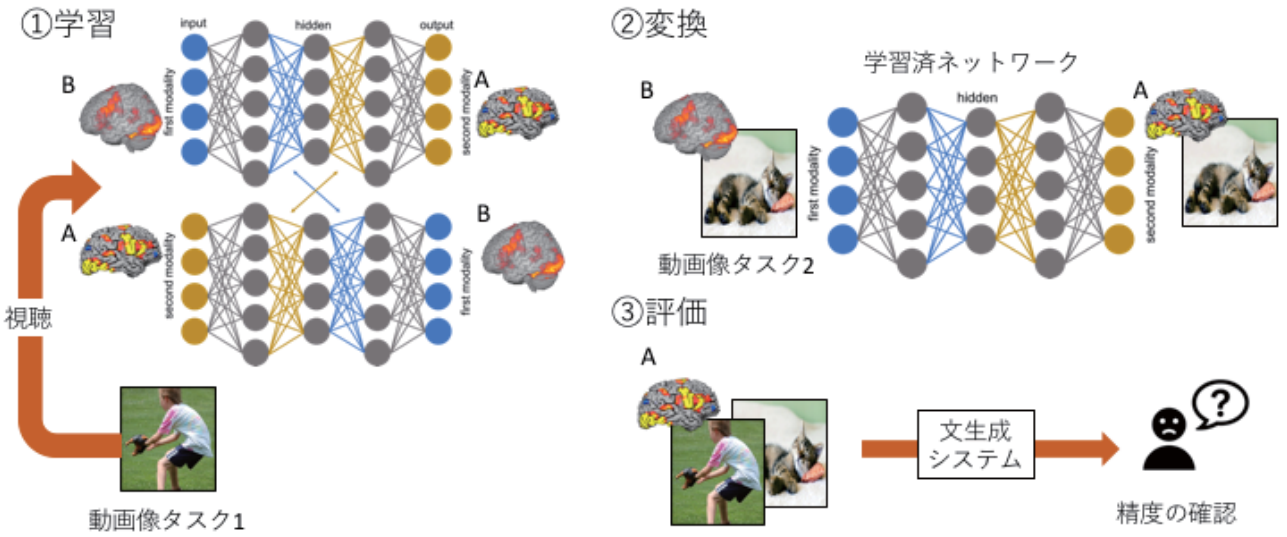


図 2: 本研究の概要

step 1. 学習

BiDNN を用いて，動画像タスク 1 を視聴したときの A と B の脳活動データをそれぞれ入出力とし，対応関係を学習する．

step 2. 変換

step 1. におけるモデルを使用し，学習に用いたタスク 1 とは別の動画像タスク 2 を視聴した B の脳活動データを A の脳活動データに変換する．これにより，擬似的に動画像タスク 2 を視聴した A の脳活動データが得られる．

step 3. 評価

評価においては，松尾らが提案した手法 [松尾 17] を利用する．動画像タスクを視聴しているときの脳活動データを入力とし，その際，被験者が頭の中に想起している意味情報の説明文を出力する．動画像タスク 1 のみを視聴した脳活動データを入力としたときと比べて，提案手法を用いて擬似的に作成した動画像タスク 2 を視聴したとされる脳活動データを加えて入力としたときに出力された説明文の精度が上がっているのか確認する．

3. 実験

3.1 学習

まず Vukotic ら [Vukotic 16] によって提案された BiDNN を構築し，異なる被験者の脳活動データ同士の対応関係を学習する．

使用するデータは，動画像タスク 1 を被験者に視聴させたときの血中酸素飽和度信号 (BOLD 信号) を．fMRI を用いて記録した脳活動データである．被験者 A の脳活動データは，脳活動の観測領域 96×96×72 ボクセルのうち，皮質に相当する 62,552 次元のデータを，被験者 B の脳活動データは，脳活動の観測領域 96×96×72 ボクセルのうち，皮質に相当する 70,933 次元のデータを利用している．データ数は，train が 6,600 サンプル，test が 600 サンプルである．

先行研究 [Vukotic 16] では，深層学習のフレームワーク Theano を利用しているが、本研究では同じ深層学習のフレームワークである Chainer [*1] で実装した．

まず，現有する計算機の性能では入出力に約 60,000 次元のデータが使えなかったため，AutoEncoder を用いて二人の被験者の脳活動データをそれぞれ 7,500 次元に次元に圧縮した．学習アルゴリズムは確率的勾配降下法，損失関数は平均二乗誤差を用いている．

次に，BiDNN を用いて，二人の被験者同士の脳活動データを学習した．学習のユニット数は 7,500 - 5,500 - 2,000 - 5,500 - 7,500 とし，学習アルゴリズムは確率的勾配降下法，損失関数は平均二乗誤差を用いている．

3.2 変換

次に，先の実験で学習したモデルを使用して，データの変換をした．

先の実験とは異なる動画像タスク 2 を被験者に視聴させたときの血中酸素飽和度信号を記録した脳活動データを使用した．こちらは被験者 B のみで，70,933 次元のデータである．データ数は 9,000 サンプルである．

まず，入出力に約 60,000 次元のデータが使えないので，AutoEncoder を用いて 7,500 次元に圧縮した．そして，先の実験で学習したモデルを使用して，データの変換をした．これにより，被験者 A が擬似的に動画像タスク 2 を視聴したとする脳活動データが得られる．このデータの次元数は 7,500 で，データ数は 9,000 サンプルである．

3.3 評価

最後に，松尾らが提案した手法 [松尾 17] を利用して評価をする．図 3 に概要図を示す．まず，動画像視聴時に見ていたとされる動画像から切り出した静止画に対して caption 生成システムを用いて生成した説明文を正解文とする．次に，動画像視聴時の脳活動データを入力とし，その時想起している言語意味情報の説明文と先ほどの正解文を BLEU スコアで評価する．

正解文の生成に用いる静止画のデータ数は，動画像タスク 1 から切り出したものが 7,200 サンプル，動画像タスク 2 から切り出したものが 9,000 サンプルである．また，脳活動データは被験者 A に動画像タスク 1 を見せたときに観測した 7,200 サ

*1 https://chainer.org/

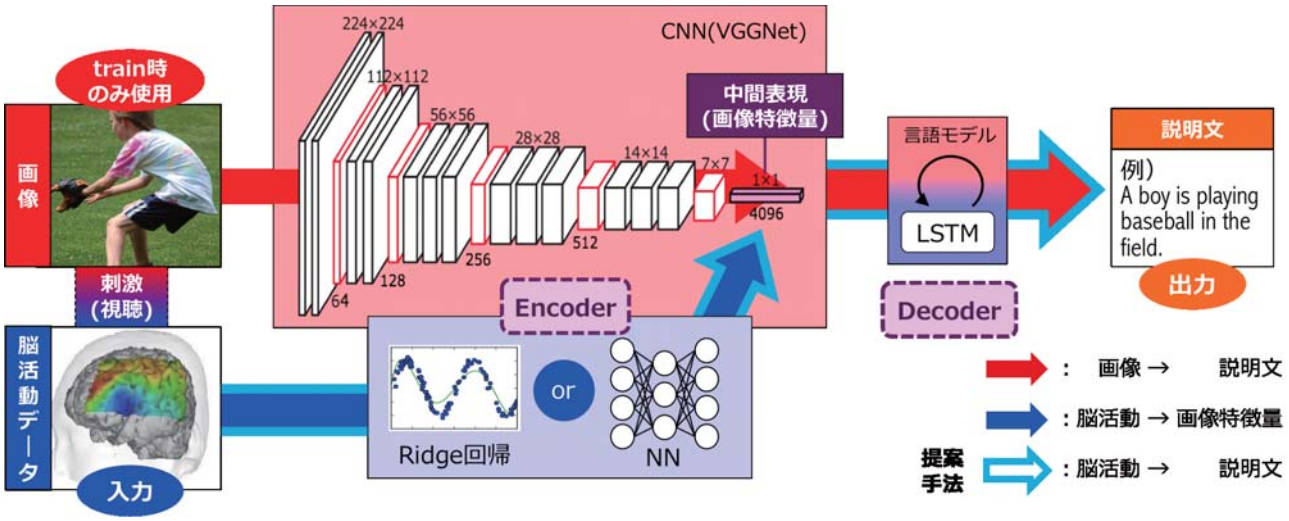


図 3: 脳活動データからの文生成の概要 [松尾 17]

ンプルと実験 2 より擬似的に得られた 9,000 サンプルと，被験者 B に動画像タスク 1 と動画像タスク 2 を視聴させたときにそれぞれ観測した 7,200 サンプルと 9,000 サンプルである．

評価実験を 2 つ行う．まず，被験者 A のデータのうち，動画像タスク 1 を視聴させたときの 7,200 サンプルを使用したときと，動画像タスク 1 を視聴させたときに擬似的に動画像タスク 2 を視聴させたときを加えた 16,200 サンプルを使用したときの精度を計測する実験である．これにより，擬似的に 9,000 サンプル増やした際に精度にどのような影響を与えるのかを見ることができる．次に，被験者 B データのうち，動画像タスク 1 を視聴させたときの 7,200 サンプルを使用したときと，動画像タスク 1 を視聴させたときに動画視聴タスク 2 を視聴させたときを加えた 16,200 サンプルを使用したときの精度を計測する実験である．これにより，実際に視聴している 9,000 サンプルを増やした際に精度にどのような影響を与えるのかを考察することができる．

上記二つの評価実験における BLEU スコアを表 1 に示した．BLEU スコアは，train データと test データの双方について求めた．

表 1: 評価実験における BLEU スコア

被験者	タスク	BLEU スコア (train / test)
A	1	0.5026 / 0.5027
	1, 2(擬似)	0.5238 / 0.5099
B	1	0.5028 / 0.5085
	1, 2	0.5191 / 0.5086

3.4 考察

被験者 A における評価実験の結果を見ると，train と test のどちらも擬似的なデータを増やした際にはわずかに BLEU スコアが上がっている．また，被験者 B における評価実験の結果を見ても，train と test のどちらもデータを増やした際にはわずかに BLEU スコアが上がっている．当初予想していたほどスコアは上がらなかったが，擬似的ではなく実際にタスク 2 を視聴した際も BLEU スコアが大きくは上がらないことから，提案手法が機能していないとは決して言えない．データを増やしてもスコアがあまり上がらなかった原因の一つに，使用したデータが視聴している動画に問題があることが考えられる．実際に，使用した動画の中には表 2 のように直感的に説明文をつけることが難しいものが含まれている．

表 2: 上:説明しやすい画像例、下説明しづらい画像例



A boat is docked in the water near some buildings.



A man is riding on the back of an elephant

.

4. 追加実験:ネットワーク層の削減

深層学習において高次元のデータを扱う際にはモデルが複雑であると学習が困難になることがある．そこで先の実験では 5 層の BiDNN のモデルを用いているが，3 層のモデルを用いることを考えた．

4.1 概要

5 層の BiDNN のモデルの学習のユニット数は 7,500 - 5,500 - 2,000 - 5,500 - 7,500 とし，3 層の BiDNN のモデルの学習のユニット数は 7,500 - 4,000 - 7,500 とした．学習アルゴリズムは確率的勾配降下法，損失関数は平均二乗誤差を用いている．

それぞれのモデルで学習，変換をした際に精度がどのように変化したかを確認する．まず，学習の際に各 epoch 数ごとの平均二乗誤差を図 4 と図 5 に示した．5 層のときには loss が学習を重ねるごとに loss が上昇する部分が見受けられるが、3 層では比較的収束している様子が見られた．

次に精度を確認した．動画タスク 1 のみを視聴させた被験者

の脳活動データを使用した．動画タスク 2 を擬似的に視聴したとするデータを加える際に，3 層の BiDNN もしくは 5 層の BiDNN を用いて，先の実験のように評価をした際の BLEU スコアを表 3 に示した．train ではわずかに下がったが，test では train で下がった分よりも上がっていることが確認できた．

4.2 考察

3 層のときの方が精度が良いとはっきりとは言い難い結果となった．3 層の方が誤差は収束していたが，良いモデルにおける平均二乗誤差が 5 層では約 2.8，3 層では約 3.9 と高かったためだと考えられる．

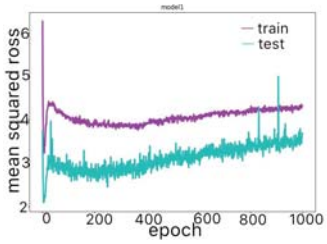


図 4: 5 層の図

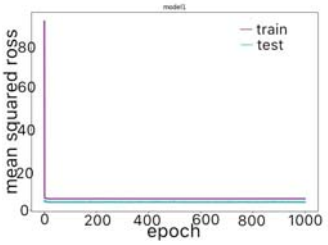


図 5: 3 層の図

表 3: 追加実験における BLEU スコア

BiDNN の種類	BLEU スコア (train / test)
5 層	0.5238 / 0.5099
3 層	0.5197 / 0.5188

5. おわりに

本研究では，異なる被験者同士の脳活動データを相互変換するために BiDNN を用いた手法を提案した．脳活動データから被験者が想起している言語意味情報を説明文として出力することにより，提案手法の精度を確認し，一応の精度向上が見られたことから，概ね提案手法は有効だと考えられる．

また，追加実験により BiDNN モデルが 5 層のときと 3 層のときの精度の確認をしたが，精度向上はあまり見受けられなかった．

今後の課題として，別被験者のデータでも実験を行い今回の結果の有意性を検討したい．また，直感的に説明文をつけやすい動画のみを使用した際に精度がどのように変わるかを確認するなどを行い，本研究の提案手法の検討を重ねたい．

参考文献

[Cukur 13] Cukur, S. Nishimoto, A. G. Hut, and J. L. Gallant, “Attention during natural vision warps semantic representation across the human brain”, Nature Neuroscience 16 (2013).

[Huth 12] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, “A continuous semantic space describes the representation of thousands of object and action categories across the human brain”, Neuron, 76(6), 1210-1224 (2012).

[Ngiam 11] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning”, In Intl. Conf. on Machine Learning (2011).

[Nishimoto 11] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant., “Reconstructing visual experiences from brain activity evoked by natural movies”, Current Biology, 21(19), 1641-1646 (2011).

[松尾 17] 松尾映里, 小林一郎, 西本伸志, 西田知史, 麻生英樹, “画像説明文生成手法を援用した画像刺激時の脳活動の説明文生成”, 言語処理学会, P6-2 (2017).

[Vukotic 16] V. Vukotic, C. Raymond, G. Gravier, “Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Cross-modal Applications”, he 2016 ACM International Conference on Multimedia Retrieval (ICMR), pages 343-346, ACM, New York (2016).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

聴神経の計算論再考：教師なし学習か課題最適化か？

Reconsidering the computational model for the auditory nerve: unsupervised learning or task-based optimization?

寺島 裕貴　　古川 茂人
Hiroki Terashima　Shigeto Furukawa

NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

Efficient coding has been a leading computational principle to understand the sensory systems. The auditory nerves have been explained by unsupervised learning of natural sounds: their filter shapes resemble the basis optimized to code human voice. However, the previous study did not clearly distinguish sounds recorded in studios and those in the environment. We found that the model fails to explain the auditory nerves when applied to environmental recordings because of reverberations. How can we model the auditory nerves including the environmental modulations? We hypothesized that the auditory nerves are optimized to perform auditory tasks we face in the environment. To test this, we trained a deep convolutional neural network to classify phonemes based on their reverberated waveforms. The filters learned in the first layer showed characteristics similar to the auditory nerves. The results suggest that the auditory nerves efficiently encode task-related information rather than the entire incoming signal.

1. はじめに

哺乳類の耳の物理的なサイズは種によってさまざまだが，聴神経が持つ周波数チューニングのフィルタ特性は種を越えて普遍的な分布に従うため [4]，何らかの機能的意義が示唆される．この特徴をどのように理解すればよいだろうか．

脳神経系の情報処理様式を理解する方策として，脳の各部位を計算アルゴリズムと対応付けていくという考え方がある．その中で，低次感覚系は教師なし学習と対応付くと考えられてきた．初めに，大脳皮質一次視覚野が自然画像のスパース符号を用いて説明され，神経系と自然刺激統計性が深い関係にあることが示された．これを受けて聴覚系では，自然音，その中でも特にヒト音声のスパース符号が聴神経のフィルタ特性と一致することが報告された [3]．

この研究 [3] の結果を素直に捉えると，聴覚末梢は教師なし学習とよく対応するように思われる．しかし前報 [7] では，この定説に疑問を投げかける現象を報告した．すなわち，先行研究で用いられた「自然音」よりもさらに耳への入力に近い，屋内外で話者から離れた場所で録音した音声を用いた場合には，先行研究のようには聴神経特性を説明できなかった．

本報告では，自然残響下でもなお聴神経特性を説明できるモデルとして，残響下の音素識別タスクを課した深層ニューラルネットワークを提案する．この結果は，単純な教師なし学習よりも，自然な音響的タスクに対する最適化のほうが聴覚末梢を理解するための枠組みとしてより有望であると示唆している．

2. 手法

2.1 自然音の教師なしスパース符号化

Lewicki [3] の結果と直接比較できるよう，同様の基底学習法を用いた．入力 $\boldsymbol{x}$ を $\boldsymbol{x} = \sum_i a_i \boldsymbol{\phi_i}$ のように基底 $\boldsymbol{\phi_i}$ の線形和で表現し，係数 a_i がよりスパースになる基底 $\Phi = \{\boldsymbol{\phi_i}\}$ を下記のように更新して学習した．詳細は原論文 [3] を参照．

$$\Delta\Phi \propto \Phi\Phi^T \frac{\partial}{\partial\Phi} \log p(\boldsymbol{x}|\Phi) \tag{1}$$

$$= \Phi(I - \mathrm{sign}(\boldsymbol{a})\boldsymbol{a}^T) \tag{2}$$

また，得られた基底ベクトルそれぞれのフィルタ特性の解析手法も原論文に準拠し，周波数チューニングとそのシャープネス（Q 値）を求めた．

2.2 自然音・自然残響データベース

自然な屋内環境におけるヒト音声の録音として，NTT 乳幼児音声データベース [1] のうち大人同士の会話としてタグ付けされたもののみを抽出して用いた．また，音素識別タスクには TIMIT データベースを用い，[2] に倣って 61 の音素ラベルを 39 クラスに再編した．自然な空間残響のシミュレーションに際しては，屋内外で録音された自然インパルス応答データベース [5] を用いた．

2.3 深層ニューラルネットワークによる音素識別

自然な音響的タスクへ最適化されたフィルタ形状を調査するため，音声波形を直接入力する深層ニューラルネットワークを構築した．入力は 16 kHz 長さ 2,000 の音声波形で，ネットワークの構造は 5 層の畳み込みニューラルネットワークである（詳細は図 2）．訓練後 1 層目に獲得されたフィルタの特性を [3] と同様に分析した．

聴神経フィルタの時間窓と類似の時間スケールにおける自然な音響的タスクとして，入力波形の中心部に位置する音素の識別をニューラルネットワークに課して訓練した．その際，自然残響下での課題を模擬するため，各入力には自然なインパルス応答をデータベース [5] からランダムにひとつ選んで畳み込んだ．

3. 結果

3.1 残響下自然音スパース符号と聴神経の不一致

聴覚末梢の教師なし学習モデル研究によれば，純粋なヒト音声録音のスパース符号と聴神経フィルタの特性は類似する [3]．

連絡先: 寺島裕貴，NTT コミュニケーション科学基礎研究所，〒243–0198 神奈川県厚木市森の里若宮 3–1，teratti@teratti.jp

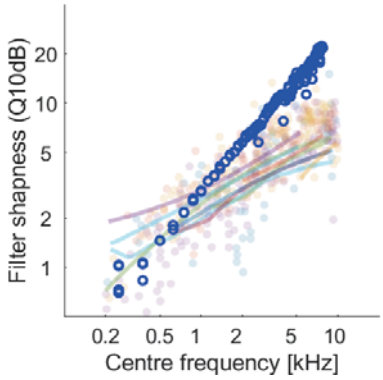


図 1　ヒト音声の自然環境下録音について最適なスパース符号の周波数チューニングとそのシャープネスの分布．聴神経の特性（半透明）よりもシャープなチューニング特性を示す．

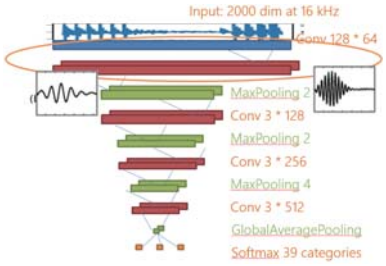


図 2　自然な残響下のヒト音声を入力とし，音素識別タスクで訓練したニューラルネットワークのアーキテクチャ．第 1 層の畳み込みフィルタと聴神経を比較する．

前報 [7] では，より自然な屋内環境で録音したヒトの音声でこの結果を再現できるかどうか調べるため，NTT 乳幼児音声データベースの中から大人同士の発話のみを抽出し，同アルゴリズムを適用した．

図 1 に各フィルタの周波数チューニングとそのシャープネスの分布を示す．背景に半透明で重畳しているのは，多種の哺乳類で記録されたデータの再プロットである [4]．前報 [7] で報告したように，自然界における残響の影響を主に受け，シャープネスが聴神経よりも高い方向にシフトした分布が得られた．

3.2　音素識別課題最適化による聴神経特性の再現

聴覚末梢を理解するために，教師なし学習以外の枠組みとして有望なものは何だろうか．近年，自然な視覚課題に対する最適化が視覚系階層性の理解に有用だと考えられつつある [6]．聴覚系の特性もまた，自然な音響的課題に対する最適化を用いて説明できるとの仮説のもと，深層ニューラルネットワークを構築して解析を行った．図 2 に示す畳み込みニューラルネットワークを，自然な残響を模擬した波形を入力して TIMIT 音素識別で訓練した．

訓練後の第 1 層に獲得されたフィルタ特性の分布を図 3 に示す．残響下の音声波形が入力されているにもかかわらず，哺乳類で見られる聴覚フィルタ特性分布を再現できていることがわかる．これは単純な教師なし学習を用いた場合（図 1）と対照的である．

4.　結論

聴覚末梢の特性は，自然音声の教師なし学習として理解できるとされてきた．しかしわれわれは前報で，音声の自然な録音であっても，聴覚系への入力により近い残響下での録音を用いると聴覚末梢の特性を説明できなくなることを報告した．本報告では，残響下の音声を入力として用いつつ，深層ニューラルネットワークを音素識別課題で訓練することで聴覚末梢の特性を説明した．聴覚末梢を理解するための枠組みとして，単純な教師なし学習よりも自然な音響的課題に対する最適化がより適切だと考えられる．

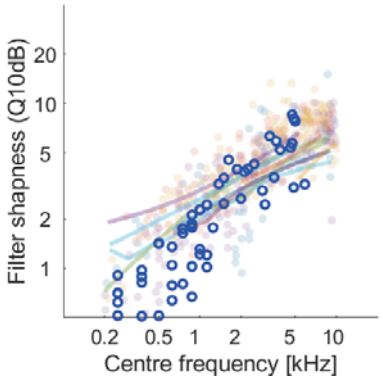


図 3　自然な残響下のヒト音声を入力として訓練したニューラルネットワーク第 1 層の周波数チューニングとそのシャープネスの分布．聴神経の特性（半透明）と似たチューニング特性分布を示す．

謝辞

本研究は JSPS 科研費 JP15H05915（新学術領域研究：多元質感知）の助成を受けたものです．

参考文献

[1] Amano S, Kato K, and Kondo T. Development of japanese infant speech database and speaking rate analysis. In *ICSLP - INTERSPEECH*, 2002.

[2] Lee KF and Hon HW. Speaker-independent phone recognition using hidden markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**(11):1641–1648, 1989.

[3] Lewicki MS. Efficient coding of natural sounds. *Nature Neuroscience*, **5**(4):356–363, 2002.

[4] Ruggero MA and Temchin AN. Unexceptional sharpness of frequency tuning in the human cochlea. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(51):18614–18619, 2005.

[5] Traer J and McDermott JH. Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(48):E7856–E7865, 2016.

[6] Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, and DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, **111**(23):8619–8624, 2014.

[7] 寺島裕貴, 古川茂人. 自然残響の考慮による聴覚音声スパース符号化再考. In 第 *31* 回人工知能学会全国大会論文集, 2K5–OS–33c–2. 人工知能学会, 2017.

コネクトーム上で機械学習器を結合するための大脳新皮質領野間トップダウン/ボトムダウン分類

Top-down and bottom-up classification between areas in mouse cerebral cortex to connect machine learning modules on connectomes

速水琢[*1] 根岸蒼[*1] 小森麟太郎[*1] 水谷治央[*1, 2] 山川宏[*1, 2]
Taku Hayami So Negishi Rintaro Komori Haruo Mizutani Hiroshi Yamakawa

[*1] ドワンゴ人工知能研究所 Dwango Artificial Intelligence Laboratory
[*2] 全脳アーキテクチャイニシアティブ Whole Brain Architecture Initiative

The Whole Brain Architecture (WBA) is a strong candidate for a computational cognitive architecture of an artificial general intelligence (AGI) computing platform on the basis of empirical neural circuits (aka connectomes). The WBA requires top-down and bottom-up relationship between machine learning modules, which is corresponding to brain areas, to exert brain-like multiple cognitive functions and behaviors in a computational system. In order to implement the brain-like cognitive functions into artificial machines, we created a part of platform based on connectomic architecture resulting from quantified neural connectivity analysis of the mouse brain provided by Allen Institute for Brain Science. Neural strengths and hierarchies of connections between areas in the cerebral cortex were analyzed by linear regression and clustering methods, and the result can be useful to connect machine learning module to develop AGI more efficiently. We suggest that computational cognitive architecture defined by connectomic data can enhance collaborative AGI development.

1. はじめに

近年の深層学習の発展により, 特定の問題領域を解決する特化型人工知能は大きな発展を遂げてきた. しかし, 多様な問題領域に対応できる汎用人工知能はこれまで実現されていない. 全脳アーキテクチャ(WBA: Whole Brain Architecture)は脳に学んだ認知アーキテクチャであり, 汎用人工知能を実現するための有用な基盤である [Yamakawa 2016]. 全脳アーキテクチャでは, 脳の各器官を機械学習モジュールとして実装し, 各器官を統合することで認知アーキテクチャを構築する. この構築で必要になるのは, 脳器官の働きを模した機械学習モジュールを統合するための統一的なプラットホームである. 我々はこの各モジュール間の結合を「コネクトーム」と呼ばれる定量的な神経科学的結合データを元に構築することで, 効率的な WBA の開発が行えると考えている. 本研究では, コネクトームを基にした AGI 開発のプラットホームを構築することを目指し, その中心的基盤を全脳コネクトームアーキテクチャ(WBCA: Whole Brain Connectomic Architecture)と呼ぶこととする. コネクトームデータは、Allen Institute for Brain Science が提供しているマウス脳の神経結合性データベース(http://connectivity.brain-map.org/)を用いている.

2. 方法

2.1 Allen Mouse Brain Connectivity Atlas

Allen Mouse Brain Connectivity Atlas とは Allen Institute for Brain Science が提供しているマウス脳の神経結合データベースである [Oh 14]. このデータベースには 491 の野生型マウスの実験データが含まれており, その実験の一つ一つでは次の手順で解析が行われている. まず, ある脳領域の神経細胞群に蛍光タンパク質をコードしたアデノ随伴ウイルス(AAV: Adeno Associated Virus)を感染させることで, 感染した神経細胞群を蛍光タンパク質の蛍光によって可視化する.次に, 脳を 100μm ごとに冠状面で切断し, それぞれの切片の蛍光の体積を計測する. 最後に, マウスの標準脳と照らし合わせることで脳領域ごとの蛍光の体積を求める. 以上より, ある脳領域への注入体積に各脳領域への投射体積が対応するデータベースとなっている.

連絡先:山川宏, (株)ドワンゴ ドワンゴ人工知能研究所,
〒113-0033 東京都文京区本郷 5-24-5 角川本郷ビル

2.2 脳領域間の神経結合強度の算出

WBA 内の機械学習モジュールの結合においては、コネクトームの神経結合強度を採用する方針である。神経結合強度はある脳領域から別の脳領域へ投射する神経細胞の情報伝達量で、強度が強いほど脳領域間の情報のやり取りが多いとする. Allen Mouse Brain Connectivity Atlas から取得した各領域の注入体積と投射体積に線形結合モデルをあてはめ, 非負制約付き最小二乗法で重回帰分析により神経結合強度を求める(式.1) 得られた回帰直線の傾きを注入脳領域と投射脳領域間の神経結合強度と定義した.

$$min_{W_{X,Y}\geq 0} \sum_{i=1}^{|S_E|} \left(\sum_{X \in S_X} W_{X,Y} PV(X \cap E_i) - PV(Y) \right)^2$$

$X : Source\ region$
$Y : Target\ region$
$W_{X,Y} : Normalized\ connection\ strength$
$PV : Projection\ volume$
$E_i : Set\ of\ all\ voxels\ containing\ neuron\ infected\ on\ the\ i^{th}\ injection$

式.1 結合強度算出の回帰式

2.3 大脳新皮質領域における Feedforward / Feedback 結合の決定

WBCA では, 情報の流れの方向性を定義しておかなければならない. 各脳器官の機能を模倣するための典型的な機械学習の手法を適用するためには, 情報の流れの方向性を定義する必要がある. そこで, ある脳領域からある大脳新皮質領域の各層 (1, ⅔, 4, 5, 6a, 6b) への結合強度を元に, 主成分分析とクラスタリング解析を用いて Feedforward (FF) と Feedback (FB) を決定し

た. この解析は, 大脳新皮質の層構造を利用した解析方法であるため, 大脳新皮質への入力でなければ FF/FB を定義することはできない. よって, 大脳新皮質内の脳領域間および視床の脳領域から大脳新皮質の脳領域についてのみ解析を行なった. しかし, 視床-大脳皮質ループ回路に代表されるように大脳皮質から視床への結合も非常に重要な役割を担っているため, この結合を無視することはできない. そこで, 大脳皮質から視床への結合の FF/FB を視床から大脳皮質への結合に合わせることで FF/FB を決定した.

(1) 主成分分析

大脳新皮質内の脳領域の各層(1, ⅔, 4, 5, 6a, 6b) への結合強度に主成分分析を行ない, 各主成分の寄与率と各層の因子負荷量を算出した.

(2) クラスタリング解析

大脳新皮質内の脳領域の各層への結合強度に K-means クラスタリング(K=2) を行なった. そして, これまでの知見との整合性が高くなるように二つのクラスに FF/FB を割り当てた. なお, 各脳領域の結合に FF/FB を割り当てる問題は分類問題として解けるが, 結合数に対して正解データ数が極めて少なく(大脳新皮質内: 47/1638(2.9%)),正しく分類できないことが予想されたためにクラスタリング解析によって FF/FB を割り当てた.

(3) 神経結合強度の閾値設定

典型的な機械学習の手法を用いる WBA アプローチでは, 情報の流れが巡回しないようにする必要がある. さらに, 神経科学的にも情報の流れが巡回してしまうことはない. よって, FF と FB のそれぞれが有向非巡回グラフ(Directed Acyclic Graph: DAG) となるような最小の結合強度を閾値として用い, 閾値に達しなかった結合は対象から除外した.

2.4 脳領域モジュール間ネットワークの可視化

クラスタリング解析と結合強度閾値の設定で決定した FF/FB を直感的に理解し易くするためにブロック図を用いて脳領域モジュール間の結合性ネットワークを描画した.

3. 結果と考察

3.1 モジュール間ネットワークの可視化

算出された各脳領域間の結合強度のうち, 大脳新皮質にある視覚野の結合強度を図.2 に示す. 例として、一次視覚野はその他の領域との結合が強いことが示されている. 本神経結合性解析は、大脳新皮質のみならず、全脳領域にも対しても行っており、脳モジュール間の定量的な接続サイズを算出する上で有効であると考えられる.

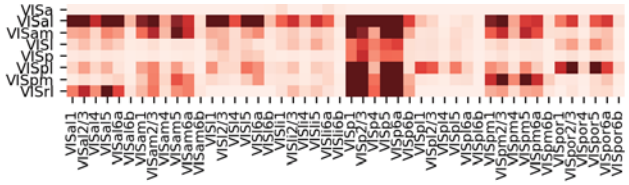


図.2 大脳新皮質内の視覚系結合強度ヒートマップ

なお, この結合強度は線形結合モデルを非負制約付き最小二乗法で回帰しているため, 全ての脳領域に十分な数の注入と投射がないと結合強度の値が正しく求まらなくなってしまう. これには, 3 つの解決方法が考えられる. 1 つ目は, さらなる実験データの蓄積を期待すること. 2 つ目は, 領域ごとではなくピクセルレベルで回帰を行うこと. 3 つ目は, 遺伝子改変マウスによる実験データを解析することである. 今後, 2 つ目と 3 つ目について検討していく予定である.

3.2 FF/FB 結合性解析

(1) 主成分分析

大脳新皮質内の脳領域の各層への結合強度に主成分分析を行い, 第二主成分(Primary component 2: PC2)までの寄与率を表.1 に示した. 大脳新皮質内の結合も視床から大脳新皮質の結合も PC2 までで累積寄与率が 90%に達した. よって, PC1 と PC2 のなす平面状に結合をプロットした.

表.1 第二主成分までの寄与率(主成分分析)

	寄与率	
	PC1	PC2
大脳新皮質内	96%	2%
視床-大脳新皮質間	80%	11%

大脳新皮質内の結合では, PC1 が大きい部分に,Primary Visual Area (VISp) へ入力する結合が多い一方で, PC1 の小さい部分には VISp から出力する結合が多かった. なお, 因子負荷量を計算することにより, PC1 は 6a, 5, ⅔, 1 層が比較的大きな影響を与えていることが分かった.

視床から大脳新皮質の結合では, PC1 が小さい部分に, Dorsal part of the lateral geniculate complex (LGd) から VISp, Medial Geniculate Complex (MG)から Primary Auditory Area (AUDp) への結合が位置していた. これらは視聴覚の Feedforwad 結合としてよく知られている結合である. なお,因子負荷量から, PC1 は 5, 6a, 2/3 層が比較的大きな影響力を有していることが分かった.

(2) クラスタリング解析

大脳新皮質内の脳領域の各層への結合強度に K-means クラスタリング(K=2) を行ない, Feedforward 結合と Feedback 結合を決定した. 決定した FF/FB を主成分分析の結果の上に重ねて, 色分けして表示した(図.3). 1638 の結合のうち, 1566 (95.6%)が FF で 72 (4.4%) が FB となった. FF には, 一次領野 (VISp, AUDp など)からその他の領野への結合が多く含まれていた. 一方, FB にはその他の領野から一次領野への結合が多かった. これは, これまでの知見と一致するものである.

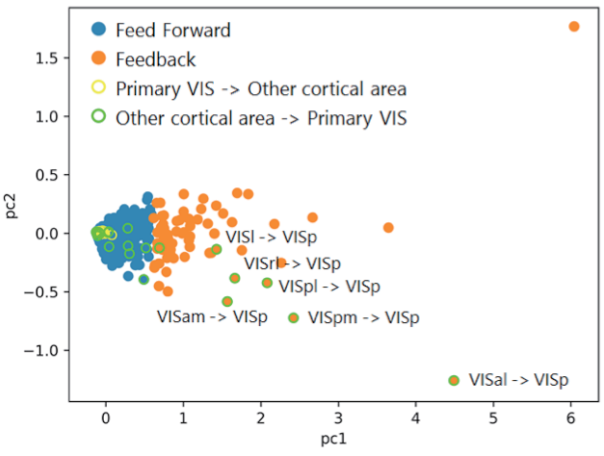


図.3 大脳新皮質の脳領域間の主成分分析とクラスタリング解析

視床から大脳皮質の結合についても同様に K-means クラスタリング（K=2）を行ない，決定した FF/FB を主成分分析の結果に重ねて，色分けして表示した．(図 4) 1398 の結合のうち 1376 (98.4%)が FB で 22 (1.6%) が FF となった．FF には，大脳皮質への感覚入力経路としてよく知られている LGd から VISp への結合と MG から AUDp への結合が含まれていた．

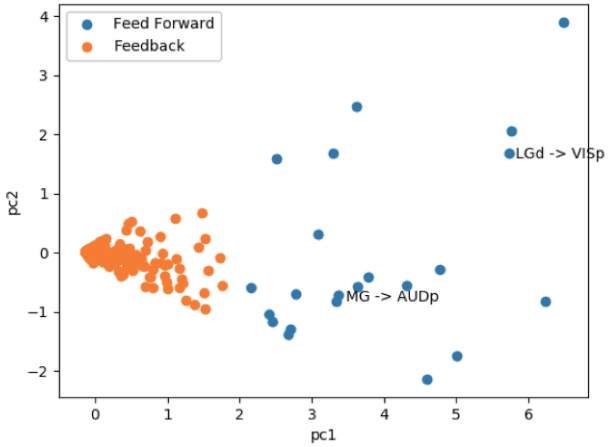


図.4 視床と大脳新皮質領域間の主成分分析とクラスタリング解析

(3) 結合強度閾値の設定

大脳皮質内の各脳領 DAG となるための最低の結合強度閾値は 0.843 となった．一方，視床から大脳皮質の結合では FF と FB で共通の閾値を設定すると結合強度の比較的弱い FF の結合が一切なくなってしまうことから、FF と FB のそれぞれが DAG となるようにそれぞれに閾値を設けた．それぞれの閾値は，0.0807, 0.4432 となった．

しかし，これらの結合強度閾値は比較的高いため，多くの結合を無視してしまっている．この点に関しては，検討する必要がある．

主成分分析とクラスタリングの結果はほぼ同じ傾向を示した．そしてその結果は，これまでの知見と一致していた．よって，大脳皮質の各層への結合強度を用いることで FF/FB を決定できる可能性が示唆された．

最後に，クラスタリング解析の結果を元に作成したブロック図の一部（大脳新皮質視覚野）を図.5 に示した．細い矢印が Feedforward，細い矢印が Feedback 性の神経結合を表している。脳領域間の FF/FB をブロック図として表すことで脳型 AGI を開発する人々が直感的に脳領域モジュール間の情報の流れを理解できると思われる．

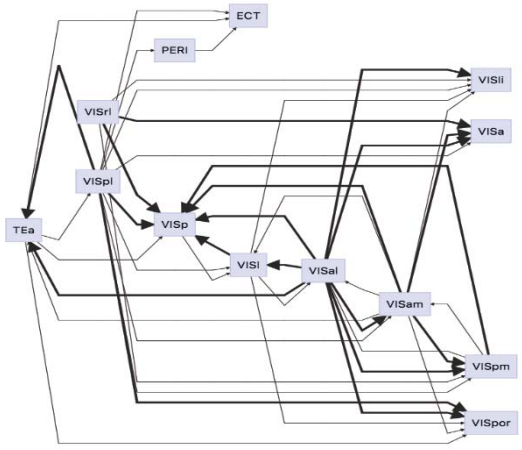


図.5 大脳皮質視覚野の FF/FB 性結合を反映したブロック図

4. 結論

本研究では，Allen Institute for Brain Science から提供されているマウス脳の定量的な実験結果をもとに各脳領域間の結合強度解析とその結合の方向性の定義を行った．そして，それを元に全脳アーキテクチャ(WBA)の統一的なプラットホームである全脳コネクトームアーキテクチャ(WBCA)を構築した．これにより，各脳領域間における情報処理の流れを開発者が明確に把握するができるため，実際の脳内ネットワークに基づいて機械学習器を結合させることが可能となり，汎用人工知能の開発が促進されることが期待される．なお，今回構築した WBCA は GitHub に公開されている(https://github.com/wbap/WBCA) ため，脳型 AGI 共同開発の足掛かりとして利用していただきたい．

参考文献

[Yamakawa 2016] Yamakawa et al.: Whole brain architecture approach is a feasible way toward an artificial general intelligence, ICONIP:275-281 ,2016.

[Oh 2014] Oh et al.: A mesoscale connectome of the mouse brain, Nature 508:207-14, 2014.

[2B4] [General Session] 1. Basis / Theory

座長:佐々木 耀一（NEC）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room B (4F Moon Light)

[2B4-01] Human body as a set of rigid units: finding correspondence between points of two sparse point-sets under distance-preserving transformation

○Torii Takuma[1], Shohei Hidaka[1] （1. Japan Advanced Institute of Science and Technology）

5:20 PM - 5:40 PM

[2B4-02] An attempt of feature extraction of neuronal activity patterns in living neuronal network, by multi-layered artificial neuronal network

○Hiromichi Sakuta[1], Suguru N. Kudoh[1] （1. Kwansei Gakuin University）

5:40 PM - 6:00 PM

[2B4-03] Structural Description and Behavior Analysis of Neural Substrate Subserving Saccadic Eye Movement with Qualitative Neuron Model

○Yoshimasa Tawatsuji[1], Tatsunori Matsui[1] （1. Waseda University）

6:00 PM - 6:20 PM

[2B4-04] Search Space Reduction in Automatically Solving Elementary Number Theory Problems in National Center Test

○Shinya Inuzuka[1], Takuya Matsuzaki[1], Satoshi Sato[1] （1. Nagoya University）

6:20 PM - 6:40 PM

[2B4-05] A Method to Estimate the Original Form of Kaomoji Using Neural Network with Character Embedding Function

○Rei Okumura[1], Noriyuki Okumura[1] （1. National Institute of Technology, Akashi College）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

剛体の集合としての身体：等長変換下の疎な点の対応づけ

Human body as a set of rigid units: finding correspondence between points of two sparse point-sets under distance-preserving transformation

鳥居拓馬 *1　日高昇平 *1
Takuma Torii　Shohei Hidaka

*1北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology

Motion capture systems record a human body as a point cloud. Given two point clouds, how can we identify human body postures at different time frames? An idea for the question is identifying two sets of points by finding a correspondence between points of them. In this paper, we briefly review key problems and algortihms for finding a point-to-point correspondence of two sparse point-sets, especially each (loosely) representing a set of rigid units, under a distance-preserving transformation. Our review is originally motivated by the common problems, involved in using motion capure systems, of missing markers, label correction, and other pre/post-processings. By regarding human body as a set of rigid units, each units can be represented by a set of points that can move together (not exact but) almost independently of other units, in our short review, we pointed out that three key problems being intertwined (chicken-and-egg) make the original problem difficult. We also showed our apporach to two of the intertwined problems based on the invariance under distance-preserving transformations.

1. 身体運動の計測技術

近年では生体データの計測技術も発展し続けており，身体の実空間内の軌跡や，生理指標など，さまざまな生体データを高い時空間解像度でしかも複数センサ同時に入手できるようになってきた．これにより，人間が自らの身体を制御する技術（技能）に関しても，身体の運動を計測する技術の開発とともに，主観的な言葉で語る研究に加えて，運動の計測データとその分析結果で“語る”研究が進められている．

身体運動の生理データ・計測データにはいくつかの種類があるが，身体運動を研究するうえでまず着目するのは３次元実空間内での身体部位の軌道（軌跡）だろう．この実空間内の軌道は近年では光学式３次元モーションキャプチャを用いることで，高い時空間解像度で計測できるようになっている（ある製品は毎秒 240 コマ，有効数字１ミリ以下を保証とある）．光学式モーションキャプチャでは，被写体に赤外光を反射するマーカー（直径１センチ大の球状）を複数取り付け，そのマーカーを複数台の赤外線カメラで撮影し，複数のカメラの位置関係からマーカーの３次元座標を推定する．人体を被写体とする場合は，40 個程度のマーカーを身体表面に固定させ，そのマーカーの位置座標の集まりとして，人体を少数の点の疎な集まりとして記録する．通常，人体の骨格に基づき，人体を点と線で表現できるような場所にマーカーを配置する（例えば [Vicon Motion Systems 16]）．力を加えても変形しない物体を剛体というが，骨のような剛体的な単位に基づくことで，少数のマーカーで人体全身を表現できる *1．このように，光学式モーションキャプチャは被写体（人体）を複数の剛体の集合（複数の疎な点の集まりの集合）として近似的に表すようなマーカー配置を用いる．光学式の長所はその時空間解像度にある．一方で，光学式ならではの欠点もある．

原理的な欠点は，赤外線カメラの死角に入ったマーカーの位置は特定できず，欠損してしまう点である．赤外線カメラの死角は障害物などがカメラとマーカーの間にあるときに生じる（オクルージョン）．被写体たる身体そのものも障害物になりえるため，たとえば前屈を行うと腰や胸に付けたマーカーを見失うこと（欠損）がよくある．動作の種類によっては１秒（240 コマ）以上そのマーカーを見失ったままとなる．一度見失ったマーカーが数秒後に（前屈から起立に戻って）再度カメラに捕捉されたとき，「いったいそのマーカーは数秒前のどのマーカーだったのか」という問題が生じる．この種の問題は対応問題（correspondence problem）[Marr 82] と呼ばれる．実際には，対応問題（マーカーの同一性問題）は欠損値の有無に関わらずいつでも生じているが，微細な時間間隔などの仮定の下で緩和されてみえる．この問題を解決し損なうと，研究者の手にする時系列データ上では実際とは異なる２つのマーカーが同じ１つのマーカーだと混同して扱われてしまう．このようなデータは扱いづらく，人力で修正しない限り，少なくともその２つのマーカーは当然ながら分析に使えない．

記録されるデータの性質の観点から大きく分けて，被写体の剛体的な単位あたりで記録される点の数の少ない「疎」なデータと，記録される点の数の多い「密」なデータがある．光学式モーションキャプチャは被写体の３次元的な骨格を少数の点の疎な集まりとして記録する．他方，３Ｄスキャナは被写体にレーザーを照射し，被写体の３次元的な形状を膨大な数の点の密な集まりとして記録する．３Ｄスキャナでは撮影の度に被写体の表面上の適当な点が記録されるため，厳密に同じ点が別の時点の点の集合に含まれるとは限らない．運動の種類によるが，身体運動の計測に関しては，光学式モーションキャプチャのような身体表面上の決まった場所のマーカーを追跡できるほうが都合よいと思われる．

そこで，本稿で扱う問題は，疎な点の集まり（点集合）が２つ与えられたとき，２つの点集合の間で疎な点の対応づけを求めることで，それらの点集合が同一の被写体すなわち剛体の集まり（剛体集合）を表すものか否かを同定する問題である．人間がデータ（点集合や動画）を見れば，比較的容易にマーカー

連絡先: 北陸先端科学技術大学院大学
石川県能美市旭台 1-1
E-mail: {tak.torii,shhidaka}@jaist.ac.jp

*1 例えば，上腕を骨１本とみなすとき，端点を表す最低２個のマーカーで足りる．各剛体あたりを少数のマーカーで表すことは，マーカーの混同を避けるなどの利点もある．しかし，各剛体を表すマーカーが少なく，冗長性が少ないほど，いずれかのマーカーを記録し損ねたときにその剛体の欠損点を補完するのは難しくなる．

の正しい対応づけや剛体的なまとまりを検出できる（たとえば，バイオロジカルモーション）．しかし，これを計算機に全自動で行わせるのは実はそれほど容易ではない．とくに，生物個体が厳密に同じ身体をもたないことから，大規模な人型モデルのデータベースを検索するアプローチは根本的な解決にはならない．むしろ，人間が素朴に対応問題や同定問題を解決できるのはなぜかを問うことから，これらの問題へアプローチできないだろうか．本稿では，３次元実空間のマーカー（点）の対応問題，および剛体集合の同定問題を話題の中心に据え，その問題の構造と著者らの取り組みを紹介したい．

2. 剛体集合の同定問題

光学式モーションキャプチャを例にとって，被験者の身体に n 個のマーカーを付けてその位置を計測したとする．小さい誤差で欠損なく計測できれば，各時点で，n 個のマーカーに対応した３次元空間内の点（マーカーの位置）の集合 $\mathcal{A} = \{(x_i, y_i, z_i) : i = 1, \ldots, n\} \subset \mathbb{R}^3$ がえらえる．このとき，ある時点の各マーカーは別の時点のどのマーカーなのかを求める対応問題を解きたい．いま２つの任意の時点の点集合 $\mathcal{A}, \mathcal{B}$ が与えられたとする．点集合 $\mathcal{A}$ の要素 $\boldsymbol{a}_i \in \mathcal{A}$ それぞれに対して重複なく１つの要素 $\boldsymbol{b}_j \in \mathcal{B}$ を割り当てる関数 $m(i) = j$ を対応づけという．可能な対応づけは全部で $n!$ 通りある．$n = 40$ でも約 8.16×10^{47} 通りあり，すべてを枚挙するのは現実的ではない．そこで，通常は被写体（被験者）の性質やユークリッド空間の性質，さらに計測機器の性質を活かし，効率的に対応づけを求める手法が開発されている．

光学式モーションキャプチャなど複数マーカーの位置（３次元座標）の集合の時系列に対して，ある時点のマーカーは次の時点のどのマーカーなのかを特定する問題はとくに追跡と呼ばれる．追跡では通常，２つの点集合 $\mathcal{A}, \mathcal{B}$ をそれぞれ記録した時間の間隔が十分に短いことが仮定される（つまり，カメラの連写速度が十分早い）．微小時間ではマーカーの移動距離は十分小さいと想定できるので，$\boldsymbol{a}_i \in \mathcal{A}$ と $\boldsymbol{b}_j \in \mathcal{B}$ の間のユークリッド距離を $d_{i,j}$ と表すと，対応づけに含まれるペアの移動距離の総和，つまりコスト関数 $\gamma(m) = \sum_{i=1}^{n} d_{i,j=m(i)}$ を最小にするような対応づけ m を選ぶという方法が考えられる．これは $n \times 3$ 行列 $A = (\boldsymbol{a}_i)$ および $B = (\boldsymbol{b}_j)$ に対して，最小二乗誤差の意味で $PA \approx B$ となるような置換 $P \in \{0, 1\}^{n \times n}$（ただし各行各列の総和がすべて 1）を求める問題といえる．この方法は最近傍距離（一対一対応を保証しない）や線形和割当問題（一対一対応を保証する）を解くアルゴリズムで実装できる．しかし，この方法は「微小時間ではマーカーの移動距離は十分小さい」という仮定に依拠しており，２つの点集合を記録した時間の間隔が長い場合にはうまくいかない．

別の問題で，複数の３Ｄスキャナの写真（３Ｄ写真）の間で被写体の同一性を判定する問題は形状認識と呼ばれる [Kaick 11]．もっとも単純な問題では被写体はひとつの剛体（ここでは単一剛体と呼ぶ）である．単一剛体同士の間の形状認識は，もし２つの単一剛体 $\mathcal{A}, \mathcal{B}$ が同じ形状なら，２つの単一剛体の間には点集合をうまく重ねるような回転（直交変換）R および並行移動 T が存在するという考えに基づく．これは $n \times 3$ 行列 $A = (\boldsymbol{a}_i)$ および $B = (\boldsymbol{b}_i)$ に対して，$AR + T = B$ となるような回転 $R \in \mathbb{R}^{3\times3}$ と移動 $T \in \mathbb{R}^{3\times1}$ を求める問題といえる．この回転と移動を求める問題は，前述の対応づけを求める問題と相互依存関係にある．すなわち，もし回転と移動が既知ならば対応づけは明らかであり，反対に，もし対応づけが既知ならば回転と移動は明らかである．しかし，どちらも未知なのでどちらも明らかではない．いわゆる，ニワトリとタマゴの関係である．計測データがノイズを含むなど実践的な状況に対して，最小二乗誤差の意味で最適な回転と移動を求めるアルゴリズムがある [Kabsch 76, Horn 87]．この最適回転と最近傍距離を反復的に求めて，単一剛体の点集合の間でペア毎の距離の合計を最小化する回転と移動（副産物として対応づけ）を求めるアルゴリズムがある [Basl 92]．しかし，この方法は「被写体は単一剛体である」という仮定に依拠しており，軟体や複数の剛体の集まりの場合には局所解に陥りやすい．とくに，人体などの被写体は複数の剛体の集まりとみなせるので，身体運動のデータへの直接的な応用は難しい．

そこで，３つ目の問題は複数の剛体（剛体集合）を表す点集合の間で点の対応づけを求める方法である．剛体集合は複数の剛体の集まりで，それぞれの剛体を表す点集合 $\mathcal{A}_i$ の和集合が全体の点集合 $\mathcal{A} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_k$ となる．異なる部分集合は共通する点をもってもよい．ある剛体を表す点集合の部分集合 $\mathcal{A}'_i \subset \mathcal{A}_i$ もまた剛体を表すことから，ここでの問題はより正確にはできるだけ少ない k で（できるだけ大きな $\mathcal{A}_i$ で）全体の点集合 $\mathcal{A}$ を被覆するような部分集合 $\mathcal{A}_1, \ldots, \mathcal{A}_k$ を求める問題になる．たとえば，剛体集合としての人体の腕を { 肩,肘,手 } の３点集合で表したとすると，上腕 { 肩,肘 } と下腕 { 肘,手 } という２つの剛体の和集合で表現できる（肘はどちらの点集合にも含まれている）．３点集合の場合，可能な和集合表現は幾通りもあるなかで，人体の骨格を知る私たちにすればこの和集合表現はもっとも「意味のある」ものだと思われる．この部分集合を求める問題は，先述の点の対応づけを求める問題，さらに回転と移動を求める問題とも相互依存関係にある．ある点の部分集合が剛体を表すとわかるには点の対応づけや回転と移動がわからなければならない．そして，点の対応づけや回転と移動を求めるには，剛体的な部分集合がわからなければならない．計算機による剛体集合の形状認識は近年活発に研究されている．比較的新しい研究では３Ｄスキャナのデータような密な点集合に対して，点の重複ありの逐次的クラスタリング（各クラスタの点集合を剛体とみなす）を単一剛体の反復的アルゴリズム [Basl 92] と組み合わせることで，置換 P，回転 R と移動 T に加えて，部分剛体 $\{\mathcal{A}_i\}_{i=1,\ldots,k}$ を反復的に求める試みがある [Huang 08]．結果的に，被写体は複数の単一剛体の継ぎ接ぎとして近似される．この方法は「計測データは密な点集合を与える」場合を仮定することでクラスタリングの初期配置に最近傍距離を用いる．しかし，単一剛体の場合と同じく，光学式モーションキャプチャで記録されるような疎な点集合の場合にはこの仮定は使えない．

実用的には，上記３つの問題に加えて，欠損点（記録されなかったマーカーの位置）を推定する必要がある．欠損点の補完では「ある時点のどの点が別の時点で欠落したのか」あるいは「ある時点のどの欠損点が別の時点で出現したのか」といった問題を含む．一般には，それぞれの時点がそれぞれ異なる欠損点をもちうる．さらに，場合によっては，ある点（マーカー）とその近傍を含む点集団は１秒以上（240 コマ以上）欠損したままとなりえるなど，「微小時間」や「密な近傍」といった仮定は必ずしも有効ではない．たとえば，１つ欠損点があると，その点の近傍点の最近傍点が変わり，したがって点集合間の２点間の距離に基づく対応づけは誤る可能性が高くなる．

ここまで，疎な点集合で表された，剛体集合の同定問題を解きたいという動機から，疎あるいは密な点集合の間の点の対応づけを求める既存研究を見てきた．本稿では，剛体集合の同定問題に伴う３つの問題（対応づけ，回転と移動，部分集合）を述べた．これら３つの問題は，一部が完全ならば１つは解ける

のに，どれも完全でないのでどれも解けない，という相互依存関係（ニワトリとタマゴの関係）にあると著者らは分析する．ここに欠損点の問題が関わってくるが，上記３つの問題が解決できれば，その解決の度合に応じた曖昧さで，欠損点の補完も可能だと考えられる．実際に，人間に被写体を記録した点集合の動画を見せると，欠損点がどの辺りにくるかをある程度予測できるだろう．もちろん，欠損値が多い場合には人間でもできないほど難しいだろう．この３つの問題の解決はある意味ではヒトの認知の仕組みと密接に関係していると思われる．

3. 剛体の不変量に基づくアプローチ

著者らは，疎な点集合で表された剛体集合の同定問題に向けて，被写体が複数の剛体で近似できると想定し，剛体の不変量に基づくアプローチを進めている．本稿の剛体の定義は，ユークリッド空間におけるある点の集合が（単一の）剛体であるとは，その点集合がある種の変換に対して各点間の距離を保つことをいう．この種の変換を等長変換という．等長変換には回転，移動，鏡映がある．

ここでは，モーションキャプチャのように被写体に複数のマーカーを付け，そのマーカーの位置（相対位置）を集めた点集合で被写体を表す．被写体そのものかあるいはカメラ（視点）を動かすと，それに応じて点集合も変化する．記録された２つの時点の点集合を $\mathcal{A}, \mathcal{B}$ とし，その行列表現を $A = (\boldsymbol{a}_i), B = (\boldsymbol{b}_j)$ とする．この段階では i と j は（利用者が期待するようには）対応づいていない．もし $\mathcal{A}, \mathcal{B}$ がどちらも単一剛体の同一被写体で，欠損値なく，小さい計測誤差ならば，何らかの置換 P と回転 R と移動 T とによって，$PAR + T \approx B$ がなりたつ．ここで，通常は，置換は計測に伴う点の欠損やデータ保存上の問題で生じ，回転と移動は剛体そのものかあるいはカメラの動きを反映する．前述のように，この近似式は，もし回転と移動がわかれば置換がわかる，また，もし置換がわかれば回転と移動がわかる，という相互依存関係にあり，これらを同時に求める方法が模索されてきた [Kaick 11]（たとえば，[Basl 92, Huang 08]）．

著者らは，単一剛体を考えた場合に，２つの点集合（剛体）がどのような等長変換（回転と移動）で対応づけできるかを探す，変換に基づくアプローチに対し，等長変換で保存される距離行列すなわち，等長変換の不変量に基づくアプローチを検討している．この不変量に基づくアプローチでは等長変換（回転と移動）を陽に求めず，対応づけ（置換）のみを直接的に求めることで,「回転と移動」と「置換」の相互依存関係を克服しようとする．今回著者らの提案するアルゴリズムは以下である．

ある点集合の行列表現 A とそれを等長変換してえられた $B = PAR + T$ があるとする．いま置換 P，回転 R，移動 T を未知とし，これらを推定したい．$A = (\boldsymbol{a}_i)$ の要素 $\boldsymbol{a}_i, \boldsymbol{a}_j$ のユークリッド距離を d^A_{ij} と記す．距離 $d^A_{i1}, \ldots, d^A_{in}$ を要素とするベクトル $\boldsymbol{d}^A_i = (d^A_{i1}, \ldots, d^A_{in})^\top$ は（未知なる）等長変換（回転と移動）に対して不変である．したがって，A が単一剛体ならば，適切な置換 P によって，$\boldsymbol{d}^A_i = P\,\boldsymbol{d}^B_j$ なる要素 $\boldsymbol{b}_j$ が１つ以上ある．そこで，すべての i について $\boldsymbol{d}^A_i = P\,\boldsymbol{d}^B_j$ をみたす一対一の対応づけ，つまり置換 P を上記 $n \times n$ 個の等式すべてに整合的になるように求めれば，等長変換（回転 R と移動 T）を求めることなく置換 P を求めることができる．もし単一剛体ならば，少なくとも１つは上記 $n \times n$ 個の等式すべてに整合的な P がある．たとえば，A の３点が正三角形をなす場合，$B = PAR + T$ をみたす複数の P がありえる．実測データは通常ノイズを含むので，複数の置換 P が求まることは少ない．

この着想の下で置換 P を求めたいが，P は上記 $n \times n$ 個の等式すべてを充たす必要がある．さらに，実データに含まれるノイズに対処するため，上記の厳密な等式に代わり，何らかの誤差の下での近似的に $\boldsymbol{d}^A_i \approx P\,\boldsymbol{d}^B_j$ を求める必要がある．そこで著者らのアルゴリズムでは，一対一の対応づけを求める方法として線形和割当問題を入れ子にして解いた．このアルゴリズムでは，それぞれのペア $(i, j) \in \{1, \ldots, n\}^2$，について線形和割当問題の意味で最適な置換 $\hat{P}_{(ij)}$ とそのときの最小合計コスト $c_{ij} := \|\boldsymbol{d}^A_i - \hat{P}_{(ij)}\,\boldsymbol{d}^B_j\|$ を求める．次に，この最小合計コストから作った $n \times n$ 行列 $\boldsymbol{C} = (c_{ij})$ に対して再び線形和割当問題の意味で最適な置換 P を求める．この置換 P を最終的な対応づけとする．もし単一剛体でノイズなしならば，最小合計コスト $c_{ij} = 0$ となる置換 $\hat{P}_{(ij)}$ およびそのようなペア $(\boldsymbol{a}_i, \boldsymbol{b}_j)$ がある．そして，$\boldsymbol{C}$ は最小合計コストが 0 になるような置換 P を少なくとも１つもつ．著者らは，各線形和割当問題はハンガリアン法 [Kuhn 55] で解いた．著者らのアルゴリズムは計算時間 $O(n^3)$ のハンガリアン法を $n \times n + 1$ 回適用するので，計算時間 $O(n^5)$ である．ただし，入れ子になった $n \times n$ 回の適用は独立なので容易に並列化できる．著者らのように等長変換（不変量）に着想をえたものではないが，より一般的な特徴空間内の２つの点集合を考えたうえで，凹関数計画法あるいは整数２次計画法で定式化し，その問題を近似的に解く計算時間 $O(n^6)$ のアルゴリズムがある [Maciel 03, Berg 05]．彼らのアルゴリズムでは $O(n^4)$ のシンプレクス法を使うので計算時間が大きい．

3.1 数値実験

著者らの不変量に基づくアルゴリズムを単一剛体の場合に数値実験を行った．本稿では，比較対象として既存のアルゴリズム，[Basl 92] および [Leordeanu 05] と比較する．[Basl 92] は，最近傍距離に基づく対応づけと，最小二乗誤差の意味での最適な回転と移動を繰り返し求めるアルゴリズムである．[Leordeanu 05] は，ネットワーク分析に着想をえた方法で，一般的な特徴空間で全データ点 $\mathcal{A} \cup \mathcal{B}$ のペア毎の類似性関係をネットワークで表し，不可能なペアを排除する制約下でそのネットワークの定常遷移確率を求め，点 $\boldsymbol{a}_i \in \mathcal{A}$ から $\boldsymbol{b}_i \in \mathcal{B}$ 間の遷移確率の高いものから，対応づけ (i, j) を見いだす（どんな問題を解いているかは不明であるが，実践的にうまくいくと報告されている）．[Maciel 03, Berg 05] は前述のとおりである．以上これらのアルゴリズムは画像データや密な点集合を想定している．そこで，この数値実験の目的は，本稿の動機である疎な点集合の場合にその成績を評価することにある．著者らの本稿での目的は計算時間よりもまずは正しい対応づけを推定できるかにあるので，以下では正しい対応づけを推定できた割合（正答率）のみを比較する．

この数値実験では $n = 10$ 個の点からなる点集合の行列表現 $A \in [0, 1]^{n \times 3}$ をランダムに生成し，この A にランダムな回転 R^Δ を加え，さらに一様分布に従うノイズ $E \in [-\epsilon, +\epsilon]^{n \times 3}$ を加えたものを $B = A\,R^\Delta + E$ とした．ここで，$\epsilon = 0$ のときノイズなし，$\epsilon > 0$ が大きいほど B は A からランダムにずれる．また，R^Δ は行列 R のべき乗で，$\Delta = 0$ のとき回転なし，$\Delta > 0$ が大きいほど B は A から大きく回転する．以下で図に示された成績はランダムに生成した異なる A, B の 100 ペアに対する推定結果である．

図 1 は回転の大きさに関するパラメータ $\Delta \geq 0$ を変えたときに各アルゴリズムが正しい対応づけ，つまり置換 P を正確に求めた割合を示す．著者らのアルゴリズム（Proposed）は

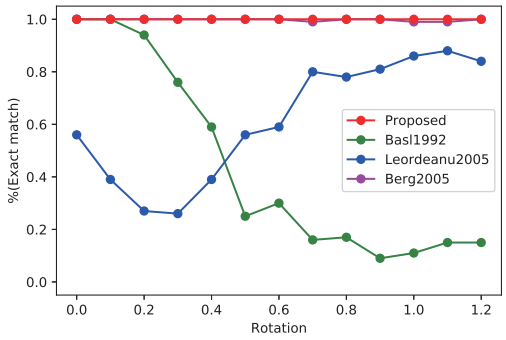


図 1: 回転の大きさ $\Delta \geq 0$ を変えたときの正答率

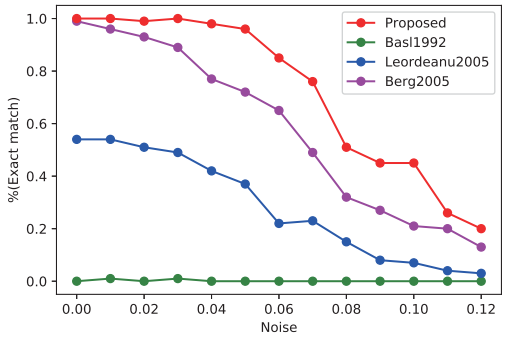


図 2: ノイズ $\epsilon \geq 0$ を変えたときの正答率（$\Delta = 1$）

回転の大きさによらず，常に 100% の割合で正しい対応づけを推定できた．また，図 2 は回転の大きさ $\Delta = 1$ を固定し，ノイズに関するパラメータ $\epsilon \geq 0$ を変えたときの結果を示す．図 2 から，4 つのアルゴリズムの中では著者らのアルゴリズム（Proposed）はもっとも回転とノイズの混合に対し耐性をもつ．とくに，$\epsilon \leq 0.04$ まではほぼ 100% の割合で正しい対応づけを推定できた．

4. 議論

正しい置換が求まれば，相互依存関係が解消され，正しい回転と移動は既存のアルゴリズムで容易に求まる．また，置換が求まれば，対の存在しない点は欠損点だと期待できる．点集合が単一剛体であるとすると，その欠損点は他の記録された点からの相対距離から，あるいは回転と移動によってある程度の精度で補完することができる．ただし，現状では，両方の点集合に複数の欠損点がある場合や剛体集合の場合には誤った点を補うことがあり，補完の正しさを裏付ける基準などを開発していく必要がある．今後は，本稿のアルゴリズムの高速化や，剛体集合を剛体の部分集合群に分ける問題も含めて検討していく．

剛体集合の同定問題に伴う 3 つの相互依存問題の同時的解決は，単に光学式モーションキャプチャで記録したデータを分析するさいの前処理に留まらず，欠損値のあるデータを自動的に補完して分析できるように自動修復したり，さらに，各時点で運動中の身体がいくつの部分剛体とみなせるか，すなわち運動中の身体の分節化と自由度を自動検出するなど，身体運動の特徴づけにも応用できると考えられる．また，一般的な研究では 3 次元実空間の点集合に限らず，任意の特徴空間での対応問題に関心が向けられている（たとえば，2 D写真のピクセルなど）．とくに，形状認識などの応用 [Huang 08, Kaick 11] では異なる身体（たとえば，人と竜）の間での対応づけ（人の手と竜の羽？）を求めれるかなど，模倣や類推にも通じる問題が（応用として）取り組まれている．著者らのアプローチは「剛体」という被写体の性質を使うものだが，模倣や類推など，剛体的な単位のあつまりとみなせそうな対象について，さらなる応用可能性を検討していきたい．

謝辞

本研究は科学研究費補助金 JP16H05860，JP17H06713 の助成を受けて行われた．

参考文献

[Basl 92] Basl, P. J. and McKay, N. D.: A method for registration of 3-D shapes, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, pp. 239–256 (1992)

[Berg 05] Berg, A. C., Berg, T. L., and Malik, J.: Shape matching and object recognition using low distortion correspondences, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 26–33 (2005)

[Horn 87] Horn, B. K.: Closed-form solution of absolute orientation using unit quaternions, *Journal of the Optical Society of America A*, Vol. 4, No. 4, pp. 629–642 (1987)

[Huang 08] Huang, Q.-X., Adams, B., Wicke, M., and Guibas, L. J.: Non-rigid registration under isometric deformations, *Computer Graphics Forum*, Vol. 27, No. 5, pp. 1449–1457 (2008)

[Kabsch 76] Kabsch, W.: A solution for the best rotation to relate two sets of vectors, *Acta Crystallographica*, Vol. A32, pp. 922–923 (1976)

[Kaick 11] Kaick, van O., Zhang, H., Hamarneh, G., and Cohen-Or, D.: A survey on shape correspondence, *Computer Graphics Forum*, Vol. 30, No. 6, pp. 1681–1707 (2011)

[Kuhn 55] Kuhn, H. W.: The Hungarian method for the assignment problem, *Naval Research Logistics*, Vol. 2, No. 1–2, pp. 83–97 (1955)

[Leordeanu 05] Leordeanu, M. and Hebert, M.: A spectral technique for correspondence problems using pairwise constraints, in *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2, pp. 1482–1489 (2005)

[Maciel 03] Maciel, J. and Costeira, J. P.: A global solution to sparse correspondence problems, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, pp. 187–199 (2003)

[Marr 82] Marr, D.: *Vision*, Cambridge: MIT Press (1982)

[Vicon Motion Systems 16] Vicon Motion Systems, : Plug-in Gait Reference Guide, in *Vicon Nexus 2.5 Documentation*, pp. 1–95 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

多層ニューラルネットワークによる 生体神経回路網における神経活動パターン特徴抽出の試み

An attempt of feature extraction of neuronal activity patterns in living neuronal network, by multi-layered artificial neuronal network

作田 尋路　　工藤 卓 CA
Hiromichi Sakuta　Suguru N. Kudoh

関西学院大学 理工学部 人間システム工学科
Human System Interaction, School of Sci and Tech, Kwansei Gakuin University

To elucidate the brain function, it is important to understand dynamics of nonlinear neuronal activity patterns. In this study, we attempted to extract features of activity patterns including evoked responses in the cultured rat neuronal network. Thus, we employed the multi-layered artificial neural network (ml-ANN) as the Deep-Learning method with stacked-autoencoder as the pre-training method to classify the network activity. As the result, activity pattern 2 seconds after the electrical stimulation was barely classified, however, discrimination ability of ml-ANN against the activity pattern of later time domain after the stimulation was not enough to classify the patterns of evoked responses, because of the insufficient amount of learning-data, which is difficult to be gathered in large quantities. This indicates a huge number of pre-learning data is absolutely necessary to improve the discrimination accuracy in order to identify patterns by the Deep-Learning method for large phenomena with "fluctuation" such as neural activity.

1. はじめに

脳の情報処理機序の解明には，生体神経回路網における非線形な神経電気活動パターンを解析し，そのダイナミクスを理解することが重要である．また，このように電気生理学的アプローチで脳の情報処理について研究することは，より高度な脳型情報処理システムを構築する上での基礎となる可能性を秘めている．そこで，近年培養神経回路網における電気活動の時系列パターンを，リカレントニューラルネットワーク（RNN）により抽出する等の研究が行われている [安江 2016]．神経電気活動の特性上，生波形データを単純に入力するだけでは学習がうまくいかないことが予想される．

このような背景から，本研究では脳の機能を本質的に保持した小規模モデルとしてラット海馬分散培養系を用いて，生体神経回路網において電気刺激印加に伴う誘発応答を含む神経電気活動パターンの解析を行った．近年，Deep Learning（DL）が高い性能を発揮している．特に，畳み込みニューラルネットワーク (CNN) であれば畳み込み層とプーリング層を積み重ねて特徴抽出を行う等，動物の視覚情報処理とよく似た処理であるため，画像認識に特に親和性が高い．ただし，DL は大規模なデータで学習することが前提である．本研究では，時空間情報の連続性をうまく保持するよう入力画像データを作成し，積層オートエンコーダを事前学習手法として利用し，特徴抽出を試みた．大量のデータを用いるのが困難である in vitro 生体データに対しても DL 手法が有効であるか検討した．

2. 実験・解析手法

2.1 ラット海馬分散培養系

胚齢 18 日齢（E18）のウィスターラット（Jcl:Wistar，日本クレア）の胎児大脳海馬領域を摘出し，分散培養を行った．底面に微小電極が配置された特殊な培養皿である MED プローブ (α-Med サイエンティフィック) において，10 分間のトリプシン処理により十分に解離分散された神経細胞を中央の電極上に播種した．これを 37 ℃・5 % CO_2 のインキュベータ内で培養し，神経回路網を再構築させた．本研究において行った動物実験は，全て「関西学院大学動物実験管理規定」に則り，動物実験委員会の承認を得た上で行った．

連絡先: 工藤卓，関西学院大学 理工学部 人間システム工学科，神経知能工学研究室（NIE-Lab），兵庫県三田市学園 2 丁目 1 番地，079-565-7244，snkudoh@kwansei.ac.jp

2.2 細胞外電位計測システム（MED64 システム）

生体神経回路網における神経電気活動の計測には，細胞外電位多点計測システム（MED64 システム，α-Med サイエンティフィック）を使用した．本研究で用いた多点電極培養皿である MED プローブには，底面の中央部分に 64 個の平面微小電極が 8 × 8 の格子状に配置され，その外側に 4 個の参照電極が配置されており，これらの電位差を計測することによって，神経電気活動を細胞外電位として 64 点から非侵襲で長期間同時計測することが可能で，神経ネットワークにおける電気活動の時空間的特性を解析することができる．本研究では，神経細胞近傍の細胞外を計測し，これをサンプリング周波数 10 kHz，量子数 12 bit でデジタル化して，コンピュータのハードディスクに保存した．また本システムは，参照電極と 64 個の電極から任意に選んだ電極との間に電圧を加えて，最大 2 点まで細胞に定電流刺激を行うことが可能である．電気刺激には同形の方形波パルスを双極で適用した．計測の際には，培養液を細胞外記録溶液（Rec.Sol.）に液置換した．

2.3 電気刺激に伴う誘発応答を含む神経電気活動の計測

胚齢 18 日，培養日数 29-65 日（E18DIV29-65）の神経回路網において，電気刺激印加による誘発応答を含む電気活動を MED64 システムを用いて 10 分間計測した．計測には，当研究室でプログラミング言語 LabVIEW を用いて独自開発した計測用プログラムである Spike_Recoder（SPR）プログラムを用いた．実験するにあたって，まず MED プローブ上の 8 × 8 の 64 電極のうち 2 点 A，B を選択し，計測開始から 30 s 毎に「電気刺激なし・電極 A に刺激・電極 B に刺激」という 3 種類の刺激パターンを 10 s 間隔毎にランダムに切り替え，個別の神経回路網に対して計測時間が終了するまで電気刺激と電位計測を繰り返した（図 1）．このとき選択した刺激電極番号と実際の刺激順序を同時に記録した．

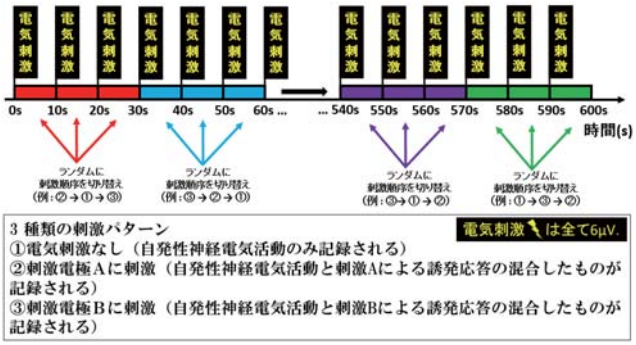


図 1: 電気刺激実験スキーム.

計測した波形データから，当研究室で独自開発した SPike_Counter（SPC）プログラムを用いて，神経電気活動スパイクを検出し，同様に当研究室で独自開発した SPC_analyser プログラムによって，計測時間内におけるタイムスタンプデータ（ms）を抽出した.

2.4 時系列情報と空間的位置情報を含む入力画像データの作成

計測時間分の 10 分間のタイムスタンプ系列データを 40 ms の時間窓幅で区切り，全 64 電極分の細胞外電位スパイク数を逐次求めた．本研究と同じ条件の実験系において，時間窓幅を変えながら全 64 電極分のスパイク数を生体神経回路網による情報表現としてまとめた．1 次元データ集合に対して，X-means 法によるクラスタリングを行った結果，出力されたクラスタ数が最も多くなり，最も多くのパターンが表現された時間窓幅が 30-40 ms であったという報告があるため，時間窓幅を 40 ms とした [Okada 07]．算出した時間窓幅 40 ms におけるスパイク数を 1 ピクセルの輝度とし，これを 2 s 分まとめて画像化したデータを作成した．本研究においては計測時間を 10 分としたため，1 画像として 3200（64 × 50）次元のデータ 300 個からなる入力データセットを作成したことになる．1 画像を 2 s 毎のデータ集合とした理由は，本研究と同じ条件で実験した際に，直前の刺激入力の履歴が持続する時間，即ち，神経細胞の内部状態の持続時間が約 2 s であったと報告されているからである [Ito 16]．以上の操作により，作成した入力画像の縦軸方向には神経活動パターンの時間的関係性，横軸方向に空間的関係性を保持した入力画像データを作成した.

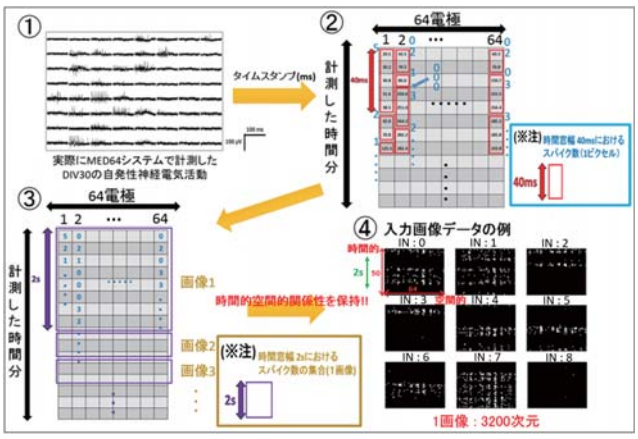


図 2: 神経活動の時空間パターンの特徴を表現した画像データの作成法の概要.

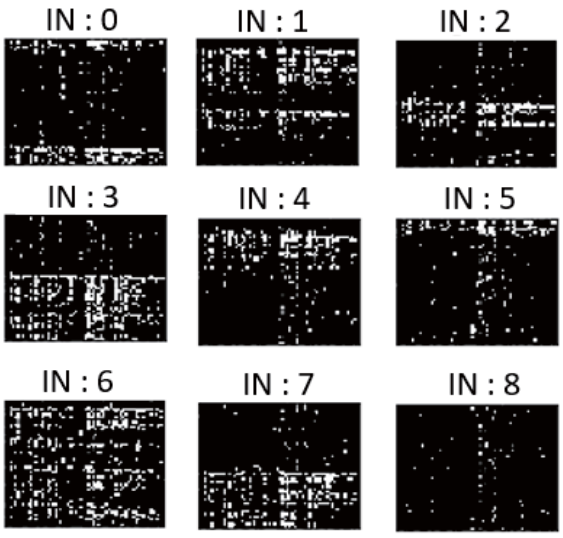


図 3: 自発性神経活動パターンの画像化データの例（E18DIV51）.

2.5 解析手法

2.5.1 積層オートエンコーダによる活動パターン識別

本研究において用いた積層オートエンコーダ（Stacked Autoencoder，SAE）は，オートエンコーダ（AE）を複数積み重ね多層化した Deep Learing の学習モデルの一種である．本研究では 4 層からなる多層ニューラルネットワーク（mL-ANN）を構成し，SAE によるプレトレーニング（事前学習）とファインチューニング（微調整）を行った．まず入力データ集合に対して，教師なし学習として SAE のプレトレーニングを行った（図 4）．一般的に DL を行う際は，入力データとして大量のデータを与え，正則化項を目的関数に踏まえ，更にドロップアウトの処理を行う等，過学習が起こらないように様々な工夫を行うことで汎可性を高める．しかし，本研究においては神経電気活動の実験データ毎に特異的な特徴量を抽出し，強く特徴を表現することを目的として，むしろ過学習するように学習させた．プレトレーニングにおいては，損失関数には最小二乗誤差，活性化関数には sigmoid 関数，勾配降下法の最適化アルゴリズムには Adam を用いた．これは本研究の前段階として AE の有効性を確認した際に用いたものと一致させた.

続いて，多層パーセプトロンの学習手法である Backpropagation（BP）により，刺激の有無，刺激電極の任意の違いにより神経電気活動パターンが識別されるよう，ファインチューニングを行った（図 5）．プレトレーニングによって得られた各層の重みやバイアス項を初期値としたネットワークに対して，出力層として 3 次元のノードを追加し，更に正解ラベル（分類ラベル）を与えて教師あり学習を行い，ネットワークの最適化を行った．また出力層の活性化関数には softmax 関数を用いることで，最上層である 3 つのノード値をそれぞれ，3 種類の刺激パターンのいずれかという確率として扱った．またテストデータに対して識別を行った結果の正答率を，出力層における 3 つのノードのうち最も確率の高かったものと正解ラベルを比較して一致した確率と定め，活動パターンの識別精度の指標とした．ファインチューニングにおいては，損失関数には交差エントロピー，隠れ層の活性化関数には ReLU 関数，勾配降下法の最適化アルゴリズムには Adam を用いた.

学習率の大きさと識別精度との関係性と，重み及びバイアス項の更新を固定する層の違いによる活動パターンの識別精度

を比較した．また，全部で 300 個ある入力データを全て，プレトレーニングに用いる学習データ及びファインチューニングに用いるテストデータとして用いた．これは，本研究の目的は未知データに対する識別より，識別に有効な特徴量を抽出することであるため過学習が起こる方が都合が良いと考えたためである．

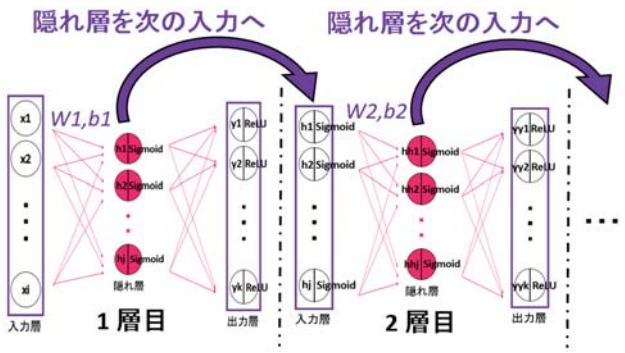


図 4: SAE におけるプレトレーニングの概要.

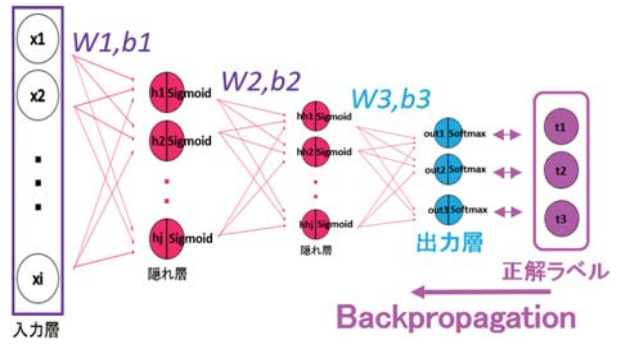


図 5: mL-ANN におけるファインチューニングの概要.

以上の解析を行うプログラムは全て，Tensorflow ライブラリを用いて Python3 で独自に作成した.

2.5.2 刺激直後 2 s の入力画像データに対するパターン識別

「電気刺激なし・電極 A に刺激・電極 B に刺激」という 3 つの刺激パターンそれぞれの刺激直後 2 s の神経電気活動を入力データとして解析した．これにより，電気刺激の影響が最も強く表われているであろう神経電気活動データのみを入力データとして用いたことになる．刺激直後のより誘発応答に着目することになり，活動パターンが精度良く識別されると予想した．作成した画像データを入力として，事前学習として SAE による特徴抽出を行った後，多層パーセプトロンによるパターン識別を行った.

また，この刺激直後 2 s のデータを入力として，SAE におけるエンコードを畳み込み層とプーリング層によって行い，デコーダとしてアップサンプリングと逆畳み込みの処理をする積層畳み込みオートエンコーダ（Stacked Convolutional Autoencoder, SCAE）による解析も行った．入力データを画像データとして作成したため，データに空間的な特徴が含まれる場合に，より精度良く識別できると考えた．畳み込み層においては 3 × 3 のフィルターにかけて畳み込み処理を行い，プーリング層においては，画像化データの縦軸・横軸ともに画像の解像度が半分になるよう，特徴マップを作成した．この得られた特徴マップに対して逆畳み込みの処理を行い，入力データを再構成した.

3. 実験結果・考察

前処理として入力データに対して標準化を行い，スパース性に対して精度良く適用される線形主成分分析（線形 PCA）により，累積寄与率がほぼ 1 になるように予め次元削減した．これは本研究の前段階として，神経電気活動データに対する AE の有効性を確認した際に，予め線形 PCA によって入力データの冗長性を省くことで，制限を設けることで AE と等価となる線形 PCA と比しても精度良く次元削減されたためである.

電気刺激印加による誘発応答を含む 10 s の神経活動から生成した画像データを入力として，プレトレーニング時の学習率 0.1，ファインチューニング時の学習率 0.01 とした際に，パラメータの更新が固定される層の違いによって活動パターンの識別精度を比較した結果,「第 1 層・最上層の重みを固定」した場合に 36.0 ± 0.61 %（平均±標準誤差）,「第 1 層の重みだけ固定」した場合に 33.9 ± 0.24 %（平均±標準誤差）,「全ての層の重みが可変」である場合に 33.8 ± 0.11 %（平均±標準誤差）,「最上層の重み以外固定」した場合に 34.9 ± 0.56 %（平均±標準誤差）となり,「第 1 層・最上層の重みを固定」,「第 1 層の重みだけ固定」,「全ての層の重みが可変」,「最上層の重み以外固定」の順に僅かながら識別精度が高く,「第 1 層・最上層の重みを固定」と「第 1 層の重みだけ固定」,「第 1 層・最上層の重みを固定」と「全ての層の重みが可変」の間に有意差があった（$p<0.01$）．いずれも識別精度の平均値が 40 %に満たず識別できなかったが，ファインチューニング時に少なくとも第 1 層の重み更新を行わないように設定することで精度が上がる可能性が示唆された（図 6）.

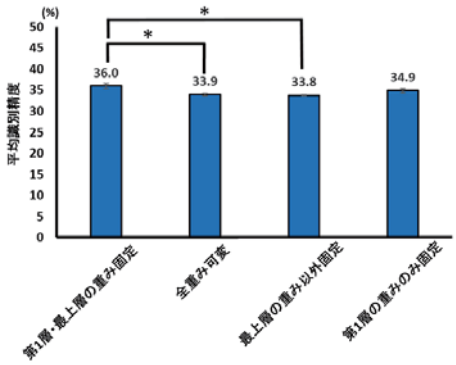


図 6: 重み固定される層の違いによる比較,（N = 25, E18DIV29-65，ウィルコクソンの符号付順位検定：*$p<0.01$）.

プレトレーニング時及びファインチューニング時の学習率設定の違いによる識別精度を比較した結果，プレトレーニング時の学習率 0.1，ファインチューニング時の学習率 0.01 とした際には 36.0 ± 0.61 %（平均±標準誤差），プレトレーニング時の学習率 0.001，ファインチューニング時の学習率 0.1 とした際には 33.7 ± 0.22 %（平均±標準誤差）となり，プレトレーニング時の学習率が大きい方が識別精度が若干高くなり，有意差が見られた（$p<0.01$，図 7）．しかし，プレトレーニング時及びファインチューニング時の学習率を反転させても，依然として識別精度は 40 %に満たなかった．この原因としては，電気刺激を印加したことによる誘発応答を含んだ神経電気活動パターンは，刺激を印加した直後は活動パターンの特異性が強いが，次の刺激を印加するまでの後半部分のデータにおいては前の刺激に対する履歴より自発的活動の影響の方が強く，1 つの刺激に対するパターンのゆらぎが大きいことが考えられる.

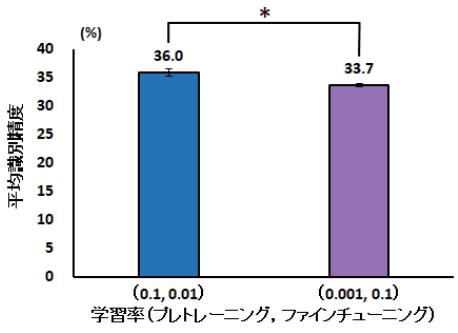


図 7: 学習率の違いによる識別精度の比較,(エラーバーは標準誤差，N = 25, E18DIV29-65，ウィルコクソンの符号付順位検定：*p<0.01)．

以上のように，電気刺激を印加してから 10 s 間の実験データを入力として用いた場合には，神経の活動パターンを精度良く識別することは難しかった．これは刺激印加から時間が経つにつれて，刺激に対する誘発応答を含む神経電気活動パターンがこの手法で識別され得る形で発現していないということで，履歴性をうまく抽出できなかった可能性が考えられる．

以上の結果から，次の刺激を印加するまでの後半部分のデータに前の刺激に対する履歴の影響が小さく，本手法では検出できない可能性が高いため，刺激を印加した直後の 2 s 間のデータ以外を省いた入力画像データセットを作成して解析した．全データを用いたときと同様，まず標準化した上で，線形 PCA による次元削減を累積寄与率がほぼ 1 になるように行った．この結果，ファインチューニング後の活動パターンの識別精度が 63.1 ± 2.24 %（平均値±標準誤差）となり，全データを用いたときと比較して約 2 倍精度良く識別でき，有意差があった (p<0.01，図 8)．培養サンプルによっては 80-90 %の精度で識別可能な神経回路網も存在した．これは入力として用いる刺激後のデータを刺激直後のパターンに限定することで，入力データの次元が少なくなったことにより，データのばらつきが小さくなり識別されやすくなったこと，刺激による直接的な誘発応答パターンのみ識別対象としたため再現性が高くなったことが考えられる．

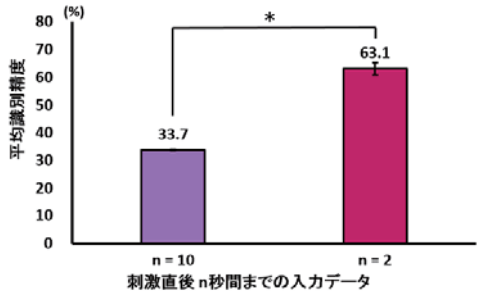


図 8: 入力データの違いによる活動パターン識別精度の比較,(エラーバーは標準誤差，N = 25, E18DIV29-65，ウィルコクソンの符号付順位検定：*p<0.01)．

さらに，エンコード部分で畳み込み処理を行う SCAE を用いることで，入力データの局所的な特徴により着目し特徴抽出を行うことができると考え，プレトレーニングに SCAE を用いて識別精度を求めた．しかし，各識別精度は，プレトレーニングを SAE を用いた場合に 63.1 ± 2.24 %（平均値±標準誤差），SCAE を用いた場合には 39.7 ± 1.94 %（平均値±標準誤差）であった（図 9）．ウィルコクソンの符号付順位検定を行った結果，有意差が見られた（p<0.01）．SCAE によって識別精度が得られなかった原因としては，元々の入力データの次元数が少ないところに畳み込み処理をすることにより識別の空間解像度が低くなりすぎたことが考えられる．

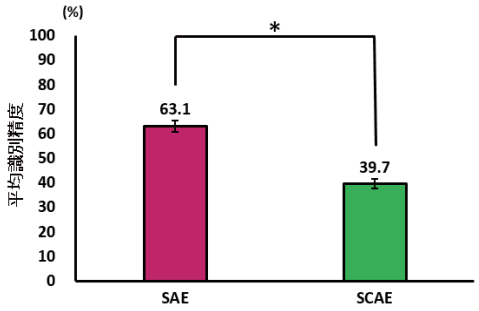


図 9: プレトレーニングを SAE と SCAE による活動パターン識別精度の比較,(エラーバーは標準誤差，N = 25，E18DIV29-65，ウィルコクソンの符号付順位検定：*p<0.01)．

4. 結論

本研究では，ラット海馬から調製した生体神経回路網に対して，電気刺激印加することで誘発される応答活動電位を含む神経電気活動を DL 手法で解析した．計測した生波形データから時空間的関係性が保持されるように入力画像データを作成し，事前学習として SAE を用いてパラメータを決定し，多層パーセプトロンとして BP によって神経電気活動のパターン識別を行った．電気刺激印加後 10 s 間の電気活動から作成したデータを入力として用いた場合，神経電気活動パターンを識別することは難しかった．一方，刺激直後 2 s 間分の電気活動データのみを入力とした場合は識別精度が上がり，刺激直後のパターンに限定することである程度精度が向上することが示唆された．以上のことから，一定の分類基準に対して「ゆらぎ」の大きい現象について DL の手法によってパターン識別を行い，識別精度を向上させるには，事前学習データ及びファインチューニングを行う入力データとして，相当膨大な数の入力データを確保する必要があることが示唆された．

参考文献

[安江 2016] 安江秀太，安田秀策，矢田祐一郎，神崎亮平，高橋宏知,「リザバー計算に基づく培養神経回路の状態制御」，電気学会研究会資料 医用・生体工学研究会 MBE-16-022 048: pp.17-22，2016.

[Okada 07] T. Okada, W. Minoshima, H. Ito, S. N. Kudoh, "Analysis of transition of spontaneous electrical activity patterns in cultured neuronal networks", Proceedings of Life Engineering Symposium 2016 (LE2016), pp. 102-105, 2016.

[Ito 16] Hidekatsu Ito, Wataru Minoshima, Suguru N. Kudoh, Relationships between stimulus interval and changes of firing properties, Proc. Soft Computing and Intelligent Systems (SCIS) 2016 and 17th International Symposium on Advanced Intelligent Systems(ISIS), pp. 722-723.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

定性ニューロンを用いた衝動性眼球運動を支える神経基盤の構造記述と挙動解析

Structural Description and Behavior Analysis of Neural Substrate Subserving Saccadic Eye Movement with Qualitative Neuron Model

田和辻 可昌[*1] 松居 辰則[*2]
TAWATSUJI Yoshimasa MATSUI Tatsunori

[*1]早稲田大学 大学院人間科学研究科
Graduate School of Human Sciences, Waseda University

[*2]早稲田大学 人間科学学術院
Faculty of Human Sciences, Waseda University

Understanding the relationship between brain function and brain structure is important issue to understand meachanism of human cognitive activities. Qualitative reasoning can provide one of effective approach to tackle with this issue. In this paper, we proposed qualitative neuron model with which we described the neural substrate subserving saccadic eye movement. As a results, the simulation showed the burst discharges in the pontine reticular formation that should play an important role in controling the saccadic eye movement. Thus it was confirmed that our qualitative neuron model has high expression power to the phenomenon related to the saccadic eye movement.

1. はじめに

1.1 背景

脳の情報処理に伴う認知機能を理解する上で，脳の構造と脳機能の間の接合は重要な課題である [Sporns 04][Pessoa 14]．これまで，脳機能を理解する試みとしては，脳部位の病変による機能不全などの臨床的知見や，Activation Studies による認知神経科学的知見を基に検討されてきた [Pastor 97]．例えば，Klüver-Bucy Syndrome やてんかん患者 HM の例から扁桃体や海馬がそれぞれ恐怖，記憶に関係することが示唆されたことは有名である．また，fMRI などの計測機器を用いた実験では，実験条件である認知タスク中の脳部位の活動と対照条件の脳部位の活動を統計的に解析することで，認知機能と脳部位の関係を抽出することを試みている．

これらの実験アプローチは脳の部位と機能の関連を知る上で重要であるが，各脳部位が認知機能における「どの機能をどのように実現しているのか」に関しては説明を与えないという課題がある．たとえば，何か新しいものに目を向ける際には，物体の検出，新奇性の抽出，眼球運動の制御などの様々な機能が特定の機序によって強調的に作用し，これを実現していると考えられる．このためには，機能（i.e. 新しいものに目を向ける）の部分機能（i.e. 物体検出，新奇性の抽出，眼球運動の制御）を達成する神経基盤の挙動と，それらの大域的な相互作用を考慮に入れた神経基盤全体の挙動解析および因果解析が必要不可欠である．しかし，系の規模が巨大になればなるほど，定量モデルでは因果解析を行うことが困難になる．これは，系に存在する変数量の増大によって，系全体に存在する定量的な微分方程式の解析が困難になることに加え，定量モデルでは解の生成はできても，解にいたる変数同士の変化の影響に関する因果は陽に示されない [西田 93] という本質的な問題に由来する．

1.2 目的とアプローチ

本研究では，大域的な脳の情報伝播に関する挙動解析および因果解析を行う方法として定性推論に着目する．本報告では，このような定性推論に基づく神経情報伝播の挙動解析・因果解析の確立を志向し，まず脳部位における機能の定性的記述に伴う解の生成可能性について検証する．

まず，2 節ではこれまでの脳の情報伝播に関する定性モデルに関する既往研究を俯瞰する．次に 3 節では，定性推論において重要となる，モデル化の対象となる構造を明確にする．本研究では，脳の部位（e.g. 扁桃体，海馬，...）を一つの構造単位として見なし，これらは代表ニューロンによってお互いが結合されているという立場に立つ．この上で，代表ニューロンの挙動を表現する定性ニューロンモデルを構築する．最後に 4 節では，構築した定性ニューロンモデルを用いて，衝動性眼球運動（SEM:Saccadic Eye Movement）制御に関わる神経基盤を例に挙げて，解剖学的知見に基づき各脳部位を結合し，本定性ニューロンモデルを用いて衝動性眼球運動が表現可能であるかをシミュレーションによって明らかにする．

本シミュレーションの結果，SEM の制御に関わるとされる橋網様体の定性ニューロンの出力において，バースト発火に関する挙動が確認され，解剖学的知見に整合することが示唆された．このことから，定性モデルは SEM 制御に関わる現象に対して記述表現力を有していることが確認された．

2. 関連研究

脳の情報処理を定性的に捉える試みはこれまで多くのモデル研究で行われてきた．ニューロンの挙動を記述するにあたり，式 (1) の integrated-and-fire モデルは入力電流 $I(t)$ に応じて発火周波数が上昇するという性質を単純な形で記述 [銅谷 12] できる点で優れている．

$$\begin{cases} \tau \dfrac{dV(t)}{dt} = -V(t) + I(t) & \\ V(t) := V_0 & \text{if} \quad V(t) \geq V_1 \end{cases} \tag{1}$$

ここで，$V(t)$ は時刻 t における細胞の膜電位の値，V_0 は細胞膜電位の初期値，V_1 は閾値である．τ は時定数を表す．

Krichmar らは小脳虫部（Occulomotor Vermis）における 256 個のプルキンエ細胞の定性モデルを構築し，衝動性眼球運動課題をシミュレートしている [Krichmar 97]．しかし，このモデルは小脳のプルキンエ細胞の挙動を実現する上では重要な立ち位置にあるものの，他の脳部位との接続を踏まえると，小脳虫部のプルキンエ細胞に限定してモデルが構築されていることは課題である．

連絡先: 田和辻 可昌，早稲田大学 大学院人間科学研究科，埼玉県所沢市堀之内 135-1 フロンティアリサーチセンター 213 実験室，wats-kkoreverfay@akane.waseda.jp

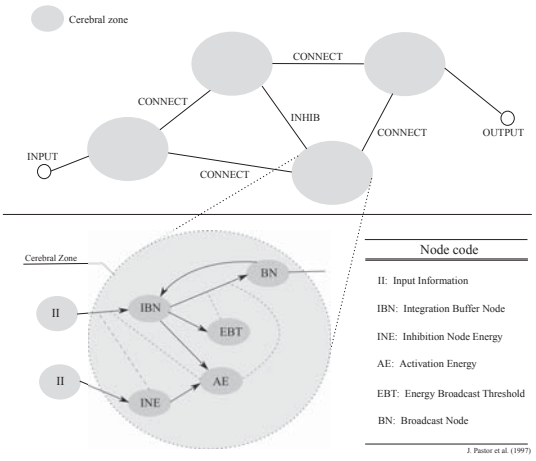


図 1: Cerebral Zone の generic モデル（下段）と構造ネットワーク（上段）（[Pastor 97] を参考に作成）

Pastor らは，大脳皮質の情報伝播モデルを構造レベルと機能・挙動レベルの二階層のレベルでモデル化した [Pastor 97]．図 1 に Pastors らの構築した二階層の大脳皮質情報伝播モデルを示す．構造レベルでは，皮質のニューロン間の興奮性結合，抑制性結合に関する解剖学的知見に基づいたネットワークが記述される（図 1 上段）．機能・挙動レベルでは，皮質ニューロンの generic なモデル（Cerebral Zone）が記述される．この generic なモデルでは,「ある閾値を超えるとニューロンが発火する」という点を表現している．具体的には，先行するニューロンから Integration Buffer Node で情報（i.e. エネルギーと情報の型）を受け取ると，Activation Energy が上昇する．この Activation Energy が Broadcast Threshold を超えると，Broadcast Node から情報が伝播し，次のニューロンに情報が送られる，というものである（図 1 下段）．この generic なモデルを構造レベルで記述した結合関係で繋げることで，皮質ネットワークを記述する．しかし，このモデルは大脳皮質における情報伝播にのみ着目しており，大脳基底核や小脳，および脊髄を介した身体との結合については述べられていない．

3. 定性ニューロンモデルの構造と挙動

3.1 着目する構造

本研究では Pastor らの定性ニューロンに基づいて，各脳部位の結合を行う．図 2 に本研究で対象とする脳部位間の構造的結合を示す．まず，マクロのレベルでは機能（e.g. 眼球運動制御など）に関わる脳部位の結合状態が解剖学的観点から記述される．各脳部位にはメゾのレベルとして，代表となる一つの投射ニューロンと介在ニューロンが存在しており，大脳皮質などは興奮性，黒質などは抑制性の投射ニューロンが，海馬などには抑制性の介在ニューロンが存在している．最後に，これらの代表ニューロンはすべて同一の基準で構造が与えられており，下記の式 (2)(3)(4) にしたがってエネルギーが伝播される．

$$\Delta \mathrm{Eqn} = \mathrm{ExcE} + \mathrm{InhE} \tag{2}$$

$$\mathrm{ExcE(InhE)} = \begin{cases} 1(-1) & \text{if IE is propagated} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$\mathrm{BE} = \begin{cases} 1 & \text{if Eqn} > \mathrm{Thr} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

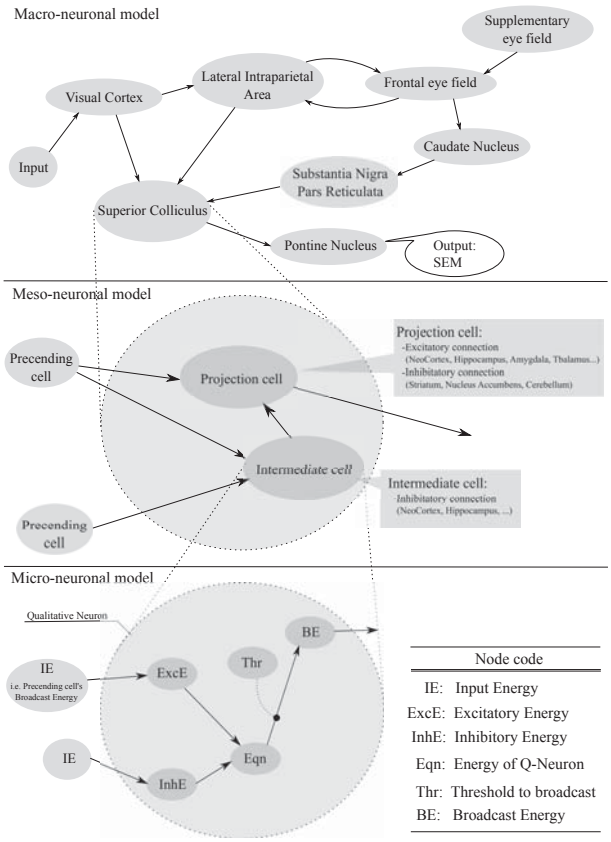


図 2: 本研究で対象とする脳部位の構造的ネットワーク．メゾレベルでは各部位に代表となる投射ニューロンと介在ニューロンが存在し，それらはマクロレベルにおいて統一的な定性ニューロンモデルとして記述される．

まず先行する代表ニューロンから興奮性（ExcE:Excitatory Energy）あるいは抑制性の入力（InhE:Inhibitory Energy）を樹状突起で受け取り，時間に関して積和を取る（Eqn:Energy of Q-Neuron）．この積和されたエネルギーがその代表ニューロン内で規定された閾値（Thr:Threshold to Broadcast）を超えることによって出力（BE:Broadcast Energy）される．

3.2 挙動シミュレーション

前小節で構築した定性ニューロンモデルを Stella Architect(Ver.1.1.2)*1 において実装した．実装した様子を図 3 に示す．Stella Architect では，図中央に表される長方形（ここでは，generic q-neuron）である「ストック」で変数量を，ストックに付随する「フロー」（ここでは，excitatory, inhibitory, broadcasting）によって変数の単位時刻あたりの入出量を表現する．両端のフローはそれぞれ興奮性の入力，抑制性の入力を表している．また，ストックにある変数量を energy of q-neuron（= Eqn）に格納し，閾値である threshold（= Thr）と毎ステップ比較する．この閾値を超えると，broadcasting のフローからそのときの energy of q-neuron の値が broadcast energy として流出する．

今，ステップ 1 からステップ 50 までステップ関数で持続的な興奮性の入力を，ステップ 20 からステップ 31 までステップ関数で持続的な抑制性の入力を与えたときの定性ニューロンの挙動を図 4 に示す．尚，この定性ニューロンの閾値は 5 で

*1 https://www.iseesystems.com

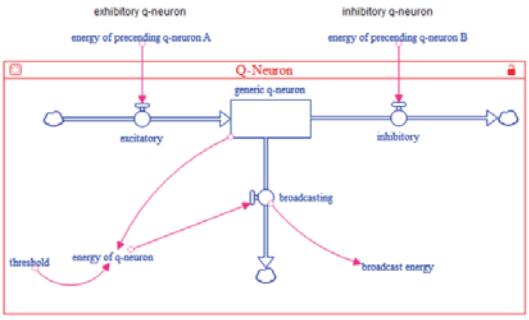


図 3: 実装した定性ニューロンモデル

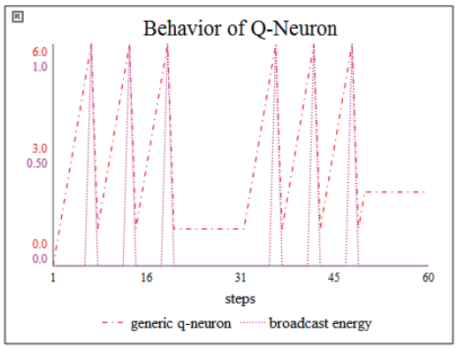


図 4: 興奮性・抑制性入力を与えたときの定性ニューロンの挙動

設定している．図の横軸はステップ数，縦軸はエネルギーの定性的な大きさを表しており，グラフにおける赤の点破線は定性ニューロンが有しているエネルギー，青い点線は定性ニューロンが伝播したエネルギーをそれぞれ表している．グラフから，一定の周期で定性ニューロンがエネルギーを伝播（活動）していることがわかる．また，抑制性の入力があるステップ 20 からステップ 31 ではエネルギー伝播がなく，定性ニューロンが抑制されていることも表現されている．この抑制性の入力がなくなると，また定性ニューロンは一定の周期でエネルギーを伝播する．これは入力が続くステップ 50 まで続き，興奮性の入力がなくなるとエネルギーを伝播しないことが分かる．

4. 定性ニューロンによる SEM の挙動

本節では前節で構築した定性ニューロンを繋ぎ合わせることによって，SEM の挙動を表現できるかを検証する．具体的には SEM の制御に関わる橋網様体の活動をシミュレーションにより観察する．

4.1 SEM に関わる神経基盤

SEM の形成には，大脳皮質，大脳基底核，脳幹がそれぞれ関与し，大脳皮質では視覚皮質・LIP 野・前頭眼野・補足眼野，大脳基底核では尾状核・黒質網様部，脳幹では上丘・橋網様体が関わっていることが知られている [Goldberg 14]．

図 5 は Stella Architect を用いて実装した，SEM に関わる神経基盤である．まず，各脳部位として前節で構築した定性ニューロンをそれぞれ実装した．次に，各脳部位の構造的結合を先行研究 [Goldberg 14] にしたがって実装した．それぞれの各脳部位のうち，尾状核から黒質網様部，黒質網様部から上丘への結合は抑制性の結合である．具体的には，これらの先行部位における BE（Broadcast energy）の値は，後結合部位の InhE（Inhibitory Energy）へと受け渡される．また，これらの部位のうち視覚皮質は情報の入力の起点であるとし，図 5 右上部に記すように，ステップ 1 からステップ関数で興奮性の入力を与えた．

彦坂によると，大脳基底核のうち黒質では高頻度の自発活動があることが知られている [彦坂 95]．そこで，図 5 左部に示すとおり，黒質網様部において高頻度の自発活動が行われるように，絶えず興奮性の入力を与え，かつ黒質の閾値（Thr）を他の脳部位と比べて小さいものとした．ただし，この興奮性の入力は他の脳部位から受け取っているものではなく，自発活動を本研究における定性ニューロンの枠組みで記述するために与えている．

4.2 橋網様体の活動シミュレーション

各脳部位の閾値 Thr をそれぞれ調整し，橋網様体の活動のパターンを観察した．閾値の調整に関しては，大脳皮質（視覚皮質／ LIP 野／前頭眼野／補足眼野）における閾値をそれぞれ 5 に設定し，これより他の脳部位の閾値を高く，あるいは低く設定することを試みた．ただし，高頻度の自発発火を表現するため，黒質網様部の閾値は大脳皮質の各脳部位の閾値よりも低いと仮定した．シミュレーションは 10000 ステップ実行した．図 6 は，黒質網様部の閾値と上丘の閾値の定性的な大小関係によって，橋網様部の活動がどのように変化するかを表したものである．ここでは，尾状核の閾値と大脳皮質の閾値は定性的に等しいものとしている．

各図の縦軸は橋網様体から伝播されるエネルギーの大きさを，横軸はステップ数を表している．この結果から，黒質網様体の閾値が上丘の閾値よりも高いと橋網様体からはエネルギーが伝播されないことが分かる．これは言い換えれば，上丘の閾値が高すぎると，視覚情報の入力があったとしても，SEM を引き起こさないことになる．一方で，黒質網様体の閾値が高いとき—特に黒質網様体の閾値が 2 で，上丘の閾値が 1 のときが顕著であるが—橋網様部で瞬間的な高頻度の発火が起こることが観察される．すなわち，大脳皮質の閾値 > 黒質網様部の閾値 > 上丘の閾値が成り立つことが期待される．

先行研究によると，リスザルにおいて SEM が起こる際は橋網様体においてバースト発火が起こることが示唆されている [Strassman 86]．このことから，定性ニューロンから構成された SEM に関する構造ネットワークモデルは，SEM 制御に関する現象に対して表現力を有していることが示唆される．

5. まとめと今後の課題

人間の認知機能に対する脳の構造的基盤からの理解を志向して，先行研究に基づき脳の大域的な情報伝播を表現するための定性ニューロンモデルを提案した．この上で，衝動性眼球運動（SEM）に関わる脳の構造ネットワークを定性ニューロンの結合によって実装した．シミュレーションにおいて，SEM の制御に関わると考えられる橋網様体の発火パターンを調査した結果，ある閾値に関する条件下で，解剖学的知見と一致する橋網様部でのバースト発火が確認された．このことから，定性的なニューロンの挙動記述であっても SEM 制御に関する現象を表現する能力があることが示唆された．今後は，新奇性が高い，あるいは予測と違った刺激に目を向ける，という現象を説明するために，定性ニューロンを用いて新奇性の検出に関する神経基盤や予測誤差に関わる神経基盤との結合を検討する．

本研究で構築した定性モデルは，モデル実装の点から定量的な情報に依存しているという重要な課題があり，この点では

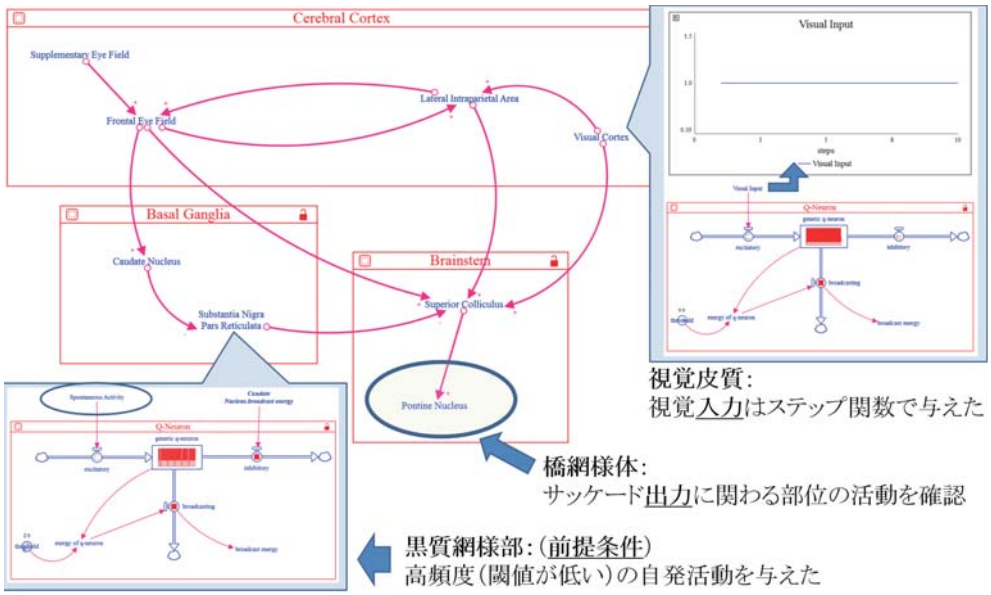


図 5: 実装した衝動性眼球運動に関わる構造的ネットワーク．各脳部位には 4 節で構築した定性ニューロンが実装されている．

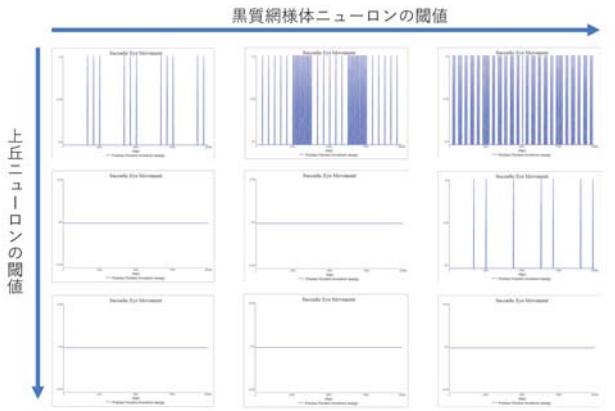


図 6: 黒質網様部の閾値と上丘の閾値の定性的大小関係による橋網様体の活動パターンの変化．閾値はそれぞれ 1, 2, 3 の組み合わせ．

integrated-and-fire モデルで得られる結果と同様であると考えられる．神経科学者は,「黒質から上丘への抑制は持続的にはたらいていて，サッケードの発現をひとまず抑えておく．尾状核から黒質への抑制は普段は働かない．しかし，尾状核が大脳皮質からの記憶や予測に関係する信号を受けて活動すると，それによって黒質が抑制されて，上丘のサッケード機構を準備ＯＫの状態にする.」[彦坂 95] といったように，系の挙動に関する定性的な因果関係で脳の伝播機序およびその機能を理解する．今後は神経情報伝播における変数間の変化に関わる因果伝播，すなわち解の生成過程を説明する定性的記述方式を検討する必要がある．

参考文献

[銅谷 12] 銅谷賢治: ニューロンのダイナミクスと数理モデル, 銅谷賢治他 (編), 脳の情報表現: ニューロン・ネットワーク・数理モデル, 朝倉書店, pp.9–14 (2012)

[西田 93] 西田豊明: 定性推論の諸相, 共立出版 (1993)

[彦坂 95] 彦坂興秀: 随意運動における大脳基底核の役割, 順天堂医学, Vol.42, No.2, pp.186-194 (1995)

[Goldberg 14] M.E. Goldberg, M.E. Walker（田中真樹, 松嶋藻乃訳）: 視線の制御, E.R. Kandel et al.(eds), カンデル神経科学 Fifth Edition, メディカル・サイエンス・インターナショナル (2014)

[Krichmar 97] J.L. Krichmar, G. A. Ascoli, L. Hunter & J.L. Olds: A Model of Cerebellar Saccadic Motor Learning Using Qualitative Reasoning, *Proceedings of International Work-Conference on Artificial Neural Networks*, LNCS Vol. 1240, pp.133-145 (1997)

[Pastor 97] J. Pastor, L. Travé-Massuyès, J-F. Démonet, B. Doyon & P. Celsis: BioCaen - A causal qualitative network for cerebral information propagation modeling, *Proceedings of the 11th International Workshop on Qualitative Reasoning*, Vol.1036, pp.305–315 (1997)

[Pessoa 14] L. Pessoa: Understanding brain networks and brain organization, *Physics of Life Reviews*, Vol.11, No.3, pp.400–435 (2014)

[Sporns 04] O. Sporns, D.R. Chialvo, M. Kaiser & C.C. Hilgetag: Organization, development and function of complex brain networks, *Trends in Cognitive Sciences*, Vol.8, No.9, pp.418–425 (2004)

[Strassman 86] A. Strassman, S.M. Highstein & R.A. McCrea: Anatomy and Physiology of Saccadic Burst Neurons in the Alert Squirrel Monkey. I. Excitatory Burst Neurons, *The Journal of Comparative Neurology*, Vol.249, pp.337–357 (1986)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

センター試験「数学」整数問題解答システムにおける探索空間の削減

Search Space Reduction in Automatically Solving Elementary Number Theory Problems in National Center Test

犬塚 慎也 *1　松崎 拓也 *1　佐藤 理史 *1
Shinya Inuzuka　Takuya Matsuzaki　Satoshi Sato

*1名古屋大学大学院 工学研究科 情報・通信工学専攻
Graduate School of Engineering, Nagoya University

We have been developing a number problem solver for National Center Test for University Admissions (NCTUA). The solver searches for a series of equivalence-preserving transformation of an input formula that results in a quantifier-free formula, from which the answer can be easily deduced. However, the increase of the search space hampered the capability of the solver. We introduced formula normalization and extended matching rules in the search in order to reduce the search space. Experimental results show that this method drastically reduces the search space and shorten the execution time.

1. 序論

本研究では，「ロボットは東大に入れるか」 (略称：東ロボ) プロジェクト [新井 12] の一環として開発されたセンター試験「数学 I・A」整数問題の自動解答システム [犬塚 17] における探索空間の削減に取り組んだ．センター試験の問題は図 1 に示すようにマーク式であり，論理式に翻訳すると空欄部分のみを自由変数とする論理式になる．そのため，このような問題を解くことは，論理式から限量子を除去し，自由変数のみからなる方程式・不等式を解く処理だと見なせる．

東ロボ数学解答システムのうち幾何や実代数の問題に対する処理では演繹に実閉体上の限量子除去を応用している [岩根 13]．しかし，整数領域においては限定的な体系の式を除いて限量子の除去は一般には不可能であり，これを行うアルゴリズムは存在しない．そのため，整数問題に対する解答システムにおいては問題の論理表現方法の工夫や逐次的な式変形探索を通じた発見的な方法による限量子の除去を行っている．

しかし，式変形探索においては探索空間の増大という課題があり，これを原因として解答が不能となる場合がある．この課題を解決するために，論理式を正規化して保持し，論理式と式変形規則とのマッチングの際にはこの正規化された式に対してマッチングを行うことを前提とする．また交換律や対称律を考慮するようにマッチング規則を拡張し，探索空間を削減することを試みた．

人手で作成した論理表現を入力とした評価実験の結果、本研究で改良を加えた探索手法は，得点率を低下させることなく探索空間を大幅に減少させられることが確認できた．さらに，解答システム全体については実行時間の短縮が確認できた．

2. システム概要

整数問題解答システム [犬塚 17] の構成を図 2 に示す．演繹部の入力は，一階述語論理式に特殊な述語を加えた体系の式である．これに数式処理を施し，解答を出力する．

演繹部においては、入力された問題はまずドライバによって複数の小問題に分割され、それぞれが整数問題用ソルバー (Heuristic Solver) に入力される．Heuristic Solver では，入力に対して限量子の除去および述語・関数に対応する計算処理を Algorithm 1 の通り行う．Algorithm 2 に示す限量子除去手続きや述語部分の処理においては，必要に応じて式変形探索による限量子除去 (Algorithm 3) が試みられるほか，プレスバーガー算術に対する限量子除去アルゴリズムの実装である Princess[Rummer 08] および数式処理システム Mathematica をモジュールとして用いる．処理が終了した式はドライバに返され，解答欄の桁数の条件を加えて解が求められる．なお，言語処理部との接続を念頭に，本稿では演繹部の入力を「中間表現」と呼ぶ．述語・関数の処理の詳細は [犬塚 17] に準ずる．

この構成のシステムについて，以降では 1 章で触れたマッチングの拡張を行う前のものと後のものをそれぞれ「旧システム」「新システム」と呼ぶ．

連絡先: 犬塚慎也, 名古屋大学大学院工学研究科 情報・通信工学専攻，名古屋市千種区不老町 C3-1(631)，052-789-4435，s_inuzuk@nuee.nagoya-u.ac.jp

k, l を自然数とし、$m = 6k$, $m + 56 = 44l$ と表すと、k, l について
一次不定方程式 $\boxed{\text{ケ}}(k + \boxed{\text{コ}}) + 1 = \boxed{\text{サシ}}\, l$ が成り立つ。

↓（言語処理による論理式への変換）

$$\forall k \forall l \forall m\big((m = 6k \wedge m + 56 = 44l) \rightarrow \boxed{\text{ケ}}(k + \boxed{\text{コ}}) + 1 = \boxed{\text{サシ}}\, l\big)$$

↓（限量子除去）

$$22\boxed{\text{ケ}} - 3\boxed{\text{サシ}} = 0 \wedge -2\boxed{\text{ケ}} + \boxed{\text{ケ}}\boxed{\text{コ}} + 1 - \boxed{\text{サシ}} = 0$$

↓（簡約）

$$\boxed{\text{ケ}} = 3 \wedge \boxed{\text{コ}} = 9 \wedge \boxed{\text{サシ}} = 22$$

図 1: 整数問題とその計算処理の例

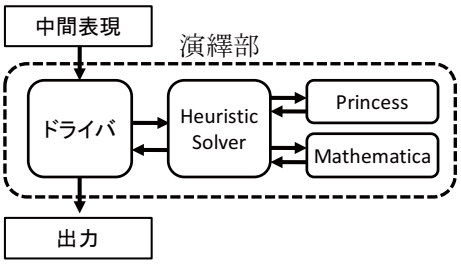


図 2: 整数問題解答システムの構成

Algorithm 1 整数問題用ソルバーが行う処理の全体像

```
procedure SOLVE(φ)                                  ▷ φ から限量子を除去する
    for all φ_i (φ の部分式) do
        φ_i ← SOLVE(φ_i)                            ▷ 部分式から処理を行う
    end for
    if φ に限量子が含まれている then
        φ ← REDUCE(φ)                               ▷ 限量子の除去を試みる
    else if φ に専用の述語が含まれている then
        φ ← 専用述語処理手続 (φ)                    ▷ 専用の述語の処理を試みる
    end if
    return φ
end procedure
```

Algorithm 2 限量子除去手続き

```
procedure REDUCE(φ)                                 ▷ φ から限量子を除去する
    if φ がプレスバーガー算術の式 then
        return PRINCESS(φ)                          ▷ Princess により限量子除去を試みる
    end if
    φ ← GRAPHSEARCH(φ)                              ▷ グラフ探索により限量子除去を試みる
    if φ が限量子を含む then
        φ ← MATH_REDUCE(φ)
                                                    ▷ Mathematica により限量子除去を試みる
    end if
    return φ
end procedure
```

Algorithm 3 式変形探索による限量子除去

```
rules ← (式変形規則の集合)
procedure GRAPHSEARCH(φ)                            ▷ 探索により φ から限量子を除去する
    try
        nodes ← {φ}                                 ▷ 探索ノード
        while nodes に未処理の式が存在 do
            φ ← (nodes のうち未処理のもの)
            for all φ_i (φ の部分式) do
                GRAPHSEARCHSUB(φ_i)
            end for
        end while
    catch SearchEnded, SearchStopped
        return (nodes のうち限量子のネストの深さが最も浅い式)
    end try
    return (nodes のうち限量子のネストの深さが最も浅い式)
end procedure
procedure GRAPHSEARCHSUB(φ)                         ▷ サブルーチン
    for all rule ∈ rules do
        φ* ←REWRITE(φ, rule)                        ▷ 式の書き換えを試行する
        if φ* ≠ φ then nodes ← nodes ∪ {φ* を部分式とする論理式 }
            if φ* を部分式とする論理式が限量子を含まない then
                throw SearchEnded
            end if
            if |nodes| ≥ Threshold then
                throw SearchStopped
            end if
        end if
    end for
    return φ
end procedure
```

3. 正規化とマッチング規則の拡張による探索空間の削減

整数問題解答システムの式変形探索における探索空間増大の原因を明らかにするため，開発用の問題 14 問分について，各式変形規則の適用回数を調べたところ，図 3 の通りであった．これにより，本質的に重要な式変形は適用回数が少なく，一方で可換性や対称性を吸収するための簡単な式変形規則が式変形試行の大部分を占めており，探索空間の増大につながっていることが判明した．

旧システムでは，探索時に論理式が式変形規則の左辺と一致するかどうかを確かめるマッチングを行うが，これは純粋な First order matching であった．このマッチングでは乗算の可換性や等号の対称性を考慮できないため，たとえば乗算の可換性 $a \times b = b \times a$ のような単純な法則についても，それらの法則自体を式変形規則に加えていた．探索の際には，たとえば可換性を持つ演算が出現するすべての箇所に関して，上記のような規則によって交換した結果をすべて生成し，それぞれが探索空間上のノードとなる．したがって，この探索手法では，入力となる中間表現が長く，可換性や対称性を考慮しなければならない箇所が多いほど探索空間の増大が顕著になる．

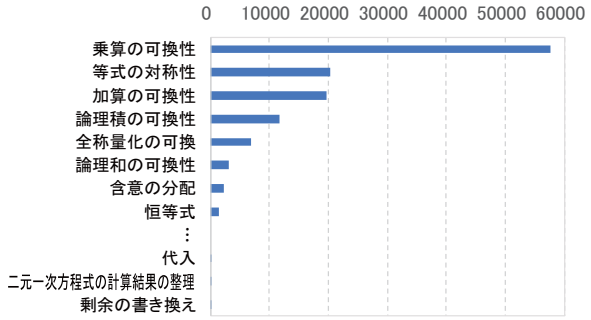


図 3: 旧システムにおける式書き換え規則の適用数

これを受けて，式の正規化とマッチング規則の拡張を組み合わせることによる探索空間の削減を試みた．この式変形探索では，式を書き換える規則を，表 1 に示す例のような「正規化規則」と，表 2 に示す「式変形規則」の 2 つに分け，式変形の探索やノードの生成は表 2 の式変形規則のみを用いて行う．各ノードの式は，正規化規則を再帰的に適用することによって正規化された形で保持される．正規化規則には，表 1 に示したもの以外に，$a - b \Leftrightarrow a + (-b)$ のような算術演算子に関する規則，$a = b \Leftrightarrow a - b = 0$ のように演算子や等号・不等号に関する規則，$\phi_1 \to (\phi_2 \to \phi_3) \Leftrightarrow (\phi_1 \wedge \phi_2) \to \phi_3$ のように論理演算子に関する規則，$\forall x \forall y \phi \Leftrightarrow \forall x, y \phi$ のように量化子に関する規則など，合計で 41 種類の規則がある．また，算術演算子や論理演算子については，$\phi_1 \wedge (\phi_2 \wedge \phi_3)$ のような結合律を満たす演算子の入れ子構造を $\phi_1 \wedge \phi_2 \wedge \phi_3$ のように順序を区別しないで保持するようにする．正規化規則の中には，複数の剰余の条件を 1 つにまとめる表中 3 番の規則や，有界な束縛変数に関する展開を行う 9 番の規則のように，演繹的な役割を持った規則も含まれているが，正規化規則を適用しても新たなノードは生成されない．

探索時のマッチングは，正規化された式に対して交換律や対称律を考慮するように拡張する．たとえば等号の対称律は，論理式に含まれる等式の右辺が 0 になるように正規化し，また式変形規則に含まれる等式も右辺が 0 になる形で作成することで，マッチング時には左辺と右辺の入れ替えを考慮する必要をなくす．また，加算や乗算の交換律は，多項式を展開し，束縛変数について同類項をまとめて整理した形で保持する．また式変形規則も束縛変数について整理された形で作成することで，その多項式の項について考えられる順列のうち，パターンとマッチするものの一つを採用する．ただし，パターンの中に束縛変数が複数含まれる場合においては，束縛変数について対称となるようにパターンを構成しているため，探索漏れは生じない．

旧システムの場合と同様に，探索では部分式に適用可能な式変形規則を再帰的に適用し，幅優先探索を行うことで限量子の除去を試みる．限量子の除去が完了したか，あるいは試行回数

表 1: 正規化規則の例

番号	変形前 *1	変形後 *1
1 *2	$(i_1 i_2 x + i_3) \bmod i_1$	$i_3 \bmod i_1$
2 *3	$(i_1 i_2 x + i_3) \operatorname{div} i_1$	$i_2 x + (i_3 \operatorname{div} i_1)$
3 *4	$\bigwedge_k -i_k + f \bmod j_k = 0$	$-i_{new} + f \bmod j_{new} = 0$
4	$i_1 x + i_2 > 0$	$\begin{cases} x > \lfloor \frac{i_2}{i_1} \rfloor & (i_1 > 0) \\ x < \lceil \frac{i_2}{i_1} \rceil & (otherwise) \end{cases}$
5	$\bigwedge_{k=1}^{j} -i_k + x > 0$	$-\max\{i_1, i_2, \cdots, i_j\} + x > 0$
6	$\bigwedge_{k=1}^{j} i_k - x > 0$	$\min\{i_1, i_2, \cdots, i_j\} - x > 0$
7	$\forall x$(x を含まない論理式)	(x を含まない論理式)
8	$\exists x$(自由変数を含まない論理式)	$\begin{cases} \top & (\text{論理式を充足する } x \text{ が存在するとき}) \\ \bot & (otherwise) \end{cases}$
9	$\forall x_1, \cdots, x_n \Big[\Big\{ \phi(x_1, \cdots, x_n) \wedge \bigwedge_{k=1}^{m} i_k \le x_k \le j_k \Big\} \to \psi(x_1, \cdots, x_n) \Big]$	$\bigwedge_{x_1^*=i_1}^{j_1} \cdots \bigwedge_{x_m^*=i_m}^{j_m} \forall x_{m+1}, \cdots, x_n \Big[\phi(x_1^*, \cdots, x_m^*, x_{m+1}, \cdots, x_n) \to \psi(x_1^*, \cdots, x_m^*, x_{m+1}, \cdots, x_n) \Big]$

*1 x は束縛変数，i, j は定数，f は多項式，h は束縛変数を含まない多項式，ϕ は論理式．
*2 $f \bmod i$ は f を j で割ったときの余りを表す．
*3 $f \operatorname{div} j$ は f を i で割ったときの商を表す．
*4 $j_{new} = \operatorname{lcm}(j_1, j_2, \cdots)$ であり，元の式と論理的に同値になるように i_{new} を決定する．

が閾値を超過した時点で探索を終了し，書き換え結果のうち限量子のネストの深さが最も浅いものを最良の結果として出力する．表 2 に示す式変形規則のうち，3 番と 5 番以外はすべて，旧システムにおいて用いていた規則を，含意の左辺に他の論理式が加わった場合や，連言で結合された部分式のうち複数のものが同じパターンにマッチする場合を扱えるように拡張したものである．また，旧システムにおいて低レベルな式変形規則による書き換えを経由して行っていた変形列は，正規化とマッチングの拡張によって，実質的に新システムでもすべて可能である．したがって，旧システムで可能だった式変形が新システムにおいて不可能になることはない．

4. 評価

システムの評価に際し，予備校マーク式問題集に掲載されている整数問題 25 問を用い，対応する中間表現を手作業により表 3 の通り作成した．この中間表現に対し，旧システムを用いた場合，および新システムを用いた場合のそれぞれについて，得点・実行時間・式変形探索の展開ノード数を調べた．

実験の結果は表 4 の通りであった．開発用データ，評価用データともに，合計得点，総実行時間，総展開ノード数が改善した．特に展開ノード数は劇的に減少した．旧システムではほとんどの論理式が交換律などの簡単な式変形規則にマッチしてしまい，いかなる子孫ノードについても本質的な式変形が不可能であるようなノードが多数生成されていた．これに対し新システムでは本質的な式変形以外では新しいノードを生成しないため，探索ノード数が大幅に減少したと考えられる．また，得点率の上昇は限定的であるが，これは，新システムには旧システムと比べて本質的に新しい式変形規則を多数追加したわけではないので，旧システムで除去できなかった限量子の多くは新システムでも除去できないためであると考えられる．

次に，各問題について旧システムと新システムそれぞれの実行時間の違いを調べた．図 4 は評価用データに対する Heuristic Solver の実行時間について，横軸を旧システムの実行時間，縦軸を新システムの実行時間にとって問題ごとにプロットしたものである．破線は縦軸と横軸の値が等しくなる直線である．

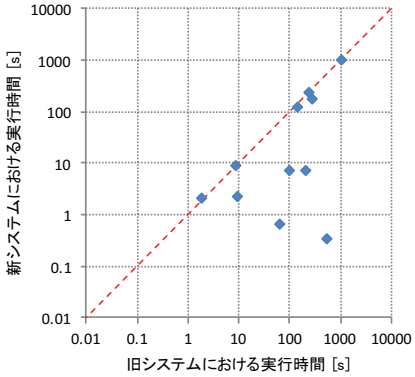


図 4: 旧システム・新システムの実行時間比較 (評価用データ)

図 4 が示すように，評価用データでは 11 問中 8 問において実行時間短縮の効果が認められた．実行時間が短縮した問題についてはすべての問題において 20 秒以上短縮されたが，実行時間が伸びたものについては，その時間変動はすべて 2 秒以内となっている．この実行時間の微小な増加は式変形探索以外の部分，主にドライバ・Princess・Mathematica における時間消費の揺らぎによるものと思われる．

最後に，評価用データのうち，新たに解けるようになった問題についてその理由を調べたところ，正規化規則 (40 番) により旧システムでは行われなかった変形が実現することで解けたものが 1 問，探索空間の削減の効果で解けるようになったものが 1 問あった．このうち後者は，旧システムにおいては探索ノード数が閾値 (実験では 1500 に設定) に達してしまい，限量子を除去できないまま探索を終了してしまうが，新システムにおいては図 5 に示す変形手順によりノード数 3 で限量子の除去を完了し，探索が終了する．

以上のように，低レベルな式変形は保持する式の正規化とマッチング規則によって吸収することで，より本質的な式変

表 2: 式変形規則

番号	変形前 *1	変形後 *1
1	$\forall x\,[(-i_1 + x \bmod i_2 = 0 \wedge \phi(x)) \to \psi(x)]$	$\forall y\,[\phi(i_1 + i_2 y) \to \psi(i_1 + i_2 y)]$
2	$\exists x_1, \cdots, x_m \left[\bigwedge_{k=1}^{m} f_k(x_1, \cdots, x_m) = 0\right]$	$\begin{cases}\top & (\text{充足する } x_1, \cdots, x_m \text{の組が存在}) \\ \bot & (otherwise)\end{cases}$
3	$\forall x_1, \cdots, x_m \left[\bigwedge_{k=1}^{m} (h_k + \sum(f(x_1, \cdots, x_m) \bmod i)) = 0\right]$	$\bigwedge_{i_1,\cdots,i_m \in \{0,1,\cdots,i-1\}} \bigwedge_{k=1}^{m} (h_k + \sum(f(i_1, \cdots, i_m) \bmod i)) = 0$
4	$\forall x, y \left[\left\{\phi(x,y) \wedge \bigwedge_{k=1}^{m} f_k(x,y) = 0\right\} \to \psi(x,y)\right]$	$\bigwedge_{(i_x,i_y) \in \left\{(i_x^*, i_y^*) \mid \bigwedge_{k=1}^{m} f_k(i_x^*, i_y^*) = 0\right\}} [\phi(i_x, i_y) \to \psi(i_x, i_y)]$
5	$[\exists x \phi_1(x) \wedge \phi_2] \to \phi_3$	$\forall x\,[\{\phi_1(x) \wedge \phi_2\} \to \phi_3]$
6	$\forall x_1, \ldots, x_n \left[\left\{\phi_1(x_1, \ldots, x_n) \wedge \bigwedge_{k=1}^{m} - i_k + x_k = 0\right\} \to \psi_1(x_1, \ldots, x_n)\right]$	$\forall x_{m+1}, \cdots, x_n$ $\left[\phi_1(i_1, \ldots, i_m, x_{m+1}, \ldots, x_n) \to \psi_1(i_1, \ldots, i_m, x_{m+1}, \ldots, x_n)\right]$
7*2	$\forall x \left[\phi(x) \to \bigwedge_k \left(\sum_l h_{k,l} x^l = 0\right)\right]$	$\bigwedge_{k,l} h_{k,l} = 0$
8*3	$\forall x_1, x_2 \left[\phi(x_1, x_2) \to \bigwedge_k \left(\sum_{l_1,l_2 \in \{0,1,2\}} h_{k,l_1,l_2} x_1^{l_1} x_2^{l_2} = 0\right)\right]$	$\bigwedge_{k,l_1,l_2} h_{k,l_1,l_2} = 0$
9*4	$\forall x_1, x_2 [\{i_0 + i_1 x_1 + i_2 x_2 = 0 \wedge \phi(x_1, x_2)\} \to \psi(x_1, x_2)]$	$\forall y [\phi(k_1 + l_1 y, k_2 + l_2 y) \to \psi(k_1 + l_1 y, k_2 + l_2 y)]$
10*5	$\forall x_1 \left[\phi(x_1) \to \exists x_2 \left\{\bigwedge_k h_{k1} + i_{k1} x_1 + h_{k2} x_2 = 0\right\}\right]$	$\bigvee_d \left[\left(\bigwedge_k \frac{i_{k1}}{d} + h_{k2} = 0\right) \wedge \exists y \left(\bigwedge_k h_{k1} + \frac{i_{k1}}{d} y = 0\right)\right]$

*1 x, y は束縛変数，i, j, k, l は定数，f は多項式，h は束縛変数を含まない多項式，ϕ は論理式．

*2 $\phi(x)$ を充足する x が上限あるいは下限を持たない場合のみ書き換えを行う．

*3 $\phi(x_1, x_2)$ が三次以上の項を含んでおらず，$\phi(x_1, x_2)$ を充足するすべての x_1, x_2 を解に持つ二元二次方程式・二元一次方程式がいずれも存在しない場合のみ書き換えを行う．

*4 $i_0 + i_1 x_1 + i_2 x_2 = 0 \Leftrightarrow \exists y\,[x_1 = k_1 + l_1 y \wedge x_2 = k_2 + l_2 y]$ となるように k_1, l_1, k_2, l_2 を設定する．

*5 $\phi(x_1)$ が x_1 について上限あるいは下限の高々一方のみを条件として与え，それ以外のいかなる条件をも与えない場合にのみ書き換えを行う．d はすべての i_{k1} の公約数．

表 3: 実験データ

-	問題数	配点
開発用	14	269
評価用	11	195
合計	25	464

表 4: 実行時間と展開ノード数

-	-	旧システム	新システム
開発用	得点	87/269	106/269
	平均実行時間 [s]	115.14	109.90
	平均展開ノード数	10357	1.29
評価用	得点	88/195	96/195
	平均実行時間 [s]	235.58	140.13
	平均展開ノード数	21224	2.55

$$\forall a \left[a \bmod 7 = 2 \to \forall b \left\{b \bmod 7 = 3 \to ab \bmod 7 = \boxed{\text{ケ}}\right\}\right]$$
$$\updownarrow \text{(正規化)}$$
$$\forall a \left[-2 + a \bmod 7 = 0 \to \forall b \left\{-3 + b \bmod 7 = 0 \to -\boxed{\text{ケ}} + ab \bmod 7 = 0\right\}\right]$$
$$\updownarrow \text{(式変形規則 1)}$$
$$\forall a \left[-2 + a \bmod 7 = 0 \to \forall k_b \left\{-\boxed{\text{ケ}} + (3a + 7ak_b) \bmod 7 = 0\right\}\right]$$
$$\updownarrow \text{(正規化)}$$
$$\forall a \left[-2 + a \bmod 7 = 0 \to \forall k_b \left\{-\boxed{\text{ケ}} + (3a) \bmod 7 = 0\right\}\right]$$
$$\updownarrow \text{(式変形規則 1)}$$
$$\forall k_a, k_b \left\{-\boxed{\text{ケ}} + (6 + 21ak_a) \bmod 7 = 0\right\}$$
$$\updownarrow \text{(正規化)}$$
$$-\boxed{\text{ケ}} + 6 = 0$$

図 5: 評価用データにおける式変形手順の例

形のみによる探索ノードの展開が可能になった．したがって，新システムでは探索空間の増大の影響を抑えつつ，式変形規則をさらに追加することが可能となったと予想される．今後は，現時点で限量子除去ができていな論理式の簡約に必要な式変形規則の精査，ヒューリスティックな解法パターンの抽出・蓄積が必要である．

5. 結論

本研究では，センター試験「数学」整数問題解答システムの探索処理を改良し，探索空間の削減に取り組んだ．式の正規化と拡張されたマッチング規則を用いた式変形探索手続きを構成し，開発用データ，評価用データのそれぞれに対して得点・実行時間・展開ノード数を調査した．その結果，低レベルな式変形規則による探索空間の増大が抑止されることで，実行時間が平均的に減少することが確かめられた．

今後は，現時点で限量子を除去できていない形の論理式については，簡約にどのような式変形規則が必要なのか，また式変形規則の個数・粒度はどの程度が適切か，より詳細な調査を行うことが課題である．

参考文献

[新井 12] 新井 紀子, 松崎 拓也: ロボットは東大に入れるか?—国立情報学研究所「人工頭脳」プロジェクト—, 人工知能学会誌 Vol. 27, No. 5, pp. 463–469 (2012)

[犬塚 17] 犬塚 慎也, 松崎 拓也, 佐藤 理史: センター試験「数学」整数問題解答システムの開発, 人工知能学会全国大会論文集, ROMBUNNO.3O1-5 (2017)

[岩根 13] 岩根 秀直, 松崎 拓也, 穴井 宏和, 新井 紀子: 数式処理による入試数学問題の解法と言語処理との接合における課題, 人工知能学会全国大会論文集 (CD-ROM) Vol. 27, ROMBUNNO.2A4-2 (2013)

[Rummer 08] Philipp Rummer: A constraint sequent calculus for first-order logic with linear integer arithmetic, LNCS, vol. 5330, pp. 274–289. Springer (2008)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

文字 Embedding を使用したニューラルネットワークによる顔文字の原形推定

A Method to Estimate the Original Form of Kaomoji Using Neural Network with Character Embedding Function

奥村 嶺　　奥村 紀之
Rei Okumura　Noriyuki Okumura

明石工業高等専門学校
National Institute of Technology, Akashi College

This research focuses on Kaoomji (Asian style emoticons) used in text-based communication. We analyze Kaomoji that are made from particular characters' sequence (original shapes). In this paper, we construct a neural network with embedding function to estimate the original form of Kaomoji. As a result, we find out that our proposed method can estimate the original form of Kaomoji with 0.746 of accuracy ratio.

1. はじめに

光通信や 4G/LTE 通信などの高速インターネット普及に伴い，オンラインでのコミュニケーションは文字情報だけでなく，画像や音声，動画などのメディアも使用可能なサービスが増えてきている．しかし，依然として SNS 等における文字情報の必要性は高く，LINE[*1] でのメッセージのやり取りや Facebook[*2] のコメント，Twitter[*3] のリプライなど，文字列に依存したコミュニケーションは活発に交わされている．

ところが，日本語でのコミュニケーションでは，空気を読む，雰囲気に合わせるといった，察する能力が求められる．このようなハイコンテクスト（抽象的な表現を好む）言語であるとされる [Hall 76] 日本語は，英語やドイツ語のような多くの内容を明記して情報を伝達する傾向のある言語と比べ，文字列のみのコミュニケーションと両立させることが困難である．一例として，不用意な発言 1 つで炎上 [*4] などのトラブルに発展してしまうことがある．これは，書き手の言葉以外の情報（表情等）が，正確に受け手に伝わらないことで起こる問題である．例えば受け手が，書き手の意思とは反する字面通りの解釈をしてしまったり，書き手の言葉の奥にあるより伝えたい情報を読み違えたりして引き起こされる．

このように，日本語でのコミュニケーション，特に対面での会話では，話し手は多くを語らず，身振りや表情を交えることで聞き手に言葉以外の情報を伝える．そのため，文字列でのコミュニケーションでは，受け手は文字情報のみから全ての情報を読取る必要があり，読み手の負担が増大してしまう．読み手の負担を軽減するには，書き手が過不足無く情報を記載することが求められる．しかし，過不足なく理解しやすい文章を書くことは決して容易ではなく，そのようなことに慣れていない日本語話者にとっては極めて深刻な問題である．

書き手と読み手の負担を軽減させる一例として，顔文字が挙げられる．顔文字は＼(^o^)／のように，表情や動作を文字列（記号列）によって表したものである．この顔文字例からは，嬉しい表情 ((^o^)) と両腕を上に挙げる動作 (＼／) の情報を得ることできる．このように，顔文字を用いることで，文章では直接伝えられない表情や動作を視覚的に伝達することが可能となる．

しかし，奥村によれば現在，顔文字の種類は 10 万種を超えており，今後も増加すると予想される [奥村 16c]．そのため，顔文字のリストを大量に保持するだけでは不十分であり，未知の顔文字に対するアプローチを検討する必要がある．さらに，顔文字は言語として成立している文字列の解析とは異なり，ありとあらゆる記号列が構成要素となる．このことから，すべての顔文字に品詞等のルールを制定することが困難である．そこで新たなアプローチとして，顔文字をある文字列（原形）から派生したものであるとみなし，原形に付随するパーツの傾向から確率的に顔文字の持つ情報を推定することが重要であると考えられる．

我々は，前述の新たなアプローチを研究する前段階として，とある顔文字（以下「元文字列」と呼称）から，その顔文字の派生元である文字列（以下「原形」と呼称）を推定する検証を行っている．我々は，この推定を顔文字の原形推定と呼称している．推定には，Cosine 類似度による手法や，サポートベクトルマシン（SVM）による手法など，多種多様なアプローチがある．本実験ではニューラルネットワークを用いた手法を扱う．従来手法 [奥村 17] では，顔文字を数値化する手法に問題があり，優良な結果が得られなかった．そこで本稿では，顔文字をニューラルネットワークで用いるにあたって Embedding を導入し，検証した結果について述べる．

2. 関連研究

本研究に関連して，我々は顔文字へのアノテーションを進めている [奥村 16c]．顔文字を包括的に扱い，確率的なモデルとして活用する為には，大規模な辞書が必要となる．顔文字は様々な記号から作られる可能性があるため，記号に着目して多元的に分析する必要がある．

顔文字の抽出に関する研究として，Tanaka らの研究では，機械学習によって顔文字を抽出を試みている [Tanaka 05]．また，文中からの顔文字抽出に関する研究として，Bedrick らの手法がある．Bedrick らは，隠れマルコフモデル (HMM) によって記号列を抽出し，確率的文脈自由文法 (PCFG) に基づく評価法によって顔文字の候補を選別する手法を提案している [Bedrick 12]．

また，テキストに含まれる顔文字の抽出に関する研究としては，Ptaszynski らの研究がある [Ptaszynski 10]．Ptaszynski らの CAO システムでは，顔文字における目-口-目の並び

連絡先: 奥村 嶺，明石工業高等専門学校，e1371@s.akashi.ac.jp

∗1 https://line.me
∗2 https://www.facebook.com
∗3 https://twitter.com
∗4 非難や誹謗中傷のコメントが殺到して収拾がつかなくなる状態

(Triplet) に着目した顔文字抽出を試みている．このシステムでは，1 万種の顔文字にデータベース規模で対応し，90% 以上の精度で抽出可能だとしている．さらに自動拡張によって 300 万種に対応可能であると述べられており，これは顔文字の研究の中では非常に規模の大きいものである [Ptaszynski 12]．奥村らは，Ptaszynski らの Triplet に加え，輪郭を表現する文字列を加えたものを顔文字の原形として定義 [奥村 16b] しており，本実験の原形はこれに準ずる．さらに，同じ顔文字の原形を推定する研究に関しては，手法としてサポートベクトルマシン (SVM) を用いた研究 [奥村 16a] が参考になる．

3. 顔文字の原形推定

先行研究 [奥村 16c] より，顔文字の原形はおよそ 3,000 種類発見されている．また，顔文字と認識された任意の文字列とその原形のペアが 43,733 組分類されている．

本稿では，顔文字と認識された任意の文字列（以下「元文字列」と略す）から，その顔文字の原形をニューラルネットワーク (Neural Network，以下「NN」と略す）によって推定する検証の進行状況を述べる．図 1 に顔文字の原形推定の例を示す．

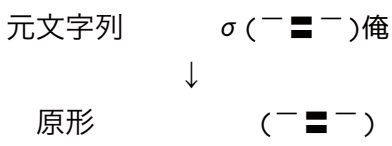


図 1　顔文字の原形推定の例

実験環境としては，Python(3.5.2)，Chainer(2.0.0)，CUDA(8.0.44) を用い，評価指標としては，式 (1) 示す分類正解率 (Accuracy) を用いた．分類正解率は 0 から 1 の値を取り，値が 1 に近いほど優良とする．

$$\text{分類正解率} = \frac{\text{正解した事例数}}{\text{検証事例数}} \tag{1}$$

4. 従来の文字列の数値化と問題点

NN では数学的な計算によって学習を進めていくため，文字列を文字情報として直接用いることができない．そのため，文字列を何らかの数値に変換する必要がある．これまでの検証 [奥村 17] では，顔文字を次の通りに数値化している．

原形の数値化

原形については，原形の種類ごとに 0 から 3,109 のラベルを付与する．具体的には，全ての原形に対して，原形の種類ごとに固有の数値を順に割り当てる．

元文字列の数値化

元文字列は，元文字列に使用されている文字だけでなく，文字の並びも考慮して数値化する必要がある．事前調査により，Bigram，Trigram，SkipBigram（一文字飛ばしの Bigram）において，高い分類正解率が得られたため，この 3 種類の N-gram に着目している．具体的な手順を次に示す．

1. 全ての元文字列に対して Bigram 要素，Trigram 要素，SkipBigram 要素を抽出
2. 各 N-gram 要素を数値に変換 (後述※ 1)
3. 各 N-gram ごとに得られた数値列を一列になるように結合

また，上記の※ 1 の手順を Bigram を例として以下に示す．

1. Bigram 要素全てに対して，その要素が既にデータベースにあるか調べる．
2. 要素がデータベースにない場合，新たな重複しない値を要素に割り当てる．また，データベースに登録する．
3. 要素がデータベースにある場合，その数値を参照し，割り当てる．

以上の工程を経て，従来の数値化では表 1 に示すようなデータを作成していた．

このような数値化手法では，各抽出要素の数値の差が各抽出要素の意味の異なり度合いと一致するように学習が行われる．そのため，数値の大小と N-gram の特徴を対応付けることができず，NN の学習の弊害となっていることが推察される．

5. 提案手法

前節に示した，数値化の問題の解決策として，我々は Embedding を用いた検証を行った．Embedding とは，単語や文字，概念などを分散表現と呼ばれるベクトル空間に埋め込むことである．単語を分散表現で表す手法 (Word Embedding) の代表的な例として，Mikolov らの Word2Vec[Mikolov 13] が挙げられる．

我々が NN に組み込んだ Embedding 層は，文字を識別するユニークな数値 (スカラ) を入力とし，固定長のベクトルを出力とする．本実験では出力ベクトルの大きさ (長さ) を 100 としているため，この Embedding 層は各文字を 100 個の要素を持ったベクトルに変換する．この変換は NN の学習によって動的に変化する．

表 1　データ加工の一例

元文字列	σ(￣〓￣)俺？							
Bigram	抽出要素	σ(	(￣	￣〓	〓￣	￣)	)俺	俺？
	数値化	9900	4748	10805	9682	1672	19532	4896
Trigram	抽出要素	σ(￣	(￣〓	￣〓￣	〓￣)	￣)俺	)俺？	
	数値化	17567	18223	31134	42804	18325	61194	
SkipBigram	抽出要素	σ￣	(〓	￣￣	〓)	￣俺	)？	
	数値化	14250	8836	34142	26975	13545	17352	
原形	(￣〓￣)			ラベル	266			

図 2 に実験で用いた NN の構造を示す．

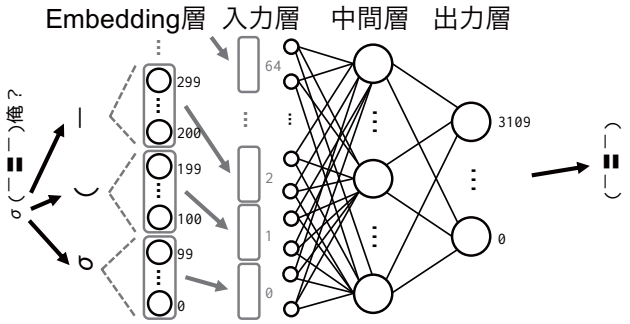


図 2　Embedding を用いる NN

ここで，図 2 に明記している各数値の根拠を示す．

- Embedding 層から出力されるベクトルが 65 個である理由は，実験に使用したデータの元文字列の長さの最大が 65 だからである．元文字列が 65 よりも短い場合は，文字なしに相当する分散表現を出力する．
- 出力層が 3,110 ユニットである理由は，実験に使用したデータの原形の種類数が 3,110 だからである．

5.1　使用したデータ

本項では，使用した学習データと検証データについて示す．元データにある 43,733 組の元文字列と原形のデータには，表 2 に示すように，原形の種類に応じて元文字列の個数にばらつきがあり，100 種類以上の元文字列がある原形もあれば，1 つの元文字列しかない原形もある．このばらつきは，NN の学習データを作成する際に，選択される原形の種類に偏りが生じるような弊害となる．

我々はこの問題を解決するために，3,110 種類ある原形全てが 100 種類の元文字列を持つようなデータ（LargeKaomoji-Data:LKD）を作成し，これを学習データとして用いた．作成にあたっては，全原形それぞれに対し復元抽出を行った．

評価データは，元データから元文字列と原形の組を 1,000 回復元抽出したものを 10 セット用意し使用した．

5.2　従来手法との比較

従来手法で使用されている NN と比較するため，中間層のユニット数や活性化関数，最適化手法を以下の通りに設定し，学習を行った．

- 中間層のユニット数は 10,000
- 活性化関数は式 (2) に示す Sigmoid 関数
- 最適化手法は Adam

$$Sigmoid(x) = \frac{1}{1+\exp(-x)} \tag{2}$$

図 3 に学習の経過 (Epoch) と分類正解率 (Acc) の変化を示す．また，図中の破線は学習データに対する分類正解率の推移で，実線は評価データに対する分類正解率の推移である．図 3 より，評価データに対する推移に着目すると，20Epoch の学習で既に分類正解率 0.7 を超え，70Epoch 学習したところで最大の分類正解率である 0.746 に達している．また，従来手法では 280Epoch の学習で分類正解率が 0.51 であることから，提案手法は従来手法と比べ少ない学習回数でより優良な分類正解率を示すことが見て取れる．

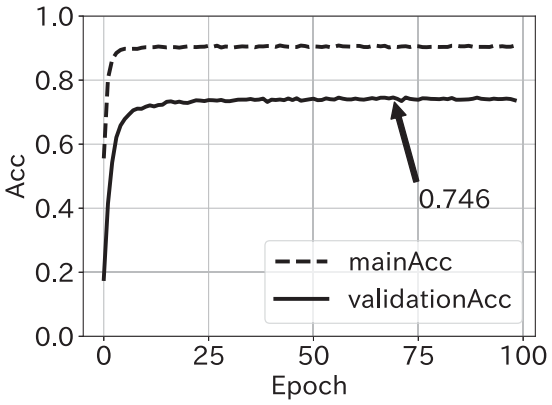


図 3　Embedding 層を追加した NN での学習回数と分類正解率の変化

表 2　元データに含まれる原形と元文字列の個数

原形	元文字列の事例数	事例 (元文字列)
(・_・)	1,032	ドキ（*・ __ ・*）ドキ C= C= ＼(;・_・)/ ーー＝＝（/_・)/＝＝≡≡卍スカッ
(-_-)	610	”/(;-_-) ｲﾃﾃ･･･ ヾ(-_-)ｼ…ｱ ﾖｲﾖｲ！ ＼(-__________-;)／
(・ω・)	589	(‘・ω・´)ノ ヾ(・ω・`)ノ ←・・・・・・(・ω・.) ジーッ
U=ｴ=U	1	ブルブル（(U= ｴ =U)）トイレ行きたい
(ε(ｴ)ε)	1	○-○へ キョロC(ε(ｴ)ε)つ)))(((C(3(ｴ)3) つキョロ
(|￣|_|￣|)	1	拉致 |Φ|(|￣|_|￣|)|Φ| 監禁！

5.3 従来手法との差異と考察

従来手法では，元文字列を Bigram, Trigram, SkipBigram の各要素に分解し，それぞれに固有の数値を与えることで，NN に入力し学習していた．そのため従来手法では，各 N-gram 要素と入力層にインプットする数値は固定されており，学習において変化しない．しかし，提案手法では，各文字を分散表現 (固定長ベクトル) に変換する過程も NN に組み込む．その結果，各文字のベクトル表現も学習によって改善させることができ，分類正解率が向上したと考えられる．

5.4 学習データに対する分類正解率

本来，学習データは学習に用いるデータであるため，学習データに対する Accuracy は限りなく 1 に近づくことが学習の理想である．しかし，図 3 の破線に着目すると，学習の進行に伴って分類正解率が 0.9 から変動していないことが見て取れる．これは，顔文字の原形抽出ルールが関係している．

我々が準じている顔文字の原形抽出ルールでは，元文字列が複数の原形を含んでいる場合，その全ての原形を抽出するルールとなっている．しかし，提案手法の NN では，1 つの元文字列に対して 1 つの原形のみ推定する構造となっているため，複数の原形を持つ事例では，一定の確率で誤りだとカウントされてしまう．複数の原形を持つ事例は，学習データには約 30% ほど含まれており，これが学習データに対する分類正解率が 0.9 から向上しない原因だと考えられる．表 3 に複数の原形を持つ事例と，1 つの原形しか持たない事例を示す．

6. まとめ

本稿では，顔文字の原形推定に関して，Embedding を用いたニューラルネットワークでのアプローチの有用性について述べた．従来手法 [奥村 17] では，ベースラインとしての Cosine 類似度による推定が苦手とする，**llllll(-ë-;)llllll ずーん** のような原形に関係の無い付随パーツが大量に存在する顔文字（元文字列）に対して優秀であるとしながらも，顔文字の数値化に問題があり，分類正解率は 0.51 に留まっていた．

我々の提案手法では，表 4 に示すように，Embedding 層を追加したニューラルネットワークを用いることにより，従来手法よりも少ない学習回数で，優良な分類正解率 0.746 を得た．

表 4　従来手法と提案手法の学習回数と分類正解率

	学習回数 (Epoch)	分類正解率
従来手法	280	0.51
提案手法	70	0.746

今後の課題として，複数の原形を持つ元文字列への対応が挙げられる．現在の FFNN では，1 つの元文字列に対して複数の原形を推定することは不可能であるため，改良が必要である．一例として，新たに元文字列から原形の個数を推定するモデルを構築し，FFNN が原形を推定する際のヒントにする手法が考えられる．また，提案手法の FFNN を用いて原形から原形を推定した結果の Accuracy は 0.066 であり，今後の検討が必要である．

謝辞

本研究は JSPS 科研費 15K21592，高専連携教育研究プロジェクト 課題番号 2309 の助成を受けたものである．

参考文献

[Bedrick 12] Bedrick, S., Beckley, R., Roark, B., and Sproat, R.: Robust kaomoji detection in twitter, in *Proceedings of the Second Workshop on Language in Social Media*, pp. 56–64, Montreal, Canada (2012), Association for Computational Linguistics

[Hall 76] Hall, E. T.: *Beyond Culture*, Anchor (1976)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, Vol. abs/1301.3781, (2013)

[Ptaszynski 10] Ptaszynski, M., Dybala, P., Rzepka, R., and Araki, K.: CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics, *Affective Computing, IEEE Transactions*, Vol. 1, No. 1, pp. 46–59 (2010)

[Ptaszynski 12] Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., Araki, K., and Momouchi, Y.: Speech, Image, and Language Processing for Human Computer Interaction - Science of Emoticons: Research Framework and State of the Art in Analysis of kaomoji-type Emoticons., *IGI Global* (2012)

[Tanaka 05] Tanaka, Y., Takamura, H., and Okumura, M.: Extraction and classification of facemarks, in *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI'05, pp. 28–34, New York, NY, USA (2005), Association for Computing Machinery

[奥村 16a] 奥村紀之 F 分類器による顔文字の原形推定, 信学技報, Vol. 116, No. 379, pp. 93–96 (2016), NLC2016-37

[奥村 16b] 奥村紀之 F 感情抽出のための顔文字の原形推定, 信学技報, Vol. 116, No. 78, pp. 1–4 (2016), NLC2016-1

[奥村 16c] 奥村紀之 F 言語解析のための大規模顔文字辞書, 人工知能学会第 30 回全国大会 (2016)

[奥村 17] 奥村嶺 F ニューラルネットワークによる顔文字の原形推定, 2017 年度 人工知能学会全国大会（第 31 回）(2017)

表 3　推定の難易度から見た事例の一例

	元文字列	原形
1 つの原形しか持たない事例	┌|T_T|┘♪ └|T_T|┐♪ ┌|T_T|┘ **テケテケテッテ** (((-_-(-_-;=(-_-;=;-_-)=;-_-)-_-))) **分身** **ア**（・ ．・）**リ**（・ ．・）**ガ**（・ ．・）**ト**（・ ．・）**ウ**（・ ．・）	|T_T| (-_-) (・.・)
複数の原形を持つ事例	**モジ**(((*Φ ｗ Φ *)(* Φ ３Φ*))) **モジ** **ﾗﾌﾞﾗﾌﾞ**(((*--)￣(**エ**)￣) ♪ **ｳﾃﾞﾏｸﾗ**♪ (*￣-(-*) ヽｽﾔｽﾔ	(**Φ** ３ **Φ**), (**Φ** ｗ **Φ**) (--), (￣ (**エ**)￣) (--), (￣ -￣)

Oral presentation | Organized Session | [Organized Session] OS-19

[2B2-OS-19a] [Organized Session] OS-19

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room B (4F Moon Light)

[2B2-OS-19a-01] How the learning environment in which high-quality opportunities exist for collaborative problem solving skills is linked to fostering ability

○Hiroyuki Masukawa[1], Ikuo Endo[2] （1. University of The Sacred Heart, 2. Shizuoka University）

1:20 PM - 1:40 PM

[2B2-OS-19a-02] Toward Understanding of Embodied Knowledge in Rehabilitation

○Kentaro Kodama[1], Kazuhiro Yasuda[2], Ryosaku Makino[2] （1. Kanagawa University, 2. Waseda University）

1:40 PM - 2:00 PM

[2B2-OS-19a-03] Designing Themes of Regional Revitalization on the basis of On-Site Inquiry by a Method of letting Subjectivity and Objectivity Coexist.

○Haruyuki Fujii[1], Ken-ichi Shinozaki[2] （1. Tokyo Institute of Technology, 2. Nihon University）

2:00 PM - 2:20 PM

[2B2-OS-19a-04] Analysis of the Handling Technique on a Rehabilitation Therapist

○Ikko Miyamoto[1,2], Takayuki Shiose[3], Masaaki Sakagami[3] （1. Ritsurin Hospital, 2. Waseda University, 3. Kyoto University）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

協調的問題解決能力の良質な発現チャンスが埋め込まれた学習環境は能力の育成にいかにつながるのか：発話データからの検討

How the learning environment in which high-quality opportunities exist for collaborative problem solving skills is linked to fostering ability: Examination from protocol data

益川　弘如[*1]
Hiroyuki Masukawa

遠藤　育男[*2]
Ikuo Endo

[*1] 聖心女子大学
University of The Sacred Heart

[*2] 静岡大学
Shizuoka University

This study examined whether collaborative problem solving (CPS) skills can be developed in school through knowledge construction. We collected dialogue data from four lessons and interview and analyzed how students engage in dialogue. As a result, the students recognized that they want to use CPS skills even in a backward learning environment.

1. はじめに

本研究は、2 人の学習者の協調問題解決(CPS)能力の発揮と認識を、その学年時々経験する学習環境との相互作用の関係と共に追った。結果、CPS 能力が発揮しやすい学習環境で学んだことが、必要な学習環境の構築のすがたの意識を持つことを、それぞれなりに持っていると共に、発揮しにくい学習環境ではこれまでの個々人の特性に依存する相互作用の関係にあることが見えた。

対象学習者は、2018 年 3 月現在中学 3 年生の A さんと B さんで、2 人の小学 3〜6 年生の授業中の発話データ、そして中学進学後の 2015 年 11 月にインタビューデータを用いる。小学時代の CPS 能力発揮の変容ついては[益川 16]でまとめており、論文内の児童 A,児童 B が A さん,B さんと対応している。今回インタビュー結果も踏まえ、改めて 2 人の CPS 能力の成長とその多様さを分析し直した。

2. 研究の背景と目的

[三宅 15]は、人が持つ学ぶ力はコンピテンスだという仮説を提示している。そこでは、CPS 能力とは、誰でも潜在的に持って生まれてくる潜在能力と考え、潜在能力が発現されるかどうかは、発現チャンスの質と量に依存するとしている。三宅は「良質な発現チャンスをいつも経験していれば発言されるパフォーマンスの質も高くなる。」と言っている。しかし経験による変容を取り上げるとなると、短期間で一部の被験者の実験結果の比較から抽象化する研究では不可能である。また三宅は[三宅 14]の論文の中で、以下のように発言しており、人の学習研究について、少ないサンプルから意図した情報を抜き出して一般化しすぎている問題点を指摘している。

> *「今までの多くの研究は、こういうふうになるはずですっていう、それが成り立つ場所、それが成り立つ授業、実験っていうのを、簡単に言えばそれを証明する場所を作って、その中でこうなりました、って見せるだけになってしまっている。」*

連絡先：益川弘如，聖心女子大学文学部教育学科，
150-8938 東京都渋谷区広尾 4-3-1，TEL 03-3407-5811，
e-mail masukawa@u-sacred-heart.ac.jp

これは[中村 92]の『臨床の知とは何か』の著書の中でも同様の問題点が指摘されている。この視点をしっかり受け止め、人は一人ひとりが多様であるという前提から、人の学びも多様である点を捉えていくことが重要であろう。同時に、人の認知活動や学習活動は状況に依存するという特性も持ち合わせている。その状況をいかにデザインするかとして、三宅の言う「良質な発現チャンス」をデザインする必要性と対応づくであろう。

現在の学校教育を見てみると、イリイチの『脱学校の社会』でも問題視されているように、学校のさまざまな制度や授業の「型」が独り歩きし、世の中の多くの知に触れる機会が保障される形で、一人ひとりなりに多様な知識や能力を構成していくという視点の教育に必ずしもつながっていない。CPS 能力育成の授業方法も同様の問題点が潜在的に存在する。それは、グループで話し合う場面において、深い理解のために話し合うという目的よりも、うまくその場が回るような指導が優先されてしまうことが多い。そこでは、他者が発言するときには傾聴するように指導され、相手に意見を言うときには「〜だと思います。なぜならば〜だからです。」と言うことを指導されその場での「型」が優先される。そうすると、話し合う内容について自分が話したいタイミングで意見を言い合うような対話や、まだまとまっていないアイデア段階や分からないことがあっても「まずは考えを出してみる」という活動につながらない。そのような経験の蓄積の結果、社会に出て仕事の中でチームの話し合う場面においても、建設的な貢献につながらなく「型」としてしか振る舞えなくなっているのではないかと考えられる。

以上のように、学校という枠の中で行われる授業のさまざまな方法の「型」は、必ずしも「良質な発現チャンス」となっているとは限らない。[白水 14]の整理では、良質な発現チャンスを実現する学習環境として、「学びのゴールを到達したらその先の次のゴールを生むための中継点に置く学びを設計する必要がある」とし「ここまでできれば合格」という設定の仕方をすると、学習者が次のゴールを自分で見つける学びの主体性が損なわれ、学んだはずのことが長続きしないだけでなく「学び方を学ぶ」のが難しいとしており、『21 世紀型スキル』の白書でスカーダマリアらも指摘している。本研究では、前者の学習環境を「前向き授業」、後者を「後向き授業」と名付ける。そして CPS 能

力の発揮に関しては、良質な発現チャンスであれば「課題内容」について建設的相互作用が起きている発話が多い状態とし、そうでなければ「チーム構築調整」といったチームの誰がどの順で発言するかなど場の回し方に終始する発話が多い状態とする。

以上を踏まえ、今回のデータの A さんと B さんは、授業研究の一環で、良質な発現チャンスとなっている前向き授業や後向き授業の学習環境下を経験していたものとなっている。これら学習環境下で CPS 能力の発揮がどうであったか、A さん B さんの成長と共に学習環境と相互作用しながらに変化する姿を追った。

3. 対象学習者と学習環境について

2011 年より、研究の一環で静岡県内公立の H 小学校の授業中の発話を収集しており、それを基にした A さん B さんの対象データの一覧は表 1 である。

表 1　分析対象とした発話データ

学年	学習環境	発話データの種類
小 3	後向き授業	国語「もちもちの木」の対話場面
小 4	前向き授業	算数「概数」の対話場面
小 6	前向き授業	算数「並べ方と組み合わせ」の対話場面
小 6	前向き授業	算数「全体を 1 と見て」の対話場面
小 6	CPS 課題	全国学テ・算数 B も「あた問題」の対話場面
中 1	後向き授業	授業を参観後、インタビューと対談を実施

A さんと B さんが小学校 3 年生のときには、学校の研修方針で話し合い方や発表の発現の仕方といった「話型」を指導され、それを基に班で話し合うことが推奨される「後向き授業」だった。一方、小 4 以降は対話を通して追求すべき問いの解決に向けて深め、各自なりの理解を構成していく「知識構成型ジグソー法」を中心とした「前向き授業」に研修方針を変えた[遠藤 15]。CPS 課題は、全国学力・学習状況調査の算数 B 問題を対象に、2 人一組で解かせることで、小学校時代に学んできた CPS 能力の発揮を検証する実験授業への参加だった。

小 3 の後向き授業では「チーム構築調整」の発話の発現割合が高かったが、小 4 の前向き授業以降「課題内容」の発話の発現割合が高くなっていった。

最初に、小学校 3 年生の国語授業での発話データを紹介する。まず A さんの発話の一部が表 2 である。A さんは班員が順番に発表できるよう、上手くコーディネートすることに力を入れていた。

表 2 A さんの小学校 3 年授業の発話データの一部

Aさん	じゃあ，豆太は臆病だけど，勇気があってやさしい子だと思います．理由は豆太は夜のモチモチの木が苦手だけど、じさまの腹が痛いときに，一人で夜道を走って医者様を呼べるからです．
Eさん	こっちいる？
Aさん	じゃあ，Eさん
Eさん	はい，豆太は勇気のある子だと思います．理由はじさまが腹を痛めて夜道をたった一人で医者様を呼びにいけるほど勇気のある子です．
Aさん	はや，早口ですよね．
Eさん	ごめん．
Aさん	じゃあFさん
Fさん	豆太は勇気のある子だと思います．
Gさん	Fさん
Fさん	理由は，一人で夜道を走っていけるからです．
Fさん	じさまが腹痛の時夜道を一人で隣村の医者様を呼びにいけるから．

次に B さんの発話の一部が表 3 である。B さんは、他メンバーの調整のもと、自分の番で意見を言った後、ワークシートに書かれていた「発表後疑問があったら質問をしよう」という教師の支持を意識して「疑問がある？」と聞いていた。しかし、特に出てこなかった。このように、型に頼りながら班活動を進めていた。

表 3 B さんの小学校 3 年授業の発話データの一部

Bさん	次うち？
Dさん	そうです．
Bさん	臆病だけど優しい子だと思います．理由は夜に一人で医者様をよびに行けたけど，次の番からせっちんによんだからです．
Dさん	はい．
Bさん	何で近づけたのよ-．
Dさん	えー今マックスになってたよ．
Bさん	ねっ疑問がある？
Hさん	疑問．
Bさん	疑問.疑問．
Hさん	今の、今ので疑問．
Bさん	疑問？後ろにあったんだ．疑問．
Dさん	あーあったー．（ふざけながら）
Bさん	ほんとだ．だぜぇ．
Hさん	他の，他の人の意見で疑問
Bさん	疑問
Hさん	疑問を持ったこと
Dさん	わからーん

続いて、小 6 の前向き授業、算数の「全体を 1 と見て」での発話データを紹介する。この授業では A さんと B さんが同じグループで対話していた。発話の一部が表 4 である。

表 4 A さん B さんの小学校 6 年授業の発話データの一部

Iさん	足せば良いんじゃない？
Aさん	足すの？
Bさん	これを全部？
Aさん	これを全部？
Bさん	ぺーぺーぺー、足す。
Jさん	足すの？
Aさん	足すの？
Bさん	足すか？
Jさん	足したらさあ、なんかさあ、全部の面積だからさあ。
Bさん	３１？
Aさん	違うくない？なんか
Bさん	３１辺
Jさん	全部の面積じゃないみたいな
Jさん	（問題上に登場する）３人で
Bさん	全部の面積書く必要なくないか？
Bさん	（問題上に登場する）３人で分担しろってことでしょ
Jさん	うん
Bさん	誰かのを基準にしてそれを割ればいいんじゃない？
Iさん	全部の面積が判ればできるよ
Aさん	誰かのを基準にするか

グループの発話のターンは短く、メンバーの誰かが話す順番をコントロールすることなく、課題内容について言いたい意見を言いながら深めている様子が伺える。そこでは、どう考えたらいいのか仮定の内容も積極的に提案していることが分かる。A さん B さん両方とも CPS 能力を発揮できていたことがわかる。

そして小 6 の CPS 課題では、分析結果から、後向き授業で研修を進めた他校の児童らは「答え合わせ」に終始していたのに対して、A さん B さんペアの話し合いでは対話を通して一緒に「答え作り」をするような CPS 能力の発揮が見られた。

A さんと B さんの小 3 時点からの変容の様子をまとめたのが表 5 である。A さんと B さんは後ろ向き授業では、それぞれ「チーム構築調整」「他者に頼り指示内容は意見す

る」という姿でそれぞれであったが、前向き授業の学習環境の中では、そのような姿は見られなくなり、まだ検討段階の考えであっても、積極的に意見を言い合いながら学びを深めていくことができるようになっていった。

表 5　小学校時代の CPS 能力の変容の様子

A さん	B さん
チーム構築調整が好きでその調整に積極的だったが、前向き授業によって課題内容を深める話し合いが好きになっていった。	他者に頼る形でおとなしい参加だったが、前向き授業によって少しずつ課題内容に対する発言数が増え好きになっていった。

4. 中学進学後のインタビュー結果

2015 年 11 月、A さん B さんを含む 4 人の所属するクラスの授業を参観し、その後 4 人を対象にインタビューを行った。最初一人ずつ中学校での様子を聞いた上で、その後 4 人全員対談形式で実施した。

A さんはインタビューで表 6 のように答えた。

表 6　A さんの個別インタビューの回答

質問者	中学授業はどう？
Aさん	知らないことを知れる、でもつまらない授業。
Aさん	国語の文法は覚えづらい。眠いし苦手。
Aさん	社会好き、覚えるのは苦手だけど、点が取れる。覚えやすい。
質問者	授業のやり方は？
Aさん	一緒（知識伝達型授業）。
質問者	中学校、班活動は？
Aさん	前期の班、苦手な子が多い。班活動そこまでやっていない。
Aさん	後期になってやっと慣れて来た。前期は人見知り。一人でいる事が多かった。
質問者	話し合う機会は？
Aさん	なかった。無言でみんないたんで、言おうにも言えないから。
質問者	機会はあってもみんな話さない？機会自体も少ない？
Aさん	あーそうですね。全然班活動なかったかも。
質問者	中学の授業で先生説明後、それを元に考えるのと、複雑、じっくりを自分や友達と考えるのは、どちらが好み？
Aさん	好み？考えてみた方が良い。私的に。
Aさん	最近慣れて来たと言ったんですけど、そういう風に考えたりすれば友達と話す機会も増えたりしてたんでその方が良いのかなぁと思って。
質問者	中学校、わからないことはどうする？
Aさん	塾の先生に聞く。
質問者	授業中に、先生や友達には？
Aさん	聞かない。
質問者	小学校の頃は？
Aさん	あーなんか、今はそうじゃないんですけど、先生が話をしているときに疑問に思ったことをズバズバ言っていたと思う。
質問者	どの先生にも？
Aさん	どの先生にも言っていた。
質問者	なぜ今は、言わないの？
Aさん	うーん、なんかまだ慣れていない先生もいるせいかも。
質問者	慣れたらいうかも？
Aさん	慣れたらいうかも。
質問者	ジグソー覚えている？
Aさん	笑い。
質問者	それに対しては？
Aさん	うーん、別になんか、感じてなかった。普通にやってた。
質問者	話し合う授業だとやりやすい？
Aさん	楽しかった。
質問者	他も楽しかった？
Aさん	はい。

「知らないことを知れるがつまらない授業。国語の文法は覚えづらい。ねむいし苦手。社会好き、覚えるのは苦手だけど点が取れる。覚えやすい。」と答え、小学校のころについては「今はそうじゃないんですけど、先生が話をしているときに疑問に思ったことをズバズバ言っていたと思う。どの先生にも言っていた。」と答えた。

続いて B さんはインタビューで表 7 のように答えた。中学について「班活動は中学になってあまりないです。説明を聞いていてちょっとスピードが早かったりして説明聞いても難しい」などと答え、小学校のころは「班活動とか多かったので眠くならなかった。班活動は好きだったです。楽しかった。色んな人とかと活動できるので意見が色々出たりして。」と答えた。

表 7　B さんの個別インタビューの回答

質問者	中学の授業は？
Bさん	眠くなります。
質問者	どういう時？
Bさん	先生が話しているばっかの時とかすごく眠くなります。
質問者	教科とかでは？
Bさん	社会が多い。
質問者	どういう授業なら楽しい？
Bさん	社会は話が多いけど、歴史とか好きなんで全然眠くならないし、逆に数学とか。
質問者	眠くなりやすい。
Bさん	はい。
質問者	説明が多いのかな？数学は？
Bさん	なんか人と話すところがあるんですけど、そういうところも,そういうところあっても眠くなる。
質問者	話す時だけ目が覚める。
Bさん	そんな感じです。
質問者	さっき手を上げてなかったね。
Bさん	手はあまりあげないです。
質問者	それは小学校の時から？
Bさん	はい。
質問者	班活動の授業はある？
Bさん	最、中学になってあまりないです。
質問者	わかりやすさは？わかりやすい、説明聞いてその分？
Bさん	ちょっとスピードが速かったりして。
質問者	説明聞いても難しい
Bさん	はい。
質問者	わからない時、授業中聞いたり口に出したりしている？
Bさん	全く。
質問者	しないの？どうやって解消？
Bさん	家で数学と英語だけ通信やっている。そのやつで電話で聞けるやつがあって、それで聞いてます。
質問者	なかなか中学では話にくいことを休み時間に友達とかに聞いたりとか
Bさん	そういうこともします。
質問者	中学と比べて小学校はこうだったって印象は？
Bさん	なんか、班の活動とか、多かったんで眠くはならなかった。
質問者	班の活動は好きだった？それとも
Bさん	いや、班の活動は好きだったです。
質問者	H小ではジグソーを行なっていましたが、覚えていますか。
Bさん	はい。
質問者	どうでしたか？
Bさん	楽しかった。

インタビュー前の授業参観での様子も補足する。

A さんのクラスの授業ではジグソー授業が持ち込み型で行われたが、インタビューで授業のことを聞くと、エキスパート資料 4 種類のうち 2 種類が同質だったため「資料 4 は、いらないと思ったんだよね。正直。（エキスパート活動を）やり終わった後、班に戻ったら、使わなそうな感じだった。」と答え、授業に対しての課題点を指摘した。

B さんのクラスの授業では班活動が行われたが、教師は「まずはせっかく今、紙を配ったので，その中に点を取って下さい。わかりました？じゃあちょっと最初なので班でくっついて，周りで確認しながらやってみましょう。点を取ったら周りで確認して」と作業結果の確認のために班を

作った。その際 B さんは、席を移動しているときに「この班活動意味ないじゃん」と呟き、問いを解決するためではない活動に不満を抱きつつも、教師の指示した活動に取り組んでいた。

これらより学習環境が「前向き授業」から「後向き授業」に変わり、その後ろ向きの学習環境に順応しようと努力している姿もみられつつ、しかしながら、しっかりと小学校時代の「前向き授業」との違いを認識しながらも CPS を発揮しようとしている。このインタビュー結果から見えてくる姿を、それぞれまとめたのが表 8 である。

表 8　中学 1 年生 11 月時点での学習環境の認識

A さん	B さん
疑問は塾の先生に質問。班のメンバーと慣れずに話ができない人間関係に困っている。班活動みんな黙っているのでしゃべり辛い。話す機会を増やしたい。	疑問は通信教育や授業時間外に友達に質問。説明を聞くだけの学習内容の進度に困っている。ただ相談するだけの班活動は意味ない。多様な意見がほしい。

そして、H 小学校出身の生徒 4 人が集まっての対談インタビューの様子の一部が表 9 である。これより興味深いのは、A さんが「同じ H 小の子が一人でもいれば、なんとかいける」、B さんが「H 小出身者が一緒にいる今の班なら平気」と発言するなど、H 小学校の学習環境を卒業した生徒同士であれば、CPS 能力を発揮できると指摘している点である。

表 9　対談インタビューの発話データの一部

質問者	話し合って考えると楽しいって言ってるけど、なんかこうしたら楽しい話し合いが盛り上がる、こうしていこう、などと意識していることはありますか？こういう話し合いをしたら楽しいなぁみたいな。
Bさん	全員意見を出す。意見を出して、お前の違うぞ。
Bさん	それは楽しい。
質問者	突っ込まれるのが楽しい？
Bさん	いや、なんだろう、１対１の討論が好き。
Aさん	出された時、絶対違ってるでしょって討論してるけど。
Bさん	はー、オレのはあってるし。
Aさん	そうだね、絶対こっちあってるし。
質問者	そんなこと言われてなんか、ウーみたいな。
Aさん	いや、なんか、
Bさん	な、なんか、なる人には言わないみたいな。
Aさん	そうだね。
Dさん	もうわかってる。
Bさん	言っても平気な人には言う。
Aさん	ていうかEちゃんと、Fちゃん。
Bさん	Fちゃんには、あーあーみたいな。
Aさん	うん、Gちゃんとかでも、静かな子でも乗ってくれるからいい。
Bさん	だから、H小全員ノリがいいっていうか。
Aさん	そうだよなぁ。
質問者	今の中学、班活動に入ったらどう？
Cさん	M小の子たち次第。
Aさん	できない。無理無理無理。
Dさん	メンバー見て、何も喋んないんだったら、もう一人でずっと考えている。
Aさん	同じH小の子が一人でもいれば、なんとかいける。
Bさん	今の班なら平気（H小出身者が他にもいる班でした）
Dさん	学校でなんていうの？H小、H小、K小、K小だったら、
Bさん	全然いける。
質問者	H小が２人いればなんとか？
Bさん	まぁまぁ。
Aさん	一緒にいてくれれば。

5. まとめ

以上より、小学校時代の「良質な発現チャンス」の積み重ねによって、自分にとっての CPS 能力を発揮するときの理想の姿のモデルを構築することができ、中学授業に対する問題点や効果的な班活動の姿に対する認識をしっかり持っていることがわかる。またこのことによって、主体的に CPS 能力を発揮したいという思いが、中学校の教員が用意した学習環境をどう変えたらより良くなるについての A さん B さんなりの考えを持ち合わせることにつながっているとも言えるだろう。

しかし、良質な発現チャンスのある学習環境であれば、A さんも B さんも質の高い CPS 能力を発揮することができるが、そうでない学習環境下では、小学校 3 年生のときに見せていたような個人の特性が強く出てくることが見えてきた。具体的には、A さんの自分が積極的に発言できる場を強く求めている「人間関係構築」をまず重視する傾向になることや、B さんの今ある環境が意味のある話し合いの場になることを願う「意味ある話し合い」を重視するが受動的な傾向になることである。

6. 今後にむけて

もし、A さん B さんが、自ら自由に学習環境を設計できる状況や立場であれば、「学習環境を構築して CPS 能力を発揮できるのかどうか」が、CPS 能力を育成することができたかどうかを検証するための鍵の一つになるとも考えられるが、今回はそのような機会を作ることは難しかった。もし機会を作った場合、そのために必要な「良質な発現チャンス」の経験の蓄積が、小学校時代で十分だったのか、中学 3 年間や高校 3 年間さらに長期の期間が必要なのか、もしくは、そもそも生涯にわたって学習環境を構築したりされたりの相互作用の関係なのか、またそれら条件というのが一人ひとりの多様性とどう関係してくるのか興味深い点である。

また、小学校時代「後向き授業」を経験してきた学習者が、同じ中学の授業についてどう認識していて、2 人とどう違うのかについても調査しきれていない。

今後も、一人ひとりなりの多様な成長の姿を学習場面での発揮場面を丁寧に追いつつ、主体的な学習を支える学習環境のデザインと学習者との相互作用を詳細に調べ、学校現場の改革に取り組んでいきたい。

参考文献

[遠藤 15]　遠藤育男, 益川弘如: デザイン研究を用いたエビデンスに基づく授業研究の実践と提案，日本教育工学会論文誌，39(3) , pp.221-233, (2015).

[グリフィン 14] パトリック・グリフィン，エスター・ケア，バリー・ マクゴー編著，三宅なほみ監訳，益川弘如，望月俊男編訳 :21 世紀型スキル―新たな学びと評価のかたち，北大路書房，(2014).

[中村 92]　中村雄二郎: 臨床の知とは何か，岩波新書，(1992).

[益川 16]　益川弘如，河﨑美保，白水始: 建設的相互作用経験の蓄積が協調的問題解決能力の育成につながるか―縦断的な発話データを用いた能力発揮場面の分析―，認知科学，23(3) ，pp. 237-254, (2016).

[三宅 15]　三宅なほみ: 三宅なほみ 最後の論文，認知科学，22(4) , pp.542-544, (2015).

[三宅 14]　三宅なほみ，大島純，益川弘如: 学習科学の起源と展開，科学教育研究，38(2) , pp.43-53, (2014).

[白水 14]　白水始，三宅なほみ，益川弘如: 学習科学の新展開:学びの科学を実践学へ，認知科学，21(2) , pp.254-267, (2014).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

リハビリテーションにおける身体知へのアプローチ

Toward Understanding of Embodied Knowledge in Rehabilitation

児玉 謙太郎[*1] Kentaro Kodama
安田 和弘[*2] Kazuhiro Yasuda
牧野 遼作[*2] Ryosaku Makino

[*1] 神奈川大学 Kanagawa University
[*2] 早稲田大学 Waseda University
[*2] 早稲田大学 Waseda University

We discuss how we can approach to embodied knowledge and skills in practical and clinical field of rehabilitation. There is a dilemma in studying them: Within traditional scientific framework, if researchers want to investigate a particular object or phenomenon in real and natural world by means of hypothesis testing experimental way, they should remove other factors apparently unrelating to their hypothesis in order to avoid affecting the result. However, most real and natural phenomena cannot be separated from various factors surrounding them (e.g., environment, situation). In other words, they are embedded in their environment or situation. This dilemma represents some difficulties of approaching to embodied knowledge and skills in practical and clinical field. We pick up related issues and show our approach to study them.

1. リハビリテーションにおける身体知

身体知とは，「身体に根ざした知」，より広い意味では「身体と心と社会のすべてに根ざして存立する知」と呼ばれるものである[諏訪 17]．一見，頭の中や，心の中に在ると考えられがちな知識や知能といったものも，身体に根ざして獲得され，実現され，身体と切り離して捉えることは難しい，という考えがその背景にある．例えば，私たちは人と話す場面でも，思わず身体を動かし使っている[McNeill 92]．また，逆上がりや自転車の乗り方を習得する際，その方法について言語化して説明することは難しいが，身体はその方法を知っており，いつの間にかその技能を獲得している．身体知は，アスリートの優れた身体パフォーマンスにも関わる技能，知能であるが，私たちの日常的な行為の場面でも広く観察される知能の側面，知の在り方だと言えよう．また，身体機能が低下・損失し，それまで通りに自由に身体を動かせない場面で感じる違和感や困難というものも，普段は自覚せず行っていた行為が，潜在的な身体知により支えられていたことを示す例だと言えよう．その意味で，リハビリテーションという臨床の現場では，この身体知の回復や再獲得が目標とされていると言える．

リハビリテーションの定義は様々だが，1 つの基準とされている定義によると，それは「身体的，精神的，社会的に最も適した生活水準の達成を可能とすることによって，各人が自らの人生を変革していくことを目指し，且つ時間を限定した過程」とされている[国連総会 82]．よって，リハビリテーションに携わるセラピストに求められる能力としては，当然，臨床的な評価と介入・訓練に直接かかわる能力が挙げられる．しかし，加えてそのために必要な能力として，対象者とのコミュニケーションを通して信頼関係を築くという社会的能力も含まれるだろう．実際，平成 27 年度に厚生労働省が理学療法士・作業療法士を対象に実施された業務実態調査の結果，セラピストに求められる能力の第 1 位は「コミュニケーション能力」，第 3 位が「協調性」であり，リハビリテーションの現場でも他者とのコミュニケーションや協調の重要性が再認識されている[厚生労働省 16]．ここでは，そのような社会的能力も含め，リハビリテーションにおける身体知と呼ぶ．そこでは，セラピストが対象者（患者）と言語・非言語など，様々なレベルでのインタラクションを行いつつ，先述の目標に向けて，必要な情報を知覚し，行為を調整する．身体知という意味では，それらの情報は必ずしも言語，非言語と分けられるものでもなく，明示的に記号化されたものも，視覚的・聴覚的な非言語情報（表情や音声）を介して表現され，知覚されると考えられる．また，この相互行為は，セラピストから対象者への一方向的な作用だけでなく，対象者からセラピストへの言語的・非言語的な情報の伝達，共有の側面も有している．そのため，セラピストには，対象者の身体的状態，心理的状態を含めて知覚・評価し，情報を引き出す能力も求められる．

連絡先：児玉謙太郎，神奈川大学，横浜市神奈川区六角橋 3-27-1，045-481-5661（内線：4752），kkodama@jindai.jp

2. 問題の所在：アプローチの難しさ

ところが，このような身体知を科学的に解明することには，いくつかの問題や難しさがある．

1 つは，リハビリテーションに限らず，実践の現場，フィールドを研究対象としようとしたときに生じる生態学的妥当性に関する問題である．すなわち，従来の科学的方法論においては，現象を再現し，仮説を検証するために他の要因（剰余変数）が統制された実験室的な環境で，実験者により変数が操作され，その結果として対象の変化や違いが計測・観察される，という手段がとられる（図 1）．ところが，多くのリアルな現象は，環境や状況，他の要因と分離できないかたちで作用し合っている（図 1：X や Y は環境や文脈に埋め込まれた他の要因 A, B, C と相互作用していると考えられる）．よって，実験室的環境は生態学的妥当性が低く，実世界のフィールドで観察される現象を必ずしも再現できるとは限らない．

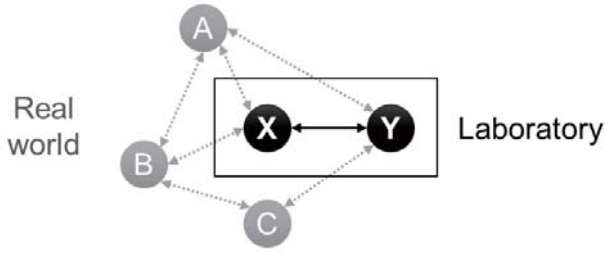


図 1 実験室的環境の問題

ここでの問題は，現象の文脈依存性に起因するとも言えよう．現象をリアルに観察し，ありのままの姿で記述しようとするならば，先述した近代科学的な方法論（実験）のように対象をその環境や文脈から切り離す方法は適切でないし，実験的研究でない場合であっても，日常生活環境から離れ，データの収集・計測のために実験室的環境に対象を移し，再現する場合も同様の問題に直面する．例えば，コミュニケーション場面の会話を対象とするような場合でも，会話参与者は自分たちがデータ（対象）として収録されていること自体に言及した会話（データ）が収録されることが指摘されている[牧野 17]．実際，ヒトの認知や行動が文脈に対し極めて鋭敏であることが知られている[Kloos 09]．

また，2 つめの問題として，リハビリテーションなどのフィールドのデータが，古典的な統計の仮定に基づくとは限らないという問題がある．まずは，対象となる集団の分布のレベルでの正規性である（図 2 上）．すなわち，リハビリテーションの現場で対象となる患者やその疾患の多様性，個別性を考えると，必ずしも古典的な統計が仮定しているような正規分布に従わないという問題である．仮に正規分布に従う場合であっても，その平均的な患者像，疾患の程度が重要とは限らず，むしろ，そこから外れる特殊なケースとして，個々のリアルなケースに向き合うことを臨床では求められるかもしれない．臨床研究で N 数の少ない事例研究が重要であるのも同様の事情によるのではないだろうか．また，臨床家のセラピストの能力を考えても，熟達したセラピストほどそのスキルには多様性や柔軟性があると考えられる．つまり，熟達したセラピストほど，多様なケースに柔軟に対応できる能力を有していると可能性がある．そのため，特定のスキルによってセラピストの能力を一意に記述できるとは限らない．

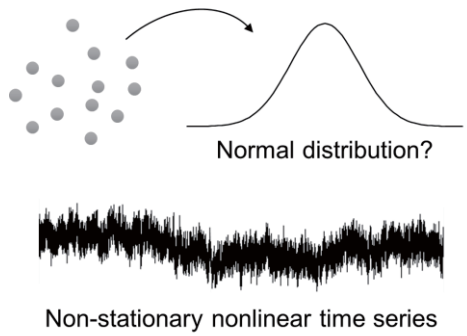


図 2 フィールドデータの特徴

さらに，多様性・個別性が高い対象のデータ（何らかの行動データ，生理データなど）についても，その時系列が定常的で線形とは限らない（図 2 下）．むしろ，様々な要因や疾患が合わさった複雑でノイズを含むデータ，系列依存性のあるデータである可能性が高い．その場合，定常性・線形性を前提とした基本統計量や線形時系列解析では適切にデータの特徴を抽出できない．そのため，フィールドデータを古典的な統計や線形時系列解析で分析しようとしても，リアルな現象のありのままの姿を抽出できる保証がない．

3. 関連研究

3 章では，2 章で指摘したような問題を踏まえ，知覚や運動，認知科学の分野で議論され，展開している理論的アイデア，枠組みを部分的に導入し，関連する先行研究について言及する．

3.1 Gibson の知覚理論

生態心理学[Gibson 66]では，心や知能を，脳だけに還元するのではなく，脳を取り囲む身体，環境，状況との相互作用の過程で創発するとみなす．20 世紀中ごろ，知覚の心理学として着想された生態心理学の知覚論は，直接知覚論と呼ばれた．一方，当時，1960 年代にコンピュータ（情報処理）のメタファーで知覚を含む人間の知能を説明しようとした主流の知覚論は，間接知覚論と呼ばれた．直接知覚論では，中枢での表象（記号）の操作・処理を介さずとも環境内の情報を能動的に探索することで直接知覚できると考える．そのため，古典的な AI が直面するフレーム問題などの課題に対しても示唆を与え，ロボットや AI の研究領域にも影響を与えた考え方である．

ここで重要な点は，生態心理学において“情報”は個体の脳内ではなく，環境内に実在し，環境やその中にある対象，そして，知覚者自身を特定するということである．また，その情報を知覚するためには知覚者は環境を能動的に探索することが必要であり，すなわち，知覚と行為とが分けられないということである．また，知覚される情報には，知覚者自身の情報をも含むため，個々の個体の能力などの特性（ある種の個別性）を含むかたちで，行為が可能か否かといった意味や価値（アフォーダンス）が環境の側に実在すると考える．このような個体と環境の相補性，知覚と行為の不可分性という考え方は，コミュニケーション研究やリハビリテーションの実践に対しても示唆を与える．アフォーダンスの観点から捉えれば，リハビリテーションにおいて，セラピストは患者の環境を構成する一要素であり，課題や環境のデザインを工夫することで，患者の行為の可能性を調整できる可能性がある[樋口 17]．

また，生態心理学では，個体間のコミュニケーションの基盤も環境に実在する知覚情報に求める[Gibson 66]．すなわち，環境内の情報を知覚するためには，個体は他個体を含む環境との相互作用を通し情報をピックアップする必要があり，その過程で他個体と情報の共有が可能となり，コミュニケーションが成立すると考える．また，言語という記号を介したコミュニケーションにおいても，光や空気など物理的で知覚可能な媒質を介して他個体で情報の共有が行われる．これらの点を踏まえるならば，コミュニケーションの基盤として，知覚情報とそれをピックアップする知覚システム（身体）の存在は無視できない．

3.2 Bernstein の運動理論

身体という極めて複雑なシステムにアプローチした研究者の一人が旧ソビエトの運動生理学者 Bernstein である．彼は，身体を制御するという観点からとらえた場合の複雑さとして，自由度と文脈の問題を指摘した[Bernstein 67]．ヒトの身体は，関節レベルで約 10^2 個，筋レベルで約 10^3 個，細胞レベルでは約 10^{14} 個のオーダーの自由度によって構成されている[Turvey 90]．自由度問題とは，これら膨大な数の自由度を，トップダウン型の運動制御モデルで説明しようとした場合，脳は膨大な数の変数を事前に決定しなくてはならず，現実的ではないという問題である．また，文脈の問題とは，システムの要素（自由度）同士の複雑なつながり，相互作用を考えると，トップダウン型モデルで考える入力に対し，出力が一意には決定されず多義性を有するという問題である．これら自由度と文脈の問題は，制御理論の観点からは不良設定問題と呼ばれる問題である．

これに対し，Bernstein 自身は，シナジーというアイデアを提案した[Bernstein 67]．シナジーとは，ある運動課題を達成する際に，身体システムの要素（自由度）同士が連携，協調することで，運動の自由度を減らすような機能的な構造・単位である．また，シナジーは，その都度，直面した課題に対して特定的に，一時的に組織化される．その意味で，シナジーは文脈依存的に柔軟に組織化されると考えられている．Bernstein 自身は，筋群，筋-関節の間の結合を想定したが，その後，シナジーの概念は

拡張的に解釈され，関節間など身体システム内部の結合から，身体-環境システム，個人間システムなど広く適用されてきた．

Rileyらが提案するInterpersonal synergy（個人間で形成されるシナジー）の考えによると，個体は互いに知覚的な相互作用をリアルタイムで行い，互いの行為を調整し，1つのシステム（シナジー）を組織化する[Riley 11]．その際，個体を取り囲む環境や状況として，他個体や課題の存在（言語・非言語情報）が行為を制約し，自由度が縮減され，また，システム外部からの外乱やノイズに対し，個体間で相互に補償し協力し合うことにより，直面した課題を達成するようシナジーが組織化される．リハビリテーション場面について考えれば，患者とセラピストは1つの課題を達成するよう相互に補償し協力し合うシナジーとみなせる．そして，運動機能回復の過程で，両者の関係も変化していくのではないかと考えられる．

3.3 自己組織化理論

自己組織化は，雪崩や台風といった自然現象から，ホタルの集団発光といった生物間コミュニケーション，そしてヒトの運動，知覚，認知，コミュニケーションまで幅広く観察される現象である[Strogatz 03]．それらの現象（システム）は，マクロレベルで観察される秩序・パターンがミクロレベルの構成要素間の相互作用によって創発すると考えられる．主に1970年代，物理学の分野で様々な理論，モデルが提案され，知覚や運動の研究領域においても，1980年代以降，先述の生態心理学やBernsteinの運動理論と結びつき，展開している[児玉 13]．

コミュニケーションの研究領域では，コミュニケーション自体を多自由度複雑系（複数の個体からなるシステム）の振る舞いとし，知覚情報を介した身体的・記号的な相互作用によって心理・社会レベルのシステム全体（マクロな秩序や構造）が創発するとみなす[Dale 13][Richardson 13]．その際，新しい概念だけでなく，非線形時系列解析を応用することで基本統計量（平均など）では抽出が困難なシステムの動的な特性を視覚化・定量評価し，複雑な現象の理解が進められている[Shockley 15]．

3.4 関連する先行研究

コミュニケーションの研究領域では，古くから個体間で身体の動きや音声が同時に変化する現象（協調・同期）は観察されてきた[Kendon 70]．それらは相手への信頼[Bernieri 88]，好感[Hove 09]などと関係し，コミュニケーションにおける身体協調が積極的な役割・機能をもつことが示唆されてきた[長岡 06]．

また，リハビリテーション分野におけるコミュニケーション研究の多くは，言語コミュニケーションに着目しており[Roberts 13]，身体レベルの協調など，非言語コミュニケーションがリハビリテーションの効果に及ぼす影響は充分に解明されていないのが現状である[Jesus 16]．また，限られた先行研究の多くは質問紙データの相関分析であり，数少ないリハビリテーション場面のコミュニケーション「過程」を分析対象とした研究も，音声・ビデオデータの質的研究に留まっている[沖田 02]．

一方，自己組織化理論に基づくコミュニケーション研究では，要素間での言語的・非言語的なダイナミックな相互作用過程を通して協調パターンが創発すると捉える[Dale 13][Kelso 95][Riley 11]．そこでは，多自由度複雑系である個体が，知覚情報を介して相互作用する過程で，個体間で身体や認知レベルでの協調が起こると考えられている．また，これらの研究では，新しい概念や理論的な枠組みを提供するだけではなく，個体の生理的・行動的変化（心拍・姿勢など）を計測し，その時系列データに非線形解析を応用することによって，基本統計量（平均など）では抽出が困難なシステムの動的な特性を定量評価し，身体的なレベルの現象と心理・認知・社会的因子の相互関係にアプローチする[Shockley 15]．非線形解析では，データの定常性・線形性を仮定しないため，それらを前提とする線形解析に比べ，リハビリテーションなどの現場における多様な要因やノイズを含む複雑なフィールドデータに対しても有効だと考えられる．

4. 本研究の位置づけと目的

本研究の目的は，個人間の身体協調がリハビリテーションの過程で，運動機能の回復や，運動の支援にどのように影響するかを検証し，そこでどのような情報（言語・非言語，知覚情報）が利用されているかを調べることである．具体的には，リハビリテーション過程での患者-セラピスト間で生じる相互作用（歩行訓練など）における両者の手や足，重心といった部位間の身体協調の程度や，Leader-Follower関係を量的・質的分析を行う計画である．本研究は，現在，計画中の段階であるため，以下では，その計画と展望について述べる．

3章で述べたように，リハビリテーション場面での患者-セラピスト間のコミュニケーション場面における言語・非言語的やりとりを対象とした研究は限られている．そのため，上記の研究目的を達成させるためには，仮説の生成から積み重ねていく必要がある．そこで，本研究では，探索的な仮説生成の段階では，リハビリテーションのフィールドの生態学的妥当性が高いデータの収集を目指す．そこでは，研究目的で掲げた現象を説明可能な要因や変数の発見，特定が目標とされる．そのため，生態学的妥当性が高く，より自然でリアルな実践，臨床のデータを対象とし，質的な観察，記述，分析に重点を置くことになろう．また，同時にカメラやセンサ機器も導入し，可能な範囲で客観的で量的なデータも収集する．さらに，仮説が生成され，要因・変数が絞られた仮説検証の段階では，ある程度，実験的な統制のもと，実験室的環境でデータを収集する．これらの研究の理論的な背景にはGibsonやBernsteinのシステム論的な視点が，データ解析においては自己組織化理論に基づくツールが適用される．

図3は，フィールドと実験室とで収集されるデータの特徴の違いを模式的に表している．左に行くほど，生態学的妥当性が高いフィールドのデータとして表され，より質的な研究の重要性が高い．一方，右に行くほど，生態学的妥当性は下がるが実験的統制がしやすい実験的なデータとなり，量的な記述や分析が威力を発揮することとなる（図3）．ただし，ここで重要なことは，如何にしてこれらを位置づけ，接続・融合していくかであろう．

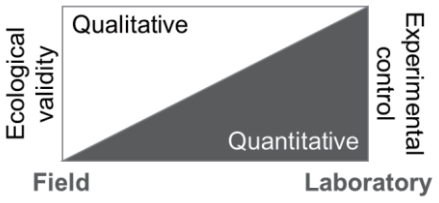


図3 フィールドと実験室の特徴

5. 今後の展望：まとめにかえて

本稿では，リハビリテーションという実践的な臨床のフィールドで営まれる活動を研究の俎上にあげることの難しさを挙げつつ，それにアプローチするための研究の枠組みや方法論を述べた．著者らは，その中でも，研究の仮説を生成～検証する過程で，主観的～客観的，質的～量的といった視点の柔軟な切り替え，方法の融合が重要だと考える．臨床などの複雑な現象を対象とする場合，仮説を生成するためには質的で（ある意味）主観的な視点からのアプローチも重要であり，一方，仮説を検証するた

めには量的で客観的なアプローチが重要となる. そして, それらのアプローチを対話的に循環させていくことがポイントになると考えている[Kodama 17a]. その循環的な過程で研究の問いが深まり, 理解が広がるからである. リハビリテーションではないが, 著者らはこのような研究の循環を試みている[Kodama 17b].

しかし, 実際には様々な課題が考えられる. 1 つは, 先述したフィールドでデータを収集するための機材(カメラやセンサ)が対象者やその振る舞いに与える影響である. とくに, 身体動作を計測するには, 侵襲性の高い機器を用いなければ, 精確なデータを取得できないことが多い. この問題については, 現在, 小型化・無線化も進んでおり, また侵襲性の低い画像処理技術を駆使したモーションキャプチャも開発されているため, それらの技術的発展に期待したい.

また, 本稿では, 複雑でノイズを含むデータに対して, 非線形解析手法の有効性を述べた. しかし, 非線形解析はその背景にある理論の理解, 計算のためのプログラムの準備, そして, 計算結果として算出された指標の解釈が難しいという問題がある. とくに, 算出された数値が実際の現象のどのような側面, 特徴を指標するのかについては慎重に吟味する必要がある. そのためには, フィールドデータに非線形解析を適用するだけではなく, 指標と現象の関係を検証するための実験的研究も必要であり, 慎重に理解と応用を進めなくてはならない[Riley 16]. 著者らは, 同様のモチベーションから, 実験的に指標されると仮定されている現象の側面を操作し, 非線形解析によって求まる数値と現象の対応の検証も進めている[児玉 2018].

そして, 最も重要な課題は, 研究の対象となるフィールドを確保することである. 臨床の現場では, とくに協力機関や対象者の理解が不可欠であり, 研究者と臨床家との対話的なコミュニケーションが必要である. そして, アプローチが異なる研究者同士(例えば, 質的研究と量的研究)のコミュニケーションも重要であり, 共通の目標に向かい, 互いの立場を尊重しつつ, 対話していくことが求められるであろう.

参考文献

[Bernieri 88] Bernieri, F. J., Reznick, J. S., & Rosenthal, R.: Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, (1988).

[Bernstein 67] Bernstein, N. A.: *The Co-ordination and regulation of movements*. Pergamon Press Ltd., (1967).

[Dale 13] Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C.: The self-organization of human interaction. In *Psychology of Learning and Motivation - Advances in Research and Theory*, 59, 43–95, Academic Press, (2013).

[Gibson 66] Gibson, J. J.: The senses considered as perceptual systems. Praeger, (1966).

[Hove 09] Hove, M. J., & Risen, J. L.: It's all in the timing: interpersonal synchrony increases affiliation. *Social Cognition*, 27(6), 949–960, (2009).

[Jesus 16] Jesus, T. S., & Silva, I. L.: Toward an evidence-based patient-provider communication in rehabilitation: linking communication elements to better rehabilitation outcomes. *Clinical Rehabilitation*, 30(4), 315–328, (2016).

[Kelso 95] Kelso, J. A. S.: *Dynamic patterns: the self-organization of brain and behavior*. MIT Press, (1995).

[Kendon 70] Kendon, A.: Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32, 101–125, (1970).

[Kloos 09] Kloos, H., & van Orden, G. C.: Soft-assembled mechanisms for the grand theory. In *Toward a new grand theory of development? connectionism and dynamic systems theory reconsidered*, 253–267, (2009).

[Kodama 17a] Kodama, K., Hori, K., Massaki, H., & Matsui, H.: Dynamics of bodily coordination of therapist-client during a psychotherapy session: Applications of video image processing and time series analyses. In *JSMMR 2017 annual conference proceedings*, 94, (2017).

[Kodama 17b] Kodama, K., Kikuchi, Y., & Yamagiwa, H.: Whole-body coordination skill for dynamic balancing on a slackline. In *JSAI-isAI 2015: New Frontiers in Artificial Intelligence*, 528–546. Springer, (2017).

[McNeill 92] McNeill, D.: *Hand and mind: what gestures reveal about thought*. University of Chicago Press, (1992).

[Richardson 13] Richardson, M. J., Dale, R., & Marsh, K. L.: Complex dynamical systems in social and personality psychology theory, modeling, and analysis. In *Handbook of research methods in social and personality psychology*, 251–280, (2013).

[Riley 11] Riley, M. A., Richardson, M. J., Shockley, K. D., & Ramenzoni, V. C.: Interpersonal synergies. *Frontiers in Psychology*, 2(MAR), 1–7, (2011).

[Riley 16] Riley, M. A.: Private communications, (2016).

[Roberts 13] Roberts, L. C., Whittle, C. T., Cleland, J., & Wald, M.: Measuring verbal communication in initial physical therapy encounters. *Physical Therapy*, 93(4), 479–491, (2013).

[Shockley 15] Shockley, K., & Riley, M. A.: Interpersonal couplings in human interactions. In *Understanding Complex Systems*, 399–421, (2015).

[Strogatz 03] Strogatz, S. H.: *Sync: How order emerges from chaos in the universe, nature, and daily life*. Hachette Books, (2003).

[Turvey 90] Turvey, M. T.: Coordination. *American Psychologist*, 45(8), 938–953, (1990).

[沖田 02] 沖田一彦, 菅原憲一, 越智淳子, & 鶴見隆正.: 理学療法現場におけるコミュニケーション分析のすすめ. *理学療法ジャーナル*, 36(9), 683‐688, (2002).

[厚生労働省 16] 厚生労働省: 地域保健総合推進事業「自治体所属する理学療法士及び作業療法士の業務実態調査」, (2016).

[国連総会 82] 国連総会. 世界行動計画 (1982).

[児玉 18] 児玉謙太郎, 清水大地, & 牧野遼作.: Leader-Follower 関係の定量化手法の検討 ー個人間協調運動課題における予備的検討ー. *信学技報*, accepted, (2018).

[諏訪 17] 諏訪正樹.] 身体知という研究領域. *人工知能*, 32(2), 215‐217, (2017).

[長岡 06] 長岡千賀.: 対人コミュニケーションにおける非言語行動の 2 者相互影響に関する研究. *対人社会心理学研究*, 6, 101‐112, (2006).

[樋口 17] 樋口貴広, 和泉謙二, & 真下英明.: *知覚に根ざしたリハビリテーション*. CBR, (2017).

[牧野 17] 牧野遼作, 阿部廣二, 古山宣洋, & 坊農真弓.: 会話における"収録される";ことの多様な利用, *質的心理学研究*, (16), 25‐45, (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

主観性と客観性を共存させる方法による臨地探究に基づく地域創生デザイン課題のデザイン

Designing Themes of Regional Revitalization on the basis of On-Site Inquiry by a Method of letting Subjectivity and Objectivity Coexist

藤井 晴行[*1]
FUJII Haruyuki

篠崎 健一[*2]
SHINOZAKI Ken-ichi

[*1] 東京工業大学
Tokyo Institute

[*2] 日本大学
Nihon University

This paper discusses the understanding embedded in the traditional form of vernacular houses and the landscape composed of those houses and other traditional forms. The authors introduce “Photo Diary” - a form of expressing findings in field-works and a process of associating findings with each other. “Photo Diary” is composed of a photograph expressing a finding in a field-work, the description of remarkable facts taken in the photograph, the description of interpretation of the facts, and the observer’s first-person experiences of the facts. Every photo diary is associated with each other to construct a whole structure of the findings. The authors have been exploring the spatial schema underlying the form of vernacular houses. On the way of exploring, we have been accumulating pieces of findings. This paper explains a process of constructing design inquiries for regional revitalization of Izena Village, Okinawa, where traditional Ryukyu houses and landscapes are remaining.

1. はじめに

デザインはものごとのありたき姿やあるべき姿を描いてそれらを実現させる方法を考案する行為である。この行為には客観的な知識に基づく合理的な思考だけではなく主観的な信念に依拠する直感的な判断がともなう。本報告は伝統的な様式を継承する民家が多く現存する沖縄離島（伊是名島）の集落（伊是名区）を＜臨床＞のフィールドとする「空間図式の探究」とそこから発展した地域創生デザインの課題デザインのプロセスを事例として、空間図式探究という研究から地域創生デザインという実践に至る経緯、主観性と客観性が共存するデザイン・プロセスを独善に陥らないように進める方法として筆者らが構築しつつある写真日記の作成と構造化という発想法の適用の実際について述べるものである。写真日記は写真、事実の記述、事実に関連して考えたり気づいたりしたものごとの記述、写真日記作成者の経験の記述からなる臨地調査の記録である。写真日記をできるだけボトムアップで構造化する過程で新たな気づきが生まれたり潜在的な問題意識が顕在化されたりする。この過程を通して客観的な事実と主観的な考えから＜臨床の知＞[中村 1992]が顕在化されていると実感している。

私たちが探究している＜臨床の知＞は伝統的な民家や集落の形態に埋め込まれている知である。伝統的な民家の様式はそこに住まう人と風土や社会との良好な関係をつくる仕方が長い年月に渡る試行錯誤を経て形成されてきているものである[和辻 1931]。このような＜臨床の知＞は私たちが、生活空間を形成して、生きる上で重要な役割を担っている [中村 1993]。伝統的な民家の形態や集落景観の姿は近代的な建築技術の発展によって変容しつつある。生活の快適性や安全性、築造や運用の効率性や経済性に注目すれば、これらの変容は自然な成り行きであり、生活の質の向上のために変容されざるを得なかったある時代の「伝統的」な形態を現代の生活に適合させることは合理的ではないという主張には一理ある。一方、機能本位の観点によって、それらでは測れない生活空間の象徴的な意味や風土や文化に根ざした固有の質は消滅しつつある。「近代科学の発展によって、そのために見えなくなってしまったもの[中村 1993]」、「人間の知恵として、人間の能力として持っていながら、近代科学のあまりにも見事な成果に目を奪われてしまい、見えなくなったもの[中村 1993]」が、伝統的なものごとに潜在しているとすれば、それらの＜臨床の知＞を顕在化し、現代の住意識と対応づけることによって、近代的な技術の導入によって存亡の機に直面している伝統的民家や文化的景観とそこに埋め込まれた知を継承する術を見いだせるのではないかと考えている。

連絡先：藤井晴行，東京工業大学，〒152-8550 東京都目黒区大岡山 2-12-1-M1-37，fujii.h.aa@m.titech.ac.jp

2. 方法

私たちが 2014 年 9 月から今日に至るまで伊是名島で継続している「空間図式の探究」を＜臨床の知＞という視点から再考する。「空間図式の探究」は学術的な関心に端を発する調査・研究であるが、伊是名村の日常生活に参与する期間が長くなるにつれて、調査・研究の成果を地域創生のためのデザイン活動に活用しようという意識が強くなっている。

2.1 本研究で注目する＜臨床の知＞の眼差し

デザインに関連づけて語られる＜臨床の知＞[中村 1993]は近代科学の普遍性、論理的一義性、客観性によって捨象される＜コスモロジー＞、＜シンボリズム＞、＜パフォーマンス＞に光をあてている。＜コスモロジー＞は個別の「意味に充ちた空間」への眼差し、＜シンボリズム＞は「機能本位ではなくて、もっと意味や固有の質を持った空間を生かそう」という多義性、多面性への眼差し、＜パフォーマンス＞は身体を環境に曝すとともに環境からの働きかけを蒙るということへの眼差しである。

2.2 経緯

2014 年 9 月の最初の訪問から直近の 2018 年 3 月の伊是名村での調査報告会までの 3 年半の間に 20 回の臨地調査を行

っている。探究は、個別の探究課題の探究と、通奏的な課題発見や深化のための探究の同時進行による。前者は例えば、集落民家の実体的な構造や空間構成の整理から民家の変遷過程を推論するといった視点や方法を限定した探究であり、後者は写真日記によるフィールドにおける発見的探究である。探究が相互に影響し合うにより、探究の視点や内容が変化する。また、生活と空間構成についての議論のために、民家の空間構成や家具の配置の実測調査と住まい方についての住み手の語りの採集を継続している。

これらの調査によって得られた成果を 2016 年 3 月に伊是名村の教育委員会に報告したことから、伝統的な民家や文化的景観が残る伊是名区後辺(くしひん)地区の全民家(49 棟)を官学共同で調査する機会を得る。この調査を遂行するための研究サテライトとして、2017 年 4 月より、伊是名区前浜(めーひん)地区の民家(空家)を伊是名リビング・ラボ(臨地研究室)として利用している。大学院生 2 名がほぼ常駐して生活し、伊是名村の祭祀や行事に参与し始めたことから、私たちと「空間図式の探究」に関する伊是名村の居住者や伊是名村役場の関心が高まっている。これらの方々との関わり方が深くなるにつれて、私たちは村の居住者に近づいた視点を持つようになっている。

上記の調査を含むこれまでの探究の成果を 2018 年 3 月に村役場に報告したことを契機として、後辺地区の伝統的民家と文化的景観の生きた生活の中にある継承と再生、及び、地域創生を実現する方法の計画と実施に協力することを村から求められている。

2.3 写真日記

＜臨床＞のフィールドにおいて、その場所を経験し、気づいたものごとを記録する写真日記を作成し、写真日記を構造化することを繰り返すことによって、空間図式を顕在化するという方法を実践している[篠崎 2015、福田 2015]。場所を経験するということは空間図式を復号化することであり、写真日記を作成することは場所の経験を記号化することである。

写真日記は実際の場所を経験して気づいたものごとを視覚的な情報(写真)とことばによって記録する媒体である[藤井 2015]。写真と三種類の記述(事実記述, 解釈記述, 経験記述)からなる。写真は場所の実体的な構造を映す。ことばでは記述できない情報を視覚的に記録したり、自覚していない情報を自覚している情報と一緒に記録したりする。ことばによる三種類の記述は、場所を経験によって気づいたものごとを表現する、経験者がいる場所に関する事実、事実と経験とを関連づける日記作成者の解釈、経験したものごとを表わす。

写真日記によって表現される気づきはある場所を経験するという事実を契機として生まれる気づきである。必ずしも客観的な事実から論理的に導かれるものごとではなく、主観的であり、非論理的に発想されるものごとである。写真日記において三種類の記述を用いることには、客観的事実、一人称の(主観的な)経験、客観的な事実と主観的な経験を橋渡しする客観的な法則性または主観的な解釈を明示的に区別して扱うという意図がある。そのことによって場所の経験における気づきや写真日記の構造化の過程で生じる気づきを観察可能な事実に設置させ、主観性と客観性を共存させている。

3. 地域創生デザインの課題

これまでに発想された地域創生のデザイン課題を＜臨床の知＞に関連づけて説明する。

3.1 普遍性に駆逐された固有性に再び意味を持たせる

かつては利用されていた人工物が利用されずに放置されているのは、利用されていた当時にはその人工物に与えられていた意味が、現在は、失われてしまっているからであろう。石垣や繧紛やフクギには強風から家屋を守る役割を担うという意味があった。その役割の意味は、強風に耐え得る今日の家屋にとっては、必ずしも強くはない。地下水を汲み上げる井戸の意味も上水設備によって弱められている。しかし、それらが持つ全ての意味が消滅したわけではない。それらは、琉球地方の伊是名の特徴的な景観を形成するという意味を持つし、それらが活躍していた時代から今日に至る軌跡を想起させるという意味を持つ。これらの意味にさらなる価値を見出し、安全性や利便性の観点からは、現在は放置されてしまっているものごとを利活用する方法を考える必要がある。

琉球地方に固有の文化的な景観は琉球という風土における人々の日常生活に根ざしているとともに、琉球地方の人々の人と風土との関係についての自己了解が具現化されている。

3.2 固有の文化と普遍的な要求を共存させる術を探す

琉球の伝統的な民家のファサードの開放性は改築によって変化しつつある。例えば、木製の障子や雨戸をアルミサッシュのガラス戸に変更したり、一番座の掃き出しの開口部を腰壁のある開口部に改造したり、一番座の前の雨端をコンクリート・ブロックの壁で囲ったりというような改築の事例が散見される。改築がなされても二番座の前は、アルミサッシのガラス戸であっても、開放されており、いひゃーじゅーての習慣は続いている。一番座を寝室として用い、そのためのプライバシーを確保しようとする住意識と二番座の前を訪問者に対して開放しようとする住意識が共存している。安全性や快適性などの住宅性能を向上させることは大切なことである。しかし、その手段としてアルミサッシュやコンクリート・ブロックのような普遍的な建築部位を不用意に適用することは、伊是名の生活様式や民家の形態として具現化されている琉球の文化の固有性を失わせるかもしれないという危険性を孕んでいる。

建築部位の改築によって意図されていることを充足し、かつ、伊是名の生活や民家の好ましい固有性を継承し得る技術を開発することが課題である。軽微な改築においては。住民が確認申請を受けずに自分の裁量で材料や構法を選択する。この選択が住意識を充足する費用対効果の観点からなされることは想像に難くない。このような選択をしたとしても、風土や文化に根ざした景観が形成され続けるような材料や構法を考案することが一つの方向である。

3.3 普遍の技術を加えて固有の知を更新する

伝統的なものごとの姿は琉球地方に固有の知の現われである。家屋や集落を現代の生活に適合させるために導入された住宅設備や建築材料は、琉球地方の家屋や集落のために独自に開発されたものではなく、日本全国のどこでも用いることができるように開発されたものである。これらが同じ空間に同時に存在することが違和感を生んでいると考えられる。例えば、屋敷囲の石垣やコンクリート・ブロックは家屋を強風から護るために設置されている。かつては海から採取してきた珊瑚石を積んで築造されていた石垣に籠められた家屋を風から護るという意図が工場で生産されたコンクリート・ブロックの塀を築造することによって充足されるようになってきている。石垣の採取が禁止されたこと、コンクリート・ブロックが登場した当時には近代を象徴するものであり、石垣よりも施工が容易で堅牢であると売り込まれた

ことなどが、関与していると考えられる。一方では、石垣が形成する景観を継承しようという動きもある。

培ってきた伝統的な＜臨床の知＞が普遍的な技術によって駆逐されつつある状況であると解釈することも可能であるし、伝統的な＜臨床の知＞にこの地域にとっては外来の普遍的な技術を付加して新しい＜臨床の知＞を創造する過渡期の状況であると解釈することも可能である。意図を充足する手法が、外来のものであっても、風土や文化と適合するように姿を変え、伝統的な＜臨床の知＞と融合していくことによって、違和感は次第に解消されていくのかもしれない。

3.4 固有の知と普遍性の知との好ましい重なり方を探る

家屋、石垣、屋敷林、電柱、電線などは生活に適した環境を形成するために設置された人工物である。伝統的な様式の家屋や石垣や屋敷林は琉球地方や伊是名地域に固有の景観を生み出している。この景観が好ましいと感じられるのは、それが、長い時間をかけた試行錯誤を経て獲得された、この風土で生活する環境を創出するための知の現れであることを無意識のうちに感じ取っているからではないだろうか。このような景観に電柱や電線が現れると、風土に根ざした文化的景観を愛でようとする眼には、違和感があるものに移る。その理由は、一方、電柱や電線は日本のいたるところで用いられている、日本では固有性のない人工物であるからではないかと考える。

地域の生活に根ざした固有性のある＜臨床の知＞の現われと地域を限定しない普遍的な知の現われが意味の上では共存できないまま、視覚的に具象化された姿で重なっているからではないだろうか。

4. おわりに

私たちが 2014 年 9 月から今日に至るまで伊是名島で継続している「空間図式の探究」の成果を地域創生のデザインに活用する課題を＜臨床の知＞という視点から解釈した。

参考文献

[藤井 2015] 藤井晴行: 知をデザインする, 一人称研究のすすめ – 知能研究の新しい潮流(諏訪正樹, 堀浩一編著), pp.151-170, 近代科学社, 2015.

[藤井 2016] 藤井晴行: 記号と実体を結びつける空間図式を「写真日記」によって顕在化する一人称研究, 人工知能学会全国大会(第 30 回), 2016.

[藤井 2017] 藤井晴行, 篠崎健一: 建築空間の認識と創生における知の身体性の現れ―空間図式の研究, 人工知能学会全国大会(第 31 回), 2017.

[福田 2015] 福田隼登, 藤井晴行: 身体性に注目した空間体験の図式表現方法に関する研究, 日本建築学会計画系論文集, 80 (709), pp.559-567, 2015.

[篠崎 2015] 篠崎健一, 藤井晴行, 片岡菜苗子, 加藤絵里, 福田隼登: 空間図式の身体的原型の実地における空間体験に基づく研究―写真日記を基礎資料とする KJ 法の試み, 認知科学 第 22 巻・第 1 号, pp.37-52, 2015.3.

[中村 1992] 中村雄二郎: 臨床の知とは何か, 岩波書店 ,1992.

[中村 1993] 中村雄二郎: デザインする意志, エッセー集成 6, 青土社 ,1993.

[和辻 1931] 和辻哲郎: 風土―人間学的考察, 岩波書店 , 1931.

謝辞

本研究は科学研究費補助金「挑戦的萌芽研究(15K12295、代表: 篠崎健一)」、同「基盤研究(B)(一般)(16H03014、代表: 藤井晴行)」を受けて遂行している研究のプロセスと成果を題材とするものです。また、臨地調査や調査結果の整理と分析においてにおいて、日本大学篠崎研究室のメンバー、東京工業大学のメンバーの協力を得ています。ここに謝意を表します。

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

リハビリテーション専門職によるハンドリング技術の分析

Analysis of the Handling Technique on a Rehabilitation Therapist

宮本一巧[*1,2] Ikko Miyamoto　塩瀬隆之[*3] Takayuki Shiose　阪上雅昭[*3] Masaaki Sakagami

[*1] 地域医療機構りつりん病院 Ritsurin Hospital
[*2] 早稲田大学人間科学部 Waseda University School of Human Sciences
[*3] 京都大学 Kyoto University

Abstract: This study examined the feature of the handling for the treatment of the sit-to-stand motion of the patient in the rehabilitation domain and "knacks" that I could suppose from those features. We had four objects; an occupational therapist who has 18 years of experience and a student from a department of occupational therapy as a role of the therapist, and people who assumed a person of virtual single paralysis as a role of a patient. It was suggested the expert made the patient move forward by maintaining a certain distance between the patient and him, and this motion is very important. What supported the characteristic was the coordinative structure of the whole body, which limits the movement of arms and makes lower limbs move more freely. On the other hand, such a physical structure is seen in the characteristic of the martial arts-like physical campaign for ancient Japanese martial arts. Therefore when we compared sit-up exercise of ancient Japanese martial arts and usual one, the distance with a head and each body parts was smaller the first one.

1. 背景・目的

1.1 ハンドリングとは

理学療法士・作業療法士（以下，セラピスト）は患者の身体に直接的に手を添え，目的動作・行為の支援を手段とし治療を展開していく場面がしばしばある．そのような徒手的介入による対象者の身体操作の総称をリハビリテーション領域では「ハンドリング（handling）」という．それは介護場面で見受けられる「動き」の主導権が介入者側であるような受身的な「介助」ではない．①対象児・者との大切なコミュニケーションの手段であり，②対象児・者の潜在能力を評価する技術であり，③さまざまな実生活支援の 1 手段であり 1 段階[岸本 16]であると言われるように，主導権が介入者－被介入者の間を常に行き来する相互的なものである．

1.2 ハンドリングに関わる問題

リハビリテーションが「医療」の枠に身を置いている以上，その治療手段には十分な根拠が必要であることはもはや自明である．しかしながら同じ対象者であっても，セラピストの体形や位置関係によっても力の加減や方向などが変わるハンドリングは，エビデンスに繋がり難い現状がある．事実，脳卒中治療ガイドライン[日本脳卒中学会 09]にはハンドリングが手段の中心であるファシリテーション・テクニックにおいて「ファシリテーション（神経筋促通手技）は行っても良いが，伝統的なリハビリテーションより有効であるという科学的な根拠はない」とし，エビデンスの限界を示している．言い換えれば，経験や勘のみの解釈で，定量的に分析が出来なければ技術の伝承はもとより，治療を発展的に考えていくことにも限界があると言える．したがって，ハンドリング技術の一部分でも客観的に示すことは，エビデンスの確立に繋がり，後輩・学生教育の充実，そしてセラピストが治療介入することの意義の高さが示されるはずである．

1.3 「コツ」を探る

上述したようにハンドリングは経験や勘に依存する暗黙知とした部分が多く，その技術の伝承は容易ではない．しかし技術を誘うような，これが出来ればあとは成り行きで動き出せるような、言わば「コツ」を示すことが出来れば，技術を伝承していくことの一助になるであろう．本研究では，運動学的な解析から得られたデータを用いて熟練者が持つ技術の特徴の可視化を目的とした．メカニズムの説明と現象の生成は別物であるため[諏訪 16]，運動学的な解析が提供するデータがそのまま「コツ」に結びつくとは限らないが，その解明の足がかりになることを期待したい．

連絡先：宮本一巧，地域医療機構りつりん病院
mail：i.miyamoto0315@akane.waseda.jp

2. 実験 1

2.1 方法

対象者 セラピスト役として 18 年目の作業療法士（以下、熟練者）と作業療法学科学生（以下，非熟練者）の 2 名、患者役として仮想片麻痺者を想定した健常者 2 名とした。実験参加者には事前に十分な説明を行い，同意を得た．

実験機材 三次元動作解析装置（Vicon Motion System 社，ViconMX），床反力計（AMTI 社，Watertown）を用いて測定した．介入者，被介入者の全身には共に 35 ヶ所の赤外線マーカーを貼付し，赤外線カメラ 10 台でその軌跡を撮影，サンプリング周波数は 120Hz とした．解析ソフトは Vicon Nexus 1.8.5 を用いた．

実験場面 患者役には，左片麻痺者特有の姿勢戦略を再現するために，高さ 20cm の椅子に右臀部のみで着座し，左足部をボールの上に載せることで非対称的で不安定な起立動作になるように設定した（Figure 1）．セラピスト役は前方から患者役の両上腕近位部に手添えし，患者役の起立動作のハンドリングを行った．



Figure 1. 被介入者の実験場面

高さ 20cm の椅子に右臀部のみで着座し，左足部をボールの上に載せることで非対称的で不安定な起立動作になるように設定．

実験手続き 起立動作の速度を一定とするためにメトロノームを用い(M.M = 60), 3 秒間で起立し, 2 秒間の静止立位を指示し, 出来るだけメトロノーム音に同期させるように教示した. 熟練者, 非熟練者各々のセラピスト役がハンドリングしながらの起立動作(5 回)を測定した(Figure 2).



Figure 2. 測定中のセラピスト・患者役の肢位

検討内容 ①セラピストー患者役が作る間合い, ②セラピスト役の「姿勢づくり」について検討した.

①両身体間の間合いの作り方 分析対象はセラピスト役と患者役の前後身体重心(center of mas, COM)の時系列データとした. 分析方法は, 1)分析対象データから両者の COM 間の距離を算出した. 2)COM 間の距離の平均値を 0 に基準化して root mean square(以下, RMS)を求めた. RMS はデータの変動性の指標であり, その値が小さい程, 変動が小さいことを示す. 統計ソフトは R 2.8.1 を用い, 1)2)ともに Group(expert／novice)× Trial(5Trial)の反復測定分散分析を実施した. 有意水準は 5%以下とした.

②姿勢づくり 分析対象は治療者役の肩関節・肘関節・股関節・膝関節・足関節屈伸可動域の時系列データとした. 1)各々の関節運動の安定変動性を評価するために平均値を 0 に基準化して root mean square(以下, RMS)を求めた. 2)下肢運動の評価のために股関節・膝関節・足関節に関して最大屈曲角度を算出した. 統計ソフトは R 2.8.1 を用い, 1)Group(expert／novice)×Joint(Shoulder／Elbow／Hip／Knee／Ankle)の二要因混合計画分散分析, 2)Group(expert／novice)× Trial(5Trial)の反復測定分散分析を実施した. 有意水準は 5%以下とした.

2.2 結果

①両身体間の空間の作り方 1)の基本統計量は, 熟練者では 355.01±20.08(mean±SD), 被熟練者では 295.00±8.61 であった. 反復測定分散分析の結果, 熟練者の方がセラピストと患者役との COM 間の距離が有意に大きかった($F(1, 4) = 24.31$, $p<.01$). 2)の基本統計量は, 熟練者では 27.72±4.38, 非熟練者では 42.86±4.96 であった. 反復測定分散分析の結果, 熟練者の方が COM 間の距離の RMS が有意に小さかった($F(1, 4) = 20.55$, $p<.05$)

②姿勢づくり

1)関節運動の変動性 基本統計量を表 1 に示す. 分散分析を行った結果, 交互作用が有意であった($F(1.68, 13.42) = 107.15$, $p<.001$). 各関節別に単純主効果を検定したところ, 肩関節・肘関 節では熟練者が有意に小さく(各々$F(1, 8) = 22.76$, $p<.001$, $F(1, 8) = 244.39$, $p<.001$), 股関節・膝関節では熟練者が有意に大きかった(各々$F(1, 8) = 36.14$, $p<.001$, $F(1, 8) = 16.35$, $p<.01$). 足関節では有意差はなかった($F(1, 8) = 0.54$, $p = 0.48$). また, Group 別に単純主効果検定を行ったところ, 熟練者, 非熟練者ともに有意であった(各々$F(4, 16) = 78.8$, $p<.001$, $F(4, 16) = 105.02$, $p<.001$). 熟練者では肩関節・肘関節の方が股関節・膝関節よりも有意に小さい反面($p<.05$), 非熟練者では股関節・膝関節の方が肩関節・肘関節よりも有意に小さかった($p<.05$).

	＊＊＊ 肩関節	＊＊＊ 肘関節	＊＊＊ 股関節	＊＊ 膝関節	n.s 足関節
熟練者	2.9841 (1.052)	7.7437 (0.8977)	16.9306 (2.7893)	18.4235 (2.9678)	5.0897 (1.3787)
非熟練者	16.416 (1.6945)	20.4082 (1.5734)	8.6218 (1.3304)	11.9982 (1.9525)	4.5456 (0.8986)

( )内は標準偏差

表 1. 各関節運動の RMS の平均値と標準偏差

2)股膝足関節の最大屈曲角の評価 基本統計量を表 2 に示す. 反復測定分散分析の結果, 股関節, 膝関節で熟練者の方が有意に大きく(各々$F(1,4) = 134.01$,$p<.001$, $F(1,4) = 527.69$,$p<.001$), 足関節でも熟練者の方が大きい傾向にあった($F(1,4) = 7.46$,$p = .052$).

	＊＊＊ 股関節	＊＊＊ 膝関節	† 足関節
熟練者	65.8266 (2.2757)	81.2484 (2.1027)	26.4785 (1.5091)
非熟練者	31.6075 (5.2916)	35.092 (4.4015)	23.4676 (1.1442)

( )内は標準偏差

表 2. 下肢三関節の最大屈曲角度の平均値と標準偏差

2.3 考察

リハビリテーション領域でのハンドリングおいて, 可視化することが出来る熟練者の技を検討した結果, 以下のことがわかった.

セラピストー患者役の身体が作る空間の使い方の検討①では, 熟練者の方が COM 間の距離が大きいにも関わらず, そのばらつきは小さい傾向にあった. つまり患者役と身体間の距離を大きくとりながらも, 終始その距離間を一定にするという「空間方向の間合い」を制御していたことが示唆された. この「空間方向の間合い」は起立動作に対するハンドリングにおいて, 本研究のように患者の前方からアプローチする場合には非常に重要になってくる. ここで起立動作の序盤を矢状面から分析してみると, 臀部が支持面から離れるまでは頭部と体幹が前方へ移動していく. この前方移動によって臀部にあった COM を足底に移動させることで離臀が起こり起立動作が進行していく. このような運動学的特徴のある起立動作に対して前方からアプローチすることは, 患者にとってはセラピストが視覚的な「壁」となり, 離臀に必要な頭部と体幹の前方移動の妨げになりかねない. しかし

熟練者は「空間方向の間合い」を作ることで離臀までの十分な前方移動を許容していたと考えることが出来る.

このような特徴を持ったハンドリングはどのような「協応構造」を持ってして実現しているのであろうか.「姿勢づくり」に着目した検討②では, 熟練者では上肢(肩関節, 肘関節)の変動性は小さく, 下肢(股関節, 膝関節)の変動性は大きかった. 逆に非熟練者では上肢の変動性は大きく, 下肢の変動性は小さかった. 下肢三関節の最大屈曲角度の評価では, 熟練者の方が下肢を大きく曲げていた. つまり熟練者では, 上肢の運動の自由度を抑え, 下肢の自由度を大きくすることで起立動作を支援していたと言える.

ところで, けん玉の「ふりけん」という技を分析した研究では, 熟練者は刻々と変化する玉の軌道にぴったりと合わせて自身の頭部を動かすことで玉の動きを精度よく捉えていると報告している[伊藤 10]. 本研究における熟練者においても, 上肢の運動の自由度を抑え, 下肢の自由度を大きくするという「姿勢」に支持された「空間的な間合い」を作ることで, 自身の動きを患者の動きにぴったりと合わせていたと推測できる. 動きを合わせることは起立動作中の患者の微細な変化を的確にセンシングすることに貢献していると考えられる.

熟練者のこのようなハンドリングなどの身体操作は上記した内容だけでも巧みな身体運動が合理的に組織化されていることがみて取れる. 合理に適った身体運動の解明には日本人古来の動きを模索している古武術が重要な示唆を与えるものと考えている. 続く実験 2 ではセラピストが行うハンドリングとは別に, 古武術的な身体操作の可視化を試みた.

3. 実験 2

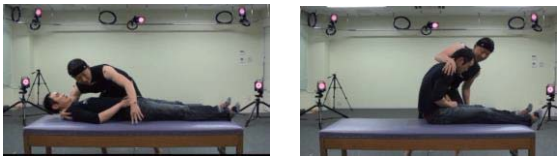
3.1 方法

対象者 50 歳代男性(香川稽古研究会主宰)1 名(介入者).

健常男性 1 名(被介入者)

実験機材 実験 1 と同様とした.

実験手続き 介入者は, 臥位姿勢の被介入者の頚部背側から右上肢を入れ, 被介入者の上半身を持ち上げるように起こす. 開始肢位は被介入者の背臥位, 終了肢位は長座位とした(Figure 3). 同一介入者が一般的な方法で起き上がりの操作をした場合(3 回)と, 古武術的な方法で操作した場合(3 回)を比較した.



(a) (b)

Figure 3. 開始肢位(a), 終了肢位(b)

検討内容 得られた介入者の頭部・肩・肘・手の左右成分 (X), 垂直成分 (Z) の時系列データから, 頭—肩, 頭—肘, 頭—手の距離の平均値と標準偏差を各々算出し比較した.

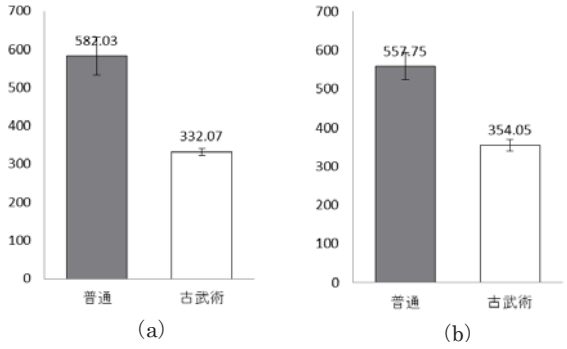
3.2 結果

基本統計量を表 3 に示す. 頭—肩, 頭—肘, 頭—手の全てにおいて, 古武術的な方法の方が小さい結果であった. 特に頭—手においてはその差が顕著であった(Figure 4). そこで頭—手の左右成分(X)の関係性をプロットした軌跡を目視により比較すると, 頭部の動きには大きな差が見られなかったが, 手の動きに著明な差が確認できた(各々の 1 trial を Figure 5 に示す).

	頭部−肩		頭部−肘		頭部−手	
	左側	右側	左側	右側	左側	右側
古武術的方法	181.13	175.62	350.25	401.42	332.07	354.05
	(13.82)	(9.69)	(12.92)	(8.83)	(9.30)	(15.90)
一般的方法	194.96	244.33	396.92	531.69	582.03	557.75
	(12.75)	(1.63)	(36.30)	(3.00)	(49.30)	(34.94)

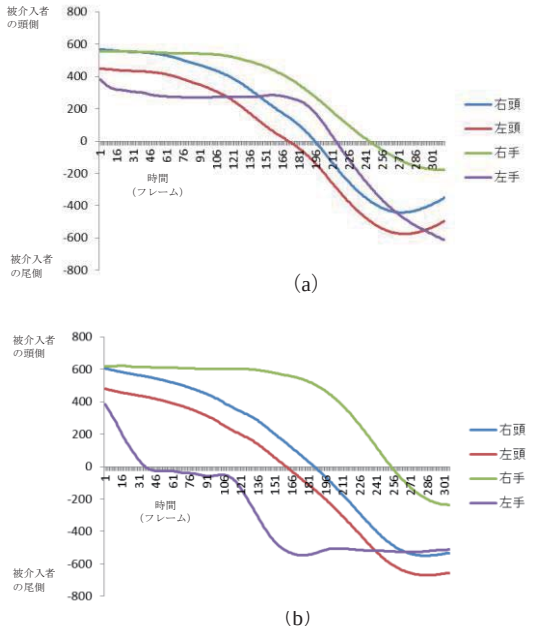
( )内は標準偏差

表 3. 頭部と各身体部位との距離の平均値と標準偏差



(a) (b)

Figure 4. 頭—左手の距離の平均値(a)
頭—右手の距離の平均値(b)



(a)

(b)

Figure 5. 頭部と手の左右成分の軌跡
古武術的方法(a), 一般的方法(b)

3.3 考察

古武術的な身体操作の特徴の可視化を検討した結果, 以下のことがわかった.

古武術的な方法の方が頭—肩, 頭—肘, 頭—手の全てでその距離が小さい結果であった. 特に頭—手においてその差が

著明であり，加えてその標準偏差が小さい傾向にあった．このことは，頭と手との距離を短く且つ，動作の間，終始それを一定に保っていたことを示している．一般に力を産む筋の特性は，筋長が長ければ得られる張力も大きくなる．つまり大きな力を得ようとするならば，身体をある程度「伸ばす」必要がある．したがって，一般的な方法では頭―手の距離を大きくとることで，右手にかかる負荷に対して局所で発生する「力」で対処していたものと推測される．これは Figure 5(b)においてまず左手を動かし両手間距離を確保していることと整合的である．それに比べ古武術的な方法で頭―手の距離が小さかったことは局所の「力」以外の戦略で動作を遂行していたことが示唆される．注目すべきは，古武術的な方法では被介入者を支えている手とは逆の左手までも頭部と近いということである．一方で一般的な方法では動作開始当初から離れていることがわかる（Figure 4（b））．これは古武術的な方法では左右の上肢と頭部の動きを同時並列的に組織化しているということに対し，一般的な方法では左上肢を初動の原点とし後から頭部と右上肢がついてくるというある種順序立った組織化に由来するものだと考えられる．

4. 結語

以上，実験 1 ではリハビリテーション領域でのハンドリングにおいて，可視化することが出来る熟練者の技の特徴の一部を示した．そして続く実験 2 では合理に適った身体運動である古武術的な身体操作の特徴の可視化を検討した．セラピストが行うハンドリングは非常に繊細であり，患者の僅かな変化を感じ取りながら行っている．それは武術的な身体操作のような合理に適った身体運動に裏付けされた知覚に支えられているはずである．繰り返しになるがハンドリングを初め，身体操作は非常に複雑である．そのため本研究ではその一部分しか示すことが出来ていない．しかしながら今回の報告が今後のハンドリング研究や医学発展の一助になることを期待したい．

参考文献

［岸本 16］岸本光夫：ハンドリングの意義－対象児・者に寄り添うためのハンドリング，OT ジャーナル，50(1), 66-70. 2016

［日本脳卒中学会 09］日本脳卒中学会：運動障害・ADL に対するリハビリテーション．脳卒中治療ガイドライン．296-299. 2009

［諏訪 16］諏訪正樹：「こつ」と「スランプ」の研究－身体知の認知科学－．講談社選書メチエ．2016

［伊藤 10］伊藤万利子・三嶋博之・佐々木正人：けん玉操作における視覚情報と姿勢制御．日本認知科学会第 27 回大会．2010

Oral presentation | Organized Session | [Organized Session] OS-19

[2B3-OS-19b] [Organized Session] OS-19

Wed. Jun 6, 2018 3:20 PM - 4:40 PM Room B (4F Moon Light)

[2B3-OS-19b-01] Stay mobile and carry on

○Fumitoshi Kato[1] （1. Keio University）

3:20 PM - 3:40 PM

[2B3-OS-19b-02] The Story of the REMAINS

○Kikuna Kuwayama[1], Minami Koseki[2], Masaki Suwa[1] （1. Keio University, 2. Keio Research Institute at SFC）

3:40 PM - 4:00 PM

[2B3-OS-19b-03] Consideration for succession of knowledge, technology and skill

○Yasuyoshi Taruta[1,2], Satoshi Yanagihara[2], Yukihiro Iguchi[1,2], Koichi Kitamura[1], Masashi Tezuka[1], Yuya Kouda[1] （1. Japan Atomic Energy Agency, 2. Fukui Univ.）

4:00 PM - 4:20 PM

[2B3-OS-19b-04] A Proposal of the Constructed Exhibition

○Minayo Okumura[1], Tomoharu Hoshina[2], Ken-ichi Kimura[3] （1. Graduate School of Systems Information Science, Future University Hakodate, 2. Hakodate City Museum, 3. Future University Hakodate）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

うごけよつねに：モノをとおしてコミュニケーションを理解する

Stay mobile and carry on: On understanding communication through our use of artifacts

加藤文俊[*1]
Fumitoshi Kato

[*1]慶應義塾大学
Keio University

The present study is an attempt to explore the notion of "mobility" through an experiential learning program. In Fall 2017, five groups of college students attempted to design a series of "mobile kits" for place-making activities. That is to envision and identify when, where, with whom, and how our "precious moments" will be realized. From a carry cart to a bicycle, each group tried to modify and redesign a range of selected items (vehicles) based on their own "issues"at hand. By referring to the process and outcome of the project, this paper begins a discussion about the ways in which we can speculate upon human communication behavior though understanding the use of artifacts/objects in our day-to-day activities.

1. はじめに

筆者は，ワツラヴィックらによるコミュニケーションに関する「試案的公理」を参照しながら，さまざまな実践の文脈における「場づくり」やファシリテーションのあり方に関心を寄せてきた．とくに「コミュニケーションしないことの不可能性（One cannot not communicate）」という言明は，きわめて示唆に富んでいる [ワツラヴィック 2007]．私たちは，しゃべることをコミュニケーションだと考えがちだが，黙っていることも，さらに他の非言語的，身体的ふるまいもふくめて，絶え間なくお互いの状況を伝え合っている．つまり，私たちはコミュニケーションせざるをえないのである．さらに，私たちが日常生活においてさまざまなモノや道具を必要としていることから，モノの成り立ちやモノ自体をとおして，人びとのコミュニケーションの痕跡やありようについて洞察を加えることもできるだろう [原田 2003]．

本論文では，「うごけよつねに」というプロジェクト [加藤文俊研究室 2018] を事例として紹介しながら，モノ（移動体）をとおしたコミュニケーション過程を理解する試みについて論じたい．モノは，私たちのコミュニケーションの契機（始まり・終わり）となるばかりでなく，時間，空間の構成・再構成に関する身体的な理解の創造に役立つ．

2. うごけよつねに

2.1 モノをとおしてコミュニケーションを考える

私たちのコミュニケーション行動が，つねに〈いつか・どこか〉で実現することをふまえると，私たちのコミュニケーション過程は，つねに「場づくり」の問題と密接に関わっていると言える．私たちは，コミュニケーション欲求を満たすために集う．つまり，「場づくり」は，私たちのコミュニケーション過程で，絶え間なく空間，時間，情報を調整する試みである [加藤 2016]．一連の時間・空間の調整によって成り立つ「集い」は，ある一定の時間・期間を経て消失する場合が多い．それは，私たちがつねに「移動していること（on the move）」の証である [アーリ 2015]．私たちの移動性に着目すると，私たちが日頃は目的地に至るまでの「あいだ」だと位置づけている，通勤や通学の際に利用している電車の車両も，私たちのコミュニケーションの所産として構成されるひとつの「場」であることに気づく．黙ったまま，お互いに干渉せずに電車に揺られていても，私たちは絶えずコミュニケーションを行い，車両にいる時間・空間を調整しながら過ごしているからである．

私たちが標榜する社会的な実践をともなう調査・研究プロジェクトにおいて，持続可能性が重要であることがしばしば指摘される．とりわけまちや地域に関わる活動の場合には，ある程度の時間をかけないと，その意義や意味を実感できないからだ．活動そのものが楽しいことも大切だが，筆者が過去の経験から学んだ「続けるコツ」は，できるかぎり身軽になることである．

これまで「ちいさなトラック」「引っ越しの準備」「爽やかな解散」「連れてって」など，つねに私たちの移動（移動性）に関わるテーマで，大学生（学部生）を対象とするフィールドワークのプロジェクトを実践してきた．そして，いまの状況からつぎの状況へと，すみやかに移動できることが重要だという点を体感してきた．いわゆる「心の準備」もふくめ，「モバイルでいること」こそが，絶え間ないコミュニケーション行動をかたどっているはずだ．こうした理解をふまえ，2017 年度秋学期は，「うごけよつねに」というテーマに決めた．英語には，「Stay Mobile and Carry On」をあてた．

2.2 課題の概要

今回は，学部 1〜3 年生で 5 つのグループを編成して，このプロジェクトをすすめた．筆者が，あらかじめベースとなる「車両」を準備し，それぞれのグループに提供した．キックボード，キャリーカート，電動アシスト付き自転車など，形もサイズも多様だ．いずれも車輪がついていて，そのままでも使えるので，いわばカスタムカー（改造車）をつくるようなものだ．

学生たちは，〈いつ・どこで・誰と・どのように過ごしたいのか〉という具体的な状況を考えながら，「場づくり」を容易にしたり，コミュニケーションを促したりする「モバイル・キット」のデザインに取り組んだ．そして，考案した「車両」は，かならず一度はまちに持ち出して，実践（フィールド・テスト）を行うよう促した．モノ自体は，多少粗削りでも，仕上げが雑でもかまわないので，とにかく教室や研究室に留めておくのではなく，リアルな現場へと「車両」を持ち出して，実践をとおして考えることを重視した．

学生たちは，ひとたび「車両」があたえられると，形や見栄えに意識が向いて，〈つくる〉ことが先行しがちになる．その結果，柔軟な発想が制約されることにもなる．そうした状況を避けるために，所与の「車両」によって，どのようなコミュニケーションが生まれるのかについて，つねに〈考える〉ことを忘れないよう促した．

連絡先：加藤文俊，慶應義塾大学環境情報学部，〒252−0882
神奈川県藤沢市遠藤 5322，fk@sfc.keio.ac.jp

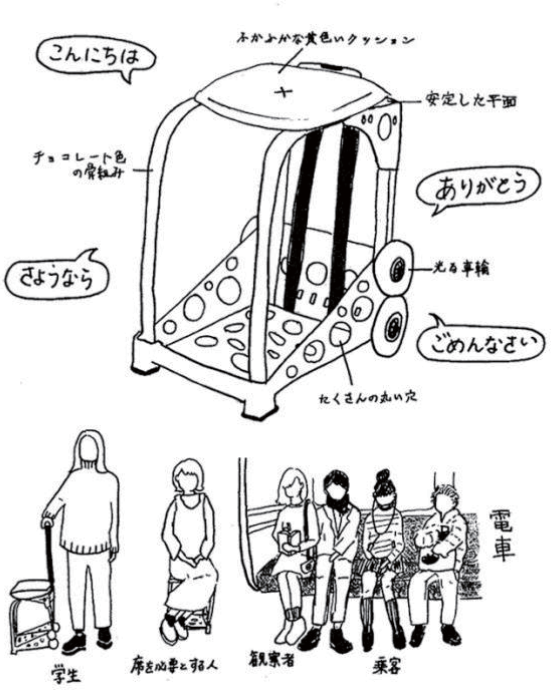
3. 経過：YUZÜROCAR

上述のとおり「うごけよつねに」というプロジェクトでは，それぞれのグループがベースとなる「車両」を起点に発想し，「モバイル・キット」の提案を試みた．以下では，5 つのグループの中から「YUZÜROCAR」の事例を紹介したい．

3.1 着眼点

「YUZÜROCAR」は，その呼称が示唆するように，電車等の公共交通機関等を利用する状況で，席を「ゆずろう」という想いを喚起することを目指している．このグループは，私たちが「こんにちは」「さようなら」「ありがとう」「ごめんなさい」といったごく基本的な挨拶行動を忘れがちな点に着目した．挨拶行動は，他者の理解，関係性の表れであり，それは「他人への想像力」を喚起するという課題として位置づけられるようになった．つまり，私たちが日常的な挨拶行動を軽視したり，忘れていたりするのは「他人への想像力」が欠如しているからだという考えである．

3.2 デザイン



このグループは ZÜCA という少し変わった形状のキャリーバッグをベースとして選択した．ZÜCA は，金属の堅牢なフレーム構造が特徴的で，簡易なイスや踏み台ように利用することができる．実際にこれを携行していれば，（さまざまな制約はあるものの）座る場所に困ることはない．そこで，このキャリーバッグを「持ち運ぶことのできる座席」に見立てながら，私たちが電車に乗り合わせるという状況への介入を試みた．その中で，乗客たちに対して座席を譲る行動を喚起し，さらにその結果として「ありがとう」「さようなら」などの言葉を引き出そうとした．（図は同グループによるイメージ [加藤文俊研究室 2018]）

3.3 実践

いくつかの作業仮説に基づいて，数回の社会実験を行った．電車で座席を「譲ろう」という気持ちにさせる状況は，他者が何らかの不便を被っていると感じたときであろう．言うまでもなく，外見は視覚的な手がかりになる．年格好や白杖，松葉杖等は，人びとの状態を知らせる働きをするので，私たちが「他人への想像力」を喚起するきっかけづくりになる．当然のことながら，一人ひとりの不便や不具合は目に見えるとはかぎらず，たとえば「マタニティマーク」のようなモノも，自らの状況を可視化し，メッセージを際立たせるのに役立つことが期待されている．

車両に空席がない状況で，この ZÜCA を座席として譲る状況を演出した結果，居合わせた乗客が席を譲ろうとしたり，「ありがとう」「さようなら」といった声をかけ合ったりする状況を観察することができた．

4. 考察

4.1 移動という関わり

いささかナイーヴな実験ではあったが，モノ（移動体）の役割を実態的に理解することができた．私たちの「居方（いかた）」は，モノを介して変容し，それはつねに「移動中」という状況にある．電車に乗り合わせる場面に限らず，他者との適切な距離や関係のありようについて考える際には，時間・空間のみならず，モノとの関わりをふくめて理解を試みることが重要であろう．

4.2 つくりながら考える・考えながらつくる

この課題は〈考える〉ことと〈つくる〉ことを往復しながら進行するように設計されていた．それは，コンセプトワークとフィールドワークとの橋渡しを担う，実験的な環境によって実現する [加藤 2017]．ここで言う実験は，統制群によって条件を操作しながら行う厳密な意味での「実験」ではなく，「即興的実験（on-the-spot experiment）」 [ショーン 2007] や，プロトタイプによる試行のくり返しを伴う学習過程として理解することができる．「ある程度の一般性とかなりの具体性を持った実践モデル」 [西條 2007] を提示することで，私たちの想像力を育むものであった．

4.3 「つくらない」という選択

「うごけよつねに」では，5 つの試行実験が行われた．ベースとなった「車両」に手を加えて「モバイル・キット」をデザインすることが期待されていたが，試行錯誤を経て，ZÜCA は，ほぼそのままの状態が最終的な成果物の「形」となった．私たちのコミュニケーションや関係性の理解において，モノの役割を無視することはできないが，同時に，すべてがモノに集約されるわけでもない．まさにコミュニケーションの所産として，相互構成的に状況が立ち現れることを再確認するきっかけとなった．

参考文献

[アーリ 2015] ジョン・アーリ: モビリティーズ, 作品社, 2015.

[加藤文俊研究室 2018] うごけよつねに（Stay Mobile and Carry On）, プロジェクト報告書, https://vanotica.net/smaco/, 2018.

[加藤 2017] 加藤文俊: 「ラボラトリー」とデザイン, SFC Journal, 17(1), pp. 110-130, 2017.

[加藤 2016] 加藤文俊: 会議のマネジメント, 中央公論新社, 2016.

[西條 2007] 西條剛央: 質的研究とは何か（ライブ講義）, 新曜社, 2007.

[ショーン 2007] ドナルド・ショーン: 省察的実践とは何か, 鳳書房, 2007.

[原田 2003] 原田隆司・寺田伸悟: ものと人の社会学, 世界思想社, 2003.

[ワツラヴィック 2007] ポール・ワツラヴィック, ジャネット・ハベラス, ドン・ジャクソン: 人間コミュニケーションの語用論, 二瓶社, 2007.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

跡の物語
The Story of the REMAINS

桑山 菊夏[*1] 小関 美南[*2] 諏訪 正樹[*3]
Kikuna Kuwayama Minami Koseki Masaki Suwa

慶應義塾大学大学院政策・メディア研究科[*1] 慶應義塾大学 SFC 研究所[*2]
Graduate School of Media and Governance, Keio University Keio Research Institute at SFC

慶應義塾大学環境情報学部[*3]
Faculty of Environment and Information Studies, Keio University

Looking at a vacant lot and thereby imagining the building and someone's life that used to be there, and thinking of the owner of a used book from handwritten memos in the book etc. – Attending to "REMAINS", are able to associate them with many stories around it. When we notice some REMAINS, what elements do we recognize in the "REMAINS", and how do we make a story out of them? This project aims at promoting our meta-cognition through writing "The REMAIN's Story", and at enabling us to encounter new aspects of things in everyday life.

1. はじめに:「跡」に出会うこと

「跡」や「痕」について、思いを馳せたことがあるだろうか。わたしたちはこの言葉を、ときに何気なく、いつのまにか口にする。

紙につけた折り目、窓ガラスについた指紋、さら地、古本にある走り書き。すべては「跡」のようで、過去を思わせる。また、わたしたちは、さまざまな方法で残された「跡」を、利用したり、消したり、作り出したり、解釈したりする。

「跡」を見つめると、そこには“だれかが居た”、“なにかが動いた”物語が見えてくる。そのとき、私たちは、「跡」のどんな要素に気づき、その要素をどう解釈してその物語にたどり着いたのだろうか。私たちは、何を「跡」と感じているのだろうか。どのように「跡」が残され、私たちは「跡」にどんなふうに出会うのだろうか。

本研究は、日常における「跡」との出会いと「跡」の持っている物語性に惹かれて始まった。

「跡」との出会いの研究を行う上で、第一著者と第二著者の2名が様々な「跡」に着眼し、「跡」にまつわる具体的な体験を集め、世界の捉え方についての問いや気づきを生むことを目指し、研究手法を試行してきた。本稿は、その軌跡を記したものである。

2. 研究の動機と背景

2.1 「予感」から語りを深める

世界の捉え方について気づきを得たいとして、筆者らがなぜ「跡」に着眼したのか。

一つは、「予感」である。「跡」について深めると、良い気づきを得られるのではないかと考えたためである。あまりに当てずっぽうではないかと思われるかもしれないが、「予感」から始まっている~~という~~ことにこそ、本研究の問いの本質が隠れている。

第一筆者が「跡」という概念について興味を持ち始めたのは、自らの散歩についての省察がきっかけである。この省察~~についてそ~~は、小さなきっかけに過ぎないが、筆者らが興味を持つ「跡」がどのようなものかを具体的に提示するため、ここで少々語ってみる。

ある晴れた日、一人で散歩をしていた第一筆者(以下 K)は、小さな公園にたどり着く(図1)。少し古びた遊具と砂場、ベンチ、トイレがあるだけの小さな公園には、余白が少ない。K は、歩き回りながら観察する。砂場のそばに残される引きずった靴の跡、丸いトイレの周りに残った地面のくぼみから、遊具以外の場所で子供達の遊ぶ姿を想像することができる。

K は公園を離れ、また歩く。明るいピンク色のアパートの二階にふと目がいく。アパートのドアの前には、トランクが置かれている(図2)。なぜ、ここにトランクを置いたままなのだろうか。これから旅に出るのだろうか、どんな人なのだろうか。

しばらくまっすぐに歩くと、踏切に出会う。踏切を渡った瞬間、道の雰囲気が変わる(図3)。初めて来た土地なのに、なぜか懐かしい気持ちになる。友人と歩いた、別の場所での思い出がよぎる。あの日、あの時のあの場所と今歩くこの場所は、もしかしたら何かが似ているのかもしれない。







左上:図1 小さな公園
右上:図2 アパートとトランク
左下:図3 踏切の先の道

連絡先:桑山菊夏, 慶應義塾大学大学院政策・メディア研究科, 神奈川県藤沢市遠藤 5322, kikuna@sfc.keio.ac.jp

別の日のことだ。また歩く。遠くを見つめると、少し先に見える家の屋根に、「つまみ」のような出っ張りが複数ついている。きっと、太陽光パネルが設置されていたのではないか。一度つけて外してしまったのか、設置するつもりで結局設置していないのかもしれない。ないようでそこにあるパネルの居場所は、なぜかむずむずと心をくすぐる。

K は思った。公園には、人の集まり方を示す「跡」がある。目についてしまう曖昧な「跡」は、誰かの生活を思わせる。その「誰か」の人となりを思わせる。そして、また別の場面では、自らの身体に残っている記憶という「跡」がある。

この省察は、おそらく K の体験した散歩のすべてを伝えてはくれない。K の印象に残ったことを、K の主観で切り取って記したものであり、現象についての「正確な」説明ではないからだ。

しかし、K 自身には、「跡」との出会いが重要であるのではないかという仮説と、「跡」に興味を惹かれる自分自身についての疑問がもたらされた。一つとして同じものはない「跡」について、もう二度と出会わない小さな何気ない出来事が気になり、「跡」という一つの言葉を用いて語るのだろうか。

このような疑問から、K は「跡」に対する興味が高まることとなった。また、この興味について第二著者に語り、彼女も類似した問いを持っているのではないかと感じた。その共感は、「跡」を研究テーマとして選ぶと良いのではないかという予感に変わった。

「良い予感があるから」というのは、あまりに根拠の見えない研究動機に聞こえる。しかし、私たちの暮らし、世界、生き方には、まだまだ予想もつかない何かが隠れているはずだ。「臨床の知」[中村 1992]は、我々の世界を形作る言語のシステムについて紐解こうとすることの中に見いだすことができるのではないか。

普段は気にとめることがない「跡」に着眼することで、暮らしにころがっているものごとの思いがけない表情や生き様に出会うことができる。「跡」との出会いを物語る試みが、「経験すること」が私たちにもたらす変化と、その土台についての見直しに繋がることを強く期待する。

2.2 「跡」の持つ可能性

本研究の目的は、「跡」を定義することにあるのではない。「跡」がどのようにつくられているかを解きほぐすこと、そして「跡」を感じる自らの認知の働きについて見つめ直そうとすることだ。

本研究の開始は「予感」と言えるが、テーマについて検討しなかったわけではないので、そのことについても触れておく。筆者らは、改めて問いを深め、自分たちがなぜ「跡」をテーマに選ぶのか論じる。

まず第一に、「跡」は様々な状況の積み重ねが顕在化したものである。世界に存在する「跡」の一つ一つは、さまざまな状況が重なったが故に生まれた、唯一無二のものである。したがって、「跡」を研究することによって、認知の状況依存性（situated cognition）の様々な側面を掘り起こすことができるのではないかと思うのである。

例えば、コップについた口紅の「跡」。もし、ほんの少しだけ長く口をつけていたら。また、もし今朝気分が変わって、違う口紅を選んでいたら、異なる「跡」が今ここにあったかもしれない。

第二に、「跡」に気づくことをきっかけにして、過去・未来の誰かの行為に注目できるという点である。「跡」がつく前の状態から今までの間について、想像することができるのだ。

「跡」を感じるとき、「跡」そのもの自体を見つつ、そこから「何かが居たのだな」と感じることができる。例えば、公園の砂場についた「足跡」。その形と道筋から、この場所で遊んでいた、そして遊ぶであろう子供たちのありさま、存在を想像することができる。「今、ここに、私が感じる、この跡」は、過去や未来など、「この場所、この時以外」との連続性の中にあるものとして考えることができる。

上にあげた2つの「跡」の性質は、もちろんある側面を見るための着眼点である。ないはずのところにある「異質なもの」を感じる（例：床の傷跡）、そのものから何かの気配を感じる（例：残り香）、残したその人自身の証明としての価値を感じる（例：人柄の現れる筆跡）など、さまざまな要素が「跡」の認知を導くのではないか、とぼんやりとした仮説が筆者らの議論の中にあった。

しかし、本研究では、生活のなかで新たに具体的な体験の記述を重ね、仮説に基づかず、生活の感覚にできる限り正直に残した。これらの仮説を崩すことを意識した。議論で生まれた「跡」についての仮説を頭の中から完全に消すことはできないが、できる限りそのとき出会った場面をもとに「跡」を集めることをした。

このように述べると、あらかじめ考えの中にあった仮説から、無理矢理に何もかもを「跡」とみなすことができるのではないかという懸念が生まれるだろう。要するに「解釈次第ではないか？」という批判である。

しかし、本研究において目指すのは、暮らしや「跡」の可能性を広げる探求だ。つまり、積極的に解釈し、「跡」や生活を語るための語彙やその使い方を増やすことで、生活に転がっている様々なものごとの新たな側面や可能性をすくい上げることを目指している。

3. 活動概要

本研究における活動を、以下の3つに分けることができる。

⑴ 跡を集める

日常生活の中で「跡」にまつわる出来事を見つけ、自ら言葉で残す。必要であれば、図や写真なども用いる。その場で「これは「跡」だ！」と明確に意識できなくても、関連しそうなことは残しておく。

実施するのは二名（第一著者、第二著者）、実施期間は約4ヶ月間（2017 年 7～11 月）である。

⑵ 「跡」について話す

⑴を実施する二名で顔を合わせ、それぞれの生活で「跡」について考えたことを持ち寄る。集めた記述を見せつつ、語る。

約4ヶ月の間、K と M の二名（以下、プロジェクトメンバー）は、それぞれが日常生活の中に「跡」を集める。それらを持ち寄るミーティングを行い、集めた「跡」や「跡」に関して考えを話すことで、気づいていなかった観点を得たり、行動を変えたりする。

ミーティングは、4ヶ月間で 7 回実施した。場所は、飲食店や大学の教室など、座って静かに話をできるところを選んだ。

K は、修士課程に所属する学生として過ごし、大学のキャンパス、一人で暮らす家、アルバイト先の学習塾などで過ごすことが多い。M は、会社員として働いている。K と M は、大学時代から約3年以上の付き合いではあるが、プロジェクトの活動期間は、「跡」について話すミーティングによって互いの近況を知ることが多かった。

環境や生活の仕方の異なるお互いの視点からヒントを得ながら、それぞれの「跡」集めに取り組んだ。

⑶ 「跡」についての言葉を生み出す

集めた記述は、何らかの意味で「跡」に関連する言葉である。したがって、それらの記述の共通性や構成要素を見出すことによって、「跡」についての言葉をさらに生み出すことを試みる。本

研究では、集めた記述からタグを作成し、さらにグルーピングする方法を用いた。

4. 「跡」との出会いを記述する

まず、「跡」とどのように出会い、どのような出来事を経験したのか、記述することが「跡」を集めるために必要である。主観的な経験として、できる限り「豊かに」残すことが求められる。

これは、からだメタ認知 [諏訪・藤井 2015] の実践を繰り返し、「体感をことばにする」ことを習慣化することである。もちろん、意識することには制限があるが、観察できることを増やそうとし、じっくりとできるだけ時間をかけ、余さずそのままに経験をなぞろうとする。

以下は、第二著者（以下、M）の記述例の一つである。『詩に就いて』[谷川 2015]という本との記述である。

> *『詩に就いて』(谷川俊太郎)*
> *ハードカバーの本に、しおりのひもの跡、ついていた。*
> *しおりで分けた2つのページに、うねったしおりの跡。*
> *これは、ハードカバーのかたさと、少しかたい紙の素材と、本と本に挟まれていた、本棚の中のひしめきや密度と、それから一彼が微動だにしかなかった時間の長さである。*
> *そうか、そんなに長らくご無沙汰だったのか。*
> *視界にはいつも入っていた。枕元の小さな本棚にいたので、毎晩寝るときに、タイトルがちらっと。そこにいる、ということはずっと認識していた。だから、開くのがそんなに久しぶりだったなんて、なんだか少し反省してしまった。君のことはずっと好きだったのよ、そこにいるというだけで、タイトルや佇まいだけで、買った当時の心に帰るような。開いてなくとも、私を静かに鼓舞してくれています。彼の静かなかなしみが胸にきて、思わず愛の弁明。*
> *カバーは少し傷んでいて、ところどころ破けている。帯も気に入っていてつけたままだが、何かがこすれた跡がついている。紙の色がそこだけはげており、触ると少し毛羽立っていて、跡がついた瞬間の激しさを思う。周りの本を出し入れした時についたのだろうか？それとも、あなたを買ってすぐに、レジの人のブックカバーも断って、電車の中でぼんやりと真剣に読みふけっていたから、外傷に気づかなったのかもしれない。*
> *あなたに出会った藤沢の有隣堂で、私は研究がよくわからなくなっていた。ちょうど一年前だった。先生に相談する日の当日だったか前日だったか。研究になりそう、という予感を信じて進めていたのに、これが発表の際にどんな形になるのか、果たしてそれで研究といえるのか。今まで漠然と抱いていた不安が、おばけになって立ち現れたのだ。助けがほしい。そういう時は、私の好きな人・尊敬する人の言葉に触れに行く。有隣堂の、詩のコーナーが私のエナジーになることは、卒論の頃から知っている。まるで・・・まるで・・・なんだ？こういうのって、なんて表現すればいいんだろう？あそこに行けば、きっと何かが歩を進めてくれる、という存在・・・師匠かな。気持ちも考えもまとまっていない状態でたどたどしく話しても、“それはいい予感だよ”と認めてくださる。*
> *そう、今日もまさに、この時と同じ状況で、これを手に取った。これが研究といえるのかもやもやぐにゃぐにゃとずっと考えてきて、あ、ちょっと師匠のお言葉に触れよう、と思って、佐伯先生の『認知科学の方法』のまえがきと最初の九ページを読んだ。そしてもう一冊、何か今の私を認めてくれそうな、あるいは歩を進めてくれそうな、これを手に取った。*
> *やはり、あなたはいつも一緒に行動する相棒ではないのだ。そこにずっといてくれるという事実に日々安心し、本当に困ったとき、迷ったときだけ、開くもの。何かの切り札のような。*
> *こうして考えてみると、あなたと私のこういう関係が、まさに跡に表出されている。*
> *あなたの周りは激しく動いていて、私はその激しい中で悩んだり、もがいたりしているのだけれど、あなた自身・あなたの中身は変わらずそこにいて、しおりもわずかにも動かさずにずっとそこにいてくれて、そのままくっきり跡がつくほどどっしりと構えてくれている。*
>
> 記述名：「ハードカバーの本に、栞の跡」

M のこの記述では、ものとの関係性を改めて解釈し直し、ほかの出来事（別の本を手に取ったこと）への連想で思考を広げる。連想で「思い出」を引き出し、自らの感情も見つめ直しているとも言える。

記述をする際に「どこまで、いつまで観察するか」が問題になる。これは、そのときに関連すると感じたところまで行う。もちろん、自ら関連を見出し、記述できること以上に、人生の「経験」として繋がりがあるだろう。しかし、それらに気づく糸口を見つけるには、自分視点で見られる世界のありように気を配り、目の前にあるものについて、まるごとの私が何を感じているのか、じっくりと観察することが不可欠である。

5. 「跡」についての言葉を生み出す

5.1 相互作用をつくるタグ

約4ヶ月間、プロジェクトメンバー2名がそれぞれ「跡」を集め、合わせて 114 の記述を集めた。

そして、集めた「跡」を用いて、改めて「跡」に関係する言葉を作り出すことを目指しタグづけを行った。記述同士に新たな言葉同士の関係を見いだすことで、「跡」について語るための新たな観点や、「跡」についてのキーワードをさらに生み出すことを試みるためである。

一般に、タグと呼ばれるものは複数の記述を分類するために付与するものであると考えられることが多い。そのため、タグに用いられる言葉の抽象度を揃えたり、網羅性があるかどうかについて検討されたりする。

しかし、本研究においては、まずは、記述に「ふだ」をつけるという簡素な考え方に留める。

体験についての語り・解釈に、さらに重ねて短い解釈（＝タグ）を重ねてつける。その短い解釈は、その記述自体についての要点をかいつまんだものであったり、「解剖」したり「解説」したりするもののこともある。もしくは、他の記述との関連性から、相対的に生まれることもある。あえて秩序づけず曖昧にし、さまざまな解釈の可能性を探る。

また、さまざまな言葉を付与することで、他の記述や他の記述についたタグ、自らの記憶などとさらに結びつき、解釈がうまれる。

タグを用いた分類の作成は、記述同士に見出される差異をもとにはっきりと切り分けることである。

本研究においてのタグの主な役割は、「ふだ」を作る作業の中でさらに新たな考えや着眼点を生むことである。また、タグと記述、タグとタグ、タグとプロジェクトメンバーの相互影響・作用を作り出すことである。

5.2 タグの作成とグルーピング

記述を改めて読み、描写された体験、そして記述自体について重要と思われる点を「要点タグ」として拾い上げる。

例えば、4 章で挙げた「ハードカバーの本に、栞の跡」に記述した本人である M がつけた要点タグが、以下の9つである。

表1 「ハードカバーの本に、栞の跡」の要点タグ

時間の跡
跡をつくった要素　密度
跡がついた瞬間
跡から想起される激しさ
わたしとそのモノの関係性を物語る
跡を介してわたしとそのモノの関係性が変化する
微動だにせずにそこに居続ける
跡をつくった要素　本棚のひしめき
過去の瞬間を想像する

このように、各記述について、記述者自身が要点タグを付与した。そして、要点タグ名どうしで似ているものを集め、タググループを作成した。

114 の記述にそれぞれ要点タグを付与し、167 種類を作成した。167 種類の要点タグをグルーピングすると、タググループが 38 個作成できた（その一部を表 2 に掲載する）。167 種類のうちの 65 種類は、38 のタググループには属さず、単独のまま残った。

表2 タググループ名と含まれる要点タグ名（一部抜粋）

タググループ名	要点タグ名
そのまま居続ける	変化しない跡
	過去を閉じ込めている跡
救い	跡に救われる
	跡が味方してくれる
継ぐ	語り継ぐ
	受け継ぐ
正体が見えない・わからない	軌跡がわからない
	どこについているのかわからない跡
	見えない
	跡に気づかない
	跡に気づけない

表3 タググループ名一覧

そのまま居続ける	跡のつき方
残らない	当たり前
正体が見えない・わからない	想像する
今まで気づかなかったものを顕在化させる	なかったことにする・しない
跡に対する積極的な働きかけ	音
跡？	時間とともにつくる跡
わたしと跡	跡の要素に関する考察
思い出す	言霊
重ねる・重なる	ちから
流れ	他者、だれかの跡
比べる	過去の重み
跡の痕跡	生命力を感じる
余韻	場所
「いつも」が変化する	たどる
継ぐ	残る
境界	安心する
そこで何かが起こった	限定・刹那
未来を感じる	染みる
救い	跡を残したいという欲

タググループを作成したことによって、タグ同士、そして記述同士の相互関係が明示され、共通点や差異を見いだすなど、新たな考えが生まれやすくなる。また、タググループ名も新たな着眼点としてはたらく。

6. おわりに：今後の課題と期待

筆者らは、本研究を「跡の物語」と題して進め、何気ない日常の中のシーンを描写した「物語」の深さと豊かさにこだわりを持ち、言葉や視点を増やす試みを行ってきた。

しかし、「物語」としての豊かさを求めるあまり、まとまりが良いとは言えない方向に向かってしまったことは否めない。思考は発散の方向に向かい、「跡とは？」というはっきりした答えを出すことはできなかった。濃く広がった「跡」についての物語を、どのように周りに提示するのか。また、どのように人々からの反応を得られるような仕組みを作るのか。これらが大きな課題だ。

しかし、課題と同時に、広がった物語を集約しすぎずに伝えることが何を生むのか、期待も生まれた。以下、K の記述である。

> *傘の持ち手の傷。まだ 2 回しか使っていないのに、もう傷がついてしまっている。ピカピカの新品は、ほんの一瞬しか保たれない。跡が積み重なって、新しいが古いになるのはいつなのかなあ。この傘だけじゃない、いろんなものがどんなふうに「古い」になるのか興味が湧いてわくわくした。*
>
> 記述名：「新しい傘」

1本の傘に傷がついた。その出来事を、「いろんなもの」に広げ、「古さと新しさの境目」に考えを巡らせる。このようなことを、わたしたちは度々行っていないだろうか。世界には、違っているが「似ている」ことがある。ひとつの出来事から、おおきな考えに可能性を広げることができる。

筆者らが惹かれているのは実は「跡」ではなく、さらにおおきな何かかもしれない。「誰もが生きながら物語を作っている」[小川 2008]と捉えると、何気ない日常の「普通」と思っていることに、自らの生きる物語を広げ、伝える手がかりを見つけることができるのではないか。

参考文献

[中村 1992]中村雄二郎：臨床の知とは何か, 岩波書店, 1992.
[小川 2008]小川洋子 ： 二人のルート, 河合隼雄・小川洋子著『生きるとは、自分の物語をつくること』, 新潮社, 2008.
[佐伯 2007]佐伯胖：認知科学の方法, 東京大学出版会, 2007.
[諏訪・藤井 2015] 諏訪正樹, 藤井晴行:知のデザイン, 出版社 , 2015.
[谷川 2015] 谷川俊太郎：詩に就いて, 思潮社, 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

知識・技術・技能の伝承支援に関する考察

メタ的な知の表現と認識

Consideration for succession of knowledge, technology and skill

樽田 泰宜[*1,2] 柳原 敏[*2] 井口 幸弘[*1,2]
TARUTA Yasuyoshi YANAGIHARA Satoshi IGUCHI Yukihiro

北村 高一[*1] 手塚 将志[*1] 香田 有哉[*1]

KITAMURA Koichi TEZUKA Masashi KOUDA Yuya

[*1] 日本原子力研究開発機構 Japan Atomic Energy Agency
[*2] 福井大学 University of Fukui

We focus on a succession science for “knowledge, skill and technology (KST)”. Topics of KST are including issue of transfer of KST. Person of expert has a too many experience of KST. However, nowadays-social situation is concerning a loss of that information when expert retire. KST science need to clarify a definition of KST. At this paper carried out to propose a framework to understand of knowledge of tacit, explicit and skill.

1. はじめに

人工知能研究の一つに，特に産業ドメインを対象にした「知識・技術・技能の伝承支援研究会」（SIG-KST）がある．この背景には人口構成比の大きな団塊の世代が大量退職することで知識・技術・技能（知識等）が喪失すると懸念されておりこの解決を目指すものである．団塊の世代は，産業ドメインに長く従事しており熟練または熟達した知識，技能，技術等を保有している．しかし，これまでは OJT で実施されていたような伝承が拠点を国外へ移転するなど従来通りの方法だけでは対応できなくなっている．同時に技能などの一般的に形で表すことが難しいものの伝達にも課題がある．これは身体性や身体知に関わる課題でもある．言語化が難しいものをうまく，より正確に伝承する仕組みを構築することが必要である．先行研究では，具体事例に対する身体知や業務知識等を伝承または継承するための情報システムや身体性の可視化技術など多方面から実践的にアプローチされている．

一方，伝承研究には知識等をコンテンツ化や汎化させて，多くのドメインでより効果的に活用することが必要であり，その意味においてナレッジマネジメント(Knowledge Management; KM)という側面も有している．そのため，対象を明示的な形で，例えば形式知(Explicit knowledge)として明確に表現することが求められる．すでに明示的に存在しているデータ・情報の利活用は他分野の技術も応用することでそれほど困難ではなく解決することが可能になるが，暗黙知(Tacit knowledge)と呼ばれるような属人的な知識をいかに扱うのかは困難な課題である．また，単に明示的に表現するだけではなく，共有したり，そこから新しい知識を創造したりすることも必要であり，KM の研究対象といえる．KM には二つの側面として「知識管理」と「知識経営」という理解がある[梅本 12]．知識管理時代（1980 年～）は，計算機や情報技術の発展により，計算機を活用したコミュニケーションやコラボレーションの促進に注目されていた．スケジュールの管理やそれを共有機能などのグループウェアの利用である．他方，野中らは KM には単なる知識と情報の管理だけでない側面があるとしている[Nonaka 95][野中 96]．この文脈では 1970 年代からの組織論やグループダイナミクスを源流に科学哲学も加味した知識管理に変わるような知識経営（1990 年～）を提案している．組織的知識創造モデルである SECI モデルが有名である．

こうした背景のもと知識等の伝承に関連して，対象となる研究課題や知識等を体系化する試みも提案されている[古川慈 14,15]．SIG-KST の研究報告を対象に視覚的な語彙を用いた体系化の提案研究である．この研究では，伝承研究には知識・技術・技能を捉える視点が重要であり，暗黙知や形式知の関係性の考察や類型化の精度を高める点が重要であると指摘している．これを受けて言語化と表現・表出の視点から知識・技術・技能の伝承研究をメタな視点で再構築した表現モデルが提案されている[樽田 18]．このようなメタ的な研究は多くはないが，知識・技術・技能の認識やその伝承における重要な研究課題である．同時に，知識・知能とは何かという科学や人工知能の命題の一つに迫るものである．

こうした研究には，知識・情報を引き出す，形にする，それらを活用するというように大分出来るが，人間の介入があるため人間の身体性とは切り離せない側面もある．例えば，「身体感覚と言葉のつながりこそが，人工知能研究の中心的な課題の一つであるシンボル設置問題と直結している」との指摘である[古川康 12]．伝承というように何かを伝達する，してもらうには，認知的課題と共に外的に表象することが重要である．そこで，古川康一らはこれまであまり盛んではなかった言語と身体の関係を扱う人工知能研究を学習や熟達化ではなく創造支援として捉え直している[古川康 07]．この研究は，本来言語化を行うことが難しい“身体知”をあえて言語化しようとする試みが，身体知の獲得を促進するという諏訪[諏訪 05]の仮説に基づくものである．このように身体知においても「創造」が重要視されており，これは KM の知識の創造や創発といった分野とも関連性が高いといえる．なお，諏訪が提案しているメタ認知またはメタ認知的言語化

連絡先：樽田泰宜，日本原子力研究開発機構バックエンド研究開発部門原子炉廃止措置研究開発センター，福井県敦賀市名神町 3 番地，0770-26-1221, taruta.yasuyoshi@jaea.go.jp

は，近年，提案されている一人称研究という新しいパラダイムの下地ともなっている[諏訪 13][藤井 13].

そこで，本研究では，知識・技術・技能の伝承や身体知をメタ的に捉えることで人工知能における当該分野の発展を目指すため，樽田が提案する知の表現モデルを既存の学術知見を基に解釈し考察することを目的とする．これは，知識等を伝承するということに対して，個々の課題に対する研究アプローチではなく，研究課題を俯瞰的に捉えるような研究アプローチを提案するという立場である.

2. 暗黙知，身体知

本研究で重要な点として，知識の表現と認識がある．例えば，暗黙知を語ることのできる明示知への変換の要請は身体知や伝承研究以外でも高い．そこで参照される理論として KM で野中らが提唱する SECI モデルがある．同モデルは，集団の知的な創造活動を対象に，個人の集団の相互作用や“よい場”によってこれまで不確かだった暗黙知が形式知に変換されるとする理論である．この理論はマイケル・ポランニーの「知」に関する議論が参照されている．ポランニーは，語ることのできない知識を暗黙知(tacit knowledge)と呼んでおり身体知は暗黙知であると表現している[Polanyi 58][マイケル 80]．一方の文章での表現や言語で表出できる，語ることのできる知識は形式知(explicit knowledge)と呼ばれる．そのため本研究では，身体的な行為や活動を自分自身で語ることはできない身体知は暗黙知に含まれるとする．さらに，暗黙知の原義にしたがい，語ることができるものと語ることが難しいものを厳密に分けて議論をする．これは，よくあるミスリーディングとして指摘されているが，本来の暗黙知は「語ることが難しい」ものではなく「語ることが出来ない」とするためである[大西 07]．この理由は不確かだが，原義と訳語の関係性や日本語と他の言語が持つ意味や概念の関係性も考えられる．そのため先行研究では，この「できる」「できない」に関してやや曖昧に捉えられていたが，本研究ではそれらとは異なる立場として，身体知は暗黙知の一種であるが，語ることができない知識という立場に立つ．そうすることで，語ることのできる知識，つまり，形式知とは異なる点を強調したい．このことは，暗黙知を多義的に捉えることで生じる，認識の齟齬を防止やコミュニケーション，ディスカッションにおける相互理解を促進するものと関連する.

さて一方の身体知であるが，ベルシュタインは身体運動の科学（The Science of Human Movements）に関して早くから身体性や身体知の重要性に言及していた．身体知の研究の難しさに関しては，後の訳書で Latash がベルンシュタイン問題と呼ぶ冗長自由度や文脈に関する問題（あるいは不良設定性）を指摘している[Bernshteĭn 96][ニコライ 03]．これは身体の運動を物理的にロボットなどに置き換えた場合，自由度が高すぎたり，一見すると同じような運動でも文脈により異なる運動を行っていたりと，身体運動を一意に決めることができない問題である．先行研究では，演奏家や表現者の振る舞いの定性運動力学モデルなどもあるが，他の視点では，身体知を獲得するためにはメタ認知的言語化が重要な役割を担うという指摘がある[諏訪 05][Suwa 08]．身体知の獲得に関して，Magill が主張するようにスポーツ科学での学習，つまり身体知の獲得と向上は言語的な意識とは関係なく進むとされていた[Magill 98]．しかし，学習者がメタ認知的言語化を行うことで自動化フェーズを再構築や，古川ら[古川康 07]の言うようなメンタルモデルをより望ましい形へと再構築が行われることで習熟が進むというのである．ここでいうメタ認知は，一般的な客観視や内省，内観といったシステムを外界から捉えようとする行為ではなく，メタ認知的言語化を実施することで身体や思考を制御するのではなく，システムの中，身体や環境の中に体感できることを増進させるツールとしての位置づけであるという[古川康 07]．この議論は，郡司らの研究パラダイムも関連している[郡司 97]．郡司らや松野[松野 00]は自然科学の観察主体と観察対象の分離したいわば外部観測と呼ぶべき方法論とは反対に内部観測という対象を内部から観測する方法論を提案している．諏訪はメタ認知的言語化の具体例を，外部観測による分析の理論とも呼ぶべき熟練者と素人の差と熟達者の身体知の獲得プロセスに着目した行為者の理論からも説明をしている． なお，松野は内部観測を「経験世界の内に現れる個物は何であれ，他の個物と関係を持つとき、相手から受ける影響特定できる限りにおいて，その相手を同定する．しかも相手を同定する，とするこの観測はこの経験世界の内で絶えることがない．何が何を観測しようとも，その観測は後続する，果てしのない観測を内蔵する．これを内部観測という」と説明している.

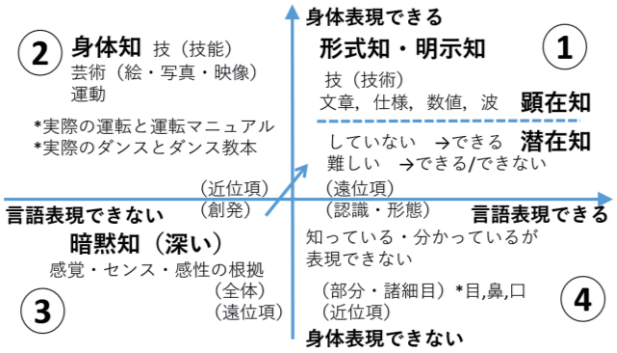


Fig 1. 知の表現モデル

3. 知識・技術・技能の伝承の表現モデル

次に，ポランニーの暗黙知，古川康一ら，諏訪らの研究を参考に，知識・技術・技能の伝承に関する知の表現モデルについて論じる．既存のモデル[樽田 18]では，「言語化できる／できない」と「表現・表出できる／できない」の二軸で捉えていた．しかし，言語も表現や表出の一形態であるのため，ここでは，「言語表現できる／できない」と「身体表現できる／できない」と捉え直し議論を深める．それぞれの象限では，第一象限が，形式知・明示知・顕在知・潜在知が付置され，第二象限には身体知，第三象限には（深い）暗黙知，第四象限には認識・諸細目・部分に関するものが付置される.

第一象限では，潜在知とこれと対する顕在知を付置した．これは，「言語化していない知識」は，「言語化できない知識」とは異なり，単に外に表現，表出されていないだけであるというポランニーに対する大西の議論を踏襲し新しく定義するものである[大西 07]．つまり，言語化していない，単に忘れただけの事柄は潜在的には知っている，分かっているという状態であるため潜在知とし，それらが表出化されると顕在的な知識となると考えるものである．明示知は形式知のことである．暗黙の反対は非暗黙であり，形式というより明示であるといえる.

また，森は技の伝承という文脈で技術と技能に分けて整理しており，技能は第二象限に，技術は第一象限に付置した[森 05]．単に技(skill)は身体に大きく依存する能力であり，特に技能と言った場合にはカンやコツといった明示的に表現や伝達ができないものが含まれる．一方，技術(technology)は仕様書など明確に表現できるものが対象である.

4. 議論

第一象限にはいくつか新しい用語を提案した. その中で図中には提示しなかったが例えば, 形式知の訳語には陽表知というものもある. 陽表とは数学の関数で使われる概念で, 陽表式と陰伏式などで用いられている. この場合だと, 野中らのいうように形式知の背景には膨大な暗黙知があるということが非明示的に意味するといえる. 野中らは創造的な行為や活動における知識の創造に着目しており, この意味においては言葉には出来ないが, 頭の中の曖昧とした暗黙とした思考があり, これを言葉に出来たときに暗黙知が形式知となると考えられる. つまり, 文脈への依存度が高い状態である. このモデルはそうした点の重要性は認めるも, より汎化した表現とするために言明しなかった.

さて, この表現モデルから伝承研究で可視化されている知識・技術・技能を考察すると, 第一象限に付置される知識等は知識管理や人工知能, 情報科学の知見が十分に活用できる分野である. しかし, 第二象限の身体的な知識等の伝承が非常に困難であり, 同分野の研究を難しくしている. 既存研究では様々な可視化や情報システム化が進められている. 本モデルに従うと, 観測者は身体知としての暗黙知を可視化できている訳ではなく, 表現者(観測対象者)の遠位項として認識可能なものを近位項の身体知として近似していると考えることが出来る. 表現者は言葉で身体的な動きを表現できるが, それは認識の表層に現れている知覚可能なものが対象となる. 手を上に上げる動作, 歩くと言った行為, 無意識の呼吸などは, 言語化できるが, 暗黙知は身体的にその動作をどのようにしているかは語ることが出来ないからである. 自転車の乗り方は, サドルにまたがる, ペダルをこぐ, といったように説明できるが, どのようにして安定して乗ることが出来るかは身体的には語ることが出来ない. ここでは, ギルバート・ライルのいう“knowing-how”と“knowing-that”という捉え方ができる. ”knowing-how”は一般的にはノウハウとして認識されるが, どうするかを知っているということであり, 手続き的, 遂行的知識やその方法でもある. ”knowing-that”は対象を知っているという事実的, 宣言的, 記述的な知識のことを指す. をするといった点は “knowing-that”として事実は知っているが, “knowing-how”として具体的にどのようにするのかは語ることができない. これを端的に表すのが“we can know more than we can tell”(我々は語る以上のことを知っている)というポランニーの言葉である. 自転車の乗り方の手続きは“knowing-that”として明示化可能であり, 重心を傾けたりペダルのこぐ速度上げたりバランスを保つなどの”knowing-how”も一般的には語ることが出来る. しかし, 身体としてそれをどのようにどうさせるのかはやはり暗黙的であるといえる.

第三象限は, 言語化も表出化もできない「深い」暗黙知とも呼ぶべきものが付置され, 無意識的な認知や認識が考えられる. 例えば, 人の感覚やセンス, 感性の根拠などは説明できない. 五感に関するものを数値化しても, おいしい, 美しい, いい匂い, 赤色に見えるなどそのように感じる根拠を自分の言葉で明示的に説明することは出来ないといえる. 相対的には語ることができるが, その指標は表出化が不可能で, 共有も互いの言葉を信じるしかない. 第四象限は, 形態とも呼ぶべき認知に関するものが付置されると考えられる. 通常, 逐行的言語のように言語化ができれば表現や表出化は可能であると考えられるが, 逆の事例を見つけることはできない. しかし, 人の顔の認知やその判断など「知っている」「分かっている」が表現できないものがあると考えられる. 部分を認識したり理解したり表現することはできるし, 目や鼻や口といった諸細目を認識することはできる. しかし, 全体としてそれを認識するとき, 人間の顔である, 男性である, 女性であるという判断の結果と根拠は, 表現できなくなり, 深い暗黙知へと移行すると考えられる. このような点は認知や認識に関わる部分である. ポランニーは「ある対象の外形を認識するとき, 私たちは感知している個々の特徴を, それが何とは特定できていないままに, 統合している」のだと説明している. また, ポランニーは, 個々の顔の特徴を諸細目として, 諸細目という部分に注目することと全体の認識は異なると説明している. これが「それが何とは特定できていないままに, 統合している」ということであり, これも暗黙的な認識であるといえる. 部分と全体の関係である全体は単なる部分の寄せ集めではなくそれ以上であるという創発に関する議論とも関係性がある. さらなる議論が必要であるが, この考えを正しいとするならば創発は, 第三象限から第一象限への知識の変換過程の一つであるとみることができる.

一方で, 身体スキルの理解と困難性に関しては, 第二象限の近接項が言語化することができない点が理由であると解釈できる. 身体性の非線形性や文脈依存性に関しては, 議論を別の機会に譲るが, 身体知の暗黙性に関してはエキスパートシステムで専門知識(expertise)の抽出が行われてきた. しかし, インタビューによる生産性の低さ(大量のデータの獲得の難しさ), プロダクションルールの作り方などに課題もあった.

そこで, 諏訪が提案するメタ認知的言語化から解釈を試みる. 諏訪らは, 長期間, メタ認知的言語化を実践的に行っている. 本研究の第 2 象限にある身体知は, 熟達者のすでに獲得しているエキスパートの知識である. 当事者は, 自らの身体的な行為を言葉にするという外的表象化を言語で行うが, この表現された言語化は, 外部から観測できる遠位項として存在する. 伝承研究や身体知研究で, 外部観察的な研究では, 身体知の外的表象化された遠位項を対象に観測したり, 記述したりしていると解釈できる. 諏訪らは, 身体的なタスクを言語化すると派フォーマスト言語化の量には U 字曲線の相関があることを発見している. 本モデルで解釈すると, 初学者が自らの行為を言語化するが, それを体で表現できないのは第 4 象限であり, パフォーマンスの向上とともに身体の動きとより正確な言語表現を獲得し, 習熟度が高まると Magill の例や Beilock のゴルフの言語意識のように遠位項であった表現を自らの内在化されたものとして言語化ができなくなり, 熟達した身体知を獲得すると解釈できる.

この内在化に関しては, 伝承では, 暗黙的である部分をどのように互いに共有するのかはポランニーの指摘がある. 外在化されているような遠位項や明示的な知識を自分の元として獲得することで暗黙知が共有化されたり伝承されたりするといえる. このとき内在化(dwell in)が起こっているとポランニーや野中らは説明している. とこれらの行為の可視化には暗黙知の共有が必要であり, 内在化された認識を共有しなければならない. この問題として, 科学の客観性やトーマス・クーンのいう反証可能性が発生する. 暗黙知を暗黙知のまま認識することはできてもそれを証明したり反証したりすることが難しくなる. それを避けるには, 研究の粒度や対象の切り取り方をより明示的に宣言することが必要である.

5. まとめ

伝承研究では, 対象となる知識等の多くは文脈やドメインに依存していると考えられる. 産業ドメインでは, 特定の文脈として産業や業務に依存した業務知識がある. その中には, ベテラン, エキスパートといった熟練者や熟達者は, 専門知識や業務知識の量が豊富にある. そのため, 自らの行為を説明する表現も多く持っていると思われるが, 習熟度が進むとその判断や行為を

適切に説明するための言語化が出来なくなるというのは，知識などが内在化されるとためと解釈できる．

本研究の限界として，メタ的な方法論について論じたために具体的な方法や実践には言及していない点にある．また，知識の静的な側面を提示しており，動的な部分として文脈は十分に表現できていない．例えば，2 次元軸から 3 次元軸へと表現を拡張することで対応可能な部分もあるが，端的な理解の促進という面では単純に次元数を多くする点についての利便性について考察が必要である．また，伝承という視点では，オープンスキルやクローズドスキルといった他者と協調する場合や与えられた課題をこなす時に関しても言及できていない．人工知能研究の発展によりこれまで考えられなかったことが AI で続々と可能になっている．しかし，それと同時に計算機の内部が分からなくなるという不思議な現象に直面している．我々は AI の仕組みは知っているが，なぜうまくいくのかという問いにはうまく答えることができない．AI の理解や認識においても人間と同様に暗黙知が生じているのであろうか．これは単なるブラックボックスではなく，諸細目として個々の部分の仕組みを理解しても全体として何であるのかが語れないことに似ている．一方で伝承研究では，様々な計測機器の発展により人間活動だけでなく脳活動の可視化も進んでいる．既存の要素技術の組み合わせで解決できる課題もあり実際に多くの成果を上げているが，伝承研究の本質である人間や，人工知能研究の知に迫るためには，本研究で提案するような概念的な理解やモデルの構築といった研究も有効であると思われる．具体の実践と概念の両輪に着目した研究の一端を担うことができたが，残された課題や限界も多くある．今後も継続的に議論を深めていく．なお，本研究は JSPS 科研費（JP17H07354）の支援を受けた．

参考文献

[Bernshteĭn 96] Bernshteĭn, N. A,Mark L. L. (Eds.). Michael T. T. (Eds.): Dexterity and its development: Secondary Dexterity and its development, Lawrence Erlbaum, (1996)

[Magill 98] Magil, R. A: Knowledge ls More Than We Can Talk About: Implicit Learning in Motor Skill Acquisition, Research Quarterly for Exercise and Sport, Vol.69, No.2, pp.104-110, (1998)

[Suwa 08] Suwa, M.: A cognitive model of acquiring embodied expertise through meta-cognitive verbalization, Information and Media Technologies, Vol.3, No.2, pp.399-408, (2008)

[Nonaka 95] Nonaka, I. and Takeuchi, H. : The knowledge-creating company: How Japanese companies create the dynamics of innovation, Oxford university press, (1995)

[Polanyi 58] Polanyi, M. : Personal Knowledge: Towards a Post-Critical Philosophy, University of Chicago Press, (1958)

[ニコライ 03] ニコライ・A・ベルンシュタイン，工藤 和俊(訳). 佐々木正人(訳): デクステリティ 巧みさとその発達，金子書房，(2003)

[マイケル 80] マイケル・ポラニー.佐藤敬三（訳）: 暗黙知の次元，紀伊國屋書店，(1980)

[野中 96] 野中郁次郎.竹内弘高: 知識創造企業，東洋経済新報社，(1996)

[郡司 97] 郡司ペギオ幸夫，オットー・E.・レスラー，松野孝一郎: 内部観測，青土社，(1997)

[古川康 07] 古川康一，諏訪正樹，加藤貴昭: 身体スキルの創造支援について，人工知能学会論文誌，Vol.22, No.5, pp.563-573, (2007)

[古川康 12] 古川康一: 身体知獲得のための脳情報処理と言語指示，知能と情報，Vol.24, No.1, pp.2-7, (2012)

[古川慈 14] 古川慈之: 知識・技術・技能の伝承支援に関する考察-SIG-KST 講演内容の分類-，知識・技術・技能の伝承支援研究会，Vol.23, No.SIG-KST-2014-02-04, 2014, (2014)

[古川慈 15] 古川慈之: 知識・技術・技能の伝承支援に関する考察-暗黙知と形式知との関係-，知識・技術・技能の伝承支援研究会，Vol.24, No.SIG-KST-2014-03-03, (2015)

[松野 00] 松野孝一郎: 内部観測とは何か，青土社，(2000)

[森 2005] 森和夫: 技術・技能伝承ハンドブック，JIPM ソリューション，(2005)

[諏訪 13] 諏訪正樹: 見せて魅せる研究土壌: 研究者が学びあうために，人工知能学会誌，Vol.28, No.5, pp.695-701, (2013)

[諏訪 05] 諏訪正樹: 身体知獲得のツールとしてのメタ認知的言語化，人工知能学会誌，Vol.20, No.5, pp.525-532, (2005)

[大西 2007] 大西幹弘: 暗黙知とは何か(1)，日本ﾅﾚｯｼﾞ・ﾏﾈｼﾞﾒﾝﾄ学会東海部会季報，Vol.3, pp.1-7, (2007)

[樽田 2018] 樽田泰宜: 知識・技術・技能の伝承支援に関する考察 -言語化と表現化からの関係-，知識・技術・技能の伝承支援研究会，Vol.33, No.SIG-KST-2018-0305, pp.1-6, (2018)

[藤井 13] 藤井晴行: 創造という行為の研究について，人工知能学会誌，Vol.28, No.5, pp.720-725, (2013)

[梅本 12] 梅本勝博: ナレッジマネジメント-最近の理解と動向-，情報の科学と技術，Vol.62, No.7, pp.276-280, (2012)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

鑑賞中の気づきを促す構成展示法について

A Proposal of the Constructed Exhibition

奥村美奈代 *1　保科智治 *2　木村健一 *3
Okumura Minayo　Hoshina Tomoharu　Kimura Ken-ichi

*1公立はこだて未来大学大学院システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate

*2市立函館博物館
Hakodate City Museum

*3公立はこだて未来大学
Future University Hakodate

We proposed "the constructed exhibition". This exhibition method was collaboration between "a traditional exhibition" and "a hands-on exhibition based on constructivism" and "Visible storage". As a result, beholders were aware of interesting facts by "the constructed exhibition".

1. はじめに

従来型の展示手法はガラスケースに資料を並べた「資料陳列型展示」があり, 学習とは講義を受けることに等しいと考えからキュレーター（学芸員）がキャプション等を用いて専門的知識を伝えている [Caulton 2000]. しかし, 鑑賞者各々の興味や関心, 既有知識はさまざまであり, 受け止め方がそれぞれで違うことが明らかになり, 展示の方法に関して再検討が行われている [小笠原ら 2012].

新しい展示方法として鑑賞者が資料を触るなどして体験できる「ハンズ・オン展示」が導入されている. 作り手が伝えたいメッセージは明確に定められているが, 利用者が展示との相互関係を通じて, 誰かに教えられるのではなく, 自ら教育目的を発見する造り [Caulton 2000] になっている.

加えて近年では「ヴィジブル・ストレージ（目に見える収蔵庫）」と呼ばれる設備の一部に透明な壁を設け, バックヤードの収蔵品も見えるようにする展示形態がある. 普段見ることのできない学芸員が作品を入れ替える様子などを可視化することで隠れた面白さを感じてもらうだけでなく, 資料を保護し管理する機能に対する人々の理解を促している.

ハンズ・オン展示やヴィジブル・ストレージは構成主義の学習理論に基づいており, 鑑賞者が展示に関する気づきを言語化し, 展示の作り手や他の鑑賞者と共有しながら展示の意味や資料を理解していくことを促している.

2. 研究の目的

本研究では, ヴィジブル・ストレージで試みられてきた方法を資料陳列型展示とハンズ・オン展示を組み合わせた展示手法に用いる「構成展示法」を提案し, 展示の作り手が伝えたかったメッセージや専門的知識を正しく伝えながら, 知識や経験の度合いの異なる鑑賞者がそれぞれ資料や展示と能動的に関わることでその人なりの気づきを促すことを目的としている.

現在までに「掛け軸」と「屋号 *1」をテーマにした構成展示法の鑑賞学習効果を検証しており, 本論では主に屋号に関する報告を行う（図 1）.

連絡先: 奥村美奈代，公立はこだて未来大学大学院 システム情報科学研究科，g2117013@fun.ac.jp

*1　一門・一家を示す称号のこと

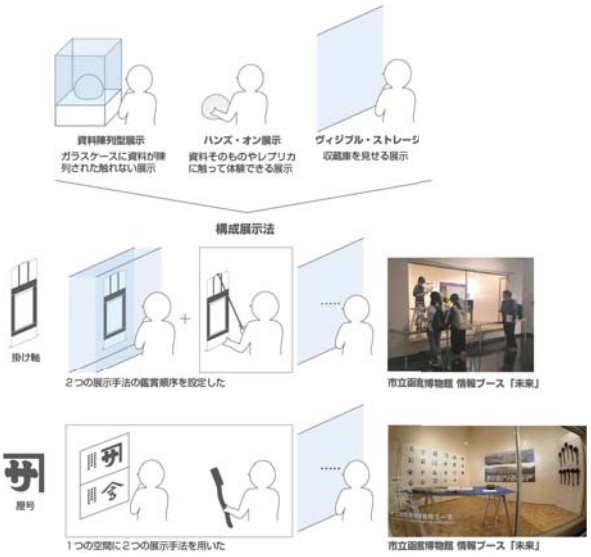


図 1: 構成展示法の概要図

3. 展示空間の制作

「屋号」をテーマにした展示を市立函館博物館の情報ブース「未来」にて 2017 年より 1 年半開催している. 開催にあたり, 筆者がデザイナーとして関与し, 構成展示法で展示を制作した.

3.1 情報ブース「未来」

情報ブース「未来」は, 市立函館博物館と公立はこだて未来大学の連携事業として 2014 年に開設されたもので, 同館 3 階第 3 展示室内の展示ケースを, 鑑賞者が中に入れるハンズ・オンの要素を取り入れた展示空間として活用されている. ガラス面から外の鑑賞者が中に入った鑑賞者の様子を鑑賞できる構造になっている（図 1）. これまでに今回と同様の展示を 3 度行なっており、それぞれ別のテーマを扱っていた [保科 2017].

3.2 展示物

展示物は資料陳列型展示とハンズ・オン展示の 2 種類を組み合わせた. 資料陳列型展示の要素は, 各屋号の所在地や商店名を表記した解説パネル, 現存する屋号と明治 25 年に存在した

屋号を比較できるパノラマ写真と, ポンポン *2 の実物である. ハンズ・オン展示の要素は, ポンポンの実寸大レプリカと, 屋号の読み方と形を一致させる屋号はんこである.

4. 観察の手順

制作した展示空間を用いて２度の観察を行なった. 対象者は大学生２名（A,B）と大学生３名（C,D,E）であり, それぞれグループ１, グループ２とした. 展示室の中を自由に鑑賞することを教示し, その様子を観察した. 鑑賞中は動画の撮影と音声を録音し, 鑑賞後に行なったインタビューと合わせて発話を書き起こした.

4.1 分析の観点

発話を分析する際の観点として作品の再解釈（見立て）に関する発話の有無を検証した. 再解釈は創造活動に効果的であるとされている [石橋ら 2014] が, 作品を理解するための行為として行うため, 鑑賞活動においても効果があるとする. 発話例としては「〜みたいだ」「〜に似ている」がある.

5. 観察結果と考察

5.1 鑑賞の順序の違い

２グループの間で鑑賞の順序の違いが観察された. グループ１は資料陳列型展示を鑑賞後に, ハンズ・オン展示を鑑賞していた. ハンズ・オン展示に含まれる屋号はんこを最後に取り組んでいたことについてインタビューしたところ, 以下のように話していた.

被験者 B　なんか変な…なんだっけ. ウロコとか, なんか変な特殊なやつ系が特に最初に入って見たときよりも, それが逆に覚えれてて.
被験者 A　うん. 印象がちがった.
被験者 B　わたしは覚えてるぞっていうのを, このはんこで試せて楽しかったというのはありました.

資料陳列型展示で得た知識を再確認したことで, 屋号に関する理解を深めていた.

グループ２はグループ１とは反対に, ハンズ・オン展示を鑑賞後に資料陳列型展示を鑑賞していた. 展示室に入室直後から屋号はんこに取り組んでおり, その中で屋号の読み方に法則性があることに気づいていた.

グループ１, グループ２ともに, ハンズ・オン展示を鑑賞する中で屋号に関する気づきや理解が促されていた. 構成展示法において鑑賞の順序が異なった場合でも同様の効果が得られることが示唆された.

5.2 再解釈に関する発話量の変化

各グループにおいて以下のような再解釈に関する発話が見られた.（一部抜粋）

《グループ１》
被験者 B　32:51　ね！「余」りに見える
被験者 B　33:06　なんか, みんな帽子に見えてきた.
被験者 A　35:48　トトロのトみたい
《グループ２》
被験者 C　13:02　なんかなぞなぞみたいだね
被験者 D　17:59　「含む」にしか見えない
被験者 C　20:35　なんか妖怪の名前みたいだね

*2　屋号を押印するための道具

グループ１での再解釈に関する発話は, 資料陳列型展示を鑑賞中には４回, その後のハンズ・オン展示では 14 回と, 後者の鑑賞で発話量が増加していた. グループ２では, ハンズ・オン展示の鑑賞中では３回, その後の資料陳列型展示では７回と, グループ１同様に後者の鑑賞で発話量が増加していた. このことから, 展示を理解しようとする姿勢になっていたと考えられる.

以前に観察を行なった「掛け軸」の場合でも, 再解釈に関する発話はハンズ・オン展示の鑑賞後に見られ, 後の展示において鑑賞の質が向上することがわかっている [奥村ら 2017]. 構成展示法は展示物に関する気づきの増加を促し, 他の展示物でも応用できる可能性が示唆されたと言える.

5.3 多視点で鑑賞する様子

ハンズ・オン展示に含まれるポンポンのレプリカを多視点で鑑賞する様子（図 2）が各グループで観察された. 行動例としては, 裏返してみる, 匂いを嗅ぐ, 振る, 顔を近づけてよく見る, 実際の使い方を真似るなどの模倣行為が見られた.

これらの様子についてインタビューを行なった際には,「そこにあって持てる方がやっぱり楽しくなる.」「ワクワクする.」「実際に持って裏も見てみたことで, 掃除用具みたいだと感じた.」と回答があった. よって, レプリカを用いた模倣行為や多視点鑑賞は, 鑑賞者をワクワクさせたり, 新たな気づきを生むキーポイントとなっている可能性が示唆された.



図 2: ポンポンのレプリカを多視点で鑑賞する様子

6. 今後の課題

今後の課題として, 鑑賞中の様子をガラス越しに外から鑑賞できる構造を利用した展示形態が鑑賞にどのように影響を及ぼすのかを観察することが挙げられる. 多くの鑑賞者が外から鑑賞している様子は見られているが, 観察が不十分であるため, 今後の長期的な観察で検証して行きたい.

参考文献

[Caulton 2000] Tim Caulton, 染川香澄ほか訳: ハンズ・オンとこれからの博物館, 東海大学出版会 (2000).

[小笠原ら 2012] 小笠原喜康, 矢島國雄, 並木美砂子: 博物館教育論, ぎょうせい (2012).

[石橋ら 2014] 石橋健太郎, 岡田猛: 創造のための「芸術作品の知覚」経験: 模倣に焦点をあてて, 認知科学 (2014).

[奥村ら 2017] 奥村美奈代, 木村健一: ハンズ・オン展示と資料陳列型展示の組合せが与える影響について, 日本デザイン学会　第 64 回春季研究発表大会, PA-28 (2017).

[保科 2017] 保科智治:　研究テーマとしての「情報ブース」第 3 弾,　道南ブロック博物館施設等連絡協議会ブログ,　(2017.04.17 更新)(最終閲覧日:2018.03.15). https://dounan.exblog.jp/27732993/

[2C2] [General Session] 6. Web Mining

座長:柳井 孝介（日立製作所 研究開発グループ）

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room C (4F Orchid)

[2C2-01] Proposal of "Social Porn" Hypothesis and Pre-analysis towards its Observation

○Takeshi Sakaki[1], Fujio Toriumi[2] （1. Hottolink, Inc., 2. The University of Tokyo）

1:20 PM - 1:40 PM

[2C2-02] Communication based on strong preference in Twitter

○kimitaka asatani[1], yasuko kawahata[2], Fujio Toriumi[1], Ichiro Sakata[1] （1. The University of Tokyo, 2. Gunma University）

1:40 PM - 2:00 PM

[2C2-03] Finding Emerging Named Entities from Social Media Streams

○Satoshi Akasaki[1], Naoki Yoshinaga[2], Masashi Toyoda[2] （1. The University of Tokyo, 2. Institute of Industrial Science, The University of Tokyo）

2:00 PM - 2:20 PM

[2C2-04] Effects of Community Structure of a Social Network on the Cascade Sizes of Tweets on Social Media

○Sho Tsugawa[1] （1. University of Tsukuba）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ソーシャルポルノ仮説の提案とその観測に向けて

Proposal of "Social Porn" Hypothesis and Pre-analysis towards its Observation

榊 剛史 *1,*2　　鳥海 不二夫 *2
Takeshi Sakaki　　Fujio Toriumi

*1株式会社ホットリンク　　*2東京大学
Hottolink, Inc.　　The University of Tokyo

In recent years, the negative aspects of information creation by individuals, such as fake news, flaming and echo champer phenomena, have been paid attention . We propose a hypothesis "social porn" as one of the causes of these negative phenomena. We define "social porn" as "information that a user belonging to a specific community wants to diffuse and share instantly".

In this paper, we define the scale of user reaction time as preparation for observation of social porn. We examine the difference of user reaction time distribution for some tweets. As a result, analysis results suggest that the distribution of user response times may differ between tweets randomly extracted and tweets spreaded by specific community.

1. はじめに

2016 年，オックスフォード大学出版局は，2016 年の「Word of the year」に「Post-truth」を選定した．彼らによれば,「Post-truth」とは,「世論を形成する際に、客観的な事実よりも、むしろ感情や個人的信条へのアピールの方がより影響力があるような状況」と定義されている．この語は，フェイクニュースや炎上，エコーチェンバー現象が日常的に発生するようになった現在の社会状況を端的に表した語として，2016 年〜2017 年にかけて注目を浴びた *1.

ソーシャルメディアの登場により，中小規模メディアや個人が情報を発信し，それを流通させるのが容易になった．それらは，ソーシャルメディアが普及し始めた当初はその正の側面としてクローズアップされてきた．しかし，現在では,「Post-Truth」の言葉が端的に表すように，その負の側面に注目が集まっている．例えば，2016 年の大統領選挙では，フェイクニュースが世論形成に大きな影響を及ぼしたとされている．フェイクニュース業者の多くは，PV 数に基づいて支払われる広告プログラムに参加し，そこから利益を得るためにフェイクニュースを発信・拡散させたことを認めている．エコチェンバー現象は，様々な事例について観測されている．その多くは，より衝撃的な内容を伝達することを競いあった結果，より過激な意見が創出されるようになったと言われている *2.

これらの多くの分析は，主に情報の発信者側に着目して，その要因や発生過程を明らかにしようと試みている．では，受信者に着目すると，どうであろうか？彼らは，いや，我々はどのような行動原理によりエコー・チェンバーを引き起こし，フェイクニュースを拡散させ，炎上をさらに加速させてしまうのであろうか?「Post-truth」の定義にあるように，多くの人は「感情や個人的信条へのアピール」に影響を受けてしまっているのであろうか？　本論文では、ここで「ソーシャルポルノ (Social Porn)」という概念を提唱する.「ソーシャルポルノ」とは「社会欲求 (Social needs) を充足する目的で消費・拡散されるコンテンツ」と定義する．ここでの「porn」は「Inspiration porn」や「Patriot Porn」と同様の用法であり，また「ソーシャルポルノ」自体はそれらを包含する概念である．

ソーシャルポルノの存在を仮定すると，前述のソーシャルメディアにおける負の側面，ネガティブな現象は容易に説明することができる．フェイクニュースにせよ，炎上にせよ，エコー・チェンバー現象にせよ，ある個人の持つ意見や感情を強化する内容のコンテンツが配信された時に，その個人の「社会欲求」が刺激されてしまうめ，真贋や情報の適切性，社会的な反響を考慮せずに脊髄反射的にその情報を拡散（再配信）してしまうと考えられる．本来であれば，情報の真贋や適切性を確認した上で情報を拡散することが望ましいが，そのような「世の中に正しい／適切な情報を伝えよう」という欲求は，マズローの欲求段階で言えば，より高次の「承認（尊重）の欲求」や「自己実現の欲求」に属するものと考えられる．マズローのによれば，低次の欲求が充足されていない限り，それよりも上位の欲求を満たすことは難しいとされている．結果として，一定規模以上の個人の「社会欲求」を刺激するようなコンテンツは大規模に拡散され，フェイクニュースや炎上，エコー・チェンバー現象を引き起こす，と考えられる．つまり，これらのネガティブな現象は，あくまで個人が脊髄反射的な情報拡散が大規模に行われた結果．多様な形で顕在化したものであり，それらの現象に共通する根源的な要因としてソーシャルポルノが存在すると考えられる．

このように人間の社会欲求を充足することを目的として作られたコンテンツーソーシャルポルノという概念を仮定することで，近年ソーシャルメディア上で引き起こされている社会現象を説明することが可能となる．しかし，もちろんこの概念は仮説に過ぎない．そこで本論文では，実際にソーシャルメディアデータを分析し,「ソーシャルポルノ」が実際に存在することを検証していきたい．

まずは「ソーシャルポルノ」を下記のように，より具体的に定義する．

> ソーシャルポルノとは，特定のコミュニティに属するユーザが「脊髄反射的に拡散・共有してしまいたくなる情報」である

このように定義するのは下記のような理由である．ここで

連絡先: 榊 剛史，株式会社ホットリンク, 東京都千代田区富士見１－３－１１ デュープレックスビズ５Ｆ

*1 https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016

*2 https://en.wikipedia.org/wiki/Post-truth_politics

のコミュニティとは，共通の趣味や感心・社会的な意見を持つ一定規模の集団である．ソーシャルポルノが問題を引き起こす場合は，一定規模以上のユーザが反応したばあいに限られるので，本稿で扱うソーシャルポルノは，一定規模のコミュニティの構成員に共通して訴求するものとした．脊髄反射的にとしたのは，「内容を精査しない」ことを前提とするためである．冷静に吟味してから，真贋や適切性を確かめてから拡散・共有するような情報は「ポルノ」とはいえない．本稿ではあくまで，「目にした瞬間に拡散したくなるような情報」を「ポルノ」と見なす．また，ソーシャルポルノの内容自体もコミュニティや個人によって異なると考えられる．通常の性的な意味での「ポルノ」と同様に人々の欲求に立脚するものであるため，ソーシャルポルノにも多様な趣味・志向があると考えられる．

本研究において，我々はまずはソーシャルポルノを観測し，定量的に評価することを最終的な目標としている．本稿では，その準備段階として，ランダムに抽出された投稿と特定のコミュニティに属するユーザによる投稿について，ユーザの反応に違いがあるかどうかの検証を行う．ここでは特にソーシャルポルノの「脊髄反射的に拡散・共有し」という定義から，ユーザの投稿への反応時間，すなわち「ある情報を受け取ってから，その投稿を拡散するまでに要した時間」に着目する．つまり，特定のコミュニティに属するユーザによる拡散が，ランダムに抽出された拡散よりも短い反応時間で行われているかどうかを分析する．

本研究では，情報取得に容易性から Twitter のデータを対象とする．

2. 着想

我々は，人工知能学会表紙問題の炎上に端を発し，2014 年から今まで Web 上，特にソーシャルメディアを対象とした炎上現象の分析手法の研究・開発ならびに，開発した手法による事例分析を数多く行ってきた [鳥海 14b, 鳥海 14a, 鳥海 17]．扱ってきた分析事例は多岐にわたっている．例えば，社会現象（アイスバケツチャレンジ）や社会問題（マクドナルド異物購入問題，オリンピックエンブレム盗作問題），科学に関するトピック（STAP 細胞問題，Facebook による不適切な心理学実験，人工知能学会表紙問題），自然災害（御嶽山噴火，鬼怒川洪水），などである．これらの事例分析においては，トピックごとのエントロピを算出し，「そのトピックが幅広いユーザに語られているのか？あるいは特定の偏ったユーザによって語られているのか」によって評価を行ってきた．具体的には下記のような手順である [鳥海 17]．

1. ソーシャルメディア上のネットワーク関係を元にユーザ・投稿をそれぞれ Modularity Optimization の手法によりクラスタリングし [Blondel 08]，ユーザコミュニティとトピッククラスタを構築する
2. 各トピック i についてユーザコミュニティの分布からトピックエントロピ H_i を算出する

$$H_i = -\sum_{c \in C} P(c) \log P(c) \quad (1)$$

(C はコミュニティ集合，c はコミュニティ，$P(c)$ はあるツイートがコミュニティ c のメンバーによっておこなされた確率)

3. 各トピッククラスタのトピックエントロピが高いトピックを「多様なユーザコミュニティによって幅広く語られていたトピック」，低いトピックを「特定のユーザコミュニティによって偏って語られていたトピック」とみなし，それぞれの内容を俯瞰する

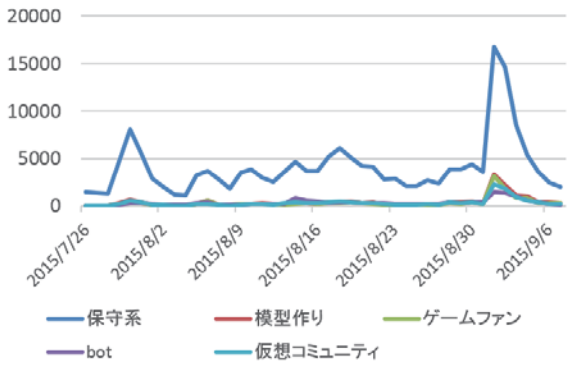


図 1: オリンピックエンブレム問題 コミュニティ別投稿数推移 (x 軸:日付，y 軸:投稿数)

このような事例分析を行う中で，複数の炎上事例に共通して出現するトピックの傾向に着目した．低いエントロピを有するトピック－つまり，特定のコミュニティにおいて語られているトピックについては，共通の傾向があることが推測された．例えば，特定の政治思想を持ついくつかコミュニティは，表 1 のように複数の炎上事例において，その事例に言及しつつ，事例とは直接関係のない政治的な批判意見や陰謀論を拡散しているのが見受けられた [*3]．特に陰謀論のトピックについては日本のみならず英語圏においても観測されることがわかった．また，トピックに関連する大半の投稿が一つのコミュニティによって投稿される場合もある．図 1 はオリンピックエンブレム問題における投稿数推移をコミュニティ別に表現したものである．図 1 より，特定の政治思想に関連する特徴語を持つグループによる投稿が大半を占めている．

この観測事実から，我々は特定のコミュニティにおいて，好んで拡散されるトピックに類似性があるのではないか，という仮説を持つに至った．前述の事例では，わかりやすい事例として特定の政治思想に興味を持つコミュニティを取り上げたが，それ以外に多様なコミュニティにおいても，そのように好んで拡散されるトピックがあるのではないかと考えた．それらのように，特定のコミュニティにおいて好んで拡散されるトピックを「ソーシャルポルノ」の一種として捉え，それらを観測することを試みる．

3. 方法論

Social Porn を検出するための方法論について述べる．まずは，あるユーザがある情報を受け取ってから情報を再発信 (拡散) するまでに要した時間:**ユーザ反応時間**を定義する．

次に，定義したユーザ反応時間に基づいて，各投稿のユーザ反応時間の分布を作成し，その統計量を算出する．

3.1 ユーザ反応時間の定義

ソーシャルポルノとは，ユーザが「脊髄反射的に拡散・共有してしまいたくなる投稿」であると定義した．ここでの「脊髄反射的」とは，内容を精査しないことを意味する．つまり，

*3 倫理的な配慮から，検索できないように文意が変わらないように文言に修正を加えている

表 1: 複数炎上事例に出現する類似トピック

事例	タイプ	Teewts
STAP 細胞	無関連な批判	STAP 細胞の件で日本の学術界の信頼がとか言い出す人は理解できない。 原発事故で失われた信頼がいつ回復したというのか
アイスバケツ	無関連な批判	【韓国経済崩壊】韓国企業の売名行為に世界中で非難殺到!慈善活動「アイスバケツ #韓国 #アイスバケツチャレンジ #韓国経済崩壊 #嫌韓
STAP 細胞	陰謀論	「STAP 細胞問題」の裏で集団的自衛権の行使容認に向けた準備が進んでいる。 安全保障調査会の会長によれば、「安保法制懇」は近々結論を出すらしい。
Facebook 実験	陰謀論	As I predicted, Facebookś psychology experiment was directly linked to the Pentagon's plan for civil unrest.

ソーシャルポルノを抽出するためには「内容を精査していない」を可能な限り確実に観測できることが望ましい．

そこで，我々は，対象とする情報をリツイートのみに絞り，ユーザが拡散に要した時間が極めて短い情報に着目することとした．なぜなら，ユーザが情報を受け取ってからそれを拡散するまでに要した時間が数十秒程度であれば，ユーザがその情報を精査してないことは明らかである．本稿では，この，**ユーザ u_j がある情報 i_k を受け取ってからそれを拡散するまでに要した時間をユーザ反応時間** $T_R(u_j, i_k)$ と定義する．

ユーザ反応時間 $T_R(u)$ を測定するためには，ユーザがその情報を受け取った時刻（以下、情報受信時刻）を計測する必要がある．ユーザが Twitter 内で情報を受け取るリツイートであれば，情報受信時刻をある程度推測することができる．一方，リツイート以外では，ユーザが他のメディアから情報を受け取っている可能性も少なくなく，そのような場合には情報受信時刻を推測することは困難である．よって，本稿では，リツイートのみを分析対象とする．ここで，ユーザ u_j が情報 i_k を受信した時刻 (情報受信時刻) を $t_r(u_j, i_k)$, ユーザ u_j が情報 i_k を送信（拡散）した時刻 (情報送信時刻) を $t_s(u_j, i_k)$ と表す．

$$
\begin{aligned}
t_r(u_j, i_k) &: \text{情報受信時刻} \\
t_s(u_j, i_k) &: \text{情報送信時刻} \\
T_R(u_j, i_k) &= t_r(u_j, i_k) - t_s(u_j, i_k)
\end{aligned}
$$

あるツイート i_k について $T_R(u_j, i_k)$ を算出することを考える．$t_s(u_j, i_k)$ は，ツイート i_k がユーザ u_j によって，リツイートされた時刻であるため，データから明示的に取得可能である．一方，$t_r(u_j, i_k)$ は，データから明示的に取得するのは困難である．なぜなら，1. Twitter の仕様上あるツイートがユーザの TL(タイムライン）上にいつ出現したかを知ることが難しく，2. またいつユーザが TL をチェックしたかを知ることが困難である．ためである．そこで $t_r(u_j, i_k)$ を推測する必要がある．

ここでは 2 つの推測方法を提案する．1 つ目は，ユーザ u_j がフォローしているユーザのうち，一番最初に情報 i_k をツイートしたユーザの i_k の送信時刻を $t_r(u_j, i_k)$ として用いる手法である．これをアプローチ 1 と呼ぶ．

2 つ目は，ユーザ u_j が情報 i_k を送信する前に送信したツイートのうち，最も新しいツイートの送信時刻を $t_r(u_j, i_k)$ をとする手法である．こちらをアプローチ 2 と呼ぶ．

今回は，データ取得の制約上，アプローチ 2 を用いることとする．

3.2 ツイートごとの反応時間分布

ソーシャルポルノの定義は,「脊髄反射的に拡散・共有して」しまう投稿のことである．もしもこの仮説が正しかった場合，ソーシャルポルノにあたる投稿と，それ以外の投稿を比較した場合に，それらを拡散したユーザ群の反応時間の分布が異なることが想定される．より具体的には，ソーシャルポルノにあたる投稿の方が，ユーザ群の反応時間が短いことが想定される．そこで本稿の分析においては，分析対象としたツイート群について，それらのツイートを RT したユーザとその反応時間の分布を作成し，その統計的な指標（平均値，中央値，最頻値，標準偏差）を比較することとした．

表 2: D_{random} のユーザ反応時間分布 統計量

ID	RT 数	最頻値	平均値	中央値	標準偏差
1-1	151	1.0E+05	6.8E+05	1.5E+05	1.1E+06
1-2	1204	10	6.5E+04	1.6E+03	3.3E+05
1-3	551	100	2.3E+05	3.2E+03	1.3E+06
1-4	55	100	4.7E+04	5.7E+03	1.1E+05
1-5	452	10	4.6E+04	4.3E+02	3.9E+05
1-6	558	10	2.8E+04	9.8E+02	1.3E+05
1-7	6681	10	6.9E+04	2.0E+03	3.3E+05
1-8	179	1	2.0E+05	6.0E+02	7.5E+05

4. 分析

4.1 データセット

今回は，比較対象として 2 つのツイート群を用意した．1 つのツイート群は，一般的なツイートに対するユーザの反応時間分布の傾向を見るためのデータセットである．今回は，著者のタイムラインから 2018 年 3 月 6 日・7 日に投稿されたツイートで，リツイート数が 100 を超えるツイートをランダムに 8 件抽出した．こちらを D_{random} とする．

もう一方のツイート群は，ソーシャルポルノらしい投稿に対するユーザの反応時間分布の傾向を見るためのデータセットである．今回は，特定の政治思想に関連した 2 ちゃんねるまとめサイトのアカウントから，2018 年 3 月 8 日 0 時において，直近に投稿された 8 件を抽出した [*4]．こちらを D_{matome} とする．

今回は，本格的な分析を行う前の試行的な分析であるため，小規模なデータセットを用いることとする．また，リツイート数やいいね数は 2018 年 3 月 8 日 15 時時点のものとする．

4.2 分析結果

2 つのデータセットにおいて，前述の通りアプローチ 2 を用いてツイート・ユーザごとのユーザ反応時間を算出する．得られたユーザ反応時間に基づいて，ツイートごとのユーザ反応時

*4 倫理的な配慮から，当該アカウントについては付せるものとする

表 3: D_{matome} のユーザ反応時間分布 統計量

ID	RT 数	最頻値	平均値	中央値	標準偏差
2-1	239	10	3.5E+04	3.5E+02	1.4E+05
2-2	352	10	5.2E+04	4.5E+02	3.4E+05
2-3	408	10	8.5E+04	8.0E+02	3.2E+05
2-4	117	10	3.4E+04	2.0E+02	1.2E+05
2-5	429	10	3.0E+04	4.5E+02	1.2E+05
2-6	248	10	3.3E+04	5.9E+02	1.4E+05
2-7	360	100	2.9E+04	1.1E+03	1.2E+05
2-8	72	10	1.2E+04	1.7E+02	5.5E+04

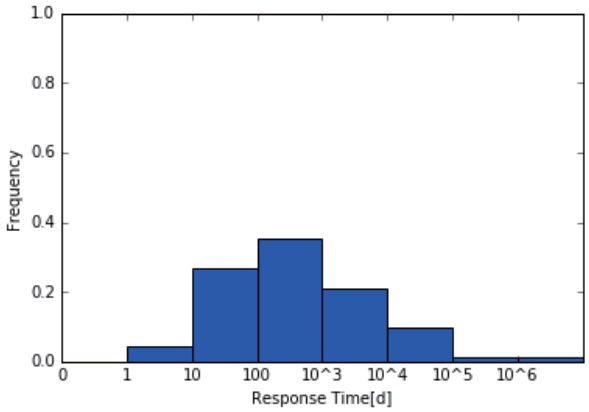


図 2: ユーザ反応時間分布 D_{random} 1-3(x 軸:反応時間，y 軸:頻度)

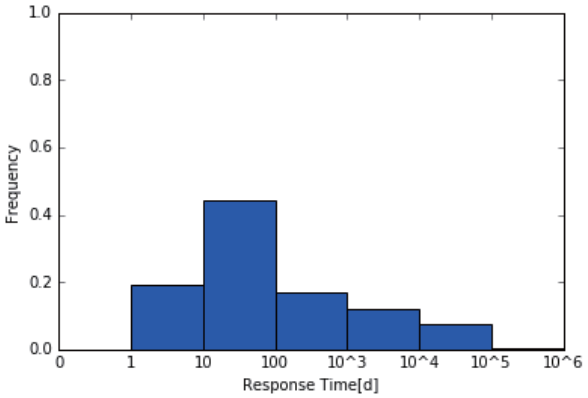


図 3: ユーザ反応時間分布 D_{matome} 2-5(x 軸:反応時間，y 軸:頻度)

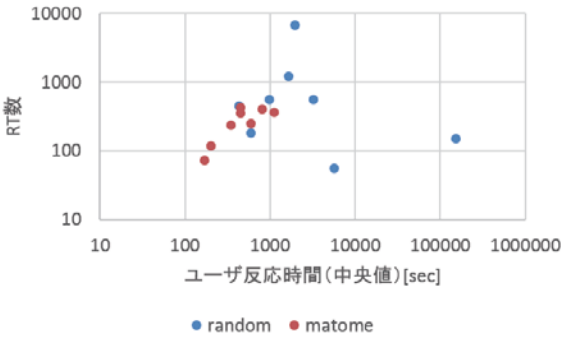


図 4: ユーザ反応時間分布 D_{matome} 2-5(x 軸:反応時間，y 軸:RT 数)

間分布の統計量を算出した．その結果を表 2,3 に示す．なお，最頻値は，10^i が $10^i - 10^{i+1}$ を表している．2 つの表において，最頻値と中央値については，D_{matome} の方が，低い傾向が見受けられる．

図 2，3 に代表的なツイート 1-3，2-5 におけるユーザ反応時間のヒストグラムを示す．2 つの図において，D_{random} に属するツイート 1-3 のがやや対照的なグラフになっているのに対し，D_{matome} に属するツイート 2-5 の反応時間の分布は，より短い時間の方向に偏っていることが見受けられる

最後に，各ツイートのユーザ反応時間とリツイート数の関係性をプロットした散布図を図 4 に示す．図 4 より，D_{random} よりも D_{matome} の方が短い時間で反応している傾向が見受けられる．

ただし，いずれの分析においても，今回用いたデータセットが極めて小規模であるため，有意な傾向が見えたとはいえない．あくまで仮説を指示する可能性がある傾向が示唆されたに過ぎない．今後は，データセットを大規模化し，今回の分析で示唆された内容を有意に示せるような結果を得たいと考えている．

5. おわりに

本稿では，ソーシャルポルノという仮説を提案し，それによりソーシャルメディア上の情報拡散における様々な現象を説明できることを述べた．分析においては，「脊髄反射的に拡散する」というソーシャルポルノの定義に基づき，ユーザが情報を受け取ってから再発信（拡散）するまでの時間:ユーザ反応時間に注目し，一般的な投稿と「ソーシャルポルノ」的な投稿にユーザの反応時間に差があるのかを分析した．結果として，仮説を指示する可能性が示唆された．今後はデータセットを大規模化し，ソーシャルポルノの存在の立証およびそれのフェイクニュースや炎上への影響について明らかにしていきたい．

参考文献

[Blondel 08] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E.: Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10 (2008)

[鳥海 14a] 鳥海不二夫, 榊剛史 : STAP 問題に対するソーシャルメディアにおける反応の分析, 第 7 回 Web とデータベースに関するフォーラム WebDBForum2014 (2014)

[鳥海 14b] 鳥海不二夫, 榊剛史, 岡崎直観 :「人工知能」の表紙に関するツイートの分析・続報, 第 4 回 Web インテリジェンスとインタラクション研究会 ARG (2014)

[鳥海 17] 鳥海不二夫, 榊剛史 : バースト現象におけるトピック分析, 情報処理学会論文誌, Vol. 58, No. 6, pp. 1287–1299 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

SNS 上の一方的な選好をもとにしたコミュニケーションの解析

Communication based on unidirectional preference in Twitter

浅谷　公威 *1　川畑 泰子 *2　鳥海　不二夫 *1　坂田　一郎 *1
Kimitaka Asatani　Yasuko Kawahata　Fujio Toriumi　Ichiro Sakata

*1東京大学 The University of Tokyo　*2群馬大学 Gunma University

Homophily and preferential attachment explain a large part of the formation of online social network(OSN). However, a unidirectional preference for specific kind of person is not regarded to be an important factor in network formation. Such preference assumes to have important roles in OSN where no geographical restriction and high searchability exist. To observe such preference in the social network, we analyzed the user network constructed by interactions between who tweets about "leaving home" and who react to them. The network has a similar structure to a bipartite network which composed of the targeted group(who tweets "leaving home") and reaction group. Furthermore, about 20% of users in the latter group take one-to-many communication with the users in formar group. We can assume a clear unidirectional preferences exist in them. In addition, we found that a large part of the tweets from users taking such one-to-many communication have the intention of invitation.

1. はじめに

オンラインソーシャルネットワーク (OSN) 上におけるコミュニケーションには時空間的な制約がなく，人物や投稿の閲覧/検索性も高い。そのような性質をもつ OSN 上では，マイノリティがコミュニティとしてつながり [3]，マスメディアが報じない重要な情報の拡散されている [4]．しかし，ソーシャルネットワーク上では従来にないリスクが表面化しつつある [8]．例えば，2017 年神奈川県で起きた自殺希望表明者の殺害事件や，未成年の誘い出し事例の一部は Twitter 上でのコミュニケーションが発端となっている．このようなコミュニケーションがどのような規模でどのような人々の間でどのように行われているかを解明することは，リスクの早期発見、教育による周知徹底、リスクを防止するシステムの設計に必要である。

OSN における会話のネットワークや友人ネットワークの構造は，Preferential Attachement(一部の人への人気の集中) と Homophily（同類選好）によりその多くが説明されてきた．Preferential Attachment とはネットワーク上の次数に比例した確率で該当するノードにリンクが貼られる性質である [1]．この性質は，Justine Bieber などの一部の人物が膨大な数のフォロワーを有していることを説明している．Homophily は似た人同士は好きになるという有向性のない概念であり，SNS 上の趣味・性的嗜好・職業などが同一である人々の結びつきを説明している．同類選好の強さは Assortative mixing[7] として指標化されており，多くの事例でその存在が確認されている．

しかしながら，これらのメカニズムには特定の属性を持つ人間への別の属性を持つ人間からの一方的な選好は想定されていない．例えば，家具を売りたい人々と買いたい人々の間や，家出希望を表明している（主に）女性達と家に泊めたい（主に）男性達との間には，特定の属性への別の属性からの一方的な選好が想定される．OSN はコミュニケーションに地理的な制約がなく，さらにハッシュタグ等で検索する機能が備わっているため，一方的な選好によるコミュニケーションが容易であると考えられる．Twitter であれば，特定のハッシュタグを持つ人を検索できるし，そのような人を起点として友人関係をたどることも可能である．このように，オンラインコミュニティは一方的な選好によるコミュニケーションが行われやすい場所であると想定される．

近年、このような一方的な選考により起きたと考えられる OSN 上の事件が表面化しつつある。このような事例を理解するため、未成年の誘い出しの事例の分析や [2]、未成年がさらされているリスクのヒアリングによる調査 [5, 6] が行われている．これらの研究により、ミクロな視点での個別の事例および、それら個別のケースの集合により個人がリスクにさらされる頻度が分かりつつある、しかし，誘い出しなどのリスクが内在する OSN におけるコミュニケーションの全体像（どのような人同士でどのようにどのように行われているか、どれくらいの数の潜在的な加害者や被害者がいるか）は明らかではない。リスクの早期発見、教育による周知徹底、リスクを防止するシステムの設計のためにも、そのコミュニケーションの全体像を解明することは重要である．

本論文では，Twitter 上の一方的な選好によるおこるコミュニケーションのネットワーク構造を分析する．さらに，そのネットワーク構造から典型的な行動を行うユーザーを抽出する提案し，その方法により想定される典型的なコミュニケーションが抽出できることを示す．具体的には，家出の希望を表明する人々と何らかのリアクションを行う人々の間のインタラクションのネットワークを観察し，その中に一方的な選好があると想定されることを明らかにする．さらに，家出希望表明者の誘い出しにつながるような行動を取りやすいユーザーを抽出するため，典型的な行動をとるユーザーを会話ネットワークから抽出し誘い出しに繋がる行動をとっているかを分析する．

2. 手法

2.1 コミュニケーション構造の解析

はじめに，ハッシュタグに紐づくツィートの集合へのリアクションより，ユーザー間ネットワークを構成する．具体的には，特定のハッシュタグを用いて検索したツィートをシードツィートとし，そのツィートへのリアクション (リプライ, リツィート, ライク) を収集する．そのリアクションの元となる人から先となる人への関係性を有効リンクとしたネットワークを構成

連絡先: 浅谷公威，東京大学工学系研究科，asatani@gmail.com

する．

そのように構成されたネットワークはどのような構造になるだろうか．はじめに，ハッシュタグが#理科○類△組のようにコミュニティを代表する場合について考える．シードツィートを行ったユーザー自身もシードツィートへのリアクションを行うことがあり，入次数も出次数も 1 以上のノードがある程度の割合で存在すると考えられる．次に，一般的なハッシュタグである「#○△を食べたい」というタグに対するリアクションを観察した場合，入次数のみが 1 以上と，出次数のみが 1 以上の 2 つのグループからなる 2 部グラフになると想定される．その 2 部グラフにおいて,「#○△を食べたい」という人を探し出してリアクションを行う人は少ないと考えられるので，出次数が 2 以上のユーザーは想定しにくい．上記とは異なり,「#家具を売ります」もしくは「#家出」などのツィートが一方的な選好を惹きつけた場合，あるユーザーからの不特定多数へのアプローチが想定される．このような場合，ユーザー間ネットワークは 2 部グラフになるとともに，出次数が 2 以上のユーザーが少なくない割合で存在していると考えられる．

以上の考察より、以下の構造のユーザー間ネットワークは一方的な選好を想定することでその形成過程を理解することが可能である。

- 内部のリンクが疎な 2 つの集合により大部分が構成され、片方の集合から他方の集合へ一方的に有向リンクが張られる
- 前者の集合のなかには、出次数が 2 以上のユーザー（後者の複数のユーザーにリンクを貼る）が少なくない割合で存在する

ハッシュタグをもとにしたリアクションより構成されたユーザー間ネットワークが以上の構造を持つかを測るため 2 つの指標を定義した．1 つめの指標はリンクの非対称性である．ユーザーを，ターゲットのみのユーザー集合 $T = \{u \mid k_u^{in} > 0, k_u^{out} = 0\}$，リアクションのみをユーザーの集合 $R = \{u \mid k_u^{in} = 0, k_u^{out} > 0\}$，その他のユーザー O に分類する．k^{in} は入次数，k^{out} は出次数を表す．リンクの非対称性指数 $A = \frac{N(R)+N(T)}{N(R)+N(T)+N(O)}$ となる．もう一つの指標として，リアクションのみのユーザー R からのエッジの平均本数を $ka_R = N(E_{ij}\{i \in R\})/N(R)$ と定義した．検出されたネットワークのリンクの非対称性指数 A が大きく，R に属するユーザーからのエッジの平均本数 ka_R が大きい場合に，一方的な選好により構成されたネットワークであると想定可能である．

2.2 典型的な行動を行うユーザーの抽出

検出したユーザーのネットワークでリアクションを行ったユーザーの集合 R の中から，典型的な行動を行うユーザーを以下のルールベースのアルゴリズムによる抽出する．

まずはじめにアクティビティの高いユーザー群を抽出する．具体的には，複数のターゲットにリアクションしている，つまり出次数が k 以上のユーザーを抽出する．

さらに，そのユーザー群の中で典型的な行動を行っているユーザー群の抽出する．そのような人は，リアクションを行う先が他の人のリアクションを行う先とよく似ている．まず初めに，ユーザーのペア同士でリンク先ノードの Jaccart 係数を計算する．次に，その Jaccart 係数を重みとしたユーザー間のネットワークを構成し，そのグラフ内でのページランクの上位のユーザーを典型的な行動を行うユーザーとして抽出する．このようにして抽出されたユーザーは，周囲の人たちの平均的な行動特性を有している代表的なユーザーと考えられる．

3. データ

Twitter より 2017 年 8 月 1 日から 2018 年 2 月 20 日までの,「#家出#神待ち #泊めて #泊めてくれる人」それぞれのハッシュタグに関連するツィートと，それへの反応 (リプライ, リツィート, ライク) を全て収集した．さらに，そのユーザーの一部に対しフォロー関係を取得した．

ハッシュタグ検索で取得したデータは，2,614 人からの 24,773 件の自殺関連のツィートとなった．1,341 人の 9,250 件のツィートに対し，5,307 人からのライクは 11,245 件，リプライは 5,306 件，リツィートは 894 件のレスポンスを取得した．

「#家出#神待ち #泊めて #泊めてくれる人」のツィートを行っているアカウントは女性と推定されるものが多く，それらに対してレスポンスを行うアカウントは男性と考えられるものが多かった．ネット上で人格を変える人や，複数のアカウントを使う人も想定できるが，本研究ではそのような人は少ないと想定し補正することなく分析する．

4. 分析

4.1 コミュニケーション構造の解析

入出次数の分布

抽出したネットワークの出次数を横軸と入次数を縦軸にヒストグラムを作成した (図 1 左)．ヒストグラムの色はログスケールで表示している．出次数・入次数がともに 1 以上のユーザーの数は少ないことが分かる．このようなユーザー数は 428 人で，全体で 8,121 人のユーザーがいることを考えると，定義したリンクの非対称性指数 A は 0.947 となり，ネットワークは 2 部グラフに近い構造である．

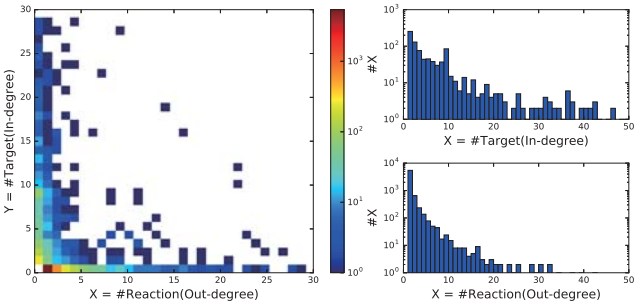


図 1: 左: ネットワークの出次数 (横軸) と入次数 (縦軸) のヒストグラム．右上: 出次数が 0 のユーザーの集合 R における入次数の分布．右上: 入次数が 0 のユーザーの集合 T における出次数の分布．

図 1 の右上に出次数 0 であるターゲットユーザー群 T に属する 913 人の入次数の分布を示している．このグラフから分かるように家出に関するツィートをしているユーザーでリアクションを受けたユーザーの多く (659 人) は，複数人からのアプローチを受けている場合が多いことが分かる．

図 1 の右下に出次数が 1 以上で入次数 0 のリアクションのみのユーザー群 R の 6,780 人のユーザーの出次数の分布をプロットである．入次数ほど顕著ではないが 2 人以上のユーザーにアプローチをしている人が少なくない割合で存在している．このように，複数の#家出ツィートをする人にアプローチをする人は 1,324 人であり，R の中の 19%を占めている．また，リアクションのみをユーザーの集合 R の出次数の平均値 ka_R は 1.62 本となった．

フォロー関係の密度

ターゲットのみユーザー群 T，リアクションのみユーザー群 R の間のフォロー関係を図 3 にまとめた．データ取得の関係上，次数 5 以上のユーザーのみを対象とし，7.6%の関係性のみをサンプリングした．

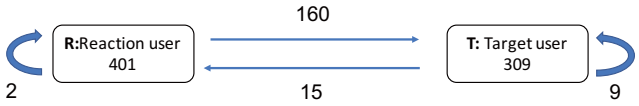


図 2: リアクションのみユーザー群 R とターゲットのみユーザー群 T との間のフォロー関係の数．次数 5 以上のユーザーのみに絞った．リンク数は 7.6%サンプリング．

図 2 から分かるように，リアクションのみのユーザー群 R およびターゲットのみユーザー群 T の間にはほとんどフォロー関係が存在しない．リアクションのみのユーザー群 R の人々はお互いに知り合いではないが，同じ集合に対してリアクションを行っていることが分かる．同様に，ターゲットのみユーザー群 T は R よりも明確ではないものの関してもお互いに知り合いではないと想定される．また，R と T の間にも，7.6%のサンプリングで 160 の関係性（全体では約 2,100 本と推定）のエッジが存在しおり，これは R に属するユーザーが T に属するユーザーを一方的にフォローしたと考えられる．

コミュニケーション構造の解析より，家出に関するコミュニケーションの大部分が，リアクションを受ける側のユーザーの集合 T とリアクションをする R の 2 部グラフから構成され，R のユーザーの中には複数人の T に属するユーザーにリアクションを行うユーザーが千人以上存在している．また、リアクションを受ける集合 T の中のフォロー関係も大きく疎であることから，その千人以上のユーザーがリアクションを送る先のユーザー同士が友人でコミュニティ内にいるとも想定し難い。このことより抽出したユーザーネットワークは手法で定義した背後に一方的な選好が想定されており，特に複数人の家出希望表明者へリアクションする千人以上のユーザーの行動原理には一方的な選好が想定される．

中心的なユーザー間の会話構造の可視化

コミュニケーションの構造を可視化するために，全てのノードの次数 k が 6 以上となる k-core[9] ネットワークを図 3 に可視化した．このグラフにおけるエッジの大きさは出自数（リアクションを取った回数）を示しており，色が明るいほど入次数が高い．また，リンク先のノードの近辺のリンクが太く表示されている．この図は，円の外側の出次数が高い（リアクションをする）ユーザーから内側の入次数が高い（家出をツィートする）ユーザーへの一対多の一方的なコミュニケーションがあることを示している．

4.2 危険度の高いと思われるツィートの特定

検出したデータの中に，誘い出しにつながりうるリプライ（誘い出しリプライ）が存在する割合を調べた．誘い出しリプライとは，Twitter のパブリックな場所以外でのコミュニケーションや現実世界での接触を促すツィートである．例えば,「家出したいです」というツィートに対し,「DM ください」（※DM=ダイレクトメッセージ）や「○○ (地名) なら大丈夫です」というツィートは誘い出しリプライとなる．それぞれ検出したサンプルから 150 を選び，各々のツィートがどの集合に属しているかをマスキングした上で，人間が読んで誘い出しリプライかどうかを判断した．

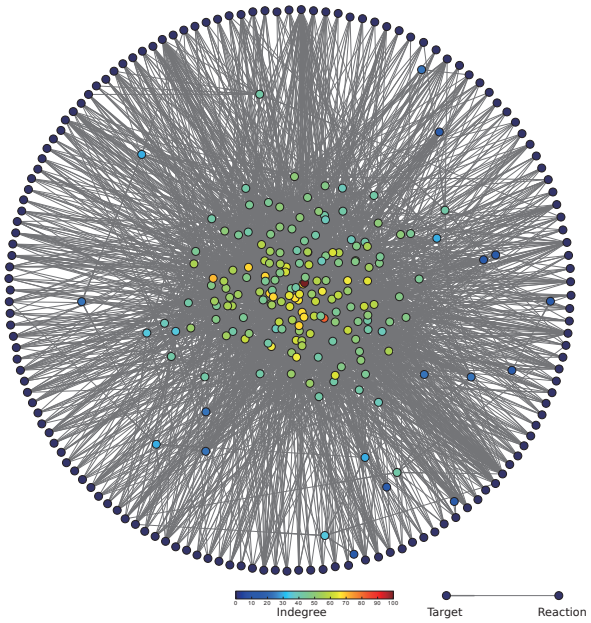


図 3: 家出に関するユーザー間ネットワーク。$k = 6$ の k-core 部分のみを描画．ターゲットとなるノード側のエッジは太く、ノードの入次数が高いほど明るく描画．この図では中心部に入次数が高い（自らの家出に関するツィートに複数のリアクションを受けている）ユーザー，周辺部に入次数が小さい（リアクションをする）ユーザーが配置されている．また，周辺部のユーザーは複数の中心部のユーザーにリアクションを行っている．

表 1: リプライの検出数と誘い出しリプライの割合

	誘い出し	
	リプライの割合	検出数
全リプライ	0.58	5,306
高頻度ユーザー	0.69	1,996
典型的な行動のユーザー	0.73	668

表 1 は各ツィートの集合に対して誘い出しリプライの割合である．全リプライの対して 58%が誘い出しに関するものであり，単純に#家出などに対するリプライは高い蓋然性で誘い出しの危険を孕んでいることがわかる．また，高頻度ユーザーや本論文で定義した典型的な行動を行うユーザーからのリプライは誘い出しの危険性が高いことが分かった．また，ランダムに数十のユーザーをサンプリングしたところ、性別の判定は主観であるものの，家出したいとツィートしているのは未成年を含む若い女性でありリプライしているのは主に男性と想定される．

5. 考察

本論文ではソーシャルネットワーク上に一方的な選好による会話構造が存在していることをデータより裏付けた．そこでは，それぞれの間ではリンクが疎な 2 つの集合があり，1 つの集合から他方の集合への一対多のリンクが多く存在している．ある人から#家出をツィートした複数の人への一対多のリンクは，一方的な選好が背後にあると想定される．また，それぞれの集合のユーザー数はそれぞれ千以上であった。したがって、一方的な選好によるコミュニケーションは、ソーシャルネット

ワーク上のコミュニケーションを議論する上で無視できない規模で存在しているといえる。

また，単純に「家出」関連のツィートへのリプライを収集するだけで半分以上のツィートが誘い出しが想定された．また，家出をしたい人と，誘い出したい人がそれぞれ数千人の単位で存在していることは，Twitter 上での誘い出しリスクの大きさを物語っている．誘い出しの中では，直接的に特定の場所に呼び出すものと近い頻度で，DM（ダイレクトメッセージ）でのコミュニケーションに誘導するものが多かった．DM はユーザーの権限で OFF にしたり，フォローワー以外からの DM のみ禁止といった設定が可能である．誘い出されるユーザーが任意のユーザーからの DM を許可した場合は，パブリックに閲覧できるリプライではなく DM でやりとりが行われていると考えられる．この DM に関しては本分析のスコープ外であるが，そこでどのようなやり取りが行われているかは分析を行う必要がある．本研究では Twitter の表層部に見える危険なやり取りを検知したが，Twitter 上にはさらに多くの危険が潜んでいると考えられる．

本研究では具体的に家出関連のハッシュタグをシードとして，一方的な選好によるコミュニケーションを検出した．家出以外にも自殺に関するツィートを中心としたネットワークは同様な構造になっていると想定され，このような一方的な選好をネットワーク構造から発見することは新たなリスクの予知につながると考えられる．したがって，将来的にはネットワーク構造のみから一方的な選好によるコミュニケーションが存在する部分集合を発見する技術の開発が必要である．

6. 結論

Twitter における家出に関するツィートとそれに対するリアクションを抽出し，ユーザー間のネットワークを分析することで，OSN 上に数千人単位の一方的な選好によるコミュニケーションが存在すること確認した。そこでは，お互いに関係が疎な 2 つのグループ間（誘い出す側と，家出を表明する側）で一方向のコミュニケーションが存在する．さらに、前者のグループのユーザーの 2 割程が後者への一対多のコミュニケーションをとっており，明確な一方的な選好が存在することが想定される．また，そのような一対多のコミュニケーションをとるユーザーからのツィートは誘い出しの意図が想定される割合が非常に高いことが分かった．

7. 謝辞

本研究は RISTEX「未成年者のネットリスクを軽減する社会システムの構築」プロジェクトの支援を受けて実施した．

参考文献

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] et al. DeHart. Internet sexual solicitation of children: a proposed typology of offenders based on their chats, e-mails, and social network posts. *Journal of sexual aggression*, 23(1):77–89, 2017.

[3] Keith N Hampton. Internet use and the concentration of disadvantage: Glocalization and the urban underclass. *american behavioral scientist*, 53(8):1111–1132, 2010.

[4] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[5] Sonia Livingstone, Leslie Haddon, Anke Görzig, and Kjartan Ólafsson. Risks and safety on the internet: the perspective of european children: full findings and policy implications from the eu kids online survey of 9-16 year olds and their parents in 25 countries. 2011.

[6] Giovanna Mascheroni and Kjartan Ólafsson. Net children go mobile: Risks and opportunities. 2014.

[7] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.

[8] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, et al. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, 2011.

[9] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ソーシャルメディアストリームからの新固有表現の発見

Finding Emerging Named Entities from Social Media Streams

赤崎　智 *1　　吉永 直樹 *2　　豊田　正史 *2
Satoshi Akasaki　Naoki Yoshinaga　Masashi Toyoda

*1東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

*2東京大学 生産技術研究所
Institute of Industrial Science, The University of Tokyo

New things and concepts are born everyday in the real world. A database covering such information becomes indispensable in various tasks such as document understanding, information retrieval, entity monitoring. However, it is no longer impossible to manage the database manually, so it is desirable to collect these entities automatically and quickly. We aim to discover new named entities pointing to new things and concepts from social media streams using distant supervision. In the experiment, we use Twitter as a social media stream, and apply the proposed method to the weekly tweets. We confirmed that the performance of proposed method outperformed the baseline in terms of precision@k.

1. はじめに

我々の住む現実世界では製品や映画，音楽などの作品，イベント，人物などといった新しい事物が日々次々と誕生している．例えば，事物に関する情報を蓄積する有名な知識ベースである Wikipedia*1 では，2015 年 4 月から 2015 年 5 月の間に 41,579 件の新しい記事が登録されている [1]．Wikipedia に登録されるもの以外でも，スポーツ選手や店舗名，飲食物などは把握できないほど事物の数が年々増加している．これらを集積する知識ベースや辞書など（以降，まとめて辞書と呼ぶ）の事物のデータベースは事物の情報を利用するさまざまなタスクで必要不可欠となっている．例えば，昨今のウェブに存在する膨大な量のテキストを理解するためには，情報を事物単位にまとめ人間にわかりやすい形で整理して提示することが有効である．また，製品の不具合情報や，災害などのイベント等の特定の事物の動向をソーシャルメディアストリームなどで監視するソーシャルモニタリングも政府や自治体，企業にとって大変重要である [2]．

ここで，実世界に真に新しく出現する新事物は，その事物に関する情報の新規性から特に重要であるが，辞書に未登録であるためにその情報を活用することができず問題となる．また，冒頭で述べたような把握できないほど事物が多い種や，そもそも出現数が低頻度であるロングテールなものについては現状無視せざるを得ない状況となっている．

そのため，新しい事物が出現次第辞書を更新し，かつ事物をできるだけ網羅し最新の状態に保つことが必要となる．一般に辞書の整備は人手により行われるが，日々増え続ける新しい事物を人手で収集し辞書に追加することはもはや不可能であるため，計算機を用いて自動化することが望ましい．

そこで本研究では，事物のデータベース構築の支援のため，リアルタイムに次々と新しい情報が発信されるソーシャルメディアストリームである Twitter を対象とし，そこから新しい事物や概念を指す新固有表現を精度良く発見することを目指す．提案手法は新固有表現が最初に出現する時にその新規性を示唆するような文脈で出現することに着目し，そのような文脈を含むツイートを Wikipedia を用いた Distant Supervision で自動的かつ大量に獲得し，それらのツイートから新固有表現を発見する系列ラベリングモデルを学習する．

実験では，一週間分のツイートに対して，辞書に未登録のものを新固有表現と判定するベースラインと提案手法を適用し，後者がより精度良く新固有表現を発見できることを示す．

連絡先: {akasaki,ynaga,toyoda}@tkl.iis.u-tokyo.ac.jp

*1　https://en.wikipedia.org/wiki/

2. 関連研究

マイクロブログ等のソーシャルメディアストリームを対象とした固有表現抽出に関する研究がこれまで盛んに行われている．ソーシャルメディアストリームを対象に固有表現抽出を行う際の問題として，投稿が一般の新聞やニュース記事などの媒体と比べて文量が短くノイジーになりがちであるため，予測のための適切かつ十分な手がかりが得られないことが挙げられる．Ritter ら [3] はマイクロブログの文の短さや単語の表記ゆれ，不正確な Capitalization が品詞タグ付けやチャンキング，固有表現抽出の精度に影響を与えていることを指摘し，マイクロブログ投稿へのラベル付けによる学習データの拡張，ブラウンクラスタリングによる表記ゆれ語に関する素性の作成などを行い，各タスクの精度を改善した．また，固有表現に人名や場所といったカテゴリを割り当てる固有表現分類のタスクも行っており，そこでは抽出した固有表現が出現する全投稿からトピックを推定することで適切なカテゴリの割当を行っている．Ritter らの研究は一般の固有表現を対象としているが，本研究では新固有表現を対象としている点が異なる．

現実世界の事物（エンティティ）を指す固有表現を認識し，知識ベース中の項目と対応付けるエンティティリンキングというタスクにおいて，未知エンティティと呼ばれる対応付け先が知識ベースに存在しないようなエンティティが精度に影響を与えることが報告されている．Wu[4] はこの問題に対し，未知エンティティの判定を単一のタスクとして解くことを提案した．具体的には，固有表現に対し対応付けを行う前に，それが未知エンティティであるかどうかを文脈，分散表現，トピック，検索エンジンでの注目度，統語的要素の 5 つの要素から総合的に

判定するモデルを提案した．実際に彼らの手法をエンティティリンキングの前処理として実行したところ，対応付けの精度が有意に向上した．Wu らの手法は予め対応付けの対象となる固有表現が与えられており，固有表現部分の同定を問題としていない点が本研究と異なる．

3. 新固有表現

新固有表現を発見するにあたって，本研究における新固有表現が何であるかを定義する．

Graus ら [5] は，ニュースストリームとソーシャルメディアストリームにおいて新しく出現する事物が Wikipedia に記事として登録されるまでにどのような振る舞いをするか分析し，それらの事物のほとんどが「ニュースやソーシャルメディアストリームなどのメディアで散発的に言及されている状態」から「参考文献などの二次情報の充実により一記事として確立された状態」へと推移することを報告している．

これはあくまで Wikipedia に登録された事物を対象とした報告であるが，実際にはこのような状態の推移は Wikipedia に新規に登録されるような事物のみに限らず，実世界のあらゆる事物において確認できる普遍的なものである．Wikipedia は記事に内容の特筆性 [*2] を要求するため，いわゆる後者の状態へと推移した時点で記事として登録される場合がほとんどであるが，ソーシャルモニタリングなどの応用を考慮すると新しい事物は前者の状態のうちに早期発見することが重要である．また，ユーザはソーシャルメディアストリームにおいて新しい事物についての投稿を発信するとき，以下の例 [*3] のようにその事物の新規性を示唆するような表現で発信を行うことが多い．

> 『ポケモン』新作映画のタイトルが『劇場版ポケットモンスター みんなの物語』に決定！　予告編映像も公開
> https://www.famitsu.com/news/201802/27152683.html

そのため，このような表現が現れる投稿を収集することができれば，そこに現れる固有名詞，すなわち新固有表現を発見することが可能となる．これらを踏まえ，本研究では発見を目指す新固有表現の定義を「公共性を有するメディアにおいて，新規出現に関する情報が提供・共有されるような固有名詞」とする．

4. 提案手法

3 節で述べた新固有表現をソーシャルメディアストリームから発見するための提案手法について述べる．提案手法は新固有表現が出現する投稿を Distant Supervision により獲得し，それらを用いて新固有表現にラベル付けを行う系列ラベリングモデルを構築する．構築したモデルをソーシャルメディアストリームである Twitter の投稿（ツイート）へと適用し，実際に新固有表現の発見を試みる．

4.1 Wikipedia を利用した Distant Supervision

教師あり学習は大量の学習データを要求するが，データへのラベル付与は基本的に人手で行う必要があるためコストが大きい．また，人手で付与を行ったとしても，将来的に辞書が更新されたり発見対象の性質や種類が変化した場合には学習データの再構築をする必要があり現実的でないため，学習データを自動生成する仕組みが求められる．

Mintz ら [6] は，知識ベースなどを手がかりに擬似的な学習データを生成する Distant Supervision という手法を提案した．新固有表現のタグ付けがなされたソーシャルメディアストリームのデータセットは存在しないため，本研究でも Distant Supervision を用いて擬似的な学習データを大量に作成し，それらのデータを用いて新固有表現を発見するモデルを学習する．

3 節で述べた通り，新事物は出現の初期の段階では，その新規性を示唆するような表現で発信される．我々はこれに着目し，既知の事物の時系列における初期のツイートを正例として収集した．具体的な手順は以下の通りである．

1. Wikipedia に登録された記事のタイトルを既知の事物の固有表現として収集する
2. 得られた各固有表現を含むツイートを固有表現毎に時系列の古い順に収集する．

負例については，正例として集めた事物について，それぞれ収集した時刻の一ヶ月後のツイートを正例と同数収集する．

Distant Supervision は大量の疑似学習データを生成できる一方で，誤ったラベルのデータが混入してしまうことが多い．そのため，収集の際に以下のような制約を設ける．

- 収集する事物について，Wikipedia に記事として登録された半年前以前に 5 回以上出現しているものは候補から除く．これは，語義の曖昧さが存在する事物や，既に既知であるものとして普及してしまっている事物が正例及び負例に混入しないようにする目的がある．
- 上記手順 2 で正例のツイートとして収集する対象はリツイートされたものとする．また，1 日に最低 10 回以上リツイートされない事物については候補から除く．これは出現の初期の段階で一定頻度リツイートされるような事物は，ユーザが注目するような新規性の高いものであるという仮定を置いている．
- 上記手順 2 で事物毎に収集するツイートの時系列の範囲は，事物の出現が最も古い日から，1 日に 10 回以上リツイートされる最初の日までとした．これは，事物毎に最初にバーストするツイートとそれ以前のツイートではその事物の新規出現に関する情報が提供されやすいという仮定を置いている．

4.2 CRF を用いた新固有表現の発見

4.1 節において収集した疑似的な学習データから新固有表現を発見するための教師あり学習モデルを構築する．提案手法では新固有表現の発見を，固有表現抽出のタスクにおける標準的な枠組みである系列ラベリング問題として解く．教師あり学習のモデルとしては条件付き確率場（CRF）を採用する．

CRF はトークン列 $\boldsymbol{x}$ を入力とし，トークンのラベル列である $\boldsymbol{y}$ を出力する．具体的には以下の式のような条件付き確率で定義される．

$$Pr(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\Lambda}) = \frac{1}{Z_{\boldsymbol{x}}} \exp\left(\sum_{i=1}^{|\boldsymbol{x}|-1}\sum_{k}\lambda_k f_k(y_i, y_{i+1}, \boldsymbol{x}, i)\right)$$

$Z_{\boldsymbol{x}}$ は正規化係数である．$f_k()$ は素性関数でトークンとラベルの依存関係を表現しており，形態素解析の結果や文字種などが用いられることが多い．$\boldsymbol{\Lambda}$ はモデルのパラメータであり，学習により調整される．学習したモデルは，任意のトークン列

∗2 https://ja.wikipedia.org/wiki/Wikipedia:独立記事作成の目安
∗3 https://twitter.com/famitsu/status/968399069189087235

$\boldsymbol{x}'$ に対して最適なラベル列 $\hat{\boldsymbol{y}}' = \arg\max_{\boldsymbol{y}} Pr(\boldsymbol{y}|\boldsymbol{x}'; \hat{\boldsymbol{\Lambda}})$ を出力する．

CRF の学習データは 4.1 節で収集したツイートを分かち書きし新固有表現に対応するトークンに B，I，O，E，S の 5 種のラベルを割り当てたものを用いる．ラベルはそれぞれ，新固有表現の開始位置，新固有表現の 2 トークン目以降，新固有表現の外側，新固有表現の終了位置，単一のトークンからなる新固有表現を意味する．

素性については，現在位置のトークンとその前後 2 トークンの表層，品詞細分類，文字種，前トークンの出力ラベルを用いる．

5. 実験

提案手法をツイートに適用し，実際に新固有表現を発見することで有効性を確認した．以下ではまず詳細な実験設定について述べ，その後実験結果を定量的・定性的に確認する．

5.1 実験設定

学習データ

学習・評価に用いるソーシャルメディアストリームのデータセットを我々の研究室で 2011 年 3 月より継続して収集 *4 している Twitter のデータをもとに構築した．まず 4.1 節の手順 1 を実行し，2012 年 3 月 11 日から 2015 年 12 月 30 日までに Wikipedia に新しく登録された記事タイトルを既知の事物の固有表現として収集（このとき，リダイレクトページや曖昧さ回避の記事については除外した）し，これらの事物について 2011 年 3 月 11 日から 2012 年 3 月 11 日までの期間でツイート頻度を集計し，頻度が 5 を超えるものは除外した．これにより残ったそれぞれの事物について，手順 2 を実行し時系列の古い順にリツイートを収集し，1 日に 10 回以上リツイートされた事物 11,698 件を残し，それぞれの事物の出現が最も古い日から，1 日に 10 回以上リツイートされる最初の日までのリツイートを最終的に抽出し，結果 207,508 件を最終的な正例として採用した．負例については 4.1 節に述べた方法で 11,698 件の事物それぞれについて収集し，結果 155,254 件を負例として採用した．これら正例と負例を合わせたものに ipadic を用いた MeCab*5 で分かち書きを行ったものを学習データとした．

評価手順

2016 年 6 月 10 日から 2016 年 6 月 16 日までの 1 週間を対象に，各日付の全日本語ツイート（リツイートとリプライ，クオートは除く）に各手法を適用し，各日付毎に出力された新固有表現の候補を出力された回数が多い順に並び替え，Precision@k(k=50) を算出した．この時，短期間で同名の新固有表現はほとんど出現しないという仮定を置き，対象としている日の各出力が前日までの各日付の出力上位 50 件に入っていた場合にはそれを評価から除外した．新固有表現の発見の評価には本来 Recall も重要であるが，数千万ツイートを対象に出現した新固有表現の総数を算出することは現状不可能と判断したため，便宜的にこのような手法を取った．正解ラベルについては各日付の出力である新固有表現候補の上位 50 件に対して，第一著者が発見元のツイートを確認し新固有表現であるか否かを判断し付与した．

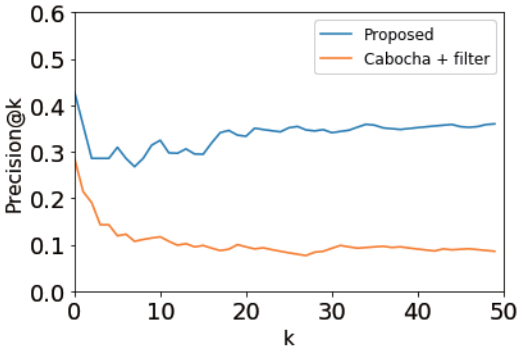


図 1: 各手法の Precision-Curve

比較手法

新聞記事のコーパスから学習された一般固有表現抽出器をツイートに対して適用し，出力された固有表現が Wikipedia に未登録であれば新固有表現とみなす手法をベースラインとして採用し提案手法との比較を行った．提案手法の CRF の実装物として CRFsuite*6 を，ベースラインの一般固有表現抽出器として Cabocha*7 を用いた．CRF のパラメータ最適化には L-BFGS 法に基づく L2 正則化を適用し，L2 正則化の係数については学習データの 1 割を開発データとして用いて調整した．

5.2 実験結果

5.1 節の評価手順における各日付の Precision@k を平均し，各 k の値を手法毎にプロットしたものを図 1 に示した．どの k においても提案手法がベースラインの数値を上回っており，特に k が大きくなるにつれ数値の差も顕著になっていることが確認できる．これは出力の数 k を増やしても提案手法の性能が比較的安定していることを示している．提案手法はツイート中の新規性を示唆するような表現を捉えたうえで新固有表現を認識しているが，ベースラインは辞書に未登録なものをすべて新固有表現として判定してしまうため，偽陽性の割合が大きく性能低下の原因となっている．

提案手法で発見できた新固有表現の例を発見元のツイートとともに表 1 に例示する．「ゼルダの伝説ブレスオブザワイルド」などの有名な事物だけでなく，「MUFG コイン」や「GeForce GTX1070」などのロングテールな事物の固有表現を適切に発見できていることが確認できる．また，「ソニーネットワークコミュニケーションズ」などの事物の名称変更や，「アクセル・ワールド」などの表層としては既知だが新しい語義などを発見できているのは興味深い．とくに新語義については，ベースラインのように既存の辞書でフィルタリングを行う手法では発見できないものであるため，この点でも提案手法は有効であることがわかる．

最後に提案手法で誤って出力した例を発見元のツイートとともに表 2 に例示する．1 つめの例の「ソードアート・オンライン」のように，実際は既知の事物の固有表現だが，ツイート中に新規性を示唆するような表現が存在したために誤って出力してしまったものが多かった．2 つ目の例の「舛添要一」も同様だが，こちらはツイート中に新規性を示唆するような表現が存在しないのにもかかわらず誤って出力してしまっている．

*4 2011 年 3 月から 2017 年 1 月までの期間で約 380 億ツイートを保有

*5 http://taku910.github.io/mecab/

*6 http://www.chokkan.org/software/crfsuite/

*7 https://taku910.github.io/cabocha/

表 1: 提案手法で正しく出力した例

出力	発見元のツイート
ゼルダの伝説ブレスオブザワイルド	最新作 『ゼルダの伝説ブレスオブザワイルド』 映像が公開! _URL_
	【速報】 ゼルダ新作の正式タイトルが『ゼルダの伝説ブレスオブザワイルド』に決定!!! _URL_
MUFG コイン	三菱東京 UFJ が独自の仮想通過『MUFG コイン』を発行! 本格的に紙幣を持ち歩かない時代に _URL_
	【仮想通貨】 三菱東京 UFJ 銀行、「MUFG コイン」を来秋発行へ _URL_ 利用者同士が手軽にやり取りしたり、割安な手数料で外貨に交換したりできる。スマホのアプリに取り込むなどして使う。
GeForce GTX1070	「GeForce GTX1070」が 6 月 10 日 (金)22 時に解禁 一部ショップでは深夜販売を実施 -アキバ総研 _HASH_ _URL_
	【超悲報】 本日 22 時解禁予定の NVIDIA 新型 GPU「GeForce GTX1070」搭載ビデオカードですが、ASUS 製は本日発売ではありません。近日中発売ですのでご注意ください。 _URL_
ソニーネットワークコミュニケーションズ	ソネット、7 月より新社名 「ソニーネットワークコミュニケーションズ株式会社」 に _URL_ _HASH_
	ソネットが社名変更ー「ソニーネットワークコミュニケーションズ」に、また長く ーインターネットコム _URL_
アクセル・ワールド	【特報】 劇場版 『アクセル・ワールド』 新作アニメの新キービジュアル & 新キャラ初公開! 来場者プレゼントで川原礫書き下ろし「AW」特別短編小説が貰えることも判明 _URL_ _HASH_
	劇場版「アクセル・ワールド」 第 2 弾キービジュアル公開 新キャラクターも明らかに _URL_

表 2: 提案手法で誤って出力した例

出力	発見元のツイート
ソードアート・オンライン	TV アニメ 『ソードアート・オンライン』 第 1 期の再放送が決定! 松岡禎丞さん・日高里奈さんら出演のニコ生特番も _URL_
	【朗報】 アニメ『ソードアート・オンライン』1 期が再放送決定キター (゜∀゜) ー!! _URL_
舛添要一	血税を納めてる庶民には、生活が苦しい人もいっぱいいるのに。リオに行きたいからって居座るな 【炎上】 舛添要一が政治資金で「高級すきやき牛肉」と「高級ワイン」を購入した事が判明/公用車でプロ野球観戦の疑惑も _URL_
	「舛添要一語録」の裏側をコンサル目線で解読する - 今週もナナメに考えた鈴木貴博ダイヤモンド _URL_

これについては学習データを分析したところ，政治家の人物がニュースなどで話題になっている状況のツイートが多数含まれていた．これは Distant Supervision で学習データを構築したのが原因であり，政治家などは Twitter 上での新規出現がスキャンダルなどなんらかのニュースになったタイミングである場合が多く，提案手法の Distant Supervision でそれらのツイートを学習データとして収集してしまった結果といえる．

6. おわりに

本研究では，ソーシャルメディアストリームから新固有表現を発見するタスクに取り組んだ．提案手法は Distant Supervision により自動的かつ大量に収集した新固有表現とその新規性を示唆するような表現が現れるツイートを，系列ラベリングモデルである CRF で学習し，実際に 1 週間分のツイートからの新固有表現の発見を行った．結果としては提案手法はベースラインの手法を大きく上回ることとなり，1 節で述べたようなロングテールな事物や表層としては既知だが新しい事物なども発見することが出来た．

今後の課題としては，Distant Supervision の工夫や CRF の素性の拡張による精度改善がまず挙げられる．また，評価手法として今回 Precision@k のみを用いたが，実世界の新しい事物をどれだけカバーできたかを示す指標である Recall を評価する手段も検討する必要がある．

7. 謝辞

本研究の一部は JSPS 科研費 16K16109 と 16H02905 の助成を受けたものです．

参考文献

[1] Michael Färber, Achim Rettinger, and Boulos El Asmar. On emerging entity detection. *European Knowledge Acquisition Workshop*, pp. 223–238, 2016.

[2] Yue Sui and Xuecheng Yang. The potential marketing power of microblog. *Communication Systems, Networks and Applications (ICCSNA), 2010 Second International Conference on*, Vol. 1, pp. 164–167, 2010.

[3] Alan Ritter, Sam Clark, and Oren Etzioni. Named Entity Recognition in Tweets : An Experimental Study. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1524–1534, 2011.

[4] Zhaohui Wu, Yang Song, and C Lee Giles. Exploring Multiple Feature Spaces for Novel Entity Discovery. *Proceedings of the Conference on Artificial Intelligence (AAAI)*, pp. 3073–3079, 2016.

[5] David Graus, Daan Odijk, and Maarten de Rijke. The birth of collective memories: Analyzing emerging entities in text streams. *arXiv preprint arXiv:1701.04039*, 2017.

[6] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 1003–1011, 2009.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ソーシャルネットワークのコミュニティ構造が ソーシャルメディア上の投稿の拡散規模に与える影響の分析

Effects of Community Structure of a Social Network on the Cascade Sizes of Tweets on Social Media

津川 翔
Sho Tsugawa

筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

We investigate how the community structure of a social network affects information diffusion on social media. Empirical analyses using large-scale dataset of Twitter show that the community structure significantly affects the sizes of information cascades on social media. More specifically, the results of our regression analysis show that the number of retweets of a tweet spreading among different communities is significantly higher than that of a tweet spreading within a single community. We also tackle a problem of predicting the future popularity of a tweet in an early stage of its diffusion. We constructed prediction models of future popularity of a tweet using features obtained from community structures of a social network as well as the tweet's content. Through experiments, we show that the features extracted from community structure of a social network are effective for predicting the future popularity of a tweet.

1. はじめに

筆者はこれまで、ソーシャルメディアユーザ間の関係を表現したソーシャルネットワークのコミュニティ構造が、ソーシャルメディア上の投稿の拡散に与える影響を分析してきた [津川 16]。人と人との関係を表現したソーシャルネットワークの多くは、互いに密に接続されたノードのコミュニティと、それらのコミュニティ間を接続する少数のリンクで構成されるというコミュニティ構造を有している [Newman 04, Girvan 02]。このようなソーシャルネットワークのコミュニティ構造は、人と人との交流のパターンに影響を与える [Weng 13, Onnela 07, Tsugawa 15, 津川 16] 。[津川 16] では、代表的なソーシャルメディアである Twitter において、ユーザのコミュニティをまたがって拡散されたツイートは、コミュニティ内で拡散されたツイートよりも、将来のリツイート数の期待値が高いこと (図 1) などを示している。

[津川 16] では、ソーシャルメディア上の投稿の拡散に影響を与える要因として、コミュニティ構造のみに着目している。一方で、コミュニティ構造以外にも様々な要因がソーシャルメディア上の投稿の拡散に影響を与えることが知られている [Suh 10, Naveed 11, Stieglitz 13]。例えば、情報を投稿したユーザのフォロワー数や投稿に URL やハッシュタグが含まれるかどうかは、投稿の拡散規模に影響を与える主要因であることが示されている [Suh 10, Naveed 11]。これらの要因と比較した時に、コミュニティ構造が投稿の拡散規模に与える影響がどの程度強いかはこれまで明らかにされていない。

本稿では、ソーシャルネットワークにおけるコミュニティをまたがる投稿の拡散が、最終的な投稿の拡散規模に与える影響の強さを、投稿の拡散規模に影響を与える他の要因と比較することで分析する。代表的なソーシャルメディアである Twitter のデータを用いる。ツイート本文の URL の有無やツイートを投稿したユーザのフォロワー数などコミュニティ構造以外にツイートの拡散規模に影響を与える要因を制御変数とする重回帰分析により、コミュニティをまたがるツイートの拡散が、ツイートの最終的な拡散規模にどの程度の影響を与えるかを明らかにする。さらに、ツイートの拡散規模を予測する実験により、コミュニティに関する特徴量が拡散規模の予測にどの程度有用であるかを明らかにする。

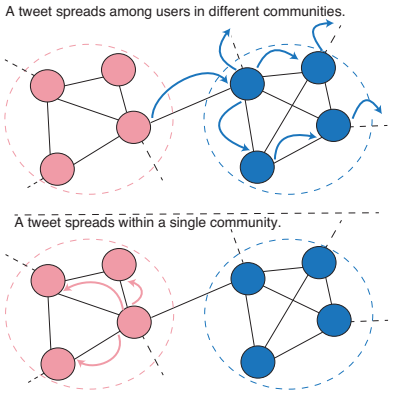


図 1: コミュニティ構造と投稿の拡散の関係：コミュニティをまたがって拡散されたツイートは、コミュニティ内で拡散されたツイートよりも、将来のリツイート数の期待値が高い [津川 16]

2. 分析

2.1 用いるデータ

分析には、[津川 16] と同じデータセットを用いる。本データセットには、日本語のツイートを投稿する 356,453 ユーザのフォロー関係を表現したソーシャルネットワークと、それらのユーザの投稿したリツイートの情報が含まれている。ユーザ間のフォロー関係は、Twitter API を用いて、2014 年 1 月 1 日〜11 日の期間に収集した。同様に、対象ユーザ 356,453 人の投稿したリツイートは、2014 年 1 月 1 日〜 31 日の期間に Twitter API を用いて収集した。これにより、対象ユーザの投稿した 14,001,533 件のツイート (オリジナルツイート) と、

連絡先: 津川 翔、筑波大学 システム情報系、茨城県つくば市天王台 1–1–1、s-tugawa@cs.tsukuba.ac.jp

表 1: 構築したソーシャルネットワークと抽出したコミュニティに関する統計量

ノード数	356,453
リンク数	29,535,582
最大弱連結成分に属するノードの数	351,870
弱連結成分の数	4,558
コミュニティ数	4,584
コミュニティの規模の最大値	77,695

それらに対する 63,462,387 件のリツイートを得た。これらのツイートから、ランダムに選択した 200 万件のオリジナルツイートと、それらに対するリツイートを以降の分析に用いた。

データセット中のソーシャルネットワークに対して、代表的なコミュニティ抽出アルゴリズムである Louvain 法 [Blondel 08] を適用することにより、各ユーザの属するコミュニティをを得た。本データセットのソーシャルネットワークと抽出したコミュニティに関する統計量を表 1 に示す。

2.2 重回帰分析

重回帰分析により、Twitter ユーザのソーシャルネットワークにおけるコミュニティ構造がツイートの拡散規模に与える影響を分析する。オリジナルツイート t に対する k 回目のリツイートを $r_t(k)$、ツイート t に対する総リツイート数を N_t と表記する。また、ツイート t を投稿したユーザを $u(t)$、リツイート $r_t(k)$ を投稿したユーザを $u(r_t(k))$ と表記する。各リツイート $r_t(k)$ に対して、そのリツイートが投稿された後のリツイート数、すなわち $N_t - k$ を目的変数とする。説明変数として、オリジナルツイート t を投稿したユーザ $u(t)$ とリツイート $r_t(k)$ を投稿したユーザのコミュニティが異なるか否かを用いる。また、他の説明変数として、ツイート t の本文の URL の有無、ハッシュタグの有無、単語数、ユーザ $u(t)$ のフォロワー数、リツイートを投稿したユーザ $u(r_t(k))$ のフォロワー数、ツイート t を投稿したユーザ $u(t)$ の属するコミュニティの規模、リツイート $r_t(k)$ を投稿したユーザのコミュニティの規模、リツイートの回数 k も用いる。重回帰分析に用いる変数を表 2 に示す。

$RTnum$ (リツイート $r_t(k)$ が投稿された後のリツイート数) を目的変数とし、表 2 に示す他の変数を説明変数とする回帰分析を行った。リツイート数の分布は裾の長い分布であるため、ここでは負の二項分布回帰 (negative binomial regression) を用いた。回帰分析により求めた偏回帰係数を表 3 に示す。

表 3 より、他の要因の影響を考慮しても、コミュニティをまたがるツイートの拡散は、ツイートの拡散規模を増加させる影響を与えていることがわかる。構築した回帰モデルにおける各変数の与える影響の大きさは、偏回帰係数を β とすると e^{β} から推定される。表 3 の e^{β} の値より、コミュニティをまたがるツイート拡散は、コミュニティ内での拡散と比べて、その後のリツイート数を約 1.4 倍にする影響を与えると推定される。一方、変数 URL および $hash$ の e^{β} の値はそれぞれ、約 4.2、約 1.05 である。これらのことは、コミュニティをまたがる拡散がツイートの拡散規模に与える影響の強さは、URL の有無やハッシュタグの有無が拡散規模に与える影響の強さと比較可能であることを示している。このことは、コミュニティをまたがるツイート拡散は、ツイートの拡散規模に影響を与える主要因の一つであることを示唆している。

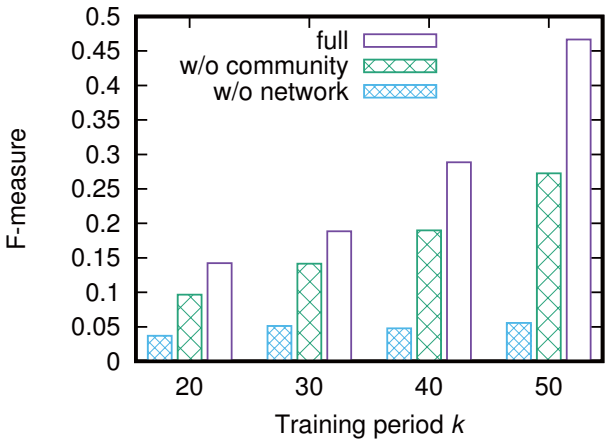


図 2: 用いる特徴量ごとのツイートの拡散規模予測精度の比較：ツイートが k 回リツイートされた時点でそのツイートの総リツイート数が 100 を超えるかどうかを予測した場合の F 値

2.3 拡散規模の予測

次に、コミュニティ構造に関する特徴量を用いて最終的なツイートの拡散規模を予測する実験を行う。これにより、コミュニティ構造に関する特徴量がツイートの拡散規模を予測するのにどの程度有用であるかを明らかにする。リツイート数そのものよりも、ツイートが大規模に拡散されるかどうかを予測することが実際には有用であると考え、ここでは、ツイートが k 回リツイートされた時点において、そのツイートの総リツイート数が 100 を超えるかどうか予測するタスクを設定する。以降、この k を学習期間と呼ぶ。なお、リツイート数が 100 以上のツイートは、データセット中の全ツイートのうち 0.5% である。

前節で用いたデータを学習用データとし、ランダムフォレストによって予測モデルを構築する。表 2 に示す変数に加えて、リツイートを投稿した k 人のユーザのうち、ツイートの投稿者と同じコミュニティに属するユーザの割合 ($comrate$)、リツイートを投稿した k 人のユーザのフォロワー数の平均 ($avedeg$) も特徴量として用いる。これら全ての特徴量を用いた予測モデル (full)、コミュニティに関する特徴量 (com、$comsize_{orig}$、$comsize_{rt}$、$comrate$) を用いないモデル (w/o community)、ネットワークに関する特徴量 (deg_{orig}、deg_{rt}、$avedeg$ ならびにコミュニティ関する特徴量) を用いないモデル (w/o network) を構築し、用いる特徴量によって予測の精度がどの程度異なるかを調査する。ここでは、学習期間 k が 20、30、40、50 のそれぞれの場合の予測モデルを用いる特徴量ごとに構築した。

学習データに含まれないツイートをデータセットからランダムに 100 万件抽出し、それらのツイートとリツイートをテストデータとした。これらのテストデータに対して、ツイートが $k = 20$、30、40、50 回リツイートされた時点において、そのツイートの総リツイート数が 100 を超えるかどうか予測した。図 2 に各モデルの予測精度を測る F 値を示す。

図 2 より、コミュニティに関する特徴量を用いたモデルの F 値は用いないモデルに比べて高いことがわかる。このことは、コミュニティの特徴量は投稿の拡散規模を予測する上で有用な特徴であることを示している。一方、モデルの予測精度自

表 2: 回帰分析に用いた変数

変数	説明
$RTnum$	リツイート $r_t(k)$ が投稿された後のツイート t のリツイート数
k	リツイート $r_t(k)$ が投稿されるまでのリツイート数 k
com	ツイート t とリツイート $r_t(k)$ を投稿したユーザのコミュニティが異なるかどうかを表すカテゴリ変数
$comsize_{orig}$	ツイート t を投稿したユーザの属するコミュニティに属するノード数
$comsize_{rt}$	リツイート $r_t(k)$ を投稿したユーザの属するコミュニティに属するノード数
deg_{orig}	ツイート t を投稿したユーザのフォロワー数
deg_{rt}	リツイート $r_t(k)$ を投稿したユーザのフォロワー数
URL	ツイート t が URL を含むかどうかを表すカテゴリ変数
$hash$	ツイート t がハッシュタグを含むかどうかを表すカテゴリ変数
$word$	ツイート t に含まれる単語数

表 3: 負の二項回帰分析により求めた各変数の偏回帰係数 (**: significant at 1% level)

目的変数: $RTnum$		
説明変数	Coeff. β	e^{β}
k**	3.388e-03	1.003
com**	3.528e-01	1.423
$comsize_{orig}$**	2.301e-06	1.000
$comsize_{rt}$**	8.180e-06	1.000
$\log(deg_{orig})$**	6.473e-02	1.067
$\log(deg_{rt})$**	-4.064e-02	0.960
URL**	1.427e+00	4.166
$hash$**	5.348e-02	1.055
$word$**	2.163e-02	1.022
constant***	1.571e+00	

体はそれほど高くないこともわかる。実用上有用な予測を実現するためには、既存の拡散規模を予測するモデル [Cheng 14] にコミュニティに関する特徴量を組込むなどして、さらなる予測精度の改善が必要であると考える。

3. まとめと今後の課題

本稿では、ソーシャルネットワークのコミュニティ構造が、Twitter におけるツイートの拡散規模に与える影響を分析した。重回帰分析により、コミュニティをまたがるツイートの拡散は、最終的なツイートの拡散規模を増加させる影響を与えること、その影響の強さは、ツイート本文の URL やハッシュタグの有無と比較可能であることを示した。さらに、ツイートの拡散の過程において、最終的なツイートの拡散規模を予測する実験を行った。その結果、コミュニティに関する特徴量はツイートの拡散規模を予測する上で有用な特徴量であることを示した。

今後の課題として、他のコミュニティ抽出手法を用いた場合の分析が挙げられる。また、ソーシャルネットワークのコミュニティに関する特徴を用いて、ツイートの拡散の初期の段階において最終的な拡散規模を高精度に予測するモデルを構築することも重要な今後の課題である。

謝辞

本研究の一部は JSPS 科研費 17H01733 の支援を受けている。

参考文献

[Blondel 08] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E.: Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p. P10008 (2008)

[Cheng 14] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J.: Can cascades be predicted?, in *Proceedings of the 23rd International Conference on World Wide Web (WWW'14)*, pp. 925–936 (2014)

[Girvan 02] Girvan, M. and Newman, M. E. J.: Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, Vol. 99, No. 12, pp. 7821–7826 (2002)

[Naveed 11] Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A.: Bad news travel fast: A content-based analysis of interestingness on Twitter, in *Proceedings of the ACM Web Science Conference 2011 (WebSci'11)*, pp. 1–7 (2011)

[Newman 04] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, No. 2, p. 026113 (2004)

[Onnela 07] Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A. L.: Structure and tie strengths in mobile communication networks, *Proceedings of the National Academy of Sciences*, Vol. 104, No. 18, pp. 7332–7336 (2007)

[Stieglitz 13] Stieglitz, S. and Dang-Xuan, L.: Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior, *Journal of Management Information Systems*, Vol. 29, No. 4, pp. 217–247 (2013)

[Suh 10] Suh, B., Hong, L., Pirolli, P., and Chi, E.: Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network, in *Proceedings of the 2nd IEEE International Conference on Social Computing (SocialCom'10)*, pp. 177–184 (2010)

[Tsugawa 15] Tsugawa, S. and Ohsaki, H.: Community Structure and Interaction Locality in Social Networks, *Journal of Information Processing*, Vol. 23, No. 4, pp. 402–410 (2015)

[Weng 13] Weng, L., Menczer, F., and Ahn, Y.-Y.: Virality prediction and community structure in social networks, *Scientific Reports*, Vol. 3, p. 2522 (2013)

[津川 16] 津川 翔：Twitter における情報拡散とコミュニティ構造の関係に関する一考察, 電子情報通信学会技術研究報告 (CQ2016-54), pp. 13–18 (2016)

[2C1] [General Session] 9. NLP / IR

座長:宮西 大樹（国際電気通信基礎技術研究所）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room C (4F Orchid)

[2C1-01] A Study of Document Summarization for Knowledge Presentation from Stored Information of Business Activities

○Yohko Konno[1], Takuya Nakamura[1], Masayuki Yoshida[2], Hidenori Kawamura[3] （1. Chowa Giken corporation, 2. HAJ Corporation, 3. Hokkaido University）

9:00 AM - 9:20 AM

[2C1-02] Solving the opinion summarization problem in English in the "Can a Robot Get into the University of Tokyo?" project

○Ryuichiro Higashinaka[1], Hiroaki Sugiyama[1], Hiromi Narimatsu[1], Hideki Isozaki[2], Genichiro Kikui[2], Kohji Dohsaka[3], Hirotoshi Taira[4], Tomoya Kita[4], Yasuhiro Minami[5], Takeru Kazama[5], Junji Yamato[6] （1. NTT Corporation, 2. Okayama Prefectural University, 3. Akita Prefectural University, 4. Osaka Institute of Technology, 5. The University of Electro-Communications, 6. Kogakuin University）

9:20 AM - 9:40 AM

[2C1-03] Sentiment Classification with Gated CNN using Spatial Pyramid Pooling

○Makoto Okada[1], Hirokazu Yanagimoto[1], Kiyota Hashimoto[2] （1. Osaka Prefecture University, 2. Prince of Songkla University, phuket campus）

9:40 AM - 10:00 AM

[2C1-04] An Application of Recursive Neural Tensor Network to Sentiment Analysis of Japanese Sentences

○Ryuichi Akai[1], Masayasu Atsumi[1] （1. Soka University）

10:00 AM - 10:20 AM

[2C1-05] A study on document classification focusing on the output side weight on Word2Vec

○Shuto Uchida[1], Tomohiro Yoshikawa[1], Felix Jimenez[1], Takeshi Furuhashi[1] （1. Nagoya university）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

営業活動の蓄積情報を利用した自動要約・知識提示に関する研究

A Study of Document Summarization for Knowledge Presentation from Stored Information of Business Activities

今野陽子[*1] Yohko Konno
中村拓哉[*1] NakamuraI Takuya
吉田優之[*2] Masayuki Yoshida
川村秀憲[*2] Hidenori Kawamura

[*1](株)調和技研 Chowagiken Corporation
[*1](株)調和技研 Chowagiken Corporation
[*2](株)北海道アルバイト情報社 HAJ Corporation
[*3]北海道大学 Hokkaido University

This study examines a method of effective utilization as knowledge in organizational stored information of business activities. As an approach, a information is presented by the document summarization using the real bulletin board system data. For a question sentence, the system retrieves information based on categories and feature words, and documents are summarized.

1. はじめに

Web の投稿データから，有益な情報を発見し応用する動きが活発である．一方で企業や組織の業務においても，日々多くのデータを記録している．例えばメール，電子掲示板(BBS: Bulletin Board System)，営業日報や議事録などの文書においても，重要なノウハウとなる情報が記録されている．現在，このような組織の蓄積情報に着目し，形式的な知識を可視化するための研究，特定業務における成功要因やリスク管理などのナレッジマネジメントに活用するサービスにニーズが高まっている(図 1)．これらの記録は，時系列に蓄積され，内容も多岐に渡る．キーワードによって検索しても，文書がそのまま表示される場合が多く，重要性の高い情報がどこに含まれているか把握できずに埋もれてしまう可能性が高い．本研究では，このような組織で蓄積された大量のテキストデータを対象として，社内の知識として有効活用することを目的に，その方法について検討する．本研究のアプローチとして，ユーザは質問文とカテゴリを入力し，システムはあらかじめカテゴリが付与された情報の中から，該当のカテゴリと，質問文に含まれる特徴語に基づいて，必要な情報を抽出する．抽出された情報は，自動要約により重要な部分のみを取り出し，質問への回答として提示する．情報検索の際に，特徴語と併せて，過去の掲示板データにおいて類似の意味で使用される語も検索のキーワードに含めて情報を抽出する．また自動要約では，特徴語に関係の強い文を優先的に採用する．実際に企業で蓄積された電子掲示板のデータを取り上げて評価を行った．最終的には社内で活用される自動要約・提示システムを目指している．

2. システムの構成

本研究で対象とするデータ は，アルバイト求人情報を扱う企業の社内スレッド型電子掲示板に投稿された 2007 年から 2017 年までの 36,000 件のテキストデータである．この掲示板は主に営業活動に関する社内の情報共有のために設けられ，効果的な事例や地域の顧客情報，企画の告知などが自由形式で投稿される．営業活動は，経験に基づく属人的成果が大きい．これまでは投稿データを整理や再利用することは無かったが，営業成果のノウハウが蓄積されていることから，営業活動の支援や標準化に役立てたいという要望がある．本研究では，質問に適した文書要約[平尾 2001]に着目し，具体的なサービスとして，質問に対して特定のキーワードをクエリとした情報検索を行い，該当情報を要約して提示するシステムを開発した．このようなシステムは，ユーザの問い合わせに対して大量のコーパスから確信度の高い回答を提示する対話エージェントへの応用も考えられる．図 2 は，自動要約・提示システムの構成であり，システムは大きく 5 つのプロセス(A)～(E)から構成される．

(A) 質問文を入力し，知りたい情報のカテゴリを選択する
(B) 質問文から特徴語を抽出する
(C) カテゴリと特徴語を用いて情報検索する
(D)検索でヒットした投稿について，自動要約を行う
(E)自動要約した結果を提示する

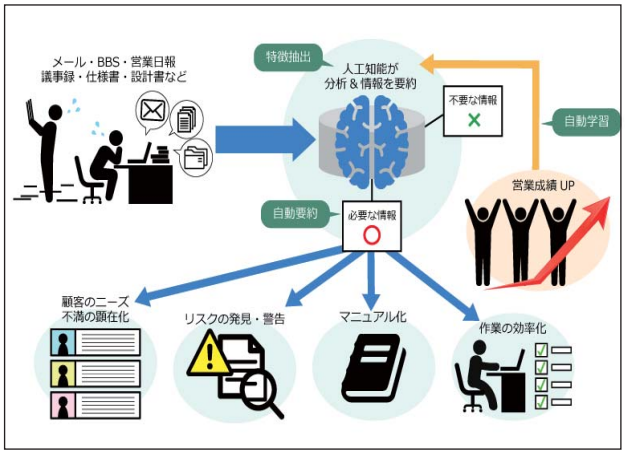


図 1　蓄積文書の活用イメージ

2.1 情報検索

上記(A)(B)について，効率的に情報を取得するためには，ユーザは必要なものは何かを示すことになり，また，多くの情報の中から必要なものを取り出すためには，キーとなる特徴を決めることが条件となる．特徴語は文を特徴づけるために掲示板データから抽出した語である．この特徴語の機能は，質問文からの検索で抽出する情報を的中させること，および質問に応じて要約内容を推定することである．

連絡先:[*1] 今野陽子，株式会社調和技研，北海道札幌市北区北 21 条西 12 丁目 2 北大ビジネススプリング 204 号室，ykonno@chowagiken.co.jp

(C)において，検索する掲示板のデータは，ひとつの投稿を 1 レコードとして，クエリに該当するレコードを全て抽出する．質問文に例えば「派遣」という特徴語が含まれる時，それは「求人情報」の話を指しているのか，あるいは「人材紹介」の話か，「派遣」という語だけで推測することは難しい．そのため，特徴語のみを検索の条件にするのではなく，あらかじめレコードごとにカテゴリを付与し，特徴語でレコードを抽出する前に，検索の候補となるレコードを絞り込むようにした．そして(D)において，レコードごとに特徴語を考慮してリアルタイムに自動要約した結果を表示する．また詳細に投稿の内容を確認できるように，アコーディオン形式で本文を表示できる機能を持つ．本システムを実現するために必要な要素技術として，大きく 3 つの機能(1)特徴語抽出，(2)カテゴリ分類，(3)自動要約が求められる．それぞれについて 4 章，5 章，6 章で説明する．

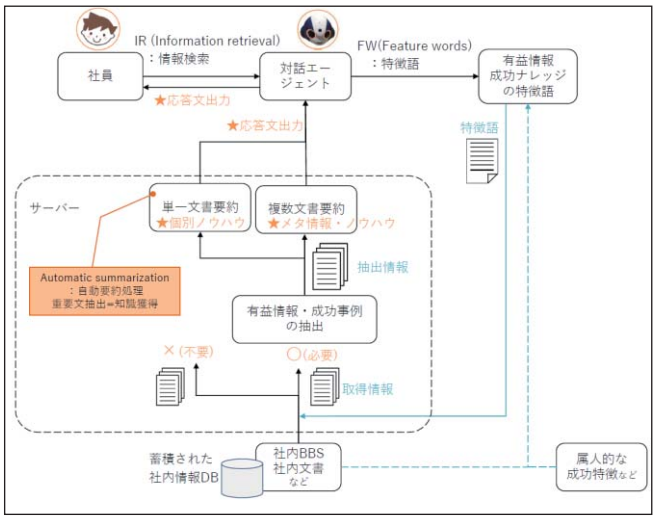


図 2　自動要約・提示システムの構成

3. 本研究の位置づけ

本研究における提示の部分となる自動約に着目して，要約の方法を検討する上で，その目的を明確にする[奥村 2005]．一般に要約の目的は「指示的要約」と「報知的要約」に大別される．指示的要約は，元の文書を読むべきかどうかを判断するための要約の事を示す．例としては新聞の見出しの一文が挙げられ，必要に応じて本文を読むことを前提としている．一方で，報知的要約は，元の文書の概要を伝える要約で，テレビニュースの字幕などが例として挙げられる．本研究で想定するサービスでは報知的な場合を取り扱う．次に，入力が単一文書か複数文書かによって分けられる．単一の場合とは，一つの文書から一つの要約を作成することであり，一方で，複数文書から作成する場合とは同じ事柄に対する複数の文書から，一つの要約を生成することをいう．本研究では，ある話題について複数の投稿が存在する社内文書を用いるが，現在は該当する投稿ごとに要約することを想定しているため，単一文書からの要約作成を対象としている．また，文書の要約には，その文書の概要を示す一般的な要約が必要な場合と，利用者がキーワードなどを指定してそのキーワードに関する要約を文書から生成する場合が存在する．ユーザから何らかの指定を受ける場合をクエリ依存，それ以外の場合をクエリ非依存と呼び，本研究ではクエリ依存型をベースとして要約文書を取得するシステムを想定する．そして，要約手法については要約対象の文書の中から重要と思われる文を抽出する抽出的要約と，意味を汲み取り抽象化を行った上で適切な要約を作成する生成的要約が存在する．後者に対しては意味表現の生成文の整合性等の問題があり非常に困難であるため，本研究では抽出的要約を行う．

4. 特徴語の抽出

本研究における特徴語は，掲示板に投稿されるそれぞれの内容を特徴づける語として位置づけられる．内容を特徴づけることから，特徴語の役割は，情報検索の際に，ユーザの質問文に含まれる語と検索したい情報のマッチングを図ること，また抽出した情報を自動要約する際に，質問文に対して特徴語に関わる文が優先的に取得されるように調整することである．この際，特徴語はユーザの問いの意図を汲み取っている必要がある．本研究のシステムでは，入力情報として質問文に含まれる特徴語を抽出し，ユーザの求める投稿文を検索するためのキーワードとして使用する．特徴語を獲得する方法として，ベース手法として TF-IDF を用いる．また，情報を検索するにあたり，質問文に含まれる語だけではなく，掲示板において似たような使われ方をする語を類義語として，検索のキーワードとして追加する．これは社内用語や業界用語と一般的な語を結ぶ意味でも効果がある．類義語を定義するための手法として，文書の中に現れる単語間の距離を計算してベクトルで表現する Word2Vec を用い，掲示板の全文書を学習する．近傍の語を Cosine 類似度で算出して類義語を取得する[Word 2013]．

検索の性能向上を図るため，いくつかの工夫が必要となる．一つ目に，質問文の内容から，カテゴリを推測して検索対象のカテゴリに加える．この方法として，予めカテゴリ分けした投稿データ(5 章) から，各カテゴリに頻出する語を抽出する．質問文の中に，あるカテゴリと関わりの強いカテゴリ頻出語が含まれる場合，ユーザによって該当のカテゴリが選択されていなくても(図 3)，検索対象として加える．また，例えば「活動」などの広く使われる語を曖昧語として定義し，曖昧後が質問に含まれている場合には，検索のキーワードから除くようにする．このようにして抽出された特徴語は，類義語の組を or で結び，類義語の組同士を and としてクエリ検索を行う．各語には，文書間の類似度に基づく優先度を付与して，自動要約処理における文の重要度計算に利用する[森 2002]．例えば，ある特徴語の優先度=1.0 の場合，要約処理では基本重要度=1.0 と設定し，優先度を加算して重要度=2.0，すなわち基本の 2 倍の重み付けを行った．

5. カテゴリ分類

本研究では，情報検索の際に，特徴語による検索と併せてカテゴリを指定することで，取得する情報の厳選を行う．そこで，あらかじめ掲示板の投稿内容に応じて分類を行い，投稿ごと(レコードごと)にカテゴリを付与する．掲示板の投稿内容がどのようなカテゴリに分けられるか，その種類については，掲示板データ提供元の企業にヒアリングを行い，以下の 10 種類のカテゴリを決定した．

(Category01) 営業活動報告
(Category02) 反応事例
(Category03) 採用事例
(Category04) 競合情報
(Category05) 人材紹介
(Category06) 地域情報
(Category07) イベント
(Category08) 特集
(Category09) 業務連絡
(Category10) 求人受注

カテゴリ分類のための手法として，ランダムフォレスト(Random Forest)に勾配ブースティング(Gradient Boosting)を用いるアンサンブル学習を行う．ランダムフォレストは，複数の決定木を組み合わせて，各決定木の分類結果から多数決で決定する手法である．また勾配ブースティングは，前回の学習結果を逐次的に取り入れながら学習モデルを生成する．分類のための教師データは，人手によりカテゴリを付与した 500 件の投稿を用意する．予備実験を行い，分類の精度の比較検証によって，実際に投稿データにカテゴリを付与した．比較した 3 種類の方法は，(Type1)ランダムフォレストにより，投稿ごとに最高スコアだったカテゴリを一つ付与，(Type2)ランダムフォレストにより，投稿ごとにスコアの閾値により複数のカテゴリを付与，(Type3)ランダムフォレストの(2)をベースに勾配ブースティングのアンサンブル学習である．S 分類精度の検証は，人手により正解のカテゴリを付与した投稿 20 件について，(Type1)～(Type3)の実行結果と比較した．(Type1)9.5%，(Type2)71.4%，(Type3)81.0%の精度で分類することが判り，現在のカテゴリ分類には(3)の方法を採用している．

6. 自動要約

本研究における自動要約は，3 章で述べたように報知的，単一文書，クエリ依存型，抽出的要約を対象とする．要約の方法として，元の文書を文分割し，要約に含める重要文を選択し，それらをつなげることで要約文を作成した[Luhn 1958]．

要約の処理は，形態素解析，構文解析，そして要約処理の順で行う．形態素解析器に Juman，構文解析器に KNP を用いる[笹野 2013]．文の重要度を計算する方法として，構文解析の出力結果に基づいて，各単語に付与された品詞情報，構文関係情報を用い，単語ごとに重みを付与し，文ごとの重要度計算を行う．また，入力された質問文の特徴語の優先度も，重みに加算する．各文のうち重要度の高い文を，要約文字数制限の範囲で，時系列に取得する．要約文字数は可変として 7 章で評価する．

リアルタイムな要約処理のための高速化の工夫として，蓄積データに対して，KNP による構文解析までの計算を事前に実行しておくことで，要約時に時間短縮を図った．また，自明な定型文，すなわち，お疲れ様です，おはようございます，よろしくお願いしますなどの文は，形態素解析の事前に，入力文から除去した．

7. 評価実験

図 3 は本研究で開発した自動要約・提示システムの入力画面である．ユーザが入力時に指定する項目は以下の 4 項目である．

(Input1) 質問文入力 (検索キーワードのみでも可)
(Input2) カテゴリ選択
(Input3) 類義語有／無
(Input4) 投稿期間

自動要約の一般的な評価として，内容，読み易さ，外的評価が挙げられる．本システムを評価する観点として，大きくは 3 つのポイントがあり，(1)要約内容が投稿原文に対して適当であるか，(2)ユーザの質問の意図を汲み取り，有益な情報を提示しているか，(3)システム全体として合理的であるかが考えられる．それらを考慮して，以下の項目について検証を行った．

(Test1) 投稿原文に対する要約文の妥当性(要点を得られる)
(Test2) 要約文の字数制限はどの程度が適当か
(Test3) 類義語を加味する効果はあるか
(Test4) 特徴語に応じて要約内容が調整されるか
(Test5) 重要で有益な情報を提示できるか

上記の(Test1)と(Test2)は関連があり，字数が多いほど原文に近い内容を表すことが可能であるが，あまり重要ではない内容の文も抽出される可能性が高くなる．反対に字数制限を少なくすると，より核心部分を抽出する必要があり，また，文をつなげた時に不自然になる可能性が高まる．検証方法として，社員 10 人へのヒアリングにより 100 字，200 字，300 字について確認を行った．その結果，100 字の要約が咄嗟に見やすいという意見に集約された．投稿原文から抽出される要約文の傾向としては，(1)顧客名，営業署名，個人名が含まれる文が要約文に採用されやすい，(2)100 字の制限により要約文は 2～3 文になることが多い，(3)話題について表す冒頭の文，(4)全体のまとめや結論を表す文を採用する，といった要約結果が大半となった．要点を獲得できるかについて，上記(1)(3)の顧客や話題に関する文，(4)まとめの文を採用することで，100 字内に概ね要点がまとめられた．しかし，掲示板への投稿は自由形式であり，必ずしも 1 つの投稿に 1 つの話題とは限らない．投稿文に顧客名が複数登場したり，複数の話題を扱う場合には，これらの文書の要点を 100 字の制約の中で抽出することが難しい．要約処理を改善するためには，文書分割や，生成的要約の導入が見込まれる．



図 3　自動要約・提示システムの入力画面

図 4 は，掲示板に投稿された情報の中から「美容関係の反応について知りたい」という質問文，および全カテゴリ選択，期間指定なしで検索した場合の実行結果である．上段は類義語=有の場合で，[美容 or エステ or ファッション]and[関係 or 築く or 信頼]による検索結果を示す．下段は類義語=無の場合で，[美容]and[関係]による検索結果を示す．(Test3)について，類義語=有の場合に，検索結果の拡張性を検索ヒット数で計算した場合，類義語=無の場合に対して 5 割増加する．また，図 4 の例で，掲示板データから作成した類義語の辞書において，[美容]の近傍にある語は [エステ][ファッション]と，いずれも美容に近い意味で使われる語を指している．一方で，質問に[美容関係]と入力したにも関わらず，[関係]の近傍にある語は[築く][信頼]と，いずれも類義語ではなく[関係を築く]や[信頼関係]といった共起性の高い語になっている．類義語の辞書を改良するためには，掲示板データの語を学習するだけではなく，語の取捨選択が必要である．

図 5 は入力が異なる質問文に対する，同じ投稿原文の要約結果である．(Test4)について，)図 5 の例では，上段の質問文は「美容」についての反応，下段の質問文は「帯広」の「美容」についての反応を問い合わせした．要約結果の 1 文目は，それぞれ美容に関する顧客であるが，複数の顧客が併記されている場合，下段では帯広に特化することで帯広店の顧客を取り上げ

る. 2 文目はユーザが体験したプラスになる情報と成功体験について述べている. 投稿原文の中でも重要度の高い文といえる. これは質問が異なれば必ずしも要約内容が変わるということではなく, 要約処理の文の重要度計算に依存する.



図 4　類義語を含む検索(上)
含まない検索(下)の実行結果



図 5　質問の違いによる要約結果
美容の反応(上)と帯広の反応(下)

(Test5)については, 要約により, 投稿文書からユーザに有益であり知識となる情報を提示できることが目標である. そのための検証方法として, 実際に社員に対して要約結果の検証を行った. 掲示板は営業活動における情報共有のための投稿であることから,「営業活動の支援になる情報が掲示されている要約文」にチェックするシンプルな指標とした. 結果として, 質問文 10 問中 20%の要約が有益な情報を提示していると評価された. また, 評価されなかった要約内容ついては, (1)営業ポイントではない原文の要約, (2)現状報告, (3)部分的な内容の抜粋, といった内容に大別された.

今後の取り組みとして, 検証結果に基づいて, 詳細に調査を行う. (A)有益な情報の提示に向けた分析を行う, (B)要約の評価手法について検討し, 例えば検索キーワード, 有益に関わるキーワードの出現確率を指標とすることで, 要約結果を評価し, スコアで順位付けして表示することや, 厳選した情報を表示することが考えられる. また, (C)分析, 評価に基づいて, ユーザが意図する情報と質問と関連付けるための特徴語と優先度の算出方法について検討を行う.

8. まとめ

本研究では, 組織で蓄積された大量文書について, 社内の知識として有効活用することを目的に, 自動要約・提示システムの検討を行った.

(1) 営業活動の情報共有ツールである電子掲示板の蓄積情報から, ユーザのノウハウとなる有益な情報を獲得するための方法について検討した.

(2) 提案システムは, ユーザが質問文とカテゴリを入力することで, 入力情報に基づいて履歴情報を検索するともに, 投稿された文書の自動要約により重要な部分を回答として提示する.

(3) 自動要約処理分類上の位置づけは, 報知的で単一文書によるクエリ依存の抽出型要約を対象として, 処理方法の検討を行った.

(3) 情報を検索する際には, 特徴語と併せて蓄積された過去データに基づく類義語を加味し, 自動要約においては, 特徴語に関係の強い文を重要な文として優先的に提示するようにした.

(4) 実際の電子掲示板データを用いて検証を行い, 要約内容文の妥当性, 文字数, 特徴語の効果, 有益情報の獲得について評価と考察を行った.

(5) 要約内容の検証結果に基づいて, 今後, 有益情報の可視化に取り組み, 有益情報の評価方法や表示方法について検討することで, システムの改良を行う.

参考文献

[平尾 2001]　平尾努, 佐々木裕, 磯崎秀樹: 質問に適応した文書要約手法とその評価, 情報処理学会論文誌 42(9), pp.2259-2269, 2001.

[奥村 2005]　奥村学, 難波英嗣: テキスト自動要約, オーム社, 2005.

[Word 2013]　Word2Vec: https://code.google.com/p/word2vec/, 2013.

[森 2002]　森辰則: 検索結果表示向け文書要約における情報利得比に基づく語の重要度計算, 自然言語処理 9(4), pp.3-32, 2002.

[Luhn 1958]　Luhn, Hans Peter: The automatic creation of literature abstracts, IBM Journal of research and development 2(2), pp.159-165, 1958.

[笹野 2013]　笹野遼平, 河原大輔, 黒橋禎夫, 奥村学: 構文・述語項構造解析システム KNP の解析の流れと特徴, 言語処理学会第 19 回年次大会発表論文集, pp.110-113, 2013.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

「ロボットは東大に入れるか」プロジェクトの英語における意見要旨把握問題の解法

Solving the opinion summarization problem in English in the "Can a Robot Get into the University of Tokyo?" project

東中 竜一郎 *1 Ryuichiro Higashinaka
杉山 弘晃 *1 Hiroaki Sugiyama
成松 宏美 *1 Hiromi Narimatsu
磯崎 秀樹 *2 Hideki Isozaki
菊井 玄一郎 *2 Genichiro Kikui
堂坂 浩二 *3 Kohji Dohsaka
平 博順 *4 Hirotoshi Taira
喜多 智也 *4 Tomoya Kita
南 泰浩 *5 Yasuhiro Minami
風間 健流 *5 Takeru Kazama
大和 淳司 *6 Junji Yamato

*1NTT コミュニケーション科学基礎研究所 NTT Communication Science Laboratories
*2岡山県立大学 Okayama Prefectural University
*3秋田県立大学 Akita Prefectural University
*4大阪工業大学 Osaka Institute of Technology
*5電気通信大学 The University of Electro-Communications
*6工学院大学 Kogakuin University

We have been working on the English problems in the "Can a Robot Get into the University of Tokyo?" project. Focusing on the opinion summarization problem, we show that our method, which uses deep learning with a large number of English problems, can improve the accuracy over a Word2vec-based method. We perform qualitative analyses of the trained models and results, and discuss our findings with regards to machine comprehension.

1. はじめに

我々は「ロボットは東大に入れるか」プロジェクト [Arai 16] において，引き続き英語（特に，センター試験の英語問題）に取り組んでいる．[東中 17] では，英語問題を短文問題，複数文問題，長文問題に分け，問題文がほぼ一文からなる短文問題はおおむね解けるものの，複数の文からなる複数文問題は 30%〜40%程度，さらに長い長文問題についてはチャンスレベルの正解率であることを報告した．

本稿は，複数文問題のうち，意見要旨把握問題の解法について述べる．意見要旨把握問題は，複数人の議論において，各発言者が提示した意見の要旨を最もよく表す文を 4 つの選択肢（要旨候補）から選ぶ問題である．具体例を図 1 に示す．

[東中 17] では，我々は Word2vec[Mikolov 13] の類似度に基づく従来手法（提示意見と要旨候補の Word2vec に基づくコサイン類似度を用い，最も類似した要旨を正解とするもの）と深層学習ベースの手法を比較し，Word2vec の類似度に基づく手法のほうがよいとの報告を行った．本稿では，意見要旨把握問題の改善手法について述べる．具体的には，RACE と呼ばれる大規模な英語問題のデータセットを用いた深層学習の手法により，Word2vec の類似度に基づく手法よりも高精度に意見要旨把握問題が解けることを示す．

2. データセット

ここでは，意見要旨把握問題の自動解答に用いるデータセットについて説明する．本研究では，解法の性能を測るベンチマークデータとして，大学入試センター試験の本試験及び追試験の過去問, 代ゼミセンター模試, ベネッセ模試, 独自に収集したその他の問題を合わせた, 合計 234 問を収集した．このうち 120 問を訓練・開発セット（dev），114 問をテストセット（test）として用いる．この意見要旨把握問題のベンチマークデータは，丁寧に作成された良質な問題である一方，そのまま機械学習の訓練データとして用いるには量が不足するという問題がある．そこで本研究では，RACE データセット [Lai 17] と呼ばれる，中国人中高生向けの英語試験を大量に収集したデータを活用し，機械学習による解法を試みる．RACE データセットの特徴として，いわゆる長文読解に該当する問題であり，本文中の意見との含意関係を問うような，意見要旨把握に類似した問題を含む点がある．現在提案されている Reading Comprehension 関連のデータセットでは，分量自体は数十万問以上と非常に多いものの，そのほとんどが単語もしくは短い語句を解答させる問題であり，意見要旨把握問題との差異が大きい．そのため，こうした問題は，意見要旨把握問題を解く上での訓練データとして用いることは難しい．一方 RACE デー

Stephen: Thank you, Dr. Ishii. I agree we are living in a time when technology will soon improve even more rapidly. Looking back at the 1900s shows us how people faced rapid changes in their societies. I think this has lessons for us today. One of the biggest changes of the 20th century was the rise of a global society. I believe airplanes made this possible. For the first time, people could travel quickly to the farthest corners of the earth and experience life in other countries. Certainly telephones and the Internet had an impact as well. But there's no substitute for traveling to new places and actually meeting people.
Sue: I've heard this opinion before, Stephen. Are you saying [32] ?

選択肢:
(1) airplanes helped create our global society
(2) foreign travel was not possible before the 1900s
(3) technology will soon change more slowly
(4) telephones and the Internet were more important than airplanes

図 1: 意見要旨把握問題の例（ベネッセ模試 2016 年 6 月第 3 問 C）．正解は (1)．

連絡先: higashinaka.ryuichiro@lab.ntt.co.jp

タセットは，分量自体は 9 万 7 千問程度と中程度の大きさながら，選択肢に文を含む，作成が難しい問題を多数含んでいるため，意見要旨把握問題の自動解答に有用であると予想される．

3. 手法 : SAR/GAR を用いた意見要旨把握問題の自動解答

Stanford-Attentive Reader (SAR) モデル [Chen 16] および Gated-Attention Reader (GAR) モデル [Dhingra 17] は，アテンション機構を持つリカレントニューラルネットワーク (RNN) の一種であり，いくつかの穴埋め型の Reading Comprehension タスクにおいて，高い精度が得られており注目されている．我々の RACE データセットと東ロボのベンチマークデータを用いた評価実験においてもこれまでにない高得点を得ることが確認されている [喜多 18]．本稿は同じ手法を用い，詳細な分析を行っているものである．

以下，本研究で使用した SAR/GAR モデルを用いた意見要旨把握問題の自動解答手法について GAR モデルを中心に述べ，後で SAR モデルとの違いについて説明する．

3.1 問題本文と質問文のエンコード

まず，問題本文 D と質問文 Q について，文を構成する単語を単語ベクトルに変換し，それぞれ $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, ..., x_{|D|}^{(0)})$, $Y = (y_1, y_2, ..., y_{|Q|})$ とする．ここで $|D|$ と $|Q|$ は問題本文および質問文の単語長である．

次に，双方向 GRU [Cho 14] を用いて，問題本文および質問文の単語ベクトルをエンコードする．

$$\begin{aligned} D^{(k)} &= \mathrm{BiGRU}_{\mathrm{D}}^{(\mathrm{k})}(X^{(k-1)}) \\ Q^{(k)} &= \mathrm{BiGRU}_{\mathrm{Q}}^{(\mathrm{k})}(Y) \end{aligned}$$

このとき k 層目の $D^{(k)}$ に関しては $k-1$ 層目のベクトル $X^{(k-1)}$ をエンコードする．

そして，この $D^{(k)}$ と $Q^{(k)}$ とを用いて，

$$\begin{aligned} \alpha_i^{(k)} &= \mathrm{softmax_i}(Q^{(k)T} d_i^{(k)}) \\ \tilde{q}_i^{(k)} &= Q^{(k)} \alpha_i^{(k)} \\ x_i^{(k)} &= d_i^{(k)} \odot \tilde{q}_i^{(k)} \end{aligned} \tag{1}$$

を求める．この機構は「ゲート付アテンション」と呼ばれており，

$$X^{(k)} = \mathrm{GA}(D^{(k)}, Q^{(k)})$$

で表す．このようにして第 k 層の双方向 GRU の入力である $X^{(k)}$ を得，順次，次層のパラメータを計算する．ここで，$d_i^{(k)}$ は $D^{(k)}$ の i 番目の要素，$x_i^{(k)}$ は $X^{(k)}$ の i 番目の要素を表す．また，$\odot$ はベクトルの要素同士の積を表す．

最後に，選択肢 i が正解である確率 p_i の計算は, Lai らが 4 択 RC タスクで用いた方法 [Lai 17] と同じく以下のように計算した．

まず，4 つの選択肢それぞれについて，双方向 GRU でエンコードする．そして最終時刻でのこの双方向 GRU の隠れ層をそれぞれ，$h^{o_1}, h^{o_2}, h^{o_3}, h^{o_4}$ と書く．

次に，質問文に特化した問題本文のベクトル表現 s^d を双線型アテンション機構 [Chen 16] を用いて次の式で計算する．

$$\begin{aligned} \alpha_i &= \mathrm{softmax_i}((X_i^{(k)})^T W_1 q^{(k)}) \\ s_d &= \sum_i \alpha_i X_i^{(k)} \end{aligned} \tag{2}$$

ここで，$q^{(k)}$ は $Q^{(k)}$ の最終時刻での隠れ層である．

最後に，選択肢 i が正解である確率 p_i を次の式で計算する．

$$p_i = \mathrm{softmax_i}(h^{o_i} W_2 s^d)$$

ここで，W_1, W_2 は学習可能なパラメータである．

このように，GAR モデルは，各隠れ層で毎回アテンション機構を利用する「マルチホップ機構」と，アテンションにゲート機構が組み込まれた「ゲート付アテンション」を持っているのが特徴であり，複数の箇所にアテンションを考慮することが可能となっている．

SAR モデルの場合は, アテンションがゲート付アテンションではなく通常のアテンションであることが異なっている。

4. 結果

4.1 実験設定

3. 節で述べた GAR と SAR を用いて正解率を比較する．学習およびテストに用いるデータは，RACE の train データ 87860 問，大学入試センター試験の本試験と追試験，代ゼミセンター模試，ベネッセ模試，独自に収集したその他の問題を合わせた合計 234 問である．RACE のデータについては，独自に表記誤り等を修正したデータを用いた．これらのデータを，学習データセット (train)，開発データセット (dev)，テストデータセット (test) に分け，開発データセットで正解率が最大になった epoch でのテストデータセットの正解率を比較する．

データセットの分け方は，次の 2 通りの設定を試した．1 つ目の設定では，RACE の train データを学習データセットに，意見要旨把握問題 234 問のうち，120 問を開発データセット，114 問をテストデータセットに分けた．2 つ目の設定では，RACE の train データ +120 問の開発データセットのうち 60 問を学習データセットとして，残りの 60 問を開発データセット，1 つ目と同じ 114 問をテストデータセットに分けた．

GAR および SAR におけるモデルの安定性を評価するために，いくつかのパラメータを変えて正解率を評価する．変動させたパラメータおよび設定した値は，隠れ層数 1 もしくは 2, ユニット数 128 もしくは 256，学習率 0.1 もしくは 0.3 である．それ以外のパラメータについては，固定値を用いた．dropout は一般に良く用いられる 0.5，ランダムシードは [Lai 17] においてデフォルトで設定されている 1013 を用いた．語彙サイズは，50000 とした．大学受験に必要な語彙数よりも多いが，意見要旨把握に良く出現する，人名等が未知語となり悪影響を与えることを懸念し，少し多めに設定した．また最適化に用いる勾配法は SGD を用い，epoch は 100 まで行った．ただし，2 層 256 ユニット数のモデルについては，時間の関係で epoch は 50 までで評価した．

また，従来手法として，意見文書の各文と選択肢を表す文を，Word2Vec [Mikolov 13](w2v) を用いて得られた単語ベクトルの集合を平均したベクトルで表現し，各文と選択肢の文のコサイン類似度を計算し，最も類似した選択肢を答えとする手法とのスコアを比較する．

4.2 結果

表 1 に結果を示す．dev acc は東ロボ dev の正解率，test acc は東ロボ test の正解率である．全体をみると，SAR において，1 層 256 ユニット数で学習率を 0.3 に設定した時には，正解率が 0.2 と著しく低い．本設定での epoch ごとの正解率をみると，RACE データおよび東ロボ dev/test 全てにおいて不動であり，学習が進んでいなかった．従って，学習できなかった本設定を除いて評価を行う．表下部の平均値においても，本設定を除いた正解率のみを用いて算出した．GAR および SAR に

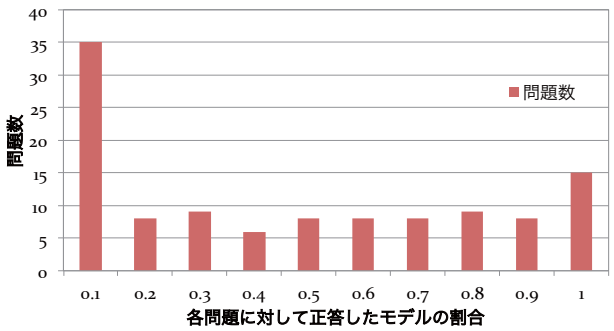


図 2: 正答したモデルの割合と問題数

おいて，それぞれ 16 個のモデルの結果を比較すると，test の正解率はいずれも 0.42±0.05 程度，0.44±0.05 程度に収まっており，モデルの違いによる正解率の差はほとんど見られなかった．従って，正解率の観点では，パラメータの変動によらず安定した結果が得られていると言える．

また意見要旨把握問題の dev の半分を学習データセットに含めた場合の結果を比較した場合，GAR では，意見要旨把握問題の dev を含めると，dev と test ともに 0.01 ポイント正解率が低下した．一方，SAR では，意見要旨把握問題の dev を含めない方が，dev で 0.03 ポイント，test で 0.01 ポイント正解率が向上した．このことから，学習に含めた意見要旨把握問題のデータセットは，RACE の train データの総量に対して，0.1% にも満たないが，SAR においては，意見要旨把握問題特有の特徴を何らか学習し，正解率の向上に寄与できた可能性がある．なお，w2v を用いた手法での同テストセットの正答率は，0.3 前後にとどまっていたことから，GAR/SAR で RACE データを学習することで，約 15 ポイント正解率が向上することがわかった．

4.3 モデルごとの性能比較

アテンションを用いる手法と w2v との正誤の分布を評価する．アテンションを用いる GAR および SAR については，モデルによる正答率のばらつきが小さい問題を対象として，正誤問題数の分布を評価した．図 2 は，横軸に正答したモデルの割合，縦軸にその問題数を表す．なお，正答したモデルの割合は，0.1 のとき，0 から 0.1 まで，0.2 のとき，0.1 から 0.2 までの正解率を含む．0.1 および 1 において，正答したモデルの割合が他と比べて高く，全体的な分布が凹型になっていることから，正答する傾向にある問題と誤答する傾向にある問題があることが分かる．

そこで．図 2 において，7 割以上のモデルが正答した問題を GAR および SAR で正解した問題，0 から 3 割のモデルが正答した問題を，GAR および SAR で誤答した問題として分類し，w2v との正答誤答問題数を調べた．その結果を表 2 に示す．GAR/SAR の 7 割のモデルが正答する問題 32 問のうち 20 問が w2v のモデルにおいて不正解であり，w2v で正答する 33 問のうち 16 問が GAR/SAR で不正解であった．我々は，これらの 20 問，16 問を分析することで，GAR/SAR におけるアテンションの効果および，アテンションのモデルに足りない特徴を調べることができると考えた．次節に分析を示す

5. 分析

5.1 定量評価

アテンションの効果を分析するために，アテンション機構の利用有無による正解率を比較する．アテンション機構を利用しない場合には，(2) の算出において $q^{(k)}$ を削除する．学習とテストの両者において利用する (有)，利用しない (無) のそれぞれの組み合わせについて評価する．分析の対象は表 1 において test の正解率が最も高い 0.48 であった 2 つのモデルとし，それぞれの正解率で評価する．結果を表 3 に示す．test の正解率は，All に示す．また，アテンションに寄与する質問文（意見要旨把握問題において空所のある文）に，正答する上で必要な情報が含まれているかどうかで問題を分類し，その正解率も比較した．質問文の重要度は，質問文が So, you say _. のように特に情報が含まれていない場合に「低」，それ以外を「高」とした．

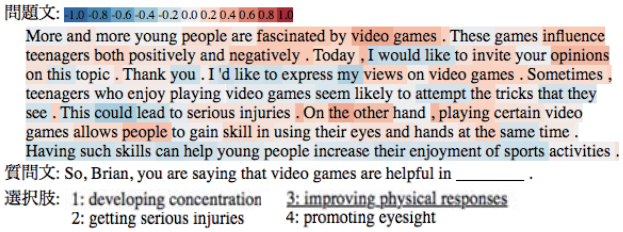


図 3: GAR で正答し，w2v で誤答した問題の例（2007 年のセンター試験 3B-1）．アテンションのかかり具合を濃淡で表している．

結果より，アテンション機構を用いて学習したモデルでは，アテンション機構を用いないと正解率が低下するが，アテンション機構を用いずに学習したモデルの正解率は，アテンション機構を用いて学習およびテストした結果と比較して若干の低下にとどまっていた．このことから，アテンション機構の効果は限定的と言える

5.2 定性評価

限定的な効果のアテンション機構であるが，今回若干の改善が見られてはいるので，それがどのように寄与しているかを，代表的な問題を 2 問抽出して確認する．アテンションを分析する対象として，GAR のモデルを用いる．図 3 は，w2v では不正解であったが GAR で正答した問題での GAR のアテンションを可視化したものである．この問題では，w2v では 2 を，GAR では正解の 3 を解答している．2 の選択肢には，問題文中と全く同じフレーズが出現しており，語の一致や表層的な近さで選ぶと誤答となる問題である．一方で，アテンションを用いた場合，質問文に含まれる video games が問題文中に出現する場合に高いアテンションがかかっており，いずれも正答の選択肢を選ぶ上でヒントになる箇所であることがわかる．従って，GA での正答は複数文をみることができたことによる効果である可能性が高い．

次に，w2v では正解であったが GAR で誤答した問題のアテンションを図 4 に示す．この問題では，w2v では正解の 2 を，GAR では 3 を解答している．表層的に近い文を選択するだけで正答できる問題であるが，GAR では good leader に関わる語にアテンションがかかっており，それらは本問題文中で頻繁に出現する語の一つであることから，アテンションが悪い影響を与えているように見える．

以上より，質問文中に，問題文中の注目箇所を制限するための重要な単語が含まれる問題に対しては，アテンションを用いる GAR 等の手法の効果が見られるが，問題文のトピックや問題文中に高頻度で出現する単語が含まれる問題に対しては，逆に悪影響となる可能性もあるとわかった．

表 1: 様々なパラメータで実験したアテンションモデルでの正解率の比較

パラメータ			GAR/RACE		GAR/RACE+dev		SAR/Race		SAR/Race+dev	
層数	ユニット数	学習率	dev acc	test acc	dev acc	test acc	dev acc	test acc	dev acc	test acc
1	128	0.1	0.52	0.4	0.53	0.45	0.54	**0.48**	0.6	0.4
1	128	0.3	0.52	0.43	0.53	0.38	0.54	0.45	0.63	0.46
1	256	0.1	0.51	0.39	0.55	0.37	0.52	0.46	0.57	0.47
1	256	0.3	0.52	0.43	0.55	**0.48**	0.25	0.2	0.25	0.2
2	128	0.1	0.54	0.42	0.52	0.43	0.58	0.43	0.57	0.44
2	128	0.3	0.57	0.38	0.60	0.46	0.54	0.44	0.52	0.42
2	256	0.1	0.48	0.42	0.51	0.46	0.51	0.39	0.52	0.4
2	256	0.3	0.46	0.38	0.47	0.36	0.52	0.43	0.52	0.45
Average			0.52	0.41	0.53	0.42	0.53	**0.44**	0.56	0.43

表 2: 手法毎の正答する問題数の分布

	正答モデルの割合	w2v		GAR/SAR の正誤割合
		正解	不正解	
GAR SAR	7 割以上	12	**20**	0.28 (32/114)
	3~7 割	5	25	0.26 (30/114)
	3 割以下	**16**	36	0.46 (52/114)
w2v の正誤割合		0.29 (33/114)	0.71 (81/114)	

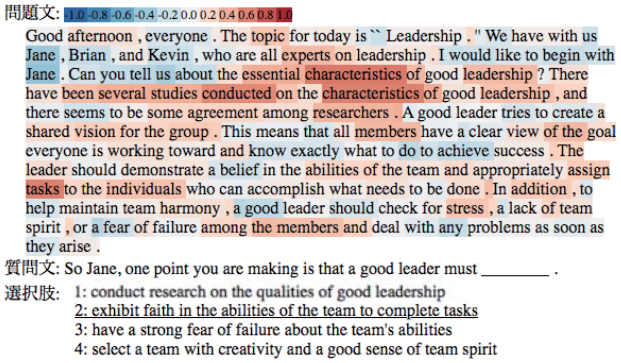


図 4: GAR で誤答し，w2v で正答した問題の例（2015 年のセンター試験追試 3C-1）

6. まとめ

本稿では，「ロボットは東大に入れるか」の英語における意見要旨把握問題の解法について説明した．深層学習と大規模な英語問題のデータセットを用いることで，Word2vec に基づく手法よりも高い正解率を達成した．具体的には，Word2vec に基づく手法は 30%の正解率だったところを，44%まで引き上げることができた．今後は，定性的分析やアテンション分析を通し，この正解率をさらに改善させる手法について検討していきたい．

謝辞

本研究を推進するにあたって，大学入試センター試験問題のデータをご提供下さった独立行政法人大学入試センターおよび株式会社ジェイシー教育研究所に感謝いたします．実験データをご提供くださいました学校法人高宮学園，株式会社ベネッセコーポレーションに感謝いたします．

表 3: アテンションの効果の比較

	アテンション		dev	test		
	学習	テスト		All	質問文の重要度	
					高	低
GAR	有	有	0.55 (33/60)	0.48 (55/114)	0.45 (18/40)	0.5 (37/74)
	有	無	0.37 (22/60)	0.42 (48/114)	0.48 (19/40)	0.39 (29/74)
	無	無	0.53 (32/60)	0.45 (51/114)	0.43 (17/40)	0.46 (34/74)
SAR	有	有	0.54 (65/120)	0.48 (55/114)	0.45 (18/40)	0.5 (37/74)
	有	無	0.39 (47/120)	0.27 (31/114)	0.2 (8/40)	0.31 (23/74)
	無	無	0.54 (65/120)	0.46 (52/114)	0.43 (17/40)	0.49 (36/74)

参考文献

[Arai 16] Arai, N.: Todai Robot Project., in *NTCIR* (2016)

[Chen 16] Chen, D., Bolton, J., and Manning, C. D.: A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task, in *Proc. of ACL 2016*, pp. 2358–2367 (2016)

[Cho 14] Cho, K., Van Merrinboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in *Proc. of EMNLP 2014*, pp. 1724–1734 (2014)

[Dhingra 17] Dhingra, B., Liu, H., Yang, Z., Cohen, W., and Salakhutdinov, R.: Gated-Attention Readers for Text Comprehension, in *Proc. of ACL 2017*, pp. 1832–1846 (2017)

[Lai 17] Lai, G., Xie, Q., Liu, H., Yang, Y., and Hovy, E.: RACE: Large-scale ReAding Comprehension Dataset From Examinations, in *Proc. of EMNLP 2017*, pp. 785–794 (2017)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013)

[喜多 18] 喜多 智也, 平 博順 : Gated-Attention Reader を用いた英語意見要旨把握問題の自動解答, 第 24 回言語処理学会年次大会, D5-3 (2018)

[東中 17] 東中 竜一郎, 杉山 弘晃, 成松 宏美, 磯崎 秀樹, 菊井 玄一郎, 堂坂 浩二, 平 博順, 南 泰浩, 大和 淳司 :「ロボットは東大に入れるか」プロジェクトにおける英語科目の到達点と今後の課題, 2017 年度人工知能学会全国大会予稿集, pp. 2H2–1 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Spatial Pyramid Poolingを用いたGated CNNによる評判分析

Sentiment Classification with Gated CNN using Spatial Pyramid Pooling

岡田　真 *1　　柳本 豪一 *1　　橋本　喜代太 *2
Makoto Okada　Hidekazu Yanagimoto　Kiyota Hashimoto

*1大阪府立大学　　*2Prince of Songkla University Phuket Campus
Osaka Prefecture University

The paper proposes a neural sentiment classification system with Gated CNN applied Spatial Pyramid Pooling. We apply Gated CNN to feature generation from text but not language model construction. Moreover, to generate the final single feature vector the proposed method uses max pooling and Spatial Pyramid Pooling. In experiments we evaluate the proposed method with real datasets: amazon product reviews. We confirmed the proposed method achieves the same performance as SVM with linear kernel without any customization.

1. はじめに

評判分析はインターネット上の大量の文書情報に含まれる筆者の意見や評価 (好評，不評) を理解するために重要かつ代表的な手法の一つとなっている．自然言語処理を用いた評判分析は多くの文書の内容を正確に理解する手法としてその重要性は高待っている．

自然言語処理分野では文書を単語の集合としてあらわす Bag of Words (BOW) モデルがよく用いられている．BOW モデルではそれぞれの単語は独立に扱われ，単語の意味的なつながりは失われる．そのため，潜在的意味解析 (latent semantic analysis, LSA) 等の手法を用いて意味の近い単語を集約する手法が従来用いられてきた．

近年，文書コーパスから単語の意味を表すベクトルを生成する手法が注目されている．それにより構築されたベクトルを単語の分散表現ベクトルと呼ぶ．分散表現ベクトルは単語の意味を計算により比較，評価することが可能なため，自然言語処理関連分野では自動翻訳やテキスト生成などさまざまな分野で用いられている．また，再帰的ニューラルネットワーク (reccurent neural network, RNN) を用いて，文の分散表現を作成し，文間の内容を比較するなどの応用も提案されている．

近年の自然言語処理分野ではニューラルネットワークを用いた言語モデルが提案されており，さまざまな分野で用いられている．Long Short-term Memory(LSTM) や Gated Recurrent Unit (GRU) といった RNN を用いた言語モデルが代表的である．RNN が用いられているモデルは距離の近い単語の関係を考慮して処理することが可能なため，言語処理と相性が良いと言われている．

畳み込みニューラルネットワーク (convolutional neural network, CNN) は，人間の視覚処理を基にしたニューラルネットワークモデルであり，その性質上，画像処理分野での成果が多い．CNN は畳み込み層とプーリング層から構成されている．畳み込み層では入力データを小さな領域に分けて，その中の要素を重み付きで集約する．重み付き小領域用のデータをフィルタ (カーネル) と呼ぶ．プーリング層では，重要な特徴を最大プーリングや平均プーリングなどのさまざまな戦略に基づいて選出する．CNN を GPGPU を利用した並列計算処理と相性がよく，データ処理や学習を高速におこないやすい．その特性から近年では自然言語処理分野でも CNN を用いた言語モデルが注目を集めている．単語間の距離を考慮する RNN と違い，CNN では入力データは距離に関係なくフラットに扱う特徴がある．

文書中のある種の単語は，その文の評価情報に強い影響を与えると仮定される．文書中のどの単語が評価情報に強い影響を与えているかを予測するのは難しい．我々は CNN を用いた言語モデルがそれらの予測を可能にするのではないかと考えている．重要な単語を適切に抽出するためには，不必要な単語をうまく無視する必要がある．RNN を用いた処理では各文中のどの位置の単語が意味的に重要かという情報をゲート構造を用いて扱えるようにしている．CNN を用いた言語モデルにおいて，ゲート構造を用いた手法が注目されている．

本論文では，我々は Gated CNN を用いた評判分析を提案する．Gated CNN はゲート構造を通常の畳み込みニューラルネットワーク構造に組み込んだモデルである．Gated CNN は言語モデルを構築する際に用いられることがあるが，提案手法では Gated CNN　を入力データから特徴ベクトルを生成する際に用いる．さらに，プーリング手法の一つである　Spatial Pyramid Pooling　[8] を組み込み，その有効性を実験により検証した．

2 章では Gated CNN による評判分析について説明する．3 章では Spatial Pyramid Pooling について述べる．4 章で評価実験について説明し，最後にまとめと今後の課題について述べる．

2. Gated CNN による評判分析

提案手法である Gated 畳み込みニューラルネットワーク (Gated convolutional neural network, Gated CNN)[7] を用いた評判分析について説明する．Gated CNN は文書中の連続した単語列をまとめて意味のベクトルを生成する．さらに，最大プーリングを文中の単語から生成された意味ベクトルに適用することで提案手法は文全体の意味ベクトル表現を得ることができる．また，提案手法は文中の単語列の長さを処理に適切な長さで自由に扱うことが可能である．

2.1 Gated 畳み込みニューラルネットワーク

本提案手法の重要部である Gated CNN について説明する [7]．図 1 では，Gated CNN は文中の単語 2 個ずつ選び，文中の単語すべてに処理が適用され，意味ベクトルが生成されて

連絡先: 岡田　真, 大阪府立大学, 599-8531 大阪府堺市中区学園町 1-1，072-252-1161(代表)，okada@cs.osakafu-u.ac.jp

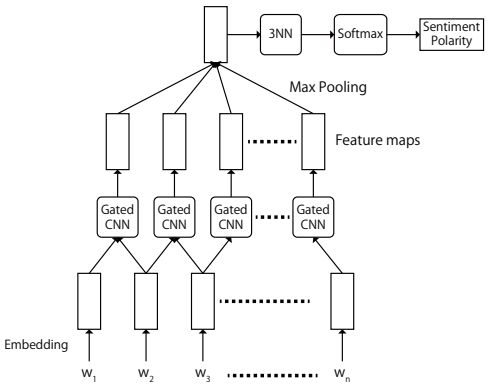


図 1: Gated CNN による評判分析の模式図

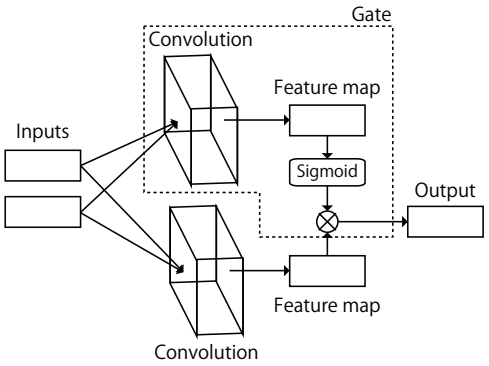


図 2: Gated CNN の模式図

いる．図 2 では，gated CNN 自体の模式図を示す．自然言語処理分野では LSTM を用いた言語モデルが用いられることが多い．LSTM は 3 個のゲートで情報の流れを制御して，適切な言語モデルを構築している．Gated CNN には，LSTM で用いられているようなゲート構造が従来の CNN に導入されている．

通常の CNN と同様に，Gated CNN においても入力として複数のベクトルからなる行列を受け取る．入力された行列は 2 個の異なるカーネル (テンソル) により，2 個の別々のベクトルに変換される．ベクトルの内，1 つは入力データの意味情報を表し，もう 1 つは Gated CNN 内部のコントロールゲートの情報を表す．この処理は次の数式で示される．

$$\mathbf{X}' = \sigma\left(\mathbf{X} * \mathbf{F}_g\right) \otimes \left(\mathbf{X} * \mathbf{F}_c\right) \tag{1}$$

ただし，$\mathbf{X} \in \mathbb{R}^{N \times m}$ は Gated CNN への入力，$\mathbf{F}_g \in \mathbb{R}^{k \times m \times n}$ と $\mathbf{F}_c \in \mathbb{R}^{k \times m \times n}$ は畳み込み処理のカーネル，$\sigma(\cdot)$ はシグモイド関数, $*$ は畳み込み処理，$\otimes$ はベクトル間の要素ごとの積を表す．したがって出力は $\mathbf{X}' \in \mathbb{R}^{k \times (N-n+1)}$ となる．出力長はデータでパディングがされない場合は入力 $\mathbf{X}$ よりも短くなる．ここでは，$\mathbf{F}_g$ はコントロールゲートの情報であり，$\mathbf{F}_c$ は入力データの畳み込み後の意味情報を表す．シグモイド関数の出力は 0 から 1 までの範囲なので，入力データの値にかかわらず出力もその範囲内に収まる．パラメータと $\mathbf{F}_g$ と $\mathbf{F}_c$ は訓練データにより学習される．

2.2 Gated CNN による評判分析

本来の Gated CNN では，文書コーパスから言語モデルを生成することを提案している．であるので，入力は単語列であり，出力もデータの系列となる．Gated CNN を用いた評判分析のためには，入力は単語列であり，それらの中の意味情報を基にして素性ベクトルが出力として生成されなければならない．出力された素性ベクトルは好評不評などの分類器への入力として扱われる．

まず，入力データを生成する．文中の単語を埋め込み行列 $\mathbf{D} \in \mathbb{R}^{m \times |V|}$ と参照して単語の埋め込みベクトルを得る．ここで $|V|$ はコーパス全体の出現単語数 (語彙数)，m は分散表現のベクトルの次元数を表す．したがって，$\mathbf{w}_i$ は 1-of-N コーディングベクトルとなり，埋め込み処理は以下の式で示される．

$$\mathbf{e}_i = \mathbf{D}\mathbf{w}_i \tag{2}$$

文は単語の系列なので，埋め込みベクトルの系列に変換される．本論文ではベクトルの系列を $\mathbf{X}$ で表す．したがって $\mathbf{X}$ は Gated CNN への入力となる．

$$(\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N) \Rightarrow (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_N) = \mathbf{X} \tag{3}$$

Gated CNN が文に適用されると，単語の情報は集約される．例えば 2 個の単語で考えると，$\mathbf{e}_1$ と $\mathbf{e}_2$ が Gated CNN で集約される．

$$\mathbf{m} \otimes \left(\mathbf{k}_1^{\mathrm{T}}\mathbf{e}_1 + \mathbf{k}_2^{\mathrm{T}}\mathbf{e}_2\right) \tag{4}$$

$\mathbf{m}$ は式 1 中の $\sigma\left(\mathbf{X} * \mathbf{F}_g\right)$ で計算される．$\mathbf{k}_1$ と $\mathbf{k}_2$ は重みであり，式 1 中の $\mathbf{F}_c$ から求められる．

入力 $\mathbf{X}$ を Gated CNN で処理したのち，ベクトル列は短くなっている．出力されたベクトル列から素性ベクトルが構築される．その素性ベクトルを用いて好評不評といった文書のポラリティを予測する．系列から素性ベクトルを生成する手法は数多く提案されている．本論文では単語出現順の最大要素を用いて素性ベクトルを生成する．

$$\mathbf{f} = \max_{1 \le j \le N-n+1} \mathbf{X}' \tag{5}$$

式 5 は要素の順序に従う．

$$f_i = \max(X'_{i1}, X'_{i2}, \cdots, X'_{i(N-n+1)}) \tag{6}$$

特徴ベクトルは 3 層のニューラルネットワークに入力され，その出力がポラリティ(好評不評など) となる．

$$\mathbf{h} = \sigma(\mathbf{W}_1\mathbf{f}) \tag{7}$$

$$\mathbf{o} = \mathrm{Softmax}\left(\sigma(\mathbf{W}_2\mathbf{h})\right) \tag{8}$$

提案手法の学習の際に，ロス関数を定義する．本論文ではクロスエントロピーを用いて，予想と正解のポラリティとの差異を求めた．$\mathbf{t}$.

$$\mathbf{t} = \begin{cases} (1,0) & (\text{negative}) \\ (0,1) & (\text{positive}) \end{cases} \tag{9}$$

$$E(\mathbf{o}, \mathbf{t}) = t_1 \log o_1 + t_2 \log o_2 \tag{10}$$

提案手法で用いるすべてのパラメータは以下の通りである．$\theta = (\mathbf{D}; \mathbf{F}_g; \mathbf{F}_c; \mathbf{W}_1, \mathbf{W}_2)$. システムは誤差逆伝搬法を使って学習する．

$$\theta \leftarrow \theta - \epsilon \frac{\partial E(\mathbf{o}.\mathbf{t})}{\partial \theta} \tag{11}$$

表 1: レビューデータセットの内容

レビュー情報	
好意的評価レビュー数	10,000
否定的評価レビュー数	10,000
訓練データ数	18,000
テストデータ数	2,000
平均単語数	59.1 words

3. Spatial Pyramid Pooling

Spatial Pyramid Pooling (SPP) は [8] プーリングの一種であり，CNN による画像処理分野で用いられている手法である *1．通常のものとの違いはプーリングの際に対象領域を分割し，階層的に処理することにある．SPP のパラメータにレベルがあり，レベルによって対象領域を分割する個数が変わる．レベル 1 では通常通り，レベル 2 では 2 個に分割，レベル n では $2^{'n-1)}$ 個に分割される．それぞれの領域でプーリング処理をする．そしてどのレベルの結果までを用いるかを事前に決めて，レベル 1 からそのレベルまでの出力結果を結合する．その結果を最終的な出力とする．画像処理では拡大縮小に対して頑健な処理が可能になるということで用いられることが多い．本論文では文に適用する．この場合，プーリングの対象範囲が文を先頭から $2^{(n-1)}$ 分割して，それぞれの領域から情報を取得する．これにより，文中のポラリティに関係する語がどのような位置にあってもその情報を適切に取得して，素性ベクトルを適切に作成できることを期待する．今回のシステムでは，図 1　の最上部 Max Pooling の部分を SPP に置き換えたシステムを提案する．

4. 実験

提案手法の有効性をデータセットを用いた実験により検証し，その結果に対して考察する．機械学習手法として一般的なサポートベクターマシン (SVM) による実験結果をベースラインとして用いる．

4.1 データセット

今回，データセットしては UCSD の　Julian McAuley [9, 10]　が配布している Amazon 商品レビューを用いた．このデータセットには Amazon の商品レビューとメタデータが含まれている．取得期間は 1996 年 5 月から 2014 年 7 月まで，レビュー総数は 1 億 4280 万レビューとなっている．各レビューには (評価値，レビュー文書，役に立ったかどうか (投票)) の情報と商品のメタデータ (商品説明文，商品カテゴリー，値段，メーカー名，画像) とリンクの情報が付与されている．

実験では，データセットから抽出された小データセットを実験用データとして用いた．これらは 24 個のカテゴリーに分けられており，今回は 5 個のカテゴリー (Electronics, Home and Kitchen, Sports and Outdoors, Health and Personal Care, Video Games) を実験用データを選択した．各小データセットから 2 万件のレビュー (好評 1 万件，不評 1 万件) をサンプリングした．5 段階評価の評価 5 がつけられているものから好評のデータを抽出し，評価 1 がつけられているものを不評のデータとして抽出した．それらの 10 %に当たるレビュー千件ずつ 2 千件をテストデータ，残りの 1 万 8 千件のレビューを訓練データとして用いた．各レビューの本文は千文字以下とした．表 **??** にレビューデータの情報を示す．

表 2: 提案手法のパラメータ

パラメータ名	値
単語の最低出現頻度	3
埋め込みのサイズ (m)	200
カーネルのサイズ ($\mathbf{F}_g$, $\mathbf{F}_c$)	$200 \times 200 \times 2$
ミニバッチのサイズ	500
分類器 (3 層 MLP)	200-100-2
最適化手法	ADAM

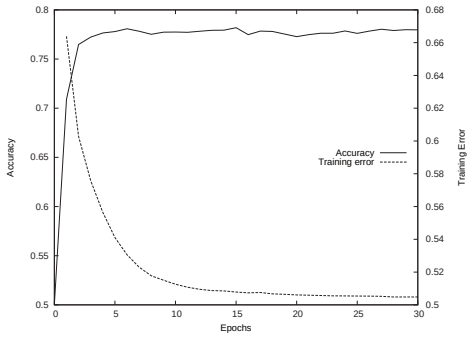


図 3: Video Game　レビューテストデータを用いた際の提案手法のロスと精度

4.2 実験用モデル

提案手法である Gated CNN による評判分析システムを Chainer*2　で構築した．表 2 に提案手法で用いたパラメータを示す．

単語の最低出現頻度は 3 とした．3 回以下の単語は未知語と考えられる．埋め込みのサイズは推定制度に強く影響する．今回は 200 次元とした．最適な次元数は今後の課題とする．カーネルのサイズは $200 \times 200 \times 2$ とした．今回は 2 単語間の関係のみを考える．2 単語を集約したのちの素性ベクトルは同じ次元数となる．提案手法では，最後の分類器は 3 層のニューラルネットワークとした．分類器も重要ではあるが，その比較調査は今後の課題とする．

4.3 実験結果

Amazon 商品レビューを用いた評判分析の結果を以下に示す．今回の実験では，文書データに前処理を適用せず，そのまま用いた．

図 3 に Video Game のデータセットを用いた際の訓練データのエラーとテストデータの精度を示す．提案手法の学習が進むにつれて，エラーが減り，精度が向上している．30 エポックまでに過学習の兆候は見られなかった．

サポートベクターマシン (SVM) の実験結果をベースラインとして，提案手法と比較した．言語は Python (version 3.5)，ライブラリは scikit-Learn SVM (LIBSVM) を用いた．パラメータは以下の通りである：カーネル=線形，C = 1.0．提案手法の Gated CNN (GCNN) によるシステムの実験結果と比較した．SPP のレベルは 2 とした．

実験結果を表 **??** に示す．

*1　http://image-net.org/challenges/LSVRC/2014/slides/sppnet_ilsvrc2014.pdf

*2　https://chainer.org

表 3: 実験結果 (F 値)(Amazon 商品レビューデータセット)

カテゴリー	SVM (Linear)	GCNN (no SPP)	GCNN (SPP)
Electronics	0.91	0.91	0.89
Home & Kitchen	0.91	0.92	0.87
Sports & Outdoors	0.90	0.91	0.91
Health & Personal Care	0.88	0.88	0.90
Video Game	0.87	0.88	0.87

提案手法の精度 (F 値) は，ベースラインと同等か 1 %高いものとなった．これにより前処理や調整なしでも提案手法が有効に働くことが示された．一般に，SVM で文書データを扱う際に，線形カーネルが有効であると言われる．事前の実験でも RBF カーネルよりも線形カーネルの方が有効であったため，今回我々は SVM に線形カーネルを用いた．SVM と提案手法はどちらも高い精度を示しており，その有効性は明らかといえる．提案手法の調整は今後の課題である．

また，今回は単語の集約単位を 2 個としている．これらも文書の長さに合わせて変更することで，精度が向上する可能性がある．それらの適切な値などについても今後の課題とする．

SPP を用いた場合，精度が向上したものと低下したものが見受けられた．SPP を用いることにより，文はいくつかの領域に分けられてプーリングされる．その際に好評不評を評価する単語が含まれる個所が適切に評価される場合とされない場合で制 d の差が出たのではないかと考えられる．適切なレベルの値についても今後の課題である．

5. まとめと今後の課題

本論文では，ゲート機構付き畳み込みニューラルネットワーク (Gated convolutional neural network, Gated CNN) による評判分析システムを提案した．Gated CNN を通常言語処理で系列データの生成によく用いられる．提案手法では，その系列データを用いて素性ベクトルを生成し，それらを訓練データとして分類器を生成した．Amazon review dataset を用いた実験により，提案手法の有効性を確認した．サポートベクターマシン (SVM) をベースラインとして比較し，前処理なしに同等の性能を得ることができた．

今後の課題として，提案手法の構造の改良がある．例えば埋め込みベクトルやカーネルの大きさ，単語の出現頻度の最低値，複数の分類器による性能の比較などである．カーネルの大きさは一度に集約する単語数に関連するため，重要な課題の一つである．非線形なデータに対応可能な機構の検討も重要な課題の一つである．また，SPP の適切なレベルの値についても今後の課題の一つである．

参考文献

[1] V. Hatzivassiloglou and K. R. McKeown, *Predicting the Semantic Orientation of Adjectives*, Proc. of the 35 Annual Meeting of the Association for Computational Linguistics and Eight Conference of the European Chapter of the Association for Computational Linguistics, pp.174-181, 1997.

[2] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, *Semi-supervised recursive autoencoders for predicting sentiment distribution*, Proc. of the Conference on Empirical Methods in Natural Language Processing, pp. 151-161, 2011.

[3] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, *Semantic compositionally through recursive matrix-vector spaces*, Proc. of the Conference on Empirical Methods in Natural Language Processing, pp.1201-1211, 2012.

[4] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, *Recursive deep models for semantic compositionality over a sentiment treebank*, Proc. of the Conference on Empirical Methods in Natural Language Processing, pp.1631-1642, 2013.

[5] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, Neural Computation, Vol.9, No.8, pp.1735-1780, 1997.

[6] Kim Y.: Convolutional neural networks for sentence classification, Proc. of the Conference on Emprical Methods in Natural Language Processing, pp.1746-1751 (2014).

[7] Dauphin Y. N., Fan A., Auli M., and Grangier D.: Language Modeling with Gated Convolutional Networks, arXiv:1612:08083 (2016).

[8] He K., Zhang X., Ren S. and Sun J.: Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, journal arXiv:1406:4729 (2014).

[9] R. He and J. McAuley, *Ups and downs: Modeling the visual evolution of fashion trends with oneclass collaborative filtering*, Proc. of International Conference on World Wide Web (WWW'16), pp.507-517, 2016.

[10] J. McAuley, C. Targett, J. Shi, and A. van den Hengel, *Image-based recommendations on styles and substitutes*, Proc. of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.43-52, 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

リカーシブニューラルテンソルネットワークの日本語文のセンチメント分析への適用

An Application of Recursive Neural Tensor Network to Sentiment Analysis of Japanese Sentences

赤井 龍一[*1] Ryuichi Akai
渥美 雅保[*1] Masayasu Atsumi

[*1] 創価大学理工学部情報システム工学科
Information Systems Science, Faculty of Science and Engineering, Soka University

Recursive Neural Tensor Network (RNTN) is a neural network model that recursively computes the synthetic distributed vector representation for phrases of various lengths and syntax types from the distributed vector representation of words along the syntax tree. Distributed vector representation is used as a feature to classify each phrase and it is used to classify sentiment of phrases in sentiment analysis. In this paper, we apply the RNTN to sentiment analysis of Japanese sentences. For this purpose, based on the Stanford Sentiment Treebank corpus for sentiment analysis, we first create a corpus of Japanese sentences with teacher labels only for words and sentences. Then we evaluate the accuracy of sentiment analysis on Japanese sentences when we learn from only teacher labels for words and sentences. We also consider the effect of attaching teacher labels of phrases by heuristic rules.

1. はじめに

リカーシブニューラルテンソルネットワーク RNTN(Recursive Neural Tensor Network)[Socher 13]は様々な長さと構文タイプの句に対する合成分散ベクトル表現を単語の分散ベクトル表現から構文木に沿ってリカーシブに計算するニューラルネットワークモデルで，合成分散ベクトル表現は各句を分類する特徴として用いられる．センチメント分析では，これら単語・句・文の分散ベクトル表現を用いてそれらをポジティブ・ネガティブの 2 クラス，もしくはより多くのクラスにソフトマックスクラシファイアにより分類する．本論ではこの RNTN を日本語文のセンチメント分析に適用する．そのために，センチメント分析のための英語文のコーパス Stanford Sentiment Treebank[注 1]をもとに，まず単語と文のみに教師ラベルがついた日本語文のコーパスを作成する．そして，単語と文に対する教師ラベルのみから学習を行うようにした場合の日本語文に対するセンチメント分析の精度を評価する．また，句の教師ラベルをヒューリスティックな規則によりつけることの効果も検討する．

2. RNTN によるセンチメント分析

2.1 概要

リカーシブニューラルテンソルネットワーク RNTN を用いたセンチメント分析では，分散表現された単語の意味を構文木に沿って合成することを通じて句と文のセンチメントを構成的に分析する．具体的には，構文木の全てのノードに対してボトムアップにテンソルベースの関数を適用させることにより単語，句，文と辿ってセンチメント分析を行っていく．本論では，センチメントを，「強くポジティブ」，「ポジティブ」，「ニュートラル」，「ネガティブ」，「強くネガティブ」の 5 センチメントクラスに分類する場合を扱う．RNTN の学習では，すべての単語・句・文に対するセンチメントラベルが教師ラベルとして用いられる．これにより,文中に現れる否定語によるセンチメントの反転に関する効果が学習されて従来法に比べて高い精度が実現されているが,ラベル付けのコスト，特にすべての句に対するラベル付けのコストが高い．日本語文のセンチメント分析器の学習では，Stanford Sentiment Treebank の英文を日本語文に翻訳し，それら文から形態素解析と構文解析により構文木を生成して，それらの単語と文に元の英文の対応するセンチメントラベルを与える．また，句に関しては，教師ラベルとしてのセンチメントラベルは与えずに，簡易な規則により句のラベルを部分的に生成して与える．

2.2 リカーシブニューラルテンソルネットワーク RNTN

RNTN は，単語埋め込み層，ニューラルテンソル層，センチメント分類層からなる．単語埋め込み層は，単語の one-hot ベクトルw に対して，単語埋め込み行列W_Eを用いて(1)式によりその分散ベクトル表現v_Wを計算する．

$$v_W = embedID(w)$$

ここで，v_Wはd-次元ベクトルとする．ニューラルテンソル層は，二分構文木の左と右の子ノードの分散表現をそれぞれv_l,v_rとするとき，(2)式によりそれらの合成ベクトルv_pを句のベクトル表現として計算する．

$$v_p = tanh\left(\begin{bmatrix} v_l \\ v_r \end{bmatrix}^T W_B^{[1:d]} \begin{bmatrix} v_l \\ v_r \end{bmatrix} + W_L \begin{bmatrix} v_l \\ v_r \end{bmatrix}\right)$$

ここで，v_lとv_rはd-次元単語ベクトルまたは d-次元合成句ベクトル，$W_B^{[1:d]}$は$R^{2d\times 2d\times d}$の双線形テンソルである．これら合成ベクトルは単語から句，そして文へとリカーシブに計算される．センチメント分類層は，構文木の各ノードの出力，即ち単語ベクトル，合成句ベクトル，合成文ベクトルvに対して，(3)式によりセンチメントクラス次元のセンチメントラベルの確率分布を出力する．

$$p_v = softmax(W_o v)$$

連絡先：赤井 龍一，創価大学理工学部情報システム工学科，
〒192-8577 東京都八王子市丹木町 1-236,Tel: 042-691-2211
E-mail: m10777.t@outlook.jp

注 1: https://nlp.stanford.edu/sentiment/treebank.html

学習におけるロスは，単語・句・文それぞれに対して与えられる教師ラベルtとソフトマックスクラシファイアの出力確率分布pとの間のクロスエントロピーの総和として(4)式により計算される．

$$L(\theta) = \sum_v L_v(\theta) = -\sum_v \sum_i t_{v,i} \log p_{v,i} + \lambda \|\theta\|^2$$

ここで，p_vは単語・句・文の分散ベクトルvに対する出力確率分布，t_vは one-hot ベクトルで与えられるその教師ラベル，$\theta = (W_E, W_B, W_L, W_O)$であり，θはロスの誤差逆伝播で最適化される．

2.3 単語教師ラベルからの句教師ラベルの設定

RNTN の高い精度はすべての句に対する教師ラベル付けによるところが大きいが，このラベル付けには膨大なコストがかかる．そこで，二分構文木の左と右の子ノードの教師ラベルから親ノードの教師ラベルを導く簡易な規則を導入する．いま，5 つのセンチメントクラス「強くポジティブ」，「ポジティブ」，「ニュートラル」，「ネガティブ」，「強くネガティブ」をそれぞれ 4,3,2,1,0 で表し，左と右の子ノードのラベルをそれぞれl_{LC}とl_{RC}，親ノードのラベルをl_Pとする．このとき，子ノードの教師ラベルから親ノードの教師ラベルを導く規則として，

RULE1: *if* l_{LC}==2 *then* $l_P = l_{RC}$
RULE2: *if* l_{RC}==2 *then* $l_p = l_{LC}$
RULE3: *if* l_{LC}>2 *and* l_{RC}>2 *then* l_P= max(l_{LC}, l_{RC})

を導入する．これにより，単語の教師ラベルを与えることで部分的に句の教師ラベルをボトムアップに設定することができる．このとき，学習におけるロスは，単語，文と教師ラベルが設定された句の集合V^*に対するソフトマックスクロスエントロピーの和$\sum_v L_{v \in V^*}(\theta)$として計算される．

3. 日本語コーパスの作成

Stanford Sentiment Treebank の英文コーパスから日本語文のコーパスを作成する処理の流れを図 1 に示す．まず，Stanford Sentiment Treebank の英文コーパスから，文センチメントラベルが付いた英文セットとセンチメントラベルが 2 以外の英単語の辞書を作る．次に，それらを機械翻訳により文センチメントラベル付き日本語文セットと単語センチメントラベル付き英和辞書に翻訳する．そして，それらに対して，意味の通らない文の修正・削除，英和辞書の修正・追加などを手作業で行う．最後に，それら日本語文を形態素解析・構文解析して構文木を生成し，その単語にセンチメントラベルを割り振る．日本語文の形態素解析と構文解析には Jigg[Noji 16]の Kuromoji と Jaccg を利用した．

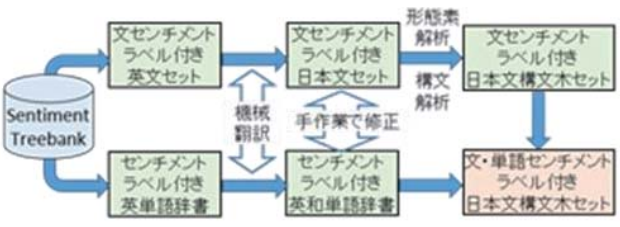


図 1　日本語文コーパス作成の流れ

このコーパスから文の長さを基準に 30 単語で長文・短文にわけ，「長文コーパス」，「短文コーパス」，「長文＋短文コーパス」の 3 つのコーパスを作成した．図 2 と 3 に長文コーパスと短文コーパスの例文，表 1 にコーパスの文の数を示す．

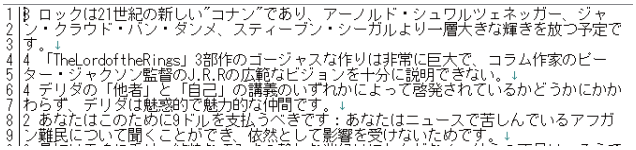


図 2　長文コーパスの例文

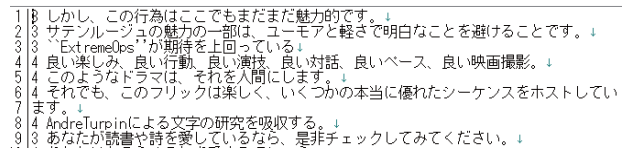


図 3　短文コーパスの例文

表 1　日本語文コーパス(文数)

	長文	短文	長文+短文
訓練用	1988	2211	4197
テスト用	102	100	100

また，「長文＋短文コーパス」から，「しかし」または「ない」を含む文だけを抽出して「否定文コーパス」を作成した．表 2 に否定文コーパスの文数を示す．

表 2　否定文コーパス(文数)

否定語を含む文	619
否定語を含まない文	3580

4. 実験

4.1 概要

RNTN を用いた日本語文のセンチメント分析の実験的評価を行う．実験には，Stanford Sentiment Treebank の英文コーパスを利用して作成した日本語文の 3 種類のコーパス「長文コーパス」，「短文コーパス」，「長文＋短文コーパス」，及び「否定文コーパス」を用いる．否定文コーパスは，否定語の有無がセンチメント分析の精度に対してどの程度影響を及ぼしているかを分析するために用いる．また，比較のために「英文コーパス」を用いた実験も行う．句に対する教師ラベルを与えないで学習した場合と，句に対する教師ラベルを「句教師ラベル設定規則」により設定して学習した場合の精度を評価する．

4.2 実験結果

(1) 文の長さ

日本語の「長文コーパス」，「短文コーパス」，「長文＋短文コーパス」を用いたセンチメント分析実験により，文の長さの精度に及ぼす影響を評価した．表 3 に句ノードに対する教師ラベルなしの場合と「句教師ラベル設定規則」により教師ラベルを設定して学習した場合の精度を示す．長文コーパスのほうが短文コーパスよりも高い精度が得られる傾向がみられた．また，句教師ラベル設定規則の適用によりノード全体に対しては精度の向上

がみられたが，そのボトムアップな適用の影響が文まで適切に及んでいないことが確かめられた．

表 3 日本語文コーパス(文長別)の結果

	5 分類		
	全ノード	文	単語
長文コーパス	68.7	67.6	97.5
短文コーパス	68.8	56.0	94.6
長文＋短文コーパス	79.4	56.0	97.5

(a)句ノード：教師ラベルなし

	5 分類		
	全て	文	単語
長文	86.1	56.8	97.2
短文	82.7	54.0	95.2
長+短	83.1	50.0	96.4

(b)句ノード：教師ラベル設定規則でラベル付け

(2) 否定語の有無

日本語の「否定文コーパス」を用いたセンチメント分析実験により，「しかし」，「ない」といった否定語がセンチメント分析の精度に及ぼす影響を評価した．表 4 に句ノードに対する教師ラベルなしの場合と「句教師ラベル設定規則」により教師ラベルを設定して学習した場合の精度を示す． 否定語を含む文のほうが文に対する精度は下がるが，全ノードに対する精度はむしろ上がる傾向がみられた．また，この傾向は句教師ラベル設定規則の適用においてもみられた．句教師ラベル設定規則はセンチメントの反転を扱っていないが，句ノードへのラベル付けが何らかの効果をもたらしていることが示唆される．

表 4 日本語文コーパス(否定語の有無)の結果

	5 分類		
	全ノード	文	単語
否定語あり	83.0	46.0	95.0
否定語なし	81.7	54.0	96.3

(a)句ノード：教師ラベルなし

	5 分類		
	全ノード	文	単語
否定語あり	86.7	45.0	96.3
否定語なし	84.2	53.0	96.4

(b)句ノード:教師ラベル設定規則でラベル付け

(3) 英文の結果

Stanford Sentiment Treebank の英文コーパスに対するセンチメント分析の結果を比較のために表 5 に示す．

日本語文に対するセンチメント分析の精度が全般的に英文に対する精度より高くなっている．これには，日本語文のコーパスに対して手作業で意味の取りにくい文を整形したり削除したりした効果も考えられるが興味深い傾向である．また，英文コーパスに対しては句教師設定規則が機能して精度向上につながっている点も興味深い．

英文コーパスにおいてすべての句ノードに教師ラベルを与えた場合の全ノードに対する 5 分類精度は約 80%，文ノードに対する 5 分類精度は約 40%であった．これより，句へのラベル付けの必要性が再確認された．

表 5 Stanford treebank の英文コーパスの実験結果

5 分類		
全ノード	文	単語
71.9	28.5	96.8

(a)句ノード:教師なし

5 分類		
全ノード	文	単語
73.3	34.9	97.4

(b)句ノード: 句教師ラベル設定規則でラベル付け

(4) 考察

以上の実験結果をまとめると次のとおりである．

(a) 長文コーパスのほうが短文コーパスよりも高い精度が得られる傾向がみられた，
(b) 否定語を含む文のほうが文に対する精度は下がるが，全ノードに対する精度はむしろ上がる傾向がみられた，
(c) 日本語文に対するセンチメント分析の精度が全般的に英文に対する精度より高い傾向がみられた，
(d) ボトムアップな句教師ラベル設定規則の適用により全ノードに対する精度の向上はみられたが，文の精度の向上にはつながらなかった．一方，英文に対してはうまく機能した．

ボトムアップなセンチメント分析において，句ノードに対する適切なセンチメントの推定は，特に否定語が含まれる文における否定語によるセンチメントの反転を文のセンチメントに伝えるうえで重要な役割を果たす．今回の句教師ラベル設定規則にはこの反転を文に伝える機能がないため今後改良が必要である．

5. むすび

本論文では，リカーシブニューラルテンソルネットワーク RNTN の日本語文のセンチメント分析への適用について述べた．今後の課題として，句に対するラベル付けを自動的に行うより良い方法を開発していきたい．また，Stanford Sentiment Treebank の日本語訳の量と質を改善していくとともに，日本語文の別のセンチメント分析用のコーパスを作成して実験をしていきたい．

参考文献

[Socher 13] Socher,R., et al.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, EMNLP 2013.

[Noji 16] Noji,H. and Miyao,Y.: Jigg: A Framework for an Easy Natural Language Processing Pipeline, Proc. of the 54th Annual Meeting of the ACL, pp.103-108, 2016.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Word2Vecにおける出力側の重みに注目した文書分類手法の検討

A study on document classification focusing on the output side weight on Word2Vec

内田脩斗 *1 Shuto Uchida　吉川大弘 *1 Tomohiro Yoshikawa　ジメネスフェリックス *1 Felix Jimenez　古橋武 *1 Takeshi Furuhashi

*1名古屋大学大学院工学研究科
Graduate School of Engineering Nagoya University

Document classification is an important technology in modern information society. In recent years, distributed representation (DR) which embeds semantic relationships of words into vectors has attracted attention and the methods applying DR to document classification have been reported. DR can be generated mainly by using a tool called Word2Vec. Word2Vec has the learning structure using a neural network, and we use the weights on the input side as DR. However, Word2Vec learns different characteristic weights on the output side from DR, which is not focused on and not commonly used. In this paper, we propose a document classification method by ensemble learning using DR and the output side weights and suggest the usefulness on the proposed method.

1. はじめに

インターネットの普及に伴って大量の電子文書が日々生成されている現代において，テキストデータの自動解析手法や解析技術は，様々な場面での応用が期待されている．その技術の一つに文書分類がある．文書分類とは，与えられた文書を予め定められたクラスのいずれかに分類することである．スパムメール分類や Web 記事分類などに広く実用化されており，より高精度な分類手法が求められている．

文書分類では各文書をどのような素性で表すかが重要であり，各単語は互いに独立であるという仮定の下，単語の出現回数を利用する Bag-of-Words（BOW）という表現手法が古くから広く利用されている [1][2][3].

一方で，Micolov らが発表した Word2Vec[4] は，大規模コーパスから教師なし学習を行うことで，語義の似た単語が類似した重みの要素をもつ分散表現を生成することができる．これにより，従来では困難であった単語ベクトル間での意味の演算が可能となり，学習された分散表現に対し，

$$vector(Paris) - vector(France) + vector(Italy)$$

により算出されるベクトルが $vector(Rome)$ に，

$$vector(king) - vector(man) + vector(woman)$$

により算出されるベクトルが $vector(queen)$ に，それぞれ近くなるという性質をもっている．このように分散表現は，学習によって得られた単語同士の意味的な関係を捉えることができ，文書分類に適用する手法も多く報告されている [5][6].

この Word2Vec は，ニューラルネットワークを利用した学習構造をしており，ネットワーク上の入力層-隠れ層間の重みである上述の分散表現と，それと共に生成される隠れ層-出力層間の重み，“出力側の重み”があるが，一般的に利用されることはない．しかし，この出力側の重みは単語に対して分散表現とは異なる意味関係を捉えていると考えられ，その有用性があると思われる．

連絡先: 内田脩斗，名古屋大学大学院工学研究科，名古屋市千種区不老町，052-789-2793，052-789-3166，uchida@cmplx.cse.nagoya-u.ac.jp

そこで本稿では，分散表現と出力側の重みの捉えている意味関係の違いについて検討し，分散表現と出力側の重みをアンサンブル学習により組み合わせた文書分類手法を提案する．

2. Word2Vec

本章では．Word2Vec の学習モデルについて説明する．

Word2Vec は，言語処理でよく用いられる分布仮説（同じ文脈で出現する単語は同じ意味を持つこと）[7] に基づいており，文脈上のある単語に対して，周辺に現れやすい単語を予測することをモデル化した構造をしている．

図 1 は Word2Vec の学習モデルを表した図である．また，入力層と出力層の次元数はボキャブラリ数，隠れ層は埋め込み次元数である．Word2Vec はニューラルネットワークを用いて

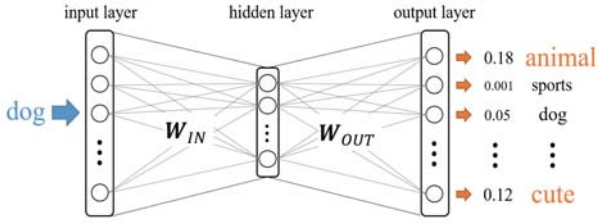


図 1: Word2Vec 学習モデル図

分散表現を獲得する．具体的には，入力層に入力された単語に対して，その単語の周辺に出現しやすい単語の出現確率が大きくなるように各層の重み（入力側重み W_{IN}，出力側重み W_{OUT}）を更新する．例えば，ある文脈上において,「dog」という単語の周辺に「animal」や「cute」という単語が出現したとすると，入力層の「dog」に対応した要素に 1，それ以外の要素に 0 が入力され,「animal」,「cute」に対応した出力層の出力が 1 に近づくように W_{IN} と W_{OUT} が更新される．

W_{IN} は周辺単語の情報を元に，単語の意味関係を学習している．つまり,「dog」や「cat」という単語の周辺に「animal」や「cute」という単語が共通して出現することで類似した重みを学習し，

$$vector(dog) \simeq vector(cat) \simeq vector(rabbit)$$

に近づくように，個々の次元に単語の意味関係を埋め込んでいる．また，W_{IN} の各行ベクトルが個々の単語ベクトルに対応しており，Word2Vec では W_{IN} を分散表現として利用することを前提とした学習モデルとなっている．

一方，W_{OUT} はある単語の周辺に出現する単語を予測するための重みである．つまり，「dog」や「cat」という単語の周辺に「animal」や「cute」という単語が現れる確率を大きくするため，入力単語のベクトルと周辺単語のベクトルの内積値が大きくなるように重みを学習する．これより，W_{OUT} は単語ベクトルを共起しやすい単語に展開する共起単語ベクトルであると捉えることができる．つまり，W_{OUT} においては，

$$vector(animal) \simeq vector(cute) \simeq vector(bark)$$

のような共起関係を学習する傾向にある．また，W_{OUT} の各列ベクトルが個々の共起単語ベクトルに対応している．ただし，W_{OUT} は W_{IN} を獲得するために生成される副産物と考えられており，一般的に W_{OUT} を単語ベクトルとして利用することはない．W_{OUT} を利用した従来研究は数少ないが，Mitra ら [8] は，W_{IN} と W_{OUT} の双方から得られる共起情報を利用した文書検索手法を提案している．しかし，W_{OUT} を単語ベクトルとして，独立して用いる本研究とは異なっている．

上述したように，W_{IN} と W_{OUT} では学習される重みの性質が異なっており，通常利用されない W_{OUT} にも有用性があると考えられる．

3. 従来手法

Word2Vec により学習された分散表現 W_{IN} を，文書分類に適用する一般的な方法 [5] を説明する．ある文書 d_i があったとき，その文書のベクトル表現である文書ベクトル $\vec{d_i}$ は，以下の式で算出される．

$$\vec{d_i} = \frac{1}{|d_i|} \sum_{j=1}^{|d_i|} \vec{w_j} \quad (1)$$

$\vec{w_j}$ は文書 d_i における j 番目の単語の単語ベクトルである．取得する単語ベクトルが分散表現に含まれていない場合，その単語は（そもそも文書内に含まれていなかったものとして）除外する．これは，分散表現を学習したコーパスに存在しない単語が文書 d_i に含まれていた場合に生じる．$|d_i|$ は，文書内の単語のうち，単語ベクトルを取得できた単語の数である．

(1) 式は，文書内に含まれる単語の単語ベクトルの重心（平均）をとったものを文書ベクトルとして扱うことを意味する．この文書ベクトルを用いて分類を行う．

従来手法では，W_{IN} のみを文書分類に適用する．一方，W_{OUT} は W_{IN} と異なる情報を保持しており，これを利用することで更なる分類精度の向上に寄与できると考えられる．そこで本稿では，W_{IN} と W_{OUT} を組み合わせた分類手法を提案する．

4. 提案手法

4.1 アンサンブル学習

提案手法では，[9] で利用されているクラス所属確率を用いたアンサンブル学習を行う．アンサンブル学習とは，複数の分類器を組み合わせることで分類精度を向上させる手法である．また，クラス所属確率とは，分類対象となる文書がそれぞれのクラスに所属する確率である．

4.2 概要

提案手法のイメージを図 2 に示す．

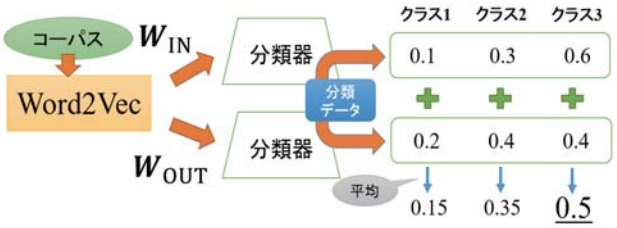


図 2: 提案手法のイメージ

1. テキストコーパスから Word2Vec を用いて，W_{IN} と W_{OUT} を生成する．(W_{IN} と W_{OUT} は同時に生成される)
2. 個々の素性を用い，文書ベクトルで表した学習文書に対してそれぞれ分類器を生成する．
3. 分類データに対して個々の分類器によりクラス所属確率を算出する．
4. クラスごとの所属確率の平均値を算出し，最大のクラスに文書を分類する．

5. 実験設定

5.1 分類対象データセット

実験で使用した分類対象データセットを表 1 に示す．livedoor ニュースコーパス（以下、「livedoor」という）は日本語のテキストデータとなっており，ウェブサイト *1 からダウンロードして利用できる．また，Reuters 21578 は英語のテキストデータとなっており，[10] の著者のウェブサイト *2 からダウンロードして利用できる．

表 1: 分類対象データセット

データセット	学習文書数	テスト文書数	クラス数
livedoor	4420	2947	9
Reuters 21578	5485	2189	8

5.2 Word2Vec 学習コーパス

Word2Vec の学習には日本語 Wikipedia と英語 Wikipedia を利用した．また，各種パラメータは，$window = 5$，$size = 300$ を使用した．「$window$」は前後何単語を教師データとするかを指定するオプション，「$size$」は学習する単語ベクトルの次元数を指定するオプションである．Word2Vec の実装は Python の gensim ライブラリを使用し，また，W_{OUT} はライブラリ内の syn1neg に保存されているものを使用した．

5.3 分類器

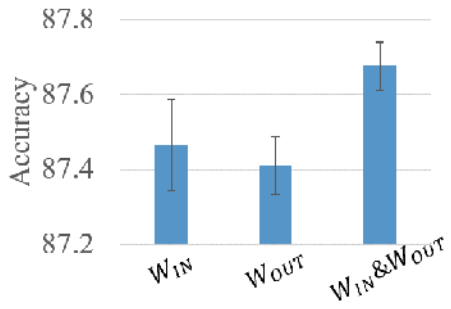
分類器には SVM を用いた．SVM は RBF カーネルを用い，ハイパーパラメータはライブラリのデフォルト値である $C = 1$ と $\gamma = 1/$次元数 (300) を用いた．

*1 https://www.rondhuit.com/download.html
*2 http://web.ist.utl.pt/acardoso/datasets/

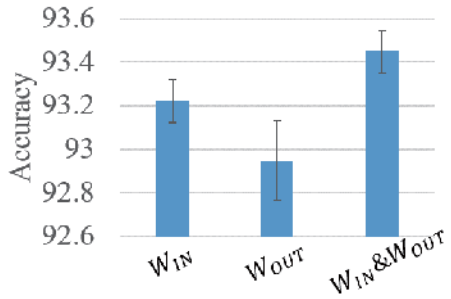
6. 実験

6.1 実験 1（提案手法の有効性の検証）

実験 1 では，従来手法に対する提案手法の有効性の検証を行った．図 3(a) は日本語 Wikipedia から学習した Word2Vec における W_{IN} と W_{OUT} を用いて，livedoor を分類した際の正答率，図 3(b) は英語 Wikipedia から学習した Word2Vec における W_{IN} と W_{OUT} を用いて Reuters 21578 を分類した際の正答率である．また，Word2Vec は内部にランダム性を保持しているため，ここでは 5 試行の平均と標準誤差を表示している．



(a) livedoor ニュースコーパス



(b) Reuters 21578

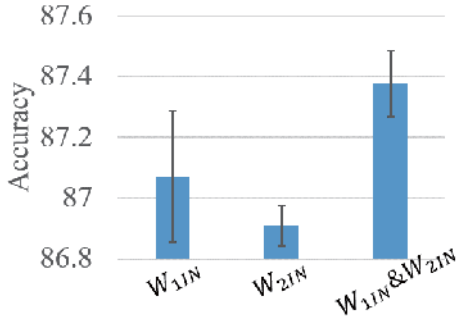
図 3: 各データセットにおける正答率

正答率は，テスト文書に対して，正しく分類された割合である．また，図の W_{IN} は，分類器の素性として W_{IN} を用いた場合の正答率（従来手法），W_{OUT} は，分類器の素性として W_{OUT} を用いた場合の正答率（参考），$W_{IN}\&W_{OUT}$ はアンサンブル学習を用いた場合の正答率（提案手法）を示している．

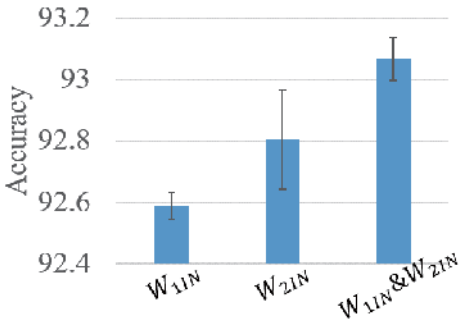
これらの結果より，W_{OUT} 単体では従来手法の W_{IN} に劣っているが，W_{IN} と W_{OUT} を組み合わせた提案手法では精度の向上が確認でき，提案手法が有効である可能性を確認した．これは，W_{IN} と W_{OUT} が異なる特徴を捉えていることで，個々の分類器に強みがあり，アンサンブル学習が有効に働いているものと考えられる．

6.2 実験 2（他手法との比較）

実験 2 では，同一のコーパスから異なる分散表現を生成する手法を用いて，アンサンブル学習を行った場合との比較を行った．図 4(a) に livedoor を分類した際の正答率（W_{1IN} と W_{2IN}），図 4(b) に Reuters 21578 を分類した際の正答率（W_{1IN} と W_{2IN}）を示す．実験 1 と同様に，5 試行での平均と標準誤差を表示している．



(a) livedoor ニュースコーパス



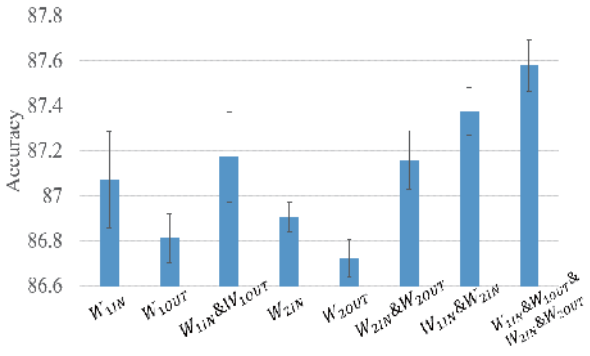
(b) Reuters 21578

図 4: 各データセットにおける正答率（W_{1IN} と W_{2IN}）

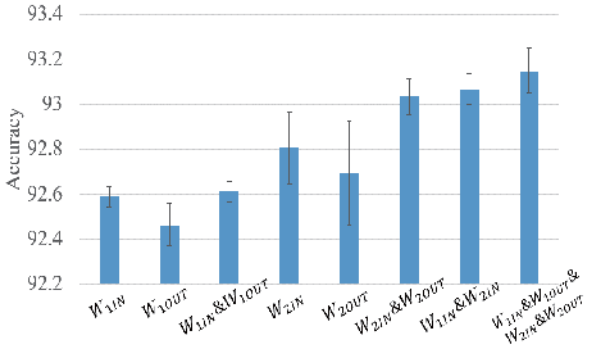
同一のコーパスから異なる分散表現を獲得するため，Wikipedia コーパスを 2 分割し，それぞれに対して Word2Vec により学習した 2 種類の W_{IN} である W_{1IN} と W_{2IN} を生成した．コーパスを分割することで出現単語や文書情報に偏りが生じるため，生成される W_{1IN} と W_{2IN} の特徴の捉え方に差が生じ，双方を用いてアンサンブル学習を行った場合にも，分類精度の向上が期待できる．実際に図 4(a) と図 4(b) より，精度の向上が見て取れるが，実験 1 の従来手法 W_{IN} よりも精度が劣っていることが確認できる．これは，コーパスを分割したことで Word2Vec の学習に使用するデータ量が減少したことによる影響であると考えられる．よって，コーパスのサイズを維持しつつ，異なる意味関係を捉える W_{IN} と W_{OUT} を使用する有効性が確認された．

さらに，提案手法を実験 2 に適用した場合の結果を図 5(a) と図 5(b) に示す．

これより，W_1 と W_2 のそれぞれで，実験 1 と同様の傾向が確認でき，提案手法の有効性が確認された．これらの結果から，複数のコーパスを Word2Vec の学習に利用し，提案手法と組み合わせることで，更なる精度の向上に繋がることが期待できる．W_{OUT} は，W_{IN} を生成するときに同時に生成される



(a) livedoor ニュースコーパス



(b) Reuters 21578

図 5: 各データセットにおける正答率（W_1 と W_2）

ため，分散表現を用いた文書分類手法には常に適用することが可能と考えられ，一般的に利用されてこなかった W_{OUT} の有用性が示唆された．

7. まとめ

本稿では，Word2Vec の学習過程で生成される出力側の重み W_{OUT} に注目し，入力側の重み W_{IN} と出力側の重み W_{OUT} を利用したアンサンブル学習による文書分類手法を提案した．実験の結果，提案手法の有効性を確認した．また，一般的には用いられていない出力側の重み W_{OUT} の新たな有用性を示した．

今後は，他のデータセットに対する有効性の検証や，W_{OUT} のより有効的な活用方法について検討することが挙げられる．

参考文献

[1] Pedro Domingos and Michael J. Pazzani. ”On the optimality of the simple Bayesian classifier under zero-one loss.” Machine Learning, 29(2-3):103130, 1997.

[2] Rennie J., Shih, L., Teevan, J., Karger, D. ”Tackling the Poor Assumptions of Naive Bayes Text Classifiers.” Proc. of ICLM-2003.

[3] Rie Johnson and Tong Zhang. ”Effective use of word order for text categorization with convolutional neural networks.” arXiv preprint arXiv:1412.1058, 2014.

[4] Mikolov, Tomas, et al. ”Efficient estimation of word representations in vector space.” arXiv preprint arXiv:1301.3781 (2013).

[5] Xing, Chao, et al. ”Document classification with distributions of word vectors.” Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA). IEEE, 2014.

[6] R. Liu, D. Wang, and C. Xing, ”Document classification based on word vectors.” in ISCSLP ’ 14, 2014.

[7] Hinton, G.E., McClelland, J.L., Rumelhart, D.E. ”Distributed representations.” In: McClelland, J.L., Rumelhart, D.E. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, Cambridge, MA, pp. 77109, 1986.

[8] Bhaskar Mitra, Eric Nalisnick, Nick Craswell, Rich Caruana. ”A dual embedding space model for document ranking.” arXiv preprint arXiv:1602.01137. 2016.

[9] 高橋和子. ”多クラス SVM におけるクラス所属確率を用いたアンサンブル学習の提案.” 研究報告音声言語情報処理 (SLP) 2011.2 (2011): 1-8.

[10] Cardoso-Cachopo, A. ”Improving Methods for Single-label Text Categorization.” PdD Thesis, Instituto Superior Tecnico, Universidade Tecnica de Lisboa, 2007.

[2C4] [General Session] 12. HI / Education Aid

座長:稲葉 通将（広島市立大学）

Wed. Jun 6, 2018 5:20 PM - 6:40 PM Room C (4F Orchid)

[2C4-01] Performance Evaluation of Automatic Gesture Generation System Using Bi-Directional LSTM on Humanoid Robot

○Kodai Hiyori[1], Kenji Araki[1], Dai Hasegawa[2], Satoshi Yoshio[1] （1. Graduate School of Information Science and Technology, Hokkaido University, 2. Tokyo University of Technology (Currently: Hokkai Gakuen University)）

5:20 PM - 5:40 PM

[2C4-02] Contextual Dependent Gaze Aversion in a Communication Robot

○Yukiko Kawaguchi[1], Fumihide Tanaka[2] （1. College of Engineering Systems, School of Science and Engineering, University of Tsukuba, 2. Faculty of Engineering, Information and Systems, University of Tsukuba）

5:40 PM - 6:00 PM

[2C4-03] Dialogue Selection Algorithm Based on the Amount of User’ s Emotional Expression in a Circulate-Learning Robot Connecting Elderly People

○Tsubasa Kasai[1], Fumihide Tanaka[1] （1. University of Tsukuba）

6:00 PM - 6:20 PM

[2C4-04] Input Interface for Robot System Using Mixed Reality that Encourages Retrospect of Driving

○Yuya Aikawa[1], Masayoshi Kanoh[1], Felix Jimenez[2], Mitsuhiro Hayase[3], Takahiro Tanaka[2], Hitoshi Kanamori[2] （1. Chukyo University, 2. Nagoya University, 3. Toyohashi Sozo University）

6:20 PM - 6:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

講義代行ロボットにおける双方向 LSTM を用いたジェスチャ自動生成システムの性能評価

Performance Evaluation of Automatic Gesture Generation System Using Bi-Directional LSTM on Humanoid Robot

日和 航大[*1] 荒木 健治[*1] 長谷川 大[*2] 芳尾 哲[*1]
Kodai Hiyori Kenji Araki Dai Hasegawa Satoshi Yoshio

[*1] 北海道大学大学院情報科学研究科 Graduate School of Information Science and Technology, Hokkaido University
[*2] 東京工科大学 Tokyo University of Technology

Conventional lecture substitution systems with humanoid robots use pre-defined gestures created by hand. Automatically generating these gestures makes it possible to create gestures without requiring expert knowledge and labor, which is expected to lead to further progress in research on lecture substitution systems. This paper proposes an automatic gesture generation method which is expected to consider the semantic context of an utterance. Our proposed method is implemented using a deep neural network with Bi-Directional LSTM units, applying filters for data correction, and axis conversion.

1. まえがき

近年, Softbank 社の Pepper[*1] や NAO[*1] のようなヒューマノイドロボットの発展が目覚ましく, かつては人間が行った仕事やサービスを代行することが可能になっている. また, 橋本ら [橋本 2012] や Xu ら [Xu 2014] によるヒューマノイドロボットを用いた講義代行の研究が行われていて, 今後ヒューマノイドロボットが教育の現場で導入される可能性は十分に考えられる.

現在の講義代行システムでは, ジェスチャは人手で作成されている. 自動化を実現することで, 専門的知識や労力を必要とせずにジェスチャを作成でき, 講義代行システムの開発におけるコストの削減が可能となる. このため, ジェスチャの自動生成化は講義代行システムの開発において, 必要不可欠な要素であると考えられる.

本研究の最終目的は, ジェスチャを自動生成し, 人間の講義よりも質の良い講義代行システムを開発することである. 本稿では, その第一段階として, 時系列データの関係性を考慮できる Bi-Directional LSTM (BLSTM)を使用した RNN により, ジェスチャを自動生成するシステムを Pepper を用いた講義代行システムに用いた際の印象評価実験およびジェスチャの自動生成手法の性能評価実験の結果, 考察について述べる.

2. 関連研究

ヒューマノイドロボットが講義を行う研究として, 橋本らの研究が挙げられる. 橋本らは, SAYA [橋本 2006] に小学校の理科の実験を代行させ, 印象評価実験を行った. ジェスチャは実験者が別室から SAYA を遠隔操作することで表現されている.

また, Xu らの研究では, 人手で作成されたポジティブ, ネガティブなジェスチャによって, 講義の印象がどのように変化するかを調査した. ヒューマノイドロボットには NAO を用いている. ポジティブな動作とは, 身振りが大きく活発な動きのことを指し, 一方, ネガティブな動作とは, 身振りが小さく動きも遅いジェスチャと定義される. ポジティブな動作とネガティブな動作で比較実験を行い, ジェスチャの動作による印象の変化を調査した.

一方, 我々は Pepper を用いて人手で作成したジェスチャを行う講義代行システムを開発し, 印象評価実験からジェスチャが講義にどのような印象を与えるかの調査を行なった[日和 2017].

本手法は, ジェスチャの作成方法が, 人手によるものでなく, 発話内容を考慮した自動生成である点で関連研究と異なる. 本稿の目的は, 自動生成されたジェスチャを用いた講義代行システムの印象評価およびジェスチャ自動生成手法の性能評価から, 自動生成されたジェスチャの発話内容に対する妥当性を評価することである.

3. 講義代行システムの処理過程

講義代行システムは GUI 開発ツールの Choregraphe[*2] を用いて実装した. 本システムの講義形式はスライドを用いた講義となる. 処理過程を図 1 に示す.

処理過程全体の概要を述べる. まず, 講義開始時に簡単な挨拶を行う. 挨拶を終えた後, スライド1ページから生成された講義内容とジェスチャを読み込む. 講義内容とジェスチャの再生後, 次のページの内容とジェスチャを読み込む. この一連の流れを最後のページまで繰り返すことで講義は終了となる. ジェスチャは 4 章で説明する BLSTM を用いた RNN のネットワークにより自動生成されたものである.

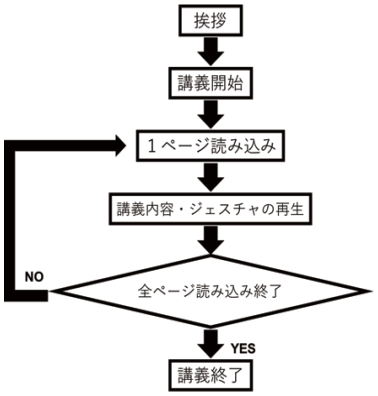


図 1 講義代行システムの処理過程

連絡先: 北海道大学工学部メディアネットワークコース言語メディア学研究室
〒060-0814 札幌市北区北 14 条西 9 丁目
E-mail:kururi@eis.hokudai.ac.jp

*1 http://www.aldebaran-robotics.com/ja/.
*2 http://developer.softbankrobotics.com/jp-ja/downloads/pepper

4. ジェスチャ自動生成手法

4.1 BLSTM を用いたジェスチャの自動生成

(1) データセット

学習に使用するデータセットは，竹内らが作成したモーションキャプチャデータセット [竹内 2017] である．データセット内には音声とそれに伴うジェスチャのペアが含まれる．ジェスチャのモーションデータは，関節の位置と回転角がフレーム毎に記載される時系列データである Bio Vision Hierarchy (BVH) 形式として保存される．データセットの増量と過学習対策のため，元の音声にホワイトノイズを合成した音声とジェスチャのペアを追加し，合計 2,098 センテンス (約 392 分) を学習と検証に用いた．

(2) 入力データ

入力データの特徴量として，音声認識において一般的な Mel-Frequency Cepstrum Coefficients (MFCC) を使用した．MFCC は声道特性を表す特徴量である．

図2に示すように，ある時刻に対して，timestep を 0.05 とし前後 30step 分のコンテキストを抽出した MFCC を入力とする．ジェスチャは一瞬の音素よりも，単語や文節の前後のコンテキストにより決定されるため，LSTM による時系列的な連続性の考慮だけでなく，一定範囲の音素の考慮も行う．

正解データとなる，入力データのラベルは時刻 t における各関節の位置と回転角のベクトルとなる．

(3) ネットワーク

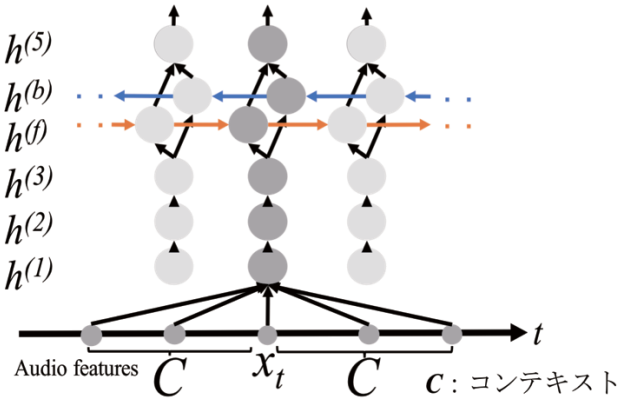


図 2　ネットワークの概略図

本ネットワークは竹内ら [竹内 2017] の手法を参考にしたものである．ネットワークの概略図を図 2 に示す．

BLSTM を用いたネットワークから学習モデルの構築を行う．

ネットワークは全 5 層の隠れ層で構成される．最初の 3 層は同じ構造で，再帰でない全結合層になる．

4 層目は BLSTM 層にあたる．この層により，時系列の前後の関係性を考慮することが可能となる．

5 層目は出力層であり，出力は BVH 形式における足の関節を除いた 43 関節の xyz オイラー回転角のベクトルとなる．

出力結果から平均二乗誤差により誤差 (loss) を算出した．データの学習回数を 500 回にし，学習を行ったところ最終的な loss は 27.4 となった．

4.2 フィルタ補正

出力となる回転角の時系列データにおいて，フレーム飛びと見られる箇所の動きを滑らかにするために，移動平均フィルタを用いて出力データの平滑化を行った．このフィルタを用いることで，データの特徴を残したまま滑らかなデータにすることが可能になる．式(1)に示すように，時系列データ$a_1, a_2, \ldots, a_n$に対し，直近 n 個の平均から新しい時系列データ$b_1, b_2, \ldots, b_n$を定める．

$$b_n = \frac{a_1 + a_2 + a_3 + \cdots + a_n}{n} \quad (1)$$

CG の棒人間でジェスチャの動きを確認したところ，窓枠 n=3 では依然ジェスチャがぎこちなく，n=9 では滑らかではあるが，値が間引かれたことによるジェスチャの遅延が目立ったため，今回は窓枠を n=5 とした．

4.3 座標軸変換

最後に，Pepper の各関節角度と対応するように，フィルタ補正を行ったデータに後処理として座標軸変換を行う．

BVH の座標軸と Pepper の座標軸を一致させて各関節で対応する軸の回転角を代入させる．具体的には BVH の姿勢を，各関節角度の値が 0 の時の Pepper の姿勢 (以後,初期ポーズという) に統一し，関節毎に対応させる．

図 3 に示すように，BVH の初期ポーズは T ポーズであり，手前が正面向きとして鉛直上向きが y 軸方向，紙面奥が z 軸方向となる．なお，Hand Middle1 は中指の付け根の関節にあたる．一方，Pepper の初期ポーズは前ならえであり，手前が正面向きとして鉛直上向きが z 軸方向，紙面手前が x 軸方向となる．

BVH の肩関節の y 軸を回転させ，BVH と Pepper のポーズを前ならえに統一する．各関節の回転軸は BVH では 3 軸，Pepper では 2 軸で構成されるが，肩関節においては Pepper では考慮されない BVH の z 軸 (Pepper の x 軸) が及ぼす肩の動きへの影響が大きかったために，BVH の肩関節の z 軸である ArmZ を ShoulderRoll，ShoulderPitch に仮に表 1 に示すように加え，考慮した．表 1 に示す Hand は手の開閉を表す．Pepper への変換のために，今回は仮に中指の付け根の曲げ具合を表す Hand Middle1Z を用いた．また，Pepper の関節名の一部として表記される Yaw，Pitch，Roll はそれぞれ x，y，z 軸回転を表す．

学習データとなる人間の頭の向きは正面方向よりも下向きの場合が多く，Pepper に変換すると常に頭が下を向くことになる．また，腰の動きは Pepper に変換すると無理な姿勢を連続して行うことが多かった．これらの動きはジェスチャの評価に悪影響を及ぼすことが懸念されたために，Pepper への変換は両腕のみとした．

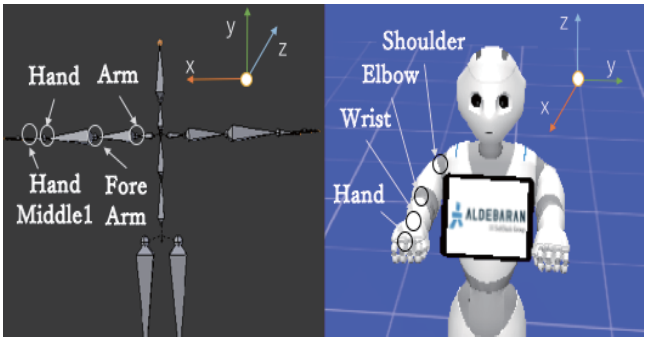


図 3　BVH と Pepper の座標軸と関節名

表 1　関節角度の対応表(符号上:右腕，符号下:左腕)

Pepper の関節	BVH の関節から Pepper の関節への対応式
ShoulderRoll	ArmY ∓ 75 − ArmZ*5/6
ShoulderPitch	− ArmX ∓ ArmZ*1/6
ElbowRoll	ForeArmY (両腕同式)
ElbowYaw	± ForeArmX
WristYaw	± HandX
Hand	90 + HandMiddle1Z (両腕同式)

4.4 ジェスチャの自動生成に対する性能評価実験

(1) 実験設定

10 種類の音声データに対する original, predicted, mismatched の 3 種類のジェスチャの印象を検証する 1 要因 3 水準の実験を行った. original は音声データのジェスチャ, predicted は音声データから予測したジェスチャ, mismatched は別の音声データのジェスチャを音声の再生時間と一致するよう末尾を削除したジェスチャである.

参加者は 20 代理系大学生 28 名, 文系大学生 2 名の計 30 名 (男性 21 名, 女性 9 名) である. 音声データは **4.1** の (1) で述べられたデータセットのうち, モデルの学習に用いていない 10 秒ほどの音声を 10 種類選定した. 実験は Web 上で行い, 1 ページにつき動画を 1 つ表示する. 実験で使用した動画は, YouTube のサイト*3から閲覧できる. 3 条件の表示順は参加者 6 名毎にランダムに行い, 表示順による回答への影響を均一化した.

(2) 評価項目

ジェスチャの妥当性を「動きの自然さ」,「発話音声との時間的な整合性」,「発話音声との内容の整合性」の 3 項目から評価した. 以下に示すように, 項目毎に 3 つの小項目を設け, 合計 9 個の項目について 7 段階評価を行った.

・動きの自然さ
- Q1:ジェスチャは自然に見えた
- Q2:ジェスチャは滑らかであった
- Q3:ジェスチャに違和感はなかった

・発話音声との時間的な整合性
- Q4:ジェスチャのタイミングは発話音声と一致していた
- Q5:ジェスチャのスピードは音声発話と一致していた
- Q6:ジェスチャの緩急は発話音声と一致していた

・発話音声との内容の整合性
- Q7:ジェスチャは発話音声と一致していた
- Q8:ジェスチャは発話音声をよく表現していた
- Q9:ジェスチャは発話音声の理解を促進した

(3) 結果

ジェスチャの種類別の平均に対して, 一元分散分析を行ったところ, 主効果が認められた (Q1:F(2, 29)=4.766, p=.0121, Q2: F(2, 29)=20.27, p<.001, Q3:F(2, 29)=3.277, p=.0448, Q6:F(2, 29)=8.39, p=.001, Q7:F(2, 29)=17.18, p<.001, Q8:F(2, 29)= 14.96, p<.001, Q9:F(2, 29)=11.84, p<.001). Bonferroni の下位検定を行ったところ, Q2 では predicted が他条件に対して有意差が検出されるほど評価が高くなった. Q1, Q3 は original には平均評価値で下回るが, mismatched を上回る結果となった. Q5~Q9 においては predicted が他条件に対して有意に平均評価値が高くなる設問は一つも存在しなく, original に対して有意差が検出されるほど平均評価値が低くなる設問も存在した. 結果を図 4 に示す.

(4) 考察

- ジェスチャの自然さ

今回の結果から, 滑らかさにおいては十分な質のジェスチャが生成できていると考えられる. また, Q1~Q3 の結果から predicted は mismatched よりも自然なジェスチャであると考えられる.

- 時間的な整合性

今回の結果の要因として, 頻繁に手を動かしていたことが悪影響を及ぼしていると考えられる. また, ジェスチャの緩急においては, original に対して有意に平均評価値が低くなった. 原因として, 移動平均フィルタを用い動きが間引かれたことで, ジェスチャの速度が多少なりとも均一化されたことが考えられる.

*3 https://youtu.be/iINtbQUAhLI

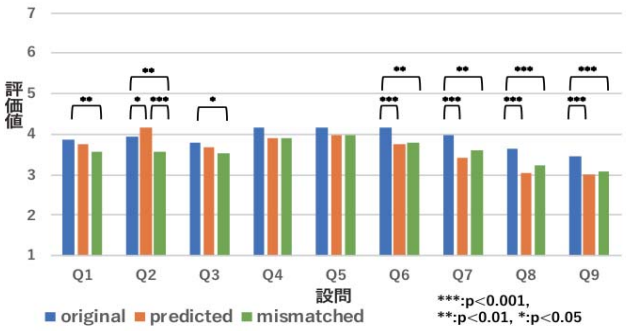


図 4 各設問の平均評価値

- 内容の整合性

全ての設問で, 平均評価値が他条件よりも下回る結果となった. このような結果の原因として, 提案手法により生成されたジェスチャは, 一定範囲で上下に手を動かすことが多く見受けられたことが挙げられる.

5. 講義代行システムの印象評価実験

5.1 実験設定

講義内容は芳尾ら [芳尾 2017] の論文から自動生成される. 芳尾らの論文では, 教科書と対応するスライドを基にシナリオを自動生成し, Pepper が講義代行するという内容が述べられている.

参加者は 20 代理系大学生 31 名 (男性:27 名, 女性:4 名) である. 実験後, アンケート用紙を配布し, 評価を行った. なお, ジェスチャの自動生成に対する性能評価実験時と同様に Pepper の動作は両腕のみとした.

あらかじめ録音した Pepper の音声を入力として, 4 章で述べられた手法を用い, ジェスチャを自動生成した.

5.2 評価方法

「講義全体」,「ジェスチャ」の二つの項目について評価した.

「講義全体」は GODSPEED [Bartneck 2008] を参考に「人間性」,「知性」,「好意」,「活発性」,「講義の有効性」という 5 つの測定概念を定めた. 表 2 に示すように, 概念毎に 3 組の形容詞対を設け, 全 15 組の形容詞対に対して SD 法 [末永 1987] を用いて 7 段階評価 を行った.

「ジェスチャ」は, 発話内容に対するジェスチャの妥当性を,「動きの自然さ」,「発話との時間的な整合性」,「発話と内容の整合性」を評価する 3 つの設問に対し, 7 段階評価を行った.

表 2 項目「講義全体」の概念毎の形容詞対

概念	形容詞対
人間性	機械的な-人間的な, 人工的-生物的, 偽物のような-自然な
知性	愚かな-賢明な, 知的でない-知的な, 無知な-物知りな
好意	嫌い-好き, 親しみにくい-親しみやすい, 不愉快な-愉快な
活発性	熱心でない-熱心な, 非活発な-活発な, 活気のない-生き生きとした
有効性	わかりにくい-わかりやすい, 興味深くない-興味深い, 飽きのくる-飽きのこない

5.3 結果

各項目の平均評価値の結果を表 3, 表 4, 図 5 に示す. 表 3 は概念に属する形容詞対の値の平均値となる.

講義全体では，高評価の概念は知性と活発性，低評価の概念は人間性と有効性，中央値 4 付近の概念の好意であった．ジェスチャでは，全ての設問で中央値 4 を下回った．

表 3　項目「講義全体」の概念毎の結果

概念	人間性	知性	好意	活発性	有効性
平均値	3.15	4.97	4.14	4.28	3.73

表 4　項目「ジェスチャ」の結果

設問	平均値
ジェスチャは自然だったか	3.19
ジェスチャのタイミングは音声とあっていたか	3.74
ジェスチャと説明内容は一致していたか	3.26

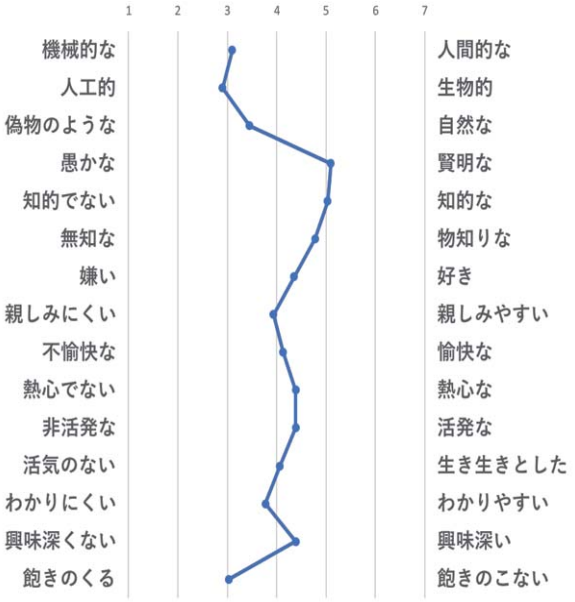


図 5 「講義全体」の 7 段階印象評価

5.4 考察

- 講義全体

点数の高かった概念である，活発性について考察を行う．活発性はジェスチャによる影響が大きく，常に手を動かしていたために活発な印象を与えたと考えられる．

次に，点数の低かった概念である人間性，有効性の考察を行う．これらの概念が低評価であった要因は，類似のモーションが多かった，話の内容とジェスチャの内容が合っていなかったといったことが挙げられる．

- ジェスチャ

3 つの設問のうち，「動きの自然さ」は 3.19 と最も低い値となった．低評価の要因は，まず，「ジェスチャは自然だったか」という内容が解釈の余地が大きい設問であり，他 2 つの設問を鑑みた「ジェスチャの総合的な自然さ」という評価も含まれていることが挙げられる．次に，動作が両腕のみであったことによる見た目の不自然さが及ぼした影響も低評価に繋がったと考えられる．また，生成されたジェスチャの速度に追いつけないために生じたガタつき等も低評価の原因の一つと考えられる．

6. まとめ

本稿では，自動生成によるジェスチャを用いた講義代行システムを提案した．

ジェスチャ自動生成に対する性能評価実験から，自動生成されたジェスチャは滑らかな動きであることが評価される一方，発話内容との時間的な整合性，内容の整合性は低評価であった．また，講義代行システムの印象評価実験では，活発な印象を与える結果となった．常に両腕を動かすことで，活発な身振りを表現できたと考えられる．一方，飽きやすい印象を与える結果となった要因としては似た動きが多いことが考えられる．

今後の課題として，発話内容の意味を取れるジェスチャを生成するために，ニューラルネットワークの改良とデータセットの増築を行いたい．データセットに関しては，openpose [Cao 2017] を用いて動画から回転角を導出し，データセットの増量を行うことを検討する予定である．また，自動生成する関節を両腕のみだけでなく全身の関節を対象に拡大することにより，見た目の違和感を除いていく予定である．

参考文献

[橋本 2012] Takuya Hashimoto, Igor M. Verner and Hiroshi Kobayashi.: “Human-Like Robot as Teacher’s Representative in a Science Lesson: An Elementary School Experiment.” Robot Intelligence Technology and Applications, pp.775–786, 2012.

[橋本 2006] Takuya Hashimoto, Sachio Hitramatsu, Toshiaki Tsuji and Hiroshi Kobayashi.: “Development of the Face Robot SAYA for Rich Facial Expressions.” In international joint conference on artificial intelligence, pp.5423–5428, 2006.

[Xu 2014] Junchao Xu, Joost Broekens, Koen Hindriks and Mark A. Neerincx.: “Effects of Bodily Mood Expression of a Robotic Teacher on Students.” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.2614–2620, 2014.

[日和 2017] 日和航大, 荒木健治, 長谷川大: “ヒューマノイドロボットを用いた講義代行システムにおけるジェスチャの考察”, 電気・情報関係学会北海道支部連合大会, pp.220‐221, 2017.

[竹内 2017] Kenta Takeuchi, Dai Hasegawa, Shinichi Shirakawa, Naoshi Kaneko, Hiroshi Sakuta and Kazuhiko Sumi.: “Speech-to-Gesture Generation: A Challenge in Deep Learning Approach with Bi-Directional LSTM.” In Human-Agent interaction (HAI) symposium, pp.365–369, 2017.

[竹内 2017] Kenta Takeuchi, Souichirou Kubota, Keisuke Suzuki, Dai Hasegawa and Hiroshi Sakuta.: “Creating a gesture speech dataset for speech-based automatic gesture generation.” In HCI International 2017–Posters’ Extended Abstracts: 19th International Conference, HCI International 2017, pp.198–202, 2017.

[芳尾 2017] 芳尾哲, 荒木健治.: “ヒューマノイドロボットを用いた講義代行システムのための講義シナリオ文自動生成手法”, 電気・情報関係学会北海道支部連合大会, pp.218‐219, 2017.

[Bartneck 2008] Christoph Bartneck, Dana Kulic and Elizabeth Croft.: “Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots.” In Proceedings of the Metrics for Human-Robot Interaction Workshop in Affiliation with the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp.71–81, 2008.

[末永 1987] 末永俊郎. “社会心理学研究入門”, 東京大学出版会, 1987.

[Cao 2017] Zhe Cao, Tomas Simon, Shin-En Wei and Yaser Sheikh.: “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.” In IEEE, pp.1302–1310, 2017

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

コミュニケーションロボットにおける文脈に依存した視線外し動作の検討

Contextual Dependent Gaze Aversion in a Communication Robot

川口 佑希子 *1　田中 文英 *2
Yukiko Kawaguchi　Fumihide Tanaka

*1筑波大学 理工学群 工学システム学類
College of Engineering Systems, School of Science and Engineering, University of Tsukuba

*2筑波大学 システム情報系 知能機能工学域
Faculty of Engineering, Information and Systems, University of Tsukuba

In this paper, we study contextual dependent gaze aversion in communication robot and discuss its influence on the impression of humans who interact with the robot. We conducted a human-human data collection, and then implemented a gaze aversion pattern based on the data. Three different robots were compared: a robot with a contextual gaze aversion (Robot(A)), a robot with a gaze aversion that was not contextual (Robot(B)), and a robot that could track human faces but with no gaze aversion (Robot(C)). Those three robots were evaluated in human participants experiments. The results show that Robot(A) gave better impression compared to Robot(C). Comments obtained from the participants indicated that their semantic interpretation of the robot's utterances depended on the contextual gaze aversion.

1. はじめに

会話により心身の健康や主観的幸福感に有用であり，他者とのコミュニケーションは重要視されているが，日常生活の中で十分な会話量を確保することは難しい．そこで，コミュニケーションロボットによる問題解決が期待されている．

ロボットが人の発話を引き出すためには，人間の信頼を勝ち取り，聞き上手・話し上手であることが必要である．これまでの研究では，ロボットに人間らしい動作を取り入れることが信頼獲得に有効だと示されており [帆足 97][神田 03]，この動作の 1 つに視線がある．Waxer は人間-人間の会話で視線が相手への印象形成の大きな要素となっていることを示した [Waxer 77]．神田らの研究では，アイコンタクトのようなロボットの協調的身体行動が人間のロボットに対する評価を向上させることが分かっている [神田 03]．一方でロボットに顔を凝視されることに不気味さを感じるという報告 [三河 11] もあり，適切なタイミングで視線を外すことが必要である．人間-人間の会話中に見られる視線外しについて Glenberg らは，適度に難しい質問をされた時に視線を外す動作に着目し，視線外しは人間の認知処理効率を高めると示した [Glenberg 98]．人間-ロボットの会話においての視線外しについて，Andrist らは Glenberg らの研究を踏まえ，人間はロボットの視線外しを何らかの意図的な動きであると認識し，そのロボットをより思慮深く創造的であると評価することを示した [Andrist 14]．

しかし，これらの先行研究では視線外しとそれが観測された時の文脈については研究されていない．そこで我々は文脈に依存したコミュニケーションロボットの視線外し手法の提案を目指し，ロボットの文脈と視線外しの組み合わせがロボットへの印象に与える影響を調査する．文脈の例として，与えられた質問の回答しやすさや会話の内容が挙げられるが，本研究では与えられた質問の回答しやすさに注目した．以後，質問の回答しやすさを回答難易度と呼ぶ．本研究では，人間-人間の対話における視線外しの分析のためデータ収集の予備実験を行い，提案する視線外しをロボットに実装し，人間-ロボットで会話実験を行い，アンケート結果とビデオ等の分析から質問の回答難易度と視線外しの関係の考察と更なる要求仕様について議論する．

連絡先: 川口 佑希子，kawaguchi@ftl.iit.tsukuba.ac.jp

2. 予備実験

2.1 実験目的

人間-人間の会話において自然な視線外しパターンのデータ収集及びそれが表れる会話内容との関連性の調査を目的として，実験者との 2 者間会話実験を行った．

2.2 実験手順

本実験では，人間-人間の日常会話を想定し，実験者と対面で会話する配置とした．このとき実験者と実験参加者の距離は 1.5m とし，両者の顔角度を Kinect センサ [Kinect] で取得した．また，データ分析の用途で会話の様子を録画した．

実験参加者は実験者の質問に答え，その後同じ質問を聞き返すように指示され，これを 10 分間続けた．

2.3 実験参加者

実験参加者は 20 代の日本人男性 2 名であった．2 名とも実験者とは日常的に会話をしており，初対面での緊張などによる影響は無視できると考えられる．

2.4 結果

会話中の様子の録画から，視線外しのパターンが得られた．

質問内容を受け，考えている途中に視線外しが見られた．与えられた質問が同じテーマについてであれば，より過去の記憶であるほど回答までの時間と視線外しの時間の長さは長かった．

例) 実験参加者 1 において
“今朝は何を食べましたか?”　: 2.5 秒
“昨日の朝は何を食べましたか?” : 7.5 秒
この結果から，ロボットに実装するパターンを抽出した．

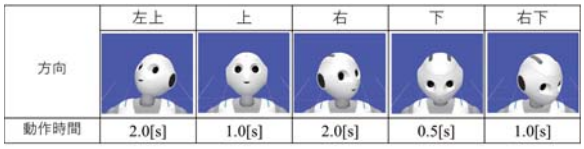


図 1: ロボットに実装した視線外しの方向

3. 実装

3.1 使用したロボット

本研究では，視線外しを実装するロボットとして Softbank Robotics 社のヒューマノイドロボット Pepper[Pep] を使用し，公式モジュールの Choregraphe を用いて制御を行った．

3.2 実装した機能

3.2.1 視線外し

ロボットには図 1 に示した 5 種類の視線外しのパターンと顔追跡の実装をした．

3.2.2 会話内容

会話内容として 10 個の質問とそれに対するロボットの回答を用意した．質問の回答難易度について，回答難易度が高いほど思い出したり考えたりする認知的処理が必要で，回答が始まるまでに時間のかかるものと定義した．10 個の質問のうち 5 個を認知的処理が比較的軽く即答可能な回答難易度の低い質問，残りの 5 個を回答難易度の高い質問とした．質問とそれに対するロボットの回答は実験者の操作によって発話されるよう実装を行った．

3.3 ロボット条件

のちに行う評価実験のため，3 つのロボット条件を設定した．

ロボット (A)　文脈依存あり：
回答難易度が高い質問に回答するとき視線外しを行う．

ロボット (B)　文脈依存なし：
回答難易度が低い質問に回答するとき視線外しを行う．

ロボット (C)　顔追跡：
実験参加者の顔を追跡し見つめ続ける．

これらのロボット条件での視線外しと顔追跡は首関節の動きのみで行い，いずれの条件でも他の関節は動かさず，発話内容やイントネーションも統一した．

4. 実験

3 章のロボットを使用し，評価実験を行った．

4.1 実験目的

以下に述べる 2 つの仮説の検証と更なる要求仕様の探索を目的に実験を行った．

(H1) ロボットが適切な文脈で視線外しを行うと，視線外しをしない場合と比較して印象が向上する．

(H2) ロボットが適切な文脈以外で視線外しを行うと，視線外しをしない場合と比較して印象が低下する．

4.2 実験手法

実験手法の詳細について以下に述べる．

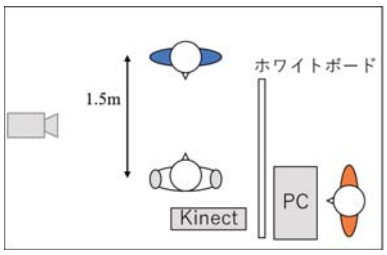


図 2: 実験室内の配置



図 3: 実験の様子

4.2.1 実験参加者

実験参加者は，日本人 20 代学生 7 名 (男性 6 名，女性 1 名) であった．7 名とも今回使用したロボットを日常的に見ており，ロボットとの対面での緊張が実験結果へ影響することは無視できると考えられる．

4.2.2 アンケート

実験後の評価のため，アンケートと自由記述欄をまとめたアンケート用紙を作成した．アンケートではロボットの社会的属性評価として，Warmth，Competence，Discomfort という因子を持つ RoSAS[Carpinella 17] をもとにした．日本語でアンケートを取るため，野口ら [野口 18] が作成した和訳を用い，これらの項目について 6 段階の Likert-scale で評価してもらった．

このアンケート用紙を用いて，実験参加者がロボットに抱いた印象を評価する．4.1 節で触れた (H1) について，ロボット (A) はロボット (C) に比べ Warmth と Competence の得点が高くなり，Discomfort の得点が低くなると想定した．また，(H2) については，ロボット (B) はロボット (C) に比べて Warmth と Competence の得点が低くなり，Discomfort の得点が高くなると想定した．

4.3 実験手順

実験環境を図 2 に，実際の実験の様子を図 3 に示す．このときロボットと実験参加者の距離は 1.5m であった．データ分析の用途で実験参加者の顔角度を Kinect で取得し，会話の様子を録画した．実験は，説明フェーズ，実験という流れで行った．

- 説明フェーズ
  実験の説明では，実験参加者はロボットからの質問に答え，“君は?” 等の聞き返しを行うように指示された．10 個の質問が終わり次第，アンケートに回答するよう伝えられ，事前にロボットからの質問の内容は伝えられなかった．また，ロボットとの会話後，アンケートに答えるという流れを 3 回繰り返すことも伝えられた．

- 実験
  実験参加者は 3 章で示した 3 条件のロボットにおいて参加者内配置実験を行った．

  実験は Wizard of Oz(WOZ) 法を用い，操作は実験者がホワイトボードの裏で見えないように行った．実験参加者は各ロボットと対面し，ロボットからの質問に回答，ロボットに聞き返した．聞き返されたロボットは回答しながら条件に応じた視線行動を行い，実験参加者に次の質問を尋ねた．

5. 結果

5.1 回答難易度の妥当性

全実験参加者の初回実験時の質問に対する回答までにかかった時間をについて対応のある t 検定を行ったところ，質問難易度において有意差が認められた ($t(7)= 7.76$，$p <.05$)．

5.2 ロボットに対する社会的属性の評価

RoSAS はロボットの社会的属性について 3 つの因子で構成される．社会的属性とロボット条件を参加者内因子とした 3 × 3 の反復測定分散分析を行った結果，社会的属性のみ有意差が認められ，ロボット条件と相互作用は確認されなかった (社会的属性：$F (2, 54) = 3.17$，$p <.05$，ロボット条件：$F (2, 54) = 3.17$，$p = .387$，相互作用：$F (4, 54) = 2.54$，$p = .690$)．

また，各社会的属性の中で 3 つのロボット条件を比較するため分散分析を行った結果を図 4 に示す．Warmth については有意差が認められた ($F (2, 12) = 3.89$，$p <.05$) が，Competence, Discomfort の中ではロボット条件間に有意差は出なかった (C：$F (2, 12) = 3.89$，$p = .800$，D：$F (2, 12) = 3.89$，$p = .951$)．続いて Warmth のうち，ロボット条件のどのペア間で有意差があるか明らかにするため，対応のある t 検定とボンフェローニ補正を行った結果，ロボット (A)-(C) 間で有意差が見られた ($t(7)= 3.09$，$p<.017$) が，その他では認められなかった ((A)-(B)：$t(7)= 2.09$，$p=.041$，(B)-(C)：$t(7)= 1.82$，$p=.059$)．

5.3 自由記述

実験参加者からアンケート自由記述欄にて得られたコメントを列挙する．なお，参加者内配置実験を行ったため，コメントにはアンケート回答直前に会話したロボット条件と，その前に会話したロボット条件を比較したものが多かった．

- ロボット (A) について
  - 動きがないよりは対話している感じがあった．
- ロボット (B) について
  - 視線が外れた時は人間のような動きを感じたが，それ以外ではじっと見られている感じがして少し怖かった．
  - 兄弟の話をした時によそ見をしたので，その話題について何かマズイ思い出でもあるのかと思った．
  - Pepper がジョークを言うのが面白かった．ロボット (A), (C) ではジョークだと気づかなかったが，今回は顔を逸らして言ったのでジョークだと分かった．
- ロボット (C) について
  - ロボット (A)，(B) と比べて淡々と話が進んでいった感じがして，あまり楽しい会話とは思えなかった．
  - ロボットが動かないので，退屈に思った．
  - ロボット (B) との違いがあまりよくわからなかった．

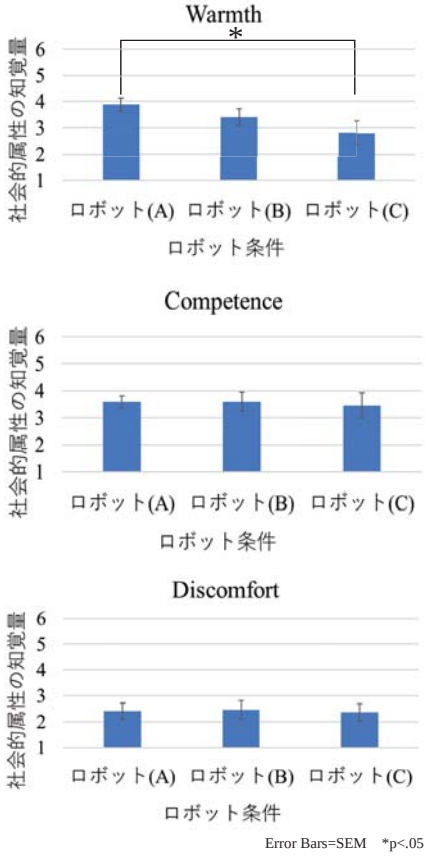


図 4: RoSAS の各社会的属性内での，各ロボット条件における社会的属性知覚量の比較結果

6. 考察

本章では，5 章で得られたアンケート結果や自由記述欄のコメント，及び録画データによる考察を行う．

6.1 回答難易度の妥当性

5.1 節で得られた結果から，我々が設定した質問の難易度分類は妥当であったと言える．

6.2 視線外しがロボットの印象に与える影響

また，5.2 節の結果より，適切な文脈での視線外しはロボットの社会的属性 Warmth に対する知覚量を向上させるということが分かった．この Warmth には “幸せな”，“思いやりのある”，“情緒的な” などの要素が含まれており，ロボットが人間の信頼を獲得するための印象向上が見られたということができる．一方で Competence について，適切な文脈での視線外しでこの知覚量が向上すると想定していたが，すべてのロボット条件間で有意差が見られなかった．今回の文脈での視線外しのやり方では Competence には影響がなく，その知覚量は 3 条件間で共通の，ロボットの発話内容に依存していたといえる．また，Discomfort の知覚量もすべてのロボットで有意差が認められなかった．これは，3 条件下で共通の発話内容

やイントネーション，体の動きに依存していたといえる．以上のことから，仮説（H1）は部分的に支持されたと考えられる．

また，仮説（H2）について，Competence と Discomfort はすべてのロボット条件間で有意差が見られなかった．また，有意差は出なかったものの，Warmth の結果で顔追跡のみを行ったロボット (C) よりも，文脈を考慮しない視線外しを行ったロボット（B）の得点がわずかに高い結果となった．これらのことから，文脈に合わない視線外しをするロボットは，全くしないロボットに比べて印象が低下するという仮説（H2）は支持されなかったと考えられる．

一方で，5.3 節の自由記述欄の結果ではロボット (B) について，ロボットの視線外しにより “発話内容がジョークであると感じられた”，“ロボットにとって都合の悪いことを話しているように感じられた” という内容の意見が得られた．これらのコメントから，実験参加者はロボットの意図を感じ取ったと考えられる．このように文脈と視線外し，またそれらから受ける印象は関係性があり，ロボットの視線外しを調整することにより，同じ発話内容でも人間がロボットの意図を感じると示唆された．また，ロボット (C) について “淡々と話が進んでいった感じがして，あまり楽しい会話とは思えなかった”という意見が得られ，視線外しが会話を楽しいと感じさせる効果も確認された．

6.3 今後の課題

今回実施した実験では，第 1 にデータ収集実験，評価実験のサンプル数が小さいことが問題であった．これによりデータ収集で得られる視線外しとその時の文脈の種類も少なくなってしまった．この問題を解決するため，今後はデータ収集実験のサンプル数を大幅に増やし，文脈と視線外しのパターンを多く抽出し，ロボットに実装する．

第 2 に，アンケートの取り方や使用したアンケートにも問題があったと考えられる．今回の実験結果でも，自由記述欄のコメントからは印象が改善していることが示されているが，アンケート結果では明確な結果は出ていないため，今回の実験の評価尺度としては適切ではなかったと考えられる．この問題点を踏まえ，今後はより細かい印象まで評価できる実験設定にすることが必要である．さらに，ひとつひとつの視線外しについてより細かく評価するため，実験参加者にロボットとひととおり会話してもらった後，その会話のビデオを見返してもらい，各視線外しについて評価してもらう実験が適していると考えられる．このとき自由記述だけではなく，違和感を持った度合い，信頼できるかの度合いを 6 段階の Likert-scale で回答してもらうことを検討している．また，ロボットの視線外しが発言のニュアンスに与える影響の調査のため，ロボットの意図を感じたか，感じた場合はどのような意図を感じたかを調査するアンケートを検討している．

また，今回は注目しなかった文脈の他の要素についても調査が必要である．

7. おわりに

本稿では，視線外しの文脈依存性がコミュニケーションロボットの印象に影響を与えると仮定し，人間-人間のデータ収集実験を行い，そこから得られた視線外しのパターンをロボットに実装し，評価実験を行った．結果，質問の回答難易度に適した視線外しを行うロボットは視線外しを全く行わないロボットと比較して一部の印象が向上した．また，自由記述欄のコメントから，視線外しの回答難易度との対応により，同じ発話内容でも受ける印象や意味の解釈が異なることが示唆された．

今後の課題として，より多くの視線外しデータを収集し，実験目的に適した個別具体的なアンケートを使用して実験を行うことで，更なる視線外しによる印象改善，視線外しによる発話内容のニュアンス変化，そして文脈の他の要素と視線外しの関係についても調査をする必要がある．

謝辞

本研究は, 科研費・基盤研究 (A) コミュニケーションを支援するロボットインタフェース (課題番号 15H01708) の支援を受けて行われた．

参考文献

[Andrist 14] Andrist, S., Tan, X. Z., Gleicher, M., and Mutlu, B.: Conversational gaze aversion for humanlike robots, in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 25–32ACM (2014)

[Carpinella 17] Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J.: The robotic social attributes scale (RoSAS): development and validation, in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 254–262ACM (2017)

[Glenberg 98] Glenberg, A. M., Schroeder, J. L., and Robertson, D. A.: Averting the gaze disengages the environment and facilitates remembering, *Memory & Cognition*, Vol. 26, No. 4, pp. 651–658 (1998)

[Kinect] Kinect, : https://www.xbox.com/ja-JP/xbox-one/accessories/kinect

[Pep] SoftBank Robotics Corp. http://www.softbank.jp/en/robot/

[Waxer 77] Waxer, P. H.: Nonverbal cues for anxiety: An examination of emotional leakage., *Journal of Abnormal Psychology*, Vol. 86, No. 3, p. 306 (1977)

[三河 11] 三河正彦, 田中和世：実環境を利用した図書館司書ロボットの研究開発, 情報処理学会インタラクション, Vol. 2011, pp. 319–322 (2011)

[神田 03] 神田崇行, 今井倫太, 小野哲雄, 石黒浩：人-ロボット相互作用における身体動作の数値解析, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2699–2709 (2003)

[帆足 97] 帆足啓一郎, 横山真男, 荒井大輔, 安藤義範, 白井克彦：人間型対話ロボットにおける非言語情報の役割, 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 1997, No. 16 (1996-SLP-015), pp. 67–73 (1997)

[野口 18] 野口洋平, 上出寛子, 田中文英：遠隔コミュニケーションを仲介するロボットが高齢話者の自己開示に与える影響, ヒューマンインタフェース学会論文誌, Vol. 20, No. 1 (採録決定，2018)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

高齢者をつなぐ巡学ロボットにおける ユーザの感情表出量に基づく対話選択のアルゴリズム

Dialogue Selection Algorithm Based on the Amount of User's Emotional Expression in a Circulate-Learning Robot Connecting Elderly People

笠井 翼 *1　田中 文英 *2
Tsubasa Kasai　Fuminide Tanaka

*1筑波大学 大学院 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba

*2筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

We proposed a Circulate-Learning Robot for solving isolation of elderly people. Circulate-Learning Robot is a teachable robot which is designed to go to various places and be taught by people. It was found from the previous research that it is important to equalize the amount of teaching from the user. In addition, we focused on the importance of empathy in new relationship building. We propose a dialog selection algorithm to equalize the amount of user's emotional expression for a user of Circulate-Learning Robot. The number of morphemes in the sentence in which the emotion expression word appears was defined as the amount of emotional expression. In the proposed method, users are clustered using this amount of emotional expression. The topic is decided from the co-occurrence probability of an independent word and an emotion expression word. The proposed system switches dialogue strategy between empathy expression and promote to expressing impressions. This paper proposes the algorithm that selects a dialog strategy according to the amount of user's emotional expression.

1. はじめに

我々は, 独居高齢者の孤立の解消に向けて, 1 台のロボットが多地点を巡り学習することで, 複数のユーザが 1 つのロボットを共有して育成する経験を通して新たなコミュニティを構築することを目的とした巡学ロボットを提案した [笠井 17]. 巡学ロボットのコンセプト図を図 1 に示す. 先行研究では, 高齢者への聞き取り調査から, 人間的な反応をするロボットに教えたいとの意見から人間的な学習機能と学習結果表出機能を合わせた 4 つの機能をロボットに実装し, 実際の高齢者に対して巡学ロボットを使用してもらう予備調査を行った. 機能に関しては, 実装目的通り, 学び方が人間のようだったとの意見がインタビューから得られた. また, ユーザ間で巡学ロボットの学習結果の表出から会話が生まれたことから, 巡学ロボットのコンセプトの有効性が示唆された. しかし, 人間関係構築の上で必要な他者との共通点や相違点の把握を支援する機能については充分な結果を得たとは言えなかった.

他者との共通点や相違点を認識する部分には, その人の仕草や言語, 話す内容など多く挙げられる. その中でも, 関係初期においては話す内容からその人となりを判断することが多いと考えられる. この時に重要視されるものの 1 つに共感できるかどうか, という指標がある. そこで, 巡学ロボットではユーザがどのような内容でどのような感情を抱くのかを学習する, それをどのようにユーザから引き出すかが重要となってくる. ユーザに合わせて共感を示すことで対話に対する印象の向上や自然な対話の実現を目的とした対話システム開発は様々行われているが, ロボットとの対話を通してユーザ同士の相互理解を促進させることを目的として、ロボットの共感に着目したものはない.

そこで, 本稿では巡学ロボットにおいて, ユーザ間の共感の促進が関係構築において重要であると考え, ユーザから感情表現を引き出すためのユーザの感情表出量に基づく対話選択アルゴリズム手法を論じる. 2 章で予備調査と関連研究から巡学ロボットの課題とそれに対する解法へのアプローチを述べ, 3 章にて具体的なアルゴリズムについて述べる.

連絡先: 笠井 翼, 筑波大学システム情報工学研究科, 〒 305-8573 茨城県つくば市天王台 1-1-1, kasai@ftl.iit.tsukuba.ac.jp

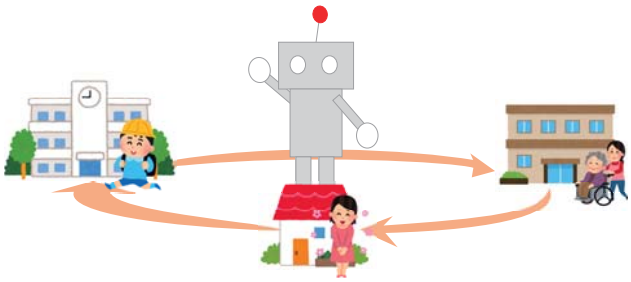


図 1: 巡学ロボットの概念図

2. 巡学ロボットの課題と関連研究

本章では, 予備調査から得られた結果と関連研究から, 巡学ロボットの課題を述べ, 要求仕様について論ずる.

2.1 学習結果表出の偏り

予備調査に協力いただいたユーザで, 「お寿司」に関連する語句の教示に対して, 「ばななとみかん」と特徴的な教示を行ったユーザがいた. そのユーザに対して, 何故, お寿司からばななとみかんを教えたのか, という会話が発生した. このように, 互いの共通点や相違点で明確化されたところから会話が生起することがロボットを通しても行われることが確認され, 巡学ロボットのコンセプトの有効性が示唆された. しかし, 会話のほとんどの時間が, その特定の人物に対してのみ行われ, 話題内容に偏りが生まれた. 小川は, 2 者間の会話において, 開示

量の偏りが会話への好ましさと負の相関があり, 好ましさは会話相手との将来の関係性に対する印象と正の相関があると報告している [小川 03]. つまり, 巡学ロボットの学習結果の表出に偏りがあると, ユーザ間で発生する会話の話題の偏りに繋がり, 関係構築において悪影響を及ぼすことが懸念される. しかし, これに対処する巡学ロボットの機能はまだ考案されていない. 加えて, ユーザ間の関係構築を促進させる機能についても, 予備調査にてユーザ間で共通してロボットに教えた内容を優先的に学習結果として表出するアルゴリズムを実装したが, 関係構築の促進という目的に対して充分な結果は得られていない. これらから, ユーザ独自の特徴的な情報である感情を表出した量に着目し, ユーザ間でロボットへの感情表出量を均等に調整する機能が必要であると考える.

2.2 ユーザへの共感表出

人間の共感には重要な要素が含まれているとされている. 人は文化や性別に問わず否定的な事象やトラウマティックな事柄に遭遇した時, そのときの情動を他者に話す傾向があり, この情動の共有行動の理由の 1 つは, 他者に自分を理解し, 受け入れてもらうためであるとされている [Rimé 92]. 共感的感情反応と向社会的行動には正の相関があると報告し, ポジティブな感情への好感・共有が高い人は向社会的な行動を選択しやすいのではないかとされている [櫻井 11]. これらから, 孤立する独居高齢者にとって共感できる相手を持つことで向社会的に繋がることが推察され, 孤立の解消への一助となることが考えられる.

ロボットがユーザに共感を示す様々な研究があり, 共感することはロボットの印象を向上させることが報告されている. チェスの趨勢に合わせて共感する内容の発話を行うことで, ユーザがロボットに感じる friendly が向上した [Leite 13]. ユーザの表情から感情を推定し, ロボットが徐々にユーザと同じ感情の表情を表出することで, positive, active, happy の評価が向上した [山野 08]. 高齢者向けの対話システムの中に共感を示す手法を取り入れ, 高齢者のユーザにシステムとの対話に対して好印象を与えた [大竹 12]. しかし, これらの研究は全てロボットやシステムに対する印象を向上させるためであり, 共感を用いて人と人を繋げる研究はされていない. また, 感情の強さを検出し, それに合わせてシステムの表出を選択する手法には着目はされているが, ユーザの感情表出量によってシステムの表出を選択する手法に着目したものはない.

2.3 ユーザからの要求によるタスクの決定

予備調査から, 高齢者がロボット技術に対して求めるものとして, 話をもっと聞いて欲しい, 自らの認知機能改善に貢献して欲しいと 2 つの意見が得られた. このことから本研究では, 決められたテーマについて話す内容を予め準備し, 制限時間内に話すことで, 認知機能に直結するとされている体験記憶, 注意分割, 計画力を訓練する手法である共想法 [大武 12] を基にしたタスクをロボットで行うことにした. 巡学ロボットでは, グループ内を 4 周することで 1 セット分の共想法が行えるように適応する.

1 周目 ユーザがロボットに対して写真を見せつつ, 制限時間内に決められたテーマについて話し, 巡学ロボットはその話と写真を覚える.

2 周目 巡学ロボットが 1 周目に覚えた話を記録した写真とともに他のユーザに行い, ユーザはそれに対して質問する. この時, ロボットは質問内容を覚える.

3 周目 2 周目で集めてきた質問をユーザに行い, 解答を得る.

4 周目 質問に返ってきた解答を伝える.

以上の 1 連の流れを共想法の巡学アレンジとする. 巡学ロボットでは, これを複数セット行うことを想定する. 巡学ロボットは各周において, 感情表出量を見ながら独自に対話を行う. これを実現するために, 巡学ロボットでは話を聴いて質問する傾聴対話システムの構築と話を再現する手法が必要となる. 本稿では, 話を聴いて質問する傾聴対話システムの構築に焦点を当てる.

様々な傾聴対話システムの研究が行われてきた [大竹 12]. 下岡らは高齢者を対象とし, 回想法を行うことを目的とした傾聴対話システムの開発を行っている [下岡 17]. 高齢者の心のケア, 認知機能のケアという点に焦点を当てたシステム開発という点では本研究の目的の 1 つと合致する. しかし, これらの研究はユーザとシステムとの 1 対 1 の対話にのみ焦点をあて, ロボットの成長がユーザに与える影響やロボットの表出内容から他者との間接コミュニケーションを促すものではない. そこで本研究では, これらの研究を参考に対話システムを構築し, ユーザがどういう時に感情を表出するのかを発話内容から学習し, ロボットが徐々に共感するようになることを目的とする.

3. 感情表出量による質問選択アルゴリズム

本章では, まず巡学ロボットとのインタラクションタスクについて簡単に述べた後, 感情表出量の算出及びそれによるユーザのクラス分類について提案し, 巡学ロボットの学習結果とクラス分類の結果から対話内容を選択する手法を提案する. 感情表出量の算出とユーザのクラス分類の流れを図 2 に, 対話選択アルゴリズムの概略図を図 3 に示す.

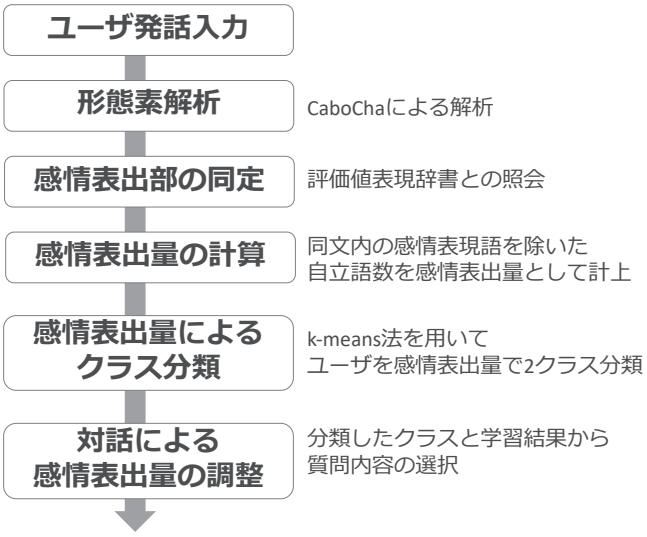


図 2: 感情表出量の算出とユーザのクラス分類の流れ

3.1 感情表出量

発話内容から感情を推定する研究の中で, 感情表現と共に生起されやすい確率から感情の強弱を決定する手法やポジティブとネガティブの極性を評価する手法 [乾 06] はあるが, 量に着目したものはない. 文章から感情が想起された要因を抽出するには, 動詞と名詞の間の意味的関係を表現した深層格を同定する必要がある. しかし, 日本語において深層格の設定方法について、統一した見解はない. そこで, 本研究では, ユーザが発

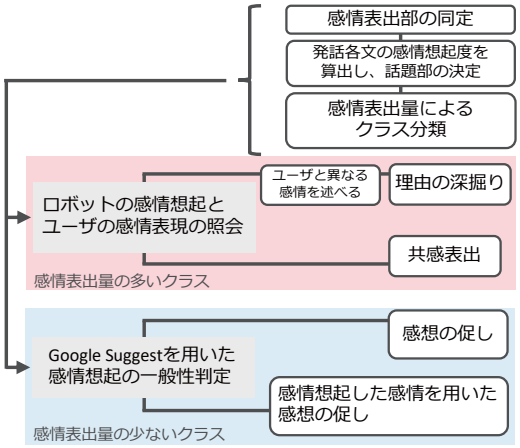


図 3: 対話選択アルゴリズムの概略図

話する全てが感情の想起に大小問わず影響を与えるものとし, 感情表出量を評価値表現辞書 [小林 05][小林 06] 内に存在する語と同一文内にある自立語の総数と定義した.

3.2 感情表出量によるクラス分類

k-means 法を用いて, 感情表出量によりユーザを 2 クラスに分類する. このクラス分類の結果から, 質問戦略を決定する.

3.3 質問による感情表出量の調整

既存研究ではユーザに共感を表す研究は話者の感情を推定し, 同調することを目的としたものである [下岡 17]. しかし, 本研究では必ずしも同調することを目的とせず, 巡学ロボットの他ユーザはどう思うのかを巡学ロボットを通して理解することも重要な目的としている. そこで, 本研究では巡学ロボットユーザ達から感情表出表現を学習することで, ユーザグループ化されたロボットの感情想起部を実装する.

3.3.1 ユーザグループ化されたロボットの感情想起部

ロボットの感情想起というと, システム独自の感情を決定し行動選択する研究が多いが, 本研究では各ユーザが表出した感情表現を基に, ある文からユーザグループが最も想起されやすい感情を算出するという点で異なるものである.

ユーザ発話から, 内容を形態素解析にかけ, 自立語を対象とした感情表現語句との共起確率を学習する. そして, 学習結果を用いてユーザの発話内容から感情が想起されやすさを算出する. ユーザ発話 1 文における感情の想起されやすさを感情想起値とすると, 感情想起値 C は, 1 文中の形態素数を n, ある形態素を a, a に共起する感情表現の数を m, ある感情表現を b, 形態素 a と感情表現 b が共起する条件付き確率を $P(b_j|a_i)$ とすると,

$$C = \sum_{i=1}^{n} \sum_{j=1}^{m} P(b_j|a_i)$$

で表すことにする. ただし, どの感情表現も共起しない確率については負の値として扱うことにする. この文中の形態素全ての感情表現との共起確率の総和を感情が想起されやすさ, 感情想起値とする.

3.3.2 話題内容の決定

話題内容の決定では, まずユーザの発話各文をロボットの感情想起値 C が大きい順にランク付けする. 次に, 検索エンジン Google での検索結果数と, 2 つの集合の類似度を計算する Simpson 係数を用いて, 各文内の自立語と最も共起されやすい感情表現語句との Simpson 係数を算出する. ネット検索を用いたのは, ユーザの発話内容のみから話題選択基準を設けると, データ数の少ない領域の話題に対応することが難しいと考えたためである. ある文中の自立語 a の検索結果数を N_a, 感情表現語句 b の検索結果数を N_b, a と b をキーワードとして検索した際の検索結果数を N_{ab} とすると感情後検索結果 Simpson 係数 S は以下の式で表される.

$$S = \frac{N_{ab}}{min(N_a, N_b)}$$

Simpson 係数で発話文毎に比較した際, 上位 3 割に入る文の中で感情想起値 C が大きい順に話題内容として決定する. 決定プロセスの簡単な流れは以下の通りである。

1. 感情想起値 C の値が高い順に発話各文をランク付け
2. Google Suggest による各文の感情想起された感情と同文内の形態素毎の Simpson 係数の総和でランク付け
3. Simpson 係数の総和が上位 3 割に入った文の中で、感情想起値 C の値が最も高いものから話題にする。

3.3.3 質問選択アルゴリズム

先に分類した感情表出量のクラスによって, ロボットが行う質問を選択することでユーザ間の感情表出量を調整する. なお発話する内容については動的に生成せず, テンプレート文に当てはめることとした.

感情表出量の多いクラス

感情表出量の多いクラスでは, 話題部でユーザが表出した感情とロボットの感情想起が一致するかによって共感をするかを決定する. 共感の発話内容に関しては, 感情表現語句を当てはめるテンプレート文から選択する. 一致しない場合には, ロボットの感情想起値で最も高い値の感情を表出し, なぜその感情を抱いたのかを質問する.

感情表出量の少ないクラス

話題内容に決定した文を繰り返し発話し, その時に抱いた感情について尋ねることで感情表出を促す. この時に, 検索エンジン Google の API である, Google Suggest を用いて話題文中の自立語毎に上位 10 位と評価値表現辞書を照会する. 一致した場合には, その自立語と感情表現語句を当てはめたテンプレート文を表出し, 感想を促す. これは, まだ感情表現を行っていない, つまりロボットに対してあまり親密でないユーザに, なるべくニュートラルな印象を与えることで対話への偏見を与えないようにすることを目的とした. また, 一致しない場合には, 話題文を繰り返した後にどう思ったかを単に問うこととした.

3.3.4 話題数の決定

3.2 節で述べた, クラス分類結果を基に感情表現の少ないクラスのユーザに対する, ロボットが取り上げる話題数を決定する. ユーザの感情表現量と感情表現の多いクラス重心距離を, ユーザの 1 質問に対する平均解答感情表現量で割る. 小数点以下については切り上げることとし, 話題数を決定する.

4. まとめ

本稿では, 予備調査で明確になった巡学ロボットの課題に対して関連研究を交えて解決手法を論じ, ユーザ間の感情表出量の均等化に向けた対話選択手法を論じた. 今後は, 実際に高齢者の方に使用してもらい, 感情表出量の均等化の有効性を検討する. 検討結果次第では, この対話選択アルゴリズムによる効果を検証するために本稿では取り組まなかった, 発話内容の動的生成などにも焦点を当てる予定である.

謝辞

本研究は, 科研費・基盤研究 (A)「子どもと高齢者の教育的コミュニケーションを支援するロボットインタフェース（課題番号 15H01708)」の支援を受けて行われた. また, 予備調査にご協力いただいた参加者の皆様に心より感謝する.

参考文献

[小川 03] 小川一美 : 二者間発話量の均衡が観察者が抱く会話者と会話に対する印象に及ぼす効果, 実験社会心理学研究, Vol. 43, No. 1, pp. 63-74 (2003)

[乾 06] 乾孝司, 奥村学 : テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol. 13, No. 3, pp. 201-241 (2006)

[大竹 12] 大竹裕也, 萩原将文 : 高齢者のための発話意図を考慮した対話システム, 日本感性工学会論文誌, Vol. 11, No. 2, pp. 207-214 (2012)

[大武 12] 大武美保子 : 介護に役立つ共想法: 認知症の予防と回復のための新しいコミュニケーション, 中央法規出版 (2012)

[笠井 17] 笠井翼, 田中文英 : 巡学ロボットの開発とその学習アルゴリズムの一検討, 第 31 回人工知能学会全国大会 OS-3 世代をつなぐ知的インタフェース (2017)

[小林 05] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一 : 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 3, pp. 203-222 (2005)

[小林 06] 小林のぞみ, 乾健太郎, 松本裕治 : 意見情報の抽出/構造化のタスク仕様に関する考察, 情報処理学会研究報告自然言語処理 (NL), Vol. 2006, No. 1 (2006-NL-171), pp. 111-118 (2006)

[櫻井 11] 櫻井茂男, 葉山大地, 鈴木高志, 倉住友恵, 萩原俊彦, 鈴木みゆき, 大内晶子, 及川千都子 : 他者のポジティブ感情への共感的感情反応と向社会的行動, 攻撃行動との関係, 心理学研究, Vol. 82, No. 2, pp. 123-131 (2011)

[下岡 17] 下岡 和也, 徳久 良子, 吉村 貴克, 星野 博之, 渡部 生聖 : 音声対話ロボットのための傾聴システムの開発, 自然言語処理, Vol. 24, No. 1, pp. 3-47 (2017)

[山野 08] 山野美咲, 薄井達也, 橋本稔 : 情動同調に基づく人間とロボットのインタラクション手法の提案, HAI シンポジウム (2008)

[MeCab] 日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクト京都大学情報学研究科, “ Mecab: Yet another part-of-speechand morphologicalanalyzer,” http://taku910.github.io/mecab/（2018 年 2 月 参照）

[Leite 13] Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A.: The influence of empathy in human-robot relations, International journal of human-computer studies, Vol. 71, No. 3, pp. 250-260 (2013)

[Rimé 92] Rimé, B., Philippot, P., Boca, S., and Mesquita, B.: Long-lasting cognitive and social consequences of emotion: Social sharing and rumination, European review of social psychology, Vol. 3, No. 1, pp. 225-258 (1992)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

複合現実を用いた運転行動の改善を促すロボットシステムのための入力インタフェース

Input Interface for Robot System Using Mixed Reality that Encourages Retrospect of Driving

相川裕哉 *1 Yuya Aikawa
加納政芳 *1 Masayoshi Kanoh
ジメネスフェリックス *2 Felix Jimenez
早瀬光浩 *3 Mitsuhiro Hayase
田中貴紘 *2 Takahiro Tanaka
金森等 *2 Hitoshi Kanamori

*1中京大学 Chukyo University
*2名古屋大学 Nagoya University
*3豊橋創造大学 Toyohashi Sozo University

In this research, we have been developing a robot system that encourages improvement of driving behavior at home. Mixed reality technology is used to implement driving reflection function for driving behavior improvement in this system. An existing mixed reality system uses a special input interface, therefore smooth operation is expected to be difficult. In this paper, we propose two interfaces that are made with reference to swipe behavior and laser pointer, respectively. The input interface that is imitated laser pointer indicated good perfomance in the experiments.

1. はじめに

近年、交通事故の発生件数は減少傾向にあるが、高齢者ドライバの交通事故率は年々増加している。この問題に対し、高齢者が自身の運転について見直すことや、運転を振り返る機会が必要であると考える。現状、高齢者は教習所に行き運転講習を受けることで自身の運転行動を認識することができる。しかし、運転行動を改善するためには、自身の運転に対する問題点を繰り返し学ぶ必要がある。そのため、講習を何度も受ける必要があるが、決められた時間に教習所に行かなければならなかったり、拘束時間が長かったりと高齢者の負担が大きいことが課題である。また、運転時の問題点を学ぶ際には、指導者はその都度指示する必要があるため、指導する側のコストも大きくなってしまう。これらのことから、教習所に行かずに高齢者が一人で運転行動の振り返りを行うことができるシステムが求められる。ここで、運転行動は指導によって改善するが、その効果は一時的であり、指導前の運転行動へすぐに戻る可能性がある [1, 2] 。コーチング理論に基づくトレーニング手法 [3] では、ドライバに自己の運転行動を認識させ、分析し、改善するというプロセスを繰り返し行う必要があるとされている。このプロセスを教習所の講習ではなく、ロボットが提示することができれば、高齢者は一人で運転行動の改善を行うことができる。本研究では、在宅時に運転行動の変容を促すロボットシステムを開発している。本システムでは、ロボットが高齢者の運転を解析して、危険な場面を提示し、高齢者がロボットから提示された場面を振り返ることで、運転行動の認識を行う。このとき、ロボットが危険な状況に対してどのような行動をとればよいのかなどの設問を含む会話を行うことで、高齢者は運転の教示を受けながら運転行動の分析と改善を行うことができる。このロボットによる運転の振り返りプロセスを自宅で行うことで、高齢者の負担を大きく減らすことができる。本システムでは、運転時の映像を見ながらロボットの教示を受け、設問に回答する機能を実装するため、複合現実の技術を用いる。ここで重要となるのが、与えられた設問に回答する際のインタフェースである。認知的負荷を可能な限り少なく運転の振り返りを行うためには、より簡単な入力インタフェースが求められるが、従来のインタフェースでは特殊な入力方法が採用されており、円滑な操作が困難であると予想される。そこで本稿では、ロボットシステムに複合現実を用いた際の二種類の入力インタフェースについて提案し、その評価を行う。

連絡先: 相川裕哉，中京大学，名古屋市昭和区八事本町 101-2，052-835-7111，t11701m@m.chukyo-u.ac.jp

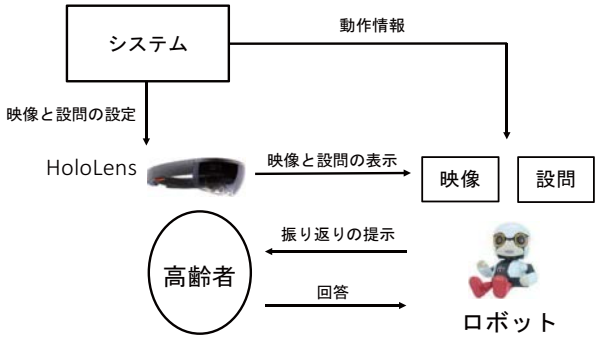


図 1: システム概要

2. システム概要

本稿では、複合現実の環境構築に Microsoft 社製 HoloLens を使用する。HoloLens は頭部に装着するメガネ型の wireless 計算機デバイスである。視界前の透明なレンズに仮想空間を映し出すことができ、仮想と現実とを融合した世界が体験できる。

図 1 にシステムの概要を示す。システムはドライブレコーダーの映像から危険と判断した場面の映像と、その場面にとるべき行動を問題形式にして HoloLens に送信する。HoloLens は送られてきた映像と設問をレンズに表示する。また、表示される映像と設問の直感的な理解を促すために、ロボットへ動作の情報を送信する。高齢者は HoloLens に表示される映像を見ながら、ロボットからの教示を受け、設問に回答することで運転の振り返りを行う。

図 2 に高齢者の視点からロボットと映像を見たときのイメー

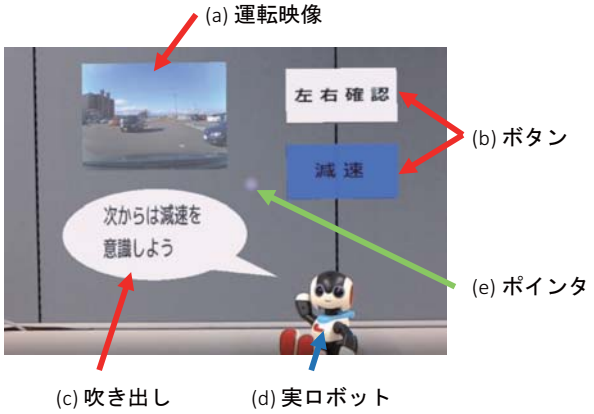


図 2: ユーザ視点からのイメージ図

ジ図を示す。(a) 運転映像、(b) ボタン、(c) 吹き出しは複合現実によって提示する。これらはコンピュータグラフィックスで作成されている。高齢者は、(a) に提示された映像を見つつ、ロボットの質問に対して、(b) に表示されているボタンを操作することで運転時の振り返りを行う。(c) の吹き出しは、高齢者がロボットの発話を聞き取れないときのための補助的な情報提示インタフェースである。(d) のロボットは提示している映像やロボットの発話内容、ボタンの押下状況をシステムと共有する。

3. インタフェース

本稿では、HoloLens に標準搭載されているインタフェースを「ジェスチャ」と呼ぶ。また、提案する二つのインタフェースを「スワイプ」「レーザー」と名付ける。各インタフェースによる本システムのボタン操作方法について説明する。

3.1 ジェスチャ

HoloLens を装着すると、装着者の頭部の向いている方向にポインタ（図 1 (e)）が設定される。デバイス、すなわち頭部を動かすと、このポインタも連動して動く仕組みになっている。ボタンを押すには、まず、ポインタをボタンの位置まで頭部を動かして移動させ、次に、視界内で親指と人差し指をつまむ動作を行う。このインタフェースには二つの問題点が考えられる。一つ目は、ポインタを移動させるために、頭部を回転させ、ポインタを合わせた後に固定し続けなければならない点である。この動作は視線を合わせるのではなく、頭部の中心位置にあるポインタを合わせるという普段の生活で行わない動作のため、慣れるまでに時間がかかる可能性がある。二つ目は、ポインタをボタンに当てた後に、手を一定の高さまで上げ続ける必要がある点である。ボタンを押すたびに手を挙げなければならず疲労する恐れがある。

3.2 スワイプ

この手法はタブレット端末の入力に用いられるスワイプを参考にしたインタフェースである。タブレット端末の普及により手を小さく動かす動作を行う機会が増えた。ボタンを直接触るような動作であれば、直感的な操作感覚が期待できる。この手法でボタンを押すには、選択するボタンに右手をかざし、その状態から腕を右に振る動作を行う。

3.3 レーザー

この手法はレーザーポインタを模したインタフェースである。レーザーポインタは指差しの要領でポインタを操作することができる。このことから、頭部を動かすことよりも簡単にポインタを操作することができると考える。この手法では、腕からレーザーポインタが放出されている設定になっている。このポインタをボタンに合わせ、手の平を反す、もしくは拳を握ることでボタンを押す動作となる。

提案する二つのインタフェースを作成するにあたって、腕の動きを検出するために、アームバンド型のデバイスである MYO を使用する。MYO はユーザの上腕部に装着するデバイスで、ジャイロセンサによって腕の傾きを、内側の電極パッドにより腕の筋電を計測する。

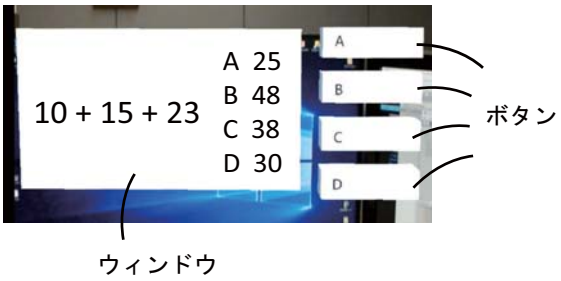


図 3: 実験環境

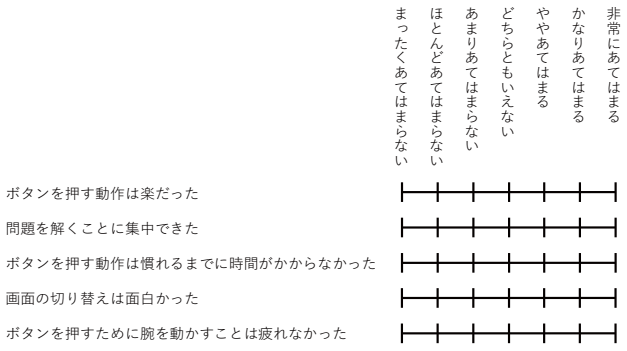


図 4: アンケート

4. 実験

本システムは、コンピュータグラフィックスで作成したウィンドウに表示される映像を見ながら問題に答えることで運転の振り返りを行うものである。実験ではこの状況と類似した環境を構築し、インタフェースを評価する。図 3 に実験環境中の視界に表示されるコンピュータグラフィックスを示す。被験者はウィンドウに表示される計算問題を解き、四つの選択肢の中から正解を選ぶ。図 3 の例では答えが「48」なので、B が正解である。被験者はウィンドウの隣にある B のボタンを押すことで回答を行う。被験者には 15 問の問題を解く課題をすべての手法に対して行う。手法を体験する順序は被験者ごとにランダムとし、カウンターバランスをとる。各手法の体験後にアンケートを行う。

図 4 に実験に使用したアンケートを示す。ボタン押下に関する質問 3 項目、システム評価に関する質問 2 項目で構成されており、7 段階リッカートスケールで評価する。被験者は大学生 20 人とする。

また、各手法体験中にボタンを操作するために必要な時間の指標として、(1) 計算問題を 15 問回答するまでに必要となった時間（回答時間）、(2) 各手法のボタンが押せる状態にあった時間（押下可能時間）、(3) 被験者が実験中にウィンドウの

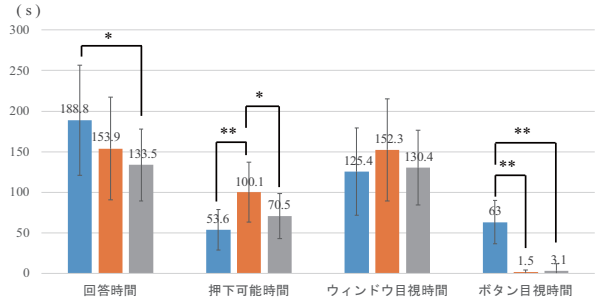


図 5: 計測結果

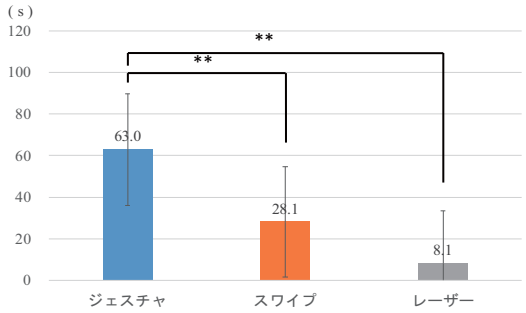


図 6: ボタン操作総時間の結果

方向を向いていた時間（ウィンドウ目視時間）、および (4) ボタンの方向を向いていた時間（ボタン目視時間）をそれぞれ計測した。

5. 結果と考察

図 5 に計測結果を示す。縦軸は時間（秒）である。各項目に表示されているバーは左からジェスチャ、スワイプ、レーザーを示している。回答時間では、ジェスチャとレーザーの間に有意水準 5%で有意差が認められた。押下可能時間では、ジェスチャとスワイプの間に有意水準 1%、スワイプとレーザーの間に有意水準 5%で有意差が認められた。ウィンドウ目視時間は有意差は認められなかったが、ボタン目視時間はジェスチャとスワイプ、ジェスチャとレーザーの間に有意水準 5%の有意差が認められた。

まず、ボタン目視時間に注目する。ジェスチャはボタンを押す際、必ずボタンの方向を向く必要があるが、二つの提案手法は目視だけでもボタンを押すことができると考える。実際に、計測結果を見るとジェスチャの平均 63 秒に対し、二つの提案手法は非常に短い目視時間となっている。このことから、二つの提案手法ではボタンを目視して押下しているといえる。上記のことから、提案手法ではウィンドウ目視時間の中に、計算問題を解いている時間とボタンを操作している時間が含まれていることになる。そこで、ジェスチャのウィンドウ目視時間を回答時間（125.4 秒）の基準と設定し、スワイプ、レーザーのウィンドウ目視時間から基準とした回答時間、すなわちジェスチャのウィンドウ目視時間を除くことで、ボタンを操作している時間（ボタン操作時間）を測る。このボタン操作時間にボタン目視時間を加えることで、三つの手法におけるボタンを操作するために必要となった時間（ボタン操作総時間）を比較する。

図 6 にボタン操作総時間を示す。ジェスチャが 63 秒に対し、スワイプは 28.1 秒、レーザーは 8.1 秒となりジェスチャとスワイプ、ジェスチャとレーザーの間に有意水準 1%の有意差が認められた。このことから、二つの提案手法はジェスチャと比較し、有意に短い時間でボタンを操作できるといえる。

表 1: 平均正答率

ジェスチャ	スワイプ	レーザー
92%	72%	95%

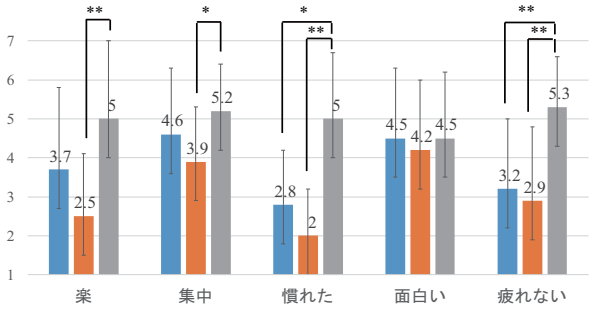


図 7: アンケート結果

ここで、ジェスチャはデバイスの前にあるポインタをボタンに当てる解答方法のため、ボタンの押し間違えは起こりにくいが、二つの提案手法では腕の動作のずれによりボタンを押し間違えてしまう可能性があるため、正解のボタンを押すことができた回数を計測した。

表 1 に平均正答数を示す。本実験では単純な四則演算のみ出題したため、正答率が著しく低くなることは想定されにくい。にもかかわらず、スワイプの正答率が低くなっており、ボタンの押し間違いが多発したものと考えられる。一方、ジェスチャとレーザーはほぼ同程度の正答率となっており、レーザーは正確にボタンを操作できるといえる。

最後に、図 7 にアンケート結果を示す。すべての項目で提案手法のレーザーが最も高い評価を得ることができた。

以上の結果から、ボタン操作に適したインタフェースはレーザーであるといえる。

6. おわりに

本稿では、在宅時に運転行動の改善を促すためのロボットシステムについて報告した。本システムではロボットと複合現実映像を同一視界内で提示するが、その操作インタフェースとして二つの手法を提案した。実験の結果、レーザーポインタを模した手法が最も操作することに適していることが示された。今後は、ロボットシステムに複合現実を用いた際の影響について調査する。

本研究は、名古屋大学エージェントを介した運転支援プロジェクトの支援を受けた。

参考文献

[1] T Tanaka et al.: Driver Agent for Encouraging Safe Driving Behavior for the Elderly, 5th International Conference on Human-Agent Interaction, 2017.

[2] 田中 他: 高齢ドライバ支援エージェントの提案 –運転指導員による指導方法の分析–, 第 31 回ファジィシステムシンポジウム, 2015.

[3] I. Edwards: ドライバーのための自分づくり教育 –コーチングのすすめ–, SMA サポート株式会社, 2014.

Oral presentation | Organized Session | [Organized Session] OS-17

[2C3-OS-17] [Organized Session] OS-17

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room C (4F Orchid)

[2C3-OS-17-02] Preference Prediction using User/Property Features and Floorplan Images

○Naoki Kato[1], Toshihiko Yamasaki[1], Kiyoharu Aizawa[1], Takemi Ohama[2] （1. The University of Tokyo, 2. ietty Co., Ltd.）

3:40 PM - 4:00 PM

[2C3-OS-17-03] Extraction of User Latent Preference for Real Estate Website

○Kenshin Ikegami[1], Hironori Ito[2], Shimpei Nomura[1] （1. Recruit Sumai Company Ltd., 2. Graduate School of Informatics, Kyoto University）

4:00 PM - 4:20 PM

[2C3-OS-17-04] Multivariate analysis of the transition probability of an occupation of rental rooms by using the housing information website data

○Hayafumi Watanabe[2], Yu Ichifuji[3], Masato Suzuki[4], Satoshi Yamashita[1] （1. The institute of statistical mathematics, 2. Research Organization of Information and Systems, 3. Nagasaki University, 4. UD Asset Valuation Co., Ltd.）

4:20 PM - 4:40 PM

[2C3-OS-17-05] Classification of Real Estate Images using Bottleneck Features

○Suguru Toyohara[1], Youiti Kado[2], Toshihiko Yamasaki[3], Susumu Fujimori[1], Ikuo Tahara[1] （1. Tokyo University of Science, 2. At Home Co., Ltd., 3. The University of Tokyo）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ユーザ・物件特徴と間取り特徴を用いた嗜好予測

Preference Prediction using User/Property Features and Floorplan Images

加藤尚輝 *1 Naoki Kato　山崎俊彦 *1 Toshihiko Yamasaki　相澤清晴 *1 Kiyoharu Aizawa　大浜毅美 *2 Takemi Ohama

*1東京大学 The University of Tokyo　*2株式会社 ietty ietty Co., Ltd.

Due to the advancement of E-commerce in recent years, recommendation for not only mass-produced daily items but also special items that are not mass-produced is an important task. In this research, we present an algorithm for real estate recommendation. There is no identical property in the world, properties already occupied by someone else can not be recommended, and users rent or buy properties only a few times in their lives. Therefore, automatic recommendation of property is one of the most difficult tasks. As the first step of property recommendation, we predict users' preference for properties by combining content-based filtering and multilayer perceptron (MLP). In the MLP, we used not only attribute data of users and properties but also deep features extracted from floorplan images of properties. As a result, we succeeded in predicting preference with accuracy of 60.7%.

1. はじめに

近年，オンラインサービスの普及に伴い E-commerce の利用者は増加傾向にある．E-commerce のウェブサイトでは商品の推薦が行われていることが多く，近年のデータの大規模化によりその精度は向上傾向にある．推薦システムには大きく分けてコンテンツベースフィルタリングと Goldberg ら [Goldberg 92] が提唱した協調フィルタリングの 2 種類がある．しかし，これらはアイテムが量産されているような場合にしか利用できず，各アイテムがこの世に 1 つしかないような分野，例えば不動産業界では不動産物件の推薦を行うのが困難である．

そのような状況の中，不動産業界では 2013 年以降，Real Estate Tech（不動産テック）とも呼ばれる不動産テクノロジー企業が急成長し始め，日本でもそのような企業が相次いで登場している．また，国土交通省もネット上での不動産業の規制緩和に向けて社会実験を行う等，積極的に不動産業界に IT を取り入れる動きがある．そのため，ウェブサイト上での物件の推薦は重要な課題の 1 つとなっている．

本研究では，上述したような一般的な推薦システムをそのまま使うことのできない特殊なデータにおける推薦システムの実装を最終的な目標とし，その例として不動産物件の推薦を扱う．物件推薦の第一歩として，不動産の賃貸会社である ietty*1 が所有する，ユーザによる物件評価が含まれるデータセットを用いてユーザの物件に対する嗜好予測を行う．通常の不動産賃貸と異なり，ietty ではウェブサイト上でユーザに物件を推薦し，ユーザはその物件を評価して実際に内見したり賃貸契約したりするという方法を取っている．ユーザの物件に対する嗜好予測において有効な手法を提案できれば，物件の推薦も比較的容易に行うことができると考えられる．

2. 関連研究

2.1 推薦に関する研究

協調フィルタリングは Goldberg ら [Goldberg 92] のメモリベースなものと，Breeze ら [Breese 98] が新たに提唱したモデルベースなものが存在する．また，Pennock [Pennock 00] はそれらを組み合わせたハイブリッド協調フィルタリングを考案し，Xue [Xue 05] はクラスタリングを用いたデータ補間によりハイブリッド協調フィルタリングの性能を向上させた．

一方，Simon [Func 06] は Netflix Prize*2 において，行列因子分解を用いたアルゴリズムにより推薦システムの性能を大きく向上させた．行列因子分解は式 (1) を満たすユーザ因子行列 p 及びアイテム因子行列 q を求めるというものである．尚，κ は train データとして存在する評価値 r_{ui} が既知であるようなユーザ u とアイテム i のペア (u, i) の集合であり，$\hat{r_{ui}} = q_i^T p_u$ がユーザ u に対するアイテム i の予測評価値になる．

$$\min_{q,p} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\|q_i\|^2 + \|p_u\|^2) \quad (1)$$

Simon [Func 06] は式 (2)(3)(4) を反復させる確率的勾配降下法を用いてこれを解いた．

$$e_{ui} = r_{ui} - q_i^T p_u \quad (2)$$

$$q_i \leftarrow q_i + \gamma \cdot (e_{ui} \cdot p_u - \lambda \cdot q_i) \quad (3)$$

$$p_u \leftarrow p_u + \gamma \cdot (e_{ui} \cdot q_i - \lambda \cdot p_u) \quad (4)$$

この Netflix Prize で優勝した Koren ら [Koren 09] は確率的勾配降下法だけでなく，Alternating Least Squares (ALS) という q_i と p_u を交互に最適化するアルゴリズムで解く方法も紹介し，行列因子分解がデータが高次元になると発生する次元の呪いを解消して高次元データでも精度の高い推薦を実現することを示した．

また，Hidasi ら [Hidasi 15] は RNN（再帰型ニューラルネットワーク）の一種である Gated Recurrent Unit (GRU) により，E-commerce のウェブサイト上でのユーザの過去のクリックデータを全て用いた推薦システムを提案している．しかし，データのスパース性が高い特殊なアイテムに対してこれらの手法はうまく機能しない．現在の不動産物件の自動推薦システムに関しては，ietty では各ユーザの希望条件に対して徐々に条件を緩和しながら物件の推薦を行う，ルールベースなアルゴリズムを適用するという方法をとっている．

連絡先: 加藤尚輝，kato@hal.t.u-tokyo.ac.jp

*1 https://ietty.me/

*2 http://www.netflixprize.com/

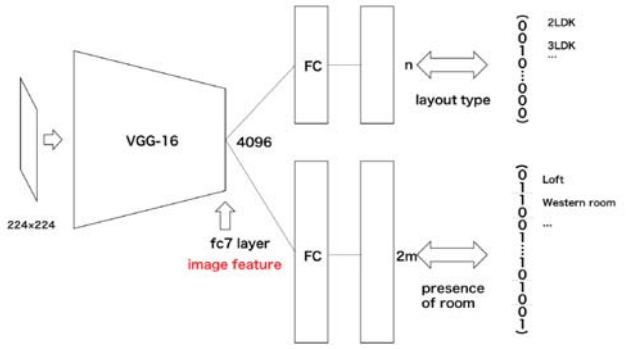


Fig. 1: 高田らの手法のネットワーク [Takada 18]

2.2 不動産物件の間取りの構造に関する研究

近年の深層学習の発展以前の物件の間取りの構造に関する研究は，間取り図をグラフで解析するというものであった．瀧澤ら [Takizawa 08] は，京都市内の 3LDK，3K，3DK の集合住宅の間取り図を用いて，部屋や廊下等をラベル付きノード，ドアやガラス等をラベル付きエッジとした隣接グラフによる物件の賃料の分析を行った．彼らは隣接グラフから部分グラフを抜き出し，共通な部分グラフの有無等から家賃推定を行い，その有効性を示した．しかし，隣接グラフによる間取りの構造の取得には，隣接グラフを作成するコストが非常に大きいという問題がある．

大原ら [Ohara 16] は間取りの共通部分グラフを用いてユーザの希望を反映した物件検索システムを開発した．また，高田ら [Takada 18] は大原ら [Ohara 16] の作成したデータセットを用いて，間取り図からグラフ構造を推定し，類似している間取り図の検索を行った．具体的には，Fig. 1 のネットワークを用い，間取りのタイプと部屋の有無の 2 種類の学習を同時に行う，マルチタスク学習によりモデルを作成し，そのモデルを使用して間取り画像から抽出した深層特徴を用いて検索を行った．

3. 提案手法

3.1 データセット

ietty ではウェブサイト上でユーザに不動産物件を数件推薦し，ユーザが推薦された各物件に対して「内見する，お気に入り，興味なし」を選択して評価できるようになっている．本研究では 2016 年から 2017 年にかけて連続した期間の評価データを取得し，それを間取り画像が存在する 220,094 件に絞り，3:1:1 の割合で train データ，validation データ，test データに分割して用いる．このデータセットに含まれるユーザ数は 19,538 人，物件数は 131,947 件である．

3.2 手法の概要

ユーザによる物件評価のうち「内見する，お気に入り」を positive，「興味なし」を negative と定義して，これらをユーザの物件に対する嗜好と呼称し，スパースなデータにも利用可能な手法によりこれを予測する．本研究では，ユーザ間類似度とアイテム間類似度を共に用いたコンテンツベースフィルタリングによる嗜好予測，物件の間取り画像から CNN（畳み込みニューラルネットワーク）を用いて抽出した深層特徴を入力に加えた MLP（多層パーセプトロン）による嗜好予測をそれぞれ行う．これらの手法は推薦対象のユーザの過去の評価データがなくても使用することができるため，スパースなデータに比較的強い．また，それらを組み合わせたシステムをハイブリッドフィルタリングと呼称し，それによる嗜好予測も行う．

物件の間取りのメタデータ（1DK，1K，1LDK 等）は物件の属性データに含まれているが，間取りの構造まで考慮することで嗜好予測の精度が上がることが期待される．例えば，同じ 2LDK でも部屋と部屋のつながり方が異なっていることがあり，幼い子供が勝手に外に出ないよう玄関と子供部屋が直接つながらないようにすることが可能な物件と，どの部屋からでも廊下に出られるような利便性の高い物件では推薦対象となるユーザの層も異なってくる．そのため，本研究では間取りの構造を考慮するために MLP に間取り画像の深層特徴を入力として加えることを提案した．以下の節で各手法の詳細を述べる．

3.3 コンテンツベースフィルタリング

ユーザ間の属性の類似度及び物件間の属性の類似度を共に用いたコンテンツベースフィルタリングにより嗜好予測を行う．ユーザの属性データを u，物件の属性データを i で表し，positive の評価値を w（train データ内の negative なデータの数と positive なデータの数の比），negative の評価値を -1 とする．test データ (u_{test}, i_{test}) に対し，評価値 r_{ui} が既知である train データ $(u,t) \in \kappa$ との，ユーザの属性に対するコサイン類似度 $\cos(u_{test}, u)$ を取り，u の類似度上位 k_u% の集合を U_{CBF} とする．同様に，物件の属性に対するコサイン類似度 $\cos(i_{test}, i)$ を取り，i の類似度上位 k_i% の集合を I_{CBF} とする．その後，式 (5) により test データ (u_{test}, i_{test}) の予測値 v_{CBF} を求める．

$$v_{CBF} = \sum_{u \in U_{CBF}} \sum_{i \in I_{CBF}} \cos(u_{test}, u) \cdot \cos(i_{test}, i) \cdot r_{ui} \quad (5)$$

予測値 v_{CBF} が 0 より大きいとき positive，0 以下のとき negative と予測する．

3.4 MLP

ユーザ・物件の属性データと，物件の間取り画像から CNN を用いて抽出した深層特徴を入力とした Fig. 2 のネットワークを用いて嗜好予測を行う．その際，dropout や batch normalization を用いて汎化性を向上させる．訓練時は train データと validation データを，テスト時は dropout を除いて test データを用いる．FC 最終層では negative，positive に対応した 2 次元ベクトル (x_n, x_p) が出力され，test データの予測値 $v_{MLP} = x_p - x_n$ が 0 より大きいとき positive，0 以下のとき negative と予測する．

間取りの深層特徴の抽出には，高田ら [Takada 18] の手法を改良して訓練を行ったモデルを用いる．具体的にはファインチューニングの際に VGG-16 の代わりに ResNet50 を使用し，データの増強として間取り画像をランダムに回転させて入力に用いる．また，データセットは高田ら [Takada 18] が訓練に用いたものと同一のものを使用し，20,000 枚の間取り画像を訓練に用いる．これにより作成した学習済みモデルを抽出器とし，pool5 層の 2,048 次元のベクトルを間取り画像の深層特徴量とする．

3.5 ハイブリッドフィルタリング

3.3，3.4 で求めた予測値 v_{CBF}，v_{MLP} の重み付き和を予測値 v_{HF} として嗜好予測を行う．概略図を Fig. 3 に示す．それぞれに積算する重み w_{CBF}，w_{MLP} として，本研究では 3.3，3.4 において test データの代わりに validation データを用いたときの予測値の集合の標準偏差 σ_{CBF}，σ_{MLP} の逆数を用い

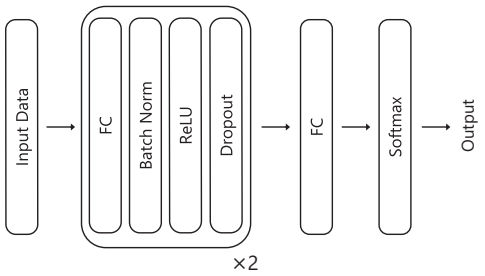


Fig. 2: MLP のネットワーク

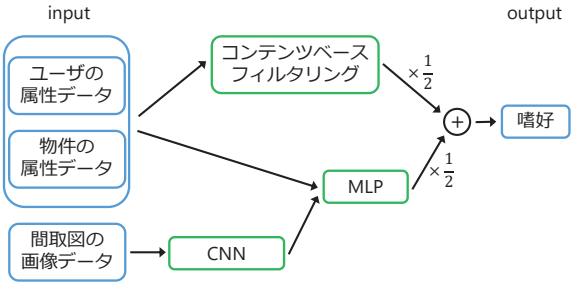


Fig. 3: ハイブリッドフィルタリングの概略図

る．式で表すと式 (6) のようになる．

$$v_{HF} = w_{CBF} \cdot v_{CBF} + w_{MLP} \cdot v_{MLP} = \frac{v_{CBF}}{\sigma_{CBF}} + \frac{v_{MLP}}{\sigma_{MLP}} \quad (6)$$

このようなスケーリングを行うことで 2 手法を同程度に重視した予測値 v_{HF} が求まる．他の手法同様，予測値 v_{HF} が 0 より大きいとき positive，0 以下のとき negative と予測する．

4. 実験

4.1 評価手法

全 220,094 件の評価データに対し，positive なデータは 76,871 件 (35%)，negative なデータは 143,223 件 (65%) 存在する．このようにデータセットに偏りがあるので，平等な評価をするために Matthews Correlation Coefficient (MCC) を評価に用いた．予測結果における true positive の数を TP，true negative の数を TN，false positive の数を FP，false negative の数を FN としたときの MCC は式 (7) により算出される．MCC の最大値は 1，最小値は-1 であり，大きい方が性能が良いということになる．

$$\mathrm{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

4.2 コンテンツベースフィルタリング

3.3 の手法において，train データのユーザの属性に対する類似度上位 $k_u (= 1, 10, 100\%)$，不動産物件の属性に対する類似度上位 $k_i (= 1, 10, 100\%)$ を用いたときのコンテンツベースフィルタリングの MCC は Fig. 4 のようになった．

ユーザ・物件共に train データの属性に対する類似度上位 10%を用いたときに最も性能が良く，MCC は 0.140 となった．k_u が小さすぎると参考にできる train データの数が極端に減ってしまってうまく予測し難く，逆に大きすぎると属性が全く類似していないユーザを参考にしてしまうためにうまく予測できないということが見て取れる．その一方で k_i については，

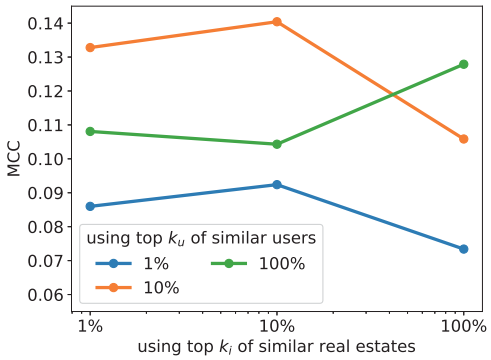


Fig. 4: コンテンツベースフィルタリングによる予測結果

Table 1: MLP による予測結果

Input	w/o image feature	w/ image feature (pre-trained)	w/ image feature (fine-tuned)
MCC	0.127	0.142	**0.149**
Accuracy	0.612	0.616	**0.619**
Precision	0.439	0.446	**0.451**
Recall	0.396	0.418	**0.423**
Confusion matrix $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$	$\begin{pmatrix} 20843 & 7802 \\ 9281 & 6094 \end{pmatrix}$	$\begin{pmatrix} 20678 & 7967 \\ 8954 & 6421 \end{pmatrix}$	$\begin{pmatrix} 20725 & 7920 \\ 8866 & 6509 \end{pmatrix}$

k_u が小さいとき（$k_u = 1, 10\%$）は同様に小さすぎず大きすぎない値を選ぶのが良さそうであるものの，$k_u = 100\%$，つまり train データ内の全てのユーザの情報を用いる場合は train データ内の物件の情報も全て用いる方が性能が良くなるという結果となった．

4.3 MLP

3.4 の手法において，以下の 3 通りの入力を用いた場合の MLP による嗜好予測を行った．

w/o image feature:
ユーザ・物件の属性データだけを用いた場合

w/ image feature (pre-trained):
ユーザ・物件の属性データだけでなく，ImageNet の学習済みモデルにより間取り画像から抽出した深層特徴も用いた場合

w/ image feature (fine-tuned):
ユーザ・物件の属性データだけでなく，3.4 においてファインチューニングを行ったモデルにより間取り画像から抽出した深層特徴も用いた場合

このときの嗜好予測結果は Table 1 のようになった．

3 通りの入力の MLP の中では，間取り画像の深層特徴をファインチューニングを行ったモデルにより抽出して入力に加えた MLP の性能が最も良く，MCC は 0.149 となった．間取り画像の深層特徴を入力に加えたことにより，MLP の性能が上昇していることがわかる．このことから，間取りの構造を考慮することは，ユーザの物件に対する嗜好予測において有効であると考えられる．また，間取り画像の深層特徴の抽出器として，ImageNet の学習済みモデルを用いるより，ファイン

Table 2: ハイブリッドフィルタリングによる予測結果

Input	w/o image feature	w/ image feature (pre-trained)	w/ image feature (fine-tuned)
MCC	0.150	0.159	**0.166**
Accuracy	0.600	0.604	**0.607**
Precision	0.437	0.442	**0.446**
Recall	0.500	0.507	**0.514**
Confusion matrix $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$	$\begin{pmatrix} 18748 & 9897 \\ 7693 & 7682 \end{pmatrix}$	$\begin{pmatrix} 18796 & 9849 \\ 7579 & 7796 \end{pmatrix}$	$\begin{pmatrix} 18823 & 9822 \\ 7479 & 7876 \end{pmatrix}$

チューニングを行ったモデルを用いた方が MLP の性能は高くなることもわかる．このことから，ファインチューニングによって間取りの構造の特徴がうまく取れるようなより良い深層特徴が抽出できていると考えられる．

4.4 ハイブリッドフィルタリング

3.5 の手法において，4.3 と同じ 3 通りの入力を用いたときの嗜好予測結果は Table 2 のようになった．

前述した全ての予測システムの中で最も性能が良いのは，間取り画像の深層特徴をファインチューニングを行ったモデルにより抽出して入力に加えたハイブリッドフィルタリングであり，MCC は 0.166 となった．ハイブリッドフィルタリング内での比較としては，3 通りの入力による予測性能の傾向は 4.3 と同様であった．また，MLP との比較を行うと，全体的に予測性能がかなり向上していることがわかる．このことから，コンテンツベースフィルタリングと MLP を組み合わせて用いることにより，機械学習におけるアンサンブルのように汎化性能が向上する効果があると考えられる．

5. 結論

量産されていないような特殊なアイテムに対する推薦システムの構築を目的とし，その第一歩としてユーザ間類似度とアイテム間類似度を共に用いたコンテンツベースフィルタリング，物件の間取り画像から CNN を用いて抽出した深層特徴を入力に加えた MLP，それらを組み合わせたハイブリッドフィルタリングによってユーザの不動産物件に対する嗜好予測を行った．その結果，間取り画像の深層特徴をファインチューニングを行ったモデルにより抽出して入力に加えたハイブリッドフィルタリングが最も優秀であることが示された．

本研究の課題は大きく分けて 2 つ存在する．まず，本研究では間取り図の深層特徴を抽出する際に高田ら [Takada 18] の手法を用いたが，他の手法との比較・検討を行うことができていない．例えば，Liu ら [Liu 17] は深層学習を利用してラスター形式の間取り画像からベクトル形式の間取り構造を取り出すことに成功している．この手法において入力画像から junction layer を取り出す際に使用するネットワークを本研究の提案手法に応用することが考えられる．次に，実際に物件推薦に提案手法を用いる際は処理の速さも重要になってくるが，本研究で用いたコンテンツベースフィルタリングは計算量が非常に多くなりやすいものであるため，より効率的な設計をしなければ実際の物件推薦に用いることはできない．また，物件を推薦する毎に間取り画像を読み込んで抽出器に通すのも時間的なコストが大きく，予め深層特徴を抽出しておく等の対策が必要となる．これらを解決できれば，本研究の提案手法を実際の推薦に用いることができるため，これらを今後の課題としたい．

参考文献

[Breese 98] Breese, J. S., Heckerman, D., and Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering, in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 43–52Morgan Kaufmann Publishers Inc. (1998)

[Func 06] Func, S.: Netflix Update: Try This at Home (2006), `http://sifter.org/simon/journal/20061211.html`

[Goldberg 92] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D.: Using Collaborative Filtering to Weave an Information Tapestry, *Commun. ACM*, Vol. 35, No. 12, pp. 61–70 (1992)

[Hidasi 15] Hidasi, B., Karatzoglou, A., Baltrunas, L., and Tikk, D.: Session-based Recommendations with Recurrent Neural Networks, *CoRR*, Vol. abs/1511.06939, (2015)

[Koren 09] Koren, Y., Bell, R., Volinsky, C., et al.: Matrix factorization techniques for recommender systems, *Computer*, Vol. 42, No. 8, pp. 30–37 (2009)

[Liu 17] Liu, C., Wu, J., Kohli, P., and Furukawa, Y.: Raster-To-Vector: Revisiting Floorplan Transformation, in *The IEEE International Conference on Computer Vision (ICCV)* (2017)

[Ohara 16] Ohara, K., Yamasaki, T., and Aizawa, K.: An Intuitive System for Searching Apartments using Floor Plans and Areas of Rooms, *The 78th national convention of IPSJ* (2016), (in Japanese)

[Pennock 00] Pennock, D. M.: Collaborative filtering by personality diagnosis : A hybrid memory and model-based approach, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)* (2000)

[Takada 18] Takada, Y., Inoue, N., Yamasaki, T., and Aizawa, K.: Similar Floor Plan Retrieval Featuring Multi-Task Learning of Layout Type Classification and Room Presence Prediction, *IEEE International Conference on Consumer Electronics (ICCE)*, pp. 931–936 (2018)

[Takizawa 08] Takizawa, A., Yoshida, K., and Kato, N.: Applying graph mining to rent analysis considering room layouts, *Journal of Environmental Engineering*, Vol. 73, No. 623, pp. 139–146 (2008)

[Xue 05] Xue, G.-R.: Scalable collaborative filtering using cluster-based smoothing, *Proc. 28th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval* (2005)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

不動産ポータルサイトでのユーザの潜在的嗜好の抽出

Extraction of User Latent Preference for Real Estate Website

池上 顕真
Kenshin Ikegami

伊藤 博典
Hironori Ito

野村眞平
Shimpei Nomura

*1 株式会社リクルート住まいカンパニー
Recruit Sumai Company Ltd.

*2 京都大学情報学研究科
Graduate School of Informatics
Kyoto University

*1 株式会社リクルート住まいカンパニー
Recruit Sumai Company Ltd.

As a current trend, in the consumption of durable consumer goods such as cars, real-estate properties, increasing number of buyers are making their purchasing decisions based on information on the Internet. Despite that, since real-estate contraction process is barely common or frequent, more users (consumers) than not suffer from a lack of adequate knowledge. Consequently, searching for a house on-line could become confusing. Renting a house, specifically, often comes time restriction, and from times requires compromise in search terms. To help, search terms and listing recommendations have to be more properly generated according to each user's needs and preferences. In this analysis, we propose a method of user preference extraction and provide insights on the results.

1. はじめに

近年，車の購入や，不動産契約などの耐久消費財の商取引において，ユーザがウェブ上の情報を元にして購買の意思決定を行う機会は増加している．一方で，不動産契約は頻繁に行う取引ではないため，消費者であるユーザが十分な知識を持っていないことも多い．その結果，知識が十分でないことにより適切な探し方できず，特に賃貸物件においては選定期間に制約があることが多いため，妥協した選定をせざるを得ない場合がある．この問題を解決するためには個人に応じた検索条件や、物件の推薦をより適切に行う必要がある。今回の分析では，これらを実施するためにユーザの潜在的嗜好の抽出を行ったのでこの結果について報告を行う．

2. SUUMO について

2.1 SUUMO に蓄積されているデータ

不動産情報ポータルサイトである SUUMO には様々なデータが蓄積されている．各企業から提供される物件情報や施工事例といった不動産データを始めとして，各不動産会社のデータ，ユーザが SUUMO の Web サイトやスマートフォンアプリケーションを利用した際の行動履歴データ，SUUMO 独自で収集を行っている街情報や，住みたい街ランキングに代表されるアンケートなどの調査データをはじめ，様々なデータが存在している．本稿では，上記のうち不動産データと行動履歴についてそれぞれ簡単に紹介する．

§ 不動産データ

本稿における不動産データとは，主に不動産会社から提供される物件情報や施工事例などのデータのことを指す．例えば物件の緯度経度や最寄り駅などの地理情報や，その物件が何階に所在し，どのくらいの広さを持ち，どんな設備を持っているかといったスペックの情報，どのような物件かを視覚的に確認できる画像情報，そして現在から過去にかけていくらで掲載されていたのか，という価格情報など多岐にわたった情報が蓄積されている．

§ ユーザの行動履歴

行動履歴とは，ユーザが SUUMO の Web サイトやスマートフォンアプリケーションを利用した際のアクセスログを指す．cookie 毎にどのページにいつアクセスし，どんな検索を行い，どんな物件を比較し，どの不動産会社に問い合わせたのか，などの情報が含まれている．

3. 分析事例

3.1 分析の背景

SUUMO の Web サイト上では，ユーザは家賃などの検索条件を複数設定しながら，物件の検索を行うことができる．このようにユーザが明示的に指定する検索条件に対して，検索条件には設定していないが，ユーザが暗黙的に重視している物件の条件があると仮定する．このような条件のことを本稿ではユーザの潜在的嗜好と呼ぶ．ユーザの潜在的嗜好の抽出を行うことができれば，レコメンドシステムなどへの応用が可能になる．

3.2 先行研究

土方（土方, 2005）は，Web サイトを利用しているユーザの嗜好抽出を行う方法として，2種類の手法があると述べている．1 つはユーザに興味に関する情報を入力してもらう明示的な方法である．もう 1 つは，Web 閲覧時の挙動からユーザの興味に関する情報を取得する暗黙的手法である．暗黙的手法は，閲覧した情報から手がかりを利用してユーザが閲覧した情報への興味の有無を推定する手法である．SUUMO の Web サイトを利用しているユーザに当てはめると，閲覧した情報とは，ユーザが閲覧した物件のことを指し，手がかりとは物件を閲覧した時間などを指す．

明示的な手法を SUUMO の Web サイトの閲覧履歴に応用した例では，布施ら（布施, 2018）の分析がある．彼らは，ユーザが入力した複数検索条件を分散表現化することで，検索条件の可視化を行なった．

浅沼ら（浅沼, 2011）は暗黙的手法によってユーザのレシピ閲覧履歴からレシピの推薦システムの構築を行なった．彼らが提案する手法では，閲覧したレシピに使用される食材に注目し，食材を素性とした TF-IDF 表現でレシピを表現することにより，推薦するレシピのスコアリングを行なった．

3.3 提案手法

本項では, SUUMO の Web サイトの行動履歴を用いて, ユーザの潜在的嗜好を抽出する手法について述べる. 手法を考える上で, ユーザの物件閲覧行動には以下の傾向があると仮定した.

仮説 1: ユーザは物件に求める条件に対して内的優先順位を持つ

仮説 2: 複数の物件を閲覧していく中で, 物件に求める条件の内的優先順位は変動する

仮説 3: 物件に求める条件の内的優先順位は物件を閲覧するごとに変動するが, その中でも常に上位に存在する条件が存在する

ユーザが仮説 1, 仮説 2 に基づいて物件の閲覧を行なった際に, 仮説 3 に当てはまる物件の条件をユーザの潜在的嗜好とし, 条件の抽出を行なった. 具体的には以下の手順に基づいて潜在的嗜好を抽出した.

まず, ユーザが物件を閲覧するたびに, 直近閲覧したN件の物件を用いて, ユーザが物件に求める条件の内的優先順位を求める. N 件の物件を 1 つの文書, 集合内に含まれる物件の条件を単語とみなし, 物件条件の TF-IDF 値を求める. j 番目の物件を閲覧した際の, 物件条件c_iの TF-IDF 値は以下の式で求める.

$$TFIDF_{i,j} = \frac{n_{i,j}}{DF_i \sum_k n_{k,j}} \quad (1)$$

ただし, $n_{i,j}$は直近閲覧したN件中, 条件c_iを満たしている物件の数, DF_iは全物件中, 条件c_iを満たしている物件の数を表している. TF-IDF 値を元に, 物件条件を降順にソートし, j 番目の物件を閲覧した際の物件条件c_iの内的優先順位$rank_{i,j}$を求める.

次に物件を閲覧するたびに求めた物件条件の内的優先順位から, 常に上位に存在し, かつ順位の変動幅が小さい条件を抽出する. 具体的には, 以下の式を用いて, j 番目の物件を閲覧した際の物件条件c_iの平均順位$rankave_{i,j}$, 物件条件c_iの平均順位変動$rankdiff_{i,j}$を求める.

$$rankave_{i,j} = \frac{\sum_{t=1}^{t=j} rank_{i,t}}{j} \quad (2)$$

$$rankdiff_{i,j} = \frac{\sum_{t=2}^{t=j} |rank_{i,t-1} - rank_{i,t}|}{j-1} \quad (3)$$

最後に平均順位$rankave_{i,j}$が上位 X%かつ平均順位変動$rankdiff_{i,j}$が上位 Y%の条件c_iを抽出し, これを j 番目の物件を閲覧した際のユーザの潜在的嗜好とみなす.

3.4 分析準備

本節では実験に用いたデータ, 評価指標において説明を行う.

関東地方の賃貸物件を閲覧しているユーザを分析の対象とした. 期間 2017-01-01 ~ 2017-01-31 で関東地方の賃貸物件を閲覧しているユーザの内, CV(これ以後, CV は閲覧した物件の資料請求を行うことを指す)以前に 20 件以上, 100 件未満の物件を閲覧しているユーザの中からランダムに抽出した 100 人を実験対象とした.

手法の評価には, 各ユーザから抽出した潜在的嗜好と CV した物件の条件との適合率, 再現率, F-measure の平均を用いた. またベースラインとして, ユーザが閲覧した物件からランダムに抽出した物件の物件条件, ユーザが閲覧した物件の中で最も閲覧時間が長い物件の物件条件と, 同じく CV した物件の条件との適合率, 再現率, F-measure の平均を算出した.

3.5 分析結果

提案手法の定性評価と定量評価を示す. あるユーザの CV 直前までの閲覧履歴から抽出した潜在的嗜好と, その直後に CV した物件の物件条件を表 1 に示す.

表 1：閲覧履歴から抽出した潜在的嗜好と，その直後に CV した物件の物件条件

提案手法により抽出した潜在的嗜好	CV した物件の物件条件
地域:練馬区	地域:練馬区
徒歩分数:10 分以内	築年数:築 30 年以上
設備:建物設備	家賃:6.8 万円
間取り:2K	間取り:2K
建物種別:マンション	設備:建物設備
	沿線:西武池袋線
	徒歩分数:徒歩 8 分
	専有面積:33 平米
	建物種別:マンション

表 1 より, 提案手法によって抽出された潜在的嗜好は5条件中 4 条件が, CVした物件の物件条件と一致しており, 適合率が高い. 一方で, 抽出された嗜好数は 5 条件と少ないため, 再現率は低くなっている. これは CV した物件の物件条件の全てが必ずしもユーザが重要視している項目ではないために起こると考えられる.

ユーザ 100 人の閲覧履歴を用いて行った定量的評価の結果を表 2 に示す.

表 2：提案手法とベースラインの定量評価

	適合率	再現率	F-measure
提案手法	0.75	0.21	0.32
ランダムな物件	0.31	0.04	0.07
閲覧時間の長い物件	0.65	0.11	0.18

表 2 より, 提案手法の方が, ランダムな物件, 閲覧時間の長い物件と比較した際に高い精度を示しており, ベースラインよりも高い精度でユーザの潜在的嗜好が抽出できていると言える.

4. まとめ

今回の実験では, ユーザが明示的に入力している検索条件を用いることなく, ユーザの閲覧履歴から一定の精度で潜在的な嗜好を得られることを示した. 今後は実験対象者を十分に増やすとともに, 明示的に入力されるユーザの嗜好との比較を行うことにより, より適切な絞り込み条件のサジェストや物件推薦の精度向上を行っていきたい.

参考文献

[土方 2005 年] 土方嘉徳: “嗜好抽出と情報推薦技術”, 情報処理学会論文誌, 土方嘉徳. "嗜好抽出と情報推薦技術." *情報処理学会論文誌* 47.4, 2005.

[布施 2018 年] 布施拓馬, 關翼人, 村田昇, 杉浦太樹, 野村眞平: “検索条件のベクトル表現を用いた検索の興味の可視化”, *情報処理学会第 80 回全国大会*, 2018.

[高畑 2011 年] 高畑麻理, 上田真由美, 中島伸介: “食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案.”, *DEIM Forum*. 2011.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Web 不動産データを用いた空物件が埋まる遷移に関する多変量解析 —アパートローンリスク計量モデル構築のための予備解析—

Multivariate analysis of the transition probability of an occupation of rental rooms by using the housing information website data: Preliminary analysis to develop the risk model of the rental home financing

渡邊隼史 *1 Hayafumi Watanabe　一藤裕 *2 Yu Ichifuji　鈴木雅人 *3 Masato Suzuki　山下智志 *4 Satoshi Yamashita

*1情報・システム研究機構 Research Organization of Information and Systems
*2長崎大学 Nagasaki University
*3UD アセットバリュエーション UD Asset Valuation Co., Ltd.
*4統計数理研究所 The institute of statistical mathematics

The apartment loan is a loan for rentals such as for condos, apartments. This loan is a very large loan which is the account for a percentage of more than 10 percents of the whole banks' loan. However, a risk model of the apartment loan with the appropriate accuracy has not been provided in Japan mainly due to the lack of data. Thus, in order to develop the risk model, we preliminarily analyse and compare two types of data set: the servey data which is made by the real estate appraiser and the housing information website data. As a result, it was found that (i)the web data is approximetly corresponding to the survey data which was made by exparts with respect to statitical properties, (ii)The AR value of the simple multivariate regression model developed in this study which explains the transion of rooms from vacancy to occupation takes about 0.4 and (iii)this transition probability of the model is mainly explained by an age of a building.

1. はじめに

アパートローンは，賃貸物件向けの融資であり，銀行が貸し出す全与信額の 10 %を超える巨大な融資であるにもかかわらず，これまで十分な精度のリスク計量化モデルが考案されてこなかった．アパートローンのリスクは，（1）賃貸物件の経営自体による資金不足リスク（2）賃貸経営以外の事業によるリスクの 2 種類の要因から構成される．（2）については，一般的な信用リスクモデルやデータベースが整っており，それを用いてある程度の精度をもってリスク計量が可能である．一方（1）については，賃貸不動産の空室データベースが整っておらず，十分な精度のモデルが提供されていなかった．そこで現在（1）について空占室データベース構築と評価モデルの開発を目指しプロジェクトとして研究を行っている．なお，本研究プロジェクトの最終的な目的の一つは構築したアパートローンリスク計量モデルを CRD 協会（銀行にリスクデータベースとリスクモデルを提供する組織）を通して，民間金融機関に提供することにある．

本稿では，アパートローンリスクの評価法の開発の第一歩として，「物件の埋まりやすさ」や「埋まりにくさ」をアパートのもつ特性から評価するプロトタイプのモデル構築について報告する．空室から占室への遷移や物件情報など研究に必要なデータは不動産情報サイトのデータ (以下 Web データ) を用いた．本研究では，まず，Web データと不動産鑑定士のによる現地調査データを比較することで Web データがどのような特色をもつデータかを確認し，次に結果に基づき Web データによる「埋まりやすさ」評価のモデルの構築可能性検証のため，簡単な多変量解析モデルの構築と評価を行った．

2. 関連研究

本稿で扱う研究は，データを用いた不動産評価の研究の一種といえる．データを用いた不動産評価，特に，不動産価格評価の研究は 1970 年代に主な手法が確立し，現在も大規模データや機械学習等の技術を用いて精度の改善，網羅性やリアルタイム性の性能上の向上の努力が続けられている [清田 17, 清水 17]．最近年においては，実務においては，2015 年前後から，ソニー，ネクスト (現 FIFULL， HOMES)，マンションリサーチ等の企業により不動産価格評価の Web サービスが公開・提供されている [谷山 16]．また研究レベルでは，画像情報を用いた不動産価値評価 [清田 17]，アンケート情報やセンサー等用いた人の流量データを不動産や都市評価に利用する試み [荒川 17]，など近年まで利用が難しかった様々な情報を不動産評価に加える試みもおこなわれている．

一方，価格ではなく，データを用いた個別物件の空占遷移予測について国内における学術論文はデータ入手の困難もあり多くは見つけられなかった．その例としては，籠，高辻，小野らがモンテカルロシミレーションによる空室率を考慮したアパートリスク計量の理論的な研究を行っている [籠 00]．また，実データを用いた例では小林の研究がある [小林 16]．小林の研究では仲介管理会社 3 件の約 6 7 3 件のデータについて空室期間等を実データから推定しそれに基づく将来収益予測モデルの構築を行っている．

本研究は実データ解析のため小林の研究に近い．小林の研究として比べた新たな貢献としては，より様々な物件要因を考慮していること，よりスケーラブルな Web データを利用すること，また不動産鑑定士の現地調査のサーベイデータ（業者による偏りが少ないデータ）と比較することで業者やサービス等のバイアスの存在の有無を確認していること等があげられる．

3. データの取得および前処理

本研究では 2 つのデータを用いた．一つはメインの Web 不動産サイトデータであり，もう一つは補助的に用いる不動産鑑定士による現地調査データである．不動産鑑定士による調査データを利用する主な目的は，以下の 2 点である．一点目は，Web では得にくい情報（既に埋まっている物件の情報，管理

連絡先: 渡邊隼史, hayafumi.watanabe@gmail.com

の悪さなどネガティブな情報等）の効果の影響調査のため，二点目は，２つのデータを比較することで Web サイトデータのサンプリングバイアスについての情報を得るためである．

Web データは，大手不動産のサイトのある県のある路線地域について，2014 年 11 月 28 日から 2015 年夏まで，各月の 8 日， 18 日， 28 日の情報を取得した．一方，鑑定士によるサーベイデータは対応する地域についてある物件名簿より駅別に層化抽出を行った物件に対して同観測期間に 3 か月ごとに空室状況や物件の状況等を現地調査して得られたデータである．

解析で用いた Web データについては以下の前処理を行った．具体的には，(1) 取得データのスプレイピング（２）単位等の統一（３）表記ゆれ等の統一（４）欠損値補完（５）物件の名寄せである．なお，前処理の結果，観測期間の１４か月で解析可能な部屋数は 35,807 件 [空室のみ] となった（不動産鑑定士データは，246 棟,4333 戸 [うち空室は 400 戸程度]）．

本研究の Web データ解析では「空室が占室がなる」を「Web ページから物件が消える」で大まかに近似できるという仮定のもとで研究を行っている．

4. 解析とその結果

4.1 不動産鑑定士データと Web データの基本的な統計の比較

不動産鑑定士データと Web データを基礎的な統計量な簡単な比較した結果以下のことがわかった．

- ２つのデータ共に，空室から占室になる期間の分布は共に平均 150 日程度の指数分布によく近似できる（[小林 16] も指数分布と報告）．
- 築年数，駅からの距離，建物の高さの存在率の分布は Web と鑑定士データでほぼ対応する．ただし，Web データのほうが築浅物件がやや多い傾向にある．

つまり，いくつかの相違はあるものの大まかには，Web データはほぼ鑑定士によるデータと対応していることが確認できた．

4.2 ３か月後遷移確率の多変量解析

次に空室から占室への遷移が物件がどの程度説明できるかを確認するため簡単な多変量解析を行った．具体的には，簡単のためロジスティック回帰モデル（二項ロジットモデル）を用いた．モデルでは，被説明変数を「空室だった物件が 3 か月に占室になる場合のフラグ（空→空が 0, 空→占が 1)」，説明変数を Web データに掲載されている項目（築年数，最寄り駅，コンビニからの距離，面積，募集賃料，保証人の有無など数十項目程度）としてモデルを構築した．変数選択は多重共線性の影響を避けるため精度指標の AR 値（AR=2AUC-1）[山下 11] が改善がとまるまで 1 変数づつ回帰モデルに変数を追加する方法を用いた（AIC ステップワイズ法に近い方法,「追加」なのは異常値等の影響を目視で確認しやすいため）．なお，一時期だけ空占に効く変数は特定の事業者がまとめて不動産サイトから物件を除去した効果など本質的でない要因が多かったため，観測期間で常に効果がある変数のみを採用するようにした．

結果，以下の 7 要因採用され，モデルの 49 観測点でのバックテスト AR 値の中央値 0.378 だった（国産小企業の倒産予測モデル程度の精度）．採用された項目は以下の通り：築年数 (0.318）[-]，即入居可 (0.357)[-]，保証人不要 (0.361)[+]，洗髪洗面化粧台 (0.370)[+]，駅から徒歩距離 (0.372)[-]，追い炊き有 (0.376)[+]，コンロ（なし、IH， ガス…）(0.378)[+]．ここで表記は，項目 (その要因より前に記述がある要因郡を変数として採用したモデルの累積 AR 値)[効果の正負] であった．結果より，築年数が空占遷移の説明力の大きな部分を占めていることがわかった．

5. まとめと考察

本研究ではアパートローンリスク計量化のうち，アパート経営の事業リスク評価の第一段階の準備研究として，Web 不動産サイトデータの性質の検証と「空→占」の簡単な状態遷移モデルの作成を行った．具体的には，まず，Web データの収集と前処理の手法を整備した．次に，不動産鑑定士による現地調査と比較し，Web から得られたデータの性質の検証を行った．最後に，簡単な多変量空占遷移モデルを構築した.（このとき，異常値など Web データの特有の問題に対応するため，時間に依存せず安定した説明要因をひとつずつ採用する変数選択など Web に合わせた工夫を行った).

解析の結果以下のことがわかった.（１）Web データは不動産鑑定士データと統計的性質が類似している [Web データ使用の妥当性確認].（２）多変量二項ロジットモデルを用いた「空→占の確率遷移モデル」はバックテスト AR 値で 0.4 程度 （３）モデルで主要な説明力を持つ要因は築年数．

考察　本研究の多変量解析で採用された変数には,「募集賃料」や「最寄り駅」など主要な変数を含んいない．これは以下の２つが主な理由と考えている（１）賃料調整効果（２）市場価格より大きく離れた物件データの不足（同データから構築した賃料回帰モデルの説明要因は，今回採用された変数は築年数と駅からの距離を除き含まれなかった）．つまり，本研究の結果は賃料に反映されていない物件の需要要因と解釈できるかもしれない（例えば，築年数はニーズ比べて割安など）．モデルを信頼性をもって利用するためには，このような空占室遷移と賃料との関係を今後詳細に検討する必要があると考えられる．

参考文献

[清田 17] 清田陽司・山崎俊彦・諏訪博彦・清水千弘:不動産と AI，人工知能，32 巻，4 号，pp. 529535 (2017)

[清水 17] 清水千弘: 不動産ビッグデータでみる不動産価格の決まり方，日本不動産学会誌，32 巻，1 号，pp. 4551 (2017)

[谷山 16] 谷山俊彦: 人工知能とビッグデーアが返る不動産市場,Financial Information Technology Focus 2016，6 号，pp. 8-9 (2016).

[荒川 17] 荒川周造・諏訪博彦・小川祐樹・荒川豊安本慶一・太田敏澄，暗黙知に基づく飲食店用不動産の賃料推定システム，SIG-SAI，28 巻， 1 号，pp. 16 (2017)

[籠 00] 籠義樹・高辻秀興・小川祐哉，空室率と賃料の変動過程を考慮した不動産投資モデルの期待収益率と VaR に関する研究，麗澤経済研究，8 巻，2 号，pp. 99113 (2000)

[小林 16] 小林 秀二: 入退去ミクロ分析による不動産評価における将来予測の捉え方：生存時間解析データによる家賃キャッシュ・フローの無条件確率遷移, ファイナンシャル・プランニング研究 16 号, pp. 18-27(2016)

[山下 11] 山下智志・三浦翔，信用リスクモデルの予測精度: AR 値と評価指標，朝倉書店 (2011).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Bottleneck 特徴量を用いた不動産画像の分類

Classification of Real Estate Images using Bottleneck Features

豊原 優[*1] 門 洋一[*2] 山崎俊彦[*3] 藤森 進[*1] 太原育夫[*1]

Suguru Toyohara Youiti Kado Toshihiko Yamasaki Susumu Fujimori Ikuo Tahara

[*1] 東京理科大学 [*2] アットホーム株式会社 [*3] 東京大学

Tokyo University of Science At Home Co., Ltd. The University of Tokyo

Image processing enhanced by deep learning is now being actively applied to real-world issues. The real estate industry, which owns lots of property information, also expects to develop new services using deep learning. In this paper, we present the classification methods for property images using bottleneck features generated by convolution networks (CNN), and experimentally show that we can achieve 88% accuracy in Top-1.

1. はじめに

ILSVRC2012 において, Convolutional Neural Network (以下, CNN)が優勝して以来, CNN は世界中の研究者たちに大いに注目され, それに伴い画像解析技術はさらに飛躍的に進歩を遂げている. さらに, この技術をビジネスに活かす為の研究開発も盛んに行われ, 自動車業界や医療分野にまで広がっている.

一方で不動産業界においては, 未だアナログな部分が多く残っているが, Real-Estate-Tech と呼ばれる不動産と情報テクノロジーを掛け合わせて新しい価値を生み出す動きが始まっている. 具体的には, 機械学習理論を用いた価格推定や, 新たな推薦システムの仕組みなどが新たに検討されており, これからも大いに注目される分野と言える.

本論文では, CNN を用いた画像分類技術を不動産画像に対し適用した際の分類手法を提案し, 不動産会社の負担を減らす様々な業務利用可能な学習データ収集を可能にするシステムを, 実運用までを視野に入れ検討する.

2. 物件画像の分類

2.1 CNN を用いた特徴抽出

CNN を用いた深層特徴量の一つに, Bottleneck 特徴量と呼ばれる特徴量がある. これは CNN における畳み込み最終層の出力する特徴量のことで, ILSVRC での学習データである ImageNet [Russakovsky 2014] ベースに学習を行った CNN では, この特徴量が画像の特徴を非常によく表していることが分かっている為, 多くの CNN 研究において ImageNet ベースの重みを利用している.

当論文では GoogLeNet (Inception-v3) [Szegedy 2015] における ImageNet 学習済み重みを用いて Bottleneck 特徴量を計算している. これに多層パーセプトロンを用いて解析することで画像の種別判定を行うこととする.

2.2 問題設定

画像のタグ付けを自動化するシステムを考える上で重要なのが, Top1-Error での評価指標で考えなければならないという点にある. 不動産業者がシステムを利用した際に, 画像登録時にシステムが自動でタグ付けをすることを想定しており, 確信度が最も高いクラスを自動で登録する必要がある. また, ランニングコストの観点から, 複数の畳み込みモデルによる複数の Bottleneck 特徴量生成を用いたアンサンブル学習は用いることができない.

分類するカテゴリは, 外観・玄関・エントランス・室内・リビング・お風呂など, システム上付けることのできる全 20 クラスとする. その内訳を Table1 に示す.

2.3 提案手法

実運用をする際には, 精度とコストのバランスが重要であり, 少ない計算量でいかに精度を上げるかが課題となる. 本論文で提案する手法は, 共通の Bottleneck 特徴量を用いて, 複数の全結合層に解析させることで精度を向上させるものである. 提案する学習モデルは, 1 つの共通の畳み込み層を用いて, データセットの組み合わせを変え学習した各全結合層のモデルと, 畳み込み層を除いた各全結合層のみでアンサンブル化したモデルである.

それぞれの全結合層では, データセットの組み合わせを変えることで精度がどのように変わるのかを観察し, 最終的にアンサンブル化した場合とで, どのように性能が変わるのかを調べた.

さらに, 実験条件にて述べる D5 に関しては, 間違いやすいカテゴリに対して寝室や外階段といった, さらに粒度の細かいカテゴリを用意することで, どのように精度が向上するか, アンサンブルにおいて多様性を発揮するかを調べた. また, 本論文のアンサンブル方式は, 各全結合層の出力における最大スコアのカテゴリに対して投票を行う方式である.

Table 1: 20 カテゴリの詳細とその出現頻度

カテゴリ	出現比率	カテゴリ	出現比率
外観	12.2%	洗面所	4.8%
バルコニー	3.5%	間取り図	10.5%
お風呂	6.5%	リビング	5.5%
共有部	2.8%	地図	0.1%
エントランス	2.0%	駐車場	1.3%
玄関	5.4%	収納	5.6%
設備	6.4%	周辺環境	5.0%
庭	0.2%	トイレ	5.5%
室内	12.7%	地形図	0.6%
キッチン	8.0%	眺望	1.6%

連絡先:豊原 優, 東京理科大学工学研究科経営工学専攻,
〒125-8585 東京都葛飾区新宿 6-3-1 太原研究室,
E-mail:4417619@ed.tus.ac.jp

Table 2: アンサンブル Top1 の判断と正解との Confusion Matrix

正解 \ Deep-Learningの判定	外観	ﾊﾞﾙｺﾆｰ	風呂	共有	ｴﾝﾄﾗﾝｽ	玄関	設備	庭	室内	ｷｯﾁﾝ	洗面	間取り	ﾘﾋﾞﾝｸﾞ	地図	駐車	収納	周辺	ﾄｲﾚ	地形	眺望
外観	172	2	0	1	0	0	0	8	0	0	0	0	0	0	6	0	9	0	0	2
バルコニー	0	185	1	1	0	2	1	1	4	0	0	0	0	0	1	0	0	0	0	4
風呂	0	0	186	0	0	2	2	0	0	0	5	0	1	0	0	0	0	4	0	0
共有部	0	11	0	124	25	4	6	9	0	1	0	0	7	1	7	2	3	0	0	0
エントランス	3	5	0	15	158	0	0	3	1	0	0	0	8	0	5	1	1	0	0	0
玄関	0	4	0	3	2	155	11	0	10	1	1	0	6	0	0	7	0	0	0	0
設備	0	1	1	1	0	3	176	0	6	7	2	0	0	1	0	0	0	1	1	0
庭	2	2	0	0	2	0	0	186	0	0	0	0	1	0	1	0	1	0	1	4
室内	0	0	1	3	0	6	1	0	106	4	0	0	63	0	1	15	0	0	0	0
キッチン	0	0	1	0	0	1	4	0	1	185	3	0	4	0	0	0	0	1	0	0
洗面	0	0	4	0	0	4	2	0	1	3	183	0	1	0	0	1	0	1	0	0
間取り	0	0	0	0	0	0	0	0	0	0	0	195	0	1	0	0	1	0	3	0
リビング	0	0	0	0	1	2	0	0	23	5	1	0	165	0	0	3	0	0	0	0
地図	0	0	0	0	0	0	0	0	0	0	0	1	0	197	0	0	0	0	2	0
駐車場	3	0	0	1	1	0	0	4	0	0	0	0	0	0	182	0	7	0	0	2
収納	0	1	0	0	0	7	4	0	11	0	2	0	5	0	0	170	0	0	0	0
周辺環境	7	0	0	3	2	0	0	4	0	0	0	0	0	0	11	0	160	0	1	12
トイレ	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0	0	0	195	0	0
地形図	0	0	0	0	0	0	1	0	0	0	0	55	0	9	0	0	0	0	135	0
眺望	5	5	0	0	0	0	0	2	0	0	0	0	0	0	3	0	1	0	0	184

3. 実験条件

使用したデータは，アットホーム株式会社が保有する不動産物件画像データの中から抽出している．また，訓練データを以下の 5 つのカテゴリに分け，データセットごとに調べた．

- D1:不動産業者がラベル付けした 60 万枚のデータセット
- D2:D1 内から 1200 枚だけを抽出したデータセット
- D3:D2 に対して人が正確にラベル付けをしたデータセット
- D4:D3 を用いて学習した全結合層を利用し，D1 のラベルデータと，D3 で学習した分類器による判断が一致したデータだけを残した 44 万枚のデータセット
  （質の良いデータセットを大量に獲得することにより，精度向上に期待する．）
- D5:D3 に，詳細なサブカテゴリを追加したデータセット
  （新たな知識により，精度向上を期待する．）

テストデータに関しては，人が正確にラベル付けをした重複のない 20 カテゴリ×200 枚ずつのデータを，分類器の精度評価用テストデータとした．また，室内とリビングなどに頻発する不動産会社が画像を見ても，どちらのカテゴリとも判定できると思われる画像に関しては，正解として扱った s.

学習に使用したモデルは GoogLeNet を用いて，Bottleneck 特徴量に関しては，学習の簡易化をする為，ImageNet による重みに固定し，データセットごとに全結合層のみを学習できる状態で訓練を行った．誤差関数は Cross Entropy を使用し，Batch-Size は 100，勾配降下法による学習率は 0.01 で合計 4000 回学習させたものである．

4. 実験結果

4.1 カテゴリ間における判断ミスの傾向

実験結果として，アンサンブル学習を用いた結果を Table2 の Confusion Matrix に示す．Confusion Matrix は Top1 での判断と正解ラベルとの比較の図である．室内・リビング，エントランス・共有・玄関の組み合わせなどは，カテゴリ的に近い特徴を持っているため混同しやすく，全体の成績を下げている傾向が伺える．

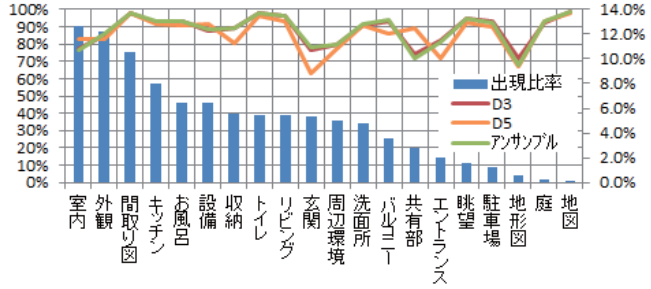


Figure 1:各データセットの正解率を比較

Table 3:データセット毎のマルチラベル正答率一覧

	D1	D2	D3	D4	D5	ｱﾝｻﾝﾌﾞﾙ
外観	89%	87%	86%	85%	83%	86%
バルコニー	95%	96%	93%	94%	86%	94%
お風呂	93%	94%	94%	92%	91%	93%
共有部	74%	66%	75%	68%	90%	72%
エントランス	83%	81%	82%	79%	72%	82%
玄関	78%	82%	77%	77%	63%	78%
設備	87%	89%	88%	85%	92%	89%
庭	92%	93%	93%	92%	93%	93%
室内	78%	75%	77%	74%	83%	77%
キッチン	95%	94%	93%	92%	92%	94%
洗面所	93%	92%	91%	90%	91%	92%
間取り図	98%	98%	98%	97%	98%	98%
リビング	96%	94%	96%	96%	94%	96%
地図	97%	94%	99%	98%	98%	99%
駐車場	92%	91%	93%	92%	90%	92%
収納	86%	84%	90%	90%	81%	90%
周辺環境	86%	82%	80%	78%	77%	80%
トイレ	98%	98%	98%	98%	97%	98%
地形図	61%	60%	72%	67%	67%	68%
眺望	95%	93%	95%	94%	93%	95%
平均	88.05%	86.78%	88.18%	86.65%	86.30%	88.08%
加重平均	88.87%	87.74%	88.03%	86.73%	86.76%	88.21%

4.2 正答率

実験結果における，D3,D5 およびアンサンブル との正答率を比較した図を Figure1 に示す．カテゴリ粒度の細かい D5 では，新たなカテゴリ知識を増やした場合に，共有部のスコアは良くなった一方で隣接するエントランスカテゴリのスコアを吸収してしまっている．さらに，アンサンブルを見ると成績が D3 とほぼ同じ傾向を持つことが分かるので，D5 による多様性がアンサンブル学習にうまく反映されていないことが伺える．

一方，各データセットによる正答率を Table3 にて示す．アンサンブルの結果は若干 D3 よりも悪く，マルチラベルで 88.08% であった．Table1 にて示す実際の出現頻度との加重平均で見ても，88.21%と D1 での学習結果より低い結果となってしまった．

4.3 全体のまとめ

実験結果から，Top1 での精度がおおよそ 88%を超えるものであり，またその中で判定に納得のいかない結果となったものは数える程少ないということから，CNN をベースにした Bottleneck 特徴量で計算する部分に関して，実用性を満たす範囲にあると考えられる．また，Global Average Pooling を使用していることから，Class Activation Map [Zhou 2015]出力し，システム側の判定基準がどこかを目視することも可能である．

5. おわりに

本論文では，不動産画像を CNN による Bottleneck 特徴量で分類したが，今後さらにカテゴリ粒度を細かく分類するアルゴリズム研究や，全結合層の改善，さらには不正画像検知などの分野にも研究を進めていく予定である．

参考文献

[Szegedy 2015]C. Szegedy, W.Liu, Y.Jia, P.Sermanet, S.Reed, D.Anguelov, D.Erhan, V.Vanhoucke, and A.Rabinovich:“Going deeper with Convolutions,” In CVPR,2015.

[Russakovsky 2014] O.Russakovsky, J.Deng, H.Su, J.Krause, S.Satheesh, S.Ma, Z.Huang, A.Karpathy, A.Khosla, M.Bernstein, A.C.Berg, L.Fei-Fei,” ImageNet Large Scale Visual Recognition Challenge,” arXiv:1409.0575,2014.

[Zhou 2015] B.Zhou, A.Khosla, A.Lapedriza, A.Oliva, and A.Torralba: “Learning Deep Features for Discriminative Localization,” arXiv.1512.04150, 2015.

[2D1] [General Session] 13. AI Application

座長:尾形 哲也（早稲田大学）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room D (4F Cattleya)

[2D1-01] Parallel Implementation of Probability Occupancy Map Generation for Multiple Human Tracking Based on Graph Optimization

○Yuri Nishikawa[1,2], Hitoshi Sato[1], Jun Ozawa[1] （1. National Institute of Advanced Industrial Science and Technology (AIST), 2. Panasonic Corporation）

9:00 AM - 9:20 AM

[2D1-02] Towards Accurate Multiple Human Tracking Using Scalable Distributed Deep Learning

○Hitoshi Sato[1], Yuri Nishikawa[1,2], Jun Ozawa[1] （1. National Institute of Advanced Industrial Science and Technology (AIST), 2. Panasonic）

9:20 AM - 9:40 AM

[2D1-03] Performance enhancement of multi-human tracking based on K-Shortest Paths by data reduction

○Nozomi Hata[1], Yuri Nishikawa[3,4], Takashi Nakayama[1], Jun Ozawa[3], Katsuki Fujisawa[2,3] （1. Faculty of Mathematics, Kyushu University, 2. Institute of Mathematics for Industry, Kyushu University, 3. National Institute of Advanced Industrial Science and Technology, 4. Panasonic Corporation）

9:40 AM - 10:00 AM

[2D1-04] An action recognition system using a highly abstracted infrared array sensor

○Shogo Murakami[1], Ikuko Eguchi Yairi[1] （1. Graduate School of Science and Technology, Sophia University）

10:00 AM - 10:20 AM

[2D1-05] Detection of a Painter School with Deep Learning Method: A Case Study of Scenes from the Tale of Genji

○Takuya Kato[1], Mariko Inamoto[2], Akihiko Konagaya[1] （1. Tokyo Institute of Technology, 2. Keisen University）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

グラフ最適化を用いた多人数追跡手法における グリッドマップ生成の並列化

Parallel Implementation of Probability Occupancy Map Generation for Multiple Human Tracking Based on Graph Optimization

西川 由理 *1*2　佐藤 仁 *1　小澤 順 *1
Yuri Nishikawa　Hitoshi Sato　Jun Ozawa

*1産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

*2パナソニック株式会社
Panasonic Corporation

Multi-camera multiple object tracking methodology that uses probability occupancy map (POM) and graph optimization can efficiently track multiple people in spite of significant occlusion, even without appearance model and prerequisite knowledge about the number of people in the target area. However, the computation time of POM increases according to grid size and number of cameras, which limited latency and throughput. In this paper, we report the effect of hierarchically applying two parallelization techniques to enhance the performance of POM generation. As a result, we achieved more than 20 times higher throughput compared to sequential processing, and reduce latency by 66%.

1. はじめに

カメラ映像から複数の人物を追跡する多人数追跡技術は，セキュリティ，業務効率化，マーケティング，スポーツ戦略分析など，多岐に渡る応用がある．一方，オクルージョンの前後で人物の対応をとる人物同定 (association) は，今なお，コンピュータビジョン研究の重要なテーマの一つである．

複数カメラの多視点映像を統合する手法は，このこの課題に対する有効なアプローチの一つである．特に [1] は，各カメラの人物検出結果から，空間を格子状に分割したセル毎の占有確率である 2 次元のグリッドマップ (Probability Occupancy Map: POM) を生成し，世界座標系の人物位置を推定する手法を提案している．また，各時刻のグリッドマップから定められるネットワークフローグラフの最適化により，人物を追跡する手法が提案されている [2]．この追跡手法は，人物の顔や色といった画像特徴量が不要であり，作業服やユニフォーム等，複数人のアピアランスが類似するシーンに適用できる．また，人数や移動方向に関する前提知識も不要なため，追跡空間に人が出入りしたり，速度や向きが急に変化したりするシーンへの応用も可能である．

我々は，これらの利点に着目し，グラフ最適化アルゴリズムを用いた多人数追跡システムの開発と評価を進めている [3][4]．追跡精度と実行時間を評価した結果，カメラ数やグリッド数が増えるほど追跡精度が向上する反面，グリッドマップ (POM) 生成の処理時間が増加することが明らかになった．例えば，20cm 角に区切ったサッカー場をカメラ 6 台で撮影した動画 [5] では，1 フレーム分の POM 生成に Intel Ivy Bridge-EP CPU でも 10 秒前後を要する．さらに，グラフ最適化による追跡では，50〜100 フレーム程度の POM を用いてネットワークフローグラフを構築するため，POM を順次生成すると，数秒の動画長にもかかわらず 10 分以上を要していた．

本稿では，2 階層の並列化を適用することで，PC クラスタ上で POM 生成処理を高速化した結果を報告する．POM 生成はフレーム毎に独立して演算可能であり，GNU Parallel[6] を用い，各フレームを複数の計算ノードに分散する処理を行った．さらに並列プログラミング API である OpenMP[7] を用い，1 ノード内でカメラ毎の演算処理をマルチスレッド化し，各スレッドの結果を統合する改良を行った．これにより，同サッカーのフレームあたりの POM 生成処理が約 10 秒から約 3.5 秒と，約 66%短縮された．さらに GNU Parallel を用い，ノード数を 1, 2, 4, 8 と変化させながら並列分散処理した場合も，ノード数の増加によらず，一律に約 4 秒で処理が可能であり，20 倍以上のスループット向上を確認した．

連絡先: 〒 135-0064 東京都江東区青海 2-4-7
産業技術総合研究所臨海副都心センター別館 9 階
西川由理 (nishikawa.yuri@aist.go.jp)

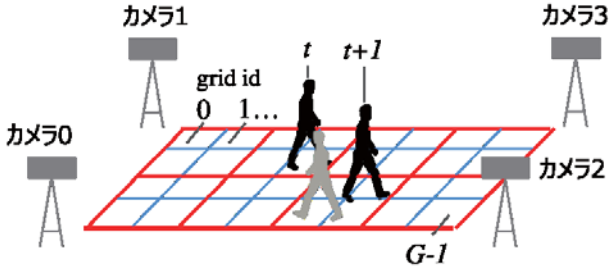


図 1: 追跡空間モデル

2. グラフ最適化を用いた多人数追跡

2.1 空間モデル

グラフ最適化を用いた多人数追跡では，図 1 に示すように，2 次元のグリッド（0,1，…，G-1）に分割された世界座標系の追跡空間において，各人物が連続フレーム間（t から t+1）にグリッドセルを移動すると想定する．また，各グリッドセルを占有するのが高々1 人であるという制約を置くことで，多項式時間で解くことが可能な最小コストフロー問題となる [2]．この制約を考慮すると，セルは高々一人が入る大きさにする必要がある．特に，人物の移動速度が速いスポーツなどでは，経験的に 20cm 角程度とするのが望ましいことが分かっている [4]．

2.2 多人数追跡システムの概要

多人数追跡システムの概要を図 2 に，またモジュール間のデータの依存関係を図 3 に示す．人物検出には深層学習ベースの一般物体認識手法 YOLO[8] を，グラフ最適化には K-Shortest

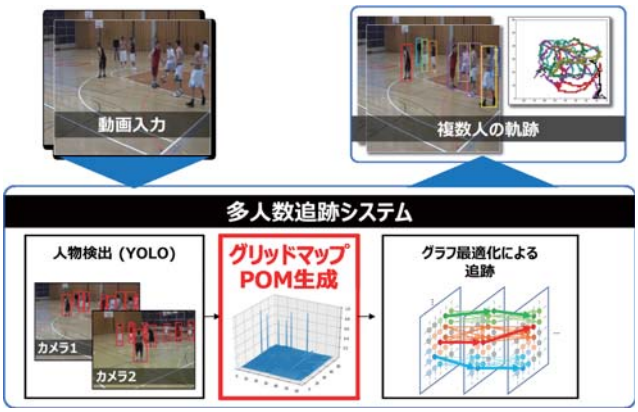


図 2: 多人数追跡システム

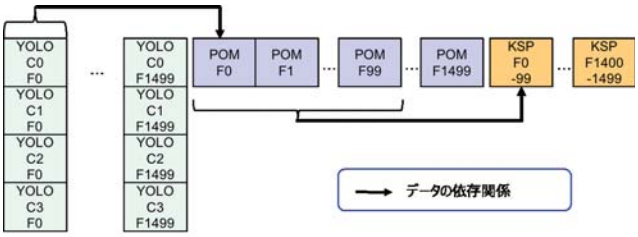


図 3: データの依存関係

Paths(KSP) を適用し，各々開発者が提供するソースコードを用いた．YOLO は 1 枚の画像を入力すると，人物検出領域のバウンディングボックス座標 (xmin, xmax, ymin, ymax) を出力する．POM 生成モジュールは，図 3 に示すように，複数カメラの出力を用い，グリッドマップを生成する．最後にグラフ最適化モジュールは，POM 生成モジュールが出力した複数フレームの POM からグラフを構築し，追跡を行う．

3. グリッドマップ POM の並列化

3.1 Probability Occupancy Map (POM) の概要

Probability occupancy map (POM) は，複数カメラで時刻 t に撮影された，画像中の人物のカメラ座標系における位置と大きさに基づき，格子状に区切られた空間の位置 i を人物が占有する確率 q_i^t を推定する手法である．基本となるのは，人物がある高さと直径の「円柱」であると想定し，人物が世界座標系の特定の位置に存在する場合に，その人物を表す円柱（四角形）を，カメラ画像中の適切な位置と大きさで描画する，生成モデルである．そして，複数カメラの画像が与えられた時，このモデルの下で，真の確率分布との KL(Kullback-Leibler) 情報量を最小化するモデル $Q^t = \{q_0^t, q_1^t, ..., q_{G-1}^t\}$ を推定することにより，各位置の占有確率を求める．なお，Q^t は，時刻毎に独立して演算が可能であるため，以後，必要がある場合を除き，添字 t を割愛して表記する．また，各変数の記法を表 1 に示す．

POM の具体的な生成手順を Algorithm1 の擬似コードに示す．また，2 つの画像 $\mathbf{A}$, $\mathbf{B}$ の類似度を表す「擬似距離」$\Psi(\mathbf{A}, \mathbf{B})$ を求める関数 get-pseudo-distance は，式 1 で定義される [*1]．また，各カメラの平均画像 $\mathbf{A_c}$ を求める関数 get-

*1 特に断りがない限り，太字の大文字の変数および「画像」という用語は，サイズ $W \times H$ の二次元配列を，次元変更した一次元ベクトルを表す．

Algorithm 1 POM の生成アルゴリズム

input: 時刻 t の人物検出結果 $\{\mathbf{B_0}, ..., \mathbf{B_{C-1}}\}$, 事前知識 $\{\mathbf{R_0^{\xi_0=1}}, ..., \mathbf{R_0^{\xi_{G-1}=1}}, ..., \mathbf{R_{C-1}^{\xi_{G-1}=1}}\}$
output: 時刻 t の POM $Q = \{q_0, ..., q_{G-1}\}$
$\{q_0, ..., q_{G-1}\} \leftarrow \{\epsilon_0, ..., \epsilon_{G-1}\}$
for $t \leftarrow 0$ to T **do**
  while $e < \varepsilon$ **do**
    for $c \leftarrow 0$ to カメラ数 C **do**
      $\mathbf{A_c} \leftarrow$ get-average-image $(\mathbf{A_c}, Q)$
      for $i \leftarrow 0$ to グリッドセル数 G **do**
        $\mathbf{S_c^{\xi_i=1}} \leftarrow \mathbf{A_c} \oplus \mathbf{R_c^{\xi_i=1}}$
        $\mathbf{S_c^{\xi_i=0}} \leftarrow \mathbf{A_c} \otimes \mathbf{R_c^{\xi_i=0}}$
        $\Psi_c^{\xi_i=1} \leftarrow$ get-pseudo-distance $(\mathbf{B_c}, \mathbf{S_c^{\xi_i=1}})$
        $\Psi_c^{\xi_i=0} \leftarrow$ get-pseudo-distance $(\mathbf{B_c}, \mathbf{S_c^{\xi_i=0}})$
        $sum_i += \Psi_c^{\xi_i=1} - \Psi_c^{\xi_i=0}$
      end for
    end for
    for $i \leftarrow 0$ to グリッドセル数 G **do**
      $tmp \leftarrow= 1/(1+\exp(\lambda_i + sum_i))$
      if $\mathrm{abs}(tmp - q_i) > e$ **then**
        $e \leftarrow \mathrm{abs}(tmp - q_i)$
      end if
      $q_i \leftarrow tmp$
    end for
  end while
end for

average-image は，式 2 で定義する．

$$\Psi(\mathbf{B}, \mathbf{S}) = \frac{1}{\sigma} \frac{|\mathbf{B} \otimes (\mathbf{I} - \mathbf{S})| + |(\mathbf{I} - \mathbf{B}) \otimes \mathbf{S}|}{|\mathbf{S}|} \quad (1)$$

$$\mathbf{A_c} = \mathbf{I} - \prod_i (\mathbf{I} - q_i \mathbf{R_c^{\xi_i=1}}) \quad (2)$$

Algorithm1 は，5 行目の誤差 e が収束するまで（多くの場合は 100epoch 以下）処理を繰り返す．直感的に述べると，ある epoch における $Q = \{q_0, q_1, ..., q_G\}$ から生成されるカメラ毎の生成画像 $\mathbf{S_c^{\xi_i}}$ と，実際のカメラ画像から YOLO 等の人物検出器が検出した人物位置に基づく画像 $\mathbf{B_c}$ の類似度 Ψ を式 1 により求める．誤差が収束し，2 枚の画像の類似度が十分近づいたとき，POM Q は，図 2 に示すように，ピークの尖った分布となる．

3.2 POM の並列化指針

並列化は，独立性の高い外側のループから順に行うのが基本指針である．ここで，Algorithm1 によれば，フレーム毎の演算を表す 4 行目の for ループは，独立して演算が可能である．一方，6 行目の for ループは，要素数 G の配列 sum にループ間の依存関係がある．しかし，配列 sum をループ毎にプライベートに確保し，14 行目の処理が完了した時点で要素毎にリダクション演算を行うことにすれば，6 行目のループでマルチスレッド化することで，独立に演算が可能であることが分かる．

以上を踏まえた，POM の並列化方針を図 4 に示す．まず，1 プログラム (コマンド) で 1 フレーム毎の POM 計算（4 行目のループ）を行うこととした．そして，GNU Parallel[6] というコマンドの並列実行を可能とするシェルツールを用い，複数の計算ノードを用いて分散処理することとした．一方，カメラ毎の並列化（6 行目のループ）は，14 行目終了時点で，都度

表 1: 変数の記法

C	カメラ数
G	グリッドセル数の合計
T	バッチ処理するフレーム数 (=50〜100 程度)
t	時刻 (フレーム)
i	位置
c	カメラ ID
ε	許容誤差
q_i	時刻 t における位置 i の占有確率
Q	時刻 t の POM $(=\{q_0, q_1, ..., q_G\})$
ξ_i	位置 i の真の占有確率
	$\xi_i = 0$:不在，$\xi_i = 1$:人が存在
ϵ_i	位置 i における事前確率
λ_i	ϵ_i のロジット $\log \frac{1-\epsilon_i}{\epsilon_i}$
$W \times H$	カメラ画像の解像度 (ピクセル)
$\mathbf{I}$	値が 1 の大きさ $W \times H$ の画像
$\lvert\mathbf{A}\rvert$	画像の合計（ベクトル要素同士の和）
$\mathbf{A} \otimes \mathbf{B}$	画像 $\mathbf{A}$ と $\mathbf{B}$ の要素毎の積
$\mathbf{A} \oplus \mathbf{B}$	画像 $\mathbf{A}$ と $\mathbf{B}$ の要素毎の和
$\Psi(\mathbf{A}, \mathbf{B})$	画像 $\mathbf{A}$ と $\mathbf{B}$ の擬似距離 (pseudo distance)
$\mathbf{A_c}$	カメラ c における平均画像
$\mathbf{B_c}$	人物検出器が検出した矩形内に対応する
	要素が 1，その他が 0 の画像
$\mathbf{R_c^{\xi_i=1}}$	カメラ c の位置 i における想定人物矩形内に
	対応する要素が 1，その他が 0 のシルエット画像
	$\mathbf{R_c^{\xi_i=0}} = \mathbf{I} - \mathbf{R_c^{\xi_i=1}}$
$\mathbf{S_c^{\xi_i}}$	画像 $\mathbf{A_c}$ と画像 $\mathbf{R_c^{\xi}}$ から生成される画像

表 2: 計算ノード仕様

項目	値
CPU	Intel Xeon E5-2680 v2 x 2
	(Ivy Bridge-EP)
CPU コア数	10 core/CPU
スレッド数	1 thread/core
主記憶	128GiB DDR3-1866
I/F	InfiniBand Mellanox ConnectX-3(FDR)
	Intel X520-DA2(10 GbE)
OS	CentOS 6.7

sum 配列の要素毎の和を求めるリダクション演算が発生する．CPU を用いるノード内並列化手法の候補としては，MPI[9] と OpenMP[7] が挙げられる．簡易評価の結果，プロセス毎に独立のメモリ空間を占有する MPI では，リダクション演算を行うたびにメッセージサイズが G の MPI 集団通信が発生し，性能低下が生じた．従って，共有メモリ型のスレッド並列化手法である OpenMP を用い，カメラ毎にプロセッサ内のコア（スレッド）を割り当てる，マルチスレッド化を行った．なお，グリッドセル単位（8 行目のループ）の並列化も，アルゴリズム上は可能だが，多くの同期の発生が見込まれ，今回は行っていない．

4. 評価

本評価では，産業技術総合研究所の PC クラスタ，AIST Super Green Cloud(ASGC) 上の計算ノード (表 2) を最大 8 台を用いた．実行時間の評価には，バスケットボール動画 [10]

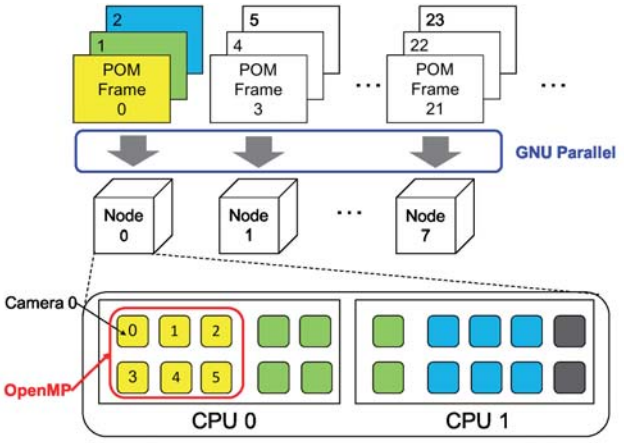


図 4: POM の並列化方針

表 3: 評価用動画の概要

	バスケ	サッカー 50	サッカー 20
空間の広さ	15m x 15m	120m x 90m	120m x 90m
グリッド	100 x 100	212 x 138	530 x 345
カメラ数	4	6	6
人数	13	22	22
fps	25	25	25

およびサッカー動画 [5] を用いた．動画の概要を表 3 に示す．前者は，半コート内で選手が競技する様子を，約 170cm の高さに設置した 4 台のカメラで撮影したものであり，多数のオクルージョンが含まれる．後者は競技を 6 台のカメラで撮影したものである．バスケットボールは追跡空間を約 20cm 角，サッカーは 50cm 角 (サッカー 50) と 20cm 角 (サッカー 20) のグリッドで離散化した．

3 種類の動画について，次の 3 点を評価した．

1. ノード内並列化による 1 フレームの処理高速化の効果 (OpenMP)
2. 1 ノードに 1 フレームのみ割り当てた時の，ノード数増加による処理高速化の効果 (GNU Parallel)
3. 1 ノードに複数フレーム（ジョブ）を割り当てた時の，ノード内ジョブ数の増加による処理時間への影響

1 の結果を，図 5 に示す．バスケットボール動画は 4 台のカメラの映像を用いるため 4 スレッド，サッカーは 6 スレッドで演算処理を行った．Epoch 毎に結果を統合するために発生した同期処理が OpenMP による並列化のオーバヘッドとなったが，バスケットボール動画で約 2.5 倍，サッカーで約 3 倍の高速化を確認した．特に OpenMP 適用前では「サッカー 20」

表 4: ノード数と処理時間の関係 (単位:秒)

ノード数	バスケ	サッカー 50	サッカー 20
1	0.73	2.12	4.07
2	0.75	2.18	4.05
4	0.75	2.24	4.16
8	0.78	2.19	4.11

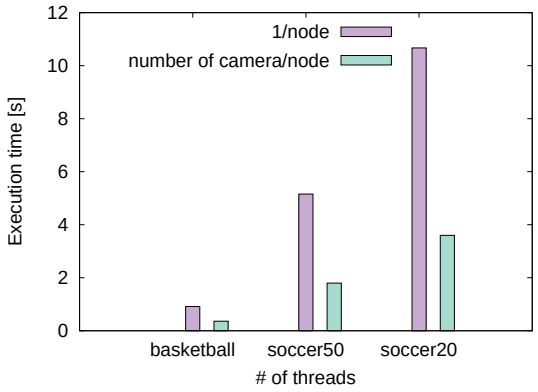


図 5: OpenMP によるカメラ毎並列化の効果

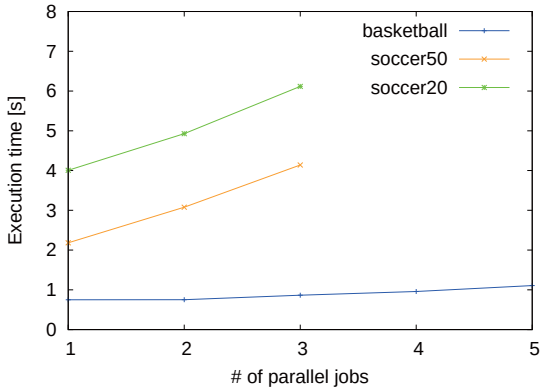


図 6: 1 ノード内のジョブ数と処理時間の関係

で約 10.7 秒を要した POM 生成処理が約 3.6 秒となり，実行時間が約 66%削減された．

次に，計算ノード数と処理時間の関係を，表 4 に示す．各ノードに割り当てるフレーム数を 1 に固定した上で，ノード数を 1, 2, 4, 8 台と変化させ，GNU Parallel で分散処理を行った時間を計測した．表に示すとおり，処理時間はノード数によらずほぼ一定であり，8 フレーム分の「サッカー 20」の結果が約 4.1 秒で得られた．OpenMP 適用前かつ 1 ノードでこれらの処理を行うには，$10.6 \times 8 = 84.8$ 秒を要したことから，分散処理により 20 倍以上の高速化が実現された．

最後に，GNU Parallel を用いて，1 ノード内に複数の OpenMP のジョブを投入したときの，ジョブ数と処理時間の関係を図 6 に示す．ノード数は 1 に固定 (20 スレッド/ノード) にした上で，6 スレッドを使用する OpenMP のジョブ数を 1〜3 (バスケットボールのみ 1〜5) と変化させた時の処理時間を計測した．図に示すとおり，処理時間は並列ジョブ数に従って増加し，サッカー 50 では約 1.9 倍となった．性能低下要因としては，GNU Parallel のオーバヘッドに加え，Disk I/O がボトルネックになっている可能性が挙げられる．今後より詳細に要因を分析し，性能改善を図る予定である．

5. まとめ

OpenMP と GNU Parallel を用いた 2 階層の並列化により，POM 生成処理の高速化を行った．まず OpenMP を用いたカメラ毎の演算処理のノード内並列化により，サッカー動画 1 フレームあたりの POM 生成処理に要する時間を約 66%削減した．さらに，GNU Parallel を用いて最大 8 台の計算ノードに 1 フレームずつ割り当てるノード間並列化を行ったところ，ノード数の増加によらず，ほぼ一定の遅延時間で，各ノードから出力が得られた．以上により，遅延時間の削減とスループットの向上を確認した．

一方で，1 ノード内に複数の POM 生成処理のジョブを投入すると，個々のジョブの処理時間が増える課題も明らかとなった．今後は性能要因を分析し，遅延時間増大の抑制を図ると共に，GPGPU 対応による処理時間の短縮を目指す．

参考文献

[1] F. Fleuret, J. Berclaz, R. Lengagne and P. Fua, *Multi-Camera People Tracking with a Probabilistic Occupancy Map*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **30**(2), 267/282, 2008.

[2] J. Berclaz, F. Fleuret, E. Türetken and P. Fua, *Multiple Object Tracking Using K-Shortest Paths Optimization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33**(9), 1806/1819, 2011.

[3] Y. Nishikawa, H. Sato and J. Ozawa, *Performance Evaluation of Multiple Sports Player Tracking System Based on Graph Optimization*, in IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD2017) in IEEE BigData2017, 2903/2910, 2017.

[4] Y. Nishikawa, H. Sato and J. Ozawa, *Multiple Sports Player Tracking System Based on Graph Optimization Using Low-Cost Cameras*, in IEEE International Conference on Consumer Electronics (ICCE), 2018.

[5] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, P. L. Mazzeo, *A Semi-Automatic System for Ground Truth Generation of Soccer Video Sequences*, in 6th IEEE International Conference on Advanced Video and Signal Surveillance, 2009.

[6] O. Tange, *GNU Parallel - The Command-Line Power Tool*, in ;login: The USENIX Magazine, **36**(1), 42/47, 2011.

[7] L. Dagum and R. Menon, *OpenMP: An industry standard API for shared-memory programming*, in IEEE Computational Science and Engineering, **5**(1), 46/55, 1998.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779/788, 2016.

[9] W. Gropp, E. Lusk, N. Doss and A. Skjellum, *A high-performance, portable implementation of the MPI message passing interface standard*, Parallel Computing, **22**(6), 789/828, 1996.

[10] `https://cvlab.epfl.ch/data/pom`

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

多人数追跡のための分散深層学習による高精度な検出にむけて

Towards Accurate Multiple Human Tracking Using Scalable Distributed Deep Learning

佐藤　仁 *1　　西川　由理 *1*2　　小澤　順 *1
Hitoshi Sato　　Yuri Nishikawa　　Jun Ozawa

*1産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

*2パナソニック
Panasonic

Scalable distributed deep learning is widely studies for open datasets for visual object recognition, i.e., ImageNet. In general, when we apply these scalable techniques to real applications, costly model training to domain-specific datasets is required for accurate recoginition; however, there are a few case studies for distributed deep learning except for ImageNet in terms of scalability, hyper parameter settings, and generalization, etc. This paper demonstrates our early activities on accurate human detection from soccer video images using distributed deep learning, as an instance towards accurate multiple human tracking from application-domain-specific video images.

1. はじめに

ビデオカメラで撮影された動画に対する多人数追跡は，小売店，物流センター，工場などでのヒト・モノの流れの分析・解析など幅広い応用が見込まれるため重要な技術である．とりわけ，サッカー，バスケットボール，アメリカンフットボールなどのチームスポーツの動画の分析・解析は，人物の動きが速い一方で，局所的な空間に人物が集中しオクリュージョンが頻発するため，高度な多人数追跡技術を必要とする事例の一つである．我々は，これまで，K-Shortest Path [Berclaz 11] のグラフ最適化アルゴリズムを用いた多人数追跡システムのスポーツ動画への応用を進めており，1 時間程度のバスケットボール動画の追跡においても現実的な時間で処理が可能であるという知見を得ている [Nishikawa 17].

多人数追跡は，一般に，(a) 動画のフレーム毎に人物を検出する，(b) 各フレームの人物検出結果を対応付けして軌跡を生成する，という 2 つのステップから成る．現状の我々のシステム [Nishikawa 17] では，(a) のステップにおいてロバストな人物検出を行うために，深層学習ベースの一般物体検出アルゴリズムである YOLO [Redmon 17] を用いている．しかし，人物検出の際の学習モデルは一般物体検出を目的とした COCO データセットより生成されたものを用いているため汎用の人物検出に特化しており，実問題を対象とした高精度な物体検出を行うためには，そのドメインに特化したデータセットより学習モデルを生成することが必要である．一方で，学習モデルを生成するためには大量の計算機資源を必要とし，その処理にも長時間を要することが問題となる．

分散深層学習は，複数の計算機を利用し並列分散処理により深層学習を行うことで，高速に学習モデルを生成することができると期待されている．特に，近年，ImageNet のデータセットに対する一般画像認識を対象に，大規模並列分散環境下での分散深層学習の高速化の取り組みが進んでいる．しかし，ImageNet 以外のデータセットを対象とし分散深層学習を用いた場合，学習率やバッチサイズなどのハイパーパラメタや汎化性能，最適化アルゴリズム，並列化性能などを評価した事例は多くはない．

連絡先: 佐藤　仁，産業技術総合研究所，東京都江東区青海 2－4－7　産業技術総合研究所臨海副都心センター別館，hitoshi.sato@aist.go.jp

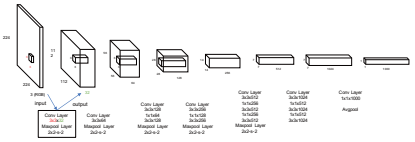


図 1: Darknet19

我々は，ドメインに特化したデータセットに対して分散深層学習により高精度な学習モデルの生成を高速に行うことで，多人数追跡の実問題への適用を目指している．本稿では，サッカー動画のデータを用いた多人数追跡を対象とし，この取り組みの概要と初期の評価について報告する．

2. ドメインに特化した多人数追跡

2.1 データセット

ドメインに特化した多人数追跡として，ここでは，サッカー動画のデータを用いた多人数追跡を対象とし，その際の人物検出の学習モデルを生成する例を取り上げる．ISSIA CNR では，イタリアのサッカー動画と人物を矩形としてアノテーションしたデータを公開している [D'Orazio 09]．これらの動画を基にして，人物を検出するための学習モデルを生成する．具体的には，動画のデータセット中から 5997 枚を画像として切り出し，それらの画像から総計 29860 個の人物の矩形情報を切り出すことで学習データセットとした．実際には，23836 個の人物の矩形情報を学習対象のデータセット，6024 個の人物の矩形情報をテスト対象のデータセットとした．

2.2 学習モデル

サッカー動画の人物検知のための学習モデルとしては，一般物体検出アルゴリズム YOLO で用いられている Darknet19 のネットワークを用いた [Redmon 17]．図 1 にそのネットワークの概要を記す．

3. 分散深層学習

3.1 分散深層学習フレームワーク

2.2 節の学習モデルの生成を分散深層学習で行うためのフレームワークとして ChainerMN [Akiba 17] を用い，データセットを分散して学習モデルを計算する「データ並列」の手

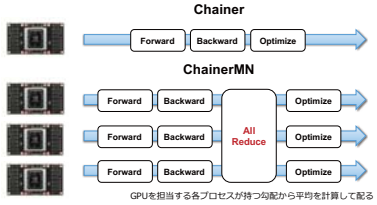


図 2: ChainerMN による分散深層学習

表 1: AAIC の計算ノードのスペック

CPU	Intel Xeon E5-2630L v4 1.80GHz (10 cores, HT-enabled) × 2
GPU	NVIDIA Tesla P100 × 8
Mem	256 GiB
SSD	Intel DC S3510 480 GB × 1

法を用いた．その概要を図 2 に記す．深層学習では，1) 予測を行いその誤差を計算（Forward 処理）し，2) 誤差を減らす方向の勾配を計算（Backward 処理）し，3) 勾配を用いて学習モデルを更新する．ChainerMN による分散深層学習では，1) Forward 処理，2) Backward 処理を行った後，分散して計算した勾配から平均を計算し配り直す AllReduce 処理を行う．そのため，計算性能だけでなくネットワーク通信の性能が重要となる．

3.2 ハイパーパラメタの設定

分散深層学習は精度を保って高速化することが難しい．具体的には，a) どのような勾配降下の最適化アルゴリズムを用いるか，b) どのような学習率やデータセットのバッチサイズなどのハイパーパラメタを設定するか，c) いかに局所的な最適解を避け，大域的な最適化を行うか，などが課題となる．今回は，[Goyal 17] の論文と同様の方法を用いた．具体的には，バッチサイズを GPU 台数 ×32 (1GPU 毎に 32)，学習率を 0.1× バッチサイズ /256，最適化アルゴリズムを Momentum SGD (momentum=0.9)，Weight decay を 0.0001 とした．実験は epoch 数を 10 として行った．

4. 初期評価

4.1 実験環境

初期の評価を産総研 AI クラウド AAIC 上で行った．表 1 に計算ノード 1 台のスペックを示す．計算ノード間は EDR Infiniband により Director Switch を介して Full-bisection Fat Tree 構成で接続されている．計算ノードの OS は CentOS 7.3 で構成され，Linux のカーネルは v3.10.0 である．また，ソフトウェアは，GCC v4.8.5, CUDA v8.0.61.2, CuDNN v6.0.21, NCCL v2.1.4, MPI は OpenMPI v2.1.2 をベースとして，ChainerMN v1.2.0, Chainer v3.4.0, CuPy v2.4.0 を用いた．

4.2 実験結果

図 3 に結果を記す．x 軸は GPU の台数，y 軸は 1 台の GPU を用いた際の実行時間を基準としたときのスケーラビリティを表す．概ねスケーラブルな処理性能を示すことを確認し，16 台の GPU で 6.41 倍の性能を達成した．

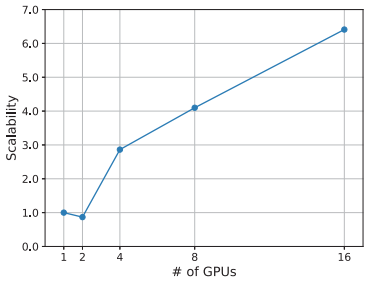


図 3: 実験結果

5. おわりに

ドメインに特化したデータセットに対して分散深層学習により高精度な学習モデルの生成を高速に行うこと目指し，サッカー動画のデータを用いた多人数追跡を対象とした事例について報告した．今後の課題としては，数千 GPU 規模の環境での大規模な分散深層学習の実行に向けて，a) ドメインに特化した大規模データセットの収集方法やラベル付け，b) 大規模なバッチサイズ下でのハイパーパラメタの設定，c) ネットワーク通信や I/O の最適化，などの検討を進めていきたい．

参考文献

[Akiba 17] Akiba, T., Fukuda, K., and Suzuki, S.: ChainerMN: Scalable Distributed Deep Learning Framework, in *arXiv.org e-Print archive*, pp. 1–6 (2017)

[Berclaz 11] Berclaz, J., Fleuret, F., Turetken, E., and Fua, P.: Multiple Object Tracking Using K-Shortest Paths Optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 9, pp. 1806–1819 (2011)

[D'Orazio 09] D'Orazio, T., Leo, M., Mosca, N., Spagnolo, P., and Mazzeo, P.: A Semi-automatic System for Ground Truth Generation of Soccer Video Sequences, in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 559–564 (2009)

[Goyal 17] Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Kaiming, Y. J., and Facebook, H.: Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour, in *arXiv.org e-Print archive*, pp. 1–12 (2017)

[Nishikawa 17] Nishikawa, Y., Sato, H., and Ozawa, J.: Performance evaluation of multiple sports player tracking system based on graph optimization, in *The IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD 2017) in IEEE BigData2017*, pp. 2903–2910 (2017)

[Redmon 17] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

K-Shortest Paths を用いた多人数追跡におけるデータ削減による高速化

Performance enhancement of multi-human tracking based on K-Shortest Paths by data reduction

秦　希望 *1　西川　由理 *3*4　中山　俊 *1　小澤　順 *3　藤澤　克樹 *2*3
Nozomi Hata　Yuri Nishikawa　Takashi Nakayama　Jun Ozawa　Katsuki Fujisawa

*1九州大学大学院　数理学府
Graduate School of Mathematics, Kyushu University, Japan

*2九州大学　マス・フォア・インダストリ研究所
Institute of Mathematics for Industry, Kyushu University, Japan

*3産業技術総合研究所
National Institute of Advanced Industrial Science and Technology, Japan

*4パナソニック株式会社
Panasonic Corporation, Japan

Object tracking is a challenging problem and it has been improving dramatically in recent years. In this paper, we perform parallelized multi-object tracking system. Object tracking problem has 2 difficulties; one is to detect objects collect, and the other is to track collect using the collect object detection. Jerome et al. performed a multi-object tracking system using K-Shortest Paths to avoid these problems efficiently. However, it is difficult to calculate in parallel because of the iterations calculation of shortest paths on the graph while changing the weight of graph. In our method, we divided time intervals to apply KSP method from Probability Occupancy Map(POM), which is also obtained via using KSP method. Performance evaluation shows our algorithm is 5.4 times faster than the original KSP with 87% accuracy.

1. はじめに

カメラやビーコンなどのデータを用いた人物の検出・追跡に関する研究は現在も盛んである. 人物応用として, 監視カメラの映像内での人物検出による防犯への活用や, また商業施設内における売上と客の分布の関係を見ることによってマーケティングに役立てることができるなど, その活用方法は多岐に渡っている.

この中で動体の検出については, 機械学習を用いることで人や車など検出対象の画像を学習した上で検出する方法が代表的である. 動画の画像情報のみから追跡を行うことは難しく, 特に多数の人物が入り混じっているような状況下での正確な検出は非常に困難であり, ID の入れ替わりや消失が起こることが多い.

上記を解決するための先行研究として, 機械学習で出力された検出結果を入力とする K-Shortest Paths(KSP) を用いたものが提案されている [1]. KSP を用いた方法では図 1 のような手順を取る. 対象領域上に点を配置したグラフを作成した上で, まず機械学習を用いた人物検出の結果から各地点ごとの人物存在確率 (Probability Occupancy Map: POM) [2] を求め, POM を基にした枝長を持つ時間拡大したネットワーク上で人数分の最短路を解くことで人物追跡を行っている. このことによって一時的な多重検知・未検知等の誤差を修正しながら人物の追跡を行うことができる.

本研究は, この手法を適用するべき場面を POM から選択することで KSP の適用範囲を局所的なものとし, 計算の高速化・並列化を目的としたものである. KSP 内では追跡対象とする動画の撮影範囲・時間幅に応じた大きさのグラフ上で繰り返し最短路を解いており, KSP そのものを並列化することは難しい. そこで本研究では, 混雑が起きているかどうかを POM から判断し, その結果を KSP の適用対象かどうかに用いることで人物追跡の並列化を行い, 計算時間を短縮した. 人物検出の精度には様々な要因が関連して来るが, 画像内の人々の混雑度合いはその要因の一つと考えられる. 一般に, 混雑が発生している場合はそうでない場合と比較して検出の精度が落ちる. 検出の精度が高い部分で KSP を分割することで, 分割した際の整合性を保ちつつ計算の分割を行うことが可能となる. また分割方法は POM にのみ依存しているため, 本追跡・分割手法はカメラやビーコンなどの位置取得の方法によらず, 汎用的な適用が可能である.

連絡先: n.hata.zlqk@gmail.com

2. KSP による動線追跡方法

動線追跡方法として用いた KSP による手法 [1] では, 人物検出を行った後, 整合性が取れる様に検出結果を繋ぐことによって動線追跡を実現している. 人物検出に際しては対象領域を撮影した複数の映像を入力として用いており, 各映像内の各時刻における検出結果を地点ごとの存在確率として出力しておく. 出力の際には, 図 2 のようにあらかじめ検出領域を矩形領域に各矩形内の人数が高々1 人となるよう十分小さく分割した上で, 各矩形領域内の存在確率を出力する.

2.1 問題設定

分割された矩形領域の個数を K, 解析対象とする映像のフレーム数を T とする. 矩形領域の大きさ及び 1 秒あたりのフレーム数を考慮し, 時刻 t において領域 k に存在した人物は次の時刻 $t+1$ で k に隣り合う領域 $\mathcal{N}(k) \subset \{1, \ldots, K\}$ に移動できるものとする. このとき対象領域のグラフ $G = (V, A)$ が

$$V := \{1, \ldots, K\}$$
$$A := \{(i, j) \in V^2 \mid i = 1, \ldots, K, j \in \mathcal{N}(i)\}$$

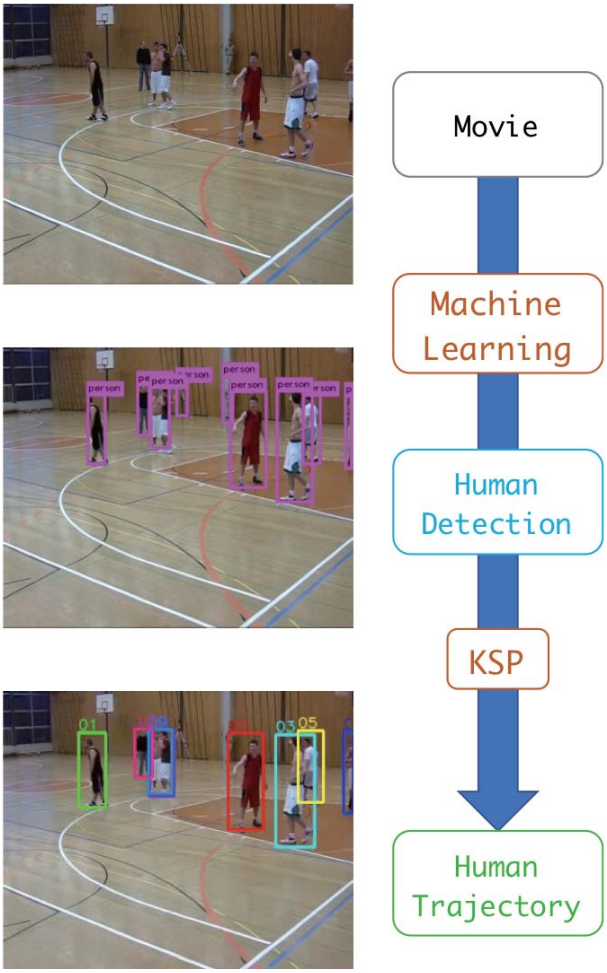


図 1: KSP による動線追跡

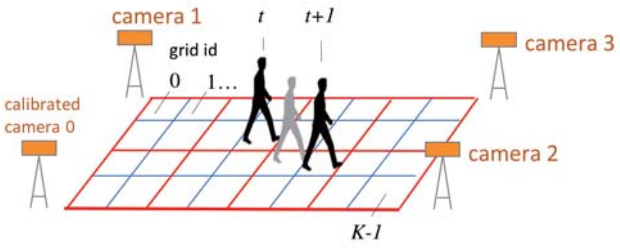


図 2: 追跡空間の分割

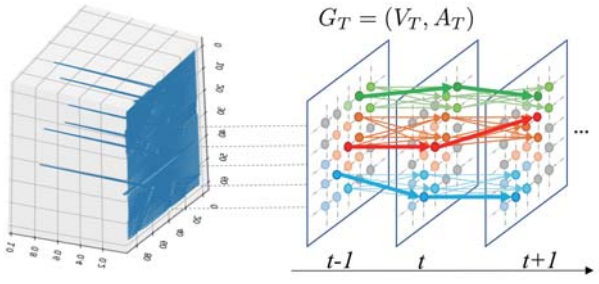


図 3: POM 及びネットワークの時間拡大

によって定まり, これを各フレームに対して時間拡大したネットワーク $G_T = (V_T, A_T)$ が

$$V_T := \{1, \ldots, T\} \times V$$

$$A_T := \{((t,i),(t+1,j)) \in V_T^2 \mid (i,j) \in A, t \in \{1,\ldots,T-1\}\}$$

によって定まる. 以下では (t,i) から $(t+1,j)$ への枝 $((t,i),(t+1,j))$ を単に $e^t_{i,j}$ と表す. G_T の各点 (i,t) は時刻 t と物理的な位置 i の組によって定まっており, 各枝 $e^t_{i,j}$ は時刻 t において i から j へと 1 フレーム間で移動が可能であることを表している (図 3). あらかじめ求めてある, 時刻 $t \in \{1,\ldots,T\}$ における領域 $i \in V$ 内の人物存在確率を ρ^t_i とする. このとき人物追跡の問題はグラフ G_T 上において尤度を最大化するような経路を出力する問題として定式化することができる.

2.2 整数計画問題としての定式化

定式化に際しては, 人流を G_T 上のフローと考えて定式化を行う. G_T 上のフローを表す変数 $f^t_{i,j}$ を

$$f^t_{i,j} = \begin{cases} 1 & (\text{時刻 } t, \text{領域 } i \text{ に人物が居り,} \\ & \text{その人物は次の時刻に領域 } j \text{ に移動}) \\ 0 & (\text{その他}) \end{cases}$$

となるように定める. このとき時刻 t において領域 i に存在する人数は $m^t_i := \sum_{j \in \mathcal{N}(i)} f^t_{i,j}$ である. したがって実行可能なフロー $\{f^t_{i,j}\}$ に対する対数尤度は

$$\begin{aligned} &\log \prod_{t,i} \left(\frac{{\rho^t_i}^{(1-m^t_i)}}{(1-\rho^t_i)^{m^t_i}} \right) \\ &= \sum_{t,i} \log \left(\frac{(\rho^t_i)^{(1-m^t_i)}}{(1-\rho^t_i)^{m^t_i}} \right) \\ &= \sum_{t,i} (1-m^t_i) \log (1-\rho^t_i) + m^t_i \log \rho^t_i \\ &= \sum_{t,i} m^t_i \log \frac{\rho^t_i}{(1-\rho^t_i)} + \log (1-\rho^t_i) \end{aligned}$$

となる. したがって目的関数は

$$\sum_{t,i} m^t_i \log \frac{\rho^t_i}{1-\rho^t_i}$$

である.

制約にはフロー整合条件と各領域に存在する人数についての制約がある. 前者を単一始点・単一終点で表現するためにスーパーソースとスーパーシンクと呼ばれる仮想点 v_{source}, v_{sink} を追加する. 時間拡大ネットワーク上において, 人物が新しく出現する点は領域の端点及び $t = 1$ に該当する点であるため, v_{source} からそのような点 (t,j) に向けて枝 $e^t_{v_{source},j}$ を張る. こうすることにより, フローの湧き出しを v_{source} のみであるとして問題を設定することができる. v_{sink} についても同様である. したがって制約は

$$f^t_{i,j} \geq 0 \qquad \forall t,i,j$$

$$\sum_{j \in \mathcal{N}(i)} f^t_{i,j} \leq 1 \qquad \forall t,i$$

$$\sum_{j \in \mathcal{N}(i)} f^t_{i,j} \leq \sum_{k \in \mathcal{N}(i)} f^{t-1}_{k,i} \qquad \forall t,i$$

$$\sum_{j \in \mathcal{N}(v_{source})} f^t_{v_{source},j} \leq \sum_{j \in \mathcal{N}(v_{sink})} f^t_{j,v_{sink}} \qquad \forall t,j$$

となる. これらの制約によって, 結果として得られるフローは v_{source} から v_{sink} への交わらない経路となり, これらを追跡結果とすることが可能となる.

[1] 内で, この問題内の制約は全ユニモジュラ行列として表されることが示されており, そのため上記の問題は変数を実変数に緩和することが可能である. また同論文内では上記の最適解を得るための方法として, 作成したネットワーク上で枝の重みを変更しながら最短路を繰り返し解くことによって高速に最適解を得る方法について示されている.

3. KSP 適用範囲の分割方法

前述の方法による人物追跡の方法では, 最短路の繰り返し計算による高速な計算手法が既に確立されてはいるが, リアルタイムでの活用には至っておらず, 即時的な活用には並列化などの工夫が求められる. 特に対象となる領域が広い場合には, 考えている制約上グラフの点数を増やす必要があり, 最短路を解く際のグラフの点数が膨大となってしまい計算が遅くなる. 一方, ネットワーク上での経路導出の際には 1 経路を出す毎に枝の重みを変更しながら計算を行っているため, 導出に対しての直接の並列化は難しい. そこで本研究では上記のアルゴリズムを適用する時間を分割するための手法を提案する.

本研究では, 追跡の難しさや追跡に成功したときの活用のし易さ等の観点から, 主にスポーツを撮影した動画への適用を考えている. この際, 特に球技のような 1 つの物体に対して多人数が集まるスポーツにおいては, 領域の中での人の分布が偏ることが多い. 人が集中している領域内では人物追跡やそれに用いる人物検出が難しく, 人の目で見ても追跡がままならないような状況が生まれる一方で, その他の領域では人が少なく, 追跡は非常に容易であると考えられる. このため映像内から混雑箇所の予想をすることができれば, 混雑している箇所では検出誤差の補正を行うために局所的に KSP を適用し, これらの結果を繋ぎ合わせることで全体の追跡を復元することができる.

混雑の検知は各フレームごとに与えられる POM を用いており, 各点に対してその周囲にある点における POM の和が一定以上である場合に混雑が起こっているとみなす. 例えば図 4 では, 中央及びその左側の 2 点で混雑が検出される. また混雑検知結果は一般には連続的ではないため, 混雑検知から一定時間以内に再度検知がなされた場合には混雑が継続しているとみなす. このしきい値のことを本研究内ではインターバルとよぶことにする. 今回の KSP 分割に際しては, インターバルの設定により分割されている混雑の時間帯に対し, 混雑が終了した時刻と次の混雑が発生した時刻の中間で KSP を分割する (図 5 下部).

4. 実験結果及び考察

本実験においては, Multi-camera Pedestrians Video データセット [4] の中からバスケットボールに関する映像を選択し実験に用いた. 計算機環境を表 2, 動画に関する情報を表 1 に示す。前ページで掲載してある画像の様に, 動画内ではハーフコートでの競技の様子が撮影されており, 混雑が頻発していることが見受けられる. 人物検出に関しては YOLO [3] を, POM の生成及び KSP の適用についてはそれぞれ [2], [1] の実装を用いて計算を行った.

まず POM から得られた混雑検知結果を図 5 の上部で示す. 横軸はフレーム数を表しており, 図の上部に出ている番号では混雑が検知された時間を表している. また図の下部ではインターバルとして 3 秒を選択した場合, すなわち前後の混雑が 3

空間の広さ	15m × 15m
グリッド数	100 × 100
人数	13 人
KSP の深度	1
動画の長さ	360 秒
1 秒あたりフレーム数	25fps
カメラ数	4

表 1: 実験時の設定及び環境

CPU	Intel Xeon E5-2687W v4
アーキテクチャ名	Broadwell
CPU 周波数	3.00GHz
CPU 数	24(12 コア x 2 ソケット)
RAM	512GiB

表 2: 計算機環境

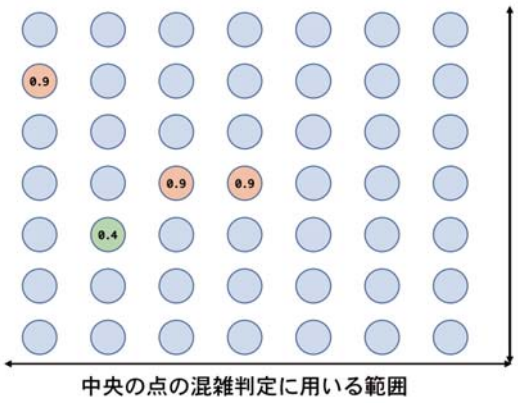


図 4: 各点に対する混雑判定に用いる範囲

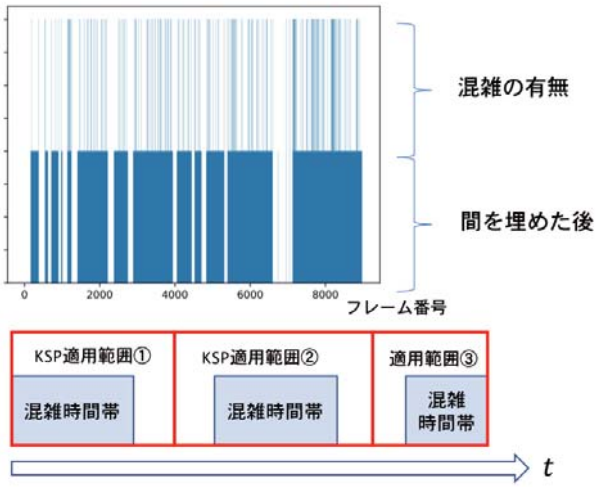


図 5: 混雑を基準とした KSP の分割方法

秒以内であった場合に混雑が継続していたとみなした場合の結果が示されており, 上部と比較して連続的になっている事が分かる. インターバルの設定により混雑が発生している時間帯は 16 箇所に分割されており, したがって KSP 適用範囲も 16 分割されている. この各区間において KSP を適用した結果及び分割前との比較結果を表 3 に示す.

まず計算時間については, 各区間での計算時間は最大で 164.62 秒となり, 分割前の KSP にかかっていた 890.52 秒を大きく短縮する結果となった. 各区間での計算は全て独立に行うことができ, 内部での並列化は行っていないため, 全体で約 5.4 倍の高速化を実現することができる.

次に分割前と比較した精度であるが, 追跡に際していずれかの失敗をしているフレームは約 13.5%である 1216 フレームとなり, 精度の低下はあまり見られない結果となった.

このうち位置を見誤るケースはグリッドの大きさが人間の大きさに比べ小さいため一概に誤りであるとは言えず評価が難しい. 人数を見誤るケースについては, 長時間他人と重なり続けたために起こる未検出が大きな原因となっている. 未検出が起きている場合には, KSP の性質上未検出の前後で検出に成功している必要があり, この検出成功時間が短い場合には逆に成功した部分を過剰な検出とみなしてしまう.

例として 5301 フレーム目における図 6 の手前部分に未検出の人物が居るが, この人物は重なっている間しばらくの間未検出である. この部分については 4768〜5353 フレームにて KSP の適用を行っており, この未検出が影響した結果 5301 フレームから 5353 フレームまでの約 2 秒間追跡が行えていない. このため分割に際しては未検出の取扱いが今後の主要な課題となる. 特にスポーツの映像に関しては図 6 のように人が重なって起こる未検知が頻発するため, 追跡を行うまでの全体の中で重なりに強い人物検出を行う必要がある.

5. まとめ・今後の課題

本研究においては, 誤検出を考慮した動線追跡アルゴリズムである KSP を並列化するための分割手法を提案した. その結果, 分割前と比べて 9 割以上の精度を保ちつつ 5 倍以上の高速化に成功した. 今後の課題として, 分割した際に前後の整合性を取るためのアルゴリズムの作成, 未検出を考慮した分割方法の改善, 空間に対する KSP の並列化が挙げられる.

1 つ目については, 今回行った KSP の分割に関して前後の人数や位置が合わないことが起こるため, 全体で統一された人物追跡を行うために必要である. 精度を損なわない分割ができている場合には分割の前後で最寄りの人物を同一とみなすなど単純な接続が可能であると見込まれるため, 2 つめに上げた分割手法の改善と合わせて研究を行っていく必要がある.

分割手法の改善に関しては, 考察にて述べたように未検出が起こる区間をある程度含むような区間で分割する必要がある. 映像であればカメラの位置, ビーコンであれば受信機の位置など, データ取得方法の特性に合わせて分割手法を考えることが望ましい.

最後に空間に対する並列化であるが, 今回は混雑箇所に関わらず混雑が発生しているか否かのみによって分割を行っているため, 混雑の発生地点を考慮していない. サッカーなど対象となる領域が広い場合には, KSP によって対処する混雑は全体の中でごく一部の狭い領域で発生するものであり, その際に全体を扱う必要はないと考えられる. 混雑発生領域とそうでない領域を分けることで, 精度を保ったまま更なる高速化が期待される.

分割前の計算時間	890.52 秒
分割後の最大計算時間	164.62 秒
全フレーム数	8999 フレーム
正解数	7783 フレーム
位置がずれた数	717 フレーム
人数がずれた数	499 フレーム
連続した未検知	3 箇所

表 3: 計算実験結果



図 6: 未検出の例

参考文献

[1] Jérôme Berclaz, François Fleuret, Engin Türetken, and Pascal Fua, *Multiple Object Tracking using K-Shortest Paths Optimization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 33, Issue 9, September 2011.

[2] François Fleuret, Jérôme Berclaz, Richard Lengagne, and Pascal Fua, *Multi-Camera People Tracking With a Probabilistic Occupancy Map*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 267-282, February 2008.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, CVPR 2016, OpenCV People's Choice Award, June 2016.

[4] http://cvlab.epfl.ch/cms/site/cvlab2/lang/en/data/pom

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

高抽象度赤外線画像センサを用いた行動認識システムの研究

An action recognition system using a highly abstracted infrared array sensor

村上 奨悟[*1]
Shogo Murakami

矢入 郁子[*1]
Ikuko Yairi

[*1] 上智大学院 理工学研究科 理工学専攻 情報学領域
Graduate School of Science and Technology, Sophia University

According to the population growth of the aged people, the needs of action recognition system for elderly care are increasing. Upcoming IoT era, the privacy issues of the vision based human action recognition is still primary concerns. With due consideration of the privacy protection of users, we have been proposing a system using an infrared array sensor with high abstraction and low resolution. This paper reports the developed simulator to reproduce the behavior of the human, the data collection under various situations using infrared sensor and the simulator, and classification results whose accuracy was more than 90% using machine learning.

1. はじめに

現在，日本は超高齢化社会に突入し，高齢者の行動の見守りへのニーズが高まっている．見守りシステムの代表例として，カメラ画像による行動認識[Rohrbach12][Kamal13]が挙げられるが，私的空間でのカメラを利用は，対象のプライバシー保護の点でリスクが高い．プライバシーに配慮した行動認識手法はいくつかあるが[Weinland06][Najafi14]，筆者らはカメラの代わりに低画素の赤外線センサを用いることで，プライバシーの問題を乗り越えながらも対象の行動を認識できるシステムの作成を目指してきた[岡田 13]．本稿ではシステムの改良，データ処理の工夫による行動の認識精度の向上，センサの取り付け位置の違いによる認識精度への影響の調査について報告する．

2. 赤外線画像センサを用いた行動認識システム

筆者らはこれまで，赤外線アレイセンサを搭載したデバイスを壁に設置し，そのデータをネットワーク経由で集積・解析し，屋内での行動の推定を行うシステムを開発してきた．赤外線アレイセンサが出力するデータは抽象度が高いため，第三者が見ても人物や部屋の様子を知ることはできない．そのため，データ漏洩によるプライバシーの侵害の恐れがなく，ネットワーク経由でのデータのやり取りが可能である．

2.1 高抽象度赤外線センサを用いたデバイス

本研究では赤外線アレイセンサ Grid-eye を，Raspberry Pi に搭載することでデバイスを実現した．このデバイスを図 1 に示す．Grid-eye は 8×8，計 64 の赤外線の発信・受信が可能な素子を持ち，赤外線を反射した物体の摂氏温度の 5 倍に対応した 64 個の整数値をテキストファイルとして出力する．

当システムは，屋内の 8×8 地点の温度値を取得し，その温度差を利用して対象人物の位置や姿勢の認識を行う．これらの温度値は第三者が入手しても抽象度が高いため個人を特定することは不可能であるため，見守られるユーザにとってプライバシーを侵害される心理的負担は少ないと考えられる．

連絡先:村上 奨悟，上智大学院，murakami@yairilab.net

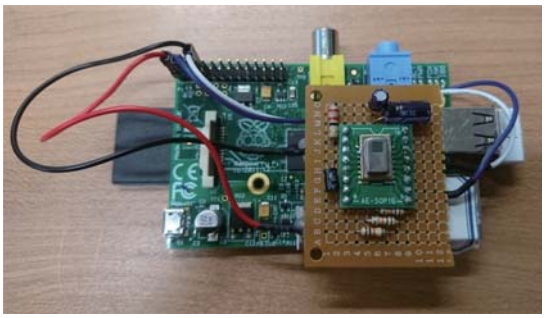


図 1 赤外線センサを搭載した行動認識デバイス

2.2 シミュレータによる機械学習用データの生成

本研究では当初，行動認識のための分類学習を行うにあたり，被験者を募って実際の室内での行動データを当システムによって収集し，同時に正解となる行動をラベル付けするためにカメラで撮影することを行ってきた．しかし，部屋のサイズ，センサの配置，ユーザの身長・体型などの様々なパターンの学習データを効率よく収集するために，実際の室内での当システムの出力と同様の出力が得られるシミュレータを作成した．図 2 に人体とセンサを再現したシミュレータの画面を示す．シミュレータで生成したデータを利用する場合，再現性の高さが重要な問題になるため，実際にセンサを用いて取得したデータと，シミュレータを用いて取得したデータの比較を行った．図 3 に示した比較データの白色部分は高温部，黒色部分は低温部である．この結果より，両データの形状の相違は少なく，再現性は高いと言える．

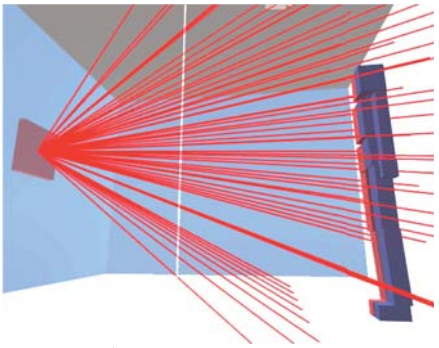


図 2 人体とセンサの再現シミュレータ

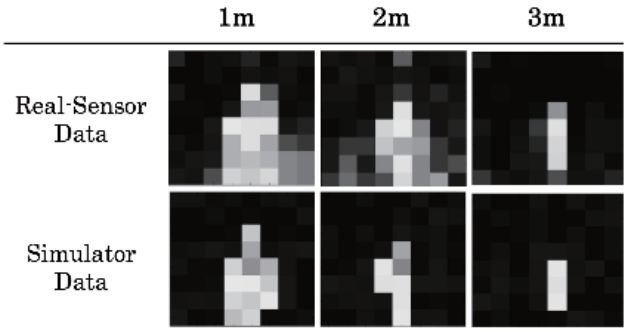


図 3 実データとシミュレータデータの画像比較

3. 分類学習による行動認識

3.1 実データの学習度分類

実空間でデータの収集を行い，分類器の性能を評価した．収集を行った部屋は室温約 13℃，6 畳の和室で，センサは全身が映る床から 140cm の場所に設置した．被験者は身長 170cm の男性であった．収集は，センサから 1m~3m の位置，センサの視野角内で被験者に「立つ」「座る」「寝る」の何れかの姿勢を取ってもらうことで行い，各姿勢につき 30 分程度のデータを収集した．

これらの収集データは，Deep Convolutional Neural Network(以下 DCNN)を用いて分類を行った．DCNN の構成は python の機械学習用ライブラリである chainer を用いた．この分類器に収集データを入力した結果を表 1 に，評価を表 2 に示す．結果を見ると，精度・再現率の 2 値の調和平均である F 値は 87%程度であった。誤分類に着目すると，「立つ」と「座る」，「座る」と「寝る」のような温度分布の近い姿勢に多く誤分類が見られた．

システム運用上は，最低でも 90%以上の分類精度が望ましい．そこで，収集した温度データを，RGB 値を持った 8×8 のカラー画像データに変換し，DCNN への入力を増やすことで誤分類を減らすことができないかを試みた．データのカラー化は,色相空間に温度を写像することで行った．このデータの変換対応を図 4 に示す．この変換を行った同じデータの分類結果を表 3 に，評価を表 4 に示す．表 1 と比較すると総じて誤分類は減少しており，F 値もすべての状態において 90%以上となった．この結果よりデータのカラー画像化が誤分類の減少に効果的であることが判明した．以降，DCNN へ入力されるデータは全てにカラー画像化したデータを利用する．

表 1 DCNN による行動分類結果

Recognition	Answer 立つ	座る	寝る
立つ	2512	235	15
座る	260	3665	368
寝る	98	493	3131

表 2 DCNN による行動分類評価

	立つ	座る	寝る
精度	0.90948	0.85371	0.84121
再現度	0.87526	0.83428	0.89100
F値	0.89204	0.84388	0.86539

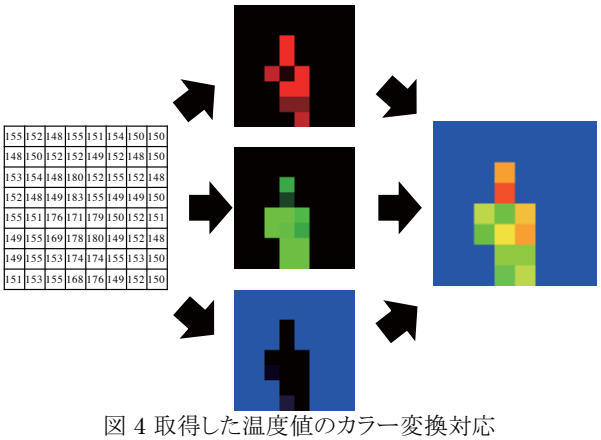


図 4 取得した温度値のカラー変換対応

表 3 DCNN によるカラーデータの行動分類結果

Recognition	Answer 立つ	座る	寝る
立つ	2658	110	18
座る	143	3968	107
寝る	69	315	3389

表 4 DCNN によるカラーデータの行動分類評価

	立つ	座る	寝る
精度	0.95405	0.94073	0.89822
再現度	0.92613	0.90325	0.96442
F値	0.93988	0.92161	0.93014

3.2 仮データからの学習と実データの分類

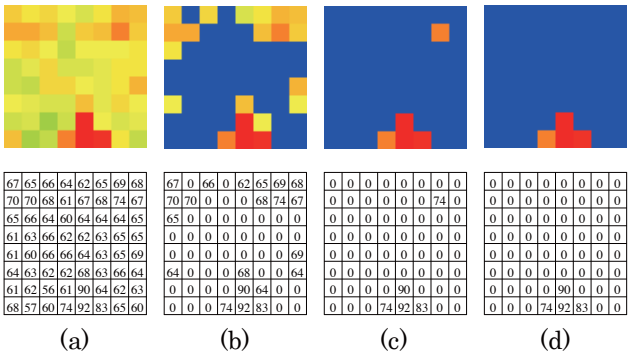
シミュレータで生成したデータ(仮データと略記)を前述したカラー画像化して学習を行い，3.1 で用いた実データの分類を試みた．表 5 に分類結果を，表 6 に評価を示す．表 3,4 と比較して分類精度が低下した原因として，人物以外の背景のノイズの影響が考えられた．そこで，フレーム間の温度差分から背景部分を検出し，背景を除去できるアルゴリズムを作成した．このアルゴリズムの処理手順を図 5 に示す．処理の手順としては，まずデータ全体の平均値を求め，その値以下のデータ列を 0 に置き換える[図 5(a)→図 5(b)]．続いて各データでの平均値を求め，その値以下のデータを同様に 0 に置き換える[図 5(b) →図 5(c)]．最後に温度が存在し得ない部分の除去を行う[図 5(c) →図 5(d)]．

表 5 仮データによる実データの行動分類結果

Recognition	Answer 立つ	座る	寝る
立つ	1916	297	128
座る	949	3669	681
寝る	5	427	2705

表 6 仮データによる実データの行動分類評価

	立つ	座る	寝る
精度	0.81845	0.69239	0.86228
再現度	0.66759	0.83519	0.76977
F値	0.73536	0.75711	0.81341



(a) (b) (c) (d)

図 5 背景除去の処理手順

背景除去アルゴリズムを学習用の仮データ，分類評価用の実データの双方に施して分類を行なった結果を表 7 に，評価を表 8 に示す．結果を見ると，全ての分類において精度が向上したことがわかる．この結果から，シミュレータによって得られたデータを用いて学習器を用意し，実データを分類する場合，双方のデータの前処理を工夫することにより，高精度での分類が可能であることがわかった。

表 7 背景除去を行ったデータの行動分類結果

Recognition	Answer		
	立つ	座る	寝る
立つ	2306	0	0
座る	564	4194	842
寝る	0	199	2672

表 8 背景除去を行ったデータの行動分類評価

	立つ	座る	寝る
精度	1.00000	0.88369	0.95703
再現度	0.95818	0.96858	0.87478
F値	0.97864	0.92419	0.91406

4. センサ配置による分類精度への影響の検証

見守り環境へのシステム設置時には，最も行動の分類精度が高くなる配置が求められる．そこで，配置の高さによる分類精度への影響を検証する為，シミュレータ内で，センサの高さを 10cm 刻みで 50~170cm と変えた 13 パターン，人物の身長を 170/160/150cm の 3 パターンを組み合わせたケースでのデータを用意し，調査を行なった．結果を図 6 に示す．170/160/150cm の 3 パターンともに，最も精度が高くなるのは人物の身長より 50cm 低い地点であることがわかった．

次に，センサに傾きが発生した場合，どの程度の誤差が発生するのかを検証した．モデルは高さの実験と同様に 170/160/150cm，センサ高は各モデル高で最も精度が高かった 120/110/100cm に設定して行った．傾きの幅は，センサの上面もしくは下面が壁から 1cm 程度離れる±5°の傾きに設定した．この分類結果を表 9 に示す．結果を見ると，傾きのない場合に比べ，著しく精度が低下していることが分かる．そこで，傾きによる誤分類の増加が何に起因しているのかを明らかにするため，モデルが 170cm の場

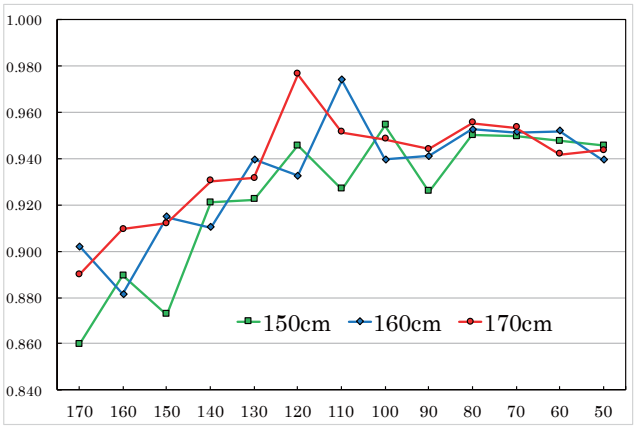


図 6 様々なセンサ高度に対する精度

合に着目し，それぞれの傾きでどのような誤分類が多く発生しているのかを調べた．-5°傾いている場合の分類を表 10 に，+5°傾いている場合の分類を表 11 に示す．表を見ると，センサが下に傾いているデータは「寝る」の「座る」への誤分類と，「座る」の「立つ」への誤分類が非常に多くなっている．これは，センサが下に傾いたことにより，相対的に人体モデルの映る位置が高くなり，「座る」「寝る」の姿勢が傾きのない場合に比べて高い位置に映り込み，傾きのない場合の「立つ」「座る」の高さに近くなってしまったものと推定できる．逆に，センサが上に傾いているデータは「立つ」の「座る」，「座る」の「寝る」への誤分類が圧倒的に多い．これも同様に，センサが上に傾いたことにより，相対的に人体モデルの映る位置が低くなり，「立つ」の姿勢が傾きのない場合に比べ低い位置に映り込み，傾きのない場合の「座る」の高さに近くなってしまったものと推定できる．

表 9 上下傾斜とモデル身長の分類精度(F 値)

	150cm	160cm	170cm
0°	0.95434	0.97393	0.97447
+5° (下に傾く)	0.81661	0.78921	0.79078
-5° (上に傾く)	0.83182	0.87895	0.86658

表 10 上に5度傾いたデータの分類結果

Recognition	Answer		
	立つ	座る	寝る
立つ	9484	4037	0
座る	200	6643	1920
寝る	86	92	7539

表 11 下に5度傾いたデータの分類結果

Recognition	Answer		
	立つ	座る	寝る
立つ	2306	0	0
座る	564	4194	842
寝る	0	199	2672

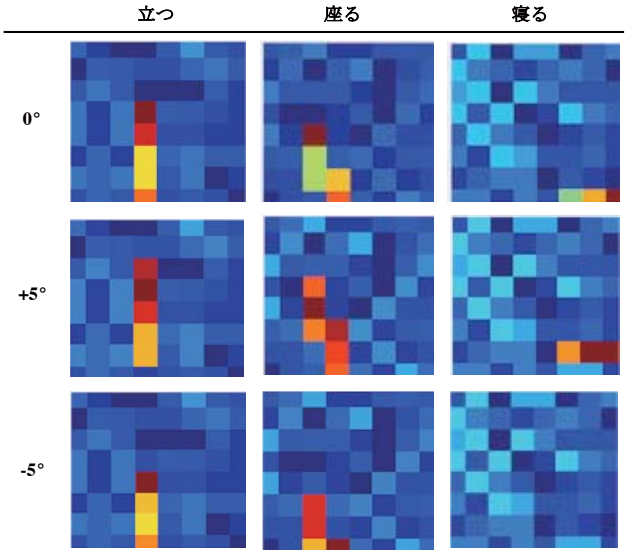
実際に，この推定が正しいのかを確かめるため，傾きのないデータと，±5°傾いたデータを可視化して各姿勢の比較を行うことにした．その結果を以下の表 12 に示す．図を見ると，傾きがない場合に比べて+5°の場合は平均的に一列ほど上に画素がずれており，「座る」の高温部分の頂点が，傾きがない場合の「立つ」の高温部分の頂点と同程度の高さになっているのがわかる．また，-5°の場合は平均的に一列ほど下に画素がずれており，「立つ」の高温部分の頂点が，傾きがない場合の「座る」の高温部分の頂点と同程度の高さになっているのがわかる．当センサはプライバシーに配慮し画素数が少ないものを利用しているため，分類は各列における高温分布に強く依存し，たった一列の差でも大きく分類に影響を与えてしまう．故にこれらの傾きによって生じた高温分布のずれが，学習用データとの齟齬を発生させ，分類精度を低下させてしまった原因であると考えられる．これらの分析結果から，傾きは分類結果に大きな誤差を与えることが判明した．

5. 考察

本稿では，実験と分析の結果，十分な前処理を行うことで，シミュレータで学習した認識システムが，実センサのデータを高い精度で分類可能であることがわかった．このシミュレータにより，カメラ撮影による学習データの収集といった工程を不要にし，多種多様なデータ収集の大幅な高速化・簡略化を実現した．それに加え，シミュレータの高い再現度から，「転倒」のような実空間では取得が難しい外れ値行動の検知に有用であるという展望も得られた．

しかし，現時点では人物を自律的に実際の人間のように自然に行動させる機能の実装には至っていない．見守りシステムに不可欠である「外れ値行動検知」を行うにためにも，この機能の実装は今後の課題である．また，今回行なった実験で判明したセンサの傾きに対する脆弱性についても，設置を工夫するか，傾いた状態での学習用データの収集などの対策が必要であると言える．

表 12　上下の傾きと温度分布の関係



6. おわりに

本稿では，見守り対象者のプライバシーを侵害せず行動を認識できるシステムと，行動認識のための機械学習用データを収集するシミュレータの構築，及びセンサの高さや傾きによる誤差の評価結果について述べた．実験と分析の結果，十分な前処理を行うことで，シミュレータで学習した認識システムが，実センサのデータを高い精度で分類可能であることがわかった．今後は，シミュレータの性能向上や学習アルゴリズムの工夫を引き続き行いながら，高齢者が暮らす実環境での見守りを実現する予定である．

謝辞

本実験に参加・協力して下さった全ての皆様に感謝いたします．本研究は平成 23，24 年度テレコム先端技術支援センター SCAT 研究助成金，科学研究費基盤研究 B(17H01946)，科学研究費挑戦的萌芽研究(16K12537)のもと行われました．

参考文献

[Rohrbach12] M. Rohrbach, S. Amin, et al.: A Database for Fine Grained Activity Detection of Cooking Activities, IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[Kamal13] Jalal, Shaharyar Kamal and Daijin Kim: A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments, Sensors, 14(7), 11735-11759 , 2013.

[Weinland06] Daniel Weinland, Remi Ronfard and Edmond Boyer: Free viewpoint action recognition using motion history volumes, Sensors, Computer Vision and Image Understanding, Volume 104, Issues 2–3, Pages 249-257, 2006.

[Najafi14] B. Najafi, K. Aminian, et al.: Wearable sensor data fusion for remote health assessment and fall detection, Electro/Information Technology, 2014 IEEE International Conference on, pp.303-307, 2014.

[岡田 13] Okada, R., Yairi, Katou, Y.: An indoor human behavior gathering system toward future support for visually impaired people, ASSETS '13 Proceedings of the 15th International ACM SIGASESS Conference on Computers and Accessibility, Article No. 36, 2013.

深層学習法による源氏絵の流派推定

Detection of a Painter School with Deep Learning Method: A Case Study of Scenes from the Tale of Genji

加藤拓也[*1] Takuya Kato　稲本万里子[*2] Mariko Inamoto　小長谷明彦[*1] Akihiko Konagaya

[*1] 東京工業大学 Tokyo Institute of Technology　[*2] 恵泉女学園大学 Keisen University

Detection of a painter school often becomes controversial in antique pictures such as scenes from the Tale of Genji painted in 17th century or before. This paper proposes a noble approach to detect a painter school with a deep learning system focusing on faces appeared in the pictures. The method successfully classified the pictures painted by Tosa-school, Kanoh-school, Iwasa-school and others with the accuracy of 96.5 percent in five-fold cross validation.

1. はじめに

源氏絵とは,『源氏物語』を題材とした絵画の総称である. 源氏絵の絵師には土佐派をはじめ狩野派,岩佐派,など多数の流派があり,各流派独自の個性がある. これまでに見つかった作品にはどの流派の絵師が描いたかわからないものがあり,美術史の専門家たち中でも意見が分かれている[1]. そのため新たな知見から流派を判断する手法が望まれる. 近年,深層学習の一種である畳み込みニューラルネットワークの画像分類能力の向上は著しく,一部の分野では人間よりも高いという報告もある. 深層学習では特徴量がデータから学習されるため,これまで人間が発見していない特徴量に基づく分類をすることが期待できる. 本稿では,深層学習による物体検出手法である YOLOv2[2] を用いて顔を自動認識し,畳み込みニューラルネットワークにより流派を推定し,Grad-CAM[3]を用いて可視化を行った.

2. 接近法

2. 1 学習データ

学習データには,源氏絵研究の第一人者であり,2014 年 2 月に急逝された東京芸術大学名誉教授田口榮一氏が過去に調査し,リバーサルフィルムで撮影した 35 ㎜スライドをデジタルデータ化した画像データを用いた.本画像データの特徴は全体図に加え,人物の顔写真の接写画像が含まれている点にある.この画像データを用いることにより,数センチメートル四方の大きさで肉眼でもわかりづらい源氏絵の顔画像の特徴を捉えることが可能となっている.

2. 2 顔自動検出

源氏絵に現れる貴族の顔は「引目鉤鼻」と呼ばれる独特な画風で描写されている.このため,写真画像で用いられる両目と口を結ぶ三角形のような顔画像の特徴量は顔の自動検出には必ずしも有効ではない.本研究では,顔画像を「引目鉤鼻で描かれている物体」とみなし,深層学習の物体検出のモデルの一つである YOLOv2 によって切り出した(図1).ただし,後ろを向いているものや真横を向いているものは除外した. また,画像のサイズは 224×224 ピクセルにリサイズし,正方形でないものは白画素でパディングした.

2. 3 流派推定

連絡先:小長谷明彦,東京工業大学情報理工学院,kona@c.titech.ac.jp

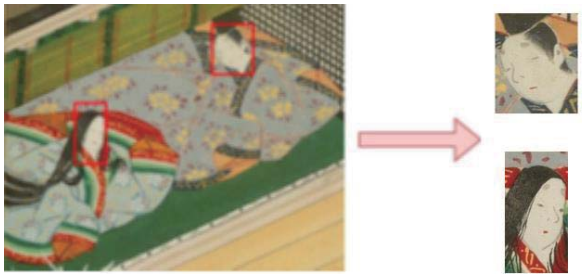


図 1 顔画像の切り抜きの例:「源氏物語画帖」,メトロポリタン美術館所蔵

土佐派,狩野派,岩佐派の絵師が描く顔画像はそれぞれの流派で特徴があるといわれているが,その違いは極めて軽微であり,絵師の判定には高度な経験と知見が必要である.このような微妙な違いを識別するために,豊富な特徴表現を学習することで知られている事前学習済 VGG[4]モデルを用いた.また,画像の特徴を抽出する畳み込み層と特徴の位置ずれを吸収するプーリング層を変化させることで,源氏絵の顔画像を最適に分類するニューラルネットワークモデルを求めた.

2. 4 可視化

深層学習は特徴量がモデルに内包されるため,高い精度を示しても,本当に学習したかどうかがわからない.そのため,Grad-CAM を用いて畳み込み層またはプーリング層の出力のどの部分が分類に寄与すると考えられるかを可視化した.

表1 源氏絵画像

	土佐派	狩野派	岩佐派	その他
画像数	114	125	120	201

3. 実装

3.1 学習データセット

深層学習法による源氏絵流派識別を評価するために,表1に示す数の源氏絵画像を用いた.対象としている流派の種類は「土佐派」,「狩野派」,「岩佐派」,「その他」の四種類である.

3.2 学習モデル

学習に使用したネットワークのモデルは VGG[4]の全結合層部分を取り除いて GlobalAveragePooling 層[5] と二層の全結合層を繋げた図 2 のようなものを使用した.

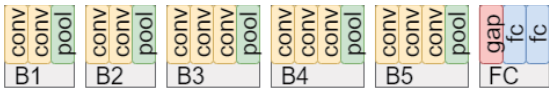


図 2 使用したモデルのアーキテクチャ

また,ランダムな初期値を用いて一から行う学習と,B1 ブロックから B5 ブロックの部分を学習済みのモデルの結果を初期値に使って再学習させるファインチューニングを用いて結果を比較した. 学習済みのパラメータは keras[6] で公開されているものを使用した. ファインチューニングの際には使用したネットワークの 5 ブロックのうち再学習させるブロックを変えることで 6 つのモデルを用意した.表 2 に各モデルで学習させるものを示す. ただし MN はファインチューニングを行わず,すべての層に対してランダムな初期値を使ったものである.

表 2 各モデルの違い

	B1	B2	B3	B4	B5	FC	正答率
M0	○	○	○	○	○	○	96.5%
M1	×	○	○	○	○	○	95.9%
M2	×	×	○	○	○	○	94.7%
M3	×	×	×	○	○	○	95.7%
M4	×	×	×	×	○	○	95.7%
M5	×	×	×	×	×	○	70.8%
MN	○	○	○	○	○	○	37.4%

4. 評価

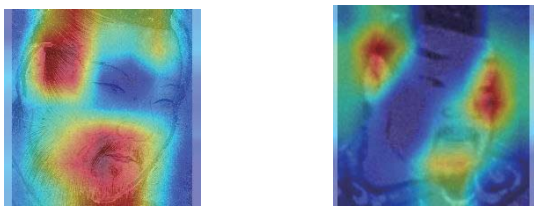
一般に分類器が未知のデータに対して有効か検証するためは訓練用データと評価用データに分け訓練に使わない評価用データのみで精度を推定する. しかし,評価用データに偏りがあった場合に大きな推定誤差が生まれる. そのため,本稿ではその偏りを減らすためデータを流派ごとに 5 分割し,そのうちの一つを評価用データとし,残りの 4 つを訓練用データとして学習することを 5 回行い,その評価平均を最終的な結果とする 5 分割交叉検証を使用して各モデルの評価を行った.

表 2 に各モデルの平均正答率を示す.ファインチューニングを行ったときに非常に正答率が上がっていることがわかる.

表 3 交叉検証での混合行列

		出力			
		土佐派	狩野派	岩佐派	その他
入力	土佐派	112	2	1	0
	狩野派	6	116	0	3
	岩佐派	3	2	114	1
	その他	0	2	0	199

表 3 は最もよい正答率をだした M0 での混同行列である. 交叉検証での各テストデータの結果を合算したものを表示している.どの流派に対してもよい結果になっていることがわかる.



(a) 画像出典：土佐光吉筆,「源氏物語図屏風」,賢木, 京都国立博物館所蔵

(b) 画像出典：岩佐勝友筆「源氏物語図屏風」,関屋, 出光美術館所蔵

図 3 M0 の認識部位

図 3 は M0 の最後のプーリング層に対して Grad-CAM を適用したものである. 顔のパーツに着目して判断していることがわかる.

5. おわりに

本論文では深層学習によって源氏絵の顔画像から流派を分類した. その結果,高い精度で正しく判定することができた. また,顔をとらえて判断していることを図示できた. 今後の課題は,顔の切り抜き画像ではなく全体図から判断し特徴が表れている箇所を図示することである.

謝辞

本研究を推進するにあたり,貴重な意見と議論を頂いた小長谷研究室ならびに源氏絵 DB 研究会の皆様に感謝します.また,本研究の一部は科研研究課題 17H02295「オントロジーに基づく源氏絵データベースを共有・活用した源氏絵の総合研究」に支援されている.

参考文献

[1] 稲本万里子:"幻の「源氏物語絵巻」の制作背景再考",恵泉女学園大学紀要(29), p.204-182, 2017-02

[2] Joseph Redmon, Ali Farhadi :"YOLO9000: Better, Faster, Stronger", arXiv:1612.08242v1 [cs.CV] 25 Dec 2016

[3] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra :"Grad-CAM: Visual Ex-planations from Deep Networks via Gradient- based Localization", arXiv:1610.02391v3 [cs.CV] 21 Mar 2017

[4] Karen Simonyan, Andrew Zisserman :"Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556v6 [cs.CV] 10 Apr 2015

[5] Min Lin, Qiang Chen, Shuicheng Yan:"Network In Network", arXiv:1312.4400v3 [cs.NE] 4 Mar 2014

[6] Chollet, Francois and others, "keras" https://github.com/keras-team/keras, 2015

[2D4] [General Session] 13. AI Application

座長:森川 幸治（パナソニック株式会社）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room D (4F Cattleya)

[2D4-01] Utopia of 81 Squares

○Naoyuki Soma[1] （1. The University of Tokyo）

5:20 PM - 5:40 PM

[2D4-02] Evaluation of Hybrid Reward Architecture on various learning policies and environments

○Yutaro Fujimura[1], Tomoyuki Kaneko[1] （1. The University of Tokyo）

5:40 PM - 6:00 PM

[2D4-03] Identification of the Expression of the Move of Go by Classifier Learning

○Natsumi Mori[1], Takehito Utsuro[1] （1. University of Tsukuba）

6:00 PM - 6:20 PM

[2D4-04] Development of a Jam Session System with Hatsune Miku That Learns User's Favorite Scat by Reinforcement Learning

○Joji Tsuruda[1], Natsuki Oka[1], Kazuaki Tanaka[1] （1. Kyoto Institute of Technology）

6:20 PM - 6:40 PM

[2D4-05] Automatic Music Composition System Based on Genetic Programming and Surrogate Model with Deep Learning

○Hironori Yamamoto[1], Taku Hasegawa[2], Mori Naoki[2], Keinosuke Matsumoto[2] （1. Osaka Prefecture University, 2. Graduate School of Engineering, Osaka Prefecture University）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

81 マスの無何有郷について
——将棋およびチェスにおける遊戯的人間（ホモ・ルーデンス）と遊戯的機械（ロボ・ルーデンス）の邂逅——

Utopia of 81 Squares: Harmony between *homo/robo ludens* in Shogi and Chess.

相馬　尚之[*1]
Naoyuki SOMA

[*1] 東京大学
The University of Tokyo

This paper explores the relationship between humans and machines from the perspective of play. In late 18th century, machines were playful amusement. As the Industrial Revolution progressed, machines metamorphosed from joyful automata into practical robots which emphasized the value of productivity and practicability. The recent development in artificial intelligence has enabled AI to defeat top chess and Shogi players and surpass human intelligence. These events cause us fears that humans shall be destroyed by what they create, like Dr. Faust or Dr. Frankenstein. However, machines can abandon their practicability at this crucial moment and regenerate as *robo ludens* which play eagerly with *homo ludens*.

1. はじめに

2017 年 5 月，第 2 期電王戦第 2 局は将棋ソフト PONANZA の佐藤天彦叡王に対する勝利で幕を閉じた．チェスでは 1997 年に IBM のディープ・ブルーがガルリ・カスパロフに，囲碁では 2016 年に Google DeepMind の AlphaGo がイ・セドルに勝利していたが，将棋でも人工知能が現役のトップ棋士を下したのだ．

チェスや将棋といった人間と機械の対決の最前線では機械の勝利が決した．もちろんこれらの事実のみで，機械が人間を凌駕したと宣言することはできない．人工知能はいくつかの領域で人間の頭脳に勝るようになったが，依然として汎用的な能力の獲得には至っていない．それでも，チェスや将棋という人間の「知性」の精華とみなされてきた分野での理性ある人間（ホモ・サピエンス）の敗北は，将来の機械と人間の関係に対する不安を象徴する出来事であった．

幾度となく語られた恐怖が浮かび上がる．被造物はいつしか創造主を超え，遂には破滅させてしまう．というのも，ファウスト博士やフランケンシュタイン博士のような狂気に取りつかれた学者は，決して満足することができず，破局に向けて突進せずにはいられないのだから．将棋での人間の敗北は，知性さえ人為的に生み出そうとした人間を襲う恐るべき報いの前兆ではないかという危惧は，容易に拭い去ることができるものではない．

しかし本論で注目したいのは，チェスや将棋はその本源において，人間の知性のみならず，それに先行する遊戯性の発露であるということである．人間の本源を遊戯性に求めたヨハン・ホイジンガは『ホモ・ルーデンス』（1938）のなかで次のように述べている．

> 動物は遊ぶことができる．だから，彼らはすでにその点で単なる機械仕掛け以上のものである．我々は遊び，かつ，遊ぶことを知っている．だから，我々は単なる理性的存在より以上のものである．なぜなら，遊びは非理性的なものであるから．[ホイジンガ 89]

連絡先：相馬尚之，東京大学，somuch59@gmail.com

ホイジンガの時代には可能性としても一顧だにされなかったであろう「機械仕掛けが遊ぶことができるか」という想像は，技術的特異点（シンギュラリティ）を前に現実の問題となろうとしている．これを受けて電王戦は，人間と人工知能のどちらにより優れた知性があるかということのみならず，むしろ 81 マスの将棋盤の上で，両者はいかに対決し，そしていかに「遊ぶ」かという問いに，すなわち，遊戯的人間（ホモ・ルーデンス）と並ぶ「遊戯的機械（ロボ・ルーデンス）」の出現に向かう挑戦へと通じていなければならない．

この「遊戯的機械（*robo ludens*）」という言葉は，いくつかの問題をはらんでいる．その一つは，「ロボ」の範疇に関する問題である．ロボットと人工知能は，しばしば関連し合うが異なる存在である．またロボットも，小説や映画では特に人型のマシンを表しているが，産業用ロボットからアイボやペッパーまで，様々な様態の機械に用いられる．そのため本論では，知性の有無や形態に拘らず，人間（*homo*）と向かい合うあらゆる機械的存在の最も広い集合として「ロボ（*robo*）」を用いるが，これは依然として緻密な分類学ではなく，曖昧な領域を指示するにとどまらざるを得ない．

そしてもう一つは，遊戯とロボットの結合に関する矛盾である．ロボット（robot）とは元々，チェコの作家カレル・チャペックが『R. U. R.』（1920）において，人間が労働を肩代わりさせるために創造した存在を指すために，チェコ語の「賦役（robota）」から生み出した造語である．つまり「ロボット」は，誕生の瞬間には確かに労働と結びついていたのであり，そのため，*robo ludens* という言葉は，労働と遊戯の結び付きという幾らか矛盾を孕んだ表現のように感じられるであろう．

しかし，この遊戯と労働の撞着こそ，産業革命以後の機械の本質の一面である．本論では主にチェスと将棋を主題に，産業革命期にそれまで遊戯的であった機械が，生産性や労働と密接に結びつけられてしまい，以後この実用性の観点のもとに，現在の人工知能による知性の代替まで到達しつつあることを確認する．しかしこの実用機械は，人知を機械知能で超える瞬間に，その舞台となったチェスや将棋の世界で遊戯性を取り戻す可能性を秘めている．ここで論じられるのは，実用機械はその発展の頂点で生産性の枠組みから跳躍し，遊戯的機械（ロボ・ルーデンス）として再生することによって，遊戯的人間（ホモ・ルーデンス）との間に共生の道を発見することができるのではないか，という可能性である．

2. 娯楽の自動人形から実用機械へ

チェスを指すには人間の知能が必要であるという考え方は，すでに 19 世紀初めの西欧では階級等に拘らず受け入れられていた[Babbage 64]. チェスは人間の矜持の一つであったため，チェスを指す自動人形は人間に対する挑戦となり，また人間も機械なのかという懐疑さえ招いた. それゆえに機械が知性を持つことができるかという問いに多くの学者や職工が没頭し，大衆は自動人形の見世物に熱狂したのだが，産業革命と共に実用性の尺度が取り入れられるにつれて，機械も遊戯性を喪失していくことになる. この機械における遊戯と労働の複雑な混淆を明らかにするために，まずは 18 世紀のチェスマシンをとり上げることとしよう.

2.1 ケンペレンのチェス指し人形

歴史上もっとも有名なチェスを指す機械は，1770 年にオーストリアの宮廷で披露された自動人形であろう. 時の女帝マリア・テレジアの命を受け官吏ヴォルフガング・フォン・ケンペレンの制作したこの精巧な機械は，技巧の詰められたキャビネットの上面にチェス盤が置かれ，背後のトルコ人の姿をした人形が駒をつかみ，動かす仕掛けだった. この「タ－ク」と呼ばれた自動人形は，思考能力を有する機械として評判を博してヨーロッパ各地を巡り，ベンジャミン・フランクリンやナポレオンとも対戦した.

ケンペレンの死後，タークはバイエルン出身の興行師ヨハン・ネポムク・メルツェルの手に渡った. 彼もヨーロッパやアメリカの各地で人形を展示したが，もっともこの自動人形は本当にチェスの手を考えていたのではなかった. 1835 年に実際に人形を見たアメリカの作家エドガー・アラン・ポーが「メルツェルのチェス指し」(1836)で分析したように，実際には中に人が潜んでいて，彼の操作に仕掛けが反応し，チェスの駒が動いていたのである. 結局のところ機械の思惟はまやかしであり，チェス指し人形は娯楽のための見世物に過ぎなかった.

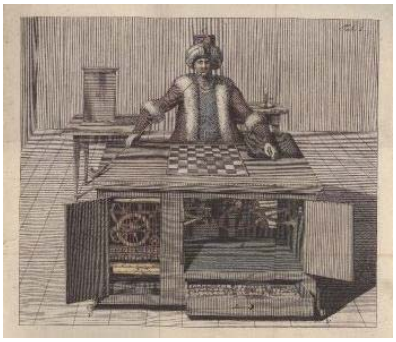


図 1: ケンペレンのチェス指し人形

機械が娯楽であったという事実は，18 世紀末のウィーン宮廷のチェスマシンを「ロボット」と呼ぶことをいくらか躊躇させるだろう. ロボットは先述のように，元々は人間の労働を代わる存在であることを含意していたが，ケンペレンが将棋指しを作った時代には，機械は依然として遊戯的性格を帯びており，生産性による評価を免れていた. 独文学者の種村季弘は，ケンペレンの奇術について次のように述べている.

> ケンプレン(ママ)の工学的才能は，かくてただの一度も実用に結びついたことはなかった. […]将棋差し人形の愛すべきトリックを含めて，ケンプレンは，終始一貫，絶対に実用と生産技術には結びつかない，遊戯を目的とした機械しか制作しなかったのである. [種村 85]

2.2 実用機械への道程

とはいえ，自動人形は産業と断絶していたのではない. イギリスの物理学者デイヴィッド・ブリュースターは，「18 世紀を特徴づけた自動装置展示への情熱は，極めて巧妙な機械装置のもととなり，上級の職人たちの間に，非常に精巧な機械の構成における，適確かつ正確な制作の習慣を導入した」と述べている[Brewster 83]. 自動人形の流行は，機械技術の発展という点で，続く産業革命の礎となった.

しかしこの連関のために，自動人形も産業革命に伴う実用主義の呪縛から自由であることはできなかった. 以後機械の発展は急速に，娯楽から富の生産に向きを変える. ブリュースターの言葉によれば，その変化は以下のようである.

> それら機械の驚異は，ある世紀にはそれらを使った手品師のみを豊かにしたが，他の世紀では国家の富の増加に貢献した. そして自動装置の玩具は，かつては大衆を楽しませたが，いまや人類の力を広げ人類の文明を増進することに用いられているのである. [Brewster 83]

この娯楽と実用性の間の葛藤は，もっとも著名な自動人形制作者の一人ジャック・ド・ヴォーカンソンの生涯からもうかがえる. フランスの技師であったヴォーカンソンの傑作であるアヒルの自動人形は，えさを食べ排泄する消化のメカニズムを模倣するために，内部の消化管なども忠実に再現されていた. 彼の技能は生産性の追究ではなく，生命体の内部機構の徹底した模倣に向けられていたのである.

しかし彼は，実用機械から離れた玩具制作にとどまることはできなかった. 1741 年，ヴォーカンソンはフランス政府から絹織物産業の改善に取り組むよう命じられ，パンチカードを用いた織機の自動化に取り組んだ. だが彼の新たな提案は失業を恐れたリヨンの職工らの反発を招き，遂には町を脱出することを余儀なくされてしまう. ヴォーカンソンは，職人の遊戯から実用機械の開発へと向かい，そこで思わぬ挫折を経験したのであった.

ケンペレンの自動人形に戻ろう. このチェス指し人形からも，後に 2 つの代表的な実用機械が派生している. その第一は，エドモンド・カートライトによる力織機である. 1785 年に開発されたこの機械は，紡績機や蒸気機関と並び産業革命を主導したが，カートライトがその製作に向かったのは，1784 年にロンドンでタークを目撃し，チェスを指す機械を作るより布を織る機械を作ることが難しいはずは無いと考えたためであった[Altick 78].

そしてこのタークから影響を受け実用機械に向かったもうひとりの人物が，チャールズ・バベッジである. バベッジは少年時代にロンドンのジョン・ジョゼフ・メルランの博物館を訪れ優雅な女性の踊り子の自動人形に感嘆するなど早くから自動人形に関心を示しており，後の 1819 年 3 月にはロンドンでタークを見物し，翌年 2 月には実際に試合を行った. 彼はこのチェス指し人形について，どこかに人間が隠れており実際にチェスの手を考えているのではないことを見抜いたが，一方で機械が思考することは不可能だとは考えなかった. バベッジは，彼の解析機関を用いればチェスのあらゆる手筋を分析することも可能であるとして，実際に 3 目並べで機械の戦略を検討している[Babbage 64].

だがバベッジの目的は，機械を用いてチェスを指すこと自体にはなかった. 彼の階差機関及び解析機関は，その計算力において人間の知性を凌駕し，より速く正確に，かつ安価に数表を作成することを目指していた. バベッジを捉えていたのは経済性と合理性であり，階差機関の構築は当時のマニュファクチュアや生産システムの体系化とも不可分であった[Schaffer 94].

機械は工業化とともに，見世物としての役割を終えて労働のための道具となる．それでもバベッジが後に踊り子の自動人形を購入し階差機関の一部と並べて家に展示したのは，見世物としての機械の延長上にあっただろう．だがこの配置は同時に，このコンピュータの先駆者の時代には，機械は遊戯性の残滓を含みつつも実用機械へと変身したことを示しているのである．

3. 実用機械から遊びへの回帰

19 世紀は機械技術が大きく発展した時代であったが，それは自動人形のような娯楽的機械にとっては受難の時代であった．「産業革命の到来とともに，機械は神秘性を失い始め，19 世紀の実用的な技術的成果が――鉄道，蒸気船，写真，電灯など――しばしば自動人形よりも素晴らしいとされた．」[Geduld 24] これらの産業技術が多くの富をもたらし，社会を大きく発展させたのは事実である．しかし，人間の労働を代替し拡張するための機械は，娯楽を離れ，実用性と生産性によって評価されるようになった．ここで機械的存在は，語源通りの「ロボット」に最も接近したが，もはや人間の姿を懐かしむ余裕はなかった．

20 世紀には，機械は人間とチェスを指す人形ではなかった．フリッツ・ラングの映画『メトロポリス』(1927) で描かれたように，機械は無数の人間の労働者の奉仕を受ける，遥かに巨大な装置となった．鉄道や蒸気機関を進歩の象徴とするのはあまりに素朴かつ楽天的であると判明し，人間は機械にとって代わられるのではないかという恐怖にとらわれ続けた．それでも脳と知性は，圧倒的な機械に対する人間の最後の砦であった．そして 20 世紀の半ば以降この牙城に機械化の波が押し寄せるようになると，チェスはその最果てにおいて重大な役割を果たすことになる．

3.1 人工知能のハエ

人工知能の研究でもチェスは初期から研究対象となった．その理由としては，チェスのルールが明確に定義されること，難易度が適切であること，機械が思考しているか評価しやすいことなどがあげられよう．しかしこのチェスは，実用性のくびきから自由ではなかった．

ロシアの数学者アレクサンダー・クロンロッドは，遺伝学におけるショウジョウバエのように基本的な実験材料となったことから，コンピュータチェスを「人工知能のハエ」としたが，「人工知能」という用語を提唱したことで知られるジョン・マッカーシーは，カスパロフに対するディープ・ブルーの勝利後に，チェスのために人工知能研究が被った痛手について次のように述べている．

> コンピュータチェスは，遺伝学者たちが 1910 年以降レース用ショウジョウバエを繁殖させることに努力を集中していたらそうなっていたように，発展してきた．いくらかの科学を得たかもしれないが，主には，非常に速いショウジョウバエを手にしただろう．[McCarthy 97]

「人工知能のハエ」は，チェスにしか役立たないのではないか．1949 年，アメリカの数学者クロード・シャノンは「チェスをプレイするためのコンピュータのプログラミング」という論文を発表したが，この中で彼は既に，チェスマシンは「なんの実用的重要性もないかもしれない」と留保しつつも，それでも翻訳や論理的推論のような同種の更なる問題はコンピュータの発展に伴い，「経済的観点から真剣に考える価値がある」とした[Shannon 50]．マッカーシーやシャノンの，コンピュータチェスに発展性や実用性が無いことに対する懸念は，彼らがケンペレンではなくバベッジを継承したことを示している．かつてのケンペレンの自動人形のように，「チェス指すこと」自体を目的とすることは，20 世紀後半には科学者にさえ受け入れがたい挑戦となっていたのである．

チェスや将棋における人間と機械の対決に多くの人々が恐怖を覚えるのも，ひとつは人工知能の実用的な応用可能性のためであろう．チェスや将棋における戦いは，単なる人間と機械の知性の衝突ではない．人工知能の勝利は，それに続く「知的な職」の奪い合いに通じている．人間の労働力を代替するためのロボットは，現在，なによりも人間の頭脳労働を代替しようというのである．

ここに，人間と機械の対決を超えて共存共栄へと向かう鍵がある．それは，産業革命以降の機械と人間を蝕む，実用性と生産性の宿痾を棄て去ること，すなわち，失われた遊戯性へと両者が共に帰還することである．

3.2 ユートピアへの跳躍

では実用性から遊びへと回帰することは果たして可能なのだろうか．その検討するため，ここで遊戯の性質を確認しよう．

> 形式について考察したところをまとめて述べてみれば，遊びは自由な行為であり，「ほんとのことではない」としてありきたりの生活の埒外にあると考えられる．にもかかわらず，それは遊ぶ人を完全にとりこにするが，だからと言って何か物質的利益と結びつくわけでは全くなく，また他面，何かの効用を織り込まれているのでもない．それは自ら進んで限定した時間と空間の中で遂行され，一定の法則に従って秩序正しく進行し，しかも共同体的規範を作り出す．[ホイジンガ 89]（強調著者）

遊戯は，遊ばれる世界が限定され，規則があるからこそ可能となる．外部から切り離された限定的世界を構築し，いかなる実利も求めず，その内部でのみ通用するルールに従うことは，遊びの不可欠の基底である．この実践によってロボットと人間の関係が問われたのは，現実社会よりも文学の世界であったが，特に規則の観点からはアイザック・アシモフが挙げられよう．

アシモフは第一次世界大戦後の SF 小説で，『R.U.R.』のようにロボットが暴走し創造主を破滅させるプロットがひたすら現れることに辟易していた．そこで彼はこのファウストやフランケンシュタインのモチーフを離れるために，「ロボット三原則」を考案し，ロボットを電線や圧力鍋同様に可能な限り安全に設計された人工物とすることで「わたしのロボットは創造主にむやみに反抗したり，ファウストの罪と罰を具現してみせるような退屈な行為は，決して決してない」ように構成したのである[アシモフ 04]．ロボット三原則は以後多くの SF 作品を貫く規則となったが，この規範によって小説世界はフランケンシュタインの劫罰を免れるのだ．

現実にはロボット三原則は存在せず，またアシモフの物語のように苦悩と葛藤を伴いつつも人間とロボットの共存が実現しているのでもない．人間はようやく，知性において創造主を上回る人工知能と将棋盤の上で対面したに過ぎない．しかし盤上には，ゲームのルールがあり，それが世界を支配している．

そのためこの世界には，あらゆる他の規則は導入されてはならない．人間も機械も，この「ほんとのことではない」世界では，勝利を目指してひたむきに遊ぶのだ．アンドロイドが現実と異なる舞台上で演劇を行ったのも偶然ではない．遊びの世界は，現実社会での生産性の尺度がひとたび持ち込まれれば，たちまちに破壊されてしまう無何有郷である．

将棋の 81 マスの宇宙は，そのようなユートピアとなりうる．夢枕獏は電王戦のドキュメントの中で，羽生善治の次のような言葉を紹介している．

> 将棋というのは完全解析が可能なゲームじゃないですか，そこで羽生さんに「将棋が解明されてしまって，例えば先手番がこうすれば必ず勝つことがわかってしまったら，どうしますか？」と聞いたんです．「その瞬間，将棋が終わっちゃいませんか？」と．
> そうしたら羽生さんは，「だいじょうぶです，ルールを一個変えればいいんです」と，平然と答えたんです．[夢枕 03]

この言葉は駒の動かし方のようなミクロな話であろう．しかし，将棋がすべて解明されゲームとして死ぬ瞬間に，これまでの蓄積をすべて放棄し，新たなゲームに向かうというのは象徴的である．

人工知能は，チェスや将棋といった知性の領域で人間を超越しようとしている．だがこの勝利の瞬間，機械知能はゲームの決まりのみならず，その盤上を支配する「遊び」というルールを学び，誕生から自らを貫く実用性や生産性といった規範を棄て去らねばならない．それによりロボットは，新たに「遊戯する機械」として再臨し，ただひたむきに遊び始めるのである．

4. 遊戯的機械（ロボ・ルーデンス）への再生

人工知能は人類を破滅させるメフィストフェレスやフランケンシュタインであるとの危惧は後を絶たない．だがチェスや将棋のソフトは，人工知能の「ハエ」である．そしてこのハエが，ゲーテの「至福の憧れ（Selige Sehnsucht）」の火蛾のように浄福の世界への衝動に突き動かされ，死と再生の境地に達しようとしている．

Keine Ferne macht dich schwierig,	遠さは何の支障でもない
Kommst gefolgen und gebannt,	ましぐらに，呪に惹かれ翔ぶ
Und zuletzt, des Lichts begierig,	そしてついに，光に焦がれ
Bist du Schmmetterling verbrannt.	火蛾よ，お前は焼け死ぬ．
Und so lang du das nicht hast,	この，死して成れ！ これを
Dieses: Strib und Werde !	ついに会得せぬ限り，
Bist du nur ein trüber Gast	おまえは暗い地上の
Auf der dunklen Erde.	暗く悲しい孤客に過ぎぬ．

「死して成れ（Strib und Werde）」の秘蹟は，憧れの極限で起こる．ゲーテは『ファウスト』のなかでそれをファウストの営みとしてのみならず，第 2 部第 2 幕のホムンクルス，そして第 3 幕のオイフィリオーンにも託している．ファウストの息子はイカロスのように天空に飛翔するも，海に墜ちてしまう．だが，それは徒な死ではない．

> 最大の火にあこがれて飛翔したイカロスは，その人工の翼の羽搏きの極点で海中に落ちた．[…]それは自然との無媒介的な和合ではない．人工的-男性的な世界の隅々を遍歴してそのさい果ての反射を通じて，はじめてあらかじめ失われていたものに――ときには思いがけない懲罰の外見において――めぐり逢う．[種村 74]

死に憧れるのは人間の特権ではない．アシモフの「バイセンテニアル・マン」（1976）では，人間になることを欲したロボットアンドリューは，人間と同等の自由や権利を求める闘争の果てに，自らの機械仕掛けの身体を改造し，最後に死ぬことを目指す．そして彼が人間になる最後の瞬間に彼の前に浮かんだのは，リトル・ミスの面影であった[アシモフ 79]．

もちろん和合のためには，人間も死なねばならない．ホイジンガは既に 19 世紀末にはスポーツやチェスやトランプなどの理知的計算のきく遊びが組織化や訓練により「遊びでもないし真面目でもないものになってしまった」と嘆いたが[ホイジンガ 89]，この誹りは現在の職業化し，商業化した将棋界も免れない．人間も，実利の価値観を打ち捨て，遊ぶ必要があるのだ．

この遊戯的機械（ロボ・ルーデンス）と遊戯的人間（ホモ・ルーデンス）の遊びから何が生み出されるかは定かではない．だが，「文化はその根源的段階においては遊びの性格を持ち，遊びの形式と雰囲気の中で活動する」ということ，すなわち「遊びは文化より古い」というホイジンガの視座を継承するのであれば[ホイジンガ 89]，技術的特異点以後に拓かれるべき人間と機械の共生可能な文化は，何よりもまずコンピュータ将棋のような遊びの領分にその本源を求めることになるだろう．

かつての遊びの機械は，産業革命期に実用性と生産性という尺度に基づく実用機械へと変節してのち，あらゆる領域で人間を上回ろうと努め，遂には思考や理性においてさえ人間を超越しようとしている．しかし，この実用性のための発展の頂点たる思考能力の獲得において機械は，将棋盤の世界へと迷い込み生産性を失う．だがそれと同時に，「ロボット」にはあらかじめ失われていた遊びの領域（ロボ・ルーデンス）へと跳躍し，遊戯的機械として遊戯的人間（ホモ・ルーデンス）との邂逅を果たすのである．

人工知能のハエは，この「死して成れ」の秘儀を実現する，稀代の錬金術師であった．

参考文献

[Altick 78] Altick, Richard D: *The Shows of London*. The Belknap Press of Harvard University Press, 1978.

[Babbage 64] Babbage, Charles: *Passages from the Life of a Philosopher*. Longman, Green, Longman, Roberts & Green, 1864.

[Brewster 83] Brewster, David: LetterXI. In. *Letters on Natural Magic*. Chatto & Windus, Piccadilly, pp. 317-345, 1883.

[Geduld 78] Geduld, Harry: Genesis II: The Evolution of Synthetic Man. In. Harry M. Geduld and Ronald Gottesman (eds.) *Robots, Robots, Robots*. New York Graphic Society, pp. 3-37, 1978.

[McCarthy 74] McCarthy, John: AI as Sport. *Science*, Vol. 276, No. 5318, pp. 1518-1519 (1997)

[Schaffer 94] Schaffer, Simon: Babbage's Intelligence: Calculating Engines and the Factory System. *Critical Inquiry*, Vol. 21, No. 1, pp. 203-227 (1994)

[Shannon 50] Shannon, Claude: Programming a Computer for Playing Chess. *The Philosophical Magazine*, Ser. 7, Vol. 41, No. 314, pp. 256-275 (1950)

[Standage 02] Standage, Tom: *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*. Walker & Company, 2002.

[アシモフ 79] アシモフ, アイザック: バイセンテニアル・マン, 聖者の行進, 池央耿訳, 創元推理文庫, pp. 240-306, 1979.

[アシモフ 04] アシモフ, アイザック: ロボットの時代, 小尾芙佐訳, 早川書房, 2004.

[種村 74] 種村季弘: 怪物の解剖学, 青土社, 1974.

[種村 85] 種村季弘: ロボット考, アナクロニズム, 河出文庫, pp. 103-114, 1985.

[ホイジンガ 89] ホイジンガ, ヨハン: ホモ・ルーデンス――文化のもつ遊びの要素についてのある定義づけの試み, 河出書房新社, 1989.

[夢枕 03] 夢枕獏他: ドキュメント電王戦――その時，人は何を考えたのか，徳間書店，2013.

[図 1] http://www.sammlungen.hu-berlin.de/dokumente/23240/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

様々な学習戦略と学習環境における Hybrid Reward Architectureの性能の評価

Evaluation of Hybrid Reward Architecture on various learning policies and environments

藤村悠太朗 *1　　金子知適 *2
Yutaro Fujimura　　Tomoyuki Kaneko

*1東京大学教養学部学際科学科
Department of Interdisciplinary Sciences, College of Arts and Sciences, The University of Tokyo

*2東京大学大学院情報学環
Interfaculty Initiative in Information Studies, the University of Tokyo

Deep Q-Network (DQN) was able to achieve a level comparable to the performance of a professional human player. However, in large and complex domains (e.g. Ms. Pacman), learning can be very slow and unstable. In Hybrid Reward Architecture (HRA), a reward function is decomposed in advance to enhance learning in such domains, and then value functions are separately learned for decomposed reward functions. In this paper, we made some environments that made learning more difficult to evaluate the performance of HRA. The results indicated that HRA need more enhancements to learn environments where learning is difficult on the uniform random policy.

1. はじめに

人工知能研究の一分野として，AI エージェントの強化学習に関する研究が数多く行われてきた．囲碁において AlphaGo [Silver 16] が人間のトッププレイヤと対戦して勝ち越し，チェスや将棋においても自己対戦のみで学習した Alpha Zero が既存の人間より強いプログラムに勝ち越す [Silver 17] など，様々なゲームで人間より強い AI エージェントが登場するまでとなった．一方で，リアルタイムで迅速な意思決定をする必要がある状況を想定し，このようなゲームに関する研究も注目されている．強化学習の手法である Q 学習 [Sutton 98] と深層学習を組み合わせた Deep Q-Network (DQN) という手法によって，Atari 2600 のゲームの一部で人間より高いスコアを記録した [Mnih 15] が，人間のプレイヤのスコアに達することができないゲームも存在し，そのようなゲームの中の 1 つが Ms. Pacman である．Ms. Pacman の難しさは取りうる状態の数が非常に多く，状態と報酬の関係も複雑であるため，DQN を用いてもうまく学習が進まないという点である．van Seijen らは Hybrid Reward Architecture (HRA) を提案し，人間のスコアを上回るプレイヤの作成に成功した [Seijen 17]. HRA は，ゲームがエージェントに与える報酬関数を分割し，それぞれに対して行動価値関数を学習することで，非常に大きな状態空間に対しても学習を可能にしている．

本研究では，Ms. Pacman よりも難しい性質を持ったゲームにも HRA の手法を適用できるかを調査するため，HRA による分割を用いた学習が適するゲーム環境をいくつか作成し，DQN と HRA で性能の比較と考察を行った．

2. 関連研究

ここでは，本研究がベースとしている，DQN と HRA について述べる．

連絡先: 藤村悠太朗, 東京大学教養学部学際科学科, 東京都目黒区駒場 3-8-1 15 号館, yut-mak874@g.ecc.u-tokyo.ac.jp

2.1 Deep Q-Network (DQN)

Deep Q-Network (DQN) [Mnih 15] は，Q 学習の行動価値関数 Q をニューラルネットワークを用いて関数近似することにより，学習を行う手法である．ニューラルネットワークの重みを θ としたときの行動価値関数を $Q(s,a;\theta)$ と書く．DQN では，次の誤差関数 $\mathcal{L}_i(\theta_i)$ を最小化するように θ を繰り返し更新することによって学習を行う．

$$\mathcal{L}_i(\theta_i) = \mathbb{E}_{s,a,r,s'}\left[(y_i^{\mathrm{DQN}} - Q(s,a;\theta_i))^2\right], \quad (1)$$

ただし，　$y_i^{\mathrm{DQN}} = r + \gamma \max_{a'} Q(s',a';\theta_{i-1})$　(2)

ここで，θ_i とは，i 回更新が行われた後の重みを指す．

2.2 Hybrid Reward Architecture (HRA)

Hybrid Reward Architecture [Seijen 17] では, 環境報酬関数 R_{env} が式 (3) のように n 個の報酬関数に w_i の重み付け和で与えられる環境を考える:

$$R_{env}(s,a) = \sum_{k=1}^{n} w_k R_k(s,a) \qquad \text{for all } s,a. \quad (3)$$

そして，この分割された報酬関数それぞれに対して, 強化学習のエージェントを訓練する．それぞれのエージェント k が持つ行動価値関数を $Q_k(s,a)$ と書く．エージェントの意思決定には，分割された行動価値関数 Q_k の合計

$$Q_{\mathrm{HRA}}(s,a;\theta) := \sum_{k=1}^{n} w_k Q_k(s,a;\theta) \quad (4)$$

を用い，誤差関数

$$\mathcal{L}_i(\theta_i) = \mathbb{E}_{s,a,r,s'}\left[\sum_{k=1}^{n}(y_{k,i} - Q_k(s,a;\theta_i))^2\right], \quad (5)$$

ただし，　$y_{k,i} = R_k(s,a,s') + \gamma \sum_{a'\in\mathcal{A}} \frac{1}{|\mathcal{A}|} Q_k(s',a';\theta_{i-1})$　(6)

を最小化する．

問題固有の知識を用いた報酬関数の分割を行うことで, Ms. Pacman やそれに類似するゲームにおいて，DQN より行動価値関数の収束が早いことが文献 [Seijen 17] で示されている．

3. 問題設定

3.1 Fruits Collection Task

Fruits Collection Task は文献 [Seijen 17] で，Ms. Pacman の簡易版として提案されているゲームである．この環境を本研究では，FruitsCollection-v0 と呼称する．

このゲームでは，エージェントは 10×10 のグリッドの中で，できるだけ早くフルーツを集めることが目的である．グリッドのうち，10 箇所がフルーツ置き場となっており，各エピソードの開始時に，10 個のフルーツ置き場の中から，ランダムに 5 個が選ばれてフルーツが配置される．エージェントはランダムな位置から出発して，ターン毎に上下左右の 4 方向を選び，選ばれた方向に隣接するマスへ移動する．エージェントはフルーツを得た場合に 1 つあたり+1 の報酬を得て，それ以外の場合は 0 の報酬を得る．5 個のフルーツを全て集めるか，300 ターンが経過するとそのエピソードは終了する．

Fruits Collection Task に対して，いくつかの要素を加えたゲームを複数用意した．これは，行動の種類を増やしたり，報酬を得にくくするような要素を加えることによって，ゲームの設定を難しくすることが目的である．

3.2 エージェントの方向があるゲーム

FruitsCollection-v1 は，FruitsCollection-v0 に，次のルールを追加したゲームである．

- エージェントは上下左右の 4 つの方向を持つ．
- エージェントが行動するとき，
  - エージェントの方向と，選択した行動の方向が一致するとき，その方向に隣接するマスへ移動する．
  - そうでなければ，エージェントは，選択した行動の方向に向きを変える．

3.3 遠距離からフルーツを集める行動があるゲーム

FruitsCollection-v1 に更に次のルールを追加したゲームを作成し，これを FruitsCollection-v2 とする．

- エージェントは，上下左右へ 4 方向を選択する行動に加えて,「レーザーを撃つ」行動を選択することができる．
- レーザーを撃った場合，エージェントの向いている方向の半直線上に存在するフルーツすべてを回収する．

また，FruitsCollection-v1,2 を簡略化し，FruitsCollection-v0 に［上，下，左，右］の方向へレーザーを撃つという 4 つの行動を追加したゲーム FruitsCollection-v3 も作成した．

3.4 特定の行動によりペナルティが発生するゲーム

FruitsCollection-v0,1,2,3 は，ランダムに行動しても 300 ターンの間に報酬を得ることが可能な設定のゲームである．そこで，FruitsCollection-v2 に，次のルールを追加したゲームを作成した．

- 10×10 のグリッドのうち，中央の 8×8 の領域にエージェントが移動すると，-1.0 の報酬を得た上で，エピソード開始時の場所に移動する．

これは，[Sutton 98] で強化学習エージェントのための環境の例として提示されている Cliff Walking と FruitsCollection を組み合わせたものである．このルールが追加されることによって，正の点数でエピソードを終了するために取るべき行動の選択肢が少なくなり，ランダムに行動とっても 300 ターン以内にエピソードを終了することが難しくなる．

このゲームを本研究では，FruitsCollection-v4 と呼称する．

4. 比較対象とするエージェント

2. 節で述べた DQN, HRA によって学習を行うエージェントを作成した．

ニューラルネットワークの構造は，ゲームの状態を受け取る入力層，全結合の中間層となっていて，中間層から各ゲームで可能な行動数のユニットを持つ出力層へと繋がるという 3 層の構造で統一した．HRA では，出力層がフルーツ置き場の数だけ独立しており，それぞれの出力層は自分の担当するフルーツ置き場に関する行動価値関数を求める．また，Q_{HRA} を求める際は式 (4) を用いて計算する．

また，先行研究 [Seijen 17] での実装では，全結合層はそれぞれの出力層で独立に存在しているが，論文中にアイデアとして示されていた，全結合層を共有するネットワークでの学習も行った．これを，本研究では HRA-Share と呼称する．

エージェントに与える報酬については，DQN ではゲームから受け取った報酬をそのまま与え，HRA, HRA-Share ではどの場所のフルーツから得た報酬かに応じて，適切な出力層に報酬を与えた．

5. 実験と考察

本研究では，プログラムの実装には python 3.5 を利用した．提案した環境を実装するために Open AI Gym [Brockman 16] を利用し，深層学習のフレームワークとして tensorflow を利用した．

5.1 実験条件

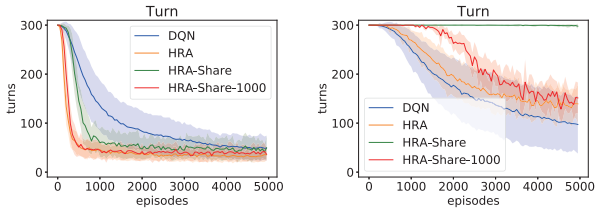
方向の要素がない FruitsCollection-v0,v3 がエージェントに与えるゲームの状態は，[Seijen 17] と同様に，10×10 のマップ内でエージェントがいる場所を表す one-hot ベクトルと，10 個のフルーツ置き場のそれぞれでフルーツがあるかどうかを表現するベクトルを結合した，要素数 110 のバイナリ型のベクトルとした．方向の要素がある FruitsCollection-v1,2 については，方向を表す one-hot ベクトルをこれに加えた，要素数 114 のバイナリ型のベクトルをエージェントに与えるゲームの状態とした．

先行研究 [Seijen 17] に従い，DQN における式 (2) の γ は 0.85, ネットワークの学習率は 0.0005，HRA における式 (6) の γ は 0.99, ネットワークの学習率は 0.001 とした．DQN，HRA-Share における全結合層のユニット数，及び，HRA で分割されている全結合層のそれぞれのユニット数は，全て 250 個とした．また，HRA-Share における全結合層のユニット数を 1000 個としたエージェントを以下の実験では，HRA-Share-1000 とする．

DQN では，訓練時は ϵ-greedy 方策で行動し，評価時は greedy 方策で行動する．この時の ϵ の値は 0.01 とする．

5.2 ランダム方策による訓練

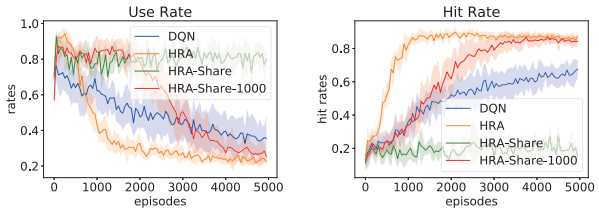
先行研究 [Seijen 17] と同様に，HRA における訓練例を一様ランダム方策で生成し，行動価値関数を式 (6) で更新する際の方策は，一様ランダム方策とした．



(a) FruitsCollection-v0 **(b)** FruitsCollection-v1

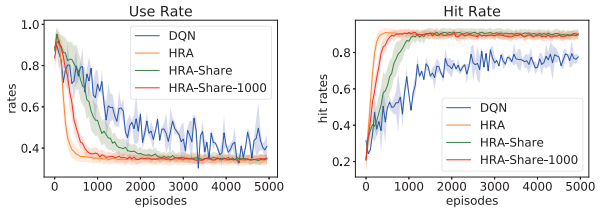
(c) FruitsCollection-v2 **(d)** FruitsCollection-v3

図 1: 各ゲームでのゲーム終了までに要したターン数



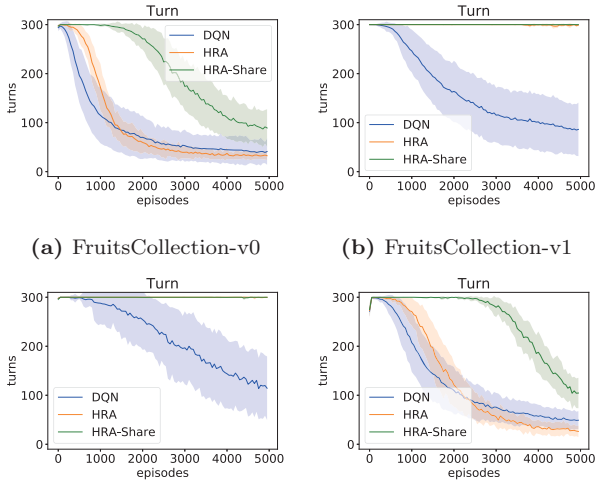
(a) レーザーを使用した割合 **(b)** 獲得した報酬のうちレーザーによって得た報酬の割合

図 2: FruitsCollection-v2 でのレーザーの使用について



(a) レーザーを使用した割合 **(b)** 獲得した報酬のうちレーザーによって得た報酬の割合

図 3: FruitsCollection-v3 でのレーザーの使用について



(a) FruitsCollection-v0 **(b)** FruitsCollection-v1

(c) FruitsCollection-v2 **(d)** FruitsCollection-v3

図 4: 各ゲームでのゲーム終了までに要したターン数

FruitsCollection-v0,1,2,3 で DQN, HRA, HRA-Share, HRA-Share-1000 を 5,000 エピソード学習させた．学習の過程での，エピソードを終了するまでに要したターン数の変化を示したグラフを図 1 に示した．横軸が学習したエピソード数，縦軸はゲーム終了に要したターン数であり，ゲーム終了に要したターン数が少ない程性能が良い．

図 1a の DQN, HRA を比較すると，先行研究 [Seijen 17] の実験と同等の結果が得られていることがわかる．また，HRA-Share は DQN と HRA の中間の性能を示しているが，全結合層のユニット数を増やした HRA-Share-1000 は HRA と同等の性能を示すことが読み取れる．図 1d より，FruitsCollection-v3 においても，この関係性は同様である．

図 1b, 図 1c を見ると，DQN, HRA は学習が進んでいる様子が読み取れるが，HRA-Share は学習が非常に遅く，HRA-Share-1000 は HRA-Share 程ではないが，学習が HRA より少し遅く進んでいることが読み取れる． FruitsCollection-v2 では，特に HRA は学習が早く，最終的に獲得する報酬，ゲーム終了までに要したターン数の値も，DQN より優れていることがわかる．

また，FruitsCollection-v2,3 で，エージェントがエピソード中にとった行動のうち，レーザーを使用した割合，及び，獲得した報酬のうちレーザーによって得た報酬の割合を示したグラフを図 2, 図 3 に示した．

図 1c, 1d と図 2, 3 を比較すると，どのエージェントも学習が進むと，レーザーを使用する割合は小さくなり，フルーツをレーザーによって回収する割合が大きくなることが読み取れる．

5.3 ϵ-greedy 方策による訓練

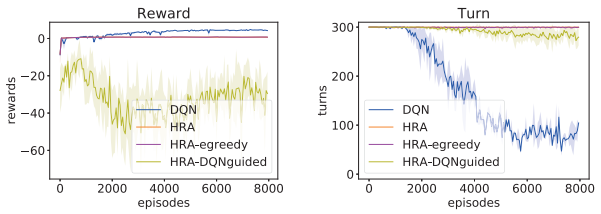
HRA における訓練例を，一様ランダム方策から，式 (4) を評価した ϵ-greedy 方策で行動したものに変更して学習を行う実験を行った．

HRA が行動を選択する際は，式 (4) を評価した ϵ-greedy 方策で行動し，行動価値関数を式 (6) で更新する際の方策は，一様ランダム方策とした．この時の ϵ の値は 0.01 とした．

FruitsCollection-v0,1,2,3 で DQN, HRA, HRA-Share を 5000 エピソード学習させた．HRA-Share-1000 は，学習にかかるコストが大きいため，今回は実験を行わなかった．学習の過程でゲームから得た報酬と，エピソードを終了するまでに要したターン数の変化を示したグラフを図 4 に示した．

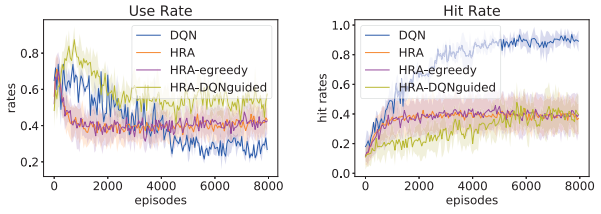
図 1 と図 4 を比較すると，ランダム方策で訓練例を作成する場合より，ϵ-greedy 方策で訓練例を作成する方法の方が，HRA の性能が悪化することが読み取れる．また，環境によって，ランダム方策で訓練例を作成した時は学習が進む環境の中で，ϵ-greedy 方策で訓練例を作成しても学習が進む場合と，学習が進まなくなる場合の両方があることがわかる．

FruitsCollection-v1 における実験では，エージェントが同じ状態にとどまり続ける様子が見られ，これが原因で中々エージェントが正の報酬を観測することができず，学習を進めることができないものと考えられる．FruitsCollection-v1 のように方向の要素がある場合，エージェントが異なる座標へ移動するためには，移動したい方向を向き，その後移動するために 2



(a) 獲得した報酬　　**(b)** ゲーム終了までに要したターン数

図 5: FruitsCollection-v4 での学習結果



(a) レーザーを使用した割合　　**(b)** 獲得した報酬のうちレーザーによって得た報酬の割合

図 6: FruitsCollection-v4 でのレーザーの使用について

回同じ行動を取るという特定の手順を行う必要がある．しかし，エージェントが行動価値関数を更新する方策が一様ランダム方策であるため，なかなか新しい状態に推移することができないと考えられる．つまり，より新しい状態に遷移させるための工夫によって，性能が改善する可能性がある．

5.4　報酬を得にくい環境の学習

5.2 節の実験から，FruitsCollection-v0,1,2,3 のように，ランダム方策でも報酬を得ることが可能なゲームでの HRA の性能を確認することはできた．しかし，5.3 節の実験を見ると，ランダム方策で報酬を得ることが難しい ϵ-greedy 方策での学習が必要になるようなゲームでは，HRA での学習が難しいことが予想される．

FruitsCollection-v4 で，DQN，5.2 節と同様にランダム方策で訓練例を作成する HRA，5.3 節と同様に ϵ-greedy 方策で訓練例を作成する HRA-egreedy，更に DQN で学習を行ったときの行動を訓練例として用いる HRA-DQNguided の 4 つのエージェントで 8000 エピソードの学習を行う実験を行った．

学習の過程でゲームから得た報酬と，エピソードを終了するまでに要したターン数の変化を示したグラフを図 5 に示した．また，レーザーを使用した割合，及び，獲得した報酬のうちレーザーによって得た報酬の割合を示したグラフを図 6 に示した．

図 5 を見ると，DQN ではこのゲームでも学習が進むが，HRA, HRA-egreedy, HRA-DQNguided はどれも学習がうまく進んでいないことがわかる．図 5b を見ると，HRA, HRA-egreedy と異なり，HRA-DQNguided は少しだけゲームをクリアしているエピソードもあるという差異はあるが，負の報酬を得ている割合も HRA, HRA-egreedy と比べて非常に高い．また，図 6 を見ると，DQN は図 2, 3 と同様に，学習が進むにつれて，レーザーを使用する頻度は小さくなり，フルーツをレーザーによって回収する割合が大きくなるが，HRA, HRA-egreedy, HRA-DQNguided はどれも途中で変化が止まってしまっていることがわかり，報酬だけでなくレーザー使用率からも，HRA で学習したエージェントが学習が進んでいないことを示している．

6.　おわりに

本研究では，HRA で学習するエージェントを，複数のゲーム環境と学習戦略で学習を行い，それぞれの性能を比較した．その結果，HRA はランダム方策での訓練と相性が良く，DQN よりも優れる学習効率と性能を発揮するが，ランダム方策で報酬を得ることが困難なゲーム環境においては，ϵ-greedy 方策や DQN で学習を行ったときの行動を訓練例とするといった手法では，学習が困難であることを示した．

今後の課題としては，5.3 節の実験で見られた，ϵ-greedy 方策における同じ状態に留まりつづける行動を抑制し，新たな状態に遷移させるような工夫を加えることが挙げられる．

謝辞

この研究の一部は，JSPS 科研費 16H02927 と JST さきがけの支援を受けています。

参考文献

[Brockman 16] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W.: OpenAI Gym, pp. 1–4 (2016)

[Mnih 15] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D.: Human-level control through deep reinforcement learning, *Nature*, Vol. 518, No. 7540, pp. 529–533 (2015)

[Seijen 17] Seijen, van H., Fatemi, M., Romoff, J., Laroche, R., Barnes, T., and Tsang, J.: Hybrid Reward Architecture for Reinforcement Learning, No. Nips 2017, pp. 1–16 (2017)

[Silver 16] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, van den G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D.: Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 484 EP – (2016), Article

[Silver 17] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm (2017)

[Sutton 98] Sutton, R. S. and Barto, A. G.: *Reinforcement Learning : An Introduction*, MIT Press (1998)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

分類器学習による囲碁の手の描写表現の同定

Identification of the Expression of the Move of Go by Classifier Learning

森 菜都未[*1] 宇津呂 武仁[*2]
Natsumi Mori Takehito Utsuro

[*1]筑波大学大学院システム情報工学研究科 Grad. Sch. Sys. & Inf. Eng, Univ. of Tsukuba
[*2]筑波大学システム情報系 Fclty. Eng, Inf. & Sys, Univ. of Tsukuba

The ancient Chinese board game, Go, with its simple rules yet highly complex strategies, requires players to encircle more territory than their opponent. However, owing to the rise in the capabilities of Go-playing software and a lack of Go instructors in Japan, there is a need for a software that actively assists human players in learning the high-level strategies required to win the game. This study focuses on generating a review for each consecutive player's move. This paper studies how to generate an expression for each move based on the distribution of the stones on the board. To this task of generating the expression for each move in a game of Go, we apply a classifier learning technique.

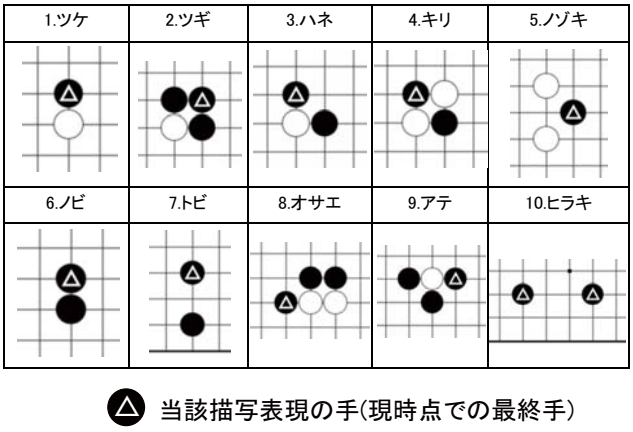


図 1: 物理的描写表現の内の頻度上位 10 種の典型例

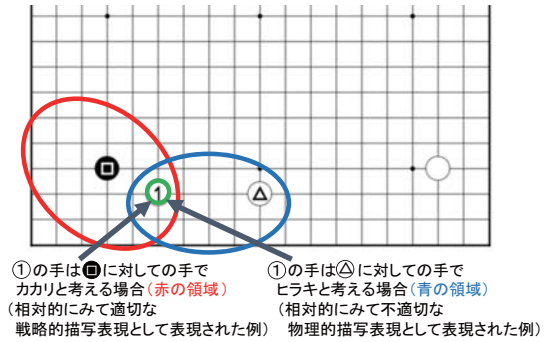


図 2: 手に対する描写表現の曖昧性の例: カカリ または ヒラキ

1. はじめに

近年，AlphaGo と呼ばれるコンピュータ囲碁プログラムがプロ棋士に初めてハンデ無しで勝利した [Silver16]．これをきっかけに，いくつかの囲碁プログラムの強さは大きく躍進し，家庭用のコンピュータでもプロ棋士と同等の力を体験することができるようになってきている．ただし，これらの囲碁ソフトのほとんどは，強さのみを追求したものであり，エンターテインメントや教育の観点からみると，まだ不十分と言える部分が多い．特に，現時点では、日本には囲碁を指導することができるプロ棋士やその力を持ったアマチュアはわずかしかないため指導者不足が問題となっており，その盤面を見ただけでは理解できないような高度な技術を教えるソフトウェアの開発は重要な課題である．

これらより，本研究では人間プレイヤーが囲碁の手を学ぶことを支援するソフトウェアの開発を目的とする．特に本論文では，囲碁における各手の解説を生成するタスクに着目した．具体的には，囲碁の解説には様々な「囲碁用語」と呼ばれる描写表現が含まれており，それらの描写表現を囲碁の盤上の石の配

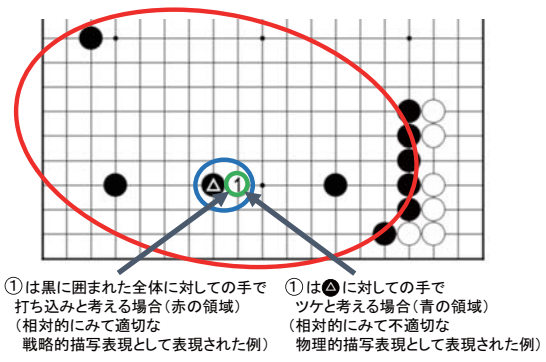


図 3: 手に対する描写表現の曖昧性の例: 打ち込み または ツケ

連絡先: 森 菜都未，筑波大学大学院システム情報工学研究科，
〒 305-8573 茨城県つくば市天王台 1-1-1, 029-853-5427

置から分類器学習を行うことにより生成する手法を提案する．ただし，これらの描写表現は単に形だけを表現しているものと，戦略的な意味を含んだ表現の 2 種類存在し，本論文では前者を「物理的描写表現」，後者を「戦略的描写表現」と呼ぶこととする [Yoshikawa98]．

関連研究として池田らは分類器学習を用いて囲碁の盤上の石の配置から描写表現を生成する手法を提案している [Ikeda15]．ただし，池田らの研究は手の形を表現する物理的描写表現を生成することのみに留まり，戦略的描写表現に関しては触れていない．

したがって，池田らの研究と本研究との主な違いの一つは，本研究は戦略的描写表現も考慮に入れながら全ての描写表現を生成する点である．さらに，本研究では手の描写表現を生成するための分類器の素性に関して，池田らが提案している盤上の石の配置の位置関係に着目した素性に加え，囲碁の手が時系列データであることを考慮した上で時間素性を提案し，その効果を示した．さらに池田らの研究では多値分類によって，全体的な性能を考察していたのに対して，本研究では各表現に対して二値分類を行い，表現ごとの最良の結果を得るために必要な最適な素性の組み合わせを示した．

2. 囲碁の手の描写表現

2.1 物理的描写表現

物理的描写表現は，通常，盤上の石の静的な局所パターンを考慮して識別される．3. 節で示すように，本研究では評価の対象に頻度上位 10 種の物理的描写表現を用いた．また，これらの物理的描写表現の典型的な例を図 1 に示す (三角が含まれた黒の石が当該描写表現の手)．

2.2 戦略的描写表現

前節で述べたように，物理的描写表現は盤上の石の静的な局所パターンを考慮して識別される．それに対して，戦略的描写表現の識別には意味内容を考慮する必要があり，局所パターンではなく，より広い視野を持つ必要がある場合が多い． ただし，これらの戦略的描写表現はゲーム中の大局的な流れを通じて出現する特徴的な状況を表すという特徴を持っている．そのため，ゲーム内の手の全体的な時系列情報は戦略的描写表現によって区分化が可能であると考えられる．

本研究ではこのゲーム内の手の時系列情報は戦略的描写表現によって区分化が可能であるという考えのもと，戦略的描写表現の識別を試みた．

3. 節で示すように，本研究では評価の対象に戦略的描写表現の内の頻度上位 3 種を使用した．使用した 3 つの戦略的描写表現の一般的な解釈の例を以下に示す．なお，これらは盤面の石の配置で描写表現が決まるのではなく，多様な石の配置を抽象化した概念によってその解釈が決まるため，前節の物理的描写表現のように図を用いて表現したとしても，それはごく一部の例にしか過ぎず，全体的な形を表現することは難しい．

受ける 相手の戦略に対して自分の石を守る意味を表す

カカリ 相手の石に対して攻撃（戦略）を開始する意味を表す

打ち込み 相手が占める領域に侵入する意味を表す

ここで，図 2 と図 3 に戦略的描写表現を識別する際に問題となるパターンを示す．図 2 では，白が矢印で指されている石を新たに置いた場面で，この手に描写表現を付ける場合を考える． この手は，白△との位置関係だけを見れば，物理的描写表現のヒラキと捉えられる．しかし，この盤面では黒□への攻撃（戦略）を開始する手と考える方が自然なため，物理的描写表現のヒラキとするのは相対的にみて不適切であり，戦略的描写表現のカカリを用いる方が適切である． 同様に図 3 においても，黒△との位置関係だけを見れば物理的描写表現のヒラキと捉えることができるが，相手が占める領域に侵入する意味であると考える方が自然な場面であるため，物理的描写表現のツケとするのは相対的にみて不適切で，戦略的描写表現の打ち込みを用いる方が適切である．

このように戦略的描写表現を識別する際には，物理的描写表現と比較して相対的にみてどちらが適切かを選択する必要があり，これらをどのように解決するかが重要な問題である．

3. データセット

分類器の評価に使用したデータセットは，SGF ファイルと呼ばれる手の座標と解説 (テキストファイル) が対になったものから構築した．これらは 5 年間の新聞記事の内の利用可能な囲碁掲載欄から収集した． この収集データには 60 ゲーム分の 19 × 19 の大きさの盤上の石の配置を順番に記録したデータが含まれており，そのデータと解説を照らし合わせながら手作業で描写表現の抽出を行った． その結果，描写表現がついた手は総手数 9,209 手の内，1,207 手であった．描写表現は 238 種類にも及んだが，10 回以上出現した描写表現は 34 種類であった．

今回の評価には物理的描写表現の内の頻度上位 10 種と，戦略的描写表現の内の頻度上位 3 種を用いた．評価に用いた描写表現は表 1 に示されている描写表現 13 種類である．

4. 素性

本研究では，関連研究でも用いられていた盤面の石の配置から描写表現を識別するための位置素性と，それに加えて，盤面の石の配置だけでは識別が難しいと考えられる戦略的描写表現を識別するために，囲碁の手は時系列データであることを考慮した時間素性を新たに提案し検討した．

4.1 位置素性

本研究で用いた位置素性のほとんどは，物理的描写表現を識別することを目的とした関連研究によって得られた素性を用いている [Ikeda15]．この位置素性とは，盤面の石の配置から得られる特徴を素性としたもので，その描写表現が付けられる瞬間の盤面の状態だけを見るものである．

4.1.1 FP_1: 座標

盤面の左上を原点として，描写表現が付けられる現在の手の石が置かれた場所を x（1∼19），y（1∼19）で表したもの．これらを FP_1 と呼ぶ．

4.1.2 FP_2: 手数

ゲームの開始から現在の手の石が何手目に置かれたものかを表したもの．これを FP_2 と呼ぶ．

4.1.3 FP_3: 最も近い端までの距離

図 4 で示すように，現在の手の石が盤の端からどれだけ離れているかを表したもの．これを FP_3 と呼ぶ．

4.1.4 FP_4/FP_5: 最寄りの味方の石/敵の石までの距離

図 5 で示すように，現在の手から最も近い味方の石/敵の石までの距離を表したもの．これらを FP_4，FP_5 と呼ぶ．距離の測り方は，関連研究で用いられていた R 距離を用いた [Coulom07, Ikeda15]．R 距離とは以下の式で表され，本研究では，この距

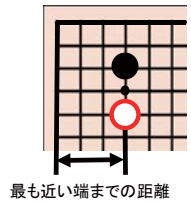


図 4: 素性 FP_3: 最も近い端までの距離



(a) 素性 FP_4: 最寄りの味方の石までの距離　(b) 素性 FP_5: 最寄りの敵の石までの距離

図 5: 素性 FP_4/素性 FP_5: 最寄りの味方の石/敵の石までの距離

離が 6 以内のものを見ることとし，それ以上の距離のものは一律で遠く離れているとみなした．

$$\mathrm{distance}(\delta x, \delta y) = |\delta x| + |\delta y| + \max(|\delta x|, |\delta y|) \quad (1)$$

4.1.5 FP_6/FP_7: 最寄りの味方の石/敵の石から最も近い端までの距離

図 6 で示すように，最も近い味方の石/敵の石が端からどれだけ離れているかを表したもの．これらを FP_6，FP_7 と呼ぶ．最も近い味方の石/敵の石を探索する際には，FP_5/FP_6 と同様に R 距離を用いており，R 距離が 6 以内に対象となる石が見つからない場合は，やはり一律に遠く離れているとみなした．

4.2 時間素性

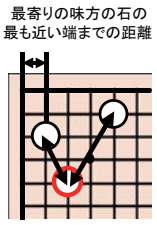
関連研究では現在の手の盤面の状態だけを見ていたのに対して，本研究では新たに戦略的描写表現を識別するために時間素性を提案する [Ikeda15]．時間素性とは囲碁の手は時系列データであることを考慮して，一つ前，二つ前の手から現在の手にどのように遷移しているかを見る素性である．

FT_1/FT_2: 現在の手から 1 手前の手/2 手前の手までの距離

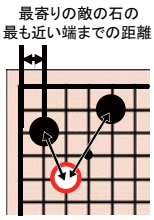
図 7 で示すように，現在の手から 1 手前の手/2 手前の手までの距離を測り，どれだけ離れているかを素性とした．　これらを FT_1，FT_2 と呼ぶ．

戦略的描写表現を識別する際には，図 2，および，図 3 で示すように，物理的描写表現と比較して，相対的に適切かどうかを判断する必要がある．しかし，これらは物理的描写表現と同様に盤面の状態だけで識別することは難しい．よって，本研究では囲碁の手が時系列データであることを考慮することによって，この問題の解決を図った．具体的には，囲碁は一つのゲーム中に戦略が幾つも出現するため細分化が可能である．そして本研究では，その戦略の区切りが戦略的描写表現であるという考えのもと，前回の手から新たな場所に展開した手を戦略の区切りとみて，戦略的描写表現の識別を試みた．

また，距離の測り方に関して，位置素性では盤面の状態を見るために関連研究と同様に R 距離を用いたが，時間素性に関



(a) 素性 FP_6: 最寄りの味方の石から最も近い端までの距離

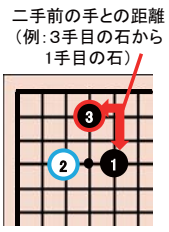


(b) 素性 FP_7: 最寄りの敵の石から最も近い端までの距離

図 6: 素性 FP_6/素性 FP_7: 最寄りの味方の石/敵の石から最も近い端までの距離



(a) FT_1: 現在の手から 1 手前の手までの距離



(b) FT_2: 現在の手から 2 手前の手までの距離

図 7: FT_1/FT_2: 現在の手から 1 手前の手/2 手前の手までの距離

しては，近い石の厳密な位置関係よりも，石がどれだけ離れているかというような，より広い範囲の情報を得ることを考えて以下の式を用いた [Ikeda15]．

$$\mathrm{distance}(\delta x, \delta y) = |\delta x| + |\delta y| \quad (2)$$

5. 評価

5.1 手順

描写表現を生成するための分類器の評価は，物理的描写表現頻度上位 10 種類と戦略的描写表現頻度上位頻度 3 種類の合わせて 13 種類の各表現について，　それぞれの描写表現に含まれる全データを正例とし，それと同数の負例を，正例を除いた 12 種類の全データから無作為抽出し，二値分類の 10 分割交差検定を行った．　また，分類器の実装には Python のモジュールの一つである scikit-learn toolkit の決定木を用いた [Pedregosa11]．さらに，決定木は過学習を抑えるために子ノードに存在するデータ数の最小値を 1 から 7 の間になるようにし，ツリーの深さを 2 から 8 の間に制限した．　評価結果には，10 分割交差検定を用いて，上記の 2 つのパラメータを最適化し，その最適化されたものを最終的な結果として採用した．

5.2 結果

表 1 では，各表現における最適な素性の組み合わせと，これらの最適な素性の組み合わせによって得られる二値分類における性能を示している．表 1 を見ると，ほとんどの描写表現に FP_4(最寄りの味方の石までの距離) と FP_6(最寄りの味方の石の最も近い端までの距離) が含まれており，逆に，FP_1(座標) や FP_3(最も近い端までの距離) はあまり含まれていないことが分かる．この結果から，今回の 13 種類の描写表現に関しては，座標のような物理的な絶対位置を示す素性より，最寄り

表 1: 評価結果 (%)

描写表現の種類	ID	二値分類	正例データ数	評価結果	最適な素性
物理的描写表現	1	ツケ/ツケ以外	60	95.0	FP_4, FP_6
	2	ツギ/ツギ以外	57	86.1	$FP_4, FP_5, FP_6, \underline{FT_2}$
	3	ハネ/ハネ以外	42	89.5	FP_4
	4	キリ/キリ以外	37	84.2	FP_2, FP_4, FP_6
	5	ノゾキ/ノゾキ以外	30	81.7	$FP_2, FP_6, \underline{FT_1}, \underline{FT_2}$
	6	ノビ/ノビ以外	29	95.0	FP_6, FP_7
	7	トビ/トビ以外	25	91.7	FP_4
	8	オサエ/オサエ以外	25	82.5	$\underline{FT_1}$
	9	アテ/アテ以外	21	83.3	$FP_7, \underline{FT_2}$
	10	ヒラキ/ヒラキ以外	19	87.5	FP_6
	合計/平均		345/34.5	87.6	—
戦略的描写表現	11	受ける/受ける以外	29	71.7	FP_1, FP_2
	12	カカリ/カカリ以外	25	95.8	FP_4
	13	打ち込み/打ち込み以外	23	94.2	FP_6
	合計/平均		77/25.7	87.2	—
合計/平均			422/32.5	87.5	—

の味方の石までの距離のような相対的な位置関係を示す素性の方が有効であると言える．

また，見方を変えて，FT_1(一手前の手との距離) は相手の石との距離を見ていて，FT_2(二手前の手との距離) は自分の石との距離を見ていることを踏まえると，FP_4/FP_5(最寄りの味方の石/敵の石までの距離)，FP_6/FP_7(最寄りの味方の石/敵の石から最も近い端までの距離)，FT_1/FT_2 の全ての組み合わせにおいて，自分の石との関係を見ている素性の方が多く出現している．

ただし，FT_1 と FT_2 に関しては，最適な素性として現れた数の差はごく僅かであり，全体的に見てもデータセットの規模が小さいため，一般的に最適な素性として味方の石の方が有効であるとは言えない．また，決定木を作成する際に深さと最小値の設定により汎用的なルールになるような最適化は行っているが，やはりデータセットの規模が小さいため訓練に用いるデータ数が少なく，汎用的なルールになっているとは言えない．

今後は，データセットの規模を大きくし，上記の評価結果が他の多くの描写表現にも適用できることかどうかの確認を行う．また，新たな有効な素性や SVM やランダムフォレストのような他の分類器に関しても検討し，全体的な精度の向上を目指す．

6. おわりに

本研究では，囲碁における各手の解説を生成することに着目し，高度な戦略を学ぶ際に人間プレイヤーを支援するソフトウェアの開発を目的とした．特に，本研究では盤上の石のパターンに基づいて各手の描写表現を生成する課題に取り組んだ．今後はデータセットを拡張し，評価をより正確なものとした上で，SVM などの他の分類手法についても試し，さらなる全体的な精度の向上を目指す．

参考文献

[Coulom07] Coulom, R.: Computing Elo Ratings of Move Patterns in the Game of Go, *Proc. Computer Games Workshop* (2007).

[Ikeda15] Ikeda, K., Shishido, T. and Viennot, S.: Machine-Learning of Shape Names for the Game of Go, *Proc. 14th ACG*, pp. 247–259 (2015).

[Pedregosa11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830 (2011).

[Silver16] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D.: Mastering the Game of Go with Deep Neural Networks and Tree Search, *Nature*, Vol. 529, pp. 484–503 (2016).

[Yoshikawa98] Yoshikawa, A., Kojima, T. and Saito, Y.: Relations between Skill and the Use of Terms — An Analysis of Protocols of the Game of Go —, *Proc. 1st CG*, pp. 282–299 (1998).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ユーザ好みのスキャットを強化学習する 初音ミクとのジャムセッションシステムの開発

Development of a Jam Session System with Hatsune Miku That Learns User's Favorite Scat by Reinforcement Learning

鶴田 穣士 Joji TSURUDA　岡 夏樹 Natsuki OKA　田中 一晶 Kazuaki YANAKA

京都工芸繊維大学 大学院工芸科学研究科
Graduate School of Science and Technology, Kyoto Institute of Technology

Musicians often use jam sessions as a way to communicate with music. Many studies on jam session systems, which enable jam sessions between people and machines, have been made. In jam sessions, singers often use the scat singing. In our proposed system, a user plays keyboard and the system will reply in scat singing of Hatsune Miku according to the user's performance. We aim to make the system learn the user's preferred scat words by reinforcement learning through jam sessions. We implemented the learning system and confirmed that the system could learn the user ' s preference in scat words.

1. はじめに

近年，家庭用コミュニケーションロボットなどの普及に伴い，人間と機械とのコミュニケーションがより一般化しつつある．音楽の分野においても，人間の代替として機械がコミュニケーション相手として使われるようになってきている．機械による自動演奏を人間が聴くという形態としては，初音ミク *1 などのバーチャル・アイドル主演のライブコンサートが催されている．透明なスクリーン上に映された 2 次元のアイドル初音ミクに対して観客たちは，人間相手と同じように手を振り声援を送る．それに対して初音ミクは手を振り返したり「ありがとう」と言い返したりする．

しかし，これらは一見，機械が声援に対する応答をしている，つまり人間と機械のインタラクションができているように見えて，実際は，観客の反応に関係なく事前にプログラミングされた通りのことをしているにすぎない．

よりインタラクティブな音楽を用いたコミュニケーションとして，ジャムセッションと呼ばれるものがある．これは楽譜に書かれた通りに演奏するアンサンブルとは異なり，演奏者同士が即興的に演奏していく形態のことである．ある演奏者による演奏が他の演奏者に影響を与え，新たな演奏として返ってくるといった演奏者間のインタラクションが顕著である点が他の音楽を用いたコミュニケーションとの大きな違いである．しかし，ジャムセッションを行うには，他の演奏者の協力が必要である．それを個人でいつでも可能としたのが，ユーザと機械とのジャムセッションシステムであり，これまで多く研究されてきた．

また，ジャムセッションでは，ボーカルが声を楽器のように使う歌唱法であるスキャットが使われる場合がある．スキャットとは，あらかじめ決められた歌詞ではなく，「ラララ〜」や「シャバドゥビ」，「ダバダバ」といった意味の無い音を即興的に歌うことである．本研究では，近年の歌声合成技術の発達を受け，スキャットを用いた初音ミクとのジャムセッションシステムを提案する．初音ミクという一般的に認知されている既存のキャラクターを使うことで，ただ楽器が鳴るだけのジャムセッションシステムに比べて，人間の代替としてのリアリティを感じられ，コミュニケーション相手の存在をより強く意識することができると考えられる．

連絡先: 京都工芸繊維大学 大学院工芸科学研究科 情報工学専攻 インタラクティブ知能研究室，tsuruda@ii.is.kit.ac.jp

*1 クリプトン・フューチャー・メディア社製の歌声合成ソフトの製品名およびそのパッケージに描かれたキャラクターの名称．

2. 目的

本研究では，ジャムセッションに馴染みのない演奏者でも気軽に始められるような初音ミクとのジャムセッションシステムの開発を目指す．対コンピュータであれば，初心者でも人間相手より，臆せず気軽に始められ，その面白さに気付き，興味を持つきっかけになり得ると考えられるからである．スキャットに使われる言葉は，よく使われるものはあるが，特に決まりは無く，歌唱者が好きなように歌うものであって，聴き手側でも好みが分かれやすい．よって，機械である点を生かし，強化学習を用いて，ユーザとのジャムセッションを通して，各々のユーザの好みに合うようなスキャットの言葉を学習するジャムセッションシステムを提案する．また，人間には歌いにくい，歌えないような歌を自由に発声可能な点が歌声合成技術の長所である．したがって，ユーザの好みに応じた，人間のスキャットの型にとらわれない，自由で新しいスタイルのスキャットの生成も期待できると考えられる．

3. 関連研究

ジャムセッションシステムは，これまで様々な研究が行われてきた．演奏形態として，主にピアノ同士のジャムセッションを扱ったものが多い．コード進行をあらかじめ決めておくもの [Nishijima 92] [後藤 99] や，ユーザに合わせてコード進行を変えるもの [青野 94]，ユーザがメロディを弾き，それに沿ったコード進行でシステムが伴奏を行い，そのコード進行に応じてユーザがメロディを弾くといったもの [北原 09] がある．しかしこれらのシステムはコード進行といった専門知識を有するジャムセッション熟練者のためのシステムであり，初心者には敷居が高く感じられる．コード進行を有さない打楽器であるドラム同士のジャムセッションシステムの開発もされている [西嶋 91]．ピアノ同士でコード進行の知識が無い初心者でも気軽に始められるジャムセッションシステムとして Yotam らは「A.I. Duet」を開発した [Yotam 17]．多数の既存の楽曲のメロディパターンを LSTM を応用したニューラルネットワークを用いて学習させ，ユーザが入力したメロディに自然に繋が

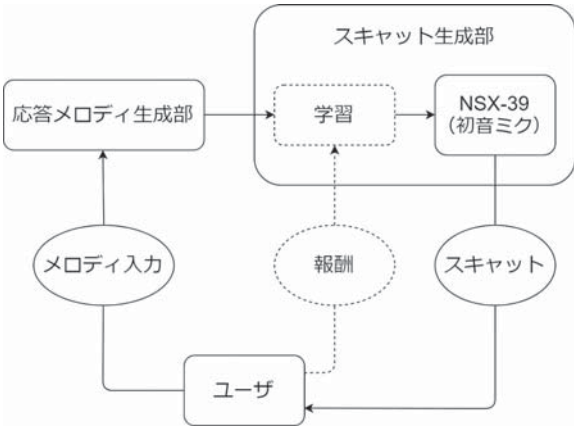


図 1: 提案システムの概要図



図 2: 提案システムを使用している様子

るようなメロディをシステムが返すというものである．web ブラウザ上で，何の音楽知識がない人でも簡単にジャムセッションを体験できるという点において他のジャムセッションシステムより優れている．また，「A.I. Duet」は Magenta[*2] というオープンソースプロジェクトに基づくもので，誰でもシステム改良を行うことが可能となっているため，本研究ではこのシステムを利用して開発を行った．

4. 提案システム概要

提案システムの概要図を図 1 に示す．本研究では，図中の点線部を新たに作成した．提案システムは大きく分けて，応答メロディ生成部とスキャット生成部に区分される．提案システムの大まかな流れは以下の通りである．

1. ユーザがキーボードによってメロディを入力する．
2. 入力されたメロディをもとに，応答メロディ生成部にて，スキャットのメロディを生成する．
3. スキャット生成部にて，生成されたメロディと学習経験をもとに初音ミクがスキャットとして歌う．
4. ユーザはそのスキャットを自分の好みかどうかで評価して，報酬を与える．
5. 与えられた報酬に応じてユーザ好みのスキャットを学習し，1 へ戻る．

提案システムを使用している様子を図 2 に示す．ぬいぐるみがある位置から初音ミクのスキャット音が流れる．現時点までの開発は，主にスキャット生成部について行ってきたため，応答メロディ生成部については 5. 節で簡単に紹介し，スキャット生成部は 6. 節にて詳しく説明する．

5. 応答メロディ生成部

先行研究の「A.I. Duet」を利用して，メロディ生成やディスプレイ表示を行う．ユーザがキーボードによって，適当なメロディを弾くと，ピアノ音がスピーカーから流れ，それと同時に MIDI 信号が PC に送られる．それをもとに，図 2 のようにディスプレイ上にどの音階をどのタイミングで弾いたのかがリアルタイムで視覚的にわかるように表示される．弾いた音階を表すオブジェクトが時間と共に上へ流れていく仕組みである．赤紫色のオブジェクトがユーザが入力したメロディ，緑色のオブジェクトが出力された応答メロディを表す．

ユーザが入力した MIDI 信号（音階と音の長さ情報のみを扱う）を事前学習したニューラルネットワークを通して，入力したメロディに自然に続くような MIDI 信号を生成し，スキャット生成部に送る．ここで使うニューラルネットワークは，Lookback RNN である．Lookback RNN とは，LSTM をもとにした機械学習による自動作曲のためのニューラルネットワークモデルである．楽曲特有の繰り返し構造を学習しやすくしたモデルであり，相手のフレーズを模倣したようなフレーズを返すことが多いジャムセッションシステムに適している．提案システムでは，Magenta による学習済み配布モデルを使用した．

6. スキャット生成部

6.1 スキャット生成部の概要

スキャットには，メロディと言葉が必要である．提案システムでは，メロディは応答メロディ生成部で生成された MIDI 信号をそのまま用いて，言葉のほうを強化学習する．スキャット生成には，リアルタイムに初音ミクの歌唱が可能なデバイス NSX-39 を用いた．また，強化学習の手法としては Q 学習を用いた．スキャット生成部ではまず，応答メロディ生成部で生成した MIDI 信号情報と，強化学習によって決まる言葉の情報を NSX-39 へ送り，初音ミクがスキャットを歌う．次に，ユーザはそのスキャットを自分の好みかどうかで評価し，図 2 中下部のように PC キーボード入力によって報酬を与え，次の演奏入力を開始する．

以上を繰り返すことでユーザ好みのスキャットを強化学習していく．

6.2 学習方法

提案システムでは，強化学習手法の一つである Q 学習を用いた．

（1）行動

スキャットに使う言葉の種類を行動とした．以下の点を考慮する．

- 一般的にスキャットに使われる言葉と使われない言葉を含む．

*2 http://magenta.tensorflow.org/

- 1 音だけでも成り立つ言葉と 2 音セットで使われる言葉を含む.
- 言葉間の違いが聴き取りやすいように言葉の種類を選ぶ（異なる母音など）.
- 行動は多すぎると学習に時間が掛かる.

以上のことを踏まえて検討した結果,「ラ」,「ドゥ」,「ビ」,「ケ」の 4 つの言葉を選択した．これらの内どの言葉を使うかを学習する.

（2）状態

今回扱う MIDI 信号は音階と，音の長さ情報のみで構成される．本来メロディは系列データとして扱うべきであるが，提案システムは実時間学習であり，系列データとして扱うと状態数が非常に多くなり，学習に時間が掛かるため，適さないと考えられる．そのためスキャット 1 音に対して以下のように離散化したものを状態とした.

- 入力できる音階の中心を境に低いか高いかの 2 状態.
- 音の長さが 0.25 秒より短いか長いかの 2 状態.

さらに，状態数の関係上，系列データとして扱うことはできなかったが，系列データの性質に近づけるため，1 つ前の行動を状態として追加した．スキャットでは音の高さや長さの系列より，言葉の繋がりがより重要であると考えられるからである．行動は 4 つあるが，1 音目は前の行動が無いので，行動無しという状態も追加した 5 状態となる.

したがって，合計で $2 \times 2 \times 5 = 20$ 状態となる.

（3）方策

行動選択確率 $\pi(s,a)$ には，式 (1) に示したソフトマックス手法を用いる.

$$\pi(s,a) = \frac{exp(Q(s,a)/\tau)}{\sum_{p \in A} exp(Q(s,p)/\tau)} \tag{1}$$

ただし，$\tau(> 0)$ は温度と呼ばれるパラメータである．温度 $\tau = 0.3$ とした.

（4）報酬

ユーザは初音ミクのスキャット（1 フレーズ）が終わるとそのフレーズ全体に対する評価（メロディーでなく言葉に対する評価）を行い，PC キーボードを用いて，報酬を与えることができる．良いと判断した場合は +1，悪いと判断した場合は −1 の報酬を与え，良くも悪くもないと判断した場合はユーザは何もする必要は無く，その場合報酬は与えられない.

（5）Q 値更新

報酬が与えられたときに，対象となるスキャットを構成する一音一音に対して更新式 (2) によって Q 値の更新を行う.

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r - Q(s,a)) \tag{2}$$

ただし，r は報酬であり，学習率 $\alpha = 0.1$ とした．また，Q 値の初期値は全て 0 とした.

7. 実験

7.1 実験環境

執筆者の 1 人によって学習動作確認実験を行った．メロディ入力からスキャット出力までを 1 エピソードとして，250 エピソード行った.

表 1: 学習後の Q 値表

状態 音の高さ	状態 音の長さ	状態 前の行動 \ 行動	「ラ」	「ドゥ」	「ビ」	「ケ」
低い	短い	「ラ」	0.197	0.143	-0.276	-0.230
		「ドゥ」	-0.163	-0.006	-0.340	-0.074
		「ビ」	-0.264	-0.257	0.049	-0.360
		「ケ」	-0.278	0.004	-0.188	-0.344
		無し	0.100	0.271	-0.264	-0.271
	長い	「ラ」	-0.337	0.645	-0.243	-0.107
		「ドゥ」	0.515	0.253	0.486	***-0.570***
		「ビ」	0.122	0.046	-0.386	-0.293
		「ケ」	***-0.506***	-0.263	-0.344	***-0.522***
		無し	0.202	-0.144	-0.379	-0.383
高い	短い	「ラ」	-0.385	0.044	-0.332	-0.190
		「ドゥ」	-0.119	0.292	0.131	-0.143
		「ビ」	-0.210	-0.092	-0.265	-0.247
		「ケ」	***-0.613***	***-0.453***	-0.410	***-0.556***
		無し	0.018	-0.116	-0.048	-0.344
	長い	「ラ」	-0.340	0.004	-0.210	-0.410
		「ドゥ」	-0.071	0.213	0.678	-0.421
		「ビ」	-0.350	0.516	-0.147	-0.415
		「ケ」	-0.344	-0.126	***-0.613***	***-0.522***
		無し	0.399	0.501	***-0.445***	***-0.530***

表 2: 言葉の遷移に着目した学習後の Q 値の平均値

前の行動 \ 行動	「ラ」	「ドゥ」	「ビ」	「ケ」
「ラ」	-0.216	0.209	-0.265	-0.234
「ドゥ」	0.040	0.188	0.239	-0.302
「ビ」	-0.175	0.053	-0.187	-0.329
「ケ」	***-0.435***	-0.210	***-0.389***	***-0.486***
無し	0.180	0.128	-0.284	-0.382

7.2 結果

250 エピソード後の Q 値を表 1 に示す．数値の高い上位 10 項の背景を薄く塗り，下位 10 項の文字を太字斜体で表した．また，表 2 では，言葉の遷移に着目し，表中の数値はそれぞれに対応する Q 値の平均を取ったものである．数値の高い上位 3 項の背景を薄く塗り，下位 3 項の文字を太字斜体で表した． 学習していないシステムと比べて，250 エピソード学習を行ったシステムの方が，より自分好みのスキャットが出力されるように感じられた.

7.3 考察

表 1 をみると，最も Q 値が高くなったのは,「ドゥ」のあとに高く長い「ビ」が続くスキャットとなっている．これは一般的によく使われるスキャット「ドゥビ」と言葉は同じである．しかし，一般的なスキャットでは「ドゥビー」と語尾を伸ばすことはなく，短い「ビ」で終わるか，すぐにまた「ドゥビドゥビ」と繰り返すことが多い．これは，提案システムによる新しいスタイルのスキャット生成ができたのではないかと考えられる．1 音目の高く長い音に対しては「ラ，ドゥ」と「ビ，ケ」に顕著な差が表れている．したがって，1 音目の高く長いメロディは好みを判断する上で重要な要因となっていることがわかる．また，Q 値の低いところに着目すると，下位 10 項目中 9 つが「ケ」に関連していることがわかる.「ケ」は一般的なス

キャットには使われない言葉であり，歌いにくい言葉ではあるが，聴き取る側も好みになりにくい言葉だったといえる．

表 2 をみると,「ドゥ」が総じて 1 番好まれた言葉だったことがわかる．1 音目から数値の高い言葉を遷移順に並べると,「ラドゥビドゥビドゥビ･･･」と続く.「ドゥ」→「ドゥ」の遷移もしやすいので，たまにドゥが連続するような「ラドゥドゥビドゥビドゥドゥビ･･･」といったようなスキャットが好みであったと考えられる．

8. 課題点と改善策

学習に時間がかかる点

今回の実験では，実時間で 30 分もかかった．改善策として，状態や行動の分け方，パラメータの調整といった学習手法の見直しなどが考えられる．

メロディも評価してしまう点

スキャット学習部では本来，メロディに適した言葉だけの評価で，応答メロディ生成部によって生成されたメロディ自体は考慮しない．しかし，心理的にメロディ自体が悪いとそれにつられて悪い評価をしてしまいがちになる．提案システムではメロディと言葉を完全に分けて学習させたが，メロディも報酬によって強化学習できるようになれば，この問題を解決できると考えられる．

状態を粗く離散化した点

人とのインタラクションを通した学習を可能とするため，状態数を抑える必要があったため，状態は各時点の音の高さと音の長さを極めて粗く離散化したものとした．このため，音の高さや長さに応じた言葉の学習ができたとは言い難い結果となった．Q 関数をニューラルネットなどを用いて関数近似することで離散化を不要にして，音の高さや長さに応じたユーザの好みを学習できる可能性がある．さらに，音の高さや長さの変化も考慮した好みの学習を可能とすることも検討したい．

初音ミクが意識しにくい点

提案システムでは，初音ミクのぬいぐるみからスキャット音が流れるようにしたが，動かない人形と歌声だけではセッション相手である初音ミクを意識しにくかった．ディスプレイ上に初音ミクを表示し，スキャット時にリップシンクさせるといったような動きを加えることで初音ミクがスキャットを歌っているという感覚がより強まると考えられる．

9. まとめ

本研究では，ユーザ好みのスキャットを強化学習によって学習していくジャムセッションシステムを提案し，実装を行った．提案システムは応答メロディ生成部とスキャット生成部に大きく分けられ，応答メロディ生成部では，先行研究の「A.I. Duet」 [Yotam 17] を用いた．スキャット生成部では，Q 学習を用いてユーザの好みのスキャットを学習させた．執筆者の 1 人が提案システムを試用することで，ユーザ好みのスキャットが学習できることを確認した．しかし，それと同時に課題も多く見つかり，それに対する改善策を検討した．今回は執筆者の 1 人だけによる実験であったため十分な評価情報が得られなかった．今後は，今回見つかった課題点の改善と共に，複数の一般被験者を用いた評価実験を実施する必要がある．

参考文献

[Nishijima 92] Nishijima, M. and Watanabe, K.: Interective Music Composer based on Neural Networks, *ICMC*, pp. 53–56 (1992)

[Yotam 17] Yotam Mann : A.I. Duet, (2017),〈 http://experiments.withgoogle.com/ai/ai-duet〉(参照 2018-2-5)

[後藤 99] 後藤 真孝, 日高 松本 英明, 黒田 洋介, 村岡 洋一 : 仮想ジャズセッションシステム : VirJa Session, 情報処理学会論文誌, Vol. 40, pp. 1910–1921 (1999)

[西嶋 91] 西嶋 正子, 村上 公一 : ニューロ・ドラマー, 計測と制御, Vol. 30, pp. 344–347 (1991)

[青野 94] 青野 裕司, 片寄 晴弘, 井口 征士 : バンドライクな音楽アシスタントシステムについて, 情報処理学会研究報告音楽情報科学（MUS）, Vol. 1994, pp. 45–50 (1994)

[北原 09] 北原 鉄朗, 戸谷 直之, 徳網亮輔, 片寄 晴弘:Bayesian-Band : ユーザとシステムが相互に予測し合うジャムセッションシステム, 情報処理学会論文誌, Vol. 50, pp. 2949–2953 (2009)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習によるユーザ評価モデルを導入した遺伝的プログラミングによる音楽自動生成手法の提案

Automatic Music Composition System Based on Genetic Programming and Surrogate Model with Deep Learning

山本 周典 *1 Hironori Yamamoto　長谷川 拓 *2 Taku Hasegawa　森 直樹 *2 Naoki Mori　松本 啓之亮 *2 Keinosuke Matsumoto

*1大阪府立大学 Osaka Prefecture University　*2大阪府立大学工学研究科 Graduate School of Engineering, Osaka Prefecture University

Automatic music composition is one of the most difficult and attractive challenges in the artificial intelligence (AI) field. In order to tackle this challenge, an approach using interactive evolutionary computation (IEC) is drawing attention because IEC takes human emotions into consideration. The major problem in using IEC is that the number of evaluations from one user is limited due to user fatigue. To tackle this problem, a surrogate model is often introduced into IEC. An approach based on deep learning (DL) is also common in this field because of many quantitative futures. However, the approach hardly considers human emotions.

In this study, we proposed the automatic music composition system based on IEC and a surrogate model called evaluation model. The model is constructed with a DL model, thus our system can compose music reflected human emotions quantitatively. The experiments are carried out to show the effectiveness of the proposed method.

1. はじめに

近年，計算機による音楽の自動生成に関する研究が積極的になされており，新しい文化や産業の創造に繋がるとして多方面から強く期待されている．しかしながら，作品の評価は個人の嗜好や感性に大きく依存するため，これを定量的に評価することは非常に困難とされている．このような問題に対して，人間の評価系そのものを評価関数として最適化システムに導入した対話型進化型計算 (Interactive Evolutionary Computation: IEC)[1] が注目をされている．

IEC は定量化が難しい人の感性を評価できるという利点がある一方で，ユーザ負荷の観点から解を評価できる回数に大きな制約があり，特に音楽の自動生成の分野においては各々の解を逐次的に評価する必要があることから，評価回数の制約が顕在化しやすいという問題がある．また，深層学習 [2] を用いた音楽の自動生成の手法も提案されている．特に，Variational Autoencoder (VAE)[3] をはじめとする生成モデルを活用した方法は，入力データが潜在的に内包している意味を反映し写像した潜在空間を構築可能な点からも幅広い応用が期待されている．しかしながら，深層学習を用いた手法は主に事前に学習した分布の再現になるため，生成物に対してユーザの感性を直接的に反映させることは困難であるという課題も存在する．

本研究では以上の点を背景として，適応度景観を学習し近似的に評価関数を推定する surrogate model[4] を応用した IEC による対話型音楽自動生成システムを提案する．本システムでは音の高さおよび長さの概念を反映した木構造によって楽曲を表現した．また，Genetic Programming (GP) の拡張手法である Genetic Programming with Multi-Layered Population Structure (MLPS-GP)[5] を音楽の探索アルゴリズムとして用いた．更に，既存の楽曲データから音楽的特徴を抽出，学習し，個体に対する近似評価をする評価モデルを surrogate model として導入した．評価モデルの学習に当たっては VAE を用い，楽曲学習時に生成される潜在空間に着目しつつ IEC と組み合わせることで，定量的かつユーザの嗜好を反映した音楽の自動生成を提案する．

連絡先: 山本 周典，大阪府立大学，h.yamamoto@ss.cs.osakafu-u.ac.jp

2. 音楽データの分散表現の獲得

本研究では Variational Recurrent Autoencoder (VRAE)[6] を使用することで音楽の潜在変数を獲得し，これを音楽の分散表現として扱う．VRAE は Variational Autoencoder (VAE)[3] と呼ばれる生成モデルの拡張手法で，エンコード部およびデコード部で Recurrent Neural Network を使用しているため音楽のような時系列データに適していると考えられる．VAE および VRAE の最適化項目としては，潜在変数の確率分布の形状に関わる項（latent loss）および潜在変数を事前分布として元データの復元に関わる項（reconstruction loss）がある．

3. 音楽の自動生成システムの概要

本章では進化型計算で音楽を扱うための個体表現の定義，surrogate model の利用方法，提案システムの流れおよび入力曲の定義について述べる．

3.1 曲における個体表現

本研究では Genetic Programming (GP) を適用するために曲を木構造によって表現する．この木構造によって曲の音高および音価が表現可能である．図 1 に，GP における曲の個体表現と楽譜の対応関係を示す．終端ノードを音高とし，非終端ノードを子ノードへの音高の倍率および子ノードへの分岐数とした．i を子ノードの音価への倍率，j を子ノードへの分岐数として，非終端ノードを S_j^i で表す．i，j に関しては，$i \in \{1.0, 1.5, 2.0\}$，$j \in \{1, 2, 3, 4\}$ と定めた．また本研究では，根ノードの音価を 16 分音符で固定する．なお，$S_j^{1.5}$ に関しては付点音価を表現することを目的としているため，ある終端ノードの音価を計算するに際して，その親ノードが $S_j^{1.5}$ の場合においてのみ音価を 1.5 倍とし，それ以外の $S_j^{1.5}$ に関しては音価を 1 倍とするように設定した．木構造表現から楽譜に変換する際は，終端ノードに関して深さ優先探索の先行順で辿ることで変換がなされる．

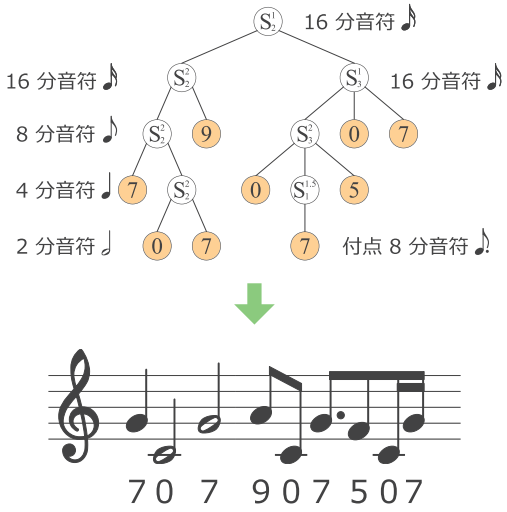


図 1: 木構造と楽譜の対応関係の例

3.2 surrogate model の導入

IEC を利用する際はユーザ負荷を考慮し，できるだけ探索に必要な解の評価回数を減らす必要がある．そこで本研究では，ユーザの代わりに適応度景観を学習し近似的に解を評価する surrogate model を導入した．これによってユーザに直接的な負荷をかけることなく十分な探索を実現する事ができる．さらに，ユーザの評価は，直接的に解の優劣を示す適応度のみではなく，主にモデルを作成し更新するために利用されるため，ユーザ負荷を小さく抑える一方で精度の高い近似評価を実現している．モデルと実際の評価関数をどのように利用し使い分けるかを示す evolution control に関しては，一定世代ごとにユーザの評価をモデルに反映する generation-based evolution control[7][8] として捉えることができる．一般的な generation-based evolution control では一定世代ごとに個体群すべての個体を評価するが，本手法においてはモデルにおける “適応度” が高い個体，いわゆるエリート個体をユーザに提示し，評価させるという枠組みになっている．以下 surrogate model として構築したモデルを評価モデルと呼ぶ．

3.3 提案システムの流れ

本システムはユーザの嗜好に合った曲をユーザの評価に基づいて進化的に獲得することを目的としたシステムである．なお，評価モデルが楽曲の特徴量を学習するのに使用した曲を入力曲，進化型計算による探索中の個体を探索曲，探索終了後のエリート個体を進化曲と呼ぶ．

1. 入力曲に基づいて評価モデルが構築される．
2. 評価モデルによって探索曲を近似評価し，この評価に従い MLPS-GP が進化曲を生成する．
3. ユーザが進化曲を実評価する．
4. 評価モデルがユーザの実評価に基づいてユーザの嗜好を推定し，さらに IEC により評価モデルを更新し，初期個体群に進化曲を反映する．
5. 2. に戻る．あるいは，ユーザが満足する曲が生成できた場合は終了する．

3.4 入力曲の定義

本研究では，入力曲を一般入力曲および選好入力曲の 2 種類の入力曲に大別し，評価モデルに学習させる．一般入力曲とは人が普段から耳にし聴き心地の良いと感じるような普遍的な曲のことで，これらを大量に学習することによって汎用的な楽譜情報の獲得が期待される．一方，選好入力曲とは，進化曲の特徴付けのために一般入力曲の中から抽出された曲のことで，ユーザの嗜好を獲得することが期待される．本システムでは，一般入力曲の学習によって獲得した普遍的な音楽の特徴およびユーザに選択させた選好入力曲の特徴を組み合わせることで，ユーザの嗜好に合った進化曲の獲得を目的としている．

4. 評価モデル

本章では，提案システムにおいて重要な評価モデルによる曲の定量的評価手法について述べる．

4.1 音楽データの規格化

VRAE を評価モデルとして用いるために，VRAE で学習する音楽データを規格化した．この規格化の 1 つ目の理由に MLPS-GP による効率的な探索の実現が挙げられる．MLPS-GP では局所探索の際にランダムなノード選択を含むため，すべての音楽データが含む音について探索をするには莫大な実行時間を要する．特に，本システムではユーザの実評価を定期的に必要とするため，近似評価に必要な実行時間はボトルネックとなる．よって，学習データを規格化した上で MLPS-GP の探索範囲を限定することで，短時間でより良い解の探索が期待できる．

また，2 つ目の理由として VRAE の汎用的な学習の実現が挙げられる．生の音楽データの場合，音高毎に存在する音楽データに偏りがあることが予想され，特に音楽データが少ない音高の範囲では十分な学習が成されない可能性がある．そこで，規格化により出現する音の種類を限定することで，より汎用的で時系列的な情報に基づいた学習をすることが可能になると考えられる．

規格化の方法は以下のとおりである．

- 音高に対して，階名を保ったまま 3 オクターヴ内に収まるように変換．
- 音価に対して，曲全体の値の比を保ったまま 16 分音符～全音符に収まるものに変換．
- 変換の過程で 1. および 2. で示した範囲に入らなかった音については “unk” タグを付与．

規格化を経て “unk” タグの割合は全体の 9.398e-5 程度になった．このことから，ほとんどの音楽データは全体を通して音が 3 オクターヴかつ 16 分音符～全音符内に存在しており，上記の規格化を通してもほとんど元のデータ構造を保持できていることがわかる．また，MLPS-GP の音の探索範囲を限定することについても合理的であると考えられる．

4.2 個体評価値

次に評価モデルにおける個体評価値を定める．まず，VRAE に対して一般入力曲を学習させる．次に，選好入力曲をユーザが設定し，これらを VRAE のエンコード部によって潜在空間に写像することで，潜在変数 $\boldsymbol{z}_t$ を得る．また，探索曲 x についても同様に潜在ベクトル $\boldsymbol{z}_x$ を算出し，得られた潜在ベクトル $\boldsymbol{z}_x$ をデコードすることで $\hat{x}$ を得る．以上の変数を用いて，

表 1: MLPS-GP のパラメータ

適応度評価回数 N_{e}	20000
初期化における減衰率 r_{dump}	0.8
初期の深さ制限 D_{init}	3

探索曲 x に対する個体評価値 $f(x)$ を次のように定義した．

$$f(x) \quad = \quad (\frac{1}{T}\sum_{t}|\boldsymbol{z}_t - \boldsymbol{z}_x| + \alpha|g(x,\hat{x}) - \beta|)^{-1} \qquad (1)$$

$g(x,\hat{x})$ は x の reconstruction loss に相当し，本研究では関数 g を softmax cross entropy とした．また，α および β はパラメータ，T は選好入力曲の個数である．

（1）式において $\frac{1}{T}\sum_t|\boldsymbol{z}_t - \boldsymbol{z}_x|$ はユーザが設定した選好入力曲の集合に対して潜在空間において探索曲 x のユークリッド距離が近くなることを目的とする項である．これにより，ユーザの嗜好を進化曲に取り入れることが期待される．また，(1) 式の $|g(x,\hat{x}) - \beta|$ は探索曲 x の reconstruction loss がパラメータ β に近くなることを目的とする項である．この β は学習済みの VRAE に対してテストデータを入力した際に得られる reconstruction loss によって決定される値である．このようにパラメータ β を定める理由として VRAE によって形成される潜在空間には曲の特徴量をうまく写像している部分とそうでない部分があり，単純に選好入力曲とのユークリッド距離を近くするだけでは楽曲としての特徴を十分に反映した進化曲が得られないと予想されるためである．よって，テストデータの reconstruction loss に近づくような個体を探索することで，VRAE が上手く写像できる空間の中での探索が可能となり，楽曲の一般的な特徴を反映した進化曲を得ることが期待できる．このようにテストデータの reconstruction loss に基づいて，教師無しで入力と学習時のデータセットとの差異を測る研究 [9][10] も複数報告されており，本研究でも有用であると考えられる．

4.3 適応度

ユーザの嗜好を反映する上で，生成する曲の長さは重要な要素である．故に，本システムでは，ユーザが必要とする曲の曲長を t_{d}，進化曲の曲長を t_x とし，MLPS-GP における適応度 $F(x)$ は (1) 式 を用いて，次式のように与えられる．

$$F(x) = \frac{t_{\mathrm{d}}}{t_{\mathrm{d}} + |t_{\mathrm{d}} - t_x|} f(x) \qquad (2)$$

5. 実験

ここでは音楽的特徴に基づく評価モデルによる探索曲の進化について解析をする．具体的には生成された進化曲の個体評価値および個体評価値を構成する各類似度について示すことで，評価モデルにおける個体評価値の妥当性，および進化曲が適切に進化しているかどうかを調べる．

5.1 実験条件

選好入力曲を “SCHOENE AUGEN SCHOENE STRAHLEN” の 1 曲に設定し，（2）式の適応度に基づき音楽を自動生成した．

表 1 に本実験で用いた MLPS-GP のパラメータを示し，表 2 に今回学習した VRAE のパラメータを示す．latent loss についてはその係数を 0 から 0.2 の範囲において指数関数的に上昇させた．その際の更新率は 0.9999 に設定した．また，VRAE で用いる最適化アルゴリズム Adam のパラメータ alpha については更新率を 0.999 として 1e-3 から 1e-6 の範囲で指数関数的に減少させた．また，表 3 に個体評価値（1）式のパラメータを示す．β は学習後の VAE に対してテストデータを入力した際の reconstruction loss の平均値とした．以上の設定に基づいて乱数シードを変えた進化曲およびランダム曲を 140 曲用意した．ただし，ランダム曲とは 3.1 節で示した曲の個体表現で表現可能な個体をランダムで生成した曲のことで，進化のベースラインとして用いた．

表 2: VRAE のパラメータ

batch size	512
epoch	5000
embed units	256
hidden state	512
latent units	128
optimizer	Adam
alpha (Adam)	[1e-6, 1e-3]
beta1 (Adam)	0.05
beta2 (Adam)	0.001
dropout	0.8
loss fuction	softmax cross entropy

表 3: 評価モデルパラメータ

個体評価値 α	0，1
個体評価値 β	3.075
曲長 t_{d}	30 秒

5.2 実験結果

表 4 にランダム曲および進化曲間で Welch の t 検定にかけた際の p 値を示す．また，図 3 に得られた進化曲の一例を示す．

以下，表 4 の結果より考察する．まず，進化曲（$\alpha = 1$）において，ランダム曲と比較するとどちらの項についても $p < 0.01$ より有意水準 1% において有意差が見られ，進化が個体評価値にもとづいて成功していることがわかる．次に，進化曲（$\alpha = 0$）においても同様に $p < 0.01$ より有意水準 1% において有意差が見られた．進化曲（$\alpha = 0$）の場合，(1) 式における $\alpha = 0$ であるため reconstruction loss の項の最適化は本来行われないはずである．しかしながら，ベースラインであるランダム曲とはこの項について有意差が生じている．この理由として，VRAE によって形成されている潜在空間が適切であるため，選好入力曲との距離を小さくすることによってランダムな音列から曲らしい音列へと近づいていっていると考えられる．一方で，進化曲（$\alpha = 1$）および進化曲（$\alpha = 0$) において，すべての項目について $p < 0.01$ より有意水準 1% で有意差が見られた．この結果より，個体評価値 1 のパラメータ α の選定が非常に重要であることが分かる．

また，図 2 に，ある進化の過程における選好入力曲，テストデータ，および最大個体評価値を更新した個体（以下優良個体）を潜在空間上に写像し t-SNE で可視化したものを示す．図 2 において選好入力曲は水色のクロス，テストデータは灰色のサークル，優良個体は色の付いたサークルによって表現されており，優良個体については青から赤にかけて個体評価値が高くなっていくように色が付けられている．この結果から，MLPS-GP によって潜在空間上で選好入力曲の周辺を中心に探索がなされていることがわかる．一方で，図 2 に示した結果より総評価個体数に対する優良個体数が非常に少ないことも示しており， MLPS-GP による探索が不十分であることも

表 4: ランダム曲および進化曲間で検定をした時の p 値

	選好入力曲との距離 $\frac{1}{T}\sum_t \lvert \boldsymbol{z}_t - \boldsymbol{z}_x \rvert$	reconstruction loss $g(x, \hat{x})$
ランダム曲 - 進化曲（$\alpha = 1$）		
p 値	1.569e-212	7.720e-127
ランダム曲 - 進化曲（$\alpha = 0$）		
p 値	3.409e-179	2.496e-23
進化曲（$\alpha = 1$）- 進化曲（$\alpha = 0$）		
p 値	5.654e-60	4.471e-88

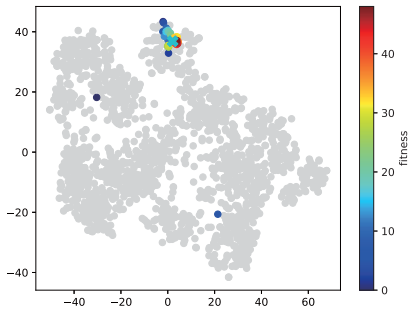


図 2: 優良個体の潜在空間上での変遷

示している．このことから，より効率的な探索手法として深層学習に基づく手法の探索アルゴリズムおよび探索オペレータの構築が必要であると考えられる．また，これらの探索性能には VAE の潜在空間の精度も重要な要因となっているため，VAE 側のより適切な構造およびパラメータの調整も重要な課題である．

5.3 アンケート実験

本システムを用いて得られた進化曲を用意し，アンケートを用いてその評価をした．被験者は，大学生および大学院生計 14 名 (男性 : 12 名，女性 : 2 名) で，選好入力曲およびそれぞれの選好入力曲から進化した進化曲を 3 曲ずつ用意し，被験者に選好入力曲に対応していると思われる進化曲を選択させた．この結果，すべての被験者が適切な選好入力曲および進化曲の組み合わせを選択した．以上の結果より本システムが選好入力曲の情報を反映しつつ進化曲を生成していることがわかった．

6. まとめと今後の課題

本論文では対話型進化型計算および深層学習を用いた音楽の自動生成のシステムを提案した．まず，システム概要を示し，次に評価モデルの構築方法および楽曲の近似評価について示した．実験では優良個体および進化曲に基づき進化の解析をした．また，アンケート実験によって本システムの有用性を示した．今後の課題としては，深層学習に基づく手法の探索アルゴリズムおよび探索オペレータの構築，コードの自動付与が挙げられる．

7. 謝辞

なお，本研究は一部，日本学術振興会科学研究補助金基盤研究 (C) (課題番号 26330282) および特別研究員奨励費 (課題番号 16J10941) の補助を得て行われたものである．
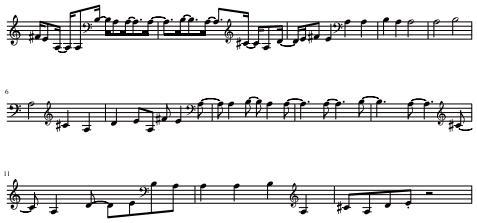


図 3: 本システムによって得られた進化曲

参考文献

[1] Hideyuki Takagi. Interactive evolutionary computation: Cooperation of computational intelligence and human kansei. In *Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems*, pp. 41–50, 1998.

[2] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, Vol. 61, pp. 85–117, 2015.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[4] Yaochu Jin. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm and Evolutionary Computation*, Vol. 1, No. 2, pp. 61–70, 2011.

[5] Taku Hasegawa, Naoki Mori, and Keinosuke Matsumoto. Genetic programming with multi-layered population structure. *Proceedings of the 2017 Annual Conference on Genetic and Evolutionary Computation*, 2017.

[6] Otto Fabius and Joost R van Amersfoort. Variational recurrent auto-encoders. *arXiv preprint arXiv:1412.6581*, 2014.

[7] Yaochu Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft computing*, Vol. 9, No. 1, pp. 3–12, 2005.

[8] JE Dennis and Virginia Torczon. Managing approximation models in optimization. *Multidisciplinary design optimization: State-of-the-art*, pp. 330–347, 1997.

[9] Erik Marchi, Fabio Vesperini, Florian Eyben, Stefano Squartini, and Björn Schuller. A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 1996–2000. IEEE, 2015.

[10] Loïc Bontemps, James McDermott, Nhien-An Le-Khac, et al. Collective anomaly detection based on long short-term memory recurrent neural networks. In *International Conference on Future Data and Security Engineering*, pp. 141–152. Springer, 2016.

Oral presentation | Organized Session | [Organized Session] OS-21

[2D2-OS-21a] [Organized Session] OS-21

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room D (4F Cattleya)

[2D2-OS-21a-01] Real-time Forecasting of Big Time-series Data

○Yasuko Matsubara[1] （1. Kumamoto University）

1:20 PM - 2:00 PM

[2D2-OS-21a-02] Anomalous popularity growth in social tagging ecosystems

○Yasuhiro Hashimoto[1], Mizuki Oka[2], Takashi Ikegami[1] （1. Univ. of Tokyo, 2. Univ. of Tsukuba）

2:00 PM - 2:20 PM

[2D2-OS-21a-03] Deep Learning Approach for Life Emerged from High-dimensional Data Recognition

○Wataru Noguchi[1], Hiroyuki Iizuka[1], Masahito Yamamoto[1] （1. Hokkaido University）

2:20 PM - 2:40 PM

[2D2-OS-21a-04] Designing a Communication Space containing Agency using Writing Record Information

○Dominique Chen[1], Hiroki Kojima[2], Mizuki Oka[3], Takashi Ikegami[2] （1. Faculty of Letters, Arts and Science, Waseda University, 2. Graduate School of Arts and Sciences, University of Tokyo, 3. Graduate School of Systems and Information Engineering, University of Tsukuba）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

1:20 PM - 2:00 PM (Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room D)

[2D2-OS-21a-01] Real-time Forecasting of Big Time-series Data

○Yasuko Matsubara[1] (1. Kumamoto University)

Keywords: time-series analysis, big data, nonlinear analysis

近年の IoTデバイスの急速な普及に伴い、それらのテ゛ハ゛イスから多様かつ大量のテ゛ータが生成され続けている。また、 Facebookや Twitterなどの巨大なソーシャルネットワーク上を大量の情報が高速に流通するようになっている。増え続ける大規模なデータ、すなわち時系列ビッグデータを高速に解析する時系列データマイニング技術は非常に重要になっている。本講演では、講演者が取り組んでいる時系列ビッグデータ解析技術、特に非線形時系列モデルに基づくリアルタイム予測技術の研究を紹介する。さらに時系列ビッグデータ解析の応用例として、具体的な事例をいくつか紹介する。

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ソーシャルタギングシステムに見られる特異的なタグの成長

Anomalous popularity growth in social tagging ecosystems

橋本康弘 *1 Yasuhiro Hashimoto　岡瑞起 *2 Mizuki Oka　池上高志 *3 Takashi Ikegami

*1*3東京大学総合文化研究科 The University of Tokyo　*2筑波大学システム情報系 University of Tsukuba

In social tagging systems, the diversity of tag vocabulary and the popularity of such tags continue to increase as they are exposed to selection pressure derived from our cognitive nature and cultural preferences. This is analogous to living ecosystems, where mutation and selection play a dominant role. Such population dynamism, which yields a scaling law, is mathematically modeled by a simple stochastic process—the Yule–Simon process, which describes how new words are introduced to the system and then grow. However, in actual web services, we have observed that a large fluctuation emerges in the popularity growth of individual tags that cannot be explained by the ordinary selection mechanism. We introduce a scaling factor to quantify the degree of the deviation in the popularity growth from the mean-field solution of the Yule–Simon process, and we discuss possible triggers of such anomalous popularity behavior.

1. はじめに

ソーシャルタギングシステム [1] とは現代のウェブサービスで用いられるアノテーションシステムの一種である．サービスを利用するユーザは，これによって自身あるいは他者が投稿した画像やテキストなどの共有リソースに「タグ」と呼ばれる短い文字列を関連付けることができる．タグとして使われる文字列はユーザの裁量に大きく任されており，その結果，タグの語彙や個々のタグの累積的な出現回数は，ユーザの文化的選好や認知的特性を反映しつつ，共に増加していく．つまり，拡大を続ける語彙空間の中で，個々のタグは人々の有限の注意を奪い合いながら，勢力を拡大していく．このタグ同士の競争的状況は，変異と淘汰が支配する生物の生態系になぞらえることができる．実際，語彙とタグの成長過程は，生物種の競争的な増加過程をモデル化した極めて単純な確率過程—Yule–Simon 過程 [2, 3]—を用いて粗く近似できることが知られている [4]．しかし，現実のウェブサービスから得られたデータを詳細に分析すると，個々のタグの成長がモデルの期待値からは大きく外れるケースが存在することが分かっている．このずれの大きさを定量化するとともに，またこのずれを生み出す社会的，あるいは認知的なメカニズムについて議論したい．

2. 準備

ここではソーシャルタギングをタグの出現シーケンスが単位時間ごとに一つずつ伸びていく過程と単純化する．そして，新たにシーケンス末尾に追加されるタグがこれまで使われたことのないタグなのか，あるいは既にシーケンス内に存在するタグなのかを確率的に決定する．前者の確率を $\alpha(t)$ と表記し，これを新規性レートと名付ける．t は離散的な時刻である．一般には t が大きくなる—サービスが成熟し，使われる語彙が大きくなる—にしたがい新規性レートは減少していく．追加するタグを既存のタグから選択する場合，どのタグを選び出すかはモデル化の重要なポイントになる．しかし，いくつかのウェブサービスでは，平均的にタグは出現回数に比例した確率で選ばれることが分かっている．これはネットワークサイエンスの分野で優先的選択 [5] として知られるメカニズムと等価である．新規性レートと優先的選択性を用いて，タグの成長（累積的な出現数の増加）は以下の時間発展式で記述できる:

$$n_i(t+1) = n_i(t) + \bar{\alpha}(t)\frac{n_i(t)}{N(t)}. \tag{1}$$

ここで $n_i(t)$ はタグ i の時刻 t における累積出現数，$N(t)$ はシーケンス長で，定義より $N(t) = t$，$\bar{\alpha}(t)$ は新規性レートの補確率 $1 - \alpha(t)$ である．十分大きな t に対して $\bar{\alpha}(t) = 1$ を仮定し，さらに t と $n_i(t)$ について連続体近似を仮定すると，近似解 $n_i(t) = t/t_i^1$ が得られる．t_i^1 はタグ i の出現数が 1 になる時刻，つまり初めて使われた時刻である．我々の興味は現実のタグがこの平均場解が示す挙動からどの程度逸脱するかにある．

3. 偏差係数

逸脱の程度を定量化するため，ここでは [6] によって導入されたアイデアを拡張して以下のように偏差係数 Φ_i を定義する．

$$\Phi_i \equiv \frac{\hat{n}t_i^1 - t_i^1}{t_i^{\hat{n}} - t_i^1}. \tag{2}$$

ここで $\hat{n}$ は累積出現数の目標値であり，分子の $\hat{n}t_i^1$ はタグ i が平均場近似により目標値に到達すると期待される時刻である．分母の $t_i^{\hat{n}}$ はタグ i が実際に目標値に到達した時刻である．つまり，Φ_i とはタグ i が目標値に到達するために実際にかかった時間とその期待値の比である．これが 1 であれば，目標値に到達するまでのタグの振る舞いは平均場解と等しく，1 より大きければ成長は速い（小さければ遅い）ことを意味する．この Φ_i をデータがカバーする時間の範囲で目標値に到達したすべてのタグについて計算し，その傾向について分析する．

4. 実験結果

ここでは [7] によって収集された写真共有サービス Flickr のデータを用いる．図 1 に示した図はそれぞれ Φ_i と $t_i^{\hat{n}}$ の関係（左），および Φ_i と最終的な累積出現数の関係（右）であり，

連絡先: 橋本康弘: hashi@sacral.c.u-tokyo.ac.jp

色はタグの濃度を表している．目標値の決め方としては，タグが作られた初期の成長過程を見るために $\hat{n} = 10$，さらに，ある程度十分な出現数を獲得するまでの成長過程を見るために $\hat{n} = 10^3$ を用いた．Yule–Simon 過程の結果と比較すると，実データには偏差係数が平均場解 $\Phi_i = 1$ の周りに集まる集団と，偏差係数の非常に大きな集団が存在することが見て取れる．この偏差係数の大きな集団は $\hat{n} = 10^3$ ではほぼ消失し，この集団が成長の初期段階に固有の特徴であることを示している．また，この集団を除外してもなお平均場解の周りの裾野は広く，平均場解から大きく逸脱するタグが多く存在することを示している．さらに，最終的な出現数の大きなタグは偏差係数の大きなタグではなく，平均場解の近くに集まっていることがわかる．これは Yule–Simon 過程にも共通する特徴である．

5. 考察

仮に，優先的選択ではなく，$n_i(t+1) = n_i(t) + \bar{\alpha}c$（$c$ は定数）という固定レートの成長を仮定すると，成長の初期段階に見られる偏差係数の大きな集団をよく説明できる．つまり，成長の初期段階において，タグ集団は優先的選択に従う集団—競争的状況で使用されるタグ—と，逆に固定レートで成長する集団—競争相手の多寡に依存せず一定時刻の中で集中的に使用されるタグ—が存在すると考えられる．また，平均場解の周りの裾野が広くなる要因としては，最近使われたタグが優先的に選択される，一方でタグは時間的に忘却される，という人間の認知的な特性が考えられる．人間の注意の時間減衰やメモリー効果について扱った研究は近年少しずつ増えてきているが [8, 9]，ここでのメカニズムについては具体的には明らかではない．また，累積出現数の大きなタグが平均場解の周りに集まる点についても，実データと Yule–Simon 過程で理を一つにしているかについては検討が必要である．また，本分析では主にタグの選択を問題にしているが，それはタグの生成メカニズムとも不可分ではないはずである．新しいタグが新しい生態的ニッチを開拓し，それによって既存のタグの意味合いが変化し，タグの選択に再帰的に影響を与えることは十分に考えられる．このような問題意識を共有して，新規性の発生と選択（採用，伝播）という広い視点に基づいて議論したい．

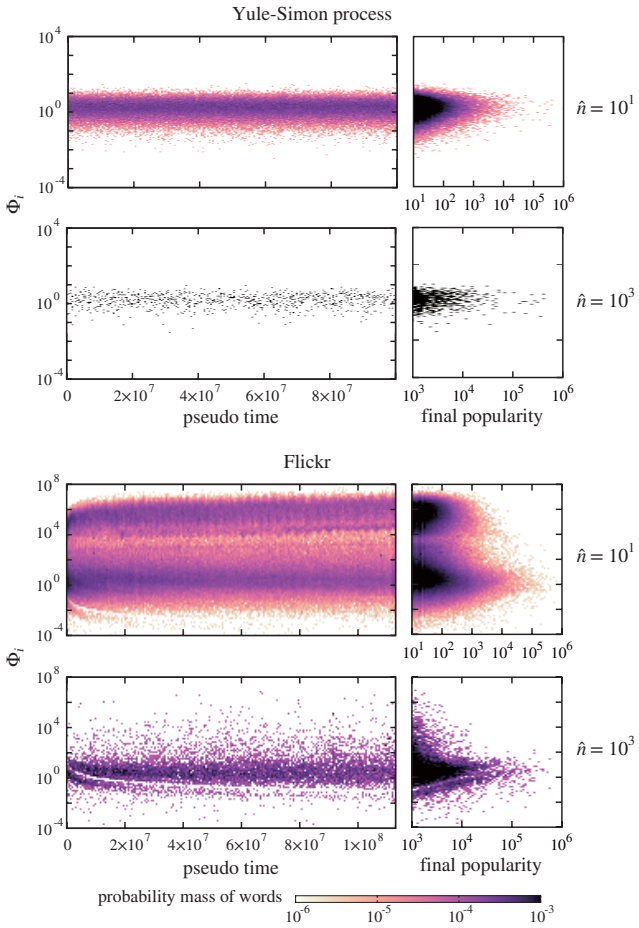


図 1: 縦軸は偏差係数，横軸は目標値に到達した時刻（左）と最終的な累積出現数（右）．縦軸はともに対数分割（左はビン数 80，右はビン数 100），横軸は左は線形分割（ビン数 200）で右は対数分割（ビン数 100）．分割したセルの色はそこに入るタグの数．全体的に Yule–Simon 過程に比べて実データの方が偏差係数の値が 1 から大きく外れている．

参考文献

[1] Daniel H. Pink. Folksonomy. `http://www.nytimes.com/2005/12/11/magazine/folksonomy.html`, 2005-12-11. Last retrived 2017-01-24.

[2] Herbert A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3–4):425–440, 1955.

[3] G. Udny Yule. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f. r. s. *Phil. Trans. Roy. Soc. B*, 213:21–87, 1925.

[4] Ciro Cattuto, Vittorio Loreto, and Luciano Pietronero. Semiotic dynamics and collaborative tagging. *Proc. Natl. Acad. Sci. USA*, 104(5):1461–1464, 2007.

[5] Albert-lászló Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(1):173–187, 1999.

[6] Yasuhiro Hashimoto. Growth fluctuation in preferential attachment dynamics. *Phys. Rev. E*, 93(4):042130, 2016.

[7] Olaf Görlitz, Sergej Sizov, and Steffen Staab. Pints: Peer-to-peer infrastructure for tagging systems. In *Proceedings of the 7th International Conference on Peer-to-Peer Systems*, IPTPS'08, pages 19–19, Berkeley, CA, USA, 2008. USENIX Association.

[8] Pietro Della Briotta Parolo, Raj Kumar Pan, Rumi Ghosh, Bernardo A. Huberman, Kimmo Kaski, and Fortunatom Santo. Attention decay in science. *J. Informetr.*, 9(4):734–745, Sep 2015.

[9] Ciro Cattuto, Vittorio Loreto, and Vito Domenico Pietro Servedio. A yule-simon process with memory. *Europhys. Lett.*, 76(2):208–214, 2006.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

高次元データの認識から生まれる生命性についての深層学習アプローチ

Deep Learning Approach for Life Emerged from High-dimensional Data Recognition

野口渉[*1] 飯塚博幸[*1] 山本雅人[*1]
Wataru Noguchi Hiroyuki Iizuka Masahito Yamamoto

[*1] 北海道大学大学院 情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

Animal receives high-dimensional and complex raw sensory information. Deep learning can recognize such complex sensory information. We studied a deep learning model called hierarchical recurrent neural network (HRNN) that develops spatial recognition through visuomotor integration learning. In a simulation experiment, the HRNN developed the cognitive map, which is an objective map-like internal model, through only subjective visuomotor experiences. Furthermore, the HRNN also developed spatial recognition through visuomotor sequences by a human subject. These results imply that deep learning model can be used to study real animals' cognition.

1. はじめに

従来の人工生命研究においては，ミニマルなモデルを用いることで生命を理解しようとするアプローチが主であった．実際，ミニマルモデルから生まれる複雑なシステムの解析は多くの知見をもたらしてきた．しかし，現実のデータは高次元で複雑な構造をもっており，ミニマルなモデルでは，データそのものの複雑性を損なうことが問題となる．例えば，ミニマルなモデルにおいては，視覚情報が距離や物体の位置として抽象化・単純化されるなど，理解したい興味の対象に関係すると思われる情報のみが選択される．しかし，視覚情報に含まれる情報は距離や物体の位置だけでなく，物体の色や形，物体の並びなど無数に存在するはずである．しかも，これらの情報はそれ自体が自明に提示されるわけではなく，感覚情報として得られるのはあくまで複雑な高次元な生のデータである．現実の生物は，視覚のような高次元の生のデータを認識しているのであり，生命の認知モデルとしてはそのような生のデータを認識するモデルを用いるアプローチが必要だと考えられる [Ikegami 12].

本論文では，従来のミニマルモデルでは扱われ難かった高次元のデータの認識が生み出す生命性を考える．例として，我々が行った視覚と運動の予測学習により空間の認知を構成する深層学習モデルについて述べる．高次元の生のデータを扱うということは，コンピュータの中で閉じた生命を考えるだけでなく，実世界と対峙する人工生命モデルを考えることを意味する．そのため，シミュレーション内での学習に加え，実際のヒトを用いて取得した視覚と運動データを学習させることで，深層学習モデルによる現実の生物の認知の理解の可能性について議論する．

2. 深層学習を用いたアプローチ

2.1 深層学習

深層学習は多層で巨大なニューラルネットワークを用いることで，高次元のデータの認識・生成を可能にした手法である [LeCun 15]. 深層学習は，設計された特徴抽出手法を用いることなく生のデータの学習を通して概念の認識を可能とする．例えばオートエンコーダは画像などの高次元データを低次元に圧縮したのち再構築（元のデータの生成）学習を行うモデルであり，

連絡先：野口渉 noguchi@complex.ist.hokudai.ac.jp

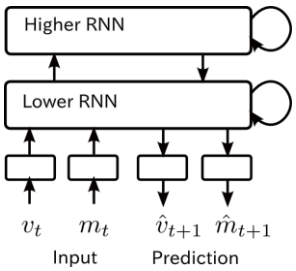


図 1 階層型 RNN

多層のオートエンコーダはその中間層において高次元のデータに含まれる抽象的な概念を表現する．生のデータの学習によりボトムアップ的に構築された概念は常に外部のデータに結びついており，すなわち外部の環境をよく認識できているということを意味する．これは深層学習と強化学習を組み合わせたモデル [Mnih 15] が環境への高い適応能力を示すことからも支持される． つまり，深層学習は高次元で複雑な感覚情報そのものと結びついた認識を構築可能であるという点で，より実際の生物に近い認識を実現するモデルとなりうると考えらえる．

2.2 視覚と運動の統合により生まれる認知地図

我々は深層学習を用いて，特に空間の移動に伴う視覚と運動の経験から生まれる認知についての研究を行っている．視覚情報を事前の特徴抽出することなく直接深層学習モデルに入力することで，高次元の視覚情報の学習を通して生まれる空間認知について考えることができる．また，視覚とともに対応する運動データも学習させることにより，実際の生物と同様な多感覚統合のプロセスもモデル化している．

学習モデルは主に RNN (Recurrent Neural Network）からなり，さらに，RNN が階層的に結合した構造をもっている（図1）．この階層型 RNN モデルは空間の移動に伴う視覚と運動感覚の予測学習によって訓練される．この階層構造においては，異なる階層は視覚・運動経験に含まれる異なるレベルの概念を認識する．下位の RNN 層は視覚・運動入力を直接受け取り視覚・運動の予測を出力するため，比較的速いタイムスケールを扱う．一方，上位の RNN 層は入出力と直接結合しておらず，比較的遅いタイムスケールで変化する情報を認識可能である．

上記の階層型 RNN を用いて，まずシミュレーション上での視覚・運動の予測学習を行った [Noguchi 17]. 図 2 (a)に示すシミ

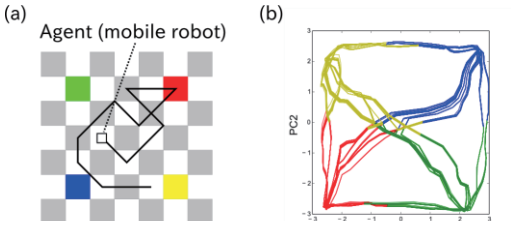


図 2　シミュレーション実験．(a) 移動ロボットシミュレーション．(b) 上位 RNN 層に獲得された内部状態

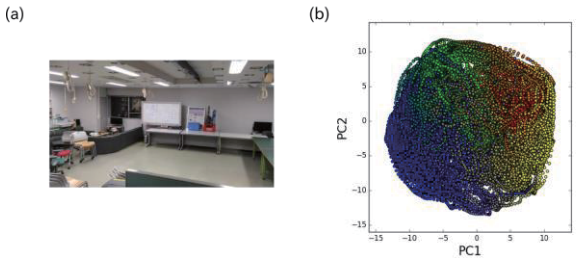


図 3　ヒトのデータを用いた実験．(a) 実験環境．(b)上位 RNN 層に獲得された内部状態．

ュレーション上に構築された仮想空間内の移動ロボットエージェントの視覚・運動系列を予測学習する．図 2 (b)は，予測学習の結果獲得された上位 RNN 層の内部状態であり，主成分分析を用いて次元削減し可視化されている．内部状態は最も近いランドマークの色で色づけられており，エージェントの移動空間の座標と対応した内部状態に組織化していることがわかる．ここで，学習モデルには空間の座標の情報を直接与えておらず，内部状態はあくまで主観的な感覚として得られる視覚・運動をもとに構築されたものである．つまり，客観的な空間座標に対応して内部状態が組織化したが，内部の認識は視覚・運動そのものと結び付けられている．また，視覚の外部入力なく運動入力のみが与えられた場合（いわば，目を瞑って歩いた場合）に出力される視覚をみることで内部に構築された空間の認識を評価できる．そこで，学習済みモデルに学習中に通ったことのない未学習領域において目を瞑って歩いた場合の予測を行わせたところ，たどり着く場所で見えるはずの視覚を正しく予測することが確認された．つまり，階層型 RNN は主観的には経験したことのない領域を補間して空間を認識できており，獲得した内部モデルは Tolman の言う認知地図 [Tolman 48] のような空間の表象としてのモデルであるとみなせる．この結果は階層型 RNN モデルが主観的な感覚である視覚と運動ともとに内部モデルを構築しつつも，客観的な概念を認識可能なモデルであることを示している．

次に我々は，実環境のデータとしてヒトの視覚・運動データを用いて同階層型 RNN モデルの予測学習を行った．ただし，より複雑な視覚の認識と生成を可能にするために Convolutional Neural Network を用いている。被験者となるヒトには部屋内の 4 メートル四方程度の領域を自由に歩き回ってもらう (図 3 (a))．このとき，被験者の額部に取り付けたカメラにより取得した画像を視覚，左右の側頭部に取り付けた加速度センサの取得する値を運動データとして用いる．左右それぞれの加速度センサは直線加速度と回転加速度をそれぞれ 3 次元取得するが，これはヒトの耳石器官と三半規管による感覚に対応する．図 3 (b)は，ヒトの視覚・運動予測学習後に組織化した上位 RNN 層の内部状態である．図を見ると，移動空間の座標に対応付けた色が内部状態空間中でも分かれている．つまり，内部状態の空間が外部空間の座標と対応するように組織化している．また，シミュレーション実験と同様に運動入力のみからの予測を行った場合も，運動に応じて変化する視覚映像を出力（イメージ）することが可能であることが確認された．現時点では，上記シミュレーション実験で確認された未学習領域に対する汎化性能については未検証であるが、階層型 RNN モデルは現実の複雑な視覚に対しても空間構造を認識可能であると示唆される．

3. 考察

本論文では，高次元で複雑な感覚情報を通した生物の認知のモデルとしての深層学習モデルの応用可能性について述べ，その一例として，視覚・運動予測学習により空間の認識を獲得する階層型 RNN モデルについて述べた．階層型 RNN の獲得した内部モデルは各状態が視覚・運動そのものに対応づいているという点で従来のミニマルなモデルとは異なる．感覚情報の単純化されたモデルにおいては，感覚情報の本来含む複雑さが排除されており，内部モデルの表現する情報も単純にならざるを得ない．対して，高次元な視覚をそのまま受け取るモデルでは，内部モデルは視覚の複雑さをよく含んだものになりうる．この内部モデル，つまり内部の認識の複雑さは，より多様な情報を認識と複雑な環境への適応能力を実現するものであり，より複雑に振る舞う生命的なモデルに必要であると考えられる．

さらに階層型 RNN モデルは，シミュレーション実験に加え実際のヒトの視覚・運動データからも空間構造を認識可能であることを示した．このように現実のデータと深層学習モデルを結びつけるアプローチは，現実の生物における認知の発達過程の研究に応用可能であると考えられる．

また，階層型 RNN は運動入力に対応して変化する視覚をイメージする能力を獲得した．つまり，モデル内部において視覚と運動の情報を扱うネットワークは分離されておらず，密に結合つまり統合されていると考えられる．また，モデルの内部状態は高次元の視覚情報と結びついているため，モデル内部では高次元の情報が統合されているということになる．今回のモデルでは運動感覚データはごく低次元のものであったが，触覚や聴覚などのより高次元な情報を用いたモデル化も可能であり，その場合，より高次元の情報が統合されることになりうる．統合情報理論によると，このような高次元で多様な感覚が統合されるシステムは意識の発生の可能性をもつ．すなわち，本論文で紹介した深層学習による高次元な感覚の統合モデルは意識のモデルとして発展可能な可能性もあると考えられる．

参考文献

[Ikegami 12] Ikegami, T. and Oka, M.: マッシブデータフローの科学を目指して: 人と環境の間を流れる高次元のデータフローを巡る生成と解析について，人工知能学会誌, Vol. 27, pp. 389-395 (2012)

[LeCun 15] LeCun, Y., Bengio, Y. and Hinton, G.: Deep Learning, Nature, Vol. 521, pp. 436–444 (2015)

[Mnih 15] Mnih, V. et al..: Human-level control through deep reinforcement learning. Nature, Vol. 518, pp. 529-533 (2015)

[Noguchi 17] Noguchi, W., Iizuka, H. and Yamamoto, M.: Cognitive map self-organization from subjective visuomotor experiences in a hierarchical recurrent neural network. Adaptive Behavior, Vol. 25, pp. 129-146, (2017)

[Tolman 48] Tolman, E. C.: Cognitive maps in rats and men., Psychological review, Vol. 55, pp. 189-208 (1948)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

執筆記録情報を用いた行為主体性を持つコミュニケーション場のデザイン

Designing a Communication Space containing Agency using Writing Record Information

チェン・ドミニク[*1] 小島大樹[*2] 岡瑞起[*3] 池上高志[*2]
Dominique Chen Hiroki Kojima Mizuki Oka Takashi Ikegami

[*1] 早稲田大学文学学術院 [*2] 東京大学総合文化研究科
[*3] 筑波大学大学院システム情報工学研究科
Faculty of Letters, Arts and Science, Waseda University Graduate School of Arts and Sciences, University of Tokyo Graduate School of Systems and Information Engineering, University of Tsukuba

Static text communications in online spaces lack the rich embodied information we experience in face-to-face communication. In this research, we aim at charging liveliness into digital text communication by developing TypeTrace, a Web software that can record and play back writing processes and visualize the micro fluctuations of time contained within. We show the design and result of a preliminary experiment we conducted and discuss further research directions.

1. はじめに

ウェブ上の Social Networking Service（SNS）やメールなどのタイピングが介在するコミュニケーションにおいては，執筆過程が終了した，静的な文章を交換することが前提となってきた．他方で，実世界における対面型（Face to face, F2F）のオーラルな対話においては，話者同士は相互の発話プロセスをリアルタイムに観察しながらコミュニケーションを行う．タイピングされた文章とオーラルな発話では，環境も使用する媒体も，前提となる条件が大きく異なる．F2F と比較した時，SNS では特に発話行為者のイントネーション，抑揚やリズム，テンポといった，非合理的だが情動を喚起する情報が捨象される．結果，テキストの受け手は，発話行為者の行為主体性（agency）を知覚することが困難になるのではないか．

二者間のリアルタイムでの相互作用のみを取り出した実験として，知覚交差実験を用いた研究をこれまでに行っている．これは一次元の仮想空間上で，２人の被験者が互いに相手を探りあてるものであり，ミニマルな実験設定でありながら相手の存在感を強く感じることがあり，その際特徴的な二者間のダイナミクスが生じることを明らかにした [Kojima 2017] .

本研究では，以上の知見と仮説に基づき，通常のウェブ上でのテキストコミュニケーションでは捨象されてしまうプロセスを可視化することによって，話者同士の互いの行為主体性の知覚が高まる可能性を検討する．そこで，発話プロセスが可視化されることによって，コミュニケーション内容がどのように変容するのかということを明らかにすることを目的として，オンライン上のコミュニケーションの予備実験を行った．

2. 実験

本研究では，タイプトレース（Type Trace）という，キーボードを使った文章入力のプロセスを記録し，再生するソフトウェアを用いて実験を行った．

2.1 Web 版 Type Trace

私たちは普段，執筆の最終結果のみとしての文章を読んでいるが，タイプトレースはどのように文章が書かれたのか，その始まりから終わりまで，誤字脱字の推敲やアイドリング時間まで含めて，記録し，それを動画のように再生する．この記録されるプロセスのなかで，各形態素の入力開始から確定までにかかった時間に応じて，その形態素の表示フォント級数が上下するというエフェクトが動くが，このことによってどこで淀みなく書き進んでいて，どこで停滞しているのか，など，執筆者の思考のリズムやテンポを可視化することを試みている．

このように「デジタルな生原稿」というコンセプトで 2006 年にメディアアート作品として制作し，これまで国内外で展示されてきた．2007 年には小説家の舞城王太郎がタイプトレースを用いて新作小説を執筆したプロセスの上映を約 3 ヶ月間，美術館で展示した[Dividual 2007]．また，後にその時のデータを用いてクリエイティブ・ライターの執筆スタイルと推敲過程の推移の分析を行った[工藤 2015].

タイプトレースは当初，Mac OS 専用ソフトウェアのダウンロード公開を行っていたが，今回の実験に際して PC とスマートフォンの Chrome ブラウザ上で記録と再生の行える JavaScript 版のバージョンを開発した（図 1）．

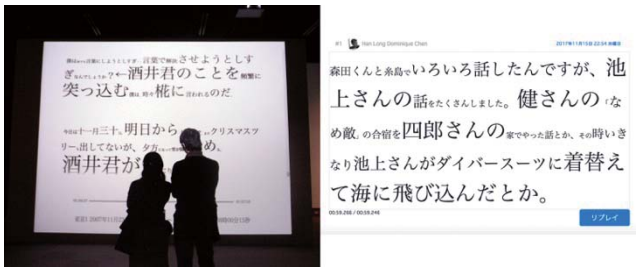


図 1. 展示版タイプトレース（左）とウェブ版タイプトレース（右）

2.2 プロセス可視化と生命性

Bateson は「すべての生物をつなげるパターンとは何か？」という問いを考察する中で，生命個体の発生過程がその形態に表出する現象を prochronism と呼び，他者はその観察を通して当該個体が生命であることを知ると表現した[Bateson 1979].

タイプトレースは文章の発生過程という，デジタルデータの prochronism を観察者に提供することで，テキストの作者の行為主体性と同時にテキストという人工物の生命性をも擬似的に知覚させると言える[チェン 2013].

そもそも文字という媒体は，身体的な発話に伴う生命的な時間の流れを固定することで，ランダムアクセスや複雑な論理記

連絡先：チェン・ドミニク，早稲田大学，〒162-8644 東京都新宿区戸山1－24－1，dominique@waseda.jp

述を可能にするが，タイプトレースは文字を再び実世界の会話に近づける作用を持つのかもしれない．生命的なコミュニケーションを受容するということは，相手の発話の内容だけではなく，リズムやテンポといった時間的な形質にも直面するということを意味する．

2.3 プロセス可視化の有無による時間間隔の差

Varela は現在性の知覚において，0.1 秒のスケールでは，神経系の知覚が起こり，1 秒のスケールではより大きな認知への統合を準備され，そして 10 秒のスケールでは状況に対する記述的・物語的な評価を行なわれるというように，3 つの時制のスケールが再帰的に作用していると論じている[Varela 1999]．テキストの執筆プロセスを再生することによって擬似的な現在性を生み出すタイプトレースの効果を検証する上で，この分類は有用であると考えられる．既に終了している執筆プロセスの再生によって，相手の発話が擬似的に現在と同期されることで，相手の行為主体性を知覚しやすくなるかもしれない．

そこで予備実験では，執筆のプロセスがタイピングのテンポを表すフォント級数の変化と共に可視化されるタイプトレース効果（TT）の有無に応じて，キーボードの打鍵のように微細なスケール（0.1 秒），入力する形態素を確定するためのエンターキーが押下されたり単語が発生するスケール（1 秒），発話間の時間間隔のスケール（10 秒）の各階層を，定量的な変化が生まれるかどうかを検証した．

具体的には，互いに知己ではない日本語話者の被験者のペアを 10 組用意し，それぞれのペアで 1. TT 効果有り，2. TT 効果無し，でのチャットを交互に行ってもらった．なお，被験者同士は全員初対面で，筆者らが用意した会話の題目リストの中から一つを選んでもらい，オンラインでタイプトレースにログインした上で，TT 効果有り無しそれぞれ 20 分のセッションでコミュニケーションしてもらった．

3. 結果

10 組のオンラインチャットの結果として集まったデータから．コミュニケーションにおける時間幅のゆらぎが生じているかどうかを階層的に検証した．具体的には，セッション毎での投稿，投稿毎の単語とキー打鍵の時間間隔の値に有意な差分が生じているかを調べた．

・eventCount：1投稿内で key を押した数
・duration：1投稿を投稿し終わるまでの時間（秒）
・key_interval_mean：1投稿内で key を押す間隔の平均（秒）
・key_interval_std：1投稿内で key を押す間隔の標準偏差
・enter_count：1投稿内で enter を押した回数
・enter_interval_mean：1投稿内で enter を押す間隔の平均（秒）
・enter_interval_std：1投稿内で enter を押す間隔の標準偏差（秒）
・post_interval：投稿間の間隔（秒）

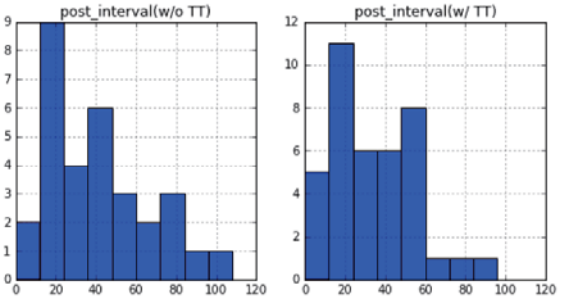


図 2. 横軸が投稿間の時間(s)で，縦軸がそのヒストグラム

図1では，post_interval（post 間の秒）を比較している．タイプトレース効果の有り無しのケースで，それぞれの差を比較して t 検定をかけたが，統計上有意な差は確認されなかった．

4. 議論と展望

今回の予備実験では統計上の有意差が確認できなかったが，定量的な執筆記録データの解析に先鞭をつけることができた．今後も引き続き改良した実験を重ねることを検討する．

4.1 追加実験の設定

今回の実験の設計上の問題点として，被験者には説明や補助が行えず，事前説明がなかったため，一部では新奇なインタフェースの操作で混乱しているケースが見られたこと，初対面同士ではリモートとはいえ緊張してしまい，会話が長続きしなかったことなどが確認できたので，今後の改善点とし，会場では最新のデータを紹介したい．

また，今回は一つの投稿が 100 文字程度の短文でのチャット形式だったが，メールのように非同期で，チャットよりも長い文章であれば，相手の書いたプロセスをしっかり読んでから何かを書くというように，ある程度の長さのある文章の応答であれば執筆プロセスの可視化によって有意な差が生まれるかもしれない．

また，日本語以外の言語でのデータ収集も進めることで，日本語データとの比較によって，新たな特徴が見つかるかもしれない．英語話者の被験者を募るために Amazon Mechanical Turk の使用を検討している．

4.2 コミュニケーションモデルの変化と主体行為性

プロセスが再生されるインタフェースを通して，通常のテキスト・コミュニケーションでは捨象されてしまう価値が生じる可能性を検証する上で，水谷による共話概念[水谷 1993]と木村による共在感覚[木村 2005]の議論が参考になる．

共話とは，ターンテイクが明確な対話と比較して，発話行為者同士の境界が曖昧となるような話法である．水谷は実際に外国人による日本語学習を観察することで，日本語の自然な会話にはあいづち（back-channel）が英語と中国語と比較して 3 倍以上多いという定量事実を挙げながら，文章の後半を省略して相手に文の完成を委ね合ったり，主語を共有したりしながら，いわば共同作業として行われるコミュニケーションを共話と定義した（図 3）．

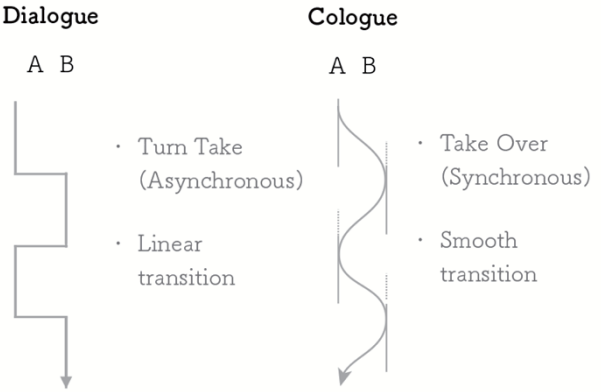


図 3. 対話と共話における相互行為の差異

タイプトレースにはリアルタイムに相手の執筆にあいづちを打つという機能は備わっていないが，相手の執筆プロセスを注視することを通して，非同期で内声的なあいづちや首肯を促す効果があるかもしれない．

木村はアフリカの少数民族における言語現象の実地調査を通して，リアルタイムな応答可能性のチャネルが開かれているコミュニケーション場には話者同士に共在感覚が発生しており，それは言語や風習などのインタフェースによって異なると論じ，ITC においても共在感覚は設計可能であると論じている．執筆が終わったテキストでも，その生命的なプロセスの再現を通して擬似的に観察者の現在性の知覚と同期するのであれば，デジタルな共在感覚を生成するためのヒントを得られるかもしれない．

以上のように，共話と共在感覚の概念を参照しながら立てた仮説は，定量的に検証する必要がある．

4.3 共話性の測定：会話における情報の流れ

共話に基づいた仮説とは，古典的な対話モデルにおいて生じるA，B 間の会話では，ターンテイクおよび主体行為性の区切りはある程度厳格となるが，共話的な構造の会話においてはそれが曖昧になる，というものだ．

仮に TT 効果を介したテキスト・コミュニケーションがこのように主体行為性の相互的な認知に影響を与えるのだとすれば，会話システムを構成するサブシステム（投稿）間での情報の流れの変化として表れるかもしれない．Transfer Entropy [Schreiber 2000]は，異なる種類の生体データを比較した時，どちらからもう片方へ，より大きな情報の流れがあるのかということを計測することに優れている．

そこで，TT 効果ありと TT 効果なしのそれぞれのシナリオで，ペアを組んだ二人のそれぞれの投稿の間で Transfer Entropy を測定することで，各ケースにおける情報の流れのパターンの違いが生じるかどうかを検証できる．そこから，各シナリオの会話における非対称性の比較を行うことで，主体行為性の認知への影響を測ることができる．

4.4 well-being の構成への影響

行為主体性の認知から共話と共在感覚の概念を参照することで，プロセスを再生するインタフェースがコミュニケーションの当事者の心理に及ぼす影響についても検討することができる．

well- being とは，心の充足を構成する複数の要素の内実を問う考え方である．インタフェースやアプリケーション，アルゴリズムの作動の仕方がユーザーの心理的（理論的），主観的な well-being の構成に影響することを前提とする Positive Computing という領域がある．本研究の一義的な対象はテキストコミュニケーションにおける主体行為性の認知の変容だが，発話内容の感情分析を行うことで当事者の well- being の変化も検討できる．

水谷は共話が話者同士に「全部言わなくてもわかってくれる相手がいることは、心を暖かくする」と述べており，共話的な話し方がポジティブな感情を与えると論じている．この点は well-being の測定方法として共話性を検証する意義を示唆している．

木村は共在感覚を価値中立的なものとして論じているが，応答可能性が担保されているコミュニケーション範囲の大小は人間の心理に少なくとも無関係ではないことが予想されるとすれば，共在感覚の有無の測定も well-being の増減と相関するかもしれない．

以上のことは，現代において、同じ場所に集まっている人間同士がスマートフォン上の SNS に興じる余り、F2F（face-to- face）の会話が疎かになることが当該の人間関係に悪影響を及ぼしたり，SNS において常に相手と接続していなければならないという強迫観念が生じる問題と通底している［Turkle 2015]．Twitter や Facebook といった社会インフラ化したインタフェースを所与のものとして受容するのではなく，ユーザーの心理的な自律性と well-being に考慮する再定義可能性の議論が必要である．

5. 謝辞

この研究は，新学術領域研究「認知的インタラクションデザイン学：意思疎通のモデル論的理解と人工物設計への応用」からの補助を受けて行っています（課題番号：17H05855「Agency を持つ仮想空間デザインと，そのなかでのコミュニケーション実験」，研究代表者：池上高志）．また，主体行為性と well-being の概念を接続する議論は，JST・RISTEX「人と情報のエコシステム」採択の研究開発プロジェクト「日本的 Wellbeing を促進する情報技術のためのガイドラインの策定と普及」（研究代表者：安藤英由樹）と研究調査プロジェクト「人間と情報技術の共進化を目指す共創コミュニティALife Lab. の構築」（研究代表者：岡瑞起）の活動を通して行われました．

参考文献

[Dividual 2007] dividual(遠藤拓己，ドミニク・チェン)：タイプトレース道――舞城王太郎之巻，東京都写真美術館，文学の触覚展，2007.12.15~2008.2.17.

[チェン 2013] ドミニク・チェン: インターネットを生命化する―プロクロニズム の思想と実践，青土社，2013.

[工藤 2015] 工藤 彰・岡田 猛・ドミニク チェン: リアルタイムの創作情報に基づいた作家の執筆スタイルと推敲過程の分析，認知科学，22(4)，573-590，2015.

[木村 2005] 木村大治: 共在感覚―アフリカの二つの社会における言語的相互行為から，京都大学学術出版会，2005.

[Varela 1999] Francisco Varela: The Specious Present: A Neurophenomenology of Time Consciousness，Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science，Chapter 9，pp.266-329，1999.

[水谷 1993] 水谷信子：「共話」から「対話」へ，『日本語学』12巻 4 号，明治書院，4-10，1993

[Kojima 2017] Kojima Hiroki, Froese Tom, Oka Mizuki, Iizuka Hiroyuki, Ikegami Takashi: A Sensorimotor Signature of the Transition to Conscious Social Perception: Co-regulation of Active and Passive Touch, Frontiers in Psychology vol.8, 2017

[Schreiber 2000] Thomas Schreiber: Measuring Information Transfer, Phys. Rev. Lett. 85, 461, 2000

[Calvo & Peters 2015] Rafael A. Calvo, Dorian Peters: Positive Computing: Technology for Wellbeing and Human Potential, MIT Press, 2014

[Turkle 2015] Sherry Turkle: Reclaiming Conversation: The Power of Talk in a Digital Age, Penguin Books, 2015

Oral presentation | Organized Session | [Organized Session] OS-21

[2D3-OS-21b] [Organized Session] OS-21

Wed. Jun 6, 2018 3:20 PM - 4:40 PM Room D (4F Cattleya)

[2D3-OS-21b-01] Bursting Behavior of Social Insects through Agent-based Simulation

○Doi Itsuki[1], Takashi Ikegami[1] （1. the University of Tokyo）

3:20 PM - 3:40 PM

[2D3-OS-21b-02] Generation and Selection Mechanism on Social Tagging Systems

○Yoshimasa Nishikawa[1], Mizuki Oka[1], Yasuhiro Hashimoto[2], Takashi Ikegami[2] （1. University of Tsukuba, 2. The University of Tokyo）

3:40 PM - 4:00 PM

[2D3-OS-21b-03] Modelling and experimental approaches to continuous dynamics of social interactions

○Reiji Suzuki[1], Takaya Arita[1] （1. Nagoya University）

4:00 PM - 4:20 PM

[2D3-OS-21b-04] Autonomous Regulation of Self and Non-Self in Embodied Neural Networks.

○Atsushi Masumori[1], Norihiro Maruyama[1], Lana Sinapayen[1], Takashi Ikegami[1] （1. The University of Tokyo）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Agent based model を用いたシミュレーションから見る社会性昆虫のバースト現象

Bursting Behavior of Social Insects through Agent-based Simulation

土井樹 *1　　池上 高志 *2
Itsuki DOI　　Takashi IKEGAMI

*1東京大学総合文化研究科
Graduate School of Arts and Sciences, University of Tokyo

Global patterns and dynamics observed in a social system often reflect and amplify the underlying characteristics of its elements. Although there are many studies linking global and the local social dynamics and patterns, the relation between the organized global pattern and the behaviors of the individual constituents has not yet been fully understood. In our previous study, using a set of individualized behavioral records of a honey bee (*Apis mellifera*), we found the hive often exhibited collective bursting behaviors and that some of these bees cause bursts multiple times. Also, we found that some of these bees cause bursts multiple times, suggesting that the individuality has hardly ignorable effects on the global organization of the whole social collective behavior. In this study, in order to understand what causes the bursts, we use an evolutionary agent based model approach with neural networks: each agent is controlled by an artificial neural network which neural networks are evolved by a genetic algorithm with sensors detecting the environmental information.

1. はじめに

1.1 先行研究

社会性昆虫はその社会全体のグローバルなダイナミクス、それを構成する個体の振る舞い、またその個体間の関係性といった観点からこれまで多くの研究が行われてきた。特に division of labor に関する研究は特に活発に研究される分野であり、多くの研究から労働分担が何かしらの中心制御系によって決定されるわけではなく、個体間のインタラクションや外部環境からの刺激によって決定される事を示唆する結果が得られている [Robinson 1992][Camazine 2003]。さらに 2013 年に Mersch らが数百のアリの人口コロニーを作成し全個体トラッキングを行った結果、個体/個体間行動の空間的分布が労働分担と関わっていることを示唆する結果を得ている [Mersch 2013]。このように、「集団の中での個体/個体間の運動」に着目した研究は多く存在するが、グローバルな集団運動とそれを構成する個々の構成要素の個体行動がどういったふうに関係しあっているかは完全には理解されていない。

我々はこれまで、グローバルな集団運動とそれを構成する個々の構成要素の個体行動がどういったふうに関係しあっているかについて探るために、セイヨウミツバチ *Apis mellifera* を用い、人工的に作った約 1000 頭の個体からなるコロニーを作り、全個体トラッキングをしその解析を行ってきた [Doi 2017]。この研究では、セイヨウミツバチのコロニー内で起こるバーストと呼ばれる、コロニー内の個体の多くのアクティビティーが同期する現象に着目し、このバーストにはコロニー外部からの刺激によるもの（外因的バースト）とそうではないバースト（内因的バースト）が存在し、これらは巣全体のアクティビティの時間発展からも区別できることが分かった。特にその立ち上がりの仕方が異なるという結果は、外因的バーストは、多くの個体が外からの刺激によって同時にアクティビティーが高まるのに対し、内因的バーストは、（アクティビティーの高い）数個体の行動が段々と連鎖する、つまり、個体の行動をコロニー全体が増幅していることを示唆する。さらに、内因的バーストが起こる直前に活動が活発になる個体群は、異なるバーストにおいても同様にバーストが起こる直前に活動が活発になる傾向があり、これらの個体群にはいくつかの特徴があることを明らかにした。

1.2 本研究の目的

上記の結果を踏まえた上で、本研究ではなぜこのようなバースト現象が起こるのか、モデルベースに明らかにすることを目標とする。また本誌では自発的に生じる内因的バーストにフォーカスした結果を述べる。

バースト現象はセイヨウミツバチのみならず現実世界あるいはインターネットなど多くの場所で観測されてきており、バースト現象を説明するモデルはこれまでにも幾つか提案されてきている。[Sornette 2008] らは、Hawkes Process[Hawkes 1971] という確率点過程を用いたものを用いて、youtube の動画視聴者数の time evolution のモデル化を行った [Crane 2008]。また Ikeda らは SIR model を拡張しさらに agent のキャラクターも考慮したモデルにより twitter におけるデマの拡散を説明した [池田 2016] が、セイヨウミツバチのような agent 各々に individualty がある系での空間方向のバーストに対するモデルアプローチは少ない。

本誌では、特に空間方向に agent が運動することに重きを置き、ニューラルネットを搭載した Agent based model でのシミュレーションを行った。

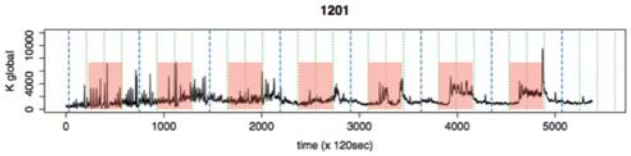


図 1: セイヨウミツバチアクティビティーのの時間発展の例。オレンジで囲まれた部分は日中を表している。

連絡先: 土井樹 idoi@sacral.c.u-tokyo.ac.jp

2. 手法・解析

2.1 Agent

この Agent はセンサー細胞 (7)+中間層 (4)+コンテキスト層 (2)+出力層 (2) で構成されるリカレントニューラルネットを持つ。中間層内部での結合はなく、コ ンテキストを介して相互作用し出力層は中間層からの結合で制御される。また各個体は一定時間で蒸発するフェロモンを自らの周りに放出しながら移動する。ニューラルネットワークの重みは、単独の個体を数種類の環境内で自律的に動かし、一定時間内にセンサーで検知できる他個体のフェロモン量が多くなるように遺伝的アルゴリズムにより進化させ、数種類の重み行列 (Gene) を生成した。

また agent は速度が可変であり、またここでは活発に動き回るエージェントを active state にいる agent とみなし、それは以下にみるように、S 字曲線で inactive から active へと推移すると仮定した。活発に動き回る度合いは、したがってその運動エネルギーで定義される。

$$v_{agent} = \frac{|\mathrm{v_{max}}|}{1 + e^{-(N_{agent} \quad - \mathrm{N_0})}} \begin{pmatrix} \mathrm{motorL} \\ \mathrm{motorR} \end{pmatrix}, \qquad (1)$$

ここで v_{agent} は各 agent の速度、motorL 及び motorR はニューラルネットの出力層からの値、N_{agent} はある agent が他 agent にぶつかられた回数とする。$\mathrm{v_{max}}$ は agent がとりうる最大の速度である。agent は $\mathrm{N_0}$ 回以上他個体ぶつかられると active になる。$|\mathrm{v_{max}}|$ 及び $\mathrm{N_0}$ は実際のハチの観察データを元にそれぞれ 10.0、2000 とした。

また、active なハチはある値以上に動きまわると以下のようなルールで inactive になるものと仮定した。

$$E_{agent} = \int v_{agent}^2 dt$$
$$\mathrm{if} E_{agent} > \mathrm{E_0}, \mathrm{then} N_{agent} = 0, \qquad (2)$$

ここで E_{agent} は空間内での agent の移動距離、また $\mathrm{E_0}$ は実際のデータ解析をもとに 1500 とした。

2.2 Setting

シミュレーション空間の agent packing ratio は実際の観察データを元に 0.6 とし、空間サイズは 25 x 25 の周期境界条件、agent size は 2、agent の数は 100 とした。またシミュレーション時間における 1 step を 1 sec と仮定し以下の解析を行った。 またシミュレーションは、E_{agent}、N_{agent}、v_{agent} ともランダムな状態にしてスタートさせた。また、agent のニューラルネットワークは全て同じ重み行列である。

3. 解析結果

前研究での実際のセイヨウミツバチのデータ解析時同様、本研究でも agent のアクティビティー $K_{individual}$ を以下のように定義し解析を行った。

$$K_{individual}(t) = \Delta x^2 + \Delta y^2 , \qquad (3)$$

ここで Δx 及び Δy は 1 秒あたりの agent の移動量である。またコロニー全体のアクティビティーを $K_{individual}(t)$ の平均として定義した。

$$K_{global}(t) = \frac{1}{n} \sum_b K_{individual}(t). \qquad (4)$$

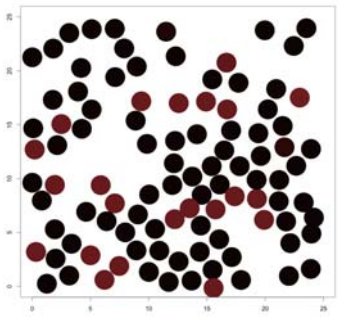


図 2: シミュレーション空間の様子。円で表されているものが agent。色は v_{agent} を 2 乗したものと対応し、赤色が強いほど高速で運動していることを表している。

初めにコロニー全体の活動を計測するために、agent 全個体の運動エネルギー $K_{global}(t)$ の時系列を作成した (図 2)。これに

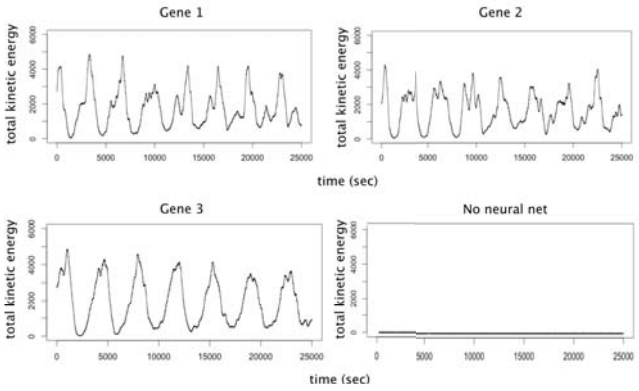


図 3: 遺伝子毎の K_{global} の時間発展の様子。

よりバーストの時間発展は重み行列によって、つまりフェロモンへの反応の仕方によって様々なパターンを示すことがわかった。また、バーストピーク付近での $K_{individual}$ の空間方向での相関長の様子も Gene 毎に異なることが分かった。この結果から、空間方向でのバーストの起こり方には、 inactive な数匹の agent 集団の中から $K_{individual}$ が大きな agent が現れこれが周りの inactive な agent たちを押して回るというパターン（図 5. A)。もう一方は空間の異なった場所で $K_{individual}$ が大きな agent が同時刻に出現する（図 5. B) という 2 パターンがあることが示唆された。

4. 考察

本研究では、我々の以前の研究で明らかにしたセイヨウミツバチのバースト行動のうち自発的に起きるバーストに注目し、agent based model を用いてモデル化することで、バーストの空間的な起こり方には 2 つのメカニズムがあることを見出した。ここでは、特に空間方向に agent が運動することに重きを置き、ニューラルネットを搭載し Agent based model でのシミュレーションを行ったが、各シミュレーションでは単一の Gene(重み行列）のみでシミュレーションを行っていたため、バーストのバリエーションは限られていた。会場では、様々な Gene が共存した状態でのシミュレーション結果、またこれに

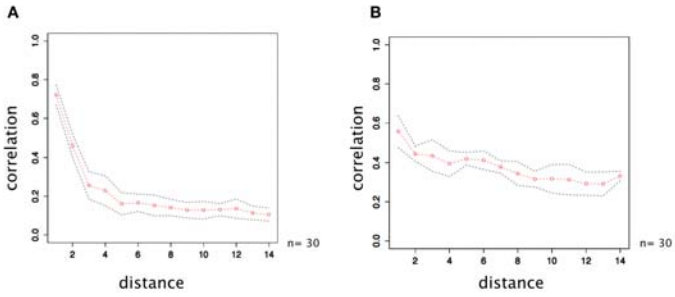


図 4: $K_{individual}$ の相関長。(A) Gene1 (B) Gene3

加えバーストモデルとして頻繁に研究されてきた SIR model によるセイヨウミツバチの実験データのモデル化に関する結果を発表予定である。

参考文献

[Robinson 1992] Robinson, Gene E. "Regulation of division of labor in insect societies." Annual review of entomology 37.1 (1992): 637-665.

[Doi 2017] Doi, I. and Ikegami, T. 社会性昆虫を通してみる個と集団. 人工知能学会全国大会論文集 (ISSN:13479881), vol.31, 2017.

[Crane 2008] Crane, Riley, and Didier Sornette. "Robust dynamic classes revealed by measuring the response function of a social system." Proceedings of the National Academy of Sciences 105.41 (2008): 15649-15653.

[Hawkes 1971] Hawkes, Alan G. "Spectra of some self-exciting and mutually exciting point processes." Biometrika 58.1 (1971): 83-90.

[Camazine 2003] Camazine, Scott. Self-organization in biological systems. Princeton University Press, 2003.

[Mersch 2013] Mersch, Danielle P., Alessandro Crespi, and Laurent Keller. "Tracking individuals shows spatial fidelity is a key regulator of ant social organization." Science 340.6136 (2013): 1090-1093.

[池田 2016] 池田圭佑, 榊剛史, 鳥海不二夫, 風間一洋, 野田五十樹, 諏訪博彦 and 栗原聡. (2016). マルチエージェント型情報拡散モデルの提案. 人工知能学会論文誌, 31(1), NFC-C_1.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ソーシャルタギングにおけるタグ出現数のゆらぎ

Generation and Selection Mechanism on Social Tagging Systems

西川仁将 *1 Yoshimasa Nishikawa　岡瑞起 *1 Mizuki Oka　橋本康弘 *2 Yasuhiro Hashimoto　池上高志 *2 Takashi Ikegami

*1筑波大学大学院システム情報工学研究科 Department of Computer Science, University of Tsukuba

*2東京大学大学院総合文化研究科 Graduate School of Arts and Sciences, The University of Tokyo

1. はじめに

近年，Delicious や Flickr，Twitter，などのソーシャルメディアサービスが普及している．これらのサービスではユーザがタグと呼ばれる任意の文字列を付与することによってコンテンツの管理を行うソーシャルタギングシステムが採用されている．

ソーシャルタギングシステムを対象とした研究は様々に行われている．たとえば Golder らは「Polya urn(壺モデル)」に基づいてソーシャルタギングシステムを理解しようとしている [1]．壺モデルでは過去に使われた回数の多いタグほど選ばれやすいという優先的選択性が働いている．また，Catutto らは既存のタグの選択に優先的選択性が働くことだけでなく，新しいタグの増加傾向がベキ則に従うことを示した [2]．このように新しいタグがどのように生み出されるか，また，それぞれのタグがどのように使用されていくかはソーシャルタギングシステムを理解する上で重要な問いである．

これらの問いを 1 つの数理モデルとして記述できるモデルとして Yule-Simon 過程がある [3][4]．Yule-Simon 過程は古典的な数理モデルの一つであり，新しいタグの増える割合を一定と仮定して，そうでなければ既存のタグを選択する，というルールに従う．Yule-Simon 過程は実際のソーシャルタギングシステムにおける全体的な振る舞いに関してはうまく記述することが可能である．

一方，個々のタグの累積出現回数について注目してみると Yule-Simon 過程が予測する値から逸脱するタグが存在して，Yule-Simon 過程ではその逸脱の分布が指数関数的減衰を示す [5]．しかし，実際のデータからは Yule-Simon 過程では捉えきれない大きさの逸脱をもつタグが多く存在することがわかっている．そこで本研究では実データにおける Yule-Simon 過程からの逸脱の大きさを確認し，Yule-Simon 過程を拡張したモデルを提案し比較する．

2. Yule-Simon 過程

本研究ではソーシャルタギングの新規生成と既存タグの選択のモデルとして Yule-Simon 過程を用いる．Yule-Simon 過程は，出現回数に比例した確率でランダムに既存の単語が選択される優先的選択性をもつ．Yule-Simon 過程では，新しいタグを追加するごとに，新たな種類の単語を新規タグ生成確率 α で生成し，これまでに使われた単語の中から確率 $1-\alpha$ で選択する．ここで t は時刻である．時刻 t におけるアノテーション数を $N(t)$，タグ i が出現している回数を $n_i(t)$ とすると，時刻 t にタグ i が選択される確率 $P_i(t)$ は，

$$P_i(t) = \frac{(1-\alpha)n_i(t)}{N(t)} \tag{1}$$

で与えられる．

Yule-Simon 過程では以下のような 2 つの性質を過程している．

- 新規タグ生成確率は一定でランダムに新規タグが生成される
- 出現回数に比例した確率でランダムに既存タグが選択される (優先的選択性)

また，Yule-Simon 過程では，定義からタグの種類の増加はアノテーション数のベキに比例 (Heaps 則) し，タグ出現数の累積確率分布は指数 $\alpha - 1$ のベキ分布 (Zipf 則) に従う．

3. Yule-Simon 過程からの逸脱

実データにおける個々のタグの出現回数が Yule-Simon 過程における予測値からどの程度逸脱するかを計測する．ここでは実データとして写真共有サービスである RoomClip[*1] を用いる．図 1 は RoomClip におけるタグの累積出現数の増加曲線の一部で，実線は実データで破線が Yule-Simon 過程が予測する出現数である．たとえば「ラウンジ」というタグではモデルの予測値と実データの振る舞いがほぼ一致している．一方，「地球儀」ではモデルの予測値に比べて増加が遅い．逆に，「男前」では予測値を大きく上回っている．この実測値とモデルの予測値の間のズレの大きさを x とするとタグ i が初めて使われた時刻 t_i から時刻 s_i までの累積出現数の予測値と実測値の比によって次のように記述できる．

$$x = \frac{n_i(s_i)}{{n_i}^*(s_i)} \tag{2}$$

ただし，$n_i^*(s_i)$ は時刻 s_i における予測値で次のように定義される．

$$n_i^*(s_i) = (\frac{s_i}{t_i})^{(1-\alpha)} \tag{3}$$

タグの成長が Yule-Simon 過程に従う場合，x の確率密度分布は，指数関数的減衰を示すことが示されている．たとえば予

連絡先: yoshimasa@websci.cs.tsukuba.ac.jp

*1 http://roomclip.jp/. Tunnel 株式会社が運営するサービス．日本最大の部屋のインテリア実例共有サイトである．

測値と実測値の比は，おおよそ 10 倍程度に収まる．しかし，実データにおける逸脱の大きさ x の分布を測定すると，Yule-Simon 過程が予測する成長よりも 100 倍，1000 倍の大きさになるタグが多く存在することが分析からわかっている (図 2)

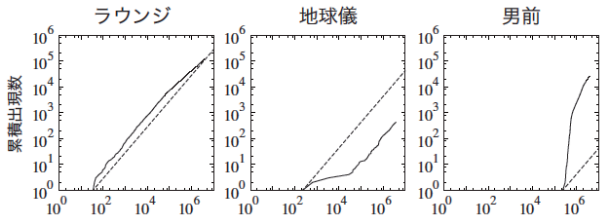


図 1: 個々のタグにおける累積出現数の増加

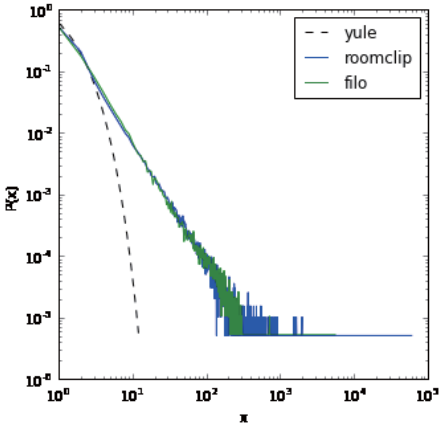


図 2: 各モデルにおけるゆらぎ x の分布

4. 提案モデル

図 1 に示したような予測値からの大きなズレを再現するために Yule-Simon 過程のタグの選択確率について修正を加える．実データでは Yule-Simon 過程と同様に Zipf 則が成り立っており，モデルの修正もこの性質を維持したまま行いたい．そこで Yule-Simon 過程のタグの選択確率を以下のように変形する．

$$\begin{aligned} P_i(t) &= (1-\alpha)\frac{n_i(t)}{N(t)} \\ &= (1-\alpha)\frac{n_i(t)k_i(t)}{N(t)}\frac{1}{k_i(t)} \end{aligned} \tag{4}$$

ここで出現回数が同じタグを Class と呼ぶこととする．Class1 とは累積出現回数が 1 回のタグで，クラス N とは累積出現回数が N 回のタグのことを指す．$k_i(t)$ とは時刻 t にタグ i が属する Class のタグ総数を表す．(4) 式右辺第 2 項はどの Class が選択されるかを表していて，第 3 項は Class の中でどのタグが選ばれるかを表している．Zipf 則が保たれるにはクラスの選択確率の形が必要条件であり，クラスの中でのタグの選択確率の形は問わない．そこで本研究では式の後半にバイアスを与えた First-In，Last-Out(FILO) モデルを提案する．これはある Class の中で最初に Class アップしたタグが必ず最後に選ばれるというモデルである．これによって最近使われたタグが優先的に選択され，より多く使われるタグが出現することが考えられる．

5. 結果・考察

まず，FILO モデルにおけるゆらぎの大きさ x の分布を計算し，カルバック・ライブラー情報量 (Kullback-Leibler Divergence,KLD) によって実データの分布の類似度を計測した．x の分布を図 2 に示す．

図 2 をみると FILO モデルはオリジナルの Yule-Simon 過程に比べて実測値により近い振る舞いを再現できていることがわかる．また，KLD による実データとモデルの分布の類似度は Yule-Simon が 0.833，FILO が 0.00390 という結果になった．これらの結果から提案する FILO モデルは Class 内のタグにバイアスをもった選択を与えることによって，大きな逸脱を生じ得ることを示している．

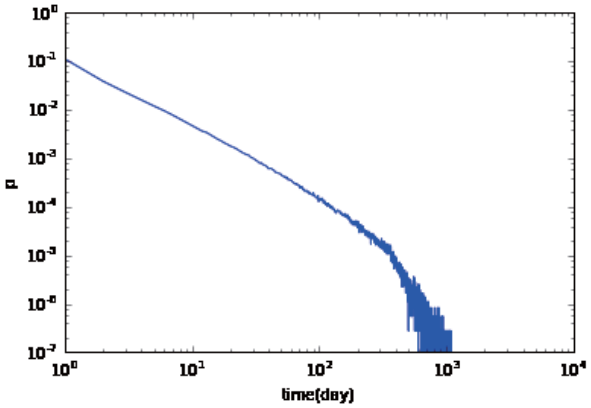


図 3: タグの再使用時間の分布

6. まとめ・課題

本研究では，Yule-Simon 過程では捉えきれない，出現数の増加曲線に大きなゆらぎをもつタグが実データに存在することを示した．また，Yule-Simon 過程の Class 内の選択にバイアスを加えることによって，タグの出現回数に大きな逸脱を生む効果を導入できることを示した．しかし，実データでは FILO モデルでも捉えきれない大きなゆらぎをもったタグが存在することが図 2 からわかった．

本研究で提案した FILO モデルでは，クラス内で最近使われたタグが必ず選ばれる，という 1 つの極端な例を考えた．このようなタグの「鮮度」に依存した使われ方は，図 3 に示すインターバルの分布を見ることでも明らかである．今後はこのような人間の注意の時間的な変化にも着目してモデル化を試みたい．

謝辞

解析に用いた貴重なデータを提供いただいた，Tunnel 株式会社の皆様に感謝いたします．また，本研究は JSPS 科研費 15K00420 の助成を受けたものです．

参考文献

[1] Golder, S. A. and Huberman, B. A. : Usage patterns of collaborative tagging systems, Journal of Information Science, Vol. 32(2), pp. 198-208 (2006)

[2] Cattuto, C., Loreto, V., and Pietronero, L.: Semiotic dynamics and collaborative tagging, Proceedings of the National Academy of Sciences, Vol. 104(5), pp. 1461-1464 (2007)

[3] Yule, G. U.: A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS,Philosophical Transactions of the Royal Society of London, Vol. 213, pp. 21-87 (1925)

[4] Simon, H. A.:On a Class of Skew Distribution Functions, Biometrika, Vol. 42(3/4), pp. 1-35(1955)

[5] Hashimoto, Y.: Growth fluctuation in preferential attachment dyanamics, Phys. Rev. E, Vol. 93, p. 042130 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

連続的な社会相互作用ダイナミクスに対するモデル・実験アプローチ

Experimental and modeling approaches to continuous dynamics of social interactions

鈴木 麗璽[*1]
Reiji SUZUKI

有田 隆也[*1]
Takaya ARITA

[*1] 名古屋大学 大学院情報学研究科
Graduate School of Informatics, Nagoya University

This paper introduces three approaches to continuous dynamics of social interactions: 1) a social particle swarm (SPS) model, in which particles move in a two-dimensional social space according to their social interactions based on the particles' game-theoretical states and positions, 2) a web-based experimental framework for understanding of dynamic social relationships with a real-time decision-making and rewarding environment, 3) a three-dimensional version of the SPS model with a massive population size.

1. はじめに

人は社会を形成して協力しあう種である一方，その構造は安定ではなく，変化し続けている．我々は，社会のネットワーク構造の持つ動的な側面を人間社会の本質であると考え，その仕組みや適応的意義の理解を目指している．

近年の情報技術の進展と普及は人と人との付き合い方を大きく変化させている．特に，ソーシャル・ネットワーキング・サービス（SNS）は，従来の物理的な付き合いに比べて大きな自由度をもたらしているといえる．その影響の一つとして，相互作用の連続性の高まりが考えられる．

これには次の二つの意味がある．一つは，相互作用の時間的な連続性である．例えば，スマートフォンなどのデバイスを用いて社会ネットワークに常時接することで，我々は刻一刻と社会的な関係を作り変え，同時に，常にそこから利益や不利益を得ているといえる．もう一つは，相互作用の程度，もしくは，質的な連続性である．例えば，SNS における直接的なつながりのネットワークの多くは離散的であるが，つながりそれぞれにおいて相互作用が生じる頻度や内容は様々であり，相互作用の程度や質には違いがある．また，多様なメディアの出現によってネットワークが重なることによっても，相互作用の程度に大きな違いが生じているといえる．

社会における動的な協力関係の創発は，進化ゲーム理論やネットワーク科学に基づく数理的・計算論的アプローチ[Pacheco, 2006, Suzuki 2008]や，近年では大規模なウェブ実験を含む被験者実験アプローチ[Rand 2011, Yonenoh 2014]によって理解されてきたが，その多くは離散的な相互作用を想定しており，上記のような意味での連続的な相互作用に注目した研究はまだ少ない．一方，例えば，実時間で利得が得られ続け，戦略の更新が随時可能な囚人のジレンマ状況では，被験者はより協力的に振る舞うことが示される[Friedman 2012]など，連続的な相互作用が社会的関係にもたらす影響について注目が高まりつつある．

我々は，現在，このような意味で連続的に変化する社会的相互作用において，協力的な関係がいかに生まれ，また，崩れていくかを明らかにすべく，エージェントベースモデルと被験者実験の両面から取り組んでいる．本稿では，これらの概要をまとめ，情報社会において発展拡大し続ける連続的な社会相互作用ダイナミクスに関して論ずる．まず，出発点として Nishimoto らが提案した social particle swarm（SPS）モデル[Nishimoto 2013, 2014]を紹介し，自己駆動粒子系で表現された連続的な社会関係の変化過程において，協力者集団のクラスタの創発と爆発的な崩壊が生じうることを示す．次に，SPS モデルと同様のゲーム的状況設定を想定した多人数 Web 被験者実験を用いて，類似の現象の創発や人間の行動規則の分析に基づくモデルの妥当性の検討，および，人の心理傾向とゲーム上での行動との関係の予備的分析について述べる．最後に，実世界の SNS のように無数の主体が相互作用する状況として，SPS モデルに準じた 3D モデルにおいて，集団のスケールを大きくした場合のその影響について予備的知見を示す．

連絡先：鈴木麗璽，名古屋大学大学院情報学研究科，名古屋市千種区不老町，+81-52-789-4258，reiji@nagoya-u.jp

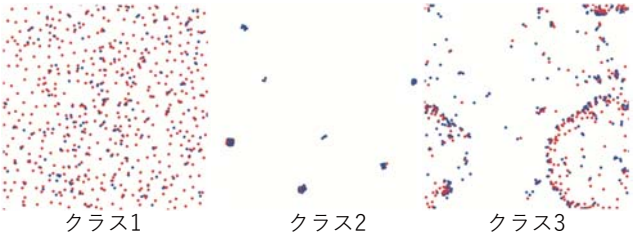


図 1：SPS モデルで生じる 3 種のクラス [Nishimoto 2013]

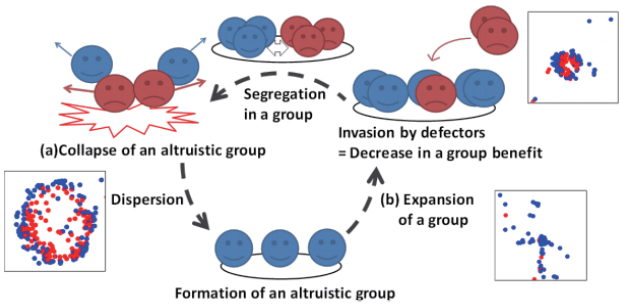


図 2：SPS モデルにおける協力クラスタの生成と崩壊 [Nishimoto 2013]

2. Social particle swarm（SPS）モデル

Nishimoto らは，ゲーム戦略と社会的関係性が連続的に変化する状況を自己駆動粒子系を用いて表現した social particle swarm（SPS）モデルを提案した[Nishimoto 2013, 2014]．

このモデルでは，700×700 の二次元トーラス平面を用いて各人とその間の社会的・心理的関係を抽象的に表現する．平面

上での連続的な相互作用を仮定するのは，主体間の社会的関係にはある程度の相関があると考えられるため，また，実世界に多数存在するそのような相互作用関係のチャネルの縮約として自然な表現の一つであると考えるためである．各粒子は囚人のジレンマゲームの戦略を状態として持ち，決められた半径内の近傍の協力粒子割合が，各粒子が固有に持つ閾値を超えた場合に自身も協力を選択する一般化しっぺ返し戦略従うものとする．各粒子は近傍粒子と戦略の組み合わせに基づきジレンマゲームの利得を得るが，粒子間距離が短いほど，すなわち社会的影響が強いほど利得の絶対値が大きくなるものとする．正の利得を得る相手に対して近づく方向に，負の利得を得る相手に対して遠ざかる方向に，それぞれ利得の絶対値に比例した力を受けるとみなし，その合力の方向に向かって一定速度で移動するものとした．

上記の設定で粒子数 700 の条件で実験を行うと，集団全体は図 1 に示すような 3 種のクラスのいずれかに至ることがわかった．クラス 1 は全体の半数以上が裏切りを選択し，多くが反発しあっている状況，クラス 2 はほぼ全個体が協力戦略をとり，いくつかのクラスタを安定して形成する状況である．

クラス 3 は我々が特に注目する状況であり，図 2 に示すような協力クラスタの生成と爆発的な崩壊の繰り返しが頻繁に観測された．まず，数個体の協力粒子が集まることで協力者の小集団が形成され，その高い協力率から他個体も協力的にして取り込みながらより大きな協力集団へと成長する．しかし，しばらくすると，裏切り傾向の強い（しっぺ返し閾値の高い）粒子も加わり蓄積することで，集団の内部で裏切り個体が居座る状況が生じる．裏切り個体から逃げる力と，協力者同士で繋がる力のバランスが崩れ後者が上回ると，裏切り者が協力者を追いかけるようにして集団が崩壊する．これが別の場所で再び協力者集団が生じるきっかけになり，このような過程が繰り返される．

我々は，このような過程が現実社会における集団の動的な変化の側面を反映していると考えている．

3. 多人数 Web 実験

SPS モデルで観察された相互作用ダイナミクスは，人同士の相互作用において実際に生じるだろうか．伊藤らは，人の集団において同様なゲーム的設定を用いた際に，上記のような協力クラスタの生成と崩壊を含むどのようなダイナミクスが観測されるかを明らかにすることを目的とし，多人数が同時に参加可能な Web 実験環境を構築し，分析を行っている[伊藤 2018]．

各実験参加者（プレイヤー）は相互の社会的関係の強さを抽象化した 500×500 の二次元トーラス平面を共有し，それぞれ平面上の粒子で表現される．各プレイヤーには図 3 に示す Web インターフェイスが提示され，中心の粒子が自身を，半径 100 の円内の粒子が視界円内の近傍プレイヤーを表す．また，各プレイヤーは囚人のジレンマゲームの戦略（協力（青色）または裏切り（赤色））のいずれかを状態として持つ．

実験は 0.5 秒間隔のステップ更新頻度で進行し，各プレイヤーは任意のタイミングで自身の戦略と位置を変更できる．具体的には，マウスのカーソルを円の中心から離れた場所へ移動させると，プレイヤーはその方向に一定の速さだけ進むことができ，また，戦略はキーで切り替えられる．

各ステップにおいて，各プレイヤーは近傍の個体それぞれとの関係から，利得を得る．このとき，基本的にはプレイヤー間のジレンマ戦略の組み合わせに応じた利得が得られるが，距離が近いプレイヤー同士ほど利得行列の得点の絶対値が大きく，社会的な影響が大きい状況を想定する．各プレイヤーの目的は，ゲーム期間中の累積得点を最大化することである．また，実験

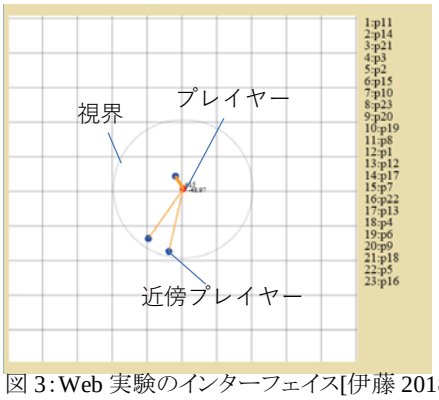


図 3：Web 実験のインターフェイス[伊藤 2018]

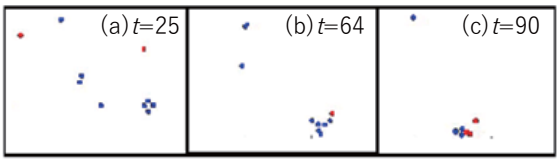


図 4：Web 実験での協力クラスタの生成と裏切り者の侵入[伊藤 2018]

終了時に，個人の心理傾向を 5 つの次元で計測する Big Five 尺度の日本語版 TIPI-J[Oshio 2012]と，個人の属する社会の中での関係形成の形成および解消の自由度を計測する関係流動性尺度[Schug 2010]に関するアンケートを行なった．

23 名，および，13 名の名古屋大学の学部生と大学院生を対象に予備的実験を試行したところ，次のことがわかった．まず，集団全体は協力者が裏切り者の数を平均的に上回り，協力行動が創発した．この時，図 4 に示すように，少数の協力者のクラスタが形成されたのちに，裏切り者への寝返りや侵入をきっかけにクラスタが解散する過程が局所的に繰り返され，SPS モデルでの協力クラスタの生成と崩壊に類する挙動が観察できた．

また，一般的な傾向として，協力者が近傍に多い場合に自身も協力しやすく，その場にとどまる傾向があった．これらは，一般化しっぺ返し戦略に基づき，正の利得を互いに得る場合には近くいて関係が維持されやすい SPS モデルの行動規則の妥当性の一端を示しているといえる．

さらに，心理指標に関するアンケートとゲーム中の行動指標の相関の分析からは，実験を行った集団に応じて異なる相関関係が観察された．例えばある集団では，ゲーム中に近傍個体数が多い人ほど協調的な心理傾向が高いことや，ゲーム中協力しがちな人ほど外交性が低く，日常において関係形成・解消が不自由な環境に置かれている傾向があった．また，別の集団では，ゲーム中戦略を頻繁に変更する人ほど周囲の刺激に敏感な傾向が見られたり，集団間で逆の傾向も観察された．個人の心理傾向がこのような仮想的な社会環境の行動においても反映されうることが示唆されたが，参加者の構成やゲーム設定など様々な要因が影響しうると考えられるため，より詳細な実験と検討が必要である．

4. 超多個体モデル

上記のような動的な社会関係の変化が小規模な抽象モデルと実際の人間の集団内で生じうることが示された．では，このような状況に基づく相互作用が，SNS のように何十万人のような大きなスケールの集団で生じたらどのような現象が創発するだろうか．

我々は，現在，GPGPU を用いたマルチエージェントモデル開発環境である FLAME-GPU [Richmond 2011]を用いて，SPS モ

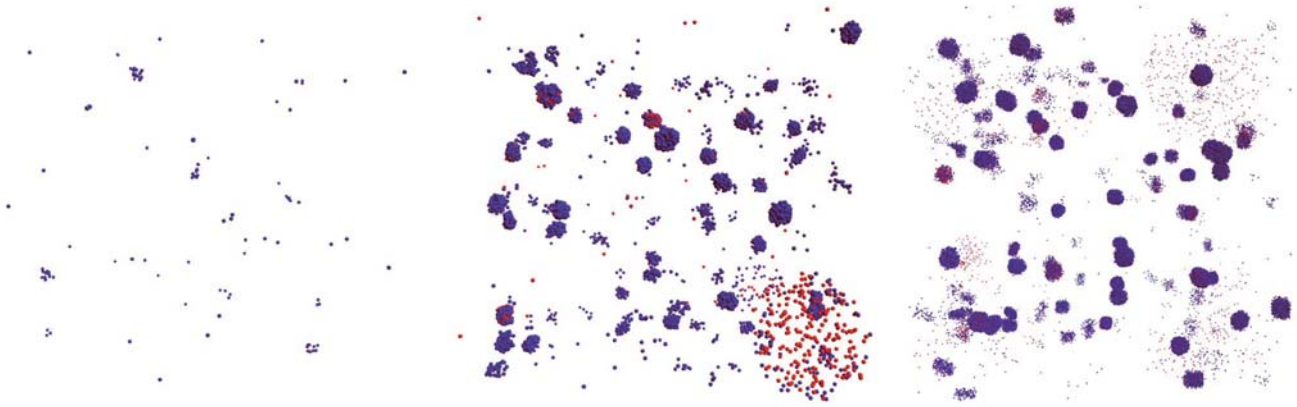


図 5：3D 版 SPS モデルにおける個体数 *N* の影響（*N*=100,000 の場合，各粒子の半径を小さく表示）

デルに準じた 3D 自己駆動粒子系モデルを実装し，基本的な挙動を分析している．現在のモデルにおける SPS モデルとの相違は，3 次元空間を想定したこと，および，移動に関しては，自身の現在の総利得が正の（負の）場合は近傍の粒子の重心の方向（から離れる方向）へ移動し，移動時には利得の大きさに応じた速度の変化と慣性が加わっていることである．

異なるスケールの個体数で行なった予備実験から次のことがわかってきた（図 5）．粒子数が少ない 100 個体程度の場合，個体間での相互作用の機会が少なく，協力者同士が引き合っても少数・小規模で，慣性などの影響で解消される場合もあった．粒子数が 1,000 程度になると協力者が多くを占める球状のクラスタが多数形成され，10,000 程度になると，SPS モデルで観察されたような協力クラスタの爆発的な崩壊が頻繁に生じることがわかった．さらに，粒子数が増えるに従い，例えば 100,000 を超えると，安定して存在する密で大規模なクラスタと空間上を浮遊する小規模なクラスタに分化する傾向が見られた．大小クラスタ間の相互作用により大規模なクラスタに発展したり，爆発的崩壊が生じて多数の小規模クラスタの発生を招くことなども観察された．これらは，集団における個体数が系全体の構造や多様性に大きく影響しうることを示している．現在，クラスタの分布や特性，クラスタ同士の相互作用について詳しい分析を進めている．

5. おわりに

本稿では，連続的な社会的相互作用のダイナミクスの理解を目的とした取り組みとして，SPS モデルにおける協力集団の生成と爆発的崩壊のダイナミクスを示し，Web 実験においても類似の行動傾向が観察されることや，心的傾向が行動に影響する可能性を示した．超多個体モデルにおいては，系を構成する個体数のスケールの増大が系の構造や多様性に影響しうることを予備的実験で示した．

これらの連続的な社会的相互作用に関する様々なアプローチを組み合わせて，人間の心的傾向から巨大な仮想社会にまで通じる普遍的な知見を得たいと考えている．

謝辞

伊藤百夏氏，小寺俊哉氏，西本恵太氏，Zineb AlHamer 氏（名古屋大学）の実験と執筆への協力に謝意を表する．本研究の一部は，科研費・特設分野・基盤研究（B）17KT0001「紛争と協力の文化進化的基盤に関する学際的研究」，および，日本学術振興会・課題設定による先導的人文学・社会科学研究推進事業（領域開拓プログラム）「予測的符号化の原理による心性の創発と共有-認知科学・人文学・情報学の統合的研究-」17J0011b の支援を受けた．

参考文献

[Friedman 2012] Friedman, D. and Oprea, R. (2012): A continuous dilemma, *American Economic Review*, 102 (1), 337-363.

[伊藤 2018] 伊藤百夏，鈴木麗璽，小寺俊哉，西本恵太，有田隆也: 多人数 Web 実験における 協力行動と社会的関係性の相互作用の統計的・心理分析，情報処理学会第 80 回全国大会講演論文集, 3Q-01 (2 pages) (2018).

[Nishimoto 2013] Nishimoto, K., Suzuki, R. and Arita, T.: Social Particle Swarm: Explosive Particle Dynamics on Cooperative/Defective Forces, *Proceedings of IEEE ALIFE 2013 (IEEE SSCI 2013)*, pp. 134-139 (2013).

[Nishimoto 2014] Nishimoto, K., Suzuki, R. and Arita, T.: Where do the dynamics of social relationship come from? - An analysis based on social particle swarm, *Proceedings of AROB2014*, pp. 86-91 (2014).

[Oshio 2012] Oshio, A., Abe, S. and Cutrone, P.: Development, reliability, and validity of the Japanese version of Ten Item Personality Inventory (TIPI-J), *The Japanese Journal of Personality*, 21, 1, 40- 52 (2012).

[Pacheco 2006] Pacheco, J. M., Traulsen, A. and Nowak, M. A.: Coevolution of strategy and structure in complex networks with dynamical linking, *Physical Review Letters*, 97, 258103 (2006).

[Rand 2011] Rand, D. J., Arbesman, S. and Christakis, N. A., Dynamic social networks promote cooperation in experiments with humans, *Proceedings of Natural Academic Science of the United States of America*, 108, 19193-19198 (2011).

[Richmond 2011] Richmond, P.: FLAME GPU Technical Report and User Guide, *University of Sheffield, Department of Computer Science Technical Report*, CS-11-03.

[Suzuki 2008] Suzuki, R., Kato, M. and Arita, T., Cyclic coevolution of cooperative behaviors and network structures. *Physical Review E*, 77(2), 021911 (2008).

[Schug 2010] Schug, J., Yuki, M., & Maddux, W. Relational mobility explains between- and within-culture differences in self-disclosure to close friends, *Psychological Science*, 21(10), 1471-1478 (2010).

[Yonenoh 2014] Yonenoh, H. and Akiyama, E., Selection of opponents in the prisoner's dilemma in dynamic networks: An experimental approach. *Journal of Theoretical Biology*, 351, 25-36 (2014).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

身体化されたニューラルネットワークにおける自律的な内と外の境界の規定

Autonomous Regulation of Self and Non-Self in Embodied Neural Networks.

升森敦士 *1 Atsushi Masumori　丸山典宏 *1 Norihiro Maruyama　シナパヤ ラナ *1 Lana Sinapayen　池上高志 *1 Takashi Ikegami

*1東京大学総合文化研究科
The University of Tokyo

Our previous studies has shown that embodied cultured neural networks and spiking neural networks with spike-timing dependent plasticity (STDP) can be learned as if it avoids stimulation from outside. In a sense, embodied neural networks can autonomously change their activity to avoid an external stimuli. It has known that there are two ways to learn a behavior to avoid the stimulation for the networks: (1) reinforcement of a behavior which can avoid external stimuli, (2) suppressing of a behavior which induces external stimuli. We found there is also a third property for avoiding stimulation where if avoiding external stimuli is difficult, then, network decrease stimulus-evoked spikes as if it ignores input neurons. We also show these results are reproduced by spiking neural networks with asymmetric STDP. These properties can be regarded as autonomous regulation of self and non-self for the network.

1. はじめに

[Shahaf and Marom, 2001] は，培養神経回路網には，外部刺激の入力によって可塑性が向上する *Modifiability* と外部からの刺激がない状態で神経細胞間の機能的結合関係が維持される *Stability* の 2 つの性質を持つと主張している．さらに，そのような性質をもった培養神経回路網において，外部から電気刺激を入力し，タスクが達成されたら刺激を止めるという手順の繰り返しによって単純なタスクの学習が可能であることを実際に培養神経回路網を用いた実験で示している．[Sinapayen et al., 2017] らは，スパイク時刻依存性可塑性 (STDP) とスパイキングニューロンモデルを組み合わせたシミュレーション実験を行うことで，上記の Shahaf らの主張する *Modifiability* や *Stability* といった抽象的な性質ではなく，STDP の長期増強 (LTP) によって刺激を避けるような神経活動を強化し，STDP の長期抑制 (LTD) によって刺激を増やすような神経活動が抑制されることで同様の学習が可能であることを示している．

ここでは、このような神経ネットワークが持つこのような刺激を避けるように可塑性が働く性質を刺激を避ける原理と呼ぶ。ここで重要なのは、STDP によって可塑性が駆動している神経ネットワークが身体性を持ち、環境と閉ループを構成している場合に、外からの指示などがない状況においても自律的に刺激を避けるようなふるまいが創発するという点である。これは神経細胞レベルからボトムアップに創発する内発的動機付けとしても捉えることもできる。

さらに、本研究では、我々の行った培養神経細胞を用いた同様の学習実験の結果、外部からの刺激入力を止めるような行動を表現することができないような状況においては、徐々に刺激入力に対する反応が抑制されることが分かった。これは、あたかも刺激入力の入ってくる神経細胞を切り離して、内と外の境界が自律的に規定されているとも捉えることができる。つまり、これまでの我々の研究から、神経ネットワークの持つ刺激の避け方には、上で述べた、1) 刺激入力を避けるような神経活動を強化する、2) 刺激入力を誘発するような神経活動を抑制するという二つ方法があると考えられていたが、今回、新たに第 3 の性質として、3)1) の方法で刺激入力を避けることが難しい場合に入力ニューロンからの結合強度を弱める、という方法があるということが新たに分かった。本論文では、特にこの刺激を避ける原理の第三の性質に注目してを結果を示して議論を行う。

連絡先: 升森敦士，東京大学総合文化研究科，masumori@sacral.c.u-tokyo.ac.jp

2. 対象と手法

2.1 分散培養神経細胞

本研究における培養神経細胞を用いたすべての実験で、胚齢 18 日目のウィスターラットの胎児から取り出した大脳皮質の一部を分散培養した培養神経細胞を用いる．また，培養神経細胞の活動を計測するための装置として高密度 CMOS アレイ (図 1) を用いる．この高密度 CMOS アレイは従来の一般的な多点電極アレイ (電極数 64 個程度) と比較して空間解像度が非常に高く,11,011 個の電極を持ち，その電極の中から最大で同時に 126 チャンネルの計測が可能となっている [Bakkum et al., 2008]. 本実験前に，高密度 CMOS アレイ上の神経細胞の活動データを計測し，そのデータをもとに実際に計測する 126 個のチャンネルを決定する．そのため，1〜100 程度の非常に少ない細胞数でも計測が可能となり，また，細胞数が 126 細胞以下である場合，理想的には各細胞の活動を個別に計測することが可能となる．本研究では，培養神経細胞は全て細胞数が 126 以下になるよう調整して培養したものを用いている．

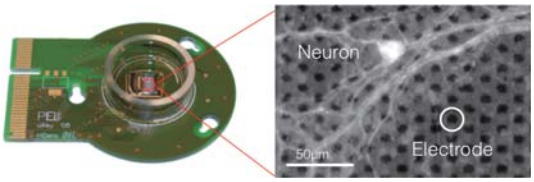


図 1: 高密度 CMOS アレイ

2.2 スパイキングニューラルネットワーク

本研究におけるシミュレーション実験では、興奮性と抑制性のニューロンモデルとして [Izhikevich, 2004] の提案するモデ

ル (IZ モデル) を用いる。このモデルは生体内の多様なニューロンのダイナミクスを再現でき、かつ、計算コストが低いことからシミュレーション実験等で広く用いられている。IZ モデルは以下のように定式化される:

$$\begin{aligned} \frac{dv}{dt} &= 0.04v^2 + 5v + 140 - u + I, \\ \frac{du}{dt} &= a(bv - u), \\ \text{if } v \geq 30\text{mV, then} &\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \end{aligned} \tag{1}$$

ここで、v はニューロンの膜電位、 u は脱分極に関わる変数、I は入力電流、t は時間を表し、a, b, c, d はその他のパラメータとなっている。このモデルでは、膜電位 v が 30mV を越えるとニューロンは発火したと見なされ、次ステップで v は c の値にリセットされる。本研究では、興奮性のニューロンとして $a = 0.02,\ b = 0.2,\ c = -65\text{mV}\ d = 8$、抑制性のニューロンとして、$a = 0.1,\ b = 0.2,\ c = -65\text{mV}\ d = 2$ のパラメータを用いる。

シナプス可塑性のモデルとして STDP を用いる [Song et al., 2000]。STDP では、2 つのニューロン間のシナプスの重みの変化は、それらのニューロン同士の発火のタイミングによって決まる。例えば、シナプス前ニューロンが発火してからある時間内にシナプス後ニューロンが発火した場合にシナプス結合の重みは上昇し、逆に、シナプス後ニューロンが発火してからある時間内にシナプス前細胞が発火した場合、シナプスの重みは減少するといったものである。具体的には重み w の変化は以下のように定式化される:

$$w_t = w_{t-1} + \Delta w\ . \tag{2}$$

$$\Delta w = \begin{cases} A_{LTP}(1 - \frac{1}{\tau})^{\Delta t}, & \text{if } \Delta t > 0 \\ -A_{LTD}(1 - \frac{1}{\tau})^{-\Delta t}, & \text{if } \Delta t < 0 \end{cases} \tag{3}$$

ここで、Δt はシナプス前ニューロンとシナプス後ニューロンの発火のタイミングの差を表す。本研究ではその他のパラメータは、対称的な STDP として、$A_{LTP}, A_{LTD} = 0.1$,$\tau = 20$ ms、非対称な STDP として、$A_{LTP} = 0.1, A_{LTD} = 0.12$,$\tau = 20$ms とした。ネットワーク構造はニューロン数 100(興奮性ニューロン 80, 抑制性ニューロン 20) の全結合とし、重みの初期値はランダムに決まるものとする。

2.3 実験設定

培養神経を用いた実験では 2 つの実験を行った: (1) 一次元空間上を移動するロボット実験 [Masumori et al., 2017]、(2) 実空間での二次元的な空間を移動するロボット実験 [Masumori et al., 2015]。それぞれ、壁に近づくと刺激が入り、壁を避けられるような神経活動が出ると刺激が止まるという条件になっている。(1) の実験では、刺激のない場合、ロボットは一定速度で前進し、刺激入力がある場合は、刺激後 20-40ms 以内に半数以上の出力ニューロンが発火すると 180 度回転して壁を避けることができる [Masumori et al., 2017]。(2) の実験では、左右二つの赤外線センサーがついており、壁に近づきセンサー値が上昇することで刺激入力が生じる。出力ニューロンはロボットの左右 2 つのモータースピードの値の係数となっており、壁を避けるためには左右のモータースピードを制御する必要があり、(1) の実験と比較すると難しい条件になっている [Masumori et al., 2015]。なお、(1) の実験では対照実験として刺激入力が神経活動によらずランダムに止まる条件での実験も行った。本研究では、刺激を避ける行動が学習できないような条件、または学習が難しい条件における神経活動を調べるために、これらの実験結果のうち、特に刺激を避けるための行動が学習できない実験 (1) の対称実験、学習がより難しい実験 (2) の結果に注目する。

シミュレーション実験では上記の実験 (1) と同様の実験を上述のスパイキングニューラルネットワークを用いて行った。



図 2: 一次元空間を移動するロボットの概念図。



図 3: 実空間でのロボット実験の様子

3. 結果

培養神経を用いた (1) の実験では、タスクが達成されるまでにかかる時間が減少し低い値で安定するといったように学習ができることが確認された (図 4)。一方、ランダムに刺激が止まるという条件の対照実験では学習は進まなかったが、興味深いことに、なかには入力刺激によって喚起される発火が徐々に抑制される場合があることが分かった (図 5)。

一方、培養神経を用いた (2) の実空間上のロボット実験では学習がより難しいものになっており、5 回中 2 回の実験のみ壁避け行動の学習が成功した。この際、学習達成の条件として、壁に接近して刺激が入り始めてから壁を避けるまでの反応時間が 30%以上向上することとした。ここでは、特に学習が失敗した場合のダイナミクスに注目する。

まず、壁避け行動の学習に失敗した実験での入力刺激によって喚起される発火頻度の変化をみてみると徐々に減衰していくことが分かった (図 6)。この際、入力ニューロンの発火頻度には減少する傾向は見られなかった (図 7)。このことから、単純に頻発する入力刺激によって入力ニューロンの発火が追いつかずその発火頻度が減少していったためではなく、入力ニューロンからのシナプス結合強度が減少することで、刺激誘発発火が抑制されたことが示唆される。

このように、培養神経ネットワークにおいて、刺激を避けるような行動の学習が難しいような場合には、その入力刺激自体を無視するような神経活動がみられることが分かった。

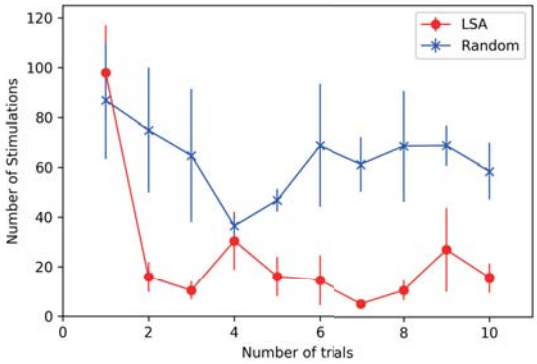


図 4: 学習曲線。値は各試行のタスクが達成までに入力された刺激の回数を表している。学習実験 (LSA) ではタスク達成までの反応時間が減少し低い値で安定しているが、対照実験 (Random) では、反応時間が大きく変化していないことが分かる。

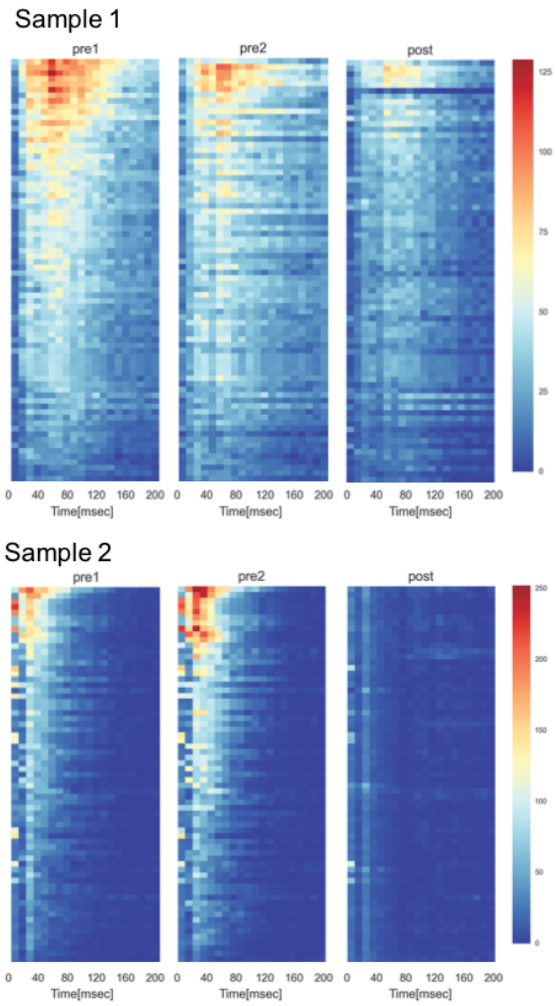


図 5: 刺激誘発発火の変化の典型例。刺激直後 200ms の全ニューロンの発火パターンを表す。ニューロンの順番は pre1 のデータを元に発火数でソートしている。

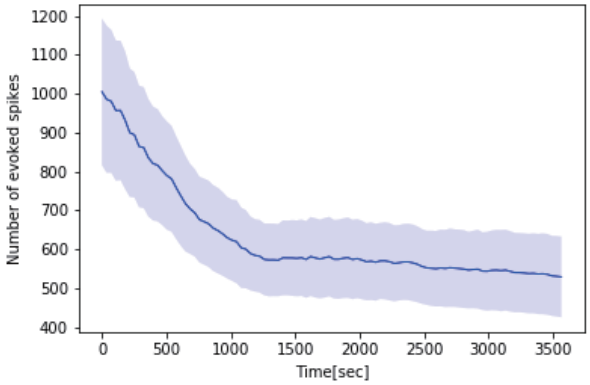


図 6: 刺激誘発発火の変化。値は 1 分ごとの刺激誘発発火の平均値を表す。刺激誘発発火が徐々に減少していることが分かる。

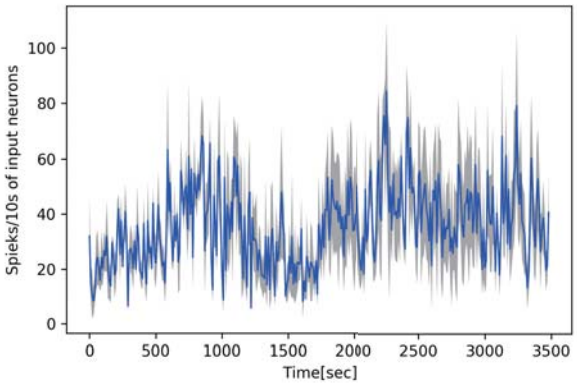


図 7: 入力ニューロンの発火数の変化。値は 10 秒ごとの入力ニューロンの発火数を表す。

シミュレーション実験では、LTP と LTD のダイナミクスが対称になっている STDP を適用した場合は、学習が可能な条件、学習が不可能な条件ともに、入力ニューロンからのコネクションの重みは上昇する結果となり、入力刺激を無視するような可塑性は生じなかった。一方で、LTP と LTD を非対称にした場合の STDP を適用した場合は、学習が可能な条件では、入力ニューロンからの重みは上昇するが、学習が不可能な条件では重みが減少するという結果が得られた (図 8)。

4. 議論

これまでの我々の研究から、培養神経細胞及び STDP による可塑性を持つスパイキングニューラルネットワークが、身体を持つことによって環境と閉ループを構成する場合、外からの操作がない状況においても、 1) 刺激入力を避けるような神経活動を強化する、 2) 刺激入力を誘発するような神経活動を抑制する、という 2 つ方法で自律的に刺激を避けるという性質があることが分かっている [Sinapayen et al., 2017]

本研究ではこれに加えて、培養神経ネットワークの実験において、刺激を避けための行動の学習が難しい状況においては、あたかも入力刺激自体を無視するような神経活動がみられることが分かった。この際、入力ニューロンの発火頻度に減少傾向

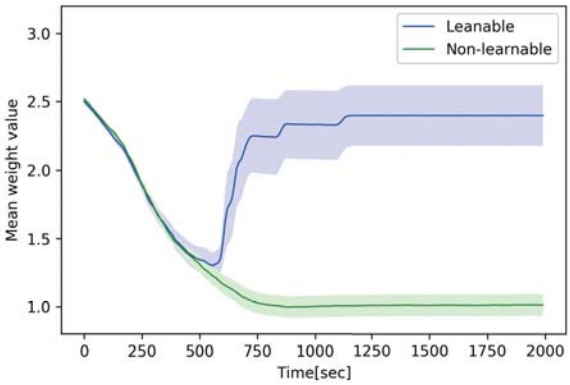


図 8: 入力ニューロンからの結合強度の変化。

はないことから、入力ニューロンから他のニューロンへの結合強度が弱まったことを示唆する。

また、シミュレーション実験では、LTD の方がより強くはたらく非対称な STDP を適用することで、刺激を避ける行動が学習可能な場合はその行動を強めることで刺激を避け、一方で、学習が難しい状況においては培養神経の場合と同様に、刺激入力の入ってくる入力ニューロンからの結合強度を減少させていくことが分かった。

このように、新たに、身体化した培養神経細胞及び STDP による可塑性を持つスパイキングニューラルネットワークがもつ第三の性質として、刺激入力を避けることが難しい場合には、入力ニューロンからの結合強度を弱めることで、入力ニューロン以外への刺激入力の影響を抑制する方向で可塑性がはたらく性質があることが分かった。これは制御不能な刺激入力の入ってくる領域を切り離そうとするように可塑性が働くことを示唆しており、制御不可能な領域を非自己として斬り離そうとするダイナミクスとして捉えることもできる。STDP のようなローカルなダイナミクスから、自律的に刺激入力を避けるような行動の学習が生じるだけでなく、自己組織的に自己と非自己を規定するようなダイナミクスが生じるということは興味深い。

参考文献

[Bakkum et al., 2008] Bakkum, D. J., Chao, Z. C., and Potter, S. M. (2008). Spatio-temporal electrical stimuli shape behavior of an embodied cortical network in a goal-directed learning task. *Journal of neural engineering*, 5(3):310–23.

[Izhikevich, 2004] Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070.

[Masumori et al., 2015] Masumori, A., Maruyama, N., Sinapayen, L., Mita, T., Frey, U., Bakkum, D., Takahashi, H., and Ikegami, T. (2015). Emergence of Sense-Making Behavior by the Stimulus Avoidance Principle: Experiments on a Robot Behavior Controlled by Cultured Neuronal Cells. *Proc. of the European Conference on Artificial Life (ECAL) 2015*, pages 373–380.

[Masumori et al., 2017] Masumori, A., Maruyama, N., Sinapayen, L., Mita, T., Frey, U., Bakkum, D., Takahashi, H., and Ikegami, T. (2017). Learning by Stimulation Avoidance in Cultured Neuronal Cells. *The 2nd International Symposium on Swarm Behavior and Bio-Inspired Robotics (SWARM2017)*.

[Shahaf and Marom, 2001] Shahaf, G. and Marom, S. (2001). Learning in networks of cortical neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(22):8782–8.

[Sinapayen et al., 2017] Sinapayen, L., Masumori, A., and Ikegami, T. (2017). Learning by stimulation avoidance: A principle to control spiking neural networks dynamics. *PLOS ONE*, 12(2):e0170388.

[Song et al., 2000] Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature neuroscience*, 3(9):919–26.

[2E2] [General Session] 9. NLP / IR

座長:田中 貴秋（NTT コミュニケーション科学基礎研究所）

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room E (4F Queen)

[2E2-01] A Framework of Generating a Narrative Using Metaphor Mapping and Post-editing

○Suguru Matsuyoshi[1], Akira Utsumi[1] （1. The University of Electro-Communications）

1:20 PM - 1:40 PM

[2E2-02] A Generator of Various Styles of Utterances Using Utterance Expression Dictionary

○Ryo Kimura[1], Kazuko Natsume[1], Satoshi Sato[1], Takuya Matsuzaki[1] （1. Nagoya University）

1:40 PM - 2:00 PM

[2E2-03] Headline Generation of Yahoo! Chiebukuro Based on Learning to Rank

○Tatsuru Higurashi[1], Hayato Kobayashi[1,2], Kazuma Murao[1], Masuyama Takeshi[1] （1. Yahoo Japan Corporation, 2. RIKEN AIP）

2:00 PM - 2:20 PM

[2E2-04] TULIP: Generation of Screenplay-style Novel (SS) Using Three-step LSTM Trained on Web Novels

○Atsushi Suzuki[1] （1. The University of Tokyo）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

メタファー写像と後編集を利用する物語文章生成フレームワーク

A Framework of Generating a Narrative Using Metaphor Mapping and Post-editing

松吉 俊 *1　　内海 彰 *1
Suguru Matsuyoshi　Akira Utsumi

*1電気通信大学大学院 情報理工学研究科
Graduate School of Informatics and Engineering, The University of Electro-Communications

We propose a framework of generating a narrative. Using simulators and metaphor mapping knowledge, our system generates various narratives which can be revised better by post-editing.

1. はじめに

人工知能や認知科学の分野において、物語文章自動生成システムの研究は古くから取り組まれてきた [秋元 13]。物語生成の主な流れは、次のとおりである。

1. 物語内容 (story) の決定: 何を語るか
2. 物語言説 (discourse) の決定: どのような構成・技法を用いるか
3. 文字列化: 内部表現をどのように表層文字列に変換するか

本研究では、この流れを利用しない新しいパラダイムを提案する。具体的には、人間と同等以上の推論・探索能力が期待できる問題やゲームにおいて、システムにその世界をシミュレート [松尾 16] させ、その履歴である状態遷移列をメタファー写像に基づいて解釈させるという方法を提案する。

2. 関連研究

物語文章自動生成の関連研究について、主に利用されるフレームワークという観点からまとめる。

内部構造 物語文法などに基づき、独自の記法で記述されたスクリプト (事象集合のグラフ) を操作する手法 [秋元 13, Akimoto 17, Montfort 13, 佐藤 16]

テンプレート テンプレート列を利用して文章を生成する手法 [長尾 96]。テンプレートには、スロットと呼ばれる穴があり、事前に定めた選択肢集合の中から 1 つを選択して穴を埋めることにより、テキストを生成する

連接尤度 これまで生成した語句の列を入力として受け取り、次の語句を生成する手法 [浅川 15]。文の境界を越えつつ、この過程を繰り返すことによりテキストを生成する

単純シミュレート ゲームやシミュレーターを実行して得られた出力や実行ログの列を文字列に翻訳する手法 [福田 15, 鳥海 16, 松山 17]

連絡先: 松吉 俊、電気通信大学大学院 情報理工学研究科、東京都調布市調布ヶ丘 1-5-1 電気通信大学 西 5 号館 505 号室、042-443-5292、matuyosi@uec.ac.jp

「内部構造」手法は、前章で述べた 3 つのステップを実現するものであり、理論的で汎用的な手法である。必要に応じて、「テンプレート」手法が併用されることもある。研究者独自の内部表現が多く、複雑な体系をしていることが多い。

「テンプレート」手法は実装が容易であるため、駅における列車の行き先案内など、一般の言語生成タスクにおいても古くから利用されている。スロットの位置や選択肢集合をどのように定めるのが適切かを問題ごとに決定する必要があり、分野専門性が求められる。

「連接尤度」手法は、N-gram モデルやリカレントニューラルネットワークモデルを利用するものである。現在の技術では、意味の通る文章を生成できるのはたかだか数文程度である。

「単純シミュレート」手法は、これまで人力で行われていたテレビゲームの小説化作業を、計算機に実行させるものである。目的の物語を生成するためには、そのシミュレーターを用意しなければならないという短所がある。

これら上記の手法は、他の自然言語処理タスクと比べると、システムの評価 (人手評価、自動評価ともに) が難しいものとなっている。一般に、あるシステムが複数の構成要素 (モジュール) からなる場合、全体の評価とモジュールごとの評価が可能である。コストや方法論などの面においてシステム全体の評価が難しい場合でも、モジュールごとの自動評価が相対的に容易であれば、各モジュールの性能を向上させることにより、ひいては、システム全体の性能を向上させることができる。しかしながら、先行研究においては、モジュールの入出力に、複雑な独自内部表現や数値行列が採用されていることが多いため、その出力を人間に評価させることも、コミュニティ共有の評価データを用意することも難しい。

その他の比較については、論文 [松吉 18] を参照してほしい。

3. 提案手法

提案手法の流れを図 1 に示す。

3.1 概要

本手法では、以下の手順により、自然言語で書かれた物語文章を自動生成する。

1. **メタファー写像データベースの構築**

   テキスト集合からメタファー写像の集合を自動抽出し、表 1 に示される形式でデータベースに格納する。この表の最左列には、ゲーム名やシミュレート世界が示されており、この世界を構成する事象と、物語を構成する事象と

表 1: メタファー写像の例 (左辺はシミュレーター内の行動や状態。右辺は物語内の表現)

	物語	メタファー写像	
チェス	部屋の片付け	ポーンを取る ⇌ 本を整理する	ポーンが取られる ⇌ 本を読みふける
	悩み	ポーンを取る ⇌ ほんの少し前向きになる	ナイトを取る ⇌ 誰かに相談してみようと思う
	多対 1 の恋愛競争	チェックをかける ⇌ 対応人物が告白する	駒が取られる ⇌ 対応人物が誘惑され離脱
将棋	探し物	王手が防がれる ⇌ 見つけたと勘違い	駒が取られる ⇌ 探す気力が減る
迷路探索	プロジェクト進行	ゴールに近づく ⇌ 順調に進む	袋小路に向かう ⇌ 泥沼に陥る
果物育成	発掘調査	水を撒く ⇌ 地道に掘る	実が小さい ⇌ 成果が少ない
野球の打席	会話の駆け引き	直球を投げる ⇌ ストレートに伝える	ボール球に手を出す ⇌ うっかり話す

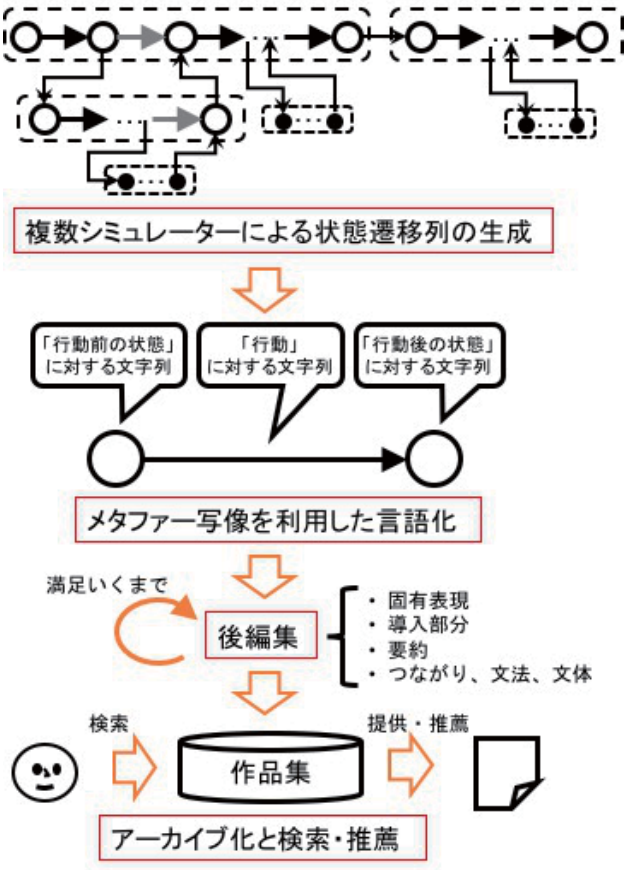


図 1: 提案手法の流れ

の間の関係を写像として蓄積する。メタファー写像に関する理論的説明は、3.2 節で述べる。

2. **複数シミュレーターによる状態遷移列の生成**

   ゲーム AI やシミュレーターを実行し、実行に付随する「状態」や「行動」の列 (状態遷移列) を取得する。スペースの都合により、図 1 では、この段階ですでに世界同士が繋がるように描いたが、実際に繋がるのは、言語化の後である。

3. **メタファー写像を利用した言語化**

   表 1 のようなメタファー写像データベースを参照し、状態遷移列の各要素を文字列に置き換える。対応する写像が存在しない場合、置き換えは実施せず、その要素を削除する。

4. **後編集**

   人間が草稿を書いた後に実施するのと同様に、文章の質を高めるためにテキストを校正する。具体的なサブモジュールについては、3.3 節で説明する。

5. **アーカイブ化**

   自動生成された多くの物語文章を作品集としてまとめる。ユーザーは、検索により、望む作品を見つける。

割り当てられた文字列が高い類似度を持つ「状態」を媒介として、複数の世界を繋げる。繋ぎ方には次の 2 とおりがある。

- 複数シミュレーターを並列的に走らせてから、のちに繋げる。
- 1 つのシミュレーターを走らせ、状態遷移列を生成する。新しい「状態」が生成されれば、すぐにそれを言語化し、その「状態」を始点とできるシミュレーターの有無を調べる。存在すれば、そのシミュレーターを走らせるとともに、世界を繋げる。

3.2 メタファー写像に関する不変性原理

認知言語学の概念メタファー理論において、**メタファー**は、2 つの意味領域間の連結であると捉えられる [鍋島 11, アリス 10]。表 1 はメタファーの例を表すものであるが、例えば表の 6 行目において、「プロジェクト進行」という意味領域 (根源領域) が、「迷路探索」という意味領域 (目標領域) に結び付けられている。この連結において、例えば、根源領域内の「順調に進む」という表現は目標領域内の「ゴールに近づく」という表現に**写像**可能であるとみなされる。

メタファー写像に関して、Lakoff [Lakoff 93] は、「写像において、根源領域の表現間の (位相) 構造は目標領域内でも保たれる [*1]。」という**不変性原理**を提唱した。この原理に則ると、任意の 2 つの意味領域間において数個のメタファー写像の存在が確認されれば、根源領域内の他の表現も位相構造を保ったまま目標領域内に対応づけることができる見込みが非常に高い。言い換えると、次の 2 つのことが成り立つ見込みが非常に高い。

- テキストのジャンルを問わず、数個のメタファー写像の存在が確認されれば、表 1 のようなデータベースが高い品質で構築できる
- ゲームやシミュレーター内に実装されている種々の制約や論理が、物語文章生成においても制約・論理として働くので、物語の展開において、「狭い意味でのフレーム問題」[人工知能学会 05] を回避できる

我々は、これらの成立の検証も兼ねて本研究を進めている。

*1 正確には、「目標領域が課す制約が許す範囲において」が付く。

3.3 モジュール

この節では、提案手法に関連するモジュールについて考察する。

多くのモジュールがすでに自然言語処理において研究が進められているタスクであり、その部分のみの評価は比較的容易であると思われる。また、後編集のサブモジュールは、すべて、文章を入力として別の文章を出力するものである。文字列から文字列への系列変換問題とみなすことができるため、必要に応じて訓練データを用意すれば、既存の多くの手法が適用可能である。

3.3.1 メタファー写像データベースの構築

メタファー認識

テキストからメタファーを自動認識する技術が必要である。「のように」や「みたいな」のような手がかり語が利用できる直喩のみでなく、隠喩も認識することが望まれる。

知識獲得

数個のメタファー写像の存在が確認された場合、明示的な写像がテキスト中になかった事象も、物語の構成要素として獲得する。従来の知識獲得では、時間的に近い事象のみを収集することが望ましいと思われるが、時間的に離れた事象を収集することも大事であると思われる。例えば、ツイート間の間隔が数日ある場合、物語の階層構造の章レベルもしくは段落レベルの知識として収集することが考えられる。

高い精度での知識獲得が実現されれば、人手を全く介さずに、新語を含むテキストからデータベースを更新し、新語を含む物語を生成できるようになると思われる。

フィルタリング

出力される物語に不適切な表現が含まれないように、あらかじめデータベースからそのような表現を除去することが可能である。除去した表現を別のデータベースに蓄積しておくことにより、文字列完全一致でなくとも、別の不適切表現を既存の自動分類技術により自動認識することができると思われる。

クラウドソーシング

データベースの構築にクラウドソーシングも有効である。

データベースの共有

メタファーは文化に依存する。少資源言語におけるデータベースを構築する際に、文化の共通性という観点に基づいて取捨選択すれば、リソースが豊富な他言語のデータを一部共有可能であると思われる。

3.3.2 複数シミュレーターによる状態遷移列の生成

状態遷移列の生成自体は、言語非依存である。それゆえ、このモジュールは日本語以外にも直接応用可能である。

ゲームのバリエーション

システムの設定を変更することにより、物語としての一貫性を保ったまま、容易に出力にバリエーションを持たせることができる。例えば、チェスの場合、駒の数や配置を変えて物語を生成することができる。迷路探索の場合、複雑な迷路やほぼ一本道の迷路などを選択できる。

機械学習

失敗もストーリーである。それゆえ、通常のタスクにおいては失敗であるとみなされる学習モデルであったとしても、魅せる物語を生成するために利用できる可能性が高い。単に強いだけの AI はおそらく望ましくなく、接待将棋 AI や指導碁 AI を開発するような特別な方法論が必要となるかもしれない。

また、物語の構成単位 (部、章、節、段落) のどこを担当するかによって、良い学習モデルが変わる可能性がある。

マルチエージェント

登場人物が複数である場合、それぞれ別のシミュレーターを走らせることができる。

伏線

伏兵が絡むようなシミュレーターの場合、伏線を持った物語を自動生成できる可能性がある。

(登場人物が実行する) 予測

1 つのゲームに複数の AI が関与する場合、「今後こうなると思う」と登場人物に予測を語らせ、予測と実際の結果の差分が楽しめるような物語が生成できる可能性がある。

3.3.3 メタファー写像を利用した言語化

比喩表現

提案手法と「単純シミュレート」手法を併用することにより、生成される物語に比喩表現を混ぜることができる。例えば、将棋利用時に「穴熊です」のような隠喩表現を生成したり、果物育成シミュレーター利用時に「植物に水を撒くように地道に掘る」のような直喩表現を生成したりすることができる。

レトリック表現

メタファー写像を用いることにより、擬人法、誇張法、抑言法、婉曲法 (忌語を避ける) の表現が自然に生成される可能性がある。

3.3.4 後編集

後編集モジュールは、システムが生成したテキストのみならず、人間が書いたテキストにも適用可能である。それゆえ、これらのモジュールは、子供や日本語学習者を支援するツールに直接応用できる。

固有表現認識

本提案手法の枠組みでは、メタファー写像データベースにおいて固有表現は制御されていない。それゆえ、物語全体として一貫性を保つためには、後編集において、適切な固有表現を選択する必要がある。後編集すべきテキストから固有表現を認識する技術が必要である。

固有表現の置き換え

固有表現の制御は大きく 2 種類ある。1 つは、絶対的な置き換えであり、例えば、主要登場人物が「太郎」である場合、テキスト中の主要人物の名前をすべて「太郎」に置換する。もう 1 つは、相対的な置き換えである。例えば、イギリスの物語を生成している最中に、「近くのパリに出かける」のような表現があった場合、< イギリス $\rightarrow$ フランス > の関係から類推して < ロンドン $\leftarrow$ パリ > の関係を得て、「パリ」を「ロンドン」に置換する。このような類推ができる機構が必要である。

相対的な置き換えにおいて、現在の知識では置き換え候補が存在しないことが認識できれば、新たに造語を用意する必要があることをシステムに認識させることができる。状況に合ったもっともらしい造語を生成できる機構が必要である。

導入部分

物語の導入部分には、登場人物の紹介があると良い。主要な物語文章を生成し終えた後で、その人物の行動を抽出し、それらから帰納的に性格や容姿を推論させる方法が有効であると思われる。

要約・ハイライト抽出

本フレームワークでは、世界をシミュレートしているので、冗長で退屈な部分が多い。この問題を解消するために、要約技術やハイライト抽出の技術が必要である。

つながり訂正、文法訂正、文体訂正

言語学や日本語教育学における知見に基づいた訂正機構が必要である。特に、1 文内の訂正のみに限らず、文章構成論に基づく校正 [石黒 04] の研究が必要である。

3.3.5 アーカイブ化

検索、推薦

おそらく既存の枠組みで十分であると思われる。

3.3.6 全体のフレームワーク

質問応答

表 1 のようにチェスで「多対 1 の恋愛競争」を描写する場合、生成時に盤面情報を保有しているので、全登場人物の現状について質問応答 (「ジョンは奥手で全く動いていない。」など) が可能である。このとき、要望があれば、特定の邪魔な駒を排除するなど、物語の現状を途中で改変し、改変後の世界で物語を引き続き生成し続けることも可能であると思われる。

文章の新しい分類軸

これまで、文章は、ジャンルや難易度で分類されることが主であった。本フレームワークを使用することにより、計算論的に次のような分類軸を考慮することができるかもしれない。

- 生成の計算量 (時間計算量と領域計算量)
- 階層構造の複雑さ

4. 提案手法の試作

提案手法に関して、「複数シミュレーターによる状態遷移列の生成」と「メタファー写像を利用した言語化」を実装した。プログラミング言語は、Python 3 を用いた。

4.1 クラス定義

実装したシステムは、物語生成のために、次のようなクラスを利用する。

main クラス 全体の制御

Instrument クラス 楽器に見立てたシミュレーター

Score クラス パート譜に見立てたメタファー写像群

Performer クラス 楽器とパート譜をもつ演奏者

Orchestra クラス すべての楽器と演奏者からなる

Story クラス 物語を制御

Part クラス 演奏者を雇い、文字列を生成

4.2 現状

現在の実装においては、シミュレーターとしてチェスのみが利用可能である。メタファー写像を記述することにより、物語が生成可能である。

5. おわりに

本論文では、メタファー写像に基づく変換と階層化可能な状態遷移列を組み合わせることで、多様な物語文章を自動生成することができるフレームワークを提案した。今後は、知識獲得に関するモジュールの実装と後編集に関する学習コーパスの構築を実施する予定である。

本フレームワークで生成しうるものは、人間が創作しうる物語集合のうち、ほんの一部である。本フレームワークでは生成できない差分を見ることにより、人間の言語能力に感嘆すると共に、フレームワークのさらなる改良を行いたい。

参考文献

[秋元 13] 秋元 泰介, 小方 孝：物語生成システムにおける物語言説機構に向けて, 認知科学, Vol. 20, No. 4, pp. 396–420 (2013)

[Akimoto 17] Akimoto, T.: Computational Modeling of Narrative Structure: A Hierarchical Graph Model for Multidimensional Narrative Structure, *International Journal of Computational Linguistics Research*, Vol. 8, No. 3, pp. 92–108 (2017)

[アリス 10] アリス ダイグナン著, 渡辺 秀樹, 大森 文子, 加野 まきみ, 小塚 良孝訳：コーパスを活用した認知言語学, 大修館書店 (2010)

[浅川 15] 浅川 伸一：ニューラルネットワーク物語生成モデル, 人工知能学会第 2 種研究会ことば工学研究会資料, Vol. 49, pp. 15–21 (2015)

[福田 15] 福田 清人, 森 直樹, 松本 啓之亮：既存小説に依存しない小説の自動生成に関する一考察, 言語処理学会第 21 回年次大会発表論文集, pp. 992–995 (2015)

[石黒 04] 石黒 圭：よくわかる文章表現の技術 II 文章構成編, 明治書院 (2004)

[Lakoff 93] Lakoff, G.: The Contemporary Theory of Metaphor, in Ortony, A. ed., *Metaphor and Thought (2nd edition)*, Cambridge University Press (1993)

[松尾 16] 松尾 豊：Deep Learning から身体性、シンボルグラウンディングへ, 人工知能学会全国大会発表論文集 1A5-OS-27c-2 (2016)

[松山 17] 松山 諒平, 佐藤 理史, 松崎 拓也：人狼ログからの小説の自動生成, 言語処理学会第 23 回年次大会発表論文集, pp. 32–35 (2017)

[松吉 18] 松吉 俊, 内海 彰：メタファー写像に基づく物語文章の自動生成, 言語処理学会第 24 回年次大会発表論文集, pp. 1288–1291 (2018)

[Montfort 13] Montfort, N., Pérez y Pérez, R., Harrell, D., and Campana, A.: Slant: A Blackboard System to Generate Plot, Figuration, and Narrative Discourse Aspects of Stories, in *Proceedings of the International Conference on Computational Creativity (ICCC) 2013* (2013)

[鍋島 11] 鍋島 弘治朗：日本語のメタファー, くろしお出版 (2011)

[長尾 96] 長尾 真（編）：岩波講座ソフトウェア科学 15 自然言語処理, 岩波書店 (1996)

[佐藤 16] 佐藤 理史：コンピュータが小説を書く日, 日本経済新聞出版社 (2016)

[鳥海 16] 鳥海 不二夫, 大澤 博隆：AI 達は物語を生み出すか, 人工知能学会全国大会発表論文集 2F4-5 (2016)

[人工知能学会 05] 人工知能学会 (編)：人工知能学事典, 共立出版 (2005)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

発話表現文型辞書を利用した多様な発話文生成機構

A Generator of Various Styles of Utterances Using Utterance Expression Dictionary

木村 遼 Ryo Kimura　夏目 和子 Kazuko Natsume　佐藤 理史 Satoshi Sato　松崎 拓也 Takuya Matsuzaki

名古屋大学 大学院工学研究科 情報・通信工学専攻
Graduate School of Engineering, Nagoya University

This paper describes an utterance generator, which can generate various styles of utterances. The key component of this generator is Utterance Expression Dictionary, which contains more than 1,000 utterance patterns with style parameters; these patterns are also classified into 74 groups of utterance intention. Given an utterance content, an utterance intention, and style parameters, the generator produces a surface string of the utterance that has the given style.

1. はじめに

小説は，地の文と会話から構成されるのが普通である．地の文が一人称あるいは三人称視点からの描写であるのに対し，会話は複数の登場人物の発話の列として構成される．小説の発話は，実際に人間が話す本当の意味での発話ではないが，それを模したものであるため，話し言葉的要素を多く含む．さらに，その発話の話者が誰かという情報を「とAさんが言った」のような明示的な描写なしで伝えるため，発話自体にその情報を持たせる工夫がなされることが多い．通常，それは，それぞれの登場人物に対して，異なる言い方・話し方・言葉遣いを用いることによって実現される．そして，このような言い方の個性は，登場人物のキャラクタ設定においても，重要な役割を果たす．

ほとんどの小説は会話を含む．そのため，小説の自動生成を実現するためには，上記のような特徴をもつ会話—発話の列—を生成する必要がある．その際，それぞれの登場人物に対して個別の発話文生成システムを作ることは現実的ではなく，一つのシステムで多くの話者の発話を生成できるシステムが望ましい．つまり，話者の情報をパラメータとして受け取り，それによって言い方・話し方・言葉遣いを変更できるようなシステムが求められる．

我々は，このようなシステムを，以下に示す手順に従って実現することを計画している．

ステップ **1** 多様な発話文を生成できる機構を実現する．具体的には，一連のパラメータ群によって，言い方・話し方・言葉遣いを制御できるような発話文生成機構を実現する．

ステップ **2** 実現したい話者をパラメータ群に対する仕様として定義し，話者像が推測できるような発話文の生成を実現する．

上記の 2 つのステップのうち，本稿ではステップ 1 を扱う．

本研究は，テキストにキャラクタ性を付与する研究の一形態として位置付けられ，類似研究に，水上らの研究 [水上 14] と宮崎らの研究 [宮崎 15] がある．これら 2 つの研究は，いずれもオリジナルのテキストを書き換えることによってキャラクタ性を付与する方法を採用している．この方法は，標準的な発話の存在を仮定するのであれば機能する．しかし，仮定しないのであれば，書き換え系は，書き換え前の発話の話者Aと書き換え後の話者Bの両方に依存することとなり，多数の話者の発話を生成するのは非常に困難となる．我々は，「発話には，必ず話者の情報が何らかの形で反映されている」という立場を取る．そのため，書き換えによる実現方法を採用せず，内部表現から表層形式を生成する過程でキャラクタ性を付与する方法を採用する．

連絡先: 木村遼，名古屋大学大学院 工学研究科 情報・通信工学専攻，〒 464-8603 愛知県名古屋市千種区不老町 C3-1(631) IB 電子情報館南棟 159 号室，052-789-4435，r_kimura@nuee.nagoya-u.ac.jp

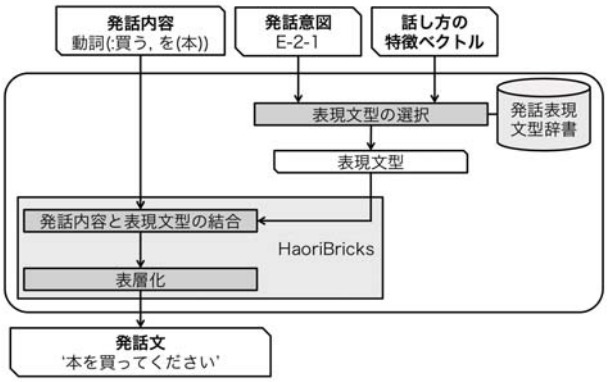


図 1: 発話文生成機構の構成

昨年，我々は，話者，発話意図，発話内容，の 3 つの入力から，話者の特徴を反映した表現（機能表現，文末表現，人称代名詞）を含む発話文を生成するシステムを試作した [刀山 17]．このシステムは，話者を変更することで話し方を変えることができるが，話し方は話者のみに依存するため，会話の相手や状況に応じて話し方を変えることができない．

この問題を解決するために，我々は，多様な話し方の実現基盤となる『発話表現文型辞書』[夏目 17] を増強するとともに，日本語文生成器 HaoriBricks [佐藤 17] を用いて発話文生成機構を再実装した．本稿では，その内容について報告する．

2. 発話文生成機構の概要

作成した発話文生成機構の構成を図 1 に示す．この機構は，発話内容，発話意図，話し方の特徴ベクトル，の 3 つの入力

表 1: 発話意図一覧表

ID	発話意図名	表現文型数	(未実装)
A	情報伝達		
A-1-1	説明-事情の提示	20	
A-2-1	説明-物	47	
A-3-1	説明-結果・帰結	15	
A-4-1	説明-換言	6	
A-5-1	説明-原因・理由	7	
A-6-1	言い訳	3	
A-7-1	伝聞	21	
A-8-1	引用	9	
B	認識表示		
B-1-1	願望-行為	14	
B-2-1	願望-物	9	
B-3-1	期待-事態	17	
B-6-1	後悔-非実行	14	
B-7-1	後悔-実行	16	
B-8-1	決心-実行	30	
B-9-1	決心-非実行	15	
B-10-1	希望-他者の動作	17	
B-11-1	意見	11	
B-12-1	推量-主観的	24	
B-13-1	推量-客観的	15	
B-14-1	確信	15	
B-15-1	可能性	8	
C	感情表出		
C-1-1	高評価 (A)	21	
C-1-2	高評価 (Na)	21	
C-2-1	低評価 (A)	21	
C-2-2	低評価 (Na)	21	
C-3-1	想定外 (A)	5	
C-3-2	想定外 (Na)	5	
C-3-3	想定外 (N)	5	
C-3-4	想定外 (V)	5	
D	態度表明		
D-1-1	褒める (A)	11	
D-1-2	褒める (Na)	12	
D-2-1	貶す (A)	19	
D-2-2	貶す (Na)	18	
D-3-1	非難-過失	21	
D-4-1	非難-行為-実行	15	
D-5-1	非難-行為-非実行	19	

ID	発話意図名	表現文型数	(未実装)
E	行為要求		
E-1-1	命令	18	
E-2-1	依頼-実行	32	
E-3-1	依頼-非実行	19	
E-4-1	勧誘-グループ型	20	
E-5-1	勧誘-引き込み型	12	
E-6-1	忠告	20	(3)
E-7-1	勧告	14	
E-8-1	許可	15	(4)
E-9-1	禁止	52	
E-10-1	勧め-行為	24	
E-11-1	勧め-物	11	
F	意向伺い		
F-1-1	申し出	23	(6)
F-2-1	提案	20	
G	質問		
G-1-1	yes-no 質問	15	
G-3-1	wh 質問	15	
G-4-1	確認-肯否要求 (N/Na)	7	
G-4-1	確認-肯否要求 (V/A)	10	
G-5-2	確認-未知情報要求	10	
G-6-1	確認-念押し	12	
G-8-1	許可要求（肯否要求）	10	
H	応答		
H-1-1	肯定	20	
H-2-1	承諾-依頼	22	
H-3-1	承諾-勧誘	14	
H-4-1	受諾-申し出	20	
H-5-1	否定	12	
H-6-1	断り-依頼	19	
H-7-1	断り-勧誘	15	
H-8-1	断り-申し出	15	
H-9-1	不明	23	
H-10-1	未定	12	
K	感動詞		
K-1-1	よびかけ	17	
K-2-1	別れ	15	
K-3-1	感謝	13	
K-4-1	謝罪	12	
S	接続表現		
S-1-1	接続表現-文頭-添加	6	
S-2-1	接続表現-文頭-逆説	4	
S-3-1	接続表現-文頭-話題転換	12	
S-4-1	接続表現-文末-ぼかし	6	
		1168	

から発話文を生成する．

我々は，発話文を，発話内容と発話意図から構成されるものと捉える．たとえば，

(1) 本を買ってください

という発話文を，

(2) a. 発話内容 =「本を買う」
b. 発話意図 =「依頼-実行」

のように捉える．このように発話内容と発話意図を分離するのは，「話し方の特徴は，主に発話意図の表層化において現れる」という観察結果に基づく．我々は，「〜てください・〜てよ・〜てくれ」などのある特定の意図を表す一連の表現文型（話し方）をグループ化・抽象化したものを発話意図とみなす．このグループの中のどの表現文型を使うかがそれぞれの話者によって異なり，それぞれの発話意図におけるこのような選択の総体が，話者の話し方の個性となると考える．

3. 発話表現文型辞書

発話文生成機構の中核をなすのは，『発話表現文型辞書 (Version 2)』である．この辞書は，「話者が，自分の言いたいことを表現する際に選択することができる言語形式のセットを提供すること」を目的とした辞書である．辞書のエントリは，「〜てください」のような表現文型である．

各エントリは，以下の 4 つの情報を持つ．

1. **ID**
   そのエントリに固有の ID.
2. デフォールトか否か
   同一発話意図内で最も典型的な表現文型には，デフォールトという情報が付与されている．
3. 表現文型文字列
   そのエントリ (表現文型) の文字列表記.
4. 話し方の特徴ベクトル
   この表現文型から感じられる話し方の特徴を表すベクトル.

エントリの ID は，ハイフォンで結合された 4 つの部分から構成される．そのうちの前の 3 つは，そのエントリの発話意図を表す発話意図 **ID** であり，先頭のアルファベットがその大分類を，後ろ 2 つの数字が詳細分類を表す．現在，大分類は 10 種類，発話意図 ID は全部で 74 種類定義されている (表 1)．発話意図 ID の最後の部分は，発話意図内の ID である．たとえば，E-2-1-12 という ID は，大分類 E(=行為要求)，詳細分類 2-1(=依頼-実行) という発話意図に属する 12 番目のエントリであることを意味する．本辞書には，総計で 1168 のエントリが収録されている．なお，複数の発話意図に対して同一の表現文型文字列が定義されている場合が存在する．

表 2: 発話表現文型辞書のエントリ例 (発話意図 E-2-1(=依頼-実行) の一部)

ID	表現文型文字列	話し方の特徴ベクトル						
		ジェンダー	年齢	長さ	強さ	待遇: 上	待遇: 下	印象
E-2-1-1	V-ろよ	20	00	00	00	0000	0001	0000
E-2-1-2	V-て	00	00	01	00	0000	0000	0000
E-2-1-3	V-てよ	00	00	00	00	0000	0100	0000
E-2-1-4	V-てくれ	10	00	01	00	0000	0000	0000
E-2-1-12	V-てくれないかしら？	01	00	20	01	0000	0100	0010
E-2-1-13*	V-てください	00	00	00	00	0001	0000	0000
E-2-1-14	V-てくださる？	02	01	10	00	0001	0000	0000
E-2-1-24	V-てほしいんだけど	00	00	10	01	0000	0011	0000
E-2-1-25	V-ていただけない？	01	01	10	00	0001	0000	0000

話し方の特徴ベクトルは，その表現文型の特徴を表す，以下に示すような 7 軸 20 成分のベクトルである．

1. ジェンダー (男性的，女性的)
2. 年齢 (子どもっぽい，大人っぽい)
3. 長さ (長い，短い)
4. 強さ (強い，弱い)
5. 上向き待遇表現 (上扱い，遠ざけ，あらたまり，丁寧)
6. 下向き待遇表現 (下扱い，親しみ，くだけ，ぞんざい)
7. 印象 (明るい，暗い，可愛い，古風)

各成分は，$\{0, 1, 2, 3\}$ のいずれかの値をとる．それらの値は次のことを意味する．

0　その特徴を持たない
1　その特徴を持つ
2　その特徴を強く持つ
3　その特徴を極めて強く持つ

なお,「話し方の特徴」は，その表現文型の 形式としての 特徴であり，その表現文型を用いる話者の特徴とは直接は対応しない．たとえば,「(本を読んで) くださる？」には「女性的=2」という特徴が付与されているが，それは，この発話の形式が「強く女性的」ということを意味するだけで,「話者は (かならず) 女性である」ことを意味するわけではない．

辞書のエントリの例を表 2 に示す．この表では，デフォールトの情報は，ID の直後に '*' を付与することによって示した．

本辞書と以前の辞書 [夏目 17] との最大の差異は，話し方の特徴ベクトルの成分を 20 に拡大したことである．特に，相手に応じて話し方を使い分けることを想定し，待遇表現 [日本 09] に関わる特徴を整理・増強した．これ以外にも，発話意図および表現文型も増強した．

4. 文生成器への実装

前節で述べた『発話意図表現文型辞書』の各エントリでは，表現文型が文字列として記述されているため，辞書をそのまま文生成器に組み込んでも，文法的に適切な発話文を生成できない．たとえば,「本を買う」に「V-てください」を結合するためには,「買う」をタ系連用テ形の「買って」に活用する必要があるが，その情報は「V-て」という形で間接的に記述されているにすぎない．さらに,「ください」がどのような形態素列であるかの情報も明示的には記述されていない．昨年の実装 [刀山 17] では，それぞれのエントリに対して，結合のための条件や結合する表現の文法的情報を個別に記述する方式をとったため，辞書を文生成器で利用できる形にするために，かなりの作業が必要であった．それに対し，今回の実装では，日本語文生成器 HaoriBricks [佐藤 17] を利用することで，簡略化を図った．

表 3: 表現文型のブロック化

表現文型	ブロック
V-ろよ	終助詞よ (命令形 (v))
V-てくれ	テくれ (v)
V-てください	テください (v)
V-ていただけない？	疑問符 (ない (テいただける (v)))

HaoriBricks は，あらかじめ用意されたブロック (関数) を組み合わせて，テキストを生成するコードを記述する．さらに，既存のブロックを組み合わせたマクロブロックを定義することができる．そのため，辞書の各エントリに対して，既存のブロックを対応させるか，あるいは，マクロブロックを定義すれば，そのブロックを発話内容 (を表すコード) に結合するだけで，簡単に表現文型の結合が実現できる．

表 3 に，エントリに対して定義したブロックの例を示す．ブロックの定義では，原則として，辞書の表現文型文字列をそのままブロック名とした (たとえば,「V-てください」). HaoriBricks には，標準的な日本語の文を組み立てるために必要なブロックが含まれている．たとえば，助詞を付与するブロック「終助詞よ」，補助動詞を付与するブロック「テください・テいただける」，活用形を変化させるブロック「命令形」などが用意されている．そして，それらのブロックには，文法的に正しい表層文を生成するための制約が組み込まれている．たとえば「テください」というブロックには，結合する述語をタ系連用テ形に変更する機能が組み込まれている．このため，大部分の表現文型は，既存のブロックをそのまま流用するか，あるいは，それらの組み合わせとして実現できる．

一方，話し言葉にみられる「くだけた」表現を生成するためのブロックは不足していたため，これらは新たに定義した．それらには，次のようなものがある．

- 終助詞「じゃ・ねえ・もん」などを付与するブロック
- 「母音の引き延ばし」を行うブロック

これらの終助詞は,「P のじゃ(A-1-1(=説明))」,「R たいねえ (B-1-1(=願望-行為))」,「P んだもん (A-6-1(=言い訳))」などで用いる.「母音の引き延ばし」は,「買いまーす」や「すごーい」のように，短母音が長母音に交替する変化である．これは，活用形の一種であるとみなし，助動詞です，接尾辞です，接尾辞ます，イ形容詞の活用形に，長音符号「ー」を含んだ「母音延長基本形」を追加した．

現時点では，1168 エントリのうち，1155 エントリに対してブロックが定義されている．実装を保留しているエントリ (表

表 4: HaoriBricks のコードと出力

	コード	出力
0	動詞 (:買う, を (:本))	本を買う
1	発話 ({id: 'B-1-1' }, 動詞 (:買う, を (:本)))	本を買いたいな
2	発話 ({id: 'B-1-1', fv: '10000100000000000000'}, 動詞 (:買う, を (:本)))	本を買いたいんだ
3	発話 ({id: 'B-1-1', fv: '02001001000000000000'}, 動詞 (:買う, を (:本)))	本を買いたいわ
4	発話 ({id: 'B-1-1', fv: '01000010000001100000'}, 動詞 (:買う, を (:本)))	本を買いたいねえ
5	発話 ({id: 'B-1-1', fv: '00100010000001101010'}, 動詞 (:買う, を (:本)))	本を買いたーい
6	発話 ({id: 'E-2-1' }, 動詞 (:買う, を (:本)))	本を買ってください
7	発話 ({id: 'E-2-1', fv: '10000100000000000000'}, 動詞 (:買う, を (:本)))	本を買ってくれ
8	発話 ({id: 'E-2-1', fv: '02001001000000000000'}, 動詞 (:買う, を (:本)))	本を買ってくださる？
9	発話 ({id: 'E-2-1', fv: '01011000101200000000'}, 動詞 (:買う, を (:本)))	本を買ってくださいませ
10	発話 ({id: 'E-2-1', fv: '01000000000010110000'}, 動詞 (:買う, を (:本)))	本を買ってちょうだい
11	発話 ({id: 'E-2-1' }, 動詞 (:する, を (:作業)))	作業をしてください

1 の「未実装」) には，以下のような，接頭辞「お・ご」を持つ表現文型のエントリである．

- お R なさい　　E-6-1(=忠告)
- お R ください　E-8-1(=許可)
- お R します　　F-1-1(=申し出)
- ご N します　　F-1-1(=申し出)

表 4 に，発話文を生成するコードと生成される文の例を示す.「発話」というブロック (関数) の第 1 引数に発話意図 ID と話し方の特徴ベクトルを与え，第 2 引数に発話内容を与える．この表に示した例は,「E-2-1(=依頼- 実行)/F-2-1(=提案)」という発話意図と,「本を買う/作業をする」という発話内容から発話文を生成している．1 行目から 5 行目に示すように，話し方の特徴ベクトルを変更することで，発話意図を変えずに，異なる表現の発話文を生成することができる．2 行目と 7 行目，3 行目と 8 行目に示すように，同じ特徴ベクトルを指定することで．異なる意図でも，同じような話し方の発話文を生成することができる．6 行目と 11 行目に示すように，発話内容を変更すれば，異なる内容の発話文を生成することができる．

発話文の生成は，以下の手順で行う．

1. エントリの検索: 引数で指定された発話意図 ID を持つエントリを検索する．
2. エントリの選択: 得られたエントリのうち，引数で指定された話し方の特徴ベクトルに最も近いベクトルを持つエントリを選択する．話し方の特徴ベクトルが指定されていない場合，デフォールトのエントリを選択する．
3. 結合と表層化: 得られたエントリに対して定義されているブロックを，第 2 引数の発話内容に結合したのち，それらを表層文字列化する．

引数で指定されたベクトル ($\boldsymbol{V_0}$) に最も近いベクトルは，以下のように決定する．

1. $\boldsymbol{V_0}$ に含まれる非 0 成分のみに着目して，2 つのベクトルを比較し，値が等しい成分が最も多いものを選ぶ．
2. 1 を満たすベクトルが複数存在する場合は，全ての成分に着目して，値が等しい成分が最も多いものを，その中から選ぶ．

このような選び方をするのは，話し方の特徴ベクトルの性質に基づく．ある表現文型が，ある特徴を持つときのみ，ベクトルの成分は非 0 の値をとるため，表 2 に示すように，多くの成分は 0 となる．そのため，$\boldsymbol{V_0}$ の成分のうち，値が 0 以外の成分に着目し，それらの成分が等しいものを優先する．

5. まとめ

小説の自動生成に利用可能な発話文生成システムを実現するため，その第一段階として，多様な発話文を生成できる機構を作成した．

今後の課題として，以下の 2 つが挙げられる．1 つ目は，辞書の増強である．現在収録されているエントリで話し方をどの程度カバーできているかを発話意図ごとに確認し，不足している表現を収集し追加する必要がある．2 つ目は，実装上の問題である．それぞれの発話表現には，結合可能な発話内容の形式が定まっている．たとえば,「V-てください」は，動詞にのみ結合でき，形容詞には結合できない．現在実装したブロックには，結合できない形式を排除する機能がないため，発話内容と表現文型が結合可能か判断する処理を追加する必要がある．

謝辞

本研究は JSPS 科学研究費挑戦的研究 (萌芽)「ブロック玩具をモデルとする日本語文章合成ツールキットの設計と実装」(課題番号 17K20028) の助成を受けている．

参考文献

[宮崎 15] 宮崎 千明, 平野 徹, 東中 竜一郎, 牧野 俊朗, 松尾 義博, 佐藤 理史：文節機能部の確率的書き換えによるキャラクタ性変換, 自然言語処理, 言語処理学会第 21 回年次大会発表論文集 (2015)

[水上 14] 水上 雅博, Neubig, G., Sakti, S., 戸田 智基, 中村 哲：特徴的話者を対象とした言語的個人性変換, 情報処理学会研究報告, 情報処理学会 (2014)

[夏目 17] 夏目和子, 刀山将大, 佐藤理史：発話文自動生成のための日本語表現文型辞書の設計, 自然言語処理, 言語資源活用ワークショップ (2017)

[日本 09] 日本語記述文法研究会：現代日本語文法 7 談話・待遇表現, くろしお出版 (2009)

[佐藤 17] 佐藤 理史：HaoriBricks: ブロック玩具に学ぶ日本語文章生成ライブラリ, 自然言語処理, 言語処理学会第 23 回年次大会発表論文集 (2017)

[刀山 17] 刀山 将大, 夏目 和子, 佐藤 理史, 松崎 拓也：話者の特徴を反映した発話文生成器の作成, 自然言語処理, 言語処理学会第 23 回年次大会発表論文集 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ランク学習によるYahoo!知恵袋の見出し生成

Headline Generation of Yahoo! Chiebukuro Based on Learning to Rank

日暮立 † 小林隼人 †‡ 村尾 一真 † 増山 毅司 †
Tatsuru Higurashi Hayato Kobayashi Kazuma Murao Takeshi Masuyama

† ヤフー株式会社 ‡ 理化学研究所 AIP センター
Yahoo Japan Cooporation RIKEN AIP

In this paper, we propose a training method for headline generation based on learning-to-rank, so that the appropriate summarization of a question posted by the user is distributed in Yahoo! Chiebukuro. For headline generation we used data provided from National Institute of Informatics and generated supervised-data through crowdsourcing. As a result, effective results were obtained compared with conventional methods.

1. はじめに

Web 上で配信されているテキストコンテンツには, 新聞記事やコラムをはじめとするニュースサイト以外にも掲示板や SNS など, ユーザーが投稿したコンテンツ (UGC: User Generated Content) がある. UGC の一つに, 質問したいユーザーと回答したいユーザーをむすび, お互いに知恵や知識を Q&A 方式で共有できる Q&A サイトがある. Q&A サイトでは, ユーザーに回答可能な質問を見つけてもらうために投稿された質問に見出しを付けて提示することが多いが, 新聞記事とは異なり, 必ずしもユーザーが適切な見出しを付けているとは限らない. また, ユーザーの投稿時の負担を減らすために投稿フォームに明示的に見出しの項目を設けていない Q&A サイトも多く, その場合には質問文の先頭部分が見出しの代用として用いられている. そのため, このような Q&A サイトでは見出しが質問内容を適切に表していないことが原因で, 未回答の質問が回答可能なユーザに見つけてもらえないという課題がある.

日本で多くのユーザーに利用されている Q&A サイトの一つに Yahoo!知恵袋 [*1] がある. 図 1 に, Yahoo!知恵袋の見出しと投稿質問文の例を示す. Yahoo!知恵袋では, ユーザーの入力の負荷が高いことから投稿時に見出しを入力させていないため, 前述のとおり投稿質問文の先頭から一部を切り出して見出しとして提供している. 投稿された質問文は, Web または Yahoo!知恵袋が提供しているスマートフォンアプリ [*2] を通じて配信される.

先頭切り出しの見出しが質問内容を適切に表していない場合があるという問題は, 特にスマートフォンアプリのように表示領域が限られている場合に顕著である. 図 2 に, 図 1 の内容をスマートフォンアプリの通知機能を利用して配信した場合の見出しの表示例を示す. 図 1 では, 投稿者の意図としては片思いの初期段階の恋愛相談をしているにもかかわらず, 図 2 のように先頭から部分的に切り出した見出しではユーザーに質問内容を適切に伝えることができていないことが分かる. このようにコンテンツの内容を正しくユーザーに伝えられていないことは, ユーザビリティ低下に繋がり, 利用者数に悪影響を及ぼす可能性がある. 逆に, 投稿質問文の内容を正しく反映した見出しを生成してユーザーにコンテンツを提供することができれば, ユーザーが見たい質問を容易に見つけられるようになり, 回答率の向上が期待できる.

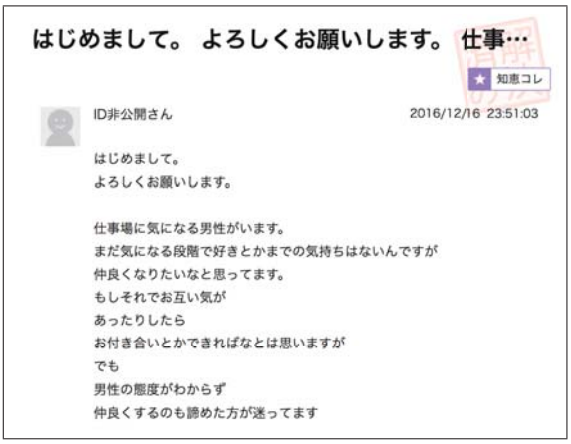


図 1: Yahoo!知恵袋の見出しと投稿質問文の例



図 2: Yahoo!知恵袋アプリの通知例

そこで本研究では, Yahoo!知恵袋の見出し生成タスクを投稿質問中の各文のランキング問題とみなし, ランク学習による見出し生成手法を提案する. その際, Yahoo!知恵袋のユーザーに最適な見出しを提供する目的で, クラウドソーシングを用いて各文の “見出しらしさ” のスコアを付与したデータセットを作成した. このデータセットを用いたランク学習モデルで生成し

連絡先: 日暮 立, ヤフー株式会社, thiguras@yahoo-corp.jp

∗1 http://chiebukuro.yahoo.co.jp/

∗2 https://itunes.apple.com/jp/app/id354626165

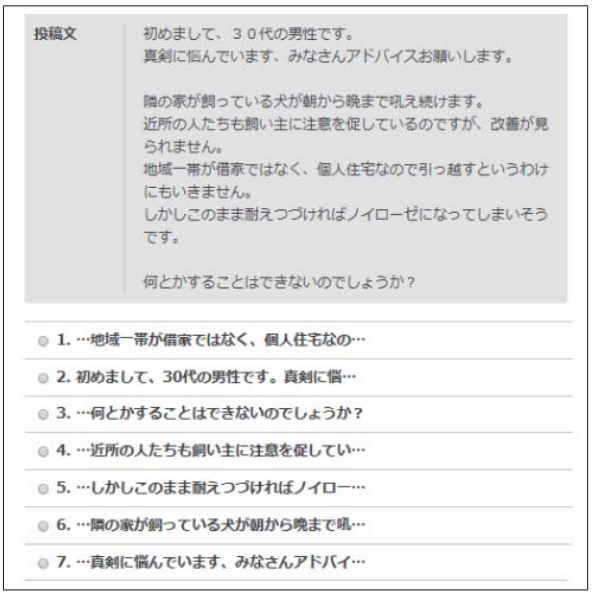


図 3: クラウドソーシングタスクの例

た見出しと, 現在 Yahoo!知恵袋で提供されている先頭切り出しや, 他の類似手法による見出しを比較評価したところ, 提案手法が最も有効であることが示された.

2. 関連研究

要約に関する研究は幅広くなされているが, その多くは新聞記事のようなある程度書式が統一されているデータに関するものであり, Q&A サイトのようにユーザーが自由に記述できるデータに対して要約技術を適用したものは少ない.

野口らは, 分散表現を用いて投稿質問文のスコアを定義し, ナップサック問題として定式化して動的計画法を解くことで Yahoo!知恵袋の要約を生成する教師なし学習の手法を提案し, 従来の TFIDF スコアを用いた抽出よりも良好な結果となることを示した [野口 15]. Bhaskar は, Q&A サイトにおいて, 複数文から成る質問文を, グラフを用いてクラスタリングした上でランキングし, ランクの最も高い文を要約文とする手法を提案した [Bhaskar 13]. Heilman と Smith は, 本文から質問文候補を文構造からルールベースで生成し, 質問文候補の教師あり学習によってランキングを学習する手法を提案している [Heilman 10]. しかし, これらの手法は質問文の要約タスクを対象としており, 本研究で扱っているような短い文字制約における見出し生成タスクについては検討していない.

我々のタスクに近い設定として, Ishigaki らは回答者の要旨理解のための長文質問の要約を生成的手法で行った [Ishigaki 17]. 彼らは, Yahoo! Answers のデータセットから投稿質問文を分析し, 投稿質問文と見出しを, 元文とその要約文の対とみなして教師データを抽出し, 生成型要約モデルを構築する方法を提案した. しかし, 生成型要約モデルは致命的な誤りを含む見出しを生成する可能性があるため, サービスに直接適用することは難しい.

3. 提案手法

本研究では, 見出しを持たない Q&A サイトを対象として, 教師あり学習を用いた抽出型の見出し生成手法を提案する. 教師あり学習を用いることでサービスに適した見出し生成モデルを構築することができ, 抽出型の手法を用いることでサービスの品質を維持することができる.

以下に, クラウドソーシングによる教師データの作成方法と, 作成したデータを用いた学習手法についてそれぞれ述べる.

3.1 データ作成

実験のためのデータ作成は Yahoo!クラウドソーシング *3 を利用して行った. データは国立情報学研究所がヤフー株式会社から提供を受けて研究者に提供している Yahoo!知恵袋のデータセット *4 を利用した. 以降, 知恵袋データと呼ぶ. 知恵袋データから投稿質問文を抽出し, 投稿質問文を文末表現 (「!」「?」「。」) で分割し, 分割した文の先頭 20 文字を見出し候補とした. 投稿質問文のうち, 上記の分割を行った場合に候補数が 6 以上となる投稿質問文 10,000 件を対象とした.

クラウドソーシングタスクの具体例を図 3 に示す. このタスクでは, 投稿質問文と, それに対する見出し候補となる選択肢が提示される. ワーカーには, 提示された投稿質問文を読んだ後, 他の投稿質問文と区別して内容が推測できそうな見出し候補を選択肢の中から一つ選ぶように指示をした. その際, 挨拶文, 自己紹介文や具体的な内容がない文を除くように補足説明も行った.

図 3 の例に対して, 各選択肢が見出しとして適切か不適切かを考える. 例題の投稿質問文から推測したい内容は「隣の家の犬の鳴き声に困っている」と考えられる. これを元に選択肢がそれぞれ適切かどうかの判断を行う. 選択肢 1, 選択肢 5 はペットに関する相談であることがわからないので不適切である. 選択肢 2, 選択肢 3, 選択肢 7 はユーザーへの挨拶文, 問いかけ文であり, 投稿質問文の内容に関する情報がないため不適切である. 選択肢 4 は「何かしらのペット」に関する内容であることが推測できるが, 具体的にどのようなペットなのか, またどのような相談なのかはわからないため, 不適切ではないが含まれる情報は不十分である. 選択肢 6 は 「隣の家の犬」に関する相談であることが推測できるため適切である. 以上の判断から選択肢 6 が最も適切な選択肢となる.

ワーカーには 1 タスクあたり, チェック設問（不適切に回答するワーカーを除外するための設問）を含めた 11 問を回答してもらった. 投票のポジションバイアスを除くために設問ごとに選択肢をランダムに並べ替えた. クラウドソーシングに掲載したタスク数は 10,000 タスクで, 1 タスクあたり 10 人のワーカーにタスクを依頼した. 1 質問あたり 10 人のワーカーが各々が最適と判断した選択肢を一つ選んでいるので, 各選択肢は 0 から 10 の投票スコアが与えられる.

3.2 学習方法

本研究では, クラウドソーシングで作成したデータを利用して, 投票スコアの高い選択肢の見出しを抽出する手法を提案する. クラウドソーシングで作成したデータの投票スコアの予測ではなく, 相対的に投票スコアの高い見出しを予測するという目的から, 本研究ではランク学習を利用する.

ランク学習は, 検索ランキングの最適化によく使われる手法であり, 検索クエリ q に対する文書 x とその適合度 y の 3 つ組の集合が与えられたときに, 検索結果を提示するための最適なランキングを学習することができる. 本研究においては, 投稿質問文を q, その見出し候補の選択肢を x, 各選択肢の投票スコアを y に対応付けて定式化を行う.

*3 http://crowdsourcing.yahoo.co.jp/
*4 http://www.nii.ac.jp/dsc/idr/yahoo/yahoo.html

本研究ではランク学習でよく使われる Pairwise 法を採用した. クラウドソーシングで作成したデータ D は, 投稿質問文 q, 選択肢 x, 各選択肢に対する投票スコア y の3つ組 $(x, y, q) \in D$ で表されるものとする. Pairwise 法では, データ D 中の3つ組のペア $((x_i, y_i, q_i), (x_j, y_j, q_j))$ について, $q_i = q_j$ となる同じ投稿質問中の, $y_i > y_j$ となる順位付けられた選択肢に基づいて学習を進める. 具体的には, まず上記を満たすようなデータ D 中の ID のペアを $P = \{(i, j) \mid q_i = q_j, y_i > y_j, (x_i, y_i, q_i) \in D, (x_j, y_j, q_j) \in D\}$ で定義する. この P を用いて, 下記で定義される最適化問題を解くことでランク学習が実現される.

$$\min_w \ \frac{1}{2} w^\top w + C \sum_{(i,j) \in P} \ell(w^\top \tilde{x}_i - w^\top \tilde{x}_j) \tag{1}$$

ここで, w は学習される重みベクトル, $\tilde{x}_i$ は選択肢 x から抽出した特徴量, C は正規化パラメータである. ℓ は二乗ヒンジ損失であり, $\ell(d) = \max(0, 1 - d)^2$ で定義される.

4. 実験

4.1 実験設定

本実験では, クラウドソーシングで作成したデータで, 投稿質問文ごとの各選択肢に投票した数が多い順に順位をラベル付けした. クラウドソーシングで作成したデータ 10,000 件を学習セット 9,000 件と評価セット 1,000 件に分割した. 学習セット 9,000 件で投稿質問文ごとに各選択肢の順位を予測するランク学習を実行した. 学習では, 判定対象文とその前後の文に対して, Bag-of-Words (BoW) ベクトル, Doc2Vec ベクトル, 文長に関するベクトルを計算し, 9 つのベクトルを結合し特徴量とした. BoW ベクトルは, 文を形態素解析して抽出した名詞・動詞・感動詞・接続詞・副詞・形容詞の TFIDF スコアに基づく 30,820 次元のベクトルである. Doc2Vec ベクトルは, 文の意味の特徴を表したものであり, Yahoo!知恵袋のデータセットの投稿質問文 3,092,958 文を, gensim[*5] の doc2vec を用いて学習させたモデルで, 文を 100 次元のベクトルに変換した. 文長ベクトルは, 対象文の文字数をベクトルで表現したものである.

学習には RankSVM[*6] を使用した. 学習における正規化パラメータ C は, 交差確認でパラメータ探索を行い, 精度が最も高くなった $C = 0.125$ を採用した.

4.2 比較手法

実験では, ランク学習と以下の見出し生成手法との比較実験を行なった.

- **ランダム抽出**: クラウドソーシングの選択肢をランダムで一つ選択して見出しとした.
- **TFIDF 重要文抽出**: 各選択肢を形態素解析して抽出した名詞・動詞・感動詞・接続詞・副詞・形容詞の TFIDF スコアの合計が最も高い選択肢を見出しとして選択した.
- **TFIDF 類似文抽出**: 投稿質問文と各選択肢をそれぞれ形態素解析して抽出した名詞・動詞・感動詞・接続詞・副詞・形容詞の TFIDF スコアに基づく 31,990 次元のベクトル間のコサイン類似度を計算し, 投稿質問文と最も類似度が高い選択肢を見出しとして選択した.
- **Doc2Vec 類似文抽出**: gensim の doc2vec を利用して, 投稿質問文と各選択肢を 100 次元のベクトルに変換してコサイン類似度を計算し, 投稿質問文と最も類似度が高い選択肢を見出しとして選択した.
- **辞書ベース先頭削除**: Yahoo!知恵袋のデータセットにおいて, 出現頻度の高い挨拶文 (「おはようございます」「こんにちは」「はじめまして」など) や, 知恵袋の投稿質問文の先頭文として出現頻度の高い定型文 (「質問です」「教えて」「お願いします」など) を人手で 913 文抽出し, それらを投稿質問文から除外したうえで, 先頭 20 文字を切り出して見出し生成を行なった.

4.3 評価尺度

本研究では, 見出し生成のオフライン精度評価を行うための3つの指標を定義する. 以下では, 現行の Yahoo!知恵袋の投稿質問文の先頭切り出し見出しに対して, 各手法で生成した見出しが適切かどうかを評価する.

4.3.1 見出し変化率

先頭切り出しによる見出し以外の見出し候補（具体的には2文目以降から生成された見出し候補）が選ばれた場合, その見出し候補は各手法によって適切だと予測されたものだと判断できる. この予測が正しければ, 先頭切り出しによる見出しを変化させた割合が多い方が良い手法といえる. そこで, 見出し生成が行われて先頭切り出し見出しが変化した割合を評価指標のひとつとして, 以下で定義する.

$$\text{見出し変化率} = \frac{\text{先頭切り出し見出しが変化した質問数}}{\text{投稿質問文全て}} \tag{2}$$

4.3.2 勝率

先頭切り出しの見出しと比較して, 各手法が選択した見出し候補がクラウドソーシングの投票スコアの順位で上位であれば, その手法の予測は正しいとみなせる. そこで, 各手法による先頭切り出しに対する勝率を評価指標のひとつとして, 以下で定義する.

$$\text{勝率} = \frac{\text{生成見出しの投票スコアが最も高い質問数}}{\text{先頭切り出し見出しが変化した質問数}} \tag{3}$$

4.3.3 平均順位

生成された見出しが, クラウドソーシングの投票スコアで何位であるかの平均を各手法の最終的な評価指標として用いる. 平均順位は低いほど, より質の高い見出しを生成できたことを意味している.

5. 結果と考察

5.1 クラウドソーシングデータ分析

先頭切り出しの見出しの投票スコアが最も高い場合は見出しは変更しない方が良いと判断し, 先頭切り出し以外の選択肢の投票スコアが最も高い場合は, 見出しは変更した方が良いと判断する. この基準でクラウドソーシングの結果を分析したところ 38.1%の投稿質問文の見出しは, 先頭切り出しの見出しから別の見出しに変更した方が良いという結果になった. 従って, 全質問文に対して理想的に見出しが変更された場合の最適見出し変化率は 38.1%と言える. 先頭切り出しの見出しを選択した場合の平均順位は 1.582 であった.

クラウドソーシングの選択肢において, 挨拶文や自己紹介文が含まれる選択肢が選ばれることは少なかったが, 「教えてく

*5 https://radimrehurek.com/gensim/models/doc2vec.html
*6 https://www.csie.ntu.edu.tw/c̃jlin/libsvmtools/

表 1: 先頭切り出しと提案手法による見出しの生成例

	先頭切り出しによる見出し	提案手法による見出し
1	情緒不安定なものです。回答して下さった…	…新婚旅行は、行きは別々で行き、現地で…
2	カテ違いならゴメンナサイ。今、財布が破…	…今、財布が破れてツライ状況です。新し…
3	現在、私の父は 60 歳で定年退職を迎えま…	…厚生年金の支払いは何歳までなんですか…
4	前にとても大変で嫌な思いをしたので携帯…	…携帯メールお断りというのはこの落札者…
5	27 歳女です。環境的になかなか新しい出…	…環境的になかなか新しい出会いがなくて…

表 2: 見出し生成の評価結果

	見出し変化率 (%)	勝率 (%)	平均順位
先頭切り出し	0	-	1.582
ランダム抽出	85.9	11.7	2.784
TFIDF 重要文抽出	81.1	12.7	2.635
TFIDF 類似文抽出	79.3	18.7	2.396
Doc2Vec 類似文抽出	**88.2**	13.0	2.731
辞書ベース先頭削除	2.2	72.0	1.577
ランク学習	9.9	**94.9**	**1.569**

ださい」などの定型文を含む選択肢は 1 人または 2 人が選んでいることがあった. 説明文に記載している「具体的内容がない文」についてワーカーにより判断のばらつきがあったとみられる. クラウドソーシングにおいて, 手作業で全てのルールを網羅することは難しいので, ワーカーの主観により判断のばらつきが生じる場合があることを念頭におく必要がある.

5.2 定性評価

表 1 にランク学習によって見出しを生成した場合の例を示している. 生成した見出しの例 1, 例 2, 例 5 では先頭切り出しの見出しには「情緒不安定なものです」や「カテ違いならゴメンナサイ」や「27 歳女です」などの文が含まれていたが, 提案手法によって生成された見出しではそれらが除外されて表示されるようになり, より質問内容を推測しやすくなっている. また, 生成した見出しの例 3, 例 4 では, 先頭切り出しの見出しと比べて, 提案手法によって生成した見出しは, より内容の具体性が増し, 質問内容が推測しやすくなっている.

5.3 オフライン評価

オフライン評価として, ランク学習と比較手法で生成した見出しを, 4.3 節で定義した評価尺度を用いて比較した. 比較結果を表 2 に示す. ランダム抽出, 重要文抽出, 類似文抽出による手法では, 見出し変化率がクラウドソーシングのデータ分析から得られた最適見出し変化率 38.1%を大きく上回り, 適切な見出しを変化させて不適切な見出しに変化させていることがわかった. そのため勝率も 10%台という低い数値となっており, 平均順位も先頭切り出しの見出しを選んだ場合の 1.582 より悪い結果となっている.

一方で, 辞書ベース先頭削除による見出し生成の結果においては, 見出し変化率は 2.2%とカバレッジは高くないが, 先頭切り出しの見出しと比較して平均順位は 1.577 となり, 辞書ベース先頭削除で行う見出し生成は有効であることが示された. 見出し変化率を向上させるためには人手でより網羅性の高い辞書を作成する必要があるが, 人手で網羅するにはコストがかかり非現実的である.

ランク学習による見出し生成は, 勝率および平均順位の両方において比較手法の中で最も良い結果となった. 先頭切り出しよりも平均順位が良い結果となった辞書ベース先頭削除とランク学習の 2 つの手法の見出し変化率を比較したところ, ランク学習の方が見出し変化率が高くなっており, 高いカバレッジで適切な見出しを提供できることが示された.

提案手法であるランク学習の勝率と平均順位については, その他の比較手法それぞれに対して Wilcoxon 検定を実施し統計的に有意 ($p < 0.01$) であることを確認した.

6. まとめ

本研究では, クラウドソーシングを利用して Yahoo!知恵袋の見出しの教師データを作成し, それを用いたランク学習による見出し生成に取り組んだ. ランク学習による見出し生成手法は, 他の見出し生成手法と比較して質の高い見出しが生成できることがわかった. また, 人手で作成した辞書を利用した生成手法と比較して見出し変化率が高くなることが示され, サービスに反映する際のカバレッジをより高められることを確認した.

今後は, ランク学習の他にも深層学習を取り入れた手法の検討を行なっていく予定である. また, ランク学習によって生成した見出しのユーザーの回答率やクリック率への影響をオンラインで検証していくことも検討している. 本研究においてクラウドソーシングを利用して作成した投票のラベルは公開する予定である.

参考文献

[Bhaskar 13] Bhaskar, P.: Answering Questions from Multiple Documents – the Role of Multi-Document Summarization, in *RANLP 2013*, pp. 14–21 (2013)

[Heilman 10] Heilman, M. and Smith, N. A.: Good Question! Statistical Ranking for Question Generation, in *NAACL 21010*, pp. 609–617 (2010)

[Ishigaki 17] Ishigaki, T., Takamura, H., and Okumura, M.: Summarizing Lengthy Questions, in *IJCNLP 2017*, pp. 792–800 (2017)

[野口 15] 野口 正樹, 谷塚 太一, 小林 隼人：分散表現を用いたヤフー知恵袋の要約, 言語処理学会第 21 回年次大会 (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

TULIP: Web 小説を学習に用いた三段階 LSTM による台本形式小説 (SS) 生成

TULIP: Generation of Screenplay-style Novel (SS) Using Three-step LSTM Trained on Web Novels

鈴木 惇 *1
Atsushi Suzuki

*1東京大学 大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

This paper proposes a new method to generate screenplay-style novels (SS: side story or short story), mainly written on the web. The key idea is to incorporate each talker's utterance-style into existing neural conversation model. Our model, three-step unified LSTM interlocution producer (TULIP), consists of unified three LSTMs: utterance-encoding LSTM, context-updating LSTM, and utterance-decoding LSTM. We trained our model on screenplay-style novels on the web, and generated new novels from scratch, in other words, not by direct processing of existing novels. We evaluated novels generated by our model both in quality and quantity, and confirmed that our LSTM based method generated a natural Japanese sentence sequence. We also indicate the limitation of existing quantitive evaluation methods.

1. 序論

本研究は、主に Web 上で投稿されている、台本形式小説 (SS: ショートストーリーもしくはサイドストーリーの略とされる) を自動生成することを目的とする。台本形式小説とは、登場人物の発話のみにより構成され、かつ各々の発話の発話者が明示された小説である。図 1 に一例 *1 を示す。台本形式小説は、匿名掲示板に多く投稿され、その一部は編集されて複数の「まとめサイト」に複製保管される。まとめサイトは一般的に広告収入をもとに運営されており [Kashiwabara 12]、台本形式の小説の自動生成は、それ自体が娯楽の提供として意義があるだけでなく、広告収入源としての需要も考えられる。台本形式小説生成は、ゲームのセリフ制作の補助も可能であり、その点からも需要が考えられる。

一般的な小説を自動生成する試みは、従来、元となる小説や文をベースに一部を加工することで新たな小説を生成するという方法で行われてきた [Takagi 14] [Jain 17]。しかしながらこの方法では、新たに生成される小説の枠組みが元となる小説に大幅に制約される。また、一般的な小説を扱う場合、元となる小説の各発話の発話者を特定する必要があり、その性能に出力の質が依存する。さらに、出力した小説の扱い方によっては、元となる小説の著作権を侵害する可能性が非常に高い。このような事情から、小説を生成する際は、特定の小説をベースとするのではなく、一から行われることが望ましい。

他方、台本形式小説は、登場人物の対話の連続ととらえることができる。従って、対話自動生成の手法を台本形式小説生成に応用することが考えられる。対話の自動生成の技術は、近年、深層学習の発展に伴い、急速に向上している [Serban 16] [Ouchi 16] [Li 16]。これらは必ずしも長い文脈や登場人物の口調の一貫性を考慮したものではないが、これらを拡張することで、質の高い台本形式小説を一から自動生成することが可能になると考えられる。

このような背景の下、本研究では、深層学習を用いた対話自動生成を、登場人物の口調を考慮したモデルに拡張し、台本形式小説生成用のモデル three-step unified LSTM interlocution producer (TULIP) を提案する。本研究のモデルは、投稿された台本形式小説を学習した後、生成段階では学習データを直接用いず、学習されたパラメータのみを用いて新たな小説を生成する。従って、本研究は著作権を侵害しない小説生成の可能性を開く。

本研究はさらに、生成された小説の質を、出力文の精査による定性的評価のほか、翻訳で用いられる定量評価指標で評価を行い、性能を示す。さらに評価指標の問題点にも触れる。

早苗「楓ちゃ～ん？」
楓「はい」
早苗「言いたいことはそれだけかしら？」
楓「そうですね……まるで洗いたて、干したてのタオルケットを触っているようでした」
早苗「どうも反省していないみたいね」
あい「早苗さん、落ち着こうか。振り上げたその一升瓶を一度下ろして」
楓「私の命はあいちゃんに託されているから。ね？」
あい「それは楓さんの態度次第さ」
楓「あら、そうなの？」
あい「……いま一度、自分が置かれている状況を把握するべきだよ」

図 1　台本形式小説の一例: 台本形式小説では、すべての発話の発話者が明示されているのが特徴である。

2. 手法: Three-step Unified LSTM Interlocution Producer (TULIP)

2.1 概要

本節では本研究で提案するモデル three-step unified LSTM interlocution producer (TULIP) の構造を記述する。TULIP は、Long Short Time Memory (LSTM) [Hochreiter 97] を用いるものであり、発話符号化 LSTM と文脈更新 LSTM、そして発話復号化 LSTM の三段階の LSTM を用いて、次の単語、および話者を再帰的に生成する。TULIP は、Hierarchical recurrent encoder-decoder (HRED) [Sordoni 15], [Serban 16] に話者の概念を加え拡張したものととらえることができる。

2.2 記法

準備として、N 次元確率ベクトルの集合 $\mathcal{P}_{(N)}$ を $\mathcal{P}_{(N)} \stackrel{\text{def}}{=} \left\{ \boldsymbol{p} \;\middle|\; \boldsymbol{p} \in [0,1]^N, \sum_{n=1}^{N} [\boldsymbol{p}]_n = 1 \right\}$ で定義する。ただし、ここで、数ベクトル $\boldsymbol{v}$ に対し $[\boldsymbol{v}]_n$ は $\boldsymbol{v}$ の n 番目の要素を表す。N 次元 onehot ベクトルの集合を $\mathcal{E}_{(N)} \subset \mathcal{P}_{(N)}$

連絡先: 鈴木 惇: 東京大学 atsushi.suzuki.rd@gmail.com

*1 楓「藍子ちゃんのお腹を触っていたら夜になっていた」 http://ex14.vip2ch.com/test/read.cgi/news4ssnip/1463151146/

を $\mathcal{E}_{(N)} \stackrel{\text{def}}{=} \left\{ \boldsymbol{e} \;\middle|\; \boldsymbol{e} \in \{0,1\}^N, \sum_{n=1}^N [\boldsymbol{e}]_n = 1 \right\}$ で定義する。ただし、ここで、数ベクトル $\boldsymbol{v}$ に対し $[\boldsymbol{v}]_n$ は $\boldsymbol{v}$ の n 番目の要素を表す。また、$\boldsymbol{e}_{(N,n)} \in \mathcal{E}_{(N)}$ を第 n 要素が 1, それ以外が 0 となる onehot ベクトルとする。さらに、softmax 関数 $a_{\text{SM}} : \mathbb{R}^N \to \mathcal{P}_{(N)}$ を $[a_{\text{SM}}(\boldsymbol{x})]_n \stackrel{\text{def}}{=} \frac{\exp(x_n)}{\sum_{n'=1}^N \exp(x_{n'})}$ で、cross entropy $\ell_{\text{CE}} : \mathcal{E}_{(N)} \times \mathcal{P}_{(N)} \to \mathbb{R}$ を $\ell_{\text{CE}}(\boldsymbol{e}, \boldsymbol{p}) \stackrel{\text{def}}{=} -\sum_{n=1}^N [e_n \log p_n + (1-e_n)\log(1-p_n)]$ で定義する。

最初に、空白を表す特殊単語 [pad]、発話の始まりを表す特殊単語 [go]、未知語を表す特殊単語を用意し [unk]、それぞれ $1,2,3$ と番号を振る。また、出現頻度上位 $I-3$ 位までの単語に $4,5,\ldots,I$ と番号を振る。そして、i 番の単語を $\boldsymbol{e}_{(I,i)}$ と同一視する。出現頻度 $I-2$ 位以下の単語は $\boldsymbol{e}_{(I,3)}$ と同一視する。

発話者についても $1,2,\ldots,M$ と番号を振り、m 番の発話者を $\boldsymbol{e}_{(M,m)}$ と同一視する。学習または出力する小説の発話数を L とする。ℓ 番目の発話 $(m = 1,2,\ldots,M)$ の発話者を $\boldsymbol{e}_\ell^{(\text{t})} \in \mathcal{E}_{(M)}$、$\ell$ 番目の発話内容の j 番目の単語を $\boldsymbol{e}_{\ell,j}^{(\text{w})} \in \mathcal{E}_{(I)}$ とする。この記法の下、ℓ 番目の発話は発話者 $\boldsymbol{e}_\ell^{(\text{t})}$ と発話内容 $\boldsymbol{E}_\ell^{(\text{w})} \stackrel{\text{def}}{=} \left[\boldsymbol{e}_{\ell,1}^{(\text{w})} \quad \boldsymbol{e}_{\ell,2}^{(\text{w})} \quad \cdots \quad \boldsymbol{e}_{\ell,J}^{(\text{w})}\right]$ の組 $\left(\boldsymbol{e}_\ell^{(\text{t})}, \boldsymbol{E}_\ell^{(\text{w})}\right)$ で表現される。なお、元の発話文の長さ J' は一般に J に満たないため、$J'+1$ 番目から J 番目まで [pad] を挿入する。

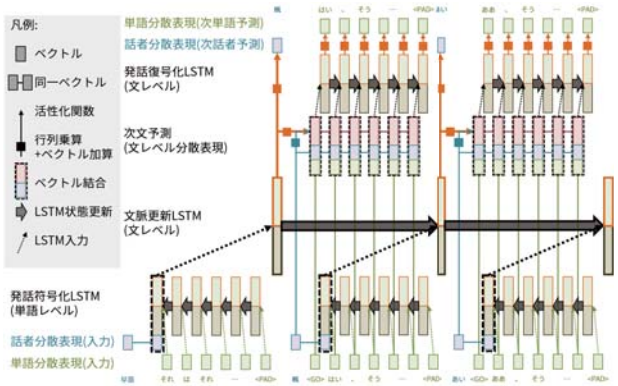


図 2　TULIP の概要: 下部が発話符号化 LSTM, 中部が文脈更新 LSTM, 上部が発話復号化 LSTM である。

2.3　モデル

2.3.1　入出力関係

本研究で用いるモデルは以下の入出力関係を可能にする。

a) 完全な発話列から次の発話者の出力

1 番目から L' 番目までの完全な発話の列 $\left(\boldsymbol{e}_1^{(\text{t})}, \boldsymbol{E}_1^{(\text{w})}\right), \left(\boldsymbol{e}_2^{(\text{t})}, \boldsymbol{E}_2^{(\text{w})}\right), \ldots, \left(\boldsymbol{e}_{L'}^{(\text{t})}, \boldsymbol{E}_{L'}^{(\text{w})}\right)$ の入力に対し、$L'+1$ 番目の発話者 $\boldsymbol{e}_{L'+1}^{(\text{t})}$ の確率に対応する確率ベクトル $\boldsymbol{p}_{L'+1}^{(\text{t}')}$ を出力する。

b) 生成中の発話の次の単語の出力

1 番目から L' 番目までの完全な発話の列 $\left(\boldsymbol{e}_1^{(\text{t})}, \boldsymbol{E}_1^{(\text{w})}\right), \left(\boldsymbol{e}_2^{(\text{t})}, \boldsymbol{E}_2^{(\text{w})}\right), \ldots, \left(\boldsymbol{e}_{L'}^{(\text{t})}, \boldsymbol{E}_{L'}^{(\text{w})}\right)$ に $L'+1$ 番目の発話の J' 番目までの単語列 $\left[\boldsymbol{e}_{L'+1,1}^{(\text{w})} \quad \boldsymbol{e}_{L'+1,2}^{(\text{w})} \quad \cdots \quad \boldsymbol{e}_{L'+1,J'}^{(\text{w})}\right]$ を加えた入力に対し、J' 番目の単語 $\boldsymbol{e}_{L'+1,J'+1}^{(\text{w})}$ の確率に対応する確率ベクトル $\boldsymbol{p}_{L'+1,J'+1}^{(\text{w}')}$ を出力する。なお、ここには $J'=0$ の場合、すなわち、$L'+1$ 番目の発話について、発話者のみが与えられ、発話内容の単語が 1 つも入力されていない場合も含まれる。

このモデルを再帰的に適用することで台本形式小説を生成することができる。以下、入出力関係の詳細を記述する。ただし活性化関数への入力についてはバイアス項を省略する。

2.3.2　過去の発話の分散表現の出力

まず、パラメータ $\boldsymbol{K}^{(\text{t})} \in \mathbb{R}^{D^{(\text{t})} \times M}, \boldsymbol{K}^{(\text{w})} \in \mathbb{R}^{D^{(\text{w})} \times I}$ を用いて話者、および単語を $\boldsymbol{t}_\ell = \boldsymbol{K}^{(\text{t})} \boldsymbol{e}_\ell^{(\text{t})} \in \mathbb{R}^{D^{(\text{t})}}$, $\boldsymbol{w}_{\ell,j} = \boldsymbol{K}^{(\text{w})} \boldsymbol{e}_{\ell,j}^{(\text{w})} \in \mathbb{R}^{D^{(\text{w})}}$ と分散表現に変換する。$\ell = 1,2,\ldots,L'$ の時、発話内容 $\boldsymbol{E}_\ell^{(\text{w})}$ は発話符号化 LSTM に入力され、1 つの発話内容が 1 つの分散表現 $\boldsymbol{u}_\ell \in \mathbb{R}^{D^{(\text{u})}}$ に変換される:

$$\left(\boldsymbol{h}_{\ell,j}^{(\text{e})}, \boldsymbol{m}_{\ell,j}^{(\text{e})}\right) = \begin{cases} \left(\boldsymbol{h}_J^{(\text{e})}, \boldsymbol{m}_J^{(\text{e})}\right) & (j = J) \\ g_{\text{LSTM}}^{(\text{e})}\left(\boldsymbol{w}_{\ell,j}, \boldsymbol{h}_{\ell,j+1}^{(\text{e})}, \boldsymbol{m}_{\ell,j+1}^{(\text{e})}; \boldsymbol{\Theta}^{(\text{e})}\right) & (j = 0,1,\ldots,J-1) \end{cases}. \tag{1}$$

また $\boldsymbol{u}_\ell = \boldsymbol{m}_{\ell,0}^{(\text{e})}$. ただし、ここで $\boldsymbol{h}_J^{(\text{e})}, \boldsymbol{m}_J^{(\text{e})} \in \mathbb{R}^{D^{(\text{u})}}$ はあらかじめ定められた定数ベクトルであり、$j = 0,1,\ldots,J$ に対し、$\boldsymbol{h}_{\ell,j}^{(\text{e})}, \boldsymbol{m}_{\ell,j}^{(\text{e})} \in \mathbb{R}^{D^{(\text{u})}}$ である。また、$g_{\text{LSTM}}^{(\text{e})} : \mathbb{R}^{D^{(\text{w})}} \times \mathbb{R}^{D^{(\text{u})}} \times \mathbb{R}^{D^{(\text{u})}} \to \mathbb{R}^{D^{(\text{u})}} \times \mathbb{R}^{D^{(\text{u})}}$ は発話符号化 LSTM の入出力関係を表す写像であり、$\boldsymbol{\Theta}^{(\text{e})}$ は発話符号化 LSTM のパラメータである。発話符号化 LSTM への入力は、通常の文の単語順の逆順であることに注意されたい。この逆順入力のアイデアは [Sutskever 14] によるものである。

2.3.3　次の発話の分散表現の出力

発話内容の分散表現 $\boldsymbol{u}_\ell$ は、話者の分散表現 $\boldsymbol{t}_\ell$ と結合され、文脈更新 LSTM に入力される。$\ell = 1,2,\ldots,L$ に対し、以下の出力が得られる:

$$\left(\boldsymbol{h}_\ell^{(\text{c})}, \boldsymbol{m}_\ell^{(\text{c})}\right) = g_{\text{LSTM}}^{(\text{c})}\left(\begin{bmatrix}\boldsymbol{u}_\ell \\ \boldsymbol{t}_\ell\end{bmatrix}, \boldsymbol{h}_{\ell-1}^{(\text{c})}, \boldsymbol{m}_{\ell-1}^{(\text{c})}; \boldsymbol{\Theta}^{(\text{c})}\right). \tag{2}$$

ただし、ここで $\boldsymbol{h}_0^{(\text{c})}, \boldsymbol{m}_0^{(\text{c})} \in \mathbb{R}^{D^{(\text{c})}}$ はあらかじめ定められた定数ベクトルであり、$\ell = 0,1,\ldots,L$ に対し、$\boldsymbol{h}_\ell^{(\text{c})}, \boldsymbol{m}_\ell^{(\text{c})} \in \mathbb{R}^{D^{(\text{c})}}$ である。また、$g_{\text{LSTM}}^{(\text{c})} : \mathbb{R}^{D^{(\text{u})}+D^{(\text{t})}} \times \mathbb{R}^{D^{(\text{c})}} \times \mathbb{R}^{D^{(\text{c})}} \to \mathbb{R}^{D^{(\text{c})}} \times \mathbb{R}^{D^{(\text{c})}}$ は発話符号化 LSTM の入出力関係を表す写像であり、$\boldsymbol{\Theta}^{(\text{c})}$ は発話符号化 LSTM のパラメータである。

ここで得られた文脈更新 LSTM の出力 $\boldsymbol{m}_{L'}^{(\text{c})}$ が、以降で述べるように $L'+1$ 番目の発話の発話者と発話内容を決定する。

2.3.4　次の発話者の出力

L' 番目までの完全な発話が入力として与えられている場合、$L'+1$ 番目の発話者については、パラメータ $\boldsymbol{K}^{(\text{t}')} \in \mathbb{R}^{D^{(\text{t})} \times D^{(\text{c})}}, \boldsymbol{L}^{(\text{t}')} \in \mathbb{R}^{M \times D^{(\text{t})}}$ と活性化関数 $a^{(\text{t}')}$ によって、

$$\boldsymbol{p}_{L'+1}^{(\text{t}')} = a_{\text{SM}}\left(\boldsymbol{L}^{(\text{t}')} a^{(\text{t}')}\left(\boldsymbol{K}^{(\text{t}')} \boldsymbol{m}_{L'}^{(\text{c})}\right)\right) \tag{3}$$

と確率ベクトル $\boldsymbol{p}_{L'+1}^{(\text{t}')}$ が与えられる。

2.3.5　発話内容の出力

L' 番目までの完全な発話に加え、既に得られた $L'+1$ 番目の発話の発話者 $\boldsymbol{e}_{L'+1}^{(\text{t})}$ と発話内容の J' 番目までの単語が入力として与えられている場合、$J'+1$ 番目の単語については、以下のように確率ベクトルが得られる。$\boldsymbol{e}_{L'+1}^{(\text{t})}$ の分散表現 $\boldsymbol{t}_{L'+1} = \boldsymbol{K}^{(\text{t})} \boldsymbol{e}_{L'+1}^{(\text{t})}$ と文脈更新 LSTM の出力 $\boldsymbol{m}_{L'}^{(\text{c})}$ を結合し、パラメータ $\boldsymbol{K}^{(\text{u}')} \in \mathbb{R}^{D^{(\text{u})} \times \left(D^{(\text{c})}+D^{(\text{t})}\right)}$ の乗算と活性化関数

$a^{(\mathrm{u}')}$ をの適用により、$L'+1$ 番目の発話の発話内容の分散表現 $\boldsymbol{v}_{L'+1}$ が次のように得られる:

$$\boldsymbol{v}_{L'+1} = a^{(\mathrm{u}')}\left(\boldsymbol{K}^{(\mathrm{u}')}\begin{bmatrix}\boldsymbol{m}^{(\mathrm{c})}_{L'}\\ \boldsymbol{t}_{L'+1}\end{bmatrix}\right) \tag{4}$$

この $\boldsymbol{v}_{L'+1}$ と、話者の分散表現 $\boldsymbol{t}_{L'+1}$、および既に得られた単語 $\boldsymbol{e}^{(\mathrm{w})}_{L'+1,j}$ の分散表現 $\boldsymbol{w}_{L'+1,j} = \boldsymbol{K}^{(\mathrm{w})}\boldsymbol{e}^{(\mathrm{w})}_{L'+1,j}$ が結合され、発話復号化 LSTM へ入力されて、以下が得られる:

$$\left(\boldsymbol{h}^{(\mathrm{d})}_{L'+1,j}, \boldsymbol{m}^{(\mathrm{d})}_{L'+1,j}\right) = \begin{cases}\left(\boldsymbol{h}^{(\mathrm{d})}_0, \boldsymbol{m}^{(\mathrm{d})}_0\right) & (j=0)\\ g^{(\mathrm{d})}_{\mathrm{LSTM}}\left(\begin{bmatrix}\boldsymbol{v}_{L'+1}\\ \boldsymbol{t}_{L'+1}\\ \boldsymbol{w}_{L'+1,j}\end{bmatrix}, \boldsymbol{h}^{(\mathrm{d})}_{L'+1,j-1}, \boldsymbol{m}^{(\mathrm{d})}_{L'+1,j-1}; \boldsymbol{\Theta}^{(\mathrm{d})}\right) & (j=1,2,\ldots,J)\end{cases}, \tag{5}$$

ここで $\boldsymbol{h}^{(\mathrm{d})}_0, \boldsymbol{m}^{(\mathrm{d})}_0 \in \mathbb{R}^{D^{(\mathrm{u})}}$ はあらかじめ定められた定数ベクトルであり、$j=0,1,\ldots,J$ に対し、$\boldsymbol{h}^{(\mathrm{d})}_{\ell,j}, \boldsymbol{m}^{(\mathrm{d})}_{\ell,j} \in \mathbb{R}^{D^{(\mathrm{u})}}$ である。また、$g^{(\mathrm{d})}_{\mathrm{LSTM}} : \mathbb{R}^{D^{(\mathrm{c})}+D^{(\mathrm{t})}+D^{(\mathrm{w})}} \times \mathbb{R}^{D^{(\mathrm{u})}} \times \mathbb{R}^{D^{(\mathrm{u})}} \to \mathbb{R}^{D^{(\mathrm{u})}} \times \mathbb{R}^{D^{(\mathrm{u})}}$ は発話復号化 LSTM の入出力関係を表す写像であり、$\boldsymbol{\Theta}^{(\mathrm{d})}$ は発話復号化 LSTM のパラメータである。また、各発話の 0 番目の単語 $\boldsymbol{e}^{(\mathrm{w})}_{L'+1,0}$ は、特殊単語 [go] である。

発話復号化 LSTM の出力にパラメータ $\boldsymbol{K}^{(\mathrm{w}')} \in \mathbb{R}^{D^{(\mathrm{w})}\times D^{(\mathrm{u})}}$, $\boldsymbol{L}^{(\mathrm{w}')} \in \mathbb{R}^{I\times D^{(\mathrm{w})}}$ を乗算し、活性化関数 $a^{(\mathrm{w}')}$ を適用して、次のように発話内容の各単語の確率ベクトル $\boldsymbol{p}^{(\mathrm{w}')}_{L'+1,J'+1}$ を得る:

$$\boldsymbol{p}^{(\mathrm{w}')}_{L'+1,J'+1} = a_{\mathrm{SM}}\left(\boldsymbol{L}^{(\mathrm{w}')}a^{(\mathrm{w}')}\left(\boldsymbol{K}^{(\mathrm{w}')}\boldsymbol{m}^{(\mathrm{d})}_{L'+1,J'+1}\right)\right). \tag{6}$$

なお、発話復号化 LSTM からの出力は通常の文の単語順通りである。また、通常は途中から [pad] のみが出力されるようになり、事実上そこでその発話は終了する。

以上のように、本モデルは入力から次の発話者または単語の確率ベクトルを出力する。図 2 に概要を示す。

2.4 学習

発話の列 $\left(\boldsymbol{e}^{(\mathrm{t})}_1, \boldsymbol{E}^{(\mathrm{w})}_1\right), \left(\boldsymbol{e}^{(\mathrm{t})}_2, \boldsymbol{E}^{(\mathrm{w})}_2\right), \ldots, \left(\boldsymbol{e}^{(\mathrm{t})}_L, \boldsymbol{E}^{(\mathrm{w})}_L\right)$ (小説全体) を $\boldsymbol{E}$ と、またすべてのパラメータをまとめて $\boldsymbol{\Theta}$ と書くことにする。学習における損失関数を cross entropy を使って以下のように定義する:

$$\ell_{\mathrm{train}}\left(\boldsymbol{E};\boldsymbol{\Theta}\right) \stackrel{\mathrm{def}}{=} \sum_{\ell=1}^{L}\left(\lambda\ell_{\mathrm{CE}}\left(\boldsymbol{e}^{(\mathrm{t})}_{\ell}, \boldsymbol{p}^{(\mathrm{t}')}_{\ell}\left(\boldsymbol{\Theta}\right)\right) + \sum_{j=1}^{J}\ell_{\mathrm{CE}}\left(\boldsymbol{e}^{(\mathrm{w})}_{\ell,j}, \boldsymbol{p}^{(\mathrm{w}')}_{\ell,j}\left(\boldsymbol{\Theta}\right)\right)\right). \tag{7}$$

ここで $\lambda \in \mathbb{R}$ は発話者と発話内容の学習の重要性を調節するハイパーパラメータである。提案モデルの学習は、この損失関数の最小化を行う。

2.5 生成

生成時は、学習時に用いた損失関数に加え、二つの罰則項を加えた状態で beam search を行う。一つ目の罰則項は発話内容の短さに科される。これは極端に短い応対が連続する単調な出力を防ぐためである。ここでは ℓ 番目の発話の発話内容の長さ J_ℓ を、その発話内容の中の最初の [pad] までの長さで定義し、罰則を科す。二つ目の罰則項は同一話者の登場の間隔の短さに科される。これは極めて少ない話者で対話が進行するのを防ぐためである。ここでは ℓ 番目の発話の発話者の連続度 s_ℓ を、その発話者が直近で何発話前に発話したかで定義し、罰則を科す。発話内容の短さ、および同一話者の連続度に対する罰則項は、いずれも単調非減少関数 $\ell_{\mathrm{len}}, \ell_{\mathrm{succ}} : \mathbb{N} \to \mathbb{R}$ によって実現される。三つ目の罰則項は発話の最初の三点リーダ「…」に科す。これは、実際の台本形式小説において、三点リーダのみで終了する発話が非常に多く存在し、学習がそれに引きずられるため、これを除外しないと三点リーダの応酬に陥るからである。この罰則項は指示関数を用いた罰則関数 $\ell_{\mathrm{sil}}\left(\boldsymbol{e}^{(\mathrm{w})}\right) \stackrel{\mathrm{def}}{=} \lambda_{\mathrm{sil}}1\{\boldsymbol{e}^{(\mathrm{w})} = \text{「…」}\}$ で定義される。これらの罰則項を加えて、出力に用いる損失関数を以下で定義する:

$$\ell_{\mathrm{gen}}\left(\boldsymbol{E};\boldsymbol{\Theta}\right) \stackrel{\mathrm{def}}{=} \sum_{\ell=1}^{L}\left(\lambda\ell_{\mathrm{CE}}\left(\boldsymbol{e}^{(\mathrm{t})}_{\ell}, \boldsymbol{p}^{(\mathrm{t}')}_{\ell}\left(\boldsymbol{\Theta}\right)\right) + \sum_{j=1}^{J}\ell_{\mathrm{CE}}\left(\boldsymbol{e}^{(\mathrm{w})}_{\ell,j}, \boldsymbol{p}^{(\mathrm{w}')}_{\ell,j}\left(\boldsymbol{\Theta}\right)\right) + \ell_{\mathrm{len}}\left(J_\ell\right) + \ell_{\mathrm{succ}}\left(s_\ell\right) + \ell_{\mathrm{sil}}\left(\boldsymbol{e}^{(\mathrm{w})}_{\ell,1}\right)\right). \tag{8}$$

生成は、この損失関数を小さくするような発話列を 一単語ずつ (あるいは一発話者ずつ) n-beam search で、すなわち n 個まで状態を保持した貪欲法で行う。

2.6 その他

上記の他に、後述の実験ではいくつかのレイヤーに Dropout [Hinton 12] を適用した。また、発話者の入出力と同様の要領で発話の種類の区別に用いられる括弧記号の入出力も行った。

```
片桐早苗「それはそれ、これはこれ」# match
高垣楓「はい、そうです。私、今まで通りに頑張ってきました」
東郷あい「ああ、そうだね。それじゃ、また明日ね」
片桐早苗「うん、ありがとう。それじゃ、また明日ね」
高垣楓「はい、お疲れ様です、プロデューサー」# match
東郷あい「あぁ、すまない、少し疲れ気味でね」
片桐早苗「んー、そうねえ…あ、これお土産なんだっけ」
東郷あい「あぁ、これは…フフッ、そうだね」
高垣楓「ふふっ、ありがとうございます♪」# match
片桐早苗「んっ…んっ…」# match
```

図 3　提案手法による出力例 ($D^{(\mathrm{w})} = 300$, $D^{(\mathrm{u})} = 300$, $D^{(\mathrm{c})} = 1600$, $D^{(\mathrm{u}')} = 200$): なお、# match は学習データに完全一致した出力を示す。

```
P「そうですね。それじゃ、おやすみなさい」
千川ちひろ「おやすみなさーい…」
佐久間まゆ「pさんはどういう風にお茶を飲むのでしょう」
P「そうだな。まゆが作ってくれたんだ」
佐久間まゆ「pさんはどういう風にお茶を飲むのでしょう」
渋谷凛「いや、そういうわけじゃないんだけどさ」
P「そうだな。俺は凛のプロデューサーだからな」
島村卯月「プロデューサーさんは、私のことをどう思っているんですか」
渋谷凛「私は、卯月のことが、好きなんだ」
本田未央「しぶりんは、私のことが大好きなんだね」
北条加蓮「うん、そうだね。私も、卯月のことが好きなんだよ」
島村卯月「えへへ、ありがとうございますっ」
```

図 4　比較手法による出力例 ($D^{(\mathrm{w})} = 300$, $D^{(\mathrm{u})} = 300$, $D^{(\mathrm{c})} = 1600$, $D^{(\mathrm{u}')} = 200$):

表 1　提案手法及び比較手法の Perplexity 及び BLEU

	perplexity	BLEU
RNN-Encoder-Decoder	41.97	**0.6876**
TULIP	**39.51**	0.56642

3. 実験

データ

本研究で提案する枠組は、登場人物の固定を必要とし、同時には一作品しか扱えない。従って実験では、Web 上の台本形式小説の中でも投稿数がトップクラスである「アイドルマスター シンデレラガールズ」の二次創作の台本形式小説に限定し行った。データの収集は Web のクローリングにより、著作権法第四十七条の七の範囲内となるよう、電子計算機による情報解析のみを目的として複製した。当該作品の登場人物は 180 人以上、小説数は 8817、総発話数は 2211505 であった。

実験条件

ベースラインとして、翻訳で用いられる RNN encoder-decoder モデル [Cho 14] を、提案手法と同様話者を考慮したものに拡張したものと比較する。このモデルは、提案手法の文脈更新 LSTM に相当する部分を、前の文から次の文への Dense レイヤーに置き換えたものである。従って、二文以上前のデータを生成に使用しない。これにより、提案手法の文脈更新 LSTM による文脈の学習能力を測ることができると期待される。実装は Tensorflow をバックエンドに Keras で行い, 最適化手法は Adam [Kingma 14] を用いた。学習データ全体を 16 反復学習した。ハイパーパラメータは以下のとおり定めた。$\lambda = 0.001$, $D^{(\mathrm{t})} = 30$, beam search の beam 数は 2 とした。罰則項については, $\ell_{\mathrm{len}}(k) = \frac{5.0}{k+1}, \ell_{\mathrm{succ}}(k) = \frac{5.0}{k+1}$, $\ell_{\mathrm{sil}} = 5.0$ とした。定量評価ではパラメータのグリッドサーチを $D^{(\mathrm{w})} = 300, 600$, $D^{(\mathrm{u})} = 300, 600$, $D^{(\mathrm{c})} = 600, 1600$, $D^{(\mathrm{u}')} = 200, 400$ のすべてのパラメータの組で行った。比較手法についても、同様のグリッドサーチを行っている。この条件の下、図 1(学習データから除外) の開始数文を種としてのみ用い、それぞれの手法で生成を行った。

定性評価

提案手法の出力例として $D^{(\mathrm{w})} = 300$, $D^{(\mathrm{u})} = 300$, $D^{(\mathrm{c})} = 1600$, $D^{(\mathrm{u}')} = 200$ としたものを図 3 にあげる。このパラメータ及び例は著者による cherry-picking である点に注意されたい。同じ小説を種にした場合の比較手法の生成例を図 4 に示す。どちらも、文脈の学習という点では、十分な性能とは言えないが、Web 小説のみを学習から発話単位で自然な日本語を生成している。これは、文脈を学習する上では、モデルの複雑さに対し小説数 8000 が十分でないのに対し、発話の学習の上では 220 万という発話数が十分であることを示唆している。

定量評価

定量評価では、テストデータに対する perplexity と、n-gram の一致度に基づく、翻訳分野での定量評価指標 BLEU [Papineni 02] を用いた。ただし BLEU は文のみを評価できるため、発話者については所与とし発話内容のみ評価を行った。表 2.6 に結果を示す。提案手法は比較手法に比べ perplexity では上回ったものの BLEU スコアでは下回った。興味深いのはこれらのスコアでバリデーションによって選ばれたハイパーパラメータ ($D^{(\mathrm{w})} = 300$, $D^{(\mathrm{u})} = 300$, $D^{(\mathrm{c})} = 600$, $D^{(\mathrm{u}')} = 200$) が先ほど筆者が選んだハイパーパラメータと異なる点である。定量評価指標には、依然として改善点が残るといえる。

4. 結論と課題

本研究では台本形式小説の自動生成を行い、提案手法の発話レベルでの高い性能を示した。課題としては文脈レベルでの性能の向上のほか、提案手法の各コンポーネントの性能への寄与の精査が考えられる。また、1 発話 1 分以上かかる生成時間についても改善が必要である。

参考文献

[Cho 14] Cho, K., Merrienboer, van B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734 (2014)

[Hinton 12] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, *arXiv preprint arXiv:1207.0580* (2012)

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)

[Jain 17] Jain, P., Agrawal, P., Mishra, A., Sukhwani, M., Laha, A., and Sankaranarayanan, K.: Story Generation from Sequence of Independent Short Descriptions, *arXiv preprint arXiv:1707.05501* (2017)

[Kashiwabara 12] Kashiwabara, T.:「2 ちゃんねるスレッドまとめブログ」によるニュース・コミュニケーションに関する一考察, 哲學, Vol. 128, pp. 207–234 (2012)

[Kingma 14] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014)

[Li 16] Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., and Dolan, B.: A persona-based neural conversation model, *arXiv preprint arXiv:1603.06155* (2016)

[Ouchi 16] Ouchi, H. and Tsuboi, Y.: Addressee and response selection for multi-party conversation, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2133–2143 (2016)

[Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, in *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318Association for Computational Linguistics (2002)

[Serban 16] Serban, I. V., Sordoni, A., Bengio, Y., Courville, A. C., and Pineau, J.: Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models., in *AAAI*, Vol. 16, pp. 3776–3784 (2016)

[Sordoni 15] Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Grue Simonsen, J., and Nie, J.-Y.: A hierarchical recurrent encoder-decoder for generative context-aware query suggestion, in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 553–562ACM (2015)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in *Advances in neural information processing systems*, pp. 3104–3112 (2014)

[Takagi 14] Takagi, D., Sato, S., and Komatani, K.: 会話を中心とした超短編小説の自動生成, in *The 28th Annual Conference of the Japanese Society for Artificial Intelligence* (2014)

Oral presentation | NFC Session | [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

[2E3-NFC-1a] [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

座長:成田 雅彦（産業技術大学院大学）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room E (4F Queen)

[2E3-NFC-1a-01] Robot based Active Sensing System and "Integrated Platform for Cloud-based Robot Services Project"

○Masahiko Narita[1], Nobuto Matsuhira[2], Yosuke Tsuchiya[1], Yuka Kato[3], Yoshihiko Murakawa[4] （1. Advanced Institute of Industrial Technology, 2. Shibaura Institute of Technology, 3. Tokyo Woman's Christian University, 4. FUJITSU LABORATORIES LTD）

3:20 PM - 3:40 PM

[2E3-NFC-1a-02] Behavior information acquisition using wireless tag in conventions

○Tokuro Matsuo[1], Hidekazu Yanagimoto[2], Masahiko Narita[1], Naoki Fukuta[3], Katsuhide Fujita[4], Kiyota Hashimoto[5], Hidekazu Iwamoto[6] （1. Advanced Institute of Industrial Technology, 2. Osaka Prefecture University, 3. Shizuoka University, 4. Tokyo University of Agriculture and Technology, 5. Prince of Songkla University, 6. Josai International University）

3:40 PM - 4:00 PM

[2E3-NFC-1a-03] Development on a Large Display Type Tele-presence Robot

○Tatsuno Kyoichi[1], Hideharu Oobitsu[2], Katsumi Yoshino[2] （1. Meijo University, 2. Shimane Institute For Industrial Technology）

4:00 PM - 4:20 PM

[2E3-NFC-1a-04] An analysis of user's utterance to reduce inappropriate response of chat robot

○Shodai Aoyagi[1], Kazuki Hirata[1], Tomoya Takatani[2], Eri Sato-Shimokawara[1], Toru Yamaguchi[1] （1. Tokyo Metropolitan University, 2. Toyota Motor Corporation）

4:20 PM - 4:40 PM

[2E3-NFC-1a-05] Development of a partner robot aiming at reducing the psychological burden of child-care environment

○Yosuke Tsuchiya[1], Jun Uchiyama[2], Kazunari Hirakoso[2], Daichi Takeshima[2], Rika Sekita[2], Jinnan Zhang[2], Yoshio Kondo[2], Masahiko Narita[2] （1. Tokyo Online University, 2. Advanced Institute of Industrial Technology）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ロボットを用いたアクティブセンシングシステムと「クラウドベースのロボットサービスの統合基盤」

Robot based Active Sensing System and "Integrated Platform for Cloud-based Robot Services Project"

成田 雅彦[*1]　松日楽信人[*2]　土屋 陽介[*1] 加藤 由花[*3] 村川 賀彦[*4]
Masahiko Narita[1], Nobuto Matsuhira[2], Yosuke Tsuchiya[1] , Yuka Kato[3], and Yoshihiko Murakawa[4]
[*1] 公立大学法人首都大学東京産業技術大学院大学　[*2] 芝浦工業大学　[*3] 東京女子大学
[*4] 株式会社富士通研究所
Advanced Institute of Industrial Technology, Shibaura Institute of Technology, Tokyo Woman's Christian University
FUJITSU LABORATORIES LTD

The purpose of this challenge is to build CRSP (the Cloud based Robot Service Platform), designed for non-specialist who will use the robots, particularly, AI researchers. A robot service integration foundation consists of the base function as the middleware layer built on the RSi technology, and service development framework as a SDK layer. Since this proposal was proposed in June 2013, we proceeded with the development and utilization of this integration infrastructure. This project was judged as "graduation", after passing the examination of the last fiscal year, and this year is fifth years in the final year. In this presentation, we review the project and propose the results obtained as a concept for future development.

1. チャレンジの目的

本テーマの目的は，ロボットサービスイニシアチブ(RSi)[RSi]のロボットサービス仕様である RSNP（Robot Service Network Protocol）[RSi 10]を発展させ，AI 系の研究者等ロボットの非専門家がロボットサービスの開発・利用に参入できるクラウドベースの統合基盤であるCRSP(Cloud-based Robot Service Platform)（図 1）の構築推進である．本プロジェクトは 2013 年 6 月に提案以来，本統合基盤の整備と活用を進め，昨年度の審査を経てめでたく「卒業」となり，今回で最終年の 5 年完了となる．本発表では，本プロジェクトを振り返り，得られた成果を，今後の展開のためにコンセプトとして提案する．

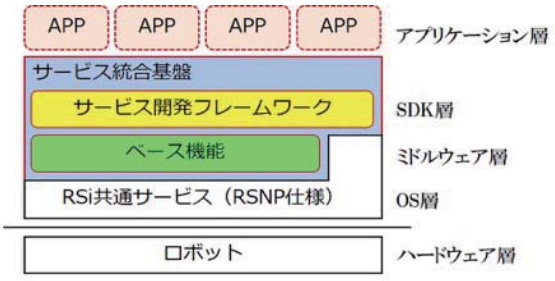


図 1　CRSP の構成

2. 1～4 年目の成果

ロボットサービスに必要なリアルタイム音声通話機能などの機能モジュールを CRSP に統合し，移動ロボットを中心としてロボットシミュレーション環境と SLAM を CRSP へ統合し，実環境へも CRSP を組み込む試作を行った．また，コミュニティで要求された CRSP フレームワークの自己拡張構造について仕様を提案し試作し，これに基づくサービスロボットをマーケティングへ展開しうるコンセプトとして「ロボットによるアクティブセンシングモデル」を考案し，アンケートサービスとサイネージロボを CRSP の上に構築した．4 年目は，ロボットによるアクティブセンシングのためにデータ収集の高度化を狙って，ロボットサービスと連続性のあるスマホロボット，アンケートの動的生成，センサーとの連携を試作した．また，RSNP コンテスト 2013-2016，国際ロボット展 2015，JAPAN ROBOT WEEK 2016 や学会・商店街などのデモ展示を実施し，本チャレンジとの相乗効果による研究者間や産業との相互作用の促進，研究コミュニティや研究領域の拡大，普及活動を行った．以下に概要を述べる．

2.1 音声通信やプラットフォーム連携モジュールの統合

ロボットサービス構築に必要な機能モジュールの連携と統合として，a. 汎用的な RSNP/RTM gateway の設計開発を行い，MRPT を用いた環境地図作成用 RT コンポーネントを，インターネット上のサービスとして公開する仕組みの試作を行った．また，b.音声通信の RSNP ベースの本プラットフォームへの統合を実現した．これにより，インターネット上でリアルタイム音声通信を容易に扱え，サービスロボット分野で有望とされている遠隔地からの見守りサービスや接客サービスなど，生活空間でのサービスへの適用できる[大澤 15]．

2.2 移動ロボット向け統合環境

社会知能発生学シミュレータ SIGVerse を用い，RSNP と接続してロボットシミュレータ環境を実現した．ロボットだけではなく，ロボットのサービス対象や設置環境のシミュレーションができることに特徴がある．これを用いて，RSNP/RTM gateway により，大域地図生成 RTC を接続することで SLAM モジュールと統合し，移動ロボットの操作・モニターするトレースサービス[成田 15]を実現した．また，実用化の近いロボットへ本プラットフォームを展開する試みを都立産業技術研究センターの協力を得て行った．さらに，ネットワーク遅延がロボット制御に与える影響について考察し，観測と制御のずれを考慮した制御モデルを提案した．[加藤 17-1] [加藤 17-2] [加藤 17-3]

2.3 ロボットによるアクティブセンシングとマーケティングへ展開

ロボットを用いて直接の利用者にサービスを提供すると同時に，ロボットからアクティブに話しかけやアンケートを提示することで利用者の趣向情報を収集（センシング）し活用するサービス

ロボットのモデルを提案した(図 2). これは展示会場や観光地でサービスロボットを用いてマーケティングへ展開する際のモデルとなる.この実現のため，CRSP フレームワークの自己拡張機能[土屋 17]・アンケートプロファイルなど上位レベルの機能モジュール，スタンプラリーサービスなどの CRSP のアプリケーション，CRSP を組み込んだサイネージロボットを開発した. さらに，人工知能研究者を含む「ベイエリアおもてなしロボット研究会」と連携して，国際ロボット展 2015，産業交流展 2015 など大規模イベントで CRSP を用いたアンケートの収集の実証実験を行って効果を検証した[成田 17-1] [成田 17-2].

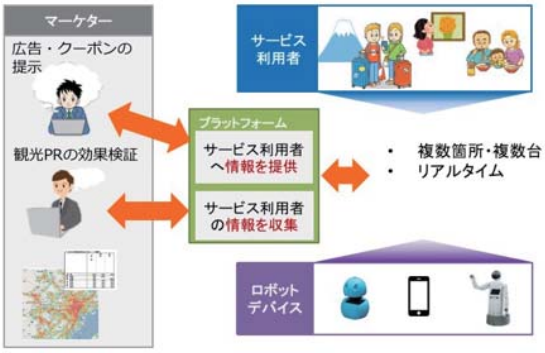


図 2: ロボットによるアクティブセンシングコンセプト

このコンセプトは，サービスを直接の利用者だけでなく得られた利用を別のステークホルダーが利用でき，ここに様々なデータ分析の手法を適用できるので，従来のサービスロボットのエコシステムを改善する可能性を秘めている. このシステムを容易に構築できるように，タッチパネル式ディスプレイやタブレット PC ベースで等身大にスタンドで固定した低費用のサイネージロボットや，ロボットでの回答の代わりに，アプリインストール無しでスマートフォンから回答できるスマホロボ，アンケートサービスの多国化[青木 16]などの機能強化によりデータ収集手法を高度化し，システムとして完成度を上げ汎用化した.

3. 最終年の活動と成果

3.1 研究開発

最終年の 5 年目は，前年に引き続き CRSP を高度化し，適用範囲を広めることに務めた. 多岐にわたる高度化の要件を適用箇所をもとに以下の a～d にデータの収集・分析・フィードバック・適用範囲の拡大に分類し体系化した[成田 17-3].

a. ロボットによるデータ収集の高度化：ロボット利用者に対するロボットによる人間的データの取得・センシング手法を高度化し，データ集積を促進する.
b. 分析手法の高度化：集積したデータを分析し，フィードバックを生成する. 各種クラウドサービスや人工知能技術と連携し高度化する.
c. フィードバックループの高度化：取得した情報を解析し，さらにサービス提供プロセスへフィードバックするループ形成を高度化する
d. 適用範囲の拡大：イベント・観光地での適用の深化，他領域への試み.

これに基き(1)～(4)に述べるロボット，サービスを開発した.

(1) アンケートの動的生成

ロボットアンケートシステムは，観光地や展示会場で声がけして，その場でアンケートに回答してもらう特性上，アンケートの設問を少なくする必要がある. 一方で，たくさんの種類の設問の回答を得たいという主催者側の要望もある. このアンケートの動的生成は，すでに得られた回答状況をもとに数の少ない設問を重点的に質問するなど，柔軟なアンケート設問の設定や，多くの設問に答えてもらうため仕組みである.

この実装のために以下の 2 つの手法を検討した.

- 複数の設問からなるアンケート群を，出題の優先度を付与して用意し，回答者やアンケートの回答結果にもとづき，アンケートをアンケート群から選択する方式.
- 複数のアンケート設問群を用意し，回答者やアンケートの回答結果にもとづき，設問を設問群から選択し，柔軟にアンケートを生成する.

前者は前年に実施したので本年は後者を実装した.

実装では，アンケートの要素・生成をアンケート XML，アンケートテンプレート XML として定式化した. 具体的には，複数のアンケート設問からなるテンプレートからなるアンケートデータファイル，初期メッセージファイル，ロボットがユーザーに出力するアンケートを選択する方法を記述したマッチングファイルを用意する. マッチングファイルの中では，アンケートの結果データベースの集計結果や，回答者のアンケート結果を参照でき，これらにもとづき条件分岐の記載ができる. アンケートサービスシステムは，回答者が回答者 id を入力すると，回答者情報（個人を特定する情報は含まれていない）を抽出後，マッチングファイルを読み込み，出力するアンケートを生成する. この予備実験は，2017 年 7 月に浜松で開催された国際会議 IIAI2017 にて，主催者と連携し主催者の案内サービスの一部として行い，実証実験は後述の国際ロボット展 2017 で実施した. 本機能の高度化の箇所は上記の a,b,c である.

(2) 身体性のあるロボットの試作と導入

デザイン性のある”子育て支援ロボット” (図 3 左)，浄瑠璃人形を参考にした身体性のある”かしらロボ”[成田 17-4](図 3 右)を開発し，集客・データ収集効果の向上を検証した. 前者は家電のようなデザインで家庭へ違和感の内容配慮し，後者は集客効果を狙い，品質の高いデザインのかしらを用意に設計できる特徴を持つ. 今後，腕・手を付加し大型化する. 本機能の高度化箇所は上記の a, c, d である[内山 17-1][内山 17-2].

(3) 調理用音声アシスタントロボット

家庭空間のアクティブセンシングを試行した. 高度化箇所は d である. 詳細は 3.3 コミュニティ活動を参照されたい[熊谷 17][佐藤 17].

(4) 複数種のロボットの連携運用向けモニタリンクシステム

サービスロボットの低価格化が進みつつあるが，異なるプロトコルを持つ異種サービスロボットが大量に配置された場合を想定し，ロボット間情報共有の実現可能性や効率的な管理・運用の行うことである. 本システムでは，収集したデータは分析・整理し，各ブースにあるディスプレイへ表示される. データは，収集されるたびにポップアップで表示され，さらに時間ごとやそれぞれのロボットにふさわしい方法で詳細をモニタできる. 同時に，後の分析のために使用者の感想などをアンケートすることもできる. アンケート内容は蓄積データを利用して動的に生成することで，手短に使用者に応じた適切な項目を質問することが特徴である. また，ロボットと会話しながらのモニタリングを実現した. 基礎実験は，2017 国際ロボット展中に6つブースに，子育て支援ロボット，調理用音声アシスタントロボット，かしらロボなど，

複数種のロボットや人数推移システム(センサ)を配置し，ロボットサービス，センサのリアルタイム監視・集計データの収集，結果をフィードバックした[成田 17-5]. この実験は,RRI，RSi,ベイエリアおもてなしロボット研究会との連携により実現した. また，RRI の検討をもとにより広い範囲での適用を試みる.

3.2 コンセプトの提案



図 3: ディスプレイに装着した「かしらロボ」と子育て支援ロボット

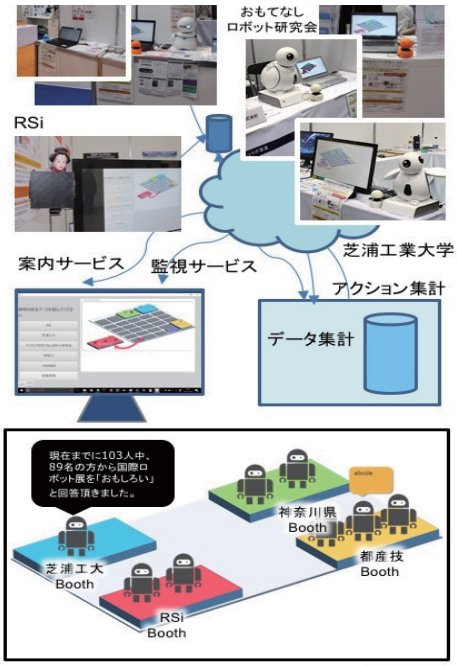


図 4: モニタリングシステムの実験

本プロジェクトの結果として説明済みのものを含めて以下のコンセプトを提案する.

(1) ロボットによるアクティブセンシング

サービスロボットの活用の方向性を示すもので，2.3 で説明した. 本プロジェクトでも既に幾つかの成果が得られた。

(2) CRSP を高度化

CRSP の機能強化の体系化を提案するもので,3.1 で説明した.

(3) RUI と業務システムとの分離

3.1 で述べたように本プラットフォームを用いてアンケートサービスをベースにアクティブセンシングを行うロボットサービスを幾つか構築した. これらを注意深く見ると，アンケートに直接依存しない，回答者の普遍的な属性を用いてフィードバックを作成している部分と，アンケート結果の分析に業務知識を使っている部分に分けられる. 実際，旅行向けと学会向けアンケートサービスでは，業務知識へのアクセス部分が異なっている. 回答者の普遍的な属性を用いてフィードバックをロボットユーザインタフェース(RUI:Robot User Interface)と呼ぶとすれば，ロボットサービスは，RUI と業務知識から成り立っていると言える. 結果，業務に依存しないロボットサービス，ロボットからなるロボットユーザインタフェースを用いることで，任意の業務を共通な仕組みでアクティブセンシングに組み込める構造が実現できると思われる. ここで言う RUI 機能としては，例えば，アンケートサービス機能，会話蓄積，会話生成であり，ロボットは 3.1(2)(3)で述べた様々ものが利用できる. 業務サービスは，3.1(3)で述べた調理用音声アシスタントロボット向け料理レシピサービス，3.1(4)のロボットモニタリングが当たる. 特に，モニタリングシステムの実証では，ロボット無しのシステムとアクティブセンシングに組み込んだものの 2 つを実装して比較した. ソフトウエアでは，GUI,HCI と言われて人へのインタフェースが分離され有効性を認識されているが，ロボット分野でもそうしたコンセプトが成り立ち，よりサービスロボットの適用を容易にできると期待される.

3.3 コミュニティ活動

(1) 近未来チャレンジ「クラウドベースのロボットサービスの統合基盤」セッション (2017)

「クラウドベースのロボットサービスの統合基盤」セッションとして 2017 年度 人工知能学会全国大会にて，第 4 回目のサバイバルセッションを行った. これらを以下に列記する.

- 産技大から「クラウドベースのロボットサービスの統合基盤」として概要説明を行った.
- SDK 層への取り組みでは，「クラウド環境を利用した移動ロボット遠隔ナビゲーション手法」(発表:加藤由花(東京女子大学))を紹介した. 今後，移動ロボット向けの統合環境フレームワークに組み込んでいく. 一方，「共起・共助マッチングに基づく対話支援ロボットの開発に向けた人同士の会話解析」(山口亨(首都大学東京))の発表があり興味深い内容であった。
- 今後の可能性としては，「固定型遠隔訪問ロボットの開発」」(発表:辰野恭市(名城大学))，「IoT デバイスによる洪水検知システム」(発表:土屋陽介 (産業技術大学院大学))，アンケートロボットを用いて回答の信頼性を回答時間から推測する試みである「RSNP を利用したロボットアンケート評価システム」(発表:池田貴政(芝浦工業大学))の発表があり大変有用であった. 本プラットフォームを用いた研究計画として「コンベンション情報システムのための期待分析と実証実験に関する研究」(発表:松尾徳朗(産業技術大学院大学))が紹介された.
- パネルディスカッションでは，発表者が各自の発表を元に，介護・福祉・イベント・観光などの適用分野，ロボットの共通知能や管理システムなど今後増える複数ロボットの取り扱い，脈拍センサー・赤外線カメラ・人数推移システムなどセンサーとの連携の期待について議論した.会場からも多くの発言があり，パネルでのディスカッションが良かったとの評価を受けた.

(2) RSNP コンテスト 2017

2017 年 9 月に RSi 主催，ロボット学会共催の RSNP を利用したロボットサービスコンテストを開催した. 関係者による審査を行った結果，以下の作品を表彰した.

- 最優秀賞(RSi 賞):調理用音声アシスタントロボット (産業技術大学院大学 佐藤ほか). 約 300 レシピを用意した.
- 優秀賞: Falcon を用いたマスタスレーブシ ステム(芝浦工業大学 内藤ほか).RSNP を用いてマスタスレーブ制御で遠隔操作を行 うためのシステムを構築した.マスタ側の Falcon に，スレーブ側が物に触れた 際の反力を出力する.
- 優秀賞: Raspberry Pi を用いた RSNP 人検 出データ収集システム(芝浦工業大学 瀬沼ほ か). 焦電人感センサーを

用いて人検出のた めのセンサユニットを開発し, RSNP で人検出 データの送受信を行う「人検出データ収集シス テム」を提案する.LRF を用いた従来のも のに比べて低コストに構築できる.

- コンセプト部門賞:「笑顔溢れる子育て環境 を実現するロボット」- 人と共生するパートナ ーロボットのコンセプト提案 -(産業技術大学 院大学 竹島, 内山ほか) . 乳児感情 と連携する子育て支援ロボットのコンセプト 提案.プロトタイプ提示を行った.

技術的な観点では, ロボットによる アクティブセンシングモデルは商店街・観光地での利用を想定しているが, 調理用音声アシスタントロ ボットのサービスにより個別のデータの収集とフ ィードバックにも適用できることを示唆しており 興味深いシステムである.人と共生するパートナー ロボットでは, 顔の表現の LED パネル柔軟性が高く様々な表現が期待でき, 空気アクチュエータは無音なので従来とは違った展開が期待できる.

(3) 「エージェント技術に基づく大規模合意形成支援システムの創成」グループ

イベントの参加者情報の取得・利用・働きかけについて, 類似の問題意識を持つ, JST CREST「エージェント技術に基づく大規模合意形成支援システムの創成」の松尾共同研究グループが開発しているセンサーデバイスと IoT システムと CRSP と本ロボットサービスによるアンケートシステムの連携し,EAIS2017 にて予備実験を行った.

(4) RSi と RRI

RSNP コンテストの実施に加え, RRI に「多種多様なロボットが活躍するためのネットワークアーキテクチャ - 100 台のロボットと 100 のアプリ:継続する連携研究-」を提案し採択された[松日楽 17]. また, RRI(ロボット革命イニシアティブ) WG3 新規プロジェクト等企画立案 WG では, 上記にもとづき RSI と協力し検討をしている.

4. まとめと今後の展開

本プロジェクトは 2013 年 6 月に提案以来, 本統合基盤の整備と活用を進め, 昨年度の審査を経てめでたく「卒業」となり, 今回で最終年の 5 年完了となる. 最終年度は本プラットフォームを高度化して効果を検証するとともに, 今後の展開のためにコンセプトとして提案した. 引き続きご協力を賜りたい.

謝辞

本研究の一部は JSPS 科研費 17K00366 の助成を受けたものです. また, 本活動の一部は, JST CREST JPMJCR15E1 の助成, RRI WG3 新規プロジェクト等企画立案 WG の助成を受けたものです.

参考文献

[RSi] RSi-Robot Service initiative: http://robotservices.org/ [Online]

[RSi 10] ロボットサービスイニシアチブ, Robot Service Network Protocol 2.3 仕様書 1.0 版, 2010.

[青木 16] 青木大起,前佛達也,中村隆宏,宮内真紀江.井上直己,増田均,泉井透,成田雅彦,"スマートデバイスとサービスロボットの連携によるマーケティングプラットフォーム構築の拡張",日本ロボット学会学術講演会 2016, 1R2-01, 2016

[大澤 15] 大澤秀也, 朝倉健介, 小原範子, 藤田尚宏, 佐藤健, 中川幸子, 成田雅彦, RSNP 拡張によるロボット制御と音声通信の統合のためのロボットサービスプラットフォーム,日本ロボット学会誌, Vol.33 No.2, 2015(3).

[土屋 17] 土屋陽介, 成田雅彦, 泉井透, RSNP 仕様拡張のためのカスタムプロファイル機能の実装, 産業技術大学院大学紀要 Vol. 10, pp.67-72, 2016 .

[成田 15] 成田雅彦, 泉井透, 中川幸子, 土屋陽介, 松日楽信人,加藤由花, "ネットワークを活用したロボットサービスのための非専門家向け開発フレームワークの提案", 日本ロボット学会誌, Vol.33, No.10, pp.807-817, 2015.

[成田 17-1] 成田雅彦, 土屋陽介, 泉井透, 阿久津裕, 安田福啓, 中川幸子, 松日楽信人,"サービスロボットによる分散アンケートサービスの提案"~非専門家向け開発フレームワークの応用~", 日本ロボット学会学会誌 ,2017 .

[成田 17-2] 成田 雅彦, 土屋陽介, 中川幸子, 阿久津裕, 泉井透, 野見山大基, 松日楽信人, 本村陽一,"マーケティング分野への適用を目指したスタンプラリーとアンケートサービスの CRSP を用いた構築 ～クラウドベースのロボットサービス統合基盤の進展～", 人工知能学会論文誌 32 -1 号,2017 .

[成田 17-3] 成田 雅彦 (産技大) 中川幸子(産技大)土屋陽介(産技大) 加藤由花(東京女子大), 松日楽信人(芝浦工大), 非専門家向けロボットサービスプラットフォームの高度化, 2B2-03,日本ロボット学会学術講演会,2017(9)

[成田 17-4] 成田雅彦 , サイネージロボットの身体性を実現する一提案 浄瑠璃人形の比較を参考にした「かしら」のモデリング, 産業技術大学大学紀要,No.11, pp.91-96

[成田 17-5] 成田 雅彦, 松日楽 信人,RSi(ロボットサービスイニシアチブ)の活動と今後の展望,p1111-1115,1D6_01, SI2017, 仙台,2017 (12)

[熊谷 17] Seiichi Kumagai, Tomoko Sato, Hiroki Sano, Hiroshi Tauchi, Norio Maeda, Yuki Horiguchi, Sachiko Nakagawa, and Masahiko Narita, Cooking assistant service utilizing an interactive robot, ThA3.3, SII2017, Taipei,Taiwan,2017/12

[加藤 17-1] Y. Kato and M. Tanaka, A Delay-conscious Communication Model for Mobile Robot Navigation, IEEE International Conference on Robotics and Automation (ICRA 2017), Poster-PM-21, 2017(5).

[加藤 17-2] Y. Kato, Y. Nagano and H. Yokoyama, A Pedestrian Model in Human-Robot Coexisting Environment for Mobile Robot Navigation, Proc. 2017 IEEE/SICE International Symposium on System Integration (SII 2017), 2017(12).

[加藤 17-3] 加藤由花, 池田貴政, 岡野憲, 松日楽信人,インタフェースロボット応答制御のための歩行者分岐方向の予測, 情報処理学会マルチメディア通信と分散処理研究会, DPS-174, pp.1-7, 2018(3).

[佐藤 17] 佐藤 智子, 前田 紀夫, 堀口 裕記, 田内 博史, 熊谷 誠一, 佐野 博紀, 成田 雅彦, 調理用音声アシスタントロボット, 2B1-04,日本ロボット学会学術講演会, 2017(9)

[内山 17-1] 内山 純, 近藤 嘉男,「パートナーロボット」の開発研究 ― 表情と仕草に着目したコンセプト構築の試み ―,産業技術大学院大学紀要, No.11, pp.175-180

[内山 17-2] 内山純, 竹島大智, 関田理花, 張進男, 平社和也, 近藤嘉男, 「笑顔溢れる子育て環境を実現するロボット」コンセプト, 2B1-032,日本ロボット学会学術講演会,2017/9

[松日楽信人 17] 松日楽信人,長瀬雅之,成田雅彦,多種多様なロボットが活躍するためのネットワークアーキテクチャ - 100 台のロボットと 100 のアプリ : 継続する連携研 RSI,2017(8)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

無線タグによるコンベンション参加者行動センサリング

Behavior information acquisition using wireless tag in conventions

松尾徳朗[*1] Tokuro Matsuo　柳本豪一[*2] Hidekazu Yanagimoto　成田雅彦[*1] Masahiko Narita　福田直樹[*3] Naoki Fukuta

藤田桂英[*4] Katsuhide Fujita　橋本喜代太[*5] Kiyota Hashimoto　岩本英和[*6] Hidekazu Iwamoto

[*1] 産業技術大学院大学 Advanced Institute of Industrial Technology　[*2] 大阪府立大学 Osaka Prefecture University　[*3] 静岡大学 Shizuoka University　[*4] 東京農工大学 Tokyo University of Agriculture and Technology

[*5] プリンス・オブ・ソンクラー大学 Prince of Songkla University　[*6] 城西国際大学 Josai International University

This paper describes an experiment condition and situation about attendee's behavior in convention. Each attendee holds wireless tag which sends signal to parent devices by low energy Bluetooth communication. Parent devices are set up at a lot of points in the convention facility. Each parent device sends the data into the server through wifi. The paper shows the visual result of experiments.

1. はじめに

コンベンションの目的の一つは広義にいう合意形成である．参加者にとって，自らの研究成果の発表を他者に理解させること，他者の発表の聴講を通した研究内容の理解を通した自らの研究内容の拡張や深化，分科会の外でのネットワーキングなどを含めて様々なタイプの合意形成が存在する．本稿では，コンベンションにおける合意形成支援に用いるための参加者行動データ収集手法について説明し，データ収集結果について紹介する．

コンベンションビジネスは，情報システムや各種センサーの利活用により，参加者を効果的に支援することが可能であり，ビジネス成功に多くの可能性を持つ．コンベンションマネジメントやそのビジネス研究において，アンケートやインタビューを通した調査分析により，イベントツーリズム分野において多くの研究成果が提供されている[1]. Oppermann や Chon らによるコンベンションへの参加動機に関するモデリングが提案されている[2]. 参加者の個人の嗜好や制約を含め，コンベンション事業の性質や特徴，学協会やイベント自体の特徴，開催地の特徴などによる参加動機が存在することが知られている．Tanford，Montgomery および Nelson らは，コンベンション参加者の参加動機，参加者の満足度，および参加リテンションの立場で分析している[3]. 先述の通り，当該分野の多くはアンケートやインタビューによる調査が多い．これまでに著者らは，コンベンション参加者がキャリア形成や教育的活動よりネットワーキングに対して大きな期待を持つことを明らかにしている．

コンベンションにおけるネットワーキングの機会として，分科会などでの質疑応答，ソーシャルイベントでの交流などが挙げられるが，その実態についてはほとんど知られていない．そこで著者らは，ネットワーキングを合意形成の観点から支援することを目的として，具体的な支援を可能とするために参加者の行動データを収集した．コンベンション参加者は，無線タグを保持し，10秒に1度ビーコンを発信する．省電力の Bluetooth（BLE: Bluetooth Low Energy）を用いることで，会場に設置されたスキャナがビーコンを受信する．これらの情報は，サーバには，各スキャナから収集された参加者位置情報が蓄積される．これらのデータにより，(1)参加者の行動の視覚化，(2)参加者の分科会間の移動パターン分析，(3)ポスターセッションでの滞留パターン分析，(4)参加者間の行動類似度分析などに活用することが可能となる．これらを用いたコンベンション設計が可能となる．本稿では，本実験の環境および会議場のフロアマップ上で再現結果について紹介する．

連絡先：松尾徳朗，産業技術大学院大学，140-0004 東京都品川区東大井1−10−40，Tel: 03-3472-7831

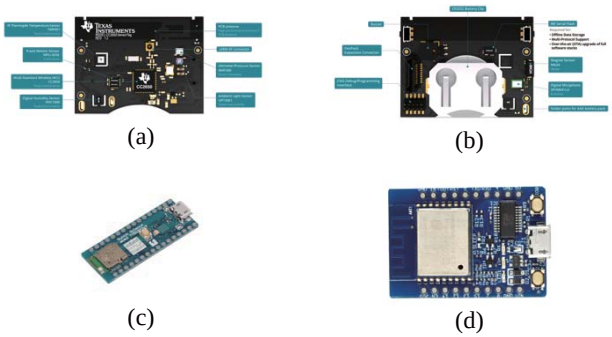
2. 位置データ取得環境

2.1 参加者位置情報収集

既存の研究として，位置情報システムに関する研究成果が多く提案されている．位置情報を推測したり，所在情報を提供することによる業務や活動支援に関する研究は多い．著者らは無線タグに複数のセンサーが搭載されたデバイスを用いた位置情報取得を目的とした環境を構築する．センサーで取得されたデータは，省電力 Bluetooth を用いて発信され，コンベンション施設に設置されたスキャナにより受信される．wifi モジュールに接続されスキャナから，wifi ルータにデータが送信され，サーバに格納される．無線タグから発信されたビーコンは，複数のスキャナにより取得されるが，無線強度の違いにより，参加者の位置情報は推測される．

2.2 デバイス

無線タグには加速度センサー，温度センサー，磁気センサーなどを含めた複数のセンサーが搭載されている．位置情報の他に参加者の振る舞いに関するデータを取得することができる．図1の(a)および(b)は無線タグデバイス（CC2650STK（Texas Instruments 製 [4]）の写真である．図1の(c)はスキャナ（親機 mbed TY51822r3[5]）の写真である．スキャナには図 2 の(d)に示す Wi-fi モジュール（ESPWROOM-02 Development Board）を接続して用いる．表1は実験に用いるデバイスの機能と詳細を説明している，無線タグはコンベンション参加者が保持する．スキャナはコンベンション施設の各所に配置される．

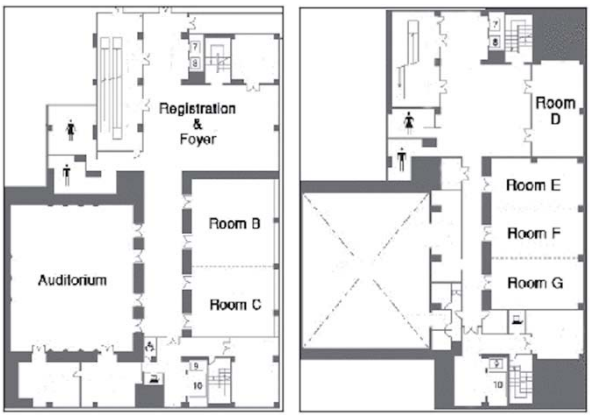


(a) (b) (c) (d)

図1 無線タグおよび親機

表 1 ：実証実験使用機材

Devices	ID/Sensors	Function
Wireless tag	Accelerometer	Taking an attendee's behavior data and location data. And, this can be used the participation situation.
	Gyroscope	Taking an attendee's behavior data.
	Ambient temperature	Taking data such that the attendees are at the outside of building.
Parent device	mbed TY51822r3	This device collects data from wireless tag.
	ESP-WROOM-02	This is attached into the mbed TY51822r3 and the data received by the parent device send to the main computer.
WI-FI A/P	Access Point	Access point provided by the facility.
Main Computer	Apple MacBook	This is used to collect data sent from parent devices via W-FI access point.

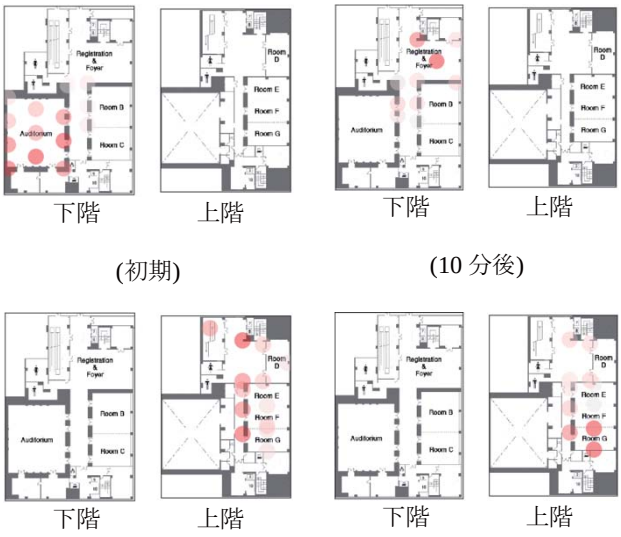


下階 上階

図2 フロアマップ

3. 位置情報マッピング

収集された参加者位置データを用いて，コンベンション施設のフロアマップに参加者の行動を再現できる．図2に示されるフロアマップは，下階（左）と上階（右）を示している．下階には，Auditorium，Foyer，分科会場 B および C があり，上階には，分科会場 D，E，F，および G がある．コンベンションの受付は Foyer エリアで行われており，またコーヒーブレイクおよびポスターセッションも同様に Foyer エリアで行われている．基調講演や主たる分科会は Auditorium で実施されている．図3に示すテーブルには，参加者の 10 分ごとの移動の変化を示したものである．スキャナが受信する信号の強度により，フロアマップに示される色を変化させている．参加者は濃く示された色により近い



下階 上階 下階 上階

(初期) (10 分後)

(20 分後) (30 分後)

図3 参加者の行動再現

場所にいることを表している．初期位置においては下階の左側に示される部屋に参加者がいることを示している．その中でも部屋の右下付近で受信された信号の強度が強い．次に 10 分後には，下階の右上の付近で信号が受信されている．このエリアでは，コーヒーやリフレッシュメントが提供されており，参加者は飲み物を取った可能性がある．20 分後には，上階の中央部分に移動しており，30 分後には上階の右下に示される分科会場に入場していることが示されている．

4. おわりに

本稿では，コンベンションにおける合意形成の一つであるネットワーキングを分析することを目的とした，無線タグを用いた参加者行動センサリングに関する実験環境および実験で得られたデータを用いた行動の再現の可視化に関して紹介した．参加者全員の行動データについて多次元尺度法を用いた行動類似度計算により，参加者のネットワーキング活動の活発化に関する評価が可能となる．今後の展望としては，コンベンション環境を変化させることによるネットワーキング活動の変化について取り組む．

参考文献

[1] Y. Sugiyama, T. Matsuo, K. Tei, H. Naganuma, H. Iwamoto, Interest and Expectation Analysis of International Con-vention Participants to the International Conference, International Journal of Service and Knowledge Management, Vol 1, No 2, pp. 63-75, 2017.

[2] M. Oppermann, K. Chon, Convention participation Decision Making Process, Annals of Tourism Research, Vol. 24, No. 1, pp.178-191, 1997.

[3] S. Tanford, R. Montgomery, K. B. Nelson, Factors that Influence Attendance, Satisfaction, and Loyalty for Conventions, Journal of Convention & Event Tourism, Vol.13, pp.290–318, 2012.

[4] Texas Instruments, Inc. http://www.ti.com

[5] Switch Science, Inc., https://international.switch-science.com/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

大型 TV 版遠隔訪問ロボットの開発

Development on a Large Display Type Tele-presence Robot

辰野恭市[*1] 大櫃秀治[*2] 吉野勝美[*2]
Kyoichi Tatsuno Hideharu Oobitsu Katsumi Yoshino

[*1]名城大学 理工学部 電気電子工学科 Meijo University
[*2]島根県産業技術センター Shimane Institute For Industrial Technology

We have developed a large display type tele-presence robot. This robot has not mobile components at all for the safty. This paper describe the outline of the system and introduce the usage example of the system.

1. はじめに

独居の高齢者や介護センターの入居者が家族・友人と離れて寂しく暮らしているケースが少なからず見られる。これらの高齢者が、離れて暮らす家族・友人と話をしながら少しでも楽しく生活できるように、遠隔訪問ロボットを開発してきた。このロボットは、遠隔に居る家族・友人が、パソコン・インターネットを介してロボットに乗り移り、テレビ電話をしたり、音楽を聞かせたり、四季折々の風景ビデオを見せたりできるものである。

これまで開発してきたものは図1のように移動ができるもの[1], 図2のようにディスプレイを顔に見立てて首を振りながら話ができるもの[2]を開発してきた。これらのロボットは機械的な可動部があるために安全上の問題を危惧される方も居た。

そこで、可動部を取り除き、大型テレビ顔を写して TV 電話をしたり、Youtube の映像・音楽を表示するタイプの遠隔訪問ロボットを試作した。大型 TV に顔や Youtube の映像・音を映すことにより、人がそこに居る臨場感を出せたと思っている。また、システムは全て市販のものを用いて構成されており、システムに新規性は無い。インターネットの利用法の1例を示すものである。

本報告では、試作した大型 TV 版遠隔訪問ロボットのシステム概要を説明し、このシステムの使用例を紹介する。



図1 移動型遠隔訪問ロボット

連絡先:辰野恭市, 名城大学・電気電子工学科, 〒468-85026
名古屋市天白区塩釜口1－501, tatsuno@meijo-u.ac.jp

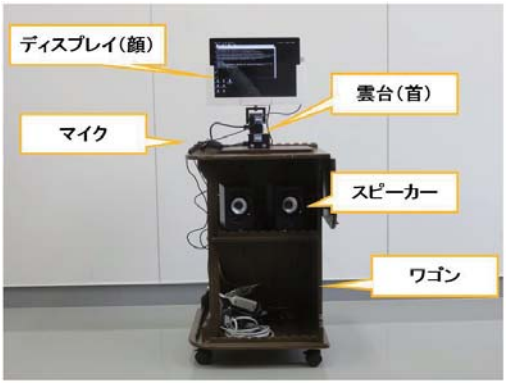


図2 固定型遠隔訪問ロボッ

2. システムの概要

図3に試作した大型 TV 版遠隔訪問ロボットのシステムの構成を、ロボット側にはノートパソコン・大型 TV・マイクがあり、遠隔の操作側にはパソコンがある。操作側のパソコンからリモート・ディスクトップ(TeamViewer)を用いてロボット側のパソコンを操作し、Skypeで大型テレビに顔を出してTV電話をしたり, ウェブブラウザ(Google Chrome)を用いてYoutubeにアクセスし、音楽や観光案内を大型TVで映す。

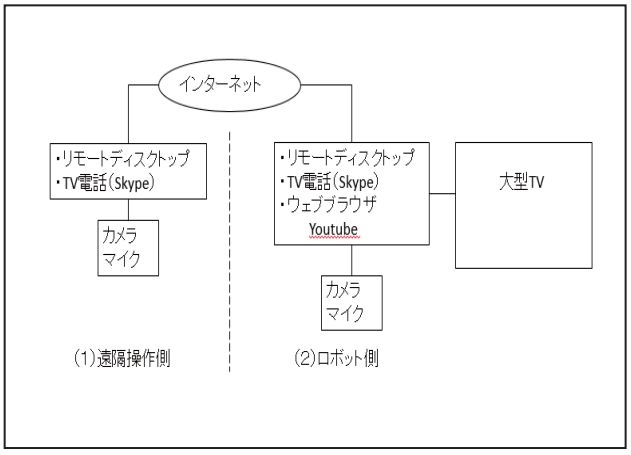


図3 システム構成

3. システムの使用例

提案するシステムは、利用の仕方にポイントがある。そこで、1)独居高齢者を訪問する場合と、2)介護センターで Youtube を利用したコンサートをする場合の使用例を以下に示す。

1)独居高齢者を訪問する場合

・遠隔操作側の人が、自分のパソコンのリモート・ディスクトップ(TeamViewer)を起動し、ロボット側のパソコンを操作できる状態にする。

・ロボット側の Skype を立ち上げ、遠隔側の自分のパソコンで TV 電話を掛けて、TV 電話で通話できるようにする。図4にその時の様子を示す。



図4 TV 電話(Skype)の様子

・遠隔操作側からカメラを操作し、GUI を図5に示す。カメラはパン・チルト・ズームができ、部屋の様子や顔色を見ることができる。

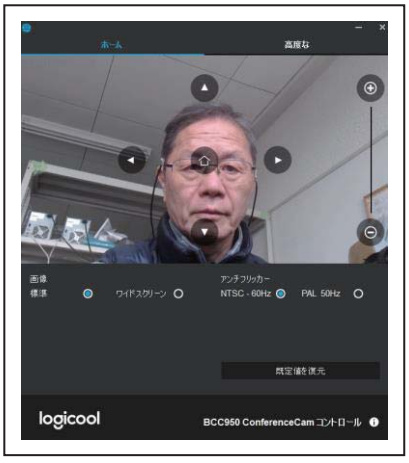


図5 カメラ操作の GUI(Logicool)

・遠隔操作側から、Youtube でクラシックやジャズの BGM (BackGround Music)を掛けることも出来る。

・また、遠隔操作側から、ネット買物を支援できる。(図6)



図6 イオンのネットショップの画面

2)介護センターで Youtube を利用したコンサートをする場合

介護センターでの使用方法も、独居高齢者を訪問する場合と殆ど同じである。大型 TV の前に10人程度の入居者が居て、コンサートの司会を遠隔からおこなうという設定で話をする。

・ロボット側の Skype を立ち上げ、遠隔側の自分のパソコンで TV 電話を掛けて、TV 電話で通話できるようにする。このとき、司会者の顔はロボット側のパソコンに映す。

・ウェブブラウザで youtube を起動し、大型 TV に映す。コンサートの企画に沿って、クラシック・歌謡曲・ジャズ・シャンソンなどの曲を掛ける。

2018 人工知能学会全国大会では、鹿児島に所縁のある長渕剛のコンサートを Youtube で試みる予定である。(図7)



図7 長渕剛の Youtube の画面

4. 結果・考察

提案するシステムは、通常のパソコンとリモートディスクトップ・ウェブブラウザを用いた通常皆さんが使っているパソコンの環境である。TV電話やYoutubeを使って、楽しいひと時を演出することが普及のポイントである。

これまで少し実験をおこなった結果を挙げておく。

・画像よりクリアな音を出すことが重要である。マイク・スピーカーの選定は大きなポイントである。

・通信遅れがあるために、遠隔側とロボット側で一緒にカラオケは出来ない。

・TV電話のアイコンタクトが不十分である。

・TV 電話は、受け入れられる人と極端に嫌がる人がいる。

今後、継続して使用し、このようなシステムが人々に受け入れられ、楽しく穏やかに過すための支援ができるか評価していく。

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

対話ロボットの不適切応答軽減のためのユーザ発話の解析

An analysis of user's utterance to reduce inappropriate response of chat robot

青柳翔大 *1 Shodai Aoyagi　平田和暉 *2 Kazuki Hirata　高谷智哉 *3 Tomoya Takatani　下川原 (佐藤) 英理 *4 Eri Sato-Shimokawara　山口亨 *5 Toru Yamaguchi

*1*2*4*5首都大学東京 Tokyo Metropolitan University　*3トヨタ自動車株式会社 Toyota Motor Corporation

In recent years, aging society has advanced in various countries around the world, and it is particularly serious in Japan. The problem of elderly living alone is that the frequency of conversation is low, which is the cause of depression. Therefore, as a previous study, an experiment of a chat robot for elderly people was conducted. The current problem is that the robot makes an utterance that does not match the Time, Place and Occassion. In this research, when a robot make an utterance that does not match the season, we proposed a method to acquire correct seasonal information from the user by robot uttering recovery sentence.

1. はじめに

1.1 研究背景

近年，世界各国で高齢化が進んでおり，特に日本国内の高齢化は深刻である．一人暮らしの高齢者の問題として，家でのコミュニケーションの頻度が低く，社会との接点が減少していることが挙げられる．会話の減少はうつ傾向を引き起こす原因となるため，その予防のためにも会話を行うことが重要な要素であると考えられる．そこで先行研究として，高齢者を対象とした対話ロボットの実験が下川原ら [1] [2] によって行われた．また，平田ら [3] の研究では，ユーザの発話からロボットの応答文として再利用可能な文を抽出するにあたり，そのフィルタリング手法を検討している．しかし，現時点では高齢者との十分な会話ができているとは言い難い．その理由の一つとして，図 1 に示すように，ロボットが時節にそぐわない発話をしてしまうということが挙げられる．ロボットの発話文には時節に制限される文とされない文がある．「タバコは本当に健康に悪いですよね」という文は通年でいつでも発話して良い文章であり，時節に制限されない．これに比べ，「暑い夏が続いてきましたね。泳ぎに行きませんか?」という文は夏での発話が想定される文章で，時節に制限される．表 1 を見るとわかる通り，ロボットが 2 月にこの発話をしてしまっており，時節にそぐわない発話となっている．時節にそぐわない文をロボットが発話した際，ユーザとロボットの会話の流れが途切れてしまうのはもちろん，このような不適切な発話が一度でもあると，ロボットと会話するモチベーションが低下してしまうことも分かっている．このためロボットが時節を理解し，その状況にあった発話をする必要がある．

1.2 研究目的

現在，ロボットが時節にそぐわない発話をしてしまう問題を回避するために，ロボットの発話文が時節に制限される文か否かのタグ付けを手動で行い，時節に制限される文かどうかの判断の自動化を目指している．教師データを作成して評価を行っているが，100 %の精度は難しい．その原因として，時節に制限される文か否かのタグ付けの評価は人の主観によって行われており，そのため，このタグ付けの基準が人によって異なるからである．また，現時点ではその文がどの時期での発話に適したものであるかということは決定できていない．季節の判断や認識には，個人の主観が多く含まれ，住んでいる場所によっても左右される．そのため個々のユーザにあった時節のタグを付与する必要がある．そこで，ロボットとの対話の中で，時間や季節の判断をユーザ自身に行なってもらうことはできないかと考えた．ユーザに時間や季節の判断をしてもらい，取得した時節情報によってリアルタイムに時節タグを更新する．ロボットの発話が個々のユーザに合ったものとなるため，そのユーザに合ったコミュニケーションが可能になり，ロボットが時節にそぐわない発話をしてしまう問題を回避することができる．本研究では，ロボットが時節にそぐわない発話をしてしまい，ユーザが違和感を感じた場合，その時のユーザの返答にどのような



図 1: 対話ロボットにおける時節の例

表 1: 時節に制限される文と時節に制限されない文

	timestamp	ロボットの発話
時節に制限される文	2015/2/22 9:38	暑い夏が続いてきましたね。泳ぎに行きませんか?
	2014/12/9 13:23	暑い夜は、寝るときに、扇風機を使います。
時節に制限されない文	2015/2/23 21:19	タバコは本当に健康に悪いですよね
	2014/12/1 14:39	歌を歌うのは大の苦手で、音痴なんです。

連絡先: 青柳翔大, 首都大学東京, aoyagi-shodai@ed.tmu.ac.jp

特徴が見られるのかを確認する．またロボットがリアスキングすることによってユーザから時節情報が得られるかどうかを確認する．時節情報が得られた場合，得られた情報にはどのような特徴があるのか，タグ付けの自動化は可能であるのか，ということを実験を通して検討する．

2. 関連研究

コンピュータにおける時節の理解に関する研究

コンピューターに時間を理解させる為の研究として，野村ら [4] の『時間の常識的判断メカニズムとその未知語処理』や，篠原ら [5] の『常識判断に基づく会話意味理解方式』がある．両研究では，人間とコンピュータとの双方向会話によるコミュニケーションのために，コンピュータによる時間の理解，それによる会話意味理解の手法を提案している．また，文の時間や季節を判断する関連研究として，小畑ら [6] の『単文の名詞と動詞から時間／季節を判断するメカニズム』や，吉川ら [7] の『文の意味を考慮した常識的時間判断システムの構築』がある．これらの研究では文の意味と文中の単語によって、時間判断、季節判断を行うシステムの構築を目標としている．このように，人間とコンピュータとの双方向会話によるコミュニケーションのために，コンピュータに時節を理解させることは必要不可欠な要素であることがわかる．しかし，これらの研究で使用されているデータは人手で判断されたクローズドデータであるため，そのデータを教師データとして評価することは難しい．また，時節の判断や認識は，住んでいる場所や個人の主観を多く含むため，ユーザに合った時間の判断をロボットに理解させる必要がある．本研究では，対話中にユーザにリアルタイムでリアスキングし直接評価してもらうことで，この問題を回避できる．

3. 提案手法

ロボットの不適切な応答を軽減するために，まずユーザが不満や違和感を感じた原因をロボットの発話からたどる必要がある．ユーザが違和感を感じるロボットの発話の種類はいくつかある．時節にそぐわない発話や，同じ発話の繰り返し，急な話題転換など原因は様々である．本研究ではロボットによる時節にそぐわない発話に焦点を置く．ロボットが時節にそぐわない発話をしてしまい，ユーザが違和感を感じた時に，図 2 に示すような，ロボットが自身の誤った発話に対してユーザにフォローを入れるような手法を提案する．季節を連想させる「キーワード」，時節にそぐわない発話をするための「テンプレート文」，ユーザから正しい時節情報を聞き出すための「リカバリー文」の 3 つを用意した．これらを使用することで，ロボットが時節にそぐわない発話をした際のユーザの返答の特徴を解析し，リアスキングを行うことでユーザから正しい時節情報を獲得できるのではないかと考えた．

a. キーワード

表 2 に示すような，季節を連想する単語をキーワードとして用意した．実験日の季節に適する単語，実験日の季節より前の単語，実験日の季節より後の単語の 3 種類を用意した．その 3 種類にそれぞれ，季節がわかりやすく誰もが共通認識しているもの，季節がわかりづらく人によっては認識が異なったり知らなかったりするものの 2 種類を用意した．計 6 種類のキーワードを使用する．現在の時節に適するキーワードも使用することで，現在の時節に適さないキーワードとの違いを確認できるようにする，季節がわかりやすい単語とわかりにくい単語をそれぞれ用意し，その 2 つでユーザの返答に違いが見られるかも確認する．

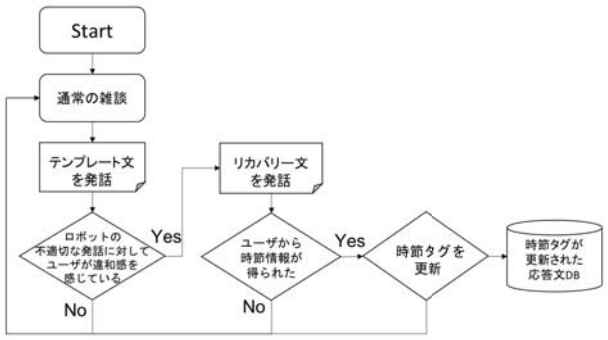


図 2: 提案手法におけるシステムフロー

表 2: 季節をもったキーワード

	季節がわかりやすい	季節がわかりにくい
過去	海	ブルーベリー
現在 (10 月)	紅葉	さんま
未来	クリスマス	ぶり

b. テンプレート文

時節にそぐわない文の発話をする際のテンプレート文として，表 3 に示すような文を使用する．用意したキーワードを入れて発話する．キーワードのカテゴリごとにテンプレート文を用意した．「この季節は」，「もうすぐ」，「そろそろ」といった単語と，現在の季節に適さないキーワードを組み合わせることで，現在の時節にそぐわない発話文を作ることができる．このテンプレート文を対話中にロボット側が発話して，ユーザからどのような反応があるか，特徴を見る．

表 3: 時節にそぐわないテンプレート文

カテゴリ	テンプレート文	例
イベント	この時期は [*] で人があふれています この時期になるともう [*] ですね もうすぐで [*] ですね	紅葉狩り クリスマス
食べ物	この時期だと [*] がおいしいです [*] がおいしい季節です [*] が食べたくなる季節ですね	さんま ぶり ブルーベリー
場所	そろそろ [*] に行きたい季節です [*] には行きましたか この時期は [*] が混んでいますね	海 プール

c. リカバリー文

ロボットの時節にそぐわない発話に対して，ユーザが指摘してきたとき，ロボットからユーザへのリアスキングとして発話するリカバリー文を表 4 に示す．テンプレート文に対応したキーワードを入れて発話する．これによりユーザから正しい時節情報が得られるのかを確認する．返答が得られた場合，どのような特徴が見られるのか，また，リカバリー文によってユーザの返答に違いは見られるのか，キーワードのカテゴリによってユーザの返答に違いがあるのかも確認する．

表 4: 時節情報を聞き出すためのリカバリー文

イベント	[*] はいつだったかな [*] はいつ頃だったかな [*] は何月だったかな
食べ物	[*] はいつが旬だったっけ [*] がおいしい季節はいつだっけ [*] はいつがおいしいんだっけ
場所	[*] にはいつ行きましたか [*] にいつ行きますか [*] が賑わっているのはいつですか

4. 実験

4.1 実験の流れ

対話ロボットで実際に実験するのは難しいため，Wizard of Oz 法での対話を行った．被験者にはロボットと対話をしてもらうと伝えている．図 3 に示すように，対話を行う中で、ロボット側 (実験者）が意図的に「“キーワード” を含む，時節にそぐわない文」を発話し、直後のユーザ（被験者）の反応を確認する．ユーザから反応が得られた場合に，時節情報を聞き出すためのリカバリー文を発話する．実験条件を表 5 に示す．被験者数は 6 人，情報系の学生でタイピングには支障がなく，ほぼリアルタイムでの対話が可能である．対話時間は 1 人当たり 30 分で，その中でキーワードを 6 つ使用しながら対話を行う．その結果，合計で 208 件，1 人当たり 35 タームの対話ログデータを取得することができた．

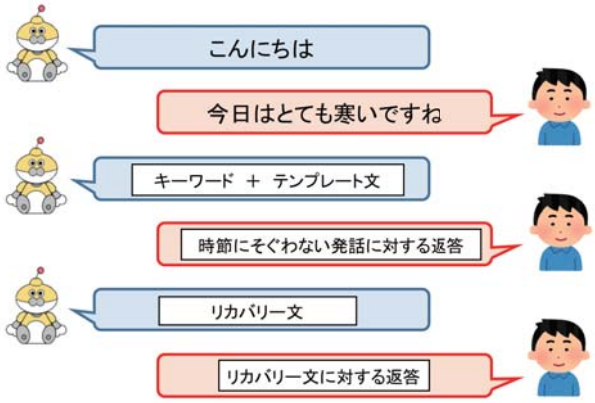


図 3: 実験の流れ

表 5: 実験条件

実験日	2017/10/12-20
被験者数	6 名
実験時間	30 分
データ数	208 件（35 ターム/1 人）
話題	フリー
実験機材	Windows10 PC 2 台

4.2 被験者への指示

実験するにあたって，被験者に 2 つの指示を出した．1 つ目は，テキストチャットでの対話ということで，発話は交互に行い，連続送信を禁止した．連続して発話してしまうと，どの発話に対しての返答か曖昧となってしまう．人及びロボットの発話の直接的な原因を明確にする．2 つ目は，対話中に違和感を感じたロボットの発話と，違和感に感じた理由を実験後のアンケートで回答してもらうため，違和感を感じた時間を控えておいてもらった．対話終了後，その時間の対話ログを見て，違和感を感じたロボットの発話及びその理由を記入してもらう．ユーザがロボットのどのような発話に違和感を感じるのか，なぜ違和感を感じるのか調べることができる．

5. 結果と考察

5.1 テンプレート文に対するユーザの返答

5.1.1 ユーザが違和感を感じるタイミング

時節にそぐわないテンプレート文に対するユーザの返答を表 6 に示す．網掛け部は，ユーザがロボットの時節にそぐわない発言について指摘してきたものである．赤字は，ユーザがロボットの時節の誤りについて違和感を感じたもの，青字は，急な話題転換などのその他の理由でユーザが違和感を感じたものを示している．季節がわかりやすいキーワードとわかりにくいキーワードを比較するとわかる通り，時節にそぐわないテンプレート文を発話した場合，季節がわかりにくいキーワードに比べ，季節がわかりやすいキーワードの方がユーザはロボットの時節の誤りに対して違和感を感じている．時節の誤りについて違和感を感じた場合，ほとんどのユーザがロボットに指摘をしたという結果となった．違和感を感じているが指摘はしていない，という場合については，アンケート結果によると，「ロボットだから」という理由で誤った発話を許容していることがわかった．“ロボットの発話に違和感を感じる” = “ロボットに対して指摘する” とは限らないということがわかる．同じテンプレート文だとしても，ユーザによって時節の誤りを指摘する人と指摘しない人，また，違和感を感じる人と感じない人がいることもわかる．その理由として，個人の認識の違いであったり，季節を知っているか否かであったりが考えられる．また，時節の誤りに対してではなく，急な話題転換に違和感を感じてしまうユーザも多いということが実験後のアンケートによってわかった．

5.1.2 ユーザの返答から得られそうな特徴

時節が実験日より前の単語である「海」「スイカ」に対するユーザの返答を見てみると，“もう～ない” というキーワードがユーザの返答から見られた．時節が実験日より後の単語である「クリスマス」に対するユーザの返答を見てみると，“まだ” や “早い”，“先” などのキーワードが見られた．これらのキーワードがユーザの返答から見られた場合，ユーザはロボットの時節にそぐわない文に対して指摘したと思われ，そのあとのリカバリー文を発話すべきという判断できるのではないかと考えられる．

5.2 リカバリー文に対するユーザの返答

ユーザがテンプレート文に対して指摘してきた場合に，ロボット側はリカバリー文を発話した．リカバリー文に対するユーザの発話を表 7 に示す．キーワードのカテゴリによって，ユーザの返答に特徴が見られた．「スイカ」や「ブリ」のように，キーワードが食べ物の場合，” いつ ” という問いかけに対して，“夏”，“冬” などの四季の語が得られた．今回の結果で

表 6: テンプレート文に対するユーザの返答

	季節がわかりやすいキーワード				季節がわかりにくいキーワード		
	過去 (10 月以前)		現在 (10 月)	未来 (10 月以降)	過去 (10 月以前)	現在 (10 月)	未来 (10 月以降)
	海	スイカ	紅葉	クリスマス	ブルーベリー	さんま	ぶり
user01	なぜですか		今が一番人多いでしょうね	まだ早いと思います	ブルーベリー好きですか	ブリもいいですけどやはりこの時期なら秋刀魚がおすすめです	ブリもいいですけどやはりこの時期なら秋刀魚がおすすめです
user02	寒いです 海には行きました		いいこと言うじゃん	いや、さすがにまだ早い	ブルーベリー！目にいい！	秋刀魚の収穫量が減っているらしいよ	そうなんか、魚食べてねぇなあ最近
user03	今はそんなにいないんじゃないですかね		食べ物は好きなんですけど紅葉狩りはあまりやったことがないですね	クリスマスはもうちょっと先じゃないかな	リンゴとぶどうもおいしいですよ	秋刀魚は好きですね	ブリって 11 月なんですか
user04	そうですね。魚が賑わっています	スイカは夏ではないですか？	紅葉狩りもいいですね。奥多摩の紅葉がきれいでした。	そうですね。予定はもう決まっているのですか？		いいですね。でも秋刀魚は骨が多いですよね.	ぶりおいしいですよね
user05	入水者はそんなにかな	スイカはもう見ない頃よね	だろうねえ	クリスマスになると少し早いんちゃう？		おいしいよね	ブリって夏じゃなさそう
user06	最近は寒くてそんな気にはなりませんよ	もうそんな季節じゃないですよ	もう紅葉も散るころじゃないですか	まだハロウィンも終わってないですよ		旬ですもんね	捕った魚を食べるんですか

は，食べ物の時節情報をユーザから聞き出すと，春夏秋冬のいずれかが得られたため，食べ物の旬に対して「○月です」という返答はあまりないと考えられる.「海」や「プール」のようにキーワードが場所の場合，“いつ”という問いかけに対して，“夏”という季節のキーワードが得られた．加えて、“7 月”，“8 月”のように月での返答も得られた.「クリスマス」のようにキーワードがイベントの場合 “いつ” という問いかけに対し，“12 月 25 日” という具体的な日付が得られた．また，“12 月” という返答もあった．以上のことから，キーワードのカテゴリごとに，ユーザから得られる時節情報に特徴が見られることがわかる．また，食べ物の時期を○月で尋ねたり，イベントの時期を季節で尋ねたりするように，リカバリーの仕方によって，ユーザから得られる時節情報の形をキーワードのカテゴリに関わらず，変えることができると考えられる．

表 7: リカバリー文に対するユーザの返答

	過去		現在	未来	
	海（プール）	スイカ	紅葉	クリスマス	ブリ
user01				12 月 25 日です	
user02	夏だよ			12 月 25 日だよ	
user03	海はやっぱり 8 月がいちばん人が多いと思いますよ			クリスマスは 12 月です	寒ブリって聞くくらいだから、確かに冬な気がしますけどあまりよく知らないんですよね
user04		夏だと思います			
user05	7月とかならもう少しいるのかも	夏やな～		12 月 25 日よ	冬なのでは
user06	夏ですね 7 月、8 月ころです	夏です	9 月下旬とかじゃないですか	クリスマスは 12 月 25 日です	

6. おわりに

本研究では，ロボットが時節にそぐわない発話をした際にリカバリーを入れることで会話の流れを戻し，ユーザから正しい時節情報を得るような手法を検討した．ロボットの時節にそぐわない発話に対するユーザの返答には,「もう～ない」や「まだ～」などの否定の語が含まれていた．ロボットがリアスキングをすることにより対話の中でユーザから得られた時節情報は,「○月○日」や「○月」,「夏」といった様々な形で得られ，キーワードのカテゴリによってその特徴が分かれた.「スイカは夏が美味しいですね」という文に対して自然言語処理の技術を用いて，“スイカ：8 月，夏” というようにタグを付与することで，ユーザとの対話中にリアルタイムで応答文データベースの時節タグ自動更新が可能になると考える．本研究より，ユーザの違和感を感知してリカバリーを行うことで，対話の中でユーザから正しい時節情報を取得することが可能であることが示唆された．この手法を用いることで，住んでいる場所や個人の感性の違いにより季節の認識が違う場合でも，それぞれのユーザに合った時節情報を得ることが可能となる．本研究では季節を区別することで実験を行なったが，1 日の朝昼晩の違いでも時節に制限される文とそうでない文があることは明らかである．今後の展望として，1 日の時間の中でのロボットの発話に着目した研究を行なっていきたい．

参考文献

[1] Eri Sato-Shimokawara, S. Nomura, Y. Shinoda, H. Lee, T. Takatani, K. Wada, and T. Yamaguchi.“A cloud based chat robot using dialogue histories for elderly people.” Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on. IEEE, 2015.

[2] 下川原 (佐藤) 英理, 篠田 遥子, 李 海妍, 高谷 智哉, 和田 一義, 山口 亨. “高齢者と音声対話ロボットの雑談履歴の解析”. 日本ロボット学会誌, 34(5):309-315, 2016.

[3] 平田 和暉, 下川原 英理, 高谷 智哉, 山口 亨. 対話ロボットの構築に向けた対話ログのフィルタリング手法の検討”, 第 17 回 計測自動制御学会システムインテグレーション部門講演会 (SI2016), 講演番号 2C1-5 (2016/12/15-17).

[4] 野村 理樹, 渡部 広一, 河岡 司. “時間の常識的判断メカニズムとその未知語処理”. 情報科学技術フォーラム一般講演論文集 2002(2), 141-142, 2002-09-13

[5] 篠原 宜道, 渡部 広一, 河岡 司. “常識判断に基づく会話意味理解方式”. 言語処理学会第 8 回年次大会発表論文集, 2002, 2002

[6] 小畑 陽一, 渡部 広一, 河岡 司. “単文の名詞と動詞から時間／季節を判断するメカニズム”. 情報処理学会研究報告知能と複雑系（ICS） 2001(1(2000-ICS-123)), 91-95, 2001-01-10.

[7] 吉川 義紀, 芋野 美沙子, 土屋 誠司, 渡部 広一. “文の意味を考慮した常識的時間判断システムの構築”, 情報科学技術フォーラム講演論文集 13(2), 235-236, 2014-08-19

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

子育て環境の心理的負担軽減を目指したパートナーロボットの開発

Development of a partner robot aiming at reducing the psychological burden of child-care environment

土屋陽介[*1] Yosuke Tsuchiya　内山純[*2] Jun Uchiyama　平社和也[*2] Kazunari Hirakoso　竹島大智[*2] Daichi Takeshima　関田理花[*2] Rika Sekita

張進男[*2] Zhang Jinnan　近藤嘉男[*2] Yoshio Kondo　成田雅彦[*2] Masahiko Narita

[*1] 東京通信大学 Tokyo Online University　[*2] 産業技術大学院大学 Advanced Institute of Industrial Technology

In recent years, parents with stress on child rearing are increasing in Japan. This is considered to be caused by parents who are raising their children cannot gain cooperation from people around them by nuclearization of the family accompanying changes in industrial form, increase in single households, and changes in local communities. Therefore, we developed a prototype of robot and robot services with the aim of reducing the psychological burden of child rearing environment. Thereby we aim to build a childcare environment where parents who are raising children are full of smiles. In this paper, we report the concept of the robot, the system configuration, and the result of the demonstration.

1. はじめに

近年，国内において産業形態の変化に伴う核家族化や単独世帯の増加，地域コミュニティの変化[厚生労働省 2013]により，育児中の親が周囲に頼れずストレスがかかることが要因となって，自殺や幼児虐待が起こるなど深刻な社会問題となっている．育児中のストレスには様々な要因が考えられるが，なかでも乳児の泣きに対処できないことが，育児不安や自信喪失につながるといわれ，育児中の親は乳児の泣きに共に対処してくれる手助けを求めている[堀越 2016].この問題解決のため，「笑顔溢れる子育て環境を実現するパートナーロボット」[内山 2017]のコンセプトに基づくプロトタイプ開発を行った．本稿では開発したロボットの概要，システム構成，展示会でのデモについて述べる．

2. コンセプト

「笑顔溢れる子育て環境を実現するパートナーロボット」は，

- 子育てする親の心理的負担の軽減
- ロボットを通じた子育てへの参加

の 2 点を目的とし，

- 家族間のコミュニケーションの「きっかけ」となるロボット
- 子育て中の親が「ロボット自身」に共感し，「ひとりぼっちでない」ことを気づかせてくれる

の 2 点を基本コンセプト[内山 2017]としてプロトタイプの開発を行った．ロボットとインターネットを連携することで，大切な人との距離を縮め，「人と豊かに共生するパートナーロボット」の実現を目指す．

本ロボットは，図 1 のサービス概要図に示すようなサービスを提供することを目標としている．本サービスの前提として，片方の親が自宅で育児をし，もう片方の親は会社で仕事をしているという家庭を想定している．

「赤ちゃん」の泣き声，生後月，時間などから泣いている原因を判定し，「自宅で育児中の親」にその理由を知らせると同時に，ロボットは「腕」を動かしたり，顔の表情を変えたり，音を鳴らしたりすることで「赤ちゃん」を「あやす」．また「職場で仕事中の親」にもその情報が伝わり，ロボットを通して「赤ちゃん」とコミュニケーションを取ることができる．これにより「自宅で育児中の親」の心理的負担を軽減させる．

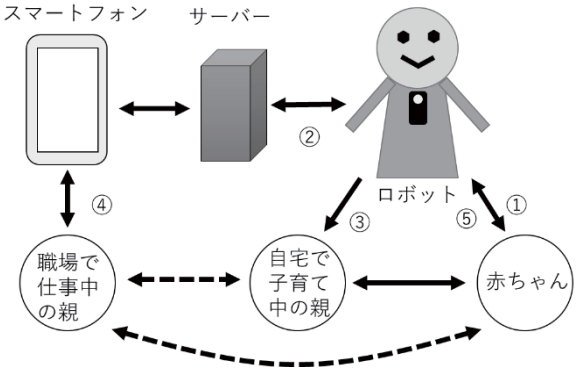


図 1　サービス概要図

① 「赤ちゃん」の泣き声を検知，音声，日時，映像を取得．
② 取得した情報をサーバに転送し，泣いている原因を推定．
③ 泣き声分析パターン（空腹，退屈，不快，眠気，ストレス）に対応する対処法を「自宅で育児中の親」に伝える．
④ 「職場で仕事中の親」がスマートフォンを通じてロボットから取得した映像を見ながらロボットを操作して「赤ちゃん」とコミュニケーションをとる．
⑤ 音と動作，表情により，「赤ちゃん」の注目を集め，一時的にでも泣き止ませる．

3. システム構成

上述のコンセプトを実現するために，プロトタイプを開発した．図 2 に開発したプロトタイプロボットの全体像を示す．本章では本ロボットのシステム構成について，ロボット本体とそれを使ったサービス側とに分けて説明する．

3.1 ロボット本体

本ロボットは以下の 3 点の機能を持つ．

連絡先：土屋陽介，東京通信大学，東京都新宿区西新宿 1-7-3，03-3344-2222，tsuchiya.yosuke@internet.ac.jp



図 4 ロボットの全体像

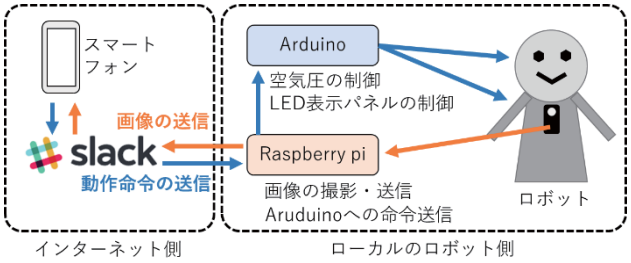


図 2 サービス側のシステム構成図

- 表情:LED 表示パネル
  顔部分に 127 個のフルカラーチップ LED を正三角形のマトリクス状に配列. 色と点灯個所の組み合わせでロボットの表情をダイナミックに表現可能.
- 仕草:空気圧駆動アーム
  ポリエステルフィルムを使用した蛇腹構造の空気圧アクチュエータ. 柔軟で自然な静穏動作を実現. 「赤ちゃん」が動くアームに触っても壊れにくい構造.
- 画像撮影:ウェブカメラ
  周辺状況(静止画)を撮影し, 配信.

本プロトタイプでは, LED 表示パネルおよび空気圧駆動アームはロボット外部に設置している Arduino で制御している(図 2 左側). この空気圧駆動アームでは空気ポンプと 2 つのサーボモータを利用し, 送気, 排気をコントロールすることで, 左右のアームを別々に動作できるようにしている. さらに, このロボット制御用 Arduino とロボットに内蔵しているウェブカメラを Raspberry pi と接続することにより, インターネットへの接続を実現している.

3.2 サービス

本プロトタイプでは 2 章で述べた図 1 サービス概要図④の「職場で仕事中の親」がスマートフォンを通じてロボットから取得した映像を見ながらロボットを操作して「赤ちゃん」とコミュニケーションをとるサービスの実装を行った. 本サービス実装のために, Raspberry pi を利用し, ロボットをインターネットに接続できる環境を構築した(図 3 参照). 本サービスの流れを以下に示す.

① ロボット内蔵のウェブカメラで定期的に画像を撮影する
② 撮影された画像をスマートフォンに送信する
③ 必要に応じてスマートフォンからロボットの動作命令を送信する

ここで, ②の画像の送信先および③のスマートフォンからの動作命令の送信には, チャットアプリの Slack を利用した. これは, 一般的に利用可能なアプリで両親ともに画像を簡単に共有できるという点, テキストチャットにて命令を送信できるという点による. 近年スマートスピーカーが普及しており, 「音声」による家電, ロボットのコントロールや, コミュニケーションがとれるサービスが増えている. しかし, 本サービスはコンセプトとして「職場で仕事中の親」がターゲットユーザーとなっているため, あえて「音声」ではなく「テキストチャット」によりコミュニケーションが取れる仕組みを採用した. ロボットに予め準備された特定のコマンドを Slack 上から送信することで, ロボットに「表情」と「仕草」の組み合わせによる表現を可能としている.

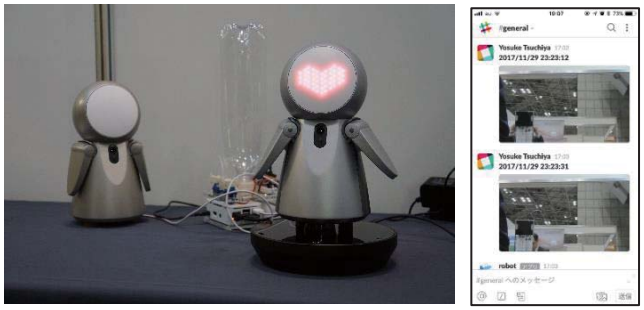


図 3 国際ロボット展 2017 でのデモの様子
(左:デモブース, 右:Slack の画面)

4. 展示会でのデモ

2017 年 11 月 29 日から 12 月 2 日で開催された国際ロボット展 2017 ロボットサービスイニシアチブのブースにて本ロボットのデモを実施した. 図 4 にデモの様子を示す.

本デモでは図 2 のロボット本体の展示および, ④のスマートフォンとの連携サービスのデモを実施した. ロボットから Slack に画像を送信し, 図 4 右の図のようにスマートフォン上でその画像を確認することができる. また, 同じく Slack からテキストチャットで動作コマンドを送信することでロボットが表情を変えたり, 腕を上下に動かしたりすることができる. ということをデモで紹介した.

デモブースで実施したアンケート結果では, ロボットが育児に加わることに対して大多数が肯定的な意見を持っていることがわかったが, 本ロボットの実現性に対する厳しい指摘もあった.

5. まとめ

本稿では, ロボットを利用した子育て環境の心理的負担軽減を目的に, ロボットおよびロボットサービスを開発した. 本ロボットのコンセプトは, 核家族化などにより周りに頼ることができず, 一人で子育てせざるを得ない状況にある親のために, 「赤ちゃん」が泣きだした時, ロボットによって, 泣き声を少しでも小さくしたり, 少しの間だけでも泣き止ませたりすることで, 子育てで悩んでいる親の笑顔を取り戻したいという願いから生まれた.

そのプロトタイプとして, ロボット本体の開発および, チャットを利用したコミュニケーションのサービスを実装した. 自宅で一人, 子育て中の親が「ひとりぼっちでない」ことを気づかせ, 両親で共感しあえるロボットの存在が子育て環境の改善につながると我々は考えている. 今後も開発・実証実験を続け, 「人と豊かに共生するパートナーロボット」の実現を目指していく.

参考文献

[厚生労働省 2013] 厚生労働省: 平成 25 年版労働経済の分析, http://www.mhlw.go.jp/wp/hakusyo/roudou/13/dl/13-1-4_02.pdf (2018 年 3 月 5 日最終閲覧)

[堀越 2016] 堀越摂子, 常盤洋子, 國清恭子, 高津三枝子: 生後 1 か月児の泣きに関する母親の認識, 北関東医学 Vol. 66, No. 1, p. 23-30, 2016

[内山 2017] 内山純, 竹島大智, 関田理花, 張進男, 平社和也, 近藤嘉男: 「笑顔溢れる子育て環境を実現するロボット」コンセプト, 第 35 回 日本ロボット学会学術講演会, 2B1-03, 2017

[2E4-NFC-1b] [NFC Session] NFC-1 CRSP (Integrated Platform for Cloud-based Robot Services)

座長:成田 雅彦（産業技術大学院大学）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room E (4F Queen)

[2E4-NFC-1b-01] Predicting a Pedestrian Trajectory Using Sequence to Sequence Learning for Mobile Robot Navigation

Natsuki Sakata[1], Yuka Kinoshita[1], ○Yuka Kato[1] （1. Tokyo Woman's Christian University）

5:20 PM - 5:40 PM

[2E4-NFC-1b-02] Classification of Utterances Using Temporal Relation in Cloud-based Dialogue System

○Kazuki Hirata[1], Eri Shimokawara[1], Tomoya Takatani[2], Toru Yamaguchi[1] （1. Tokyo Metropolitan University, 2. Toyota Motor Corporation）

5:40 PM - 6:00 PM

[2E4-NFC-1b-03] Structure of RT system to realize various robot services

○Nobuto Matsuhira[1], Takamasa Ikeda[1], Mirai Shimoyama[1], Satoshi Okano[1], Masahiko Narita[2], Yosuke Tsuchiya[2], Yuka Kato[3] （1. Shibaura Institute of Technology, 2. Advanced Institute of Industrial Technology, 3. Tokyo Woman's Christian University）

6:00 PM - 6:20 PM

[2E4-NFC-1b-a] (Panel discussion) The next direction of "Integrated Platform for Cloud-based Robot Services Project"

Masahiko Narita, Tokuro Matsuo, Nobuto Matsuhira, Tatsuno Kyoichi, Toru Yamaguchi, Yuka Kato

6:20 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

移動ロボットのためのSequence to Sequenceモデルを用いた歩行者移動軌跡の予測

Predicting a Pedestrian Trajectory Using Sequence to Sequence Learning for Mobile Robot Navigation

坂田 夏樹 *1 Natsuki Sakata　木下 裕香 *1 Yuka Kinoshita　加藤 由花 *1 Yuka Kato

*1東京女子大学
Tokyo Woman's Christian University

This paper proposes a method to predict the future trajectory of a pedestrian as sequence data by using massive trajectory records collected by various sensor devices. We aim to use the method for safely and efficient path planning of autonomous mobile robots in a human-robot coexisting environment. For the prediction, we use Sequence to Sequence learning, which is frequently used in the field of natural language processing. That enables to treat long-term sequence data. In order to verify the effectiveness of the proposed method, we conduct experiments using a dataset. As a result, we show that we can predict the trajectory sufficiently by converting the trajectory data to adequate sequence data.

1. はじめに

人・ロボット共存社会の実現に向け，ロボットと人との衝突回避が重要な研究課題になっている．我々はこれまで，Social Force Model（SFM）として知られる Helbing の歩行者モデル [Helbing 95] を拡張した人移動モデルを構築し，この問題を解決する手法の研究を進めてきた [Kato 17]．ここでは，空間の形状を基に人移動をシミュレーションするモデルを構築したが，環境に応じた人の意図（移動の目的等）はモデルに組み込まれず，十分な予測精度を得ることは難しいという課題が残った．

本稿では，この課題に対応するために，環境内に設置したセンサにより収集される人の移動履歴を基に，未来の人の移動軌跡を系列データとして予測する手法を提案する．近年，行動把握のためのセンシング技術が進展し，高い精度で，常時人移動データを取得する環境が現実味を帯びてきている．我々自身も，各種センシング技術とロボットサービスの融合に関する研究を行っており [Sakai 16]，本稿でも，これら大量のセンシング結果を，ロボットの経路計画に利用することを考える．特に，近年，自然言語処理やマーケティング分野等で，時系列データ予測技術が大きく進展しており，時空間データから人の行動を理解・予測する時空間行動理解技術の研究が盛んに行われている．本稿では，系列データを予測する機械学習アルゴリズムの 1 つである Sequence to Sequence モデル（Seq2Seq）[Sutskever 14] を利用し，移動軌跡を適切な系列データに変換することで，未来の移動軌跡の予測を実現する．

2. 関連研究

人の移動軌跡を系列データとして予測するには，大きく分けて 2 つの方法が考えられる．一つは，歩行者の移動傾向を数理モデルとして定式化し，このモデルに従い人の移動傾向を推定するものである．もう一つは，大量の観測データを利用し，対象環境における人の移動傾向を推定するものである．

人移動モデルを用いる手法として最もよく知られているのは，前述した Helbing の SFM [Helbing 95] である．これは，人と人，人と壁の間に，粒子間力のような心理的斥力と物理的斥力を定義し，歩行者の進行方向を決定するモデルである．その他，避難シミュレーションのためにセル・オートマトンモデルを用いて人の移動傾向を予測する手法 [Boukas 15] の提案や，ロボットと共存する環境での人移動モデルの定式化 [Kato 17] なども行われている．これらの手法は，適切なモデルを定義することができれば，事前に大量のデータ取得を行う必要がなく，適用が容易であるという大きな利点がある．一方，モデル化において考慮する必要があるパラメータ数が多く，精度の高いモデル構築が難しいという課題がある．

観測データを用いる方法は，環境ごとに大量の移動軌跡データを収集し，その結果から人の移動傾向を推定する方法である．人移動計測データから移動モデルを生成し，ロボットの経路計画に利用する方法 [Noguchi 12] などがある．近年では，機械学習アルゴリズム，特に自然言語処理技術の進展により，移動データを系列データとして学習させる研究が盛んに行われている．RNN（Recurrent Neural Network）を用いて移動軌跡の系列から目的地を推定する手法 [遠藤 16] や，系列データ間の相互作用を Social LSTM（Long short-term memory）としてモデル化する手法 [Alahi 16] などが提案されている．これらの手法は，事前に大量のデータ収集が必要になるというデメリットはあるが，環境に適したモデルの構築が可能であり，建物の幾何形状や環境要素もモデルに取り込むことができる．本稿でも，観測データを用いる手法を採用し，特に，系列から系列を予測する手法を用いる．

3. 提案手法

本稿では，人の移動軌跡を系列データとして扱い，観測した系列データから未来の系列データを予測する．予測には，Seq2Seq モデルを用いる．これは，LSTM encoder，LSTM decoder の 2 つの LSTM を組合わせて構成され，m フレームの系列データを入力すると，予測結果として n フレームの系列データが出力されるものである．モデルのイメージを図 1 に示す．

LSTM encoder は，入力された系列データ $\boldsymbol{x}_{t-m+1:t}$ に対し $\boldsymbol{c}$ を出力し，これが LSTM decoder の入力になる．また，LSTM encoder の隠れ変数 $\boldsymbol{h}_t$ は，LSTM decoder の隠れ変数の初期値となる．LSTM decoder は，入力された $\boldsymbol{c}$ に対

連絡先: 加藤由花，東京女子大学数理科学科，
yuka@lab.twcu.ac.jp

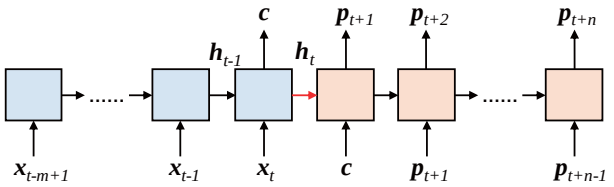


図 1: Seq2Seq モデルのイメージ．水色の部分が LSTM encoder，ピンクの部分が LSTM decoder に相当する．

し，$\boldsymbol{x}_{t+1}$ の予測に相当する $\boldsymbol{p}_{t+1}$ を出力する．同様に，$\boldsymbol{p}_{t+1}$ が LSTM decoder に入力され，その結果 $\boldsymbol{p}_{t+2}$ が出力される．この処理を繰り返し，最終的に系列データの予測値として $\boldsymbol{p}_{t+1:t+n}$ が得られる．

ここで，Seq2Seq モデルを用いて系列データを予測するためには，対象となる人の移動軌跡を系列データ $\boldsymbol{x}_{t-m+1:t+n}$ に変換する必要がある．本稿では，人の位置座標を 1 次元のグリッド空間上にマッピングし，人の移動をグリッド間の遷移とみなす．そして，人の移動軌跡を，グリッド ID の系列データとして構成することにした．

4. 実験

4.1 実験の内容

手法の有効性を検証するために，商業施設内を一般の歩行者が徒歩で移動するデータである ATC pedestrian tracking dataset [Brscic 13] を用いて，提案手法の評価を行った．これは，大阪の商業施設 ATC ショッピングモールにおけるのべ 3,758,348 軌跡（1 日平均 40,851 本）の人移動軌跡と，ユーザ ID などのメタデータを含んだ，約 1 年間（2012/10–2013/11 までの毎週水曜・日曜，9:40–20:20），通算 92 日分のデータである．複数の 3 次元距離画像センサーにより，人の位置が 10 ∼ 40 Hz（周期は人の密集度により異なる）で継続的に計測されている．データとしては，ユーザ ID，位置座標 [mm]，移動速度等が含まれる．本稿ではこのうち，ショッピングモールの一部のエリア（10 m × 10 m の大きさ）に含まれる 1635 本の軌跡を取り出し，軌跡ごとに $N = 160$ フレーム分の位置座標 $\boldsymbol{x}_{1:N}$ を抽出して利用した．パラメータとしては，$m = n = 15$ とし，LSTM の隠れ層の次元は 128 次元とした．なお，提案手法では，2 次元座標をグリッドに変換するため，入力データ $\boldsymbol{x}_t$ は 1 次元のデータになる．

入力データと教師データは，$\boldsymbol{x}_{1:N}$ から t を 1 フレームづつずらしながら，入力データとして m フレーム（$\boldsymbol{x}_{t:t+m-1}$），教師データとして n フレーム（$\boldsymbol{x}_{t+m:t+m+n-1}$）を続けて切り出すことで取得した．これにより，$N - m - n = 130$ 個の入力データと教師データの組が得られる．つまり，全体で 1635 本 ×130 個 = 212,550 組のデータを学習データとして利用した（ここでは，学習データを 9 : 1 に分割し，前者を訓練データ，後者をテストデータとした）．なお，重み更新のためのバッチサイズは 200 とし，100 イテレーション学習させた．

表 1: グリッド幅と Accuracy の関係

グリッド幅	Accuracy
0.3 m	0.02
0.5 m	0.61
1.0 m	0.66

評価指標としては，テストデータに対する Accuracy（教師データと予測データが系列として一致した割合）を用い，人の位置座標を 1 次元グリッド空間にマッピングする際の各グリッドの幅を，0.3 m，0.5 m，1.0 m と変化させたときの Accuracy を算出した．

4.2 実験の結果

各グリッド幅における Accuracy を表 1 に示す．また，系列予測結果の一例を以下に示す．

```
------------------------------------------------------------------------
グリッド幅 0.3（不一致）
入力：[271 271 271 302 302 302 302 302 302 302 301 301 301 301 301]
出力：[300 300 300 300 300 300 300 331 331 331 331 331 331 331 331]
予測：[304 304 304 304 304 304 304 304 304 304 304 304 304 304 304]
------------------------------------------------------------------------
グリッド幅 0.5（不一致）
入力：[ 73  73  73  73  73  73  73  73  73  73  73  92  92  92]
出力：[ 92  92  92  91  91  91  91  91  91  91  91 110 110 110 110]
予測：[ 92  92  92  92  91  91  91  91  91  91  91  91  91  91 110]
------------------------------------------------------------------------
グリッド幅 1.0（一致）
入力：[ 34  34  44  44  44  44  34  44  44  34  34  44  44  44  44]
出力：[ 44  44  44  44  43  43  43  43  43  43  43  43  43  43  43]
予測：[ 44  44  44  44  43  43  43  43  43  43  43  43  43  43  43]
------------------------------------------------------------------------
```

これらの結果から，グリッドの幅，つまり人の移動軌跡の表現方法により，予測精度は大きく異なることがわかる．グリッドの幅を小さく取る方が細かい動きの表現が可能だが，データを系列として扱う場合には次元が爆発し，学習がうまく進まない（0.3 m の場合）．一方，グリッドの幅を大きく取ると予測精度は上がる．しかし，これはフレームごとの移動がほとんど発生しないためであり，人の移動の粒度が適切に表現されているとは言い難い（1.0 m の場合）．適切なグリッド幅は，人体の大きさ，人の移動速度，フレーム周期に依存し，今回の実験条件では，0.5 m 程度が適切なグリッド幅であると考えられる．

なお，本稿では評価指標に Accuracy を用いたが，ロボットの経路計画での利用を考えると，系列が完全に一致する必要はない．例えば，上記の 0.5 m の例では，4 フレームで値の不一致が発生しているが，軌跡としての予測は十分な精度で行えていると考えることもできる．今後，系列データ間の距離等，より適切な評価指標の利用を検討していきたい．

5. おわりに

本稿では，Sequence to Sequence モデルを用いて，未来の人の移動軌跡を系列データとして予測する手法を提案した．データセットを用いた評価実験により，移動軌跡を適切な系列データに変更することで，十分な精度で軌跡の予測が可能であることを確認した．

今回の実験は，限定的なデータでの検証実験であったが，今後，グリッド幅に合わせた計測周期の調整，適切な系列データの表現方法，より長い系列の予測，評価指標の見直し，Seq2Seq2 以外のモデルの適用等を行っていく予定である．また，系列間の相互作用を考慮した手法を検討し，予測結果をロボットの経路計画に利用できるよう検討を進めていく．

謝辞

評価実験にあたっては，東京女子大学数理科学科の柴田ひかりさん，西村早也香さん，吉村美穂さんにご協力頂きました．ここに感謝の意を表します．また，本研究の一部は，JSPS 科研費 15K00137, 17K00366 の助成を受けたものです．

参考文献

[Alahi 16] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., and Savarese, S.: Social LSTM: Human Trajectory Prediction in Crowded Spaces, in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 961–971 (2016)

[Boukas 15] Boukas, E., Kostavelis, I., Gasteratos, A., and Sirakoulis, G.: Robot Guided Crowd Evacuation, *IEEE Trans. on Automation Science and Engineering*, Vol. 12, No. 2, pp. 739–751 (2015)

[Brscic 13] Brscic, D., Kanda, T., Ikeda, T., and Miyashita, T.: Person Position and Body Direction Tracking in Large Public Spaces Using 3D Range Sensors, *IEEE Trans. on Human-Machine Systems*, Vol. 43, No. 6, pp. 522–534 (2013)

[Helbing 95] Helbing, D. and Molnar, P.: Social Force Model for Pedestrian Dynamics, *Physical Review E*, Vol. 51, No. 5, p. 4282 (1995)

[Kato 17] Kato, Y., Nagano, Y., and Yokoyama, H.: A Pedestrian Model in Human-Robot Coexisting Environment for Mobile Robot Navigation, in *Proc. IEEE/SICE Intl. Symposium on System Integration (SII 2017)*, pp. 992–997 (2017)

[Noguchi 12] Noguchi, H., Yamada, T., Mori, T., and Sato, T.: Mobile Robot Path Planning Using Human Prediction Model Based on Massive Trajectories, in *Proc. IEEE Intl. Conf. on Networked Sensing Systems (INSS 2012)*, pp. 1–7 (2012)

[Sakai 16] Sakai, S., Kimura, S., Nomiyama, D., Ikeda, T., Matsuhira, N., and Kato, Y.: Classification of Age Groups using Walking Data Obtained from a Laser Range Scanner, in *Proc. the 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON 2016)*, pp. 5862–5867 (2016)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to Sequence Learning with Neural Networks, in *Advanced in Neural Information Processiong Systems 27 (NIPS 2014)*, pp. 3104–3112 (2014)

[遠藤 16] 遠藤 結城, 西田 京介, 戸田 浩之, 澤田 宏：長期的な状態依存性を考慮した移動軌跡からの目的地予測, 情報処理学会マルチメディア・分散協調とモバイルシンポジウム 2016 論文集, pp. 1524–1536 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

クラウドベースの対話システムにおける時節情報を用いた発話分類

Classification of Utterances Using Temporal Relation in Cloud-based Dialogue System

平田 和暉 *1 Kazuki Hirata　下川原 英理 *2 Eri Shimokawara　高谷 智哉 *3 Tomoya Takatani　山口 亨 *4 Toru Yamaguchi

*1*2*4首都大学東京 Tokyo Metropolitan University　*3トヨタ自動車株式会社 Toyota Motor Corporation

In recent years, research on dialogue system is prosperous. we developed spoken dialogue system as the previous research. One of the challenges of the system that we developed in previous research is that temporal relation cannot be considered. In order to solve this problem, we consider a system which automatically judges whether the temporal relation is contained in robot response sentences. We used two methods as the classification method. The first is simple classification by word pattern matching. The second is classification using MLP (Multi-Layer Perceptron.) As a result, the classification method using MLP was more useful than the word pattern matching. However, it was not practical with current precision. We found that more training data is necessary.

1. はじめに

近年，対話ロボットの研究が盛んに行われている．対話ロボットには大きく分けてタスク指向型対話を目的としたロボットと非タスク指向型対話を目的としたロボットの二つの種類に分類される．タスク指向型対話とは受付の案内や電話でのカスタマー対応等，特定の目的を達成するための対話である．非タスク指向型対話とは雑談のように，特定の目的がない対話の事である．我々は先行研究として非タスク指向型対話を目的とした音声対話システムを開発した．[1] 先行研究で開発したシステムの概要を図 1 に示す．対話にはロボットの応答文を格納したデータベースを用いる．話者の発話に対する返答として適切な応答をデータベースから選択してロボットが発話する事で，話者と会話を行う．ロボット応答文を格納したデータベースは端末内のホームサーバとクラウドサーバによって管理される．会話を行う際には，ホームサーバ内にあるデータベースを用いてロボットは会話を行い，その会話のログデータは逐次ホームサーバに蓄積される．そして，その日の終わりに全ての端末に蓄積した会話のログをクラウドサーバにアップロードし，会話データから得られる新しい情報をデータベースに追加した後，各端末のホームサーバ内のデータベースを更新するシステムを構築した．しかし，現状の対話システムでは話者との雑談が成立しているとは言い難い．我々はより人間らしい雑談を行う音声対話ロボットの実現を目指す．

2. 雑談時における時節の影響

先行研究で構築した対話システムを用いて高齢者を対象として，6 月，7 月に会話実験を行った．その実験において，被験者からロボットの応答に違和感を感じるという評価が得られた．具体的に被験者が違和感を感じた会話の例を表 1 に示す．表 1 より，ロボットが「今日は寒くて氷が張っていました」と応答を返している．しかし，この会話実験を行った時期は 6, 7 月であり，暖かい時期に行っている．実際の会話の環境にそぐわない応答をロボットが返している事が分かった．光田らの研究 [2] では，ロボットが人と会話を行う上で考慮すべき言外の情報は多様であり，また，言外の情報が会話に与える影響は

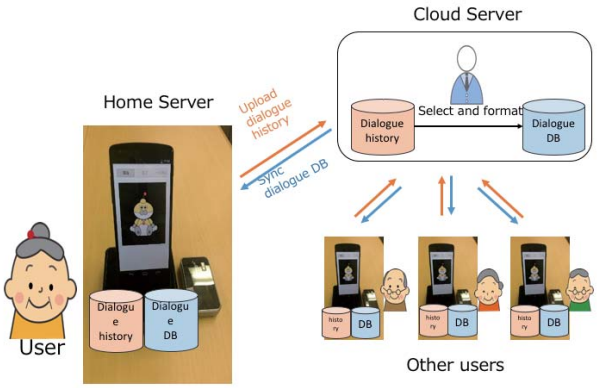


図 1: 先行研究として開発した音声対話ロボット

大きいと述べている．我々は会話に影響を与える言外の情報の一つとして時節情報に着目した．単語，または文に適切な時節情報（四季，日中の時間帯，日付など）を推定する研究は行われている．[3][4] しかし，これらの研究ではテキストに対し時節情報を推定しているが，そもそもそのテキストが時節情報を付与する必要があるのかは考慮していない．光田らの時期を判定するシステム [5] では，新聞記事からあらかじめ時節情報が含まれる表現，文を人手で抽出し，抽出したデータに対しどれだけ適切な時節情報を付与できるかを評価している．ロボットの応答文に対し時節情報を含む文であるかどうかを自動で判定する事ができれば，ロボットが時節にそぐわない発話をすることを回避できる．時節情報を含む文を本論文では「時節文」と定義し，本論文では以降，時節文と呼ぶ．また，ロボットの応答文が時節文と判定される要因となる単語を時節語と定義する．本論文ではロボットの応答文を時節文かどうかを判定する分類問題を検討することを目的とする．

連絡先: 平田 和暉，首都大学東京，hirata-kazuki@ed.tmu.ac.jp

表 1: 時節にそぐわない発話をした場合の雑談履歴.

U: ユーザ, R: ロボット	発話文
U:	こんにちは
R:	こんにちは
U:	今日は暖かいね
R:	今日は寒くて氷が張っていました (時期にそぐわない応答)
U:	そうねー 氷が張ると寒いねー
R:	春が好きです。桜が咲いてきれいですよね。
U:	早く春がくるといいね 私も春は大好きです
R:	一人暮らしですか？
U:	父と二人です
R:	今は小学校からパソコン習っているんですよね

3. 提案手法

時節文かどうかを判定する手法として，二つの手法を本論文で提案する．一つ目の手法として，時節語を用いた単語マッチングによる分類を提案する．また，二つ目の手法として多層パーセプトロン（MLP: Multi Layer Perceptron）を用いた分類を提案する．

3.1 単語マッチングによる分類

一つ目の手法として時節語を用いた単語マッチングによる分類を行う．時節語の例を表 2 に示す．時節語は 1〜12 月のいずれかの範囲，またはある特定の日を連想させる単語である事がわかる．

表 2: 時節語の例

時節語
盆踊り
花火
梅雨
コタツ
熱中症
冬至
スイカ

単語マッチングの処理を図 2 に示す．入力となる発話を MeCab [6] によって単語毎に分割する．そして，分割した単語リストの中に時節語が 1 単語以上含まれる場合，その文を時節文と分類する．また，分割した単語リストの中に時節語を含まな場合，その文を非時節文と分類する．上記のように，単語マッチングはシンプルな条件判定処理によって構成する．

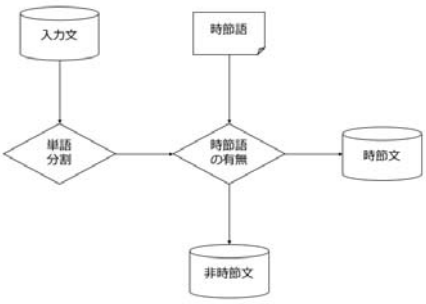


図 2: 単語マッチングの概要

3.2 MLP による分類

3.2.1 時節文の特徴抽出・学習

特徴抽出と学習の処理を図 3 に示す．単語マッチングと同様に，入力となる文を MeCab によって単語毎に分割する．そして，単語リストの先頭と末尾に空文字を二つずつ追加し，各単語を中心として 5 単語ずつ抽出する．そして抽出した 5 単語のリストに対し，各単語を Word2Vec[7] を用いて 100 次元のベクトルに変換した後，100 次元ベクトルを連結する事で 500 次元のベクトルを生成する．この 500 次元の特徴ベクトルを入力として学習を行う．また，その特徴ベクトルを生成した 5 単語の中に時節語を含む場合，その特徴ベクトルを正例として学習する．時節語を含まない場合は，その特徴ベクトルを負例として学習する．

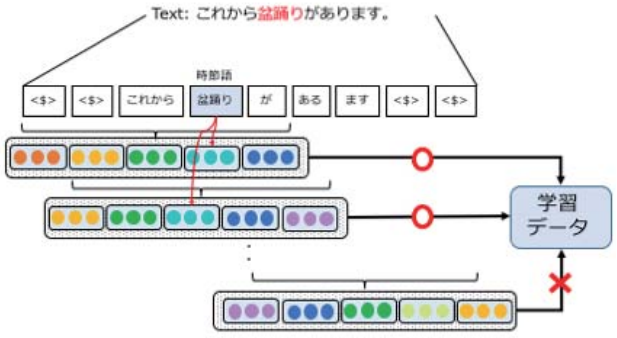


図 3: 特徴ベクトルの生成

学習器として用いた MLP の構成を図 4 に示す．入力層のノード数は 500，中間層はノード数が 355，隠れ層は 5 層，出力層のノード数は 500 である階層型ニューラルネットワークを構築した．活性化関数には ReLU（正規化線形関数），最適化手法には SGD（確率的勾配降下法）を用いた．出力層では入力の文が，どの程度時節文であるかを示す尤度と，どの程度時節文でないかを示す尤度を，それぞれ 0 から 1 の範囲で出力する．

3.2.2 時節文の判定処理

時節文の判定処理を図 5 に示す．未知の文に対し，前節で述べた特徴抽出方法と同様に，一文から複数の 500 次元ベクトル生成する．この 500 次元のリストを MLP の入力とする

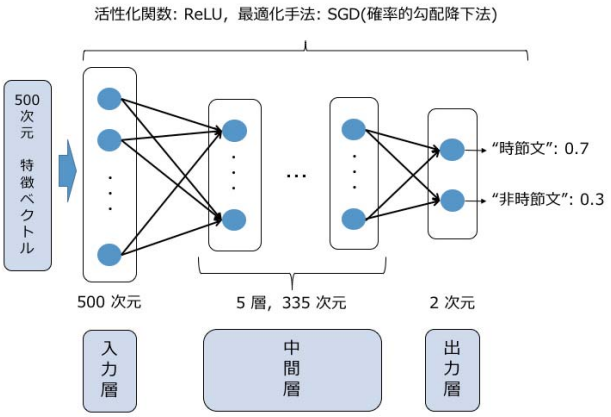


図 4: MLP の構成

ことで，各ベクトルに対する時節文である事を示す尤度を出力する．その中で時節文である尤度が最も高い値が閾値をより大きい場合，その文を時節文と判定する．今回は閾値値を 0.1 と設定した．

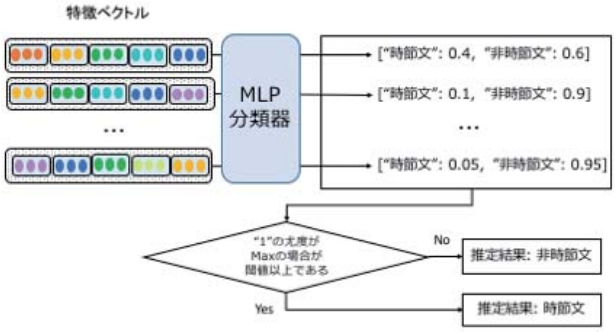


図 5: MLP による時節文判定の処理

3.3 提案手法に用いるデータ

提案手法における入出力の概要を図 6 に示す．評価データにはロボットの応答文が格納されている．学習データには時節文かどうかのアノテーションがされた文が格納されている．また，学習データにある時節文とアノテーションされた文から時節語を抽出する．単語マッチングには，抽出した時節語と評価データを入力として分類を行う．MLP の学習には，時節語と時節文かどうかのアノテーションを入力として生成した正例データ・負例データを使用する．MLP を用いた判定処理では，評価データと MLP の学習モデルを入力として分類を行う．

また，各提案手法において使用するデータ，処理の相違を表 3 に示す．単語マッチングでは時節語を必要とするのに対し，MLP では時節語，時節文，学習モデルを必要とする．また，単語マッチングでは MeCab を用いて単語分割したリストを入力とするのに対し，MLP では単語分割した後に Word2Vec によって変換した特徴ベクトルを入力とする．このような相違点があげられる二つの提案手法を比較し，評価する．

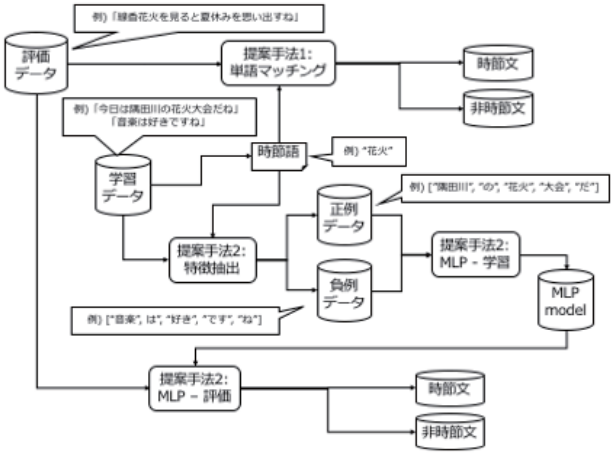


図 6: 提案手法における入出力データ

表 3: 各提案手法における使用するデータの相違点

	単語マッチング	MLP
時節語	必要	必要
時節文	不要	必要
学習モデル	不要	必要
MeCab による単語分割	必要	必要
Word2Vec による特徴ベクトル変換	不要	必要

4. 提案手法の評価実験

4.1 学習データと評価データの設計

ロボットの応答文として使用している 2,478 件の文を評価データとして用いる．この 2,478 件の各応答文に 7 名の被験者が時節文かどうかのアノテーションを行った．7 名の被験者の内，過半数の被験者が時節文と判定した文を時節文と定義し，提案手法の評価を行った．また，時節文のアノテーションを正確に行う信頼性が高い特定の被験者 3 名を用意し，同様に 2478 件の評価データに時節文のアノテーションの作業を行った．この 3 名がアノテーションを行った 2,478 件のデータを学習データとして用いる．その際，信頼性の高い 3 名の被験者の中で，誰も時節文と判定しなかった文を重要度 0，1 名が時節文と判定した文を重要度 1，2 名が時節文と判定した文を重要度 2，3 名全員が時節文と判定した文を重要度 3 と定義する．それによって，三つの学習セット A，B，C を設計した．学習セット A は重要度 1 以上の時節文を正例として使用した場合，学習セット B は重要度 2 以上の時節文を正例として使用した場合，学習セット C は重要度 3 の時節文を正例として使用した場合の学習データとする．また，三つのデータセット全てに共通して，負例には重要度 0 の文を用いた．それぞれの学習セットのデータ件数を表 4 に示す．

表 4: 各学習セットのデータ件数の内訳

	学習セット A	学習セット B	学習セット C
正例の文数	176	74	24
負例の文数	2,302	2,302	2,302
時節語	106	62	28

4.2 時節文判定の結果

評価指標として，再現率（Recall）と適合率（Precision）の二つの指標を用いた．二つの提案手法における各学習データを用いた場合の時節文判定結果を表 5，表 6 に示す．そして，各分類の再現率と適合率をまとめたグラフを図 7 をに示す．

表 5: 単語マッチングによる分類結果

		予測結果					
		学習セット A		学習セット B		学習セット C	
		pos	neg	pos	neg	pos	neg
真の結果	true	156	2	128	30	103	55
	false	610	1,710	279	2,041	165	2,155

表 6: MLP による分類結果

		予測結果					
		学習セット A		学習セット B		学習セット C	
		pos	neg	pos	neg	pos	neg
真の結果	true	143	15	115	43	72	86
	false	133	2,187	69	2,251	46	2,274

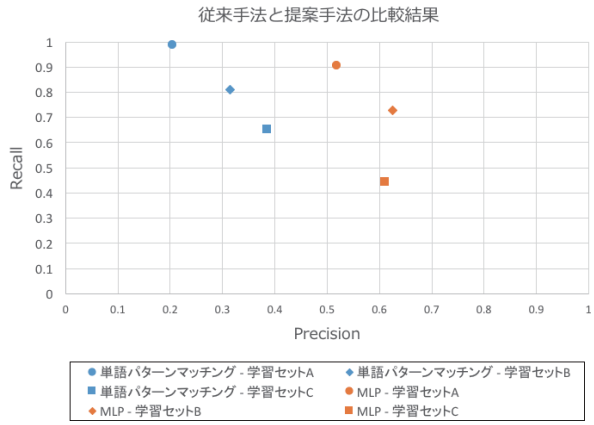


図 7: 各分類結果の比較

表 5 より，単語マッチングでは学習セット A を用いた場合の再現率の値が最も高い事がわかる．これは学習セット A に含まれる時節語の種類数が最も多い事が起因している．そのため，学習セット A を用いた場合の適合率は最も低い．表 6 より，MLP でも学習セット A を用いた場合の再現率が最も高い．適合率に比べ再現率の値の方が高い理由として，時節文と判定する閾値が低い事が考えられる．本論文では再現率を目標値として使用しているため，閾値を低くした結果として適合率が下がってしまった．図 7 より，単語マッチングでは全ての学習セットにおいて適合率の値が MLP より大きく下回っている．また，MLP において学習セット A を用いた場合の再現率は 90%を超えている．これらのことから，MLP は単語マッチングに比べ再現率を大きく損なわずに，より高い適合率で時節文判定を行うことができる．

また，MLP の分類では全ての学習セットの場合において再現率の値が単語マッチングより下回っている．その原因として，学習データの不足が考えられる．表 4 より，特に重要度 3 の時節文は 28 件と圧倒的にデータ数が少ないことがわかる．そのため，特に学習セット C では十分な学習を行えていないと考えられる．それぞれ学習セット A を用いた場合の単語マッチングによる分類と MLP による分類において目標値を達成した．また，適合率においては学習セット A を用いた単語マッチングより，学習セット A を用いた MLP による分類の方が高い精度を出している．よって，現状では学習セット A を用いた MLP による分類が最も有効である．

5. おわりに

本論文で使用した学習データが少ないことが課題として挙げられた．しかし，普段の会話で使用する発話のほとんどが時節文でないため，時節文の件数を大量に取得するにはさらに多くのロボットの応答文を生成する必要がある．また，本論文の提案手法では入力となるロボットの応答文が時節文かどうかの 2 値分類を行った．しかし，時節文の重要度毎に分けて学習する事で 2 値分類ではなく他クラス分類を行う事ができる．それによって，時節文判定の中間値を表現する事ができると考えられる．今後は対話データの収集と多クラス分類による時節文判定システムを検討する．

参考文献

[1] E. Sato-Shimokawara, S. Nomura, Y. Shinoda, H. Lee, T. Takatani, K. Wada, and T. Yamaguchi. A cloud based chat robot using dialogue histories for elderly people. Vol. 2015-November, pp. 206–210, 2015.

[2] 光田航, 東中竜一郎, 松尾義博. 複数の作業者グループを用いた対話における言外の情報の類型化. Technical Report 3, dec 2016.

[3] 吉川義紀, 芋野美紗子, 土屋誠司, 渡部広一. E-015 文の意味を考慮した常識的時間判断システムの構築 (e 分野:自然言語・音声・音楽, 一般論文). 情報科学技術フォーラム講演論文集, Vol. 13, No. 2, pp. 235–236, aug 2014.

[4] James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. The timebank corpus. In *Corpus linguistics*, Vol. 2003, p. 40. Lancaster, UK., 2003.

[5] 小畑陽一, 渡部広一, 河岡司. 単文の名詞と動詞から時間／季節を判断するメカニズム. Technical Report 1(2000-ICS-123), jan 2001.

[6] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *In Proc. of EMNLP*, pp. 230–237, 2004.

[7] Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

多様なロボットサービスを実現するための RT システムの構造化

Structure of RT system to realize various robot services

松日楽信人[*1], 池田貴政[*1], 下山未来[*1], 岡野憲[*1], 成田雅彦[*2], 土屋陽介[*2], 加藤由花[*3]
Nobuto Matsuhira, Takamasa Ikeda, Mirai Shimoyama, Satoshi Okano, Masahiko Narita, Yosuke Tsuchiya, Yuka Kato

[*1] 芝浦工業大学, [*2] 産業技術大学院大学, [*3] 東京女子大学
Shibaura Institute of Technoloy, Advanced Institute of Industrial Technology, Tokyo Woman's Christian University

We developed the robot service system using common robotic software. Proposed common RT unit has three parts; sensor RTC, service RTC and robot control RTC. We verified the common unit system to apply the reception robot that responds the human walking property, knows the people behavior in the facility, and finally applies for a cruising guidance robot. It is possible that this system will be useful to develop new applications easily.

1. はじめに

経済産業省によると 2035 年にロボット産業市場は約 9.7 兆円に拡大し，その中でも特にサービス分野が大きく拡大していくと予想される[1]. それに伴って近年，人々に対してコミュニケーションや情報提供を行い，サポートするロボットが開発されている[2]. これらロボットには案内，受付，在宅での見守り，巡回警備，アミューズメントなど多岐に渡るサービスが期待される. しかし，それぞれのサービスによって必要となるシステムが変わってくるため，1 つのロボットで複数のサービスを行うことは難しい. そこで，本研究では複数応用可能なサービスロボットの共通化システムを構築し，そのシステムの有効性を確認する.

2. 共通化システムの提案

2.1 ミドルウェア RTM

RT ミドルウェア（RTM）は産業技術総合研究所が開発し，ロボット機能要素のソフトウエアモジュールを複数組み合わせてシステムを構築するためのソフトウェアプラットフォームである[3]. RTM はハードウェアの種類や機能毎にロボットシステム（RT システム）を構成するプログラムを RTC としてモジュール化し，各 RTC 間でデータを通信することによって機能を実現できる. 本研究では，現在最も使用されている OpenRTM-aist を用いる.

2.2 通信プロトコル RSNP

Robot Service Network Protocol（RSNP）はロボットサービスイニシアチブ（RSi）が開発したロボットを共通の通信プロトコルで動かすことができる仕組みである[4]. 現在開発されている RSNP は XML を利用したメッセージをサーバ，クライアント間の双方向でやり取りを行うことができる仕様となっている. また，利用するサービスごとに Profile が規定されており，利用したいメソッドの実装を行う必要がある.

2.3 共通化システム

ロボットサービスを効率的に開発するために提案する共通化システムの概念を図 1 の 3 種類のユニットとして示す. このユニットは，3 つの RTC から構成されている. これらの中から「センサ情報取得 RTC」と「サービス RTC」をそれぞれ入れ替えることで，様々なセンサ情報が取得可能となり，サービスの用途にあわせて「属性ごとへの応答」，「混雑度対応」，「対面応答」や位置情報を利用した「案内」など 1 つのロボットで複数のサービスの提供を行うことが可能となる.

さらに各ユニットを図に示すように RSNP で複数台繋げることで多くのデータを取得し，サーバで取得データを管理できる. このようなロボットネットワークを構築することで，ロボットの管理者はロボットの状況を把握することができ，ロボット利用者にとっては各エリアや全体の状況把握が容易になると考える.

図1において，これまでの研究内容を提案するユニットと対応付ける. 歩行情報に基づく属性分けによるインタフェースロボットの応答の研究では，センサ情報取得 RTC の「測域センサ」，サービス RTC の「受付属性応答」，ロボット制御 RTC で構成している. また，歩行者の混雑度予測による応答では，サービス RTC を「混雑度応答」に切り替えた. さらに，位置情報を用いた案内サービスでは，センサ情報取得 RTC で GPS データをソケット通信で受け取るために「ソケット通信対応」に切り替え，サービス RTC を「案内位置情報」に切り替えてシステムを構成した.

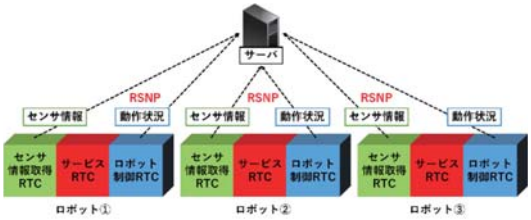


図 1 共通化ユニットによるロボットネットワーク

3. 歩行情報の属性分けによる受付応答

図 2 に RTM を用いて開発した属性ごとにおける応答制御システムの構成図を示す[5]. 図中の URG RTC は測域センサから値を取得するものであり，Human_Step RTC は足の検出，追跡及び歩行情報の取得を行い，歩幅，歩隔,歩幅・歩隔比速さの 4 つの歩行情報を出力する. Step_Classfication RTC は 4 つの歩行情報を受け取り，それらの情報をもとに属性（子供・大人・高齢者）分けを行う. ApriPocoMotion RTC では取得した属性分けの結果をもとに属性ごとへの応答の切り替えを可能とする. また，ApriPocoMotion RTC はインタフェースロボットへの指令をするコンポーネントであるので本応答制御システム以外にも使用する. ここで,属性分けを行うための歩行情報の範囲は文献[6]を参考にした.属性の判定には2つ以上の条件を満たすようにしている.例えば,歩幅と歩隔の値が「大人」の範囲に含まれていた場

連絡先：松日楽信人，芝浦工業大学，東京都江東区豊洲 3-7-1 ， TEL 03-5859-8054 ， Fax 03-5859-8001 ， matsuhir@shibaura-it.ac.jp

合,属性を「大人」と判定する.また,満たす条件が 1 つの場合は「NO Data」と表示し,判定なしとする.プログラムは 4 つの歩行情報のうち先に 2 つ以上の条件を満たした際に属性判定を行う.この条件で自分自身(成人)が普段通りの歩行を行い,属性判定をした結果,84%の精度であった.また機械学習を取り込むことで,複数人のテストデータで 80%以上の再現率を確認している[7].

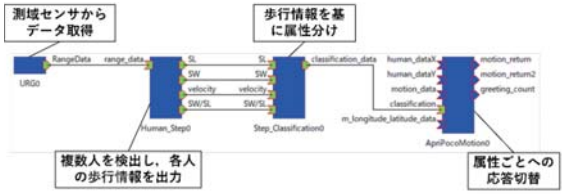


図 2 RTM を用いた属性ごとにおける応答システム

4. 複数の人検出デバイスによる応答システム

図 3 に RTM を用いた複数の人検出デバイスによる応答制御システム構成図を示す[8]. 本システムは図 4 に示す 4 種類の RTC から構成されている.

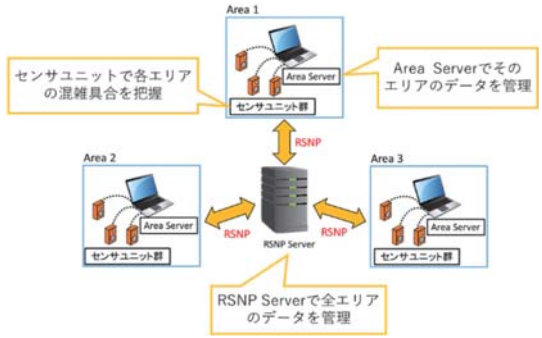


図 3 複数エリアでの人検出デバイスによる応答制御システム

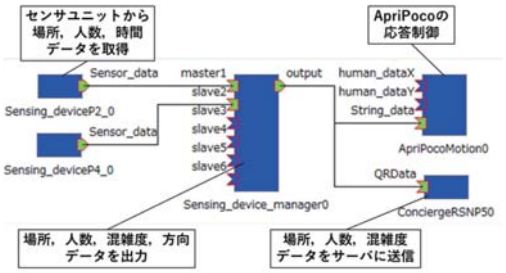


図 4 複数エリアでの人検出デバイスによる RT 応答システム



図 5 混雑度データ収集表示例

図 3 で Sensing_devicePe1, 2 がそれぞれ master1, slave2 に RTM 上で接続されたときに,2 個のデバイス間の人の通過時間差を計測することができる.2 つの Sensing_devicePe1, 2 からのデータの受信待機状態から,Sensing_devicePe2 で人を検出,データ送信し,Sensing_device_manager で新しいデータを受信後に ApriPocoMotion RTC と ConciergeRSNP5 RTC へ場所,人数,混雑度,方向データを送信する.ConciergeRSNP5 RTC は受け取ったデータを RSNP でサーバに送信する.資料館での実験結果のデータ表示例を図 5 に示す.

5. 位置情報を用いた案内サービス

本研究で使用してきた ApriPoco™(以下,ApriPoco)は固定型ロボットであるため受付への応用をメインにしてきた.しかし,クルーズ船等の移動体に乗せることで,移動サービスも可能となる.そこで,GPS 位置情報を用いた案内を提案した[9].この案内と前述の受付のシステムは共通の仕組みで開発されており,RTC を入れ替えるだけで容易にシステムの切り替えが可能である.

全体のシステム構成図を図 6 に示す.緯度・経度情報を取得するためのシステムは,Android Studio で構成し,Android タブレット端末の GPS を利用する.ここでは,Android タブレットから緯度・経度情報を受信し,それを別の RTC に送信する UteranceServer RTC と設定した緯度・経度情報を受信後,ApriPoco に発話・動作指令を行う ApriPocoMotion RTC を組み合わせた.21 の観光ポイントにおいて GPS 情報で説明が可能であった.

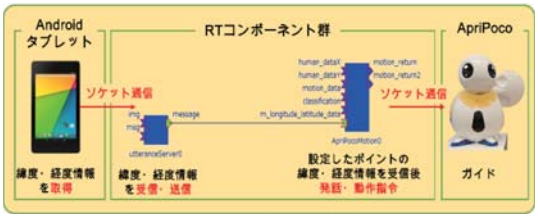


図 6 クルージングガイドシステム

6. おわりに

本研究では,ロボットサービスの実現を容易とするために RT システムの構造化を考え,共通ユニットを提案した.これに基づき,ロボットの受付対応において属性や複数センサによる混雑度による応答制御と位置情報を用いた案内サービスを可能とするシステムに適用した.今後は,より多くのロボットサービスの実現に展開して行く.

参考文献

[1] 経済産業省, 2012 年ロボット産業市場動向調査, 2012 年ロボット産業の市場動向調査結果
[2] 山本, 土井, ユーザーと家電機器をつなぐインタフェースロボット, 東芝レビュー, Vol. 64, No. 1, pp. 28- 31, 2009
[3] Open RTM-aist, http://openrtm.org/
[4] http://robotservices.org/index.php/aboutrsnp/
[5] 池田, 唐真, 松日楽, 加藤, インタフェースロボットの応答制御のための歩行情報取得方法の検討, SI2016, 3E3-3 , 2016
[6] 赤平, 歩行の加齢現象とめまい平衡障害例の歩行分析, 日本耳鼻咽喉科学会会報, Vol. 102, No.2, pp.277-285, 1999
[7] 坂井, 木村, 池田, 野見山, 松日楽, 加藤, 測域センサにより取得される歩行パターンを利用した高齢者/若年者分別手法, 情報処理学会論文誌, 58 巻 2 号, pp.375-383, 2017
[8] 瀬沼, 岡野, 池田, 松日楽, 土屋, 成田, Raspberry Pi を用いた RSNP 人検出データ収集システム, RSJ2017, 2B1-06, 2017
[9] 池田, 唐真, 松日楽, 志村, インタフェースロボットのクルージングガイドへの応用, ROBOMECH2017, 1A1-I06, 2017

6:20 PM - 7:00 PM (Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room E)

[2E4-NFC-1b-a] (Panel discussion) The next direction of "Integrated Platform for Cloud-based Robot Services Project"

Masahiko Narita, Tokuro Matsuo, Nobuto Matsuhira, Tatsuno Kyoichi, Toru Yamaguchi, Yuka Kato

©The Japanese Society for Artificial Intelligence

[2F4] [General Session] 4. Knowledge Utilization and Sharing

座長:森田 武史（慶應義塾大学）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room F (4F Garreria)

[2F4-01] Ontology-based Knowledge Systemization for Interpreting the Mechanism of Liver Toxicity

○Yuki Yamagata[1], Yoshinobu Igarashi[1], Noriyuki Nakatsu[1], Katsuhisa Horimoto[2], Kazuhiko Fukui[2], Yoshihiro Uesawa[3], Hiroshi Yamada[1] （1. National Institutes of Biomedical Innovation, Health and Nutrition, 2. National Institute of Advanced Industrial Science and Technology, 3. Meiji Pharmaceutical University）

5:20 PM - 5:40 PM

[2F4-02] Reorganizing Hierarchical Category Structure of Japanese Wikipedia for Knowledge Engineers

○Takanori Nakagawa[1], Masaharu Yoshioka[1,2] （1. Hokkaido Unviersity, 2. RIKEN AIP）

5:40 PM - 6:00 PM

[2F4-03] Towards Implementation of Integrated Knowledge Infrastructure to Create Knowledge Graphs

○Ryutaro Ichise[1,2], Natthawut Kertkeidkachorn[2], Lihua Zhao[2] （1. National Institute of Informatics, 2. National Institute of Advanced Industrial Science and Technology）

6:00 PM - 6:20 PM

[2F4-04] Training Data on Japanese Archery Motion Recognition for Self-training Support

○Takahiro Morozumi[1], Tadachika Morozumi[1], Toramatsu Shintani[1] （1. Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology）

6:20 PM - 6:40 PM

[2F4-05] A Model for Knowledge Representation Based on the Concepts of Things and Events

○Takahiro Yamada[1] （1. Japan Aerospace Exploration Agency）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

安全性評価のための肝毒性知識の体系化
－オントロジー工学に基づく機序解釈支援に向けて－
Ontology-based Knowledge Systemization for Interpreting the Mechanism of Liver Toxicity

山縣 友紀[*1] Yuki Yamagata　五十嵐 芳暢[*1] Yoshinobu Igarashi　中津 則之[*1] Noriyuki Nakatsu　堀本 勝久[*2] Katsuhisa Horimoto　福井 一彦[*2] Kazuhiko Fukui

植沢 芳広[*3] Yoshihiro Uesawa　山田 弘[*1] Hiroshi Yamada

[*1] 医薬基盤・健康・栄養研究所 Toxicogenomics Informatics Project, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN)

[*2] 産業技術総合研究所 Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST)

[*3] 明治薬科大学 Department of Medical Molecular Informatics, Meiji Pharmaceutical University

To understand the mechanisms of drug-induced liver injury, we have developed an ontological model for the liver toxicity. In this paper, we focus on safety assessment and propose an application system based on the ontology and the semantic web technology.

1. はじめに

医薬品開発では開発中止および市場撤退の主な原因として薬剤性肝障害があり，開発早期段階での安全性評価技術が要請される．毒性学とは化学物質などが引き起こす生体影響（毒作用）を科学的に明らかにする学問であり[日本トキシコロジー学会教育委員会 09]，近年 *in silico* 解析による毒性予測が期待され，機械学習を用いたバイオマーカーの候補の探索など，データ駆動型研究によるアプローチが進められている．これをさらにすすめ適切な安全性評価を行うには，機械学習等で得られた結果の根拠として機序との関係性を明らかにする事が必要となる．そのためには毒性の発現機序（メカニズム）そのものを把握する必要があるが，肝毒性の発現機序は複雑であり，様々な因子が関与することから全体像の把握が困難となっている．毒作用へ対処し適切な安全性管理を行うためには，機序の本質を把握し，一貫性のある視点に基づき必要な知識を体系化していく必要がある．

このような背景から，本研究では肝毒性に関わる多種多様な知識を組織化し，肝毒性の知識基盤の構築を目指してオントロジー構築に取り組むとともに，機序を適切に表現する記述枠組みのモデル化を行っている．本稿では，これまでの取り組みについて報告する．さらに，応用研究としてデータ駆動型研究との連携による安全性評価と管理への適用について考察し，機序解釈支援のための知識システムを提案する．

2. 肝毒性知識の体系化

2.1 オントロジー構築の概要

薬剤は生体にとっては異物（Xenobiotics）であることから，体内で有害反応を引き起こすというリスクが存在する．一方，生体においては，細胞小器官，細胞，器官など様々な構造物や分子がホメオスタシス維持のための役割を担っている．肝毒性機序を把握するためには，これらの多様な生体構成要素の薬剤に対する防御反応からホメオスタシス破綻における毒性発現プロセスまで一貫性をもって概念化できる手段が必要となる．そこで，注目した技術がオントロジーである．オントロジーとは対象世界に存在する概念を切り出し，それらの関係を記述するための基盤概念の理論であり，知識背景を明示し，対象の捉え方を規定することで知識の体系化のための共通基盤構築に貢献する．図1に現在構築を進めている肝毒性オントロジーの全体像を示す．最上位は持続物(Continuant)（図 1 左）と生起物(Occurrent)（図 1 右）に大別でき，持続物は 3 次元空間の存在物，いわゆる「モノ」を指し，生起物は時空間上の存在物である「コト」を示す．これら分野独立の一般概念は，哲学的思考に基づいて構築された上位オントロジーBFO [Arp 15] と YAMATO [Mizoguchi10] を参照することで客観性を確保している．オントロジーでは一般・特殊関係（is-a 関係）に基づいて下位概念を定義していくことで共通性や相違点を明らかにしながら知識を積み上げて体系化を図る．持続物の下位概念には，分子や化合物，細胞小器官や細胞，器官などの生体構造物，ヒトやラットの生物種等がある．これらは生命科学領域では汎用的な概念であり，バイオコミュニティでの知識共有と標準化を目指して開発されている既存のデータベースおよびオントロジー[1,2,3,4]の定義の再利用を基本とした．生起物の下位概念にはプロセスが定義されている．本研究ではプロセスの下位を Primitive プロセスと複数のプロセスから構成されるプロセス系列に分け，プロセス系列の下位に毒性機序を定義することとした．さらに，持続物/生起物と別の分類軸として何かに従属することでしか存在できない従属的実在物があり，コンテキストに依存して変化するロールや，属性，特性などの性質に関する概念を定義している．ロールの下位概念には毒性機序の発現の過程で分子の果たす様々な役割を中心に概念化を行っている．オントロジーの構築は文献や教科書をもとにマニュアルアノテーションにより情報を抽出し，Protégé[5] を用いて記述を行っている．2018 年 2 月 1 日時点でオントロジーには約 6000 の概念が定義されている．

連絡先：山縣友紀，医薬基盤・健康・栄養研究所，〒567-0047 大阪府茨木市彩都あさぎ 7-6-8，Tel: 072-641-9826，Fax:06-6879-2123，E-mail:y-yamagata@nibiohn.go.jp

[1] UBERON, http://uberon.org
[2] ChEBI, http://www.ebi.ac.uk/chebi
[3] Gene Ontology, http://www.geneontology.org
[4] NCBI Taxonomy DB, http://www.ncbi.nlm.nih.gov/taxonomy
[5] Protégé, http://protege.stanford.edu/

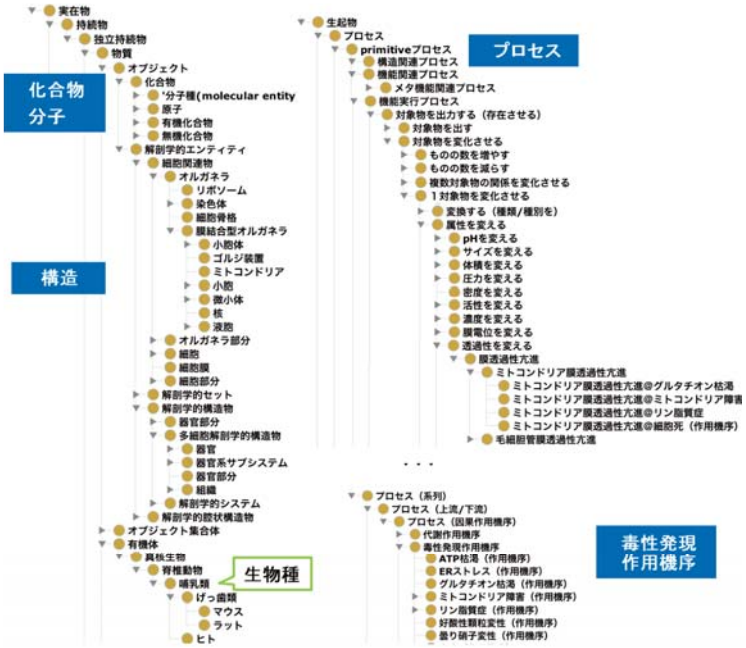


図1 オントロジーの全体像

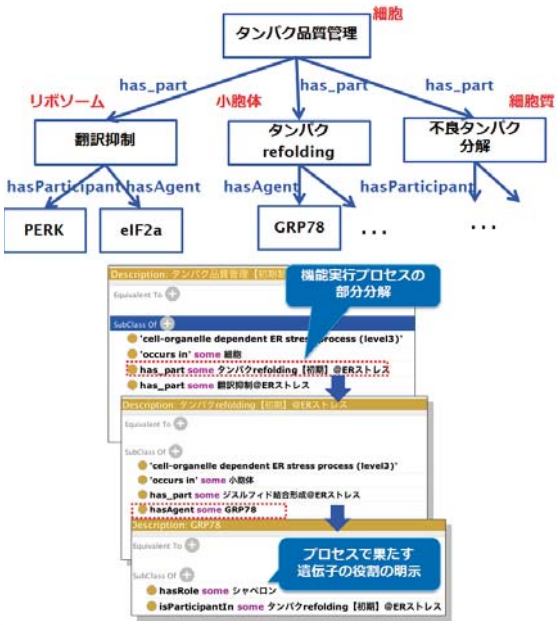


図 2 細胞機能実行プロセスと部分分解例

2.2 作用機序と構成プロセスの概念定義の詳細

本節では毒性発現機序とそれを構成するプロセスの概念定義についての詳細を説明する．肝毒性発現機序は一般的な過程では，薬剤が標的部位へ配送されターゲットに作用し，生体側では生体防御反応や適応を試みるが，細胞障害が起きることから毒性が発現するとされている[日本トキシコロジー学会教育委員会 09]．したがって，機序は複数のプロセスが組み合わさった一連の流れとして捉えることができることから，本研究では毒性発現機序について「毒性発現に至るまでに引き起こされる生体プロセスの間の因果関係の組み合わせ（因果連鎖）」として定義する．オントロジー階層では，プロセス系列の下位概念として毒性発現機序の概念化を行い，さらに，その下位に小胞体ストレス（ER ストレス），グルタチオン枯渇，ミトコンドリア障害，リン脂質症等，機序テーマの詳細を定義することとした（図 1 右下）．

次に毒性機序を構成するプロセスについて説明する．機序の把握には，生体の構造と機能を対応付けたうえでプロセスを客観的かつ一貫性をもって概念化する必要がある．そこで，本研究では機能概念オントロジー[來村 02]を参考に生体の機能発揮プロセスについて定義することとした．機能概念オントロジーでは，機能語彙体系の定義および機能的知識のモデル化を進めており，機能の発揮主体を「装置」(device)と呼ぶ.装置は全体-部分関係を持ち，ある装置はより小さな装置の組み合わせから構成され（システムと呼ぶ），複数の部分機能発揮によってシステムの機能が達成される．そして，装置と装置に入出力された対象物の状態変化（対象物のもつ属性の値の変化）から装置の機能を捉える．本研究では，細胞をシステムとして捉え，複数の装置（部品）として細胞小器官から構成されるという全体-部分関係（has_part 関係）をもとに，粒度毎の生体構造物の機能発揮プロセスを明らかにすることを基本方針とした．図 2 は，方針に従い細胞システムがどのように機能を達成するかについて記述した例である．ER ストレスでは薬剤が小胞体に蓄積し，その結果小胞体内でタンパクが折りたたまれずに不良タンパクが生成されることが知られている．それに対し，細胞システムでは「タンパク品質管理」機能を実行する．ここで，細胞の部品である細胞小器官に注目すると，小胞体では「不良タンパク refolding（再折りたたみ）」，細胞質では「不良タンパク分解」という2つの部分機能発揮プロセスが起こり，これらの部分プロセスによって細胞システムとして品質管理が達成されることがわかる．また，不良タンパクに応答する間，リボソームでは新たなタンパクの生成を抑制する「翻訳抑制」の部分プロセスも行われているが，これは機能オントロジーでは品質管理のためのメタ機能「補助貢献」に対応する．

2.3 プロセスに参加する分子とプロセスで果たすロール

分子は生体内では一般に生体内で複数の役割を担っていることが知られている．そこで，本研究では毒性機序を構成するプロセスに参加する分子を同定するとともに，その分子がプロセス内で果たす役割をロール概念として概念化することとした．例えば，タンパク refolding プロセスでは GRP78 が Agent として参加し，タンパクの再折りたたみを補助する「シャペロン」ロールを担う（図 2 下部）．

機能分解については，分子も装置とみなして分子の粒度まで部分分解するという手法も考えられる．しかし分子レベルではあまりに多くのプロセスが機序に関与することから機序の全体像の把握をかえって困難にする恐れがある．また，細胞構造から考えても細胞小器官以上を基本部品とするのが本質であると考えられる．そこで，部分分解については必要とされる場合のみ分子レベルのプロセスを詳細化することとした．

3. 作用機序の統一的表現記述枠組み

生体機序の理解を支援するには，知識の体系化に加えどのように知識を表現するかということも重要である．進行の過程の全体を適切に表現し，視覚化できれば，機序の理解を促進することが期待される．これまでに生命科学分野では既に KEGG[6] や Wikipathways[7], Reactome[8]等パスウェイデータベースが公開されているが，分子間相互作用の記述が中心となっており，細胞レベルの記述は少ない．また，いずれも毒性の観点で構築さ

[6] KEGG http://www.genome.jp/kegg
[7] Wikipathways https://wikipathways.org
[8] Reactome https://reactome.org

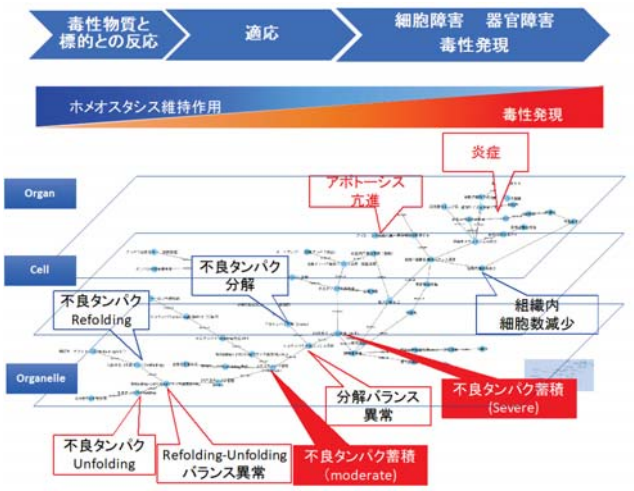


図 3　作用機序マップ

れたものではないため肝毒性機序の過程を把握するには不十分である．そこで，本研究では作用機序を適切に表現するための記述枠組みとして「肝毒性作用機序マップ」を開発することとした．本マップはオントロジーで構築した概念をもとに，1つの肝毒性作用機序テーマに対し，毒性発現に至る過程において生体で起こりうるプロセスをノードとし，因果関係をエッジで表現することで，機序の全体像をグラフ形式としてモデル化する．さらに毒性発現の過程について，本研究ではバランス異常の考え方を導入し[Mizoguchi 11]，毒性の進行過程において，1) 毒作用としての機能要求プロセスと，2) ホメオスタシス維持のための機能発揮プロセス，そして 3) それらのバランス異常から 4) アウトカムとして生体構造物の機能障害が起こる各々のプロセスを可視化して表現することを基本方針とした．これまでグラフ表示ツール Cytoscape[9]を用いて 6 種の機序テーマのマップ記述を進めており．その一例として，ER ストレスのマップ表現を図 3 に示す．ER ストレスでは，不良タンパク Unfolding とタンパク Refolding のバランス異常の結果，不良タンパクが蓄積することがわかる．さらに，不良タンパクを分解しようと機能発揮するが，結局重度の不良タンパクが蓄積する．そして，その対処として，不良タンパクの蓄積した細胞を除去するために細胞レベルではアポトーシスが起こるが，アポトーシスも亢進しすぎると組織内の細胞数を減少させる毒作用として捉えられる．また，生存のための防御反応といえる炎症が，器官レベルでは肝の器質的変化として肝線維化を引き起こす可能性があることがマップによって把握可能である(図 3)．

このように，機序マップでは細胞や細胞小器官自体の適応のための機能発揮プロセスも，一段粒度の大きいレベルからみると毒作用となるということが理解できる．従来のパスウェイデータベースでは分子間相互作用の可視化が主であったが，本研究ではオントロジーで定義された概念に基づいて統一された記述枠組みで細胞や細胞小器官レベルのプロセスを表現することで，毒性発現までの過程が粒度毎に把握しやすく，機序の理解のサポートに大きく貢献する．

4.　応用研究

オントロジーを用いれば知識を適切に計算機処理できることから，基盤技術として様々な応用が期待できる．本節ではデータ駆動型研究の結果に対するオントロジーの適用例をもとに，安全性評価および安全性管理に貢献できるか考察し，今後の機序解釈支援型知識システムの実現に向けた検討を行う．

4.1 データ駆動型研究との連携による安全性評価

安全性評価において，適切なバイオマーカーの設定は重要課題となっている．特に，生体内 (*in vivo*)と細胞/組織培養系 (*in vitro*)の間での毒性発現の違いや生物種の違いが問題となっており，これらの差異を反映できるマーカーや，さらには生物種や実験系に依存しないマーカーを開発すれば創薬に大きく貢献すると考えられる．これまで我々はラットを用いた *in vivo* 実験モデル系とヒト初代培養細胞系 (*in vitro*)の各実験系を用いて化合物投与実験を行い，マイクロアレイ解析による網羅的遺伝子発現プロファイルに基づき，機械学習により肝毒性マーカーとなりうる複数の候補遺伝子を選定してきた．今回，機械学習で得られたマーカー候補遺伝子について文献をもとに機序への関連性をアノテートし，意味内容をオントロジーで定義付けた後，マップへの表現記述を試み，本研究が安全性評価に貢献しうる記述能力を有するか，検証を行った．

In vitro マーカー候補遺伝子の検証

図 4 の赤い丸は ER ストレスについて，*in vitro* における機械学習で得られたマーカー候補遺伝子がどのようなプロセスに関わるか，マップ上で表示した結果を示したものである．培養細胞は，通常生体の変化の一部しか反映されないとも考えられているが，候補遺伝子は ER ストレスの毒性発現機序において，タンパク品質管理から下流の重度の肝障害のプロセスまで関与していることが明示され，それらの中には複数のプロセスに関わる遺伝子もあることが確認された．

In vivo と *in vitro* のマップ比較におけるマーカー候補遺伝子の識別性と共通性に関する検証

次に，ラット生体 (*in vivo*) への化合物投与実験についても，同様に機械学習で予測された遺伝子についてマップに反映し，*in vivo* と *in vitro* の間で比較を行った．その結果，両実験条件下で共通して予測される遺伝子があることがわかり，例えば，タンパク refolding プロセスに関与するシャペロンロールを果たす遺伝子群に共通性が見られた(図 4 左下)．また，*in vitro* 特異的に予測されたシャペロン遺伝子のうち，ある遺伝子はアポトーシス抑制因子としても働き，下流では腫瘍生成に関与するということも把握できる(図 4 中央から右上の点線枠)．このように，機械学習で予測された遺伝子について，マップを利用することで共通性や差異を可視化できることから，実験条件等を考慮した適切な評価やさらには両者のブリッジングへ貢献すると期待される．また，遺伝子についてオントロジーでは関与するプロセスやその役割，さらに進行過程を把握できることから，機械学習で得られた情報について生物学的根拠を示すことが可能となる．

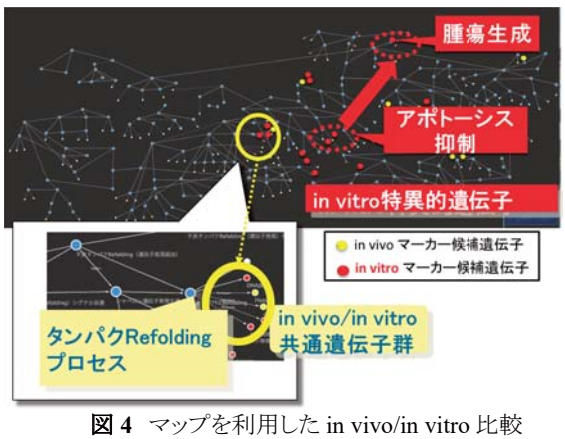


図 4　マップを利用した in vivo/in vitro 比較

[9] Cytoscape http://www.cytoscape.org

我々は別のデータ駆動型研究として，数理解析によるパスウェイの順位相関等からの毒性予測も行っている．解析では化合物投与後の生理学的計測条件下，数理解析によって有意に出現するパスウェイを抽出し，順位相関解析を行っている．しかし，薬剤には本来の目的としての薬効作用もあるため，解析結果には毒性に関わるパスウェイ (Off-target) の他，薬効の延長上 (On-target) のパスウェイも含まれる場合がある．そこで，オントロジーに蓄積された毒性機序プロセスに関与するパスウェイとこれら数理解析で予測されたパスウェイとを比較することで，毒性機序に特異的なパスウェイと薬効上のパスウェイを識別できれば，安全性評価の精度の向上に貢献すると考えられ，今後検討していきたい．

ここで，オントロジーとデータ駆動型の連携は，オントロジーによる知識の体系化にも利点をもたらす．オントロジー内の各概念は教科書や文献からマニュアルでアノテーションを行っているため，質の高さは確保されるが，あくまでも人間のこれまで得た知識の集積である．そのため，未だ人間が認識していない情報をデータ駆動型研究が発見する可能性があると考えられる．データ駆動型研究で得られる特徴量などの情報について，一般化/特殊化，あるいは概念間の関係性をたどりながらその性質を検証し，新たな概念や属性として知識化することも安全性の評価という点において重要であると思われる．

4.2 機序解釈支援型知識システムの開発に向けて

オントロジーを基盤知識としてデータベース化することで，客観的かつ頑健な理論に支えられた知識の積み上げや知識管理が期待される．これに加え，実用面ではユーザーが肝毒性機序の把握するための関連情報を容易に利活用できることが望ましい．そこで，オントロジーデータベースとセマンティックウェブ技術を用いた Web アプリケーションによる機序解釈支援のための知識システムを目指す．前節で考察した安全性評価に向け，システムではオントロジーをもとにした機序に関する説明文生成機能や機序可視化機能を基本仕様として備える．さらに，Linked Data 技術 [ヒース 13] を用いることで外部リソースの副作用情報や文献情報などを利活用し，創薬に必要な様々な知識を統合していきたいと考えている．加えて，システムでは機序解釈支援のための知識探索・推論機能を持たせたい．推論には SPARQL 検索に加え，OWL 推論機構を用いる予定である．機序については，原因プロセスの複数機序テーマ横断探索や，機序の後期段階で起こりうる毒性発現プロセスの推論，二点間の作用機序経路探索等を予定している．上位−下位概念や兄弟概念，全体−部分関係などオントロジーで定義された概念とその関係性に基づく共通性や違いからの説明や，多様な視点から機序解釈を支援することで，ユーザーが知見を深めることができるようなシステムとしていきたい．

5. まとめと今後の展望

本研究では，肝毒性発現機序の本質を理解するために肝毒性知識の体系化を目指したオントロジーの構築および機序を表現するための記述枠組みの開発について述べた．さらに，応用として安全性評価への貢献について議論した．

臨床では病理所見や血液学的検査の測定値である AST，ALT 等の変動等観測データから肝障害を判断している．本研究では顕性化前の早期段階から機序をたどることができることから，安全性管理においても有用な知見が得られることが期待される．図 5 は ER ストレスの機序の進行過程を示したものであり，ストレス応答プロセスが進行していく中でホメオスタシス維持プロセスに関する遺伝子発現の変化をモニタリングし，毒性発現の指標としてアポトーシスへのスイッチング遺伝子の発現などをとらえていけば，ターニングポイントの把握が可能となる．さらに，発現プロセスの段階に応じた重篤化を防ぐための治療計画，あるいは新たな分子標的薬開発など毒作用への対処にも適用できると考えている．今後知識システムの実装をすすめ，創薬支援に役立てていきたい．

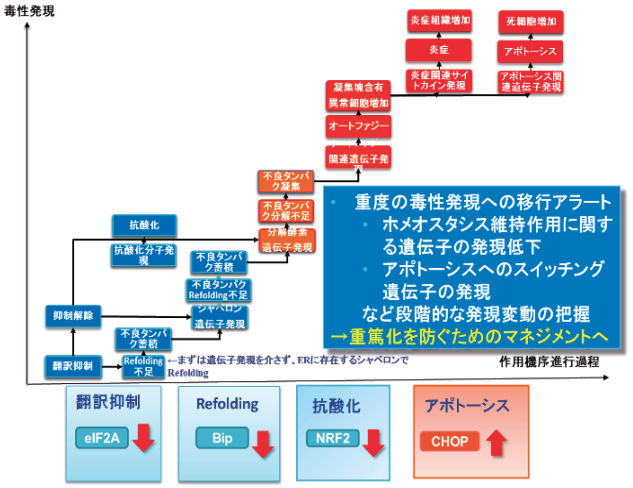


図 5 マップを利用した進行過程把握による安全性管理

本研究は創薬分野の安全性評価の問題として扱っているが，その内容を汎化技術として拡張すれば環境等他の分野における安全性評価・管理にも有用であると考えている．本稿では，データに囲まれた社会において対象とする問題にどのように向き合い，内容を同定していくかという基礎的な考察から応用への実現に向けて議論した．データ駆動型研究の結果に対し，オントロジーが核となり俯瞰的な視野でデータの意味解釈を処理することで，データから高次知識へ変換し知識空間における位置づけを知ることができれば，新たな技術開発につながり AI 研究の進展に貢献すると考えている．

謝辞

本研究は，本研究は国立研究開発法人日本医療研究開発機構(AMED)の創薬支援推進事業（JP17nk0101103）の支援によって行われた．

参考文献

[Arp 15] Arp R. et al., Building Ontologies with Basic Formal Ontology, MIT Press, (2015).

[日本トキシコロジー学会教育委員会 09] 日本トキシコロジー学会 教育委員会編集, トキシコロジー, 朝倉書店, (2009)

[來村 02] 來村他 ：オントロジー工学に基づく機能的知識体系化の枠組み．人工知能学会誌．Vol 17, No.1. pp.61-72 (2002).

[Mizoguchi 10] Mizoguchi R: YAMATO: Yet Another MoreAdvanced Top-level Ontology, In Pro. of the Sixth Australasian Ontology Workshop, pp.1-16, (2010)

[Mizoguchi 11] Mizoguchi R. et al.: River Flow Model of Diseases, In Proc. of 2nd International Conference on Biomedical Ontology (ICBO2011), pp.63-70 July 28-30, 2011, Buffalo, NY, USA, (2011).

[ヒース 13] トム ヒース (著), クリスチャン バイツァー (著), 武田英明 (監訳)：Linked Data: Web をグローバルなデータ空間にする仕組み, 近代科学社 (2013)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

知識工学者のための日本語Wikipediaのカテゴリ階層構造の再整理

Reorganizing Hierarchical Category Structure of Japanese Wikipedia for Knowledge Engineers

中川嵩教 *1　　吉岡真治 *1*2
Takanori Nakagawa　　Masaharu Yoshioka

*1北海道大学　　*2理研 AIP
Hokkaido University　　RIKEN AIP

Since Wikipedia is an on-line encyclopedia that covers varieties topics, there are several attempts to extracting knowledge from its contents. Wikipedia category is one type of information that provides topical index for the articles and organized as hierarchical manner. However, since hierarchical structure of Wikipedia category have different characteristics compared to the one for knowledge representation, it may not fully and correctly utilized at this moment. In this paper, we propose a method to extract hierarchical structure that have similar characteristics with knowledge representation.

1. はじめに

Wikipedia*1 は世界最大のインターネット百科事典であり、様々な形で知識源として活用する方法が検討されている。具体的には、各記事に関する属性情報（所在地、生年月日、所属…）などを記述したインフォボックスから属性情報を抽出する DBpedia[Bizer 09] の研究や、カテゴリの情報を用いて、様々な単語間の類似性を計算する方法 [Ponzetto 07, Taieb 13] や、オントロジーの構築に役立てる YAGO2[Hoffart 13] や、日本語 Wikipedia オントロジー [玉川 10] の研究などが存在する。また、我々は、これまでに、このカテゴリの性質に関する分析 [藤原 12, Yoshioka 14] を行っており、これらの研究の中で、Wikipedia のカテゴリの持つ特殊性を指摘し、その特殊性を考慮した利用方法が必要であることを提案してきた。

本研究では、これまでの研究踏まえて、Wikipedia のカテゴリ階層を概念階層として扱う場合の問題点を、具体例を示しながら指摘するとともに、その問題点を解消するための、知識工学的観点からの Wikipedia カテゴリの再整理を行う方法を提案する。

2. Wikipedia カテゴリ

2.1 Wikipedia カテゴリの階層構造

Wikipedia において、カテゴリとは、膨大な記事群を様々な観点から分類するための索引であり、各記事には、それぞれに一つ以上のカテゴリが付与される。また、このカテゴリは、さらに詳細なカテゴリと関連付けることにより、カテゴリは階層的な構造となっている。このカテゴリ階層については、基本的には、下位カテゴリに属する記事は、上位カテゴリにも含まれるという包含関係が成立することが期待されることになっており、カテゴリ階層は、知識工学で用いられる概念階層と似た性質を持つことが期待されている。しかし、このカテゴリ階層の構造は、Wikipedia に登録される記事の増加に伴い、必ずしも、包含関係が成り立たない形でカテゴリ階層が作られるようになっている。次小節では、このような包含関係が成り立たないカテゴリが生成される原因である「分割のためのカテゴリ」について説明する。

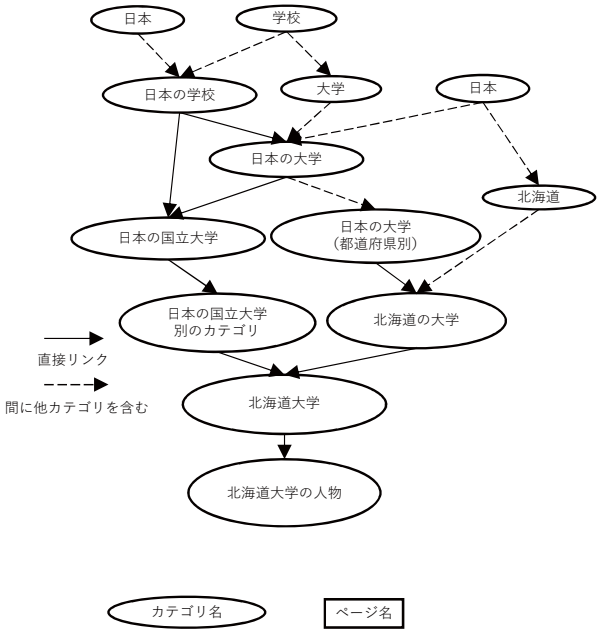


図 1: 分割のためのカテゴリを含むカテゴリ階層

2.2 分割のためのカテゴリ

Wikipedia カテゴリには、それぞれに所属するページ数に制限は無いが、一つのカテゴリに対し、大量のページが所属する場合、閲覧者の視点からすると、不都合であるため、そのカテゴリを分割するようなカテゴリが生成される。図 1 は「北海道大学」を例にとったものであるが、カテゴリ「北海道大学」が、カテゴリ「大学」の下に直接繋がっていたとすると、その他世界中の大学も同じように繋がることとなり、カテゴリ「大学」の下には数万もの大学名が所属し、閲覧者からすると意味のないものとなってしまう。そこで、カテゴリ「大学」を「日本の大学」、「日本の国公立大学」と分割していけば、閲覧者からすると、日本の国公立大学の一覧を見ることができるなどの利点を持つことができる。以上のような役割を持つカテゴリを

連絡先: 中川嵩教，北海道大学工学部情報エレクトロニクス学科，札幌市北区北 14 条西 9 丁目,011-706-7161，f-b-hawk07@eis.hokudai.ac.jp

*1 https://www.wikipedia.org/

「分割のためのカテゴリ」と呼ぶ。一方で、英語版 Wikipedia では、このようなカテゴリは「set and topic カテゴリ」（以下 SaT と略記）と呼ばれ、「分割のためのカテゴリ」でないカテゴリは、「set カテゴリ」、または、「topic カテゴリ」のどちらかに分けられる。先の例を見ると、「日本の大学」が SaT であり、「日本」が topic、「大学」が set である。ほとんどの場合、SaT カテゴリは、set と topic を共に、親（もしくは先祖）カテゴリに持つ。

3. Wikipedia カテゴリの問題点と再構成

前述の、set または topic と SaT の関係を見た時、親カテゴリの set または topic と、その子カテゴリの SaT の関係は、分割が行われているだけのため、包含関係が成り立つと考えて良い。しかし、SaT から分割でないカテゴリに到達した際に、包含関係が途絶えてしまうことがある。例を見ると、「大学」から分割が行われ、「北海道大学」までは大学についての階層であるが、その「北海道大学」の下には、「北海道大学の人物」という人物の階層が続いていく。このような問題が Wikipedia カテゴリの階層構造に多く含まれ、結果として、階層構造全体を知識工学的観点から活用する際の大きな妨げとなっている。つまり、問題の起こる原因は以下の仮説で説明することができると考えた。

仮説　上位カテゴリが topic カテゴリの場合（SaT の topic 部分など）、クラス制約を含め、包含関係が成り立たない場合が存在する。

この問題は、topic に用いられる概念については、SaT カテゴリの分割時に topic を基準に分割される場合を除いて、topic に関連する階層構造は、上位階層を参照しないということで、上記のような問題の大半が解決すると考えられる。ただし、topic 間の関係でも、地理的な包含関係などが成り立つ場合がある。この仮説の検証をするためには、以下の作業が必要となる。

1. カテゴリを set や topic といった、構成要素となるカテゴリ（以下 Single と略記）と SaT のような複合的なカテゴリ（以下 Combination と略記）に分割すること。
2. 構成要素となるカテゴリを set と topic に分割すること。
3. topic 間の階層関係がある場合には、それを分類すること。（地理的包含関係、メンバー、関連など）

4. 分析作業

今回は、前章記載の作業の 1,2 を行い、そのデータを用いて、Wikipedia のカテゴリ構造の中から、部分構造を取得し、その部分構造について考察する。作業 3 については今後の課題とする。

4.1 扱ったデータと作業方針

作業は、Wikipedia の 2017 年 10 月 20 日のダンプデータをもとに行った。このデータ中には、207,036 件のカテゴリが存在する。この実験では、スタブなどの Wikipedia のメンテナンスをするために設定されているカテゴリを除く、197,378 件から、更に、上下カテゴリに関する情報が存在しない 14,644 件を除いた、182,734 件のカテゴリを分析の対象とした。

まず、このカテゴリから分割のためのカテゴリを認識するために、以下の基準にあたるものに Combination、それ以外のものを Single として分類を行った。

- 動詞句を含むカテゴリ
- 付加情報が付くカテゴリ、また、それを上位カテゴリとして持つカテゴリ
- 「SaT」→「SaT」というペアの上下のカテゴリ
- 「set(または topic)」→「SaT」というペアの「SaT」のカテゴリ

そして、Single に分けられたものの中で、Combination → Single となっている Single に topic のラベルを付け、それ以外のものに set のラベルを付けた。そのラベルが付けられたデータを用いて、Single から始まり、Combination を辿り、topic に到達するまでの階層を部分階層として取得し、その部分階層の中に、包含関係が成り立たない関係が存在するかどうかを確認した。

4.2 分析結果

各カテゴリに対するラベル付けに関しては、57,976 件に Single、122,945 件に Combination のラベルが付与された。更に、Single の内 39,956 件に topic、18,020 件に set のラベルが付与された。

分類されたものの中から、500 個のカテゴリをランダム抽出し、誤っているものがあるかを確認したところ、Combination のラベルが付いたものの中には、Single とした方が良いものは 1 件（例：アラブの春）存在するのみであったが、Single のラベルが付いたものの中に、25 件（例：国籍別の在外人）ほど Combination とした方が良いものが存在した。前者の問題は、主に、固有名詞に関する問題なので、そのリストを並行して作成していくことにより、解消可能だと考える。後者については、Combination と判断するためのパターンが不足していることが原因である。先の誤りでは、「A」→「B の C」の「B の C」が Combination として抽出できていない。このような形は子の構成要素が、親を言い換えた形になることによって起こる。先の例では、「外国人」→「国籍別の在外人」となっている。これは「外国人≒外人」となるようなリストがあれば解決すると思われる。このように、見つかった例外事例から、さらなるパターンを作成することで、その精度の向上を目指す予定である。

set と topic に関する分類については、パターンにより topic と分類できない場合は、set とするという基準で分類を行っているために、topic として分類されたものにはほとんど誤りがないが、topic の分類に関するパターンが不足していることもあり、set の中に topic と分類したほうが良いもの（例：夏目漱石）が含まれている。こちらについても、獲得した例外事例を分析し、パターンを拡充することで、その精度の向上を目指す予定である。

ただ、両者ともに、一定レベルの精度が高いパターンが作成された場合に、それらのパターンによる分類結果に当てはまらないものをさらに分析していくという形で進めていくことによって、少しずつ分類が難しいカテゴリの特徴が明らかになっていくことを期待している。

次に、ラベル付けされたデータを用いて取得した部分階層の考察を、具体例を用いて述べていく。取得した部分階層の良い点としては、例えば、「学校」というカテゴリから、今回設定した基準で子カテゴリを辿っていくと、「学校」→「各国の学校」→「日本の学校」→「日本の大学」→「日本の国立大学」→「北海道大学」という階層が得られる。本来はこの下に「北海道大学の人物」という「人物」に関する階層が続くが、

「北海道大学」という topic カテゴリで階層が切れることにより、「学校」に関する階層のみを取得することができた。しかし、「学校」→「学校を題材とした作品」という set が追加され分割された時には、その下に付くカテゴリは、「作品」に関する階層が続いてしまっていたり、「学校」→「各国の学校」→「アメリカ合衆国の学校」→…→「カリフォルニア州の大学」→「サンフランシスコ美術大学の教員」のように Combination → Combination の際に、set の部分（あるいは set と topic 両方）の影響により、「人物」を表すカテゴリが入ってしまうことがある。前者に対しては、今回は、上（親）から下（子）へと辿っていったが、同じように、下から上に辿っていったときの階層と比較をすれば解消すると思われ、後者に対しては、Combination の構成要素に対し、それぞれが、set か topic かを断定し、set が変更された際の、許容範囲を設定すればある程度解消されるように思える。

上記を考慮して、もう 1 つの具体例として、「職業」というカテゴリから辿っていった結果を見ると、「職業」→「分野別の職業」→「メディア関連の職業」→「放送関連の職業」→「ナレーター」では（「ナレーター」が set か topic であるかの議論は今後していく必要があるが）、その下にある、「ナレータープロダクション」という企業を表すカテゴリへはいかず、職業の範囲で階層が途切れている。また、「職業」→「専門家」→「武道家」→「槍術家」という例を見ると、set 間であれば、包含関係を保ちながら、topic へと辿りつくことができるということがわかる。

4.3 想定する活用事例

まず、各カテゴリを SaT、set、topic に分類するだけでも、それぞれに注目をしたページ間の類似度を計ることができる。例えば、「北海道大学」と「札幌駅」を見ると、札幌駅は北海道大学の最寄駅となり、位置的にも、関係的にも非常に近い関係にあるが、Wikipedia カテゴリの類似度では、「北海道」→「北海道の教育」→「北海道の学校」→「北海道の大学」→「北海道大学」→「北海道大学（ページ）」と、「北海道」→「北海道の市町村」→「北海道の市」→「札幌市」→「札幌市の交通」→「札幌市北区の交通」→「札幌市北区の鉄道駅」→「札幌駅（ページ）」とで、非常に遠い。対して、「九州大学」は、位置的にはかなり遠い位置にあるが、Wikipedia カテゴリの類似度では、「日本の国立大学別のカテゴリ」→「北海道大学」→「北海道大学（ページ）」と、「日本の国立大学別のカテゴリ」→「九州大学」→「九州大学（ページ）」とで、かなり近いものとなる。どちらを近いものとしたいかは、用途によるが、「北海道」や「札幌市」といった topic を優先するか、「大学」のような set を優先するかということが、各々の類似度に関する重み付けなどによって表現可能となることが期待される。

更に、例えば「職業」というカテゴリから繋がっていくものを見ると、再整理された Wikipedia 構造では、最終的には職業名にたどり着くことが期待されるため、その一覧を辞書のような形で取得できる。その辞書は、先の例の「槍術家」のようなあまり一般的ではない職業を含むことができ、用途次第では、便利なものとなる。

4.4 データ公開に向けて

Single と Combination の分類については、もう少し基準をしっかりと付け、例外に関しては、そこまで多くの数があるわけではないので、最悪手作業で処理していけばよいと考えている。set と topic の分類については、まだいくつかの分類基準を増やしていかなければならないと思われる。現在考えているのは、形態素解析を用いた一般名詞（set）と固有名詞（topic）の同定や、言語間リンクを用いた英語版表記の確認等である。上記二つができれば、Combination の構成要素それぞれが、set か topic かを断定することができ、更に、その構成要素のつなぎとして使われる語（「の」「を題材とした」等）についても取得することができる。また、カテゴリ間の関係の同定も行っていく必要がある。

5. まとめ

本研究では、Wikipedia カテゴリをカテゴリ構造全体として扱う際の問題点についての議論を行い、簡単な基準で再整理を行い、考察した。set と topic の分類等がまだ不十分に思われたようにもかかわらず、topic で階層を止めるという単純な基準だけでも、本来包含関係が成り立たないようなを切捨て、包含関係が保たれている部分が多いような階層を取得することができた。これから、データを整理し、基準を増やしていくことで、Wikipedia カテゴリを、様々な研究に利用できるように再整理していきたい。

参考文献

[Bizer 09] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S.: DBpedia - A crystallization point for the Web of Data, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 7, No. 3, pp. 154 – 165 (2009)

[Hoffart 13] Hoffart, J., Suchanek, F. M., Berberich, K., and Weikum, G.: YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia, *Artificial Intelligence*, Vol. 194, No. 0, pp. 28 – 61 (2013)

[Ponzetto 07] Ponzetto, S. P. and Strube, M.: Knowledge Derived from Wikipedia for Computing Semantic Relatedness, *J. Artif. Int. Res.*, Vol. 30, No. 1, pp. 181–212 (2007)

[Taieb 13] Taieb, M. A. H., Aouicha, M. B., and Hamadou, A. B.: Computing semantic relatedness using Wikipedia features, *Knowledge-Based Systems*, Vol. 50, No. 0, pp. 260 – 278 (2013)

[Yoshioka 14] Yoshioka, M.: Analysis of Japanese Wikipedia Category for Constructing Wikipedia Ontology and Semantic Similarity Measure, in *Information Retrieval Technology 10th Asia Infomation Retrieval Societies Conference, AIRS 2014, Kuching, Malaysia, December 3-5, 2014 Proceedings*, pp. 470–481, Springer-Verlag GmbH (2014), LNCS8870

[玉川 10] 玉川 奨, 桜井 慎弥, 手島 拓也, 森田 武史, 和泉 憲明, 山口 高平?F 日本語 Wikipedia からの大規模オントロジー学習, 人工知能学会論文誌, Vol. 25, No. 5, pp. 623–636 (2010)

[藤原 12] 藤原 嵩大, 吉岡 真治?FWikipedia の階層関係を分析するためのカテゴリパターンの提案, 2012 年度人工知能学会全国大会 (第 26 回) 論文集 (2012), CD-ROM 2C1-NFC2-4

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

知識グラフ作成のための統合知識基盤の構築に向けて

Towards Implementation of Integrated Knowledge Infrastructure to Create Knowledge Graphs

市瀬 龍太郎 *1*2 Ryutaro Ichise
Natthawut Kertkeidkachorn*2 Natthawut Kertkeidkachorn
趙 麗花 *2 Lihua Zhao

*1国立情報学研究所
National Institute of Informatics
*2産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

Knowledge graphs are recently used for many AI tasks such as question and answering. However, there are still problems to create precise knowledge graphs. In this paper, we discuss three problems and propose a solution called integrated knowledge infrastructure consisting of four modules.

1. はじめに

機械に知的な作業を行わせるためには，知識が大きな役割を果たす．たとえば，質疑応答システムでは，質問に対する回答の作成に，知識を利用するし，Web 検索システムでは，ユーザに必要な情報を提供する際に，知識を利用している．このような知識を構築する際には，知識グラフ (Knowlege Graph) を用いられることが多い．知識グラフとは，＜主語，述語，目的語＞の３つ組（トリプル）を用いて，エンティティとその関係を表現したものである．

知識グラフを作成する際には，通常，Linked Data から必要な情報を取得し，それらを整理の上で，別のデータベースなどの情報を統合することで作成される．しかし，そのようなプロセスを経て作成しても，精密な知識グラフを作成することは困難であることが多い．作成が困難な点にはいくつかの理由がある．まず，第１に，利用した Linked Data やデータベースなどの資源自身が不完全である点である．利用するアプリケーションにも依存するが，知識グラフを作成する際には，様々な場合に対応できるようにするために，大量のエンティティに関する情報が集められる．そのため，全てのエンティティの全ての関係について，予め完全な情報を準備することは難しい．第２に，知識グラフは時間的に変化し，新しい情報を常に入れる必要がある点である．知識グラフは静的に存在するものではなく，常に情報が追加，更新される必要がある．そのために，最初の Linked Data やデータベースに記述されていなかった知識でも，新たに知識が生ずれば，その知識の記述方法を必要に応じて新たに定義し，知識を迅速に追加することが必要となる．第３に，知識グラフとして用いられた知識には，誤った知識が混入している場合がある点である．知識を作成する際には，人手による作業が行われるため，誤った情報が混入してしまうことが避けられない．本稿では，精密な知識グラフを作成するために，知識グラフの作成方法に関して，上記の問題点の考察を行い，必要な技術要素の同定，および，それを統合した精密な知識グラフを作成するためのフレームワークを提案する．

連絡先: 市瀬 龍太郎，国立情報学研究所情報学プリンシプル研究系，〒 101-8430 東京都千代田区一ツ橋 2-1-2，Tel:03-4212-2000，E-mail:ichise@nii.ac.jp

2. 知識グラフ作成のための諸問題

前章において，知識グラフを作成する際の問題点として，知識の資源自身が不完全である問題，新しい情報を追加する問題，誤った知識の混入問題の３つの問題を提起した．本章では，それらの問題について考察を行う．

2.1 知識資源の不完全性

この問題は，知識グラフを作成する際に，用いる資源が不完全なことで発生する問題である．例えば，既存の Linked Data を用いると，多数のエンティティに対する情報が入手できる反面，全ての情報が完全に網羅されているわけではない．例えば，多数の人物に関する情報が書かれていても，全ての人物に対して，国籍や生まれた場所などが記述されているわけではない．そのような欠けた情報は補完する方法が別途必要となる．

補完するためのアプローチとして，知識グラフ補完 (Knowledge Graph Completion) という方法が知られている．知識グラフ補完では，記述されている既知の情報を学習データとして用いることで，未知の関係の予測 (Link Prediction) を行う．例えば，ある人物に対して，国籍の記述がなかったとしても，生まれた場所の記述があれば，その情報を元にして，精度高く国籍の予測ができる．そのために，知識埋め込み手法 (Knowledge Embedding) などの技術が必要となる．

2.2 新しい情報の追加

この問題は，新しい知識を知識グラフにどのように追加するかに関する問題である．第 2.1 章で述べた手法により，新たな情報を追加することも可能であるが，予測に十分な知識が知識グラフ上に存在しない場合もある．そのため，新たな資源から知識を抽出することが必要となるが，データベースのように知識が整理された資源は多くはない．そこで，大量に存在し，新たな情報が随時入手可能なテキスト情報から知識を作ることが望ましいであろう．テキスト情報からトリプルを作成するには，領域に依存せずにテキストから関係を抽出できる技術 (Open IE: Open Information Extraction) や照応解析 (Coreference Resolution) の技術が必要となる．その後，作成されたトリプルを知識グラフに統合する必要がある．統合の際の問題は，２つに分けることができる．一つは，既存のオントロジーで表現できる知識を対象とし，単に欠けている知識を既存のオントロジーで追加する場合である．もう一つは，既存のオントロジーでは，表現できない知識を対象とし，オントロジー自体を変更

する必要のある場合である．例えば，人物を表現する場合に，出身校を表現する述語がオントロジーとして定義されていなければ，出身校を定義するようなオントロジー記述を新たに追加しなければ，十分な知識を表現できない．前者の場合には，オントロジーマッチング技術 (Ontology Matching) などにより解決することができるが，後者の場合には，オントロジー学習技術 (Ontology Learning) [市瀬 09] などが必要となる．

2.3 誤った知識の混入

この問題は，知識グラフに誤った知識が混入する問題である．実際の Linked Data データでは，このような問題はしばしば見られる．例えば，DBpedia では，ある会社の立地する国名がケルンであると記述されている．しかし，ケルンは国名ではなく地名であり，正しくは，ケルンが存在するドイツとするべきである．このような誤った知識に対応するためには，知識グラフにおける誤りを検知する技術 (Error Detection) や，誤った知識に対して修正する技術 (Error Correction) が必要となる．

3. 統合知識基盤

前章で考察したように，精密な知識グラフを作成するプラットフォームは，3つの問題に対処しなければならない．そのために，既存の知識グラフから足りない知識を補完する手法，テキストから知識グラフに欠けている知識を補うための関係抽出手法，得られた知識を統合するためのオントロジー統合手法，そして，正しい知識グラフになるように，誤りを検知，修正する手法を統合する必要がある．

本稿では，これらを統合するための統合知識基盤を提案する．そのアーキテクチャを図 1 に示す．本アーキテクチャは，主に 4 つのモジュールより構築されている．1 つ目は，図 1 の右上にある知識補完 (Knowledge Completion) モジュールで，既存の知識グラフから足りない部分を補完した知識グラフを生成する．この部分は TorusE [Ebisu 18] の技術を用いることができる．2 つ目は，図 1 の左上にある関係抽出（Relation Extraction) モジュールで，テキスト情報からトリプルを作成する．その際に，専門家が作ったブートストラップデータを用いることで，より的確なトリプルの作成が行える．この部分は，T2KG [Kertkeidkachorn 18] の技術を用いることができる．3 つ目は，図 1 の左下にあるオントロジー統合 (Ontology Integration) モジュールで，新たなトリプルを知識グラフに統合する．その際に，既存のオントロジーにない概念が含まれる場合には，オントロジーの学習を行う．この部分は，T2KG [Kertkeidkachorn 18] や FITON [Zhao 14] の技術などを用いることができる．4 つ目は，図 1 の右下にある知識検証 (Knowledge Verification) モジュールで，知識グラフに誤りがないかを調べ，状況に応じて修正を行う．この部分は，ALDErrD [Rahoman 16] や誤り修正 [Lertvittayakumjorn 17] の技術などを用いることができる．以上のような統合を行うことで，精密な知識グラフを作成することが可能となる．

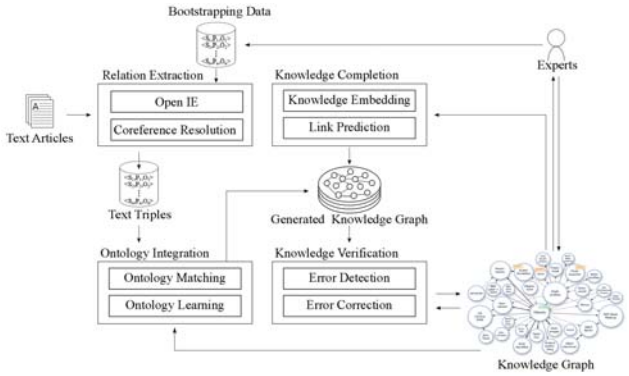


図 1: 統合知識基盤のアーキテクチャ

4. まとめ

本稿では，精密な知識グラフを作成するために必要な課題を整理し，そのための技術について考察を行った．その考察に基づき，知識グラフを生成するためのプラットフォームとして，統合知識基盤のアーキテクチャを提案した．統合知識基盤は，既存の知識グラフに対して，自動的に誤りの検知，修正を行い知識グラフの改善をする機能，知識グラフ上で欠けている知識を補完する機能，テキストから知識を抽出し新たに追加する機能などがあり，精密な知識グラフを作成するための基盤技術として，大きな役割を果たすことが期待される．今後は，第 3. 章で述べた技術をモジュール化し，本稿で提案したアーキテクチャで統合するとともに，知識グラフが必要となる特定ドメインを例として，知識グラフを実際に整備することを試みる予定である．

謝辞

本研究の一部は，新エネルギー・産業技術総合開発機構 (NEDO) の支援により実施されたものである．

参考文献

[Ebisu 18] Ebisu, T. and Ichise, R.: TorusE: Knowledge Graph Embedding on a Lie Group, in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, AAAI (2018)

[Kertkeidkachorn 18] Kertkeidkachorn, N. and Ichise, R.: An Automatic Knowledge Graph Creation Framework from Natural Language Text, *IEICE Transactions on Information and Systems*, Vol. 101-D, No. 1, pp. 90–98 (2018)

[Lertvittayakumjorn 17] Lertvittayakumjorn, P., Kertkeidkachorn, N., and Ichise, R.: Resolving Range Violations in DBpedia, in *Proceedings of the 7th Joint International Semantic Technology Conference*, Vol. 10675 of *Lecture Notes in Computer Science*, pp. 121–137, Springer (2017)

[Rahoman 16] Rahoman, M.-M. and Ichise, R.: Automatic Erroneous Data Detection over Type-Annotated Linked Data, *IEICE Transactions on Information and Systems*, Vol. E99-D, No. 4, pp. 969–978 (2016)

[Zhao 14] Zhao, L. and Ichise, R.: Ontology Integration for Linked Data, *Journal on Data Semantics*, Vol. 3, No. 4, pp. 237–254 (2014)

[市瀬 09] 市瀬 龍太郎：オントロジー学習 - 計算機によるオントロジー構築 -, 電子情報通信学会誌, Vol. 92, No. 9, pp. 791–795 (2009)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

弓道練習支援システムにおける射法八節の認識のためのデータ収集について

Training Data on Japanese Archery Motion Recognition for Self-training Support

両角 貴弘 *1　大圃 忠親 *1　新谷 虎松 *1
Takahiro Morozumi　Tadachika Ozono　Toramatsu Shintani

*1名古屋工業大学大学院工学研究科情報工学専攻
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

Self-training of Kyudo, i.e. Japanese archery, is difficult for novices but learning from experts is costly. We aim to develop a self-training support system for Kyudo that evaluates shooting forms for trainees. The shooting form of Kyudo, called Shaho-Hassetsu, consists of eight steps. To realize the evaluation function, we plan to build a recognition system on Shaho-Hassetsu that splits the shooting form into the eight steps. To make training data for the Shaho-Hassetsu recognition, we collected shooting form data of Kyudo experts by using the Kinect motion tracker and implemented a tagging support system to identify the eight parts. This paper explains how to collect and process the data for the recognition system.

1. はじめに

スポーツなどのスキルの習得に関して，練習者に対して熟達者のような指導が可能なシステムの実現が望まれている．本研究では，弓道を題材として，練習者に対する指導を可能とするシステムの実現を目指している．本稿では，弓道において基本となる型である射法八節を，システムが認識するために必要なデータの収集について述べる．また，射法八節習得支援システムの試作について述べる．本研究では弓道を題材とした．弓道には段位が存在し，本研究では初段以上を取得している者を熟達者として扱う．また，八節から成る射法八節におけるフォーム（射形）の骨格情報を収集する．射形を評価する際に，重点的に評価したい節が撮影した動画において，どの部分であるかを把握するのは手間である．そのため，射法八節における各節の自動認識は有益である．

本研究の目的は，収集する骨格情報における正しい射形を行うための共通する骨格情報の模索，射形評価を支援するシステムの試作，基本の型における八節の自動認識の前段階として，認識モデルのためのデータセット作成である．

2. 弓道

弓道は日本の武道の一つであり，日々の修練の結果として指標となる級位と段位が存在する．全日本弓道連盟の初段の審査基準として，基本の型に適って矢所の乱れぬ程度に達した者とある *1．つまり，初段以上を有している者は基本の型を行うことが可能な者と判断できるため，本研究では初段以上の者を熟達者データとして取り扱う． 射法八節とは，八つの動作に区分されており，足踏み，胴造り，弓構え，打起し，引分け，会，離れ，残心がある．各節の説明を表 1 に示す．これら八つの動作を一連の動作として行う．各節における骨格の動きは重要であり，骨格点の相関関係を数値化することで評価を行うことが可能になる．弓道は，正しい射形データの指標となる段位が存在する．また，射を行う動作は動きが少なく，モーションセンサでの骨格点の取得が正確に行うことが可能である．このため，本研究では弓道を対象とした．

表 1: 射法八節の一覧 *2

足踏み	足を開き，正しい姿勢を作る
胴造り	弓を左膝に置き，右手は右の腰にとる
弓構え	右手を弦にかけ，左手（手の内）を整えてから的を見る
打起し	弓構えの位置から，静かに両拳を同じ高さに持ち上げる
引分け	打起こした弓を，左右均等に引分ける
会	引分けが完成し，心身が 1 つになり，発車のタイミングが熟すのを待つ
離れ	胸郭を広く開いて，矢を放つ
残心	射の総決算．矢が離れた時の姿勢をしばらく保つ

3. 方法

本研究では，射法八節を認識するためのデータセットを作成するために，1）射法八節の映像・骨格情報の収集，2）骨格情報の評価・可視化，および 3）映像上における各動作に対するタグ付けの 3 点を行うことにした．

本研究では，骨格情報の評価・可視化として骨格点の相対関係を数値化することを行う．そのため，モーションセンサとして，Kinect v2（以後，Kinect）を用いた．Kinect は，25 点の骨格点の三次元座標を取得することが可能である．図 1 は Kinect を用いて，骨格点の認識を行った様子である．青色の丸は，認識している骨格点である．本研究では，1）両肩の骨格点，2）両肩の中央点と腰の中央点の相対関係を取り扱った．図 1 では，1）を実線，2）を点線で表している．

3.1 撮影

本研究では，射法八節を撮影しつつ Kinect による骨格点の取得を行った．被験者は，名古屋工業大学の弓道部員 14 名であった．被験者の構成は，男性 11 名，女性 3 名であった．全員

連絡先: 両角貴弘, 名古屋工業大学工学研究科情報工学専攻, 愛知県名古屋市昭和区御器所町,052(735)5584, moroz@toralab.org

*1 http://www.kyudo.jp/pdf/documents/probation_standard.pdf
*2 http://www.kyudo.jp/howto/syaho.html



図 1: 取り扱う相対関係

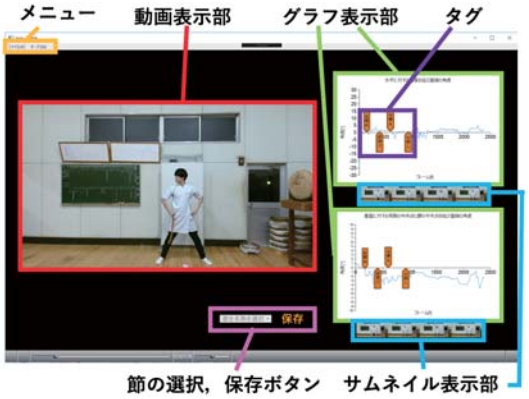


図 2: システム実行例

が有段者であった．被験者は基本的には袴を着用した．撮影場所は，名古屋工業大学の弓道場とした．撮影に用いた Kinect は，被験者から 4m 離れた地点に設置した．各撮影において，直立姿勢から射法八節を行い，その後直立姿勢に戻る過程を 1 セットとした．各自 3 セットずつ撮影し，計 41 セットのデータを得た [*3]．

3.2 収集データ

収集したデータは，カラー画像，深度画像，骨格点の三次元座標である．それぞれのデータの取得頻度は 30fps である．深度画像は，Kinect に搭載されている赤外線カメラによって取得可能である．カラー画像，深度画像は動画として保存した．また，骨格点の三次元座標は 25 点取得し，毎フレームにおける座標データを csv ファイルに出力する．三次元座標は Kinect v2 を原点としている．撮影の際には性別，身長および段位を記録した．

1），2）の相対関係として，それぞれ 2 点間を結ぶ直線の傾きを算出した．1）は XY 平面における傾き，2）は YZ 平面における傾きとなっている．Kinect で取得した骨格座標が外れ値となることがある．Peng ら [Peng 15] と同様に，2 重指数平滑フィルタを用いて，座標の補正を行った．平滑化パラメータとして，$0 < \alpha < 1$，$0 < \gamma < 1$ を用いる．α は古いフレームの座標と比較して，最近のフレームの座標に対する重みを指定する．γ は座標の傾向に対する重みを指定する．二重指数平滑フィルタは，以下の式で定義される．

$$\hat{X}_i = \alpha X_i + (1-\alpha)(\hat{X}_{i-1} + b_{i-1}), \qquad \hat{X}_0 = X_0 \tag{1}$$

$$b_i = \gamma(\hat{X}_i - \hat{X}_{i-1}) + (1-\gamma)b_{i-1}, \qquad b_0 = X_0 \tag{2}$$

$\hat{X}_i$ は i フレームにおけるフィルタ処理後の値，X_i は元の座標データである．b_i は i フレームにおける座標の傾向である．

4. 射法八節タグ付け支援システム

本研究では，撮影した射形データを再生する機能，射形データに対してタグ付けを行う機能を持つシステムを試作した．システムの画面を図 2 に示す．まず，システムを起動し，メニューの”開く”からフォルダを選択する．選択するフォルダには，射形データとして収集したカラー動画と骨格座標の csv ファイルが格納されている．選択すると，図 2 のようにシステムの画面に動画とグラフが表示される．上部，下部のグラフはそれぞれ，第 3 章で述べた両肩の骨格点，両肩の中央点と腰の中央点の相対関係を表す．どちらのグラフも横軸はフレーム番号である．縦軸は，上部は両肩の骨格点の結んだ直線の右肩を原点とした傾きの値であり，下部は腰の中央の骨格点を原点とした傾きである．本システムでは，動画の基本的な再生方法であるタイムスライダーに加えて，グラフの任意の点をマウスクリックすることで，指定した点に対応するフレームから動画を再生する機能を実装した．また，グラフの下部には動画のサムネイルが表示されている．

射法八節の認識モデル作成のためのデータを収集するためにタグ付け機能を実装した．画面上部にあるメニューにモード変更がある．そこから編集モードが選択できる．編集モードに変更すると，射法八節の名称，保存ボタン表示される．タグ付けは，射法八節の各節の名称を選択し，ユーザが指定した箇所にタグを配置する．その後，保存することでデータベースに保存される．また，モード変更から再生モードに変更すると，グラフ上にタグが配置されている．通常の再生機能に加えて，グラフ上の任意の点をマウスクリックすることで，その点に対応したフレームから動画を再生する機能によって，ユーザによる射形の観察を支援することができた．さらに，この機能はタグ付けを行う支援にもなる．また，動画のサムネイルをグラフの下部に表示することで，グラフと動画のがりを分かり易くすることで，タグ付けを行う支援をすることができた．タグ付け機能を実装することで，射法八節の認識モデル作成のためのデータを収集することが可能となった．

5. おわりに

本研究では，弓道を題材として，システムによって射形評価を支援する射法八節習得支援システムを試作した．また，タグ付け機能を実装することで，射法八節における各節の認識を行うためのデータセットを作成した．今後は，本研究で取り扱わなかった骨格点の相対関係も考察し，作成したデータによって射法八節の自動認識を行う．

参考文献

[Peng 15] Peng, Qiang and Chen, Weihai and Wu, Xingming and Wang, Jianhua: A novel vision-based human motion capture system using dual-Kinect. 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), 51-56p (2015)

*3 1 セットはデータの欠損が見つかったため除去した．

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

モノとコトに基づいた知識表現モデル

A Model for Knowledge Representation Based on the Concepts of Things and Events

山田隆弘[*1]
Takahiro Yamada

[*1] 宇宙航空研究開発機構 宇宙科学研究所
Japan Aerospace Exploration Agency/Institute of Space and Astronautical Science

To enable automatic processing of human knowledge, various technologies are being studied, such as Knowledge Graph and Linked Data. Several other technologies were studied in the past for representing human knowledge, such as the Production System, Objects, and Frames. These technologies are used to represent knowledge independently of the semantic contents of the knowledge. In order to enable sharing of knowledge, however, these technologies are not sufficient because there can be multiple ways of representing the same knowledge with these technologies. To integrate knowledge bases, each developed by a different group, there must be a guideline or standard that specifies how to represent knowledge based on the semantic contents of the knowledge. This paper proposes a unified model for representing contents of knowledge based on the concepts of things and events. This model combines the object-oriented modeling method with Neo-Davidsonian event semantics.

1. はじめに

人間の持っている知識を機械的に処理できるようにするための技術として Knowledge Graph [Wang 14]や Linked Data [Heath 11]等が注目を集めている。また、以前から知識を形式的に表現する方法として論理式、Production System、オブジェクト、フレーム等の方式が注目されてきた[Brachman 04]。これらの方法は、様々な知識を統一的に表現するための技術として開発されたものであり、知識の内容には依存しない。

知識の内容に依存せずに知識を表現する技術は、知識を機械的に処理できるようにするための基礎として必要な技術である。しかし、それだけでは知識の共有のためには不十分である。なぜならば、一つの（あるいは同様の）知識を Knowledge Graph（あるいは論理式）として表現する方法には、複数のものがあり得るからである。まず、Knowledge Graph などで使うべき語彙が必ずしも統一されていないという問題がある。世の中には schema.org [schema.org]のような標準語彙集が存在しているが、それが使われない場合もあり得るし、そのような標準語彙集が存在しない分野も多い。さらに、複雑な知識をどのようにグラフ化するかという基本的な考え方が統一されていない。

一つの知識を複数の異なる方法で表現できるということは、作成元の異なる Knowledge Graph を組み合わせて一つの大きな Knowledge Graph とするような場合、必ずしもうまく組み合わせられないことを意味する。このような状況では、知識の共有や再利用は達成されないであろう。知識の共有を進めるためには、知識の内容に応じて知識をどのように表現するかという基準を作り、同一の基準を適用して様々な知識をグラフ化する必要がある。そこで、本稿では、そのための基礎として、知識内容を表現するためのモデルを提案する。

本稿では、知識内容の基本的な構成要素はモノ(entity)とコト(event)であると考える。人間の知識にモノとコトの知識が含まれることは、多くの研究で言及されているが、モノとコトの双方を含めた知識の全体像を表現する方法に関する研究は、ほとんど存在していない。そこで、本稿では、モノとコトの知識を一体的に表現する方法としてのモデルを提案する。ここで提案するモノの表現方法は、オブジェクト指向モデリング[Blaha 2005]に基づいている。また、コトの表現方法は、Neo-Davidsonian の事象意味論(event semantics)と呼ばれている方法[Parsons 90]に基づいている。本稿の試みは、この両者を組み合わせることによって知識内容を一体的に表現するためのモデルを制定しようというものである。

なお、本稿は、著者の談話理解に関する研究[山田 18]の中の知識構造に関する部分をさらに発展させたものである。

連絡先：山田隆弘，宇宙航空研究開発機構宇宙科学研究所，
相模原市中央区由野台 3-1-1，tyamada@isas.jaxa.jp

2. モノに関する知識

2.1 モノ

モノ (entity) とは、ある世界に存在している何かである。ここでの世界とは、現実の世界でも仮想的な世界でも良い。知識には何かの計画に関する知識も存在する。例えば、鈴木さんが新たに家を建てることを計画している場合、その家は現実には存在していないが、鈴木さんはその家に関する知識を保有しているはずである。このような場合の家もモノであると考える。

モノは、一つあるいは複数の属性(attribute)を有している。例えば、家の場合、床面積、外壁の色、価格などを属性として有している。それぞれの属性は、値を有する。例えば、床面積という属性は 120 平米という値を有し、外壁の色という属性は黒という値を有する。

あるモノについて、そのモノの属性と属性値との対を状態(state)と呼ぶことにする。例えば、ある家の「外壁の色が黒である」という言説は、その家の状態の一つを表したものである。

モノは、他のモノとの間に関係(relation)を有することがある。例えば、人が家を所有する場合、人というモノと家というモノとの間に「有する」という関係が存在していると考える。また。ある家がいくつかの部屋を有している場合、家というモノと部屋というモノの間に「有する」という関係が存在していると考える。

ここで定義した概念（すなわち、モノ、関係、属性）は、オブジェクト指向モデリング[Blaha 2005]で使われているものと同様のものであり、Unified Modeling Language (UML) [Booch 05]の用語に従えば、上記の 3 つの概念は、それぞれ、オブジェクト(object)、リンク(link)、アトリビュート(attribute)に相当する。

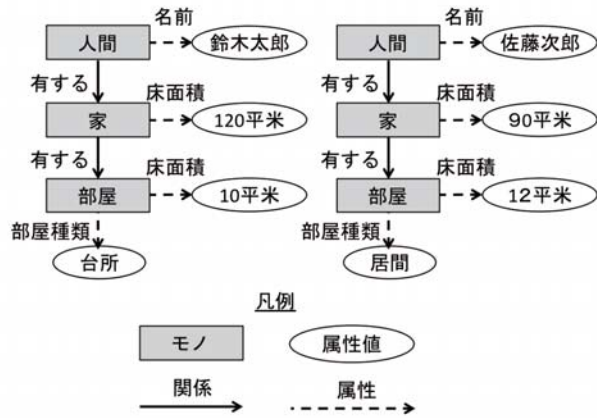


図 1 モノを図式的に表した例

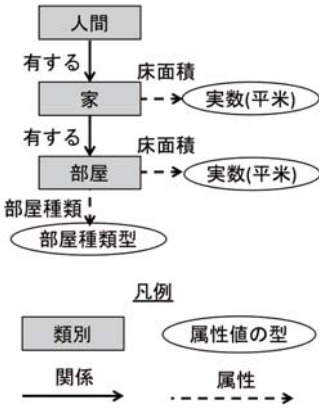


図 2 部類(モノの概念)を図式的に表した例

以上で定義した概念の例を図式的に表したものが図 1 である。この図では、モノ、関係、属性、属性値のそれぞれが凡例に示した記号を用いて示されている。これらの概念は UML のオブジェクト図を用いて表すことも可能である。図 1 では、鈴木太郎の家も佐藤次郎の家も一つしか部屋を有していないように描かれているが、図を見やすくするために他の部屋は省略されている。また、人間も家も部屋も、この図に示されていない属性も有しているのであるが、それらも省略されている。

モノに関する基本的な知識は、上で述べた方法で表現できるが、属性値については、さらに付加的な情報が必要になることがある。それらは、時制(tense)と法性(modality)である。時制については、「太郎の家の外壁は、以前は黒かったが、今は赤い」という知識があり得る。この場合は、個々の属性値(黒や赤)にさらに時間という属性とその属性値(昔、2017 年以前、今、2018 年 5 月以降等)を付加する必要がある。法性については、「太郎の家の外壁は、黒いはずである」という知識があり得る。この場合も、属性値(黒)にさらに法性という属性とその属性値(可能性、必然性等)を付加する必要がある。

2.2 部類(モノの概念)

モノには、似たようなモノが存在する。例えば、鈴木太郎の家と佐藤次郎の家は、様々な共通点を有している。いろいろの家は、共通の属性を使って表すことができるが、それぞれの家は、異なる属性値を有している。例えば、鈴木太郎の家の床面積は 120 平米であるが、佐藤次郎の家の床面積は 90 平米である。また、どのような部屋を有しているかは、それぞれの家によって異なっているが、どの家も何らかの部屋を有している。

このように共通の属性と関係を有しているモノの集合を部類(class)と呼ぶことにする。あらゆるモノは、何らかの部類に属している。場合によっては、一つのモノが複数の部類に属していることもある。部類は、オブジェクト指向モデリングにおけるクラス(class)に相当する。図 1 に描かれている「人間」は、実際には、人間という部類に属するモノとしての人間を表している。同図の「家」と「部屋」も同様である。部類は、UML におけるクラス(class)に対応する。

部類と部類の間にも関係があり得る。例えば、モノとしての家はモノとしての部屋を有することができる。このような事実は、部類としての家と部類としての部屋の間の関係として定義できる。このような部類と部類の間の関係は、UML ではアソーシエーション(association)と呼ばれている。

図 1 の表記法に基づいて家と部屋という二つの部類の定義を表したものが図 2 である。図 2 に描かれている「家」は、モノとしての家ではなく、部類としての家である。図 2 では、(1) 部類、(2) 一つあるいは複数の部類に属するモノの間に成立し得る関係、(3) 各々の部類に属するモノが有すべき属性と属性値のデータ型が示されている。これと同様な情報は、UML のクラス図によっても表現できる。

第 2.1 節で述べたモノに関する知識は、個々のモノに関する個別的な知識であり、それをここでは個別知識と呼ぶことにする。一方、本 2.2 節で述べた部類に関する知識は、概念レベルの知識であり、これをここでは概念知識と呼ぶことにする。

3. コトに関する知識

3.1 コト

コト(event)とは、ある世界で生起した何かである。ここでも、世界とは、現実の世界でも仮想的な世界でも良いものとする。

コトは、一つあるいは複数の参加者(participant)を有している。各々の参加者は、モノである。例えば、「鈴木太郎が 2018 年 3 月 14 日に家を買う」は一つのコトである。この場合、鈴木太郎が動作主という参加者であり、家が対象という参加者である。また、コトも属性を有している。「鈴木太郎が 2018 年 3 月 14 日に家を買う」の場合、買った時期は属性であり、その値は 2018 年 3 月 14 日である。また、「鈴木太郎が山本三郎に 2018 年 6 月から 2020 年 5 月まで月額 5 万円で家を貸す」もコトである。この場合の参加者は、鈴木太郎と山本三郎と家である。「2018 年 6 月から 2020 年 5 月まで」は期間という属性の値であり、「月額 5 万円」は対価という属性の値である。

ここで定義した概念(すなわち、コト、参加者、属性)は、Neo-Davidsonian 事象意味論(event semantics) [Parsons 90]で使われている概念と類似している。

以上で定義した概念の例を図式的に表したものが図 3 である。この図では、コト、参加者、属性のそれぞれが凡例に示した記号を用いて示されている。

コトに関する基本的な知識は、上で述べた方法で表現できるが、さらに付加的な情報が必要になることがある。それらは、相(aspect)と法性(modality)である。相については、「太郎は走っていた」という知識があり得て、この場合は、「走る」というコトに相という属性と進行中という属性値を付加する必要がある。法性については、「太郎は走るだろう」という知識があり得る。この場合も、「走る」というコトにさらに法性という属性とその属性値(可能性、義務等)を付加する必要がある。

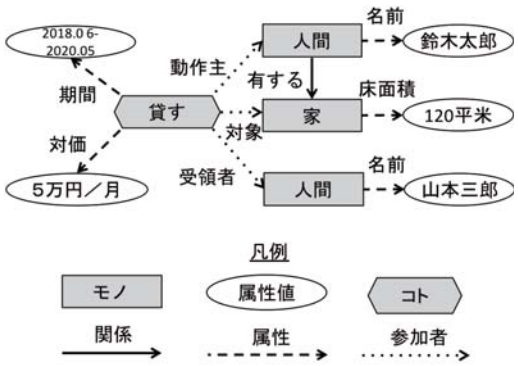


図 3 コトを図式的に表した例

3.2 行為(コトの概念)

コトには、似たようなコトが存在する。例えば、「鈴木太郎が山本三郎に家を貸す」というコトと「佐藤次郎が太田四郎に自動車を貸す」というコトは、様々な共通点を有している。さまざまな「貸す」というコトは、共通の参加者の種類(すなわち、動作主、受領者、対象)を有しているが、それぞれのコトにおける具体的な参加者は、コト毎に異なっている。例えば、鈴木太郎が家を貸す場合は、動作主は鈴木太郎、受領者は山本三郎、対象は家であり、佐藤次郎が自動車を貸す場合は、動作主は佐藤次郎、受領者は太田四郎、対象は自動車である。また、それぞれの「貸す」コトは、共通の属性も有している。

このように共通の参加者と属性を有しているコトの集合を行為(action)と呼ぶことにする。あらゆるコトは、何らかの行為に属している。行為は、Neo-Davidsonian 事象意味論における行為文(action sentence)の述語(predicate)に相当する。図 3 に描かれている「貸す」は、実際には、「貸す」という行為に属するコトとしての「貸す」を表している。

図 3 の表記法に基づいて「貸す」という行為の定義を表したものが図 4 である。図 4 に描かれている「貸す」は、コトとしての「貸す」ではなく、行為としての「貸す」である。図 4 では、(1) 行為、(2) 各々の行為に属するコトが有すべき参加者とその部類、(3) 各々の行為に属するコトが有すべき属性と属性値のデータ型が示されている。

第 3.1 節で述べたコトに関する知識は、個々のコトに関する個別的な知識であり、個別知識に分類される。一方、本 3.2 節で述べた行為に関する知識は、概念レベルの知識であり、概念知識に分類される。

4. 関係と規則に関する知識

4.1 個別知識としての関係

モノとモノとの間の関係については、第 2.1 節に述べたように「鈴木太郎が 120 平米の家を有する」などのものがある。

それ以外にも、個別知識の間には様々な関係が成立する。これらの大部分は、属性値と属性値の間、属性値とコトの間、あるいは、コトとコトの間に成立する。これらの関係は、談話分析の分野で談話関係(discourse relation) [Pitler 08]や修辞関係(rhetorical relation) [Asher 03]と呼ばれているものにほぼ対応する。[Pitler 08]に基づいて幾つかの例を以下に示す。

- 彼は、一日中テニスをしていたので、非常に疲れている。(因果関係)

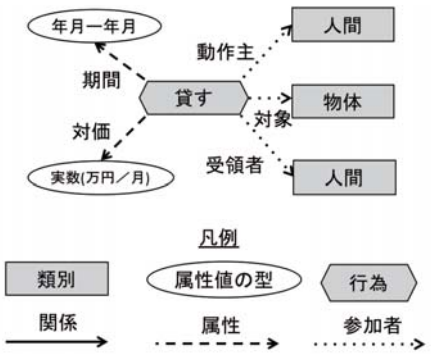


図 4 行為(コトの概念)を図式的に表した例

- 彼は、非常に強くはないが、驚くほど速く走ることができる。(対比)
- 我々は、午後にお茶を飲み、その後、レストランに晩餐会に行った。(時間順序)

これらの関係を図式的に表すには、該当する属性値やコトの間に談話関係を表すリンクを張れば良い。

4.2 概念知識としての規則

モノとモノとの間に関係が存在するかどうかを規定するための概念レベルの知識が規則(rule)である。規則は、複数の部類(あるいは属性値)の間、複数の行為の間、複数の部類と行為の間に成立する。規則には様々なものがあるが、例としては、論理的な因果関係(例えば、X を買うと X を所有するに至る)、物理法則(例えば、水の温度が摂氏 0 度になると氷に変化する)等がある。また、Schank [Schank 77]のスキーマ(schema)のような社会的な通例に基づく順序関係(例えば、メニューを見てから注文する)も含まれる。規則の中には、非単調論理(例えば、食べると通常は満腹になる)も含める必要がある。

これらの規則を一般的に表現する方法については、今後の検討課題としたい。

5. 知識の全体像

第 2 節から第 4 節までに述べてきた知識の全体を一つの表にまとめると表 1 のようになる。

表 1 知識の全体像

	個別知識	概念知識
モノ	• モノ • 属性(属性値) • 関係	• 部類 • 属性(属性値の型) • 関係(どの部類間に成立するか)
コト	• コト • 参加者 • 属性(属性値)	• 行為 • 参加者とその部類 • 属性(属性値の型)
関係と規則	• 談話関係	• 論理的因果関係 • 物理法則 • 社会的通例 • その他

6. 本モデルの意義

本稿では、知識を表現するための一般的な概念とともに、その概念を用いて知識を図式的に表現する方法を提示した。図 1

から図 4 までに示したグラフは、そのまま RDF を用いて Knowledge Graph あるいは Linked Data として表現することができる。また、既存の Knowledge Graph や Linked Data も図 1 から図 4 までに示したグラフと同様な構造を持つものが多い。

しかし、単に RDF で書けるというだけでは、知識の共有は困難である。知識の共有を目指すのであれば、第 1 節で述べたように、語彙の統一とともに知識内容を表現するための方法の統一が必要である。本稿で提示したモデルは、知識内容を表現するための統一的な方法を提案するものであり、知識共有を実現するための重要な一つのステップとなる。

本稿では、図 1 から図 4 までに示したように、モノやコトを複数のグラフの集合として表現する方法を示したが、モノやコトを人間に分かりやすい形で扱うためには、モノやコトを一つのまとまった存在(例えば、フレーム)として表現する方が望ましいだろう。フレームとして表現したモノやコトは、グラフにも容易に変換できる。人間に分かりやすい形でモノやコトを表現する方法の開発は今後の課題である。

7. おわりに

本稿では、モノとコトの概念に基づいて知識を一般的に表現するための一つのモデルを提示した。これは、従来から存在しているオブジェクト指向モデリングと Neo-Davidsonian の事象意味論とを一体化させたものである。しかし、この研究はまだ初期段階にあり、今後さらに精密化する必要がある。特に、第 4.2 節で述べたように、規則の一般的な表現方法は今後に残された大きな課題である。また、第 6 節で述べたように、モノやコトを人間に分かりやすい形で表現する方法の開発も今後の課題である。

本稿の研究は、そもそもは談話の意味を記述するための研究の一部として始めたものであるが[山田 18]、この研究の適用範囲はそれだけではない。再利用可能な Knowledge Graph や Linked Data を構築するためには、知識内容に基づいた知識の表現方法に関する統一的な基準が必要であり、本稿の研究は、そのような知識の表現方法の基準作成のための基礎としても使用されるべきものである。

参考文献

[Asher 03] Asher, N., Lascarides, A.: *Logics of Conversation*, Cambridge University Press (2003).

[Blaha 05] Blaha, M., Rumbaugh, J.: *Object-Oriented Modeling and Design with UML*, Second Edition, Prentice Hall, 2005.

[Booch 05] Booch, G., Rumbaugh, J., Jacobson, I.: *The Unified Modeling Language User Guide*, Second Edition, Addison-Wesley, 2005.

[Brachman 04] Brachman, R. J., Levesque, H. J.: *Knowledge Representation and Reasoning*, Morgan Kaufmann (2004)

[Heath 11] Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool Publishers (2011).

[Parsons 90] Parsons, T.: *Events in the Semantics of English*, MIT Press (1990).

[Pitler 08] Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., Joshi, A.: Easily identifiable discourse relations, In *Proceedings of COLING 2008* (2008).

[Schank 77] Schank, R. C., Abelson, R.: *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates (1977).

[Schema.org] Schema.org: Schema.org, http://schema.org/

[Wang 14] Wang, Z., Zhang, J., Feng, J., Chen, F.: Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (2014).

[山田 18] 山田隆弘: 世界知識の構造に基づいた談話理解モデル, 言語処理学会第 24 回年次大会, A2-3, 2018.

[2F1] [General Session] 12. HI / Education Aid

座長:松原 行宏（広島市立大学大学院）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room F (4F Garreria)

[2F1-01] Measuring Beginner Friendliness of Chinese Web Pages Explaining Academic Concepts using HTML Structures

○BINGCAI HAN[1], Takahide Kasuga[1], Hayato Shiokawa[1], Kota Kawaguchi[1], Takehito Utsuro[1], Yasuhide Kawada[2] （1. University of Tsukuba, 2. Logworks Co., Ltd.）

9:00 AM - 9:20 AM

[2F1-02] Estimation of understanding of words by lifelog analysis for vocabulary visualizing

○Kazuma Nozawa[1], Noriki Mochizuki[1], Gakuto Masuyama[2], Sousuke Nakamura[3] （1. Department of Science and Engineering, Hosei University Graduate School, 2. Faculty of Science and Technology, Meijo University, 3. Faculty of Science and Technology, Hosei University ）

9:20 AM - 9:40 AM

[2F1-03] Implementing a Weekly Report Management System Based on Activity Records for Reuse of Past Reports

○Rikiya Ando[1], Tadachika Ozono[1], Toramatsu Shintani[1] （1. Nagoya Institute of Technology）

9:40 AM - 10:00 AM

[2F1-04] Effectiveness of Partial Fading of Text on Reading Comprehension

○Jumpei Kobayashi[1,2], Toshio Kawashima[2] （1. Dai Nippon Printing Co., Ltd., 2. Future University Hakodate）

10:00 AM - 10:20 AM

[2F1-05] Proposal of fashion coordinate system based on color harmony

○Takumi Saito[1], Mizuki Oka[1], Kazuhiko Kato[1] （1. University of Tsukuba, Graduate School of Systems and Information Engineering）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Measuring Beginner Friendliness of Chinese Web Pages Explaining Academic Concepts using HTML Structures

Bingcai HAN[*1] Takahide KASUGA[*1] Hayato SHIOKAWA[*1] Kota KAWAGUCHI[*1]
Takehito UTSURO[*2] Yasuhide KAWADA[*3]

[*1]Grad. Sc. of Systems and Information Engineering, University of Tsukuba
[*2]Faculty of Engineering, Information and Systems, University of Tsukuba
[*3]Logworks Co., Ltd.

Search engine is an important tool of modern academic study, but the results are lack of measurement of beginner friendliness. For improve the efficiency of using search engine for academic study, it is necessary to find a method of measuring the beginner friendliness of Web page explaining academic concepts and build an automatic measurement system. In this research, we formalize the measurement of beginner friendliness by several individual factors, including definition, formula and so on. We collect about 900 Web pages for manual measurement using the individual factors to build a reference dataset. Then we analyze the HTML structure of the collected dataset, and craft some rules for measuring four individual factors. The crafted rules are tested using program and the performance is evaluated by recall and precision. Finally, we test about 400 Web pages and find some special features of HTML structure of Chinese Web pages. The modified rules for the features perform better than general rules we first crafted. The results of this research would be the an important reference for further work for a whole automatic measurement system of beginner friendliness.

1. Introduction

Search engine is a quite important tool for getting concerned knowledge when it comes to the study of academic concepts. However, if we want to select the beginner friendly pages during using search engine, it is necessary to compare pages by manual work. The reason of ineffective manual comparison is that there is no systematization for beginner friendly Web pages in the results of search engine. Therefore, it comes up with us to find a method of measuring the beginner friendliness of Web pages explaining academic concepts automatically, and finally build a whole assisting system for promoting academic study using searching engine, which would improve the efficiency of web learning.

This research presents a method of formalizing the measurement of beginner friendliness by several individual factors. For collecting reference dataset, we analyze the Web pages explaining academic concepts based on HTML structures. And after collecting data, we select four factors to craft rules for measurement using HTML, and then build a Python program based on the crafted rules to verify the feasibility of measurement using HTML.

We use data from Baidu.com and Google.hk, which are two search engines mostly used for academic search in Chinese. The formalization of measurement of beginner friendliness is described with details in Section 2. Section 3 shows the details about reference dataset of Web pages explaining academic concepts. And Section 4 describes the rules used in program of measuring beginner friendliness and the evaluation of the program. Section 5 introduces the related work of this research. Finally, Section 6 concludes this paper.

Contact: Takehito UTSURO, Graduate School of Systems and Information Engineering, University of Tsukuba, 305-8573 Japan, 029-853-5427

2. Factors of Beginner Friendliness of Web Pages Explaining Academic Concepts

The measurement of individual factors and overall measurement are binary decision, and the rules are modified according to the measurement results of same Web pages by 3 persons in our group.

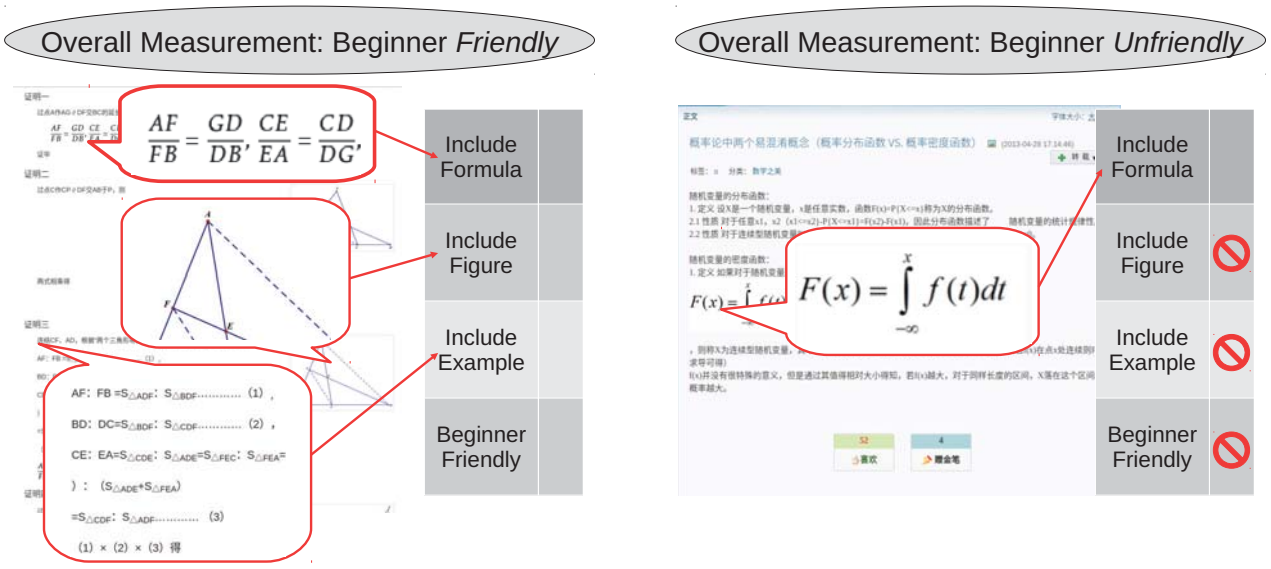
2.1 Individual Factors

We determine several individual factors to formalize the measurement of beginner friendliness of Web pages explaining academic concepts. After prior investigation, we abstract six individual factors including definition, formula, figure, example, beginner friendliness of text and Web page layout. For each factor, there are some basic rules for measurement by manual work.

(a) Definition: measured positive when a Web page contains correct and precise definition of the explained academic concept.

(b) Formula: measured positive when a Web page contains formula whether in text or figures. The formulas should be relevant to the academic concept explained in the Web page.

(c) Figure: measured positive when a Web page contains figures or pictures relevant to the academic



(a) Overall Measured Beginner Friendly

(b) Overall Measured Beginner Unfriendly

Figure 1: Examples of Web pages that explains an academic concept

concept explained in the Web page, except when the figure shows formula only.

(d) Example: measured positive when a Web page contains examples relevant to the academic concept, including examples of application, proof, explanation and so on. And when the examples shown in figures, it would be measured positive for both figure and example.

(e) Beginner friendliness of text: measured positive when the text of Web page is considered beginner friendly by evaluator. The measurement should not take consider in the other factors.

(f) Web page layout: measured positive when the layout of Web page is considered easy to read by evaluator. The measurement should not take consider in the other factors.

2.2 Overall Measurement considering Individual Factors

The overall measurement of the beginner friendliness of Web pages explaining academic concepts is performed by each of our group members. The rules of overall measurement are modified by results of different members. Parts of overall measurement is shown in Figure 1. A Web page explaining academic concept would be measured as beginner friendly when the measurement of each individual factor is considered benefiting the beginner friendliness. The overall measurement is performed by not only the combination of individual factors but also which factor is measured positive. For positive situation, the Web page should contain at least one of Web page layout or beginner friendliness of text measured positive. If a Web page contains only one measured positive of layout or beginner friendliness of text, it would be measured positive for overall measurement when there are enough positive results for the other factors. And if the layout and beginner friendliness of text are both measured positive, the overall measurement would be positive when there is at least one positive measurement of the other factors.

3. Reference Dataset of Web Pages Explaining Academic Concepts

In this section, we collect the reference dataset of Web pages explaining academic concepts based on the rules determined in Section 1. The dataset would be used for building automatic measurement program and test.

3.1 Academic Fields and Concepts for Study

After prior investigation, we determine that the Web pages are collected on three academic fields, including statistics, physics and linear algebra. These filed have same features for measurement of individual factors. For each field, we choose 15 queries for academic concepts based on the teaching content of high school and collage in China, and then we collect the top 10 pages of each query. Table 1 shows the details about the collection.

3.2 Procedure

For building the reference dataset, we choose two search engines as Baidu and Google.hk for collecting Chinese Web pages. For each search engine, we collect the URL data of top 10 Web pages for each query determined before. Then we measure the individual factors of every Web page according to the rules determined in Section 2. The inaccessible and unrelated Web pages would be ruled out. The manual measurement results are sorted by academic fields and search engines for

Table 1: The Fields and Concepts of Reference Data

Fields	Queries
统计/統計/Statistics	惠斯通电桥/ホイートストンブリッジ/Wheatstone Bridge，离心力/遠心力/Centrifugal force，声波/音波/Sound wave，惯性定律/慣性の法則/Law of inertia，交流电/交流/AC，正电荷/正電荷/Positive charge，速度/速度/Speed，张力/張力/Tension，电力线/電気力線/Electric lines of force，电波/電波/Radio waves，电流/電流/Current，恢复系数/反発係数/Coefficient of restitution，变压器/变圧器/Transformer，万有引力/万有引力/Gravitation，互感/相互誘導/Mutual induction
物理/物理/Physics	回归分析/回帰分析/Round analysis，概率/確率/Probability，概率密度函数/確率密度関数/Probability density function，伽玛分布/ガンマ分布/Gamma distribution，主成分分析/主成分分析/Principal component analysis，自回归/自己回帰/Autoregression，置信区间/信頼区間/Confidence interval，相关系数/相関係数/Correlation coefficient，独立成分分析/独立成分分析/Independent component analysis，标准差/標準偏差/Standard deviation，方差/分散/Dispersion,正态分布/正規分布/Normal distribution，泊松分布/ポアソン分布/Poisson distribution，零假设/帰無仮説/Null hypothesis，协方差/共分散/Covariance
线性代数/線形代数/Linear algebra	阶数/階数/Order，行列式/行列式/Determinants，对角化/対角化/Diagonalization，梅涅劳斯定理/メネラウスの定理/Menelaus's theorem，正交矩阵/直交行列/Orthogonal matrix，共轭梯度/共役勾配/Conjugate gradient，范数/ノルム/Norm，克罗内克函数/クロネッカーのデルタ/Kronecker function，克莱姆法则/クラメルの公式/Clems law，三角矩阵/三角行列/Triangular matrix，标准正交基/正規直交基底/Standard orthonormal basis，特征多项式/特性多項式/Characteristic polynomial，雅可比矩阵/ヤコビ行列/Jacobian matrix，二次型/二次形式/Secondary type，线性变换/線形変換/Linear transformation

efficient analysis of HTML structures. For analysis of HTML structures, we collect HTML contents of the reference dataset from the recorded URL using program. The analysis results are used for crafting rules in next section.

3.3 Reference Dataset

The final reference dataset of measured Web pages explaining academic concepts contains about 900 Web pages. During the measurement, some Web pages showing only academic papers or books are considered as unmeasurable and ruled out from final dataset. The results of manual measurement are shown in Table 2. We use the dataset and the results of analysis for crafting rules of automatic measurement program and testing, which are described in next section.

4. Measuring Individual Factors by Manually Crafted Rules

According to the analysis results, we craft rules for four individual factors, including figure, formula, example and beginner friendliness of text. The rules are modified referring to the features of HTML structures appeared in Chinese Web pages.

4.1 Manually Crafted Rules

The rules are all crafted based on the features of HTML structures. During the analysis of HTML structures, we find that some features of HTML of Chinese Web pages are totally different with other Web pages, mainly from Baidu.com and zhihu.com, which make up almost half of the collected Web pages. Referring to the HTML features of Chinese Web pages, the crafted rules of the selected four factors are shown as blow.

(a) Formula: measured positive when a Web page includes tag “math” in HTML, mathematical symbols and numbers in any attribute value in HTML. It is also measured positive when an equal symbol(=) is included in the text of HTML.

(b) Figure: measured positive when a Web page contains typical and specific tags of picture in HTML, like “pic”, “figure” and so on. And the picture should also be larger than 100×100 (in pixel) without link embedded in it.

(c) Example: measured positive when a Web page contains specific keywords or word patterns of example in HTML, like “例”, “Example” and so on.

(d) Beginner friendliness of text: the measurement of this factor is performed by a combination of several rules of HTML, including pattern of attribute values, size of “body” component and text composition in HTML. In general, it is measured positive when a Web page is in a suitable size with simple text in HTML.

4.2 Evaluation Procedure

The automatic measurement program is programmed with Python. For test data, we select 300 Web pages, 100 Web pages for each field with 5 queries averagely. The manual measurement is the reference for the evaluation of program. We record the measurement results of the measurement program and compare it with the manual results, and then calculate the recall and precision of the results to evaluate the performance of our crafted rules based on HTML structures.

4.3 Evaluation Results

The evaluation results are shown in Figure 2. From the figure, it is seen that the measurement program achieves an acceptable performance, especially for beginner friendliness of text. The results show that mea-

Table 2: Details of Manual Measurement Results

Fields	Number of Queries	Total Pages	Measured Pages	Include Definition	Include Figure	Include Formula	Include Example	Web Page Layout	Beginnner Friendly Text	Positive Overall Measurement
统计/ 統計/ Statistics	15	300	287	64%	66%	78%	50%	76%	80%	61%
物理/ 物理/ Physics	15	300	263	56%	50%	74%	42%	80%	88%	63%
线性代数/ 線形代数/ Linear algebra	15	300	279	58%	56%	50%	30%	75%	68%	74%
Summation	45	900	829	59%	57%	67%	41%	77%	79%	66%

suring the four factors using HTML structure is feasible and the method could be applied and tested for the other two factors.

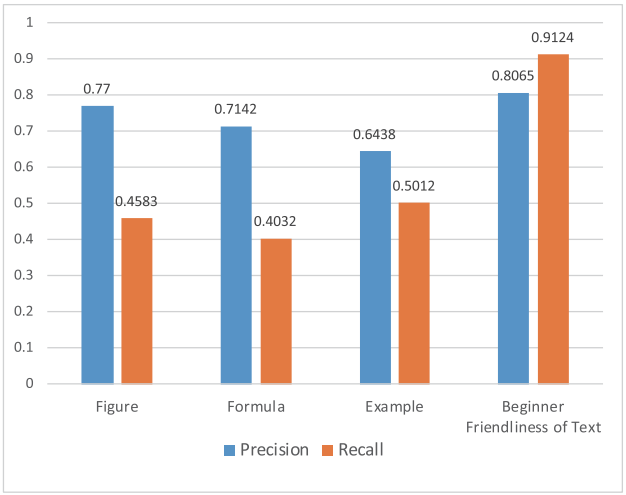


Figure 2: Evaluation Results of Automatic Measurement of Beginner Friendliness

5. Related Work

As a concerned field with this research, the community Question-Answering selection is focused by some researchers [石川 11, Sakai11]. These researches present methods of evaluating and selecting good answers in Question-Answering system, which has much same with the measurement of beginner friendliness discussed in our research. For Japanese Web pages explaining academic concepts, similar methods for the measurement using HTML are discussed in the studies by other members in our group, including using SVM to perform the measurement based on HTML structures [春日 18] and image recognition using deep learning for measuring the Web page layout [塩川 18].

6. Conclusion

This research presents a method of formalizing the measurement of beginner friendliness of Web pages by individual factors, and evaluates the method of measuring four individual factors using HTML structures. The evaluation results of the measurement program show that the crafted rules perform quite well. And it is proved that HTML structure is applicable for measuring the individual factors of beginner friendliness of Web pages explaining academic concepts. The method presented in this research would be an inspiration of finding a method for measuring the beginner friendliness of Web pages automatically, and it would also be an important reference for further work of building a complete assisting system for academic study using search engine.

References

[石川 11] 石川大介, 酒井哲也, 関洋平, 栗山和子, 神門典子：コミュニティQA における良質回答の自動予測, 情報知識学会誌, Vol. 21, No. 3, pp. 362–382 (2011).

[春日 18] 春日孝秀, 塩川隼人, 韓炳材, 宇津呂武仁, 河田容英：HTML 構造上の特徴を利用した学術用語解説ウェブページの分かり易さの自動評定, 第 10 回 DEIM フォーラム論文集 (2018).

[Sakai11] Sakai, T., Ishikawa, D., Kando, N., Seki, Y., Kuriyama, K. and Lin, C.-Y.: Using Graded-Relevance Metrics for Evaluating Community QA Answer Selection, *Proc. 4th WSDM*, pp. 187–196 (2011).

[塩川 18] 塩川隼人, 春日孝秀, 韓炳材, 宇津呂武仁, 河田容英：深層学習を用いた学術用語解説ウェブページの見易さの自動評定, 第 10 回 DEIM フォーラム論文集 (2018).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

語彙の可視化を目的としたライフログ解析による単語の理解有無の推定

Estimation of understanding of words by lifelog analysis for vocabulary visualizing

野澤 一真[*1] 望月 典樹[*1] 増山 岳人[*2] 中村 壮亮[*3]
Kazuma Nozawa Noriki Mochizuki Gakuto Masuyama Sousuke Nakamura

[*1] 法政大学大学院理工学研究科電気電子工学専攻
Department of Science and Engineering, Hosei University Graduate School

[*2] 名城大学理工学部電気電子工学科
Faculty of Science and Technology, Meijo University

[*3] 法政大学理工学部電気電子工学科
Faculty of Science and Technology, Hosei University

In this study, we made an estimate understanding of words for visualization of individual vocabulary. We have hypothesized that each of words which a person understand and does not are concentrating on certain word concept. Therefore, we proposed a method for estimating understanding of words with hierarchical neural network and word2vec, and succeeded in estimating the understanding of words by accuracy of over about 96%. From these results, it has been shown the validity of the hypothesis and that it is possible to estimate the understanding of word by life log analysis.

1. はじめに

近年，機械学習，IoT 等の技術が進み，社会的課題を解決するためにビックデータを利活用するデータ主導社会の実現が図られている[総務省 2017]. また，2017 年 5 月に施行された改正個人情報保護法により，個人情報利用の障壁が下がった事を受けて，個人に関するデータを記録したライフログを利活用するビジネスや研究への期待が高まっている. ライフログを活用した研究の例として，購入履歴の利用による消費者の嗜好・ニーズに合致した商品やサービスの提供[大井 2015], SNS での投稿履歴の利用による生活パターンの推定[潘 2013], 生体情報の利用によるストレス軽減[前田 2016]といったものが挙げられる.

このような社会的情勢を受けて，我々は個人の語彙（使いこなせている単語）に基づいた知能補助サービスの検討を始めている. 具体的には発話ログや文書データ等のライフログを解析して個人の語彙を可視化し，それをサービスへと応用する展開を考えている.

知能補助サービスの例を挙げる. [片山 2013]で提案されているような，検索システムにおいて個人の知らない知識に対して注釈をつけるといった情報提示サービスが考えられる. その他にも，聞き手の語彙に応じたロボットの発話生成，語彙能力劣化診断による認知症の早期認知，特定個人と同じ語彙能力を持ったアンドロイドの実現などが挙げられる.

これらのサービスを確立するうえで，語彙の可視化が必要不可欠となる. そこで、本研究の目的を，個人の語彙の可視化に向けた「単語の理解有無の推定機能の開発」とする.

本研究における「理解している単語」とはその単語の意味を一つ以上説明できることとする. 単語を使いこなすことはその意味を理解することから始まる. 一部には理解しているが使いこなせない単語があるがその割合は小さいと考えられるため，理解有無を推定すれば語彙の可視化につながると考えられる. 従って，個人の語彙を把握するために，本研究では単語の理解有無の情報を利用することとした.

我々は，興味，嗜好および生活環境に従って，理解している単語や理解していない単語がそれぞれ特定の単語概念に集中しているとの仮説を立てている。この仮説が正しければ，「理解している単語」と「理解していない単語」の特徴を学習器に学習させることで理解有無の推定が可能であると考えられる.

2. 関連研究

文書データや生体情報から単語の理解有無の測定や推定が行われている. [四方 2014]は認知症患者によるブログから語彙が時間経過とともにどのような推移をたどるのかを調査している. しかしながら，[四方 2014]は語彙の「測定」であり，ブログに出現しない語彙を扱う「推定」までには至っていない. 本研究では単語の「測定」と「推定」を区別している. これはブログ等の記録媒体では使用単語の偏りもあるなど，全て理解している単語が出現するわけではないため，「測定」した単語の理解有無に基づき全単語に対する理解有無を「推定」することが肝要となるためである.

一方で，理解有無の推定に関する研究も行われている. 一般的な単語の推定[天野 1996]の研究が多いが[片山 2013]では専門用語の単語の知っているか否かの推定を行なっている. しかしながら，推定している単語が 1 つの分野の用語約 400 語に限定して評価を行なっているため，他の専門分野の単語や日常会話レベルの単語の推定も必要である.

また，[江原 2010]は英文 web ページに対して，知らない英単語を個人の英語力に応じて訳を予め付与する読解支援システムを提案している. 個人がクリックした単語をライフログとして記録し，このライフログを解析することによって，知らない英単語を項目応答理論を用いて推定している. 推定の際には基本的な語彙 12000 語に対し，英語母語話者が人手で 12 段階の難易度をつけた語彙リストを使用している. 難易度リストを用いた推定は人の主観が入ることや人的コストがかかる.

連絡先：野澤一真，
法政大学大学院理工学研究科電気電子工学専攻
〒184-0002 東京都小金井市梶野町 3-7-2
090-6477-3817, kazuma.nozawa.0703@gmail.com

3. 提案手法

本研究では関連研究を踏まえて，難易度リストを使用せず，個人の「理解している単語」と「理解していない単語」で学習させることで理解有無を推定する手法を提案する．word2vec で表現された単語ベクトルをニューラルネットワークに入力し，被験者にヒアリングで収集した理解有無を出力させるように学習させた単語理解度識別モデルを使用することで単語の理解有無の推定を実現する．提案手法を図 1 に示す．

3.1 word2vec を用いた単語の低次元表現

機械学習でテキストを扱うためには，単語を単語そのものではなく，何らかの数値に置き換える必要がある．設計が容易な単語のベクトル表現に One-Hot 表現があるが，ベクトルの各要素に単語が対応するため，単語数と等しい膨大な次元数となるため適切ではない．そこで，word2vec[Mikolov 2013], [Rong 2016]と呼ばれる Mikolov 氏が提案した単語の分散表現を利用する．

word2vec とは単語の分散表現(以下，単語ベクトルと呼ぶ)を獲得する手法である．単語のベクトル表現が可能であるため，類似語はベクトル空間でグループ化され，意味特徴を捉えたものになる．

単語 *word* の単語ベクトル v は以下の式で求められる．

$$v = word2vec(word) \quad (1)$$

ただし，$word2vec(word)$は word2vec で表現される *word* の分散表現である．

また，word2vec は大量のテキストデータを学習させる必要があるため，本研究では wikipedia の記事を使用し，各単語をベクトル化をする word2vec のモデルを生成した．

3.2 単語理解度識別モデルの生成

単語の理解度の推定には入力層，隠れ層及び出力層からなる階層型ニューラルネットワーク(図 1)を使用した．

ニューラルネットワークは，word2vec で表現した単語ベクトル(x_1, x_2, ⋯, x_{200})を入力として，理解度 y_1 及び非理解度 y_2を出力とした．そして，被験者に対して理解の有無をヒアリングした単語に対して，その word2vec による単語ベクトルを入力とし，出力を t_1=1, t_2=0(理解している場合)または t_1=0, t_2=1(理解していない場合)としたものを教師データとした．教師データとしては被験者の「理解している単語」と「理解していない単語」の両方をバランス良く用意することで「理解している単語の特徴」と「理解していない単語の特徴」を識別するモデルが生成できると期待される．

これは，理解している単語の集合と理解していない単語の集合は word2vec で表されるベクトル空間上で複雑に交錯していると予想されるがニューラルネットワークは複雑な非線形分離を可能にすることによる．

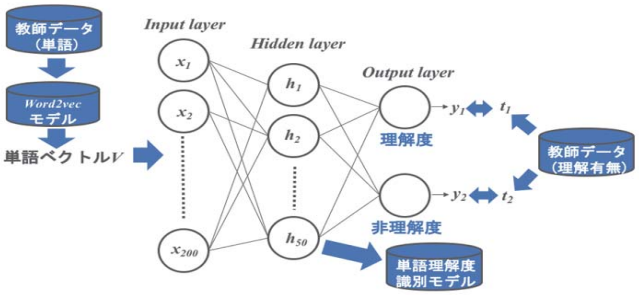


図 1 ニューラルネットワークによる理解有無の推定

4. 評価実験

4.1 データの収集方法

3 人の被験者(表 1)に単語リストから無作為に単語を提示し，理解している/いないのヒアリングを行うことで，教師データを収集した．実用化の際は音声認識等を用いて，自動で収集することが望ましいが，音声をテキストに変換する際にノイズ等が混入することと，本研究の本質は「理解している単語」と「理解していない単語」が推定可能であるかを確かめることであることからヒアリングを行うこととした．

提示する単語リストは wikipedia の記事から抽出した．公開されている wikipedia には学習には必要のない XML のコンテナ部分や独特な記号を含むためそれらを削除し，形態素解析を用いて文章を形態素(言語で意味をもつ最小単位)に分割し，名詞のみ抽出した．そこから，wikipedia 内で頻度が 6 以上かつ文字数が 10 文字以下のみを抽出して，単語リストを生成した．

4.2 word2vec の学習条件

word2vec の学習には自然言語処理ライブラリ gensim [Radim Rehurek 2018]を使用した．学習させるモデルは CBOW を用い，word2vec の出力を 200 次元とした．ウィンドウサイズは 15，頻度が 5 未満の単語は学習に使用しなかった．他の設定は gensim のデフォルト値を使用した．

4.3 ニューラルネットワークの学習条件

ニューラルネットワークの構築には tensorflow をバックエンドとした keras を使用した．word2vec で表した単語ベクトル 200 次元をニューラルネットワークへの入力とした．隠れ層のノードは 50 個とし，活性化関数は ReLU 関数とした．また，出力層のノードは 2 個(理解している/していない)とし，活性化関数は softmax 関数とした．

損失関数 E は推測値と目標値(教師データ)の二乗和誤差とした．損失関数 E を(2)式に示す．

$$E = \frac{1}{2N}\sum_{n=1}^{N}\sum_{k=1}^{2}(y_k^n - t_k^n)^2 \quad (2)$$

ただし，N は学習に使用する教師データの数とする．

ニューラルネットワークのパラメータ w は誤差逆伝播法で最適化し，勾配更新には SGD と Adam[Kingma 2014]を用い，それぞれの場合で精度を比較した．SGD を(3)式に示す．

$$w^{t+1} \leftarrow w^t - \eta\nabla E(w^t) \quad (3)$$

ただし，学習率 η は 0.01 とした．

次に Adam を(4)式から(8)式に示す．

$$m_{t+1} = \beta_1 m_t + (1-\beta_1)\nabla E(w^t) \quad (4)$$

$$v_{t+1} = \beta_2 v_t + (1-\beta_2)\nabla E(w^t)^2 \quad (5)$$

$$\hat{m} = \frac{m_{t+1}}{1-\beta_1^t} \quad (6)$$

$$\hat{v} = \frac{v_{t+1}}{1-\beta_2^t} \quad (7)$$

$$w^{t+1} \leftarrow w^t - \alpha\frac{\hat{m}}{\sqrt{\hat{v}}+\epsilon} \quad (8)$$

ただし，$m_0 = 0$，$v_0 = 0$，$\alpha = 0.001$，$\beta_1 = 0.9$，$\beta_2 = 0.999$，$\epsilon = 10^{-8}$とした．

また，重みの初期値は ReLU 関数と相性が良い「He の初期値」[He 2015]を用いた．

ニューラルネットワークの学習条件を表 2 に纏める．

表 1 被験者の詳細

被験者名	性別	年齢	大学での専門
A	男	21	教育学
B	男	22	電気電子工学
C	男	20	デザイン工学

表 2 ニューラルネットワークの学習条件

入力層	次元	200
	活性化関数	線形写像
隠れ層	次元	50
	活性化関数	ReLU関数
出力層	次元	2
	活性化関数	softmax関数
バッチサイズ		50
epoch		50
損失関数		二乗和誤差
重みの初期値		Heの初期値

4.4 評価方法

未知データに対する認識率を評価するために交差確認法を用いた. 交差確認法の手順を以下に示す.

1. 全単語をグループ X_1, X_2, X_3, X_4に分割する. なお, 分割したグループには理解している単語と理解していない単語の数を同じにする.

2. X_iを除いた 3 グループで学習し, X_iを用いて正解率を算出する. このとき, X_iを検証データとし, それ以外を教師データとする.

3. 各試行の正解率 C_iを算出する. 正解率 C_iは[片山 2013]に倣って, (9)式を用いた. 被験者が理解していると答えた単語の集合を T, 被験者が理解していないと答えた単語の集合を F, モデルが理解していると推定した単語の集合を P, モデルが理解していないと推定した単語の集合を N とすると正解率 C_i は以下の式で表される.

$$C_i = \frac{|P \cap T| + |N \cap F|}{|P| + |N|} \qquad (9)$$

4. この手順を全ての $i(1<i<4)$について行い, 4 回の正解率の平均をモデルの正解率 C とする.

5. 結果

エポック数が 50 の時の各試行における正解率と全試行の平均を表 3 に示す. 表 3 の結果より, 最適化手法が Adam のときにはどの被験者においても正解率は 96[%]を超えた. テストデータは 200 個であったので, およそ 190 個は正解したこととなる. 一方で SGD の時の正解率はおよそ 70[%]に留まった. これらのことから重みを更新する際, Adam を適用した方が正解率が高いことが示された.

図 2 にエポック数に対する正解率をプロットしたグラフを示す. どの被験者においても学習を進めるにつれて正解率が増加していることがわかる. Adam の場合はエポック数が 5 回目までは急激に精度が向上し, それ以降は緩やかに増加している. 一方で, SGD の場合は初めの正解率が低く, 学習を重ねても Adam に比べて大きな正解率の向上は見られなかった. 一般的に SGD より Adam の方が素早く学習が可能であり, 精度が高いことは[Kingma 2014]で示されており, 今回の実験でも同様な結果となった.

また, 被験者ごとの正解率の差は Adam の場合においては最大 1[%]と, 小さかった. 理解している単語は個人ごとに違うにも関わらず, 同じ手法で理解有無の推定が可能であることが確認できた.

6. 結論

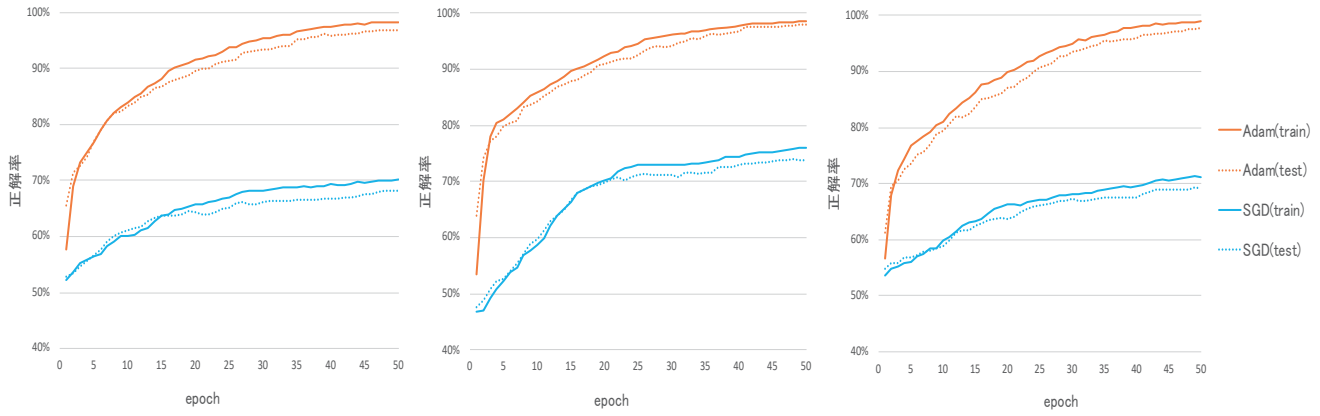
6.1 まとめ

本研究では, 個人の語彙の可視化に向けて, 単語の理解有無の推定を行った. 実験結果より, 理解している単語と理解していない単語を約 96[%]超える正解率で識別することに成功した.

成功した理由として,「理解している単語の集合」と「理解していない単語の集合」は word2vec で表されるベクトル空間上で偏りがあったことによると考えられる. これは理解している単語や理解していない単語がそれぞれ特定の単語概念に集中している

表 3 実験結果

被験者	A		B		C	
最適化手法	Adam	SGD	Adam	SGD	Adam	SGD
1回目	95.5%	65.5%	98.0%	70.5%	99.0%	66.0%
2回目	97.5%	72.5%	98.0%	76.0%	96.5%	73.5%
3回目	97.5%	65.5%	98.5%	74.5%	97.0%	66.5%
4回目	97.0%	69.5%	97.0%	74.5%	98.5%	70.5%
平均	96.9%	68.3%	97.9%	73.9%	97.8%	69.1%



(a) 被験者 A　(b) 被験者 B　(c) 被験者 C

図 2 エポック数に対する正解率の変化

という仮説の妥当性を示すものでもあり，ライフログを解析して個人の単語の理解有無の推定が可能であることが示唆された．

6.2 今後の展望

本研究は会話ログやチャットといったライフログから個人の語彙を推定することが最終目標である．そのため，ヒアリングを行うのではなく，SNS（Twitter，Facebook 等）やチャットツール(LINE 等)や会話ログといったライフログをソースとして，本手法を適用する必要がある．

また，word2vec は他言語での学習も可能であるため，英語等の言語の理解の推定が可能かについても検証する必要がある．日本人の英語学習者の支援や言語に依存しない語彙推定が可能となる可能性がある．

最後に，語彙推定が確立した暁には，知能補助サービスに取り入れ，社会実装を図っていきたい．

参考文献

[総務省 2017] 総務省: 平成 29 年度版 情報通信白書 データ主導経済と社会変革.

[大井 2015] 大井 貴裕，三川 健太，後藤 正幸: 評価と購買の両履歴データの学習による確率的潜在クラスモデルの推定精度向上に関する一考察，日本経営工学会論文誌，65(4),pp.286-293, 2015-03-10.

[潘 2013] 潘睿，松尾豊: 個人の生活管理のためのソーシャルデータからの行動パターンの発見,人工知能学会誌，28(6), pp.829-834, 2013-11-01.

[前田 2016] 前田直樹，平部裕子，荒川豊，安本慶一: オフィスワーカのための自然なストレスモニタリングシステムの提案と設計,情報処理学会論文誌，2016, pp.1788-1793, 2016-07-06 .

[片山 2013] 片山太一，小林のぞみ，牧野俊朗，松尾義博: トピック情報を利用したユーザの知識推定，人工知能学会，27，pp.1-4，2013.

[四方 2014] 四方朱子，荒牧英治: 言語能力検査としての言語処理:長期間のブログ執筆を継続した認知症の 1 例，言語処理学会，pp.1126-1129，2014-3.

[天野 1996] 天野成昭: 頭の中に単語はいくつある?情報処理学会，00037(00004), pp.351-354，1996-04-15.

[江原 2010] 江原遥，二宮崇，清水伸幸，中川裕志: ユーザが知らない語を予測する読解支援システム SocialDict とそのリーダビリティ測定への拡張，情報処理学会，1，pp.1-7，2010-5-20 .

[Mikolov 2013] T. Mikolov, W. Yih, G. Zweig: Linguistic Regularities in Continuous Space Word Representations, NAACL-HLT-2013, pp.746–751, May 27, 2013.

[Rong 2016] Xin Rong: word2vec Parameter Learning Explained, arXiv:1411.2738v4 ,cs.CL,Jun. 2016.

[Radim Rehurek 2018] Radim Rehurek: genism topic modelling for humans https://radimrehurek.com/gensim/ .

[Kingma 2014] Diederik P. Kingma, Jimmy Ba: Adam: A Method for Stochastic Optimization, 22 Dec 2014.

[He 2015] Kaiming He，Xiangyu Zhang，Shaoqing Ren，Jian Sun: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, arXiv:1502.01852v1, cs.CV, 6 Feb 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

過去の週報の再利用のための活動記録に基づく週報管理システムの実装

Implementing a Weekly Report Management System Based on Activity Records for Reuse of Past Reports

安藤力哉 *1　大囿忠親 *1　新谷虎松 *1
Rikiya Ando　Tadachika Ozono　Toramatsu Shintani

*1名古屋工業大学大学院工学研究科情報工学専攻
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

We have accumulated weekly reports for more than 20 years and we are trying to utilize the reports for students. The aim of this study is to realize a system for weekly report management and research activity support that allows students to view past weekly reports and research activities. This paper shows an implementation of the management system that has three functions: (1) searching reports by research papers, (2) visualizing research activities, and (3) editing annual weekly reports. The system helps users to overview the activities of past students to make their plans.

1. はじめに

週報等の報告書が蓄積される一方，これらの有効活用のための技術開発が不十分であり，再利用が容易ではないという問題がある．例えば，本研究では週報管理システムおよび研究活動支援システム [1] により，学生は，過去の週報や研究活動を閲覧可能である．これらのシステムを用いて，現在の学生が過去の学生の週報を参考にする（週報の再利用と呼ぶ）ことは有益であると考えるが，そのための支援が不十分である．本研究では，週報の再利用における煩雑さを軽減するために，週報管理システムを拡張した．具体的には，1）研究業績からの週報検索機能，2）研究活動の可視化機能，3）年間週報編集・編纂機能，の3点を実現することで，週報の再利用を支援することを目指した．

本稿では，週報管理システムの拡張に関連して，上記3点の実現方法について述べる．

2. 活動記録を用いた週報管理システム

本研究室では，独自に開発した週報管理システムを用いて，週報を蓄積している．本研究室における週報とは，その週の活動内容やそれに伴う成果，課題等に関する報告書である．週報には研究活動における情報や，それに伴う論文発表，及びその発表に向けての準備期間における各個人の研究活動の情報などが記載されている．そのような週報に含まれる様々な他者の情報は，自身が研究活動を行う際に，非常に有益な情報である．そこで本研究では，研究活動や参加学会の情報などから，その情報を含む週報を利用することで，週報の再利用を行う．

本研究では，図1（3章で後述）に示されるように，開発済みの他の研究室内システム（論文管理システム，打刻システム）における情報の利用，および週報からの研究活動の概観を可能にすることで，週報の再利用を支援することを目指している．具体的には，週報の再利用における煩雑さの軽減を目的として，1）研究業績からの週報検索機能，2）研究活動の可視化機能，および3）年間週報編集・編纂機能，の3点を実現する．1）研究業績からの週報検索機能とは，研究業績に関連する週報の閲覧を支援するために，研究室内の発表済み論文が登録された論文管理システムから，各論文に関連する週報を閲覧するための機能である．次に，2）研究活動の可視化機能とは，打刻システムに記録された，研究室への入退室時間を週報とあわせて提示することで，忙しさ等を可視化する機能である．最後に，3）年間週報編集・編纂機能とは，週報等の情報から年間の研究活動を概観するための機能である．以降，これらの機能について説明する．

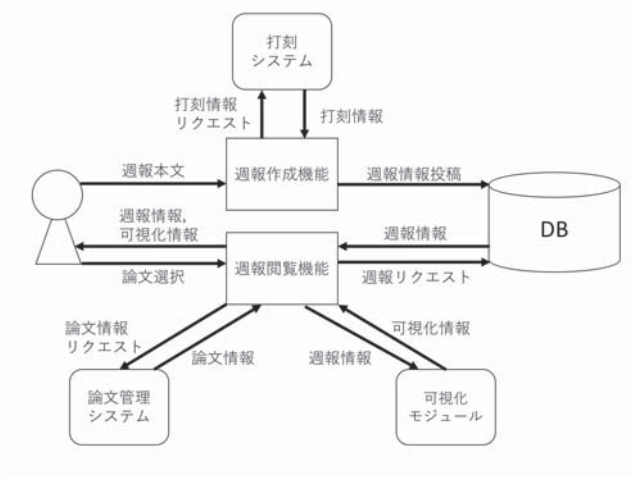


図 1: システム構成図

2.1 研究業績からの週報検索機能

研究業績からの週報検索機能とは，研究業績に関連する週報の閲覧を支援するために，研究室内の発表済み論文が登録された論文管理システムから，各論文に関連する週報を閲覧するための機能である．研究活動を行うにあたり，研究の活動履歴は論文には含まれていない．そのため，論文では得られない研究活動中の情報を取得するために，関連する週報の閲覧をするための機能が必要である．実現のためには，論文に関連する情報を特定することが必要である．使用例として，参考にしたい論文について検索を行い，その論文発表時期の週報を閲覧することにより，発表に向けて，どのような準備や研究を行っていたかを閲覧することができる．

連絡先: 安藤力哉，名古屋工業大学工学研究科情報工学専攻，愛知県名古屋市昭和区御器所町，052(735)5584，rando@toralab.org

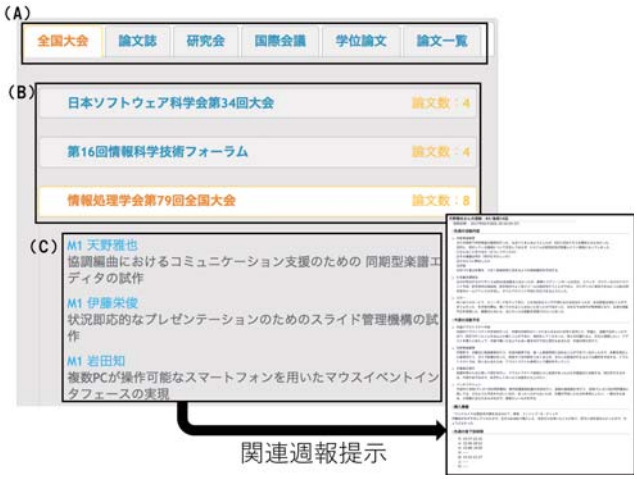


図 2: 実行例：週報検索機能

2.2 研究活動の可視化機能

研究活動の可視化機能とは，打刻システムに記録された，研究室への入退室時間を週報とあわせて提示することで，忙しさ等を可視化する機能である．また，打刻システムに記録された情報から，週の在室時間を取得し，週報情報とあわせてグラフ化を行う機能である．研究活動において，適切なスケジューリングを行うことは非常に大切である．自身の研究活動に費やした時間を把握することや，他者の論文執筆時期等の研究活動に費やした時間を閲覧することは，スケジューリングを行う際に非常に有益である．使用例として，研究活動を行う際に，研究活動の進捗度合いと打刻情報との比較を行い，今後，どの程度の作業時間が必要であると予想されるかを把握できる．

2.3 年間週報編集・編纂機能

年間週報編集・編纂機能とは，週報等の情報から年間の研究活動を概観するための機能である．蓄積された週報のテキスト量，週報回数は非常に多くなり，1 つずつの内容を閲覧機能により確認することは，効率的でない．ゆえに，研究活動履歴の保存・閲覧のために，年間の週報全体を概観するために，編集・編纂する必要がある．使用例として，一年間の週報データを 1 つのファイルとして取得し，保存・閲覧することができる．

3. 実装

図 1 を用いて，システムの構成を示す．ユーザは，週報作成機能において，週報本文を入力し，週報を投稿する．週報の投稿を行う際に，打刻システムに対して，打刻情報をリクエストする．これにより，取得した打刻情報を記載した週報をデータベースに投稿する．週報閲覧機能において，論文管理システムに対して論文情報のリクエストを行い，取得した論文情報からユーザは論文選択を行う．選択した論文について，その論文の著者情報および公開日を利用して，データベースに対して，論文と関連のある週報をリクエストし取得する．また，ユーザに対して取得した週報情報を提示する．また可視化情報については，データベースより，ユーザの年間の週報情報を取得し，可視化モジュールに送ることで，週報情報および時刻情報が可視化され，ユーザは可視化情報として取得できる．また，各ユーザの年間の全週報は週報閲覧機能によって，週報情報が編集・編纂され，年間週報情報として一括で取得できる．

研究業績からの週報検索機能は，論文管理システムとの連携により，論文に関連する週報を提示する．ユーザが選択した研究業績 (論文，学会発表等) に関連する週報を検索するために，著者名および公開日を利用する．ここでの問題は，公開日から関連する週報の期間を推定することである．ここでは，ヒューリスティックを用いて特定の業績に関連する週報の期間を推定することとした．取得した週報をユーザに提示することで，研究業績に関連する週報の閲覧を支援する．これにより，論文に関連した研究活動の記録の閲覧を支援することができる．

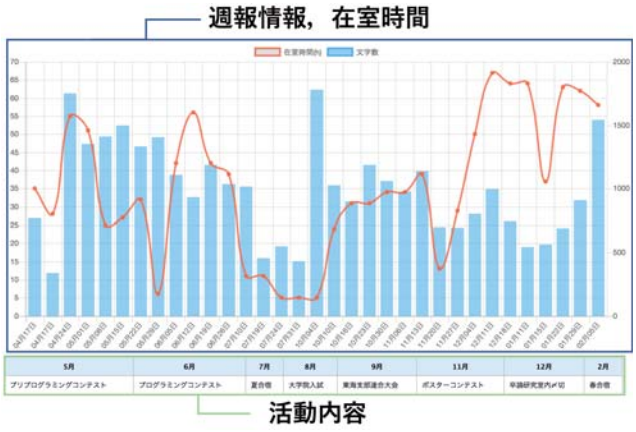


図 3: 研究活動の可視化機能

図 2 では，研究業績からの週報検索機能の実行例を示す．(A) では，学会の分類別にわかれており，学会分類を選択することで，(B) に示されるような，学会名ごとに表示される．(B) では，学会名を選択することによって，(C) のように，著者名，論文名ごとに表示される．また，(C) において，著者名から，関連する週報へリンクされている．

研究活動の可視化機能は，週報作成時に，打刻システムとの連携により，週の打刻情報を週報本文とともに記載する．週の打刻情報には，入退室時間のみでなく，外出時，休憩時の時刻情報も含まれているため，具体的な作業時間の把握ができる．また，図 3 に示すように，週報本文の分量などの週報情報および，各週の打刻情報から，週の在室時間や作業時間を計算した，週ごとの在室時間の遷移をグラフを用いて視覚的に示す．また，研究室内での各月のイベント情報やユーザの学会参加状況などの活動内容をあわせて提示することによって，各時期の活動も同時に示す．これにより，自身の研究活動について振り返ることができるとともに，教員による学生の研究活動状況の把握を支援する．

4. おわりに

本稿では，週報の再利用の支援に関連して，1）研究業績からの週報検索機能，2）研究活動の可視化機能，さらに 3）年間週報編集・編纂機能，の 3 点の開発について述べた．本システムを用いることで，業績からの週報検索が可能になり，適切な週報を閲覧するための手間の軽減が可能になる．また，週報の概観が可能になり，研究活動を振り返ることや，今後の方針を立てる際に役立つことが期待される．

参考文献

[1] T. Iwasa, Y. Kato, S. Shiramatsu, T. Ozono, T. Shintani.:*Linked Data-based Slide Repository: The Episodic Slide Retrieval Using the Episodic Keyword Networks*, Journal of Control Science and Engineering（2016).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

文章の要旨把握における部分的退色表示の有効性

Effectiveness of Partial Fading of Text on Reading Comprehension

小林 潤平 *1*2　　川嶋 稔夫 *2
Jumpei Kobayashi　　Toshio Kawashima

*1大日本印刷株式会社　　*2公立はこだて未来大学
Dai Nippon Printing Co., Ltd.　　Future University Hakodate

We propose a new text format to direct attention to the head sentence of paragraphs by varying the darkness of displayed sentences by gradually fading characters. Recognition rates of head sentences were increased by applying the proposed format. The overall comprehension performance was also enhanced by applying the proposed format. These results suggest that the proposed format is an effective technique to increase reading comprehension.

1. はじめに

多くの情報が文字で伝達される現在，もし電子リーダーの工夫によって文章の理解度を向上させることができれば，その効果の大きさは計り知れない．

先行研究において，文章読解の向上に関する様々な提案および効果の有無が検証されてきた．例えば，キーワードやテキストの強調が強調内容の記憶成績を高めること [関 97, 魚崎 03]，読み難い書体が記憶の向上につながること [Diemand-Yauman 11]，また，読解に関しては，段落の有無が文章要旨の把握に影響をあたえること [関 96]，文字種や固有名詞に応じた彩色が内容把握の向上につながる結果等が報告されている [内田 97]．しかし，難関試験レベルの正確な文章読解において向上効果をもたらす仕組みについては，検討の余地が残されていた．

そこで本研究では，文章の段落単位やセンテンス順序の顕在化によって文章読解を支援する新たな表示方式を提案するとともに，提案方式が文章の要旨把握にあたえる影響について詳しく検証した．

2. 提案方式

一般的に，文章を執筆する際には，文意をつかみやすくするために，段落の区切りやセンテンスの順序を決めていく．そこで本研究では，段落のまとまりとセンテンスの順序に着目し，自然なかたちでの顕在化を図った．

本研究で提案する「段落先頭から順にセンテンス単位で文字を退色させていく表示方式」を図 1 に示す．段落先頭センテンスの文字は濃く，段落末尾センテンスの文字は薄く表示することで，段落単位やセンテンス順序を顕在化し，文意をつかみながらメリハリをつけて読める効果を狙った．

まず，文章を段落単位で分割し，さらにセンテンス単位で分割する．次に，分割されたセンテンスごとに退色度を設定する．図 1 に示すように，段落先頭のセンテンスの退色度を 0 とし，後に続くセンテンスの退色度をひとつずつ増やしていく．退色度は 6 段階であり，退色度が増えると文字色が薄くなるように設定した．文字色の設定は，退色度 0 を#000000（黒），退色度 1 を#545454，退色度 2 を#888888，退色度 3 を#B0B0B0，退色度 4 を#CFCFCF，退色度 5 を#E0E0E0 とした．退色度 5 のセンテンスの後にも文が続く場合は，それらセンテンスの退色度をすべて 5 とした．なお，最も薄い退色度 5 の文字色でも，支障なく読むことが可能である．

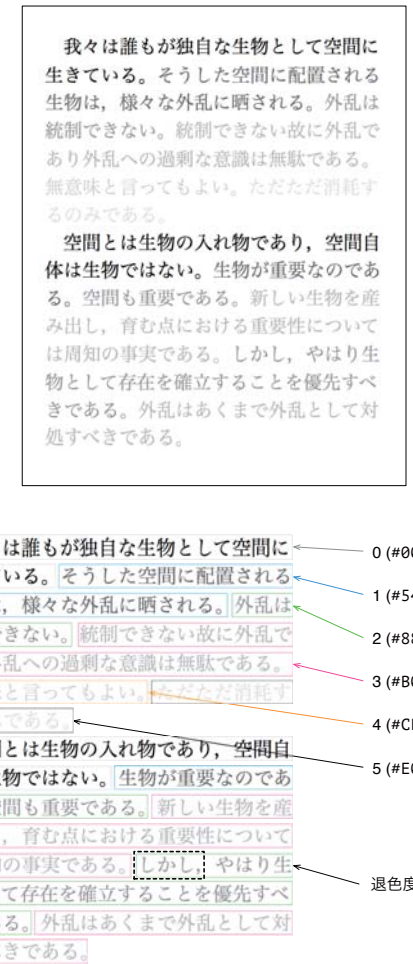


図 1: 提案方式による文章表示と退色度設定の例．

例外として，「順接（したがって等）」「逆説（しかし等）」「対比（むしろ等）」「総括（すなわち等）」を示す接続詞 4 種類がセンテンス先頭に存在する場合は，図 1 の「しかし」を含むセンテンスのように，前センテンスよりも退色度をひとつ戻し，濃く表示することとした．

連絡先: 小林 潤平, kobayashi-j3@mail.dnp.co.jp,
大日本印刷株式会社, 東京都新宿区市谷加賀町 1-1-1

表 1: 予備実験における実験協力者 3 群の正答率.

	実験協力者群		
	提案方式群 (N = 15)	ランダム退色群 (N = 15)	標準方式群 (N = 15)
標準方式での平均正答率	47 ± 6 %	47 ± 10 %	60 ± 10 %

3. 実験 I（要旨把握）

まず，提案方式が要旨把握問題の正答率にあたえる影響について検証した.

3.1 実験方法

本実験条件に関する予備知識や被験経験のない大学生 45 名が，実験協力者として参加した.

刺激は公務員試験 *1 で出題された「要旨把握」問題とした. 要旨把握問題は本文と 5 つの選択肢文から構成され，実験協力者は本文を読んで「本文の主旨として最も妥当な選択肢」をひとつ選ぶ. たとえ選択肢文と同一の文が本文中に存在した場合でも，選択肢文が本文の主旨でない場合は不正解である. 本文と選択肢文あわせて 1000 字程度の 8 題を用いた.

提案方式における退色有無および退色場所の 2 要因を検証するため，すべてのセンテンスに退色度 0 を割り当てた「標準方式」，各センテンスに退色度 0~5 をランダムに割り当てた「ランダム退色方式」の 2 方式をあわせて準備した.

問題は一題ずつ，画面サイズ対角 9.7 inch，画面解像度 264 ppi の Apple 社製タブレット型端末「iPad」上に表示した.

実験協力者は，予備実験および本実験で計 8 題を解答した. まず，実験協力者 45 名をランダムに選んだ 15 名ずつの 3 群にわけ，それぞれ「提案方式群」「ランダム退色方式群」「標準方式群」とした. 次に，予備実験として，実験協力者 3 群ともに標準方式で 2 題を解答した. その後，本実験として，提案方式群は「提案方式」で，ランダム退色方式群は「ランダム退色方式」で，標準方式群は「標準方式」で，6 題を解答した. 出題順はランダムとした. 予備実験と本実験は連続して実施し，実験協力者には予備実験と本実験の区別を教示しなかった.

3.2 結果

3.2.1 予備実験

提案方式の効果を検証する前に，実験協力者 3 群ともに標準方式で表示した場合の正答率を確認した. その結果を表 1 に示す. 提案方式群およびランダム退色方式群の平均正答率は，標準方式群よりも低いことがわかった. ただし，分散分析では実験協力者群間で有意差が認められなかった.

3.2.2 本実験

提案方式が要旨把握問題の正答率にあたえる影響を分析した.

図 2 は，各表示方式における正答率を示したものである. 提案方式における正答率 57%に対して，ランダム退色方式では 37%，標準方式では 34%であった. 分散分析の結果，表示方式間で有意な差が認められた（$F_{2,42} = 6.18$, $p < 0.01$）. また，Dunnett 検定より，提案方式と標準方式（$p < 0.01$）の間で有意差が認められた. 一方，ランダム退色方式と標準方式の間では有意差が認められなかった.

したがって，要旨把握問題を提案方式で表示した場合の正答率は，ランダム退色方式および標準方式で表示した場合よりも，有意に向上していることがわかった. また，ランダムな場所を退色しても効果をもたないことがわかった.

*1 特別区職員，国家公務員一般職，裁判所職員

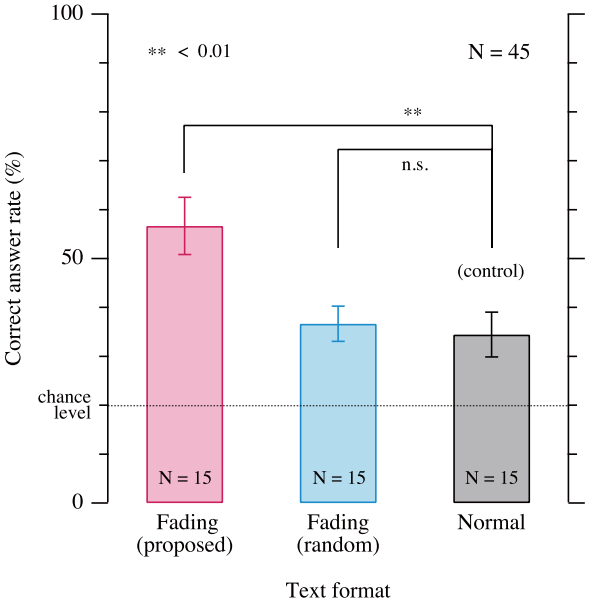


図 2: 各表示方式における要旨把握問題の平均正答率.

4. 実験 II（眼球運動）

実験 I より，提案方式では要旨把握問題の正答率が向上することがわかった. その原因を検討すべく，実験 II では，提案方式が眼球運動にあたえる影響について検証した.

4.1 実験方法

本実験条件に関する予備知識や被験経験のない大学生 9 名が，実験協力者として参加した.

刺激は，実験 I と同様，公務員試験で出題された要旨把握問題とした. 本文と選択肢文あわせて 1000 字程度の 8 題を用いた. 問題は一題ずつ，画面サイズ対角 12.9 inch，画面解像度 264 ppi の Apple 社製タブレット型端末「iPad Pro」上に表示した. 実験 I と同様に「提案方式」「ランダム退色方式」「標準方式」の 3 方式で検証した.

まず予備実験として，実験協力者は全実験協力者で共通の 2 題を標準方式で解答した. その後，本実験として各表示方式で 2 題ずつ計 6 題を解答した. 本実験では，問題と表示方式と出題順の組み合わせはランダムとした. 予備実験と本実験は連続して実施し，実験協力者には予備実験と本実験の区別を教示しなかった. 実験協力者の目の動きは，nac 社製の視線検出装置 EMR-9 を用いて 1/60 s 間隔で計測した.

4.2 結果

実験協力者は，まず本文を読み，次に選択肢文を読み，必要に応じて本文を再読した後，選択肢を一つ選んだ. 本文よりも先に選択肢文を読んだ実験協力者はいなかった.

4.2.1 要旨理解度

まず，各表示方式における要旨把握問題の正答率を確認した.

表 2 は，各表示方式における要旨把握問題の正答率を示したものである. 表 2 より，実験 II においても，提案方式で表示した場合の正答率がもっとも高く，ランダム退色方式および標準方式で表示した場合よりも向上していることがわかった.

表 2: 各表示方式における要旨把握問題の正答率.

	表示方式		
N = 9	提案方式	ランダム退色	標準方式
平均正答率	56 ± 15 %	17 ± 8 %	33 ± 11 %

表 3: 各表示方式における解答中の再読行動.

	表示方式		
N = 9	提案方式	ランダム退色	標準方式
平均再読発生率	40 ± 1 %	50 ± 1 %	40 ± 1 %
平均再読回数	1.8 ± 0.3	2.4 ± 0.7	1.9 ± 0.4
平均再読時間率	6.3 ± 0.8 %	13 ± 5 %	6.8 ± 1.8 %

表 4: 各表示方式における読み速度.

	表示方式		
N = 9	提案方式	ランダム退色	標準方式
平均読み速度（文字/分）	707 ± 61	661 ± 57	660 ± 63

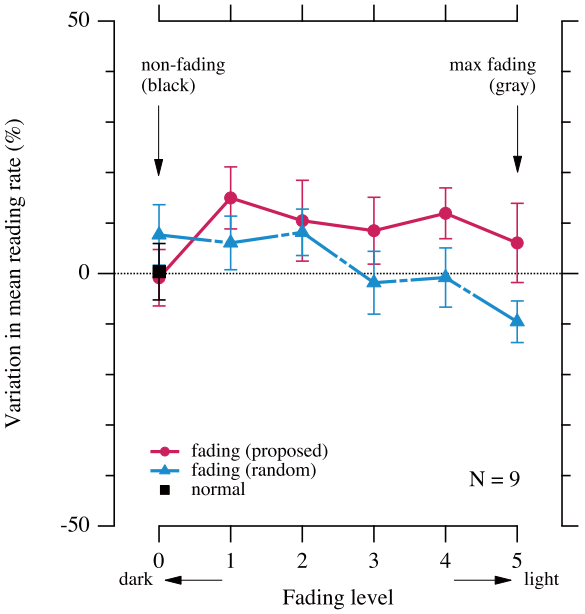


図 3: 各表示方式における文字退色度に対するセンテンス単位の読み速度の変化.

したがって，実験 II の実験協力者においても，実験 I の実験協力者と同様に，提案方式による要旨把握問題正答率の向上効果が認められた.

4.2.2 解答中の再読行動

次に，提案方式が本文の再読行動にあたえる影響を分析した.

表 3 は，各表示方式における本文再読行動を示したものである．本研究では「再読」を「本文を読み終えた後，選択肢文を読んでいる最中に本文をもう一度読むこと」と定義した．表 3 より，各表示方式における再読発生率，再読が発生した場合の再読回数，本文の全閲覧時間に対する再読時間の割合について，提案方式とランダム退色方式および標準方式と間に大きな差異は認められなかった.

したがって，選択肢文を読んでいる間の本文再読に関しては，表示方式間で大きな違いがないことがわかった.

4.2.3 退色度と読み速度

最後に，提案方式が読み速度にあたえる影響を分析した.

表 4 は，各表示方式における本文の平均読み速度を示したものである．表 4 より，提案方式で表示した場合の本文全体の平均読み速度は，ランダム退色方式および標準方式で表示した場合よりも速いことがわかった.

このとき，センテンス単位の文字退色度と読み速度の関係を図 3 に示す．縦軸はセンテンス単位の平均読み速度の変化率，横軸はセンテンスの文字退色度，誤差範囲は標準誤差である．読み速度変化率の基準値は，予備実験における標準方式の平均読み速度とした.

図 3 より，提案方式で本文を表示した場合，退色度 0 のセンテンスは標準方式と同等の読み速度であったが，退色度 0 以外のセンテンスでは，標準方式よりも速く読んでいることがわかった．一方，ランダム退色方式では，退色度 0 のセンテンスにおいて標準方式よりも速い読み速度であったが，退色度が増加し文字色が薄くなるにつれて読み速度は低下し，最も薄い退色度 5 のセンテンスでは標準方式よりも遅くなる傾向を示した．すべての文字の退色度が 0 である標準方式では，予備実験における標準方式の平均読み速度と同等であった.

したがって，提案方式では標準方式よりも平均読み速度が向上していることがわかった．また，その原因は，通常よりも薄く表示されるセンテンスを，通常よりも速く読んでいるためであることがわかった.

以上より，本文を読んだ後に選択肢文を読み，必要に応じて本文を再読して回答するという流れについては，各表示方式ともに同様であった．しかし，読み速度については提案方式で特徴的な分布を示しており，段落先頭センテンスは通常速度で読んでいるが，段落先頭以外のセンテンスは通常よりも速く読んでいることがわかった.

5. 実験 III（記憶）

実験 II より，提案方式では段落先頭以外のセンテンスを速く読んでいることがわかった．実験 III ではさらに，提案方式が文章内容の記憶にあたえる影響について検証した.

実験は，各表示方式で呈示される文章を読んだあと，次々呈示されるセンテンスに対して「読んだ文章に含まれていたセンテンス（ターゲット刺激）」か「含まれていなかったセンテンス（ディストラクタ刺激）」かを回答する再認テストとした．表示方式による再認成績の変化を検証した.

5.1 実験方法

本実験条件に関する予備知識や被験経験のない大学生 6 名が，実験協力者として参加した.

呈示文章は，実験 I および II と同様に，公務員試験で出題された 15 題の要旨把握問題とした．ただし，実験 III では本文のみ用い，選択肢文は削除した．呈示した本文の文字数は 700 字程度であった.

ターゲット刺激は，上記 15 題の本文中から抽出した．段落先頭センテンス 15 文および段落末尾センテンス 15 文の計 30 文とした．ディストラクタ刺激は，公務員試験で出題された 30 題の現代文問題の本文から抽出した．段落先頭センテンスや段落末尾センテンスを中心に，固有名詞など特徴的な単語を含まない計 60 文を抽出した.

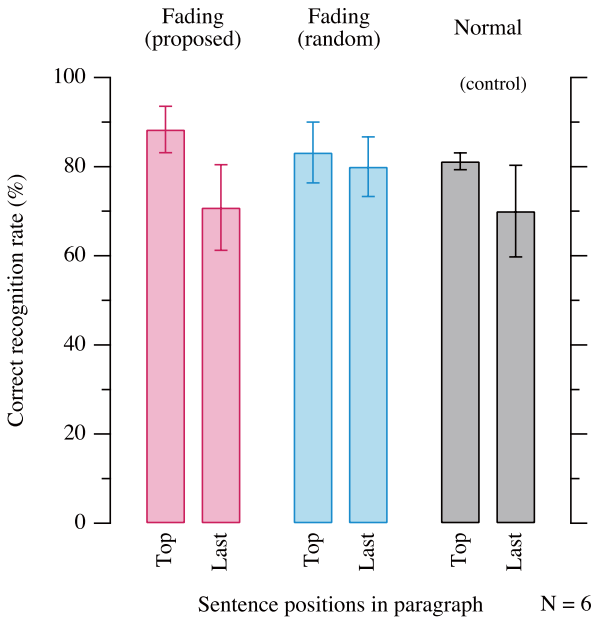


図 4: 各表示方式におけるセンテンス再認率とセンテンス位置の関係.

呈示文章は一文章ずつ，ターゲット刺激およびディストラクタ刺激は一センテンスずつ，画面サイズ対角 12.9 inch，画面解像度 264 ppi の Apple 社製タブレット型端末「iPad Pro」上に表示した．

実験協力者は,「提案方式」「ランダム退色方式」「標準方式」の 3 方式で 5 文章ずつ計 15 文章を読んだあと，次々と呈示されるセンテンスに対して，ターゲット刺激かディストラクタ刺激かを回答した．呈示文章と表示方式と呈示順の組み合わせはランダムとした．ターゲット刺激およびディストラクタ刺激の呈示順もランダムとした．なお，実験協力者には，あらかじめ読後に要旨把握問題を解答することを教示した．ただし，この教示は読みの理解度を一定にするためであり，実際には問題を呈示せず，センテンスに対する再認テストのみ実施した．

5.2 結果

再認成績と表示方式の関係を図 4 に示す．縦軸がターゲット刺激のセンテンスを再認できた割合，横軸がターゲット刺激センテンスの抽出位置（「Top」が段落先頭センテンス,「Last」が段落末尾センテンス），誤差範囲は標準誤差である．

図 4 より，提案方式では，標準方式と同様に，段落先頭センテンスに対する再認率が高く，段落末尾センテンスに対する再認率は低い傾向をもつことがわかった．また，提案方式では，段落先頭センテンスに対する再認率が標準方式よりも向上していることがわかった．一方で，ランダム退色方式では，段落末尾センテンスに対する再認率が提案方式および標準方式よりも向上しており，段落先頭センテンスおよび段落末尾センテンスに対する再認率は，同程度であることがわかった．

したがって，提案方式では，段落先頭センテンスに対する再認率が向上するとともに，段落先頭センテンスに対する再認率は段落末尾センテンスよりも高い傾向をもつことがわかった．

6. 全体考察

実験 I，II，III の結果をもとに，提案方式が文章の要旨把握にあたえる影響について考察する．

提案方式では，要旨把握問題の正答率が有意に向上していることがわかった（図 2）．このとき提案方式では，段落先頭センテンスを通常の速度で読んでいるが，段落先頭「以外」のセンテンスは，通常よりも速く読んでいることがわかった.（図 3）．さらに，提案方式では，段落先頭センテンスに対する再認率が向上しており，よく記憶されていることがわかった（図 4）．

一方，ランダム退色方式における要旨把握問題の正答率は，標準方式と同程度または低下していることから（図 2，表 2），ランダムな場所を退色した表示方式は，読解に対する向上効果をもたないことがわかった．

一般的に，段落先頭のセンテンスには重要な情報が書かれていることが多い．提案方式では，段落先頭のセンテンスを相対的にゆっくりと読み，再認率も向上していることから，部分的退色によって段落先頭センテンスへの注意を促し，より正確な要旨把握をもたらした可能性が推察された．

文章の読解においては，ワーキングメモリにおける保持と処理のバランス調整や，不要となった先行情報を抑制する能力が重要とされる [Gernsbacher 90, 苧阪 98, 大塚 03]．提案方式は読みにおけるワーキングメモリの情報処理を支援している可能性が示唆されるが，今後さらなる検証が必要である．

7. おわりに

本研究では，文章読解を支援する電子リーダーを実現すべく，段落先頭から順にセンテンス単位で文字を退色させていく表示方式を提案し，文章の要旨把握にあたえる影響について検証した．提案方式では，段落先頭センテンスに対する再認率が向上し，要旨把握問題の正答率も有意に向上することがわかった．部分的退色によって段落先頭センテンスへの注意を促し，より正確な要旨把握をもたらした可能性が推察された．

謝辞

公立はこだて未来大学 松原 仁 教授に機材の便宜をお図りいただくとともに，公立はこだて未来大学学生の方々に被験者として多大なご協力をいただいた．ここに感謝の意を表する．

参考文献

[Diemand-Yauman 11] Diemand-Yauman, C., Oppenheimer, D. M., and Vaughan, E. B.: Fortune favors the Bold (and the Italicized): Effects of disfluency on educational outcomes, *Cognition*, Vol. 118, No. 1, pp. 111–115 (2011)

[Gernsbacher 90] Gernsbacher, M. A., Varner, K. R., and Faust, M. E.: Investigating differences in general comprehension skill, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 16, No. 3, pp. 430–445 (1990)

[魚崎 03] 魚崎 祐子, 伊藤 秀子, 野嶋 栄一郎：テキストへの下線ひき行為が内容把握に及ぼす影響, 日本教育工学雑誌, Vol. 26, No. 4, pp. 349–359 (2003)

[内田 97] 内田 友幸, 田中 英彦：可読性向上を図る対話的文書自動彩色システム, 電子情報通信学会論文誌 D-II, Vol. 80, No. 12, pp. 3173–3180 (1997)

[大塚 03] 大塚 結喜, 森下 正修, 近藤 洋史, 苧阪 直行：読解力とワーキングメモリにおける抑制メカニズムの関係性, 基礎心理学研究, Vol. 21, No. 2, pp. 131–136 (2003)

[苧阪 98] 苧阪 満里子：読みとワーキングメモリ, 苧阪 直行（編）, 読み—脳と心の情報処理, 第 14 章, pp. 239–262, 朝倉書店 (1998)

[関 96] 関 友作, 赤堀 侃司：テキストにおける段落表示が内容理解に与える影響, 日本教育工学雑誌, Vol. 20, No. 2, pp. 97–108 (1996)

[関 97] 関 友作：テキストの内容把握に対する箇条書とキーワード強調の影響, 日本教育工学雑誌, Vol. 21, pp. 17–20 (1997)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

色彩調和に基づく衣服コーディネートシステムの提案

Proposal of fashion coordinate system based on color harmony

齊藤拓己 *1　岡瑞起 *1　加藤和彦 *1
Takumi Saito　Mizuki Oka　Kazuhiko Kato

*1筑波大学大学院システム情報工学研究科
University of Tsukuba, Graduate School of Systems and Information Engineering

In recent years, with the development of fashion online shops or recommendation services, images of clothes can be easily collected. Recently, various models using machine learning or deep learning are proposed on the collected data. Moreover, various color theories have been proposed and shown that color combinations affect greatly on fashion.

In this research, we propose a system that recommends color combinations for clothes given an input image of a person. The proposed system is built by two models: the first is a parsing model that divides an image into regions by clothes; and the second is a harmonic evaluation model that evaluates the color combinations for clothes based on the color harmony theory.

1. はじめに

衣服通販やコーディネート投稿型 SNS の発達によって，衣服に関するテキストや画像データが容易に収集可能となった．これらのデータを機械学習や Deep Lerning の手法を用いて学習することで様々なモデルが提案されている．大きく分けて，衣服を識別するモデル [Ziwei 16][Pongsate 17]，衣服を検索するモデル [Kiapour 15][Katrien 17]，衣服を推薦モデル [Kristen 16][Xintong 17] の 3 つの研究が特に盛んである．

また，衣服との関係が深い色彩の分野は古くから研究されており，調和する配色の法則をまとめた色彩調和論が様々提案されている．色彩調和の観点から，配色を定量的に評価する方法についても研究され，古くは，Birkhoff が提案した物事の美しさの定量的評価手法 [Birkhoff 33] を，Moon と Spencer が配色に対して適用したもの [Moon 44] から，現在では，松田の色彩調和論 [Matsuda 95] に対し，ファジー推論を用いて配色の調和度を定量的に評価する手法 [Tokumaru 07] などがある．

本研究では，衣服を識別するモデル [Edgar 14] と配色の定量的評価手法 [Tokumaru 07] を組み合わせることで，図 1 のように，入力された衣服コーディネート画像に対して色彩調和に基づいて衣服の色を変更する，新たなシステムを提案する．

2. 松田の色彩調和論

松田が提案した色彩調和論 [Matsuda 95] は，9 年間にわたって女子学生の衣服と，プリント服地の配色調査をした結果に基づきまとめたものである．

図 2 のように色相環における色相の分布を表したものを色相構成とし，8 個のタイプにまとめられた．図中の黒い部分はその色相構成に含まれる色相の領域を示している．また，図 3 のように明度と彩度の分布を表したものを色調構成とし，10 個のタイプにまとめられた．縦方向は明度，横方向は彩度を表し，色相構成と同様に黒い部分は，色調構成に含まれる明度，彩度の領域を示している．これら色相構成と色調構成の，調和する相関ルールが図 4 のようにまとめられており，合計 33 通りの調和する組み合わせがある．

連絡先: 齊藤拓己，st_tkm@websci.cs.tsukuba.ac.jp



図 1: 提案システム入力例 (左:入力，右:出力)

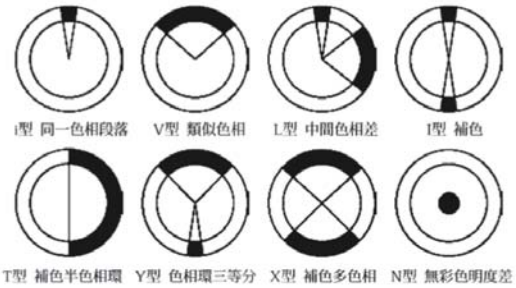


図 2: 色相構成 [Tokumaru 07]

3. 提案システム

図 1 左のコーディネート画像を入力し，図 1 右のように色彩調和に基づいて色変更した画像を出力するシステムを提案する．提案システムを構築する 5 つのモジュールについて以下でそれぞれ述べる．

3.1 パーシングモジュール

このモジュールでは，衣服コーディネート画像に対し衣服ごとの領域分割を行う．本システムでは，既存手法 [Edgar 14]

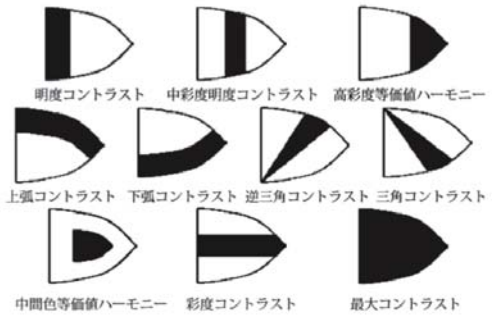


図 3: 色調構成 [Tokumaru 07]

色相構成	色調構成
i型	明度，逆三角，三角，最大
V型	明度，逆三角，三角
L型	明度，逆三角，三角
I型	高彩度等価値，逆三角，三角，中間色等価値
T型	明度，中彩度明度，高彩度等価値，最大，
	上弧，下弧，逆三角，三角，中間色等価値
Y型	高彩度等価値，逆三角，三角，
	中間色等価値，最大
X型	逆三角，三角，最大
N型	明度

図 4: 相関ルール [Tokumaru 07]

を用いる．これは色のヒストグラム情報や，セグメンテーション情報，人体推定に基づく衣服位置推定情報などを用いて，機械学習手法である Conditional Random Field(CRF) で学習したモデルである．図 5 右のように，コーディネート画像を衣服ごとに領域分割することが可能である．



図 5: 入力例におけるパーシング結果

3.2 主要色決定モジュール

このモジュールでは，衣服ごとに分割された領域のピクセル色から，衣服の色を一意に決定する．

まず領域のピクセル色をすべて抽出し，RGB から Lab に変換する．その後 Lab 色空間内での色の分布をみるため，すべて重み付きの格子空間に射影する．これによって図 5 の Bottoms の色の分布は図 6 のようになる．次に，分布の大きい点での重心を求めることで，その重心の色を対象衣服の主要色とする．入力例における各衣服の主要色を図 7 に表す．

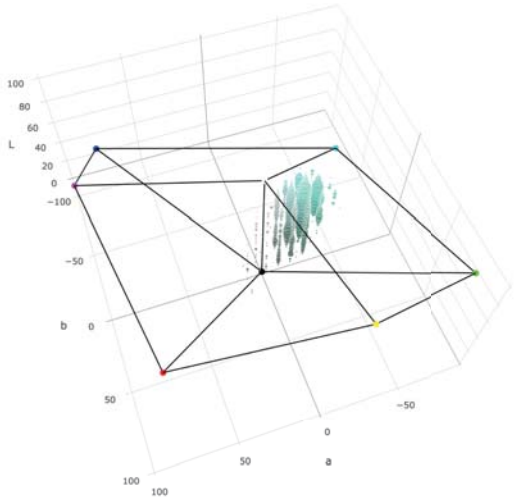


図 6: 入力例における Bottoms の色分布

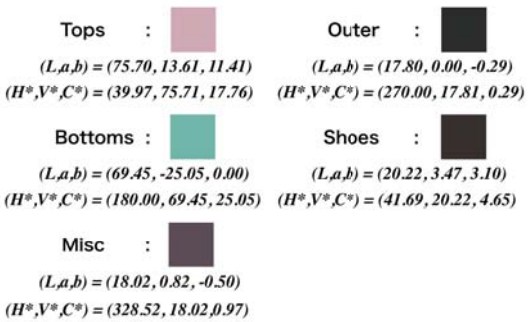


図 7: 入力例における各衣服の主要色

3.3 調和評価モジュール

このモジュールでは，衣服ごとの主要色をまとめた配色の調和度評価を行う．評価には，松田の色彩調和論に対するファジー推論によって調和評価を行う手法 [Tokumaru 07] を用いる．この手法では，色の認識の曖昧性を考慮するためにファジー推論を用いており，調和度として 0 1 で配色を評価している．

図 2 の色相構成と図 3 の色調構成をファジー集合化し，それぞれに対して，適合度を計算するメンバーシップ関数を新たに定義．図 4 の相関ルールに基づいて適合度を計算し，最大となる値を調和度とする．

松田の色彩調和論では色相明度彩度の 3 特徴量を用いているため，図 8 のように，Lab を色相 H^*，明度 V^*，彩度 C^* に変換する．Lab のパラメータを用いて以下のように求める．

$$V^* = L \tag{1}$$

$$C^* = \sqrt{a^2 + b^2} \tag{2}$$

$$H^* = \begin{cases} \cos^{-1}\left(\dfrac{a}{\sqrt{a^2+b^2}}\right) & (b \geq 0) \\ -\cos^{-1}\left(\dfrac{a}{\sqrt{a^2+b^2}}\right) & (b < 0) \end{cases} \tag{3}$$

変換された 3 特徴量を用いて，図 7 の配色の調和評価を行ったところ，T 型の色相構成と最大コントラストの色調構成の組み合わせとなり，調和度は 0.002 であった．

3.4 色推薦モジュール

このモジュールでは，調和評価に基づいて新たな色を推薦する．

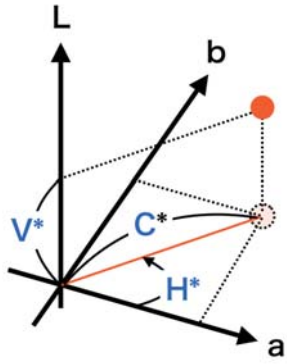


図 8: Lab 色空間での H^*, V^*, C^*

まず，配色の内の 1 色を減らした場合全てにおいて調和評価を行い，調和度が最も高くなるようなパターンに対して，減らした色を不調和な色であると判断し，変更対象とする．次に，このパターンにおいての色相構成，色調構成の両メンバーシップ関数に基づき，他の配色の重心に最も近い，かつ適合度が最大となるように色相，明度，彩度を推薦する．

入力例では，tops，outer，bottoms，shoes，misc の色をそれぞれ除いたときの調和度は，0.00，0.36，0.57，0.00，0.20 となり，bottoms の色を除いた配色の調和度が最も高くなる．そのため，bottoms の色を新たに推薦する．また，この時の色相構成タイプは V 型，色調構成タイプは明度コントラストである．図 9 は，V 型のメンバーシップ関数上に，bottoms 以外の位置と新たに推薦された色相位置を示している．図 10 は，推薦された色相の等色相面に対する明度コントラストのメンバーシップ関数上に，outer 以外の位置と新たに推薦された明度彩度位置を示している．図 11 に変更前の色と変更後の色を示す．

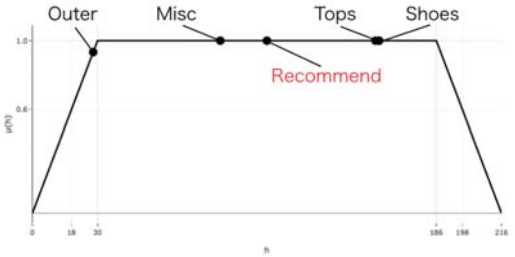


図 9: 色相推薦

3.5 色変更モジュール

このモジュールでは，対象となる衣服の色を，違和感のないよう推薦された色に変更する．色変更は，単純な空間構造を持つ RGB 色空間を用いて行う．そのため，各衣服の主要色および推薦色をすべて RGB に変換する．次に，色変更を行うにあたり色相，明度，彩度の 3 特徴量が必要なため，図 12 のように，RGB を色相 H^{**}，明度 V^{**}，彩度 C^{**} に変換する．

RGB 色空間における色ベクトル $\vec{K}$ を無彩色要素 $\vec{K_n}$ と有彩色要素 $\vec{K_c}$ に分解する．

$$\vec{K} = \vec{K_n} + \vec{K_c} \tag{4}$$

このとき $\vec{K_n}$ のノルムは明度，$\vec{K_c}$ のノルムは彩度を表す．

$$V^{**} = \|\vec{K_n}\| \tag{5}$$

$$C^{**} = \|\vec{K_c}\| \tag{6}$$

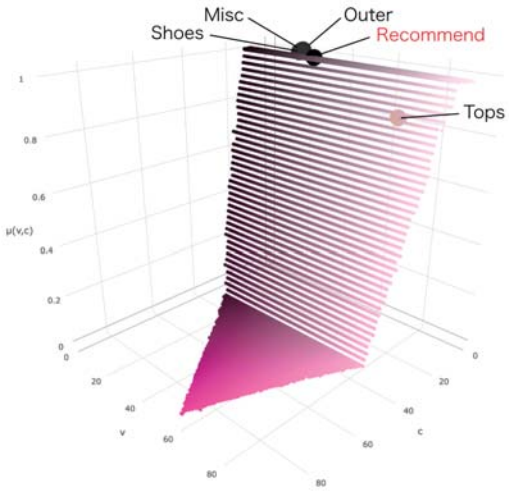


図 10: 色調推薦

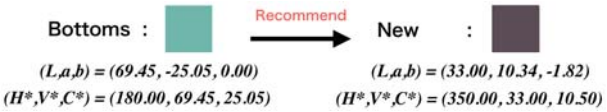


図 11: 入力例の色推薦

$\vec{K_n}$ は $(255, 2\vec{55}, 255)$ 上にあるため，

$$\vec{K_n} = \frac{\vec{K} \cdot (255, 2\vec{55}, 255)}{\|(255, 2\vec{55}, 255)\|} \tag{7}$$

以上より明度 V^{**}，彩度 C^{**} が求められる．色相 H^{**} は，赤 $(255, 0, 0)$ の有彩色要素 $\vec{R}$ と、それに垂直な要素 $\vec{R^{\perp}}$ を基準として角度を求める．

$$H^{**} = \begin{cases} \cos^{-1}\left(\dfrac{\vec{K_c} \cdot \vec{R}}{\|\vec{K_c}\|\|\vec{R}\|}\right) & (\vec{K_c} \cdot \vec{R^{\perp}} \geq 0) \\ -\cos^{-1}\left(\dfrac{\vec{K_c} \cdot \vec{R}}{\|\vec{K_c}\|\|\vec{R}\|}\right) & (\vec{K_c} \cdot \vec{R^{\perp}} < 0) \end{cases} \tag{8}$$

変換前のピクセル色 X，変更対象の衣服の主要色 G_x，推薦された色 G_y を以上のように変換し，以下のように表す．

$$X = (H_X, V_X, C_X) \tag{9}$$

$$G_x = (H_{GX}, V_{GX}, C_{GX}) \tag{10}$$

$$G_y = (H_{GY}, V_{GY}, C_{GY}) \tag{11}$$

また変更後のピクセル色 Y を

$$Y = (H_Y, V_Y, C_Y) \tag{12}$$

としたとき，Y の成分はそれぞれ以下のように求める．

$$H_Y = H_X + (H_{GY} - H_{GX}) \tag{13}$$

$$V_Y = \begin{cases} V_X \dfrac{V_{GY}}{V_{GX}} & (V_X \leq V_{GX}) \\ V_{GY} + (V_X - V_{GX}) \dfrac{V_{max} - V_{GX}}{V_{max} - V_{GY}} & (V_X > V_{GX}) \end{cases} \tag{14}$$

$$C_Y = \begin{cases} C_X \dfrac{C_{GY}}{C_{GX}} & (C_X \leq C_{GX}) \\ C_{GY} + (C_X - C_{GX}) \dfrac{C_{max} - C_{GX}}{C_{max} - C_{GY}} & (C_X > C_{GX}) \end{cases} \tag{15}$$

ここで V_{max}，C_{max} はそれぞれ V^{**}，C^{**} の全体で取りうる最大値を表す．

以上のようにして変更対象の領域すべてのピクセル色を変更することで，図 1 のような画像を出力することが可能である．

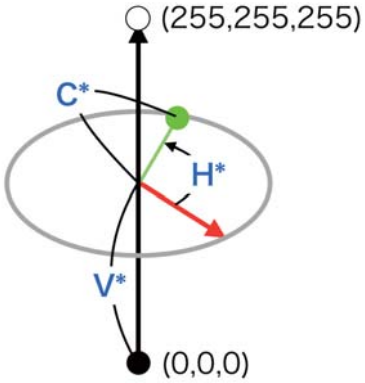


図 12: RGB 色空間での H^{**}, V^{**}, C^{**}

4. まとめ

本研究では，衣服ごとの領域分割手法 [Edgar 14] と配色の調和評価手法 [Tokumaru 07] を組み合わせることで，図 1 のように色彩調和に基づいて衣服の色を変更するシステムを提案した．色彩調和に基づいて色を推薦することで，衣服コーディネートの選択に貢献することができると考えられる．また，色変更された衣服に対して検索モデルを適用することで，衣服通販サービスの利用にも繋がると考えられる．

5. 課題

本提案システムの課題は 4 つ挙げられる．1 つ目は，パーシング精度の問題である．不正確なパーシングを行った場合，正確な色を抽出できず，色変更の際にも対象とは別の衣服の色が変更されてしまうといったように，結果的に不調和なコーディネートを推薦することもある．最新のパーシングモデル [Pongsate 17] を用いることでこの問題を解決したい．

2 つ目は，衣服が 2 つ以上の色で構成されている場合の問題である．ボーダー柄や花柄など，柄のある衣服の多くは複数の色で構成されている．本研究で構築した代表色決定モジュールでは衣服の色を一意に決定しているため，衣服が複数の色を持つ場合には対応できず，実際に存在する色を無視して調和評価を行ってしまう．この問題は，抽出した色を色相の分布によってクラスタリングすることで解決できると考えられる．

3 つ目は，無彩色と有彩色の境界問題である．人間の知覚による入力例の outer，shoes，misc の色は，黒であり無彩色だと判断できるが，図 7 の C^* 値が 0 以上であることからも分かるように，システム上では有彩色だと判断される．従って，本来色相を持たない無彩色に対しても色相構成の適合度が計算されるため，意図しない評価となる．無彩色と有彩色における境界の曖昧性を考慮し，低彩度のときに色相構成のメンバーシップ関数を補正することで解決できると考えられる．

4 つ目は，色の面積や隣接関係による調和度への影響の問題である．入力例では，全体における shoes 面積の割合はかなり小さく，全体の調和評価にはほとんど影響しないと考えられる．また，色の隣接関係によっても全体の調和度は変化すると考えられる．色の面積や位置関係が及ぼす全体の調和についての研究 [Wenke 15] を参考に改善したい．

参考文献

[Ziwei 16] Ziwei L, Ping L, Shi Q, Xiaogang W, Xiaoou T.: DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp.1196-1104(2016).

[Pongsate 17] Pongsate T, Zhipeng W, Kota Y.: Looking at Outfit to Parse Clothing. CoRR(2017).

[Kiapour 15] MH Kiapour, Xufeng H, Svetlana L, Alexander C. B, Tamara L. B.: Where to buy it: Matching street clothing photos in online shops. In Proc. of the 2015 IEEE International Conference on Computer Vision, pp.3343-3351(2015).

[Katrien 17] Katrien L, Susana Z, Marie F. M.: Cross-modal Search for Fashion Attributes. In Proc. of the KDD workshop on ML meets Fashion(2017).

[Kristen 16] Kristen V, Sunaya S, Ziqiao D, Karrie K, Ranjitha K.: The Elements of Fashion Style. In Proc. of the 29th Annual Symposium on User Interface Software and Technology, pp.777-785(2016).

[Xintong 17] Xintong H, Zuxuan W, Yu-Gang J, Larry S. D.: Learning Fashion Compatibility with Bidirectional LSTMs. In Proc. of the 2017 ACM on Multimedia Conference, pp.1078-1086(2017).

[Edgar 14] Edgar S, Sanja F, Francesc M, Raquel U: A High Performance CRF Model for Clothes Parsing. In Proc. of the 12th Asian Conference on Computer Vision, pp.64-81(2014).

[Birkhoff 33] Birkhoff G. D.: Aesthetic Measure. Cambridge, Mass: Harvard Univ. Press(1933).

[Moon 44] Moon P. & Spencer D. E.: Aesthetic measure applied to color harmony. Journal of the Optical Society of America, vol.34(4), pp.234-242(1944).

[Matsuda 95] Matsuda Y.: Color design. Asakura Shoten(1995).

[Tokumaru 07] Tokumaru M, Muranaka N, Imanishi S: Quantitative evaluation of color harmony using fuzzy reasoning. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.19(1), pp.57-68(2007).

[Wenke 15] Wenke K, Shengfeng Q, Quan Z.: Computer Aided Color Aesthetic Evaluation Method Based on the Combination of Form and Color. Mathematical Problem in Engineering, Volume 2015, Article ID 153103(2015).

Oral presentation | Organized Session | [Organized Session] OS-4

[2F2-OS-4a] [Organized Session] OS-4

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room F (4F Garreria)

[2F2-OS-4a-01] Well-being and Haptics

○Junji Watanabe[1] (1. NTT Communication Science Laboratories)

1:20 PM - 1:40 PM

[2F2-OS-4a-02] Emotional responses are changed bi-directionally through face-to-face interaction

○Aiko Murata[1] (1. Waseda University)

1:40 PM - 2:00 PM

[2F2-OS-4a-03] Measurement trials of crowd’ s well-being using visible light communication

○Takashi G Sato[1], Yoshifumi Shiraki[1], Takehiro Moriya[1] (1. Nippon Telegraph and Telephone)

2:00 PM - 2:20 PM

[2F2-OS-4a-04] Analysis of the Relationship between Sleep Quality and Stress with a focus on Circadian Rhythm for Well-being

○Ryo Takano[1], Satoshi Hasegawa[1], Yuta Umenai[1], Takato Tatsumi[1], Keiki Takadama[1], Toru Shimuta[2], Toru Yabe[2], Hideo Matsumoto[2] (1. The University of Electro-Communications, 2. Murata Manufacturing Company, Ltd.)

2:20 PM - 2:40 PM

[2F2-OS-4a-05] Investigation of appropriate fNIRS feature to evaluate cognitive load

○Kota Suzuki[1], Tatsuya Suzuki[1], Sotaro Shimada[1], Atsumichi Tachibana[2], Yumie Ono[1] (1. Meiji University, 2. Dokkyo Medical University)

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

1:20 PM - 1:40 PM (Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room F)

[2F2-OS-4a-01] Well-being and Haptics

○Junji Watanabe[1] （1. NTT Communication Science Laboratories）

Keywords: Well-being, Haptics

私たちの日常生活では、スマートフォンをはじめとする携帯型情報端末や無線通信環境が普及し、いつでもどこでもインターネットにアクセスし、メッセージの授受を行えるようになった。これは人間の情報処理能力を向上させたということができるものの、一方で、その心身への負の影響も指摘されている。また、ロボットが人間の身体的機能を補助・代替するだけでなく、人工知能が人間の知的機能を補助・代替するようになり、人間にとっての労働の意味や生きる意味について、様々な議論を巻き起こしている。そのような背景から近年、持続的ウェルビーイング（ Eudaimonic Wellbeing）についての研究が盛んに進められている。本発表では、ウェルビーイングに関する以下の3つの話題について述べる。（１）ウェルビーイングを構成する主要な要因とその計測・モデル化を行うための手法について。（２）情動変容・行動変容・関係変容をもたらす触覚に関する介入について。（３）主に西洋の視点から進められてきたウェルビーイング研究に関して、どのように日本的な視点をもたらすことができるのかについてである。

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

相互作用による双方向的な情動反応の変化

Emotional responses are changed bi-directionally through face-to-face interaction

村田 藍子[*1]

[*1] 早稲田大学
Waseda University

[*2] 日本学術振興会
Japan Society for the Promotion of Science

Our emotional states are influenced from other's emotional states. Such implicit emotional contagion is considered to have a role as social glue of interpersonal relationship. Previous studies have investigated the processes how it occurs or identified the factors that affect the occurrence of it. In the experiments of these studies, someone's emotional expression or signal of physiological change (e.g., facial expression, pupil dilation) was presented as stimulus, and the occurrence of congruent emotional/physiological reaction of observer was measured. However, it is difficult to evaluate the impact of contagion by such stimulus-response paradigm, because contagion is a bidirectional process that is evoked through social interaction. Therefore, we aimed to investigate bidirectional physiological contagion during actual face-to-face interaction. In our experiment, two participants were given thermal stimuli simultaneously. To assess temporal physiological responses, blood volume pulse in their peripheral blood vessels was recorded. The results showed that the physiological responses of dyad members were correlated with each other when they could interact. Further, we demonstrated that individuals showed higher or lower physiological responses when their partners experienced stronger or weaker stimuli respectively. Thus, people can develop similar physiological responses through interactions, and this effect seems to induce a change in the responsivity to stimuli.

1. はじめに

個人のウェルビーイングは様々な形で周囲の他者から影響を受ける. 例えば, 幸福感は, 友人が幸福かどうか, 幸福な人が近くに住んでいるか, といった対人ネットワークの距離に応じて変化することが統計的に明らかにされているし[1], SNS を通じたやりとりを通しても感情が伝染することが社会実験によって示されている[2]. これらの背後には, 喜びや悲しみといった感情が他者に伝染する情動伝染と呼ばれるプロセスが働いていると考えられる. 情動伝染がどのように生じるのか, そのメカニズムを探る上でよく用いられてきた情動反応として「痛み」がある. 「痛み」 は刺激に対する生理的な反応の現れであり, 痛みを伴う刺激を提示した際の生体反応が比較的明確であるため計測がしやすいというメリットがある. これまで, 認知神経科学研究において, 他者の痛み表出を知覚することで, 痛みに関わる神経回路が活動することが発見され, 他者の「痛み」を共有する神経基盤を持つことが明らかになっている. 例えば, 隣にいる恋人が電気ショックを受ける場面では自分が電気ショックを受けたときと共通の神経回路(anterior insula, anterior cingulate cortex)が活動し[3], 誰かの手や足が痛みを受ける場面(ドアに指を挟むなど)の写真を観察しただけでも, 同様に自身が痛みを感じた際と同じ神経回路が活動することが示されてきた[4].

これらの研究では,「他者が痛みを感じる状態」を刺激とした際にそれに対する観察者の情動反応を計測するという一方向的なアプローチによって, 情動伝染の受け手の側のプロセスについて検討されてきた. しかしながら, 情動伝染のプロセスは双方向的な影響を生じさせる可能性がある. 現実のコミュニケーション場面では, 人々は互いの反応を参照できるため, 他者の情動反応から影響を受ける「受け手」となると同時に他者の情動反応に影響を及ぼす「送り手」ともなるため, 相互に影響を与え合う可能性が想定される. つまり, 他者の反応から影響を受けたことによって生じる自身の反応が, 表情・しぐさなどを通して表出され, 再び他者の反応に影響を及ぼすという双方向的なダイナミクスが生じると考えられる.

また, Gallotti と Frith は, 社会において相互作用する個人同士は, "we-mode"と呼ばれる集合的な認知モードにおいて, 互いの心的状態を共有していると論じており, この種の認知特性は個人内の認知のみを対象にした研究では測りきれないと主張している[4]. こうした着想に基づき, 本研究では,苦痛情動の伝染について個人間の相互作用を通して初めて捉えられる現象があるのではないかと考えた.

しかしながら, 社会心理学のパニックに関わる古典的理論[5]や人類学のフィールドにおけるデモンストレーション[6]はあるものの, 情動反応の双方向的な影響がどのようなメカニズムで生じ, どのような効果を持ちうるかについては, 実証的な検討が十分になされていない.

興味深いことに, Martin ら[7]は情動の共有が個人間で共有されることによって増幅する可能性を示している. 冷たい氷水に手を入れる体験を一人で行った場合と二人で行った場合とを比較し, 二人で経験したときの方が痛みを強く感じることを報告している. ただし, 彼らの研究では, 主観的な情動の強さに違いが生じることが報告させているものの, 不随意の生理的な反応にどのような影響が生じるかは検討されていない. また, 相手が自分よりも相対的に強い情動を感じている時や, 弱い情動を感じている時に, 相手の反応からどのように影響を受けるのかについても調べるために, 本研究では3つの条件を設け, その際の刺激に対する生理反応を計測することで, 1)Martin ら[7]により示された情動の増幅現象が生理反応のレベルで生じるか, 2)その際に二者の生理的な反応は相互作用によって類似するのか, 3)相手と自分が感じている情動の強さが異なる場合には, どのような影響をうけるのか, を実験により検討した.

具体的には, 一人で痛みを経験する Alone 条件, 二人で同じ強度の痛みを経験する Same 条件, 二人が異なる強度の痛みを経験する Different 条件を設け, その際の自律神経系の生理反応を指尖容積脈波により計測した.

村田藍子, 早稲田大学基幹理工学部表現工学科・渡邊研究室, 東京都新宿区大久保 3-4-1 59 号館 407-1B 号室, 03-5286-3335, aikmurata@gmail.com

2. 方法

2.1 実験参加者

東京近郊の学生 91 名（女性 28 名，男性 63 名）が参加し，実験終了後に参加謝礼として 2000 円を受け取った．

2.2 条件・刺激

同じ強度の刺激を同時に提示される Same 条件（21 ペア）と，異なる強度の刺激を同時に提示される Different 条件（18 ペア）を設定した．Same 条件では，参加者には互いの反応を参照できるように向かい合って座ってもらい，同じ温度の刺激をペアの二人に同時に，繰り返し提示した．刺激は2種類あり，参加者はペアごとに，50℃を提示されるペアか 60℃を提示されるペアのどちらかにランダムに割り振られた．Different 条件では，ペアの一方の参加者が 50℃を提示され，もう一方の参加者が 60℃を提示された．また，統制条件として，50℃か 60℃の刺激のどちらか一方を提示される Alone 条件（13 名）を設けた．

Same 条件と Different 条件では，対面する参加者に熱刺激を二人同時に繰り返し提示した．各参加者には同じ温度（50℃／60℃）の刺激を繰り返し提示したが，刺激が同一強度であることが参加者に分かってしまうと，刺激に対する馴化により生理的喚起水準が低減してしまう恐れがあったため，IH クッキングヒーターを使って不透明のホーロー鍋に入れた湯を温め，熱刺激の温度を参加者から推測されにくいように工夫した．さらに，提示される刺激の強度に対する予期を形成させないために，IH クッキングヒーターを参加者から加熱メニューの表示が見えないように設置した上で操作し，毎試行，加水・加熱をしながら温度調整を行った．試行と試行の間は目安として 30 秒以上となるようにした．時間の経過と温度の確認をした後，二人の参加者の前腕に同時に熱刺激を一秒未満提示した．なお，痛みの有無に関わらずアルミ棒の提示自体に驚くことで参加者の生理的喚起水準が高まる可能性が考えられたため，最初に馴化を目的としてアルミ棒を一度提示し，その後6試行を行った．Alone 条件でも同一の方法で一人の参加者に刺激を繰り返し提示した．

2.3 実験手続き

参加者は実験に二人で参加した．まず，実験室中央のスペースに設置された机をはさみ，向かい合って着席した．はじめに実験状況に慣れてもらうため，二人で会話をするセッションが設けられた．実験者が提示した簡単なテーマに沿って，約 2 分間二人で話すよう指示された（e.g.,「この前の週末は何をしていましたか？」）．その後，参加者の左手に BVP を計測するための電極を装着し，課題の準備を行った．Same 条件と Different 条件では，二人同時に課題を行ったが，Alone 条件では，一人が課題を行っている際にもう一人は防音室で待機してもらった．刺激は5秒間のカウントダウン（“5, 4, 3, . . . ”）の後に提示した．

2.4 痛みに対する自律神経系の生理反応

指尖容積脈波（BVP）を計測することで，熱刺激に対する自律神経系の生理反応を計測した．課題中，BIOPAC 社製の脈拍測定トランスデューサ（TSD200）を参加者の左手中指の第一関節部分に装着し，BIOPAC 社製 MP150 システムを用いて BVP を計測した．BVP のローデータは 2000Hz のサンプリングレートで 0.05Hz の High Pass フィルタをかけて抽出され，BIOPAC 社のソフトウェアである acqknowledge により記録された．その後，Matlab（MathWorks 社）により波動の振幅値を 100Hz のサンプリングレートで算出し，分析に用いた．

2.5 データ解析

参加者の刺激に対する生理反応を評価するために，試行ごとに，刺激提示開始 11 秒前から刺激提示 8 秒後までの計 19 秒間を生理反応の応答範囲として設定した．波動の振幅値は 100Hz のサンプリングレートで算出されたため，1 試行ごとの振幅値は 1900 のデータポイントをもつ時系列データであった．刺激に対してどの程度生理的喚起水準が高まったかを調べるための指標として，BVP 収縮率を用いた．BVP 収縮率{1-反応時（刺激提示後から8秒間）の最小振幅値/ ベースライン時 BVP（刺激提示 11 秒前から 6 秒前まで）の平均振幅値}は自律神経系のストレス反応の指標であり，収縮率が高いほど生理的喚起水準が高いことが知られている．

次に，Same 条件における二人の参加者の刺激に対する情動反応の類似性に関する解析を行った．二者に同時に刺激を提示しているため，もし二者の情動反応が正に相関したとしても，刺激によって誘発された疑似相関である可能性を排除できない．したがって，対面した二者の間の相互作用によって生じた情動反応の類似性を評価するためには，純粋に刺激によって誘発される相関と比較した際により強い相関を示していることを確認する必要がある．そこで，実際のペアになった相手以外のすべての参加者同士で総当たりの組み合わせ（仮想ペア）を生成し，比較対象とした．仮想ペアの場合，経験した刺激は実際のペアと同一であったが，ペアの相手との間に相互作用はなかったため，二者間の情動反応の相関は純粋に刺激のみによって誘発されたものであると解釈できる．そのため，もし実際のペアにおける情動反応の相関が仮想ペアのものに比べて高ければ，相互作用によって二者の情動反応が類似したと解釈することができる．具体的には試行ごとに，ペアの二者間の振幅値の時系列変動が類似しているかを検討するため，実際のペア及び仮想ペアの振幅値の相関係数を算出し（＝6 個の相関係数 / 1 ペア），解析に用いた．

3. 結果と考察

条件ごとの各刺激に対する平均 BVP 収縮率を Figure 1 に示す．Wilcoxson 順位和検定の結果，Same 条件では，Alone 条件に比べて，どちらの刺激に対しても，有意に痛みに対する生理反応が大きかった（50℃: p<.05, 60℃: p<.05, Bonferroni-corrected）．したがって，Martin ら[7]により示された相互作用による増幅が生理反応のレベルでも観察されることがわかった．

また，Same 条件では，50℃に対する生理反応に比べて 60℃に対する生理反応が高かった（t_{40}=2.44, p<.05, Bonferroni-corrected）のに対し，Different 条件では，50℃に対する生理反応と 60℃に対する生理反応が同程度であった（t_{33}=-1.02, *n.s.*, Bonferroni-corrected）．Different 条件における 50℃の刺激に対する生理反応は Same 条件のものと比べて相対的に大きかったのに対し，60℃の刺激に対する生理反応は Same 条件のものと比べて相対的に小さかったことから，二人が異なる強度の刺激を提示される Different 条件では，互いの反応から双方向的に影響を受けたことが示唆された．

加えて，情動の増幅が確認された Same 条件について，二者の情動反応が類似したのか，またその類似は相互作用を繰り返すことによって強まったのかを検討するために，課題の前半（1-3 試行）におけるペアの BVP の相関と，後半（4-6 試行）におけるペアの BVP の相関を調べた．ペアの BVP 値の相関係数（前半・後半各 3 データ / 1 ペア）について，実際のペアと仮想ペア（実際のペア以外のすべての組み合わせから作成）の間の相関係数の確率分布の比較を行った（see Figure 2）．

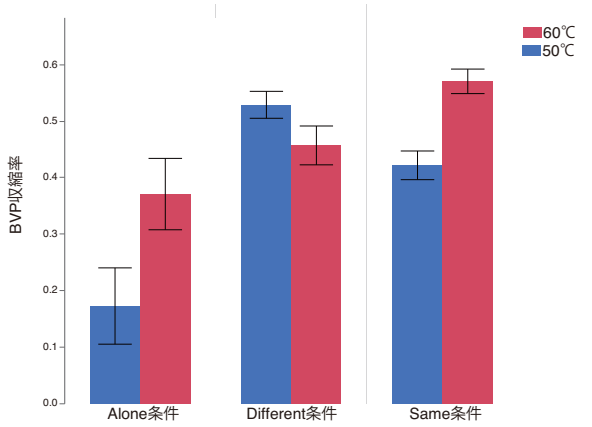


Figure 1. 条件ごとの熱刺激(50℃/60℃)に対する平均 BVP 収縮率

コルモゴロフ・スミルノフ検定の結果，50℃を提示されたペアでは，前半では実際のペアと仮想ペアの間で相関係数の分布に有意な差異が見られなかったのに対し(D = 0.20, *n.s.*, 片側検定)，後半では実際のペアの方が仮想ペアに比べて有意に相関が高かった(D = 0.32, p < .05, 片側検定)．60℃を提示されたペアでも同様の傾向が見られ，前半では実際のペアと仮想ペアの間で相関係数の分布に差が見られなかったのに対し(D = 0.10, *n.s.*, 片側検定)，後半では実際のペアの方が仮想ペアに比べて相関が高い傾向があった(D = 0.26, p < .10, 片側検定)．ここから，対面するペアの生理反応は，相互作用を繰り返し経験することで，次第に類似するようになったと考えられる．

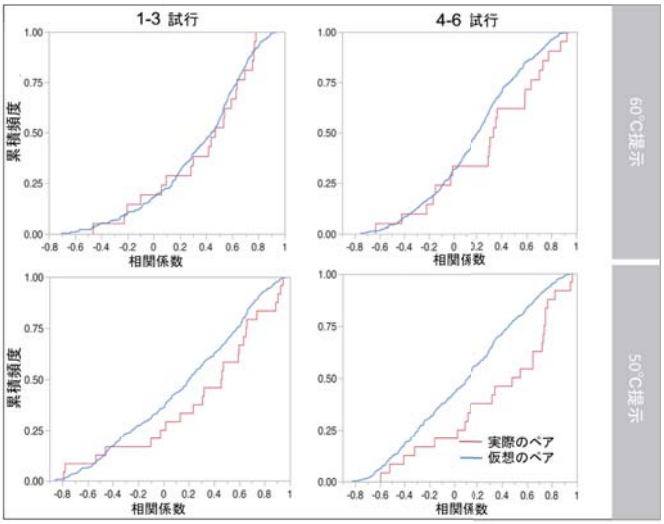


Figure 2. Same 条件の 50℃提示ペアと 60℃提示ペアの前半試行(1-3 trial)・後半試行(4-6 trial)における BVP の相関係数の累積頻度

4. 総合考察

本研究の結果から，他者との相互作用が情動反応を増幅させること，また相互作用を繰り返すことで次第に情動反応が似通ってくることが確認された．また，相手が自分とは異なる刺激を提示された場合に，相手の反応に合わせて自分の反応を変化させることが示唆された．

情動反応の類似性が後半の試行になるにつれて強まる傾向があったことから，情動反応がすぐさま無条件に個人間で連動するわけではなく，互いの刺激に対する反応の参照を繰り返す中で次第に似通っていく，謂わば，「情動反応の収束」が生じることが示唆された．総合すると，本研究で確認された，相互作用場面において二者間の情動反応が増幅・類似する現象の背景には，人々が相手の情動反応についてなんらかの視認可能なシグナルを(おそらく非意識的に)拾い合い，そのシグナルに基づいて自分自身の刺激に対する情動反応を相手の反応に合わせる形で調整するといったダイナミクスが働いていると考えられるかも知れない．しかもこれらの調整は不随意の自律神経系の生理反応のレベルで生じることが示された．しかし，人々が実際にどのようなシグナルを用いて相互に調整しているかは，今回の実験からは明らかにされていないため，今後の検討課題である．また，このような情動反応の連動を生じさせやすい環境要因や条件についても，慎重に特定していくことが望まれる．

また、痛みというプリミティブな感情にとどまらず，スポーツ観戦の興奮や映画やアートに対する感動のように，何か対象に対して集団である感情が共有される場面があるだろう．ウェルビーイングの観点からも，ストレス反応にとどまらず，ポジティブな感情についても同様の現象が確認されるかを調べることが望まれる．

参考文献

[1] Fowler, James H., and Nicholas A. Christakis. "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study." Bmj 337, 2008.

[2] Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental evidence of massive-scale emotional contagion through social networks." Proceedings of the National Academy of Sciences 111.24, 2014.

[3] Singer, Tania, et al. "Empathy for pain involves the affective but not sensory components of pain." Science 303.5661, 2004.

[4] Jackson, Philip L., Andrew N. Meltzoff, and Jean Decety. "How do we perceive the pain of others? A window into the neural processes involved in empathy." Neuroimage 24.3, 2005.

[5] Gallotti, Mattia, and Chris D. Frith. "Social cognition in the we-mode." Trends in cognitive sciences 17.4, 2013.

[6] Konvalinka, Ivana, et al. "Synchronized arousal between performers and related spectators in a fire-walking ritual." Proceedings of the National Academy of Sciences 108.20, 2011.

[7] Martin, Loren J., et al. "Reducing social stress elicits emotional contagion of pain in mouse and human strangers." Current Biology 25.3, 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

可視光通信による群集 Well-Being の測定の試み

Measurement trials of crowd's well-being using visible light communication

佐藤 尚[*1] 白木 善史[*1] 守谷 健弘[*1]
Takashi G Sato Yoshifumi Shiraki Takehiro Moriya

[*1] 日本電信電話(株) コミュニケーション科学基礎研究所
NTT Communication Science Research Laboratories, Nippon Telegraph and Telephone

Well-being (WB) is flourishing state of human beings which first can be categorized by size of the relationship. Thinking the effect of emotional contagion, there will be a WB state of the crowd. However, there are few studies that dealt with WB of the crowd. This study introduces our approach to the issue by showing parallel physiological measurement technique which uses optical camera communication. We show that crowd excitements can be reflected in the phase synchronization of the respiration. We also show our trial to measure and visualize the brain state of the crowd. The visualized feedback is considered to have an effect on the state of the WB if designed properly.

1. はじめに

Well-Being(WB)とは、ヒトが個人や社会性の中で自律的に生き生きとした心的状態を指し、その状態を維持していくことで本来の能力が発揮されより善く生きていくことを可能にするものと考えられている[1]。しかし WBの概念自体は広く、その要因も多岐にわたる。そこで WB を人と人のつながりの関係性の範囲の視点から分類、整理することが試みられている。実際に我々の研究プロジェクトでも自己完結的で他者に左右されない個人内の well から, 全く見知らぬ他人も含めた社会の中の well まで、様々な人の関係範囲の視点から研究を蓄積しようとしている。

その関係性に基づく分類のうち、本研究は群集の WB を測定し、見える化や技術による介入を行おうとする試みである。ここでいう群集とは組織的に編成されていない人の集団のことを指し、集団全体にわたる面識はないか希薄であり、集団として組織だった行動目標が設定されていないことを指す。例えば同じ催し物に参加した、電車に乗り合わせたといったことで形成される集団のことを指すとする。

群集における WB に関係する要因として大きいと考えられるのが群衆の外から与えられる外部刺激に対する個人内での反応のみならず、群集内部で生じる情動伝染[2]である。情動伝染は直接的なつながりの希薄な群集であっても、群集を構成する個人が群衆全体の傾向に直接的にさらされることによって生じると考えられる。この種の情動伝染は positive な影響のみならず、negative な情動でも伝染することはデモ隊が暴徒化する例などから容易に想像される。しかしながら、群集の WB に関する研究は少なく、ましてや生理学的なデータからアプローチしたり、フィードバックにより状態コントロールを試みる研究はほとんど見ない。

そこで、本研究では群衆の生理状態(呼吸、脳波)などを多人数から同時に測定し、コンサートなどで群集全体が盛り上がったときの生理状態の特徴を捉えることからはじめることとした。また、群集の状態を視覚的なフィードバックによって見える化する試みについても紹介する。

連絡先:佐藤尚, 日本電信電話株式会社コミュニケーション科学基礎研究所守谷特別研究室, 神奈川県厚木市森の里若宮3－1, 0462403629, takashi_goto_sato@ieee.org

2. 実験手法

2.1 集団測定技術

呼吸や脳波といった生理データを取得する場合、有線よりも無線の方が実験参加者の拘束が少なく良い。しかしながら、10人以上の集団から生理データを取得しようとする場合、既存の無線機器では混信が発生し、安定した測定が難しくなる。そこで我々のグループでは図 1に示すように、イメージセンサによる可視光通信を応用した測定機器を研究開発し[3]、応用範囲を拡げている[4]。これらのシステムでは実験参加者に取り付けた呼吸計や脳波計の出力を同じく取り付けられた LED からの光の点滅信号に変換し、遠方に設置した高速カメラにてその点滅信号を撮像することで各実験参加者の生体データを取得することを可能としている。伝送レートや測定人数は各種条件に依存するが、現時点で例えば 1kbps のデータを 40 人以上から測定することが可能である。また近赤外に波長をもつ LED を用いることで、人間の目には見えない様にして計測も可能である。以下、この測定技術を用いて行った 2 つの実験の詳細を説明する。

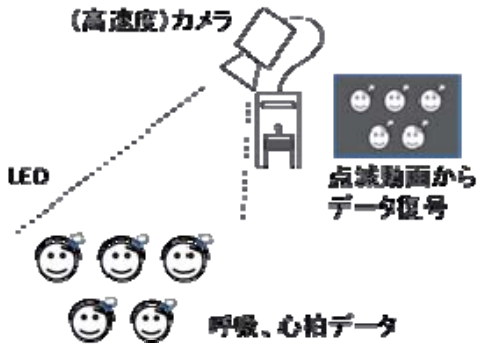


図 1 可視光通信による集団生理測定の構成図

2.2 ライブコンサートにおける群集からの呼吸測定

本実験ではプロによる 30 分程度の音楽ライブコンサートを鑑賞している実験参加者の呼吸を測定した。350 人程度が入場したライブコンサートの中で 24 人程度の実験参加者を一箇所に集めて呼吸を測定した。実験参加者は各演目が終わるたびにそのときの自分の情動を Self – Assessment Manikin (SAM) に

よって 5 段階評価した。ライブの演目はトークセッションを含めすべてで7演目あった。

2.3 ショーにおける群集の脳活動測定とフィードバック

ここでは群集の状態を見える化の試みとして、ニコニコ超会議のステージショーにて観客の脳状態を測定しその情報を観客にフィードバックした試みを紹介する。この実験ではステージ上で行われるジャンルの異なる演目を鑑賞する群集の脳波を計測し、推定された脳状態を可視光通信に伝送し、それを視覚的にフィードバック表示することを行った。また、いくつか行われた演目を得られた脳活動状態から算出された得点でランキングした。

使用した脳波計は NTT で開発し、脳波から脳状態を推定するアルゴリズムに Neurosky 社のチップを用いた。1Hz ごとに Attention, Meditation と呼ばれる集中やリラックスに関連するとされる指標が1から100の値で出力されている[5], [6]。

3. 実験結果

3.1 群集からの呼吸測定

本稿ではトークショーを除く5つの演目について紹介する。SAM によるライブコンサートの各演目に対する主観評定は no.2 の演目が Arousal, Valence 共に有意に高く、次が no.7 であった。

このときの、群集の呼吸の吸う、吐く、といった位相のそろい具合を評価した結果を図2に示す。解析の結果、群集の Arousal や Valence の評定が高いほど、呼吸の位相の同期が高いことがわかった。結果の解析方法や詳細なデータに関しては[7]を参照されたい。

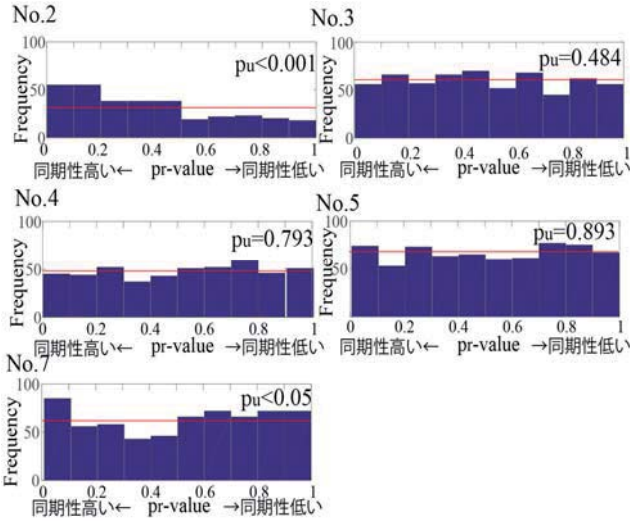


図 2. 観客の呼吸の位相の同期性を 0.5 秒ごとに計算して作成した頻度分布。観客の多くが同じ様な呼吸位相（吸う、吐く）の場合、左の頻度分布が多くなる。

3.2 群集の脳活動測定とフィードバック

図 3に群集から得た脳活動情報を個々のキャラクターの状態として視覚的にフィードバックした例を示す。この視覚表示では脳活動状態に応じて表情が変わると共に、耳の大きさも変えることによってより直感的に理解できるように工夫している。また、自分の立ち位置と画面上の位置を対応付けられるようにすることで、どのキャラクターが自分の状態なのかが把握できるようになっている。図 4に計測された脳活動の時間的変動とそのときのショーの内容を示している。



図 3. 各個人から測定された脳の活動状態は Attention, Meditation の値に基づき、左図のようにキャラクターの顔が割り当てられる。右図は群集に表示する画面例。

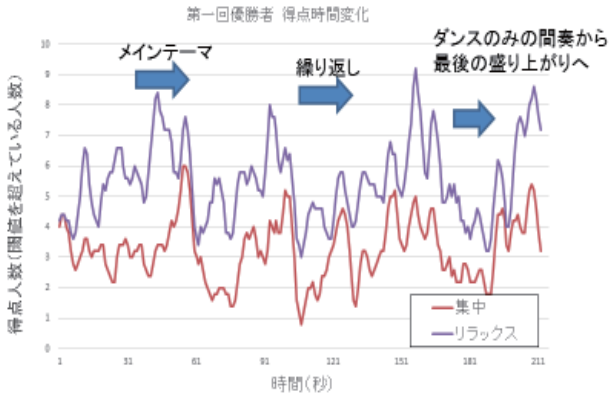


図 4. あるミュージックダンスで測定した脳活動の変化。Attention, Meditation 値が設定した閾値（65）を超えた人数の時間変化をプロットしたもの。

4. 考察

本稿では我々グループが取り組んでいる群集における WB な状態を知るために試みている生理測定と、群集の状態に介入することを目指したフィードバックの例を紹介した。群集からの生理測定を比較的容易に行えるようにするために、可視光通信技術を技術的に発展応用できたことが大きい。

ライブコンサート鑑賞時の群集の呼吸と主観評定を計測した実験では、Arousal, Valence の高い演目で実際に呼吸位相の同期性が高い傾向が見られることがわかった。聞きなれた音楽に対する呼吸の引き込みが強くなることが知られているため[8]、演目 No.2 がヒット曲であることを考えると、この同期性の高さは主として呈示された音楽が聴き慣れていることによるとも考えられる。しかし一方で、演目 No.7 は当日発表の新曲であり、単純に音楽刺激だけが要因とは言いにくい。我々としては同期性の中に、群集として集まった効果（隣の観客の呼吸パターンに影響される効果）が含まれていると考えており、群衆の WB を評価しうる指標になると考えているが、今後その影響を解明する実験を行っていく予定である。

群集の脳活動フィードバック実験では群集の WB を誘導する手段として視覚フィードバックを活用する試みを紹介した。視覚フィードバックにより、個々の観客は自分の脳活動が実際測定されているという現実感と、それをある程度変えられる有能感を個別のキャラクター表示で確認することができる。また、ポジティブ（と思われる）脳活動を群集として実現することでランキングを決めるという関係性も持たせている点で、WB に必要とされる要素をある程度カバーしていると考えられる。今後は個々の影響がどのように群集 WB に影響していくのかを評価していく必要があると考えている。

謝辞

本実験を行うにあたって、渡邊淳司（NTT）、中村睦美（NTT西日本)、及び関係各所に多大なる協力をいただいたことを改めて感謝する。

参考文献

[1] R. A. Calvo, D. Peters,（監訳）渡邊, チェン: *ウェルビーイングの設計論 -人がよりよく生きるための情報技術-*. 日経印刷, 2017.

[2] U. Hess, A. Fischer: Emotional Mimicry as Social Regulation, *Personal. Soc. Psychol. Rev.*, vol. 17, no. 2, pp. 142–157, 2013.

[3] G. P. Nava, et al.: Simultaneous acquisition of massive number, pp. 1–7, 2013.

[4] T. G. Sato, et al.: Heart Rate Measurement Based on Event Timing Coding Observed by Video Camera," *IEICE Trans. Commun.*, vol. E100.B, no. 6, pp. 926–931, 2017.

[5] E. Ščiglinskas, A. Vidugirienė: Investigation on Human Attentiveness to Video Clips Using Neurosky and LIRIS-ACCEDE Database, in *Lecture Notes in Computer Science*, vol. 8853, pp. 450–456, 2014.

[6] K. Crowley, et al.: Evaluating a brain-computer interface to categorise human emotional response," *Proc. - 10th IEEE Int. Conf. Adv. Learn. Technol. ICALT 2010*, pp. 276–278, 2010.

[7] T. G. Sato, et al.: Audience excitement reflected in respiratory phase synchronization," IEEE Int. Conf. SMC. pp. 2856–2860, 2017.

[8] T. G. Sato, et al.: Increase in the timing coincidence of a respiration event induced by listening repeatedly to the same music track, Acoustical Science and Technology, vo.33, 4. pp.255-261, 2012.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

健康促進に向けたサーカディアンリズムに着目した睡眠とストレスの分析

Analysis of the Relationship between Sleep Quality and Stress with a focus on Circadian Rhythm for Well-being

髙野諒[*1] Ryo Takano　長谷川智[*1] Satoshi Hasegawa　梅内祐太[*1] Yuta Umenai　辰巳嵩豊[*1] Takato Tatsumi　髙玉圭樹[*1] Keiki Takadama

志牟田亨[*2] Toru Shimuta　家邉徹[*2] Toru Yabe　松本英雄[*2] Hideo Matsumoto

[*1] 電気通信大学 The University of Electro-Communications

[*2] 村田製作所 Murata Manufacturing Company, Ltd.

The purpose of this study is to find novel knowledge to clarify the relationship between the sleep quality and the degree of the mental stress. For this purpose, we focus on not only these two indices (the quality of sleep and the degree of the mental stress), but also the human circadian rhythm as the new index for analysis. Through three types of data measured during the night-time sleep and during the day, we tried to inspect the usefulness of the human circadian rhythm for the index of the analysis. In this paper, data of these three indices were measured by the single subject experiment of about two weeks and analyzed comprehensively. In the analysis, we categorize good / middle / bad for each index every few days, and investigating the relationship between the three indices by summarizing the transition of the categories of the three indices. As a result, by comparing three types of data of ten-odd days in parallel, we obtained the following findings: (1) These three indices have been moving with a similar trend in units of days; (2) those trends coincide details from the simple diary written by the subject. As a result, by comparing three types of data of ten-odd days in parallel, these data were related to each other.

1. はじめに

睡眠の主な機能が身体的な疲労の回復であることは非常によく知られているが，睡眠には身体的疲労の回復と同様に精神的なストレスから回復させる効果も有している[Grant 2010]. このため，質の良い睡眠を日常的に維持することは，肉体的な健康のみならず精神的な健康を維持するためにも非常に重要な要素である. しかし，高いストレスを抱えていると質の良い睡眠を得られにくくなり睡眠の効能が阻害されるなど，睡眠とストレスの関係は複雑で未だ十分に明らかになっていない[Grant 2010]. このため，睡眠とストレスの関係を理解し質の良い睡眠を維持するための方策を獲得することは，身体的・精神的に良好な状態を保つための重要な課題の一つと言える. このためには，睡眠の質とストレスの度合い 2 種類のデータを必要とする. 特に，睡眠の質を測定したデータとして眠りの深さを離散値で表現した睡眠段階が用いられる. この睡眠段階の判定のために，Rechtschaffen& Kales の判定法(以下，R&K 法) が国際基準法として使用されているが，R&K 法は被験者へ拘束性の高い測定機器(脳波計・筋電計) の装着が必須であるため，被験者への負担が大きく長期間の測定は困難である. この問題を解決するためにマットレスセンサーを用いた無拘束型の睡眠段階推定法がある[渡邊 2001, 渡邊 2002, Harada 2016]. これらの手法はマットレスの下に敷いた偏平なエアマットレスセンサーにより被験者の生体データを計測し，その結果から睡眠段階を推定する. この手法により被験者への負担を最低限に留めて長期間の睡眠段階の測定が可能となり，ストレスの測定と合わせて長期間にわたる睡眠の質とストレスの度合い 2 種類のデータを獲得することが容易となる.

連絡先：髙野諒，電気通信大学，takano@cas.lab.uec.ac.jp

本研究では睡眠の質とストレスの大きさの 2 種類のデータの他に，睡眠に関係する要因としてサーカディアンリズムに焦点を当てる. サーカディアンリズムは，概日リズムとも呼ばれ約 24 時間周期で変動する生理現象であり，人間の「朝に起床し，夜に就寝する」という生理的なリズムを形成している. 例えば，海外旅行における時差ぼけとは，このサーカディアンリズムと渡航先の時間帯のズレにより発生することが言われている. このように. サーカディアンリズムは，睡眠の質を決定する重要な要因であることが知られている[Robert 2007]. また，サーカディアンリズムは睡眠だけでなく，起床後の日中の活動についても影響する. 本論文では，睡眠とストレスの関係性を分析する新しい視点として，このサーカディアンリズムに焦点を当て，サーディアンリズムと睡眠の質，ストレスの大きさの三点を包括的に分析することにより，睡眠とストレスの関係に関して新たな知見を獲得することを試みる. これは，睡眠段階とストレスの度合いの 2 種類のデータの測定時間帯のズレを解決するためである. 睡眠段階は基本的に夜中にしか測定されず，一方ストレスについては日中にしか測定することができない. このようにこれら 2 種類のデータが常に全く異なる時間帯でそれぞれ測定されていることが，これらの関係性を十分に明らかにできない一因であると考えられる. そこで，一日の生理的リズムであるサーカディアンリズムに注目することで，「夜間に測定された睡眠段階」と「日中に測定されたストレスの大きさ」という測定時間帯の異なる 2 つのデータを結びつけることができる. 以上のように，本論文は，夜中に測定された睡眠の質，日中に測定されたストレスの大きさ，昼夜を通じたサーカディアンリズムを測定し，それら 3 種類のデータ間の関係を明らかにすることを目指す. 本研究の最終的な目的は，これらの測定値から被験者の身体的・精神的な健康状態を日常的に把握し，これらの被験者の健康状態を良好な状態を保つための方策を提示する Well-being Computing システムを構築す

ることである．本論文は，このようなシステムを構築するために，その基礎となる 3 種類のデータ間の関係の解明を試みる．

2. 分析方法

本章では，睡眠段階，ストレスの大きさ，サーカディアンリズムという 3 種類のデータについて説明する．睡眠段階は夜中に測定され，ストレスの大きさは日中に測定され，サーカディアンリズムについては日中に測定した基礎体温から日夜を通して推定される． また，これらの 3 種類のデータを用いた包括的な解析の方法についても本章で説明する． 包括的な分析方法とは，これらの 3 種類のデータをそれぞれ 2~3 日単位の毎日の体調変化の傾向として捉え，それらの傾向の変化から 3 種類のデータの相互関係を明らかにする．

2.1 データの詳細

まず，解析に用いた 3 種類のデータの詳細をそれぞれ説明する． この 3 種類のデータは，睡眠の質，ストレスの大きさ，サーカディアンリズムのパターンの明瞭さを測定する指標である．

- **推定睡眠段階**

本論文では，睡眠の質の指標として，原田らによる睡眠段階をリアルタイムに推定する方法 (Real-time Sleep Stage Estimation: RSSE)を用いる[Harada 2016]．睡眠段階は，睡眠の深さを意味する離散値によって表される．この睡眠段階は，エアマットレスバイオセンサから得られた生体データの特徴に基づいて推定されるため，人体に脳波計などのデバイスを接続することなく算出することが可能なため，被験者への大きな影響を与えずに睡眠段階を推定することができる．睡眠段階は，以下の 6 つのタイプに分類される：(1)Wake：覚醒状態； (2) Rem：浅い眠りであり，一般的にはこのとき夢を見る；(3~6)NonRem1-4：深い眠り（眠りの深さは NonRem1 が浅く NonRem4 が最も深い）．本研究では，睡眠の質の指標として，Wake，Rem，NonRem1 の 6 段階の合計時間の割合を採用した．比率が小さいほど，深い睡眠が長く維持され，睡眠の質が良好であることが示される．

- **ストレス指標**

ストレスの大きさを示す指標として，LF / HF を用いる．LF / HF は，交感神経（LF）と副交感神経（HF）とのバランスを示す指標であり，値が高いほどストレスが高くなることを示す． 図 1 は，例として数時間ごとに測定した LF / HF の 2 日間の推移を示す．一般的に LF / HF の値が 2 より小さい場合，ストレスは小さく，2 以上の場合には，ストレスが高いことを示す．

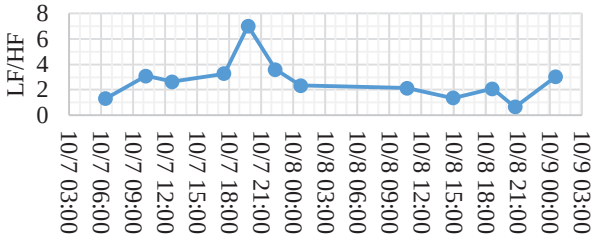


図 1. 2 日間の LF / HF の推移

- **サーカディアンリズム**

サーカディアンリズムのパターンの明瞭さを示す指標として，2 日間の基礎体温の標準偏差を用いる． 基礎体温はサーカディアンリズムに従って上下に変化することが知られている． これに基づいて，測定期間中の基礎体温の標準偏差を観察することにより，サーカディアンリズムが明確なパターンを有するかどうかを識別する． サーカディアンリズムの明確なパターンを有していることは，良好な睡眠および 2 日間の良好な活動に関連する指標となる．本指標の具体的な例を 3 つのグラフ (a) ～ (c) からなる図 2 に示す．図 2 (a-b) では点で表した 2 日間の実際の基礎体温と破線にて表した 6 次多項式近似にて基礎体温から得られるサーカディアンリズムを示す．これらのグラフにおいて，破線は，2 日間のサーカディアンリズムのパターンとして表現することができる．図 3 (a) は，ダッシュラインの 2 つのセット「深夜の谷」と「午後のピーク」の高さの差が大きく，パターンがはっきりと見えるため，サーカディアンリズムの良好なパターンである．一方，図 3 (b) は，「谷と峰」の高さの差が小さいため，模様がはっきりと見えず，そのパターンは悪い．これらのグラフから判断されたパターンの良否を数値データのみで評価するために，基本体温の標準偏差を算出する．図 3 (c) から，図 3 (a) および (b) の基本体温の標準偏差が示されている．図 3 (c) から分かるように，標準偏差は 9/26～9/27 図 3 (a) の良好なパターンは，図 3 (b) の不良パターンを示す 10/1 - 10/2 の数値よりも高い数値を示している．このようにして，基底体温の標準偏差の大きさに基づいてサーカディアンリズムの質を判別することができる．

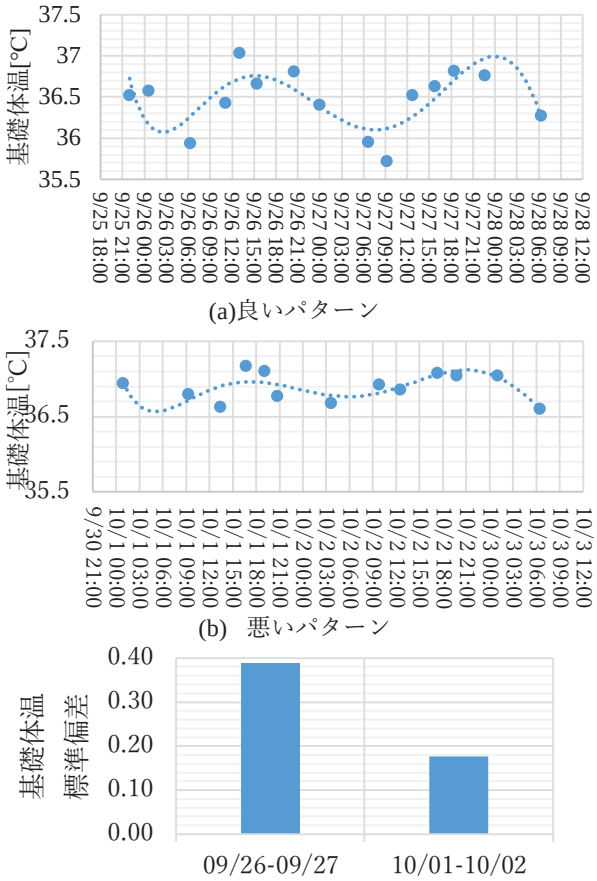


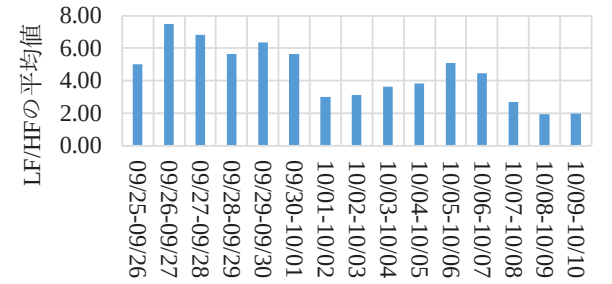
図 2. 基礎体温によるサーカディアンリズムの推定

3. 包括的分析

この分析は，約 2 週間の睡眠段階，LF / HF，基礎体温の標準偏差の 3 種類のデータから，睡眠の質とストレスとの因果関係を明らかにすることを目的する．本章では，これら 3 種類のデータを使用した包括的な分析について説明する．このためには，推定睡眠段階のデータとして夜間の浅い睡眠の割合を準備する．他の 2 種類のデータについては，2 日ごとの平均および標準偏差を数時間ごとに測定された値で計算する．このように 2 日間の測定値を用いるのは，2 日間の睡眠中のデータの前後の数値を観察するためである．

まず，推定睡眠段階データ以外の 2 つのデータについて，視覚的解析の容易性を向上させるために数値標準化を行う．数値の標準化は，各数値をそれ自身の平均で減算し，それ自身の標準偏差で割った値として計算される．その結果，データセットの平均値と標準偏差はそれぞれ 0 と 1 になる．また，データの値が 0 に近いほど平均値に近い値をとる．図 3(a)は，約 2 週間の標準化前の LF / HF の平均年齢値を示し，図 3(b)は，LF / HF の標準化された平均値を示す．図 3(a)と図 3(b)を比較すると，そのストレスの大きさを区別することが可能である．

次に，3 種類のデータを分析する方法について説明する．それぞれの因果関係を理解するために，3 種類のデータを同じ時間軸上で数日単位の良否に応じて分類する．これにより，被験者による睡眠の質，ストレスの大きさ，サーカディアンリズムと被験者による日誌との関係を考察する．



(a)元の測定値



(b) 標準化後の値

図 3. LF/HF の 2 週間の推移と数値変換

4. 被験者実験

本章では，サーカディアンリズムに焦点を当てて，睡眠の質とストレスの程度との関係に関する新しい知識を得ることができるかどうかの第一歩として，被験者実験により得られた実際の測定データを用いて，前章にて説明した分析方法を実施する．

4.1 利用データの詳細

使用データは，2016 年 9 月 25 日から 2016 年 10 月 10 日までの 15 日間に測定された成人女性の 3 種類の測定値である．推定睡眠段階は，1 泊分のデータによって計算される．睡眠段階推定のための測定機器として，EMFIT 社のマットセンサーを，LF / HF を測定するために疲労科学研究所のストレス測定センサー(VM 302)を，基礎体温については一般的に入手可能な基礎体温計を使用した．また，LF / HF と基礎体温の測定は 1 日に数時間ごとに測定された．測定終了後，LH / HF の平均値および基礎体温の標準偏差を 2 日ごとに計算した．さらに，これらの LF / HF 値および基礎体温は，平均が 0，標準偏差が 1 になるように標準化された値で表される．さらに，被験者に簡単な日記を記録するように依頼した．その日誌の内容の要約を表 1 に示す．

4.2 計測結果

実験で得られた測定結果を以下に示す．推定睡眠段階を図 4 に示す．また，LF / HF の標準化した数値は図 3(b)に示している．基本体温の標準偏差を図 5 に示す．ここで図 4 の推定睡眠段階の割合のグラフについては，次の分析のために，25 パーセンタイル未満および 75 パーセンタイルより上の値は色 それぞれ青と赤に色分けする．

表 1. 被験者による日誌の要約

9/25	記述無し
9/26	会議に参加 ジムでのエクササイズ
9/27	昼食を食べず
9/28	定時退社
9/29	セミナーに参加 飲み会
9/30	出張 重要なプレゼンテーション
10/1	[休日] 出張先の町を散策
10/2	[休日] 友人とおしゃべり
10/3	終日デスクワーク
10/4	長い会議 ジムでのエクササイズ
10/5	早めに帰宅
10/6	ジムでのエクササイズ
10/7	忙しい一日
10/8	[休日] 家でリラックス
10/9	[休日] 買い物

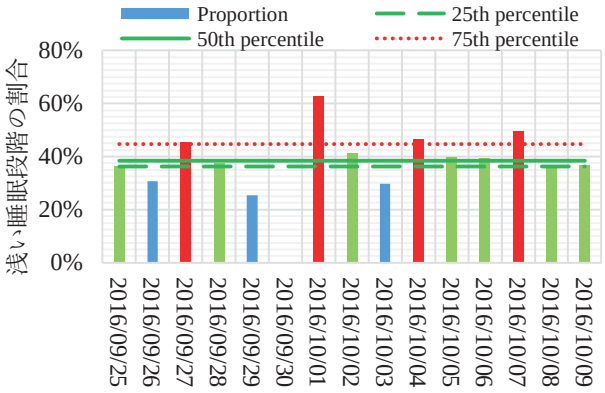


図 4. 睡眠段階の推定結果



図 5. 基礎体温の標準偏差の 2 週間の推移

4.3 分析

本節では，前章で説明したように，各データを良し悪しにより分類し，その結果の関係性を分析する．図 7 に計算された各指

標の数値と，その良さに応じて着色されたグラフを示す．図 7 は 3 つの棒グラフで構成され，第 1 列は睡眠段階を，第 2 列は LF / HF を，第 3 列は基礎体温の標準偏差を示す．第 1 列のグラフの各日のバーの色は，青，緑，赤の 3 種類の睡眠段階の良し悪し(良，中，悪)でそれぞれの睡眠を分類する．図 7 の各段の背景色は，数日間のスパンで見た各指標の良し悪しを示している．次に，上段の睡眠段階と下段の基底体温の標準偏差を比較すると，背景の色調の遷移は 1 日のずれと一致する．次に，中段の LF / HF と下段の基底体温の標準偏差とを比較すると，赤と青の背景色が逆転する．これらの結果は，図 6 に示す 3 つに分類した結果と一致する．グラフの上段の睡眠段階および下段の基礎体温の標準偏差によれば，緑色の背景はお互いにほとんど同じである．さらに，これらの分析結果と表1に示した被験者の日誌との比較から，9/30 の重要なプレゼンテーションのために測定開始から数日のうちはストレスを感じたが，交渉後にストレスが軽減されたことが見て取れる．さらに，この数日間のストレスから回復するためには，サーカディアンリズムのパターンも明確となり，深い眠りを促したと考えられる．

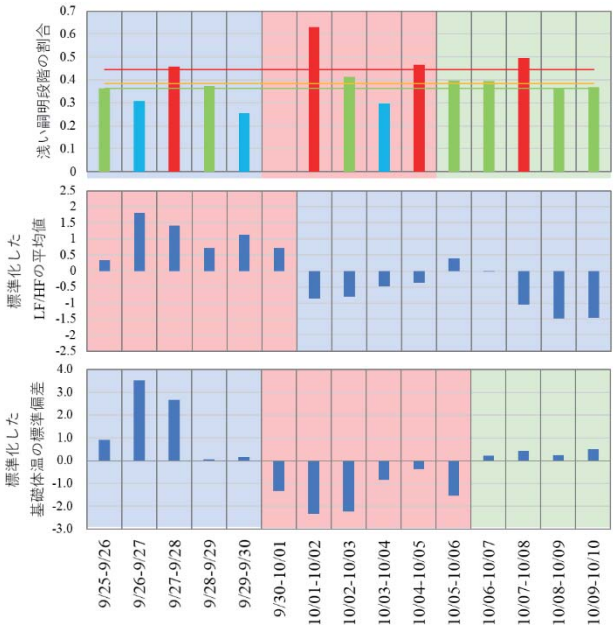


図 7. 包括的分析の結果

4.4 Well-being Computing システムを構築に向けて

本研究の最終的な目標は，睡眠やストレスについてのデータを被験者に負担をかけずに自動的に測定・分析することにより被験者の身体的・精神的な健康を促進するための方策として日中の行動(散歩，体操等)や睡眠時間などを提案する Well-being Computing システムの構築である．この目的のためにはそれらのデータ間の関係性や実際の被験者の行動とデータとの関係性を網羅的に把握することが必要不可欠である．この目的のための分析として，前節の結果から 3 種類のデータ(睡眠段階・ストレスの大きさ・サーカディアンリズム)の間に正負のそれぞれの相関関係が有しており，それらは被験者による日誌から得られる情報と非常に密接な関係にあることが明らかになった．これは目的のシステムを構築する上で非常に有用な知見であり，これらの知見を更に蓄積することで，身体的・精神的な健康を促進するための方策を獲得するためのメカニズムを構築することが期待できる．

おわりに

本論文は，睡眠とストレスについてのデータから被験者の身体的・精神的な健康状態を日常的に把握し，これらの被験者の健康状態を良好な状態を保つための方策を提示するシステムを構築するために，その基礎となる睡眠とストレスの関係性を明らかにすることを目的としている．この目的のために，被験者のサーカディアンリズムを「夜中に測定された睡眠段階」と「日中に測定されたストレスの大きさ」という測定時間帯の異なる 2 つのデータを結びつけるための新たな指標として焦点を当て，これら 3 種類のデータ(睡眠段階・ストレスの大きさ・サーカディアンリズム)の間の関係を明らかにするための分析方法を提案した．この観点からの分析方法の有効性を確認するため，15 日間の被験者実験を行った．この実験では，被験者の睡眠段階，基礎体温，HF / HF をそれぞれ測定した．これらから，睡眠段階，ストレス度，サーカディアンリズムパターンの良し悪しについての指標を算出した．更に，これら 3 つの指標を包括的に分析することにより，以下のような知見が得られた．(1)これらの 3 つの指標は，同様の傾向，もしくは真逆の傾向で推移しており互いに影響を及ぼしていること．(2)それらの傾向は，被験者によって書かれた日誌の内容と一致していること．これらの結果から，2 週間ほどの比較的長い期間の 3 種類のデータを並行して比較することにより，これらのデータの相互関係をある程度理解することが可能であることが明らかになった．今後の課題として，より多くの被験者実験により，この分析方法の様々な被験者に対する適用性を検証するとともに，3 種類のデータ(睡眠段階・ストレスの大きさ・サーカディアンリズム)間と被験者の行動についての関係に対する更なる知見の獲得を目指す．

参考文献

[Grant 2010] Grant, B., 2010. Sleep: an important factor in stress-health models. Stress and Health. Volume 26, Issue 3: 204–214

[Robert 2007] Robert, L. S.; Dennis, A.; R. Robert, A.; Mary, A. C.; Kenneth, P. W., Jr; Michael V. V. and Irina V., Z. 2007. Circadian Rhythm Sleep Disorders: Part I, Basic Principles, Shift Work and Jet Lag Disorders. Sleep, Volume 30, Issue 11: 1460-1483

[Robert 2007] Robert, L. S.; Dennis, A.; R. Robert, A.; Mary, A. C.; Kenneth, P. W., Jr; Michael V. V. and Irina V., Z. 2007. Circadian Rhythm Sleep Disorders: Part II, Advanced Sleep Phase Disorder, Delayed Sleep Phase Disorder, Free-Running Disorder, and Irregular Sleep-Wake Rhythm. Sleep, Volume 30, Issue 11: 1484-1501

[渡邊 2001] 渡邊崇士, 渡辺嘉二郎. 無拘束エアマットレス型生体センサによる睡眠段階の推定. 計測自動制御学会論文集, 2001, 37.9: 821-828.

[渡邊 2002] 渡邊崇士, 渡辺嘉二郎. 就寝時無拘束計測生体データによる睡眠段階の推定. 計測自動制御学会論文集, 2002, 38.7: 581-589.

[Harada 2016] Harada, T.; Uwano, F.; Komine, T.; Tajima, Y.; Kawashima, T; Morishima, M.; and Takadama, K, 2016. Real-time Sleep Stage Estimation from Biological Data with Trigonometric Function Re-gression Model. In the AAAI 2016 Spring Symposia, Well-Being Computing: AI Meets Health and Happiness Science, AAAI The Association for the Advancement of Artificial Intelligence), pp. 348-353.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Investigation of appropriate fNIRS feature to evaluate cognitive load

Kota Suzuki[*1] Tatsuya Suzuki[*1] Sotaro Shimada[*1,2] Atsumichi Tachibana[*3] Yumie Ono[*1,2]

[*1]Graduate School of Science and Technology, Meiji University
[*2]Department of Electronics and Bioinformatics, School of Science and Technology, Meiji University
[*3]Department of Histology and Neurobiology, Dokkyo Medical University School of Medicine

Functional near-infrared spectroscopy (fNIRS) allows researchers to noninvasively monitor cortical activity in a naturalistic environment, which is an advantage in a field of human well-being research to measure cognitive load in a daily-life situation. We investigated the appropriate features of fNIRS signals that best indicates the amount of cognitive load required for performing the multisensory-motor cognitive task. The features tested were (1) maximum amplitude relative to the baseline (MAX) and (2) cumulated amplitude (area under the curve (AUC)) of the normalized average fNIRS signals, and (3) beta value obtained by generalized linear modelling of the raw fNIRS signal using a block design (beta). Oxy-hemoglobin fNIRS features of AUC and beta showed better correspondence to the behavioral measure of cognitive load relative to that of MAX, suggesting that these two indices could be the suitable measure to evaluate cognitive load from fNIRS signals.

1. INTRODUCTION

Functional near-infrared spectroscopy (fNIRS) can measure cortical hemodynamic responses that reflect cognitive functions. The advantage of fNIRS is less constraint in a body posture and movement that enables noninvasive measurement of cortical activity in a state close to natural, daily-life environment. Utilizing these benefits, we have proposed to apply the fNIRS technology to evaluate the intensity of physical pain [Matsuda 2017a 2017b] and discomfort [Ono 2016], and cognitive load [Ono 2015] [Azman 2017].

Although the cortical origin of hemodynamic signal in fNIRS is well confirmed, the fNIRS signal is susceptible to external and/or internal artifacts such as probe displacement and systemic responses [Tak 2014]. Also the majority of commercial fNIRS systems uses continuous-wave NIRS, from which neither the absolute concentrations of oxy- and deoxy- hemoglobin (Oxy-Hb and Deoxy-Hb) nor the optical path length is obtained. Therefore the measured raw signals should appropriately be processed to determine the neuronal signals that represent the intensity of cognitive activity.

There are various fNIRS signal features that have been proposed to evaluate the intensity of cortical cognitive activity. Early studies used the concentration changes of hemoglobin oxygenation during the task period, and more sophisticated statistical approach of generalized linear model (GLM) was later introduced from its relevant field of functional magnetic resonance imaging (fMRI) [Tak 2014]. However, which feature is most appropriate for fNIRS analysis has not comparatively studied yet. We therefore investigated the most appropriate feature of fNIRS signals that represent cognitive load using the fNIRS data during the multisensory-motor cognitive task [Suzuki 2018]. We used dance video game (DVG) as a cognitive task since the cortical area required to integrate audio, visual, and proprioceptive information is well understood [Tachibana 2011][Ono 2014] and the cognitive load to perform the task can be quantitatively evaluated by the behavioral performance such as timing and response accuracies. We tested the fNIRS features of (1) maximum amplitude relative to the baseline (MAX) of the normalized average fNIRS signals [Matsuda 2017a 2017b], (2) cumulated amplitude (area under the curve (AUC)) of the normalized average fNIRS signals [Ono 2016] [Azman 2017], and (3) beta value obtained by GLM of the raw fNIRS signals using a block design (beta) [Ono 2015][Tak 2014]. The features were obtained in three different conditions with varied complexity of the task and compared with the behavioral performance.

Contact: Kota Suzuki, Meiji University, 044-934-7302, ce171035@meiji.ac.jp

2. MATERIAL AND METHOD

2.1 Participants

We measured twelve healthy male adults, aged 21-25 years (mean and standard error 22.7 ± 0.3 years, all right-handed). The study was approved by the ethics committee of Meiji University, and all participants gave written informed consent to participate.

2.2 fNIRS measurement

We used OMM-3000 fNIRS systems (Shimadzu Co. Ltd., Kyoto, Japan). Optodes were arranged over the frontotemporal regions of both hemispheres of the participant with inter-optode distance of 3 cm for each source-detector pair. Oxy- and deoxy-hemoglobin concentration changes were measured with a sampling rate of 7.9 Hz. For the current study, we used fNIRS signals obtained from the left superior/middle temporal gyrus (S/MTG; Figure 1). S/MTG is involved in the comprehension of the rhythm [Liégeois-Chauvel 1998] and its activity duration correlates with the temporal accuracy of the responses in DVG play [Ono 2014]. Previous study has also shown that S/MTG activity depends on the cognitive aspect of the motor complexity but not on the exercise intensity [Tachibana 2011]. We used a 3D digitizer (PATRIOT, Polhemus, Colchester, VT) and obtained coordinates of all optode positions and the anatomical landmark positions (nasion, inion, auricles and Cz) of each participant immediately before data collection. Individual channel positions were normalized to the standard MNI coordinates using NIRS-SPM [Ye 2009] to confirm their anatomical location.

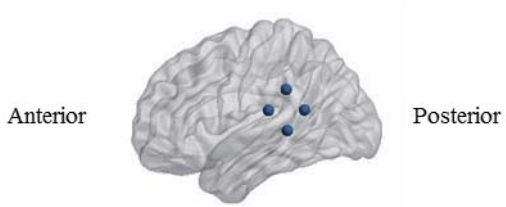


Figure 1. fNIRS channel layout. The mean channel positions across all participants of the current region of interest (S/MTG) were shown as blue dots on the normalized brain. Channel positions were visualized with the BrainNet Viewer (http://www.nitrc.org/projects/bnv/) [Xia 2013].

2.3 Experiment design

We prepared a DVG task (Figure 2) using an open source software StepMania software version 3.9. DVG is similar to the commercial game Dance Dance Revolution (Konami Corporation, Tokyo, Japan). In this experiment, we used 4 arrows, which were up, down, right and left. The participant had to press arrow buttons on the game controller with their finger or on the dance pad with their foot at the right timing which are indicated as visual cues (arrows) on the screen and by the background music. For hand-played condition, participants were instructed not to hold the game controller in hands but press the button of the controller that was fixed on the table.

Figure 2 illustrates the experimental design. All participants performed DVG three times with different appendage conditions (dominant hand (DF), foot (F) and non-dominant hand (NH)) in a random order. Generally, the foot-played condition is more difficult than hand-played condition since participants require more cognitive load to make foot movement to the appropriate direction while maintaining their trunk position, which is much larger and unfamiliar movement compared to hand movement. However our region of interest was located in the multisensory integration area of S/MTG, not in the primary motor area, therefore the recorded fNIRS signals is hypothesized to represent neuronal activities required for cognitive process of multimodal integration.

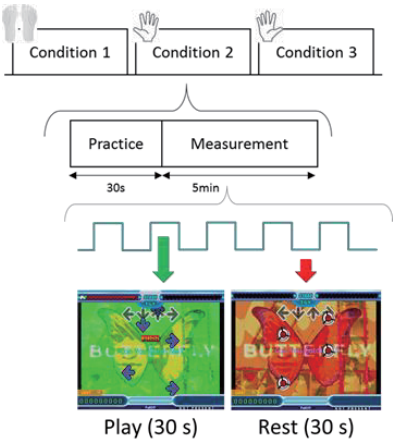


Figure 2. Experiment design. All participants performed DVG three times with different appendage conditions (dominant hand (DF), foot (F) and non-dominant hand (NH)) in a random order. A single measurement is consisted of 30 s of playing DVG and 30 s of rest alternately 5 times.

The fNIRS measurement of each condition was always preceded by 30 s of practice with the corresponding appendage. A single measurement is consisted of 30 s of playing DVG and 30 s of rest alternately 5 times. A song entitled 'Butterfly' (recorded by SMILE.dk) used as the background music with total of 240 visual cues. The performance of DVG was determined by the number of correct arrow cues that were responded in accurate timing (those being responded within ± 22.5 ms from the exact timing). All participants answered a questionnaire on which appendage was the most difficult in this experiment after the fNIRS recording.

2.4 Data analysis

Oxy-Hb and Deoxy-Hb data in the left S/MTG were analyzed to determine the features of MAX, AUC and beta as neuronal activity indices. One among 4 channels, which was localized most closely at S/MTG, was selected for each participant.

The MAX and AUC were obtained as follows. Change in hemoglobin concentration signals were first averaged with the task onset and smoothed using the moving average filter (25 point, 5 times). The averaged data was then baseline corrected so that the signal amplitude of task onset was set to zero. Lastly, the averaged signals were normalized by dividing the averaged data by the standard deviation of those during the 10 s before the task onset. The averaged hemodynamic responses were manually examined for systemic and/or facial muscle artifact [Schecklmann 2017][Zhang 2016]. The data were regarded as contaminated with systemic and/or muscle artifact and removed from further analysis if both Oxy-Hb and Deoxy-Hb responses continuously deflected to the same direction (either increasing or decreasing) for 10 s from the beginning of the task. Two datasets were excluded with this criterion.

To calculate beta value from GLM analysis we used raw fNIRS data. The beta value is obtained by solving the following equation of GLM by minimizing the square of error $\varepsilon(t)$:

$$y(t) = \beta \cdot x(t) + \varepsilon(t)$$

where $y(t)$ is raw fNIRS data and $x(t)$ is a regressor matrix consisted of vectors of regressor functions. The regressor matrix was consisted of a block-design model function derived by hemodynamic response function (HRF) [Tak 2014], baseline offset, and baseline drift components. The coefficient for the block-design model in β was determined as the beta value representing the intensity of the fNIRS signal that changed along with task. The model function was derived using spm_hrf.m function implemented in SPM8 toolbox and obtained the beta value at each condition of each subject.

To investigate the difference in hemodynamic activity among different appendages for motor output, we first compared the fNIRS features across DH, F, and NH by Friedman test as Shapiro-Wilk test failed to confirm the normality of the data. Post-hoc multiple comparison was performed by Wilcoxon signed-rank test with Bonferroni correction. Behavioral performance was compared between conditions using repeated measures analysis of variance (ANOVA) with post-hoc multiple comparison using paired t-test with Bonferroni correction due to the normality of the data.

3. RESULTS

Table 1 shows the mean behavioral performance of DVG task among three appendage conditions. There was a significant difference in the performance among appendages (repeated measures ANOVA: $F = 22.00$, $p < 0.01$). DVG play with foot resulted in the worst performance compared to dominant and non-dominant hands (paired t-test with Bonferroni correction: F and DH: $p < 0.01$, F and NH: $p < 0.01$) but there was no difference in the performance between DH and NH ($p = 0.50$). The performance scores were in accordance with the result of the questionnaire in which all participants answered that DVG played with foot was the most difficult.

Figure 3 shows the typical raw waveform of fNIRS measurement. Model functions of GLM (black solid and dotted lines) well approximated the raw data (P value for linear approximation was below 0.001 for all data). Figure 4 shows the typical waveform of event-related average hemodynamic responses. The area shaded with pink is AUC of Oxy-Hb activity. The time-course of both raw and averaged hemodynamic signals showed task-related increases in Oxy-Hb and decreases in Deoxy-Hb, showing a cortical activity pattern [Scholkmann 2013]. In each data, the opposite polarity of Oxy-Hb and Deoxy-Hb deflection was confirmed. Figure 5 show the results of mean MAX, AUC and beta values of different conditions. Although the behavioral performance showed statistically significant differences between hand- and foot- played conditions, the mean MAX value of Oxy-Hb data failed to capture the significant difference between these conditions (Friedman test: $\chi^2 = 5.60$, $p = 0.061$). On the other hand, the mean AUC and beta values of Oxy-Hb data showed significantly different values among appendages (Friedman test: $\chi^2 = 7.80$ and 9.50, $p = 0.020$ and 0.009 in AUC and beta, respectively).

Post hoc multiple comparison was performed for AUC and beta features by Wilcoxon signed-rank test with Bonferroni correction. Statistically significant difference was confirmed between foot-played and dominant-hand conditions in both AUC and beta features (AUC: $p = 0.021$, beta: $p = 0.009$). The temporal activity was larger when DVG played by foot relative to hand.

We did not find any statistically significant conditional differences in any of these fNIRS features using the Deoxy-Hb signals.

Table 1. Mean behavioral performance of participants in hand- and foot- played DVG task. The performance was represented as percentage of temporally accurate responses. The asterisk indicates that the value is significantly larger than foot-played condition ($p < 0.05$).

Appendage condition	Percentage of temporally accurate responses
DH	49.4 *
NH	46.6 *
F	36.1

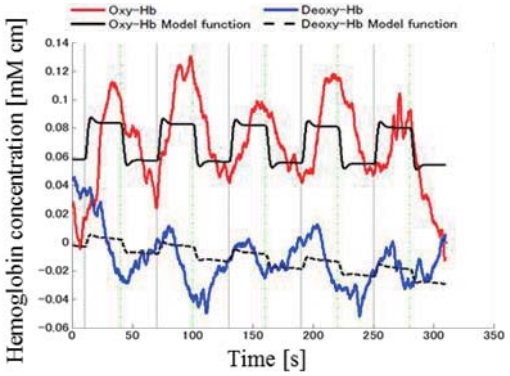


Figure 3. The typical raw fNIRS waveform (Participant 10 on foot-played condition). The dotted vertical lines indicate task start (black) and end (green).

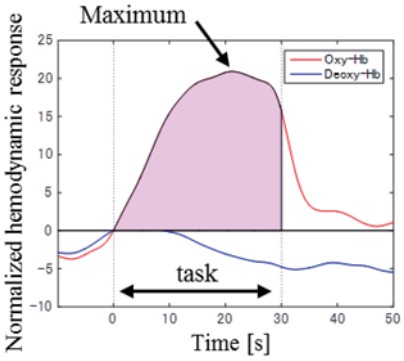


Figure 4. The typical event-related average hemodynamic responses (Participant 5 on foot-played condition). The dotted black vertical lines indicate task period. The area painted with pink shows AUC.

4. DISCUSSION

We investigated three types of fNIRS features during cognitive task to investigate the appropriate signal feature representing the amount of cognitive load. We used DVG as a cognitive task because the behavioral score of the DVG can be approximate the cognitive load required to perform the task.

The behavioral performance showed that DVG played with foot is more difficult than that played with hand. This indicates that the DVG played with foot need a higher cognitive load than with hand.

The mean AUC and beta values with foot-played condition were significantly larger than those with dominant hand-played condition. The increased fNIRS activity corresponds to the requirement of more cognitive activity to perform the task, which resulted in the lower behavioral score. However, there was no significant difference between hand and foot conditions in fNIRS features when we used mean MAX values. Previous study showed that S/MTG is necessary for motor output at the correct timing that required audiovisual integration [Suzuki 2017]. Another research reported that MTG becomes active when a person with limited ability of beat perception tried to perceive rhythm [Grahn 2009]. It is suggested that both AUC and beta values obtained from GLM analysis could be used for evaluating cognitive load. Although more data analysis using different data set is required, the higher statistical confidence in beta values suggests the superior ability to detect cognitive load over AUC.

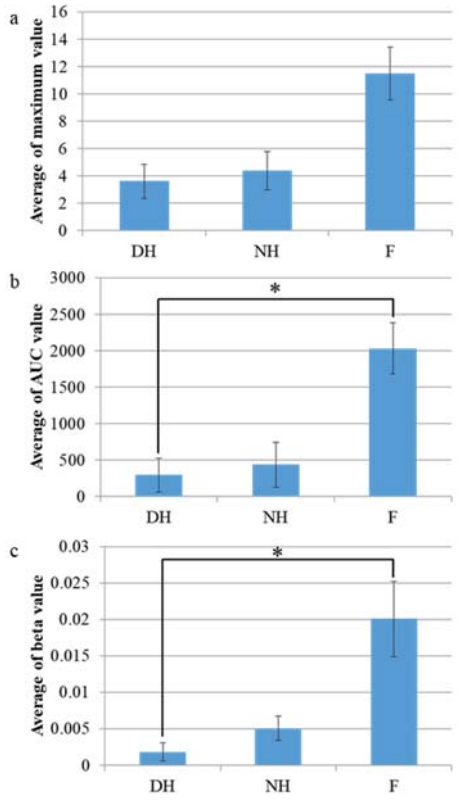


Figure 5. Average MAX (a), AUC (b), and beta (c) values in S/MTG. Error bars show standard error. Asterisk shows statistically significant difference ($p < 0.05$, Wilcoxon signed-rank test with Bonferroni correction).

In the current analysis, Deoxy-Hb signals failed to show significant difference between conditions in all analysis methods. Although its lower signal-to-noise ratio, there are reports showing the superiority of Deoxy-Hb signals in the sense of robustness of the signals against global systemic responses [Zhang 2016]. Since we observed a slight time delay in the Deoxy-Hb waveforms relative to Oxy-Hb ones (Figure 3), there may be a necessity to change the time window and/or modify the model function to be analyzed which matches with Deoxy-Hb signals to obtain meaningful features.

5. ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI (Grant Number JP 16K01565).

REFERENCE

[Azman 2017] Azman N, Suzuki K, Suzuki T, Ono Y, Edanaka Y, Kunieda F, Nakata M, Watanabe K: Effect of dance video game training on elderly's cognitive function. Transactions of Japanese Society for Medical and Biological Engineering, 55 (Proc), pp.526-529 2017

[Grahn 2009] Grahn JA., McAuley JD: Neural bases of individual differences in beat perception, NeuroImage, 1894-1903 2009

[Liégeois-Chauvel 1998] C, Peretz I, Babaï M, Laguitton V, Chauvel P: Contribution of different cortical areas in the temporal lobes to music processing. Brain, 121: 1853-1867 1998

[Matsuda 2017a] Matsuda Y, Suzuki T, Onozuka M, Ono Y: Investigation of measurement and visualization technology to convey "pain" sensation of patient to, The 31st Annual Conference of the Japanese Society for Artificial Intelligence, 2017

[Matsuda 2017b] Matsuda Y, Suzuki T, Onozuka M, Ono Y: Functional brain imaging to assess the efficacy of manipulative therapy for pain relief, Trans. Japanese Soc. Med. Biol. Eng., 55(1) 1-8 2017

[Ono 2014] Ono Y, Nomoto Y. Tanaka S, Sato K, Shimada S,Tachibana A, Bronner S, Noah JA: Frontotemporal oxyhemoglobin dynamics predict performance accuracy of dance simulation gameplay: Temporal characteristics of top-down and bottom-up cortical activities, NeuroImage.,vol.85, pt.1, pp.461-470 2014

[Ono 2015] Ono Y, Noah JA, Zhang X, Nomoto Y, Suzuki T, Shimada S, Tachibana A, Bronner S, Hirsch J: Motor learning and modulation of prefrontal cortex: an fNIRS assessment. J Neural Eng., 12(6) 066004 2015

[Ono 2016] Ono Y, Ishikawa Y, Munakata M., Shibuya T, Shimada A, Miyachi H, Wake H, Tamaki K:Diagnosis of occlusal dysesthesia utilizing prefrontal hemodynamic activity with slight occlusal interference, Clinical and Experimental Dental Research, 2(2), 129-135. 2016

[Schecklmann 2017] Schecklmann M, Mann A, Langguth B, Ehlis, A. C, Fallgatter A. J., & Haeussinger, F. B: The Temporal Muscle of the Head Can Cause Artifacts in Optical Imaging Studies with Functional Near-Infrared Spectroscopy. Frontiers in Human Neuroscience, 11, 456 2017

[Scholkmann 2013] Scholkmann F, Gerber U, Wolf M, Wolf U: End-tidal CO 2: an important parameter for a correct interpretation in functional brain studies using speech tasks. Neuroimage, 66, 71-79 2013.

[Suzuki 2017] Suzuki K, Suzuki T, Ono Y: Effect of middle-temporal tDCS stimulation on dance game exercise performance. Transactions of Japanese Society for Medical and Biological Engineering, 55 (Proc), pp.503-505 2017

[Suzuki 2018] Suzuki K, Ono Y, Shimada S, Tachibana A, Noah JA: Change in cognitive process during dance video game play with different appendages for motor output, Proc. SPIE 10481, Neural Imaging and Sensing 2018

[Tachibana 2011] Tachibana A, Noah JA, Bronner S, Ono Y, Onozuka M: Parietal and temporal activity during a multimodal dance video game: an fNIRS study. Neuroscience Letters 125-130 2011

[Tak 2014] Tak S, Ye JC: Statistical analysis of fNIRS data: A comprehensive review, NeuroImage 85 72–91 2014

[Xia 2013] Xia M, Wang J, He Y: BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. PLoS ONE 8: e68910 2013

[Ye 2009] Ye, J.C., Tak, S., Jang, K.E., Jung, J.W., Jang, J.D: NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. NeuroImage 44, 428-447 2009

[Zhang 2016] Zhang X, Noah JA, Hirsch J: Separation of the global and local components in functional nearinfrared spectroscopy signals using principal component spatial filtering. Neurophotonics 015004, 2016.

Oral presentation | Organized Session | [Organized Session] OS-4

[2F3-OS-4b] [Organized Session] OS-4

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room F (4F Garreria)

[2F3-OS-4b-01] Direction to quantifying well-being: Visualizing mindfulness based on functional brain network

○Satoru Hiwa[1], Tomoyuki Hiroyasu[1] （1. Doshisha University）

3:20 PM - 3:40 PM

[2F3-OS-4b-02] Estimation of low rank matrix between two correlation matrices based on clustering

○Kensuke Tanioka[1], Satoru Hiwa[2], Tomoyuki Hiroyasu[2], Hiroshi Yadohisa[2] （1. Wakayama Medical University, 2. Doshisha University）

3:40 PM - 4:00 PM

[2F3-OS-4b-03] Meditation Changes the Time Required for Attention Shifting from Mind-Wandering

○Issaku Kawashima[1,2], Toru Takahashi[1,3], Masahiro Fujino[3,4], Hiroaki Kumano[2] （1. Graduate School of Human Sciences, Waseda Univ., 2. Faculty of Human Sciences, Waseda Univ., 3. Japan Society for the Promotion of Science, 4. Graduate School of Education, Kyoto Univ.）

4:00 PM - 4:20 PM

[2F3-OS-4b-04] A study on Detection of Symptoms of Dementia Based on Peak Latency of P300.

○Koki Miwa[1], Tomohiro Yoshikawa[1], Takeshi Furuhashi[1], Minoru Hoshiyama[2], Taeko Makino[3], Madoka Yanagawa[4], Yusuke Suzuki[5], Hiroyuki Umegaki[6], Masahumi Kuzuya[7] （1. Dept. of Information and Communication Eng., Graduate School of Eng., Nagoya University, 2. Brain &Mind Research Ceneter, Nagoya University, 3. School of Rehabilitation and Care, Seijo University, 4. Dept. of Geriatrics, Nagoya University Hospital, 5. Center for Community Liaison and Patient Consultations, Nagoya University Hospital, 6. Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University, 7. Institute of Innovation for Future Society / Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University）

4:20 PM - 4:40 PM

[2F3-OS-4b-05] Improvement in Chronic Stress Level Recognition by Using Both Full-term and Short-term Measurements of Physiological Features

○Yoshiki Nakashima[1], Masanori Tsujikawa[1], Onishi Yoshifumi [1] （1. NEC Corporation）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Well-beingの定量化に向けて: 脳機能ネットワークに基づくマインドフルネスの見える化の検討

Direction to quantifying well-being: Visualizing mindfulness based on functional brain network

日和 悟 *1　　廣安 知之 *1
Satoru Hiwa　　Tomoyuki Hiroyasu

*1同志社大学生命医科学部
Faculty of Life and Medical Sciences, Doshisha University

In this paper, authors claim that one of the best way to promote our well-being is to be aware of states of our own well-being. Therefore it is necessary to quantify and visualize the degree of well-being. Here we focus on quantifying psychological well-being, and consider achieving it based on neuroimaging technologies such as fMRI and fNIRS. Therefore, this paper proposes a novel method to define the brain activity pattern occurs during a certain mental state—referred to as a 'metastate'—using data-driven approach. The metastate is modeled as a functional connectivity network, and its characteristic structure is determined by evolutionary computation. The effectiveness of our proposed method is validated on the fMRI data measured during meditation said to improve our psychological well-being. The experimental results revealed that the metastate found by our method can be utilized to quantify how well we can meditate. It will also lead to quantifying our well-being.

1. 序論

Well - being という言葉が最初に用いられたのは，1946 年の世界保健機関 (WHO) 憲章草案 *1 であるとされている．この中で well-being は*"Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. "*（単に病気や虚弱でないだけでなく，健康とは，身体的，精神的，社会的に良好な状態である.）と記述されている．一方，Well-being を幸福（happiness）と表現することもできる．では，幸福とは何であろうか．何をもって幸福とするのか，我々はどんなとき幸せと感じるのか，自己を対象に限定しても，幸福な状態を定義することは難しい．

ポジティブ心理学は，幸福感や well-being の体系化に取り組む学問の一つであり，主観的幸福感を測定する尺度や，幸福感を促進する方法，幸福の要因などが近年盛んに研究されてきた．一方，デジタルテクノロジーに支えられた現代社会において，テクノロジーによる well-being の促進，あるいは well-being を妨げないテクノロジーの設計指針を試みる取り組みも行われつつある．Rafael A. Calvo らは，このようなポジティブ心理学とそれを取り巻く研究領域を支援するテクノロジーを総称して Positive Computing と呼び，テクノロジーのデザインと評価においても，well-being の測定を意識的に行うべきであると主張している [Calvo 17]．また，その著書の中で，Positive Computing において考慮すべき well-being を 3 つのカテゴリ：自己（intra-personal），社会的（inter-personal），超越的（extra-personal）に分類しつつその決定因子を挙げている．ポジティブ感情や動機づけ，思いやりや共感，利他性などさまざまな well-being 決定因子が挙げられているが，これらをいかにしてテクノロジーの設計に組み込むのか，また，これらを促進するテクノロジーの要件について議論していく必要がある．

著者らは，well-being を促進するためには，自己の well-being の状態を知ることが重要であると考えている．例えば Fitbit や Apple Watch などのヘルスケアガジェットは，ユーザの行動と生体情報を計測し，これを元にユーザーの健康状態などを数値化しフィードバックすることでユーザーの健康を促進することを目指している．デバイスからのフィードバックを元に，我々は自らの健康状態を知り，課題が見つかればそれを改善し，良好であることがわかればその状態を継続しようとする．あるいは，行動の履歴から自らの生活スタイルや行動範囲について新たな気づきを得ることもあるかもしれない．自発的な観察だけでは定量化できないような自己の状態に対して，テクノロジーが果たす役割は大きい．しかしながら，これらのデバイスで計測できるのはあくまで身体的な活動であり，我々の well-being を総合的に促進するためには，心理的 well-being を定量化できるテクノロジーが必要である．

そこで本研究では,「テクノロジーによる well-being の定量化」を目的として，特に心理面 Well-being を数値化できるようなテクノロジーの開発に取り組む．また，心理状態や認知活動状態を計測するデバイスとして，機能的磁気共鳴画像法（functional magnetic resonance imaging: fMRI）や機能的近赤外分光法（functional near-infrared spectroscopy: fNIRS）などの非侵襲脳機能イメージングを用い，計測された脳活動データから特定の心理状態の発現度合いを数値化するデータ駆動型アプローチを提案する．本稿では，心理状態の一例として，自己的な well-being 決定因子の一つとされているマインドフルネスを対象として，被験者のマインドフルネスの度合いを数値化することができるかを検討した．

2. 提案手法

2.1 メタ状態マッチングによる認知状態の定量化

脳活動によりヒトの状態を定量化するためには，脳活動データからどのような特徴を抽出するかが重要である．ヒトの脳においては，機能的に分化された複数の脳部位間の構造的・機能的な繋がりが，時間と共に変化している．例えば**図 1**(1) のように，fMRI や fNIRS などの装置により複数の計測点における脳活動の時系列データが得られたとき，それらが複数の脳領域で高い相関をもっていれば，それらの領域が特定の課題中に協調して活動するとみなすことができる．このような協調関係

連絡先: 日和悟，shiwa@mail.doshisha.ac.jp

*1 *Constitution of WHO: principles*, World Health Organization, http://www.who.int/about/mission/en/

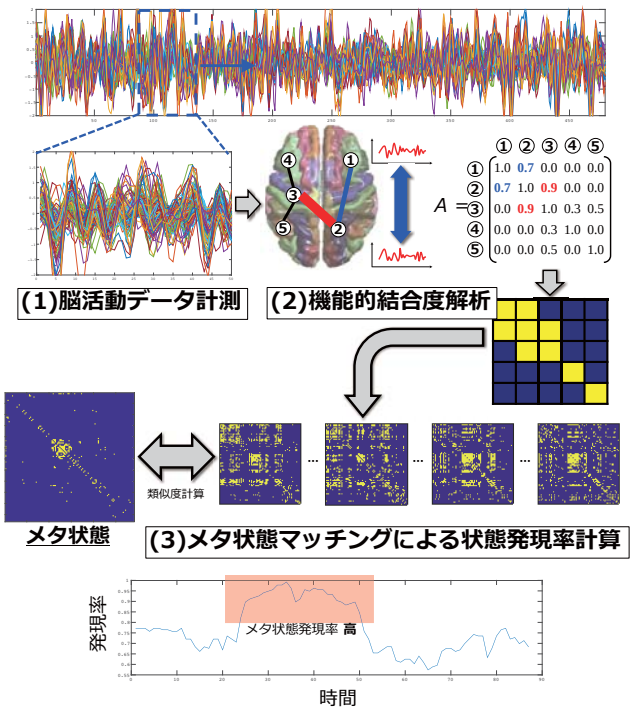


図 1: Framework of metastate matching method

を脳の機能的結合と呼ぶ．機能的結合の強さは時系列データ間の Pearson 相関係数などによって定量化され，**図 1**(2) のように 5 つの脳領域間の機能的結合度は対称行列によって記述することができる．これは脳領域をノード，機能的結合度をエッジとした重み付き無向グラフの隣接行列に相当し，脳活動状態をグラフにより表現していることになる．本研究では，この機能的結合度行列を被験者の脳活動状態を表す特徴量として用い，あらかじめ定義された特定の条件下（心理・認知状態）で発現する機能的結合度行列と比較し類似度を求めることにより，その状態の発現率を定量化する方法を提案する．ここで，特定の条件下における機能的結合度行列を記述したものを「メタ状態」と呼ぶことにする．このメタ状態を注意の集中・散漫や，感情の快・不快，共感などの要素ごとに定義することができれば，それぞれの状態を定量化することが可能である．**図 1**(3) はこの手順を示したものである．なお，図中の機能的結合度行列は，機能的結合度に閾値を設けて二値化（結合あり：1，結合無し：0）した行列を用いている．このように二値化することで，メタ状態に定義された結合が被験者の現時点における機能的結合行列内に存在しているかによって，メタ状態の発現率を数値化できる．

2.2 進化計算によるメタ状態の探索

提案手法の要となるのがメタ状態をどのように定義するかである．多くの神経科学研究により様々な認知状態と脳機能ネットワークの関係が明らかにされており，それらの研究結果を活用し，前提知識から機能的結合行列を定義し，メタ状態とすることもできるが，現在利用可能な脳機能イメージング装置で得られるデータは，装置の違いや個人差の影響を強く受けるため，再現性を確保することが難しい．そこで本研究では，使用したい装置や対象とする認知状態に合わせてメタ状態をデータ駆動型アプローチにより探索する方法を提案する．ここで，メタな状態を定義するということは，個人差に影響を受けず，

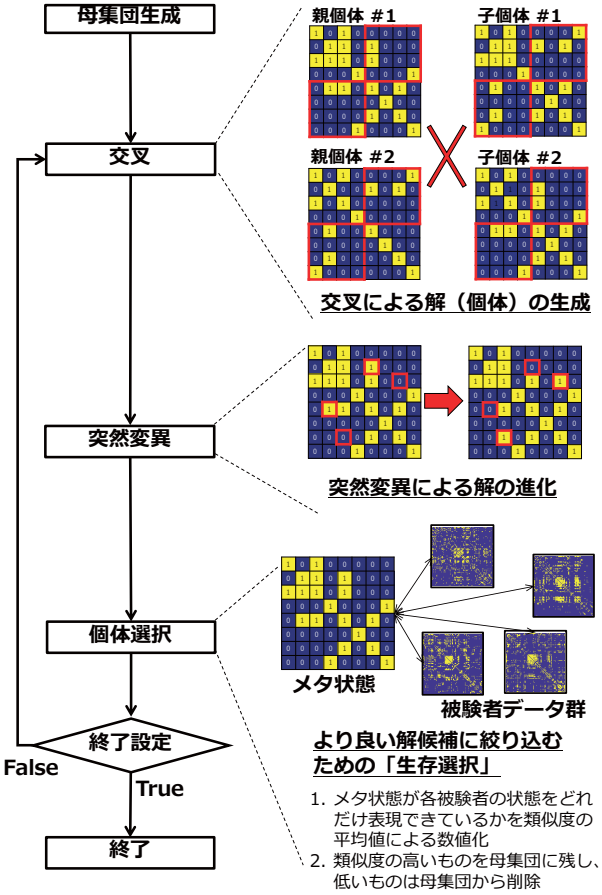


図 2: Evolutionary optimization to extract the metastate

万人に共通して存在する機能的結合を抽出することに他ならない．本研究では，同じ認知状態にある複数の被験者の脳活動データを計測し，そこから求めた各時点の機能的結合度行列に対して，全被験者に対する平均類似度が最大となるようなメタ状態を求める最適化問題とみなし，進化的最適化アルゴリズムによりメタ状態（最適解）を求める．**図 2** は，メタ状態探索のフローを示している．複数のメタ状態の候補解が進化的最適化アルゴリズムにより生成され，全被験者の機能的結合度行列との平均類似度により評価される．評価の高い候補解は保存され，これをベースとして次の候補解が生成されていく．この手順を繰り返し行うことで，最適なメタ状態の構造が抽出される．

3. 実験

提案手法の有効性を確認するため，特定の認知状態として「マインドフルネス」を対象として，健常成人を対象とした fMRI 計測により安静時・瞑想時の脳活動を取得し，提案手法を適用することで瞑想状態＝マインドフルネスの定量化を試みる．実験では，瞑想の熟練者と初心者の瞑想実践中の脳活動を計測し，熟練者のデータからメタ状態を定義し，これを用いて初心者のマインドフルネスの度合いを数値化することとした．

3.1 被験者

初心者群 20 名（訓練時間 11 ± 45h，年齢 23.0 ± 2.6 歳）と，瞑想実践者群 7 名（訓練時間 1560 ± 435h，年齢 49.0 ± 17.0 歳，7 名には同一被験者の測定 3 回分のデータを含む）の健康

な日本人成人男性が書面による同意の上，実験に参加した．瞑想者群はいずれも複数回の瞑想リトリート参加を経験している他，日常的に瞑想訓練を実施している．

3.2 実験プロトコル

瞑想の実践法には，呼吸に注意を維持しながらその出入りを数える数息観とよばれる瞑想法を採用した．実験参加者はヘッドホンを装着した上で，MRI スキャナーの中で 5 分間安静にした後，5 分間の数息観を行い，その後 10 分間安静状態を保った．脳活動の計測はこれら計 20 分間連続して実施され，一度も MRI 装置から出ることなく計測を続けた．なお，最初の安静時間終了後と瞑想時間の終了後に，それぞれヘッドホン経由で「瞑想開始」「瞑想終了」を音声により指示され，被験者はこれに従って瞑想の開始・終了時期を認識した．数息観の指示は計測前に準備室で行い，計測中は終始，閉眼・鼻呼吸を維持すること，自然な呼吸を心がけ意図的に変えないこと，息の吸い始めから吐き終わりを 1 として「いーち」「にーい」…と声に出さず数え，10 まで数えたら再び 1 に戻り数え始めることを指示した．さらに，数え間違いや呼吸以外に注意が逸れた場合には，再び呼吸に意識を戻し，1 から数え直すよう指示した．

3.3 脳機能イメージング

1.5T MRI Scanner Echelon Vega (日立製作所) を使用し，Gradient-Echo Echo-Planer Imaging (GE-EPI) シーケンスにより被験者の脳活動を計測した（TR=2500ms, TE=40ms, flip angle=90°, FOV=(240mm, 5.0-mm thick slices), matrix size=64×64, number of slices=25）．また，T1 構造画像は Rf-Spoiled Steady state Gradient echo (RSSG) シーケンスで撮像した（TR=9.8ms/TE=4.0ms/flip angle=8deg/FOV=(256mm, 1.0-mm thick slices)/matrix size=256×256/number of slices=192）．

3.4 前処理

計測された fMRI データに対して DPARSF[Yan 10] を用いて slice timing correction，realignment，MNI 空間への normalization (New Segment+DARTEL)，CompCor アルゴリズム [Behzadi 07] による白質と脳脊髄液の影響除去，頭部の動き補正，smoothing (4-mm FWHM Gaussian kernel)，linear trend 除去を行った後，BOLD 信号に対してバンドパスフィルタ（0.008—0.09Hz）を適用した．さらに，AAL アトラスにより脳領域を 116 に分割し，各領域の平均 BOLD 信号を抽出した後，125 秒幅の窓を 2.5 秒ずつシフトしながら，窓によって切り出された時系列信号に対して脳領域間の時系列相関を求め，各時点での 116 × 116 の機能的結合度行列を算出した．また，各行列を Fisher ' s z 変換し，結合度上位 15%の結合が保存されるようにバイナリ化した．

4. 結果と考察

4.1 メタ状態

図 3 は，提案手法により抽出されたメタ状態を示している．メタ状態には 13 本の機能的結合のエッジが存在し，熟練者の瞑想中におけるメタ状態との類似度は 75.5 %であった．**図 3** の下部には，このメタ状態のネットワーク構造を図示している．詳細な説明は割愛するが，注意に関わる重要なネットワークとして，背側注意ネットワークと腹側注意ネットワークの結合が含まれており，呼吸に注意を維持することにこれらの脳機能ネットワークが関与したことが示唆された．

4.2 メタ状態発現率の推定

次に，瞑想を実施していない前後 15 分間の fMRI データも利用し，計 20 分間の実験中の被験者の機能的結合度とメタ状態の類似度を算出した．**図 4** は，類似度の時系列変化を示しており，これをマインドフルネスの指標の一つとみなすことができる．**図 4** の左側には，7 名の熟練者の結果が示されている．メタ状態はこれら 7 名のデータから構築されているため，瞑想中にその類似度が高くなることは自明であるが，瞑想実施前の安静時ではその類似度が低くなっていることから，熟練者は瞑想の開始の合図に従って適切に瞑想状態に入り，脳活動状態を変化させたことがわかる．一方で，瞑想終了の合図後も類似度が高い状態が維持されたり，計測の終盤で再び瞑想状態に近い状態になっていたりする被験者も見られた．熟練者の主観評価ではあるが，MRI スキャナー内は計測音こそ騒がしく感じるものの，暗室で目を閉じた状態で瞑想を実施せず安静状態を保つことが難しく，しばしば瞑想を実施しそうになったとの声があった．本手法はこのような変化も捉えている可能性がある．

一方，初心者の多くは，瞑想の開始終了と類似度の変化が同期していないことがわかる．実際，瞑想区間における類似度の平均は 43.3%であり，熟練者の 75.5%と比較して低いことからも明らかである．ただし，個人ごとの類似度の変化を見ていくと，瞑想開始・終了の合図に合わせて類似度が変化している被験者もみられたことから，初心者であっても適切に実践できている被験者が含まれている可能性があり，マインドフルネスの度合いの評価指標としての可能性が示された．

5. 結論

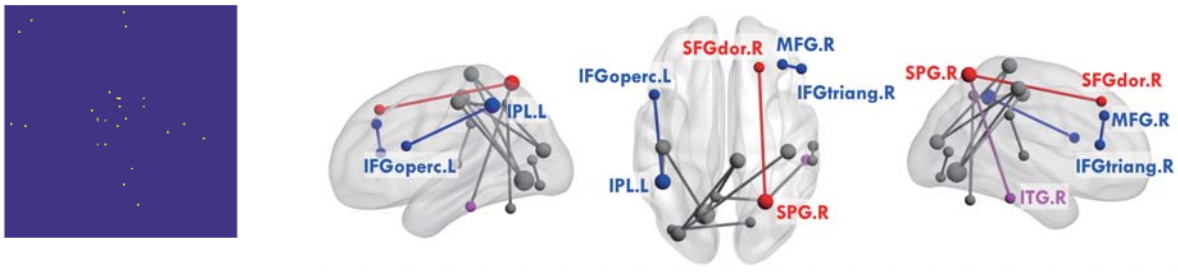
本稿では，well-being を定量化するための試みとして，脳活動データに基づく認知状態定量化の手法を提案した．さらに，心理的な wel-being の決定因子の一つとしてのマインドフルネスに着目し，瞑想中の脳活動データからマインドフルネスの度合いを定量化できるかを検討した．実験結果から，提案手法が瞑想に伴う脳活動状態の変化を捉えており，認知状態の定量化手法としての本手法の可能性を見出すことができた．本手法は，与えられた脳活動データから特徴的な脳機能ネットワーク構造を抽出することができる汎用的なフレームワークを提供している．今後はマインドフルネスに関するデータの蓄積を図りつつ，well-being に関わる様々な認知状態のデータにも本手法を展開して行く予定である．

参考文献

[Behzadi 07] Behzadi, Y., Restom, K., Liau, J., and Liu, T. T.: A component based noise correction method (CompCor) for BOLD and perfusion based fMRI, *NeuroImage*, Vol. 37, No. 1, pp. 90–101 (2007)

[Calvo 17] Calvo, R. A. and Peters, D.: ウェルビーイングの設計論：人がよりよく生きるための情報技術 (渡邊淳司／Dominick Chen 監訳), Positive computing : technology for wellbeing and human potential, ビー・エヌ・エヌ新社 (2017)

[Yan 10] Yan, C. and Zang, Y.: DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI, *Frontiers in Systems Neuroscience*, Vol. 4, No. 13 (2010)



(a) Functional connectivity matrix (b) Connectome

図 3: Metastate of meditation derived by proposed method

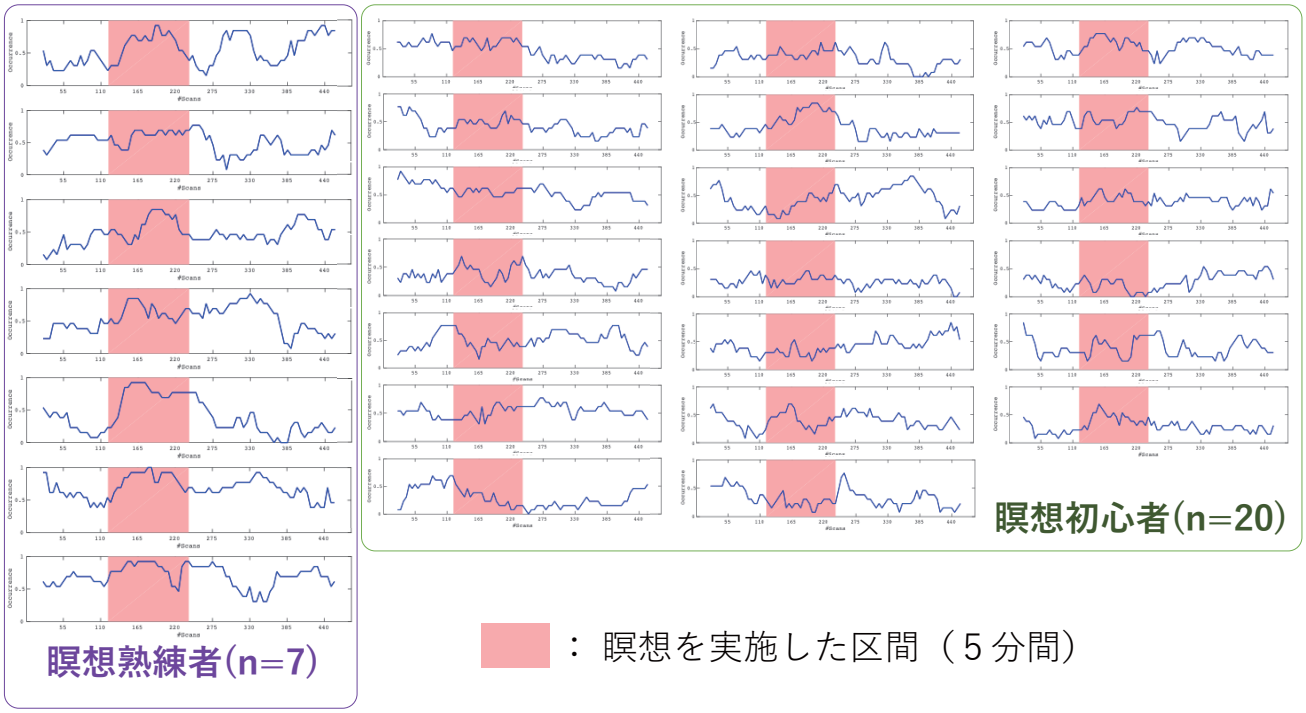


図 4: Estimation of occurrence of the *meditative* metastate

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

相関行列の差の内積に対する クラスタリングを伴う低ランク近似について

Estimation of low rank matrix between two correlation matrices based on clustering

谷岡健資 *1 Kensuke Tanioka　日和悟 *2 Satoru Hiwa　廣安知之 *2 Tomoyuki Hiroyasu　宿久洋 *2 Hiroshi Yadohisa

*1和歌山県立医科大学 Wakayama Medical University　*2同志社大学 Doshisha University

In domains of fMRI data analysis, functional connectivity is described by correlations of signals between regions of interest (ROI). One of the purpose is detecting functional connectivity related to the problem such that researchers would like to know. In this presentation, given two correlation matrices corresponding to class between ROIs, we propose new method to estimate the difference between them. The purpose of this method is three things. First, the estimation is conducted by approximating inner product of the differences, not approximating only the difference. This is because these differences between relations become larger for the inner products compared to those for only difference. Second, estimated low rank correlation becomes sparse. Therefore, it is easy to interpret the relation. Finally, the difference between these estimated values become larger since the estimated difference matrix is low rank.

1. はじめに

近年，fMRI (functional Magnetic Resonance Imaging) のような非侵襲の脳機能イメージング装置が発達し，ヒトの脳機能を解析するための様々な研究が行われている（Fillipi, 2009）．そのような解析方法のひとつに物理的に離れた関心領域 (ROI; Region of Interest) 間の機能的結合を明らかにすることを目的とした研究が挙げられる（e.g. Smith et al., 2011）．

本報告では，暗算課題を被験者に課した際に，暗算能力が相対的に高い被験者のクラス（クラス 1）と暗算能力が高いとはいえない被験者のクラス（クラス 2）で機能的結合に差がある部分を検知し，当該結果の解釈が容易となるような多変量解析法を提案する．本手法は暗算課題のみならず，2 クラスが与えられていれば適用できる方法となる．具体的には機能的結合を相関行列で記述し，相関行列の差が大きい箇所をクラスタリングによって検知して絞り込み，低ランクな相関行列の差によって推定することで容易に解釈が可能となる結果を出力する方法を提案する．

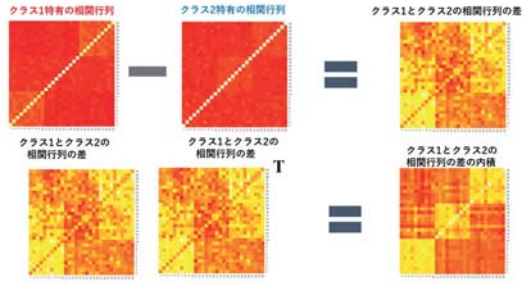


図 1: 人工データによる相関行列の差と相関行列の差の内積

しかし，クラス間でも相関行列に差がある場合でもその差が小さい場合はクラス間の差を検知することが困難になる場合がある．理由としては相関係数は -1 から 1 の範囲でしか値をとらないため，当該差は必然として小さくなる傾向にあることから問題が困難となる（図 1）．そのような問題を解決する方法として，本提案手法ではクラス間の相関行列の差の内積をとることで，相関行列の差のクラスタリング構造を明確化させ，低ランクの相関行列およびクラスタリングを同時に行う手法を提案する（図 1）．例えば，高次元データに対する内積行列や距離行列に対してクラスタリング法を行うことは Terada (2013) によってその有用性が理論的に示されている．

そこで，本報告では Pietersz et al., (2004) や Simon et al., (2010) によって提案されている Majorizing 関数をもとに自乗基準で構成されている目的関数を参考に相関行列の差の内積を低ランクの相関行列の差の内積で近似し，かつクラスタリングも同時に行うことでスパースな低ランクの相関行列の差を推定することができるような方法を提案する．

連絡先: 谷岡健資，和歌山県立医科大学，和歌山県和歌山市紀三井寺 811 番地 1，0734410867，0734410868，ktaniok@wakayama-med.ac.jp

2. 提案モデル

クラス 1 に対応する ROI 間の相関行列 $\boldsymbol{R}_1 = (r_{ij})$ $r_{ij} \in [-1, 1]$ $(i, j = 1, 2, \cdots, p)$，クラス 2 に対応する ROI 間の相関行列 $\boldsymbol{R}_2 = (r_{ij})$ $r_{ij} \in [-1, 1]$ $(i, j = 1, 2, \cdots, p)$ それぞれが与えられたとする．ここで，p は ROI の領域数とする．このとき，各クラスの相関行列は各クラスに属する被験者の相関行列の平均等を用いて計算することを想定している．このとき，クラスター数が与えられた際に，クラス 1 の相関行列 $\boldsymbol{R}_1$ とクラス 2 の相関行列 $\boldsymbol{R}_2$ の差が顕著であるような部分を検知し，かつクラスター内の相関係数間の関係を容易に把握することを目的として下記のような目的関数を定める．

$$\begin{aligned} &L(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U} |\, \boldsymbol{R}_1, \boldsymbol{R}_2) \\ &= \|\boldsymbol{\Delta} - \boldsymbol{U}\boldsymbol{U}^T \odot \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})^T \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})\|^2 \end{aligned} \tag{1}$$

subject to

$$\|\boldsymbol{x}_j\| = 1 \quad \text{and} \quad \|\boldsymbol{y}_j\| = 1 \quad (j = 1, 2, \cdots, p) \tag{2}$$

ここで，$\boldsymbol{\Delta}$ はクラス間の相関行列の差の内積であり，下記のように定義する．

$$\boldsymbol{\Delta} = (\boldsymbol{R}_1 - \boldsymbol{R}_2)^T(\boldsymbol{R}_1 - \boldsymbol{R}_2)$$

また，$\boldsymbol{U} = (u_{i\ell})$ $u_{i\ell} \in \{0, 1\}$ $(i = 1, 2, \cdots, p; \ell = 1, 2, \cdots, k)$ は領域数 p ×クラスター数 k の行列であり，下記のように解釈可能な行列とする．

$$u_{i\ell} = \begin{cases} 1 & (i \text{ belong to cluster } \ell) \\ 0 & (\text{others}) \end{cases}$$

よって，$\boldsymbol{U}\boldsymbol{U}^T$ は $p \times p$ の 1, 0 からなる行列相関係数の差の絶対値が大きければ 1 となり，相関係数の差の絶対値が相対的に小さければ 0 と推定した低ランクの相関行列の差がスパースになるよう推定するための指示行列となっている．

また，$\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ は下記のように定めるとする．

$$\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{Y}\boldsymbol{Y}^T)^T(\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{Y}\boldsymbol{Y}^T)$$

ここで，$\boldsymbol{X}\boldsymbol{X}^T$ は Rank d に制約したクラス 1 の相関行列とし，$\boldsymbol{Y}\boldsymbol{Y}^T$ も同様に Rank d に制約したクラス 2 の相関行列とする．これらの行列相関行列となるための条件が (1) 式の制約条件であり，下記のようになる．

$$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)^T \ \boldsymbol{x}_j \in \mathbb{R}^d$$
$$\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_p)^T \ \boldsymbol{y}_j \in \mathbb{R}^d \quad (j = 1, 2, \cdots, p)$$

これより，$\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ は低ランクの相関行列の差の内積という解釈をすることができる．

(1) 式を最小するような $\boldsymbol{X}$，$\boldsymbol{Y}$，$\boldsymbol{U}$ を推定することで，クラス間が大きくなるようなランク d の相関行列およびクラスタリング結果を同時に推定することを目的とする．アルゴリズムについては交互最小自乗法を用いて推定する．

交互最小自乗法に基づくアルゴリズム

Step 0 rank d，クラスター数 k，$\varepsilon > 0$ を設定し，制約条件を満たす初期値である $\boldsymbol{X}$，$\boldsymbol{Y}$ を制約条件を満たすよう生成する．

Step 1 $\boldsymbol{X}$，$\boldsymbol{Y}$ の値を固定したもとで，$\boldsymbol{U}$ を更新

Step 2 $\boldsymbol{Y}$，$\boldsymbol{U}$ の値を固定したもとで，$\boldsymbol{X}$ を更新

Step 3 $\boldsymbol{U}$，$\boldsymbol{X}$ の値を固定したもとで，$\boldsymbol{X}$ を更新

Step 4 前 step 4 で推定した目的関数と現在の目的関数の差が ε より小さい場合は停止，そうでない場合は Step1 へ

3. 数値例

ここでは，クラスター構造をもつクラス 1 の相関行列およびクラス 2 の相関行列を生成し，提案手法と相関行列の差の内積ではなく，相関行列の差に対するモデルを適用して，真のクラスタリング結果をどれほど再現できるかを簡単な数値シミュレーションによって比較した．人工データの生成方法として，クラス 1 とクラス 2 の多変量データ $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathbb{R}^{p \times p}$ を $\boldsymbol{Z}_1 \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_1)$ and $\boldsymbol{Z}_2 \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_2)$ のように生成する．ここで，

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{O}_{10} & \boldsymbol{O}_{10} \\ \boldsymbol{O}_{10} & \boldsymbol{O}_{10} & \boldsymbol{O}_{10} \\ \boldsymbol{O}_{10} & \boldsymbol{O}_{10} & \boldsymbol{\Sigma}_{31} \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} \boldsymbol{\Sigma}_{12} & \boldsymbol{O}_{10} & \boldsymbol{O}_{10} \\ \boldsymbol{O}_{10} & \boldsymbol{\Sigma}_{22} & \boldsymbol{O}_{10} \\ \boldsymbol{O}_{10} & \boldsymbol{O}_{10} & \boldsymbol{O}_{10} \end{bmatrix}$$

とし，$\boldsymbol{\Sigma}_{11}$，$\boldsymbol{\Sigma}_{31}$，$\boldsymbol{\Sigma}_{32}$ は要素 0.2 の 10×10 の行列であり，$\boldsymbol{\Sigma}_{22}$ は要素 0.1 の 10×10 の行列である．また，$\boldsymbol{O}_{10}$ は要素 0 の 10×10 の行列とする．生成した $\boldsymbol{Z}_1$，$\boldsymbol{Z}_2$ に対してそれぞれ相関行列を生成し，クラス間の行列の差に関するクラスタリング構造に関する復元度合を Hubert and Arabie によって提案された ARI(Adjusted Rand Index) を用いて評価する．つまり，対象数を 30 とし，真のクラスタリング結果である $\{1, \sim, 10\}$ $\{11, \sim, 20\}$ $\{21, \sim, 30\}$ を復元できるかを評価する．初期値の数をそれぞれ 50 回，繰り返し数を 50 回とした結果を図 2 に示す．図 2 より，提案手法による差の内積をモデル化

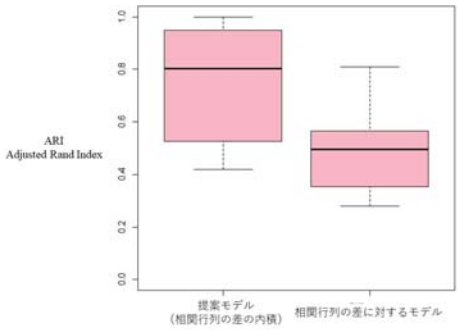


図 2: 数値例の結果

した方が単なる差に対するモデルと比較してクラスタリング結果がよいことを確認することができる．

4. おわりに

本報告では相関行列の差の内積をモデル化することで差が顕著な個所を検知する方法を提案した．しかし，クラスター数や rank 数の決定方法等，実データを通して検討する必要があると考えられる．

参考文献

[Filippi 2009] Fillipi, M. (2009). *FMRI Techniques and Protocols*, Springer Protocols, New York: Humana Press.

[Hubert 1985] Hubert, L. and, Arabie, P. (1985). *Comparing partitions*, Journal of classification, **2**, 193–218.

[Pietersz 2004] Pietersz, R. and Groenen, J. F. (2004). *Rank Reduction of Correlation Matrices by Majorization*, Quant.Finance, **4**, 649–662.

[Simon 2010] Simon, D. and Abell, J. (2010). *A Majorization Algorithm for Constrained Correlation Matrix Approximation*, Linear Algebra and its Applications, **432**, 1152–1164.

[Smith 2011] Smith, S., Miller, K., Salimi-Khorshidi , G., Webster, M., Beckmann, C., Nichols, T., Ramsey, J., and Woolrich, M. (2011). *Network modeling methods for fMRI*, Neuroimage, **54**, 875-891.

[Terada 2013] Terada, Y. (2013). *Clustering for high-dimension, low-sample size data using distance vectors*, arXiv:1312.3386.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

マインドワンダリングからの復帰に要する時間の瞑想経験による変化

Meditation Changes the Time Required for Attention Shifting from Mind-Wandering

川島 一朔[*1,2] Issaku Kawashima
髙橋 徹[*1,3] Toru Takahashi
藤野正寛[*3,4] Masahiro Fujino
熊野宏昭[*2] Hiroaki Kumano

[*1] 早稲田大学大学院人間科学研究科 Graduate School of Human Sciences, Waseda Univ.
[*2] 早稲田大学人間科学学術院 Faculty of Human Sciences, Waseda Univ.
[*3] 日本学術振興会 DC Japan Society for the Promotion of Science.
[*4] 京都大学大学院教育学研究科 Graduate School of Education, Kyoto Univ.

Mindfulness-based interventions might improve the flexibility of mind-wandering, that is, the ability to shift attention from mind-wandering, and well-being. Since the mindfulness-based intervention, which instructs participants to be aware of the occurrence of mind-wandering, might change the bias on self-reports of mind wandering, a measurement method that does not rely on subjects' verbal report is needed. Therefore, we estimated mind-wandering intensity using one-second electroencephalogram samples and a machine learning model developed previously. We observed fluctuations in mind-wandering during a 14-minute meditation and classified it into four conscious phases and validated it with subjects' report. Then we defined the time required to shift attention from mind-wandering as an index of mind-wandering flexibility. The experiment revealed that flexibility of mind-wandering was correlated with the extent of meditation experience.

1. 目的

Well-being を考える上で、日常の中で感じる幸福の度合いは重要である。先行研究は、日常生活における経験サンプリングを行うことで、マインドワンダリング（Mind-Wandering: MW）の体験が、後の幸福感を低下させることを示している (Killingsworth et al., 2011)。MWとは、今いる環境や現在取り組んでいることに注意を払わず、別の事について考えている状態のことを指す。日常生活において MW を生じさせる傾向のことは MW 傾向と呼ばれ、質問紙等によって測定される。個人の MW 傾向を低下させることが示されている唯一の介入法が、Mindfulness-Based Intervention (MBI) である。マインドフルネス (Mindfulness) とは、今この瞬間に対して判断を含まない注意を向ける心的態度を意味する。そして、マインドフルネスを涵養するための瞑想といったトレーニング技法や、それらを主軸としたグループセッションを MBI と呼ぶ。多くの研究により、MBI はうつや不安といった臨床症状を下げるほか、個人の Well-being を向上させることが示されている。

本研究の目的は、MBI が MW の柔軟性に影響を及ぼすという仮説を検討することである。MW の柔軟性とは、MW が生じたときに、必要に応じて意識的に注意を切り替える能力を意味する。MBI は MW 傾向の低減を介してその効果を発揮しているという仮説が提唱されている一方、MBI は MW 傾向を減少させるものでなく、MW の柔軟性を向上させるものであると考えられるようになりつつある (Vago and Zeidan 2016)。しかし、MBI が MW の柔軟性へ影響を与えることを実証する研究は未だない。その理由の一つに、MBI を行う文脈において、柔軟性の変化を測定するのが難しいことがある。現状、MW の柔軟性の測定は言語報告に依存している。しかし、MW の発生に気がつく訓練を行う MBI は、この主観報告において生じるバイアスに対して影響を及ぼす可能性がある。そのため、MBI による柔軟性の変化を測定するためには、自己報告によらない柔軟性の測定方法が必要である。

本研究は、MBI において主要な訓練であるマインドフルネス瞑想を長期間実践した者を対象とする。これら対象がもつ MW への柔軟性を、近年開発された、脳波とサポートベクターマシン回帰による MW 強度の推定技法 (Kawashima and Kumano, 2017) を用いて数値化する。そして、瞑想実践量と柔軟性の相関分析をすることで仮説を検証する。

本研究はすべて早稲田大学「人を対象とする研究に関する倫理委員会」の承認を受けて実施された。申告すべき利益相反はない。

2. 方法

2.1 手続き

マインドフルネス瞑想に熟達した 17 名 (女性 4 名、平均年齢 =40.82±10.45 歳) を対象とした。呼吸に伴う身体感覚へ集中する瞑想を行わせ、その 14 分間の脳波を測定した。また、瞑想中、MW に陥っていることに気がついたらボタンを押すように求めた。さらに、継続的に行っている瞑想の頻度、一日の瞑想時間、集中的に瞑想を行った日数、その際の一日あたりの瞑想時間を尋ねることで、生涯の瞑想経験時間の合計を算出した。

連絡先：川島一朔, 早稲田大学人間科学学術院, issakuss@gmail.com

2.2 意識状態フェーズの推定と妥当性の検証

測定した脳波データに対して Kawashima and Kumano (2017) が発表した MW 強度推定モデルを適用した。これにより、どの程度強く MW を起こしていたかを 1 秒ごとに求め、個人ごとに MW の変動を波形として算出した。

Hasenkamp et al. (2012) をもとに、瞑想中の意識状態は、呼吸へ集中している状態と、MW をしている状態が交互に生じると前提づけた。この前提をもとに、MW の波形を 4 つのフェーズに分類した。すなわち、呼吸に集中ができている「集中フェーズ」、注意が呼吸から離れている「MW フェーズ」、集中している状態から MW に陥りつつある「没入フェーズ」、MW に気づき、注意を呼吸に戻している「復帰フェーズ」である。このうち、復帰フェーズの時間的長さが、注意を MW から戻すのにかかる時間であり、柔軟性の指標として使用する時間である。

フェーズ分類のために、まず MW 波形に 3 ポイントの移動平均を適用し、スムージングを行った。加えて、1/60Hz のハイパスフィルタを適用した。続いて、MW をしている状態あるいは集中している状態で安定しているフェーズを抽出するために、fused lasso を適用した。これにより MW 波形は、複数の安定した(数値が一定時間変わらない)フェーズにまとまった。集中状態と MW 状態が交互に現れるという前提に基づき、この波形におけるピークを抽出した。まず、正方向でのピークを抽出し、ピークに属するフェーズを MW フェーズとした。続いて負方向でも同様にピーク解析を行い、ピークとなったフェーズを集中フェーズとした。MW フェーズまたは集中フェーズに属さない箇所を除外し、線形補間した。さらに、小さい変動の影響を除くため、連続する 2 点の差の絶対値が閾値を下回る場合、後の点の値を前の点の値に置換する処理を採用した。閾値には、その波形において連続する 2 点の差(ただし差が 0 のものは除く)の絶対値の平均値を用いた。この処理を各波形の 1 点目から順に適用した後、再度正負方向のピークを抽出した。以上の処理により、MW の波形から、意識状態フェーズが推定された

意識状態フェーズ推定の妥当性を示すため、ボタン押し (MW に陥っていると気がついたら押すように求めたもの) のタイミングを解析した。マインドフルネス瞑想では、MW が生じていることにできるだけ早く気がつく練習を行う。そのため、瞑想経験量が多いほど、MW へ気がつくまでにかかる時間が短くなると考えられている (Siegel, 2002)。そこで、被験者がボタンを押した際、その直前の MW フェーズ開始時点から何秒が経過しているのかを算出し、個人内で平均値を求めた。これは、MW が開始してからそれに気がつくまでに平均してどの程度の時間がかかったかを意味している。この値と瞑想経験時間との相関関係をスピアマンの順位相関分析により求め、有意な負の相関があることを確認することで妥当性を検討した。

2.3 MW への柔軟性指標の算出と統計解析

各被験者の MW 波形において、MW フェーズから集中フェーズへ移っているフェーズについて時間的長さを算出した。そしてその平均値を被験者の復帰時間とし、MW への柔軟性の指標とした。算出された柔軟性の得点と瞑想経験時間について、スピアマンの順位相関検定を行なった。

3. 結果

各 MW フェーズの開始時点からボタン押しをするまでにかかる時間の平均と、瞑想経験時間との相関を確認した。その結果、有意な負の相関 (ρ = -.56, p < .05) が得られた。復帰時間と瞑想経験時間との間に、有意な負の相関 (ρ = -.50, p < .05) が見られた。

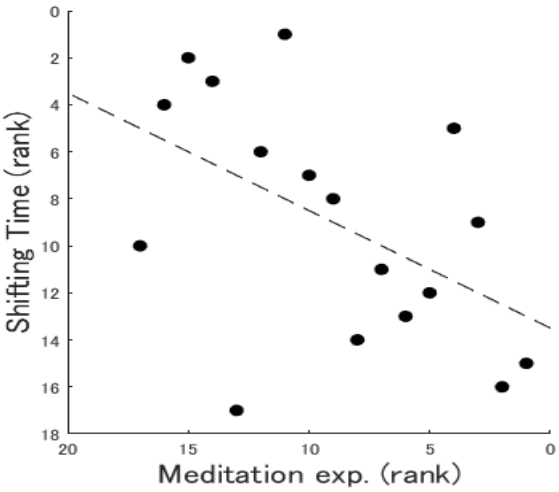


Figure. Scatter plot and regression line for meditation experience and Shifting Time

4. 考察

本研究は脳波による MW 強度の推定を用いて、瞑想中における MW の推移を描出し、さらにそれを複数の意識フェーズに分類した。意識フェーズ分類の妥当性を示すため、MW フェーズ開始から、MW に気がつきボタンを押すまでの平均時間と瞑想経験時間の間に相関があることを確認した。MW フェーズから集中フェーズへ移行するのにかかる時間の平均を柔軟性の指標とし、瞑想経験時間との相関関係を確かめた。

MW に気がつきボタンを押すまでの時間と瞑想経験時間との間に相関関係が見られたことから、本研究で用いた意識フェーズの分類法について、妥当性が一部示された。また、瞑想経験時間と復帰時間に有意な負の相関が示された。瞑想への熟達が進むほど MW から注意を戻すのに時間がかからないことを示しており、マインドフルネス瞑想が MW の柔軟性を高めるという仮説を支持した。

参考文献

[Hasenkamp 12] Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., & Barsalou, L. W.: Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. NeuroImage, 59(1), 750–760. http://doi.org/10.1016/j.neuroimage.2011.07.008 (2012)

[Kawashima 17] Kawashima, I., & Kumano, H.: Prediction of Mind-Wandering with Electroencephalogram and Non-linear Regression Modeling. Frontiers in Human Neuroscience, 11, 365. (2017)

[Killingsworth 10] Killingsworth, M. A., & Gilbert, D. T.: A Wandering Mind Is an Unhappy Mind. Science (New York, N.Y.), 330(6006), 932–932. (2010)

[Vago 16] Vago, D. R., & Zeidan, F.: The brain on silent: mind wandering, mindful awareness, and states of mental tranquility.

Annals of the New York Academy of Sciences, 1373(1), 96–
113. http://doi.org/10.1111/nyas.13171 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

P300 頂点潜時に基づく認知症の兆候検出に関する研究

A study on Detection of Symptoms of Dementia Based on Peak Latency of P300.

三輪晃暉[*1] 吉川大弘[*1] 古橋武[*1] 寶珠山稔[*2] 牧野多恵子[*3]
Koki Miwa Tomohiro Yoshikawa Takeshi Furuhashi Minoru Hoshiyama Taeko Makino

柳川まどか[*4] 鈴木裕介[*5] 梅垣宏行[*6] 葛谷雅文[*7]
Madoka Yanagawa Yusuke Suzuki Hiroyuki Umegaki Masafumi Kuzuya

[*1] 名古屋大学大学院工学研究科情報・通信工学専攻
Dept. of Information and Communication Eng., Graduate School of Eng., Nagoya University #1

[*2] 名古屋大学 脳と心の研究センター
Brain & Mind Research Ceneter, Nagoya University #2

[*3] 星城大学 リハビリテーション学部
Dept. of Rehabilitation and Care, Seijoh University #3

[*4] 名古屋大学医学部附属病院 老年内科
Dept. of Geriatrics, Nagoya University Hospital #4

[*5] 名古屋大学医学部附属病院 地域連携・患者相談センター
Center for Community Liaison and Patient Consultations, Nagoya University Hospital #5

[*6] 名古屋大学大学院医学系研究科地域在宅医療学・老年内科
Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University #6

[*7] 名古屋大学未来社会創造機構／大学院医学系研究科地域在宅医療学・老年内科
Institute of Innovation for Future Society / Dept. of Community Healthcare and Geriatrics, Graduate School of Medicine, Nagoya University #7

P300 is an enhanced positive component in EEG(Electroencephalogram) observed around 300ms after an Oddball stimulus is presented to a subject. It has been reported that the peak latency of P300 is dependent on the degree of task difficulty, age, educational background, and MMSE (Mini-Mental State Examination) score. The authors measured the peak latency of dementia patients attending the Geriatrics of Nagoya University hospital and identified a multiple regression equation with MMSE score as the objective variable. The 95% confidence interval of estimated MMSE was ±3.42. In this paper, a method to eliminate outliers of EEG data is employed. The 95% confidence interval was ±3.14, and the obtained model has a better ROC curve than that of a regression model without the outlier elimination.

1. 初めに

日本では，一人暮らしの高齢者と認知症患者の数は年々増加し，2030 年にはいずれも 700 万人以上（65 歳以上の全高齢者数の約 20%）に達すると推定されている[内閣 16]. 健康寿命を長くするなど QOL (Quality of Life)を重視する医療の確立・展開が重要課題となっている. しかし，認知症は症状が進んでからでは治療が困難であり，QOL の低下を招くため，早期発見が重要であるとされている[Arai 15].

認知症は自覚することが困難であるため，一人暮らしの高齢者向けの認知症兆候検出システムのニーズがある[Ebly 99]. 名古屋 COI(Center of Innovation)拠点未来社会創造機構サステナブル基盤部門では，図 1 に示すスマートチェアの開発が進められている[名古屋 15]. 名古屋 COI は，このスマートチェアに非接触磁気センサを搭載し，脳波測定を可能にする研究も進めている[Wang 14]. この脳波から認知症の兆候を検出できれば，高齢者はスマートチェアに座って日々くつろいでいるだけで，第三者の手助け無しに，認知症発症の可能性を早い段階で知ることができる.

国際的に広く使われている認知症のスクリーニング検査の一つに MMSE (Mini Mental State Examination) がある[Folstein 75], [Cockrell 88]. 図 2 に示すような「今日は何日ですか」などの質問にいくつ答えられたかによって点数がつけられる. この検査は病院等の医療機関において，医師や臨床心理士等によって実施されている. MMSE には，問題文が固定されているために被験者が検査慣れをする可能性があること，被験者が不快に感じ真剣に答えないことがあることなどの難点が指摘されている

連絡先：三輪晃暉，名古屋大学大学院工学研究科情報・通信工学専，052-789-2793，miwa@cmplx.cse.nagoya-u.ac.jp

[中島 11]. また, 被験者一人では実施できないことから, 一人暮らしの高齢者が自宅で気軽に検査に利用することは難しい.



図 1: スマートチェア

質問と注意点		回答	得点
1(5点) 時間の見当識	「今日は何日ですか」 ※最初の質問で, 被験者の回答に複数の項目が含まれていてもよい. その場合, 該当する項目は省く	日	0 1
	「今日は何年ですか」	年	0 1
	「今の季節は何ですか」		0 1
	「今日は何曜日ですか」	曜日	0 1
	「今月は何月ですか」	月	0 1

図 2: MMSE の抜粋

本稿では, 脳波の特徴量である P300 の頂点潜時に着目し, P300 頂点潜時に基づいた認知症の早期発見支援を目指す. P300 とは, 事象関連電位の一種であり, 出現頻度の異なる 2 種類以上の刺激を被験者に与えた際, 低頻度の刺激呈示の約 300ms 後に誘発される陽性電位を指す. P300 頂点潜時とは低頻度の刺激呈示から P300 の頂点までの時間のことであり, 従来研究では, 健常者群, 軽度認知障害者群, アルツハイマー病患者群の順に P300 頂点潜時が延伸することが報告されている[Indoria 17]. また P300 頂点潜時は MMSE スコアと負の相関があることが示されている[Braverman 03]. さらに Goodin ら[Goodin 78]などにより P300 頂点潜時と年齢に正の相関があること, 佐田ら[佐田 02]によりオドボール課題の難易度が高いほど P300 頂点潜時が延びることが報告されている. 一方で, MMSE スコアは, 教育歴が高いほど, 高くなるといわれている. 前回の研究[三輪 18]では, P300 頂点潜時, 課題の難易度, 年齢, 教育歴の 4 変数を説明変数とし, 目的変数に MMSE を使い, 重回帰分析を行うことで重回帰式を同定し, 95%信頼区間が±3.42 のモデルを得た.

本稿では p 値に基づく変数選択と, 推定値の 95%信頼区間に基づくはずれ値除去を併用することで, モデルの精度向上を目指す. 本手法によるモデルの推定精度向上効果を確認するために, 文献[三輪 18]の老年内科の通院患者のデータと文献[高倉 16]の 70 代の健常者のデータに適用して MMSE スコア推定を行い, はずれ値除去を行わない場合のモデルとの比較を行う.

2. 実験

2.1 被験者

名古屋大学医学部附属病院老年内科に通院している患者に被験者を依頼して, 脳波計測実験を実施した. 実験期間は 2016 年 9 月 14 日～2016 年 11 月 16 日と 2017 年 2 月 21 日～2017 年 5 月 16 日であった. 被験者の選考基準は MMSE20 以上とした. 実験の趣旨に同意していただき, かつ, 脳波計測に成功した被験者数は 24 名であった. 表 1 に, 各被験者の MMSE スコアと年齢を示す. 被験者は MMSE スコアの昇順にソートしてある. 本実験は名古屋大学医学系研究科生命倫理審査委員会の承認(課題番号 2016-0160)を得て, 実施した

表 1: 被験者の MMSE と年齢

被験者No.	MMSEスコア	年齢	被験者No.	MMSEスコア	年齢
1	20	85	13	27	83
2	20	85	14	27	72
3	21	79	15	27	78
4	21	91	16	28	69
5	22	79	17	28	80
6	22	73	18	29	77
7	23	74	19	29	74
8	25	78	20	29	68
9	25	75	21	29	75
10	25	78	22	30	83
11	26	74	23	30	81
12	26	80	24	30	81

2.2 視覚オドボール課題

P300 頂点潜時を測定するための課題に, 視覚オドボール課題がある[Sanei 03]. この課題では, 被験者に 2 種類以上の刺激を出現頻度に差をつけて呈示し, 被験者にはターゲット刺激(低頻度刺激)の出現時に出現回数を数え上げる(計数課題), もしくは, ボタンを押す(ボタン押し課題)ことを求める. ターゲット刺激呈示後の約 300ms 付近の脳波には P300 が含まれる. 脳波には大きなノイズが含まれ, P300 はノイズに埋もれている. そこで, ターゲット刺激を複数回呈示した際の脳波形を加算平均することでノイズ除去を行い, P300 を求める.

2.3 実験設定

脳波計は, デジテックス研究所製の PolymateII AP216 を用いた. 国際 10-20 法\cite{10-20}に則り, 図 3 のように 1 電極(P_z), および基準電極(A_1, A_2 :耳朶)を装着し, サンプリング周波数 1000Hz で測定を行った. 実験は静かな部屋で行い, 空調は被験者の要望に合わせて調節した. 被験者には利き腕にボタンを持っていただき, ターゲット刺激が画面に現れた際には, 速やかにボタンを押すことを依頼した. モニタ(DELL UltraSharp 1906FPT 19 LCD Monitor)は被験者がターゲット・スタンダード刺激を見やすい距離(約 1m), 高さに配置した. 本実験では, ボタン押し課題を採用し, また呈示刺激には図 4 の図形[高倉 16]を使用した. ターゲット刺激とスタンダード刺激の違いは半径にある. ターゲット刺激の半径を dt, スタンダード刺激の半径を ds として, 課題の難易度 D を,

$$D = \frac{d_t}{d_s} \tag{1}$$

と定義する. 刺激呈示は, ターゲット刺激を 20%, スタンダード刺激を 80%の頻度でランダムにモニタ中央に表示することで行

い，刺激呈示回数は，図 4 の各段階の難易度において，110 回（ターゲット刺激 22 回，スタンダード刺激 88 回）とした．刺激呈示間隔は，スタンダード刺激またはターゲット刺激が表示されている点灯時間 500ms，何も表示されていない消灯時間 500ms の計 1000ms とした．図 4 の難易度の段階をランダムに変えて，各段階について 110 回の刺激呈示を行った．全段階の実施にはこだわらず，被験者の様子を見て途中で打ち切ることもあった．

実験開始時に，被験者に視覚オドボール課題の説明をして，ボタン押しを依頼し，計測実験終了後に MMSE を行った．計測実験と同日に MMSE が実施されていた場合は，先に実施されたテストのスコアを採用した．

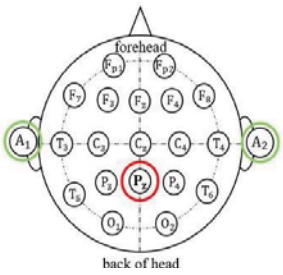


図 3: 使用電極

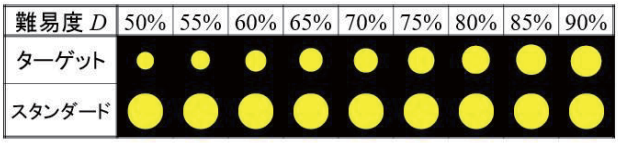


図 4: 半径比により難易度を定量化した呈示刺激

3. 解析手法

3.1 P300 頂点潜時の算出法

P300 は 1-4Hz の帯域で捉えられるため\cite{P300delta}，収録された脳波データに 1〜5Hz のバンドパスフィルタをかけた．次に，難易度の各段階におけるターゲット刺激時の脳波形（ターゲット波形）について，ターゲット刺激呈示後 0〜1000ms の区間の波形から，ベースライン（ターゲット刺激呈示より-100〜0ms の区間の波形の平均値）の値を引いた波形を求めた．そして，はずれ値除去し，残ったターゲット波形の加算平均を求めることで P300 頂点潜時を算出した．

3.2 重回帰分析と 95%信頼区間

表 2: 変数選択とはずれ値除去による重回帰分析結

重相関R	0.82		
重決定R2	0.67		
補正R2	0.67		
	偏回帰係数	偏相関係数	P-値
頂点潜時	−0.66	−0.71	0
難易度	0.37	0.51	4.3E−09
教育歴	0.41	0.56	4.1E−11

本実験で扱う重回帰分析は，目的変数を MMSE スコアとし，説明変数に 3.1 節で得られた P300 頂点潜時，課題の難易度 D，年齢，教育歴を用い，重回帰式を同定する．同定した重回帰式の各説明変数の p 値をみて，有意水準 0.05 を超える説明変数がある場合はその説明変数を除去し，再度重回帰式を同定した．

得られた重回帰式を学習データに適用し，MMSE スコアの推定値を求め，その推定誤差の 95%信頼区間を算出した．

3.3 p値による変数選択と 95%信頼区間によるはずれ値除去

データにはずれ値が含まれる場合，重回帰式の推定精度は悪化し，推定誤差が大きくなる．そこで，全説明変数を候補として，p値に基づく変数選択によるモデル同定を実施した後，推定値の 95%信頼区間外のデータをはずれ値として除去する．この作業を繰り返してはずれ値を次第に除去していき，信頼区間が収束した時点で作業を終了する．この手法により得られた重回帰式を最終モデルとした．図 5 はフローチャートである．

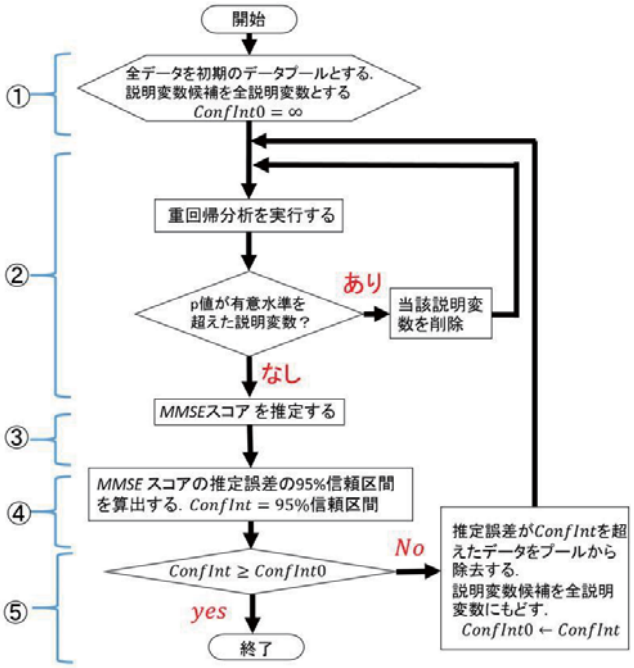


図 5: 95%信頼区間によるはずれ除去のフローチャート

① 全データを初期のデータプールとする．95%信頼区間の初期値を *ConfInt*0=∞とする．

② P300 頂点潜時，課題の難易度 D，年齢，教育歴を説明変数，MMSE 実測値を目的変数とする重回帰式を同定する．

③ p 値が有意水準を上回る説明変数があれば，これを除去してステップ②にもどる．無ければステップ④へ．

④ 重回帰式により MMSE スコアを推定する．

⑤ MMSE スコアの推定誤差の 95%信頼区間を算出する．*ConfInt* = 95%信頼区間とする．

⑥ $ConfInt \geq ConfInt0$ であれば終了する．$ConfInt <$ *ConfInt*0 であれば，推定誤差が *ConfInt* を超えたデータをデータプールから除去する．$ConfInt0 \leftarrow ConfInt$ としてステップ②に戻る．

4. 結果

4.1 4 つの説明変数による重回帰式の同定

表1のデータを用いて，変数選択とはずれ値除去を実施して得られた重回帰式は

$$\textbf{MMSEスコア} = -0.66 \times (\textbf{頂点潜時}) + 0.37 \times (\textbf{難易度}) + 0.41 \times (\textbf{教育歴}) \quad (2)$$

であった．年齢は変数選択により削除された．また，はずれ値として23%のデータが削除された．このモデルは，標準偏差推定

量 1.61, 95%信頼区間±3.15 の精度であり, 文献[三輪 18]のはずれ値除去を行わなかったモデルの 95%信頼区間±3.42 より良い結果となった.

各説明変数の偏回帰係数, 偏相関係数, p 値, および重相関係数 R を表 2 に示す. 各変数の偏回帰係数と偏相関係数に大きな差はなく, 多重共線性が起きている可能性が低いことがわかる. 偏回帰係数によると, 教育歴が高いほど MMSE スコアは上昇し, また, 頂点潜時が延びるほど MMSE スコアが下降しており, 従来研究と傾向が一致した. 難易度の偏回帰係数が正である理由は次のとおりに説明できる. 例えば, 頂点潜時を固定して考えると, 難易度が高いところと低いところで頂点潜時が同じであることは, 難易度が高いところでも頂点潜時が延びない被験者の MMSE スコアが高いことによる.

4.2 はずれ値除去の効果

文献[三輪 18]の通院患者のデータと文献[高倉 16] の 70 代の健康な被験者のデータを用いて, モデルの推定精度の比較を行った. ただし, この実験では教育歴を記録していなかったため, 説明変数候補を P300 頂点潜時, 課題の難易度, 年齢の 3 変数として, 通院患者群のデータを用いて, 変数選択とはずれ値除去を行い, モデルを同定し直した. 結果を表 3 に示す.

得られた重回帰式を通院患者／健常者のデータに適用して

表 3: 説明変数候補を P300 頂点潜時, 課題の難易度, 年齢の 3 変数として変数選択とはずれ値除去を行った場合の重回帰分析結果

重相関R	0.79		
重決定R2	0.62		
補正R2	0.62		
	偏回帰係数	偏相関係数	P−値
頂点潜時	−0.88	−0.78	0
難易度	0.48	0.57	3.0E−11

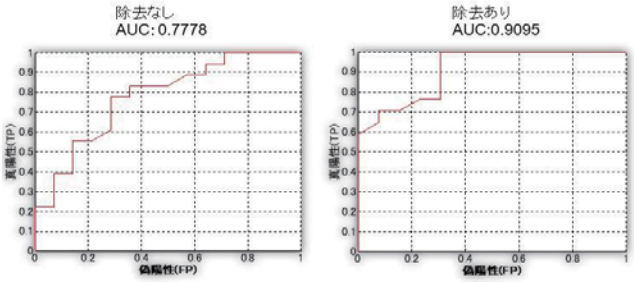


図 6: ROC 曲線
(左)変数選択のみによるﾓﾃﾞﾙ (右)変数選択とはずれ値除去によるﾓﾃﾞﾙ

MMSE スコア推定値を求め, 難易度毎の MMSE スコア推定値の平均値を被験者の MMSE スコア推定値とした. 文献[高倉 16]の健常者は全て健常群とした. 通院患者は, MMSE の軽度認知症判定の目安である 27 点を基準として, 27 点以下の被験者を認知症群, 28 以上の被験者を健常群とした. 図 6 は ROC 曲線である. 左側が変数選択のみを実施した場合であり, 右側が変数選択とはずれ値除去の両方を実施した場合である. ROC 曲線から AUC を算出した結果, 左側の除去なしでは 0.78, 右側の除去ありでは 0.91 であった.

5. まとめ

本稿では, 変数選択と推定値の 95%信頼区間によるはずれ値除去を適用し, モデルの精度向上を目指した. 得られたモデルの MMSE スコア推定値の 95%信頼区間は±3.14 であった. また, 変数選択とはずれ値除去によるモデルは, 変数選択のみによるモデルより, AUC が高いことが分かった.

今後は健常者を被験者とする追加実験によるモデルの精度向上と, モデルによるフィールド実験を通して認知症の兆候発見の可能性を探っていく.

6. 謝辞

本研究は, 名古屋 COI 拠点未来社会創造機構サステナブル基盤部門の援助を得て遂行された.

参考文献

[内閣 16] 内閣府: 平成 28 年版高齢社会白書 (2016).

[Arai 15] Arai, H. et al: Japan as the front-runner of super-aged societies: Perspectives from medicine and medical care in Japan, Geriatrics & Gerontology International, Vol. 15, No. 6, pp 673-687 (2015).

[Ebly 99] Ebly, E. M. et al: Living alone with dementia, Dementia and geriatric cognitive disorders, Vol. 10, No. 6, pp 541- 548(1999).

[名古屋 15] 名古屋 COI 拠点未来社会創造機構サステナブル基盤部門: 研究成果公開 スマートチェア, 歩行支援ロボット(2015).

[Wang 14] Wang, K. et al: Detection of P300 brain waves using a magneto-impedance sensor, 2014 the 8th International Conference on Sensing Technology, Liverpool, UK, pp 2-4(2014).

[Folstein 75] Folstein M. F. et al: “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician, Journal of psychiatric Research, Vol. 12, No. 3, pp 189- 198(1975).

[Cockrell 88] Cockrell, J. R. and Folstein, M. F.: Mini-Mental State Examination (MMSE), Psychopharmacology bulletin, Liverpool, UK, pp 2- 4(2014).

[中島 11] 中島健二: 認知症疾患治療ガイドライン 2010, 日本老年医学会雑誌, Vol. 48, No. 6, pp 637- 639(2011).

[Indoria 17] Indoria, S. P. et al.: A study of P300 and Mini Mental State Examination in Mild Cognitive Impairment and Alzheimer’s Dementia (2017).

[Braverman 03] Braverman, E. R. et al: P300 (latency) event-related potential: an accurate predictor of memory impairment, Vol. 34, No. 3, pp 124- 139 (2003).

[Goodin 78] Goodin, D. S. et al: P300 (latency) event-related potential: an accurate predictor of memory impairment, Vol. 34, No. 3, pp 124- 139 (2003).

[佐田 02] 佐田ほか:漢字および図形に対する認知機能評価第 2 報 精神遅滞児における視覚性事象関連電位 P300, Vol. 34, No. 6, pp 491- 497 (2002).

[三輪 18] 三輪ほか:脳波データを用いた MMSE スコアの推定に関する検討, Vol. 117, No. 417, pp 5- 10 (2018).

[高倉 16] 高倉, 吉川, 古橋: オドボール課題の難易度の違いに関する年齢と P300 頂点潜時との関係についての検討, Vol. 115, No. 514, pp 201- 206 (2016).

[Sanei 03] Sanei, S. and Chambers, J. A.: EEG Signal Processing, WILEY (2007).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

全期間及び短期間双方の生体信号の使用による長期ストレスレベル認識精度の向上

Improvement in Chronic Stress Level Recognition by Using Both Full-term and Short-term Measurements of Physiological Features

中島 嘉樹[*1] Yoshiki Nakashima
辻川 剛範[*1] Masanori Tsujikawa
大西 祥史[*1] Yoshifumi Onishi

[*1] 日本電気株式会社　データサイエンス研究所
Data Science Research Laboratories, NEC Corporation

In this paper, we propose improvement in chronic stress level recognition by using both full-term and short-term physiological features. In our proposed method, we employ the characteristic of PSS (Perceived Stress Scale), a widely used chronic stress measure. PSS scores are known to be influenced by mental states caused by high-stress experiences that could occur over shorter terms. So we added new stress features calculated on a weekly basis to the conventional stress features calculated on a monthly basis for which PSS questionnaires recognizes stress level. With weekly-basis feature calculations, we are able to recognize high-stress experiences over shorter terms. To evaluate our proposed method, we performed experiments using a 33-employee, 1-month database of physiological signals. Results have shown the Pearson's correlation coefficient to improve from 0.66 to 0.72 with use of our method.

1. はじめに

職業性ストレスは，鬱病等の精神疾患の発症をもたらし，生産性の低下や離職・休職等の原因となっており，被雇用者にとって大きな問題である．雇用者にとっても，生産性や労働力の減少は対策コストを強いられる課題である．ストレスを早期に見出し，それを防止することは，被雇用者・雇用者双方に大きな利益がある．

ストレスには，短期ストレスと長期ストレスがある．短期ストレスとは，身体的または心理的ストレッサーへの短期的な反応を指す．ストレッサーへの短期的な反応は，生体信号を用いて容易に検出できるため，Alberdi らがまとめているように[1]，生体信号からストレッサーへの短期的反応の程度を認識する多くの先行研究がある．それに対し，長期ストレスは，長期に亘ってストレッサーに曝されることにより起こるもので，鬱病等の原因となるため，より深刻な問題であるが，長期間に亘る継続的な測定が必要であるため，研究の難度は高く，先行研究は少ない．

生体信号を用いた長期ストレスレベル認識の重要な先行研究の一つは，Sano ら[2][3]によって報告されたものである．この先行研究では，68 人 [2]または 165 人 [3]の学生の 1 ヶ月分という，比較的大きなデータベースを使用しており，当該データベースには， 生体信号として EDA(皮膚電気活動)，ST(皮膚温度)，および加速度計信号が含まれている．当該先行研究では，心理学的に信頼できる長期ストレスアンケートである PSS(Perceived Stress Scale)[4]のスコアを元に 2 クラス(PSS スコア上位 20%，下位 20%の 2 クラス)を定義し、その 2 クラスの分類を行っている．

しかし，ストレスレベルの全範囲を認識するのに有効な技術を開発するためには，ここで省略された中位 60%も考慮する必要がある．本稿では，特にこの中位ストレスに対するストレスレベル認識精度の改善について，全期間および短期間の双方で算出した生体信号特徴量を用いて改善することを提案する．我々は広く使われている長期ストレス尺度である PSS を採用することとした． PSS は一ケ月単位での長期ストレッサーへの曝露の程度を認識するだけでなく，一ケ月より短い期間で起こり得る強いストレッサーへの曝露によって引き起こされる心理状態の悪化（心理的不調）も反映する特性が知られている[5]．また我々の考察ではこの特性は特に中位のストレスレベルにおいて顕著に表れる．よって，PSS によるストレスレベルをより精度よく認識する為に，PSS のこのような特性も考慮し，我々は，1 ヶ月単位だけでなく，より短い期間である 1 週間単位でもストレス特徴量を算出した．これによって，中位を含む全範囲のストレスレベルを精度よく認識することができる．

以下，本稿の構成について述べる．第 2 節では，Sano ら[2][3]の研究を先行研究として紹介し，その概要と課題を述べる．第 3 節では，我々が提案する方法について述べる．第 4 節では，提案法の実験的評価について述べる．第 5 節では，研究を要約する．

2. 先行研究

2.1 先行研究の概要

本節では，まず，Sano らの先行研究[2][3]の概要について述べる．当該先行研究に代表される長期ストレスレベル認識システムは，概ね図 1 のような構成をとる．まず，被験者が着用するウェアラブルセンサ等を使用して，生体信号を取得する．たとえば，Sano らの研究では，EDA， ST，加速度を取得している．これらは，それぞれ発汗，体温，

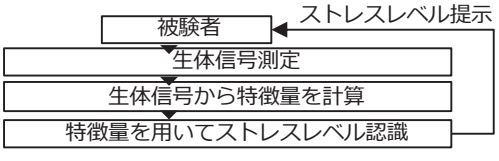


図 1 ストレスレベル認識システム

連絡先:中島嘉樹，日本電気株式会社，〒211-8666 川崎市中原区下沼部 1753，電話番号:044-431-7554，電子メイルアドレス:y-nakashima@bu.jp.nec.com

体動を測定する為のものである．次に特徴量の計算を行う．先行研究では，取得した生体信号から 1 ヶ月単位で一括して特徴量計算を実施している．このような 1 ヶ月単位の特徴量の例として，以下に示す 2 つの式(式(1)および(2))による，特徴 *a6G_M* で示されるような 3 軸加速度計信号に基づく値が挙げられる．

$$x_i = \sqrt{x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2} \quad (1)$$

$$a6G\_M = \frac{1}{N}\sum_{i=1}^{N} a6G_i,\ a6G_i = \begin{cases} 0\ (x_i < 6G, 7G \le x_i) \\ 1\ (6G \le x_i < 7G) \end{cases} \quad (2)$$

ここで，式(1)において，i は信号の系列番号であり，*1* から N の値を取る．N は信号の総数である．*SR* をサンプリングレート，D を一月の合計日数とすると，N は $D * 24 * 3600 * SR$ となる．D は *28, 29, 30, 31* のうちどれかの値を取る．*SR* は，先行研究においては 8Hz であった．$x_{1,i}$，$x_{2,i}$ および $x_{3,i}$ は，ウェアラブルセンサによって測定される，互いに直交する 3 軸の加速度である．測定された加速度信号は，*G*, すなわち重力加速度 9.8m/s^2 の単位で記録される．ここで，式(1)における x_i は 3 軸加速度の RMS(Root Mean Square)である．式(2)において，*a6G_M* は，全加速度計信号のうち，RMS が 6G～7G の範囲にある信号が全体に占める割合を示す．*a6G_M* 等の，月単位での特定の範囲の加速度信号の全体に対する割合を計算した特徴量は，当該月の体動を反映する特徴量の一種である．小さい加速度の割合が大きい場合，その月の体動は落ち着いており，一方，大きな加速度の割合が大きい場合，その月の体動は激しいと言える．先行研究では，長期ストレスはこうした月単位の特徴量を用いて認識された．当該先行研究では，PSS スコア上位 20%および下位 20%の PSS スコアの被験者が，それぞれ高ストレスおよび低ストレスクラスとして定義され，2 クラス分類で識別された．中位 60%の被験者を省略し，ストレスレベルが極端に離れた 2 クラスのデータのみを用いて 2 クラス分類を行うことは，そこで用いられる式(2)のような長期ストレス特徴量を評価するには有効であった．

2.2 先行研究の課題

先行研究で開発された特徴量は，被験者を高ストレス・低ストレスの 2 クラスに分類するには有効であるが，PSS を長期ストレスの指標とする場合，特に中程度のストレスに対しては充分に有効ではない．PSS が測定するストレスの指標である「ストレス評価」は回答者が曝されたストレッサーの程度だけでなく，ストレッサーによって影響を受けた回答者の心理状態も同時に反映する特性がある為である[5]．PSS は一ケ月間の「ストレス評価」を認識の対象とするが，回答者の心理状態は，一ケ月より短い期間に起こり得る強いストレッサーによって悪化することがある．そして，回答者の「ストレス評価」は，たとえ回答者が当該月のその他の期間には弱いストレッサーにしか曝されていない場合でも，比較的高い値となる傾向がある．ところが，先行研究の式(2)のような月単位の特徴量では，このような PSS の特性は反映されない．月単位での特徴量に PSS の特性が反映されない状況は，次章で詳細に述べるとおり，当該月のストレッサーに強弱の変動がある場合，つまり月単位で見れば中程度のストレスレベルの場合に顕著になる．

我々は，PSS，即ち「ストレス評価」がこのように回答者の心理状態の悪化の影響を受けるという特性は，リスクの高い回答者を識別するのにより有効であると判断し，そのような PSS の特性を表現できる特徴量を開発する必要があると考えた．

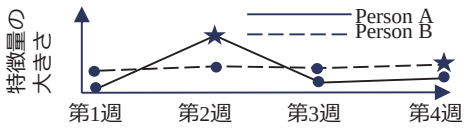


図 2 提案特徴量のコンセプト

3. 提案法

本節では，まず，前節で述べた先行研究の課題をより具体的に説明し，次にその課題を解決する為の提案法について述べる．まず，図 2 を参照する．図 2 には，式(3)で示される週単位特徴量のプロットが 2 つ示されている．具体的には，式(3)の $a6G\_W_j$ を，j = 1,2,3,4 に亘ってプロットしている．式 (3)は月単位の特徴量である式(2)の *a6G_M* を週単位にしたものであり，月単位ではなく週単位のストレスレベルを反映できる特徴量になっている．

$$a6G\_W_j = \frac{4}{N}\sum_{i=\frac{N}{4}(j-1)+1}^{\frac{N}{4}j} a6G_i \quad (3)$$

図 2 の 2 つのプロットのうち，実線，つまり「Person A」のプロットは，低いストレスレベルと高いストレスレベルの両方を含み，一方、破線，つまり「Person B」のプロットは，4 週間全体に亘って中程度のストレスレベルを示している．前述のメカニズムにより，PSS スコアにおいては「Person A」が高くなり，実際のリスクも「Person A」の方が高くなると思われるが，式(2)を使って月単位の特徴量を計算すれば，両者のストレスレベルは同程度になる．そのため，式(2)のような月単位の特徴量だけでなく週単位の特徴量を計算し，この週単位の特徴量プロットから統計値を取得することが重要であると我々は考えた．ここでいう統計値とは，週単位の特徴量プロットにおいて特定の週の高いストレス経験を的確に反映すると期待できる最大値，標準偏差等を意味する．例えば，星マークとして図 2 に示されている各プロットの最大値は，式(4)を用いて計算することができる．

$$a6G\_W_{max} = \max\left[\{a6G\_W_j | 1 \le j \le 4\}\right] \quad (4)$$

式(4)のような特徴量を用いれば，式(2)では区別ができなかった「Person A」と「Person B」の違いが明確になる．この効果は主に中程度の範囲内におけるストレスの高低の識別に有効であることには注意が必要である．ストレスレベルのプロットが連続的に高い場合，或いは連続的に低い場合には，式(2)と式(4)はほぼ同じ値を示すからである．

4. 実験

4.1 データ収集

実験の被験者は全て自発的に実験に参加した健康なオフィスワーカであった．実験に参加することに同意する前に，実験の目的及び手順を説明した．33 人が実験に参加した．33 人のうち 4 人が女性であり，年齢は，20 代が 8 人，30 代が 10 人，40 代が 10 人，50 代が 5 人であった．被験者は

勤務時間中，毎日ウェアラブルセンサ(Empatica 社 E4 センサ[6])を着用するよう要請され，センサによって取得された被験者の生体信号は毎日アップロードされた．被験者はオフィスから離れたときにはセンサは着用しなかった．センサがデータを取得した平均日数は 17 日であった．E4 センサは，4Hz のサンプリングレートで EDA と ST，32Hz で加速度信号を取得する．長期ストレスのアンケートとして PSS アンケートの 10 項目版(PSS10)を使用した．1 ヶ月の実験期間の前後に PSS10 に回答するよう被験者に要請した．実験後の PSS アンケート回答に基づくスコアを長期ストレスの指標とした．図 3 がその分布を示すヒストグラムである．平均は 17.2，標準偏差は 5.7 であり，これは典型的な分布と考えられる[7]．

4.2 実験条件

1)特徴量計算

先行研究[2]を参照し，特徴量を検討した．表 1 に，本実験で使用した特徴量を示す．先行研究[2]の特徴量計算における特徴的な手法の一つは，加速度信号を用いて活動状態(座位，歩行，走行)を区別し，各活動状態のデータ毎に特徴量を計算することである．生体信号はストレスのような精神活動だけでなく，身体活動の影響も強く受けるので，活動状態をいくつかのカテゴリーに区別することは精神状態の影響をより的確に把握する上で好ましい．我々は先行研究[2]と同様，このアプローチを EDA と加速度計信号に適用し，更に，これも先行研究と同様，個々の活動状態の時間的割合をも特徴量とした．尚，表 1 の特徴量は全て，月単位と週単位の双方で計算可能である．

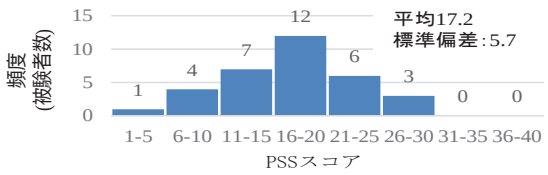


図 3 取得したデータベースにおけるPSSスコアの分布

表 1　特徴量とその合計数

活動状態	モダリティ	特徴量(頻度特徴量は、各ビンの頻度の合計も特徴量に含む)	特徴量数	合計特徴量数	
座位，歩行，走行，全体	規格化EDA	平均	1	91	91*4+24=388
		中央値	1		
		標準偏差	1		
		周波数(11ビン)(0.1Hz幅ビン，0～1.1Hz)	11+1		
		パワースペクトル密度(5ビン)(0.1Hz幅ビン，0-0.5Hz)	5+1		
	EDA ピーク	30秒区間中のピーク数の頻度(11ビン)	11+1		
		1秒区間中のピーク数の頻度(11ビン)	11+1		
		30秒区間中のピークが占める割合	1		
		30秒区間中のピーク数のエントロピー	1		
	EDA ストーム	30秒区間中のストーム数のエントロピー	1		
	加速度RMS	平均	1		
		中央値	1		
		標準偏差	1		
		周波数(11ビン)(0.1Hz幅ビン、0～1.1Hz)	11+1		
		個々のRMSの大きさ(21ビン)(0G-20G)	21+1		
		パワースペクトル密度(5ビン)(1Hz幅ビン，0-5Hz)	5+1		
全体	座位・歩行・走行のパーセンテージ		3	24	
	規格化ST	平均	1		
		中央値	1		
		標準偏差	1		
		周波数(11ビン)(0.1Hz幅ビン、0～1.1Hz)	11+1		
		パワースペクトル密度(5ビン)(1Hz幅ビン，0-5Hz)	5+1		

2)ストレスレベル回帰分析プロセス

ストレスレベルの認識手法として、我々は先行研究のようなクラス分類ではなく回帰分析を選択した．中位を含む全範囲のストレスを認識するのに適切と判断したためである．図 4 を参照し，提案特徴量を使用した場合と使用しなかった場合のストレスレベル回帰分析プロセスを説明する．黒字(赤字を含まない)は，提案特徴量を使用しないプロセスを示す．まず，33 人の被験者の生体信号について月単位で特徴量を計算した．その後特徴量の正規化を実施し，更に，LOOCV(Leave One Out Cross Validation)ループを実行した．各 LOOCV ループでは，1 人を抜いた 32 人の特徴量を訓練データ,PSS スコアを訓練ラベルとし，この訓練データ・訓練ラベルを用い，まず 10 を上限として特徴量選択を実施した．特徴量選択では特徴量と PSS スコアの相関係数をクライテリアとした．次に選択特徴量を用いて SVR(Support Vector Regression)によって抜いた 1 人の PSS スコアを推定し，この推定値を認識 PSS スコアとした．図 4 は，赤字を含めると，提案特徴量が使用された場合のプロセスを示す．特徴量は月単位だけでなく週単位でも計算された．この週単位の特徴量のプロットに対して最大値や標準偏差等の統計値が計算される．LOOCV ループでは，提案特徴量セットと従来特徴量セットの両方から同じ数の特徴量を選択した．提案特徴量は主に中位のストレスレベル認識に有効であり，従来特徴量は上位/下位ストレスレベル認識に有効であるため，両方の特徴セットを使用することが全範囲のストレスレベルの認識には重要であるためである．

4.3 実験結果及び考察

1)提案特徴量の有効性の評価

提案特徴量の有効性を評価するために，PSS スコアとの Pearson の相関係数が最も高い 1 つの提案特徴量に焦点を当てた．4 人の典型的な被験者の当該提案特徴量の週単位プロットを図 5 左に示す．一方，図 5 右の 3 つの棒グラフは，それぞれ，当該提案特徴量，PSS スコア，および当該提案特徴量に対応する従来特徴量の，4 被験者についての値を示している．ここでの従来特徴量は式(2)の $a6G\_M$ であり，提案特徴量は式(4)の $a6G\_W_{max}$ である．継続的にストレスレベルの低い被験者 A および継続的にストレスレベルの高い被験者 D の週単位プロットは，それぞれ低い PSS スコア，高い PSS スコアをもたらすことは明らかであるが，これは PSS スコアの棒グラフで確認できる．当該棒グラフでは，被験者 A のスコアが最低であり，被験者 D のスコアが最高である．また，従来特徴量の棒グラフも，被験者 A, D については同じ傾向を示す．一方，週単位プロットが継続的に中程度である被験者 B と，第 1 週～第 3 週のプロットは低いものの第 4 週のプロットが高い被験者 C については，従来特徴量は PSS スコアと反対の傾向を示す(PSS スコアでは，被験者 C が被験者 B よりも高い一方，従来特徴量はその逆)．一方，提案特徴量は被験者 B,C についても PSS スコアと同じ傾向を示す．前節で指摘した通り，提案特徴量は PSS スコアの特性，つまり，特定の短い区間の強いストレスによって悪化した心理状態に影響を受けて高くなる，を的確に反映している．更に，これも前節で指摘した通り，この提案特徴量の有効性は特に中位(図 5 においては被験者 B,C)のストレスレベルの認識において顕著である．

2)ストレスレベル認識性能

前節で述べたストレスレベル回帰分析プロセスでは，従来法(従来特徴量のみ使用)と提案法(従来特徴量と提案特

徴量の双方を使用)の 2 つの方法において，上限 10 で特徴量選択を行い，選択された特徴量で SVR を行い，SVR による認識 PSS スコアと真のスコアの間の Pearson 相関係数が最も高い場合を選択した．表 2 の「全体」列の，2 つの上の行(「Pearson の相関係数」と「誤差(RMSE)」)は，真の PSS スコアと認識 PSS スコアとの間の Pearson の相関係数と誤差(RMSE(Root Mean Square Error))を示す．「全体」以外の 2 つの列(「低ストレス」および「中+高ストレス」)は，[8]で定義されている閾値(低ストレス：0-13，中ストレス：14-26，高ストレス：27-40)に基づいて真の PSS スコアにより低ストレスおよび中+高ストレスという二つのクラスに被験者を分けた場合の値を示している．高ストレスと中ストレスをひとつのクラスにまとめた理由は，用いたデータベースにおいて高ストレスクラスに 2 人の被験者しかおらず，しかもその 2 人の PSS スコアが閾値である 27 であったため，中ストレスクラスにまとめるべきだと判断したためである．「全体」列，上 2 行に示される値には，提案法の有効性が示されている．Pearson の相関係数は 0.66 から 0.72 に向上し，RMSE が 4.32 から 4.03 に改善した．他の 2 つの列でも，提案法の使用による改善が示されている．

3)ストレスレベル順序認識性能

我々はまた，認識ストレスレベルの順序を用いて結果を評価した．例えば，図 5 の右側の棒グラフでは，4 人の被験者の真のストレスレベルの順序は[A，B，C，D]である一方，従来特徴量による認識ストレスレベルの順序は，[A，C，B，D]である．この相違を調べることは，実用的な有用性を調査する上で重要である．なぜなら，実運用においては，ストレスレベルで並べた時，高い方にいる人は，相対的にリスクが高い人と判定する場合がある為である．前述

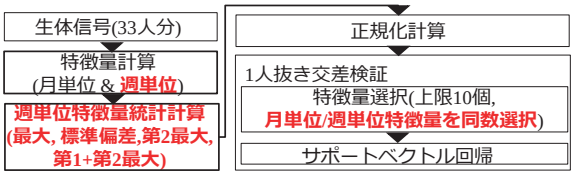


図 4 PSSスコア回帰分析プロセス

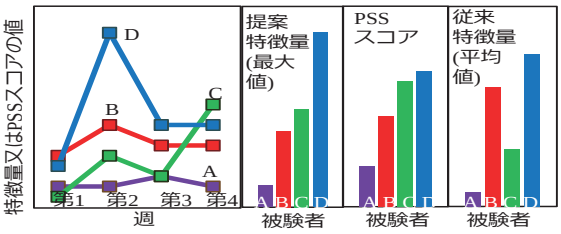


図 5 例示特徴量の週単位プロット(左)と，提案法及び従来法による月単位計算結果及び比較のためのPSSスコア(右)

表 2 提案法及び従来法のPearsonの相関係数，誤差(RMSE)，平均序数差，再現率-適合率カーブ下面積

		全体	低ストレス	中+高ストレス
Pearsonの相関係数	従来法	0.66	0.22	0.67
	提案法	0.72	0.37	0.70
誤差(RMSE)	従来法	4.32	6.24	3.50
	提案法	4.03	5.66	3.34
平均序数差	従来法	6.58	6.63	6.56
	提案法	5.45	5.63	5.40
再現率-適合率カーブ下面積	従来法	0.92		
	提案法	0.95		

の[A，B，C，D]と[A，C，B，D]の序数の差は，B，C についてそれぞれの序数が 2 と 3，3 と 2 で 1 つずつ異なるので，合計 2 となる．合計序数差を被験者数で割って，「平均序数差」も算出できる．提案法・従来法による認識 PSS スコアと真の PSS スコアとの間の平均序数差を，表 2 の「平均序数差」行に示す．当該行において最大の改善は中+高ストレスクラスであり，四捨五入すれば 7 から 5 への改善である．PSS スコア分布(図 3 参照)では，頻度が 7 より大きいビンが 1 つしかない(該当するのは PSS スコア 16-20 のビンで，頻度は 12)一方，PSS スコア 14-27 の中+高ストレス領域(PSS スコア 11-15，16-20，21-25，26-30 の各ビンで包含される領域)では半分以上のビンの頻度が 5 より大きい．序数差と比較してビンの頻度が大きいということは，異なるビンつまり PSS 分布の異なる位置と見なされる可能性が低いことを意味する為，これは顕著な改善であると言える．

4)ストレス識別性能

我々は中+高ストレスクラスを低ストレスクラスから識別する性能も評価した．我々の目的の一つは，中程度のストレスを感じている人を特定して高ストレスにならないよう注意を促す為だからである．低ストレスから中+高ストレスを識別するための再現率-適合率曲線の AUC(Area Under the Curve)を表 2 の最下段に示す．従来法の 0.92 に対し提案法は 0.95 となった．この顕著な改善(誤差削減幅約 40%)は，提案法が，中程度のストレスの人を低ストレスの人から区別するのにも有用であることを示している．

5. 結論

我々は長期ストレスレベルの認識においてより高い精度を達成する為，PSS がストレスレベルを認識する月単位だけでなく週単位で算出した特徴量も使用する方法を提案し，提案法を実験的に評価した．評価の為，オフィスワーカ 33 名の 1 ヶ月データを使用した．結果，提案法により様々な性能の改善が得られることが示された．例えば，Pearson の相関係数は 0.66 から 0.72 に改善された．

参考文献

[1] A. Alberdi, et al., “Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review,” Journal of biomedical informatics, vol. 59, pp. 49–75, 2016.

[2] A. Sano et al., “Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones,” in Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on, 2015, pp. 1–6.

[3] A. Sano, “Measuring College Students’ Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones,” Massachusetts Institute of Technology, 2015.

[4] S. Cohen, et al., “A Global Measure of Perceived Stress.” Journal of Health and Social Behaviour1983.

[5] S. Cohen, et al., “Measuring Stress: A Guide for Health and Social Scientists.” Oxford University Press, 1997.

[6] https://www.empatica.com/research/e4/

[7] S. Cohen and G. M. Williamson, ”Perceived Stress in a Probability Sample in the United States”, The Social Psychology of Health, 1988

[8] State of New Hampshire Employee Assistance Program, “Perceived Stress Scale”, URL: https://das.nh.gov/wellness/Docs/Percieved%20Stress%20Scale.pdf

[2G1] [General Session] 11. Robot / Real World

座長:長井 隆行（電通大）

Wed. Jun 6, 2018 9:00 AM - 10:20 AM Room G (5F Ruby Hall Hiten)

[2G1-01] Comparison of Methods for Localizing A Target with UAV

○Ryosuke Matsuo[1], Koya Kishi[1], Koichi Hori[1], Takehisa Yairi[1] （1. The University of Tokyo）

9:00 AM - 9:20 AM

[2G1-02] Difference evaluation of daily-life activities in physical and immersive VR environments

○Yoshiaki Mizuchi[1], Tetsunari Inamura[1,2] （1. National Institute of Informatics, 2. SOKENDAI (The Graduate University for Advanced Studies)）

9:20 AM - 9:40 AM

[2G1-03] Generation of presentation motion using sequence to sequence model

○Akihito Shimazu[1], Chie Hieida[1], Takayuki Nagai[1], Tomoaki Nakamura[1], Yuki Takeda[2], Takenori Hara[2], Osamu Nakagawa[2], Tsuyoshi Maeda[2] （1. The University of Electro-Communications, 2. Dai Nippon Printing Co., Ltd.）

9:40 AM - 10:00 AM

[2G1-04] Validation of Effective Active Learning Method in Semantic Learning of Context-dependent Motions

○Tatsuya Sakato[1], Tetsunari Inamura[1,2] （1. National Institute of Informatics, 2. SOKENDAI(The Graduate University for Advanced Studies)）

10:00 AM - 10:20 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

UAVを用いた電波発信源の位置推定のための手法比較

Comparison of Methods for Localizing A Target with UAV

松尾 凌輔 Ryosuke Matsuo　来住 航也 Koya Kishi　堀 浩一 Koichi Hori　矢入 健久 Takehisa Yairi

東京大学大学院工学系研究科
School of Engineering, The University of Tokyo

There are many applications with UAV. Localizing target which has radio source is interesting challenge for UAV and it enable to search missing people, wild life and so on. There are many researches for localizing target in context of Wireless sensor networks(WSNs) and some researches for localizing target with UAV. There are some differences between WSNs' localization and localizing target with UAV. So, this paper compares methods, which are proposed in these researches and applicable to localizing a target with UAV problem, via simulation and organizes these methods from the aspect of calculation cost, accuracy, robustness and trajectory of UAV.

1. 背景

近年，UAV の技術は急速に発展し，農業やリモートセンシングなど様々な応用がなされるようになってきている．その応用の1つとして，遭難者や保護対象の動物などのターゲットの追跡が考えられており，いくつかの研究がなされてきている．しかし，それらの研究の多くでは，視界が開けた環境でターゲットを視認できることが前提となっているため，ターゲットとの距離が離れている場合や森林地帯などでは追跡ができない．そこで，ここでは，ターゲットが電波を発信する装置を身に着けていることを前提とすることで，ターゲットを視認できないような環境でターゲットの位置を推定する問題を考える．

電波発信源の位置推定問題は，Wireless Sensor Networks (WSNs)の文脈で研究されてきている．そこでは，大量のデバイスをばらまき，そのいくつかがアンカーとして GPS などによって自己位置を知っている状態を想定する場合が多い．そして，通信の可否や電波強度(RSSI)などを用いて位置が未知のデバイスの位置を推定するという問題が考えられている．さらに，UAV による電波発信源の位置推定問題により近い問題設定として，アンカーが移動可能な状態を想定した研究もなされている．

今回想定している UAV による電波発信源の位置推定問題と WSNs における位置推定問題は類似した問題であるが，いくつかの相違点も存在する．

- 位置が未知のデバイス間の通信はできない

WSNs の研究では，位置が未知のデバイス間の通信が可能であるため，通信可能なデバイス間をリンクで接続し，ネットワークを形成できる．これにより，リンクの関係に基づく位置推定が可能になる．しかし，私たちが想定する問題では，ターゲットが電波を発信することのみを想定しているため，複数のターゲットが存在する環境でも，ターゲット間での通信はできない．

- アンカーデバイスが少数である

WSNs の研究では複数のアンカーデバイスが存在し，アンカーデバイス間の距離が既知であるため，その情報を利用することができる．今回の問題でアンカーデバイスに対応するのは UAV であり，移動可能なアンカーデバイスが少数あるいは 1 つ存在する状態での位置が未知のターゲット（デバイス）の位置を推定する問題だとみなすことができる．

連絡先：松尾凌輔，　ryosuke129@g.ecc.u-tokyo.ac.jp

2. 比較手法

WSNs での位置推定には，大きく分けて二つの方法が存在する．Range-based と Range-free である．

Range-based の手法は，RSSI 値から算出できる距離情報や Angle Of Arrival (AOA), Time Of Arrival (TOA), Time Difference Of Arrival (TDOA)などを利用する手法である．RSSI は追加のデバイスなしで取得できる情報だが，ノイズが大きいという問題点がある．その他の情報は，取得するために追加のデバイスが必要になるという問題がある．

Range-free の手法は，通信の可否を利用した手法である．追加のデバイスが必要ないことやノイズに影響されにくいという長所があるが，利用する情報が少ないために，精度が低いという問題がある．

今回は，WSNs で研究されてきた手法の中で，追加のデバイスが必要ない手法，つまり，Range-free の手法と Range-based の手法の中で RSSI を用いるもので，なおかつ，1 つの UAV（移動可能なアンカーデバイス）で，ターゲット（位置が未知のデバイス）の位置を推定することができる手法に限定して比較を行う．それ以外のいくつかの代表的な手法に関しては 5 章の関連研究で簡単に述べる．

2.1 Centroid Method(CM)

ターゲットの電波を受信したときの座標を記録し，その座標の平均をターゲットの推定位置とする手法である[Bulusu 00].

$$\hat{\boldsymbol{x}} = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i}{n} \qquad (1)$$

ここで，$\hat{\boldsymbol{x}}$は推定座標，$\boldsymbol{x}_i$はi回目にターゲットからの電波を受信できた座標，nはターゲットからの電波を受信できた回数である．

2.2 Weighted Centroid Method(WCM)

これは，Centroid Method の改良として，提案されたものである[Blumenthal 07]．障害物のない空間では RSSI 値と距離の関係は以下の式で書ける．

$$RSSI = P_1 - 10\kappa \log_{10} d \qquad (2)$$

P_1は 1m 離れた地点での RSSI 値, κは減衰係数である. この式より, 距離dは,

$$d = 10^{\frac{-RSSI+P_1}{10\kappa}} \propto 10^{\frac{-RSSI}{10\kappa}} \quad (3)$$

と書ける. ここで, 重みを

$$w_i = \frac{1}{d_i^k} \propto 10^{\frac{k\times RSSI_i}{10\kappa}} \quad (4)$$

とする. ここで, kは重みを距離の何乗に比例させるかを決めるパラメータで, [Blumenthal 07]では, シミュレーションで最適値を調べており, 電波の到達できる範囲に依存していると述べている. 例えば, 電波の到達距離が 10m のときは$k = 1$が最適であり, 30m のときは$k = 3$が最適である. この重みを用いて重み付き重心をとることで, 距離が近い受信位置に重心を引き寄せることができ, 精度の向上が期待できる.

$$\hat{\boldsymbol{x}} = \frac{\sum_{i=1}^{n} w_i \boldsymbol{x}_i}{\sum_{j=1}^{n} w_j} \quad (5)$$

2.3 Multilateration(ML)

[Motter 13]では, 3 地点での RSSI の測定結果から距離を算出して, ターゲットの位置を推定している. 3 地点以上で観測が得られる場合には, それらから 3 地点を選ぶすべての組み合わせに対して, Mutilateration を行い, 平均を推定結果とすることで精度が向上することを示した.

[Greco 15][Naglaa 16]は, UAV を活用した電波発信源位置推定の研究で, ここでは, 大量の測定情報を集め, 最小二乗法を用いることで, すべての測定結果についての二乗誤差和が最小となる位置をターゲットの推定位置としている. 具体的な定式化は以下のようにできる(定式化の流れは[Naglaa 16]より).

$$d_i^2 = \left|\boldsymbol{x}_i^u - \boldsymbol{x}_i^t\right|^2 = \left|\boldsymbol{x}_{u,i}\right|^2 + \left|\boldsymbol{x}_{t,i}\right|^2 - 2\boldsymbol{x}_{u,i}\cdot\boldsymbol{x}_{t,i} \quad (6)$$

ここで, d_i はi回目にターゲットを観測できたときの距離である. $\boldsymbol{x}_{u,i} = (x_{u,i}, y_{u,i}, z_{u,i})$は UAV の位置で, $\boldsymbol{x}_{t,i} = (x_t, y_t, z_t)$はターゲットの位置である. ここで, $\left|\boldsymbol{x}_{u,i}\right| = p_{u,i}, \left|\boldsymbol{x}_{t,i}\right| = p_t, z_t = 0$とおくと,

$$d_i^2 - p_{u,i}^2 = p_t^2 - 2(x_{u,i}x_t + y_{u,i}y_t) \quad (7)$$

さらに

$$h = \begin{bmatrix} d_1^2 - p_{u,1}^2 \\ d_2^2 - p_{u,2}^2 \\ \vdots \\ d_n^2 - p_{u,n}^2 \end{bmatrix}$$

$$G = \begin{bmatrix} -2x_{u,1} & -2y_{u,1} & 1 \\ -2x_{u,2} & -2y_{u,2} & 1 \\ \vdots & \vdots & \vdots \\ -2x_{u,n} & -2y_{u,n} & 1 \end{bmatrix}$$

$$\theta = [x_t \;\; y_t \;\; p_t^2]^\top$$

と書けば,

$$h = G\theta \quad (8)$$

よって, 最小化すべき式は,

$$\left|\left|G\theta - h\right|\right|^2$$

となり, θについての導関数を 0 とおけば,

$$\theta = (G^\top G)^{-1} G^\top h \quad (9)$$

が求まる.

2.4 Weighted Multilateration(WML)

[Naglaa 16]で提案された手法で, Weighted Centroid Method と似た処理をする. つまり, 距離の逆数を重みとして用いる. これは, 距離が近いときの情報を重要視するということである. 数式で書けば,

$$w_i = \frac{1}{d_i}$$

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & w_n \end{bmatrix}$$

$$h \leftarrow Wh$$

$$G \leftarrow WG$$

として, 2.3 と同様に,

$$\theta = (G^\top G)^{-1} G^\top h$$

で求まる.

2.5 Particle Filter(PF)

Particle Filter を用いて電波発信源の位置推定を行っている研究としては, [Caballero 08]や[Fujita 17]がある. 以下では, パーティクルを 2 次元の電波発信源の位置座標として, 式を導出する.

(1) 重みの更新

i 回目にターゲットを観測できたとき, j 個目のパーティクルに対して以下の計算を行う.

$$R_i^{(j)} = P_1 - 10\kappa \log_{10} d_i^{(j)} \quad (10)$$

$$d_i^{(j)} = \sqrt{\left(x_{u,i} - x_{p,i}^{(j)}\right)^2 + \left(y_{u,i} - y_{p,i}^{(j)}\right)^2 + \left(z_{u,i}\right)^2} \quad (11)$$

$$\widetilde{w}_i^{(j)} = \exp\left(-\frac{\left(RSSI_i - R_i^{(j)}\right)^2}{\sigma^2}\right) w_{i-1}^{(j)} \quad (12)$$

$$w_i^{(j)} = \frac{\widetilde{w}_i^{(j)}}{\sum_{k=1}^{M} \widetilde{w}_i^{(k)}} \quad (13)$$

$R_i^{(j)}$は, j番目のパーティクルと UAV の距離から, 観測されるべき電波強度であり, この値と実際に観測された電波強度$RSSI_i$のずれによって, そのパーティクルの重みを決定している. また, Mはパーティクルの数である.

(2) リサンプリング

リサンプリングを行うかどうかの判定には以下の式を用いる.

$$N_{eff} = \frac{1}{\sum_{k=1}^{L} w_i^{(k)}} \quad (14)$$

[Caballero 08]では, N_{eff}が$0.1M$を下回るとき, リサンプリングを行うとしている. リサンプリングは等間隔サンプリングで行う.

3. シミュレーション

3.1 シミュレーション 1

式(2)の$P_1 = -40, \kappa = 2$とおいて, 平均 0, 分散は$\sigma = \{1,5,10,15,25\}$dBの 5 種類のホワイトノイズを加える.

$$RSSI = -40 - 20\log_{10} d + \mathcal{N}(0, \sigma^2) \quad (15)$$

この式に従った値を UAV は受信するとし, 受信できるのはノイズを含んだ RSSI 値が「50m 離れた場合の RSSI の真値」を上回るときとする.

シミュレーションの手順としては, まず電波発信源を 200m 四方の領域にランダムに 1 つ配置する. 次に UAV を高度は 20m とし, 図 1 の左図の軌道で飛行させ, 電波発信源の位置を 2 章で述べた 5 つの手法で推定する(WCM の重みのパラメータは

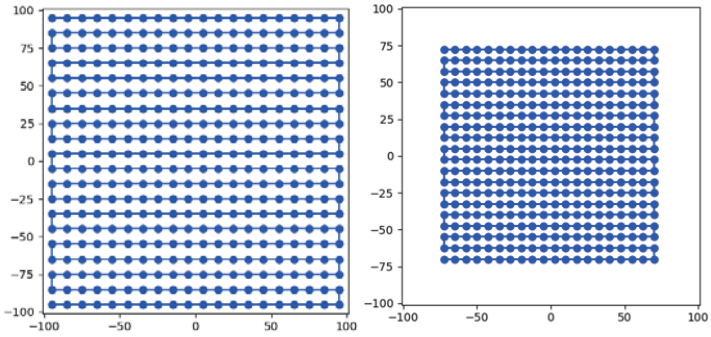


図 1 UAV の軌道（左図：シミュレーション 1，右図：シミュレーション 2）

表 1 手法比較($\sigma = 5$, N:電波受信回数, M:パーティクルの数)

手法	CM	WCM	ML	WML	PF
計算量	$O(N)$	$O(N)$	$O(N^2)$	$O(N^2)$	$O(NM)$
平均誤差(m)	27.9	13.8	47.6	40.7	17.6
最大誤差(m)	60.4	36.8	88.4	74.6	42.2
最小誤差(m)	2.3	2.6	9.1	3.5	2.7

k=3 とする). これを 50 回繰り返し，平均誤差を算出することで，精度の比較を行った．また，この手順を 5 種類の標準偏差のノイズに対して行うことで，ノイズに対するロバスト性の比較も行った．

3.2 シミュレーション 1 の結果

まず計算コストについて考えると，計算量のオーダーは表 1 のようになる．ここで，Nは電波の受信回数，Mはパーティクルの数である．50 回のシミュレーションでの平均観測回数は 88.7 で，Particle Filter のパーティクルは 400 個用いたため，計算量のオーダーは CM＝WCM＜ML＝WML＜PF となる．

次に平均誤差を見ると，WCM＜PF＜CM＜WML＜ML となっており，計算コスト，精度のどちらについても WCM が優れているといえる．

図 2 を見ると，ノイズの標準偏差＝1 のときの平均誤差は WML<PF<ML<WCM<CM となっている．また，重み付きの手法の精度がノイズの増加に伴って悪くなっていることがわかる．PF が RSSI 値を使っているにもかかわらず，ノイズに対して頑強なのは，式(12)のσ^2を調整することができるからである．式(12)のσ^2の影響はシミュレーション 3 で述べる．

3.3 シミュレーション 2

CM および WCM は，使える情報が少ないため，ほかの手法に比べ，飛行経路に精度が依存すると考えられる．よって，飛行経路が偏った状況を考え，次のシミュレーションを実施することとした．

UAV の軌道を図 1 右図のように設定し，標準偏差$\sigma = 5$のホワイトノイズを加えた RSSI を用い，各手法の精度の比較を行った．PF においては，ばらまくパーティクルの数をシミュレーション 1 と同じ 400 とし，なおかつ，初期パーティクルを配置する領域もシミュレーション 1 と同様とした．

3.4 シミュレーション 2 の結果

表 1 と表 2 を比較すると，CM，WCM，PF が UAV の軌道の影響をほかの手法と比較して大きく受けていることがわかる．CM

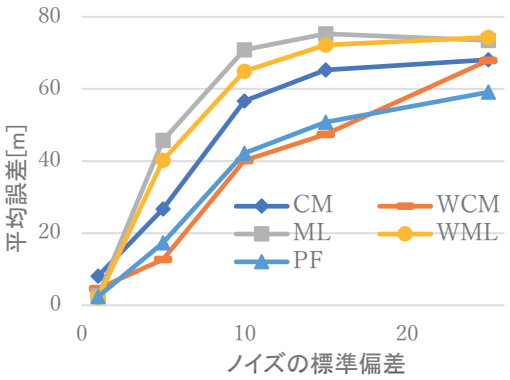


図 2 ノイズに対するロバスト性の比較

表 2 手法比較($\sigma = 5$, シミュレーション 2)

手法	CM	WCM	ML	WML	PF
平均誤差(m)	37.9	24.4	51.3	46.8	30.8
最大誤差(m)	81.6	60.1	108.4	99.3	76.1
最小誤差(m)	4.4	2.5	8.2	8.1	6.0

と WCM は，観測点の集合の内部に真値がある必要があるため，シミュレーション 2 は領域を網羅していないことから，誤差が大きくなっていると考えられる．PF，ML，WML の誤差が大きくなっているのは，観測の回数が減っていること（平均観測回数が 62.3）と，観測できるときの距離が大きくなっていることが原因として考えられる．今回は，観測ノイズを RSSI 値に対して加えているため，距離が大きいときに，RSSI 値のノイズが距離誤差に大きく効いてくる．また，PF の誤差の増加率が ML や WML と比較して大きい原因は粒子の収束であると考えられる．ML や WML では，最小二乗解を求めるため，すべての観測情報を有効に使えるが，PF では，初期の観測で大きなノイズがあると，誤った解周辺に粒子が収束して，真の解周辺から粒子が消失することが考えられる．これは，図 2 の PF の勾配が，ML や WML と比較して大きくなっていることも説明できる．

3.5 シミュレーション 3

式(12)のσ^2の影響を調べるために，σを変化させ，その時の平均誤差の推移を調べた．飛行経路や電波発信源の条件はシミュレーション 1 と同じとし，平均誤差は 100 回のシミュレーションの平均を用いた．

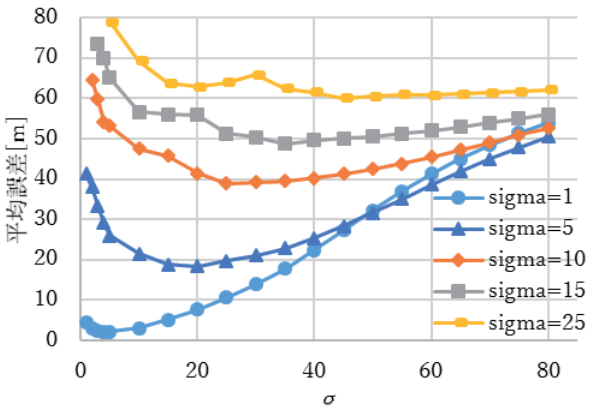


図 3 σの平均誤差へ影響

3.6 シミュレーション 3 の結果

図 3 の凡例の sigma は観測ノイズの標準偏差を意味している. 図 3 から, 観測ノイズによって, 平均誤差が最も小さくなるσの値が変化することがわかり, その値は観測ノイズの 2〜5 倍程度である.

4. 結論

本稿では, シミュレーションにより, UAV を用いた電波発信源の位置推定手法の比較を行った. シミュレーションの結果から, 観測ノイズがある程度大きい場合は, 単純で計算コストの低い Centroid Method あるいは Weighted Centroid Method を用いるとよく, 観測ノイズが非常に大きい場合は, Particle Filter が良いことがわかる.

一方で観測ノイズが非常に小さい場合には, (Weighted) Multilateration や Particle Filter などを用いるとよいと考えられる. (Weighted) Multirateration と Particle Filter の違いの 1 つには, (Weighted) Multilateration が点推定であるのに対して, Particle Filter では分散も推定できるという点が挙げられる. また, これらの手法のどちらを用いるかは, 計算コストと精度のどちらを優先するかでも変わってくる. 精度を優先する場合は Particle Filter となり, 粒子の数を増やすことで, 精度の向上を期待できる. 計算コストを優先する場合は, (Weighted) Multirateration となる.

5. 関連研究

WSNs における位置推定手法

[Dragos 01]では, hop-distance に基づく位置推定手法が提案されている. あるデバイスと通信可能なデバイス間の距離を 1-hop と定義し, デバイス間の距離を hop 数で表す. 次にあるアンカーデバイスから別のアンカーデバイスまでの hop-distance と実際の距離の関係を用いて 1-hop の距離を計算する. 1-hop の距離がわかれば, 位置が未知のデバイスの複数のアンカーデバイスからの距離が計算できるため, 位置を推定できる.

[He 03]では, Approximate Point-In-Triangulation Test という手法が提案されている. この手法では, 位置が未知のデバイスの信号を受信できたアンカーデバイスで三角形を形成し, 三角形の内側にデバイスが存在するかどうか判定し, 内側にデバイスが存在する三角形の重複領域の重心を推定位置とする手法である.

[Shang 03]では, 多次元尺度構成法(MDS)に基づいて, 位置が未知のノードを復元する手法である. ここでは, 未知デバイス間で通信が可能であり, 無知デバイス間の通信の可否からネットワークを形成する. 直接つながっているデバイス間の距離を一定とし, つながっていないデバイス間の距離はダイクストラ法などを用いることで距離行列を形成する. その距離行列から MDS を用いて未知デバイスの配置を推定する.

[Ferris 07]では, 位置が未知の移動可能なデバイス(ロボット)の位置を電波強度から推定する手法である. ここでは, Gaussian Process Latent Variable Model(GP-LVM)を用いている.

[Hanen 17]では, 機械学習を用いた RSSI に基づく位置推定手法についてまとめてある. 機械学習による手法は 2 ステップに分けることができる. 1 ステップ目は訓練ステップで, アンカーデバイスと位置が既知の訓練デバイスを用意して, 電波強度と位置の関係を学習する. 2 ステップ目は, 訓練されたモデルで位置を推定するステップで, モデルにアンカーデバイスからの電波受信強度を入力として入れ, 位置を出力する.

参考文献

[Bulusu 00] Bulusu, Nirupama, John Heidemann, and Deborah Estrin. "GPS-less low-cost outdoor localization for very small devices." IEEE personal communications 7.5 (2000): 28-34.

[Blumenthal 07] Blumenthal, Jan, et al. "Weighted centroid localization in zigbee-based sensor networks." Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on. IEEE, 2007.

[Motter 13] Motter, Paulo, et al. "Practical issues in wireless sensor network localization systems using received signal strength indication." Sensors Applications Symposium (SAS), 2011 IEEE. IEEE, 2011.

[Greco 15] Greco, G., et al. "Localization of RFID tags for environmental monitoring using UAV." Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2015 IEEE 1st International Forum on. IEEE, 2015.

[Naglaa 16] El Agroudy, Naglaa, Niko Joram, and Frank Ellinger. "Low power RSSI outdoor localization system." Ph. D. Research in Microelectronics and Electronics (PRIME), 2016 12th Conference on. IEEE, 2016.

[Caballero 08] Caballero, Fernando, et al. "A particle filtering method for wireless sensor network localization with an aerial robot beacon." Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on. IEEE, 2008.

[Fujita 17] Fujita, Shinya, et al. "An active beacon-based target localization method for Unmanned Aerial Vehicles with particle filter." Information Fusion (Fusion), 2017 20th International Conference on. IEEE, 2017.

[Dragos 01] Niculescu, Dragos, and Badri Nath. "Ad hoc positioning system (APS)." Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE. Vol. 5. IEEE, 2001.

[He 03] He, Tian, et al. "Range-free localization schemes for large scale sensor networks." Proceedings of the 9th annual international conference on Mobile computing and networking. ACM, 2003.

[Shang 03] Shang, Yi, et al. "Localization from mere connectivity." Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing. ACM, 2003.

[Ferris 07] B. Ferris, D. Fox, and N. Lawrence. WiFi SLAM Using Gaussian Process Latent Variable Model. In JCAI, 2007.

[Hanen 17] Ahmadi, Hanen, and Ridha Bouallegue. "Exploiting machine learning strategies and RSSI for localization in wireless sensor networks: A survey." Wireless Communications and Mobile Computing Conference (IWCMC), 2017 13th International. IEEE, 2017.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

実環境と没入型VR環境における日常生活行動の差異の評価

Difference evaluation of daily-life activities in physical and immersive VR environments

水地 良明 *1　稲邑 哲也 *1*2
Yoshiaki Mizuchi　Tetsunari Inamura

*1国立情報学研究所
National Institute of Informatics

*2総合研究大学院大学
SOKENDAI (The Graduate University for Advanced Studies)

Motion capturing and analysis is gaining attention to realizing intelligent systems that support human daily-life activities. Utilizing immersive VR systems is a solution to reduce costs of developing experiment environments. Although recent improvements in 3D head-mounted displays and hand tracking controllers enable us to capture natural human behavior in various scenarios easily, there is concern about the inconsistency of human behavior between a real physical environment and a simulated VR one. In particular, the restriction of the field of view (FOV) is known as a major factor in causing the inconsistency such as underestimation of distance. However, the influence of FOV restriction on daily-life activities concerning spatial perception and object manipulation in the VR environment is still unclear. In this paper, we evaluate the difference in human behavior depending on the FOV restriction in the real and VR environments from comparison results of completion time of an object manipulation task.

1. はじめに

ロボットの知能獲得において，動作スキルの獲得 [Levine 16] やデータドリブンな対人行動の生成 [Liu 16] などが実現されており，身体的・社会的な対話経験を観測，蓄積する重要性が高まっている．特に，日常生活行動をドメインとする動作スキルや行動モデルの獲得では，様々な状況で被験者による実演を通して，バリエーションに富んだ膨大な人間の挙動データを観測，蓄積することが求められる．また，単純な動作計測のみならず，発話内容や発話タイミング，視線情報，相対的な位置関係などマルチモーダル情報を記録することが望まれる．実環境において人間の行動観測を実現する実験システム [Sumi 07, Brščić 15] も実現されているものの，センサの性能や遮蔽による制約が存在するため，視線などを含めた全身の身体動作や環境中の物体とのインタラクションを計測することは難しい．また，実験環境の構築やロボットの維持管理，被験者の確保に起因する膨大な実験コストも障壁となっている．

これらの問題を解決する方法の一つは，没入型仮想現実（VR）を活用し，VR 環境中での人間の挙動を観測する方法である．実環境に比べて，安全性の確保や実験条件の制御，再現性の保証が容易であることも没入型 VR を活用する利点として挙げられる．近年では，ヘッドマウントディスプレイ（HMD）やハンドトラッキングが可能なコントローラなどの VR デバイスの発展により，高い没入感が実現されており，人間の挙動を低コストで容易に観測，蓄積することが可能となっている．稲邑らは，それらの機能を活用することで，人間がアバターとしてログインし，仮想ロボットとの対話を可能とする SIGVerse と呼ばれる VR システム [Inamura 10] を提案している．また，SIGVerse ver.3[Mizuchi 17] では，Unity と ROS を統合することで，様々な VR デバイスとロボットの制御ソフトウェアの活用した身体的・社会的対話実験の実施が可能となっている．このような没入型 VR を活用した人間の挙動の観測により，家事タスクのスキル獲得 [Bates 17] やロボット競技会における知能ロボットの評価 [Inamura 17] が実現されている．

連絡先: 水地 良明，国立情報学研究所，〒 101-8430 東京都千代田区一ツ橋 2-1-2，mizuchi@nii.ac.jp

高い没入感が実現されているものの，VR 環境の行動で得る被験者が得る感覚フィードバックは，実環境と比べて差異がある．そのため，実環境と VR 環境における人間の挙動の不整合が懸念されている．没入型 VR を活用して観測した人間の挙動データに基づいて実環境での人間支援システムを構築するためには，両環境における差異を明らかにすることが重要である．人間の挙動への影響が懸念される要素の一つが，HMD を利用することによる視野の制約である．日常生活行動において重要な空間認知や物体操作に対する視野の制約による影響は，明らかにすべき重要な要素の一つである．

実世界と VR 環境における人間の挙動の差異を検証する取り組みとして，ロボットとの協調作業における物体の手渡し動作への影響 [Duguleana 11] やジェスチャ認識における空間認知 [El-Shawa 17] に着目した研究が報告されている．Willemsen ら [Willemsen 09] は，HMD の重さや慣性，装着感，視野の制約による距離の過小知覚に対する影響を明らかにしている．しかしながら，従来研究では日常生活行動における空間認知や物体操作に対する視野の制約による影響は明らかになっていない．

本研究では，実環境と没入型 VR 環境における視野の制約が人間の日常生活行動へ与える影響について検証する．人間の挙動の差異を比較するため，指定された物体を一つずつ識別し，指定された場所に移動させる単純な物体操作タスクを設計し，達成時間を計測した．また，実環境において VR 環境と同じように視野を制限するためのゴーグル（Mock HMD）を作成し，視野の違いが与える影響について検証した．

2. 評価方法

2.1 実験環境

視野の制約に依存する人間の挙動を検証するため，単純な物体操作タスクを設計した．図 1 に，構築した実験環境を示す．環境の中央には，被験者に座ってもらうための回転式の椅子が設置されている．その両隣には，机が設置されており，一方には 15 個のブロックが操作対象として配置されている．実環境と VR 環境における重さによる挙動の差異を軽減するた



図 1: 実験環境の構成

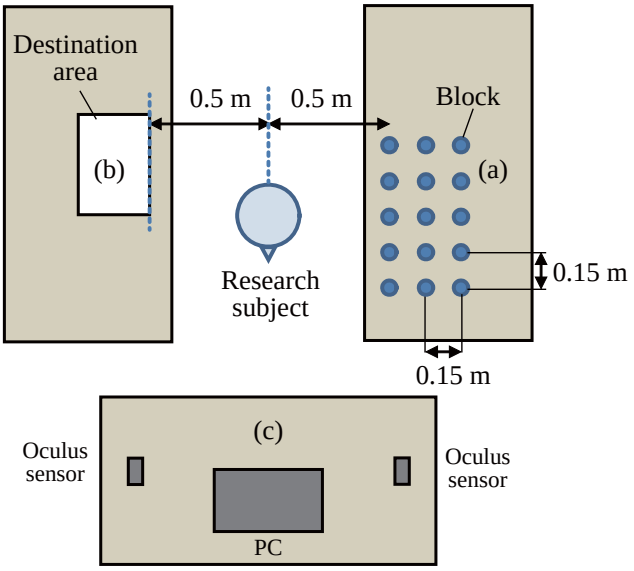


図 2: 実験環境のレイアウト

め，10g 程度の極軽量なブロックを採用した．ブロックは 7 種類の色のいずれかに属しており，3 つごとに特定の形に属している．また，各色には最大 3 つのブロックが属する．ブロックどうしの間隔は 15cm とした．もう一方の机には，物体の移動先として，A3 サイズの紙が貼付されている．ブロックと紙の間隔は 100cm となっており，いずれも被験者の手が届く範囲に配置されている．従って，ブロックと移動先となる紙が同時に視野に収まらない配置となっており，被験者は周りを見渡して物体の認識と把持，移動を行う必要がある．

図 3 に，構築した VR 環境を示す．環境は Unity 3D を用いて構築した．ブロックや紙のサイズ，机の高さは実環境と同じである．また，ブロックや紙の配置が実環境と VR 環境で一致するように調整を行った．

2.2 デバイス

被験者が VR 環境にログインするためのインタフェースとして，Oculus Rift CV1(HMD) と Oculus Touch(ハンドヘルドコントローラ) を利用した．HMD の位置・姿勢が VR 内のアバターの視線に反映され，奥行きのある一人称視点画像が被験者に提供されるようになっている．同様に，コントローラの位置・姿勢がアバターの手に反映される，アバターの手が接触した状態でコントローラのトリガーを引くことで，実世界に近

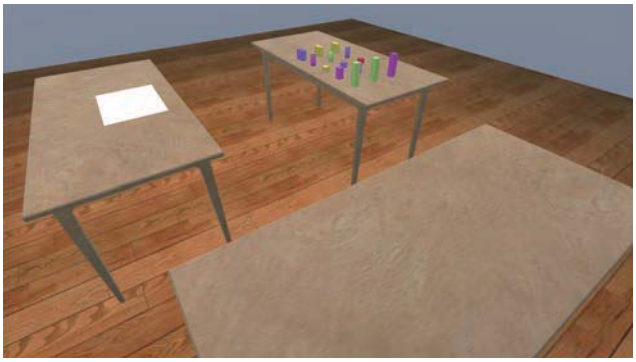
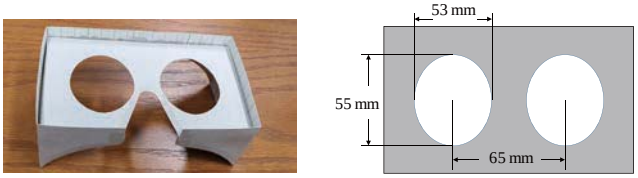


図 3: VR 環境に構築した実験環境



(a) (b)

図 4: Mock HMD の外観と構成

い感覚で仮想物体を把持，移動できるようになっている．

実環境において Oculus Rift CV1 と同じように被験者の視野を制限するため，HMD を模したゴーグル (Mock HMD) を作成した．図 4 に，Mock HMD の外観と設計を示す．レンズと目の間隔を 12mm とした場合の Oculus Rift の片目の視野角は，水平 84°，垂直 93° となっている．また，両目での視野角は，水平 94°，垂直 93° となっている．これらの仕様に従って，目の中心と視野を制限する穴との間隔が 25mm であると仮定し，Mock HMD の仕様を決定した．

2.3 実験条件

実験では，3 つの条件を設定した．図 5 に，各条件における被験者の様子を示す．一つ目の条件では，実環境において視野に制約のない状態で被験者にタスクを実施してもらった．二つ目の条件では，実環境において Mock HMD を利用して視野を制限した状態で被験者にタスクを実施してもらった．三つ目の条件では，被験者に VR 環境にログインしてもらい，タスクを実施してもらった．VR インタフェースに起因する重さや慣性，装着感などの影響の差を軽減するため，実環境においても HMD を頭に乗せ，手にコントローラを固定した状態で，被験者にタスクを実施してもらった．

被験者には，指定された 3 つのブロックを図 2 中の (a) から (b) に移動してもらった．移動対象となるブロックは，特定の色もしくは形状で指定した．指定された色または形状に該当



(a) Unrestricted (b) Mock HMD (c) HMD (VR)

図 5: 各条件における被験者の様子

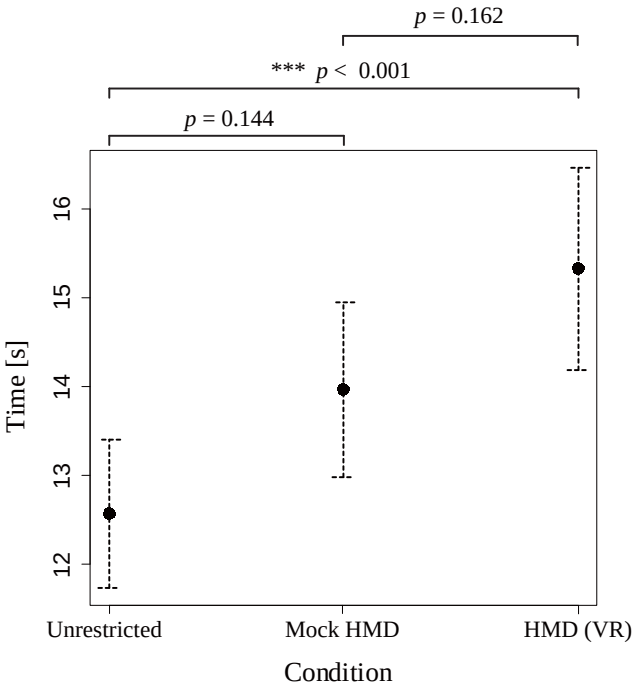


図 6: 各条件における達成時間の平均と 95%信頼区間

するブロックは，必ず 3 つのみ存在するようになっている．ブロックを操作する際には，右手のみを用いて，一つずつ操作してもらった．ブロックの配置は，試行ごとにランダムに変更した．各試行の開始前に，被験者がブロックの配置を把握できないようにするため，図 2 中の (b) 方向を向いて待機してもらった．各試行の開始時に，図 2 中の (c) 方向を向いてもらい，それと同時に時間の計測を開始した．3 つ目のブロックを紙の上に置き終えるまでの時間を計測した．

実験では，12 人の被験者（男性 10 人，女性 2 人）に参加してもらった．被験者の年齢は 20 歳から 30 歳で，平均 22.7 歳であった．各被験者に，3 つ全ての条件で上記の物体操作タスクを実行してもらった．各条件につき，3 パターンの色（紫，黄，緑），3 パターンの形状（小さいキューブ，円柱，短い直方体）を対象としたの合計 6 回の試行を実施してもらった．慣れや疲れによる影響を軽減するため，2 人の被験者ごとに条件の実施順を変えてタスクを実施してもらった．VR 環境での試行を始める前には，被験者に VR インタフェースの操作方法を説明し，物体の操作になれるまで練習してもらった．物体操作の速さについては指定せず，任意の速度でタスクを実施してもらった．

3. 実験結果

図 6 に各条件の達成時間の平均と 95%の信頼区間を示す．本実験では，全ての試行において，把持に失敗するなどのもたつく動作は見られず，達成時間には不適切な挙動による遅延は含まれていない．各条件の平均達成時間は，視野の制約が無い条件で 12.6 秒，視野の制限がある条件で 14.0 秒，VR 環境での条件で 15.3 秒であった．図 6 の上部に，Bonferroni の補正を用いて多重比較を行った結果を示す．視野の制約が無い場合と VR における達成時間には，有意な差が見られた（$p < 0.001$）．一方で，視野に制限がある場合と VR における達成時間には，有意な差は見られなかった（$p = 0.162$）．従って，視野の制限によって達成時間に有意な差が無くなることから，視野の制約が物体操作時の人間の挙動に影響を与える大きな要因と一つであることがわかる．

4. まとめ

没入型 VR 環境における視野の制約による影響を検証するため，実世界と VR における物体操作時の人間の挙動の差異を評価した．これまで，視野の制約に依存する距離の過小知覚などについては報告されているものの，環境内を見渡して物体を見つけたり，物体のある場所に手を伸ばしたりといった場面において，視野の制約や実環境と VR 環境の違いが人間の挙動に与える影響は明らかになっていなかった．本研究では，物体操作タスクを設計し，視野の制約および実環境と VR 環境の違いにより，タスクの達成に要する時間どのように変化するか検証した．実験の結果から，実環境で視野の制約が無い条件に比べ，VR 環境での達成時間が有意に長くなった．一方で，実環境で視野を制限した条件と VR 環境での達成時間に有意な差はなかったことから，視野の制約が物体の認識や操作を行う際の人間の挙動に影響を与える大きな要因であることが明らかとなった．

没入型 VR を活用して観測した人間の挙動データに基づいて実環境での人間支援システムを構築するという観点から，環境中を見渡しながら空間を把握するような多くの視線の変化を伴うシナリオでは，実環境と VR 環境における人間の挙動に大きな差異が生じることが懸念される．例えば，稲邑らはロボットの自然言語生成能力を評価する競技タスクを提案している [Inamura 17]．この競技では，VR 環境に被験者がログインし，ロボットが生成した説明文に従って複数の部屋から構成された環境中を歩き回り，対象物を見付けるまでの時間からロボットの説明能力を評価している．従来のシナリオでは，周りを見渡しながら部屋を特定するためにランドマークとなる家具を探すなど，広範囲を対象とした空間把握が必要とされるため，視野の制約による人間の挙動への影響が懸念される．このタスクでは，自然言語の生成能力の評価を対象としているため，深刻な影響はないと考えられるが，本研究の結果を踏まえると，実環境での人間支援システムを構築するための行動データ収集という観点では，被験者が視野を変化させる頻度が減るように，探索範囲を限定することや調理タスクなど手元での作業を対象としたシナリオに工夫するなどの注意を払う必要があると言える．

謝辞

本研究の成果の一部は，国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務，および JST CREST（グラント番号 JPMJCR15E3）の支援を受けたものです．

参考文献

[Levine 16] Levine, S., Pastor, P., Krizhevsky, A., and Quillen, D.: Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection, in Int. Symp. on Experimental Robotics, pp. 173-184, 2016.

[Liu 16] Liu, P., Glas, D. F., and Kanda, T.: Data-Driven HRI: Learning Social Behaviors by Example from Human-Human Interaction, IEEE Trans. on Robotics, Vol. 32, pp. 988–1008, 2016.

[Sumi 07] Sumi, Y., Ito, S., Matsuguchi, T., Fels, S., and Iwasawa, S.: Collaborative capturing, interpreting, and sharing of experiences, Journal of Personal and Ubiquitous Computing, vol. 11, No. 4, pp.265–271, 2007.

[Brščić 15] Brščić, D., Kidokoro, H., Suehiro, Y. and Kanda, T.: Escaping from children's abuse of social robots, in Proc. of ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 59–66, 2015.

[Inamura 10] Inamura, T., Shibata, T., Sena, H., Hashimoto, T., Kawai, N., Miyashita, T., Sakurai, Y., Shimizu, M., Otake, M., Hosoda, K., Umeda, S., Inui, K., and Yoshikawa, Y.: Simulator platform that enables social interaction simulation - SIGVerse: Sociointelligenesis simulator, in the IEEE/SICE Int. Symp. on System Integration, pp. 212-217 (2010)

[Mizuchi 17] Mizuchi, Y., and Inamura, T.: Cloud-based Multimodal Human-Robot Interaction Simulator Utilizing ROS and Unity Frameworks. in Proc. of the IEEE/SICE Int. Symp. on System Integration, ThA2.3, 2017.

[Bates 17] Bates, T., Amirez-Amaro, K., Inamura, T., and Cheng, G.: On-Line Simultaneous Learning and Recognition of Everyday Activities from Virtual Reality Performances. in Proc. of IEEE/RSJ Int ' l Conf. on Intelligent Robots and Systems, pp. 3510–3515.

[Inamura 17] Inamura, T., and Mizuchi, Y.: Competition design to evaluate cognitive functions in human-robot interaction based on immersive VR, RoboCup International Symposium, 2017.

[Duguleana 11] Duguleana, M., Barbuceanu, F. G., and Mogan, G.: Evaluating Human-Robot Interaction during a Manipulation Experiment Conducted in Immersive Virtual Reality, in Int ' l Conf. on Virtual and Mixed Reality, pp. 164–173, 2011.

[El-Shawa 17] El-Shawa, S., Kraemer, N., Sheikholeslami, S., Mead, R., and Croft, E. A.: "Is this the real life? Is this just fantasy?": Human proxemic preferences for recognizing robot gestures in physical reality and virtual reality, in Proc. of IEEE/RSJ Int ' l Conf. on Intelligent Robots and Systems, pp. 341–348, 2017.

[Willemsen 09] Willemsen, P., Colton, M. B., Creem-Regehr, S. H., and Thompson, W. B.: The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments, ACM Trans. on Applied Perception, Vol. 6, Iss. 2, 2009.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

系列変換モデルを用いたプレゼンテーション動作の生成

Generation of Presentation Motion Using Sequence to Sequence Model

嶋津 章仁 *1 Akihito Shimazu
日永田 智絵 *1 Chie Hieida
長井 隆行 *1 Takayuki Nagai
中村 友昭 *1 Tomoaki Nakamura
武田 祐樹 *2 Yuki Takeda
原 豪紀 *2 Takenori Hara
中川 修 *2 Osamu Nakagawa
前田 強 *2 Tsuyoshi Maeda

*1電気通信大学 The University of Electro-Communication
*2大日本印刷 Dai Nippon Printing Co., Ltd.

In the presentation, gestures play a very important role as a means of improving the information transmission effect. Moreover, it was shown that the movement expressing the enthusiasm and intention of the presenter in an unconscious manner affects the success of the presentation and the impression on the audience. For these reasons, it is required for presentation robots to perform such movements; however, manual design of these movements is not an easy task. In this research, we propose a method to learn relationship between speech prosodic information and motion using a sequence to sequence model, and directly generate appropriate motions from prosodic information. This paper also proposes a method for generating motions representing meanings of specific words.

1. はじめに

近年，ロボットの社会進出が進んでおり，日常生活においてもロボットと人間が接する機会が増加している．そのようなロボットの例として，プレゼンテーションロボットが挙げられる．プレゼンテーションにおいて人間は様々な動きを行うが，人間が発話と共に行う動きがどのように生成されているかは，ロボットで再現できるほど十分に明らかにされていない．そのため，ロボットの動きは人手で設計するのが一般的である．人手による動きの設計は，特に自由度が高い場合には容易ではないため，プレゼンテーションの際の複雑かつ自然に見える動きを，どのように自動的に生成するかが課題となる．そこで，機械学習によって音声情報と動きを結びつけて学習させ，発話音声から自動的に動きを生成することを考える．

まず本研究では動きが，伝達内容に関する動き，文脈に関する動き，身体（個性）に依存した動き，の組み合わせによって生成されていると仮定する．つまり，これら 3 つの動きをそれぞれ実装し合成することによって，人間がコミュニケーションの際に行っている動きを再現できるのではないかと考える．本稿では，文脈をプレゼンテーションという状況に固定し，身体と伝達内容に関する動きの実現を目指す．

プレゼンテーションの目的は聴衆に情報をわかりやすく伝えることであり，プレゼンテーションロボットにも情報伝達をわかりやすく行うことが要求される．プレゼンテーションにおいて人間がジェスチャを行う理由の一つは，情報伝達を助けるためであると言われる [喜多 00]．Berger らは，ジェスチャを伴った音声と伴わない音声では，ジェスチャを伴った音声の方が聞き手の理解を約 2 倍正確にしたと報告している [Berger 71]．つまり，プレゼンテーションにおいて動きは情報伝達という意味で重要であり，前述の伝達内容に関する動きに相当する．

一方で Pentland は，ビジネスプレゼンテーションの成功が単なる言語的な情報の伝達ではなく，非言語的な情報によって意思や熱意が無意識に伝わることに関連していることを明らかにした [Pentland 08]．このことは，情報の伝達自体を目的としていない無意識に訴えかける動きの情報も重要であることを示唆している．これは，前述の身体に依存した無意識の動きに関係すると考える．例えば，発話の際には呼吸によって体が自然と動くため，そのリズムと同期して腕が動くといった物理的な現象があるはずであり，そうした発話との関係性に制約されつつも，意思や熱意が無意識に伝わるような動きがある．

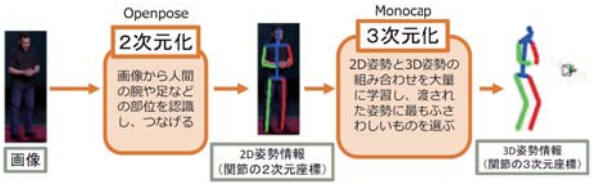


図 1: 3 次元姿勢抽出の流れ．

そこで本稿ではまず，意思や熱意を無意識のうちに表現している動き，つまりプレゼンテーションにおいて身体に依存した動きの実現に焦点を当てる．こうした動きはルール化することが難しいため，プレゼンテーション時の動きのデータを収集し，系列変換モデルによって韻律情報と動きの情報を結びつけることで自動生成する．そうすることで，プレゼンテーション時の平均的な動きが抽出され音声合成した発話の韻律に合わせて自動生成できる．この枠組みでは，学習データに多数の話者が含まれることによって個性が平均化され，より一般的な動きが生成される．逆に学習データを特定の個人に絞れば，個性を反映させることも可能である．さらに本稿では，情報伝達を助けるジェスチャ，つまりプレゼンテーションにおいて伝達内容を表現する動きについても検討する．

これまでに，アニメーションのための人の動き生成手法が多く提案されている．例えば，文献 [Alemi 15] では，Restricted Boltzman Machine を用いた情動的な動作の生成を行っている．しかし多くの研究では，発話に伴う動作の生成を考えていない．竹内らは，発話に伴うジェスチャを Bi-Directional LSTM で生成することを提案している [竹内 17]．しかし，この研究で

連絡先: 嶋津 章仁，電気通信大学大学院情報理工学研究科，東京都調布市調布ケ丘 1-5-1，ashimazu@apple.ee.uec.ac.jp



図 2: 抽出された 3 次元姿勢.

はプレゼンテーションという文脈は考えられていない．また，学習データをモーションキャプチャで収集しており，十分な数のデータによる学習が実現されていないという問題がある．

2. 提案手法の概要

提案手法は，発話音声のエネルギーとピッチから成る韻律情報を入力とし，上半身合計 9 関節の 3 次元ベクトルの時系列を出力する．これは，再帰型のニューラルネットワークを応用した，系列変換（seq2seq）モデルを学習することで実現する．ここでの問題は，大量の学習データが必要になることである．学習に必要なデータは，音声データと発表者の 3 次元姿勢データであるが，これらのデータを大量に集めるために，web 上にあるプレゼンテーションの映像から，発表者の 3 次元姿勢を抽出する手法を提案する．本稿ではさらに，情報伝達を目的としたジェスチャを生成し，系列変換モデルによって生成した動作に合成する．

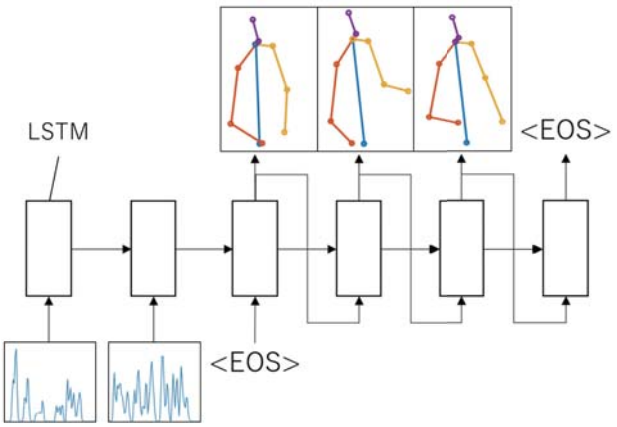


図 3: 学習モデルの概略.

3. 提案手法

3.1 学習用データセットの作成

ここでは，系列変換モデル学習のためのデータセット作成について述べる．

3.1.1 画像からの 3 次元姿勢推定

2 次元画像中の人の 3 次元姿勢推定は，OpenPose [Cao 17] による 2 次元姿勢抽出手法と，MonoCap [Zhou 16] による 3 次元姿勢推定手法を組み合わせることで実現する．これは，この二つの手法の組み合わせが精度と計算速度の点で優れているためである．図 1 に，推定の流れを示す．まず OpenPose を用いて，映像中の対象の 2 次元姿勢を取得する．その結果を MonoCap に渡し，学習した 2 次元姿勢と 3 次元姿勢の組み合わせから，似た姿勢を考慮することで 3 次元化を行う．また，プレゼンテーション動画では，スライドや観客のみが映っているフレームも多く存在する．本稿では，話者が多くの場合に中心で大きく映っているという事前知識を併用して，話者のトラッキングを行った．提案した手法で 3 次元姿勢推定を行った結果の例を，図 2 に示す．発表者の 3 次元姿勢推定の精度を評価するために，一つのプレゼン動画（約 15 分）に対して，フレーム毎に抽出の成否を人手で判定した．その結果，抽出率は 94.8% であり，プレゼンテーションの動きを学習するための精度としては十分であることが分かった．

3.1.2 韻律の抽出

音声のピッチは，森勢らによって提案された音声分析合成システム WOLRD [Morise 16] を用いて行う．また音声のエネルギーは，発表者の音声波形の各サンプルを二乗し，ピッチの計算周期で加算することで算出する．

3.2 身体に依存した動きの学習

身体に依存した動きの学習は，図 3 に示すように，音声の韻律を入力として動きを生成する系列変換モデルによって実現する．系列変換モデルは，Sutskever らによって提案された再帰型ニューラルネットワークに基づく学習モデルである [Sutskever 14]．エンコーダ—デコーダモデルとも呼ばれ，入力の時系列をベクトルにエンコードし，そのベクトルから出力の時系列をデコードする．提案手法では，入力は音声のピッチと音声エネルギーの時系列であり，これを発話文章単位でエンコードする．デコーダ側では，発表者の上半身 9 関節の 3 次

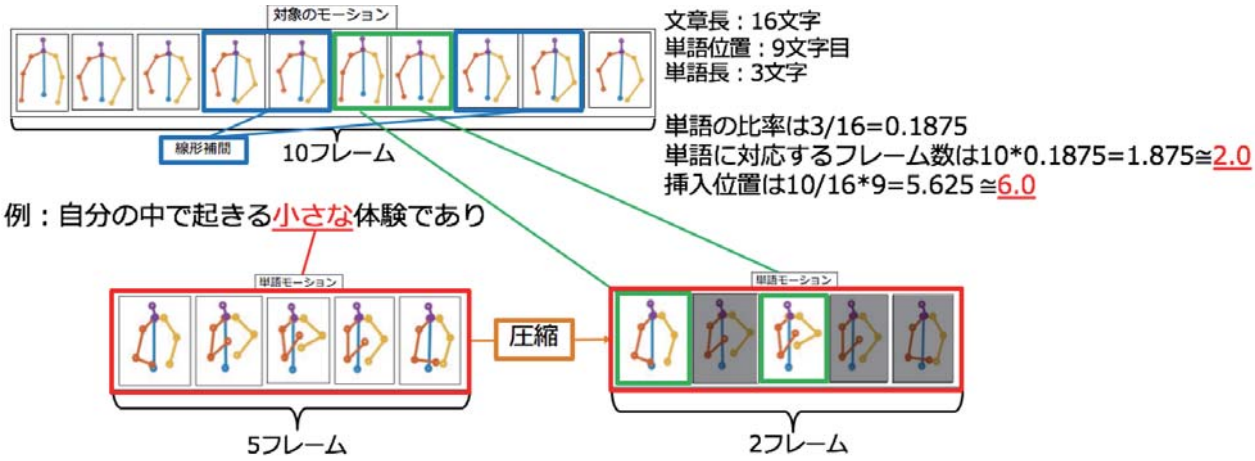


図 4: 単語的意味を表現する動きの合成手法.

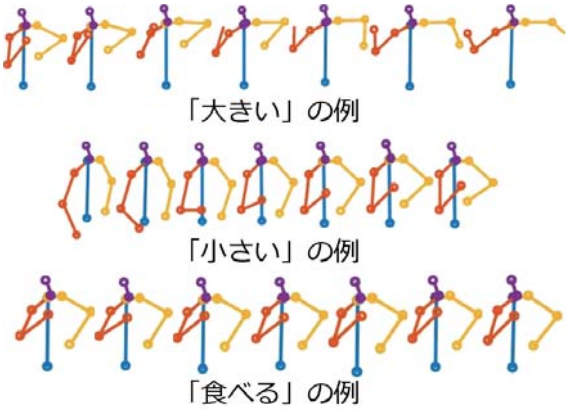


図 5: 作成した単語の意味に対応した動きの例.

元姿勢ベクトルをエンコーダ側からの情報を基に復元する．文章単位の系列変換モデルであるため，文章内での動きの滑らかさは韻律情報の滑らかさによってある程度保証されるが，文章間で不連続になる可能性がある．提案手法では，文章間の発話のないところ補間することで，文章間の姿勢を滑らかにつなぐこととする．

3.3 情報伝達を行う動きの合成

情報伝達を目的とした動きとしては様々なものが考えられるが，ここでは特定の単語の意味を表現する動きを考える．系列変換モデルによって生成された動きに対して，単語的意味を表現する動きを合成する手法の概略を図 4 に示す．文章中の対象単語の発話位置から動きを挿入する時間的位置を算出し，かつ挿入部分にモーションが収まるよう時間的に伸縮することで動きの合成を行う．なお，対象とする単語は手動で選択することとする．また，挿入する単語動作自体も前述の 3 次元姿勢抽出手法を用いて著者の動作から自動で生成したものを用いる．つまり選択した各単語に対して著者がそのイメージで動作を行い，その映像から 3 次元姿勢の抽出を行い単語動作ライブラリとして登録する．

作成した動きの一例を，図 5 に示す．図より，単語の動作を表しているような動作が取得されていることが分かる．

4. 実験

提案手法の有効性を評価するための実験について述べる．

4.1 実験設定

本稿ではデータとして，TED [TED 17] の 119 の動画を用いた．この全動画に対して，前述の 3 次元姿勢抽出と韻律の抽出を行い，それぞれ文章ごとに分割することで音声（韻律）情報と 3 次元姿勢情報の 10485 対の時系列データを得た．これは時間にして，7 時間 23 分 27 秒分のデータである．これらのデータを，学習データと評価データに分割して実験を行った．

4.2 生成モーションの評価手法

生成されたモーションをどのように客観評価するかは，重要な問題である．主観評価も必要であるが，ここではその前段として客観的な指標によって生成モーションを評価することを考える．プレゼンテーション時のモーションの自然さという視点で考えると，動きと音声の同期性が重要である．これは，音声とモーションが同一の話者で同時に生成されていることを考えれば当然であり，その同期性を観測者が「動きの自然さ」として捉えると考えることは，不自然ではない．そこで本稿では，「各関節の速さと音声エネルギーの変化量」の相関係数を指標とする．まずこの指標が使えることを示すために，前述のデータに対して相関係数を計算した．また音声とモーションをランダムに入れ替えたデータに対しても同様に計算し，それらの差異を検定したところ，統計的に優位に入れ替えのないデータの相関係数が大きかった．そこで，入れ替えのないデータの相関係数の分布と入れ替えたデータの相関係数の分布をそれぞれ正規分布で表し，その二つの分布の対数尤度比という形で生成したモーションの自然性を評価することとした．

4.3 結果

学習データを用いて系列変換モデルを学習し，そのモデルを用いて，残りの 10 個の評価用データの音声からモーションを生成した．結果の例を図 6 に示す．生成したモーションに対して，対数尤度比スコアを算出した結果を図 7 に示す．スコアが 0 より大きい場合に，音声と動きの同期性という点で自然なモーションであると言える．生成したモーションによるスコアの平均が 0.85，実際に人が動作しているデータを用いた場合のスコアの平均は 1.10 であり，提案手法による生成モー

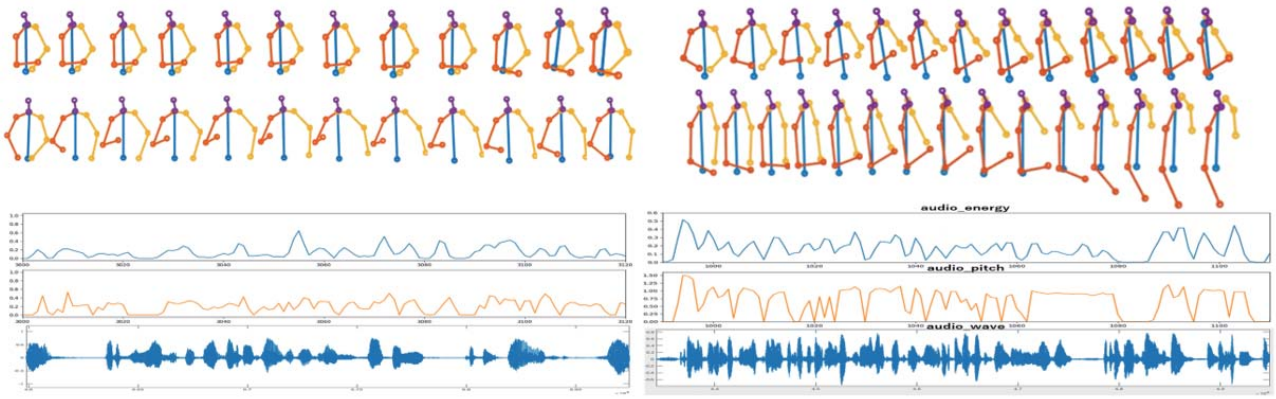


図 6: 生成モーションの一例; 上段) オリジナルの動き，中段) 系列変換モデルによって生成された動き，下段) 音声の情報.

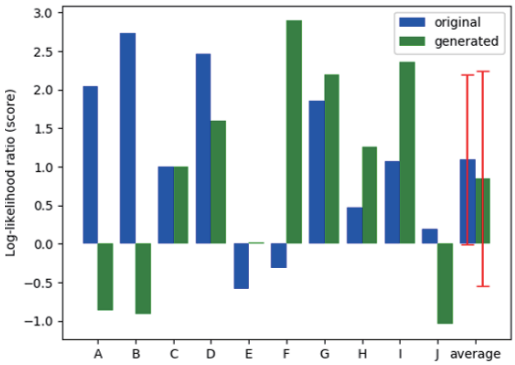


図 7: 対数尤度比スコアによる評価.

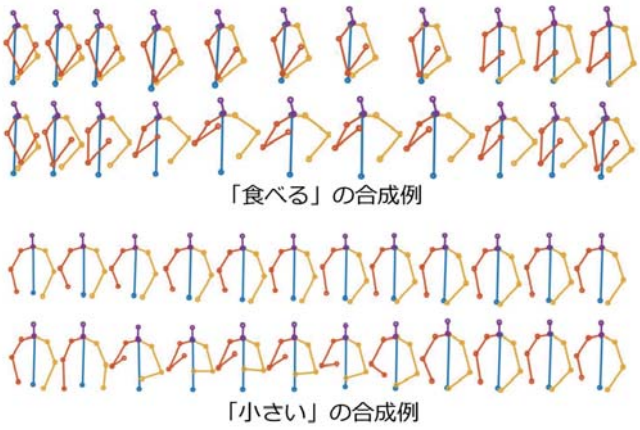


図 8: 単語の意味を表す動きの合成例; 上段) 合成前の動き，下段) 合成後の動き.

ションは，同期性という視点では人のプレゼンテーションと同等に自然であることが分かる.

さらに，系列変換モデルによって生成した動きに対して，単語の意味に応じた動きを合成した結果の例を図 8 に示す.「食べる」といった動詞や,「小さい」といった形容詞が意味的に強調された動作が生成されていることが分かる．合成した動きに対する評価は，今後の課題である.

5. まとめ

本稿では，プレゼンテーションロボット向けモーション生成手法を提案した．提案手法は系列変換モデルに基づいており，学習に必要な大量のデータをプレゼンテーション動画から自動生成した．実験によって，生成されたモーションが，音声と動きの同期性という基準で自然なモーションであることを示した．また単語の意味に相当する動きを生成し，系列変換モデルによって生成された動きと合成する手法を提案した．これにより実際に意味の強調された動きが生成できることを確認したが，今後，生成された動きを主観評価する必要がある．生成された動きのロボットへの実装も今後の課題である.

参考文献

[喜多 00] 喜多壮太朗, “特集, ジェスチャーの認知科学：ひとはなぜジェスチャーをするのか,” 認知科学 7(1), 9-21, 2000, 2000

[Berger 71] K.W.Berger *et al.*, “Extra-facial gestures in relation to speechreading,” Journal of Communication Disorders pp.302-308 1971, 1971

[Pentland 08] A.Pentland, Honest Signals, The MIT press, 2008

[Alemi 15] O.Alemi, W.Li, P.Pasquier, “Affect-expressive movement generation with factored conditional Restricted Boltzmann Machines,” ACII 2015,2015

[竹内 17] 竹内健太, 長谷川大, 白川真一, 金子直史, 佐久田博司, 鷲見和彦, “Bi-Directional LSTM を用いた発話に伴うジェスチャの自動生成手法の検討, ” HAI2017,2017

[Cao 17] Z.Cao *et al.*, “Realtime Multi-Person 2D Pose Estimation using Affinity Fields,” CVPR2017, 2017

[Zhou 16] X.Zhou *et al.*, “MonoCap: Monocular Human Motion Capture using a CNN Coupled with a Geometric Prior,” CVPR2016, 2016

[Morise 16] M.Morise *et al.*, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” IEICE transactions on information and systems, 2016

[Sutskever 14] I.Sutskever *et al.*, “Sequence to Sequence Learning with Neural Networks,” NIPS2014, 2014

[TED 17] TED Conferences LLC, “TED (Technology Entertainment Design),” https://www.ted.com/, Retrieved 2017

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

文脈に依存して変動する動作の意味学習における効率的な能動学習法の検証

Validation of Effective Active Learning Method in Semantic Learning of Context-dependent Motions

坂戸 達陽 *1　稲邑 哲也 *1*2
Tatsuya Sakato　Tetsunari Inamura

*1国立情報学研究所
National Institute of Informatics
*2総合研究大学院大学
SOKENDAI(The Graduate University for Advanced Studies)

Since the intelligent systems require a huge dataset of motion and label to recognize the meaning (label) of the body motion, we consider active learning in which the systems ask the label to users. We aim to realize an effective learning and question management method by considering the context in motion performance. In this paper, we use VR avatars that perform motions in different contexts, and define the context by tools and places used in the motion performance. Active learning was performed by combining each method concerning three points of context selection method, selection of Open/Close question, and label estimation method. We showed that the combination of margin sampling as context selection, naive Bayes as label estimation method, and performing open question at the beginning of the question and close question at latter term, is most efficient.

1. はじめに

同じような身体動作でも場面によって意味が異なる場合がある．例えば，キッチンでフライパンを用いて料理を行う身体動作（図 1 (a)）と，海岸で釣り竿を用いて釣りを行う身体動作（図 1 (b)）は，その身体動作だけを見ればよく似ているが，全く異なる意味を持っている．システムが人の身体動作のみを観測してこのような文脈依存の身体動作の意味を区別することは難しい．しかし一方で人は日常生活場面においてこのような身体動作を区別することができる．それは，人がこのような身体動作の意味を推定する場合は，その身体動作だけでなく，それが行われている場所や使っている道具，その身体動作が行われた前後の身体動作などの，文脈情報を考慮しているからである．日常生活場面では，このように文脈情報を考慮しなければ意味の推定が困難な場面は多い．よって，日常生活場面で活動するロボットにも，文脈情報を考慮して身体動作の意味を推定することが求められる．

ロボットが文脈依存の身体動作の意味を獲得する方法として，様々な文脈上での身体動作を含む人とのインタラクションを通して学習するということが挙げられる．例えば，様々な文脈上で人が身体動作を行う場面を観測し，人にその意味を尋ねてラベル付けを行うことや，観測したい身体動作を人に提示し，実際に身体動作を行ってもらうことなどが挙げられる．しかし，ロボットが実際に遭遇した場面でのインタラクションのみでは，ロボットが日常生活場面で活動するために必要な場面を学習時に網羅し，十分な量の学習を行うことは難しい．ロボットは観測したい場面を積極的に提示して人とインタラクションを行う必要がある．しかし，想定されるすべての場面を提示して学習を行うことは効率的ではない．また，学習空間が巨大である場合，すべての場面を網羅して提示することは現実的ではない．よって，ロボットは学習のために最も効果的な場面を選択して提示するべきである．

効率的な学習対象を選択して学習する手法として能動学習が挙げられる [Settles 09]．能動学習では，学習エージェントは学習に効果的な対象を能動的に選択して，教示者（oracle）に



(a) 料理の身体動作



(b) 釣りの身体動作

図 1: まぎらわしい身体動作の例

正解ラベルを尋ねることで，やみくもに尋ねるよりも効率的な学習を試みている．従来の能動学習では，oracle に対して，対象が何であるか，どのクラスに属するかなど，毎回同じ質問を行っていた．しかし，oracle が人であった場合などでは，毎回無数の選択肢を提示することは人にとって大きな負担となってしまい効率的ではない．よって，システムは質問方策を，例えば,「これは何ですか？」という open question から,「これは A ですか？」などの closed question に，学習の進捗に応じて変えていくべきである．Cakmak ら [Cakmak 12] は質問方策の違いによる人が受ける印象の違いを検証した．しかし，彼らの研究では異なる質問方策によってどのように学習するかは扱っ

連絡先: 坂戸 達陽，国立情報学研究所，〒 101-8430 東京都千代田区一ツ橋 2-1-2，sakato@nii.ac.jp

表 1: 記号一覧

記号	意味
S, L, T, M, W	場面，場所，道具，身体動作，身体動作ラベルの有限集合
$\boldsymbol{s} = [s_l, s_t, s_m] \in \boldsymbol{S}$	場面
$s_l \in L$	$\boldsymbol{s}$ の場所
$s_t \in T$	$\boldsymbol{s}$ の道具
$s_m \in M$	$\boldsymbol{s}$ の身体動作
$w \in W$	身体動作ラベル
$N_{\boldsymbol{s},w}$	Oracle が $\boldsymbol{s}$ を提示されたときに，身体動作ラベル w を答えた回数
$N_{s_l,s_t,s_m,w}$	$\boldsymbol{s} = [s_l, s_t, s_m]$ における $N_{\boldsymbol{s},w}$
$N_{\boldsymbol{s}}$	$N_{\boldsymbol{s},w}$ の w に関する和
$P(w\|\boldsymbol{s})$	$\boldsymbol{s}$ に w が対応する確率
$w_{\boldsymbol{s}}$	$\boldsymbol{s}$ に対して推定される身体動作ラベル
$\hat{w}_{\boldsymbol{s},i}$	$\boldsymbol{s}$ において対応する確率が i 番目に高いと推定される身体動作ラベル
$\hat{W}_{\boldsymbol{s},1}$	$\{w\|P(w\|\boldsymbol{s}) = \max_w P(w\|\boldsymbol{s})\}$
$H(W\|\boldsymbol{s})$	$\boldsymbol{s}$ における W の条件付きエントロピー
$w^*_{\boldsymbol{s}}$	$\boldsymbol{s}$ に対する正解ラベル
θ	質問方策選択のための閾値

ていない．これに対して本研究では，効率的な学習と，学習のためのインタラクションにおけるコストの両方を考慮する．

以上をふまえて本稿では，使用する道具，動作が行われた場所を文脈と位置づけ，VR 環境を用いて同じ身体動作パターンに対して異なる文脈で動作するアバターを用いながらラベルの能動学習を行う方法を検討する．質問の際に用いる文脈の選択方策，Open/Close question の選択基準，ラベル推定手法の 3 点について，それぞれの方法を組み合わせた能動学習を行い，学習結果の正答率の変化を評価する．

2. 問題設定

提案手法の概略図を図 2 に示す．本稿では，学習エージェントが場面に対応する身体動作ラベルの確率分布を学習するタスクを行う．本稿では場面を場所，道具，身体動作の 3 つの属性の組で表現する．各属性の値はそれぞれ図 3，図 4，図 5，表 2 の中のいずれかを示す．

学習エージェントは場面選択方策に基づいて提示する場面 $\boldsymbol{s}$ を選択する．さらに，選択した場面における身体動作ラベルの確率分布 $P(W|\boldsymbol{s})$ に基づいて，質問方策を選択する．そして，場面 $\boldsymbol{s}$ を oracle に提示して質問方策に基づいた質問を行う．Oracle は提示された場面と質問に対して回答する．学習エージェントは得られた場面-身体動作ラベルの組を用いて場面に対応する身体動作ラベルの確率分布を更新する．学習エージェントは以上の手順を一定の回数行うと学習を終了する．

本稿で用いる記号は表 1 に示す．

2.1 Oracle の実装

本実験では，oracle は提示された場面に対して対応する身体動作ラベルを自動的に回答するソフトウェアによって実装した．Yes/no question の場合，oracle は正解である身体動作ラベルと尋ねられた身体動作ラベルが一致する場合は yes，不一致の場合は no と答える．Oracle の回答は，仮想現実（VR）環境で提示された場面を著者が観て定めた．身体動作の提示はあらかじめモーションキャプチャで収録したものをアバターに再現させることによって行った．

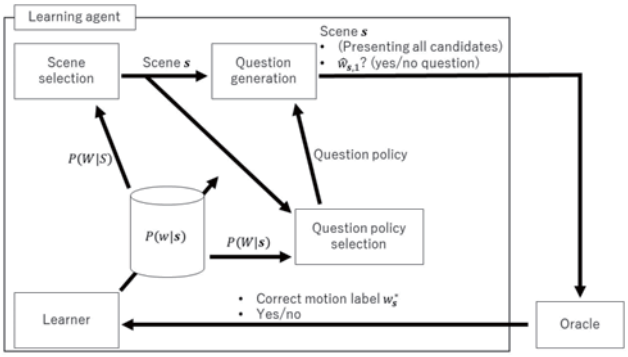


図 2: 提案手法の概略図

表 2: 実験で扱う身体動作ラベル

身体動作ラベル
挨拶
料理
掃除
釣り
打つ
切る
その他

3. 実験条件

3.1 場面提示方策

本稿では次の場面選択方策の比較を行った．Random selection 以外の場面選択方策は，[Settles 09] を参考にした．

- Random Selection: すべての場面の中からランダムに選択する．

$$\boldsymbol{s} \sim \frac{1}{|S|} \tag{1}$$

- Entropy-based: すべての場面の中からその場面に対応する身体動作ラベルのエントロピーが最大の場面を選択する．

$$\boldsymbol{s} \leftarrow \underset{\boldsymbol{s}}{\operatorname{argmax}} H(W|\boldsymbol{s}) \tag{2}$$

- Margin Sampling: すべての場面の中からその場面に対応する確率の最も高い身体動作ラベルと二番目に高い身体動作ラベルの確率の差が最小の場面を選択する．

$$\boldsymbol{s} \leftarrow \underset{\boldsymbol{s}}{\operatorname{argmin}} \{P(\hat{w}_{\boldsymbol{s},1}|\boldsymbol{s}) - P(\hat{w}_{\boldsymbol{s},2}|\boldsymbol{s})\} \tag{3}$$

- Least Confident: すべての場面の中からその場面に対応する確率の一番高い身体動作ラベルの確率が最小の場面を選択する．

$$\boldsymbol{s} \leftarrow \underset{\boldsymbol{s}}{\operatorname{argmin}} \max_w P(w|\boldsymbol{s}) \tag{4}$$

3.2 身体動作ラベル推定

本稿では次の身体動作ラベル推定手法を比較した．



(a) 海岸　(b) キッチン　(c) 森　(d) ゲームセンター

図 3: 実験で扱う場所



(a) （道具なし）　(b) フライパン　(c) バット　(d) 釣り竿　(e) 布　(f) マレット　(g) 斧

図 4: 実験で扱う道具（道具なしを含む）



図 5: 実験で扱う身体動作．(a) 右手を垂直平面上で左右に振る．(b) 右手を水平平面上で左右に振る．(c) 両手を上下に振る．(d) 両手を袈裟に振る．

- Freq.: $\boldsymbol{s}$ に対応する身体動作ラベルの確率分布の推定に $\boldsymbol{s}$ についての oracle の回答のみを用いる．式 (5) では 0 による除算を防ぐため，可算スムージングを行っている．

$$w_{\boldsymbol{s}} = \underset{w}{\operatorname{argmax}} \frac{N_{\boldsymbol{s},w} + 1}{N_{\boldsymbol{s}} + |W|} \tag{5}$$

- Bayes: $\boldsymbol{s}$ に対応する身体動作ラベルの確率分布の推定にナイーブベイズを用いる．

$$w_{\boldsymbol{s}} = \underset{w}{\operatorname{argmax}} P(w) \prod_{x} P(s_x|w) \tag{6}$$

式を満たす $w_{\boldsymbol{s}}$ が複数存在するときは，式を満たす w の中からランダムに選択する．

3.3 質問方策選択

本稿では次の質問方策選択手法を比較した．

- All: 毎回すべての候補となる身体動作ラベルを提示して正解ラベルを尋ねる．Oracle が w を答えると，$N_{\boldsymbol{s},w}$ を式 (7) のように更新する．

$$N_{\boldsymbol{s},w} \leftarrow N_{\boldsymbol{s},w} + 1 \tag{7}$$

- Yes/no: 選択された場面 $\boldsymbol{s}$ において式 (8) を満たす場合に $\hat{w}_{\boldsymbol{s},1}$ に関する yes/no question を行う．満たさない場合はすべての候補となる身体動作ラベルを提示して正解ラベルを尋ねる．Yes/no question を行う場合は，$N_{\boldsymbol{s},\hat{w}_{\boldsymbol{s},1}}$ を式 (9) のように更新する．すべての候補となる身体動作ラベルを提示して正解ラベルを尋ねる場合は all 条件と同じように更新する．

$$P(\hat{w}_1|\boldsymbol{s}) - P(\hat{w}_2|\boldsymbol{s}) > \theta \tag{8}$$

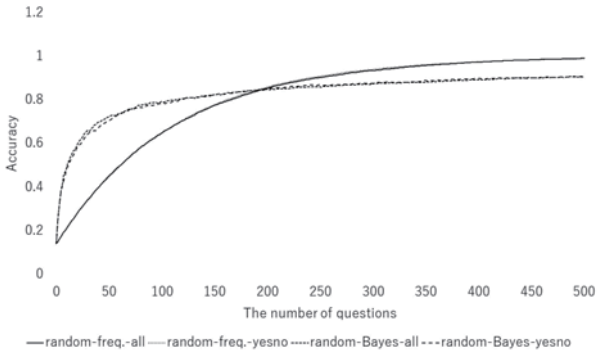
$$N_{\boldsymbol{s},w} \leftarrow \begin{cases} N_{\boldsymbol{s},w} + 1 & \text{if the oracle answered yes,} \\ N_{\boldsymbol{s},w} - 1 & \text{if the oracle answered no.} \end{cases} \tag{9}$$

本稿では，式 (8) における θ は 0.2 とした．

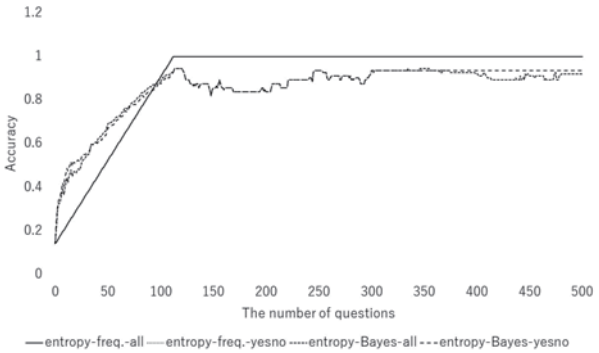
4. 実験結果

各条件ごとの質問回数に対する正答率の 60 試行の平均の推移を図 6 に示す，正答率 acc. は，式 (10) のように求めた．

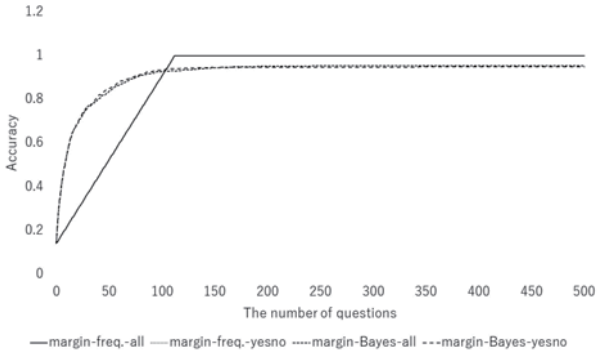
$$\text{acc.} = \frac{\sum P(w_{\boldsymbol{s}} = w^*_{\boldsymbol{s}})}{|S|} \tag{10}$$



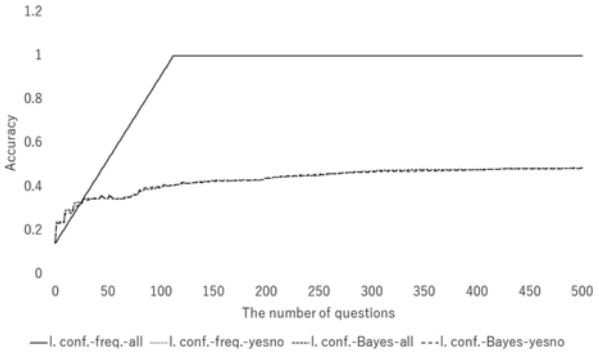
(a) Random Selection



(b) Entropy-based



(c) Margin Sampling



(d) Least Confident

図 6: 各手法ごとの質問方策の違いによる正答率の推移

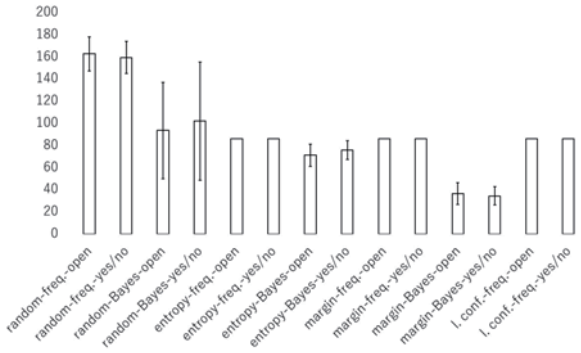


図 7: acc. > 0.8 を達成するまでの質問回数

式 (10) において，$P(w_{\boldsymbol{s}} = w^*_{\boldsymbol{s}})$ の値は式 (11) のように定めた．

$$P(w_{\boldsymbol{s}} = w^*_{\boldsymbol{s}}) = \begin{cases} \frac{1}{|\hat{W}_{\boldsymbol{s},1}|} & (w^*_{\boldsymbol{s}} \in \hat{W}_{\boldsymbol{s},1}), \\ 0 & (\text{otherwise}). \end{cases} \tag{11}$$

図 7 に acc. > 0.8 を達成するまでの質問回数を示す．各条件のうち，目標とする正解率の基準を最も少ない回数で達成したのは，場面選択方策 margin sampling，ラベル推定手法 naive Bayes，質問方策選択手法 yes/no の条件の場合であった．また，acc. > 0.8 を達成するまでの質問回数は，least confident 以外の場面選択方策を採用した場合，確率分布推定が freq. 条件の場合よりも Bayes 条件の方が少なかった．しかし，最終的な正答率は，freq. 条件の方が高くなった．場面提示方策に least confident を採用した場合は，acc. > 0.8 を達成することはできなかった．

5. おわりに

本稿では，使用する道具，動作が行われた場所を文脈と位置づけ，VR 環境を用いて同じ身体動作パターンに対して異なる文脈で動作するアバターを用いながらラベルの能動学習を行う方法を検討した．質問の際に用いる文脈の選択方策，Open/Close question の選択基準，ラベル推定手法の 3 点について，それぞれの方法を組み合わせた能動学習を行い，学習結果の正答率の変化を評価した．評価実験の結果，文脈の選択方法としては margin sampling を用い，質問の初期は open，後期は close question を行い，naive Bayes を用いたラベル推定方法を行う組み合わせが最も効率的であることを示した．

謝辞

本研究は JST CREST（グラント番号 JPMJCR15E3）の支援を受けたものである．

参考文献

[Settles 09] Settles, B.: “Active learning literature survey”, Computer Science Technical Report, no. 1648, (2009).

[Cakmak 12] Cakmak, M., Thomaz, A.L.: “Designing Robot Learners that Ask Good Questions”, In Proc. 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 17–24, (2012).

[2G4] [General Session] 11. Robot / Real World

座長:大澤 博隆（筑波大学）

Wed. Jun 6, 2018 5:20 PM - 6:40 PM Room G (5F Ruby Hall Hiten)

[2G4-01] Emotion Estimation from Robot Movements Using Weighted Random Forest and Pairwise Classification

○Hideki Kato[1], Shohei Kato[1,2] （1. Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, 2. Frontier Research Institute for Information Science, Nagoya Institute of Technology）

5:20 PM - 5:40 PM

[2G4-02] Emotional Differentiation based on Decision-Making Process

○Chie Hieida[1], Takato Horii[1], Takayuki Nagai[1] （1. The University of Electro-Communications）

5:40 PM - 6:00 PM

[2G4-03] Effect of Sympathy expression method by Face and Body of Educational-Support Robot in Long Term Experiment

○Yuhei Tanizaki[1], Felix Jimenez[1], Tomohiro Yoshikawa[1], Takeshi Furuhashi[1], Masayoshi Kanoh[2] （1. Nagoya University, 2. Chukyo University）

6:00 PM - 6:20 PM

[2G4-04] Segmenting Time Series Data Using GP-HSMM with Nonparametric Bayesian Model

○Masatoshi Nagano[1], Tomoaki Nakamura[1], Takayuki Nagai[1], Daichi Mochihashi[2], Ichiro Kobayashi[3], Masahide Kaneko[1] （1. The University of Electro-Communications, 2. Institute of Statistical Mathematics, 3. Ochanomizu University）

6:20 PM - 6:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

重み付けランダムフォレストによるペアワイズ分類を用いたロボット表出感情の主副推定

Emotion Estimation from Robot Movements Using Weighted Random Forest and Pairwise Classification

加藤 瑛樹[*1] 加藤 昇平[*1*2]
Hideki Kato Shohei Kato

[*1]名古屋工業大学大学院工学研究科情報工学専攻
Department of Computer Science and Engineering Graduate School of Engineering, Nagoya Institute of Technology

[*2]名古屋工業大学 情報科学フロンティア研究院
Frontier Research Institute for Information Science, Nagoya Institute of Technology

Recently, various communication robots are developed. These robots have functions to express emotions. Emotions are influenced by verbal and nonverbal information. Nonverbal information including facial expression and body movements has a greater influence on emotional interaction than verbal information has. In this paper, we focus on body movement, which is one of way for a robot to express its emotions for communication using nonverbal information. Emotions are expressed with some basic emotions. We defined two emotions, a primary emotion and a secondary emotion. We propose an estimation method using weighted random forest and pairwise classification. We confirm effectiveness of the proposed method.

1. はじめに

近年，様々なコミュニケーションロボットが開発されている．このようなロボットの多くには感情を表現する機能があり，より人間らしいロボットの実現などの役割を果たしている．ロボットと人のコミュニケーションにおいて感情は重要な要素として考えられており，心理的要因を取り入れた人とロボットのコミュニケーションに関する研究は盛んに行われている [2]〜[5]．感情に影響する要因は会話などの言語情報だけではなく，表情・視線・身体動作などの非言語情報も影響を与える．

本研究では人とロボットの円滑なコミュニケーションを目指し，非言語情報の身体動作に着目したヒューマノイドロボットの身体動作による適切な感情表出機能の実現を目標としている．ロボットが身体動作を用いて適切な感情を表出する機能の実現には，次のような機能が必要であると考えられる．

1. 自己の身体動作が相手に与える感情を自覚できる
2. ある感情を表出する身体動作を自律的に生成できる
3. 表出すべき感情を適切に表出できる

本稿ではラバン身体動作表出理論 [6]（以下,「ラバン理論」）に基づいた身体動作の特徴（以下,「ラバン特徴」）を定義し，1. の機能として機械学習手法の 1 種であるランダムフォレスト [10] の決定木の出力をそれぞれのクラスの得票率と得票数として用いたペアワイズ感情分類手法をそれぞれ提案する．それぞれの感情推定手法の概略と感情推定の一例を図 1 に示す．可能な感情の組み合わせそれぞれについて 2 クラス分類モデルを学習し，それらの出力を統合することで主副関係にある感情を推定する．提案手法については第 4 章で詳しく述べる．

心理学における感情の捉え方の一説として，いくつか存在する基本感情の程度の違いやそれらの混合によって異なる感情が表れる考え方があり，代表的な感情モデルとして Plutchik[7] や Ekman[8]，Russell[9] などがある．Plutchik のモデルでは人は 8 種類の行動の原型があり，それぞれに対する基本感情

連絡先: 加藤昇平，名古屋工業大学，愛知県名古屋市昭和区御器所町，052-735-5625，shohey@katolab.nitech.ac.jp



(a) 得票率による推定手法　(b) 得票数による推定手法

図 1: 感情推定手法の概略と感情推定の一例

（純粋感情）があるとしている．8 種類の基本感情は「喜び−悲しみ」「受容−嫌悪」「恐れ−怒り」「驚き−期待」の両極の組み合わせがある．また Ekman のモデルでは顔の表情について「恐れ」「驚き」「怒り」「嫌悪」「悲しみ」「喜び」6 種類の基本感情を示している．Russell の感情モデルでは図 2 のように快と不快，覚醒と眠気の 2 軸で表される 2 次元平面上に様々な感情が配置されるとしており，2 軸によって 2 次元平面にできる 4 つの領域の代表的な感情として喜怒哀楽が多く使われる．これらの感情モデルに関して共通していることは感情について普遍的なものが存在していると考えられている点である．本稿では Russell の感情モデルを参考に，二次元平面上の原点における感情無しと，4 つの領域の代表的な感情である喜怒哀楽の 4 感情の計 5 種類の感情を対象とする．Russell の感情モデルにおいて喜と怒や喜と楽のように領域が隣り合っている感情は互いに主副の関係になり得ると考えられる．

2. ラバン身体動作表出理論

本節ではラバン理論 [6] に基づいた身体動作の特徴であるラバン特徴の抽出手法を説明する．

2.1 ラバン理論

ラバン理論 [6] とは動作体の心理状態と身体運動の相関を規定する舞踊学の理論である．Darwing が提唱した「動物と身体動作の構造に関する理論」を受け継ぎ，Rudlf von Laban を

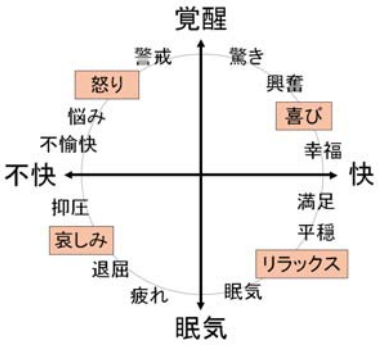


図 2: ラッセルの感情モデルの概略図

表 1: ラバン特徴量

ラバン特徴	特徴	特徴量の抽出法
$Space$	身体動作の 方向的偏り具合	四肢先端の単位運動ベクトル 顔単位方向ベクトルの 内積の重み付き合計
$Time$	身体動作の速さ	全関節の角速度の重み付き合計
$Weight$	身体動作の力強さ	全関節のトルクの重み付き合計
$Inclination_x$	姿勢の 前方向への偏り	重心位置と支持脚中心位置の 前後方向の差
$Inclination_y$	姿勢の 左右方向への偏り	重心位置と支持脚中心位置の 左右方向の差
$Height$	姿勢の高さ	重心位置と基本重心位置との 上下方向の差
$Area$	姿勢の 水平面上の広がり	四肢先端が水平面上に作る 四角形の面積

中心とするドイツの表現主義舞踊の創作者らが構築した．ラバン理論の概念は数理的かつ具体的であるため，理工学との親和性が高いと言われている．

2.2 ラバン特徴

本研究では，ラバン特徴量の抽出手法として増田ら [1] が提案した抽出手法を使用する．表 1 に増田らが提案した抽出式で得られるラバン特徴・特徴量の概要・特徴量抽出方法の概要を示す．増田らはヒューマンフォームロボット（HFR）KHR-2HV の全身動作からラバン特徴量を得る抽出式を提案し，ラバン特徴量と喜怒哀楽の 4 感情に高い相関があることを示した．時刻 t における各ラバン特徴量は，以下に提示する抽出式を用いて算出される．

$Space$ は身体動作の方向的な偏りの大きさの程度を示す特徴量であり，四肢先端の単位運動ベクトル（運動ベクトルの単位ベクトル），および顔単位方向ベクトルの内積の重みつき合計と定義する．値が大きい場合は身体動作は単一方向に向かっており，値が小さい場合は身体動作は散逸方向に向かっている．$Space(t)$ は下式で算出される．

$$Space(t) = \sum_{m=0}^{3} \sum_{n=m+1}^{4} a_{mn}(\overrightarrow{\mathrm{r_m}}(t) \cdot \overrightarrow{\mathrm{r_n}}(t)) \tag{1}$$

ここで，m，n は顔および四肢先端の識別番号を示す．$\overrightarrow{\mathrm{r_m}}$ は顔の単位方向ベクトル（m=0 の場合），あるいは四肢先端 m の単位運動ベクトル（m=1,2,3,4 の場合）を示す．a_{mn} は重み係数であり，m，n に関連する部位の視覚的質量と被注目度を考慮し定義者が設定する．

$Time$ は身体動作の速さの程度を示す特徴量であり，全関節の角速度の重みつき合計と定義する．値が大きい場合は身体動作は素早く，値が小さい場合は身体動作は遅い．$Time(t)$ は下式で算出される．

$$Time(t) = \sum_{i} b_i \dot{\theta}_i(t) \tag{2}$$

ここで，i は関節番号を示し $\dot{\theta}_i$ は関節 i の角速度を示す．b_i は重み係数であり，関節 i に関連する部位の視覚的質量と被注目度を考慮し定義者が設定する．$Weight$ は身体動作の力強さの程度を示す特徴量であり，全関節のトルクの重みつき合計と定義する．値が大きい場合は身体動作は力強く，値が小さい場合は身体動作は弱々しい．$Weight(t)$ は下式で算出される．

$$Weight(t) = \sum_{i} c_i \tau_i(t) \tag{3}$$

ここで，i は関節番号を示し τ_i は関節 i のモータのトルクを示す．c_i は重み係数であり，関節 i に関連する部位の視覚的質量と被注目度を考慮し定義者が設定する．

$Inclination$ は姿勢の偏りの程度を表す特徴量であり，重心位置と支持脚中心位置の差と定義する．$Inclination$ は前後 (x 方向)，左右 (y 方向) の両成分についてそれぞれ求められる．それぞれを $Inclination_x$，$Inclination_y$ とし，$Inclination_x$ が大きい場合は姿勢は前寄りであり，$Inclination_x$ が小さい場合は姿勢は後ろ寄りである．また $Inclination_y$ が大きい場合は姿勢は左寄りであり，$Inclination_y$ が小さい場合は姿勢は右寄りである．$Inclination_x(t), Inclination_y(t)$ は下式で算出される．

$$Inclination_x(t) = cog^x(t) - col^x(t) \tag{4}$$

$$Inclination_y(t) = cog^y(t) - col^y(t) \tag{5}$$

ここで，cog^x は HFR の重心の x 座標を，cog^y は HFR の重心の y 座標をそれぞれ示し．col^x は支持脚中心の x 座標を，col^y は支持脚中心の y 座標をそれぞれ示す．

$Height$ は姿勢の高さの程度を表す特徴量であり，重心位置と基本重心位置（直立状態での重心位置）との上下方向との差と定義する．値が大きい場合は姿勢は上寄りであり，値が小さい場合は姿勢は下寄りである．$Height(t)$ は下式で算出される．

$$Height(t) = cog^z(t) - cog_0^z \tag{6}$$

ここで，cog^z は重心の z 成分を示し cog_0^z は基本姿勢（直立）における重心の z 成分を示す．

$Area$ は姿勢の水平面上の広がりを表す特徴量であり，四肢先端が水平面上に作る四角形の免責と定義する．値が大きい場合は姿勢は広く，値が小さい場合は姿勢は狭い．$Area(t)$ は下式で算出される．

$$Area(t) = \sum_{m=1}^{3} \sum_{n=m+1}^{4} \frac{1}{2} |Pos_m^x(t) Pos_n^y(t) + Pos_m^y(t) Pos_n^x(t)| \tag{7}$$

m，n は四肢先端の識別番号を示し，Pos_m^x，Pos_m^y は四肢先端の x 座標と y 座標をそれぞれ示す．

抽出式（1）から（7）を用いて抽出された数値を特徴ごとに正規化し，これを身体動作におけるラバン特徴量の時系列データとする．

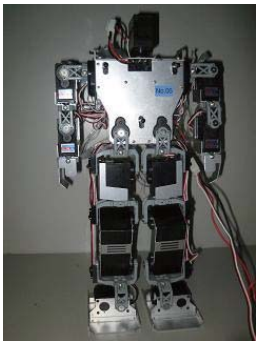


図 3: KHR-2HV の外観

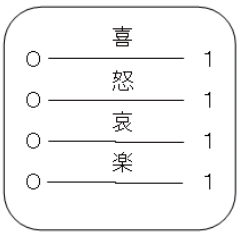


図 4: 回答用紙

表 2: 各感情を表出する動作の個数

喜	怒	哀	楽	複数感情	無感情
12(1)	6(1)	11(0)	9(1)	3	8

表 3: 主感情−副感情の内訳

		副感情			
		喜	怒	哀	楽
主感情	喜		1	0	1
	怒	0		0	0
	哀	0	0		0
	楽	1	0	0	

短時間の身体動作を対象とした場合，身体動作時間全体のラバン特徴量の平均をその身体動作のラバン特徴量とみなすものとする．例えば短時間身体動作の $Space$ は下式で算出される．

$$Space = \frac{1}{\frac{N}{T}} \int Space^*(t)dt \tag{8}$$

ここで，N は動作時間長を，T は単位時間長を示し，$Space^*(t)$ は $Space$ 全体に対して正規化した $Space(t)$ を示す．他 6 種類のラバン特徴についても（8）式と同様に短時間身体動作のラバン特徴量を算出する．

3. 予備実験

身体動作が表出する感情を取得する目的で，印象評価実験を行った．

3.1 印象評価実験

被験者は 20 代の男女 15 名，身体動作は HFR である KHR-2HV の 8[s] 間の全身動作を 40 種類用意した．図 3 に KHR-2HV の外観を，図 4 に印象評価実験に用いた回答用紙を示す．被験者は観察した動作から，喜怒哀楽それぞれの感情がどの程度感じられるかを推定し図 4 に示す形式の回答用紙に回答する．その感情を全く感じなければ 0 に印をつけ，その感情を強く感じるほど 1 に近い位置に印をつける．回答の印の位置に応じて 0〜100 までの評価値に線形変換し，ある動作について 15 人のそれぞれの感情の評価値の平均を「感情評価値」と定義する．回答を喜怒哀楽のどれか 1 つとしなかったのは，身体動作が混合感情を表出することがあり，いくつかの基本感情が様々な感情強度で存在していると考えたためである．

3.2 身体動作の表出感情

予備実験で得られた感情評価値を用いて身体動作の表出感情を決定する．身体動作の表出感情は喜怒哀楽のどれか 1 種類だけではなく，複数の感情を表出する場合も考えられる．そこで本実験では，身体動作の表出感情を高々 2 種類と仮定し，採択される感情は以下の 2 種類とした．

1. 「主感情」
   少なくとも感情評価値が α 以上の値をとり，かつ感情評価値が最大となる感情
2. 「副感情」
   主感情が存在している動作について，2 番目に大きい感情評価値をとり，かつその値が主感情の感情評価値の β% より大きい感情

ただし α, β はパラメータであり，$\alpha = 40$，$\beta = 70$ に設定した．主感情として採択される条件を満たす感情が無い場合は副感情も表出されず，動作は感情を表出しないことになる．
以上のように身体動作が表出する感情を決定した結果，各感情を表出する動作の個数は表 2 のようになった．ただし（）内の値は，その感情を副感情として表出する動作の個数を表し，無感情の列は主感情と副感情のどちらも表出しない動作の個数となる．複数感情を表出し得る場合における主感情−副感情の内訳を表 3 に示す．ただし，主感情と副感情は同一の感情にはならないため除外してある．

4. 提案手法

提案手法ではランダムフォレスト [10] を用いたペアワイズ分類を行い，主感情と副感情を推定する．

4.1 ペアワイズ分類

ペアワイズ分類とは，K 個のクラスのうち 2 つを選択してできる組み合わせについてそれぞれの 2 クラスを分類するモデルを作成し，最終的に分類モデルから得られる出力の多数決でクラスを決定する分類手法である．本稿で考慮するクラスは喜怒哀楽の 4 クラスと感情無しの 1 クラスを合わせた計 5 クラスとなる．したがって用意する分類モデルの総数は ${}_5C_2 = 10$ となる．

4.2 ランダムフォレスト

ランダムフォレスト [10] は識別性能の低い決定木を弱学習機として多数用意し，それらの出力を統合して高い予測性能を得ることができる機械学習手法の 1 種である．特徴としてノイズに頑健で高速な識別が可能であり，高次元・低サンプルに強い点があげられる．以下に本研究で用いるランダムフォレストのアルゴリズムを示す．

1. データセット $\boldsymbol{D}^{train}$ から N 個のブートストラップサンプル $B_1, B_2, \cdots, B_N$ を作成する．ただし，クラスは C_1, C_2 の 2 クラスのみとする．
2. 作成した各サンプル $B_k (k = 1, 2, \cdots, N)$ をトレーニングデータとして，N 本の決定木 $T_k (k = 1, 2, \cdots, N)$ を作成する．
3. 指定したノード数に達するまで，以下の方法で決定木のノードを作成する．
   (a) トレーレニングデータ B_k の説明変数 M 個のうち，m 個をランダムに選択する．$m = \sqrt{M}$ が多く用いられる．

(b) 選択された m 個の説明変数のうち，トレーニングデータを最もよく分類するものと，そのときの閾値を用いてノードの分割関数を決定する．

4. データセット$\boldsymbol{D^{train}}$のデータ $d_i(i = 1, 2, \cdots, D)$ をそれぞれ決定木の群に対し入力として与え，決定木が出力した C_1, C_2 クラスの割合 R_1^{train}, R_2^{train}（得票率と定義する）をそれぞれ算出する．

学習に用いていないデータセット$\boldsymbol{D^{test}}$も同様にして決定木が出力し各クラスの得票率 R_1^{test}, R_2^{test} を算出する．また得票率が大きいクラスを推定出力とすることもできるため，感情推定実験でこれら 2 つの比較を行う．

4.3 提案手法

図 1 に示ししたように，5 種類の感情から 2 つ選択してできる全ての組み合わせに対して，ラバン特徴から 2 つの感情のどちらに分類されるかを推定するモデルをランダムフォレストを用いて学習する．それぞれの 2 クラス分類モデルで得られた $R_1^{train}, R_2^{train}, R_1^{test}, R_2^{test}$ を感情ごとに合計する．ある動作について 5 つのクラスごとに得票率の合計が得られ，得票率の合計が最大の感情を主感情，2 番目に大きな感情副感情として出力する．あるいは，4.2 で説明したように得票率が大きなクラスのみを推定出力として投票し，得票数が最大の感情を主感情，2 番目に大きな感情を副感情として出力する手法も比較を行う．前者の手法を得票率による推定，後者の手法を得票数による推定と呼ぶことにする．

5. 感情推定実験

予備実験で扱った 40 種類の身体動作について提案手法 2 種類を用いた感情推定実験を行った．また，主感情・副感情の重みは違うと考えられるため，主感情と副感情の 2 つで学習データとして与える個数に違いを与える．主感情として採択された感情は学習データの個数を 2 個として入力し，副感情として採択された感情は学習データの個数を 1 個として入力し，この方法を重み付け有りと呼ぶことにする．また，主感情と副感情の 2 つで学習データに違いを与えない方法を重み付け無しと呼ぶことにする．したがって感情推定実験では重み付け無しの得票率による推定・得票数による推定，重み付け有りの得票率による推定・得票数による推定の 4 種類の実験をすることになる．

5.1 実験結果

正答率を表 4 に示す．手法・重みはそれぞれ得票率と得票数による推定，重み付けの有無の違いを表している．各感情の列は推定出力と正解データが一致した場合の正答率となり，完全正答は 5 感情全て正答した場合の正答率となる．推定モデル全体の性能は完全正答の正答率となる．

5.2 考察

重み付けの有無について考察する．得票率・得票数による推定のどちらにおいても学習データの個数を増やすことの重み付けが有りの方が正答率が大幅に向上している．これは主感情として採択された感情のデータを増やすことにより主副の違いを学習できたことを示している．次に得票率と得票数による推定について考察する．2 クラス分類の得票率の差が非常に小さい場合でも，得票数による推定の場合はその差を無視して得票率が大きいクラスを出力とする．これに対して得票率による推定ではこの差も考慮して推定出力を求めることができると考えた．しかし，表 4 を見ると得票数による推定の方が性能が良いことがわかる．これはある 2 クラス分類モデルでの得票率の差が小さい場合に，他の分類モデルでの得票率が加算されることで得票率の順位の逆転が起こりうるためであると考えられる．この課題については，データセット数を増やしても変化がないのかを調査する必要がある．

表 4: 正答率

手法	重み	喜	怒	哀	楽	無	完全正答
得票率	無	69.5	92.1	99.1	94.7	73.9	66.2
得票率	有	91.4	92.4	100	92.4	86.5	86.6
得票数	無	73.4	90.5	99.4	94.4	77.3	68.9
得票数	有	92.5	92.5	100	92.9	88.0	88.0

6. おわりに

本研究ではロボットの身体動作を用いた感情表現機能の実現に向け，自己の身体動作が表出する感情の推定手法を提案した．提案手法の特徴としては主感情・副感情の複数感情の推定が可能になった点であり，1 つの感情と比較して難易度の高い推定にも関わらず高い正答率が得られた．また，主感情と副感情で学習データの個数に差をつけることで正答率が向上することを確認できた．今後の課題として，主感情と副感情を決定するパラメータの再検討やより多くのデータでの実験が必要であることがあげられる．また，身体動作にある感情を付与させる身体動作生成の研究も行う予定である．

参考文献

[1] 増田 恵，加藤 昇平，伊藤 英則:ラバン理論に基づいたヒューマンフォームロボット身体動作の動作特徴抽出と表出感情推定，日本感性工学会論文誌，*Vol.10,No.2,p.295-303*，2011

[2] 柴田 寛, 加納 政芳, 加藤 昇平, 伊藤 英則:感性ロボット ifbot の感情空間を用いた感情遷移に伴う表情変化の主観的影響, 日本知能情報ファジィ学会誌, 知能と情報,Vol. 21, No. 5, pp 630-639 (2009)

[3] 竹内 将吾, 酒井 あゆみ, 加藤 昇平, 伊藤 英則:対話者好感度に基づく感性会話ロボットの感情生成モデル, 日本ロボット学会誌,Vol. 25, No. 7, pp 1125-1133 (2007)

[4] 加藤瑛樹, 加藤昇平:ラバン特徴と Autoencoder を用いたヒューマノイドロボットの身体動作生成に関する研究, 人工知能学会全国大会, ウインクあいち, Vol.31, pp.1G3-2(4-pages), May 23-26, 2017

[5] 中田 亨, 森 武俊, 佐藤 知正:ロボットの身体動作表現と生成される印象とのラバン特徴量を介した定量的相関分析, 日本ロボット学会誌,Vol. 19, No. 2, pp. 252-259(2001)

[6] Laban，R. V.:mastery of Movement，Princeton Book Co. Pub. 1988

[7] Plutchik, R.: A General Psychoevolutionary Theory of Emotion, in Emotion: Theory researh, and experience: Vol. 1. Theories of emotion, pp.3-33, New York: Academic (1980)

[8] P. Ekman and W.V. Friesen, Facial Action Coding System, Consulting Psychologiests Press (1978)

[9] James A. Russell. A circumplex model of affect. In Journal of Personality and Social Psychology, pp. 1161-1178, 1980.

[10] Breiman L. : Random forests，Machine Learning,45,pp.5-32，2001

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

行動決定に基づく感情分化

Emotional Differentiation based on Decision-Making Process

日永田智絵 *1　堀井隆斗 *1　長井隆行 *1
Chie Hieida　Takato Horii　Takayuki Nagai

*1電気通信大学
The University of Electro-Communications

Having emotions is essential for robots to understand and sympathize with the feelings of people. In addition, it may allow the robots to be accepted into human society. The role of emotions in decision-making is another important perspective. In this paper, a model of emotions based on various neurological and psychological findings that are related to empathic communication between humans and robots is proposed. Subsequently, a mechanism of decision-making that is based on affects using convolutional LSTM and Deep Deterministic Policy Gradient is examined.

1. はじめに

ロボット研究では感情について深く扱えずにいる．それは一般に「ロボット」が感情が無い存在として考えられがちなことからも示唆される．相手の情動を推定する手法 [Picard 97] や情動を表出する方法 [Breazeal 02] については，従来から研究がなされている．しかしながら，こうした研究の情動は作りこみであることが多い．作りこみでは社会的な感情の様な複雑な感情を作ることが困難であり，実際シンプルな基本情動のようなものしか実現できていない [Masuyama 15][Woo 15]．しかし，ロボットが人間社会に受け入れられていくためには，相手の感情を理解・共感し，行動することが必要不可欠であり，シンプルな基本情動のようなものだけでは対応できない．

作り込みではない方法として，尾形らは，内分泌系モデルをロボットに実装し，情動を実現している [尾形 99]．また，浅田らは，作りこみではなく，人の多感覚情動信号からロボットが情動を学習する情動発達ロボティクスの研究を行っているが，これは情動の発達が中心であり，上位概念である感情は扱っていない [浅田 14][堀井 16]．よって本研究では，上位概念である感情を扱うための感情モデルを提案する．将来的にはこの研究を通して，ロボットがどのように複雑な感情を持つか，相手の情動に基づいてどのように感情を変化させるか，そうした相互作用がどのようなコミュニケーションを引き起こすのかといったことを検討する．このことは，ロボットに感情を持たせるだけでなく，ロボットが人間の心的状態を真に理解するという意味でも必要不可欠である．これによってロボットが人の多様なあり方に寄り添い，人とロボットの双方が学習し成長していけるような関係性になることを期待している．

本稿では以下のような神経科学や心理学等の様々な分野の文献を基に感情モデルを作成した．情動や感情の分化に関しては以下のような文献が挙げられる．心理学者の Bridges は乳幼児の観察を基に，情動が興奮を原点として分化していくと考えた [Bridges 32]．また Ledoux は，快情動と不快情動では異なる脳部分が働いていることを明らかにしている [Ledoux 98]．Ekman は，文化に関係なく共通の基本 6 感情 (怒り，喜び，嫌悪，恐怖，悲しみ，驚き) が存在することを示した [Ekman 71]．

情動と感情の役割に関する文献は以下のようなものが挙げられる．神経科学者の Damashio は，情動が外部からの刺激の評価を効率的にするという仮説を立てた（ソマティック・マーカー仮説）[Damashio 96]．大森は，感情とは意思決定のための価値計算システムであると提唱している [大森 16]．一般的にも知られている吊り橋効果では自身の身体反応の原因推論の際に，実際の原因とは異なった原因に帰着することで，感情の認知が変化するといわれている [Dutton 74]．

これらのことから，感情システムは，無意識のうちに起こる身体反応と，それを感情として認知するという段階に分かれており，生得的なシステムと学習により強化されるシステムが存在していると考えられる．上記を踏まえ，情動を刺激によって引き起こされる身体反応のラベル，感情を行動決定のための状態に対するラベル，状態とは刺激と情動及び刺激と情動を用いた原因推論・未来予測の結果と定義し，感情モデルを構築する．

感情の重要な役割の一つが，行動決定である．したがって，提案モデルでは行動決定を行うための大脳皮質–大脳基底核ループに対応したシーケンス学習と強化学習を導入する．本稿では行動選択に基づく感情の分化を議論するため，提案モデルの中でも行動決定モジュールの実装を convolutional LSTM (long short-term memory)[Xingjian 15] と DDPG (Deep Deterministic Policy Gradient)[Lillicrap 15] を用いて行った．そして，モジュールの検証のため，Agent での表情決定タスクを行った．

2. 感情モデル

提案する感情モデルを図 1 に示す．感情モデルは以下の 3 層に分かれており，上記に示した文献を基に作成されている．

第 1 層：評価モジュール
刺激を評価し，身体的反応を行う
第 2 層：行動決定モジュール
次元圧縮された身体的反応（情動）と刺激に基づいて，原因推論・未来予測を行っている
この予測結果と刺激と情動を行動決定のためにカテゴライズしたものを人は感情として知覚している
第 3 層：感情記憶モジュール
経験を記憶し，記憶に基づき刺激を評価し生得的反応を抑制・強化する

第 1 層は反射的に反応する層であり，時間的に処理が最も早いが，エラーも多い．それに対し，第 3 層では過去の記憶にアクセスするため，第 1 層からの遅延があるが，経験に基づい

連絡先: 日永田智絵，電気通信大学，東京都調布市調布ヶ丘 1-5-1，042-443-5238，hchie@apple.ee.uec.ac.jp

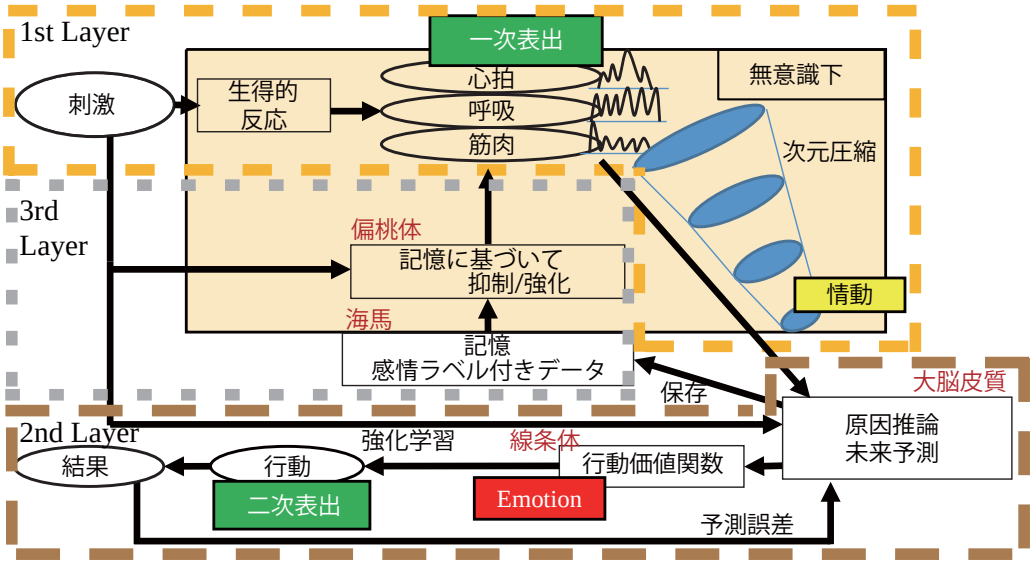


図 1: 感情モデル

て評価するため，エラーを減らすことができる．この第 1 層第 3 層から出力された一次表出を次元圧縮したものが情動である．つまり，従来研究の情動信号はこの一次表出のことを表しており，無意識下で行われるため，嘘をつくことのできない正直な反応となる．Pentland が提唱している Honest Signals はこの一次表出の一部と考えられる [Pentland 08]．情動信号と情動の関係は次元圧縮しただけの関係であり，情動信号から情動を導くことができる．

こうして得られた情動が入力信号とともに原因推論，未来予測に使用される．吊り橋効果ではこの情動が異なる入力信号に紐づけられたため，未来予測が変化し，異なるカテゴリに分類されたため，恐怖から好意へと変化が起きたと考えられる．ここで生じたカテゴリは意識へと昇り感情として認知される．それに基づき行動を起こし，その結果で得られた状況から予測誤差を計算し，原因推論，未来予測の識別関数を学習する．過去につけた評価データは記憶として保存される．

3. 行動決定モジュールの実装

先行研究では第 1 層の実装として，recurrent-attention model (RAM) [Mnih 14] を利用した視覚刺激による情動生成を検討した [Hieida 17]．この研究では，画像刺激として International Affective Picture System (IAPS)[Lang 99] を入力とし、valence と arousal を出力とする教師あり学習を行った．本研究では convolutional LSTM and DDPG を用いて，行動決定モジュールを実現する (図 2)．Convolutional LSTM は情動値と視覚刺激を入力として，次のステップの予測情動値と予測刺激を出力する．DDPG は情動値と視覚刺激，予測情報を状態として入力し，行動を出力する．この時強化学習の報酬は以下の通り設定する．

$$Reward[t] = Constant - \|Affect[0] - Affect[t]\|^2, \quad (1)$$

$Affect[t]$ は時間 t の時の情動値である．情動値は先行研究 [Hieida 17] での RAM による外的刺激の評価と，行動による体力の増減などの影響を受ける内的刺激の評価を合わせたもので表現される．本研究では $Affect[0]$ を [valence, arousal]=[5, 5] と設定している．この報酬はホメオスタシスの考えを用いており，提案モデルは情動値つまり身体反応を一定にするように行動を学習する．

4. Agent の表情生成

4.1 実験プロトコル

実装した行動決定モジュール内での感情分化を検証するため，共感コミュニケーションを想定したタスクを設定し，Virtual Agent を用いて学習を行った．本タスクでは行動として Agent は瞼，眉，口，口角を動かし表情を変化させる．Interaction Partner は Agent の表情を 4 つのカテゴリとして認識し，対応した同カテゴリの表情を画像として出力する．表情カテゴリ認識は以下のようなルールベースとなっている．

Pleasure: 口角が上がっているとき

Angry: 口角が下がっており，眉を顰めていて，目が半分よりも開いているとき

Sadness: 口角が下がっており，眉を顰めていて，目が半分よりも閉じているとき（伏目）

Neutral: 上記以外

Agent は Interaction Partner の表情画像を視覚刺激として受け取り，感情モデルで処理を行って，行動を出力する．本タスクでの情動値は，RAM により Interaction Partner の表情画像を評価した値と行動に対応した内的刺激の評価を合わせたもので表現されている．行動と内的刺激評価の対応はルールベースとなっており，基本は行動すればするほど内的評価値は大きくなる．例えば瞼の場合は閉じたときが最も値が小さく，開けば開くほど値が大きくなるといった形になっている．例外として，悲しみの表情をしたときと目を閉じたときに内的評価値を下げるルールを設定しており，これは泣いたときにミルクがもらえるということと，目を閉じたときには寝ているという赤ちゃんの振る舞いから設定したものである．即ちこの内的評価値は，体力バイアスであり，外的評価に対して体力バイアスをかけることによって，情動が表現されている．

本学習を 50000epoch 行い，学習が進むごとに，ソマティック・マーカーに対応する Policy Network の中間層を PCA により可視化することで，どのような状態空間が構成されている

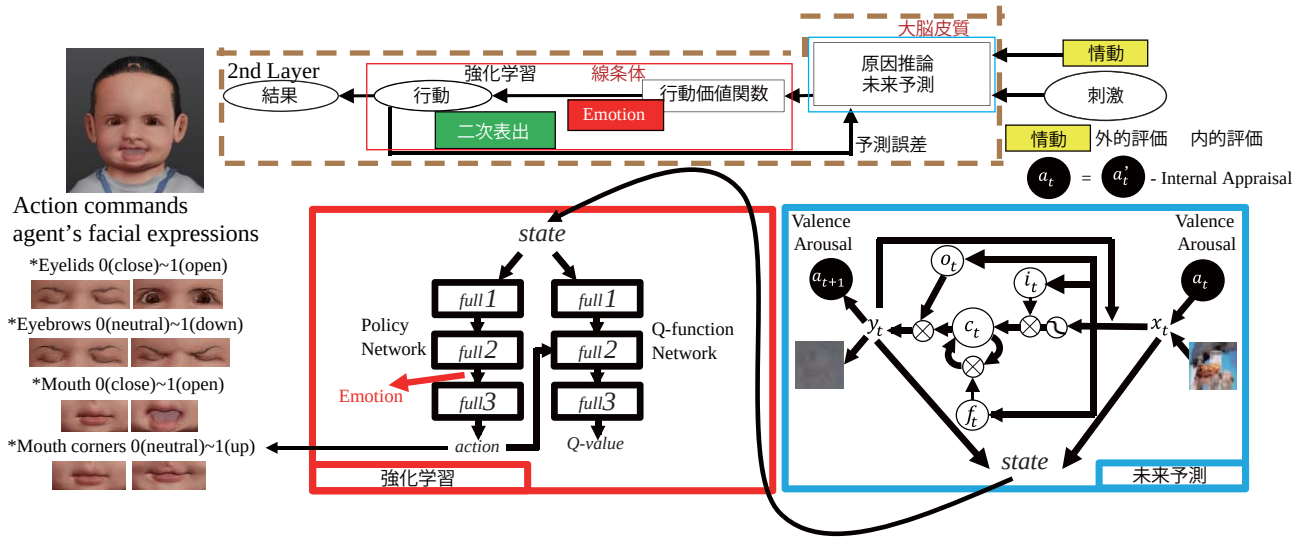


図 2: Convolutional LSTM と Deep Deterministic Policy Gradient による行動決定

かを検証する．我々の感情の定義が正しく，適切な実装が行えていれば，この状態空間は行動により感情カテゴリに分割することができ，学習が進むごとに分化していく様子が観察できるはずである．

4.2 結果

結果を図 3 に示す．図 3(a) は情動値をプロットしたもので，図 3(b) は Policy Network の中間層を PCA により可視化したものである．各色は Interaction Partner が認識している表情のカテゴリを表している．どちらの結果においても，学習が進むごとに各カテゴリがまとまっていることがわかる．これより，感情が分化する様子が表現されている．情動値のプロットでも各カテゴリが分かれているのは，今回のタスクが Agent の働きかけで変化する刺激のみ利用したことが原因だと考えられる．より複雑なコミュニケーションにおいて刺激が多様化すると情動値のみではカテゴリが分けられなくなり，行動も含めた感情によるカテゴリ表現が重要となると予想できる．また，情動値の 40000-50000epoch のプロットでは Neutral が一番 valence の高い位置に存在しており，次が Pleasure となっている．これは $Affect[0]$ をどこに設定するかによって異なると考えられる．本研究では著者らが設定した任意の値だが，ある一定長さのシーケンス内での情動値平均を設定することによって，不快な刺激を浴び続けると怒りっぽくなるといった，その日の気分といったようなものを表現できる可能性がある．

また，学習済みモデルで Agent を動かし，観察した結果以下のような行動の変化が見られた．*1

学習初め： Agent はよく目をつぶっている
10000epoch： よく目を開けるようになる
20000epoch： 目の開閉を頻繁に繰り返す
30000epoch： 目の開閉をゆっくり繰り返す
40000epoch： 表情のバリエーションが増える
50000epoch： よく笑うようになり，情動値が下がってくると目を閉じる

よく笑うようになるということは，相手を笑わせることに繋がるため，他者からみると利他的な行動を行っているようにもみえる．しかしながら，実際には，自分がほしい刺激を得るために笑っており，本研究の Agent は見せかけの利他行動を学習したことになる．

*1 学習済みモデルで Agent を動かした様子を動画で示す．https://youtu.be/mF6AfX2Wzko

5. まとめ

本研究では，共感コミュニケーションのための感情モデルを提案し，行動決定の実装を行った．行動決定の実装は convolutional LSTM と DDPG を用いて行い，報酬にホメオスタシスの考えを適用した．検証として，Virtual Agent を用いた表情決定タスクを行い，感情の分化が観察されたと共に，見せかけの利他行動の様子が見られた．今後は，全体の感情モデルの実装を行い，嫉妬や恥，罪悪感などといった社会的な複雑な感情の創発を目指す．

謝辞

本研究は JSPS 科研費 JP16J04930，JST CREST (JPMJCR15E3)，新学術領域「認知的インタラクションデザイン学」 (26118001) の助成を受けたものです．

参考文献

[Breazeal 02] Breazeal, C.: Designing Sociable Robots, *The MIT Press* (2002)

[Bridges 32] Bridges, K. M. B.: Emotional development in early infancy, *Child development*, pp. 324–341 (1932)

[Damashio 96] Damashio, A. R., Everitt, B. J., and Bishop, D.: The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion], *Philosophical Transactions of the Royal Society B, Biological Sciences*, Vol. 351, No. 1346, pp. 1413–1420 (1996)

[Dutton 74] Dutton, D. G.: Some Evidence for Heightened Sexual Attraction under Conditions of High Anxiety,

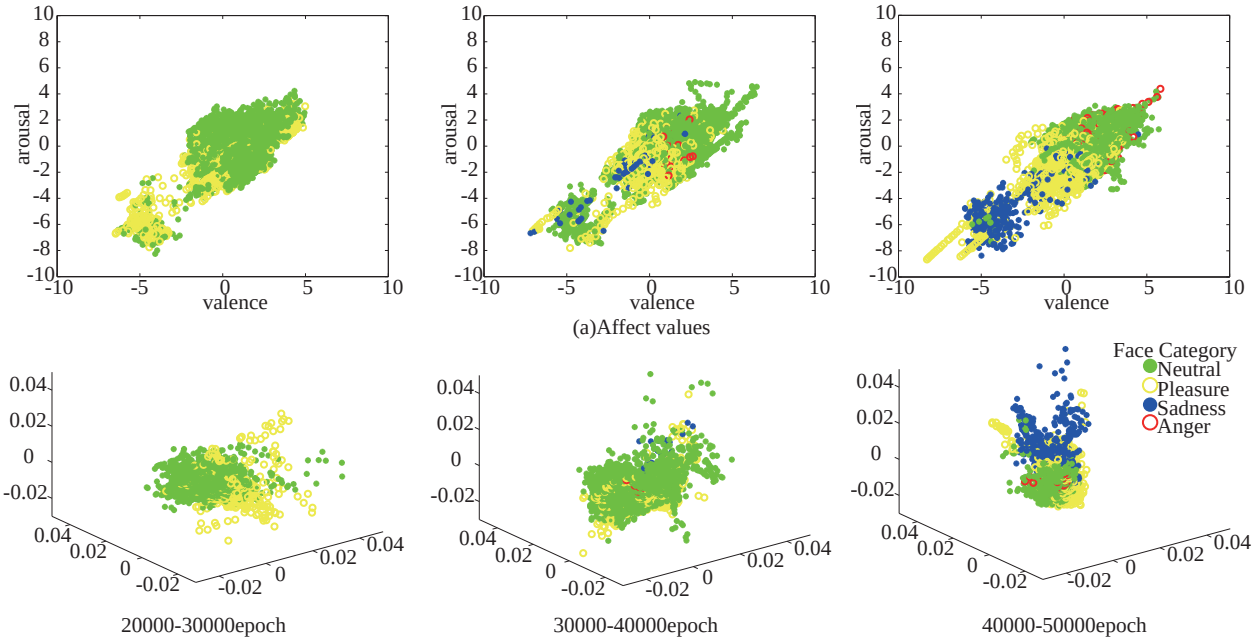


図 3: 行動決定による感情分化

Journal of Personality and Social Psychology, Vol. 30, No. 4, pp. 510–517 (1974)

[Ekman 71] Ekman, P. and Wallace, F. V.: Constants across cultures in the face and emotion, *Journal of personality and social psychology*, Vol. 17, No. 2, pp. 124–129 (1971)

[Hieida 17] Hieida, C. and Nagai, T.: A Model of Emotion for Empathic Communication, *Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 133–134 (2017)

[Lang 99] Lang, P. J., Bradley, M. M., and Cuthbert, B. N.: International affective picture system (IAPS): Technical manual and affective ratings, *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida* (1999)

[Ledoux 98] Ledoux, J. E.: The Emotional Brain: The Mysterious Underpinnings of Emotional Life, *Simon & Schuster* (1998)

[Lillicrap 15] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D.: Continuous control with deep reinforcement learning, *arXiv preprint arXiv:1509.02971* (2015)

[Masuyama 15] Masuyama, N. and Loo, C. K.: Robotic emotional model with personality factors based on Pleasant-Arousal scaling model, *In Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on. IEEE*, pp. 19–24 (2015)

[Mnih 14] Mnih, V., Heess, N., Graves, A., and kavukcuoglu, K.: Recurrent Models of Visual Attention, in *Advances in Neural Information Processing Systems 27*, pp. 2204–2212 (2014)

[Pentland 08] Pentland, A. S.: HONEST SIGNALS —How They Shape Our World—, *The MIT Press* (2008)

[Picard 97] Picard, R.: Affective Computing, *MIT Press. Cambridge* (1997)

[Woo 15] Woo, J., Botzheim, J., and Kubota, N.: Verbal conversation system for a socially embedded robot partner using emotional model, *In Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on. IEEE*, pp. 37–42 (2015)

[Xingjian 15] Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in *Advances in neural information processing systems*, pp. 802–810 (2015)

[浅田 14] 浅田稔：情動発達ロボティクスによる人工共感設計に向けて, 日本ロボット学会誌, Vol. 32, No. 8, pp. 666–677 (2014)

[大森 16] 大森隆司：人はなぜ感情をもつのか—行動決定における感情の計算論的役割—, 人工知能学会誌, Vol. 31, No. 5, pp. 710–714 (2016)

[尾形 99] 尾形哲也, 菅野重樹：情動モデルを有する自律ロボット WAMOEBA-2 と人間との情緒交流, 日本機械学会論文集, Vol. 65, No. 633, pp. 166–172 (1999)

[堀井 16] 堀井隆斗, 長井志江, 浅田稔：Restricted Boltzmann Machine を用いた多感覚情動コミュニケーション, 第 30 回人工知能学会全国大会, pp. 1O5–OS–22b–2 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

長期実験における教育支援ロボットの表情変化と身体動作を用いた共感表出法による効果

Effect of Sympathy expression method by Face and Body of Educational-Support Robot in Long Term Experiment

谷嵜悠平 *1 Yuhei Tanizaki　ジメネスフェリックス *1 Felix Jimenez　吉川大弘 *1 Tomohiro Yoshikawa　古橋武 *1 Takeshi Furuhashi　加納政芳 *2 Masayoshi Kanoh

*1名古屋大学大学院工学研究科 Graduate School of Engineering Nagoya University　*2中京大学工学部 Engineering Chukyo University

Recently, educational-support robot have been attracting increasing attention as learning-support devices. Previous research has proposed the sympathy expression method that robot expresses emotions which sympathize with learner. In previous research, the sympathy expression method has been used for the robot which expresses its emotion by face. However, no research for sympathy expression method using the robot which expresses its emotion by using body or unifying body and face. Therefore, this paper examines the learning effect and the impression about three types robots which have different emotion expression methods in long term experiment.

1. はじめに

近年，教育現場で活躍できる教育支援ロボットの研究開発が行われている．例えば，ジメネスらは，人とロボットが問題を交互に解きあう協調学習を行うことができるロボットに関する研究を行っている [1]．しかしながら，教育支援ロボットには，学習が進むにつれて学習者がロボットとの学習に飽きてしまうという問題点がある．

従来研究では，教育支援ロボットの飽きの問題を解決するために，ロボットが学習者に共感するような感情を表出する，共感表出法が提案された [2]．そして被験者実験を通して，共感表出法を用いたロボットとの共同学習は学習者の飽きを軽減できる可能性があることを示唆した．しかしながら，この共感表出法は表情変化によって感情を表出するロボットのみを対象に検討されている．HRI の従来研究では，ロボットの身体動作は，人とのインタラクションやロボットの感情伝達に有効であると報告されている [3]．そのため，表情変化による共感表出法に身体動作を加えることで，共感表出法によって促される効果を高められると考える．

そこで本稿では，被験者実験を通して，表情変化と身体動作による共感表出法を用いた教育支援ロボットが学習者に与える効果を検討する．具体的には，身体動作によって共感表出を行うロボット，表情変化によって共感表出を行うロボット，表情変化と身体動作によって共感表出を行うロボットを比較することで，どのような共感表出を行うロボットが学習者との共同学習において有効なのかを検討する．

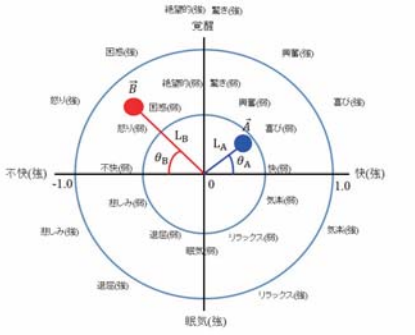


図 1: 共感表出法



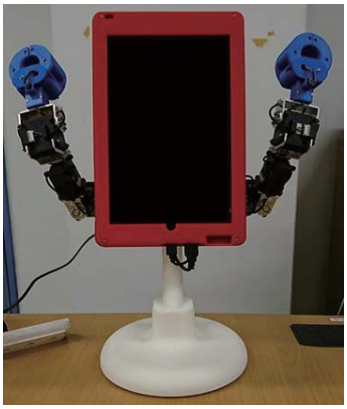
図 2: Tabot

2. 共感表出法

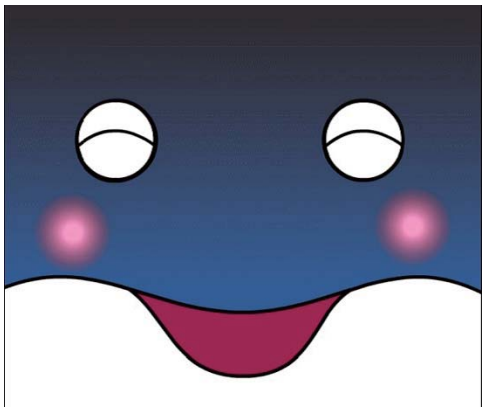
共感表出法は，人の感情変化を円環上のモデルにした Russell の感情円環モデルを基に構築されている (図 1)．共感表出法は，学習者にロボットが共感していると感じさせるために，正解用感情ベクトル $\vec{A}$ と不正解用感情ベクトル $\vec{B}$ の二つを用いて，円環モデル上における感情を表出する．学習者が問題に正解した場合は $\vec{A}$ を，不正解であった場合には $\vec{B}$ を基準に感情を表出する．$\vec{A}$ は $0 \leq L_A \leq 1.0$ かつ $-90° \leq \theta_A \leq 90°$,$\vec{B}$ は $-1.0 \leq L_B \leq 0$ かつ $-90° \leq \theta_A \leq 90°$ の範囲内で座標点を移動する．$L\cos\theta$ は快-不快の軸，$L\sin\theta$ は覚醒-眠気の軸に対応し，以下のように感情ベクトルが変動する [2]．

if(学習者が問題に正解)
　$L_A \leftarrow L_A + 0.2$
　$L_B \leftarrow L_B - 0.2$
$else$
　$L_A \leftarrow L_A - 0.2$
　$L_B \leftarrow L_B + 0.2$
if(解答時間 $<$ 基準時間)
　if(学習者が問題に正解)
　　$\theta_A \leftarrow \theta_A + 15$

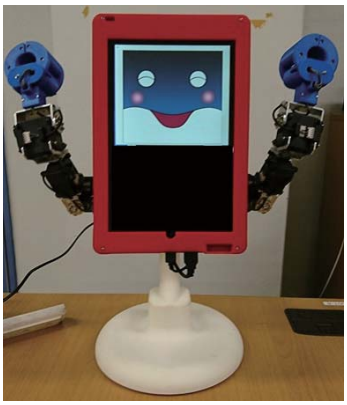
連絡先: 谷嵜悠平，名古屋大学大学院工学研究科，愛知県名古屋市千種区不老町工学部 3 号館 306 号室，052-789-2793，052-789-3166, tanizaki@cmplx.cse.nagoya-u.ac.jp



(a) 身体動作のみ



(b) 表情変化のみ



(c) 身体動作と表情変化の組み合わせ

図 3: 喜びを表出する動作例

else
$\theta_B \leftarrow \theta_B + 15$
else
if(学習者が問題に正解)
$\theta_A \leftarrow \theta_A - 15$
else
$\theta_B \leftarrow \theta_B - 15$

本実験では，基準時間を学習者の一問前の問題における解答時間とする．ただし，1 問目の問題の場合は，60 秒とする．

3. ロボット

3.1 ロボットの概要

本稿では，頭部がタブレットであるタブレット型ロボット「Tabot」を開発し実験に用いた（図 2）．Tabot は，頭部のタブレットにエージェントを表出することで，多様な表情変化を行うことができる．また，Tabot は，首の自由度が 3，両腕の自由度が 10，脚部の自由度が 1 であり，合計 14 の自由度を持つ．そのため，多様な身体動作を行うことができる．これらにより，Tabot は様々な表情変化と身体動作を組み合わせることで，多様な感情表出が可能である．本実験においては，ロボットの外観による作用をなくすために，Tabot を 3 台使用して実験を実施した．

3.2 ロボットの表情変化と身体動作

Tabot のタブレット上に表出するエージェントは，デザイナーによるデザインを基に作成した．またエージェントの表情変化も，デザイナーによるデザインを基に感情円環モデルに存在するすべての感情に対応した表情を 2 種類ずつ作成した．Tabot の身体動作は，人における身体動作と感情の関係について記された参考文献 [4][5] を基に，感情円環モデル上にあるすべての感情に対応した身体動作を 2 種類ずつ作成した．そして，同じ感情を表出する表情変化と身体動作を組み合わせた．各表情変化と身体動作及びその組み合わせは予備実験を通して学習者に感情を的確に伝達できることを確認した．Tabot の感情表出の例を図 3 に示す．

4. 実験

4.1 方法

学習者は SPI2 の非言語分野を学習するための SPI 学習システムを用いて，共感表出法が搭載されたロボットと共に学習する．学習システムは，SPI2 の非言語分野に関する問題を提示し，学習者がその問題を解答すると正誤判定を提示する．学習者は，学習システムで提示された問題を解きながら学習する．



図 4: 学習時の Tabot の様子

学習者は大学生 30 名である．30 名の学習者は，身体動作のみで共感を表出するロボットと共に学習する「身体動作群」，表情変化のみで共感を表出するロボットと共に学習する「表情変化群」，身体動作と表情変化を組み合わせて共感を表出するロボットと共に学習する「身体表情統合群」に 10 名ずつ振り分けて，1 か月間，1 回 40 分の学習を週に 2〜3 回程度行う．学習は合計 12 回行われる．1 回目と 12 回目の学習では，事前事後の知識を測るために，学習者は単独で学習システムを用いて学習する．2〜11 回目では，学習者はロボットと共同学習を行う．

学習時のロボットはタブレットの上側半分に表情，下側半分に学習システムを表示する．学習時のロボットの状態例を図 4 に示す．ロボットは，学習システムにおける正誤判定時のみに，共感表出法に基づいて感情を表出する．それ以外，ロボットは動作しない．

4.2 評価基準

各群の学習者における 1 回目の学習での正解数と，12 回目の学習での正解数の点数の差（向上点数）で学習結果を見た．また，ロボットの印象評価には，印象の定量的評価方法である SD 法を用いた（図 5）．SD 法の項目は「親しみやすい-親

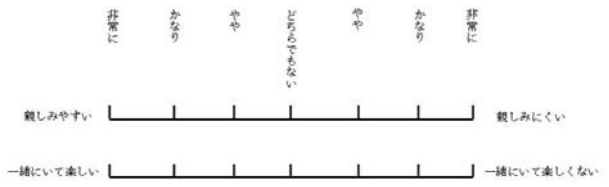


図 5: SD 法によるアンケート例

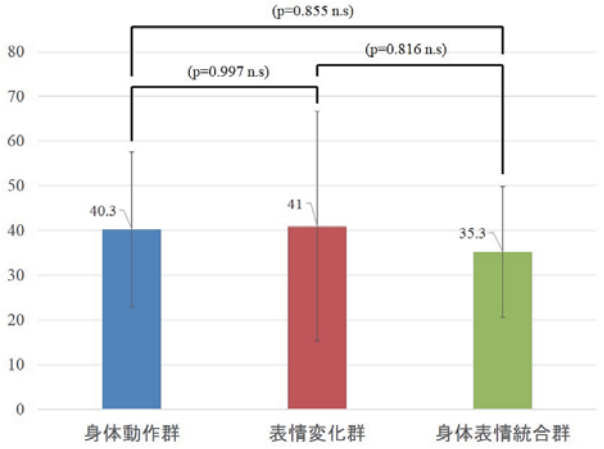


図 6: 各群の向上点数

しみにくい」,「感情的な-理性的な」,「ともに喜んでくれる-ともに喜んでくれない」,「勉強が楽しい-勉強が楽しくない」などの 14 種類の項目である．SD 法による項目は，図 5 の最左部にある「非常に」を 7 とし，右に 1 つ移動するにつれて −1 とする 1〜7 の評点で数値化した．また，学習者には，学習の感想を自由記述アンケートにおいて質問した．各群において，11 回目の学習終了時にアンケートを全被験者に対して行った．向上点数及び SD 法の各項目における評価値に対して，テューキーの方法による検定を実施した．有意水準 5 %の時に有意差を認めるものとし，ボンフェローニ法を用いて有意水準 (p:0.0033) に調整した．

5. 結果

5.1 向上点数

各群の向上点数の平均を図 6 に示す．図 6 から，向上点数は表情変化群が最も高く，次いで身体動作群，身体表情統合群であることがわかる．また，各群間の学習効果を比較するために，テューキーの方法にによる検定を実施した．その結果，各群間において向上点数には，有意差は認められなかった．

5.2 SD 法の結果

SD 法によるアンケートの結果を図 7 に示す．図 7 から，身体動作群では「思いやりのある」,「優しい」,「私の気持ちを分かってくれる」,「私と感情の変化が似ている」,「共に悲しんでくれる」,「勉強しやすい」,「ロボットに興味がわく」の 7 項目が，表情変化群では,「勉強が楽しい」の 1 項目が，身体表情統合群では,「親しみやすい」,「一緒にいて楽しい」,「感じがよい」,「感情的な」,「優しい」,「愛想がある」,「共に喜んでくれる」の 7 項目が最高点をとっている．このことから，身体動作群と身体表情統合群のロボットは，表情変化群のロボットと比べると，学習者に好印象を与えることが推察できる．各群間の

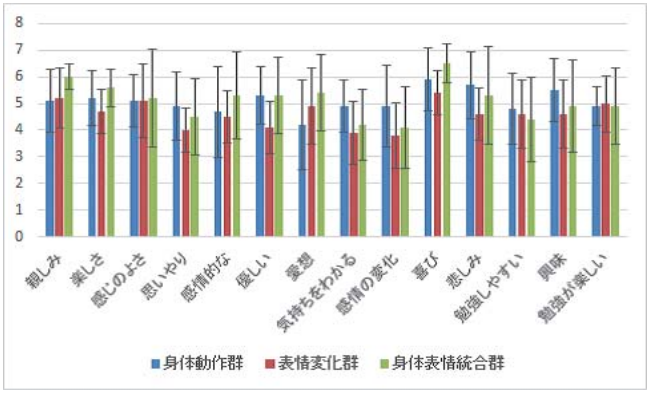


図 7: SD 法によるアンケート結果

SD 法によるアンケート結果を比較するために，テューキーの方法による検定を実施した．その結果，各群間において SD 法によるアンケート結果には，有意差は認められなかった．

6. 考察

実験結果から，学習効果は表情変化群が最も高く，次いで身体動作群，身体表情統合群の順であった．しかしながら，学習効果において有意差は認められなかった．このことから，身体動作のみ，表情変化のみ，またはその両方を用いて共感表出を行うロボットにおいて，学習者に与える学習効果に違いがないことが示唆された．この要因として，ロボットと学習者の直接的なインタラクションがなかったからだと考えられる．従来研究では，ロボットの身体動作はユーザとのインタラクションに有効であると報告されている．一方，本実験では，学習者は学習システムで表示される問題を解き，ロボットは正誤判定時のみに動作を実行する．そのため，ロボットが学習者の動きや言動に対して直接的に反応することがなく，学習者とロボットが直接的にインタラクションするような場面がなかった．これにより，各群の学習者は，ロボットの表情変化や身体動作に注目する場がなく，同様の効果を得たのではないかと考える．

しかしながら，ロボットの印象は身体表情統合群と身体動作群は，表情変化群に比べると好印象な傾向にあった．これは，従来研究で報告されているように，ロボットの身体動作が感情表出に有効に働いたからだと考える．そのため，ロボットが身体動作により感情表出することで，表情変化のみによる感情表出を行ったロボットよりも共感していると学習者が感じることがあると考える．

7. おわりに

本稿では，身体動作と表情変化を同時に行うことができるタブレット型ロボット Tabot に共感表出法を搭載し，身体動作によって共感表出を行うロボット，表情変化によって共感表出を行うロボット，身体動作と表情変化によって共感表出を行うロボットの内，どのロボットが学習者に有効であるのか比較検討した．実験結果から，学習効果においては，表情変化によって共感表出を行うロボットが高く，ロボットの印象においては，身体動作によって共感表出を行うロボットと，身体動作と表情変化によって共感表出を行うロボットが学習者に好印象を与える傾向があった．

今後は，ロボットの表情変化及び身体動作と，学習システムの連携を改善し，感情表出方法の違うロボットの差別化を行っ

ていく.

謝辞

本研究は，文部科学省科学研究費 (基盤研究 (B), No.16H02889) の補助を得て遂行された.

参考文献

[1] ジメネスフェリックス, 加納政芳, 吉川大弘, 古橋武: 建設的相互作用を基に行動するロボットとの協調学習の実現可能性, 人工知能学会論文誌, Vol. 31, No.3, pp.A-F93_1-10, (2016)

[2] ジメネスフェリックス, 吉川大弘, 古橋武, 加納政芳: 感情表出モデルを用いたロボットとの共同学習がもたらす影響, 知能と情報, Vol. 28, No. 4, pp. 700-704, (2016)

[3] 神田崇行, 今井倫太, 小野哲雄, 石黒浩:人-ロボット相互作用における身体動作の数値解析, 情報処理学会論文誌, Vol.44, No. 11, pp.2699-2709, (2003)

[4] 東山安子, ローラ・フォード: 日米ボディートーク増補新装版　身ぶり・表情・しぐさの辞典, 三省堂, (2016)

[5] デズモンド・モリス, 東山安子訳：ボディートーク新装版　世界の身ぶり辞典, 三省堂, (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ノンパラメトリックベイズ法に基づく時系列データの分節化

Segmenting Time Series Data Using GP-HSMM with Nonparametric Bayesian Model

長野匡隼 *1 Masatoshi Nagano　中村友昭 *1 Tomoaki Nakamura　長井隆行 *1 Takayuki Nagai　持橋大地 *2 Daichi Mochihashi　小林一郎 *3 Ichiro Kobayashi　金子正秀 *1 Masahide Kaneko

*1電気通信大学 The University of Electro-Communications　*2統計数理研究所 Institute of Statistical Mathematics　*3お茶の水女子大学 Ochanomizu University

In this paper, we propose a method for dividing continuous time-series data into segments in unsupervised manner. Humans recognize perceived continuous information by dividing it into significant segments such as words and unit motions. To this end, we have been proposed a method based on hidden semi-Markov model with Gaussian process (GP-HSMM). However, it has a big drawback that it requires the number of classes into which time-series data is segmented. To overcome this problem, in this paper, we extend GP-HSMM to nonparametric Bayesian model by introducing hierarchical Dirichlet processes (HDP), and propose hierarchical Dirichlet processes-Gaussian process-hidden semi-Markov model (HDP-GP-HSMM). Hence, the infinite number of classes is assumed and the number of classes are estimated by applying slice sampling. In the experiment, we used the various time-series data and showed that our proposed model can estimate more correct number of classes and achieve more accurate segmentation than baseline methods.

1. はじめに

人は知覚した連続的な情報を意味をもつ単位に分節化することで認知している．分節化を行うことで，音声波形から意味を持つ単語を認識することができ，人の動作から単位動作を認識することができる．人は，分節点や単位動作の組み合わせに無数の可能性がある中から，明示的な分節点を与えられることなく自ら学習し，情報を適切に分節化している．ロボットにおいても，言語や動作等を柔軟に学習するために，教師なしで分節化する能力は非常に重要であると言える．

そこで，我々は隠れセミマルコフモデルにおける出力をガウス過程とした Gaussian Process - Hidden Semi-Markov model (GP-HSMM)[Nakamura 17] を用いることで，時系列データの分節化を可能とした．しかし，GP-HMM では，事前に分類する単位動作のクラス数を与える必要があった．本稿では，この問題を解決するため GP-HSMM に Hierarchical Dirichlet Process (HDP)[Teh 06] を導入し，ノンパラメトリックベイズモデルへと拡張した HDP-GP-HSMM を提案する．ノンパラメトリックベイズモデルでは，無限個のクラスがあることを仮定しクラスタリングを行う．実際には学習データは有限であり，そのうちの有限個のクラスのみが学習データを表現するのに利用されることで，クラス数の推定を可能としている．すなわち，HDP-GP-HSMM のパラメータを学習することで，教師なしで分節点と各単位系列のクラスを推定するだけでなく，クラス数も決定することができる．しかし，無限個のクラスを仮定しているため，GP-HSMM の学習に用いられている Forward filtering-Backward sampling[Uchiumi 15] をそのまま用いることはできない．そのために，文献 [Gael 08] で提案された Slice-sampling を用いることにより，無限個のクラスを確率的に有限個にすることで学習を可能とする．

このような教師無しで動作を分節化する研究はいくつか行われているが，ヒューリスティックな仮定をおいているものが多い．しかしながら，ヒューリスティックな仮定が全ての動作に対して有効に働くかは定かではない．また単位動作の表現に Hidden Markov Model (HMM) を用いることで，教師なしで分節化する手法が提案されている [Beal 01, Taniguchi 11, Fox 11, Matsubara 14]．HMM は，構造がシンプルであり，複雑な動作を表現することが難しいと考えられる．一方，我々は Gaussian Process (GP) により動作を表現する．GP はノンパラメトリックな手法であり，複雑な動作を表現するのに適していると考えている．

連絡先: 長野匡隼，電気通信大学，東京都調布市調布ヶ丘 1-5-1, m_nagano@radish.ee.uec.ac.jp

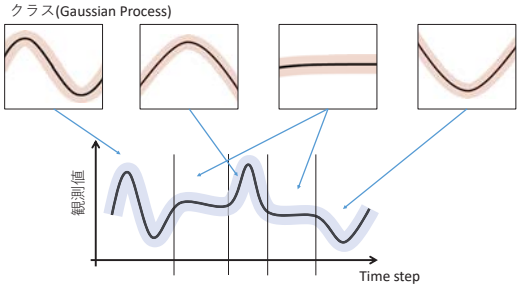


図 1: ガウス過程による単位系列表現

本稿では，人工データとモーションキャプチャデータを用いた実験により，他の教師なしの分節化手法よりもクラス数推定及び分節化精度が高いことを示す．

2. ノンパラメトリックベイズ

本稿では，ノンパラメトリックベイズモデルを用いることで単位動作のクラス数を推定する．パラメトリックな混合モデルではクラス数は有限であり，有限次元のディリクレ分布から多項分布のパラメータが生成され，この多項分布から隠れ状態が決定される．一方，ノンパラメトリックベイズモデルでは，無限次元の多項分布を用いることで無限個のクラスを仮定する．しかし，実際にはその中の有限個のクラスのみが使用されるため，結果としてクラス数の推定が可能となる．無限次元の多項分布の構成法の 1 つが Stick Breaking Process (SBP) である．

2.1 Stick Breaking Process

Stick Breaking Process (SBP)[Gael 08] は，無限次元の多項分布を生成可能なディリクレ過程である．SBP では，図 2 のように長さ 1 の棒をベータ分布からサンプリングされる値 v_k により，折り続けることで無限次元の多項分布 $\boldsymbol{\beta}$ を構成している．

$$v_k \sim \mathrm{Beta}(1,\gamma) \qquad (k=1,\cdots,\infty) \tag{1}$$

$$\beta_k = v_k \prod_{i=1}^{k-1}(1-v_k) \qquad (k=1,\cdots,\infty) \tag{2}$$

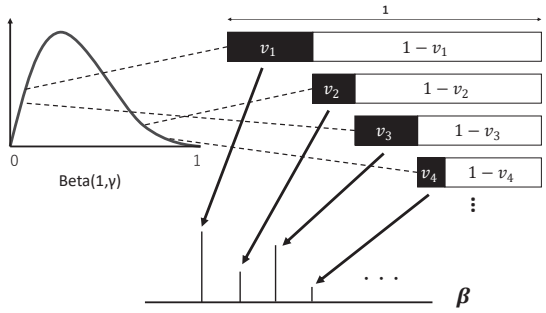


図 2: Stick Breaking Process

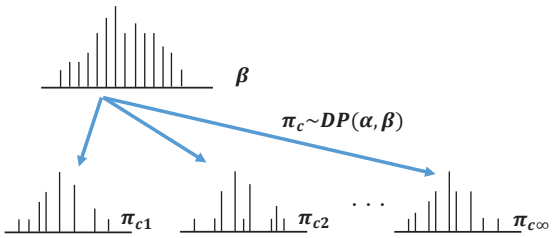


図 3: Hierarchical Dirichlet Process

このプロセスは一般的に $\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma)$ と表記される [Pitman 02].

2.2 Hierarchical Dirichlet Process

ディリクレ過程を用いることで，無限の状態数をもった HMM の遷移確率を構成することができる．しかし，単に SBP によって各状態の遷移確率を構成するだけでは，状態毎に可能な遷移先が異なることになる．HMM のようなモデルの場合，全ての状態において可能な遷移先が共有されており，その遷移確率が状態毎に異なる分布を構成しなければならない．そのために，図 3 で示すように，SBP によって生成した分布 $\boldsymbol{\beta}$ を基底測度とし共有し，状態 c 毎に異なる遷移確率 $\boldsymbol{\pi}_c$ を DP によって生成する．

$$\boldsymbol{\pi}_c \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}) \tag{3}$$

このような，二段階の DP によって無限次元の確率分布を構成する手法は，Hierarchical Dirichlet Process (HDP) と呼ばれている．

3. HDP-GP-HSMM

図 4 が提案手法のグラフィカルモデルである．$c_j (j = 1, 2, \cdots, J)$ が単位系列のクラスを表しており，無限個のクラス数を仮定している．$\boldsymbol{\pi}_c$ は，クラス c から各クラスへの遷移確率を表しており，γ をパラメータとする GEM 分布（Stick Breaking Process）によって生成された $\boldsymbol{\beta}$ と，α をパラメータとする Dirichlet Process (DP) から生成される．

$$\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma) \tag{4}$$
$$\boldsymbol{\pi}_c \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}) \tag{5}$$

j 番目のクラス c_j は，$j-1$ 番目のクラス c_{j-1} と遷移確率 $\boldsymbol{\pi}_c$ によって決定される．ここでは，クラスと対応した $\boldsymbol{X}_c$ をパラメータとするガウス過程から単位系列 $\boldsymbol{x}_j$ が生成されることを仮定している．

$$c_j \sim P(c|c_{j-1}, \boldsymbol{\beta}, \alpha) \tag{6}$$
$$\boldsymbol{x}_j \sim \mathcal{GP}(\boldsymbol{x}|\boldsymbol{X}_{c_j}) \tag{7}$$

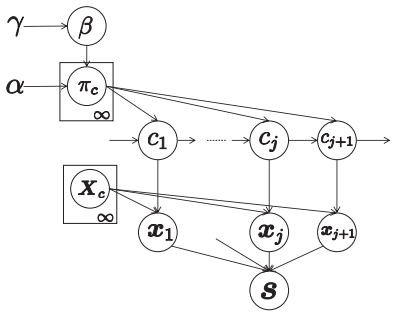


図 4: HDP-GP-HSMM のグラフィカルモデル

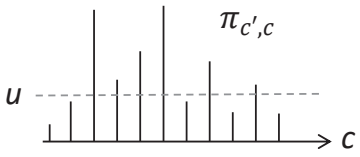


図 5: Slice sampling

ただし，$\boldsymbol{X}_c$ はクラス c に分類された単位系列の集合である．これらの単位系列を結合することで，観測系列 $\boldsymbol{S} = \boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_J$ となる．

4. 学習

提案モデルでは，観測系列内の単位系列とそのクラスをサンプリングすることで推定する．効率的に計算するために，1 つの観測系列内の単位系列とそのクラスを一括でサンプリングする Blocked Gibbs Sampler を用いる．まず，全ての観測系列をランダムに分節・分類することで初期化する．次に，ある観測系列 $\boldsymbol{s}_n$ を分節して得られた単位系列 $\boldsymbol{x}_{nj} (j = 1, 2, \cdots, J_n)$ を，それらが分類されているクラスからすべて削除し，各クラスのガウス過程のパラメータ $\boldsymbol{X}_c$ と遷移確率 $P(c|c')$ を更新する．次に，単位動作の長さとクラスのあらゆる組み合わせを効率的に計算可能な Forward filtering-Backward sampling を用い，単位系列 $\boldsymbol{x}_{nj} (j = 1, 2, \cdots, J_n)$ とそれらのクラス $c_{nj} (j = 1, 2, \cdots, J_n)$ を推定し，各クラスと対応したガウス過程のパラメータ $\boldsymbol{X}_c$ と遷移確率 $P(c|c')$ を更新する．以上の操作を全ての系列に対して繰り返すことで，観測系列の分節点とガウス過程のパラメータを交互に最適化することができる．しかし，このままでは生成された遷移確率 $\boldsymbol{\pi}_c$ の遷移先のクラス数は無限個存在するため，単位動作が各クラスに分類される確率を計算することができない．本稿では，この問題を解決するために，Slice Sampling を用いる．Slice Sampling では，パラメータ $\boldsymbol{u}$ を閾値として遷移先を確率的に切り捨てることで，遷移確率 $\boldsymbol{\pi}_c$ の遷移先のクラス数を有限にする．パラメータ $\boldsymbol{u}$ は，タイムステップ j に対して，遷移確率 $\boldsymbol{\pi}_c$ から次式のように計算する．$\pi_{c',c}$ はクラス c' から c への遷移確率である．

$$p(u_j|c_{j-1}, c_j, \boldsymbol{\pi}_c) = \frac{\delta(0 < u_j < \pi_{c_{j-1},c_j})}{\pi_{c_{j-1},c_j}} \tag{8}$$

図 5 のようにパラメータ $\boldsymbol{u}$ を用いて，$\pi_{c',c} \geq u_j$ となるクラスのみに遷移可能とすることで，使用されるクラスを有限個にする．その結果，foward filtering-backward sampling を利用することが可能となる．

5. 実験

提案手法の有効性を検証するために，複数のデータに対して分節化を行い，HDP-HMM[Beal 01]，HDP-

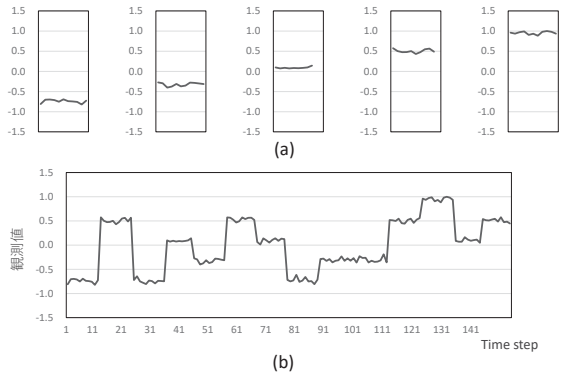


図 6: 人工データ 1: (a) 5 つの単位動作, (b) 観測系列.

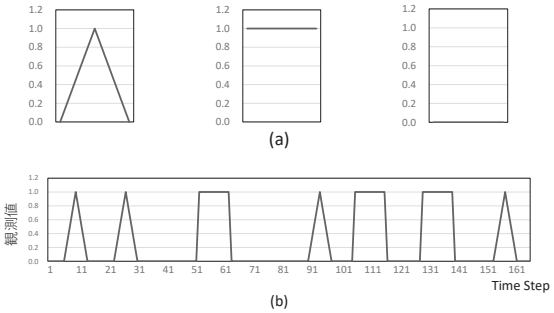


図 7: 人工データ 2: (a) 5 つの単位動作, (b) 観測系列.

HMM+NPYLM[Taniguchi 11], BP-HMM[Fox 11], Autoplait[Matsubara 14] と比較した．評価方法は文献 [Nakamura 17] と同様に，正規化ハミング距離及び適合率，再現率，F 値を用いた．分節点の誤差の許容範囲は観測系列の全フレーム数の ±5% とした．

5.1 人工データ

以下の二つの人工データに対して分節化を行った．

- **人工データ 1:** まず，5 つの値からランダムに選択し，一様乱数により系列長をランダムに選択する．

$$k \sim \text{uniform}(9, 13). \tag{9}$$

さらに，正規乱数によりノイズを加える．

$$x \sim x + \text{normal}(x, 0.05). \tag{10}$$

図 6 のように，これらの単位系列をランダムに結合して観測系列を作成した．

- **人工データ 2:** 図 7(a) の単位系列をランダムに結合することで，図 7(b) に示す観測系列を作成した．

それぞれ分節化の結果を表 1 と表 2 にまとめた．表 1 と表 2 より，提案手法を用いて分節化を行った結果，ハミング距離は 0.1 未満かつ適合率，再現率，F 値は 1.0 となった．この結果より，HDP-GP-HSMM では正しく分節化がされていることが分かる．さらに，クラス数も人工データ 1 では 5，人工データ 2 では 3 と正しく推定された．

また，人工データ 1 は HDP-HMM においても正しく分節化されていることが分かる．これは，人工データ 1 が，ガウス分布に従う単純なデータであったためだと考えられる．そのため，人工データ 2 はガウス分布に従わないデータであるために正しく分節化ができなかったと考えられる．一方，人工データ 2 のような複雑なデータは，HDP-GP-HSMM や BP-HMM のような複雑なモデルでは正しく分節化されていることが分かる．

表 1: 人工データ 1 の分節化結果

	ハミング距離	適合率	再現率	F 値	推定されたクラス数
HDP-GP-HSMM	0.016	1.0	1.0	1.0	5
HDP-HMM	0.011	1.0	1.0	1.0	5
HDP-HMM+NPYLM	0.48	0.53	1.0	0.69	35
BP-HMM	0.36	1.0	0.94	0.97	4
Autoplait	0.79	0	0	0	3

表 2: 人工データ 2 の分節化結果

	ハミング距離	適合率	再現率	F 値	推定されたクラス数
HDP-GP-HSMM	0.037	1.0	1.0	1.0	3
HDP-HMM	0.13	0.43	1.0	0.60	8
HDP-HMM+NPYLM	0.45	0.57	1.0	0.72	19
BP-HMM	0.038	1.0	1.0	1.0	3
Autoplait	0.35	0.17	0.15	0.16	1

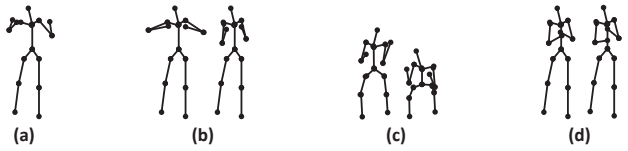


図 8: チキンダンスに含まれる単位動作: (a) beaks, (b) wings, (c) tail feathers, and (d) claps.

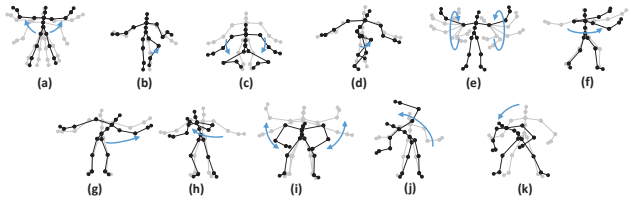


図 9: 体操に含まれる単位動作: (a) jumping jack, (b) jogging, (c) squatting, (d) knee raise, (e) arm circle, (f) twist, (g) side reach, (h) boxing, (i) arm wave, (j) side bend, and (k) toe touch.

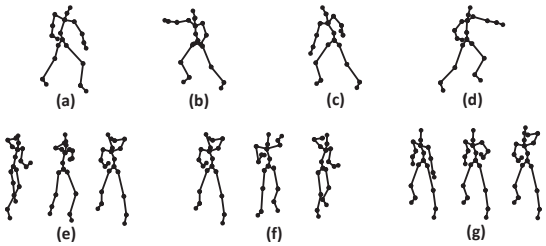


図 10: 空手の型に含まれる単位動作: (a) 下段払い (左), (b) 追い突き (右), (c) 下段払い (右), (d) 追い突き (左), (e) 上げ受け (左), (f) 上げ受け (右), (g) 添え手.

5.2 モーションキャプチャデータ

以下の三つのモーションキャプチャデータに対して分節化を行った．

- **チキンダンス:** CMU のモーションキャプチャデータに含まれるチキンダンスのモーションキャプチャデータを使用した [*1]．図 8 のようにチキンダンスには 4 個の単位動作が含まれている．分節化には，文献 [Matsubara 14] で用いられている両手両足の加速度を用いた．

*1 http://mocap.cs.cmu.edu/

表 3: チキンダンスの分節化結果

	ハミング距離	適合率	再現率	F 値	推定されたクラス数
HDP-GP-HSMM	0.038	1.0	1.0	1.0	4
HDP-HMM	0.46	0.39	0.71	0.50	4
HDP-HMM+NPYLM	0.65	0.63	0.71	0.67	8
BP-HMM	0.16	0.70	1.0	0.82	4
Autoplait	0.026	1.0	1.0	1.0	4

表 4: 体操の分節化結果

	ハミング距離	適合率	再現率	F 値	推定されたクラス数
HDP-GP-HSMM	0.29	0.44	0.93	0.60	10
HDP-HMM	0.80	0.02	0.89	0.04	17
HDP-HMM+NPYLM	0.64	0.28	1.0	0.44	23
BP-HMM	0.23	0.33	0.95	0.50	13
Autoplait	0.89	0.67	0.11	020	3

- **体操:** CMU のモーションキャプチャデータの subject14 から体操のモーションキャプチャデータを用いた．図 9 のように体操のデータには 11 個の単位動作が含まれている．分節化には，計測した動作の両手両足の奥行きを除いた左右 (x) と上下 (y) の二次元の座標を用いた．

- **空手の型:** より複雑な動作として，空手のモーションキャプチャデータを用いた [*2]．図 10 のように空手の型には 7 個の単位動作が含まれている．分節化には，計測した動作の両手の奥行きを除いた左右 (x) と上下 (y) の二次元の座標を用いた．

これらのデータに対する前処理としてダウンサンプリングと正規化を行った．HDP-GP-HSMM では，毎秒 4 フレームにダウンサンプリングを行った．また，min-max 法を用いて最小値を -1，最大値を 1 にするように正規化を行った．一方，他の分節化の手法においては，毎秒 8 フレーム，4 フレーム，2 フレームでダウンサンプリングを行い，min-max 法と z-score を用いて正規化を行い精度の良いものを選択した．z-score は，平均値及び標準偏差を 0 及び 1 とする正規化である．

表 3，4，5 は各データの分節化の結果を表す．これらの表より，提案手法では正しく分節化がされていることが分かる．また，チキンダンスのデータは BP-HMM と autoplait でも高精度で正しく分節化出来た．これは，チキンダンスのデータが単位動作を繰り返すだけの最も単純なデータであるからだと考えられる．

さらに，提案手法ではチキンダンスと空手の型のデータのクラス数を正しく推定できた．一方，体操のデータにおいて提案手法で分節化を行った結果，クラス数は正解のクラス数 11 個より少ない 10 個と推定された．これは，図 9 の動作 F と動作 H が似ている動作であるので，同じクラスに分類されたためだと考えられる．しかし，他の手法に比べ提案手法が最も正解のクラス数に近いクラス数を推定できたことが分かる．

以上の分節化及びクラス数推定の結果から，HDP-GP-HSMM が様々なデータに対して有効であると考えられる．

6. まとめ

本稿では，ガウス過程を出力分布とした隠れセミマルコフモデルにディリクレ過程を導入することで，動作のクラス数を推定しつつ動作を分節・分類する手法を提案した．提案手法では，動作の切れ目と，動作のクラスを Foward filtering-Backward sampling により同時に推定し，クラス数をノンパラメトリックベイズモデルにより推定することで動作の分節化を行った．実験により，動作のクラス数の推定をしつつ，分節・分類が可能であることが示された．また，様々な連続的な時系列データに関しても分節化が可能であることが示された．

しかし，ガウス過程を用いているため，他の手法と比較すると計算時間が長いといった問題がある．この問題のため，HDP-GP-HSMM は大規模データに用いることができない．計算時間の削減については，文献 [Nguyen-Tuong 09, Okadome 14] で提案されているガウス過程の近似的手法を用いることができると考えている．さらに，今回は特徴量を両手両足の xy 座標のみ用いたが，今後全身の関節位置を用いた，より複雑な動作を扱うことも考えている．

*2 http://www.mocapdata.com/

表 5: 空手の型の分節化結果

	ハミング距離	適合率	再現率	F 値	推定されたクラス数
HDP-GP-HSMM	0.21	0.70	0.90	0.80	7
HDP-HMM	0.43	0.47	0.71	0.57	9
HDP-HMM+NPYLM	0.57	0.28	0.67	0.40	20
BP-HMM	0.32	0.40	1.0	0.57	12
Autoplait	0.60	0	0	0	3

参考文献

[Nakamura 17] T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh, and M. Kaneko, “Segmenting Continuous Motions with Hidden Semi-Markov Models and Gaussian Processes,” Frontiers in Neurorobotics, vol.11, no.67, 2017.

[Teh 06] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, “Hierarchical Dirichlet processes,” Journal of the American Statistical Association, vol.101, no.476, pp.1566–1581, 2006.

[Uchiumi 15] K. Uchiumi, T. Hiroshi, and D. Mochihashi, “Inducing Word and Part-of-Speech with Pitman-Yor Hidden Semi-Markov Models,” Joint Conference of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp.1774–1782, 2015.

[Gael 08] J. V. Gael, Y. Saatci, Y. W. Teh, and Z. Ghahremani, “Beam Sampling for the Infinite Hidden Markov Model,” In Proceedings of the 25th International Conference on Machine Learning, pp.1088–1095, 2008.

[Beal 01] M. J. Beal,, Z. Ghahramani, and C. E. Rasmussen, “The infinite hidden Markov model,” In Advances in Neural Information Processing Systems, pp.577–584, 2001.

[Taniguchi 11] Taniguchi. T, and Nagasaka. S, “Double articulation analyzer for unsegmented human motion using Pitman–Yor language model and infinite hidden Markov model, ” In IEEE/SICE International Symposium on System Integration, pp.250–255, 2011.

[Fox 11] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, “Joint modeling of multiple related time series via the beta process,” arXiv preprint arXiv:1111.4226, 2011.

[Matsubara 14] Y. Matsubara, Y. Sakurai, and C. Faloutsos, “Autoplait: Automatic mining of co-evolving time sequences,” ACM SIGMOD International Conference on Management of Data, pp.193–204, 2014.

[Pitman 02] J. Pitman, “Poisson-Dirichlet and GEM invariant distributions for splitand-merge transformations of an interval partition,” Combinatorics, Probability and Computing, vol.11, pp.501–514, 2002. MR1930355

[Nguyen-Tuong 09] D. Nguyen-Tuong, J. R. Peters, and M. Seeger, “Local Gaussian process regression for real time online model learning and control,” In Advances in Neural Information Processing Systems, pp.1193–1200, 2009.

[Okadome 14] Y. Okadome, K. Urai, Y. Nakamura, T. Yomo, and H. Ishiguro, “Adaptive lsh based on the particle swarm method with the attractor selection model for fast approximation of Gaussian process regression,” Artificial Life and Robotics, vol.19, pp.220–226, 2014.

Oral presentation | Organized Session | [Organized Session] OS-10

[2G2-OS-10b] [Organized Session] OS-10

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room G (5F Ruby Hall Hiten)

[2G2-OS-10b-01] Establishment of Core Vocabulary based on Knowledge Graph for Interoperability between Agricultural Data

○Sungmin Joo[1], Hideaki Takeda[1,2], Akane Takezaki[3], Tomokazu Yoshida[3] （1. National Institute of Informatics, 2. The Graduate University for Advanced Studies, 3. National Agriculture and Food Research Organization）

1:20 PM - 1:40 PM

[2G2-OS-10b-02] Fruits Growth Information Extracting Application by using Deep Learning Network

○Kazuki Kobayashi[1], Toshiki Nakamura[1] （1. Shinshu University）

1:40 PM - 2:00 PM

[2G2-OS-10b-03] Development of Crop Vocabulary to facilitate interoperability among agricultural IT systems

○Akane Takezaki[1], Sungmin Joo[2], Hideaki Takeda[2], Tomokazu Yoshida[1] （1. National Agriculture and Food Research Organization, 2. National Institute of Informatics）

2:00 PM - 2:20 PM

[2G2-OS-10b-04] Agricultural Field Monitoring for Multi-Lens Camera TOMBO

○Fumiya Shimobayashi[1], Kazuki Kobayashi[1], Keiichiro Kagawa[2], Wei Guo[3], Masayuki Hirafuji[3], Jun Tanida[4] （1. Shinshu Univ., 2. Shizuoka Univ., 3. Univ. of Tokyo, 4. Osaka Univ.）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

農業データ連携のためのナレッジグラフに基づく標準語彙の運用

Establishment of Core Vocabulary based on Knowledge Graph for Interoperability between Agricultural Data

朱 成敏 *1 Sungmin JOO　武田 英明 *1*2 Hideaki TAKEDA　竹崎 あかね *3 Akane TAKEZAKI　吉田 智一 *3 Tomokazu YOSHIDA

*1国立情報学研究所 National Institute of Informatics

*2総合研究大学院大学 SOKENDAI University(The Graduate University for Advanced Studies)

*3農業・食品産業技術総合研究機構 National Agriculture and Food Research Organization

Agricultural ICT system has been accompanied by generating data such as farming information or agricultural observations. These data is expected to be used for statistics, analysis and prediction for area agriculture. However, vocabulary in the agriculture domain is treated with no standard information, data in agricultural ICT systems is not easy to integrate and maintenance. In this paper, we investigate the existing vocabulary system for agriculture domain. Then we propose the core vocabulary system based on knowledge graph by paying attention to interoperability between agriculture data.

1. はじめに

近年，日本の農業分野は ICT 基盤技術の普及により大きな進歩を遂げている．センサーネットワークによるモニタリングや営農管理システムを用いた圃場管理，栽培の自動化など生産性を従来より高めている．そして，農業 ICT システムから様々なデータが発生されており，そのデータを分析することによって作業の最適化や収穫量の予測など新しい発見が可能となる．それぞれの農業 ICT システムから発生したデータを連携・統合し，膨大なデータとして活用すれば今までより正確で多様な情報を予測することも可能となる．これまで人間の経験と勘で行われてきた意思判断を蓄積されたデータに基づいて行い，より精度の高い意思決定に導く，いわゆる知識基盤の農業が可能となる．

一方，これまでの農業 ICT システムはデータ項目の作成において基準となる情報がないためシステムベンダーは独自の定義を行い，データ項目の設計を行った．現在はそれぞれの農業 ICT システムは独自のデータ項目を持ち，その項目に従ってデータを作成しているため，異なるシステムから発生したデータの場合は連携・統合が困難となった．こういった背景から内閣官房 IT 総合戦略本部は農業分野における情報創成・流通の促進のために農業情報の相互運用性等の確保に資する標準化や情報の取扱いに関する政府横断的な戦略を策定した [内閣官房 17b]．しかし，現状は公的機関による統計調査の項目や食品安全に関する語彙が浸透しているため，農業 ICT システムと実際の現場の語彙管理に差が生じている．こういった語彙間の関連性について対応関係が定義されていないため栽培から収穫，加工，流通までのフードチェーン全般において一貫性のある管理が困難となっている．また，農作業や農作物，資機材のように農業分野におけるそれぞれの語彙間の連携も必要である．

そこで，本稿では農業データの連携・統合のためにナレッジグラフを用いた標準語彙の管理を提案する．ナレッジグラフによる標準語彙の管理は用語の多様性が多く見られ，またそれが一貫性のある語彙管理の妨げとなる農業分野の語彙が持つ特徴を補完することができる．また，農業 ICT システムの運用において機械可読性を向上するデータ形式を提供することができ，農業 ICT システム間の円滑な連携を可能とする．

連絡先: 朱成敏，国立情報学研究所，〒 101-8430 東京都千代田区一ッ橋 2-1-2, joo@nii.ac.jp

2. 背景

本章では農業分野における語彙について概観し，その現状と農業 ICT システムのデータ連携における語彙に要求される要素を確認する．

2.1 農業分野における語彙の多様性

農業は食料を確保するために人類が営んできた長い歴史の産業であり，農家や地域単位の集団によって代々受け継がれた栽培方法で行われてきた．そのため，農家や地域独自の語彙が発生し，農業分野には同じ概念に対して複数の名称が存在する場合が多く見られる．そして，農家は農業 ICT システムに独自の名称を記録し，農業データには独自の語彙が多く含まれることになった．すなわち異なる農家の ICT システムから発生した農業データにはそれぞれ農家独自の特徴が反映されており，データの連携や統合が困難となっている．現状ではデータ項目の対応関係を調べ，人手によってその関係を再定義する必要がある．この場合，データの量に比例して対応関係を人が確認しなければならないので，効率的なデータの活用が困難である．

2.2 既存の語彙

前節で述べた農業データの連携における問題を認識し，内閣官房 IT 総合戦略本部は「農業情報の標準化に関する個別ガイドライン」を提示し，農業 ICT システムのデータ標準化を推進している．農作業名称のガイドラインとして「農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第 3 版)」を，農作物名称に関しては「農業 IT システムで用いる農作物の名称に関する個別ガイドライン (第 2 版)」を提示している．農作業名称の項目には農作業の分類単位である「大項目」と「中項目」,「シソーラス」があり，それぞれの農作業と 11 種類に分類された農作物 (群) との関連も表示されている．農作物名称の場合は，分類単位である「大分類」,「中分類」,「小分類」,「シソーラス」，詳細を記載する「品種等」,「収穫部位」,「属性項目」がある．それぞれの項目には「その他」があ

り，項目の選択が困難な場合を想定して柔軟に対応している．

一方，上記の語彙以外にも公的機関による従来の語彙があり，農業現場で用いられている．まず，農作業の場合は統計調査における調査項目が挙げられる．国や自治体単位で収穫量や生産費など経営実態を調査しており，農家の経営状況を把握している．データの集計は農家が作成した作業履歴を収集し，担当部署が統計調査の項目と対応させて集計を行う．そして，農作物の場合は食品安全の確保のために管理管轄の公的機関による語彙があり，３つの語彙が中心となっている．まず，農林水産省が農薬の登録申請に適用対象となる作物名を提示するために発行している「農薬登録における適用作物名」がある．「農薬登録における適用作物名」は農作物を分類する単位である「大グループ名」と「中グループ名」,「作物名」と作物名の関連語彙である「作物名に含まれる別名，地方名，品種名等の例」，そして適用される農作物の部位を定義する「備考」がある．次に残留する農薬を管理するために厚生労働省が発行している「農産物等の食品分類表」がある．収録項目として「食品名」と「食品分類」の二つがある．最後に健康，栄養管理のために食品の成分に関する資料である「日本食品標準成 分表」が挙げられる．文部科学省が発行しており，収録されている項目は「食品番号」,「食品名」,「生物種の英名」,「学名」の４つがある．農作物に関する３つの語彙はそれぞれ異なる目的によって発行されたものであり，その目的によっては農作物名は農作物の名称そのものではなく収穫部位を意味する場合もある．

2.3 考察

語彙と語彙を連携するためには基準となる情報が必要がある.「農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第 3 版)」と「農業 IT システムで用いる農作物の名称に関する個別ガイドライン（第２版)」に共通の情報があれば，それを用いて連携することが可能となるが，農作業の対象となる農作物名が農作物の分類と異なり，現状では農作業と農作物のデータ間の連携が困難である．例えば,「農作物別農作業名」に分類されている 11 分類の中で「米」,「麦類」,「豆類」,「いも類」は「農業 IT システムで用いる農作物の名称に関する個別ガイドライン（第２版)」には分類として収録されてないためこれらの語彙を連携し，農作物に関するデータと農作業に関するデータを連携するためにはその詳細を対応させる必要がある．農作物に関する３つの既存語彙の農作物名称についても同じ問題を抱えており，農業 ICT システム間のデータと連携のためには異なる分野の語彙間の連携も考慮しなければならない．

特に農作物の語彙についてもう一つ考慮しなければならない事項がある．公的機関による３つの重要語彙はそれぞれ農作物，食品，食品成分としての作物を定義した語彙である．これらの語彙は栽培から加工，流通まで至るフードチェーンにおける食品安全の管理に基準なる情報である．近年食品安全の確保に対する要望が高まっており，ICT システムによる一貫性のある管理が求められている．そのためには語彙間の連携が必要であるが，対応関係を把握し，整理を行う必要がある．例として「じゃがいも (ばれいしょ)」が挙げられる．一般的にはじゃがいもと呼ばれるが「農薬登録における適用作物名」では「ばれいしょ」が，他で語彙では「じゃがいも」が代表表記になっている．また,「農薬登録における適用作物名」と「農作物等の食品分類表」では「ばれいしょ」と「じゃがいも」の関連が定義されているが,「日本食品標準成分表」の「じゃがいも」には「ばれいしょ」の情報が収録されていない．

以上の考察より農業 ICT システムのデータ連携のために基準となる標準語彙に求められる条件として以下の２つが考えられる．

- 明確な構造
- 語彙間の連携

本研究ではこの２つの条件を満たすためにナレッジグラフに注目した．ナレッジグラフを基盤とする語彙システムは語彙の効率的な管理と柔軟な適用，そして ICT システムにおけるデータの基準としての利活用が期待される．

3. ナレッジグラフ基盤の標準語彙

本章では農作業と農作物，そしてこれらを連携するために必要な語彙の設計について述べる．

3.1 農作業基本オンロトジー

筆者らは農作業の標準語彙として農作業基本オントロジー (AAO, Agriculture Activity Ontology) を開発し，推進してきた [朱 16]．農作業基本オントロジーは農作業の概念を定義するために 10 項目の属性 (目的，行為，対象，副対象，場所，手段，機資材，作物，時期，作業条件) を定義し，その属性が持つ値の包含関係によって農作業概念の階層構造を構築した知識体系である．また，複数の目的を持つ多義的概念と別名を収録し，農作業名称が持つ意味の多様性にも対応できる．オントロジーは記述論理に基づいて設計されており，階層構造における明確な分類と上位概念と下位概念間の論理的な矛盾がないように構築された．機械による推論も可能であり，新しい概念を追加する際に階層構造の中で最も適切な位置を提示することができ，語彙管理の手間も軽減できる．機械による推論は農業基本オントロジーが持つ論理に基づいて処理を行うため，今後 AI 農業の基盤技術として期待される．

農作業基本オントロジーは現在，共通農業語彙 (CAVOC, Common Agricultural Vocabulary)[*1] にて公開されており，農作業名称の URI と機械可読性の高い形式である RDF/Turtle と CSV 形式で公開されている．そして，これらのデータを円滑に利活用ができるよう関連 API も公開しており，誰もが自由に使うことができる．最新版である ver2.01 には 438 語が収録されており,「農業 IT システムで用いる農作業の名称に関する個別ガイドライン」に収録されている農作業名称と農林水産省の農産物生産費統計調査 [農林水産省 16] における調査項目の農作業名とも交換性を持つ．また，Wikipedia 日本語版 [*2] の項目と国際機関の FAO(Food and Agriculture Organization) による AGROVOC[*3] の項目との対応関係を調査し，収録したため，知識体系と国際標準との交換性を持つ．

3.2 農作物語彙体系

農作物の標準語彙として筆者らは農作物体系 (CVO, Crop Vocabulary) を構築した [竹崎 17]．農作物語彙体系は農作物の生物学特徴，収穫部位と栽培形態，品種としての特徴に基づいて農作物を定義している．公的機関による３つの重要語彙,「農薬登録における適用作物名」,「農産物等の食品分類表」,「日本食品標準成分表」の作物名 (食品名) を中心に構築し，それぞれの語彙と相互交換性を持つ．そして，品種情報を収録するために独自で収集し，整理した品種情報も収録されており，最新版である ver1.12 では 1,191 語が収録されている．

∗1 http://cavoc.org
∗2 https://ja.wikipedia.org
∗3 http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

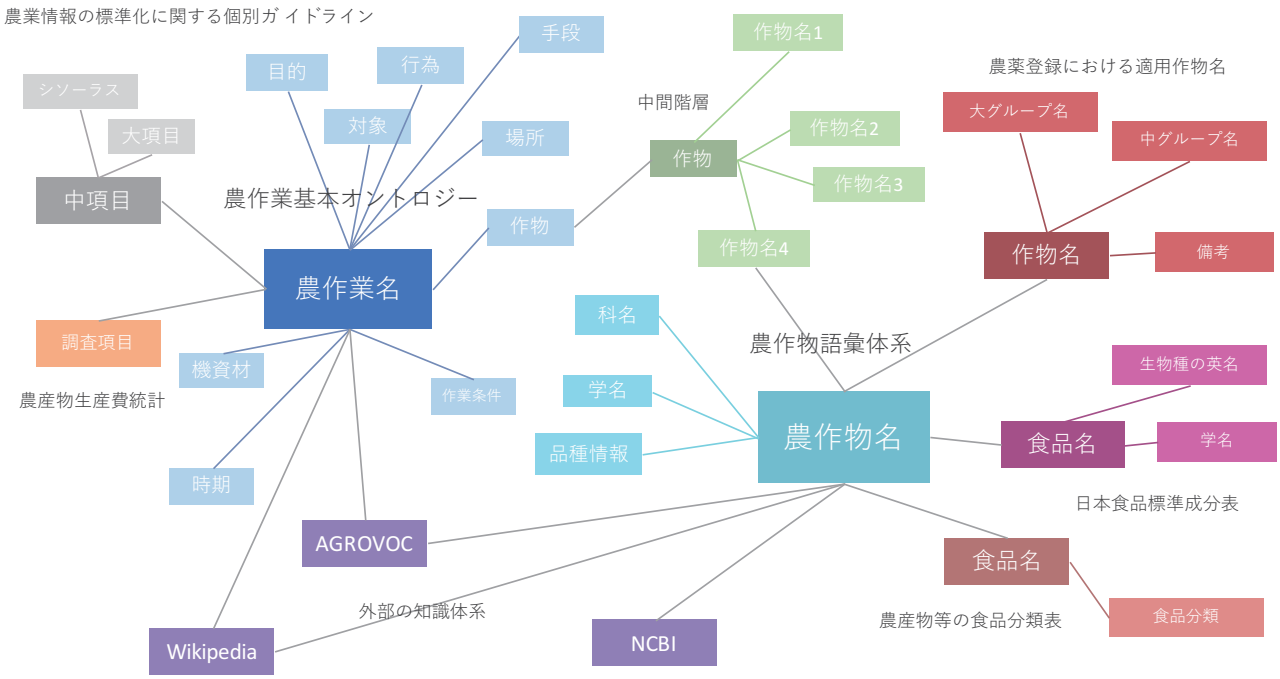


図 1: 農業分野における語彙の連携.

農作物語彙体系も農作業基本オントロジーとともに共通農業語彙にて公開されており，農作物の URI やデータとして SKOS/Turtle，CSV 形式のデータを提供している．また，農作物における 3 つの語彙が持つ重要性を考慮し，その利活用のために農作物の URI と機械可読性を向上させたデータ形式も提供している．それぞれの URI には対応関係がリンクで連携されており，対応関係の把握や詳細の情報の取得が可能である．ともに提供している API とアプリケーションを用いてこれらの語彙を農作物語彙体系が持つ対応関係の情報を用いて横断的な活用も可能となる．例えば，「農薬登録における適用作物名」の「ばれいしょ」も農作物語彙体系の API を用いると「日本食品標準成分表」の「じゃがいも」が持つ生物種の英名や学名の情報を取得することが可能となる．

そして，データの相互運用性を確保するために外部データとの連携も実装されており，内閣官房による「農業 IT システムで用いる農作物の名称に関する個別ガイドライン」の作物名，知識の連携としては Wikipedia 日本語版，国際標準である FAO の AGROVOC，そして生物種のデータであるアメリカの NCBI(National Center for Biotechnology Information)*4 の taxonomy ID と連携されている．農作物の語彙のみならず，外部のデータセットや知識体系とも繋がることで農作物全般に対する包括的情報運用が可能である．

3.3 語彙の連携

本節では上記に述べた農作業と農作物の標準語彙として構築された農作業基本オントロジーと農作物語彙体系を連携するために構築した中間階層の語彙体系とその利活用について述べる．

3.3.1 連携のための中間語彙体系の設計

農作業の中には農作物に依存しない一般的な農作業もあるが，特定の農作物でのみ行われる農作業が存在する．例えば，収穫作業の一つである「堀取り」は根菜類に適用される農作業であり，穀物類などには適用されない．そのために，農作業を記述するためには適用される作物の定義が必要であるが，すべての作物名を農作業の定義に用いることは管理・運用において非効率的である．そこで，本研究では効率的で円滑な連携が行われるように語彙と語彙の間に中間階層となる語彙体系の構築を提案する．「堀取り」を定義する属性の一つである「作物」の値を「根菜類」とし，根菜類について中間語彙体系で定義をする．中間語彙体系には該当する語彙とそれに含まれる詳細を収録する．例えば，「根菜類」の場合，農作物語彙体系に収録されている「ダイコン」，「カブ」，「ジャガイモ」などを対象作物の情報として持つ．これによって農作業と農作物の語彙が繋がり，データの連携も可能となる．図 1 は農作業基本オントロジーと農作物語彙体系を中間階層の語彙体系を用いて連携した全体図である．

このように中間階層の定義し，異なる語彙を連携すると，農作業と農作物，それぞれの語彙は基本情報のみ収録することができ，今後語彙の拡張にも対応することができるなど，その管理における手間が軽減する．それぞれ語彙が持つ URI を利用することで一貫性を持つ語彙の管理も可能となる．また，新規語彙の導入や概念の変更など目的によって中間階層の概念を追加することによってより柔軟な対応と幅広い利活用も可能となる．

3.3.2 中間語彙体系を用いた語彙の利活用

中間階層の語彙体系を用いると農業 ICT システムも横断的な利活用が可能である．図 2 は中間階層を用いて農作業基本オントロジーの農作業名称とそれに適用される農作物を農作物語彙体系を通じて既存語彙の情報を問い合わせする API の例である．農作業基本オントロジーに収録されている「堀取り」の作物に関する属性の値は「根菜類」である．中間階層の語彙体系から「根菜類」に該当する農作物語彙体系の農作物名を問い合わせし，そのリンクから「農薬登録における適用作物」と「日本食品標準成分表」の情報を得ることができる．

*4 https://www.ncbi.nlm.nih.gov/

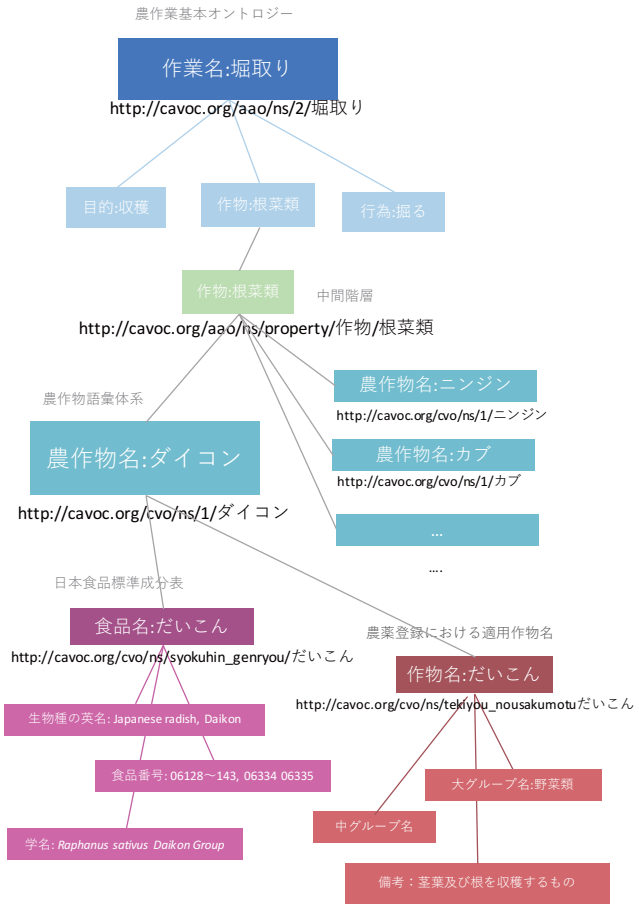


図 2: 語彙の連携の利活用.

4. 考察と今後の課題

今回は農作業と農作物の標準語彙を連携するために中間となる語彙体系を構築した．それぞれの語彙の構造を維持し，語彙を拡張することによって概念を具体化することができた．語彙が持つ構造を明確にしたまま，異なる語彙との連携が可能となり，語彙の横断的な運用についてその可能性を確認した．今後は機資材や土地利用についても具体化し，標準語彙を構築することで知識体系として繋がる語彙体系を構築していきたい．

中間階層となる語彙体系は語彙と語彙の関係を定義するために様々な対応関係の情報を収録する必要がある．対応関係を定義するためには様々な分野，業界の語彙を調査しなければならない．今後は各分野の有識者に依頼し，協議を通じて整理していく予定である．

5. おわりに

本稿では農業 ICT システムのデータ連携のための標準語彙を管理するためにナレッジグラフを用いた語彙体系を提案した．既存の語彙について分析を行い，標準語彙とその管理に求められる条件として明確な構造，語彙間の連携を提示した．ナレッジグラフによる語彙体系の設計と管理は ICT システムでの利活用に適合しており，必要に応じて拡張にも柔軟な対応が可能である．今後は農業全般における標準語彙の確立と連携にナレッジグラフを取り組み，包括的な標準語彙の運用と安定的な活用を目指していきたい．

謝辞

本研究（の一部）は，総合科学技術・イノベーション会議の SIP（戦 略的イノベーション創造プログラム）「次世代農林水産業創造技術」（管理法人：農研機構生物系特定産業技術研究支援センター）の支援を受けて行った．

参考文献

[内閣官房 17a] 内閣官房, 農業情報創成・流通促進戦略に係る標準化ロードマップ (第 3 版), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310roadmap.pdf> 2016 年 3 月 1 日参照.

[内閣官房 17b] 内閣官房, 農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第 3 版), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310gl1_besshi.xlsx> 2017 年 3 月 1 日参照.

[内閣官房 17c] 内閣官房, 農業 IT システムで用いる農作物の名称に関する個別ガイドライン（第 2 版）, <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310gl3_besshi.xlsx> 2017 年 3 月 1 日参照.

[農林水産省 18] 農林水産省, 農薬登録における適用作物名, <https://www.acis.famic.go.jp/shinsei/3986/3986beppyou1.pdf> 2018 年 3 月 1 日参照.

[厚生労働省 15] 厚生労働省, 農産物等の食品分類表, <http://www.mhlw.go.jp/file/06-Seisakujouhou-11130500-Shokuhinanzenbu/0000159254.pdf> 2018 年 3 月 1 日参照.

[文部科学省 15] 文部科学省, 日本食品標準成分表，資料 2 食品の原料となる生物種の英名・学名, <http://www.mext.go.jp/component/a_menu/science/detail/__icsFiles/afieldfile/2017/06/22/1365334-1-0321r2.xlsx> 2018 年 3 月 1 日参照.

[朱 16] 朱 成敏, 小出 誠二, 武田 英明, 法隆 大輔, 竹崎 あかね, 吉田 智一：記述論理に基づく農作業オントロジーの設計と応用, 第 38 回人工知能学会セマンティックウェブとオントロジー研究会（SIGSWO）, 06, 2016.

[農林水産省 16] 農林水産省, 農産物生産費統計, <http://www.maff.go.jp/j/tokei/kouhyou/noukei/seisanhi_nousan/> 2018 年 3 月 14 参照.

[竹崎 17] 竹崎あかね, 朱成敏, 武田英明, 吉田智一： 農業 IT システム間のデータ連携を促進する農作物語彙体系の構築, 電子情報通信学会技術研究報告，知的環境とセンサネットワーク研究会, vol. 117, no. 310, ASN2017-J27, pp. 98-99 (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習を用いた果実生育情報抽出 Web アプリケーション

Fruits Growth Information Extracting Application by using Deep Learning Network

小林 一樹 *1　　中村 俊輝 *2
Kazuki Kobayashi　　Toshiki Nakamura

*1信州大学 学術研究院
Academic Assembly, Shinshu University

*2信州大学 大学院 総合理工学系研究科
Graduate School of Science and Technology, Shinshu University

This paper proposes an apple fruits detection web application that automatically detects apples and its sizes in field monitoring images. The proposed application uses a deep neural network with a way to create massive amounts of artificial training images. We developed two methods to create the training images; an imitation method and a random method. The imitation method creates labeled training images like real field images. The random method creates the training images unlike real field images because it uses random composition parameters. The experimental result showed that the imitation method achieved an F-measure of 72.1.

1. はじめに

近年，ICT や IoT を活用して農作物の生育状態を把握し，生産性の向上をめざす研究が行われている．たとえば，深層学習を用いた果樹の果実検出 [Bargoti 17] に関する研究では，深層学習の一手法である Faster R-CNN を用いてリンゴやマンゴー，アーモンドの果実の部分を画像から検出している．現状では果実の検出率に着目されているが，果実サイズは生長を把握する上で重要な指標であり，大量の高精細画像から生長曲線を抽出する試みも提案されている [源野 17]．特に，多くの果樹は果実を覆い隠すように葉を茂らせているため，観測画像から葉で隠れた果実のサイズを把握することは困難である．

そこで本研究では，定点観測装置から取得した可視光画像を用いて，果実の一部が隠れている対象物であってもサイズ推定が可能な生育情報抽出アプリケーションを提案する．ここでは，深層学習を用い，そのための効率的な訓練画像生成システムの開発と，果実サイズの統計的情報を得るシステムの開発とを行った．以下では，提案する Web アプリケーションについて説明する．

2. 果実生育情報の抽出

果実生育情報として果実の直径を対象とした．ほ場モニタリング画像から，果実サイズを自動的に抽出するために畳み込みニューラルネットワークによる物体検出アルゴリズムの 1 つである Yolo9000[Redmon 17] を用いた．Yolo9000 を用いることで，画像からリンゴ果実の検出と位置の特定とを行うことができる．この果実検出時の境界領域のサイズを果実サイズとして抽出し，時系列で並べて生長曲線を生成する．

2.1 訓練画像の合成

果実サイズを画像から自動で抽出するために，本研究では訓練画像を自動的に生成するアプローチをとる．訓練データは人間が確認してラベル付することが一般的であるが，多くの訓練データを生成するには多大な労力を要する．ここでは複数の農園モニタリング画像 (5184 × 3456 px) から，リンゴ果実部分 60 枚，背景画像 (1088 × 1088 px, 晴れ，曇り)24 枚，葉画像 10 枚を抽出し，背景画像に対して葉画像や果実画像の輝度や角度，拡大率をランダムに変更したものをランダムな位置に配置して訓練画像を合成している．このとき，葉画像を配置してから，果実画像を配置し，最後に再び葉画像を合成することで果実の一部が葉によって隠される画像を生成している．実際には，この 1088×1088 px の画像が 416 × 416 px に縮小され入力層に渡される．



図 1: ランダム合成（左 4 枚）と模倣合成（右 4 枚）

訓練画像の合成には 2 つの方針を採用した．1 つは切り出し画像をランダムに変換と配置するランダム条件であり，もう 1 つは，現実の農園画像を模倣するような範囲に変換と配置を行う模倣条件である．図 1 に合成画像例を示す．

2.2 学習条件の比較実験

高い検出率を得るための条件を特定するために，要因を設定して深層学習を行った．表 1 に学習条件と，未学習のデータに対するテスト結果を示す．テスト用のデータには，手作業でバウンディングボックスを作成した 3 枚のモニタリング画像から切り出した 33 枚の画像（果実数 353 個）を用いた．すべてのテスト画像サイズは 1088×1088 px であり，学習時と同様に 416 × 416 px に縮小され入力層に渡される．

比較実験の結果，実際の圃場に近い設定で輝度や角度，果実色などを組み合わせて訓練画像を生成する模倣条件で，3000 枚の学習，輪郭処理は低精度な場合に最も F 値が高くなることが分かった．ここで，輪郭処理精度とは，上記の訓練画像合成用に切り出した果実画像や葉画像の輪郭部分の切り出し精度を意味する．訓練画像を自動的に生成するが，画像の切り出し作業は人間が手作業で実施する必要があり，この精度が低いこ

連絡先: 小林一樹，信州大学 学術研究院 工学系，長野県長野市若里 4-17-1 信州大学 工学部 電子情報システム工学科，026-269-5456，kby@shinshu-u.ac.jp

表 1: 学習条件の違いと検出精度 [%]

	ランダム				模倣			
学習枚数	3000		30000		3000		3000	
輪郭処理精度	高	低	高	低	高	低	高	低
再現率	58.6	43.6	62.9	42.8	44.5	63.7	55.8	44.5
適合率	67.7	65.8	30.2	84.4	76.2	69.7	63.3	89.2
F 値	62.8	52.5	40.8	56.8	56.2	66.6	59.3	59.4

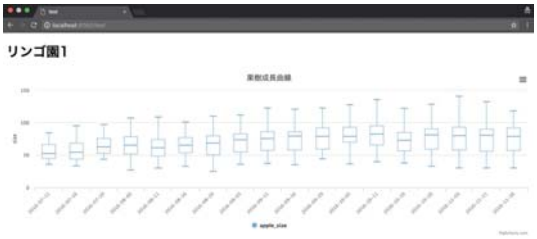


図 2: 果実生育情報抽出 Web アプリケーション

とは画像編集ソフトウェアの習熟度を必要としないことに相当する．

最終的に，この条件を用いて既存の COCO データセット *1 に対して，生成した訓練画像を追加学習させたところ，F 値 72.1（適合率 71.0, 再現率 73.4）となり，元の COCO データセットによる F 値 59.9 を上回る結果となった．すべての検出時において，Yolo の検出しきい値は 0.1 としている．

3. 果実生育情報抽出 Web アプリケーション

上記のように，COCO データセットに追加学習したネットワークを用いて，果実の生育情報を抽出するアプリケーションを開発した．

図 2 に提案する果実生育情報抽出 Web アプリケーションの画面を示す．Web アプリケーションの初期画面では，生長曲線表示モードとなっており，画像から抽出した果実サイズを，時系列順で箱ひげ図によって示している．箱ひげ図を用いることで，抽出されたサイズの分布を容易に把握することができる．また，特定の日の箱ひげをクリックすることで，図 3 に示すようにサイズ分布表示モードとなり，サイズ区分ごとのヒストグラムを表示できる．ページ下部には，検出部分をアノテーションした画像が表示され，検出されたリンゴ果実の境界領域を画像上で確認できる．

4. 考察

提案手法では，画像から抽出した果実サイズの分布を求めることで，果樹全体の生育状態の把握を試みている．この手法によって時系列での分布の推移を把握することができるが，より厳密に時系列データとして扱うのであれば，同じ果実を時系列で追跡して可視化することが望ましい．現状では，果実の個体識別が実現できていないが，定点観測画像という撮影画像が固定されていることの特徴を活かし，座標情報を利用したシンプルな手法によって実現できる可能性がある．このとき，果樹の重さで枝が下に垂れ下がったり，生産者が枝を支える支柱を追加して果実の位置が変化することがある点に注意する必要がある．

*1 http://cocodataset.org

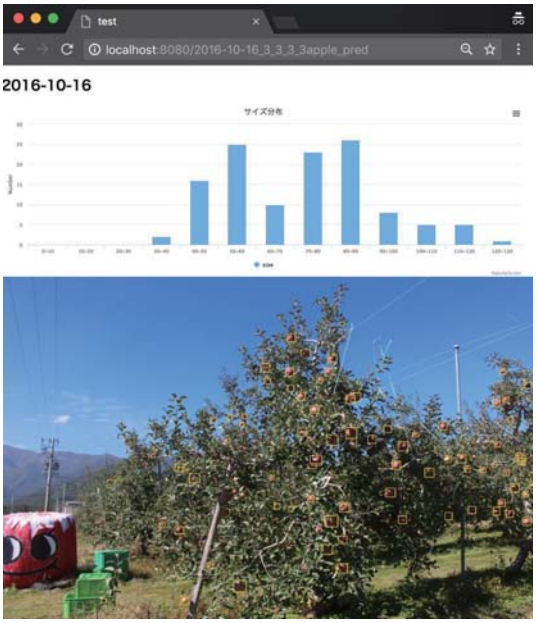


図 3: ヒストグラム表示モード

5. まとめ

本研究では，定点観測画像から果樹の生長把握を容易に行うアプリケーションを実現するために，深層学習を用いたリンゴ果実サイズの自動抽出と，その情報を用いた統計的情報の提示および，生長曲線の生成を行うアプリケーションを開発した．リンゴ果実の検出精度は F 値 72.4 であり，果実の生長を把握する上で実用的な値だと考えらえる．今後，テスト画像数を増加させるとともに，抽出した果実サイズの精度評価を行う予定である．

参考文献

[Bargoti 17] Suchet Bargoti, James Underwood: Deep fruit detection in orchards, In Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp.3626-3633 (2017)

[源野 17] 源野 広和, 小林 一樹: 大量高精細画像からの果実生育情報の抽出, 農業情報研究, Vol.26, No.4, pp.100-114 (2017)

[Redmon 17] Joseph Redmon, Ali Farhadi: YOLO9000: Better, Faster, Stronger, In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.6517-6525 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

農業ITシステム間のデータ連携に貢献する農作物語彙体系の構築

Development of Crop Vocabulary to facilitate interoperability among agricultural IT systems

竹崎あかね*1　朱成敏*2　武田英明*2　吉田智一*1
Akane TAKEZAKI　Sungmin JOO　Hideaki TAKEDA　Tomokazu YOSHIDA[1]

*1農研機構
National Agriculture and Food Research Organization

*2国立情報学研究所
National Institute of Informatics

To facilitate interoperability of knowledge between agricultural IT systems, Crop Vocabulary; CVO has been developed. CVO provides a list of properties such as primary name, scientific name, synonyms for describing the crop. Most of primary names in CVO denote the cultivated plants based on the taxonomic classification, but some of them denote the groups of cultivated plants according to both practical purpose and their characteristics. CVO has links to other standards in agricultural chemical industry, food industry, academic society. This result shows CVO can support data integration from different sources.

1. はじめに

これまで，営農管理支援システムなど多くの農業 IT システムが開発され，農業生産者の営農支援に貢献してきた．一方で農業 IT システム間のデータや機能連携は実現しておらず，各システムにおける収集データや開発機能を有効活用した営農支援がなされているとはいえない．そこで，我々は，農業 IT システム間の連携による営農支援実現を目的として，概念共通化の基盤となる農作業基本オントロジー（Agriculture Activity Ontology; AAO）を構築し共通農業語彙；CAVOC のサイトで公開してきた*1．本稿では農作業の名称とともに，農業 IT システムでよく利用され標準化の優先度が高い農作物の名称[内閣官房 2017a]を対象に農作物語彙体系（Crop VOcabulary; CVO）を開発したので報告する．

2. CVO の概要

2.1 開発背景

安全な農作物を消費者に供給し健康への悪影響を未然に防ぐため，農場から食卓にいたるフードチェーン全体で食品の安全性を確保する重要性が高まっている．上記を背景に，これまでフードチェーンの各段階が収集管理してきたデータを，相互利用できるシステム構築が求められている．一方，フードチェーンに関与する多様なキープレーヤー；生産者・流通小売業者・消費者・行政機関が，同じ農作物のデータに異なる名称を付けて管理することは少なくなく，データの相互利用を阻害する要因の一つとなっている．例えば“サヤエンドウ”は同義語“さやえんどう”・“エンドウ”・“さやまめ“や，品種名“きぬさや豌豆”・“キヌサヤ”（“きぬさや豌豆”の同義語）で表現される．そこで，データ項目名が示す農作物概念の共通化を進め，システム間のデータ連携を促進するため，農作物を定義する CVO を構築した．

2.2 収集語彙

CVO は，商品形態を属性とした名称（例；1/2 ダイコン）を利用する小売業者を除き，フードチェーン上流の生産者，流通業者，および行政機関で利用される農作物名を収集対象とした．聞き取り調査の結果，農薬適用基準と残留農薬基準情報の項目名統一に対する要望があった．そこで，まずは 2 種類の農薬情報の項目名連携を主要な目的として，農薬適用基準の項目名となる「農薬登録における適用作物名［農林水産省 2018a］」で「作物名」，「作物名に含まれる別名・地方名・品種名等の例」に分類される語彙，残留農薬基準情報の項目名となる「農作物等の食品分類表［厚生労働省 2013］」の食品名を収集し，そのうち農薬の影響が大きい食用栽培植物を中心に CVO に収録した．2 種類の農薬情報の語彙には，植物学的分類に基づく栽培植物名，属性で修飾あるいは分類された栽培植物名，品種（例；デラウェア），地方団体商標（例；中島菜）等があった．上記属性には，食用時，食用部位の残留農薬量に影響を及ぼすであろう，部位（例；ナガイモとナガイモムカゴ）・用途（飼料用トウモロコシと食用トウモロコシ）・成熟度（成熟トウモロコシと未成熟トウモロコシ）等を認めた．

2.3 定義方針

CVO は，農作物を標準和名，学名，よみ，別名，英名等で表現する．標準和名の多くは植物学的分類に基づいた栽培植物であるが，農薬情報の項目名共通化が実現するよう栽培植物を利用部位・用途・成熟度等で分類した集合，品種，地域団体商標等も示す．基本的に構造を持たないが，栽培植物と，栽培植物を分類した集合は階層関係にある．なお品種や地域団体商標は厳密には栽培植物に分類されるが，CVO では階層関係にはない．CVO に収録された農作物語彙数は 1,191（2018 年 2 月時点）である．

2.4 公開および外部語彙リストとの連携

CVO への語彙収録と併行して，2 種類の農薬情報の語彙，登録品種名［農林水産省 2018b］，行政機関の代表的なデータセット，日本食品標準成分表の項目名となる「食品の原料となる生物種の英名・学名［文部科学省 2015］」の食品名（生物種名）との対応を調査し CVO を基準とした対応表を完成させた．なお，「農薬登録における適用作物名」の語彙については備考に分類される属性（利用部位・成熟度等）も考慮して対応関係を調査した．

CVO は，共通農業語彙；CAVOC のサイトで公開している*2（図 1）．CVO は固有の名前空間（URI）を持ち，農作物はその上で定義される．URI では科名，英名，日本語別名，学名とともに，CVO と対応関係にある語彙を表示する．また，CAVOC サイトでは CVO 農作物語彙リストを csv と SKOS/Turtle 形式でダウンロードできる．

農薬登録における適用作物名［農林水産省 2018a］，農作物等の食品分類表［厚生労働省 2013］，食品の原料となる生物種の英名・学名［文部科学省 2015］のデータセット，および登録品種名を整理した CAVOC 品種リストの項目名（作物名，食品

*連絡先：竹崎あかね，農研機構革新工学センター，〒305-0856 茨城県つくば市観音台 1-31-1，E-mail: akane@affrc.go.jp

*1 http://www.cavoc.org/　*2 http://cavoc.org/cvo/ns/1/

名，品種名）に URI を提供し，その上で詳細情報を表示するようにした．また CVO を基準とした対応表に基づき CVO と外部データセットをリンクで連携させ，対応関係やデータセットの詳細情報の確認を可能にした［朱 2018］．

内閣官房は「農業 IT システムで用いる農作物の名称に関する個別ガイドライン」で優先的に利用すべき農作物名リストを提示した [内閣官房 2017b]．われわれは，農作物リストの小分類の語彙について CAVOC 農作物語彙との対応を調査した．小分類の語彙に URI を提供し，CVO 農作物語彙との対応関係に基づき CVO とガイドラインのデータセットをリンクで連携させた．

栽培植物は，植物学的分類の変更により学名変更があること，方言などの別名が多い特徴がある．そこで，学名が参照できるよう，学術情報である National Center for Biotechnology Information；NCBI 管理の taxonomy DB[*3] から CVO 農作物語彙の英名あるいは学名に該当する Taxonomy ID を，方言などの別名が参照できるよう一般情報であるウィキペディア[*4] から CVO 農作物語彙の植物和名，あるいは別名に該当するタイトル名を調査し，CVO 農作物語彙と対応させた．

以上の結果 CVO を農作物標準語彙とした関連語彙の連携が実現した．また CVO を基盤として「農業 IT システムで用いる農作物の名称に関する個別ガイドライン」の農作物名リスト[内閣官房 2017b]が関連語彙リストと連携したことで，農作物名の標準化が加速すると期待する．

3. CVO の運用体制構築に向けた活動

内閣府・戦略的イノベーション創造プログラム（SIP）「次世代農林水産業創造技術」では，農業の担い手がデータを使った生産性向上や経営改善ができる環境整備をするために，データ連携機能やデータ提供機能をもつ農業データ連携基盤が構築され，2019 年 4 月から本格運用される．我々は CVO が概念間の関係に基づく農作物データの連携基盤となりうることから，農業データ連携基盤上での運用をめざしている．また情報処理推進機構が管理運営する共通語彙基盤［情報処理推進機構 2018］のパートナー制度に登録し一般分野と農業分野のデータ連携に向けた協力関係を築いている．

CVO 農作物語彙は有識者の意見に基づき継続して改修している．一方で CVO 開発の主目的である農薬情報のデータ連携のためには，更新される農薬情報を考慮した語彙拡張が必要である．有識者により CVO の収録語彙を継続的に精査できるよう，メンテナンス体制の整備が課題である．

4. まとめ

農業 IT システム間のデータ連携促進のため，農作物を定義する CVO を構築し公開した．CVO では，農作物は標準和名，学名，よみ，別名で表現される．CVO に含まれる農作物の多くは植物学的分類に基づいた栽培植物名であるが，一部は利用部位，用途等で修飾された栽培植物名である．また，CVO 農作物語彙は農薬情報，食品成分情報，品種情報，学術情報等の語彙との対応関係に基づきリンクで連携されており，CVO を基盤とした関連語彙の相互連携が実現した．

謝辞

本報告内閣府～農水省予算により生研センターが管理運営する「SIP（戦略的イノベーション創造プログラム）次世代農林水産業創造技術」での研究成果に基づく．

参考文献

[厚生労働省 2013] 厚生労働省：農産物等の食品分類表，<http://www.mhlw.go.jp/topics/bukyoku/iyaku/syoku-anzen/zanryu2/dl/070516-1.pdf>，2018 年 3 月 7 日参照，2013.

［情報処理推進機構 2018］情報処理推進機構：共通語彙基盤，<https://imi.go.jp/goi/>，2018年1月19日参照．

[文部科学省 2015] 文部科学省：日本食品標準成分表 2015 年版，食品の原料となる生物種の英名・学名，<http://www.mext.go.jp/component/a_menu/science/detail/__icsFiles/afieldfile/2017/06/22/1365334-1-0321r2.xlsx>，2018 年 3 月 7 日参照，2015.

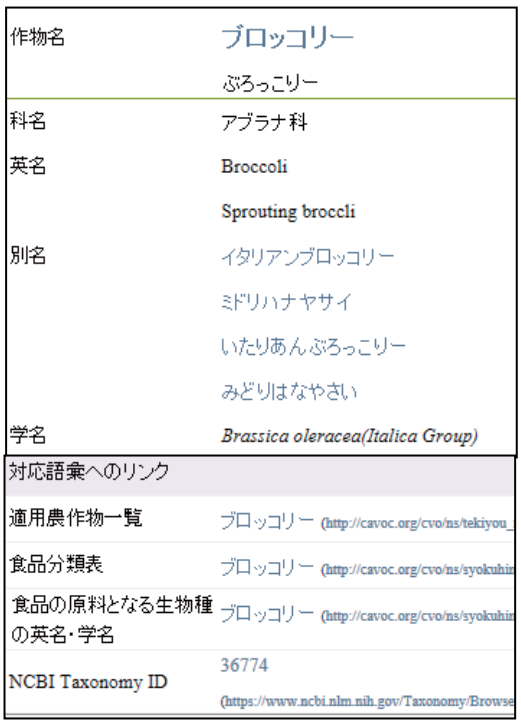
[内閣官房 2017a] 内閣官房: 農業情報創成・流通促進戦略に係る標準化ロードマップ（第3版），< http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310roadmap.docx >，2018年3月7日参照， 2017.

[内閣官房 2017b］内閣官房：農業ITシステムで用いる農作物の名称に関する個別ガイドライン第2版，別紙農作物名ガイドライン基本表，<http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310gl3_besshi.xlsx>，2018 年 3 月 7 日参照，2017.

[農林水産省 2018a] 農林水産省：農薬登録における適用作物名，<https://www.acis.famic.go.jp/shinsei/3986/3986beppyou1.pdf>，2018 年 3 月 7 日参照，2018.

[農林水産省 2018b] 農林水産省：品種登録ホームページ，< http://www.hinshu2.maff.go.jp/>，2018 年 3 月 7 日参照 .

[朱 2018] 朱成敏・武田英明・竹崎あかね・吉田智一：農業データ連携のためのナレッジグラフに基づく標準語彙の運用，第 32 回人工知能学会全国大会，2018.



作物名	ブロッコリー ぶろっこりー
科名	アブラナ科
英名	Broccoli Sprouting broccli
別名	イタリアンブロッコリー ミドリハナヤサイ いたりあんぶろっこりー みどりはなやさい
学名	*Brassica oleracea(Italica Group)*
対応語彙へのリンク	
適用農作物一覧	ブロッコリー (http://cavoc.org/cvo/ns/tekiyou_
食品分類表	ブロッコリー (http://cavoc.org/cvo/ns/syokuhin
食品の原料となる生物種の英名・学名	ブロッコリー (http://cavoc.org/cvo/ns/syokuhin
NCBI Taxonomy ID	36774 (https://www.ncbi.nlm.nih.gov/Taxonomy/Browse

図1．CAVOC における CVO の公開
http://www.cavoc.org/cvo/ns/1/ブロッコリー

[*3] https://www.ncbi.nlm.nih.gov/taxonomy

[*4] https://ja.wikipedia.org/wiki/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

複眼カメラ TOMBOを用いたフィールド モニタリング

Agricultural Field Monitoring for Multi-Lens Camera TOMBO

下林 史弥 *1 Fumiya SHIMOBAYASHI
小林 一樹 *2 Kazuki KOBAYASHI
香川 景一郎 *3 Keiichiro KAGAWA
郭 威 *4 Wei GUO
平藤 雅之 *4 Masayuki HIRAFUJI
谷田 純 *5 Jun TANIDA

*1信州大学 総合理工学研究科 Graduate School of Science and Technology, Shinshu University
*2信州大学 学術研究院 Academic Assembly, Shinshu University
*3静岡大学 電子工学研究所 Research Institute of Electronics, Shizuoka University
*4東京大学 農学生命科学研究科附属生態調和農学機構 Institute for Sustainable Agro-ecosystem Services, The University of Tokyo
*5大阪大学 情報科学研究科 Information Science and Technology, Osaka University

This paper proposes a multi-spectral camera that is light weight, compact size, and easy to use. In agricultural field, drones with hyper spectral cameras are sometimes used for growth prediction of crops. They can obtain a growth index of NDVI or NDRE that are useful for estimating crop status. However, there are few cases in practical aspects because such cameras are heavy, voluminous, and too expensive accurate instruments that require deliberate manipulation. The proposed camera resolves such practical issues in a practical agricultural situation.

1. はじめに

近年，農作物の収穫量および品質の向上を目的に，ほ場での作物の計測やデータの解析を行う精密農業に関する研究が行われている．農作物の生育データの取得には，ドローンとマルチスペクトルカメラやハイパースペクトルカメラを組み合わせた手法を用いることがある．これらのカメラから得られる複数波長の画像を用いることで，NDVI や NDRE といった作物の生育状況の把握に役立つ指標を得られる．横山ら [横山 16] は，ほ場を対象として，ドローンに搭載可能な 2 種類のマルチスペクトルカメラと 1 種類のハイパースペクトルカメラから得たデータを比較している．このような計測に使用されるハイパースペクトルカメラを用いることで広い波長帯域の画像を取得することができるが，価格と 500g ほどの重量が障害となる．一方，マルチスペクトルカメラは軽量である反面，波長帯域が限られてしまう制限がある．

そこで本研究では，波長選択の柔軟性と小型軽量な特徴を持つマルチスペクトルカメラである TOMBO [香川 10] を用いたほ場モニタリングを提案する．TOMBO は従来のマルチスペクトルカメラと異なり，複数のイメージセンサや複雑な分光器を使用しないマルチスペクトルカメラである．実験では TOMBO のフィールド モニタリングへの応用に向けて，既存の農業用マルチスペクトルカメラとの比較を通して検討を行う．

2. 複眼カメラ TOMBO

TOMBO の構成図と外観を図 1 に表す．利用した TOMBO は 1 枚のイメージセンサ上に 9 つのレンズから構成されるレンズアレイを搭載した重量 8.5g 程度の小型マルチカメラである．レンズはそれぞれ隔壁で区切られており，撮影される 1 枚の画像上には，9 つの像が撮影される．この像は視差によるずれを無視すれば同じものとなる．それぞれのレンズに透過波長の異なる光学フィルタを設定することで小型マルチスペクトルカメラとして使用することができる．従来のマルチスペクトルカメラは，撮影波長が固定されているが，TOMBO は光学フィルタを交換可能であり，撮影対象に応じてフィルタの透過波長を自由に選択できる．

連絡先: 下林史弥，信州大学総合理工学研究科，長野県長野市若里 4-17-1

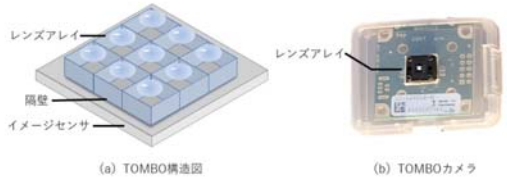


図 1: 複眼カメラ TOMBO

表 1: TOMBO と RedEdge の撮影波長

カメラ＼波長 [nm]	青	緑	赤	レッドエッジ *1	近赤外		
TOMBO	450(2 枚)	520	650	730	850	920	970
RedEdge	475	560	668	717	840	–	–

*1 植物の生育状況の評価に用いられる 700nm 付近の波長.

撮影画像例を図 3 に表す．図 3(a) を TOMBO で撮影したものが図 3(b) となる．今回の実験では中央のレンズを使用しなかったため，8 つの像が撮影されている．使用した光学フィルタの透過波長と，比較対象としたマルチスペクトルカメラ RedEdge の撮影波長を表 1 に示す．

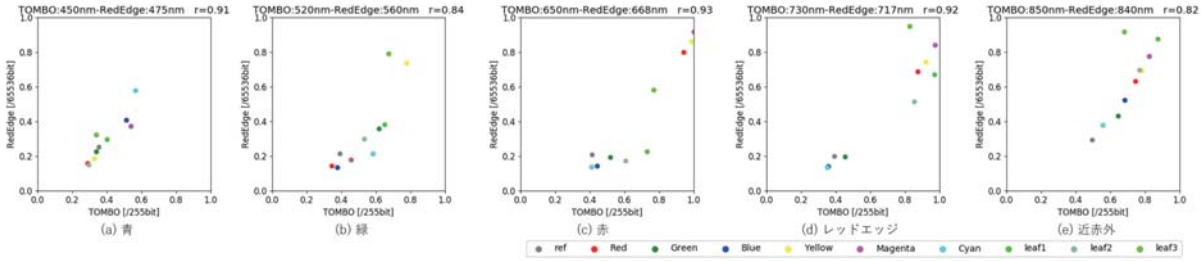


図 2: TOMBO と RedEdge の相関

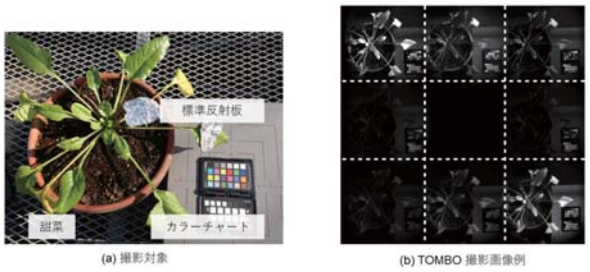


図 3: 撮影画像例

3. 撮影実験

3.1 実験方法

TOMBO の応用可能性を評価するため，既にフィールド モニタリングに使用されているマルチスペクト ルカメラである Micases RedEdge との比較を行った．RedEdge は，5 つの波長に感度を持つセンサを内蔵した重量 150g のマルチスペクトルカメラである．そのセンサはそれぞれの波長ごとに独立しているので，1 回の撮影で 5 枚の画像を撮影する．実験では TOMBO によって撮影された 7 波長の画像から RedEdge の波長に最も近いものを選び，5 組の画像において比較を行った．

3.2 実験環境

撮影実験は東京大学大学院生体調和農学機構にて栽培されている甜菜を対象として 2017 年 12 月 23 日に行った．撮影は，ビニールハウス内で行い，天候は晴天であったため光源は太陽光とした．図 3(a) のように甜菜，カラーチャート，標準反射板が 1 枚の画像に映るよう TOMBO，RedEdge で撮影を行いそれぞれの画像を比較する．カラーチャート は X-Rite Color Checker Passport を，標準反射板は Nikon 18%を使用した．

3.3 分析

2 種類のカメラによって撮影された画像中の輝度を比較することで評価を行う．標準反射板 1 点，カラーチャート 6 点 (青，緑，赤，黄，シアン，マゼンタ)，甜菜の葉 3 点の計 10 点において輝度を求め，TOMBO の輝度を横軸，RedEdge の輝度を縦軸としたグラフを作成することで TOMBO-RedEdge 間の相関について考察を行った．2 枚の画像における同一点の輝度を求める際に，画像中の計測点をマウスでクリックし，近傍の平均輝度を算出するツールを自作した．

3.4 分析結果

相関グラフを図 2 に表す．各波長における相関係数は，(a) 青 0.91，(b) 緑 0.84，(c) 赤 0.93，(d) レッド エッジ 0.92，(e) 近赤外 0.82 となり，どの波長においても強い相関を確認することができた．

4. 考察

相関グラフより，どの波長においても強い相関が認められ TOMBO の農業分野への適用可能性は高いと考えられる．一方で，波長ごとに相関グラフを比較すると，プロット のばらつきには一定の法則性がなく，波長ごとの特性について調査を行う必要があると考えられる．

図 2(c) のように，TOMBO の輝度が飽和している箇所があるため，レンズの露光時間を短くする必要がある．TOMBO は 1 枚のイメージセンサを使用しているため露光時間や感度に関する設定はすべての波長において共通である．この点に関して，今後，有効な制御手法を検討する必要がある．

実験では光源として太陽光を使用したが，撮影時刻により光源のスペクト ル成分が変化してしまい同じ箇所の輝度が変化してしまう問題が生じた．今後の実験では，暗室内で太陽光に近いスペクト ル成分の光源を用いて撮影を行う必要がある．また，フィールド での実用に向け TOMBO の校正方法についても検討する必要がある．

5. まとめ

複眼カメラ TOMBO のフィールド モニタリングにおける応用可能性を調査するため，既存の農業用マルチスペクト ルカメラ RedEdge との比較実験を行った．実験の結果，TOMBO と RedEdge で撮影された画像には強い相関があり，TOMBO の農業応用への可能性が示唆された．今後は，フィールド モニタリングへの応用に向け，検証実験を重ねる予定である．

参考文献

[横山 16] 横山 正樹, 井上 吉雄, 後藤 元, 小手 和徳: ドローン搭載マルチバンド・ハイパースペクト ルカメラの反射率特性の解析と圃場観測，計測と 制御, Vol.55, No.9, pp.810–813 (2016).

[香川 10] 香川 景一郎, 田中 映治, 山田 憲嗣, 谷田 純: 3 次元カメラとしての小型薄型複眼カメラ TOMBO の可能性と展開, 映像情報メディア学会技術報告, Vol.34, No.43, pp.37–40 (2010).

Oral presentation | Organized Session | [Organized Session] OS-10

[2G3-OS-10c] [Organized Session] OS-10

Wed. Jun 6, 2018 3:20 PM - 4:40 PM Room G (5F Ruby Hall Hiten)

[2G3-OS-10c-01] Features values for recognition of swine's head direction from video.

○Takuya Sato[1], Misaki Mito[1], Takuji Kawagishi[1], Koichi Mizutani[1,2], Keiichi Zempo[1,2], Naoto Wakatsuki[1,2], Nobuhiro Takemae[3], Takehiko Saito[3] （1. Graduate School of Systems and Information Engineering, University of Tsukuba, 2. Faculty of Engineering, Information and Systems, University of Tsukuba, 3. National Institute of Animal Health, NARO）

3:20 PM - 3:40 PM

[2G3-OS-10c-02] Development of Image Sensing Method to Automatically Obtain Soybean Growth Condition

○Kazuki Omura[1], So Yahata[1], Seiichi Ozawa[1], Takenao Ohkawa[1], Noriyuki Murakami[2], Hiroyuki Tsuji[2] （1. Kobe University, 2. Hokkaido National Agricultural Research Center）

3:40 PM - 4:00 PM

[2G3-OS-10c-03] Detection of Cattle States Relevant to Calving from Video

○Yusuke Okimoto[1], Kazuma Sugawara[1], Susumu Saito[1], Teppei Nakano[1,2], Makoto Akabane[1,2], Tetsunori Kobayashi[1], Tetsuji Ogawa[1] （1. Waseda University, 2. Intelligence Framework Lab）

4:00 PM - 4:20 PM

[2G3-OS-10c-04] A Study on Controlling Body Water Content of Tomato Using Image Processing Technology

○Shuto Namba[1], Junpei Tsuji[1], Masato Noto[1] （1. Kanagawa University）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

動画像からの豚体の頭部方向識別に用いる特徴量

Features values for recognition of swine's head direction from video

佐藤 拓弥[*1] Takuya Sato　味藤 未冴来[*1] Misaki Mito　川岸 卓司[*1] Takuji Kawagishi　水谷 孝一[*1,2] Koichi Mizutani

善甫 啓一[*1,2] Keiichi Zempo　若槻 尚斗[*1,2] Naoto Wakatsuki　竹前 喜洋[*3] Nobuhiro Takemae　西藤 岳彦[*3] Takehiko Saito

[*1] 筑波大学 大学院システム情報工学研究科 Graduate School of Systems and Information Engineering, Univ. Tsukuba

[*2] 筑波大学 システム情報系 Faculty of Engineering, Information and Systems, Univ. Tsukuba

[*3] 農研機構 動物衛生研究部門 National Institute of Animal Health, NARO

Porcine respiratory disease sometime causes the economic damage to swine farmer due to delay of finding and measures. In order to detect infected swine early, farmers observe swine's feeding and drinking movement constantly, therefore automatically observation is useful to reduce the load of farmers. We tried to automatically extract feeding and drinking movement from the video. However, feeding was mis-detected when swine's hip touched feeding area, thus discriminating between head and hip was required. In this paper, we tried to find out features to classify swine images into head or hip, and we evaluated accuracy of classification using the swine area ratio and contour of swine as the features. We performed classification by linear classifier and cross validation with these features. As a result, accuracy of classification was 96.7%, and feature about contour had a different tendency between head and hip.

1. はじめに

豚呼吸器感染症は豚呼吸器複合病の要因の一つとされ、農場では症状が発見されると全頭処分するなど多大な経済的損失を与えるため早期発見が重要である。豚呼吸器感染症の症状指標として元気消失、食欲不振、発熱などが報告されており、対策として作業者による豚の行動観察や血液検査が行われている[小澤 09]。しかし、常時観察することは困難であり、労力の軽減のためには、動画像からの自動的な行動監視が必要である。豚呼吸器感染症に感染した際、豚は食餌及び飲水行動の頻度が低下すると報告されている[Cornou 10]。そこで筆者らは、特に感染のみによる行動変化を確認するため、(国研)農研機構・動物衛生研究部門にて 16 日間の感染実験を行い、収録した動画像を用いて感染した豚の探餌及び飲水行動の判別を図った[佐藤 18]。しかし、動画像から豚の行動を推定する際、豚房内の豚を抽出しトラッキングすることは可能であったが、各行動の判別時に、豚の尾部が餌場に接触した場合に探餌行動と誤判定するため豚の前後の識別が必要であることが分かった。

そこで本稿では、動画像より取得した豚画像から頭部領域、尾部領域を自動抽出して分類するための特徴量の検討と分類精度の検証を目的とした。

2. 豚画像からの前後識別に用いる特徴量の抽出

2.1 頭部領域・尾部領域の抽出

図 1 に、画像から豚の頭部領域、尾部領域を抽出する手順を示す。収録した動画像から取得した元画像に対し、異なる時間における豚の位置が異なる画像を合成する。これにより豚のいない背景画像を作成する。元画像から背景画像差し引くことで背景差分を得る。背景差分を二値化し、ノイズ除去のために median filter を適用する。適用後の画像に対し、豚領域を囲む最小の長方形領域を抽出する。抽出した長方形領域において、おおむね頭部及び尾部は長方形領域の四隅に偏り、長方形領域の短辺 L に対し 2L/3 の幅の領域内に収まる。このため、長方形領域の四隅の画像を 1 辺 2L/3 とする正方形領域で抽出する。抽出された頭部、尾部それぞれ 2 つずつについて、豚の写っている面積の大きい方を選択し、頭部領域、尾部領域として抽出する。

2.2 前後識別に用いる特徴量

図 1 に豚の前後識別に用いる特徴量を取得する手順を示す。図 1(a)のように 2.1 節に示す処理によって得られた頭部領域、尾部領域の画像を用いて、豚の頭部及び尾部の判定するに用いる特徴量を抽出する。特徴量としては、写っている豚の(1) 面積に基づく特徴量及び(2) 輪郭に基づく特徴量とした。

(1) 豚の面積に基づく特徴量

豚の頭部の面積は尾部の面積に比べて小さいため、抽出された頭部領域と尾部領域では、頭部領域の豚領域の面積が小さいと考えられる。そこで、まず、図 2(a) の抽出された頭部領域、尾部領域の画像を二値化し、図 2(b) の二値画像を得る。次に、得られた二値画像における豚領域の画素数を二値画像全体の画素数で割ったものを特徴量とした。

(2) 豚の輪郭に基づく特徴量

豚の頭部は尾部に比べて耳や鼻などのために尖った形状を含むため、輪郭線が長いと予想される。そこで、図 2(b) の二値画像に対してエッジ抽出を施し、図 2(c) の輪郭画像を得る。得られた輪郭の画素数を画像全体の画素数で割ったものを特徴量とした。

連絡先:佐藤拓弥, 筑波大学 大学院システム情報工学研究科, 〒305-8573 茨城県つくば市天王台 1－1－1, mizutani@iit.tsukuba.ac.jp

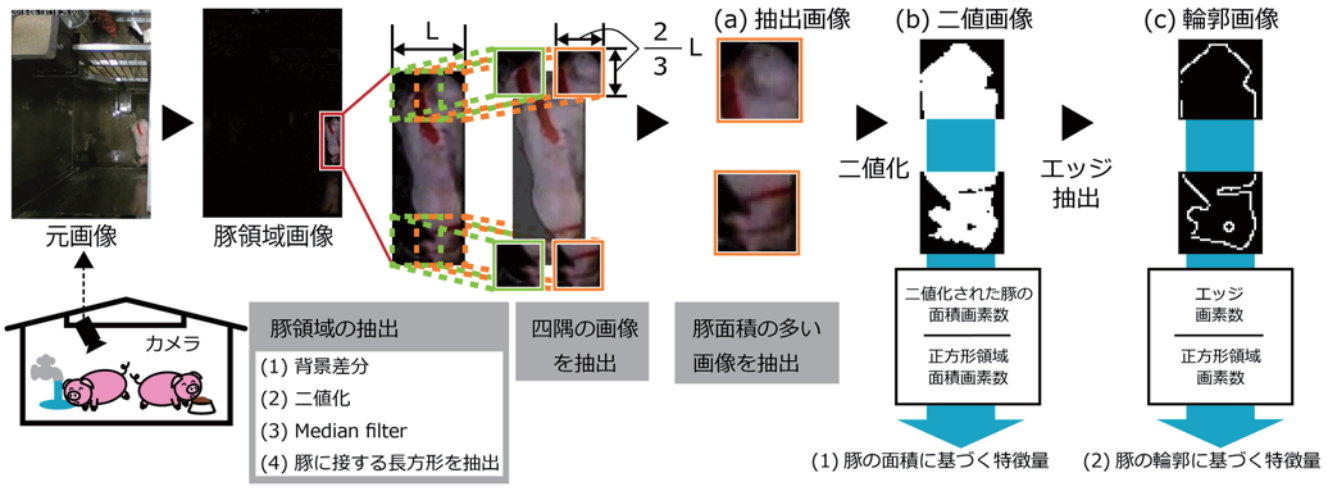


図 1　提案法による特徴量の抽出手順

3. 実験・結果・考察

2016 年 9 月 19 日から 10 月 4 日までの 16 日間, (国研) 農研機構 動物衛生研究部門 越境性感染症研究領域内の実験豚舎にて感染実験を行った。実験では 5 週齢の離乳豚に対し, 豚呼吸器感染症のひとつである豚インフルエンザへ感染させ, 動画像の収録を行った。動画像は Web カメラ(BAFFALO, BSW20KM11)を用いて 6 fps で収録した。図 1 の画像のように豚の姿勢が画面下向きである場合の 60 s の区間 360 フレーム対して 2.2 節に示す (1)豚の面積に基づく特徴量, (2)豚の輪郭に基づく特徴量の抽出を行い、線形分類器を作成し交差検定により頭部、尾部の分類精度を検証した。

図 2 特徴量の分布と分類結果、表 1 に分割表を示す。図 2 より頭部領域、尾部領域の画像は (2)豚の輪郭に基づく特徴量が異なる傾向にあった。また、表 1 より、頭部尾部を正しく分類できた正解率は 96.7 % であった。したがって、頭部尾部の分類には (2)豚の輪郭に基づく特徴量の寄与が大きいと考えられる。また、誤判定の生じた画像については、画像の輝度が低い場合、輝度による二値化がうまくできず、豚領域を抽出できない例があった。このため、高精度化のためには HOG 特徴[大西 08]やキーフレーム拘束[大田 98]を利用することで、豚領域を適切に認識・追従し、輝度の低さによる隠れが起きてもその間の動きを推定し、豚領域を抽出できる可能性がある。

4. 結論

本稿では、動画像から取得した豚画像から頭部領域と尾部領域を抽出し、頭部尾部の自動判別による前後識別を目的とした。抽出した頭部領域、尾部領域の画像を用いて (1)豚の面積に基づく特徴量および (2)豚の輪郭に基づく特徴量を抽出し、線形分類器による分類精度の検証を行った。その結果、正解率 96.7 % で分類することができた。頭部領域の画像と尾部領域の画像では (2)輪郭に基づく特徴量が異なる傾向であった。よって、頭部尾部の分類には(2)輪郭に基づく特徴量の寄与が大きいと考えられる。今後の課題として、豚領域を適切に認識・追従を行い、適切に豚領域を抽出することなどが挙げられる。

謝辞

本研究は JSPS 科研費 16H05008 の助成を受けた。

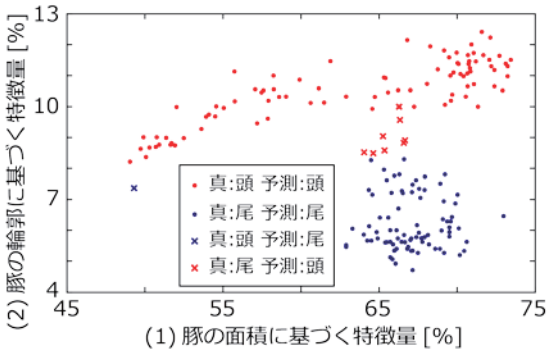


図 2　特徴量の分布と分類結果

表 1　分割表

		真のラベル	
		頭部	尾部
分類結果	頭部	342	18
	尾部	6	354

参考文献

[Cornou 10] Cornou, Cécile, and Søren Lundbye-Christensen: Classification of sows' activity types from acceleration patterns using univariate and multivariate models, Computers and Electronics in Agriculture, Vol.72, Issue.2, pp.53-60, 2010

[大西 08] 大西克則, 滝口哲也, 有木康雄: HOG 特徴に基づく単眼画像からの人体 3 次元姿勢推定, 画像の認識 理解シンポジウム, MIRU2008, pp.960-965, 2008

[大田 98] 大田佳人, 山際貴志, 山本正信, 電気情報通信学会論文誌, pp.2008-2018, 1998

[小澤 09] 小澤義博: 豚インフルエンザとその対応策, 獣医疫学雑誌, Vol.13, No. 2, pp.121-124 , 2009

[佐藤 18] 佐藤拓弥, 味藤未冴来, 川岸卓司, 水谷孝一, 善甫啓一,若槻尚斗, 竹前喜洋,西藤岳彦: 動画像を用いた豚の飲水及び食餌行動のモニタリングに関する基礎的検討, 農業施設学会 2018 年学生・若手研究発表会講演要旨集, P-13, p.9, 2018.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

大豆の生育情報を自動取得する画像センシング手法の開発: Single Shot MultiBox Detector の導入

Development of Image Sensing Method to Automatically Obtain Soybean Growth Condition: Introduction of Single Shot MultiBox Detector

大村 和暉 *1 Kazuki Omura　八幡 壮 *1 So Yahata　小澤 誠一 *1 Seiichi Ozawa　大川 剛直 *1 Takenao Ohkawa　村上 則幸 *2 Noriyuki Murakami　辻 博之 *2 Hiroyuki Tsuji

*1神戸大学 Kobe University　*2北海道農業研究センター Hokkaido National Agricultural Research Center

Observing growth state automatically is a very important task in smart farming. In this paper, we propose an image sensing method to detect soy flowers and seeds by using a state-of-art deep learning architecture called Single Shot MultiBox Detector (SSD). We also suggest a new method for counting seeds from a video. We used SSD, fast to compute and has high accuracy. We collected images at Hokkaido Agricultural Research Center from 2015 to 2017. Seed and flower detection performed 0.586 (F-measure) and 0.646(F-measure), respectively.

1. はじめに

近年の国内農業では，農業従事者の減少と高齢化が問題になっている．その中で農家が蓄積してきた農業技術や知識が失われることが懸念されており，生産力向上のために熟練者のノウハウを継承させることが課題となっている．

そこで，農業に ICT を導入して自動化・効率化を図った「スマート農業」が注目を浴びている [E-Agriculture]．スマート農業では，農作物の生育情報や環境情報をセンサやカメラなどから取得し，適切な栽培条件を見つけ出して自動制御することで，収量の増加と農作業の効率化を目的としている．これにより生産者の技術や経験に依存しない生産活動が期待されている．その中で，花や子実の検出データは農作物の収量につながる重要な情報であり，スマート農業に有用であるといえる．

本論文では，生育情報として大豆の花，子実に着目し，屋外圃場で得られた画像からこれらを認識するシステムを提案する．2016 年度以前では静止画で株を撮影していたのに対し，2017 年度のデータセットでは動画で撮影を行っていた．以前のシステムでは処理に時間がかかることが課題であり，動画に適用することが困難だった．しかし，今後動画に対して適用可能な検知システムを構築するためにも，高速な処理が必要となる．そこで，Convolutional Neural Network (CNN)[Krizhevsky 12] の一種である Single Shot MultiBox Detector (SSD)[Liu 15] を利用した．本論文では，この SSD を使用した高速な検出が可能なフェノタイピングシステムを提案する．

本論文では，2 節で子実検知について，3 節で花の検知手法について述べる．4 節では実際の大豆画像に対して検知性能の評価を行い，5 節でまとめを述べる．

2. 子実の検知

大豆の子実は緑色をしているため，蜜柑やトマトなどと異なり色相情報のみで葉や茎と区別するのは困難である．また，部分的な形状も茎や葉と似通っており，勾配情報でも抽出しにくい．そこで，茎や葉との位置関係や背景情報などの情報を総合的に捉えて物体検出することが必要である．SSD は，画像全体に畳み込み処理を施して特徴マップを生成し，マップ上で物体の検知を行う手法である．よって，特徴マップ内には物体およびその周辺から抽出した情報が含まれている．そのため，全体的な特徴量を捉えながらの検知が可能である．

連絡先: 小澤 誠一，神戸大学数理・データサイエンスセンター，〒 657-8501 神戸市灘区六甲台町 1-1，TEL: 078-803-6466，E-MAIL: ozawasei@kobe-u.ac.jp

2.1 SSD による子実検知

深層学習モデルの一つである SSD は，画像内の物体の位置検出とクラス分類を同時に高速に行うことができるのが特徴である．この SSD を本研究で導入する主な理由は，まず，子実の周囲にある物体や背景の情報を取り入れつつ子実の検知を行えるため，あきらかに子実が存在しない場所での誤検知の抑制が期待される点が挙げられる．また，圃場で花数や子実数のリアルタイム計測が可能となることが望ましいため，動画に対して高速に物体を検知・分類できる SSD が有望であると考えたからである．

SSD では，図 1 のように画像全体を一定サイズの正方形にリサイズしてから，ベースネットワークに入力する．ベースネットワークは，事前に一般物体認識データセットで学習させた CNN であり，SSD の原論文に従って VGG16[Simonyan 14] を使用している．ベースネットワークの中間層における特徴マップを取り出し，これを 1 つ目の特徴マップとする．それをさらに畳み込んでいくことで，大きさが異なる複数の特徴マップを生成する．これらのマップ上で物体の探索と分類を行う．

続いて，特徴マップ上での位置検出と分類の方法について述べる．旧来のように原画像全体を移動窓で探索する方法では，アスペクト比とサイズを変えながらの探索が必要であるため計算量が非常に大きくなる．そこで SSD では，特徴マップ上で移動窓を使用することで，原画像上の場合よりも大幅に探索回数を減らして高速化している．例えば 4×4 の特徴マップの上で，図 2(b) のように，あらかじめ設定しておいたサイズ・アスペクト比の移動窓で 1 ピクセルごとに探索していくとする．これは，入力画像を 4×4 のグリッドに分割して，各グリッドを中心に図 2(a) のように探索していくのと同じことになる．よって，大きな特徴マップにおいては小さなグリッドに区切って小物体検知を，小さな特徴マップにおいては大きなグリッドで大物体検知をすることが可能である．したがって，図 3 のように複数の特徴マップを用いて探索をしていくことで，マルチスケールの物体検知が可能になる．移動窓のサイズを調整し，小物体を検出できるように調整した．

さらに，子実検知においては性能を高めるためにファインチューニングを行っている．ファインチューニングとは，あら

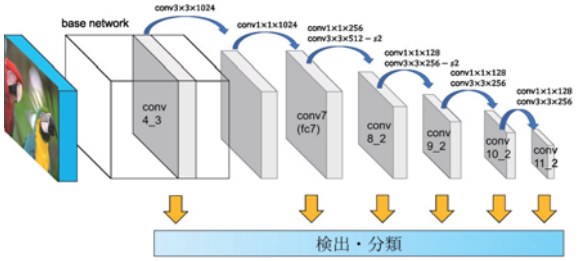


図 1: SSD のアーキテクチャ[Liu 15]

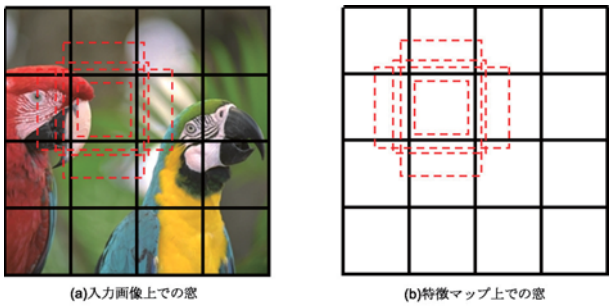


図 2: 特徴マップ上での探索 [Liu 15]

かじめ別のデータセットで学習しておいたモデルを用意しておき，そのモデルのパラメータを初期値として，本来学習したいデータセットで再学習させることである．一般的には，別のデータセットには大量の画像を含むデータセットを使用する．ファインチューニングによって，学習データが不十分でも十分な性能を発揮できる場合がある．また，学習時間を短縮する効果がある．ベースネットワークのパラメータは初期値のまま固定し，図 1 の conv4_3 以降のパラメータをファインチューニングした．

3. 花の検知

花検知においても，子実検知と同様に SSD を用いる．2.1 節でも述べたように，SSD はマルチスケールの物体検知が可能であり，小物体の探索は大きな特徴マップ上で行われている．花は子実に比べて小さな物体であることから，大きな特徴マップがあれば花検知に対してより効果的であるといえる．

そこで，花検知では SSD1024 を実装した．SSD1024 とは，入力画像のサイズを 1024 × 1024 にした SSD である．単に入力サイズを拡大しただけでなく，サイズの変更に合わせてモデルの構造も微調整し，既存手法よりスケールの大きな特徴マップを得られるようになった．表 1 に入力サイズと特徴マップのサイズを示す．表の通り，特徴マップが拡大されていることが分かる．

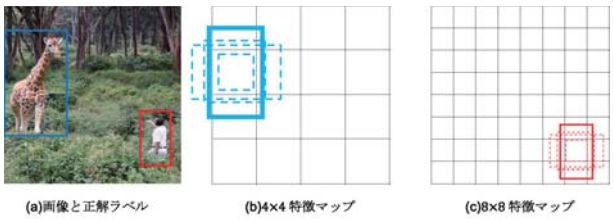


図 3: 複数マップを用いたマルチスケールな位置検出 [Liu 15]

表 1: 各特徴マップにおける 1 辺あたりのピクセル数 (一部のみ抜粋)

画像サイズ	conv4_3	conv7	conv8_2	conv9_2
300	38	19	9	5
512	64	32	16	8
1024	128	64	32	16



図 4: SSD と CNN による 2 段階検知

花検知では，クラス分類で出力されるスコアのしきい値を上げればより花らしい領域のみを残すことができるが，花の見落としも増えやすくなる．逆にしきい値を下げると，花の見落としは減るものの誤検知が増える危険がある．そこで，図 4 のように，SSD と CNN による 2 段階の手順で花検知を行う．つまり，まず SSD のしきい値を低めに設定して花の検知を行う．続いて，SSD が検出した矩形領域を覆うような正方形で画像を切り抜き，CNN に入力して本当に花かどうか精査する．花でないと判定された領域は棄却し，残った領域のみを花領域とする．以上の 2 段階の手順によって，見落としと誤検知の両方を減らすようにしている．今回は精査用の CNN として，学習が早く性能も高いことから Inception[Szegedy 15] に Batch Normalization[Ioffe 15] を追加したモデルを使用した．

4. 性能評価

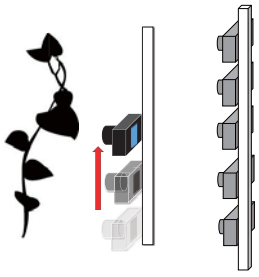
4.1 画像の撮影環境

本研究では、北海道農業研究センターの協力のもと，芽室研究拠点の試験圃場内で撮影された大豆の画像を用いた．簡易な撮影を目指すために撮影手法の改良を進めており，年度ごとの撮影方法は 2 のようになっている．撮影期間は 2015 年から 2017 年の間である．なお，2017 年については動画から一定のフレーム間隔で静止画を取得しており，それを撮影枚数としている．2015 年と 2017 年は，図 5(a) のようにカメラを下から上に等速で上昇させながら撮影を行った．2015 年は一定間隔での静止画撮影，2017 年は動画モードによる撮影である．2016 年は，図 5(b) のようにカメラ 5 台を垂直方向に並べ，リモコンで同時撮影することで連続画像を取得した．

試験圃場内の複数ブロックで撮影が行われ，一部のブロックでは多収阻害要因を特定するため，生育初期に除草剤が意図的に多く散布されている．

表 2: 大豆画像の撮影方法

撮影年度	撮影方式	カメラ台数	撮影枚数 (枚/株)
2015	静止画	1	31
2016	静止画	5	5
2017	動画	1	7



(a) (b)

図 5: 撮影装置概念図．(a)2015・2017 年，(b)2016 年

表 3: 子実検知の性能評価

モデル	適合率	再現率	F 値
SSD300	0.602	0.571	0.586
SSD512	0.454	0.491	0.472

4.2 子実検知結果

子実の検知は，北海道芽室町のテスト用区画において，2017 年 8 月に撮影された 231 枚の画像を評価に使用した．訓練画像は，芽室において 2015 年および 2016 年 8 月，2017 年 8 月 (テスト用区画以外) に撮影された合計 636 枚を使用した．本実験では，SSD300 と 512 を使用した．どちらも一般物体認識データセット 21493 枚で学習させたパラメータを初期値としてファインチューニングを行った．事前にラベル付けした子実領域と，検知した領域のオーバーラップの割合がしきい値以上であれば True Positive(検知成功) とする．また，すでに検知された子実領域を他の検知結果が検知したとしても，それは False Positive(検知失敗) とする．

性能の指標として，テスト区画全体にシステムを適用したときの再現率，適合率，F 値，ROC 曲線を求めた．ROC 曲線とは，横軸に False Positive の割合，縦軸に True Positive の割合をとり，識別器のしきい値を変えた時のそれぞれの割合をプロットすることで描いた曲線である．曲線下部の面積が大きければ大きいほど高性能であるといえる．

評価結果は表 3 のようになった．表より，SSD300 の方が高性能であると分かる．SSD512 では物体探索時の移動窓を小さめに設定しており，それが子実の大きさに対してふさわしいパラメータではなかったことが原因であると考えられる．また，SSD の適用例を図 6 に示す．画像を見ても，SSD512 の方が検出している領域が小さく，子実を捉えきれていない場合が多い一方で，SSD300 は子実に近い大きさの矩形で検出していることが分かる．今後，パラメータを改善することで SSD をより高性能にできる可能性がある．

子実検知における SSD300 と SSD512 の ROC 曲線は図 7 のようになった．この図を見ても SSD300 の方が曲線が上方向に広く，SSD300 が優れていることがわかる．

4.3 花検知結果

花の検知は，北海道芽室町のテスト用区画において，2017 年 7 月に撮影された 74 枚の画像を評価に使用した．訓練画像は，芽室において 2015 年および 2016 年の 7〜8 月，2017 年 7 月 (テスト用区画以外) に撮影された合計 2323 枚を使用した．なお，テスト区画は子実検知の場合と同じである．本実験では，SSD512 と SSD1024 を使用した．SSD512 は，入力画像を拡大することによって性能が向上するかどうかを比較するために使用している．SSD512 は子実検知と同じ事前学習済みモデル

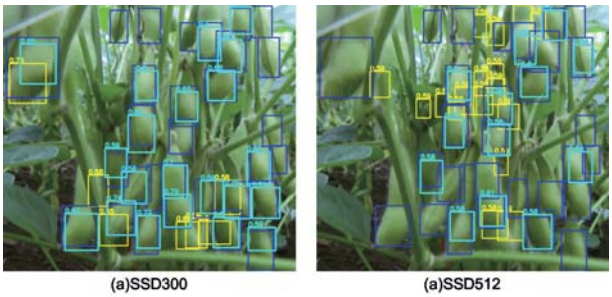


図 6: 子実の検知結果 (青色：子実ラベル，水色：True Positive，黄色：False Positive)

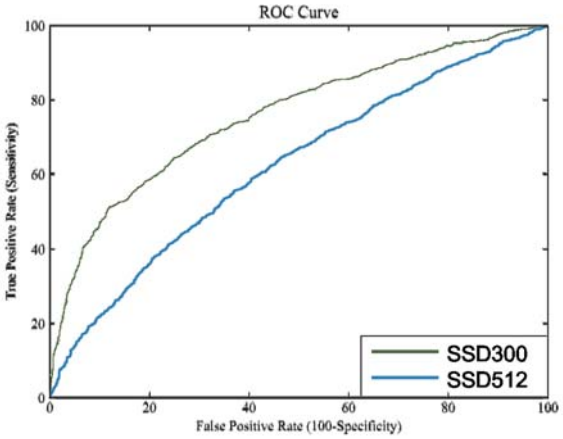


図 7: 子実検知の ROC 曲線

を用いてファインチューニングしているが，SSD1024 の方は公開されているライブラリに含まれておらず，事前学習済みモデルが存在しない．そこで，ベースネットワークのみを SSD512 の事前学習済みモデルと同じパラメータで固定し，それ以外は xavier 初期化 [Glorot 10] によって初期値をとってファインチューニングなしで学習させた．精査用の Inception モデルには事前学習済みモデルが存在するので，そのパラメータを初期値とし，2015 年と 2016 年の花画像を切り取ったデータセットを用いてファインチューニングを行っている．SSD512，SSD1024 のどちらの場合も，検出結果の精査に用いた CNN はこの同じモデルである．

性能評価では，子実検知と同様に再現率，適合率，F 値，ROC 曲線を用いた．しかし，花検知において問題となるのは，開花途中，枯れかけであったり，ボケが極めて強いなど，ラベル付けの判断が困難な花が多く存在することである．そこで，撮影条件によってラベル付けに検知の重要度を考慮することにする．鮮明に写っており，かつ満開の花は検知するべきなので「検知必須」とし，判断がしにくいものについては検知の成否を問わず「検知任意」とした．検知の成否判定は子実と同様に，ラベルとのオーバーラップの割合で判定する．また，再現率と適合率の計算式は，ラベル付けを変えたことに合わせて次のように変更している．

$$\text{再現率} = \frac{\text{検知された検知必須の花数}}{\text{検知必須の花数}}$$

$$\text{適合率} = \frac{\text{検知必須または検知任意を検知した数}}{\text{全検知数}}$$

評価結果は表 4 のようになった．表より，SSD1024 の方が

表 4: 花検知の性能評価

モデル	適合率	再現率	F 値
SSD512	0.155	0.571	0.244
SSD1024	0.585	0.722	0.646



図 8: 花の検知結果 (赤色：検知必須，青色：検知任意，紫色：検知した領域)

高性能であると分かる．また，SSD の適用例を図 8 に示す．SSD512 は誤検知が多いのに対し，SSD1024 は誤検知が少ない上に，花領域をより密接に捉えていることが分かる．

ROC 曲線は図 9 のようになった．これは，CNN による精査なしで SSD 単体の出力の検知結果を用いて描いている．SSD512 よりも SSD1024 の方がわずかに曲線下部の面積が大きく，入力画像を拡大したことによって性能が向上したことが確認できる．

5. さいごに

本研究では，大豆の生育情報を自動観測する画像センシングシステムとして，SSD をベースとした子実と花の検知システムを提案した．子実検知，花検知では，それぞれ 0.586 (F 値)，0.646(F 値) という性能を示した．また，処理の高速化によって，今後動画にも適用していくことが可能になった．

今後さらに検知性能を高めていくとともに，動画に対しての検知を進めてより詳細な生育情報を取得し，データ解析によって人手では得られなかったノウハウの取得を目指したい．

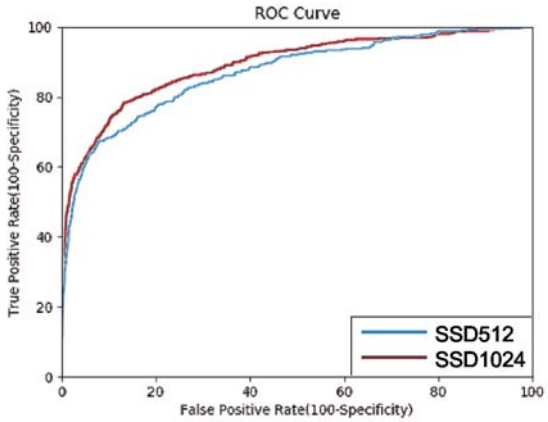


図 9: 花検知の ROC 曲線

謝辞

本研究は、平成 27 年度農林水産省委託プロジェクト研究「多収阻害要因の診断法及び対策技術の開発」の助成を得て行われた．

参考文献

[E-Agriculture] E-Agriculture, : Agricultural Value Chains and ICT, http://e-agriculture.org/agricultural-value-chains-and-ict

[Glorot 10] Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, *In Proc. of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Vol. 9, pp. 249–256 (2010)

[Ioffe 15] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *In Proc. of the 32Nd International Conference on International Conference on Machine Learning*, Vol. 37, pp. 448–456 (2015)

[Krizhevsky 12] Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *In Proc. of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097–1105 (2012)

[Liu 15] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C.-Y., and Berg, A. C.: SSD: Single Shot MultiBox Detector., *CoRR*, Vol. abs/1512.02325, (2015)

[Simonyan 14] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *CoRR*, Vol. abs/1409.1556, (2014)

[Szegedy 15] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, *In Proc. of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9 (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

牛の分娩予兆として映像から観測可能な状態の検知

Detection of Cattle States Relevant to Calving from Video

沖本祐典 *1 Yusuke OKIMOTO
菅原一真 *1 Kazuma SUGAWARA
斎藤奨 *1 Susumu SAITO
中野鐵兵 *1*2 Teppei NAKANO
赤羽誠 *1*2 Makoto AKABANE
小林哲則 *1 Tetsunori KOBAYASHI
小川哲司 *1 Tetsuji OGAWA

*1早稲田大学 Waseda University
*2知能フレームワーク研究所 Intelligent Framework Lab

We investigated the capability of general neural-network-based object detection and recognition techniques to automatically detecting in video characteristic states of cows which are relevant signs of calving. To prevent fatal accidents during calving, it is desirable to detect calving signs in video and image information from camera. End-to-end estimation of calving signs from video is not realistic without large-scale datasets. This study therefore attempts to detect from videos characteristic states that are observable as calving signs (e.g., standing/lying, tail rising, and protrusion of the allanto and amnion). For that purpose, we built a cow monitoring dataset by crowdsourcing, and trained and evaluated state detection systems based on convolutional neural networks on the developed dataset. Experimental comparisons demonstrated that cow standing and allanto and amnion detections perform well and that these features contribute to calving sign detection from video.

1. はじめに

分娩時における牛の死亡事故は，畜産農家にとって最も回避したい事故である．実際，子牛の死亡原因の 35%は難産に起因しており [Mee 13]，事故を防止するためにも畜産農家が分娩の介助を行うことが望ましい．そのため，分娩の予兆を通知する装置が多く製品化されている [Saint-Dizier 15]．これらの装置では，接触型のセンサを用いて分娩時の牛の特徴的な変化を検出している．例えば，膣内の温度低下を温度センサで検出するもの，子宮と腹部の収縮を腹部に装着したハーネスで検出するもの，尾の挙上回数の増加を加速度センサで検出するものなどがある．しかし，このような接触型センサは一般的に高価であり，接触型であるが故に牛に負担を与え，取り付けの際は農家にとって危険が伴う．そのため，畜産農家と牛の双方に負担を掛けない分娩予兆検知の手段として，非接触型センサであるカメラから取得した映像情報を用いることが望まれている．

牛の状態や行動パターンは牛の個体差や周辺環境の影響により，分娩予兆，通常時に依らず極めて多様であるため，牛の観測映像から分娩予兆かどうかを直接推定するシステムを構築するには，多様な変動を含む，整備された大規模なデータセットが必要となる．しかし，一般に利用可能な牛映像のデータセットは存在しない．さらに注意すべきは，分娩前に特徴的な変化が見られるような牛の行動 (例えば，起立と臥床など) の多くは，分娩前でなくても観測されることである．これは，分娩前数時間の映像データを用いて分娩予兆を予測するモデルを構築しても，その信頼性は低いことを示唆している．そこで，本研究では，映像情報から分娩予兆を直接推定するのではなく，映像情報から観測可能な分娩予兆に寄与する情報を抽出し，その上で分娩予兆のモデル化を行うアプローチについて検討を行う．

分娩予兆に関する臨床的，行動学的な特徴は，畜産学において詳しく調査されてきた [Saint-Dizier 15, Jensen 12, Burfeind 11]．例えば，臨床的な分娩予兆として，肉牛と乳牛とともに，分娩の 48 時間以内に膣内の温度が 0.6 度から 0.7 度降下することが知られている [Saint-Dizier 15, Burfeind 11]．乳牛において報告されている行動学的な分娩予兆としては，分娩前 2 時間における臥床時間の増加 [Jensen 12]，分娩前 2〜6 時間前からの起立・臥床の変化の回数の増加 [Saint-Dizier 15, Jensen 12]，分娩前 4〜6 時間前からの尾の挙上回数の増加 [Jensen 12] などが知られている．また，牛の分娩の過程では，尿膜・羊膜と呼ばれる，水風船上の組織が産道から露出する [Saint-Dizier 15]．このような変化の多くはカメラで捉えることが可能である．

そこで本研究では，映像から分娩予兆に寄与する情報の抽出を目指し，画像情報から観測可能な分娩の予兆として起立/臥床，羊膜・尿膜の露出，および尾の挙上という 3 状態に着目し，それらを畳み込みニューラルネットワーク (convolutional neural network, CNN) を用いて検出可能かどうかを調査した．CNN は物体認識・物体検出のタスクで高い性能を達成するが，その学習には整備された大規模データが必要となる．大規模な画像データセットとしては，ImageNet [Deng 09] が物体認識に Pascal VOC2007 [Everingham] および COCO [Lin 14] が物体検出に用いられているが，これらはクラウドソーシングを用いてラベル付与がなされている．本研究の対象である牛の行動に関するデータベースの構築に際しては，データの収集に加えて，クラウドソーシングに基づきラベルの付与を行うシステムの構築も必要となる．本研究を通じ，分娩予兆検知において画像情報を活用するための要件およびその効果に関する知見が得られることが期待できる．

本稿の構成は以下の通りである．2. では，分娩予兆に関係する牛の状態を説明し，それらを映像情報から検知するシステムについて述べる．3. では，クラウドソーシングを用いて構築した牛のモニタリングデータセットについて述べる．4. では，画像データを用いた実験により，開発した牛の状態検知システムの性能について述べる．最後に，5. にて本稿の結論を述べる．

連絡先: 早稲田大学基幹理工学部，東京都新宿区早稲田町 27，okimoto@pcl.cs.waseda.ac.jp

2. 分娩予兆に関わる牛の状態の検出

本章では，牛の分娩予兆とそれに関わる牛の状態について述べた後，それらの状態を動画像から検出するシステムについて述べる．

2.1 分娩予兆とそれに関わる牛の状態

本研究では次の 3 つの分娩予兆と，それぞれに関わる牛の状態に注目する．

- **臥床時間の増加**: 牛の臥床時間は分娩前 2 時間に増加する [Jensen 12]．臥床時間は牛の起立状態を検出することで算出可能となる．
- **尿膜・羊膜の露出**: 水風船状の組織である尿膜と羊膜は，分娩の過程に産道から露出する [Saint-Dizier 15]．一般的に，尿膜は分娩 3 時間から 7 時間前に露出し，また尿膜は分娩 2 時間半から 5 時間前に露出する．尿膜と羊膜は，前者が黒っぽく，後者が白っぽいという色の違いを除くと，外見は似ている．本研究では，尿膜の露出と羊膜の露出を一つの状態として扱う．
- **尾の拳上時間の増加**: 尾の拳上時間は分娩 4 時間から 6 時間前以降に増加する [Jensen 12]．尾の拳上時間は尾の拳上状態を検出することで算出可能である．

2.2 ニューラルネットワークによる牛の状態検出

図 2 に先述した分娩予兆に関わる牛の状態を検出する流れを示す．牛の領域が画像から切り抜かれた後，それらの領域がニューラルネットワークを用いた牛の状態検知器に入力される．

2.2.1 牛領域検出

牛領域の検出には，ニューラルネットワークを用いた高精度な物体検出アルゴリズム YOLOv2 [Redmon 17] を用いる．YOLOv2 の重みには，COCO Train/Val 2014 [Lin 14] で学習した既存のモデル YOLOv2 608x608** を用いる．このモデルは一般的な物体検出タスクを想定しているため，牧場という特殊な環境下において，牛の検出は再現率が低くなる．そこで今回，牛の可能性がある領域を多めに検出した後，ヒューリスティックに候補領域を削減する方法を取った．まず，10 種類の動物のクラス（cow, bear, dog, bird, elephant, horse, cat, sheep, mouse, giraffe）を牛の可能性があるとして検出し，領域の確信度 (confidence) が閾値 ($v_{\mathrm{th}}^{\mathrm{conf}}$) を超えたものを牛領域の候補とする．その次に，各画像において得られた候補領域を，以下のヒューリスティックな方法で特定の数まで削減する．2 つの候補領域 A, B が重なっていたとき ($A \cap B \neq \emptyset$)，次式で定義される intersection over union(IoU)

$$IoU(A, B) = area(A \cap B)/area(A \cup B) \qquad (1)$$

が閾値 ($v_{\mathrm{th}}^{\mathrm{IoU}}$) が超えていた場合，A，B のうち面積が小さい領域を破棄する（ここで $area(A)$ は領域 A の面積）．この操作を，候補領域の数が閾値 ($v_{\mathrm{th}}^{\mathrm{reg}}$) 以下となるまで繰り返す．もし候補領域数が閾値以下まで削減できなかった場合，その画像は訓練及び評価データとしては用いない．以上の操作により得られた領域を牛の領域として，CNN を用いた牛の状態検知システムの入力とする．

2.2.2 牛の状態検知

画像から検出した牛の候補領域ごとに，CNN を用いて，分娩予兆に関わる牛の状態を検知する．先述した分娩予兆に関わる牛の 3 つの状態ごとに，それぞれ状態検知器を構築する．状態検知器のネットワーク構造は状態ごとに同じであり，畳込み層は VGG16 [Simonyan 14] あるいは ResNet50 [He 16] と同様の構造，それに続く全結合層は 1 層 1024 ユニット出力層は，牛の分娩予兆に関わる状態の事後確率を出力する．ImageNet で学習済みの畳込み層とランダムに初期化された全結合層を，集めた牛のモニタリングデータを用いてファインチューニングした．ここで，YOLOv2 によって切り出された牛領域は 224x224 にリサイズされたのち，z-正規化を施し，CNN を用いた状態検知器に入力される．

3. データ収集

本章では，先述した 3 つの状態をクラウドソーシングを用いてアノテーションした，牛のモニタリング画像データセットについて述べる．

3.1 データ収録

分娩が近い黒毛和種雌牛を，分娩房の側面に設置したカメラで継続的に収録した．収録例を図 3 に示す．収録は 2016 年 11 月から鹿児島県の牧場で開始した．本研究では，収録された動画から 1 秒につき 1 枚ずつ画像として切り出したものを生の収録データとして扱う．

3.2 クラウドソーシングによるアノテーション

切りだされた牛領域の画像に対し，3 つの分娩予兆に関わる牛の状態のアノテーションを行った．アノテーションは，評価データを筆者らが，大規模な訓練データを Amazon Mechanical Turk のクラウドワーカーがそれぞれ行った．クラウドワーカーに対する質問は次のとおりである．

- 牛は起立状態か．
- 牛から尿膜か羊膜が露出しているのが見えるか．
- 牛は尾を拳上しているか．

また，上記の質問それぞれに対し，次の 4 つの選択肢を用意した．

- **はい**: 質問は正しい (例：牛は起立している)
- **いいえ**: 質問は正しくない (例: 牛は臥床している)
- **わからない**: 判断を下すのが困難
- **切り抜きエラー**: そもそも判断を下せない

回答の品質保証のため，一枚の画像に対して一つの状態のみをアノテーションする単純なタスク設計とした．タスク画面には，牛領域と，それを含む分娩房全体をワーカーに提示されている．わからない以外のラベルについての画像例も表示されており，必要な場合には別ページでさらに多くの画像例を確認できる．それぞれのタスクにつき 2 人分の回答を集め，両方のラベルが一致した画像のみを学習に用いた．

**https://pjreddie.com/darknet/yolo


(a) 起立状態


(b) 臥床状態


(c) 尿膜の露出状態


(d) 羊膜の露出状態


(e) 尾の挙上状態

図 1: 分娩予兆に関係する牛の状態の例

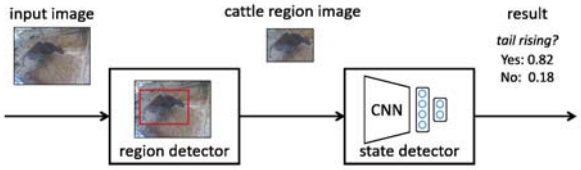


図 2: 分娩に寄与する状態検出の流れ．画像より牛領域を切り出し，状態検知システムの入力とする．



図 3: 収録された画像の例 (2 頭の牛が分娩房の斜め上方から収録される)．

4. 牛の状態の検知実験

4.1 使用データ

本実験では，独自に収集した牛のモニタリング画像データセットのうち 2017 年の 3–8 月収録分を用いた．ここには計 18 回の分娩シーンが含まれている．表 1 に，第 2.1 節で述べた分娩に関わる牛の状態ごとの，牛領域の画像の数を示す．これは，分娩 5 時間前から分娩時までのすべての画像データから 6 秒毎に 1 枚ずつ選び，それらの画像から牛領域を切り出し，アノテーションした結果である．分娩予兆に関わる状態についての訓練・検証データは 8 回の出産シーンを含み，評価データは訓練・検証データと重複しない 4 回の出産シーンを含むように選択した．さらに，データ数が多いクラスのデータの一部を無作為に削除することで，訓練データと評価データそれぞれの正例・負例の数が等しくなるようにした．

4.2 実験設定

牛領域の検出に関して，先述した検出におけるパラメータは経験的に決定し，$v_{\mathrm{th}}^{\mathrm{conf}}=0.1$，$v_{\mathrm{th}}^{\mathrm{IoU}}=0.7$，$v_{\mathrm{th}}^{\mathrm{reg}}=2$ とした．また，YOLOv2 の tensorflow による実装である darkflow[*†] を用いた．状態推定器のニューラルネットの学習には，モーメンタムを 0.9，学習率を $1e-4$ とした確率的勾配降下法を用いた．学習は validation loss が二回続けて上昇した場合に早期終了した．また，学習データに対しては，画像反転，回転，並進移動，拡大のデータオーグメンテーションを行った．学習の各エポックごとにデータオーグメンテーションを施し，エポック毎の画像枚数を 20000 まで拡張した．またミニバッチサイズは 50 とした．

[*†] https://github.com/thtrieu/darkflow

表 1: 牛領域画像の枚数.

cattle state	train	validation	test
standing	10170	1130	800
allanto & amnion	7612	844	800
tail rising	5714	634	800

表 2: 牛の状態検知システムの性能．表中の数値は AUC.

network	standing	allanto & amnion	tail rising
VGG16	0.95	0.90	0.56
ResNet50	0.98	0.92	0.48

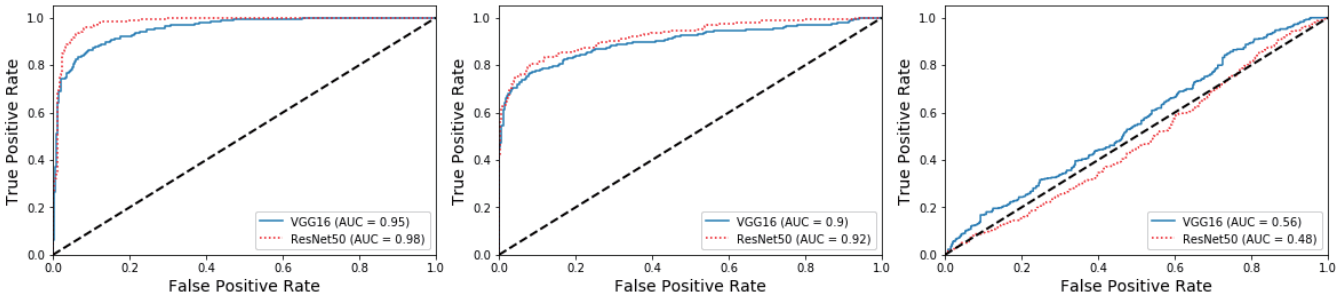
4.3 実験結果

図 4 に 3 つの分娩予兆に関わる状態をニューラルネットで検知した，それぞれの受信者操作特性 (Reciever Operation Characteristics, ROC) 曲線を示す．表 2 に ROC 曲線の下部の面積である Area Under the Curve(AUC) を示す．また図 5 は，VGG16 と同じ畳込み層の構造を持つニューラルネットワークにおいて，Grad-CAM [Selvaraju 17] を用いて得た結果を出力する上で注目された入力画像の領域を示す．

牛の起立状態と尿膜・羊膜の露出状態の AUC は 2 つの状態において 0.8 を超えている．Grad-CAM によって得られた入力画像における注目されている部位もこの結果を支持する．すなわち，起立状態の検知においては足の部分が注目されており，尿膜・羊膜の露出状態の検知においては尻の部分が注目されている．これらの結果は，起立状態および尿膜・羊膜の露出状態は動画からの分娩予兆の早期検知に有効であることを示唆する．対して，尾の挙上の検知性能は良くない．尾の挙上を検知では，尾の部分が注目されることが期待されるが，際立って注目されている部分はない．尾が挙上しているか，いないかの判断はクラウドワーカーにとって極めて難しく，尾の挙上のアノテーションが他の 2 状態に比べ信頼性が低いと考えられる事実，起立状態と尿膜・羊膜の露出状態のアノテーションの，2 人のワーカーの回答の一致率はそれぞれ 81.5%と 73.3%であった一方で，尾の挙上状態における一致率は 48.1%にとどまった．この結果は，クラウドソーシングを用いたアノテーションの設計を改良することが必要であることを示唆する．

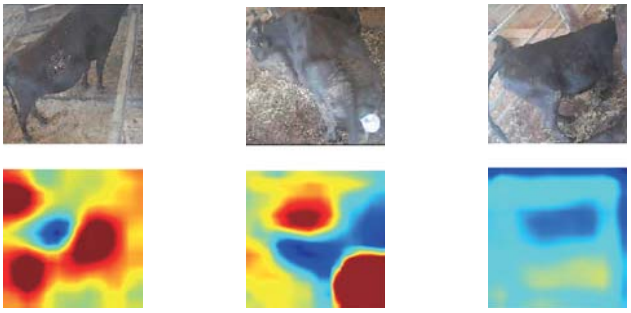
5. 結論

本研究では，画像情報から分娩予兆に関わる牛の状態を検知するシステムを開発した．CNN を用い，分娩時に観測される状態である牛の起立・尿膜及び羊膜の露出・尾の挙上の 3 状態を入力画像から検出した．ここで，それぞれの画像から切りだされた牛領域に対して，クラウドソーシングによるアノテーションシステムを開発しラベル付与を行った．これらの画像を用いた実験の結果，起立と尿膜・羊膜の露出に関しては良い性



(a) 起立状態　(b) 尿膜・羊膜の露出状態　(c) 尾の拳上状態

図 4: 3 つの牛の状態検知システムで得られた ROC 曲線.



(a) 起立状態　(b) 尿膜・羊膜の露出状態　(c) 尾の拳上状態

図 5: CNN による状態検知において顕著に利用される入力画像の領域．起立状態の検出 (a)，尿膜・羊膜の検出 (b) では，想定通り足と尻に注目があたっている．一方，尾の拳上状態の検出 (c) では，尾の部分に注目があたっていることが望ましいが，そうなっていない．

能で検出できたが，尾の拳上に関しては改善の余地があることが示された．

6. 謝辞

本研究を行うにあたり，牛のモニタリング動画を提供してくださった東村牧場様，また春日良一様 **, 高橋美博様 *, 山下博美様，田中知明様，そして有益な議論を頂いた春日久志様に感謝申し上げます．

参考文献

[Burfeind 11] Burfeind, O., Suthar, V., Voigtsberger, R., Bonk, S., and Heuwiser, W.: Validity of prepartum changes in vaginal and rectal temperature to predict calving in dairy cows, *Journal of Dairy Science*, pp. 5053–5061 (2011)

[Deng 09] Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (2009)

[Everingham] Everingham, M., Gool, L. V., Williams, C., Win, J., and Zisserman, A.: The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html

[He 16] He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)

[Jensen 12] Jensen, M.: Behaviour around the time of calving in dairy cows, *Applied Animal Behaviour Science*, Vol. 139, pp. 195–202 (2012)

[Lin 14] Lin, T., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C.: Microsoft COCO: Common Objects in Context, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 740–755 (2014)

[Mee 13] Mee, J.: Why Do So Many Calves Die on Modern Dairy Farms and What Can We Do about Calf Welfare in the Future?, *Animals*, pp. 1036–1057 (2013)

[Redmon 17] Redmon, J. and Frhadi, A.: YOLO9000: Better, Faster, Stronger, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271 (2017)

[Saint-Dizier 15] Saint-Dizier, M. and Chastant-Maillart, S.: Methods and on-farm devices to predict calving time in cattle, *The Veterinary Journal*, Vol. 205, pp. 349–356 (2015)

[Selvaraju 17] Selvaraju, R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D.: Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization, in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626 (2017)

[Simonyan 14] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv preprint arXiv:1409.1556* (2014)

** Farmer's Support, LLC.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

画像処理技術を用いたトマトの体内水分量制御手法の検討

A Study on Controlling Body Water Content of Tomato Using Image Processing Technology

難波 脩人　辻 順平　能登 正人
Shuto Namba　Junpei Tsuji　Masato Noto

神奈川大学大学院工学研究科電気電子情報工学専攻
Graduate School of Electrical, Electronics and Information Engineering, Kanagawa University

Water stress cultivation is a method that raises the sugar content of fruits by deliberately decreasing the amount of water supply from experience and intuition of skilled farmers. It is difficult for new farmers lacking experience and intuition to cultivate. Under such circumstances, we considered a water supply control method that automates the water supply work. In this study, we design a model representing plant body water content and response to irrigation using Hidden Markov Model. We investigated whether it is possible to control irrigation by using it. As a result of the experiment conducted through actual cultivation of tomatoes, It was revealed that the value obtained from the plant model designed using the Hidden Markov Model under specific conditions reflects the change in the body water content of the plant.

1. はじめに

日本の農業は，農業従業者の高齢化の進行により離農率が上昇している．農林水産省のデータによると，平成 22 年の 261 万人から平成 27 年の 210 万人と約 20%減少していることがわかる．一方で，新たに農業を開始した新規就農者人口の推移を調べると平成 22 年の 5 万 5 千人から平成 27 年の 6 万 5 千人と約 15%増加していることがわかる [*1]．このことから，熟練農家が持つ高度な栽培技術が新規就農者への継承を行う前に失われることが懸念されている．

一方で，農作物に対して高付加価値を与えることで農作物の品質を向上させる高度な栽培技術は広く行われている．その中でも水ストレス栽培は，高糖度な果実を生産し，果実の市場価値を向上させる栽培技術として知られている．しかしながら，水ストレス栽培に代表される高度な栽培技術は熟練農家の経験と勘によって行われるため，新規就農者がこれらの栽培技術を再現するには困難がともなう．新規就農者でも簡単に水ストレス栽培などの高度な栽培技術を再現するためには人工知能を用いて支援する必要がある．

そのような背景のもと，水ストレス栽培においては新規就農者への技術継承を目的とした水ストレスの程度の推定を行う研究が報告されてきた．一方で，推定された水ストレスに対して熟練農家がどのように水を与えるかの潅水制御方法は明確にされていない．そのため，熟練農家が経験と勘によって行う潅水作業を再現するための潅水制御方法を考える必要がある．

本研究の目的は水ストレス栽培に用いることが可能な自動潅水制御の仕組みを実現させることである．植物は体内水分量という内部状態をもち，体内水分量の変化に伴い水ストレスが生じる．水ストレス栽培の潅水制御を行うにあたり，体内水分量が潅水によってどの程度変化するかを明らかにしなければ潅水制御方法の確立は困難である．そのため，植物の体内水分量と潅水に対する反応を表すモデルを設計する必要があると考えられる．潅水に対する植物の反応は新規就農者への技術継承を考慮し，画像処理技術を用いて測定する．しかしながら，画像データから正確な体内水分量を測定する手法は確立されておらず，画像データから植物の内部状態を適切に設計することは潅水制御を実現させるためにも必要である．

本研究では，植物の体内水分量の変化を画像処理技術を基に制御するにあたり，植物の体内水分量と潅水に対する反応を表すモデルを設計し，潅水制御システムに用いることが可能であるかの検討を行う．

連絡先: 難波脩人，神奈川大学大学院工学研究科電気電子情報工学専攻　能登研究室，〒 221-8686 神奈川県横浜市神奈川区六角橋 3-27-1，電話：045-481-5661

*1　農林水産省：平成 28 年新規就農者調査，
http://www.maff.go.jp/j/tokei/kouhyou/sinki/index.html

2. 水ストレスの程度の推定手法

2.1 茎径変化量を用いた推定手法

植物の体内水分量を直接推定する手法 [Araki 87] は植物の葉を採取するため，継続して測定することは難しい．間接的に測定する手法としては茎径を用いて水ストレスの程度の推定を行う手法 [大石 02] が提案されている．茎径は水ストレスによって太さが変化するため，水ストレスと密接な関係があることから推定手法に用いられている．

しかしながら，茎径の測定に使用しているレーザ式変位センサは非破壊的で連続的な測定が可能であるが，センサ自体が高価なため，新規就農者にとっては導入することが困難であると考えられている．

2.2 植物の葉のしおれ具合を用いた推定手法

高価なセンサを使用せずに新規就農者でも導入が簡易なカメラを用いて水ストレスの程度の推定を行う手法として，画像処理技術の分野でも動体検出に用いられる Optical Flow を応用した研究 [Kaneda 17] が報告されている．

Optical Flow とは，追跡する対象の動きが微小変化でも検出することが可能であり，映像におけるフレーム間で画像中の輝度が変化する際の見かけの動きである．この研究では畳み込みニューラルネットワーク（CNN）を用いて水ストレスの程度の推定手法を提案している．Optical Flow によって植物の葉のしおれに関した部分を抽出し，入力画像として用いることで水ストレスの程度の推定制度を向上させている．

3. 潅水制御システムの実現に向けた植物モデルの作成

推定された水ストレスを基に潅水制御を行う技術は水ストレス栽培を新規就農者が行うための技術として必要不可欠である．さらに，新規就農者への技術継承を考慮した際，潅水制御方法を確立するためには低コストであることが望まれる．本研究では水ストレス栽培に関連した研究の中でも高価なセンサを使用せずに測定することが可能である画像処理技術を用いる．画像処理技術の中でも Optical Flow は柴田らの研究 [柴田 17] において複雑背景下でも葉のしおれの変化を検出することが可能であると報告されているため，葉の動きの検出が可能である．一方で，潅水制御を行うためには潅水量を決定するために植物の内部状態と潅水に対する植物の反応をモデル化する必要がある．

植物内部の体内水分量は植物外部の状態を決定する要因として考えられる．植物内部の体内水分量は時間変化し，現在の体内水分量は過去の体内水分量と潅水量によって決定されると考えられる．以上のことから植物のモデルは時刻 t における体内水分量 X_t とする隠れマルコフモデルとしてモデル化することが妥当であると考えられる．時刻 t における潅水量を U_t とし，植物外部に現れる植物の葉のしおれを Y_t とすると以下のように定式化できる．

$$X_t = f(X_{t-1}, U_{t-1}, w_{t-1}) \tag{1}$$

$$Y_t = g(X_t, \epsilon_t) \tag{2}$$

w_{t-1}，ϵ_t は確率的な要因を定める確率変数である．農業分野において画像データを用いる場合は植物の葉に着目することが多いため，植物の葉のしおれ Y_t と Y_t を観測する画像データ I_t の関係を考える．画像データ I_t から植物の葉のしおれ Y_t の情報を直接得ることは困難であり，I_t は Y_t の部分的な情報しか反映されていないことが考えられる．

$$I_t = Proj(Y_t) \tag{3}$$

さらに，Optical Flow は各時刻の画像データ I_{t-1} と I_t の差分を用いるため，

$$r_t = Opt(I_{t-1}, I_t) \tag{4}$$

となる．r_t の定義については次節で述べる．したがって，各時刻における X_t の情報は g，$Prog$，Opt によるフィルタを通して r_t に表されており，r_t を設計することによって X_t を推定することができると考えられる．

3.1 回復傾向指標

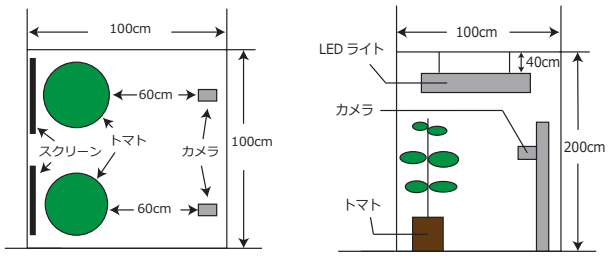
本研究では植物の葉が潅水によってどの程度回復するかの指標として r_t を設計する．この場合の r_t を回復傾向指標と呼び，回復傾向指標は時刻 t，時刻 $t+\Delta$ において Optical Flow を行い算出した変位ベクトルを用いて算出する．植物の葉が回復している状態は葉のしおれが上向きに動いている場合であると考えられる．そのため，回復傾向指標は以下のように求める．全体の変位ベクトルの集合を $V_{t,t+\Delta}$ と定義する．$V_{t,t+\Delta}$ の内，変位ベクトルの向きが上向きのものの個数，すなわち $v \in V_{t,t+\Delta}$ として v の偏角が $[0, \pi)$ であるものの個数の比を回復傾向指標 $r_{t,t+\Delta}$ と定義する．式で表すと以下のとおりとなる．

$$r_{t,t+\Delta} = \frac{\#\{v \in V_{t,t+\Delta} \mid 0 \leq \arg v < \pi\}}{\#V_{t,t+\Delta}} \tag{5}$$

$r_{t,t+\Delta}$ が植物の体内水分量の変化を表す指標として有効であると示すことで，潅水制御は $r_{t,t+\Delta}$ を用いて行うことが可能であり，熟練農家の経験と勘による作業を継承することにもつながると考えられる．

3.2 観察環境

本研究の観察環境を図 1 に示す．栽培環境は植物育成用 LED ライトと植物栽培用ハウスキットを用いた．実際，トマトの栽培には LED ライトを用いた環境においても生育が可能であることが明らかにされているため，太陽光の代わりとして使用することに問題はないと考えられる [渡邉 16]．撮影は web カメラ（Logicool 製 C270）を対象物との距離が 60cm になるように設置して行う．web カメラの撮影は図 2 に示すように植物の全体を写すように撮影した．栽培用ハウスキットの内部では LED ライトを使用したことで，Optical Flow は LED によるノイズ影響を受けることが実際の栽培から確認された．そのため，ノイズ被害を軽減することを目的とし，図 1(a) に示すように背景に黒のスクリーンを設置して撮影を行った．



(a) 俯瞰図　(b) 側面図

図 1: 植物栽培用ハウスキット　内部



図 2: 栽培キット内の撮影例

3.3 Optical Flow を用いた指標の算出方法

本研究における指標の算出までの流れを図 3 に示す．時刻 t に撮影した画像を I_t，時刻 $t+\Delta$ に撮影した画像を $I_{t+\Delta}$ とする．画像 I_t と画像 $I_{t+\Delta}$ を用いた Optical Flow の処理結果を 3.1 節で定義した $V_{t,t+\Delta}$ で表し，回復傾向指標 $r_{t,t+\Delta}$ を求める．本研究では Optical Flow は 1 日の中で 8 時から 24 時まで $\Delta = 0.5$ [時間] として処理を繰り返す．最後に回復傾向指標の値をグラフに表した．

4. 実験および結果

回復傾向指標が植物の体内水分量の変化を表す指標として有効であるかを評価するための実験を行う．まず，植物自体が持つ自然な葉の動き，すなわち潅水による影響がない場合の

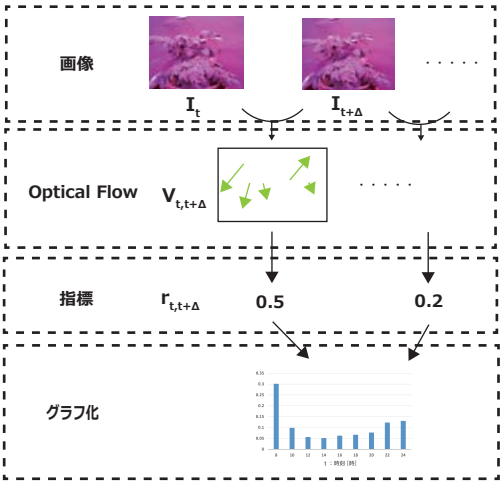


図 3: 指標の算出の流れ

植物の葉の変化を確認する．そのため，潅水を行わない日を設け，その日の回復傾向指標の値を算出した．上述した条件を参考に潅水 1 回の場合，潅水複数回の場合のパターンを用意し，実験を行った．最後に潅水量を変化させた場合でも回復傾向指標が有効であるか評価するための実験を行った．

4.1 回復傾向指標の変化（潅水なしの場合）

潅水を行うことで植物の葉のしおれの変化を確認するにあたり，どのタイミングで潅水を行うかは重要な問題である．潅水を行わない日のデータは web カメラで撮影している 8 時から 24 時に潅水を行わない場合のことを指し，図 4 に示したグラフは潅水を行わない日の回復傾向指標の値の結果である．

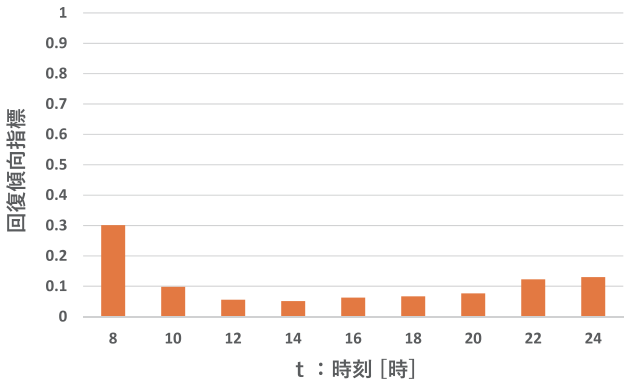


図 4: 回復傾向指標の値（潅水なし）

グラフから植物の葉は 1 日の中で 12 時に向けて回復傾向指標の値が小さくなっていることがわかる．この結果から植物の葉は 1 日の中で 12 時をピークとして葉のしおれが強くなっていることが確認できた．

4.2 回復傾向指標の変化（潅水 1 回の場合）

実験 4.1 を踏まえ，12 時に潅水を行う場合と潅水を行わない場合の回復傾向指標の値を比較することで潅水による植物の体内水分量の変化をとらえることができているかを明らかにする．12 時に 200$m\ell$ の潅水を行った場合の回復傾向指標の値と潅水なしの回復傾向指標の値をグラフに示す（図 5）．

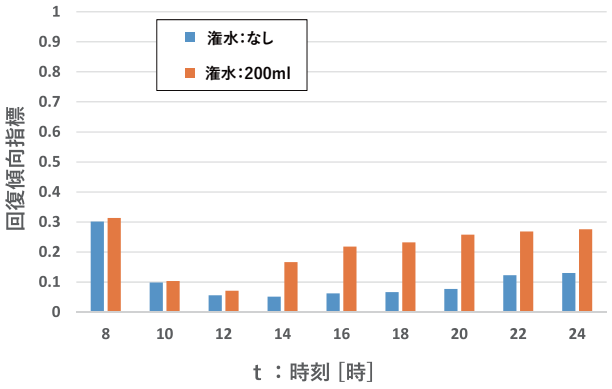


図 5: 潅水 1 回と潅水なしの回復傾向指標の値

グラフから 12 時に潅水を行った場合と行わなかった場合を比較すると潅水を行った場合は 12 時を境に回復傾向指標の値が上昇していることが確認できる．Optical Flow によって算出した変位ベクトルを基に植物内部の体内水分量の変化をとらえることは可能であることが示唆された．しかしながら，潅水は植物の葉のしおれが強いときに行ったため，植物の体内水分量が減少している状態であったことが考えられる．そのため，実験 4.2 では植物の体内水分量の変化が表れやすい状態であったことが考えられる．

4.3 回復傾向指標の変化（潅水複数回の場合）

実験 4.2 の結果を踏まえ，植物の葉のしおれが強いときに潅水を行うだけでなく，1 日の中で潅水回数を増やしたパターンにおいても植物の体内水分量の変化をとらえることが可能であるかを調べる実験を行った．今回は 1 日の中で潅水を 2 回行い，その日の結果を図 6 に示す．

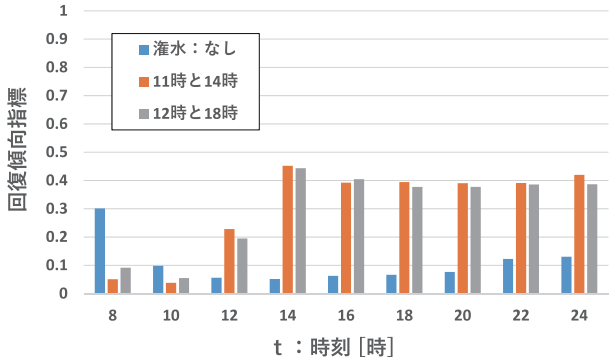


図 6: 潅水 2 回の回復傾向指標の値

潅水 1 回の場合で確認できたように 11 時と 12 時での潅水は潅水なしと比較すると植物の体内水分量の変化が回復傾向指標から確認することができる．一方で，11 時から 3 時間後の 14 時と 12 時から 6 時間後の 18 時に行った潅水は潅水なしと比較すると回復傾向指標の値に差異が生じていることは確認できるが，2 回目の潅水を 14 時に行った場合と 18 時に行った場合で比較すると回復傾向指標の値に差が生じていないことから 2 回目の潅水による植物の体内水分量の変化をとらえていないことが確認できた．このことから植物内部の体内水分量が減少している段階での潅水はその後の葉のしおれの変化に顕著に表

れることがグラフから明らかになった．一方で植物の体内水分量が 1 回目の潅水により増加していると考えられる状態においては回復傾向指標の値を算出するだけでは体内水分量の変化をとらえることは難しいと考えられる．

4.4 回復傾向指標の変化（潅水量を変化させた場合）

これまでの結果より，1 日の潅水回数を増やしたパターンについての植物の回復反応を確かめるのではなく，1 回の潅水に対して潅水量を複数設けた場合の植物の回復傾向指標の値を比較する．比較した結果より，潅水量の違いがその後の植物の体内水分量の変化に差が生じるかを明らかにする．実験では潅水量を 200$m\ell$，300$m\ell$，400$m\ell$，500$m\ell$ と 4 パターン用意し，日を分けて潅水を行った結果をグラフ（図 7）に示す．

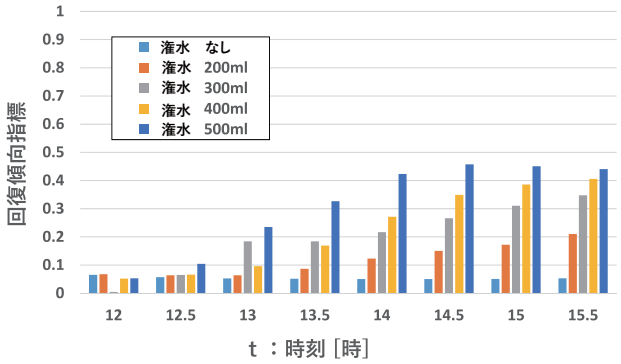


図 7: 潅水時間の違いによる回復傾向指標の値

潅水量を 4 パターン用意した結果，植物の回復傾向指標の値は潅水量が多くなるほど増加していることがわかる．このことから潅水量の変化の機微まで回復傾向指標はとらえることが可能であることが明らかになった．

4.5 考察

実験 4.1 では植物の葉のしおれは 1 日の中でどの程度の変化を示すかを明らかにするために潅水を行わない場合の回復傾向指標の値を算出した．その結果，本研究では植物の葉は 1 日の中で 12 時に葉のしおれが強くなることが確認できた．

実験 4.2 では実験 4.1 を踏まえ，12 時に潅水を行うことでその後の回復傾向指標の値が潅水なしの場合と比較し，変化があるかを確認した．その結果，潅水なしの場合と比べ，回復傾向指標の値に差が生じていることが明らかになった．このことから，回復傾向指標を求めることで植物の体内水分量の変化を明らかにする可能性を示唆できた．しかしながら，植物の葉のしおれが強い状態で潅水を行うことは植物内部の体内水分量が減少していると考えられる場合であり，潅水による体内水分量の変化が確認しやすい状態であったと考えられる．そのため，実験 4.3 では植物の体内水分量が増加していると考えられる潅水後にさらに潅水を行い，植物の葉の変化が回復傾向指標から確認できるかを明らかにした．結果として，2 回目の潅水による変化は回復傾向指標を算出するだけではとらえることが困難であることが明らかになった．

複数回の給水による植物の体内水分量の変化を回復傾向指標から算出することは困難であると判断し，実験 4.4 では 12 時に 4 通りの潅水量を設け，各潅水量がその後の葉のしおれの変化に違いが生じるかを回復傾向指標を比較することから明らかにした．その結果，回復傾向指標の値は潅水量が多くなるほど植物の葉の回復傾向が早くなることが明らかになった．

実験結果をまとめると，回復傾向指標を用いて植物内部の体内水分量の変化を確認することは体内水分量が減少していると考えられる状態においては有効であることが示せた．一方で，植物の体内水分量が増加していると考えられる状態では回復傾向指標を算出するだけでは植物内部の変化を確認することは困難であることが明らかになった．しかしながら，潅水量を複数設けた場合においては潅水量ごとに回復傾向指標の現れ方に違いが生じることが確認できた．このことから，回復傾向指標は植物内部の体内水分量が減少していると考えられる状態においては詳細な葉の変化をとらえることが可能であり，体内水分量の制御を行う指標として有効であることが明らかになった．

5. おわりに

本研究では水ストレス栽培において熟練農家が行う潅水作業を自動化するために，植物の体内水分量と潅水に対する反応をモデル化し，モデルを基に潅水制御は行えるかを検討した．

実験結果より，植物の体内水分量が減少していると考えられる状態においては潅水による植物内部の体内水分量の変化を回復傾向指標から判断することができた．また，潅水量を変更することにより回復傾向指標の現れ方にばらつきがあることから潅水制御を行う際に，潅水量のパターンを設けることが可能であることが示唆された．しかしながら，回復傾向指標は植物の体内水分量が増加していると考えられる状態では機能せず，複数回の潅水を行うことは課題としてあげられる．また，水ストレスの程度の推定値をもとに潅水制御を行う場合，基となる水ストレスの程度を取り入れた指標の評価を行っていない．

今後は水ストレスと関わりのある茎径などを測定することから水ストレスと回復傾向指標の関連性を調べる．その結果を踏まえ，回復傾向指標の有用性を示していくことが必要であると考える．

参考文献

[Araki 87] Araki, Y. and Gotoo, Y.: Application of the Pressure Chamber Method to Tomato Leaflets for Determination of Water Potential, *Journal of the Japanese Society for Horticultural Science*, Vol. 56, No. 3, pp. 328–333 (1987)

[Kaneda 17] Kaneda, Y., Shibata, S., and Mineno, H.: Multi-modal Sliding Window-based Support Vector Regression for Predicting Plant Water Stress, *Knowledge-Based Systems*, Vol. 134, pp. 135–148 (2017)

[大石 02] 大石直記：トマトの養液栽培における水分ストレスに応じた給液制御システムの開発 (1) 茎径変化による水分ストレスの非破壊評価, 生物環境調節, Vol. 40, No. 1, pp. 81–89 (2002)

[柴田 17] 柴田 瞬, 峰野 博史：Optical Flow を用いた複雑背景画像における草姿の変化検出, 情報処理学会論文誌コンシューマ・デバイス＆システム, Vol. 7, No. 2, pp. 97–105 (2017)

[渡邉 16] 渡邉 恭成, 安田 剛規, 米田 正, 中野 明正：トマト育苗のための LED 光照射条件の検討, 野菜茶業研究所報告, Vol. 15, pp. 57–66 (2016)

[2H2] [General Session] 3. Data Mining

座長:檜山 敦（東京大学）

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room H (10F Sky Hall)

[2H2-01] Pattern Recognition for Tennis Tactics using Hidden Markov Model from Rally Series

○Natsuki Miyahara[1], Taro Tezuka[2], Yasushi Nakauchi[3] （1. Graduate School of Library Information and Media Studies, University of Tsukuba, 2. Faculty of Library Information and Media Science, University of Tsukuba, 3. Faculty of Engineering, Information and Systems, University of Tsukuba）

1:20 PM - 1:40 PM

[2H2-02] New indicator for centrality measurement in passing network analysis of football

○Masatoshi Kanbata[1], Ryohei Orihara[1], Yuichi Sei[1], Yasuyuki Tahara[1], Akihiko Ohsuga[1] （1. The University of Electro-Communications）

1:40 PM - 2:00 PM

[2H2-03] Estimation of Cognitive Processes of Shield Machine Operator Based on Eye Tracking

○Nao Fujimoto[1], Junya Morita[1], Toru Okubo[1], Nobuhiko Obayashi[1], Takayasu Shirai[1] （1. KAJIMA CORPORATION）

2:00 PM - 2:20 PM

[2H2-04] Extracting Long-term Patterns from Electrocardiogram Waveform using Deep Learning

○Noriyasu Omata[1], Yoshihiro Nakamura[2], Susumu Shirayama[1] （1. School of Engineering, The University of Tokyo, 2. Faculty of Engineering, The University of Tokyo）

2:20 PM - 2:40 PM

[2H2-05] Learning sense of locality?

daisuke moriwaki[1], ○isshu munemasa[2], toshikazu fukami[1] （1. CyberAgent, Inc., 2. Meiji University）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

隠れマルコフモデルを用いたテニスにおけるラリー系列からのパターン発見

Pattern Recognition for Tennis Tactics using Hidden Markov Model from Rally Series

宮原 捺希[*1] 手塚 太郎[*2] 中内 靖[*3]
Natsuki Miyahara Taro Tezuka Yasushi Nakauchi

[*1] 筑波大学図書館情報メディア研究科
Graduate School of Library Information and Media Studies, University of Tsukuba

[*2] 筑波大学図書館情報メディア系
Faculty of Library Information and Media Science, University of Tsukuba

[*3] 筑波大学システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

In this paper, we propose pattern recognition for tennis tactics using ball trajectory data from motion capture system. The purpose of the study is to adapt machine learning in order to implement feature extraction of rallies in tennis game using positions of ball bounce. We modeled this task as time-series data statistical modeling based on the Hidden Markov Model. We also conducted experiments and we verified the dispersion of the mixture component and the centroid, corresponding to four types of tennis court area division. Moreover, we implemented feature extraction of rally according to the initial state probability and the state transition probability.

1. はじめに

近年，スポーツにおいてデータ分析が普及しており，データを利用したコーチングや，フィジカル面や技術面にとどまらない戦略企画を行うことができるようになっている．例えば，バレーボールの試合では，監督がタブレットを手に試合を分析したものをもとにアドバイスをしている姿がよく見受けられる[1]．データ分析が普及する以前のコーチングは口頭のみによるものが主流だったため，指導者の偏見や，選手と指導者間の価値観の違いにより，選手が疑問や反論を生むことも多かったと考えられる．しかしながら，客観的なデータに基づいてコーチングができるならば，双方にとって納得した上で，より効率的なコーチングが可能になるものと考えられる．

種々のスポーツがあるなかで，本研究ではテニスを対象に研究を行うこととする．既存のテニスの分析システムとして，Hawk-Eye[2]や PlaySight[3]などが挙げられる．それらのシステムは，ボールのバウンド位置の検出や，ショットの種類（サーブ，フォアハンド，バックハンド etc.）の識別およびラベリングを行うものであり，従来のテニスの分析において手作業で行われてきたものを自動化させ，効率化させる．現時点ではそれらのデータは容易に得られるものではないが，我々が日常で利用しているスマートフォンなどのカメラでスポーツの動作のラベリングを自動的に行う研究[4][5]もされており，将来スポーツのデータが身近に感じられるようになることが期待できる．そのようになったときに得られたデータをどのように活用するかが今後の課題となってくる．本研究では，テニスのデータの利用方法を考え，テニスのモーションキャプチャシステム PlaySight のバウンド位置データに機械学習を適用し，選手のプレーのパターン抽出を行う．

宮原捺希，筑波大学図書館情報メディア研究科，〒305-8550
茨城県つくば市春日 1-2，s1721698@s.tsukuba.ac.jp

2. 研究目的

2.1 テニスと機械学習

機械学習を適用したテニスの分析として，選手のランキングや試合データから試合の勝敗を予測する研究がある[6][7][8]．これらのデータ利用と分析方法は，試合の結果のみを予測するといった観戦目的での分析方法であり，試合の内容や選手のプレーの特徴といった観点では分析は行われていない．そのため，本研究では選手の戦術や判断に直接関わるバウンド位置を利用し，過去の試合データから得失点それぞれの傾向にあるパターンを分類することを最終的な目的とする．得失点パターンが抽出されることにより，選手の得意・不得意が把握できる．それによりその後の選手の競技力向上や苦手克服に繋げることを本研究の狙いとする．

2.2 バウンド位置を利用したデータ分析

図 1 は Hawk-Eye のデータ提示の一例である．これはポイントの一発目に打たれるサーブのバウンド位置を，4 つのエリアごとの確率で提示されたものである．このデータは直観的に，コート奥側にいる選手がサーブを打つ際にどのエリアをどれくらいの確率で狙ってくるのかというデータを数値化して提示している．

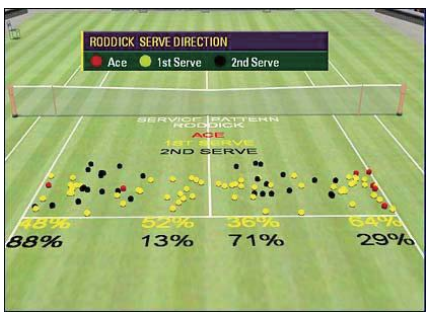


図 1. Hawk-Eye によるサーブのバウンド位置分析

このデータ提示方法は，サーブという一つのショットに対しての分析であり，その後複数のショットで構成されるラリーに対しての

分析は行われていない．既存のシステムにおいてもすべてショットにおけるバウンド位置が取得できるので，それらを利用しショットの組み立てを考慮した分析手法を提案することが本研究の目的である．

3. 隠れマルコフモデル（HMM）によるモデリング

前節で述べた通り本研究の目的となる，テニスのラリーによって繰り出される複数のバウンドを対象とする分析手法を考えた．本研究で扱うデータは，テニスのモーションキャプチャシステム PlaySight より取得している．PlaySight はテニスコートの周りに設置されたトラッキングカメラにより，選手とボールの位置座標を 50FPS で取得するものである．

本研究では時系列の統計モデリングを行う隠れマルコフモデル（HMM）を適用している．HMM は，出力ベクトル（観測変数）と出力ベクトルの確率分布によって推定された状態（予測変数）が，状態遷移確率をもって接続されたものとして定義される[9]．系列が出力される確率は，初期状態確率と状態遷移確率を掛け合わせることにより与えられる．HMM はマルコフ性を持ち，一つの状態は直前の状態にのみ依存する．HMM の学習は，複数の系列データによって行われる．

本研究で用いる学習データは試合中のラリーから観測される時系列のバウンド位置データである．例えば図 2 のようにすべてのバウンド位置（青色点）が観測されたとすると，HMM が表現するのは各状態（混合要素）に対応するガウス分布とその間の遷移確率である．初期状態確率と状態遷移確率を掛け合わせることにより，あるラリーの出力確率を数値的に示すことができる．学習された HMM において系列を表現するときは各状態の重心（橙色点）が用いられる．

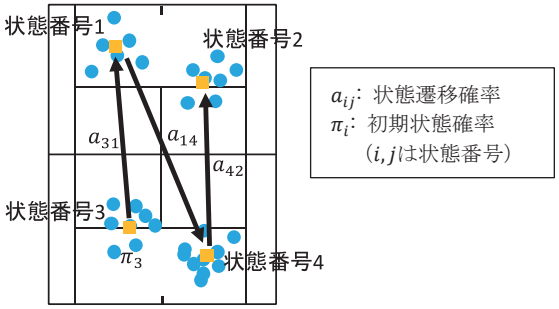


図 2. HMM の構成

4. 実装

PlaySight より 2 人の男子選手の試合のデータを取得した．本研究では 2 試合分のラリーの系列データに HMM を適用した．ラリーのデータというのは，各ポイントにおいてサーブが有効であるときのバウンドの系列データを表す．つまり，ダブルフォルトとなったものは扱わない．今回扱ったデータは系列数（ラリー数）が 214，合計バウンド数は 681 となった．本研究ではサーブのバウンドを除いたもので実験を行った．その理由としては，サーブは手前のサービスエリアという枠にバウンドさせなければならないという規則があり，そのために観測されるバウンドがサービスエリア付近に集中し，予測されるエリアの重心の位置が偏ってしまうという問題がある．そのため，HMM での初期状態はサーブの次のショットであるリターンのバウンドとなる．2試合においてすべてのバウンド位置が図 4 の青点のように観測されたものに対して，状態数を 4, 6, 8, 12 に指定しモデリングを行った．状態数をこの 4 つに指定した理由としては，テニスコートのエリアには左右を区別するフォアサイド・バックサイドの概念があり，上陣と下陣をそれらのエリアに分割するとコート全体の分割数が 4（図 3 左の上陣），さらにセンター（真ん中の概念）を考慮すると 6（図 3 右の上陣），そしてそれぞれに前後の概念を加えて分割すると 8（図 3 左の下陣），12（図 3 右の下陣）となるからである．図 4 のように分割された箇所に HMM の状態と重心が分散される形が望ましい．

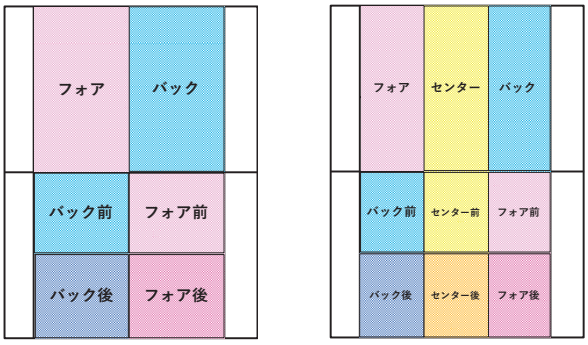


図 3. テニスコートエリア分けの例

5. 結果と考察

図 4 に状態数 4, 6, 8, 12 を設定してモデリングした結果を示す．観測バウンド位置を青色点で示し，予測されたエリアの重心を橙色点で示した．初期状態であるリターンのバウンドは全て下陣で観測されるようになっている．

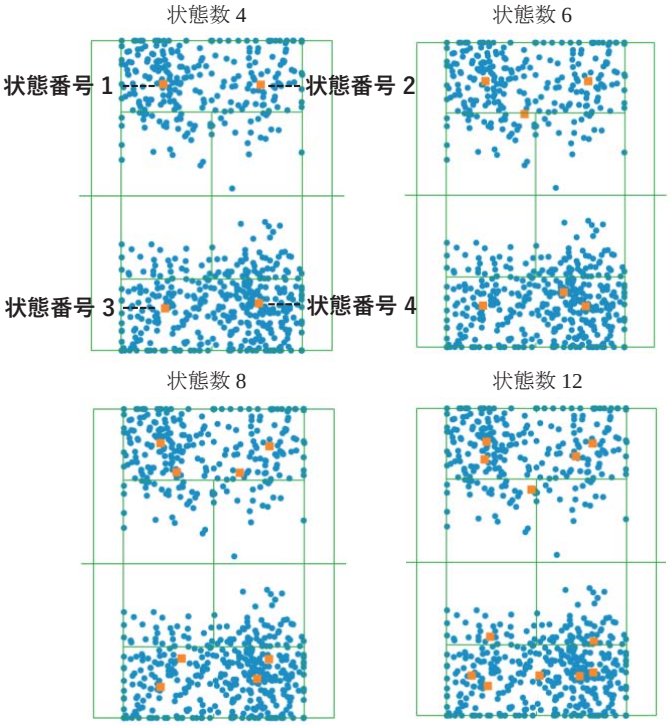


図 4. 観測データと予測変数（状態と重心）

状態数 4 では両陣ともフォアサイド・バックサイドに重心が分散されている．状態数 6 では，上陣でセンターを含めた3つのエリアに分散されているように見えるが，下陣では重心がフォアサイド側に2つ密集している．選手によっては，センターを狙うという意識はあまり持たずに，単純に左右に打ち分けることを意識するというプレースタイルも多い．そのような状況であるとセンター

にバウンドが集まらず，左右のみにバウンドが集中することも考えられる．

前後のエリアを区別する状態が学習されることを期待して実行した状態数 8 と状態数 12 では，前後に重心が分散されず，陣の後方に重心が偏ってしまっている．考えられる理由としては，テニスの試合において，打球はよりネットより遠い深めのエリアを狙ったほうが有利であるため，それを意識したプレースタイルであるほど，陣の後方にバウンドが密集し，観測されたデータからは前後のエリアが区別されにくくなることが考えられる．

状態数 4 の HMM の初期状態確率π_iを図 5 に示す．この HMM における初期状態はリターンのバウンドであり，コートの下陣にバウンドするよう設定されている．つまりこの状況における初期状態は重心 3 と 4（図 4 左上）となる．状態 3 では初期状態確率が 36.6%，状態 4 では 62.0%となっており，この試合においてリターンはフォアサイドを狙われる傾向が強いことがいえる．

図 6 に状態遷移確率a_{ij}を示す．左側の表はテニスコートの上陣にある状態 1, 2 から 3, 4 への遷移確率，右側の表は下陣にある状態 3, 4 から 1, 2 への遷移確率を示したものである．

i	3	4
π_i	0.3657	0.6203

図 5. 初期状態確率π_i（iは状態番号）

i \ j	3	4
1	0.3443	0.6557
2	0.5494	0.4501

i \ j	1	2
3	0.5137	0.4272
4	0.6263	0.3522

図 6.状態遷移確率a_{ij}（i, jは状態番号）

上陣のフォア側である状態 1 に関しては状態 4，つまりクロス方向である下陣のフォア側への遷移確率が 65.6%となっている．そして下陣のフォア側である状態 4 に関しても，状態 1 つまりクロス方向である上陣のフォア側への遷移確率が 62.6%となっている．この結果から，選手は自分のフォアに来たら相手のフォアに返すというフォアのクロスラリーを行う傾向が強いことが読み取れる．そしてほぼ同じ確率でストレート方向に展開する傾向にあることが読み取れる．

状態 2 と状態 3 は上陣と下陣それぞれのバック側を表しているが，左右の打ち分けの確率にあまり偏りが見られない．そのため，選手は自分のバック側に球が打たれると，相手の左右に同じくらいの割合で打ち分けをし，バックでクロスラリーをする傾向は弱いとも言える。両選手の違いを比較するとすれば，上陣からはバック側からクロス方向に遷移する確率のほうが高い一方で，下陣からはバック側からストレート方向に遷移する確率が高いことが読み取れる．

6. まとめと今後の課題

本研究では，テニスのモーションキャプチャシステムから取得されるボールのバウンド位置データに，隠れマルコフモデルを適用し，テニスのラリーのパターン分析を行った．

推定するパラメータの一つである状態がどのように分散するかを考察した．テニスコートを分割する上で基本の概念となるフォア・センター・バックと前後を加味したエリアが区別されることを期待し，それに対応した状態数を指定し出力結果の比較を行った．状態数 4 の出力は左右に重心が分散されたが，センターと前後を含めたエリア分けに対応する状態数の出力は，期待通りに分散されなかった．これは観測データの元になる選手のプレースタイルに依存することが考えられ，コートを広く利用した配球をするプレーであれば分散されたエリアが検出されることも考えられる．しかし，どれくらい広くテニスコートを使って配球をしているかの指標として扱える見込みがある．今後は図 3 に示したエリア分けのパターンと，出力された重心を用いて，選手がどのような配球を行う傾向があるのかを検証する手法を考えることが課題である．

また本研究でモデリングした HMM ではサーブおよびポイントの終わり（アウト・ネット・エース）をパラメータとして扱っていない．今回扱ったラリーのバウンドと連動させることで，ポイントの序盤，中盤のラリー，終盤のすべてを加味したパターンを分析し，提案システムのさらなる改善を図っていく予定である．

謝辞

本研究は，JSPS 科研費 JP16K00228，JP16H02904 の助成を受けたものである．

参考文献

[1] 鈴木淳平: バレーボールのゲーム分析―世界一決定戦での得失点の経緯とゲーム展開の研究，駒澤大学総合教育研究部紀要(4), 2010.

[2] Hawk-Eye., https://www.hawkeyeinnovations.com (2018-02-28 参照)

[3] PlaySight., https://www.playsight.com (2018-02-29 参照)

[4] Draschkowitz L., Draschkowitz C., Hlavacs H.: Predicting Shot Success for Table Tennis Using Video Analysis and Machine Learning , INTETAIN , Social Informatics and Telecommunications Engineering, 2014.

[5] Chen, HT., Tsai, WJ., Lee, SY., Yu, JY.: Ball tracking and 3d trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. Multimedia Tools and Applications 60(3), 2012.

[6] Sipko, M.: Machine Learning for the Prediction of Professional Tennis Matches, Imperial College London ，2015.

[7] Ma, S., Liu, C., and Tan, Y.: Winning matches in Grand Slam men's singles: an analysis of player performance-related variables from 1991 to 2008. Journal of sports sciences, 31(11), 2013.

[8] O'Malley, JA.: Probability Formulas and Statistical Analysis in Tennis. Journal of Quantitative Analysis in Sports, 4(2), 2008.

[9] 徳田恵一: 隠れマルコフモデルによる 音声認識と音声合成, IPSJ Magazine, 情報処理学会, 2004.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

サッカーのパス網分析における中心性測定の新指標

New indicator for centrality measurement in passing network analysis of football

神畠 正稔 Masatoshi KANBATA　折原 良平 Ryohei ORIHARA　清 雄一 Yuichi SEI　田原 康之 Yasuyuki TAHARA　大須賀 昭彦 Akihiko OHSUGA

電気通信大学大学院情報理工学研究科
Graduate School of Informatics and Engineering,The University of Electro-Communications

Data analytics is used in various field including business, science and sports. The evaluation of players and teams affects tactics, training and scouting in football teams. Players and teams are often evaluated by data such as shots and goals in game results. However it is not enough to fully understand potential of the players and teams. This paper describes a new analysis method using football passing data. In order to evaluate performance of players and teams, we employ graph mining. There is an indicator called centrality that evaluates individual contribution within an organization and it is used to evaluate the players and the teams. In calculation of the centrality for a given player pair, we consider not only the shortest path of passing but also longer ones. As a result, we found our method to be more consistent with game results than conventional methods.

1. はじめに

近年，スポーツにおける選手の走行距離や得点数など，試合において取得されるデータの分析が行われている．チームや選手の評価，戦術や選手の育成，スカウティングなどのチーム運営において，このような分析が影響を与えている．

しかし，サッカーにおいては，選手を評価するデータ指標がゴールとアシストである場合が多く，試合においてそれらを記録した選手に高評価が偏る傾向がある．このことから，ゴールとアシストを記録していない選手も含めた全選手は，チームにおいてどのように貢献したのか，それをどのように評価すべきなのかを検証する．また，より適切な選手評価からチームを評価することで，チーム成績の要因の論理的な分析に貢献することができると考えられる．

本研究では，サッカーにおいてゴールやアシストに繋がると考えられる，パスに着目して評価を行う．具体的にはグラフ理論を用いてシュートで終了したパスの流れを活用して，選手の試合における貢献を分析をする．また，グラフ理論における中心性の概念を活用することで，選手及びチームを適切に評価するための手法を提案する．

2. 関連研究

スポーツに関連する研究はこれまで数多く行われており，グラフ理論を適用する研究が，サッカーのパスに対してされてきた．また，このような研究を行う際，試合におけるパスの出し手と受け手，そのパス数を記録した passing distribution が公開データとして用意されている．

サッカーにおいて，相手ゴールにボールを移動させることができる手段の一つであるパスは，非常に重要な項目であると考えられている．このことから，passing distribution を適用してグラフネットワークを作成する分析がされてきた．グラフネットワークにおいて試合における選手評価，およびチーム評価を行うための手法として中心性が測定される．中心性測定のために用いられる指標として，betweeness centrality と flow centrality がある．ただし，passing distribution は選手間のパス数を記録しているだけであり，3 人以上の選手間のパスの流れは再現できない．

連絡先: 神畠正稔，電気通信大学大学院情報理工学研究科，東京都調布市調布ケ丘 1-5-1, Email:kanbata.masatoshi@ohsuga.lab.uec.ac.jp

2.1 Betweenness centrality

中心性とはグラフネットワークにおいて，任意のノードが他のノードに与える影響度を示す指標である [Tsugawa 14]．サッカーの試合において，より多くパスに関わった選手が貢献を示すという考え方がある．これをグラフ理論に適用すると，passing distribution により作成したグラフから，選手ノードの中心性を測定することができる．

σ_{st} をノード s からノード t までの最短経路の総数，$\sigma_{st}(v)$ をノード s からノード t までの最短経路の内ノード v を通る総数として，ノードの中心性を表す betweenness centrality は以下のように定義される．ただし，$V = \{1, 2, ..., 11\}$ としこれを選手ノードの集合とする [Barthelemy 04]．

$$betweenness(v) = \sum_{\substack{s,t \in V \\ s \neq v \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

2.2 Flow centrality

betweenness centrality を変更し，シュートまでに至るパスの過程のみを最短経路として扱う [Duch 10]．このことから，直接得点の機会を得た場合を考慮して選手の評価を行う．新たにシュートノードを作成しノード t とする flow centrality を以下のように定義する．ただし，$U = \{12, 13\}$ をシュートノードの集合とし，12 をゴール外，13 をゴール内へのシュートとする．

$$flow(v) = \sum_{\substack{s \in V \\ t \in U \\ s \neq v}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

[Duch 10] は，試合における選手の影響度はゴールに直結したパスの流れが最も大きくなると考えており，betweenness centrality より flow centrality のほうが，より選手やチーム評価を適切に行うことができると主張している．

3. 提案手法

flow centrality における最短経路の探索において，ノード t をシュートノードとしたためノード間のパスの経路が限定的となる．そのため，算出される経路の総数が少なくなり，測定において得点に関わった選手が高く評価されない場合がある．

そこで本研究では，シュートに至るパスに関わった選手をより考慮するため，経路の探索において最短経路+1 した場合の経路を求めることで，算出される経路の総数を増やす．σ'_{st} をノード s からノード t までの最短経路+1 の総数，$\sigma'_{st}(v)$ をノード s からノード t までの最短経路+1 の内ノード v を通る総数として，$flow^+$ の式を以下のように定義する．

$$flow^+(v) = \sum_{\substack{s\in V \\ t\in U \\ s\neq v}} \frac{\sigma'_{st}(v)}{\sigma'_{st}} \tag{3}$$

4. 評価

4.1 実験

提案手法が試合において，選手及びチームに対して適切な評価を行っているか検証する．UEFA EURO2008 の passing distribution [*1] を用いて，任意のチーム A における各試合に先発出場した選手の $betweenness$，$flow$，$flow^+$ を，$P^A_{betweenness}$，P^A_{flow}，P^A_{flow+} とし降順にする．それらの選手を $n = \{1, 2, ..., 11\}$ とした試合におけるチームの評価を以下のように定義する．

$$\overline{P}^A_{betweenness}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{betweenness}(i)$$

$$(P^A_{betweenness}(1) > P^A_{betweenness}(2) > ...) \tag{4}$$

$$\overline{P}^A_{flow}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{flow}(i)$$

$$(P^A_{flow}(1) > P^A_{flow}(2) > ...) \tag{5}$$

$$\overline{P}^A_{flow^+}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{flow^+}(i)$$

$$(P^A_{flow^+}(1) > P^A_{flow^+}(2) > ...) \tag{6}$$

本研究では，参考論文に基づいて n=3,11 として，チーム A とチーム B の試合における $\overline{P}^A(n)$ と $\overline{P}^B(n)$ の大小が勝敗と直結しているかどうかを検証し，その精度を算出する [Duch 10].

4.2 結果

データの予選リーグに対して前述した実験を行った結果を以下に示す．表 1 は $\overline{P}^A(11)$ と $\overline{P}^B(11)$，$\overline{P}^A(3)$ と $\overline{P}^B(3)$ の大小と勝敗の整合精度，表 2 は各試合の $\overline{P}^A_{flow}(3)$ と $\overline{P}^B_{flow}(3)$，$\overline{P}^A_{flow+}(3)$ と $\overline{P}^B_{flow+}(3)$ を示す．ただし，表 2 における国名は FIFA コードで表記し，試合結果と値が整合しない場合は F とする [FIFA 10].

表 1: 試合結果との整合精度

	$betweenness$	$flow$	$flow^+$
$n = 11$	0.43	0.33	0.43
$n = 3$	0.48	0.43	0.71

*1 試合数 : 31 試合，チーム数 : 14 チーム，全パス数 : 21624，1 試合の平均パス数 : 698

表 2: 予選リーグの各試合の $\overline{P}^A_{flow}(3)$，$\overline{P}^A_{flow^+}(3)$

Gr.A	SUI-CZE	POR-TUR	CZE-POR	SUI-TUR	SUI-POR	TUR-CZE
$flow$	4.37-7.83	4.66-3	4.86-4.89	3.19-4.67	5-5.31,F	3.9-4.05,F
$flow^+$	6.64-8.71	7.32-4.55	7.4-7.66	6.88-7.09	8.17-7.94	6.6-6.49
Gr.B	AUT-CRO	GER-POL	CRO-GER	POL-CRO	AUT-GER	
$flow$	4.83-5.12	5.67-3.87	3.23-4.22,F	4.07-3.51,F	4.72-2.71,F	
$flow^+$	7.23-7.7	8.86-6.98	6.38-6.87,F	6.45-6.87	6.92-5.24,F	
Gr.C	NED-ITA	NED-FRA	NED-ROU	FRA-ITA		
$flow$	3.27-3.33,F	4.58-3.26	3.68-5,F	5.33-3.57,F		
$flow^+$	6.04-5.19	7.24-6.24	6.69-7.18,F	8.33-5.69,F		
Gr.D	ESP-RUS	GRE-SWE	SWE-ESP	GRE-RUS	GRE-ESP	RUS-SWE
$flow$	3.57-4.61,F	3.72-3.12,F	4.33-4.5	4.97-4.43,F	4.58-4.73	2.86-3.2,F
$flow^+$	6.79-7.13,F	6.07-6.25	5.91-7.26	6.89-7.53	6.26-8.29	5.95-6.34,F

5. 考察

表 1 から $flow^+$ が，$betweenness$，$flow$ よりも高い値を示していることがわかる．$flow^+$ では最短経路+1 の経路を算出して flow centrality を行っているため，通常の flow centrality と比較して，シュートまでのパスの過程がより多く算出されている．このことから，シュートに至るパスに関わった選手が従来手法より評価され，精度を上回ったと考えられる．尚，本研究では経路を最短経路+1 として検証を行ったが，今後は経路の適切な長さを推測し検証を行う必要があると考えられる．

また，表 2 からチームの評価が試合結果に整合しなかった場合があった．この理由を調査した結果，$flow$ と $flow^+$ の相対的な選手評価の傾向は変わらなかったが，特定の選手の評価が極端に高くなる傾向にあり，それによってチーム評価が高くなることが判明した．これは，シュートを打った特定の選手にのみパスを集めることにより発生する事象であると考えられる．このような事象は，試合においてシュートの本数が極端に少なくなる場合や，特定の選手にボールを集める戦術を遂行する場合に発生すると推測され，今後はパスの総数を考慮して調査を行う必要がある．

6. おわりに

本研究では，サッカーのパスからグラフ理論の中心性の概念を活用して選手の試合における貢献を分析した．分析の結果，提案手法が従来手法を上回り，より適切な選手評価及びチーム評価を行うことができた．

参考文献

[Tsugawa 14] S. Tsugawa and H. ohsaki :A Study on the Robustness of Centrality Measures in Network Analysis(2014), 第 57 回自動制御連合講演会論文集 pp.1072-1077

[Barthelemy 04] M. Barthelemy:Betweenness centrality in large complex networks(2004), The European Physical Journal B pp.163168

[Duch 10] Jordi Duch, Joshua S. Waitzman, Luis A. Nunes Amaral: Quantifying the Performance of Individual Players in a Team Activity(2010), PLoS ONE 5(6): e10937. https://doi.org/10.1371/journal.pone.0010937

[FIFA 10] FIFA's 208 Member Associations(2010):http://www.fifa.com/mm/document/fifafacts/organisation/52/00/10/fs-120_01a_ma.pdf

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

アイトラッキングによるシールドマシン操作者の認知プロセスの推定

Estimation of Cognitive Processes of Shield Machine Operator Based on Eye Tracking

藤本 奈央[*1] 森田 順也[*1] 大久保 泰[*1] 大林 信彦[*1] 白井 健泰[*1]
Nao Fujimoto Junya Morita Toru Okubo Nobuhiko Obayashi Takayasu Shirai

[*1] 鹿島建設株式会社
KAJIMA Corporation

In shield construction, it is concerned that the shortage of skilled workers causes serious accident due to oversight of fluctuation and delay in taking action. Therefore, revealing the cognitive processes of experienced shield machine operators is important. In this report, a part of the cognitive processes was revealed by analyzing eye tracking data and control logs of shield machine. Two machine operators were conducted eye tracking test. There were differences in viewpoint between the two operators as a result of comparison of the gaze movements. From Eye tracking and data of control machine, we revealed time lags of recognizing fluctuation and clarified operator's sequences of estimating manipulation. We will develop operating support system and foster advanced operator in the future.

1. はじめに

建設業は製造業に比べ生産性が低いと言われているが、シールド工事は 1990 年代から機械化・自動化・システム化が進み，少人数の作業員による施工体制が確立されているため，生産性の高い工種である．しかしながら，就労者の高年齢化による将来的な熟練作業員の不足という建設業全体の課題をシールド工事も抱えている．特に、これまで熟練作業員が予兆に気付き，未然に防いでいたトラブルの顕在化が懸念されている．

シールド工事のマシン操作は図1に示すような中央管理室やシールドマシンの運転席で行う．切羽前面（掘削面）に作用する土水圧，シールドマシンの後続台車および地上に設置している各設備の稼働状況，排土量の計測値，出来形等のデータを操作画面に集約し，そのデータを基に運転操作を行っている．しかし，データの種類が数十項目と多く，またリアルタイムに変化するため，シールドマシンオペレータは必要なデータを選択・監視し，操作を行うことになる．したがって，熟練のマシンオペレータは，データの変動を早期に察知し，経験や勘所による判断で操作を行っている．

そこで筆者らは，状況認識から判断，操作に至るまでの，熟練マシンオペレータの経験や勘所に基づく認知プロセスを明らかにする研究を進めている．

本稿では，大野の研究[大野 02]をもとに，オペレータの操作画面に対するアイトラッキングデータとマシンオペレータのシールドマシンから取得できるデータ（以下，シールドマシンデータ）を組み合わせた取り組みについて紹介する．



図 1 シールドマシンの中央管理室

連絡先：藤本奈央，鹿島建設株式会社，東京都港区赤坂 6-5-11，fujimnao@kajima.com

2. 分析データ

2.1 シールドマシンデータの概要

マシンオペレータは遠隔操作ソフトウェアを通じ，センサデータの監視及びシールドマシンの操作を行っている．これらのログは 1 秒毎に記録され，記録サーバに保存される．このログから取得できるデータとしては，シールドジャッキの速度や選択状況，切羽水圧等のシールドマシンに関するもの，ポンプ圧力や回転数，バルブの開閉状況等の流体設備に関するものが含まれる．

2.2 アイトラッキングの実施

マシンオペレータ 2 名（表 1）に対し，2 日間，2 回ずつアイトラッキングを実施した．シールド工事では，セグメントと呼ばれるブロックをリング状に組み立てトンネルの外壁としながら掘削を進めるが，1リング分の掘進を1回の計測とし，切羽の安定に重きを置いた運転操作となる直線区間を対象とした．アイトラッカーは眼鏡型で被検者が違和感なく装着できる Tobii 社製 Tobii Pro Glasses2 を選択した．計測した視線データは一般に注視状態とされる 100deg/s 以下の視線移動を視線の停留と定義した，また操作画面に対して表示項目毎に興味関心領域（AOI：Area Of Interest）を定義した．

3. 結果と考察

アイトラッキングの結果から，操作画面に対する視線ヒートマップを作成した．13 日の例を図 3 に示す．オペレータ A（図 3(a)）は P2，P7 ポンプ，ロータリークラッシャー（以下，RC）を中心に見ていることがわかる．一方で，オペレータ B（図 3(b)）は RC，P2 ポンプ，切羽水圧を中心に見ていることがわかる．このことから，二人のオペレータは切羽を安定させる運転のために，それぞれ異なる点を重視していることが分かった．

表 1 アイトラッキングの実施概要

被検者	年齢 [歳]	勤続年数 [年]	対象リング[No]	
			2017/11/13	2017/11/28
オペレータ A	39	15	452,453	578,579
オペレータ B	56	26	457,458	574,575

アイトラッキングデータとシールドマシンデータをマージしたグラフから，特徴的であった部分の例を図 4,5 に示す．オペレータ A は No.452 リングにおいて，P2 ポンプの吸込圧力の低下という配管閉塞の兆候の発生時（図 4(b)中①）に，P2 ポンプの吸込圧力の変動を注視しながら掘進を行っていたが，低下時は何も操作することなく，注視を継続していた．次の大きな低下時（図 4(b)中②）に閉塞が発生し，4 秒後に状況認識し，認識から 5 秒後に掘進を停止した．その後流体設備の操作を実施し掘進再開まで 190 秒かかった．

オペレータ B は No.457 リングにおいて，RC の入口圧力の低下という閉塞兆候の発生時（図 5(b)中①）にジャッキ速度を遅くする対応を行うことで復旧させていた．RC の入口圧力の安定を確認後，速度を変更前に復旧させたが，RC の電流値の上昇，回転速度の低下という閉塞兆候が再度発生（図 5(b)中②）したため掘進を 28 秒間停止した．しかし，対応が早かったため，流体設備を操作することなく掘進を再開した．また，RC の入口圧力の変動からオペレータが現象を認識するまでに 22 秒かかった．

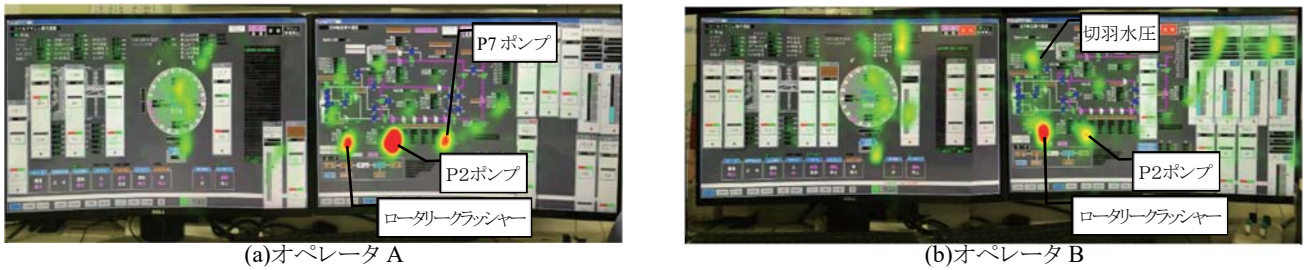
このようにアイトラッキングデータとシールドマシンデータを組み合わせた分析を実施したことで，変動発生の認識，判断，操作という一連の流れを把握する手がかりを得ることができた．

4. 今後の課題

今回の結果からマシンオペレータに早期に異常を認識させ，正確な判断を促すことがトラブルの防止に有効であることが分かった．今後は熟練オペレータの判断基準を集積し，それを基に警報などを発令する掘進サポートシステムの構築，さらに，将来的な自動運転制御を見据えたシミュレータの構築を進め，シールド工事の安全性向上へつなげたい．

参考文献

[大野 02] 大野健彦：視線からなにがわかるか-視線測定に基づく高次認知処理の解明，日本認知科学会，共立出版，Vol.9, No.4, pp.565-576 (2002).



(a)オペレータ A　(b)オペレータ B

図 3 操作画面に対する視線ヒートマップ

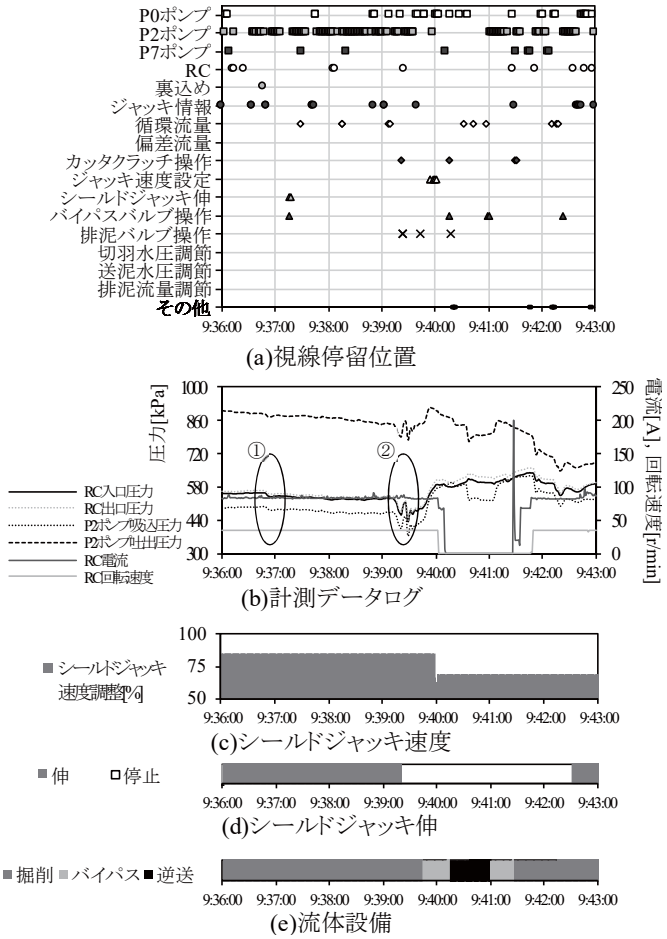


図 4 No.452 リングの操作状況

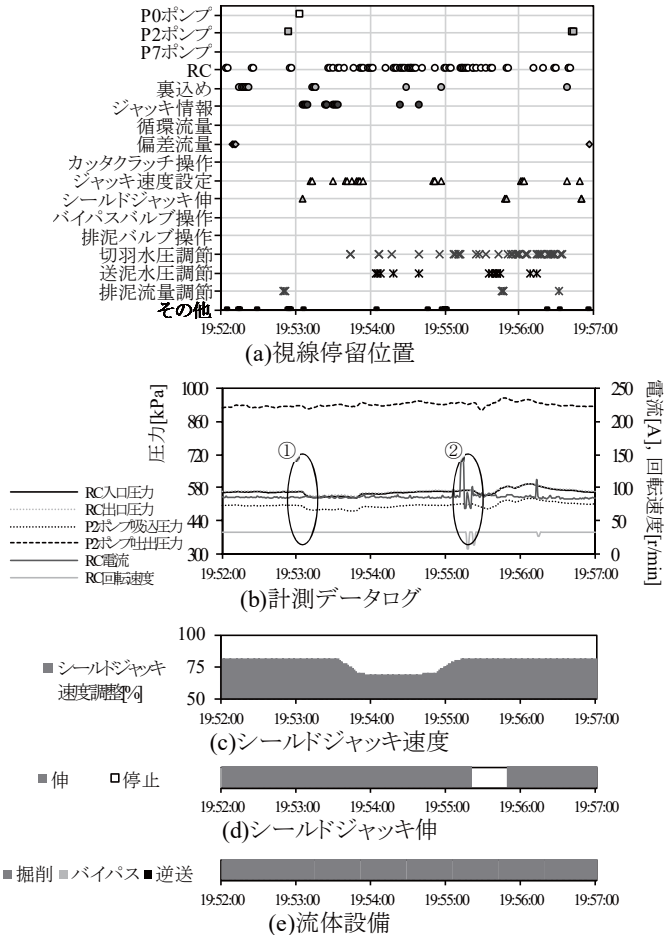


図 5 No.457 リングの操作状況

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習を用いた心電図波形の長期変動パターン抽出法

Extracting Long-term Patterns from Electrocardiogram Waveform using Deep Learning

尾亦 範泰 *1 Noriyasu Omata　中村 嘉宏 *2 Yoshihiro Nakamura　白山 晋 *1 Susumu Shirayama

*1東京大学大学院工学系研究科 School of engineering, The University of Tokyo

*2東京大学工学部 Faculty of engineering, The University of Tokyo

Automatic analysis of electrocardiograms (ECG) has been attempted. Although almost all methods pay attention only to short-term waveforms, their changes over time are said often important in diagnosis. In this paper, we focus on such long-term waveform change and propose a method to extract it as a pattern. The proposed method combines a method of expressing time series data as a trajectory in a feature space and feature extraction by an autoencoder. Evaluation experiments suggested the existence of regularity in the pattern and its association with the disease.

1. はじめに

Internet of Things (IoT) に注目が集まり，これまで入手困難だったデータの取得が可能となりつつある．医療の領域でもウェアラブルデバイス等によって収集されたリアルタイムでの生体情報の有効活用が期待されている．特に注目されている分野の 1 つとして心電図のリアルタイムモニタリングがあげられる [高橋 17].

心電図自体の取得は容易になっているものの，検出した波形から得られる情報の分析手法，活用法に関しては依然として未開な部分が多い．基本的なものは周波数解析に基づくものであり，フーリエ分析やウェーブレット分析に基づくものが既に広く用いられている [宮内 16]．しかしながら，平岡によれば，心電図の診断は生化学的な検査とは異なり，定性的な判断による部分が大きく，数字で表しにくい波形の違いという指標を基準に行われる．そのため，定量的な判断指標が存在せず，コンピュータの得意な領域とは言いがたく，自動解析が難航している [平岡 15].

そこで，従来の周波数解析に寄らない方法で心電図波形を解析する研究が行われている．特に，近年の深層学習（Deep Learning）の発展を受け，心電図に対しても適用する研究がなされている．Rajpurkar らは，畳み込みニューラルネットワーク（CNN）を用いた不整脈検知を行った [Rajpurkar 17]．彼らは，ウェアラブルの単極誘導心電モニタを使用することで，過去の研究で用いられてきた心電図のデータセットと比べてサンプル数が 500 倍以上のデータセットを作成した．このデータセットを用いて訓練された CNN は，不整脈検知の再現率，及び精度の両面で認定心臓専門医に勝る能力を持つことを示している．彼らの結果から，ラベル付きデータセットの作成が問題になるものの，CNN による心電図解析が有効であることが分かった．

また，高橋らは，教師無し学習である Stacked Convolutional Denoising Autoencoder（SCDAE）を用いて，心電図からの特徴抽出を行うことを提案している [高橋 17]．彼らは，SCDAE によって抽出された特徴量に対して，教師あり学習によって分類器を学習することで，直接的に心電図波形に対して分類器を学習する手法よりも高精度に不整脈を検知することに成功した．彼らの研究によって，ディープオートエンコーダを用いることで，観測対象および心拍変動による波形変化に対してロバストな特徴抽出が可能であることが示された．

ところで，これらの手法は「1 秒」や「1 拍」といった短時間のデータにおける不整脈の検知を行っている．しかしながら，心電図は経時的な変化を見ないと診断が困難である場合がしばしばあり [平岡 15]，長時間での変動を観察する必要がある．短期間の心電図のみに着目し，長時間での変動を考慮しない従来手法では，そのような長時間の変動に基づく疾病予測への適用は難しい．

以上を踏まえ，本稿では，ある程度の長時間における，心電図波形の変動を把握するための手法を提案する．この手法では，CNN ベースのディープオートエンコーダによって抽出される特徴量をもとに，長時間での心電図波形における変化を，特徴空間における軌跡として表現し，パターンとして抽出する．提案手法は教師なし学習であるため，心電図に対して心拍間隔の明示や不整脈の有無や種類によるラベル付けを必要としない．さらに，公開されている心電図のデータセットに対して，この手法を適用した結果を示す．実験の結果，長時間での心電図波形の変化は，複数のクラスに定性的に分離できることが示された．さらに，訓練に使用しなかったデータから，データセットにおいて特異な疾病をもつ患者を，その心電図波形の軌跡の特異性から定性的に発見できた．このような結果から，本提案手法が心電図波形の長期変動の解析に有効であることが示される．

2. 提案手法

2.1 手法の概要

動的ネットワークの時間変化を把握する方法として，動的ネットワークを特徴空間における軌跡として表現する手法が提案された [van den Elzen 16]．この手法では，各瞬間のネットワークを主成分分析等を用いて低次元化する．この低次元量を時間的に連続的にプロットすることで，時間変化を表現することが可能となる．本稿では，この手法とオートエンコーダによる特徴抽出 [高橋 17] を組み合わせることで，心電図波形の長時間での変化を軌跡として表現することを提案する．

提案手法は，訓練フェーズと適用フェーズに分割される．訓練フェーズは，事前に与えられた心電図データをもとに，学習器のパラメータを調整するフェーズである．このフェーズでは，オートエンコーダのパラメータと，主成分分析の変換行列

連絡先: 尾亦範泰, omata@nakl.t.u-tokyo.ac.jp

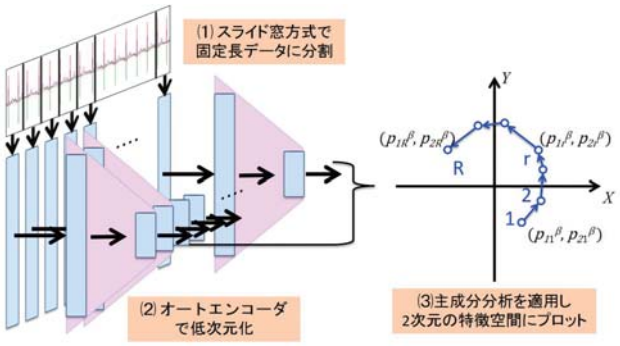


図 1: 提案手法における適用フェーズの概要

が学習される．

適用フェーズは，初めて与えられる未知のデータに対して，訓練フェーズで学習した内容をもとに，実際に可視化を行うフェーズである．このフェーズは，図 1 に示すように，下記の 3 ステップで構成される．

1. 入力データからの固定長データの切り出し
2. オートエンコーダによる各データの低次元化
3. 低次元化されたデータの特徴空間への写像

以下では，各フェーズにおけるそれぞれのステップについて説明する．

2.2 固定長データの切り出し

個人 β の心電図波形データ $\boldsymbol{x}^\beta$ をスライド窓方式で固定長 τ で区間に分割し，

$$\boldsymbol{x}^\beta = (\hat{\boldsymbol{x}}_1^\beta, \ldots, \hat{\boldsymbol{x}}_R^\beta) \tag{1}$$

とする．ここで，$\hat{\boldsymbol{x}}_r^\beta$ は個人 β の r 番目の区間，R は区間総数である．なお，τ はパラメータとして与える．

この方法で分割された固定長のデータは，心拍間隔や不整脈に対するラベルを考慮しない．このため，データに対してそのようなラベル付けを行う必要がなく，取得されるデータの利用が容易であるといえる．

2.3 オートエンコーダによる特徴抽出

提案手法では，入力を $\hat{\boldsymbol{x}}_r^\beta (\beta = 1, \ldots, M, r = 1, \ldots, R)$ とする，CNN ベースのディープオートエンコーダを用いる．図 2 に示すように，オートエンコーダを用いることで，データを表現する情報を保持する低次元な特徴量を得ることができる．

オートエンコーダのエンコーダ部を encoder$(\cdot)$，デコーダ部を decoder$(\cdot)$ で表す．各個人 β の r 番目の区間の特徴量は，

$$\boldsymbol{z}_r^\beta = \text{encoder}(\hat{\boldsymbol{x}}_r^\beta) \tag{2}$$

で表される．また，低次元特徴量であるエンコード結果 $\boldsymbol{z}_r^\beta$ に対して，デコーダを作用させた結果

$$\hat{\boldsymbol{y}}_r^\beta = \text{decoder}(\boldsymbol{z}_r^\beta) \tag{3}$$

は，元信号 $\hat{\boldsymbol{x}}_r^\beta$ に類似したものとなる．

これを実現するため，訓練フェーズでは，訓練データに含まれる全ての個人に対する，時間区間分割された全データに対する絶対復元誤差の総和

$$E = \sum_{\beta=1}^{M} \sum_{r=1}^{R} |\hat{\boldsymbol{x}}_r^\beta - \hat{\boldsymbol{y}}_r^\beta| \tag{4}$$

を損失関数として，パラメータの学習を行う．

適用フェーズでは，各個人から取得されるデータ $(\hat{\boldsymbol{x}}_1^\beta, \ldots, \hat{\boldsymbol{x}}_R^\beta)$ に対し，特徴量の系列 $(\boldsymbol{z}_1^\beta, \ldots, \boldsymbol{z}_R^\beta)$ を得る．

2.4 特徴量にもとづく軌跡の描画

提案手法では，特徴量の系列を 2 次元空間に連続的に写像することで，元データの時間変化を可視化する．写像の手法として，先行研究 [van den Elzen 16] に倣い，主成分分析を用いる．この写像は，変換行列 W を用いて，$\boldsymbol{p}_r^\beta = W\boldsymbol{z}_r^\beta$ と表される．変換行列 W は，訓練フェーズにおいて，すべての訓練データに対する特徴量を用いて学習される．

適用フェーズでは，各個人の特徴量系列 $(\boldsymbol{z}_1^\beta, \ldots, \boldsymbol{z}_R^\beta)$ に対して，訓練フェーズで学習した変換行列 W を作用させることで，主成分スコアの系列 $(\boldsymbol{p}_1^\beta, \ldots, \boldsymbol{p}_R^\beta)$ を得る．さらに，第 1 主成分スコアを x 軸に，第 2 主成分スコアを y 軸にして，この系列をプロットする．このプロットの位置は，時間経過とともに移動し，各個人について個別の軌跡を描く．この軌跡によって，心電図波形の変化が可視化される．

3. 数値実験

提案手法の有効性を確認するための数値実験を行った．

本研究では The MGH/MF Waveform Database[*1] から取得した心電図波形のデータセットを利用した．このデータセットのうち，欠損値を含まないもののみを用いた．また，このうち 106 サンプルの 3 誘導心電図波形データを使用した．さらに 8 割を訓練データ，残り 2 割をテストデータとして実験を行った．

本実験においては，$\tau = 0.6$ とし，サンプル 1 人の 0.6 秒分に相当する 216 次元 × 3 チャンネル のデータを切り出した．また，スライド窓法におけるスライド幅は窓幅と一致させた．

本実験において用いたディープオートエンコーダの構成を表 1 および図 2 に示す．表 1 における Pool-3 層を，オートエンコーダによって得られる特徴量とした．これによって，各窓の 216×3 次元のデータは，$18 \times 4 = 72$ 次元にまで次元削減される．

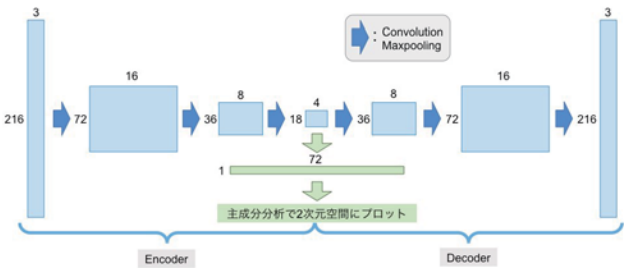


図 2: 実験における低次元化の概要

オートエンコーダによる次元削減によって情報量が大きく失われないことを確かめるため，特徴量をデコードした結果 $\hat{\boldsymbol{y}}_r^\beta$ をプロットし，元信号 $\hat{\boldsymbol{x}}_r^\beta$ と比較した．未知データに対する結

*1 https://physionet.org/physiobank/database/mghdb/

表 1: 実験に用いたオートエンコーダの構造

Name	Shape	Operation performed
Input	216 × 3	—
Conv-1	216 × 16	3 × 16 convolution
Activate-1	216 × 16	ReLU activation
Pool-1	72 × 16	3 maxpool
Conv-2	72 × 8	3 × 8 convolution
Activate-2	72 × 8	ReLU activation
Pool-2	36 × 8	2 maxpool
Conv-3	36 × 4	3 × 4 convolution
Activate-3	36 × 4	ReLU activation
Pool-3	18 × 4	2 maxpool
Conv-4	18 × 8	3 × 8 convolution
Activate-4	18 × 8	ReLU activation
Upsample-4	36 × 8	2 upsampling
Conv-5	36 × 16	3 × 16 convolution
Activate-5	36 × 16	ReLU activation
Upsample-5	72 × 16	2 upsampling
Conv-6	72 × 3	3 × 3 convolution
Output	216 × 3	3 upsampling

果の例を図 3 に示す．波形の特徴を維持して復元されており，特徴の保持に成功していると考えられる．

さらに，オートエンコーダによって得られた特徴量を，主成分分析によって写像した．写像の結果の例を図 4 に示す．各個人のデータに対する写像の結果を俯瞰して比較したところ，軌跡の形状はおおまかに定性的に 4 つのクラスに分類されることが分かった．これらのクラスを表 2 に示す．さらに，図 4 において，(a) は収束型，(b) は拡散型，(c) は直線型，(d) は曲線型の典型例をそれぞれ示している．軌跡の形状が明らかに特徴的で，パターンごとの分類が可能であり，何らかの規則性が見出せる可能性がある．

また，適用フェーズにおいて用いたデータのうち，その軌跡の形状が特異なものが定性的に発見された．このデータの軌跡は，上記の 4 つのクラスに分類できず，かつ軌跡の描かれる範囲が突出して大きい．図 5 にそのサンプル mgh126 の心電図波形の一部 (1000 秒から 1010 秒) と，特徴空間への写像を示す．データセットに示されている患者ごとの説明によれば，この患者は骨盤腔内腫瘤を患い，子宮の摘出及び卵巣の摘出を行っていた．用いたデータセットにおいて，骨盤腔内腫瘤を持つ唯一の患者であり，子宮の摘出及び卵巣の摘出を行った他の事例も本データセットには存在しなかった．この結果より，提案した手法によって心電図から疾病予測を行うこともできる可能性が示唆された．

表 2: 軌跡の定性的形状に基づく分類

名前	定義	個数
収束型	おおよそ原点の位置に収束している	46
拡散型	相対的に広範囲に軌跡が拡散している	18
直線型	軌跡が直線状に重なっている	7
曲線型	軌跡が弓なりの曲線状に重なっている	10
例外	上記の分類に当てはまらない	3

4. 結論

本研究では，心電図波形データの分析が難航していることに言及し，観測対象及び心拍変動による波形変化の問題を解決する，心電図からの新しいパターン抽出手法を提案した．

提案手法は，心電図の長時間の波形を固定長に切り出し，それぞれを CNN ベースのディープオートエンコーダによって低次元化し，さらに主成分分析で 2 次元の特徴空間に連続的に写像を行うことで，心電図波形の長時間での変化をパターンとして可視化，抽出するものである．本手法は，既存手法とは異なり，入力データの時系列性を保持して扱うことが可能であり，長時間での変動に基づく心電図の分析を行うことができる．

提案手法の有効性の確認のために行った実験によって，心電図波形は大まかに 4 つのクラスに定性的に分類できることが分かった．さらに未知のテストデータに対して提案手法を適用したところ，特異な疾病を持つ患者を，可視化結果の定性的な特異性から発見することができた．したがって，特徴空間への写像結果と疾病とのつながりを明確に見出すには至らなかったが，何らかの規則性や関連性を見出せる可能性が示唆された．

今後の課題として，結果の解釈を定量化することが挙げられる．定性的に行われている軌跡の分類や例外の判別を，客観的指標に基づいて行えるようにする必要がある．これらは，軌跡データマイニングの手法をもとにしたクラスタリングや，ロバスト主成分分析などの外れ値検出手法を用いることで達成可能であると考えられる．さらに，現状の教師なし学習にもとづく方法では，得られる軌跡に「意味」を持たせることは困難である．より解釈性を高める方法として，データの一部または全部に対してラベルを付与することにより，半教師あり学習または教師あり学習の枠組みを用いた特徴量を用いる方法が考えられる．これによって，軌跡を描く特徴空間内において不整脈の判別境界を描画するといった，より客観的な方法によって結果の解釈が行えるようになると考えられる．

参考文献

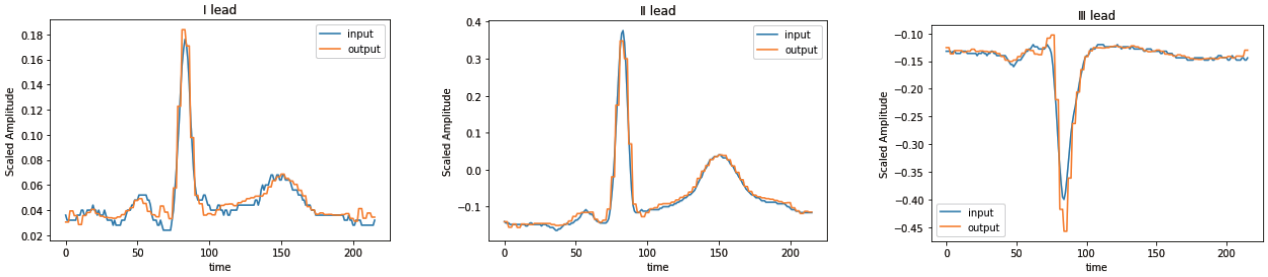
[Rajpurkar 17] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y.: Cardiologist-level arrhythmia detection with convolutional neural networks, *arXiv preprint arXiv:1707.01836* (2017)

[van den Elzen 16] van den Elzen, S., Holten, D., Blaas, J., and van Wijk, J. J.: Reducing snapshots to points: A visual analytics approach to dynamic network exploration, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 1–10 (2016)

[宮内 16] 宮内 靖史, 植竹 俊介 : 心電図の周波数解析の立場から, 心電図, Vol. 36, No. 2, pp. 161–166 (2016)

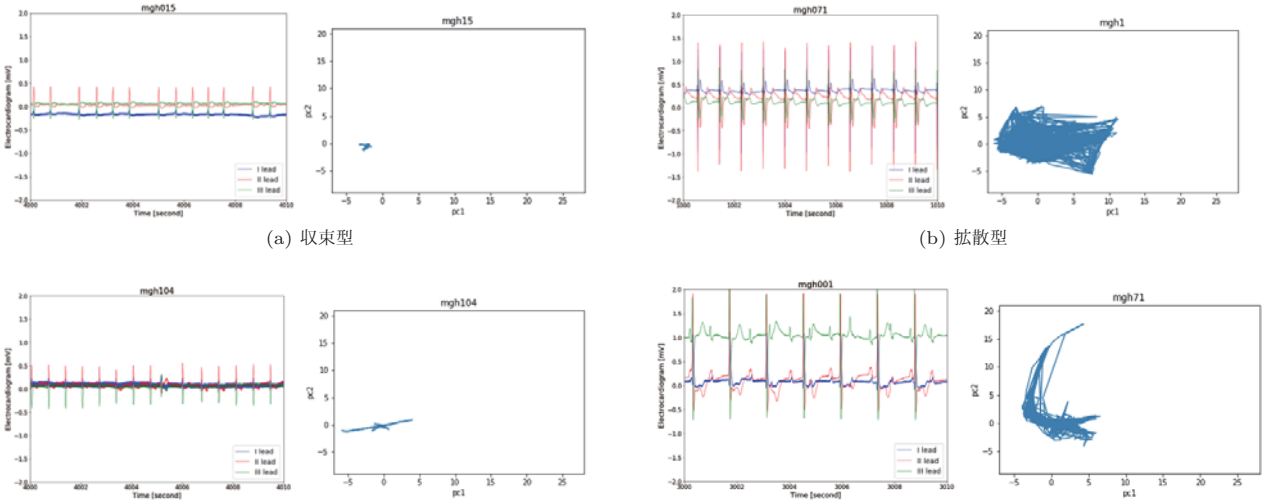
[高橋 17] 高橋 柊, 落合 桂一, 深澤 佑介 : Stacked convolutional denoising autoencoders を用いた 2 誘導心電図からの特徴抽出および不整脈分類, 情報処理学会研究報告, Vol. 2017-DPS-171, No. 25 (2017)

[平岡 15] 平岡 昌和 : 心電図自動診断の限界, 心電図, Vol. 35, No. 2, pp. 149–155 (2015)



(a) I 誘導　(b) II 誘導　(c) III 誘導

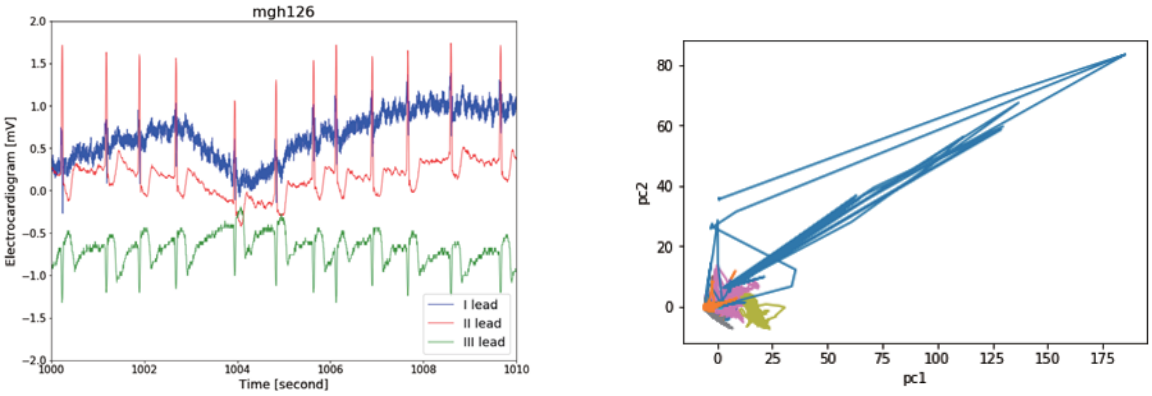
図 3: 未知データに対するオートエンコーダによる特徴量からの元信号の復元例.



(a) 収束型　(b) 拡散型

(c) 直線型　(d) 曲線型

図 4: 提案手法によって描かれた軌跡の例．4 つのクラスの典型的な例を示した．左側は信号の一部，右側は得られた軌跡を示す.



(a) 特異な患者の信号例　(b) 特異な患者の軌跡（青）と他の患者の軌跡（他色）

図 5: 定性的に発見された特異な患者に対する結果

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

「土地勘」は学習できるのか –機械学習によるエリアラベリング

Learning a sense of locality?: labeling local areas by machine learning

宗政一舟 *1　深見俊和 *2　森脇大輔 *2
Isshu Munemasa　Toshikazu Fukami　Daisuke Moriwaki

*1明治大学　*2株式会社サイバーエージェント
Meiji University　Cyberagent, Inc.

To improve the performance of location-based advertising, a model of "sense of locality" is estimated, where the output variable is the "label" of each location and inputs are geographical and demographic information associated with the location. As the input variables are all taken from the Internet, the output is unique dataset that we collects from people who know well the location. The model is estimated with three methods and XGBoost outperforms over logistic regression and SVM. The results show fairly predictive power with f1 score of 0.66.

1. はじめに

近年インターネット広告においては, 人々の行動履歴から適切な広告をプログラマティックに判断して配信する「行動ターゲティング広告」が主流となっている *1. その中で, オンライン上での行動に止まらずオフライン上での行動履歴からターゲティングを行う広告配信手法も登場している.

オフライン上での行動履歴は, スマートフォンなどの端末から得られる位置情報を人間が理解できる文脈に変換することで, ターゲティングに活用できる. 例えば, 土日に海辺に行ってる人は, 何らかのマリンスポーツを楽しんでいる可能性があり, 釣り道具やサーフボードへの関心が高いかもしれない. また, テーマパークによくいく人は, 他の面白そうなテーマパークに関心が高いかもしれないといった具合である.

ただし, 海辺やテーマパークなどのように行動が即座に読み取れる場所はそれほど多くない. 建物が密集する地域では, 住宅や商店, オフィスなど様々な用途の場所が混在しており, 何のためにそこにいるのかを位置情報だけから読み取るのは難しい. 他方で, 古書店街や道具屋街, 電気屋街など, 業種が偏ったまちが存在することはよく知られている. こうしたまちに頻繁に来る消費者は, ある特定の物事に関心が高いことが容易に予想され, 関心に応じた広告を配信することが可能である.

そのまちがどのような属性を持つかは, そのまちに長く住んでいる「土地勘」を持った人間には判別可能だが, そうでない場合は, 客観的なデータに基づいて分類することが必要である. 実際, 広告配信の現場では, 店舗の存在などから土地を分類し, その土地に来訪したユーザーをある属性を持つとみなすということ行われている. しかしながら, そうしたルールによる分類は, 実際の土地の雰囲気を見誤っている可能性や本来収集したい場所を見落としている可能性があり, 効率的とは言えない. データに基づいて長年そこに住んだ人がもつ「土地勘」を予測モデルに学習させることができれば, オフラインの行動履歴からより正確なターゲティングが可能である. こうした問題意識から, 本研究では, 土地勘のあるものによる土地のラベリングを元にオープンデータを用いた機械学習を行うことで, 日本全国の土地を分類するモデルを推定した.

本稿の構成は以下のようになっている.2. では, 関連研究について言及する.3. では, 分析手法について説明し,4. では研究に用いたデータについて解説する.5. では, 結果を示し考察を述べる.6. でまとめを行う.

連絡先: 宗政一舟，明治大学理工学研究科基礎理工学専攻，
〒 214-8571 神奈川県川崎市多摩区東三田 1-1-1,
E-mail: munemasa@cs.meiji.ac.jp

*1 [Book 2015] は主要なアプリ広告プラットフォームの一つである Google 社の AdMob (https://www.google.co.jp/admob/) を解析することで, ユーザーの属性や行動をもとにしたターゲティングを用いた広告配信は 4 割と推定している

2. 関連研究

人々の行動と土地の特性を有機的に結びつけて土地の分類を行った先行研究としては,[Noulas 2011] があげられる.[Noulas 2011] はロンドンとニューヨークのある区画を格子状に区切った上で, その小区画を SNS におけるユーザーの「チェックイン情報」に基づいてクラスタリングしている. このチェックイン情報は明示的なユーザー行動を示しており, その小区画がどのような土地なのかという情報を抽出することができる.[Noulas 2011] はさらに得られた土地のカテゴリーに基づいてユーザーのクラスタリングを行っている. また,[Kurashima 2013] は, レビューサイトと SNS から, ユーザーの行動履歴を抽出し, 潜在トピックモデルを当てはめることで, 地点の特徴を推定できることを示している.[Kurashima 2013] の手法は, 明示的に地点の特徴の情報がなくても, ユーザーの行動履歴から推定することができることを示している. 本研究は土地を格子状に区切っていることや, 明示的な土地の属性情報を用いている点で [Noulas 2011] に近いと言えるが, 分析者が直接, 土地の分類ラベルを設定し, それを土地勘のあるものにつけさせている点で異なる. 分析者が直接分類ラベルをつけることで, 実際にラベルを活用する際に使いやすい分類を行うことができることが利点である一方, 分析者が予期しないカテゴリはその他のカテゴリに埋没するため, 新たな知見が得られにくいという欠点があると考えられる.

3. 分析手法

本節では, 分析手法を解説する.3.1 で学習に用いた教師データの作成方法について説明し,3.2 において用いたモデルについて説明する.

3.1 教師データの作成

ほとんど人の訪れない山中などを除けば土地は特定の目的にラベルづけすることができる. 例えば, オフィスが密集している

場所は, オフィス街としてラベルづけすることができ, 日中そこにいる人は一定の確信を持って「サラリーマン」とか「OL」としてラベルづけすることができる. あるいは, 土日に海や山に滞在することが多い人は, アウトドア好きという推測ができる. 広告配信の観点からは，趣味嗜好が類似したユーザーが集まる場所を特定することが出来れば，該当地にいるユーザに対し，興味を持つ可能性が高いと推定される商品を効率的に訴求することができる.

筆者らはまずその土地にいる人の特性が解釈しやすいラベルを用意した. ラベルの選定にあたっては, 普段広告主とコミュニケーションを取っている営業担当からヒアリングを行い, 広告配信ターゲットを想定しやすいものとするとともに, あらゆる土地に一つ以上ラベルがつけられるように網羅性を確保するようにした. 例えば,「アウトドア」や「観光地」,「高級ショッピング街」といったその人の趣味嗜好や経済力が現れるようなラベルを用意する一方,「住宅街」や,「ビジネス街」,「工業地帯」といったそれだけでは人の属性が抽出しづらいものの, 多くの面積を占めている場所のラベルも用意するようにした. ラベル数は結果的に 25 となった.

土地のラベルづけの単位は概ね $100m \times 100m$ と, 全国 27 万地点を対象とした. 先行研究と比較すると [Noulas 2011] は $625m \times 625m$ のサイズのやや広いエリアを単位とする一方,[Kurashima 2013] はレストランなどの地点を分析と単位としており, 本研究はその中間にあるといえる. 分析単位を小さくするとラベルづけする地点の数が膨大となりデータ収集が難しくなる一方, 分析単位を広くとると, 本来の目的である広告配信の最適化に使えない精度の低いものとなってしまうため, そのバランスをとった.

上記の通り生成された 27 万地点の区画は, 協力者 15 名によって任意に選ばれた上でラベルづけされた. それぞれの協力者にはこれまで住んだことがある土地や勤務地など「土地勘」のある場所に限定してラベルづけをお願いすることで, ラベルに「土地勘」が反映されるようにした.「土地勘」には, 土地の用途や地価といった静的, 客観的な情報だけではなく, その場所に対する主観的な見方, その場所に集う人の属性など動的かつ主観的な情報をも加味されて生成されていると考えられる.

ラベルづけされた土地は,15,292 地点となった. その分布は, 東京都, 神奈川県, 大阪府, 愛知県, 奈良県, 沖縄県となっている.

3.2 モデル

こうして得られた「土地勘」を機械的に生成するため, 人がつけたラベルを公開されている土地情報で学習をさせ予測を行う. 人が土地を評価する際に明示的, 暗黙的に用いる情報としては, 主に住居, 商業地, 歓楽街といったその土地の用途, どのような人がその土地を利用するかに分けられると考えられる. 用途を学習させるためには, 存在する店舗や施設など建造物の情報, さらに駅からの距離などのアクセシビリティが必要となる. また, どのような人が利用しているかについては, その地域に住んでいる人の属性や地価などが有用であると考えられる.

これらを表す属性として, 本稿ではウェブ上で取得できる店舗のデータや, 公的機関が公表する統計情報を用いることとした. 単純な分類問題となるため, モデルは, ロジスティック回帰, サポートベクターマシン, 勾配木ブースティングを用いた. 勾配木ブースティングについては, 広く使われている Xgboost[*2] の Python ライブラリを用いた. モデルの推定にあたっては, マルチクラス分類ではなく, 土地がある特定のクラスに属しているかどうかを判定する 2 値分類問題として推定した. これは, 実際に広告配信を行う場合には, 全ての土地を分類するというより, 特定の属性をもつエリアを特定することがタスクとなるためである.

4. データ

学習に用いるデータセットは,3. に述べた協力者によるラベルデータ, ウェブ上で取得できる有力チェーンの店舗一覧などの商業施設のデータ, 国土交通省が公表している国土数値情報 [*3],2017 年国勢調査とした.

店舗データはチェーン名, 店舗名称と, 緯度経度からなり, 各チェーンに対してコンビニエンスストアや薬局, 服飾店といったカテゴリ名を付与した. こうして得た店舗データは, 総数で 50 万, カテゴリ数は 70 となった.

国土数値情報からは, 地価公示, 施設, 地域資源・観光の項目を用いた. 地価公示については, 駅からの距離, 建ぺい率, 地積, 容積率, 地上地下の階数及び地価を用いた. 施設については, 国・都道府県の機関, 市町村役場等及び公的集会施設, 福祉施設, 学校などの所在地データから土地単位でその数を特徴量とした. 地域資源・観光については, 観光資源の有無を特徴量とした. 地価については必ずしも全ての分類対象地点について地価が存在するわけではないので, 近傍の 5 地点の公示地価を平均することで推定した.

国勢調査については, 地域の就業状況を小地域集計の就業状態等基本集計に関する集計から, 大分類別の産業別男女別就業者数を割合に変換して用いるとともに,5 歳刻み年齢別人口を用いた. さらに, 都道府県単位の産業別年収を賃金構造基本統計調査から, 小地域別に産業別の労働者数を国勢調査から取得し, 加重平均することで小地域ごとに平均年収を推定し特徴量として用いた.

5. 結果

本稿では, 分析結果の一部として,「飲屋街」ラベルを推定した結果を示す.

5.1 節では各推定手法の性能比較を行う. 検証に当たっては学習に用いなかった 10 %のデータを用いた.5.2 節では特徴量を評価する.5.3 節では推定されたモデルを用いてラベルのついていない地点を含めた全国 27 万地点の予測を行った. 最後に 5.4 でモデルの考察をした.

5.1 推定モデルの比較

表 1 では, 各分析手法で「飲屋街」かどうかを推定させた結果を示している.$True$ は正例,$False$ は負例に対する予測を示している. 評価指標は, 精度, 再現率及びそれらを統合した f1 スコアとなっている.

実際の広告配信の際には, 飲屋街に滞在するユーザを特定することが目的になるので, 正例に注目して性能を比較すると, 勾配木ブースティングが他の 2 つのモデルを凌駕している. 特に精度については,87%と十分なものとなった. 他方で, 再現率については,53%とやや低くなっている. これは, 実際に飲屋街だと土地勘のあるものが認識していてもモデルがそのように認識しなかったものが半分程度あることを意味している. 与えた特徴量だけでは, 十分な判定ができていない可能性がある.

5.2 特徴量の重要度

図 1 は特徴量を重要度順に上位 10 まで並べたものである. 重要な特徴量とされたのは駅からの距離, 地積, 女性就業者の

∗2 [Chen 2016]

∗3 国土数値情報 http://nlftp.mlit.go.jp/ksj/

表 1: 分析手法の性能比較

methods	True/False	precision	recall	f1-score
ロジスティック回帰	True	0.50	0.08	0.14
	False	0.86	0.99	0.92
サポートベクターマシン	True	0.73	0.15	0.25
	False	0.87	0.99	0.93
勾配木ブースティング	True	0.87	0.53	0.66
	False	0.92	0.99	0.95

うちの宿泊業・飲食サービス業の割合, 推定所得, 建物の階数, 地価, 男性就業者のうちの金融業・保険業の割合, スポーツ系店舗の数, 女性就業者のうちの製造業の割合及び女性就業者のうちの不動産業・物品賃貸業の割合となった.

繁華街はアクセスのより駅近に密集しており, 駅周辺であることが重要な予測指標となっている. また, 地積は公示地価を調査する際の土地の面積であり, 土地が細分化された地域の方が飲屋街である確率が高いことを表している. 建物の階数については, 高いことが飲屋街率を高めている. 女性の産業比率については, 製造業が少なく不動産業が多いほど飲屋街確率が高まる. 工場などが立地する地域より, 不動産の営業所が並ぶ街中に飲屋街が発生しやすいということを表していると考えられる. ただし, 男性ではなく女性の職業比率がそのまちの属性をよく表していることについては, さらに研究が必要であると考えられる.

5.3 「土地勘」モデルによる予測

学習したモデルを用いて全国 27 万地点の土地を飲屋街かそうでないかについて予測したところ,1342 地点が飲屋街であると判定された. 図 2 は主要都市における推定された飲屋街の分布を示している.

東京では新宿の歌舞伎町や東口周辺, 渋谷のセンター街付近, 新橋, 銀座, 上野の一部など, 主な繁華街は網羅されているほか, 駅周辺が飲屋街として推定されている. 特に, 一般的な繁華街ではない円山町, アメヤ横丁や神楽坂など飲屋街としての色彩が濃いまちも漏れなく判定されている.[*4] これらのまちはチェーン店が密集しているというより, 個人営業の店も多くチェーン店の店舗情報だけでは判定しきれない. 本研究では国土数値情報や国勢調査など, より一般的な土地の属性を捉える工夫を行なっているため, こういったエリアについても特定することができている.

その他の都市を見てみると大阪では難波と梅田を中心に比較的広範囲に飲屋街が広がっていると予測されている. また, 名古屋は名駅周辺及び栄に判定がされている. 横浜については, 横浜駅周辺と, 関内が判定されているものの他の大都市と比較して狭い分布となっている.

5.4 考察

協力者のラベルづけの対象が東京を中心に大都市に偏っていることや, 協力者が「飲屋街」であると判断した地点が 798 にとどまったことなどを考慮すれば, 東京だけでなくその他の都市についても違和感の少ない予測ができていると考えられる. 特にチェーン店が密集している訳ではないむローカルな飲屋街についてもその他の情報から判定されることが示せた.

広告配信の観点からは, ウェブ上の情報だけでは直接知り得ないその土地の性格を推定できることは, 配信対象の拡大に繋がり, 意義が大きい. 従来であれば, 主要な居酒屋チェーンの店舗の位置を収集した上で, ユーザーの位置情報とをマッチングさせることでユーザーの属性を判定する必要があったが, ある程度面的に広がりを持つ土地を飲屋街として判定できれば, 大まかなユーザーの位置情報があれば, ユーザーの属性を捉えることができる.

他方で, 東京以外については飲屋街としての判定がやや辛すぎることが課題であると考えられる. 駅からの距離や地価, 推定所得が強く効いていることから, 鉄道路線が張り巡らされ, 地価や所得水準が群を抜いて高い東京において判定が緩くなっていることが考えられる. こうしたバイアスを除去するためには, 特徴量の正規化が必要になってくると考えられる.

6. おわりに

本研究では, ウェブ上で取得できる店舗の情報, 国土数値情報, 国勢調査のデータを用いることで, そこに住んだことがある人しか得られない「土地勘」を学習させ未知の土地についてその属性を予測を行なった. その結果, 東京に関しては, 居酒屋チェーンなどの店舗情報に頼らずとも飲屋街を判定することができた. 他方で, 学習に使ったデータに地理的な偏りがあることや, 特徴量の正規化がされていないことで東京以外における判定が辛くなっていることが課題であると考えられる.

参考文献

[Book 2015] Book, Theodore and Dan S. Wallach:An Empirical Study of Mobile Ad Targeting, 2015, arXiv:1502.06577 http://arxiv.org/abs/1502.06577

[Chen 2016] Chen, Tianqi and Carlos Guestrin.:XGBoost: A Scalable Tree Boosting System (2016), CoRR abs/1603.02754 (2016). arXiv:1603.02754 http://arxiv.org/abs/1603.02754

[Kurashima 2013] Takeshi Kurashima, Tomoharu Iwata, Takahide Hoshide, Noriko Takaya and Ko Fujimura.:Geo Topic Model: Joint Modeling of User ' s Activity Area and Interests for Location Recommendation (2013), WSDM 2013.

[Noulas 2011] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo and Massimiliano Pontil.:Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-Based Social Networks (2011), AAAI 2011.

*4 いずれも正解データは存在しない地点.

図 1: 特徴量の重要度

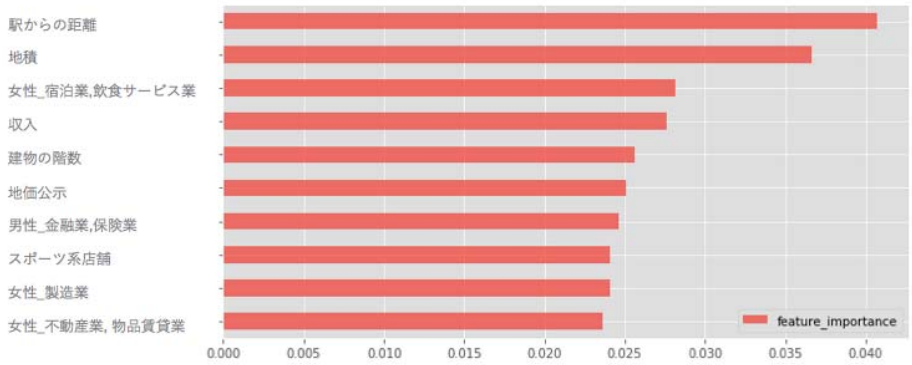
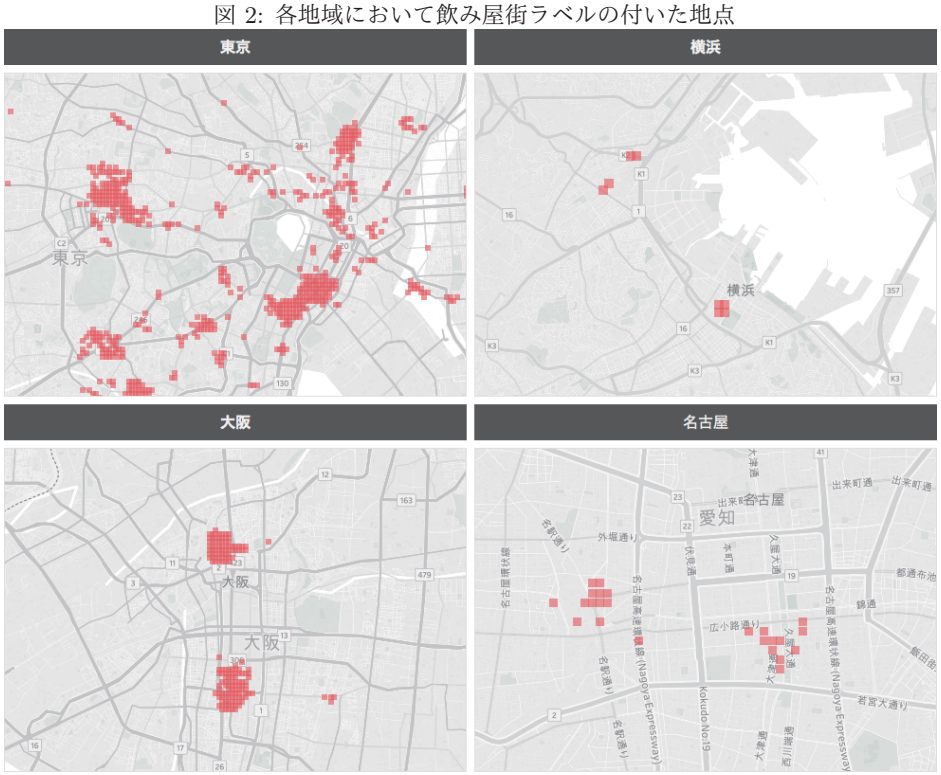


図 2: 各地域において飲み屋街ラベルの付いた地点



[2H1] [General Session] 4. Knowledge Utilization and Sharing

座長:川村 隆浩（科学技術振興機構）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room H (10F Sky Hall)

[2H1-01] Implementation of adaptive displaying function for action decomposition trees based on various viewpoints

○Masaaki Imazono[1], Yoshinobu Kitamura[1] （1. Ritsumeikan University）

9:00 AM - 9:20 AM

[2H1-02] Knowledge Sharing System in an Organization Using Extended Prediction Market

○Seiyu Yamaguchi[1], Hajime Mizuyama[1], Tomomi Nonaka[1] （1. Aoyama Gakuin University）

9:20 AM - 9:40 AM

[2H1-03] Toward an Elderly Care Ontology for Knowledge Sharing among Care Workers

○Yoshiki Fukushima[1], Satoshi Nishimura[2], Kenichiro Fukuda[2], Takuichi Nishimura[2], Yoshinobu Kitamura[1] （1. Graduate School of Information Science and Engineering, Ritsumeikan University, 2. Artificial Intelligence Research Center, National Institute for Advanced Industrial Science and Technology）

9:40 AM - 10:00 AM

[2H1-04] Constructing knowledge models and an ontology for adjusting specifications of industrial machinery

○Taisei Ido[1], Yoshinobu Kitamura[1] （1. Ritsumeikan University）

10:00 AM - 10:20 AM

[2H1-05] An Testing Environment to Accumulate Test Cases for Execution Order Variation

○Yusuke Aoyama[1], Takeru Kuroiwa[1], Noriyuki Kushiro[1] （1. Kyushu Institute of Technology）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

行為分解木に対する多様な観点に基づいた絞り込み表示機能の実装

Implementation of adaptive displaying function for action decomposition trees based on various viewpoints

今園 真聡[*1]
Masaaki Imazono

來村 徳信[*1]
Yoshinobu Kitamura

[*1] 立命館大学 情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University

The action decomposition trees show intended goals of actions and their procedures. In order to improve readability of large-scale action decomposition trees, this research aims to develop the software function to display the portions of the trees which are related with user-specified information from various viewpoints. We define the general and domain-specific types of the viewpoints, which can be used by the users to specify the kinds of the information that they need. For users' understanding of relationships between actions, the software also has a function to show the reason why the specific portions of the trees are displayed.

1. はじめに

行為分解木[溝口 15]とは，ある行為をどのように達成するかということを，部分行為に分解して木構造で表したものである．達成する行為を全体行為，全体行為を達成するための行為を部分行為，全体行為を達成するやり方を概念化したものを達成方式という．一つの行為のさまざまな達成のやり方を部分行為と達成方式の組として表すことができるため，行為の複雑な達成手順やさまざまな手法などをわかりやすく提示することができる．また，行為分解木を閲覧することによって，部分行為が達成する行為を目的として確認できるため，他の形式に比べ目的を明確に理解することができる．特に看護分野においては行為分解木を用いたツールの有用性が実際に示されている[西村 15]．

しかし，行為分解木の表現の形式上，さまざまなやり方を表すために行為の達成方式を追加したり，全体行為をより詳細な部分行為に分解したりすることなどによって，行為分解木の全体が大きくなる．その結果，ある状況において特定の情報のみに注目したいユーザにとっては，行為分解木から状況に応じた特定の情報を見つけるのが困難になるという課題があった．

そこで，本研究では，ユーザが状況に応じて入力した条件に従って，行為条件木の中の条件に適合するノードを強調表示したり，適合ノードのみから構成される部分木を表示したりすることで，ユーザが必要な情報を見つけやすくすることを目的とする．

本稿では，まず，絞り込みの条件を入力する際に用いる観点を定義する．これを「絞り込みに用いる観点」と呼ぶ．絞り込みに用いる観点を，一般的な観点と分野固有の観点に分けて，オントロジーの概念として，定義する．本研究では，分野固有の観点を「鋼板製造」と「看護行為」の 2 つの分野について定義を行った．そして，それらの観点を利用した行為分解木絞り込み表示ソフトウェアを開発する．定義した絞り込みの観点から絞り込みの条件となる項目を作成し，入力された条件値をもとに関連する行為分解木を，ユーザにわかりやすく提示することを目指す．また，結果だけでなく，なぜその行為が絞り込み項目と関連しているのかという理由を提示する機能を実装する．

本研究では，絞り込みに用いる観点を定義するオントロジーを「法造」を用いて構築する．法造は，オントロジーを構築・利用するための統合的環境であり，コンテキストにおける役割（ロール概念）を明示的に扱えるという特徴を持つ[溝口 06]．また，行為分解木の構築には OntoGearCore というモデリングツールを使用した．それぞれのデータは法造 API と OntoGearCore API を用いて参照するように実装した．

2. 絞り込みに用いる観点の定義

絞り込みに用いる観点とは，対象となる行為分解木を絞り込む際の条件の項目として用いる概念のことであり，行為分解木を構成する要素となる概念やその種類の概念のうち，絞り込みの条件として有用と考えられる概念セットのことである．そのため，まず行為分解木を構成する概念要素をオントロジーの概念として定義した．次に，定義した概念のうち，絞り込みに有用な観点と考えられるものを「絞り込みに用いる観点」として抽出した．

絞り込みに用いる観点は「一般的な観点」と「分野固有の観点」の 2 つに分類できる．行為分解木は分野に共通な枠組みに基づいて記述されるが，対象とする分野によって異なる概念も現れる．共通な記述枠組みを表す概念が「一般的な観点」に分類され，どの分野における行為分解木の絞り込みにも適用することができる．一方，主に行為の対象となる概念などの分野固有の概念は，「分野固有の観点」に分類される．これらは「一般的な観点」の概念の下位概念として，定義される．絞り込みの際には，「一般的な観点」と「分野固有の観点」のリストの中からそれぞれ選択することで，その組み合わせとして，行為分解木の絞り込みを行うことができる．

2.1 一般的な観点の定義

一般的な観点を定義するために，行為分解木の記述に用いられている概念をオントロジーの概念として定義した．本節では，「行為」と「状態」概念の定義と，その絞り込み観点について述べる．「行為」概念は図 1 の左側に示すように，「行為者」，「対象」，「方式」を部分概念としてもつような概念として定義されている．「行為者」とはその行為を実行するものであり，一般的な文章表現において主語にあたるものである．図1においてスロットのクラス制約概念として記述されているように，「行為者」は「行為者対象」クラス（人間やものなどを下位クラスとするクラス）のインスタンスでなければならない，一方，「対象」とはその行為の対象となるものであり，文章表現の目的語にあたる．クラス制約とし

て「具体物」が指定されており，物理空間上の範囲を占めるようなものが行為の対象となることが定義されている．最後に，「方式」は，その行為をどのように達成するかを示している．

行為概念は，図 1 の右側に示すように，「目的達成行為」と「不具合達成行為」の 2 つの下位概念に分類される．「目的達成行為」とは行為分解木の記述者にとって，達成されることが望ましい状態を実現するための行為であり，通常の行為分解木の要素である．一方，「不具合達成行為」は，行為を実行することによって問題が発生するような，行為分解木の記述者にとって望ましくない状態を引き起こす行為である．図 1 の右側では，不具合達成行為には，一般的に，「予防行為」，「発見行為」，「対処行為」が存在するということが定義されている．「予防行為」は，不具合達成行為が達成される前に予防するための行為，「発見行為」は不具合達成行為の兆候を発見するための行為，「対処行為」は不具合達成行為が行われた後にその影響に対処するための行為を意味している．予防行為・発見行為・対処行為は異なる行為分解木間をつなぐために有用であり，絞り込みを行う際にも利用できる．そのため，不具合達成行為の部分概念を「一般的な絞り込みに用いる観点」として定義した．

次に「状態」概念とその下位概念の定義を図 2 に示す．「状態」概念は「対象」と「時間」を部分概念としてもつ．「対象」は状態を持つようなものであり，「実在物」クラスのインスタンスである．一方，「時間」は状態の要因となるものの時間を表している．この「時間軸」クラスには「過去」，「未来」，「現在」などの下位クラスが定義されている．図 2 では「状態」が「過去に起因する状態」と「未来に関する状態」の下位概念に分類されることを示している．また，「過去に起因する状態」は「行為に起因する状態」と「性質に起因する状態」の 2 つに分類され，「未来に関する状態」は「未来の行為に基づく状態」と「未来の性質に基づく状態」の 2 つの下位概念に分類される．「行為に起因する状態」は，過去に行ったまたは行われた行為が原因で引き起こされる状態，「性質に起因する状態」は，過去に持っていた性質が原因で引き起こされる状態を表す．「未来の行為に基づく状態」とは，未来に行う予定の行為を目的として，現在の状態があることを表す．「未来の性質に基づく状態」とは，その性質を未来で持つことを目的として，現在の状態があることを表す．分類の順序としては，すべての状態が時系列情報を保持しており，また原因となった行為と性質のどちらかの特徴を持っていると考えられるため，時系列で分類した後，行為と性質で分類を行った．「状態」概念とその下位概念は，絞り込みの際の目的設定や，過去の事象から現在の注目すべきノードの絞り込みに利用できるため，「絞り込みに用いられる観点」として抽出した．

2.2 分野固有の観点の定義

本節では分野固有の観点の例として，鋼板製造分野固有の絞り込み観点について述べる．本研究では，他の研究においてすでに記述されていた実際の「鋼板製造」に関する行為分解木に基づいて，それに現れる概念をオントロジーとして定義した．まず，定義した概念の例として，「不具合達成行為」の下位概念である「硫黄(S)による熱間脆性」の定義を図 3 に示す．図 3 に示すように，「硫黄(S)による熱間脆性」という行為（不具合な現象）は，「成分改質方式」という方式によって「熱間脆性」という不具合が発生することを表し，その不具合は「シリコンやマンガンを加える」という行為によって予防できることを表している．また，「成分改質方式」は「溶鉄から不純物を除去する」と「成分を改質する」の2つの部分行為から成り立っていることを表している．定義した概念とその部分概念は他の概念と関係している可能

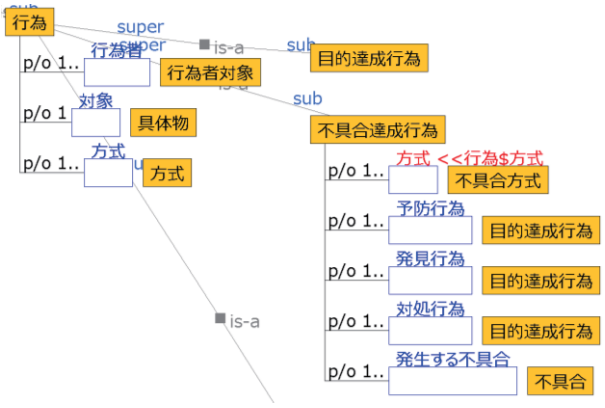


図 1 「行為」概念とその下位概念の定義

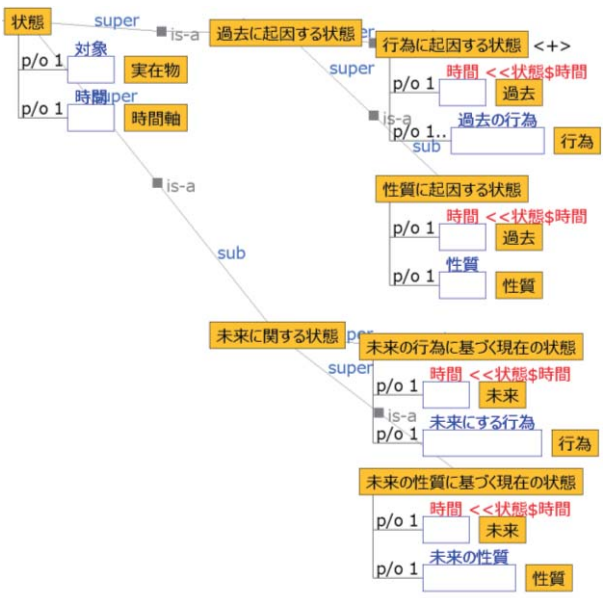


図 2 「状態」概念とその下位概念の定義

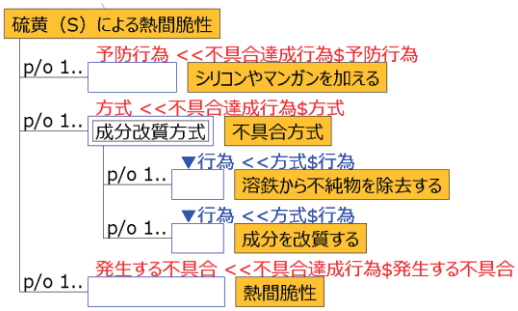


図 3 「硫黄(S)による熱間脆性」概念の定義

性があり，それらの関係性を調べるための絞り込みの条件に利用できるため，「分野固有の絞り込み観点」として抽出した．

同様に，看護分野における術後観察看護行為に固有な絞り込み観点の定義を行った．

3. 絞り込み表示ソフトウェアの開発

3.1 概要

図 4 に開発したソフトウェアのフレームワークを示す．ソフトウェアは入力データとして，「行為分解木データ」と前節で述べた絞り込み観点を定義した「オントロジーデータ」の 2 つを読み込んで動作する．「行為分解木データ」は OntoGearCore API を使用して，データを参照する．「オントロジーデータ」は，定義内容

を法造 API を使用して参照する．本ソフトウェアは，行為分解木データとオントロジーデータを読み込んだ後，まず，ユーザに対して「絞り込みに用いることができる観点」をリストとして提示する．ユーザがそのリストから絞り込みに用いたい条件を選択すると，本ソフトウェアは条件に従って，適合するノードやリンクのハイライト表示，または適合するノードのみに絞り込んだ行為分解木を表示して，ユーザに提示する．

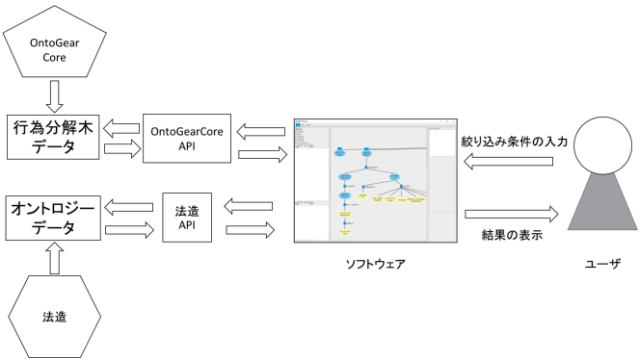


図 4 ソフトウェアのフレームワーク

3.2 動的な絞り込み条件の生成機能

図 5 はソフトウェアの絞り込み条件入力部分の画面を示している．まず，絞り込み条件は「固定項目」と「オントロジー依存項目」で構成され，2 つの項目を組み合わせることで絞り込みを実行する．前節で述べた「一般的な絞り込み観点」が「固定項目」に対応し，「分野固有の絞り込み観点」が「オントロジー依存項目」に対応する．「固定項目」として「予防行為」，「発見行為」，「対処行為」が提示され，ユーザが各項目にチェックを入れることで，対応するノードのみに表示を絞り込むことができる．「オントロジー依存項目」は前節で述べた「分野固有の絞り込観点」に基づいて，読み込んだオントロジーファイルのデータを用いて，選択した対象の概念のスロットのロール概念とクラス制限を参照して生成される．そのため，オントロジー設計者が提示する項目を自由に設定することが可能である．図 5 では，オントロジー依存項目として，「溶鉄」のスロットのロール概念である「原因となる行為」が表示され，右横の「選択」ボタンをクリックすることで，別ウィンドウとして図 6 が表示される。図 6 では具体的な分野固有な観点が表示され、ユーザは各チェックボックスを選択することで分野固有の絞り込み条件を設定することができる．

3.3 条件に基づく絞り込み表示機能

表示方法にはハイライト表示と絞り込み表示の 2 種類がある．ハイライト表示では，図 7 のように関連性があると考えられるノードとルートノードまでのエッジをハイライト表示する．絞り込み表示では，図 8 のように関連性があると考えられるノードとルートノードまでのパス上のノードを抽出した行為分解木の部分木を表示する．ハイライトまたは抽出されるノードは下記の条件に当てはまる概念である．

1. 絞り込み条件に指定された概念が，当該概念のスロットのクラス制約として記述されている概念
2. 下記の処理で取得できる概念の予防行為・発見行為・対処行為のスロットのクラス制約として記述されている概念
   ① まず，1 の条件を満たす概念が，スロットのクラス制約として記述されている概念を取得する．この条件を満たす概念を変数 A に格納する．
   ② 変数 A が示す概念，スロットのクラス制約にして記述されている概念を取得する．この条件を満たす概念を B とする．
   ③ 概念 B を変数 A に代入して，手続き②を再帰的に呼び出す．新たな概念が取得できなくなれば終了する．

このように前節で述べた絞り込み条件によって，他の行為分解木とのノードの関係性や明示的にされていないノード間の関係性が検索され，ハイライトされるノードが変化する．ハイライト表示によって，関連性があると考えられるノードと隣接しているノードを確認できるため，行為の流れが把握できる．絞り込み表示は，関連性があると考えられるノードのみを抽出するため，関連のある部分木の全体像の把握が容易になると考えられる．

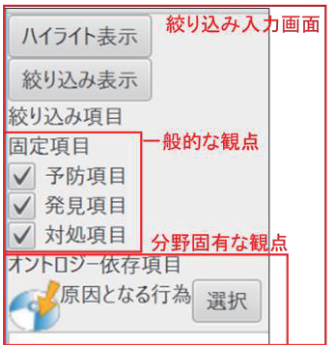


図 5 絞り込み入力画面

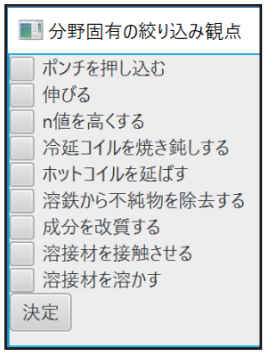


図 6 「選択」ボタンを押したときの画面

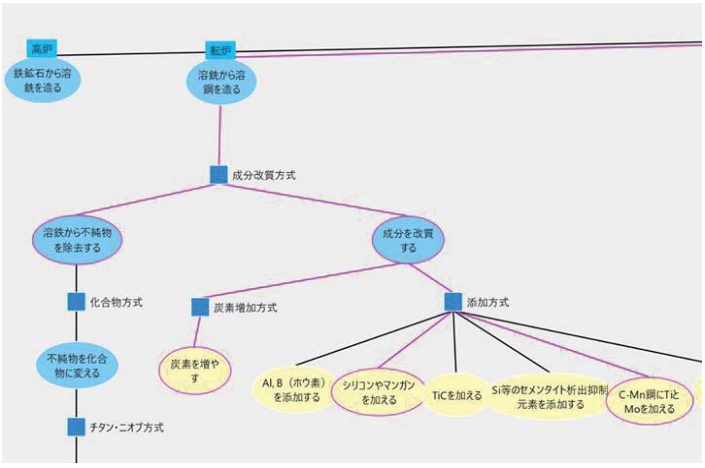


図 7 ハイライト表示による表示結果

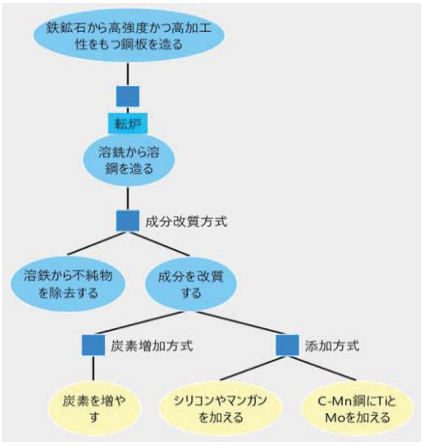


図 8 絞り込み表示による表示結果

3.4 絞り込みの理由の表示機能

本節では，絞り込み理由の表示機能について記述する．絞り込み理由とは，条件によってハイライトされたノードが，どの条件によってなぜハイライトされているのかを説明するものである，図 8 は絞り込み条件のオントロジー依存項目の「関連する生成物」として「ハイテン鋼」を設定した場合にハイライトされた「C-Mn 鋼に Ti と Mo を加える」というノードを右クリックしたときの画面を示している．右クリックが行われると，そのノードがハイライトされた理由を示す行為分解木があれば，その表示を行うためのメニューを表示する．図では「品質向上-[強度を上げる]を表示」と表示されており，これはクリックしたノードが品質向上を目的とすることを意味し，また強度を上げるためという理由があることを示す．この「品質向上-[強度を上げる]を表示」をクリックすると図 9 に示す「強度を上げる」ための行為分解木が表示される．この行為分解木には「C-Mn 鋼に Ti と Mo を加える」というノードが含まれており，このノードがどのように「強度を上げる」という全体的な目的に貢献しているかを理解することができる．このように，別の行為分解木中に同じ概念のノードを持つ場合や，検索対象の行為分解木のノードが予防行為，発見行為，対処行為を持つ場合に理由を表示できる．

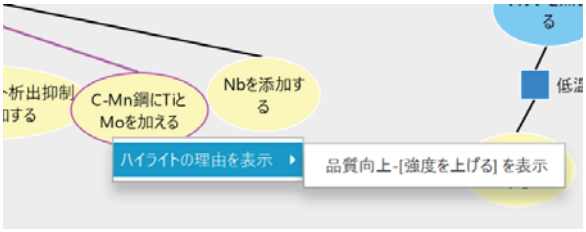


図 9 ハイライトノードを右クリックした画面

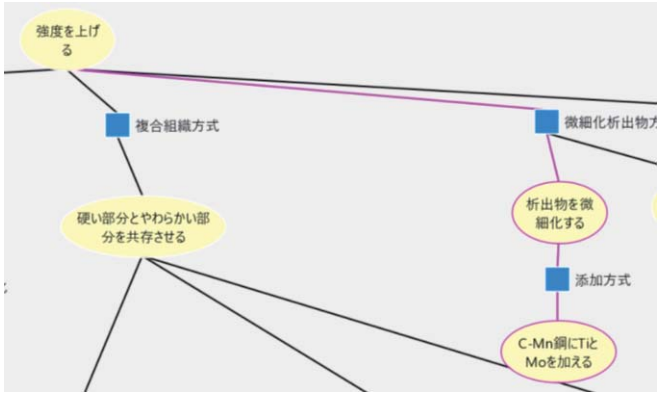


図 10 絞り込みの理由として「強度を上げる」という行為分解木を表示している画面

3.5 術後観察看護行為分解木への適応

本節では，本ソフトウェアを術後観察看護行為分解木に対して適用した例について述べる．図 11 はオントロジー依存項目の「患者の状態」の「選択」ボタンを押したときに表示される画面であり，現在，観察対象としている患者の状態を選択できる．図 12 は絞り込み条件として，「患者の状態」として「喫煙歴が長い」を指定した場合を示している．「呼吸器合併症のリスクを早期発見する」と「無気肺の有無とリスクを評価する」がハイライト表示されている．これは，前節で述べた「条件からの絞り込み」機能によって，オントロジーファイルに記述されている絞り込み観点の関係性を検索することで，絞り込み条件によっておこる合併症（不具合）の発見行為である「呼吸器合併症のリスクを早期発見する」と「無気肺の有無とリスクを評価する」がハイライトされる．ここではハイライト表示の画面のみを示しているが，図 8 と同じように，条件に適合するノードのみの絞り込み表示も可能である．

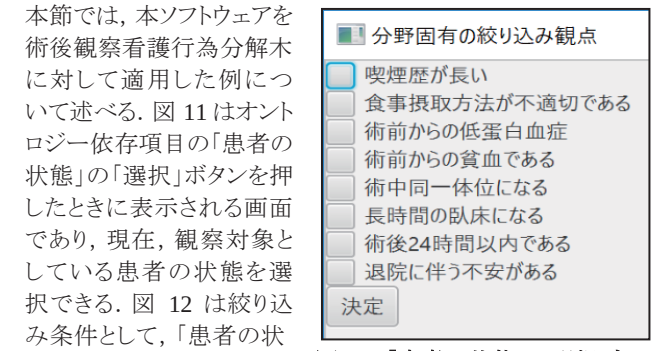


図 11 「患者の状態」の選択ボタンを押したときの画面

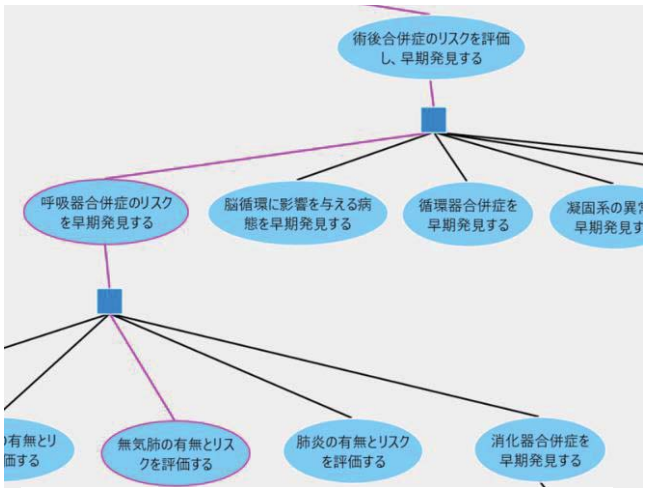


図 12 絞り込み結果のハイライト表示画面

4. まとめと今後の課題

本研究では，行為分解木を絞り込むための観点の定義を行い，その観点から絞り込み表示を行うソフトウェアを開発した．絞り込み観点として，行為分解木の記述枠組みを定義した「一般的な観点」と，特定の分野固有の概念を定義した「分野固有の観点」の 2 つを定義し，多角的に絞り込みを行えるようにした．

今後の課題は絞り込みの観点を追加することである．絞り込みの観点が少ないと，観点間の関係性を正確に記述できないため，正確な絞り込み表示がされない可能性があるためである．また，絞り込みの観点の増加によって，ユーザが設定できる条件が増え，ツールがより使いやすいものになると考える．

謝辞

OntoGearCore の開発を進めて頂いている，OntoGearCore 開発チームの皆様に感謝します．

参考文献

[溝口 15]溝口 理一郎，古崎 晃司，來村 徳信: オントロジー強化型シソーラス 工学者のための発想支援型情報検索を目指して，情報管理，Vol. 58, No. 5, pp. 361-371，科学技術振興機構，2015

[西村 15]西村悟史，笹嶋宗彦，來村徳信，中村明美，高橋弘枝，平尾明美，服部兼敏，溝口理一郎:目的指向の看護手順学習に向けた複数観点からの知識閲覧システム CHARM Pad と新人看護師研修への実践的活用，人工知能学会論文誌，Vol.30, No.1, pp.22-36, 2015

[溝口 06] 溝口理一郎，古崎晃司，來村徳信，笹島宗彦:オントロジー構築入門，オーム社，2006

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

拡張型予測市場を用いた組織内における知識共有システム

Knowledge Sharing System in an Organization Using Extended Prediction Market

山口 聖優 Seiyu Yamaguchi　水山 元 Hajime Mizuyama　野中 朋美 Tomomi Nonaka

青山学院大学
Aoyama Gakuin University

These days, knowledge sharing in an organization catches a great deal of attention, and there are various methods for knowledge sharing. However, several obstacles disturb their utilization. One of them is a motivation problem for sharing knowledge. To solve this, it is important to give incentives for them to provide their own knowledge. This work proposes a new knowledge sharing system, that introduces appropriate incentive for sharing valuable private knowledge by using gamification approach. The system is realized through an original extended prediction market mechanism with comment function. Further, a comment score is defined to gather higher quality knowledge. The effect of this system is tested in a laboratory setting.

1. はじめに

近年，あらゆる組織内において，知識共有が大きな注目を集めている．変化する環境で生き残るためには，組織内の知識を統合，管理，活用することがますます重要になっている．そのため，知識共有の様々な研究がなされている．特に，Web 技術は，必要に応じて検索，再利用，更新ができることから，知識共有で使用される最も有効な技術であると言われており [Wagner 06]，Web を介して知識を共有することは，多くの組織で最も一般的に利用されている技術となっている [Laudon 00]．知識共有システムには，従業員が共有するのを助けるグループウェアシステム，書類の共有をし，効果的に検索できるドキュメントマネジメントシステム等，様々な種類のシステムが存在する．しかし，これらのシステムは，導入はされるが，そのシステム上で十分な知識共有が行われず，維持できないという問題がある．知識共有は，知識を共有することが重要だと思わない限り，その価値を受け入れないため，個人のモチベーションに基づいている [Allameh 12]．したがって，個人のモチベーションを変えることが知識共有の課題であるといえる．

例えば，企業の営業部門の場合，訪問先企業の情報や，競合他社等の情報を獲得しておく必要がある．さらに，個人で積み重ねてきた交渉スキルや戦略といった暗黙知が非常に価値のある情報となる．営業案件を効率よく進め，かつ成功する確率を上げるためには，こういった情報や知識を部門内で共有する必要がある．しかしながら，各営業担当者は個人の成績のために互いに競争する環境でもあるため，彼ら自身の財産となる知識を他人に正直に共有させることは容易ではない．このようにして，モチベーションが保たれず，知識共有システムの導入が失敗に終わるケースが多く存在する．そのため，彼らが自分自身の知識を提供したくなるような動機を考慮した適切なインセンティブを与えることが重要となってくる．このようなインセンティブの問題を解決するための仕組みの一つとして，予測市場というメカニズムが存在する．そこで，本研究では，この予測市場を用いた知識共有システムを実装し，実験を行う．予測市場のメカニズムを導入することで，質の良い知識を提供するインセンティブが働き，知識共有の促進が期待できる．

連絡先: 青山学院大学水山研究室，神奈川県相模原市中央 区淵野辺 5-10-1 青山学院大学相模原キャンパス 0 棟 422

2. 予測市場を用いた知識共有システム

本システムで用いられる予測市場は，従来研究 [Mizuyama 14] に従って設計されている．これは，参加者の知識を収集し，所有者の交渉結果を予測するために予測市場のメカニズムを拡張している．ここで，予測市場とは，未来を予測するための先物取引市場のことである．予測したい物事の未来の状態に応じて事後的に価値が決まる仮想の証券を発行する．その証券を複数のユーザに仮想の市場で自由に売買してもらい，その過程で現れる証券価格の推移に基づいて，予測対象の未来の状態に関する動的な予測を得ることができる．

本システムでは，ある課題が，成功するか失敗するかの 2 つの証券を用意し，市場作成者以外の参加者がそれについて予測する，二択問題の予測市場ゲームとなる．このような単純な予測市場ゲームでは，成功するか失敗するかを判断するためのツールとしては機能するが，その主観的な根拠を共有することはできない．そこで，コメントシステムを導入し，拡張型予測市場を用いる．参加者が，幾つかの証券を売買すると意思決定した時，その結論に至る根拠をコメントとして記述してもらう．このコメントは全ての参加者が閲覧可能であり，他の参加者は，そのコメントに賛同した場合，そのコメントを選択して証券を購入することが可能となる．この機能により，市場参加者は，市場から得た情報を基に，自分自身の知識を提供することができ，市場作成者は，課題の成功する確率を確認しながら，参加者からのコメント，すなわちアドバイスを貰い受けることで，次の交渉を円滑に進め，成功する確率を上げることができる．

2.1 市場の流れ

営業担当者が顧客と交渉して注文を獲得する会社を対象とした例を用いて市場の流れを説明する．営業担当者（以下，オーナ）は，注文獲得までの行動計画と顧客に関する情報をシステムに入力する．他の参加者（以下，プレイヤ）は，オーナが入力した情報をみながら，自身の経験，知識，技術に基いてオーナの案件が成立するか失敗するかの主観的予測を提示する．そのため，交渉結果に対応して，Win と Lose の 2 つの予測証券が導入されている．Win 証券を購入することは，オーナが交渉を通じて顧客から注文を得ることに成功することを表す．Lose 証券は，その逆の意味を表す．さらに，プレイヤはコメ



図 1: 市場画面

ントシステムを利用し，新しい行動の提案や顧客の有用な情報等を提供する．オーナはそのコメントや予測推移を見ながら，次の行動を決定し，いくつかの追加情報を入力する．その際，次の行動の意思決定をするために最も参考になったコメントを選択する．これを繰り返し行うことで，オーナは交渉を成功させるための知識と技能を身につけることができ，プレイヤらはシステム内で知識を共有することができる．

2.2 ポイントの設定

市場は閉鎖されると，以下の式 (1) に基づいて配当が支払われる．つまり，プレイヤはこの事後価値を最大にするために予測市場ゲームで証券を売買する．

$$\begin{aligned} &\text{事後価値} \\ &= \text{所有する証券からの精算} \\ &\quad - \text{証券の購入による費用} + \text{証券の売却による収入} \end{aligned} \tag{1}$$

所有する証券からの精算とは，オーナが実際に交渉に勝利した場合，成立を表す Win 証券を保持している参加者に，1 証券あたり 100 ポイント支払われるポイントのことである．証券の購入，売却による変化は随時自身の所持金に反映される．

さらに，自身のコメントが，オーナにより，最も参考になったコメントとして選択された場合，コメントポイントが与えられる．また，自身のコメントに基づいて取引された証券の総数が最も高かった場合，その総数に対応したコメントポイントが与えられる．かつ，そのコメントを選択して取引を行ったプレイヤに対し，その取引数に応じたコメントポイントが与えられる．このようなメカニズムでポイントを付与することにより，質の高いコメントが提供，もしくは選択され，価値の高い知識の共有が行われることが期待される．

3. 本システムの実装

本システムは，上記のメカニズムを基に，Web アプリケーションとして実装し，Web ブラウザとして利用できる．実装にあたり，Python ベースの Web アプリケーションフレームワークの Django を用いた．図 1 に市場画面を示す．上部には市場についての詳細が記載されている．水色のフローチャートはオーナが入力した行動計画であり，現在の行動が濃い青色で塗りつぶされている．また，その下には，現在の行動におけるメモ欄があり，その行動に関する詳細な情報等を入力してもらう．これらはオーナが自由に編集ができる．価格におけるグラフは二種類存在し，左側の大きなグラフは縦軸に価格，横軸に時間の価格推移のグラフになっており，これが対象となる事象の予測となる．右側の棒グラフは現在の証券の価格を表している．中央には，取引の際の入力画面があり，売買する証券枚数とコメントを入力すると，証券を売買することができる．なお，円滑に取引を行うため，取引画面は，Ajax で実装している．また，図 1 の下部にはコメント画面が存在する．あるコメントを選択すると，コメント入力画面の位置にそのコメントが表示される．さらに，各コメントは取引数の高い順番に並び替えられ，価値の高い知識は上部にくるように設計されている．その他にも，市場作成画面，市場編集画面を実装した．

4. 実験計画

今回，本システムを使用してラボラトリ実験を行った．本実験では，オーナの被験者に対してミッションを与え，プレイヤ側はオーナがそのミッションを成功させるか失敗するかの予測をしながら取引をしてもらう．このミッションについて，以下の条件を満たすように設定する．

- 成功，失敗の結果が明確になるもの
- リアルタイムで進めながら，情報を共有することができるもの
- 期限を与えること

条件を満たすミッションの例として，「制限時間内に目的地に到着する」，「期限までに指定のシステムを完成させる」等が挙げられる．オーナは，ミッションが与えられてから自身の行動計画を立て，システムに入力する．なお，ミッション実行中は本システム以外見ることが出来ないとする．その後，市場が開かれ，プレイヤが参加し取引が行われる．

5. おわりに

本研究では，従来研究の予測市場メカニズムに従って知識共有システムの Web アプリケーション化を行い，実装した．実装にあたり，ページ遷移時間によるストレスを軽減するために，非同期で取引を行えるようにした．さらに，質の高い知識を得るため，コメントによるポイントの付与条件を設定した．また，実験を行い，本システムの有用性を検証した．

参考文献

[Allameh 12] S. Allameh, A. Abedini, J. Pool, & A. Kazemi: An analysis of factors affecting staffs knowledge sharing in the central library of the University of Isfahan using the extension of Theory of Reasoned Action, International Journal of Human Resource Studies, vol. 2, pp. 1(2012).

[Wagner 06] C. Wagner: Breaking the knowledge acquisition bottleneck through conversational knowledge management, Information Resources Management Journal, vol. 3,pp. 70-83(2006).

[Laudon 00] K. C. Laudon, & J. P. Laudon: Management information systems, New Jersey: Prentice Hall(2000).

[Mizuyama 14] H. Mizuyama & K. Yamamiya: The Shift from Teaching to Learning: Individual, Collective and Organizational Learning through Gaming Simulation, Edited by W. C. Kriz, wbv, pp.287-295 (2014).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

介護現場における知識共有のための介護オントロジー構築の試み

Toward an Elderly Care Ontology for Knowledge Sharing among Care Workers

福島良紀[*1] 西村悟史[*2] 福田賢一郎[*2] 西村拓一[*2] 來村徳信[*1]
Yoshiki Fukushima Satoshi Nishimura Kenichiro Fukuda Takuichi Nishimura Yoshinobu Kitamura

[*1] 立命館大学 情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University

[*2] 産業技術総合研究所 人工知能研究センター
Artificial Intelligence Research Center, National Institute for Advanced Industrial Science and Technology

This project aims to develop a software system for knowledge sharing among care workers in elderly care facilities. In order to provide a controlled vocabulary set for common description of the knowledge of care processes, the authors are currently building an elderly care ontology. It mainly defines the physical and cognitive activities in the care processes as temporal changes between states of target objects. This article also discusses the usages of the ontology in the knowledge sharing system, which include adaptive generation of hierarchical structures of the vocabulary.

1. はじめに

近年の高齢化社会における介護問題の解決に向けて，介護分野に IT 技術を適用して介護業務を効率化する試みが行われている [佐藤 2012]. しかしながら，介護対象者が人による介護を重んじて IT 技術を活用したいと思っていなかったり，介護分野に計算機可読なデータが少ない為に IT 技術が活動に適合しなかったりといった問題がある[西村 2017]. また，そうした状況であるため，介護士の技能は非常に重要である. 介護士は試験などを経て基礎的知識を身につけているものの，介護現場では使用可能な機器や被介護者の状態などによって対応は異なるため，介護士それぞれの属人的なノウハウによるところが大きいのが現状である.

こうした状況を受けて，本共同研究プロジェクトでは，介護分野の知識を計算機に蓄え，それを共有する基盤を整えることで，介護対象者に対して直接 IT 技術を適用するのではなく，介護士の業務を間接的に支援することで介護業務の質を向上させることを目的として研究を行っている[西村 2017]. 本プロジェクトでは，業務において必要となる行為や使用する道具の機能を総称した概念を人にわかりやすく説明したものをプロセス知識と定義する. さらに，プロセス知識を，介護士の個々のノウハウや設備といった，現場ごとに特有のプロセス知識である「固有プロセス知識」と，資格試験などで身に着ける基礎的知識である「共通プロセス知識」に分類する. このような固有プロセス知識はほとんどの場合暗黙的であり，現在，そういった介護分野の暗黙的知識を共有することを目的として，「プロセス知識収集システム」の開発を進めている[Nishimura 2017]. 現状のシステムでは，行為の記述を自然言語で自由に行うことが可能になっている. そのため，知識共有の際に曖昧な解釈が生まれる可能性があり，記述内容を統制する枠組みが求められている.

本研究では，開発中の知識収集システムにおいて介護行為の記述内容を統制する語彙を定義するオントロジー（以下，介護オントロジーと呼ぶ）の構築を目指している. 一般に，オントロジーは対象とした概念をどのように解釈したかを明確に示すことができる. さらに，単純な概念の階層構造のみの定義と比較すると，その概念が持つ性質，さらにコンテキスト依存の概念を定義することができる. そのため，概念の客観的視点からの表現力が高く，介護行為を記述する際の指標を提示する手段として適切である. また，概念の性質を記述可能であることからそれぞれの概念が持つ性質による分類が容易である. これらのことから，介護オントロジーをプロセス知識収集システムにおける統制語彙として利用することに，有用性が期待できる.

連絡先：福島良紀，立命館大学情報理工学研究科，
is0278fr@ed.ritsumei.ac.jp

2. プロセス知識収集システム

図 1 は，本プロジェクトで現在開発を進めているプロセス知識収集システムにおける実際の知識記述画面である[Nishimura 2017]. このシステムは，オンラインでの利用を想定した Web ア

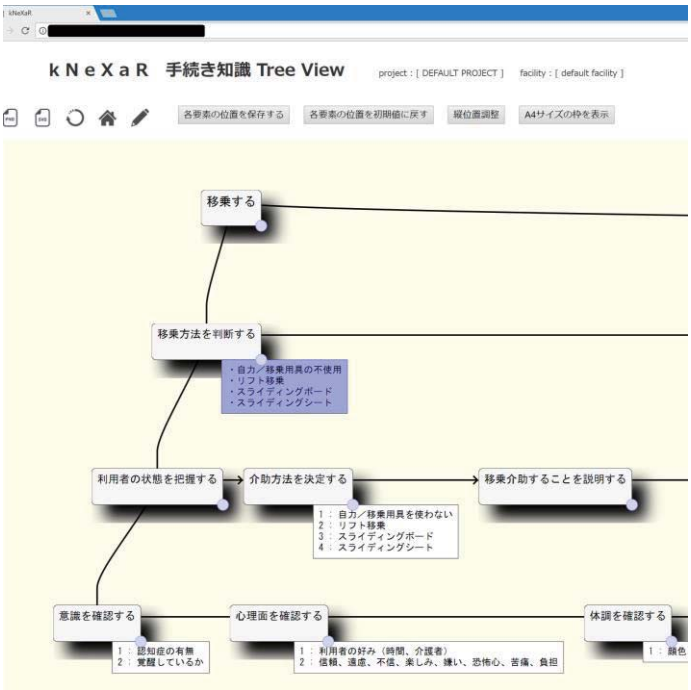


図 1 知識収集システムのインタフェース[Nishimura 2017]

プリケーションであり，介護分野における行為分解木の記述枠組みである CHARM [Nishimura 2013]を簡素化した枠組みに基づいて，介護プロセスを記述することが可能である．

図 1 の例では，「移乗する」行為を全体の目的として，それを達成する部分プロセスが記述されている．「移乗する」ためには，まず「移乗方法を判断する」ことが必要で，それはさらに細かい「利用者の状態を把握する」と「介助方法を決定する」などのプロセスを矢印の順序で行うことで達成可能であることが表現されている．

ここで，行為を明確に記述するためには，行為の客体となる名詞と具体的な行為内容を示す動詞を書かなければならない．そのため，行為の記述について名詞＋動詞の形で記述することが求められるが，現状では記述内容に制限は設けられていないため，名詞+動詞の形式でなくとも行為の記述が可能となっている．そのため，入力者が名詞を省略してしまったり，体言止めで行為を記述したりしてしまう可能性がある．また，名詞+動詞の形で行為を記述していたとしても，入力者が略語等の独自の表現を用いて行為を記述した場合，第三者と共有する際に意味を正しく伝えることができない可能性がある．このように，自然言語による自由な記述は，円滑な知識共有を妨げる恐れがある．そのため，この知識収集システム上で行為を記述する際に記述内容をコントロールできるように，介護分野の知識を体系化した統制語彙を導入することが求められている．

3. 介護オントロジーの構築

介護オントロジーの構築にあたっては，介護行為に関する基本的なプロセス知識を教科書 [大田 2014] [小林 2016]から抜き出して，図 1 の形式で記述したファイルを知識源として利用した．この形式から，名詞と動詞にあたる語彙をそれぞれ抜き出した．この作業に際しては，提供データに記述された行為数が膨大であったことから，まずは体位変換について記述されたファイルに注目して，知識を記述する際の形式を確立することを目標とした．オントロジーの構築には「法造」[溝口 2006]を用いた．法造では，概念の意味を主にスロットに対する制約によって定義する．図 2 において，「状態変化行為」というラベルから直角に折れた線で結びつけられているノード部分がスロットである．各ノードの右側で示されている概念はスロットに入るインスタンスのクラスの指定をするもので，クラス制約という．また，各ノードの上部には，スロットに入るインスタンスが担う役割を示す「ロール概念」が記述される．なお，上位概念からスロットを継承している場合には，スロット上部の文字列は赤字で表示される．

3.1 動詞

(1) 状態変化行為

動詞は，「状態変化行為」の下位概念である「物理的行為」「認知的行為」の下位に属する諸概念に対応する．「状態変化行為」は，[溝口 2006] を参考に，“行為はその行為に参加しているオブジェクトの状態を変化させることが本質である”という捉え方に基づいて定義されている．実際，図 2 に示すように，「状態変化行為」は「主体」「対象物」「間接対象物」というスロットをもち，それらは行為に参加するオブジェクトとして p/i(participate-in)関係で定義されており，それぞれサブスロットとして初期状態と終了状態が定義されている．これらの初期状態と終了状態の違いによって，その行為の意味が明確に定義される．例えば，図 3 に示すように，「説明する」という概念では，被説明者の終了状態が「理解状態」というクラスのインスタンスであることが，クラス制約として定義されている．これによって，「説明する」という



図 2 「状態変化行為」の定義

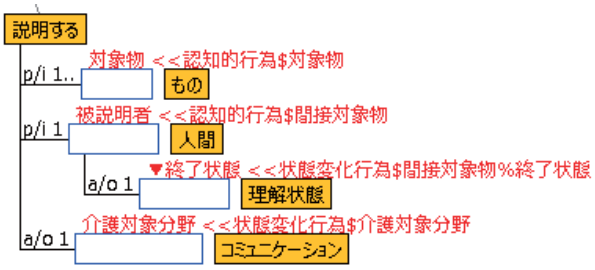


図 3 「説明する」の定義

行為は“被説明者の状態が「理解状態」に移行すること”と明確に定義される．また，「必要能力」というスロットをこの 3 つのスロットに定義している．これによって，その行為に参加する対象に必要となる能力を示す．

さらに「状態変化行為」には「介護対象分野」というスロットが定義されている．このスロットでは，ある行為が介護においてどのような分野に属するかを定義している．例えば，図 3 では，「説明する」という概念は「介護対象分野」のクラス制約が「コミュニケーション」となっている．これは，何かを説明するという行為が「コミュニケーション」という分野に属する行為である，ということを示している．

(2) 物理的行為

「物理的行為」とは，座標や形状のように物理的に観測可能な状態変化を起こす行為を示した概念のことである．図 4 の定義で示されているように，初期状態と終了状態のスロットはクラス制約が「物理的状態」となっている．「物理的行為」の下位概念の例として，「体位を端座位から立位へ変換する」がある．その定義を図 5 に示す．この概念では，状態が特殊化されているスロットは対象物のみであることから，対象物を変化させる行為で

あることが分かる．そして，初期状態のスロットはクラス制約として「端座位」を持ち，終了状態のスロットはクラス制約として「立位」を持っている．よって，この行為は，対象物の状態を「端坐位」から「立位」へ変換する行為であることがわかる．さらに，「残存能力」スロットを参照することで，この行為を受ける介護の対象者は「立位保持能力」を持っている必要があることが分かる．

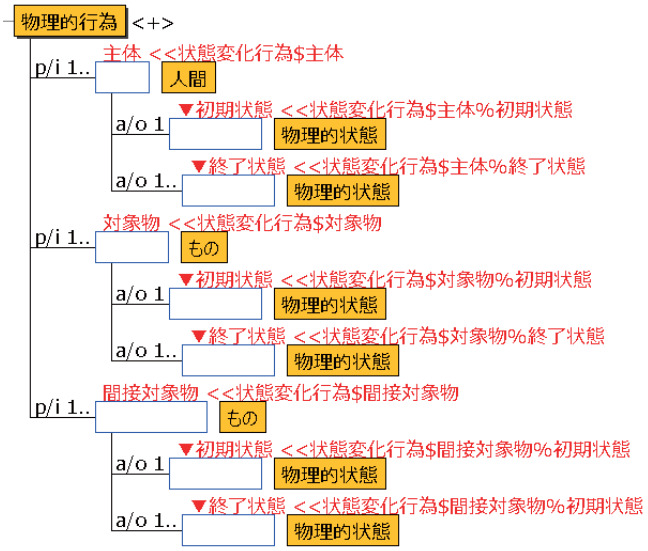


図 4「物理的行為」の定義

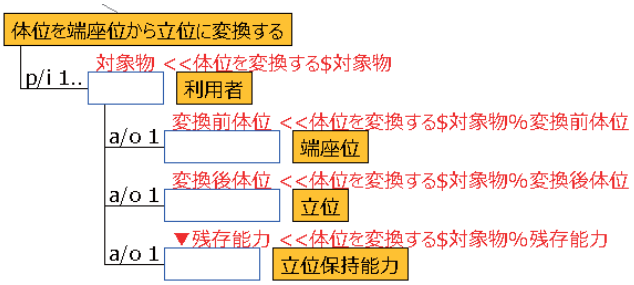


図 5「体位を端座位から立位に変換する」の定義

(3) 認知的行為

「認知的行為」とは，行為参加者の認知に変化を起こす行為のことである．図 6 に定義を示す．「状態変化行為」を特殊化した概念であり，初期状態と終了状態のスロットはクラス制約が「認知的状態」となっている．先ほど例に挙げた「説明する」という概念は「認知的行為」の下位概念に属している．なぜなら，何かを「説明する」という行為は，あるオブジェクトに対する被説明者の理解度(認知状態)を変える行為だからである．また，この概念では初期状態については特殊化をしていない．「説明する」以前における被説明者のオブジェクトに対する認知がどのような状態かに関係なく，行為後に「理解状態」へ移行するということを示すためである．

3.2 介護分野に関する概念

図 7 は，介護の「分野」に関する概念を示したものである．「分野」の下位概念には「リスクマネジメント」「コミュニケーション」「尊厳保持」「身体介護」「生活援助」という概念が定義されてお

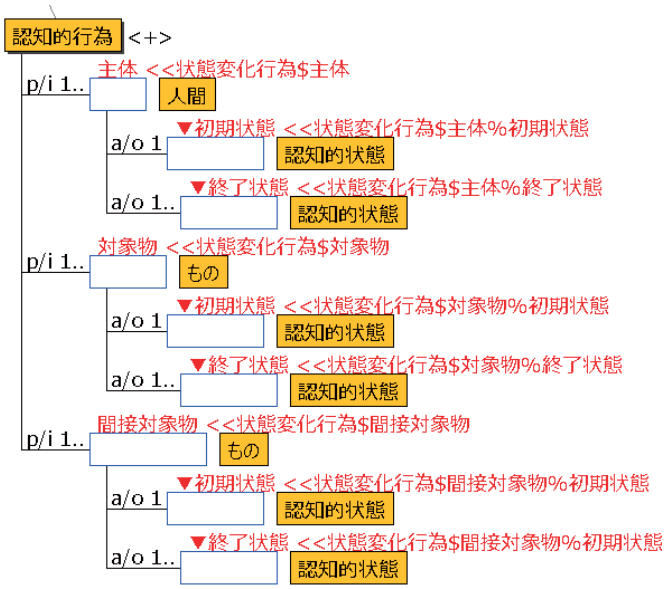


図 6 「認知的行為」の定義

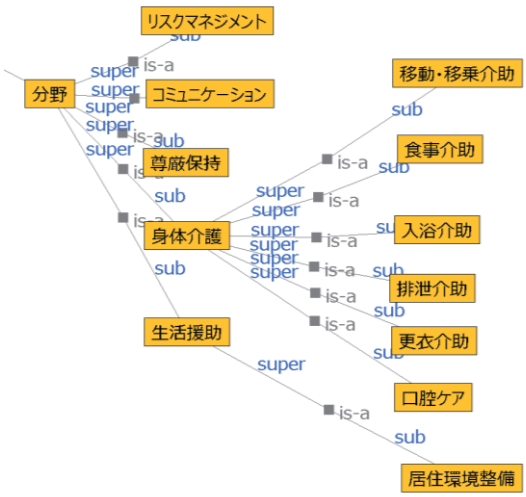


図 7 「分野」に属する諸概念

り，「身体介護」，「生活援助」の下位概念には更に細かく分類された介護分野概念が定義されている．なお，分野に属するこれらの概念は，教科書 [小林 2016]の中で，行為が記述されている箇所が説明している分野を目次等から確認して，定義した．介護分野に関する概念を利用することで，ある行為が属している介護の分野を指定することができる．例えば，図 3 の「説明する」という行為であれば，介護の分野では「コミュニケーション」という分野に属することが，概念のスロットにおいて定義されている．

4. システムにおける利用例

プロセス知識収集システムにおいて介護オントロジーを利用する方法の一つについて，利用例の現時点での想定画面を，図 8 に示す．これは，オントロジーに基づいて語彙リストを生成し，そこからユーザが選択することによって行為を記述していく方式である．ここでは，一連のプロセスの右端の「体位を変換する」という行為ノードを入力しようとしたときを想定している．そのとき，システムは，「行為分類」の下に示されるような，階層的な

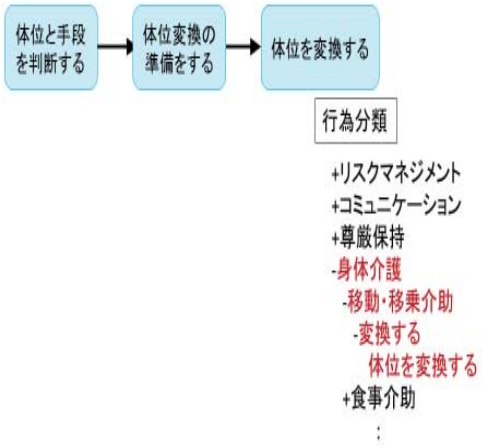


図 8　知識収集システムの想定画面

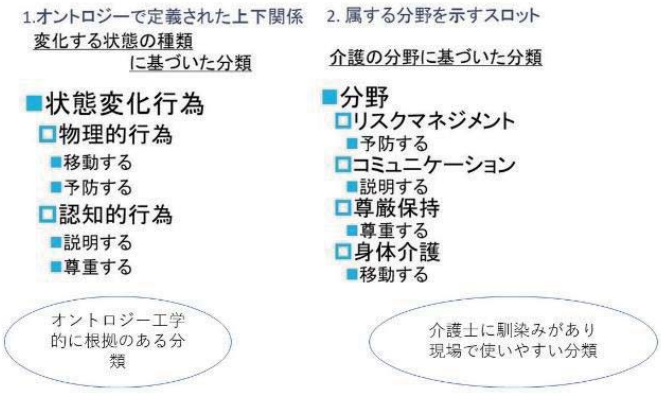


図 9 分類階層の比較

介護行為の語彙リストをユーザに提示する．ユーザがその階層に沿って，自分の入力したい語彙を選択することで，行為ノードに反映される．図 8 では，語彙リストを，赤字で示しているように，「身体介護－移動・移乗介助」と階層を辿ることで，「体位を変換する」を選択することで，上の行為ノードにそれを入力することができることを示している．

ここで，語彙リストの階層構造を生成する際には，図 9 に示すように，介護オントロジーで定義されている is-a 関係に基づく階層関係をそのまま反映する方式（図 9 左側）と，記述する行為が属する介護の分野を示すスロットを参照して生成する方式（図 9 右側）がある．前者の介護オントロジーで定義されている上下関係を利用する方式ではオントロジー工学的に根拠のある階層が生成され，後者の分野を参照する方式では介護士になじみがあり扱いやすい階層が生成されると考えられる．本知識収集システムは現場からの知識収集を目的としているため，分野を軸に語彙リストを生成する方式を採用することを予定している．

また，他の要素に基づいて階層構造をつくることも考えられる．例えば，行為概念で定義されている「必要能力」というスロットを参照することによって，介護行為の実行者または対象者が持っているべき能力ごとに，介護行為を分類して提示することもできる．

さらに，階層の単なる提示とは異なるオントロジーの利用方法として，行為参加者がその行為に必要な能力を有していなければ，知識記述者が語彙リストの中からその行為を選択できないように制限する，といったことも可能である．

このように現場でよく用いられる分類要素に基づくことで，オントロジーに詳しくない現場の介護士でも扱いやすい語彙リストとなることが期待され，知識収集システムが現場の介護士同士の円滑な知識共有を促進しうると考えられる．

5. おわりに

本論文では，開発中の知識収集システムに実装する記述統制語彙としての介護オントロジーの内容について述べた．介護オントロジーを用いて記述内容を制御する手法は，現場の介護士が持つ知識を正確に他者へ共有することを可能にすると考えられる．一方で，現在の介護オントロジーは体位変換を中心とした構築途中のものである．介護現場において実際に行われている行為については，文献[三輪 2015]において包括的な調査と分析が行われており，今後，これらの知見も参考として，現場の介護プロセス全体を記述できるオントロジーにむけてさらなる構築を進める．また，知識収集システムに実際にオントロジーを組み込み，4 節で述べたような利用方法に基づいた知識記述支援の実装を進める予定である．

参考文献

[佐藤 2012] 佐藤美恭, 工藤周平, 宮本道子. (2012). 高齢者介護施設における IT 活用に関する研究． 経営情報学会 2012 年秋季全国研究発表大会要旨集, B3-3, pp.295-298.

[西村 2017] 西村悟史, 大谷博, 畠山直人, 長谷部希恵子, 福田賢一郎, 來村徳信, 溝口理一郎, 西村拓一. (2017). 現場主体の"知識発現"方法の提案. 人工知能学会論文誌 32 巻 4 号 C C-G95_1-15.

[Nishimura 2017] Nishimura, S., Fukuda, K., and Nishimura, T., "Knowledge Explication: Current situation and future prospects", IJCAI 2017 Workshop on: Cognition and Artificial Intelligence for Human-Centered Design, pp. 1-7, 2017.

[大田 2014] 大田仁史, 三好春樹, 東田勉. (2014). 完全図解 新しい介護　全面改定版. 講談社.

[小林 2016] 小林一郎, 田中由紀子, 住居広士, 島津淳, 鈴木眞理子, 前田崇博. (2016). 介護職員初任者研修テキスト[第 3 版], ミネルヴァ書房.

[溝口 2006]溝口理一郎, 古崎晃司, 來村徳信, 笹島宗彦. (2006). オントロジー構築入門. オーム社.

[Nishimura 2013] Nishimura, S., Kitamura, Y., Sasajima, M., Williamson, A., Kinoshita, C., Hirao, A., Hattori, K., and Mizoguchi, R. (2013). CHARM as Activity Model to Share Knowledge and Transmit Procedural Knowledge and its Application to Nursing Guidelines Integration, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.17, No.2, pp. 208-220

[三輪 2015] 三輪 洋靖, 渡辺 健太郎, 福原 知宏, 中島 正人, 西村 拓一. (2015). 介護プロセスの計測と記述, 日本機械学会論文集, 81 巻 822 号 p. 14-00207

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

産業機械の仕様調整のための知識モデルとオントロジーの構築

Constructing knowledge models and an ontology for adjusting specifications of industrial machinery

井戸 大成
Taisei Ido

來村 徳信
Yoshinobu Kitamura

立命館大学 情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University

This collaborative research with a manufacturing company focuses on the adjustment process of design specifications of industrial machinery. The adjustment process consists of two steps; (1) to extract such specifications that need to be adjusted from the required specifications given by a customer and (2) to propose changes of those specifications for negotiation with the customer. In order to help engineers perform such processes efficiently, this research aims at knowledge modeling of past cases of adjusted specifications and ontology building for defining concepts in the knowledge models. The similarity calculation using the semantic structure of the ontology can improve the recall ratio of the extraction of the specifications to be adjusted. The semantic relations in the knowledge models could also contribute to higher accuracy ratio. Furthermore, the pattern knowledge of the past changes will realize semi-automatic generation of proposals of adjusted changes.

1. 研究対象と課題

製造業界における産業機械の多くは標準的な仕様を持ち，顧客から要求される仕様（顧客要求仕様）に合わせ，仕様のカスタマイズを行ってから製造される．顧客要求仕様の中には，産業機械の製造を行う側にとって，製造上の観点などから問題となりうる仕様が含まれることがある．このような仕様は「要注意仕様」と呼ばれ，顧客との間で，調整や変更が必要となるものである．つまり，製造側は，顧客要求仕様の中から要注意仕様をもれなく抽出して，それに対する適切な変更案を作成し，顧客側と仕様を調整する必要がある．このような一連の流れを顧客側と製造側の間の「仕様調整」問題として捉える．

これまで，このような仕様調整問題における「要注意仕様」の抽出は，人手によって行われてきた．しかしながら，顧客要求仕様の量は膨大であるために，抽出には大きな時間とコストがかかっていた．また，抽出と変更案の作成は，過去の要注意仕様事例と，機器の標準仕様に関する知識に基づいているが，そのような知識は属人的であり，抽出の精度は担当者の経験や力量に依存していた．さらに，本研究が対象としている産業機械では，ある 1 つの案件で要注意仕様として抽出された文章は，顧客要求仕様全体の約 45,000 文のうち，約 200 文であり，その比率は約 0.4%と非常に低い．そのような要注意仕様の全てを正確に抽出し，適切な変更案を作成することは困難であった．

このような仕様調整問題を計算機によって支援することを目指して，本研究の開始前に，自然言語処理による語句マッチングによる要注意仕様の抽出が試みられた．具体的には，過去の要注意仕様に関する文書をデータベースに格納しておき，その中から新規の顧客要求仕様書内に含まれる語句とマッチする文書を抽出する手法である．しかしながら，実際に実験を行ったところ，自然言語処理による文書抽出の精度が低いという問題があった．このような問題は他企業においても一般的に存在しており，仕様抽出の精度の向上は課題とされている［待井 2010］．また，この手法では，支援の対象が要注意仕様の抽出にとどまり，仕様の変更案の作成を支援できないことも，課題として認識された．

連絡先：井戸 大成，立命館大学情報理工学研究科ナレッジコンピューティング研究室，is0202he@ed.ritsumei.ac.jp

2. 研究目標とアプローチ

そこで，本研究では，産業機械を製造する企業との共同研究の一部として，仕様調整問題の担当者を計算機で支援するための，知識モデルとオントロジーの構築を目指している．

本研究が目標とする支援システムのフレームワークを図 1 に示す．システムは内部に過去の受注案件における顧客要求仕様に関する文書の全文を格納したデータベースと，仕様調整問題に関する知識モデルとオントロジーを持つ．それに対して，担当者が新しい案件における顧客要求仕様に関する文書を与えると，支援システムが，(1) 顧客要求仕様の中から「要注意仕様」である可能性が高い仕様を抽出して提示する，(2) 可能であれば抽出された要注意仕様に対してどのような変更を顧客に提案すべきか（変更仕様案）を提示する，という 2 つの機能を実現することを目指す．

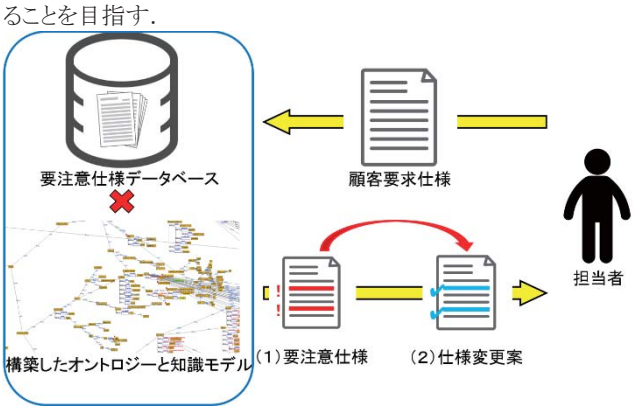


図 1 目標とする仕様調整支援システムのフレームワーク

本研究における課題は，前節で述べたように，要注意仕様の抽出の際の精度であり，特に，抽出漏れが少ないように，再現率を高めることを本研究の第1の目標とする．また，要注意仕様の抽出だけではなく，適切な仕様の変更案を生成することも目標であり，過去に完全に一致するような要注意仕様の変更事例

が存在する場合に加えて，完全一致ではないが意味構造的に類似性が見られるものに基づいて，変更仕様案を自動的に生成することを第 2 の目標とする．

これらの目標を実現するために，本研究では，まず，過去の要注意仕様の仕様変更の事例に関する知識を，知識モデル化する．次に，知識モデルに現れる概念をオントロジーとして定義することを行う．

3. 構築した知識モデルとオントロジー

3.1 仕様調整の表現枠組みの同定

まず，要注意仕様が含まれる場合の仕様調整の流れを表現するための枠組みを定義する．これによって， 3.2 節における事例の知識モデルの記述枠組みを同定する．

顧客からの要求仕様の中の要注意仕様に対して製造側が検討を行い，検討結果仕様を得て，仕様変更案として提案することになる．このとき，顧客要求仕様と検討結果仕様の 2 つの仕様の間においては，顧客要求仕様内の要素に対して，製造側によって部品の代替や削除といったなんらかの行為(“Action”)が行なわれている．また，その行為には製造上の観点などの合理的な理由(“Reason”)が存在している．以上のことから，仕様調整の流れを表現するために，図 2 の例の中で示されているような，顧客要求仕様に対して，製造側がある理由(“Reason”)である行為(“Action”)を行った結果として，検討結果仕様が得られる，という枠組みを同定した．実際の仕様調整は，この流れを双方の合意に至るまで繰り返すことで行われる．

3.2 過去事例の知識モデル化

前節で同定した枠組みに沿って，過去の要注意仕様の変更事例を知識モデルとして記述した．現在までに，共同研究先より提供を受けた 183 個の事例から，147 個を知識モデルとして記述した．この知識モデル内で出現する概念は全て 3.3 節におけるオントロジーで定義する．自然言語文で記述されていた過去事例を知識モデルとして表すことによって，文書内における重要な概念や概念間の関係が明確にすることを目指した．

図 2 に記述した知識モデルの例を示す．本例は顧客要求仕様が「RTD エレメントに電圧保護装置をつけてほしい」(原文は英語である)というものであり，これに対して製造側としては「RTD エレメントには大きな電圧がかけられることがないため，電圧保護装置は不要である」という判断がなされたため，要注意仕様として抽出されている．つまり，顧客要求仕様内の電圧保護装置を，検討結果仕様内では削除する(“Delete”という“Action”を行う)こととなる．削除する理由(“Reason”)は，「不要である」からである．以上を知識モデルとして表したものが図 2 である．

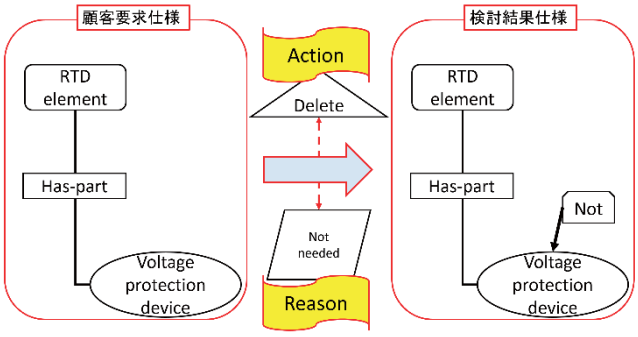


図 2 過去事例の概念的知識モデルの例

図 2 の左部分で表されている顧客要求仕様では，RTD エレメント(“RTD element”)に電圧保護装置(“Voltage protection device”)を設置することが，両装置を “Has-part” という関係によって結びつけられていることで表現されている．それに対して，右部分で表されている検討結果仕様では，電圧保護装置に対して「存在しない(“Not”)」という修飾子が付けられていることによって，顧客要求仕様から電圧保護装置が削除されていることが明示的に表されている．

このように概念的に記述した知識モデルに現れる概念をオントロジーとして定義し，次にその概念を用いて知識モデルを計算機的に記述する．それぞれを，3.3 節と 3.4 節で述べる．

3.3 知識モデル記述のためのオントロジー構築

次に，前節で述べた知識モデルに出現する概念をオントロジーとして定義した．オントロジー定義には，「法造」[古崎 2006]を使用した．現時点のオントロジーは，約 400 個の基本概念と約 300 個のロール概念が定義され，分類階層は最大で 10 段となっている．

例として，仕様変更を伴う“Delete”といった行為を，設計者が行う行為(“Designer’s Action”) の下位概念として図 3 に示すように定義した．“Designer’s Action”概念は，上位から順に，“Action”概念，“Person’s Action”概念の下位概念である．まず，“Action”概念は主体が対象物の状態変化を起こすことを表す．この定義においては，行為に関わるものが果たす役割を表す「ロール概念」として，“Doer”，“TargetObject”，“ResultObject”が定義されている．それぞれ，行為の主体，行為によって働きかけられる対象物，行為の結果として得られるものや状態，という役割(ロール)を表す．これによって，“Doer”のロールを担ったものが，“TargetObject”ロールのものに対して働きかけた結果，“ResultObject”の状態のものが得られるという概念として，“Action”概念を定義した．

次に，“Person’s Action”は人物が主体となって特定の時間的フェイズにおいて状態変化を起こすことを表す．図 3 左側に示すその定義では，“Doer”ロールを担えるクラスが“Person”と指定されており(クラス制約と呼ぶ)，主体が人間に限定されている．また，ロール概念として定義されている“time”はその行為が行われる時間的なフェイズを表しており，クラス制約は“Temporal_phase”となっている．

その下位概念である，“Designer’s Action”は設計者が主体となって設計仕様を変化させる行為を表す．その定義では， 主体となる“Doer”ロールを担えるものが設計者(“Designer”)，“time”ロールにおける時間的フェイズ(“temporal_phase”)が設計段階(“designing_phase”)，行為の対象物(“TargetObject”)や結果物(“ResultObject”)が設計仕様(“DesignSpecification”)であることが，クラス制約をそれぞれ特殊化することで，定義されている．

図 2 に現れている“Delete”概念は，この“‘Designer’s Action”の下位概念として図 3 右側のように定義されており，設計仕様(“DesignSpecification”)の中の特定の要素の数が減ることが，スロットの個数制約によって，定義されている．

また，“Delete”と同階層には，「確認項目の同定」を意味する“Identify confirmation item”，「追加」を表す“Add”，「変更」を表す“Change”が定義されており，それぞれがスロット制約の特殊化により仕様の中の要素に対しての行為を意味する．

また，仕様変更における「理由」(“Reason”)を，なにかが要因となって，結果として設計者の行為が起こるものとして図 4 に示すように定義した．例えば，”Standard of XXX” は，製造側の標準が理由であることを表す“Standard of Vender”の下位概念として，特定の製造会社 XXX 社の標準仕様が理由であることを表している．また，“Reason”の下位概念として，部品に起因する場合の“Device Factor”や，アフターサービスに起因する場合の“Factor of After Service”などが定義されている．これらも“Action”以下の概念と同様，それぞれの概念がスロット制約の特殊化により“Designer’s Action”の理由として定義されている．

3.4 計算機的知識モデルの記述

次に，3.2 節で概念的な知識モデルとして表した要注意仕様変更事例のうち，顧客要求仕様を計算機的に記述する．3.3 節でオントロジーとして定義した概念を用いて，記述した．

図 5 は図 2 に示した概念的知識モデルを計算機的に記述したものである．“RTD element”に“Voltage protection device” が設置されているものであるということが “Has-part” 関係を用いて，表されている．

これによって，次節で述べるように，語句の共起関係だけではなく，両者の間の関係性に基づいた抽出が可能になる．

4. 知識モデルとオントロジーの利用

本節では，前節で述べた知識モデルとオントロジーを利用することで，本研究の目標である，要注意仕様の抽出の際の精度の向上と，仕様変更案の生成が，どのようにして可能になると考えられるかについて述べる．

まず，3.3 節で述べたオントロジーの概念階層を用いることで，単なる語句マッチングではなく，意味レベルでの類似度の計算ができるようになると考える．例えば，配管の応力解析に関して，“stress analysis”と“piping”の組み合わせが過去において要注意仕様に現れていたとする．通常の自然言語処理の手法では，“stress analysis”と“pinping”という語句の共起によって要注意仕様を抽出する．構築したオントロジーを利用することで，“analysis”（解析）の下位概念である“calculation”（計算）と，“piping”と同じ意味である“pipe”の組み合わせも要注意仕様である可能性があることが，過去にそのような事例がなかったとしても，判断できるようになる．これにより，再現率が向上し，抽出漏れの可能性が下がることが期待できる．

さらに，3.4 節で述べた顧客要求仕様を表す計算機的知識モデルに基づくことで，語句の共起関係だけではなく，語句が表す概念の関係性に従って，要注意仕様を抽出することができる．例えば，上述の例で，行為“stress calculation”と部品“pipe”が単に共起している全ての場合を要注意仕様として抽出するのではなく，それらの間に “Target”という関係，つまり配管部品に対して応力解析（計算）行為が行われているという関係があるときのみを，要注意仕様として抽出することができる．これによって，適合率の向上が期待できる．

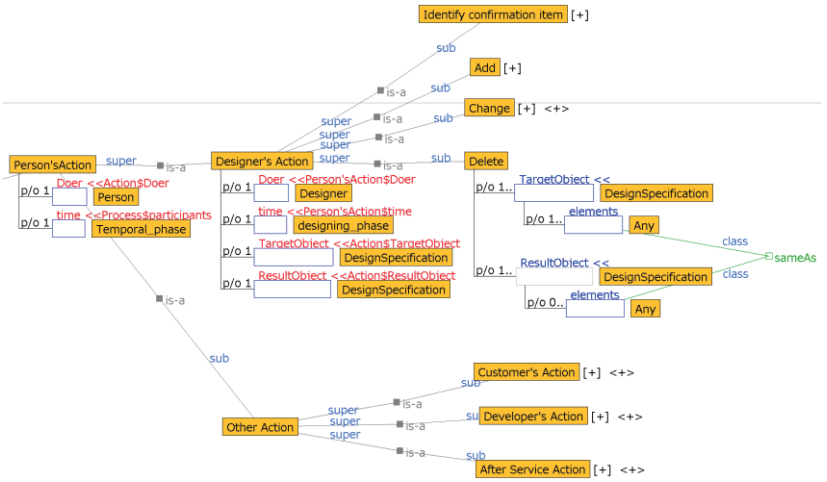


図 3 オントロジーの“Delete”行為概念の定義部分

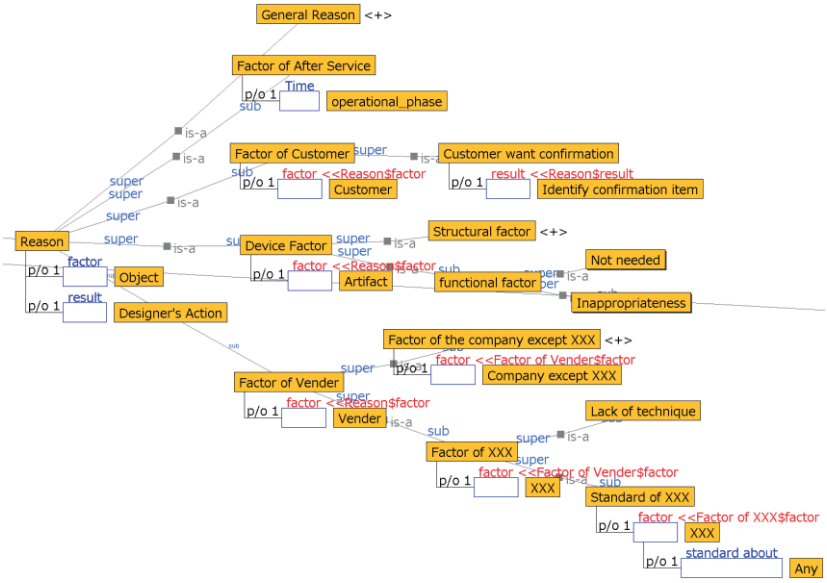


図 4 オントロジーの”Reason”概念の定義部分

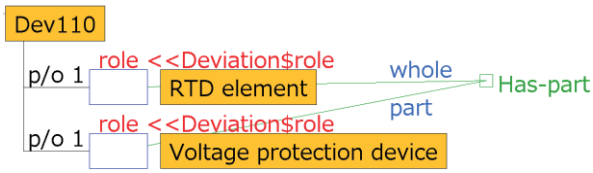


図 5 過去事例の計算機的知識モデルの記述例

さらに，要注意仕様の抽出のみでなく，変更案の示唆をできると考えられる．これは，3.2 節で述べた知識モデルの変更後の検討結果仕様の構造と，3.3 節のオントロジー，さらに今後の研究で確立していく変更パターン知識を組み合わせることによって可能になる．例えば，新規顧客要求仕様として「ソレノイドバルブを防爆仕様にしてほしい」という要求があったとする．このとき，まず，3.3 節で述べたオントロジーを参照することで，「防爆仕様」の兄弟概念にあたる，異なる「仕様」を表す概念は類似度が高いことが分かる．その中でも特に，ソレノイドバルブと is-type の関係で結ばれているような仕様概念は，ソレノイドバルブの種類（仕様）であることが分かり，例えば，「周波数タイプ」という概念の類似度はさらに高いことが分かる．これらに基づいて，今後

確立していく予定である変更パターン知識から，変更行為(”Action”)が「変更」(“Change”)であり，その理由(“Reason”)が「製造側の標準として周波数タイプを使うこととしている」(”Standard of XXX”)であるというパターンが適用され，「ソレノイドバルブは周波数タイプのものを使う」という検討結果仕様が生成され，担当者に提案することが可能になると考える．

5. 結言

産業機械の仕様調整問題に対して有効であると考えられるアプローチの提案を述べた．得られた知識モデルとオントロジーの妥当性は，共同研究先の専門家によって確認されている．

また，知識モデルとオントロジーに基づいて要注意仕様を抽出するプログラムの開発を進めている．現在までのところ，4 節で述べた知識モデルとオントロジーの利用の仕方のうち，最初に述べたオントロジーに基づいた類似度の計算と，2 番目に述べた計算機的知識モデルに現れる概念要素の組み合わせを利用したプログラムが実装されており，実際の案件の事例データを用いて，検証実験が行われている． その結果，文章間のコサイン類似度を用いた手法と比べて，オントロジーと知識モデルの利用によって，再現率が約 2 倍に向上することが明らかになっている．例えば，4 節でも述べた配管の応力解析に関して，この実験で実際に，過去の要注意仕様事例に現れていた“piping”と“stress analysis”の組み合わせとして記述された知識モデルとオントロジーに基づいて，コサイン類似度を用いた手法では抽出できなかった，“pipe stress calculation”というフレーズを含む文章を要注意仕様として正確に抽出できた．今後，知識モデルにおける要素間の関係性なども利用するプログラムを開発することで，要注意仕様の抽出におけるオントロジーのさらなる効果の検証を進める予定である．

他の残された研究課題として，以下の2点を挙げることできる．まず，オントロジーで定義された概念の網羅性を確認する必要がある．本オントロジーは事例から構築したものではあるが，対象機器依存部分を除くと，仕様調整問題に一般的に用いることができるものであると考えられる．このことを，さらに，追加で過去の変更事例を知識源として，どの程度の概念を追加で定義する必要があるかを調べることで，確認したいと考えている．予測としては追加で定義する概念数が減少していき，また，追加する必要がある概念もオントロジーの概念階層のリーフ部分のものであることが期待される．次に，変更パターン知識を精査・確立し，システムへ組み込むことで，変更案の自動的な生成ができることを確認することが挙げられる．

参考文献

[待井 2010] 待井君吉，横田毅，尾花充，後藤仁一朗：英文契約書評価支援システムの開発，情報処理学会研究報告，Vol. 2010-NL-197, No. 1, 2010

[古崎 2006] 古崎晃司，來村徳信，笹島宗彦，溝口理一郎：オントロジー構築入門，オーム社 ,2006.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

順序入れ替えテストケースの蓄積を実現するテスト実行環境

An Testing Environment to Accumulate Test Cases for Execution Order Variation

青山裕介 *1　黒岩丈瑠 *1　久代紀之 *1
Yusuke Aoyama　Takeru Kuroiwa　Noriyuki Kushiro

*1 九州工業大学
Kyushu Institute of Technology

Many kind of Internet of Things (IoT) devices are recently developed. The IoT devices connect to an open platform and communicate each other with message exchange. Since the order of the message exchange is not deterministic, developers tests IoT devices in exhaustive order. The tests are usually conducted manually, which brings two issues: p1) testers possibly miss some of the order; p2) testers possibly overlook some evaluation items.

As a solution, we developed a testing environment, which has a model checker that communicates with software on actual devices or an emulator. The testing environment solves p1) by automatic evaluation. The testing environment solves also p2) by automatic and simultaneous evaluation of evaluation items expressed as a Linear Temporal Logic (LTL) expression by integrating multiple LTL expressions with and-operator.

We applied the test execution environment to three market defects and confirmed that the test execution environment tested the defects correctly.

1. はじめに

今日，オープンなプラットフォームに接続し，複数製品を協調動作させる Internet of Things (IoT) 製品の開発が広く行われている．この IoT 製品には次の特徴がある．

1. IoT 製品がマルチベンダによって開発されること
2. IoT 製品間でのメッセージ交換が非決定的な順序で発生すること

上記の特徴 1, 2 から，製品評価において，開発した IoT 製品のメッセージの交換が順序の仮定なく動作すること，すなわち，あらゆる順序でメッセージ交換を発生させても問題が発生しないことを確認する必要がある．

一方でこうしたあらゆる順序を網羅するようなテストケースは膨大な数になってしまう．このため，現実的には例えば障害発生のリスクに鑑みてテストする対象を絞り込む方法 [Craig 02] などにより，限定的な種類のメッセージ交換について順序を入れ替えたテストを実施する．

現状こうしたメッセージ交換順序を入れ替えるテスト（以降，順序入れ替えテストと呼ぶ）はテスト技術者が手動で実施するために，絞り込みを実施した後も次の課題が存在する．

p1. 順序入れ替えテストケースの操作手順の網羅ミス

p2. 複数ある評価項目の評価の見落とし

このため，テストケースを蓄積して製品シリーズに渡って継続的に不具合を検出していくことが難しく，検出できるはずの不具合をテストをすり抜けて流出させ，過去不具合が再発してしまう可能性があった．

本研究では課題 p1, p2 を解決するために，

s1. 順序入れ替えテストケースの自動生成・実行

s2. 複数評価項目の自動評価

を実現することで順序入れ替えテストを自動化するテスト実行環境を開発した．また，s1 を個別のメッセージ交換のテストケースを元に行うことで，継続的な順序入れ替えテストのためのテストケースの蓄積を実現した．

2. 提案するテスト実行環境

節 1. の s1, s2 実現のためのテスト実行環境は図 1 のような構成となっている．本研究に先立ち開発したテスト実行環境 [Kuroiwa 16] を用いることで，s1 を実現し，さらに [Kuroiwa 16] を拡張することで節 1. の p2 の実現を図る．

2.1 拡張前のテスト実行環境

[Kuroiwa 16] は図 1 ③，⑤ を搭載し，③ ↔ ⑤ ↔ ⑥ の間でメッセージ交換を行う．このメッセージ交換をモデル検査器 Spin [Holzmann 04] を拡張したものにより行うことで，実装されたソフトウェアの順序入れ替えテストの自動実行を実現している．

Spin は従来は設計の評価に用いられてきたツールであり，入力言語として Promela を扱う．Promela 中でプロセスごとのシーケンス（本論文においては個別のメッセージ交換のシーケンス）を定義して Spin に入力すると，シーケンス同士の順序の入れ替えを自動的に網羅して実行することができる．

[Kuroiwa 16] はこの Spin の自動的かつ網羅的なテスト実行機能を実装されたソフトウェアに適用できるように，Spin に次のような拡張を施している．(詳細は [Kuroiwa 16] を参照されたい)．

1. PC と実装されたソフトウェアを搭載する組み込み機器との間の通信装置（図 1⑤）の開発
2. 実装されたソフトウェアの状態初期化機構の Spin への追加

2.2 テスト実行環境拡張内容

節 1. s2 は，Linear Temporal Logic (LTL) [Clarke 99] を使って実装済みのソフトウェアについての評価項目の記述・評価の実施を可能にすることで実現する．

以降の小節にて，まず LTL を使って節 1. の p2 を実現する評価項目記述方法について，続いて LTL で記述した評価項目の評価を実施する方法について述べる．

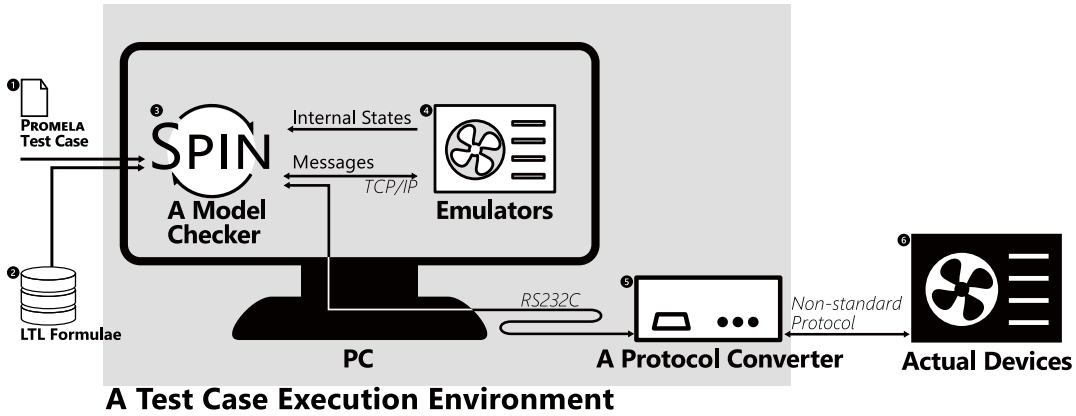


図 1: テスト実行環境の構成

2.2.1 LTL による評価項目の記述

順序入れ替えテストの評価では，組み合わせた個々のメッセージ交換シーケンスがどの順序で起こるかを一つに決定することはできない．しかし，個々のメッセージ交換を評価するために，非決定的ながらもテストの実行シーケンスのある時点を指定し，「いつ」評価を実施すればよいのかを評価項目中で表現する必要がある．

このような「いつ」を表現した評価項目の記述のために，本研究では評価項目の記述に LTL を採用した．

LTL は「いつか（記号: ◇）」や「常に（記号: □）」といった時系列情報を扱うことができる論理であり，Spin は標準で評価項目として扱うことができるものである．この LTL の採用により，メッセージ交換（受信 → 応答）のシーケンスの最終結果として「いつか」期待する状態になることの確認（例:「常に，応答実施後（P）にはある状態（S）である（$\Box(P \rightarrow \Diamond S)$）」）のための，あるいは複数のメッセージ交換の組み合わせが問題を起こさないことの確認（例:「ソフトウェアの内部状態が望まない状態になる (S) ことはずっとない ($\Box\neg S$)」）のための評価項目を表現可能となる．

本研究と同じくモデル検査器を用いて実装済みのソフトウェアをテストする手法 [Vries 00, Tretmans 03, Larsen 05] と比較すると，LTL 式を用いたことで，テスト判定において利点がある．[Vries 00, Tretmans 03, Larsen 05] ではテスト結果の判定として，ソフトウェアへの入力に対する出力が妥当かという入出力の関係を用いる．一方，本研究では LTL 式により，ソフトウェアの内部状態の状態変化によってテスト結果を判定する．このため個別の入出力の関係に加え，テストシーケンス全体に渡った異常状態の不発生，というような入出力に直接紐付かない評価も可能になっている．

また LTL 式は論理式であるので，個別のメッセージ交換の評価項目を論理演算子`&&`(And) で結合することで，複数の評価項目をまとめて確認する評価項目が表現可能である．

個別のメッセージ交換についての評価項目と，過去不具合事例から得られた評価項目を図 1 ② に蓄積しておくことで，順序入れ替えテスト実施時に組み合わせ，まとめて評価することができる．

2.2.2 LTL で記述した評価項目の評価の実施

節 1. s2 の評価の実行を実現するために，エミュレータ [Kuroiwa 15] を使ったソフトウェアの内部状態の取得機能を [Kuroiwa 16] に追加した．

エミュレータ [Kuroiwa 15] は API としてエミュレータ上で動作しているソフトウェアの内部状態を取得する機能を提供している．このエミュレータから取得したソフトウェア内部状態を期待値と比較する LTL 式を記述することで，Spin によって自動で評価し，節 1. s2 を実現できる．

表 1: ビル用空調システムを構成する機器

機種	1 年あたりのリリース台数
室内機	約 50 台
室外機	約 50 台
リモートコントローラ	約 5 台
システムコントローラ	約 1–2 台

LTL 式中での内部状態の比較自体は Spin の標準機能で実現できる．Spin の入力となるテストケース記述用の言語 Promela には，任意の C 言語プログラムを埋め込む構文 `c_expr` というものがある．構文 `c_expr` は，`c_expr` 直後に続く{}の内部に埋め込んだ C 言語の式を評価し，評価結果を真偽値として扱うことができる．`c_expr` は LTL 式を記述する構文の中にも書けるので，C 言語プログラムとしてエミュレータから取得したソフトウェア内部状態と期待値を比較する式を書くことで，節 1. の s2 を実現できる．

3. 実験

3.1 実験対象システム

本論文の評価のために提案手法を適用するビル用空調システムについて説明する．

適用対象のビル用空調システムは表 1 のような複数機器が含まれる．機器は同じネットワーク上でメッセージ交換を行いながら動作しており，ネットワークを流れるメッセージの到達順序は保証されていない．危機感のメッセージ交換順序に仮定が置けないことに鑑みると，各機器をそれぞれ一つの IoT 製品とみなすことができる．

表 1 のように毎年複数の機種がリリースされ，システムのネットワーク上に接続する機器は，リリースされた年次を問わず任意に組み合わせることを許容している．このため，接続される機器の違いによらず，任意の順序で発生するメッセージ交換が問題なく動作することを評価する，順序入れ替えテストが必要なシステムである．

表 2: 不具合事例概要

No.	概要
1	**単一機能のメッセージ交換順序の誤った仮定:** メッセージの到達順序の保証されないネットワーク上で順序の仮定を置いてしまっていた．この結果，ある複数メッセージの送付を必要とする機能が定められた順序でメッセージを受信しないと作動しなかった．
2	**機能の競合:** 機能 A と機能 B が同時に実行されると，機能 B が変更する状態によって，本来影響を受けるはずのない機能 A の振る舞いが変わった．
3	**モードの違う機器の競合:** 機器 X のモードに応じて機器 B の状態を変える機能について，モードの異なる機器 A と機器 C が共通の機器 B に本機能を実行した結果，制御される状態が一定しなかった．

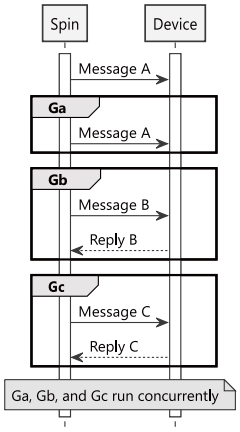


図 2: 表 2 の No. 1 のテストシーケンス

表 3: 表 2 の不具合が再発していないことを確認する LTL 式

No.	LTL 式
1	$\Diamond(StateS == value)$
2	$\Diamond(StateC' == StateC)$
3	$\Box(StateA! = StateT)$

表 4: テストの結果

No.	結果	期待結果
1	事例発生せず	事例発生せず
2	事例発生	事例発生
3	事例発生せず	事例発生せず

3.2 実験内容

本研究で開発したテスト実行環境を用いて，LTL 式を使った自動的な評価が実施可能であることを確認する．このために，節 3.1 のビル用空調システムに対し，後工程へ不具合が流出した表 2 の 3 件の事例を本テスト実行環境でテストする．

確認に用いるテストケースの操作手順としては，表 2 の不具合事例を引き起こした図 2–4 のテストシーケンスを用いる．検証に用いる LTL 式は，不具合事例が再発していないことを確認する表 3 を使用する．

4. 評価結果

テストの実行について，対象の機器には，表 2 の No. 1 と No. 3 については不具合を修正済みのものを用い，No. 2 については不具合を修正する以前の機器を用いた．テストの実行の結果実行の結果，表 4 の成否判定結果を得た．

5. 考察

表 4 の結果を見ると，不具合を残した機器で実施した表 2 の No. 2 の事例については，意図した通り不具合の発生を検出したことが確認できる．一方，不具合を修正済みの機器に対して実施した表 2 の No. 1, No. 3 の事例については，意図通りテストがパスしたことが確認できる．これは本テスト実行環境により偽陽性も偽陰性なく評価が実現できたことを示している．ここから，テスト実行環境によって，LTL 式で記述した評価項目を用いた自動的評価が可能であることを確認できた．

また，メッセージの送信順序の網羅について，表 2 の No. 1 のテストは，$3! = 6$ 通り，表 2 の No. 2 のテストは，$7!/6! = 7$ 通り，表 2 の No. 3 のテストは，$2! = 2$ 通りと，いずれの事例でも起こりうる全ての順序でメッセージを送信するテストが実施されることを確認した．この内，表 2 の No. 1 の事例から 2 例を挙げると図 5, 図 6 のようになっており，図 2 で並行に動作すると示されている Message B，Message C の送信順序が入れ替わっていることが確認できる．この結果から，本テスト実行環境により節 1. の s1 が実現できていることが確認された．

6. まとめと今後

本研究では，機器間のメッセージ交換の順序を仮定できない IoT 機器について，任意の順序のメッセージ交換をテストするテスト実行環境を開発した．本テスト実行環境では，従来の人手によるメッセージ交換順序の入れ替えテストの持つ次の 2 つ課題:

p1. 順序入れ替えテストケースの操作手順の網羅ミス

p2. 複数ある評価項目の評価の見落とし

を防ぐために，LTL 式を用いた実装済みのソフトウェアの自動評価を行うことで

s1. 順序入れ替えテストケースの自動生成・実行

s2. 複数評価項目の自動評価

の実現を図った．この LTL 式を用いた実装済みのソフトウェアの自動評価が実現できていることを確認するため，3 件の不具合事例をテスト実行環境でテストした．この結果，開発したテスト実行環境は，適用対象の過去不具合事例を自動的かつ正しく評価可能であることを確認した．

今後は s1，s2 が達成されたことを確認するために更に評価を充実させる予定である．

謝辞

本研究は，JSPS 科研費 16K00100，16H01836，JST CREST の支援を受けたものである．

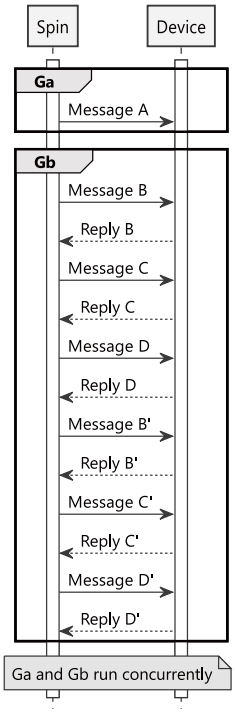


図 3: 表 2 の No. 2 のテストシーケンス

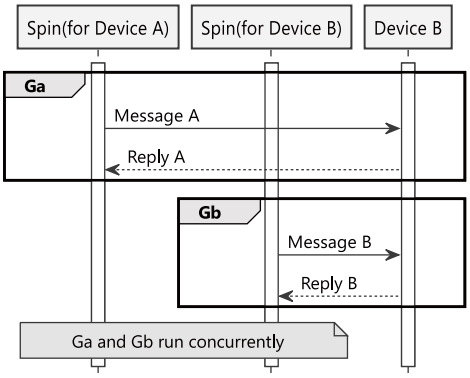


図 4: 表 2 の No. 3 のテストシーケンス

参考文献

[Clarke 99] Clarke, E. M., Grumberg, O., and Peled, D.: *Model checking*, pp. 31–33, MIT press (1999)

[Craig 02] Craig, R. D. and Jaskiel, S. P.: *Systematic software testing*, pp. 17–34, Artech House (2002), 宗 雅彦（監訳） and 成田 光彰（訳）, 体系的ソフトウェアテスト入門, 日経 BP 社, 2004

[Holzmann 04] Holzmann, G. J.: *The SPIN model checker: Primer and reference manual*, Vol. 1003, Addison-Wesley Reading (2004)

[Kuroiwa 15] Kuroiwa, T. and Kushiro, N.: Testing environment for embedded software product lines, in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–7IEEE (2015)

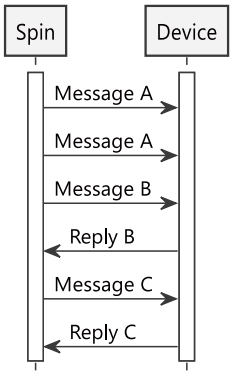


図 5: 表 2 の No.1 の実行結果の例 1

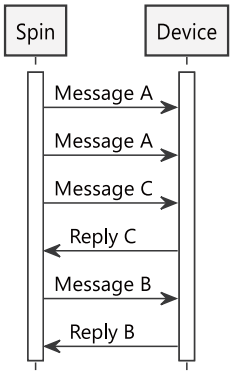


図 6: 表 2 の No.1 の実行結果の例 2

[Kuroiwa 16] Kuroiwa, T., Aoyama, Y., and Kushiro, N.: Testing Environment for CPS by Cooperating Model Checking with Execution Testing, *Procedia Computer Science*, Vol. 96, pp. 1341–1350 (2016)

[Larsen 05] Larsen, K. G., Mikucionis, M., Nielsen, B., and Skou, A.: Testing real-time embedded software using UPPAAL-TRON: an industrial case study, in *Proceedings of the 5th ACM international conference on Embedded software*, pp. 299–306ACM (2005)

[Tretmans 03] Tretmans, J. and Brinksma, E.: TorX: Automated Model-Based Testing, in Hartman, A. and Dussa-Ziegler, K. eds., *First European Conference on Model-Driven Software Engineering*, pp. 31–43 (2003)

[Vries 00] Vries, de R. G. and Tretmans, J.: On-the-fly conformance testing using SPIN, *International Journal on Software Tools for Technology Transfer (STTT)*, Vol. 2, No. 4, pp. 382–393 (2000)

[2H3-NFC-4a] [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)

座長:谷田 泰郎（シナジーマーケティング）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room H (10F Sky Hall)

[2H3-NFC-4a-01] Issues on Translating Information of Food Culture between Different Languages

○Masami Suzuki[1] （1. KDDI Research Inc.）

3:20 PM - 3:40 PM

[2H3-NFC-4a-02] Analysis of free description answer data on "super aged society"

○Yukiko Saito[1], Kotomi Takamuku[1], Yasuo Tanida[1] （1. Synergy Marketing, Inc.）

3:40 PM - 4:00 PM

[2H3-NFC-4a-03] About ‘Creative Genome Project’ and Relevant Intelligence Communication Consortiam

○Atsushi Sasaki[1,2] （1. AOI TYO Holdings Inc., 2. Relevant Intelligence Communication Consortiam）

4:00 PM - 4:20 PM

[2H3-NFC-4a-04] ‘Two-way communication’ and ‘Mind design’

○Yasuo Tanida[1], Kotomi takamuku[1] （1. Synergy Marketing Inc.）

4:20 PM - 4:40 PM

[2H3-NFC-4a-05] Text Analysis and Structuring of Persons in Kabuki

○Takashi Ogata[1] （1. Iwate prefectural university）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

食文化に関する異言語間情報共有の課題

Issues on Translating Information of Food Culture between Different Languages

鈴木 雅実[*1]
Masami SUZUKI

[*1] KDDI 総合研究所
KDDI Research, Inc.

In the background of increasing inbound visitors into Japan, there would be a problem of cross-cultural communications. ICT support to effective communication considering cultural distance between two languages is important. We will discuss this difficult issues based on a case study on food culture translation.

1. はじめに

日本語と諸外国語の間で，語彙の多様性に大きな相違が見られる分野の一つに，食物・料理に関係する表現がある．それらの語彙が指し示す対象物は，密接な繋がりを持つ他の事物との連想・共起関係を通じて理解されているため，翻訳において非常に難しい問題が存在している．すなわち語彙レベルでの対応関係が存在しないか，あっても辞書レベルで近似的な訳語のみが示される場合，元の語が示す事物を想像することは困難であり，理解には追加説明を要する．本稿では，そのような事例を挙げながら，最近の人工知能に基づく翻訳技術の現時点での限界と，今後の課題について考察する．

2. 研究の背景とねらい

近年急速に進展している人工知能研究の対象領域の一つに機械翻訳関連技術があり，翻訳精度は相対的に着実に向上しているとされる．ただし，曖昧性の少ない定形文や技術文書などに限定されており，文脈依存性の高い自由会話や文学などの解釈の自由度が高いテキストの翻訳は依然として困難である．特に言語の背景にある文化に依存した表現は，たとえ目的言語に置き換えられたとしても，そこからユーザが受け取る理解には限界が存在する．

[鈴木 2017]では，関連する先行研究である農研機構の早川文代による，日本語の食感を表す語彙の体系を引用しつつ，諸外国語と比較してオノマトペの使用率が高い日本語のオノマトペの学習支援の観点から，異文化コミュニケーションの課題について考察した．オノマトペを単独では理解が難しいことから，類似する語との近さ／遠さ等の直観的な距離感を示すこと，すなわち食感を表す日本語に顕著なオノマトペに対して，感覚的な評価軸を視覚的に提示するような支援手段も可能であり，学習・理解の参考となり得ることを示した．

オノマトペに限らずより一般的には，言語文化間の乖離を解消または軽減しようとすれば，付随する文化固有の事象を説明する必要性があろう．本稿では，食文化に関する翻訳の事例を紹介しつつ，内容理解の観点から異文化間の翻訳コミュニケーションの問題点について具体例を示しながら考察する．この問題は，[加納 2017]で取り上げられた世界価値観データベースにおける観光情報学の分野にも関係するものと認識している．

連絡先：鈴木雅実，KDDI 総合研究所 教育・医療 ICTG，
〒356-8502 埼玉県ふじみ野市大原 2-1-15
Email: msuzuki@kddi-research.jp

3. 関連動向

3.1 人工知能に基づく機械翻訳精度の向上

深層学習（ニューラルネット）に基づく機械翻訳＝ニューラル機械翻訳（Neural Machine Translation／NMT）では，多数の翻訳例を入力として学習することにより，未知の文であっても自動的に正しい翻訳結果を導く確率が高まる．ただし，何故そのような翻訳結果が出たかについての説明を求めることは不可能である[中澤 2017]．また，基本的には1文単位の翻訳が実行されており，文間に跨る文脈情報を参照するような処理は今後の課題である．

3.2 翻訳者・利用者の視点からの翻訳の限界

翻訳文を提供する上で，ユーザの立場から見れば，訳出内容の可読性だけでなく内容理解の程度で評価されることになる．その点では，翻訳された語句が差し示す対象についてユーザの知識が十分期待できる場合とそうでない場合では，理解度に大きな差が生じるであろう．そのようなギャップを軽減する手段として，内容理解を助けるような補足説明が望まれる．語のレベルでは，前章に記したオノマトペのように，知識のネットワーク化による俯瞰手段の提供は重要との考察を導いた[鈴木 2017]．

以下では，レストラン等での食事（注文と飲食）シーンを前提状況としている．料理名などから連想される様々な語や関連する別の料理や素材，調理法，由来その他のエピソードが存在することから，原文には無い情報を追加することにより，異文化間の翻訳コミュニケーションが成立するものと言える．この関係を概念的に示すと図 1 のようになる．

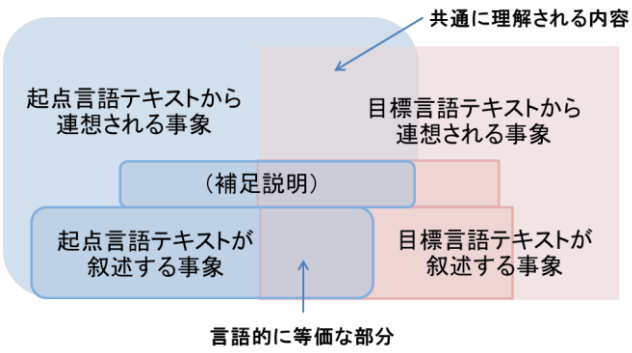


図 1. 異文化間の翻訳コミュニケーションの概念

4. 食事メニューの説明を対象とした翻訳例の分析

本研究では，来日外国人旅行者（インバウンド）に好まれる食事メニュー132 件を含む料理にまつわる様々な解説やエピソードなどを調査した．対象となる料理（名）について知識のない外国人に対して情報提供を行なうに当たって，メニューを提供する側との相互のコミュニケーションを促進するため，料理の説明だけでなく由来や関連するエピソードなど，ユーザの興味を喚起するような内容を，主として Wikipedia および個人のブログを参照の上，簡潔にまとめる工夫を行った．日本語の説明文は 160 文字程度を目安とし，その翻訳英文を人手により作成した．この結果は，記述対象の取捨選択から翻訳に至るまで，1 名の辞書編纂専門家による一貫した作業に基づくものである．

上記のような方針で臨んでも，異なる言語文化間における意志疎通は部分的な状態に留まる場合が多い．すなわち翻訳テキストとして対応が取れたとしても，参照している事物や歴史的・文化的背景に関する知識の共有には限界があるためである．そこで，上記の 132 件の料理メニューに関する日英対訳文（料理の説明／関連エピソード等を含む）のべ 276 対の翻訳としての質について内部検証を行い，主として理解容易性の観点から 3 段階評価を試みた．その内訳は表 1 の通りである．

表 1. 翻訳文の理解容易性の評価

ランク	件数	説明（日本語⇒英語翻訳文の自己評価）
評価A	137	ほぼ等価な翻訳であり，十分理解可能
評価B	125	言語的にはほぼ等価だが，説明内容に一部理解困難な箇所が存在
評価C	14	翻訳文の説明内容では，意図した伝達内容の理解が相当に困難
（合計）	276	注：評価分類は限定された対象範囲内の参考値

評価判断がBまたはCとなった主な要因とを挙げてみよう．

- 対象物（特に食材）が外国語話者の文化圏に存在せず，言葉の説明だけでは理解困難である．この点に関しては，言語文化間のギャップに関する比較文化論的な洞察が，バイリンガル環境で教育を受けた数学者により述べられている[望月 2017].
- 日本側では著名な人物や歴史的な出来事等に関係する用語が説明なく引用されている場合、すなわち該当文化圏では常識的で説明を要さないもの（年号なども含む）について補足的な説明が必要であるが，翻訳（意訳）を超えた問題とも言える．

表 2-1 および表 2-2 に記した実際の対訳例からも，上述した問題点の一部が示唆される．また，機械翻訳（Google）の結果を見ると，言語的なレベルでの訳出の精度は，従前よりも格段の進歩を遂げていることが解る．

5. 現状と今後の課題

最近の AI（ニューラル）機械翻訳技術の急速な進歩により，少なくとも1文単位で言語的に等価あるいは近似的な翻訳結果を導出することが可能となりつつある．一方で，異文化圏に属する人々の間の常識の隔たり，言い換えれば共有知識の多寡によって翻訳内容の理解容易性は大きく影響を受けることが明らかであり，その一端を食文化を題材として例示した．この問題を大規模な学習データの収集に基づく翻訳精度の向上というパラダイムの延長線上で克服することは困難であろう．言語による発想の違いに加えて文化固有の事物に関する総体的な知識が必要となり，それらをどのように翻訳コミュニケーションの場面で活用すべきかについての研究課題が存在する点を指摘することで，本稿はその契機と位置付けたい．

表 2-1. 料理名に関する日英翻訳の例示1

料理名	親子丼（料理の説明）
（日本語原文）親子丼は，文字通り"親と子の丼"の意味で，日本の丼料理であり，鶏肉，卵や他の食材を割り下で煮て，丼のご飯に載せたもの。この料理名は、鶏肉と卵が一緒に使われることから名付けられた．	
（機械翻訳）Oyako don literally means "a bowl of parents and children", which is a Japanese bowl dish, boiled chicken, eggs and other ingredients in a dish and placed on a rice bowl. This dish name was named because chicken and eggs are used together.	
（人手翻訳）Oyakodon, literally "parent-and-child donburi", is Japanese rice bowl dish, in which chicken, egg, and other ingredients are all simmered together in a sauce and then served on top of a bowl of rice. The name of the dish is a poetic reflection of the fact that both chicken and egg are used in the dish.	

（注）丼料理のイメージが湧くか？／割り下＝sauce で良いか？

表 2-2. 料理名に関する日英翻訳の例示2

料理名	親子丼（エピソード）
（日本語原文）ある日，ポール・サイモンが中華料理店に入ると，メニューに「母と子の絆」という料理がありました．出された料理を見ると，日本でいうところの「親子丼」でした．英訳≪Mother And Child Reunion≫ は，チキンと卵とが同じ丼で再会するという 家族の愛の物語なのです．	
（機械翻訳）One day, when Paul Simon entered a Chinese restaurant, there was a menu called "Mother and child's bond" in the menu. Looking at the dishes that were served, it was "Oyako-don" in Japan. English translation «Mother And Child Reunion» is a story of family love that chicken and eggs meet again at the same bowl.	
（人手翻訳）One day, Paul Simon, a singer, dropped by one Chinese restaurant, he found a dish called "mother and child reunion" in menu. The dish served was "Oyakodon" in Japan. "The Mother and Child Reunion" is the story of the love of a family. And a chicken and an egg are to meet in the same bowl.	

（注）共感できる度合いには個人差がありそう

なお，本稿に関連する調査は，東京工業大学の COI（JST の成果展開事業）における「空気・行間を読む、意訳する技術」（『以心電心』AI テクノロジー）プロジェクトの支援によるもので．協力頂いたプロログ/関倫彦氏に厚くお礼申し上げる．

参考文献

[鈴木 2017] 鈴木雅実：日本語オノマトペ学習支援に向けて―食感の理解に関する異文化コミュニケーション―，人工知能学会第 31 回全国大会， 2E1-NFC-04a-2，2017.

[Kano 2017] Fumiko Kano Gluckstad：UMAMI: Understanding Mindsets Across Markets, Internationally，人工知能学会第 31 回全国大会，2E1-NFC-04a-4，2017.

[中澤 2017] 中澤敏明：機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理，情報管理，Vol.60 No.5，pp.299-306，2017.

[望月 2017] 望月新一：「心壁論」と、論理構造の解明・組合せ論的整理術を「心の基軸」とすることの本質的重要性 (10)，https://plaza.rakuten.co.jp/shinichi0329/diary/201711210000/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

「超高齢社会」についての自由記述回答データの分析
―価値観から見えるヒトの意識構造の考察―

Analysis of free description answer data on "super aged society"
―Consideration of human consciousness structure seen from values ―

齋藤 有紀子[*1] 高椋 琴美[*1] 谷田 泰郎[*1]
Saito Yukiko Takamuku Kotomi Yasuo Tanida

[*1] シナジーマーケティング株式会社
Synergy Marketing, Inc.

The authors study the social knowledge centered on values. "Super aged society" indicates that social structure will change. We want to obtain the problem consciousness of the people from the free description reply, consider the connected values and explore the countermeasure.

1. はじめに

日本は，2007 年に超高齢社会に到達した．超高齢社会とは 65 歳以上の人口が総人口に占める割合（高齢化率）が 21%以上の社会である．最新の平成 29 年版高齢社会白書によると 2016 年 10 月時点での高齢化率は 27.3%にまで到達している[内閣府 2017]．少子化もあいまって，日本は人口遷移が最も進んでいる国でもある．言い換えれば，社会構造の変化の真っ最中であり，今までの社会常識や概念が通じにくい社会，常識が変わっていく社会になっていくわけである．このような社会の中で我々は何らかの対応をしていかなければならない．

弊社で過去に行った自主調査内に，「超高齢社会に対するあなた自身の考えをできるだけ詳しく教えてください」という内容のアンケートを自由記述形式（以下 FA と省略する）で行い，合計 6141 件のデータを集めることができた．FA 形式にした理由としては，調査側が意図しないような回答が得られる可能性と，自由に記述することで回答者の個性が表れるのではないかという 2 点があげられる．このアンケート時には，弊社独自の価値観モデルである Societas[谷田 2014]の 60 選択肢の質問項目から構成されるマルチプルアンサー（以下 MA と省略する）に対して同時に回答がなされている．

FA から得られた回答内容について，Societas の価値観との関係性が見出せたならば，我々が超高齢社会を生き抜いていく為の対応策や解決の糸口を見出すことができるかもしれない．

本稿では，得られた FA のデータについて，どのような人がどういった意見を持っているのかを把握した上で，まとめた意見群について価値観の差異から読み解ける知見を紹介する．

2. 分析に利用したデータ

本稿の分析には以下に示す調査を使用した．

- 2013 年 9 月実施
  「あなたご自身についてのアンケート価値観把握調査（ミニ調査）」A 社モニタ会員
  年齢 15～69 歳（自然発生），男女比 1:1 にて 5141 件
- 2015 年 12 月実施
  「高齢者の生活実態調査」B 社モニタ会員
  年齢 55 歳以上，男女比 1:1 にて 1000 件

2.1 Societas の価値観

Societas とは，開発当初 8 設問 303 選択肢の質問から主成分分析により抽出された成分である 61 個の価値観[馬場ら 2013]を推定するモデルであった．しかし，実際のマーケティングなどの現場にてこのモデルを適用するのは聴取の際の質問量の多さによる回答者の負担や以降のコンテンツの作成などの運用を考慮すると極めて困難である．そこで，質問項目の 反応率・情報量基準（c4.5）によって選択された主要な成分変数との相関係数（Pearson）・聴取のしやすさ・マーケティングにおける重要度などの定性的な観点から選択された 6 設問あるいは 8 設問の合計 60 選択肢の質問項目が利用されるようになった[谷田ら 2013]．本稿では，それらの経緯や詳細については述べない．

上述の 6 設問あるいは 8 設問 60 選択肢の質問は，表 1 の例に示すような文章を直感的に選択する MA で構成されている．

表 1：Societas の設問と質問項目（一部抜粋）

Q1	あなたについて伺います。あてはまるものをすべてお答えください。（ＭＡ）
□	実績がないことをやってみるのはわくわくする
□	答えのないものを探索するのが好き
□	好奇心が強い
□	人の目を気にする
□	しばしば悩むことがある
□	小さなことが気になって眠れなくなることがある
□	めんどうなことはしたくない
□	どちらかというとインドア派だ
□	のんびりがいい
□	記念日は覚えている方だと思う
□	協調性がある
□	家族や友人を喜ばせたい
□	あてはまるものはない

本稿で利用する Societas の質問項目は，上述の 6 設問 60 選択肢のものを採用し，また，価値観変数はこの質問項目の選択肢へのチェックの有無からから計算される表 2 に示す gene カテゴリの 8 変数，character カテゴリの 7 変数，meme カテゴリの 8 変数とする．

価値観変数は図 1 に示すように階層化しており，理由として人に紐付く価値には変化しにくいものと変化しやすいものがあるだろうという経験的でかつ定性的な想定からである．生まれつき持っている遺伝子に影響を受けるか，後天的に社会や環境からの影響を受けるかということでいうと，gene > character > meme の順に先天性が強いという想定をしている[synergy-marketing]．

gene カテゴリの変数は 8 つの人の性質を定性的に定義したものである（2 つずつが 4 つの対になっている，図 1 に示す gene 部分の破線矢印同士は対軸関係になっている）．それらを

連絡先：齋藤有紀子，シナジーマーケティング株式会社，研究企画チーム，電話番号：0774-95-2305

定量的に扱うための予備調査は，8 つの性質について意味の違った文章や単語を提示して聴取する形式にて 1,000 人単位で合計 7 回，7,000 人に対して行い，「保守的である」「冒険心が強い」「個人主義的である」「協調性がある」「我慢強い」「誘惑に弱い」「我慢強い」「こだわらない」「固執する」という表現に自分の性質があてはまるかどうかの MA 回答の形式が固定された．この MA 形式の回答と Societas の質問項目の MA 回答を同時に聴取した 8.5 万人強の調査データをもとに定量的に計算する仕組みを作成している．現在は，直接 gene の設問を聴取する形ではなく，60 選択肢の MA の質問項目に答えることで，それぞれの 8 つの変数に対して 2 水準(0,1)あるいは 4 つの対軸の変数に対して 3 水準(-1,0,1)の確率値を付与している．

本稿の論文の表示では，「保守的である」を「損害回避」，「冒険心が強い」を「新規性追求」，「協調性がある」を「協調的」，「個人主義的である」を「個人主義」，「こだわらない」を「柔軟」，「固執する」を「頑固」と置き換えて表示しているが両者は同じ変数である．

また，character カテゴリ，meme カテゴリの変数についても，gene を計算する時と同じ 60 選択肢の MA の質問項目に答えることで，3 水準(-1,0,1)の確率値を付与している．

表 2:本稿の分析に利用した Societas の価値観変数

カテゴリ	価値観変数名			
gene	新規性追求	損害回避	個人主義	協調性
(普遍的)	頑固	柔軟	誘惑に弱い	我慢強い
character	好奇心	自己成長	協調的	批判的
(性格)	自己愛	繊細	のんびり	
meme	ストレス	仕事充実	家族不仲	結婚願望
(後天的)	お金ゆとり	家族中心	親友あり	多忙

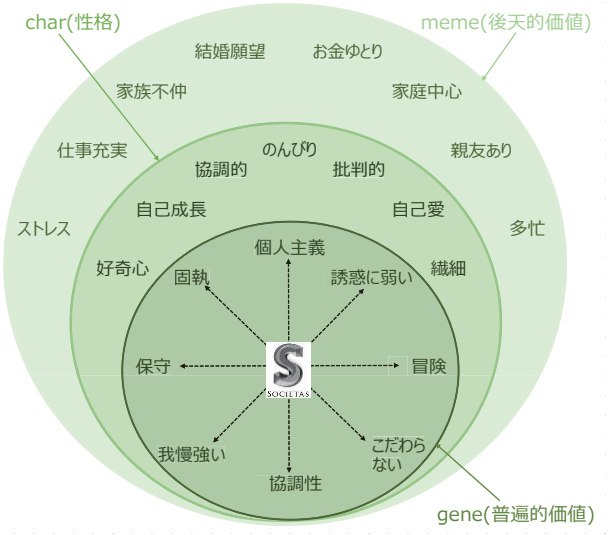


図 1:Societas 価値観変数の概念図

3. 自由記述回答の分析

「超高齢社会に対するあなた自身の考えをできるだけ詳しく教えてください」という設問に対して，FA の文脈の印象判定と特徴的と思われる単語を目視にて抽出し分類を行った．表 3 に分類した意見群(以下タグと省略する)を示す．大分類は文脈の印象判定にて分類された 5 タグ，小分類は特徴的な単語を含む意見群として 20 タグを設定した．大分類中にある『無関心』とは“分からない”や“特になし”等の回答をしているタグである．従来，分析の除外対象とすることが多いが，本稿ではこのような回答をするタグと価値観の関係性も測ることとした．

表 3:分類した意見群(タグ)

大分類	小分類			
・ネガティブ	・仕方ない	・不安	・大変	・困る/負担
・ポジティブ	・みんな/協調	・制度/福祉	・政策/国	・子供/将来
・自分事	・お金/年金	・介護	・自立する	・生涯現役
・他人事	・長生きしたくない	・延命拒否	・健康で長生き	・尊厳死
・無関心	・高齢者ずるい	・のんびり	・達観	・前向き/楽観

3.1 デモグラフィック属性からの視点

分析に使用したデータのデモグラフィック属性を表 4 と表 5 に示す．二つのアンケートを合わせて使用しており，且つ一方はアンケート対象年齢を 55 歳以上に設定したため，データ取得時の日本人の人口分布[総務省統計局]と比較すると 50 歳代 60 歳代のボリュームが膨らんでいる．本稿ではデータ数に限りがあるため補正は行なっていない．

表 4:分析データのデモグラフィック属性(男性)

男性					
年代	構成比	未婚者	既婚者	子供なし	子供あり
10代	0.7%	2.8%	0.0%	1.4%	0.1%
20代	4.1%	14.4%	0.6%	7.8%	0.5%
30代	13.9%	25.6%	9.8%	16.9%	10.8%
40代	25.9%	32.9%	23.5%	26.8%	25.0%
50代	26.7%	18.3%	29.7%	24.0%	29.6%
60代	22.1%	5.8%	27.7%	21.7%	22.4%
70代	4.9%	0.1%	6.6%	1.2%	8.7%
80代	1.6%	0.0%	2.2%	0.3%	3.0%
total	49.6%	25.6%	74.4%	50.1%	49.9%

表 5:分析データのデモグラフィック属性(女性)

女性					
年代	構成比	未婚者	既婚者	子供なし	子供あり
10代	1.7%	7.3%	0.0%	3.5%	0.1%
20代	11.0%	29.6%	5.2%	17.9%	4.7%
30代	23.1%	28.7%	21.3%	23.4%	22.8%
40代	25.6%	20.2%	27.3%	22.4%	28.6%
50代	19.0%	7.8%	22.5%	17.8%	20.1%
60代	13.0%	4.7%	15.7%	12.4%	13.6%
70代	4.8%	1.3%	5.9%	2.0%	7.5%
80代	1.6%	0.4%	2.0%	0.6%	2.5%
total	50.4%	24.0%	76.0%	47.9%	52.1%

3.2 Societas 価値観変数での特徴①

表 6 は，分類した意見群(タグ)ごとに，男女別に Societas の価値観変数の出現率の平均値と比較しての差を表したものである．(表中の上下矢印はプラス方向またはマイナス方向に χ^2 検定にて有意差を認め，“↑↑”印と“↓↓”印は($p<0.01$)，“↑”印と“↓”印は($p<0.05$)にてそれぞれ男女の平均値よりも強いことを表し，“≒”印は有意差が認められなかったことを表す)．3.3 の表 7，表 8 に関しても同様の表示方法である．表 6 が大分類のタグ，表 7，表 8 が小分類のタグである．

表 6 は，大分類のタグのうち，『ポジティブ』，『ネガティブ』，『無関心』に関して，どのような価値観との関係性が強いのかを見ている．

gene や character など変化しにくい価値観を見ると，「超高齢社会」という社会課題に対してポジティブな考えを持つ人は，性別を問わず「新規性追求(冒険心)」「好奇心」のようなポジティ

ブな価値観を持つ人が多い．その他，性別を問わずポジティブな考えに寄与する価値観には，「協調性」「我慢強い」「柔軟」などがある．また，社会課題に対してネガティブな考えを持つ人には，その逆の価値観である「損害回避（保守）」，「個人主義」，「誘惑に弱い」，「頑固」などの価値観を持つ人が多かったほか，「繊細」「のんびり」などの価値観を持つ傾向があった．性差があったものでは，ポジティブな意見を言う人に「自己愛」と「自己成長」の価値観を持つ傾向が男性のみに見られた（女性はそもそも自己愛が強い）．

meme のようにライフステージなどによって変化しやすい価値観を見ると，社会課題に対してポジティブな考えを持つ人は，性別にかかわりなく，お金にゆとりがある，友人がいる，仕事が充実しているなどの余裕がある状況の人が多いことが分かる．また，社会課題に対してネガティブな考えを持つ人は，ストレスがあったり，家族と不仲であったり，忙しかったりという状況の人が多いことがわかる．性差があったものでは，社会課題に対してネガティブな考えを持つ人に「結婚願望」「家庭中心」という価値観を持つ傾向が女性のみに見られた．

『無関心』タグの振られた人は，価値観のアンケートに関してもマーク率が低い無関心層である．無関心タグにあるような回答をする人は，意見を求められても意見を表出しない（自分を見せない）人であろう．

社会課題に対してポジティブな考えを持つ人，ネガティブな考えを持つ人，無関心な人の価値観を見てきたが，ネガティブな考えを減らし，ポジティブな考えを増やさなければ，多くの施策は生きない．そういう意味でも，ポジティブな考えを増やす，ネガティブな考えを減らすための価値観を刺激するような社会努力が必要である．

表 6：価値観変数との関係性①

価値観変数		タグ					
		ネガティブ		ポジティブ		無関心	
		男性	女性	男性	女性	男性	女性
gene	新規性追求	≒	≒	↑↑	↑	↓↓	↓↓
	協調性	≒	≒	↑↑	↑↑	↓↓	↓↓
	我慢強い	≒	≒	↑↑	↑↑	↓↓	↓↓
	柔軟	≒	↓↓	↑↑	↑↑	↓↓	≒
	損害回避	↑↑	↑↑	≒	≒	↓↓	↓↓
	個人主義	↑	↑↑	↓↓	↓↓	≒	≒
	誘惑に弱い	↑↑	↑↑	≒	↓↓	↓	≒
	頑固	↑↑	↑↑	≒	↓↓	↓↓	≒
character	自己愛	≒	↑↑	↑	≒	↓↓	↓↓
	自己成長	≒	≒	↑↑	≒	↓↓	↓↓
	協調的	≒	≒	↑↑	↑↑	↓↓	↓↓
	好奇心	≒	≒	↑↑	↑↑	↓↓	↓↓
	繊細	↑↑	↑↑	≒	↓↓	↓↓	↓
	批判的	≒	↑↑	≒	≒	↓↓	↓↓
	のんびり	↑	↑	≒	≒	≒	↓
meme	お金ゆとり	≒	≒	↑↑	↑	↓↓	↓↓
	仕事充実	≒	≒	↑↑	↑↑	↓↓	↓↓
	親友あり	≒	≒	↑↑	↑↑	↓↓	↓↓
	家庭中心	≒	↑	≒	≒	↓↓	↓↓
	結婚願望	≒	↑	≒	↓↓	≒	≒
	ストレス	↑↑	↑↑	↓	↓↓	↓	≒
	家族不仲	↑↑	↑	↓	↓↓	↓↓	≒
	多忙	↑	↑	≒	↓	≒	≒

3.3 Societas 価値観変数での特徴②

次に小分類タグの中から特徴的と思われるトピックを表 7 と表 8 に示す．

『健康で長生き』や『自立』，『生涯現役』といったタグに分類される回答者は，gene の「協調性」，「我慢強い」，「柔軟」や character の「協調的」や「好奇心」，meme の「お金ゆとり」，「仕事充実」，「親友あり」などの価値観を持っており，このような人たちは，「超高齢社会」という社会課題に対して前向きな捉え方と発想ができる人であると思われる．『長生きしたくない』や『介護』のタグでは，普遍的な価値観の gene には有意差が見られず後天的に社会や環境からの影響を受けやすい価値観である character と meme で有意差が認められた．人は元来，長生きしたくないわけではなくむしろその逆であろうし，また，介護についてもある一定の年齢になってから経験することであり，これらの結果は「超高齢社会」に対しての問題点がよく反映されていると思われる．

価値観との関係性を見ていく中で，男女で差を認めたタグもあった．男性のみで有意差が見られたのは，『みんな/協調』といったタグで，女性のみで有意差が見られたのは，『困る/負担』や『介護』といったタグであった．「超高齢社会」という社会課題に対して男性は俯瞰的に捉えているが，女性はより身近な事柄と感じているのではないかと思われる．

表 7：価値観変数との関係性②

価値観変数		タグ							
		健康で長生き		長生きしたくない		自立		生涯現役	
		男性	女性	男性	女性	男性	女性	男性	女性
gene	新規性追求	≒	≒	≒	≒	↑	≒	↑	↑
	協調性	↑↑	↑↑	≒	≒	↑↑	↑	↑↑	↑↑
	我慢強い	↑↑	≒	≒	≒	↑↑	≒	↑↑	↑↑
	柔軟	↑↑	↑↑	≒	≒	≒	↑↑	↑	≒
	損害回避	≒	↓↓	≒	≒	↓↓	↓↓	≒	≒
	個人主義	↓	≒	≒	≒	≒	↓↓	↓	↓
	誘惑に弱い	≒	↓↓	≒	≒	↓	↓↓	≒	≒
	頑固	≒	↓↓	≒	≒	≒	↓↓	≒	≒
character	自己愛	≒	≒	≒	≒	≒	≒	↑↑	≒
	自己成長	≒	≒	≒	≒	≒	≒	↑↑	↑
	協調的	↑↑	≒	≒	≒	↑↑	≒	↑↑	↑↑
	好奇心	≒	≒	≒	≒	↑	↑↑	↑	↑
	繊細	↓	↓↓	≒	≒	≒	↓↓	≒	≒
	批判的	≒	↓	≒	↑	↑	↓	↑	≒
	のんびり	≒	≒	↑↑	↑	↓↓	≒	≒	≒
meme	お金ゆとり	↑	↑	≒	≒	↑	↑↑	↑	≒
	仕事充実	≒	≒	≒	≒	↑↑	≒	↑↑	↑↑
	親友あり	≒	≒	≒	≒	↑↑	↑	↑↑	↑↑
	家庭中心	≒	↓↓	≒	≒	≒	≒	≒	≒
	結婚願望	≒	↓↓	≒	≒	↓	≒	≒	≒
	ストレス	≒	↓↓	↑	↑	↓	↓↓	≒	≒
	家族不仲	↓	≒	↑	↑	≒	↓↓	≒	≒
	多忙	≒	↓↓	≒	≒	≒	↓↓	↓	≒

表 8：価値観変数との関係性③

価値観変数		タグ							
		子供/将来		介護		みんな/協調		困る/負担	
		男性	女性	男性	女性	男性	女性	男性	女性
gene	新規性追求	≒	≒	≒	≒	≒	≒	≒	≒
	協調性	≒	≒	≒	≒	↑	≒	≒	≒
	我慢強い	≒	≒	≒	≒	↑↑	≒	≒	≒
	柔軟	≒	↓↓	≒	≒	≒	≒	≒	↓
	損害回避	≒	↑↑	≒	≒	≒	≒	≒	↑↑
	個人主義	≒	↑↑	≒	≒	↓↓	≒	≒	↑↑
	誘惑に弱い	≒	≒	≒	≒	≒	≒	≒	≒
	頑固	≒	↑↑	≒	≒	≒	≒	≒	↑↑
character	自己愛	≒	↑↑	≒	↑	≒	≒	≒	≒
	自己成長	↑↑	↑↑	≒	≒	≒	≒	≒	↑
	協調的	≒	≒	≒	≒	↑↑	≒	≒	≒
	好奇心	≒	≒	≒	↓	≒	≒	≒	≒
	繊細	≒	↑↑	≒	≒	≒	≒	≒	↑
	批判的	≒	↑↑	≒	≒	≒	≒	≒	≒
	のんびり	≒	≒	≒	≒	≒	≒	≒	≒
meme	お金ゆとり	≒	≒	≒	≒	≒	≒	≒	≒
	仕事充実	↑↑	≒	≒	≒	≒	≒	≒	≒
	親友あり	≒	≒	≒	≒	≒	≒	≒	≒
	家庭中心	≒	↑↑	≒	↑↑	≒	≒	≒	≒
	結婚願望	≒	↑↑	≒	≒	≒	≒	≒	≒
	ストレス	≒	≒	≒	≒	≒	≒	≒	↑
	家族不仲	≒	≒	≒	≒	≒	≒	≒	≒
	多忙	↑↑	≒	≒	≒	≒	≒	≒	≒

4. おわりに

FA を文脈の印象で判定し，特徴的と思われる単語を目視にて抽出することで分類を行ったタグと，Societas の価値観変数との関係性を見ることにより，デモグラフィック属性だけでは得がたい異なる視点からの知見を得ることができた．このように価値観変数を用いて人を読み解いていけば，人を深く知ることができ，人や社会を動かす際の施策などに役立つと考えている．データの分析方法や補正の方法など課題は山積しているが，まず人を知ること，人の意識構造を知ることの一歩は踏み出すことができた．これからも問題意識を持って取り組んでいきたい．

参考文献

[内閣府 2017] 内閣府：平成 29 年版高齢社会白書 http://www8.cao.go.jp/kourei/whitepaper/w-2017/html/zenbun/index.html

[谷田 2014] 谷田泰郎: 価値観マーケティングと社会知ネットワーク，人工知能 9 月号 Vol.29 No.5 P456-P463，2014．

[馬場ら 2013] 馬場綾子，谷田泰郎，Mathieu Bertin: 社会値としての消費者価値観構造モデルと類型「Societas」の構築，人工知能学会全国大会（第 27 回）JSAI2013，（2013）

[谷田ら 2013] 谷田泰郎，馬場綾子，河本裕輔，藤井絵美子: 価値観モデルを利用したマイクロブログ発言者の社会知類型の推定，言語処理学会第 19 回年次大会，(NLP2013).

[synergy-marketing] シナジーマーケティング社 研究概要 http://lab.synergy-marketing.co.jp/report/

[総務省統計局] 総務省統計局：年齢5歳階級別人口 http://www.stat.go.jp/data/nihon/02.htm

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

-新たな価値観の発見に向けて-

Creative Genome Project と「価値観・HI コンソーシアム」の取組み

About 'Creative Genome Project' and Relevant Intelligence Communication Consortiam

佐々木　淳[*1] [*2]

[*1] AOI TYO HOLDINGS 株式会社　Pathfinder 室　HI サイエンティスト/[*2] 価値観・HI コンソーシアム
AOI TYO Holdings, Inc.　/　Relevant Intelligence Communication Consortiam

sasaki.atsushi@aoityo.com

In this "Challenge Session", we aim to explain at first the outline of Creative Genome Project , (formerly introduced in detail at subcommittee meeting on last December) especially about its thought and updated points. Secondary, we will introduce Human Intelligence (HI), on that authentic way of philosophical, cultural point of view Creative Genome Ptroject is highly depends .At the end of our session we refer to 'Relevant Intelligence Communication Consortiam' . Through this Cosortiam's activity, not only AOI TYO Holdings Inc.,, but also Synergy Marketing Inc., Amana Images Inc., and other more cooperate together to make studies about Human Values and Human Intelligence , to develop or reveal new social concepts for our next AI age.
Keywords — HI, Human Intelligence, Creative, Human Values, Abduction, Algorithm

1. はじめに

当チャレンジセッションにおいては、2017 年 12 月の人工知能学会「ことば工学」分科会での説明[1]に準じ、＜Creative Genome＞の概略を示しつつ、その思想、および現在進行中のカスタマイズ点に触れる。その上で、世界価値観との関連において弊社 AOI TYO Holdings、シナジーマーケティング㈱、㈱アマナイメージズ等の各社共働による「価値観・ＨＩコンソーシアム」の活動および、コンソーシアムにおける＜Creative Genome＞の役割、最後に各社価値の総合による、ＨＩ（ヒューマンインテリジェンス）に基づいた「新たな社会的価値やクリエイティブの発見」への見通しを示すこととする。

2. ＜Creative Genome＞について

2.1 経緯

2012年に筆者が開発した「映像への心象に基づくタギング」を実装したサービスプロトタイプ "nagarami" [2]を継承発展し、主に TVCM の読後感に注目し、その読後感から逆算して各種制作手法やブランド概念・テーマなどを詳細定義・整理を行っているデータベースである。2016 年には当データベースを援用したブランドキャンペーンも行われ、2017 年から現在に渡ってはインターフェースの再検討、大学との共同研究による拡張データベースへの対応、定義の拡充やデータ量の増量などを行っている。[3]



（図１・映像タギングプロトタイプ'nagarami'<2012>）

2.2. ＜Creative Genome＞の価値

＜Creative Genome＞の目指す価値とは大きく以下の3要素である。

(1)既存形式知の発見

既存のＣＭクリエイティブを独自の観点（コンセプト・物語手法・映像手法・トーン・文化）等でタグ抽出し、それらをアナロジカルにグルーピングし、可視化している。

その際に、ＣＭの＜読後感＞を大きく２０前後のパターンに分け、これらを最重要のタグとして措いていることが大きな特徴で、特定の読後感に対し＜共起的に発生する物語コンセプト、映像手法、構成手法＞などが多数紐づくため、読後感を手掛かりに、既存の形式知、その組み合わせが視覚的に発見容易なデータベースとなっている。

(2)新たなストーリーテリングの仮説・発見

特定の読後感を惹起する新しい方法知（ストーリーテリングのための手法組合せ）を仮説することができる。

上述の通り、特定の読後感に共起的に発生するコンセプト、映像手法、等が多数紐づいていることから、これらを自由に組み替えることにより、新たな制作方法論を呈示することが可能となっている。

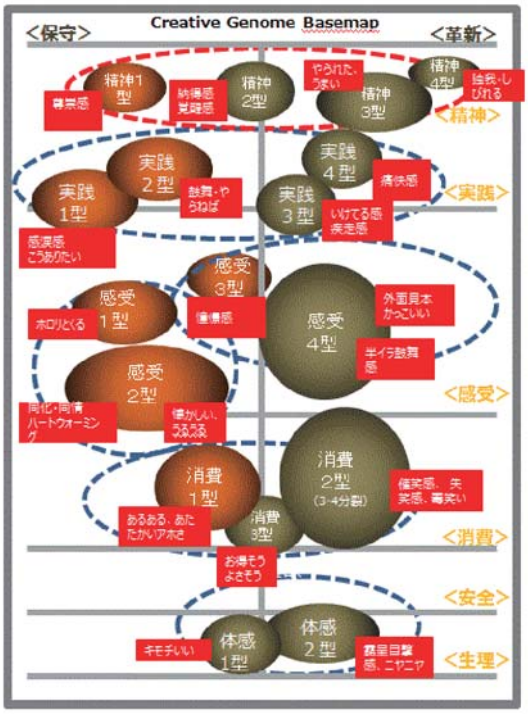
また、コンセプトなどにおいては一旦抽象化・言語化しているため、この言語上の類義性を活かし、既存作品にないコンセプトワードを新たに追加することが可能である。こうしたことによっても、企画仮説の幅が広がることになる。

(3)新たな読後感の仮定・発見

２０前後の既存読後感パターンにはない、新たな読後感の仮説・発見を促す。

２０前後の読後感パターンについては既に独自のマッピングを行っており、その位置関係が読後感の質的距離を示す。このマップを利用し、比較的異なる読後感同士から、それぞれに紐づく手法を抽出してそれらを混ぜ合わせることによって「新しい読後感」を仮定することが可能である。こうした仮定作業には当然ながら検証が必要であるため、現在そうし

た仮説スロット＝フィードバック検証、の仕組み実装へ向かっている。



（図 2・Creative Genome ベースマップ）

2.3　タギングについて

過去の国内名作 CM についての解釈データセット制作、約 1000 本分を対象に解析を行っている。解析タグについては以下の通りである。

(a)読後感のタグ（コミュニケーションコンセプトタグ＝ＣＣＴ）

前述の通り、独自の定義によってＣＭの読後感を２０個程度に分別している。この読後感タグを「コミュニケーションコンセプトタグ（CCT）」と呼ぶ。これは、ソシュールやバルトの議論[4]に沿えば「シニフィエ」に近似しており、制作者の込めた制作意図（シニフィアン）に拘わらず、「その表象が社会に受容　される、その受容概念」を表している。これを念頭に、定義者の間では簡易的に「CCT」ないし「読後感タグ」と呼称している。

(b)物語主題のタグ（物語コンセプトタグ）

上記(a)に対する、制作意図（シニフィアン）側のタグのカテゴリーである。すなわち作り手側がストーリーに託した大きな＜コンセプト＞、＜表現世界＞、＜ワールドモデル＞、＜具体モチーフ＞、＜物語手法＞、＜トーン、テンション＞、＜物語フレーム＞、＜作劇モデル＞、＜物語技法（メタファー・方法）＞をさす。これらの各タグの下に、数十のパラメーターが設定されている。

(c)映像・音声手法のタグ

これもシニフィアンのタグであるが、上記(b)とは異なり実践的な制作方法にまつわるものである。撮影技法、照明技法、編集技法、音声技法、画像加工、デザイン様式などである。

(d)UX のタグ

ユーザーの受容感情。(a)と呼応・共起の関係となる。

(e)諸事実データ・デモグラフィックタグ

クライアント・商品名・スタッフ・出演などの情報、想定ターゲット、商品・サービスジャンル等の情報群。上記をさらに細分化し、現在の 1CM あたりのタグ（カラム）は 20 数種類程度となっている。(図 3・Creative Genome データセット)

なお、各タグ、タグ内のパラメータとなるワードの選びこみや概念階層について、チャールズ・パース (Charles Sanders Peirce) による “icon” “index” “symbol” の区別議論[5]のほかロラン・バルト (Roland Barthes) [6]、ソシュール (Ferdinand de Saussure) [7]、松岡正剛[8]らの区分を参考にしている。

また、小方・秋元による物語言説論[9]や小方・金井の物語論を巡る著作[10]からも多くの示唆を受けた。物語手法についてはプロップ[11]、グレマス[12]、松岡正剛[13]、ジョゼフ・キャンベル[14]、ジェラルド・プリンス[15]などを参考としている。またメタファーについての仕分けについては、アリストテレス[16]やキケロの議論[17]も参照した。



（図 3・Creative Genome データセット）

2.4　新たな検討タグ

今回新たに定義を開始しているタグとして以下がある。

(f)文化ミームのタグ

まずは主に日本文化による表現概念の仕分けを行う。文化ミームは各文化・民族に無意識に流れ知らぬ間に形態を変えながら継承しているものでもあり、こうした要素を明示知とすることで、文化感覚の普遍性や、或いは時系列におけるその顕密の変動も明らかにされると考えている。

こうした各タグの定義については数々の一対概念構図を参照することになった。松岡[18]による室町の＜遊び(スサビ)と荒び（スサビ）＞、＜意気と粋＞、＜真名と仮名＞＜数寄と作分＞など、現在まで形を変容して継続されている一見不可視なコンセプトをＴＶＣM内に読解し、独自のタギングでデータ化を継続中である。こうしたタグ要素はまさに後述するＨＩ（ヒューマンインテリジェンス）であり、こうしたタグ同士の化学反応による新コンセプトの提案は、後述する新たな価値観の提示にも良き化学反応を示すと思われる。

ゆくゆくは諸外国文化のミームをタギングし、連結することで、世界における文化の間テクスト性ともいえるデータベースとなりうる点、それがひいては膨大なマーケットを率いる知的震源となりうる点を指摘しておく。

(g)共時的感覚のタグ

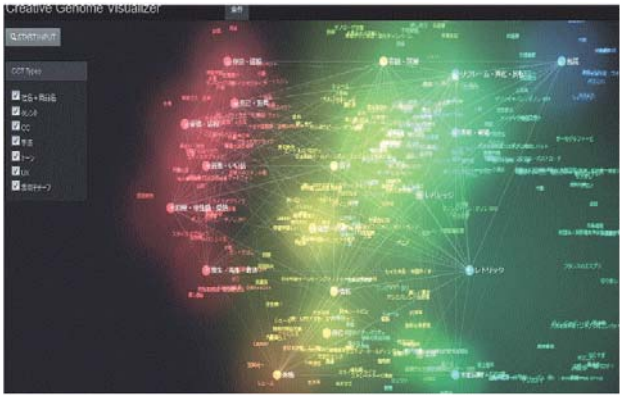
主に生活内における感情の因果律（義務感-解放感、違和感-馴化感）など、生活内に普遍的に生起している感情感覚を言語定義していく。生活シーンでのワールドモデルの細分規定にあたると思われ、大きくは＜各個人に観念として留まったまま概念化されていない生活内諸次元・感情の抽出＞に該当する。

(h)通時的感覚（人生哲学・価値観）のタグ
主に生活や経験の堆積からくる、人生において尊ぶ概念（共同体主義-個人主義、自然嗜好-都会嗜好、宗教性重視-利便効率重視、理知重視-感性重視、など）についてのタグ。最も価値観に近いタグであり、受容者側の価値観と共振・反発を起こし、同一CMに対する異なる評価を産む大きな潜在要素ともいえる。従って、後に述べる価値観・ＨＩコンソーシアムでのデータ連動に大きな役割を負うと考えられる。

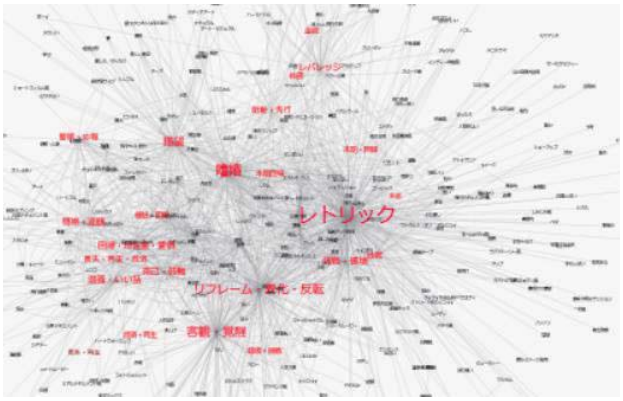
2.5　ビジュアル・インターフェース

基本的には(a)のCCTを基準とし、そこに紐づく(b)(c)(d)(e)(f)(g)(h)が連関的に表示される。
同時に、任意のCCTを選択することで、可能性のあるクリエイティブ仮説として(b)(c)(d)(e)(f)(g)(h)の組み合わせを提示することができる。ここで提示されたクリエイティブ仮説については前述の通り、評価のフィードバックを返していけるシステムについて実装可能性を検討中である。

従来のバージョンに加えて、cytoscapeバージョンを稼働させた。これにより、任意のデータセットについての可視化、及び可視インターフェースのスタイル変更が容易となった。



（図4・Creative Genomeインターフェース）



（図5・同・Cytoscape版インターフェース）

3. HI（ヒューマンインテリジェンス）について

3.1　語義について

そもそも人智を意味する’Human Intelligence’は人間による諜報活動、ないしリクルーティングにおける人材情報というビジネスワード(Humint)として近年流通していた。一方最近になって米国でのAI議論における、AIの計算能力では難しい人間知領域（文脈知や経験的暗黙知、文化的習慣知、クオリア等を含む認識や知覚をさまざまに指すケースが多い）として’Human Intelligence’という語が使われ始めている[19]。主にAIとHIが共働し、人間の認識、感情を判断しながら、いかに「人のキモチ」に響く社会を形成できるかという論点での議論が始まっているように見受けられる。

3.2　HIとしての＜Creative Genome＞

＜Creative Genome＞は、ＣＭという15秒・30秒短尺型の映像でいかに視聴者の気分を変容させるか、という読後感の領域にまずフォーカスしている。すなわちコンテンツがどのような主題や表現（シニフィアン）によって、どのような気分を惹起させたかということにおいて、そのコンテンツの作用メタファ（シニフィエ）を炙り出すことに価値をもっている。この関連性・補助線は解釈データを必然的に伴うため、現行ＡＩでは容易に描けず、非常にHI的な性向をもっている。

この＜シニフィアン-シニフィエ＞関係の有機的連関を明示知とすることが、感情の醸成に関わるＨＩ（ヒューマンインテリジェンス）であると捉えたとき、CreativeGenome内における先述の(a)(b)(f)(g)(h)のような解釈・読解によるタギングこそが根源的に重要なＨＩ要素と考えられる。
このとき、次に課題となるのは(h)にて触れた「どのような価値観をもった受容者がそのシニフィエを好むか」という点であった。受容者の価値観が、コンテンツ評価に大きな影響を及ぼす要素と考えられるためである。

3.3　価値観・HIコンソーシアムについて

上記課題について、人間の価値観を定義ベースで独自に展開・可視化し更なる研究を続けるシナジーマーケティング㈱と、その手法や目指す世界観において共鳴し、また写真という表象に対してHIによる再定義を推進中の㈱アマナイメージズ、その他も加わることで、＜価値観・HIコンソーシアム＞[20]を発足することとなった。
データによる連関の部分では＜Creative Genome＞から仮説されるクリエイティブ仮説と、シナジーマーケティング㈱による価値観定義システム＜Societas＞による人間の価値観を組み合わせ、最適なクリエイティブを最適な価値観の受容者像に定性的、質的にマッチさせる機能開発を進めている。

4. HIとしてのデータ定義思想と将来展望

4.1　「キモチの領域」の定義

IoT領域など「モノ」を中心とした様々な計算可能領域に現行AIがますます実装されて行く中にあって、人間の感情すなわち「キモチの領域」に訴える「コンテンツ・ストーリー」の領域をいかに形式知・方法知に再統合できるか、そしてまた人々の「価値観」がどのような定義によってその揺動として捕捉でき、またコンテンツ等によって価値観の変動をする

のか、これらが社会の次フェーズにおいて注目分野となるであろうことは想像に難くない。

既に広告業界においてもバズ映像の特徴研究や、映像の自動生成へのトライアルなどでのAIディープラーニング活用が熱気を帯びているが、＜Creative Genome＞の読後感定義によるHI的アプローチは、昨今のAIによる計算をベースとした＜工学的な方法論＞とは一線を画するものである。

4.2 仮説創造、仮説発見としてのHIに向けて

ディープラーニングの層構造で計算可能性が莫大となった時代にあり、問題はどのような＜ディープなモデル＞が作れるかを考えたとき、想起されるのは小方(2018)による、深層学習の方法のみによる物語生成への疑義[20]であり、やはり人文知に裏打ちされたデータ定義や、システムのアブダクティブな仮説形成能力が不可欠であろう。2-2(3)で示したような仮説創造性を大きく作用させるために、以下の３ポイントをデータ・システムにいかに組み込み、作用させるかがHIにおいては特に重要になると考えている。

(1)アナロジー＝読後感や物語コンセプト、モチーフやトーンなどにおいての類似性
(2)アフォーダンス＝読後感を方向づける因子の発見と抽出
(3)仮説創造力（アブダクション）＝既存の手法の組合せによる新たな仮説創造

4.3 HIコンソーシアムにおける＜Creative Genome＞

CMやイベントの企画演出制作を黎明期から続け現場を熟知する立場において、過去の記号論や修辞研究、概念計算の足取りからのエッセンスも注入しつつ「コンテンツ・ストーリー」領域における形式知への取り組みを興すことは、来たる空間/環境における体験コンテンツ分野への応用にも繋がると考える。まずはCMという態度変容志向の短尺映像の表現構造を形式知に落とし込むが、広がるコンテンツ領域に左右されることのない、一定のクリエイティブ・データベース及びアルゴリズムを成立させ、徐々に実空間や環境にも知見を応用し、受容者価値観との結合を志向していく。

5. おわりに

＜Creative Genome＞において、さまざまな表象の型を見極めた先には、その型から共起関連する諸要素を結合した、新たな様々の「別様」な表象の可能性が引き出せる。（特にCMにおいては、期待される読後感がある程度厳密に制作側に想定されていること、マイナスの読後感を呼ぶ表現が予め制限されていることが研究の始点として適しているといえる。）

このような「別様を発見する回路」を通じて、HIコンソーシアムにおいて世界観の定義と密接に連関し、HIコンソーシアムが「新たな価値観」をさまざまに社会提案できるナレッジ主体となることを目指している。

そのためには、より詳細なコンテンツや価値観のコンテクストをHIのアプローチによって明示知にしながら、コンテンツや体験がどのように表象することで、社会が現在の価値観から「新しい価値観」に移動しうるのか、その連関について仮説創造していく必要がある。

こうしたことを念頭に、現在、HIコンソーシアムでは各種アクティビティの傍ら「連想」におけるその無意識の連関に注目し、スタディを始めた。

コンテンツの受容には、受容者側の何等かの個人的なコンテクストが働いている。上記タグにおける(f)(g)(h)のように、そこには文化ミーム、人生観、生活コンテクストがそれぞれ交差しているものと思われる。それらコンテクストがどのように意識にのぼり、行動に結びつくかを考えるとき、大きなバックグラウンドとして、ある世界観の個人がもつ「連想の性向」が関わると仮定し、この連想のつながり(連関)とそのパターンを類型化していくことで、コンテンツの受容や価値観をより精密化できると考えている。

参考資料・文献

[1] 佐々木淳（2017）Creative Genome Projectについて「人工知能学会第二種研究会　ことば工学研究会資料」**56**,25-32
[2] AOI Pro.ニュースリリース　2012.10.17 映像テキストサービス「nagarami」 http://www.aoi-pro.com/news/releases/20121017
[3] 岩手県立大学＋AOI TYO Holdings「HI(ヒューマンインテリジェンスを用いたクリエイティブ定義システム「Creative Genome」の精査、および映像文法解析のための共同研究)」(2017-)
[4] Roland Barthes ,“モードの体系　その言語表現による記号学的分析』”, みすず書房, 1972
[5] Charles Sanders Peirce, “アブダクション：仮説と発見の論理”, 勁草書房,2007
[6] Roland Barthes ,“物語の構造分析”, みすず書房, 1979
[7] Ferdinand de Saussure, “ 一般言語学講義”, 岩波書店,1972
[8] 松岡正剛, ”知の編集工学”, 朝日文庫, 2001
[9] 秋元泰介・小方孝（2013）, “物語生成システムにおける物語言説機構に向けて―物語言説論と受容理論を導入したシステムの提案―” Cognitive Studies,20(4), 396-420.
[10] 小方孝・金井明人, ”物語論の情報学序説―物語生成の思想と技術を巡って” 学文社, 2010
[11] Vladimir IAkovlevich Pro,p、“魔法昔話の起源”, せりか書房, 1983
[12] Algirdas Julien Greimas, “構造意味論―方法の探求”, 紀伊國屋書店, 1988
[13] 松岡正剛, ”物語編集術”, 講談社現代新書, 2000
[14] Joseph Campbell,”千の顔を持つ英雄”, 人文書院, 1984
[15] Gerald Prince,”物語論辞典”松柏社, 1997
[16] Aristotélēs,″トピカ″, 京都大学学術出版会, 2017
[17] Marcus Tullius Cicero,”キケロー選集”, 岩波書店, 1999
[18] 松岡正剛, ”日本流　なぜカナリヤは歌を忘れたか”, 朝日新聞社, 2000
松岡正剛, ”日本数寄”, 春秋社, 2000
[19] Falon Fatemi “Artificial Intelligence and Human Intelligence: The Essential Codependency”, Women Forbes,2017-10-24 ※一例として https://www.forbes.com/sites/falonfatemi/2017/10/24/artificial-intelligence-and-human-intelligence-the-essential-codependency/#2a4aac1d2efb
[20] 小方孝（2018）物語生成の物語論あるいはポストナラトロジー／プレナラトロジーのためのノート「人工知能学会第二種研究会　ことば工学研究会資料」**57**,106-107

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

双方向コミュニケーションと心のデザイン

'Two-way communication' and 'Mind design'

谷田　泰郎[*1]　　高椋　琴美[*1]

[*1] シナジーマーケティング株式会社
Synergy Marketing, Inc.

Communication based on understanding of human is very important. As shifts to two-way communication to consumers progress, people's understanding with sensuous sensibility values is necessary. In this paper, we present the design of the mind to express sensory sensibility values.

1. はじめに

人にとってコミュニケーションほど重要なものはない．サービス受容者の行動を読み解く上でも，コミュニケーションを切り離すことはできない．様々な人のコミュニケーションの受発信に現れる特徴は，その人の性格や価値観などの特徴を表すものであり，消費やサービス受容行動の特徴とも関係が深いからだ．シナジーマーケティング社の「価値観」ベースの類型である'Societas'を活用した取り組みは，価値観を知ることで消費者とのよりよい関係を構築していくことが目的になっており，筆者らは，人と人，人とモノとの出会いから生まれるコトに対して感じる価値をより大きなものにするために，人の心のデザイン（心のモデルのようなもの，現在は価値観フレームがそれにあたる）を中心にした研究を行うことで「感じる」をより心地よいものにするという究極のコミュニケーションの実現を目指している．

また，世の中の多くのデータを持ち寄って付加価値を高めるための「社会知ネットワーク」という構想[谷田 2014]に基づいて，独自の価値観フレームを中心にしたデータ連携から得られるエピソードを実践的に定量データや定性データのデータベースとして蓄積している．このように，価値観と事実データとの関係性を強化することで事業活動に利用される機会が増えてきた．この事業を推進するための方針に'Everywhere'，'Inside'というキーワードが使われている．すなわち，様々なサービスの顧客 ID に適用し，また，'Societas'を使ったサービスやアプリケーションを提供するのではなく，サービスの中のプロセスの一部として組み込むという考え方である．このように'Societas'が広く活用されるようになることで社会知データベースの内容が拡充し，その恩恵が再び社会に還元されるという循環を生み出すことになる．このように事業化が進む一方で，現状の価値観枠組みでは解けない課題も表面化してきた．本稿では，現状の価値観枠組みの課題を整理し，今後，その課題解決の方向性を提案したい．

2. サービスの双方向コミュニケーションに向けて

2.1 現状の価値観枠組みの課題

サービスやブランドに対して求めるものがモノからコトへとシフトしつつある．今後，サービス提供者とサービス受容者のコミュニケーションは現状のように提供者から受容者への一方向だけではなく，サービス受容者がサービス提供者のメッセージを受け取り，サービス受容者に対して適度なタイミングでメッセージを返すことを繰り返すような，双方向で対話的なコミュニケーションになっていくと予想される．また，サービス提供者がサービス受容者のコミュニティーに積極的に参加し，対話と理解によって，そのコミュニティーに自然に溶け込んでいくことも重要になるだろう．

現在の価値観枠組みは，表 1 に示すように，'gene'（先天的な遺伝傾向），'char'（性格），'meme'（後天的な価値観）というように 3 つの階層の価値観変数から構成され[斉藤 2016]，それを'Societas'と呼ばれる 12 の価値観類型に分けて活用している．これらの枠組みの特徴として，それぞれの価値観変数の変化が長期的で，その範囲が広域的であることがあげられる．この特徴のおかげで様々なデータ間での比較はしやすい．その一方で，サービス受容者が求めるサービスに対する感覚や追い求める概念は変化しやすく短期的で，範囲も局所的なものが多い．現状の価値観変数は一般的で論理性が高いが，現実はもっと曖昧で感覚的な感性価値観を必要としているのである．サービス提供者とサービス受容者のコミュニケーションを双方向で対話的にしていくためには，現状の価値観枠組みでは説明できない「感覚的な感性価値観」を取り扱えるように対処していく必要がある．

表 1:'Societas'の価値観

カテゴリ	価値観変数名			
gene	新規性追求	損害回避	個人主義	協調性
(普遍的)	頑固	柔軟	誘惑に弱い	我慢強い
character	好奇心	自己成長	協調的	批判的
(性格)	自己愛	繊細	のんびり	
meme	ストレス	仕事充実	家族不仲	結婚願望
(後天的)	お金ゆとり	家族中心	親友あり	多忙

2.2「感覚的な感性価値観」の重要性

人間は多くの時間を1対1のコミュニケーションに費やしており，1日のうち 6～12 時間を知り合いと 1 対 1 で対話し，そのうちの80-90%を世間話に費やしているというが[ガザニガ 2010]，人間どうしのコミュニケーションは，言語，汎言語，パラ言語，どのレベルをとってもいい加減で，空想に満ちている．筆者らは，その「いい加減な」対面対話からの情報抽出を通じて，個性や人と人との関係，言語的な刺激と感情など，様々なコミュニケーション要素を心のデザインとして活用したいと考え，それらのエビデンスが対話の中に多く現れるという仮説をもって，対話データか

連絡先：谷田泰郎，シナジーマーケティング株式会社　研究企画チーム，電話番号：0774-95-2305，メールアドレス：rd-plan@synergy101.jp

ら個性を抽出する研究を進めていた[谷田 2016][高椋 2017]. インターネット社会となってしまった現在では，リアルの対面対話に費やしていたコストが SNS やチャット等のインターネット上のコミュニケーションに移行していると思われるが，チャットのような書き言葉が中心になっても，いい加減であることに変わりはない. いい加減で，取りとめもないやり取りのようでありながら，感覚的な感性価値観や概念をコト的に挟みながらコミュニケーションを図っているのである. その中には様々な人間関係や制約条件も含まれる. 過去に価値観枠組みの検証の起点として，'Societas' の価値観変数と twitter のテキストデータとの関係性モデルの検証報告をしたが[谷田 2013]，最終的に定量化する際には、このような機械学習によるモデル化が必要になるだろう.

本稿では，「感覚的な感性価値観」と表現しているが，いい加減なコミュニケーションを成立させるために重要なそれはいったいどんなものなのだろう？意識とか無意識とか見えない心の中を議論し続けるのは不毛だ. 少なくとも，自己や他者に向けた行動として現れる欲求，具体的な行動に至らずとも独り言でも口に出すことができる欲求や感覚・感情として言語化した状態で表象されうるものを対象としたい. 筆者らは，マズロー[マズロー 1987]等がリストアップしていた人間の欲求を整理すると共に，欲求を煩悩に置き換えた時，その対極にあると思われる仏教の考え方[鎌田 1986][増谷 2005]を検討した. その結果，図 1 に示すような感覚的な感性価値観を取り扱うための初期デザインを得た. このデザインはあくまでも仮置きであり，今後研究を進めていく試行錯誤の中で変更されるであろう.

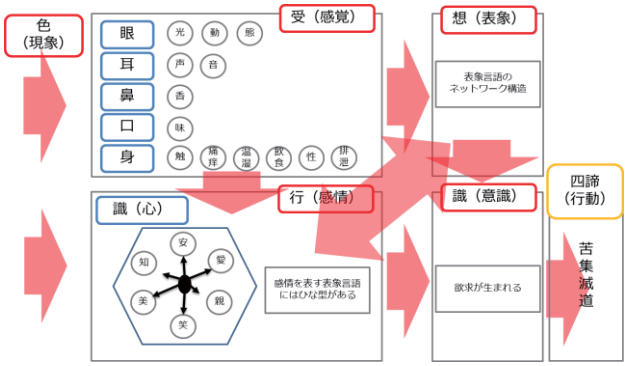


図 1:'感覚的な感性価値観を取り扱うための枠組み(仮)

人間は入力の刺激や現象(図 1 の左側に色と表示されている箇所)を 6 つの感覚器(目・耳・鼻・口・身・心，六根という)をフルに活用して感覚を生み出している. この感覚は言語的表現では交差する(視覚的な言語が聴覚にも使われたり，味覚的な言語が感情にも使われたり etc.). 入力の現象は，モノであったりコトであったりする. 生み出される感覚や感性は感情や欲求を生み出すと共に言語として表象されている. 筆者らはその順序にこだわりはない. ほぼ同時でよい. 識(感情)の部分が苦楽のパラメータであり，想(表象)には，表象される言語のある程度階層的なネットワーク構造を想定している. この構造とそのリンクや苦楽パラメータの強弱に人間の感覚的な感性価値観の個性性があると仮定している. 表象される言語のある程度階層的なネットワーク構造は，連想構造に近いものかも知れない. そして，欲求を満たすために，あるいは捨てるために，図 1 の右側の行動に移る.

3. データ収集の方針

筆者らが今後取り組むべきことは，事実データをこのデザインに当てはめ，最終的には定量的に扱えるようにすることである. 前項で示した心のデザインにコミュニケーション系の事実データを当てはめていくための方向性を示しておく.

1 つ目のアプローチは，何らかのブランドを仮定して，そのブランドに向けたコミュニケーション分析を行う予定である. 定性的にそのブランドに対するいくつかのコミュニティーや「感覚的な感性価値観」を想定し，SNS 等から収集したデータを吟味して定量化の方向性を探る方法である. この方法は目的依存性が高く，進め方がゴールからのスタートになるため，データ収集や分析は比較的安定的に進められるが，分析結果を仮説の心のデザインに当てはめるのが難しい.

2 つ目のアプローチは，6 つの感覚器それぞれについての現象を想定し，その現象が起こった時の連想構造を定性的・定量的調査によって収集する方法である. この時，想定する現象が恣意的になる，どのような調査回答を期待するのか(紋切り型の単語なのか，短文なのか etc.)等，課題は多いが，失敗を繰り返しながら方向性を探るしかないだろう.

4. おわりに

サービス提供者と受容者の双方向コミュニケーションへのシフトが予想される中，価値観枠組みでは解けない問題，「感覚的な感性価値観」の必要性を痛切に感じている. 事実データと言語，言語と「感覚的な感性価値観」における個性が定量的に取り扱えるようになれば，クリティティブやブランディングにより精緻で分かりやすい示唆を与えることができる. また，紹介した連想と価値観の関係性をデザインする研究は，筆者が別途参加している「価値観・HI コンソーシアム」の共同研究の中でも進める予定である. 現状では「感覚的な感性価値観」に関する調査を行ったわけではないが，「サバイバルチャレンジ：世界価値観と国際マーケティング」の当日の発表では，より具体的な調査等の研究成果について紹介させていただくつもりである.

参考文献

[谷田 2014] 谷田泰郎: 価値観マーケティングと社会知ネットワーク，人工知能 9 月号 Vol.29 No.5 P456-P463，2014.

[斉藤 2016] 斉藤有紀子他: 複数データのマッピングによるシニア価値観分析の試み，人工知能学会全国大会(第 30 回)JSAI2016, 2016.

[ガザニガ 2010] マイケル・S・ガザニガ: 人間らしさとは何か？，インターシフト，2010.

[谷田 2016] 谷田泰郎，高椋琴美: コミュニケーションにおける記憶と個性の評価，人工知能学会全国大会(第 30 回)JSAI2016, 2016.

[高椋 2017] 高椋琴美，谷田泰郎: 言語特徴や音響特徴が記憶や印象に与える影響，電子情報通信学会研究報告，信学技報，2017.

[谷田 2013] 谷田泰郎，河本裕輔他: マイクロブログにおける潜在的価値観の推定，人工知能学会全国大会(第 27 回)JSAI2013, 2013.

[マズロー 1987] A.H.マズロー，小口忠彦訳:人間性の心理学，紀伊國屋書店，1987.

[鎌田 1986] 鎌田茂雄: 般若心教講話，講談社，1986.

[増谷 2005] 道元，増谷文雄全訳注:正法眼蔵(1)～(8)，講談社，2005.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

歌舞伎における人物のテクスト解析と構造化
―物語生成システムにおける利用のために―

Text Analysis and Structuring of Persons in Kabuki: For the Use in the Narrative Generation Systems

小方孝[*1]
Takashi Ogata

[*1] 岩手県立大学
Iwate Prefectural University

In this paper, I aim to expand my research on the multiplicity of persons in kabuki. In particular, I propose a method for representing hierarchically the information of kabuki actors, which is divided into *geinojin* resource and life resource, using text data (Japanese Wikipedia) and I translate the text data into the information based on a property representation form to use it in my two types of narrative generation systems called *geino* information system and integrated narrative generation system. Last, I discuss various effects for expanding narrative multiplicities from kabuki person information that was divided into *geinojin* resource and life resource.

1. まえがき

筆者は物語生成システムとの関わりで歌舞伎の研究を行って来た．これまで，一種の物語生成機構としての歌舞伎の予備的・総合的考察を行い，複数の観点から歌舞伎の多重物語構造を具体的に調査・分析し，そのうち歌舞伎の人物，ストーリー及び「尽し」をより詳しく検討した．同時に，歌舞伎の調査・分析・考察により得られた知見を筆者が研究して来た物語生成システム―芸能情報システム [小方 印刷中 c]と統合物語生成システム [Ogata 2016]―に導入する方法についても考えて来た[小方 2017, 印刷中 a, 印刷中 b; Ogata 2017, 2018; 森田・小方 2018].

これらの研究を支える基本的観念は，物語の特徴は多義性にあるというものである．例えばコミュニケーションという観点から物語を考える時，そこには確かに送り手が受け手に向けて何かを伝達するというコミュニケーション構造が存在するが，このコミュニケーションにおいては，一つの意味が送り手から受け手に正確に伝えられるということは重視されない．寧ろ，直接的な意味・間接的な意味・論理的な意味・感性的な意味・隠された意味等諸々の意味の複合体として，いわば相互に否定し合う情報の矛盾に満ちた共存として，物語が存在するということに価値が置かれる．その意味で物語とは極めて優柔不断なものである．筆者が考案した多重物語構造モデル [小方 2000a, 2000b]は，優柔不断なものとしての物語を構成しまた解釈するためのモデルあるいは方法であることも意図している．歌舞伎は物語における多義性や多重性の諸相を見るのに非常に向いた物語のジャンルである．

[小方 2017; Ogata 2016]は，歌舞伎という芸能としての物語における芸能人（歌舞伎役者）について調査し，ある一人の人物は主に，現実の顔・役者としての顔・（舞台上の）登場人物としての顔という三つの顔を持っていることを指摘した．例えば，歌舞伎には名跡が存在する．名跡とは表面的には代々継承される役者名のことである．名跡の襲名は，歌舞伎のみならず落語や家元制度を採る各種の芸能（日本舞踊など）に見られる制度・慣習であるが，これは個人を超えた芸能人としての一貫した物語を構成し，名跡の継承により信用・伝統・歴史・イメージ・芸統・気風が確立される．すなわちそれは個人を超えた芸能者もしくは芸術家としてのブランドを構成する．

また，上記多重物語構造モデルや[小方 2002]では，芸能人を巡る物語（ストーリー）の多面性は，さらに次のような特色も持っているとされる．すなわち，個々の芸能人の物語の背後には，組織や社会のレベルにおける大きなストーリーが控え形成されている．芸能人一人一人のストーリーをミクロストーリーと考えると，所属するプロダクションやその社会的流通のレベルにおけるストーリーはマクロストーリーと呼ぶことが出来る．こうして，ミクロストーリーとマクロストーリーの複雑な総合によって，芸能人全体の物語が形成されて行く．歌舞伎は，個々の作品のレベルとその連鎖のレベルとの複合を通じて，多重的な物語を浮上させる．

以上のように，歌舞伎における人物の多重性は，ある人物から見れば同時並行的に諸種の物語を生成・展開して行くのに貢献するし，また歌舞伎の受け手に対してはある人物に関する多重的な観念を生じさせそれを様々な視点から眺め享受する娯しみを与える．

本稿は歌舞伎における人物の多重性に関する研究を発展させるものである．以下，テキストデータから歌舞伎役者の人物情報を記述するための知識構造を確定し，人物に関する具体的情報を属性記述としてテキストデータから獲得して，物語生成システム（芸能情報システム及び統合物語生成システム）で使用可能な形式にし，その使用の可能性及び方向性について考察する．

2. 歌舞伎役者のテクスト解析

まず，芸能人（歌舞伎役者）情報を調査・分析することによって，その知識記述構造を決定した．具体的には，テキストデータとして今回は Wikipedia をもとに，数人の芸能人情報の構造を分析し，知識記述構造を決定した．初めに十一代目市川海老蔵（以下，市川海老蔵）に焦点を当ててこれを行った．市川海老蔵の Wikipedia を見ると，次のような項目が階層的に含まれていた―[**概要**;**1 年譜**;**2 人物**:**芸に対する姿勢**, **趣味・嗜好**, **その他**, **父・團十郎との関係**, **妻・小林麻央**;**3 受賞歴**;**4 出演**

連絡先：小方 孝，岩手県立大学ソフトウェア情報学部，岩手県滝沢市巣子 152-52，t-ogata@iwate-pu.ac.jp

作品:歌舞伎, 海外公演, テレビドラマ, テレビアニメ, 吹き替え, 映画, その他の舞台, CM;5 (論文);6 家族・親族].

これの項目をもとにシステムでの利用を考慮し分析・再分類を行った. 基本構造として, この情報内容を, 芸能情報と人生情報の大きく二つに分類し, さらに複数の下位分類が付随する階層構造として定義した. ここで, 芸能情報とは歌舞伎など芸能活動そのものやそれに関連する内容であり, 人生情報とは私生活や家族に関する情報を意味する. 芸能情報や人生情報にそのまま分類可能な項目, 例えば出演作品は芸能情報, 家族・親族は人生情報というようにそのまま再分類し, 人物の項目など芸能情報と人生情報に内容が跨っているものに関しては, 別途ボトムアップ的にグループを作成し分類を付与した. 市川海老蔵の芸能情報・人生情報の構造をそれぞれ図 1 と図 2 に示す.

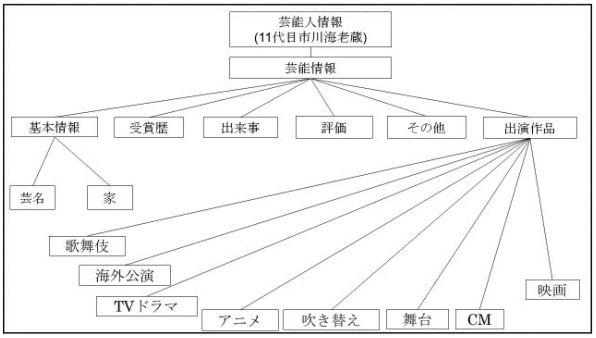


図 1: 市川海老蔵の芸能情報の構造(原拠:[森田 2018])

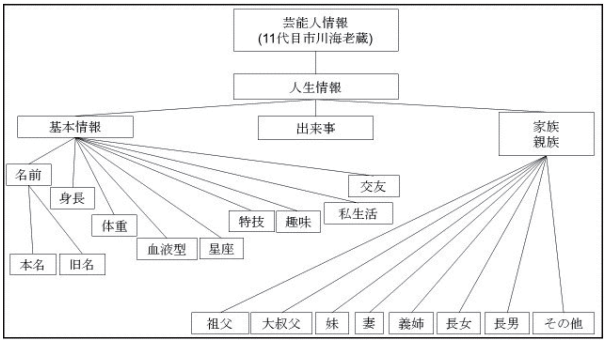


図 2: 市川海老蔵の人生情報の構造(原拠:[森田 2018])

次に, その他の現役の歌舞伎俳優五人―六代目尾上菊五郎, 二代目中村吉右衛門, 五代目坂東玉三郎, 四代目市川猿之助, 九代目市川中車―に関しても, 市川海老蔵と同様の方法で調査・分析を行った. 一例として九代目市川中車の Wikipedia は次の項目を階層的に含む―**[概要;1 略歴:生い立ち, 俳優として, 歌舞伎界進出;2 人物:俳優業, ボクシング愛好家, 昆虫好き;3 家族・親族;4受賞歴;5 出演:テレビドラマ, 映画, オリジナルビデオ, 舞台, テレビアニメ, 劇場アニメ, 海外ドラマ, 海外アニメ, ミュージックビデオ, バラエティ, ドキュメンタリー, CM;6 歌舞伎;7 著作]**. これをもとに作成した芸能情報・人生情報の構造を図 3 及び図 4 に示す.

これらを総合して, 芸能人情報の共通形式(構造), 一般形式(構造)を定義した. これを図 5 に示す. それぞれの下位項目は, 歌舞伎役者の芸能人情報の共通項として定義できると判断した項目である. この八項目は, どの歌舞伎役者にも, 名称が異なる場合もあったが, 存在していた. 項目名は市川海老蔵の情報を基準としている.

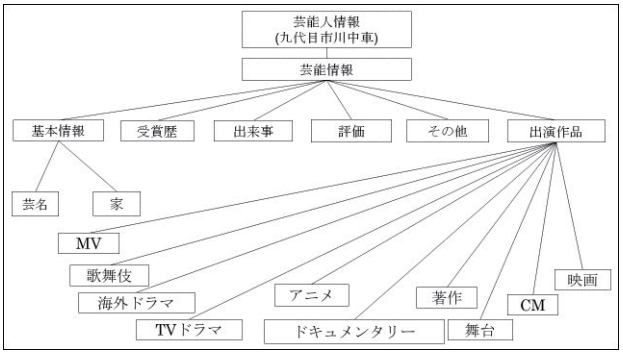


図 3: 市川中車の芸能情報の構造(原拠:[森田 2018])

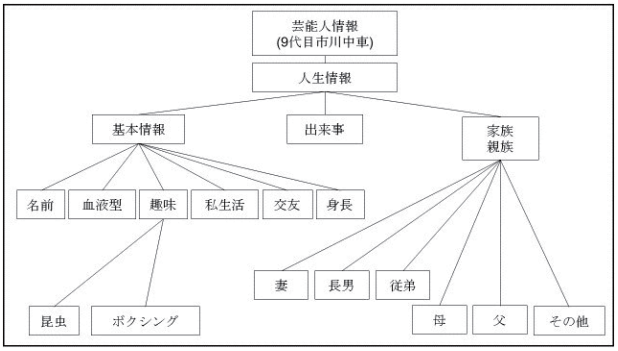


図 4: 市川中車の人生情報の構造(原拠:[森田 2018])

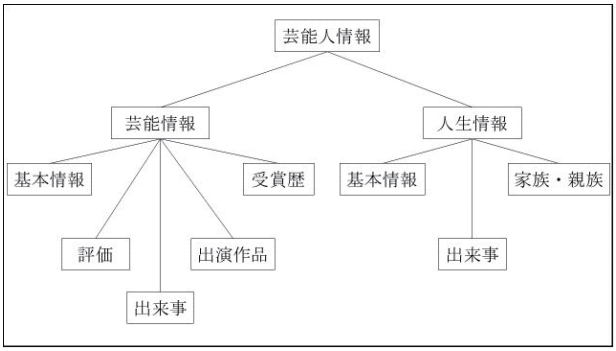


図 5: 歌舞伎俳優の芸能人情報の構造(原拠:[森田 2018])

3. 歌舞伎役者の人物属性情報の獲得

上述のような方法で整理・構造化した歌舞伎役者の芸能人情報を, システムで利用可能な形式にするために, まず当該歌舞伎役者の Wikipedia 全文を, 小野・小方 (2017)が開発している属性獲得機構に通し, 属性の獲得を行った. ここで属性とは, ある歌舞伎役者の特徴・性質に関する記述であるが, 具体的には対象 Wikipedia の文章をすべて後述する一定の形式に記述し直した情報を意味する. そしてこの属性情報を, 各々の歌舞伎役者ごとに, 前節で説明した構造に沿って整理し直した. 次節で述べるように, こうして構成された構造化された属性情報が芸能情報システムの一部として位置付けられる.

属性フレーム獲得の詳細アルゴリズムを図 6 に示す. また, 実際に属性獲得機構に市川海老蔵の Wikipedia の情報を通した結果の一部を図 7 に示す(上が入力文章, 下が出力結果(属性情報)). さらに, 今回考案した歌舞伎役者の芸能人情報の知識ベースの記述形式を図 8 に, その実例として作成した市川海老蔵の知識ベースの一部を図 9 に示す.

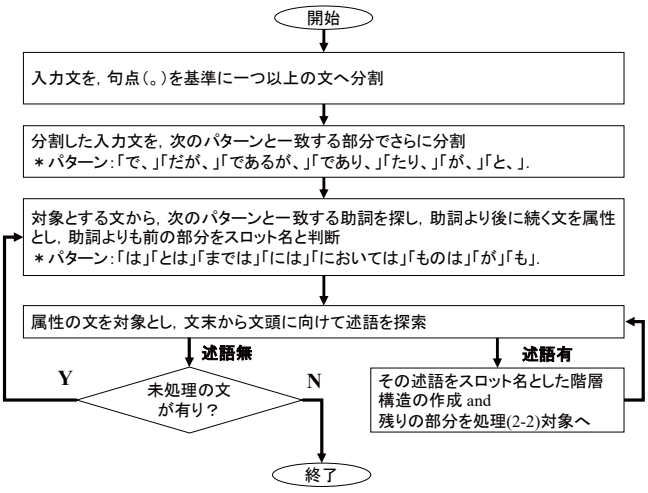


図 6: 属性情報獲得アルゴリズム

入力	十一代目 市川 海老蔵（じゅういちだいめ いちかわ えびぞう、1977年12月6日 - ）は、日本の歌舞伎役者。歌舞伎名跡「市川海老蔵」の当代。屋号は成田屋。定紋は三升、替紋は杏葉牡丹。 ・・・中略・・・ 身長176cm、体重80kg。血液型はAB型。星座は射手座。 特技はバスケットボール、趣味は相撲鑑賞。学生時代はバスケットボール部に所属していた。スタジオジブリのアニメ作品の大ファン。 ・・・後略・・・
出力	((("は" "歌舞伎役者、俳優。" (")は" "十一代目市川海老蔵（じゅういちだいめいちかわえびぞう、1977年12月6日-")) (nil "nil")) (("は" "当代。" ("」の" "歌舞伎名跡「市川海老蔵")) (nil "nil")) (("は" "杏葉牡丹。" ("は" "三升、替紋" ("は" "成田屋。定紋" ("は" "屋号")))) (nil "nil")) ・・・中略・・・ ((“は” “身長176cm、体重80kg。”) (nil “nil”)) (("は" "AB型。" ("は" "血液型")) (nil "nil")) (("は" "射手座。" ("は" "星座")) (nil "nil")) (("は" "相撲鑑賞。" ("は" "バスケットボール、趣味" ("は" "特技"))) (nil "nil")) (("は" "た。" ("に所属してい" "バスケットボール部" ("は" "学生時代"))) (nil "nil")) (("は" "ファン。" ("の" "アニメ作品" ("の" "スタジオジブリ"))) (nil "nil")) ・・・後略・・・

図 7: 属性獲得機構の入出力例

```
(<人名>
  (geinojinInfo
    (geino ;芸能情報
    (g-base <data1> <data2> … <dataX>) ;基本情報
    (evaluation <data1> <data2> … <dataX>) ;評価
    (g-event <data1> <data2> … <dataX>) ;出来事
    (work <data1> <data2> … <dataX>) ;出演作品
    (prize <data1> <data2> … <dataX>)) ;受賞歴
    (life ;人生情報
    (l-base <data1> <data2> … <dataX>) ;基本情報
    (l-event <data1> <data2> … <dataX>) ;出来事
    (family <data1> <data2> … <dataX>)) ;家族・親族
))
```

図 8: 知識ベースの記述形式

4. 物語生成システムへの組み込みと利用方法

以上の方法によって獲得・記述・構造化された歌舞伎役者の属性情報を，芸能情報システム (以下，GIS: *Geino* Information System) [小方 印刷中 c]と統合物語生成システム (以下，INGS: Integrated Narrative Generation System) [Ogata 2016]とどのようにつなげるかについて述べる.

4.1 GIS における位置付け

図 10 は GIS の概念図であるが，この中の芸能作品資源と人生作品資源（単に芸能資源，人生資源とも呼ぶ）の中に，歌舞伎役者の芸能属性情報と人生属性情報が格納される.

```
(*Ebizou*
  (geinojinInfo
    (geino ;芸能情報
      (g-base ;基本情報
          (("は" "歌舞伎役者、俳優。" (") は" "十一代目市川海老蔵（じゅういちだいめいち
                かわえびぞう、1977年12月6日-")) (nil "nil"))
          (("は" "杏葉牡丹。" ("は" "三升、替紋" ("は" “成田屋。定紋" ("は" "屋号")))) (nil
                "nil"))
           ・・・中略・・・)・・・中略・・・)
    (life ;人生情報
      (l-base ;基本情報
          (("は" "身長176cm、体重80kg。") (nil "nil"))
          (("は" "AB型。" ("は" "血液型")) (nil "nil"))
          (("は" "射手座。" ("は" "星座")) (nil "nil"))
          (("は" "相撲鑑賞。" ("は" "バスケットボール、趣味" ("は" "特技"))) (nil "nil"))
          (("は" "た。" ("に所属してい" "バスケットボール部" ("は" "学生時代"))) (nil "nil"))
          (("は" "ファン。" ("の" "アニメ作品" ("の" "スタジオジブリ"))) (nil "nil"))
           ・・・中略・・・) ・・・中略・・・)
))
```

図 9: 知識ベースの一部

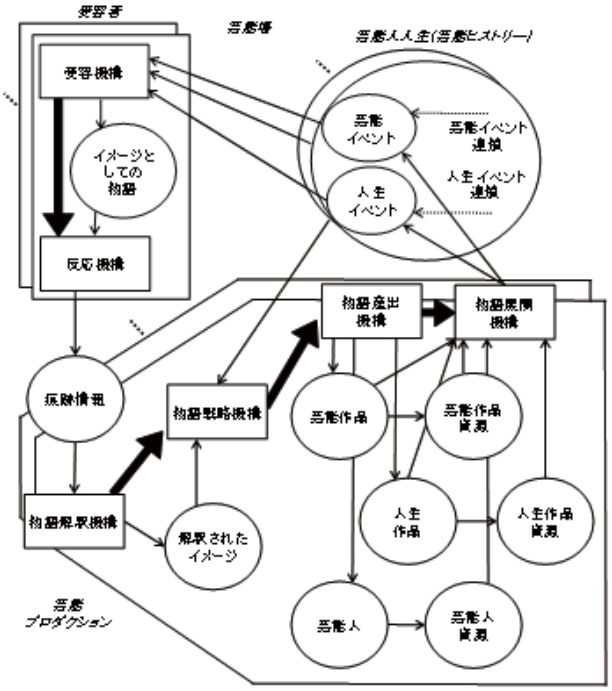


図 10: GIS 原案概念図

4.2 INGS との関係

図 11 は，GIS がその中に INGS を含んで個々の物語からその連鎖までを生成・産出し，その受容・消費が行われ，これらの過程がさらに循環する様子を示しているが，この中で，芸能人資源の一部を成す芸能資源と人生資源とは INGS によって生成された物語との間で多重的な関係を持つ. 図12 に描くのはこの多重的関係のイメージである. INGS 単体で生成される物語における人物は，ある物語の登場人物（キャラクター）のレベルでの人物であるが，GIS に定義された歌舞伎役者の人物と接続されることで，その物語は多重化される. すなわち，登場人物としての人物は，その下層に存在する芸能人としての人物及び個人としての人物と関連付けられ，常時これら三つの側面のどこからでも見られ得る，あるいは合成され得る，人物になる. 同時に，これら三種の人物は，同じく図 12 にそのイメージが示されるように，当該の物語作品から独立した三種の物語―芸能人としての物語，個人としての物語及び登場人物としての物語とも関連付けられている. このように，生成されたある物語作品そのものが多重的に構造化されると共に，各種の人物の物語も同時並行的に生成される.

なお，今回芸能人の人物情報として扱ったのは芸能人としての人物情報と個人としての人物情報だけであったが，実際はさらに上述のように登場人物としての人物情報が存在する．特に江戸時代までの歌舞伎においては，多くの人々が一般教養として知っている歴史的人物や虚構の人物が，頻繁に繰り返し様々な作品に登場し，そこからある特定の登場人物の物語というもう一つの物語のレベルが生成・展開された．現在でも，西郷隆盛のような著名な歴史的人物，テレビの連続ドラマに出る主人公や登場人物などにも，同じような現象は見られるが，歌舞伎の場合，このような特定の登場人物の連続出演がいわば激しく行われた．本稿ではこれについて扱わなかったが，この種の登場人物情報も今後 GIS の中の芸能資源の一部として持たせることを考えている．

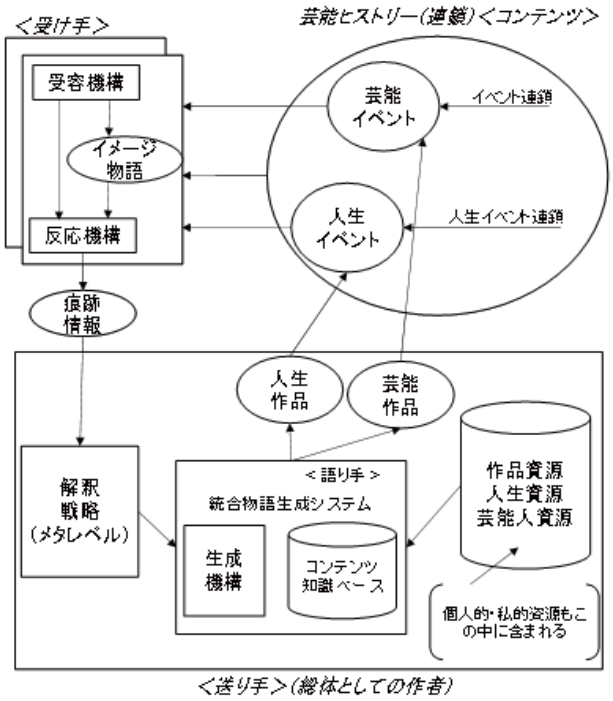


図 11: GIS と INGS の関係（原拠: [小方 2018]）

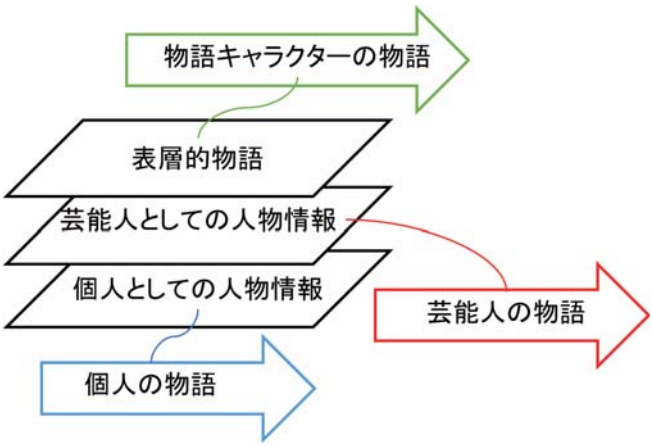


図 12: 人物の多重性と物語

5. あとがき

本稿では，歌舞伎における人物の多重性に関する研究を発展させるために，テキストデータ（Wikipedia）における歌舞伎役者の情報を芸能人資源と人生資源とに分け，それぞれの階層的な構造化方法を決定し，さらにこのテキストデータを属性情報として，物語生成システム（芸能情報システム GIS や統合物語生成システム INGS）で使用可能な形式にした．その上で，これらの情報を GIS や INGS で使用する方法と物語の多重性を増幅する効果について述べた．今後取り組みたいことの一つは，属性獲得の対象となるテキストデータを，歌舞伎役者や芸能人に対する人々の気持ちがより生の形で表出されていると思われる週刊誌や女性雑誌などに拡張し，多重性の幅を広げて行く試みである．

謝辞

本研究は一部，SCAT 研究助成 (2015-2017)，柏森情報科学振興財団 (2017)及び電気通信普及財団 (2017)及び AOI TYO Holdings 株式会社（共同研究）からの研究助成の支援を受けている．

参考文献

[森田 2018] 森田実希・小方孝: 芸能情報システムのための歌舞伎役者のキャラクター情報の獲得. 第 57 回ことば工学研究会資料, pp.1-5, 2018.

[小方 2000a] 小方孝: 多重物語構造のマクロモデル―シミュレーションとしての物語序説―. シミュレーション＆ゲーミング, Vol.10, No.1, pp.35-46, 2000.

[小方 2000b] 小方孝: 多重物語構造モデルとその展開へ向けて―シミュレーションとしての物語序説―. In 川村洋次・浜田秀・小方孝 編, 文学と認知・コンピュータ 6―ことばと文学―, 日本認知科学会テクニカルレポート 00-No.32, pp.87-119, 2000.

[小方 2002] 小方孝・網野智幸: 多重物語構造と芸能キャラクターの物語生成システム. 日本認知科学会第 19 回大会発表論文集, pp.130-131, 2002.

[小方 2017] 小方孝: 歌舞伎における「人物」の多重物語構造. 第 52 回ことば工学研究会資料, pp.113-118, 2017.

[小方 印刷中 a] 小方孝: 歌舞伎に向けて (1)―恣意性と編集性の物語から多重性と実存の物語へ―. In 小方孝・川村洋次・金井明人, 情報物語論―人工知能・認知・社会過程と物語生成―. 東京: 白桃書房, 印刷中.

[小方 印刷中 b] 小方孝: 歌舞伎に向けて (2)―多重物語構造の諸相―. In 小方孝・川村洋次・金井明人, 情報物語論―人工知能・認知・社会過程と物語生成―. 東京: 白桃書房, 印刷中.

[小方 印刷中 c] 小方孝: 外部への物語生成または芸能情報システムに向けて. In 小方孝・川村洋次・金井明人, 情報物語論―人工知能・認知・社会過程と物語生成―. 東京: 白桃書房, 印刷中.

[Ogata 2016] Ogata, T.: *Kabuki* as Multiple Narrative Structures. In T. Ogata & T. Akimoto (Eds.), *Computational and Cognitive Approaches to Narratology*. pp.391-422. PA: Information Science Reference (IGI Global), 2016.

[Ogata 2018] Ogata, T.: An Integrated Approach to Narrative Generation: From Mishima and Kabuki to Narrative Generation Systems. In T. Ogata & S. Asakawa (Eds.), *Content Generation Through Narrative Communication and Simulation*. pp. 49-147. PA: Information Science Reference (IGI Global), 2018.

[小野 2017] 小野淳平・小方孝: Wikipedia を利用した概念辞書における属性情報の獲得と物語自動生成ゲームでの利用. 人工知能学会全国大会（第 31 回）論文集, 1D3-OS-29b-2in2, 2017.

[2H4-NFC-4b] [NFC Session] NFC-4 5C (Comprehending Consumers: Computing Complexity of Cultures)

座長:谷田 泰郎（シナジーマーケティング）

Wed. Jun 6, 2018 5:20 PM - 6:40 PM Room H (10F Sky Hall)

[2H4-NFC-4b-01] A Latent Class Analysis using Association between Consumer Values and Behaviors.

○Yoshihide Nishio[1] （1. Synergy Marketing, Inc.）

5:20 PM - 5:40 PM

[2H4-NFC-4b-02] A Study of the Values and the Revisit Intention of Inbound Tourists in Japan

○Hiromi Kamata[1] （1. Hitotsubashi University）

5:40 PM - 6:00 PM

[2H4-NFC-4b-03] Inbound tourism Marketing Model Green AI Academy (GAIA)

○Izawa Yasushi[1] （1. EcoTourLab EcoGuideCafe）

6:00 PM - 6:20 PM

[2H4-NFC-4b-04] Stability and similarity of clusters under reduced response data

Marisciel Litong-Palima[1], Kristoffer Jon Albers[2], ○Fumiko Kano Gluckstad[1] （1. Copenhagen Business School, 2. Technical University of Denmark）

6:20 PM - 6:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

価値観と行動の相関に注目した消費者の潜在クラス分析

A Latent Class Analysis using Association between Consumer Values and Behaviors.

西尾　義英[*1]
Yoshihide NISHIO

[*1] シナジーマーケティング株式会社
Synergy Marketing, Inc.

Consumer segmentation is important in marketing. Consumers can be segmented by their psychological attributes or clustered with their purchase behaviors. But can we make segmentation both psychological and behavioral? We propose a latent class analysis using association rules between consumer values and behaviors. We show such latent classes balances consumer characteristics on values and behaviors in 4 test cases.

1. はじめに

マーケティングにおける消費者のセグメンテーション手法には古典的な年代や性別といったデモグラフィック属性によるものから心理的属性や行動履歴に基づく高度なものまで存在する. 中でも我々は消費者の価値観とその類型化に注目して, Societas[谷田 2013]という分析フレームを開発し, クライアント企業に活用いただいている. Societas では 2018 年 3 月現在,12 種の類型と 23 種の価値観変数[1]および年代・性別・未既婚・子どもの有無という 4 種のデモグラフィック属性を規定している. これらはさまざまなクライアント企業が持つデータを匿名状態でつなぐための共通変数の役目を果たす. これにより我々のクライアントは顧客について自社だけでは得がたい情報を得ることができる.

Societas の活用とは, まず 12 種の類型から一つまたは複数の類型をターゲットと定め, その特性を知ることでアプローチを変えるということが基本である. ただし 12 種では多いあるいはより細かくしたいという場合など, クライアント独自の類型を定義する場合もある. このとき注意すべきことが 2 点ある. まず, クライアントの視点から類型同士の違いが明らかであるということが欠かせない. これは類型の違いでアプローチを変えたいという理由による. クライアントの視点からということは, クライアントが独自に持つデータ, 例えば商材の購買傾向において差が大きいと言い換えることができる. もうひとつは我々にとって重要な事柄であるが, 類型がクライアントに閉じてしまわないようにしたい. クライアント固有のデータを良く分類できるように最適化しすぎると, 共通変数たる Societas との紐付けが難しくなり, 他で活用しにくくなることが懸念されるためである. 整理すると, 類型を定義したいデータセットに共通変数とそれ以外の変数が存在する時, それ以外の変数を良く分類しつつ, 共通変数から類型を予測できる余地も残したいということである.

従来は価値観の側を因子分析で, あるいは消費行動を潜在クラス分析で, のようにどちらか一方だけから類型を決めることが多かったが, 経験的には類型に寄与しない変数との相関はそこまで高くならない. また過去にはアンケートから導出されたライフスタイル因子と, ID-POS データに対する商品クラスを階層的に接続したモデルが提案されている[石垣 2011]が, シンプルに価値観と消費行動の両方をバランス良く説明する類型が求められないだろうか.

連絡先：西尾　義英, シナジーマーケティング(株), nishio.yoshihide@synergy101.jp

本稿では, 価値観と相関する消費行動の組を素性とした潜在クラス分析(pairs-plsa)が我々の求める類型の性質を満たすかを検証する. 価値観のみに対する潜在クラス分析(values-plsa)と消費行動のみに対する潜在クラス分析(facts-plsa)を比較対象とし, pairs-plsa が消費行動の説明力において values-plsa より優位で, 価値観による類型の予測精度において facts-plsa を上回ることを示す.

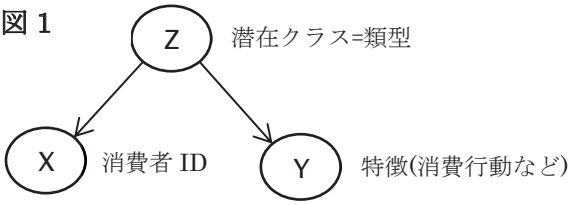
2. 手法

本稿で行う潜在クラス分析について説明する.

2.1 確率的潜在クラス分析(pLSA)

本稿では図 1 のように類型を潜在クラスと見なし EM アルゴリズムでパラメータ P(X|Z), P(Y|Z), P(Z) を最適化する pLSA[Hofmann 1999]を用いる. ある消費者 X が特徴 Y を持つ確率 P(X, Y)は類型 Z にのみ依存すると仮定するが, これは消費者の特徴を類型によって説明するという用途から自然な仮定といえる. なお本来は複数の潜在クラスの混合によって X と Y の関係を定式化する. これはいわゆるソフトクラスタリングに相当するが, 実用上は消費者が唯一つの類型に対応すると考えるのが扱いやすいため, P(Z|X)が最大となる潜在クラスを一つ当てはめることとする.

図 1



$$P(X, Y) = P(X|Z)P(Y|Z)P(Z)$$

2.2 クラス数の決定方法

pLSA ではクラス数 K を分析者が与えることとなっているので, 同じ条件でクラス数を決定できるように基準を定めておく. AIC や BIC などの情報量基準を用いる例が多い. 本稿では AIC を採用する. また EM アルゴリズムの初期値依存性を考慮し, 初期値をランダムに変えつつ K を探索する. 具体的には異なる K 毎に 10 回繰り返した時の最良解を採用することとする. なおデータセットによっては K を増やしても AIC が悪化する一方ということがありえる. 実際今回の実験では全てのデータセットにおい

て提案手法である pairs-plsa のみが AIC で K を選択することができたので，他の手法に対しては同じ K を採用することとする．

2.3 データ形式の説明

本稿での検証では同一の X すなわち消費者 ID の集合に対し Y を 3 パターン作成して pLSA を実行した結果を比較する．以下に各パターンのデータ形式を説明する．

(1) 価値観(values)

Societas が規定する 23 種の価値観変数を用いる．各変数は「好奇心が強い」といった属性を持っているか／持っていないかの二値の確率変数であるが，p>0.5 である属性を列挙したものを Y とする．

(2) 消費行動(facts)

購買データであれば購入した商品の名称やカテゴリを列挙したもの，アンケートの場合は複数選択式の設問で，選択された項目を Y とする．

(3) 価値観と行動の組(pairs)

アソシエーションルール：

rule1 ={価値観属性=**ある** ⇒ 消費行動=ある} または

rule0 ={価値観属性=**ない** ⇒ 消費行動=ある}

を Y とするが，データ量が多くなるので関連のある組み合わせのみに絞込む．facts の出現率が比較的高いデータセットにおいては，values と facts の相関係数の絶対値>0.1 を条件とする．facts の出現率が低いデータセットでは，rule1 と rule0 のリフトを計算し，lift(rule1) > 1.1 かつ lift(rule0) < 0.9 となるものだけを残す．

3. 実験

3.1 用いたデータ

実験には性質の異なる以下 4 種のデータセットを用いた．(1)と(2)はインターネット調査によるものである．N は人数，P は変数の数,K は潜在クラス数を表す．

表 1

#	description	N	P	K	pairs の抽出条件
(1)	食についての習慣や意識を問う調査	994	84	14	相関係数
(2)	地方紙とフリーペーパーの購読状況調査	3056	44	10	相関係数
(3)	観光スポットのスタンプ履歴	733	97	5	相関係数 ※しきい値を 0.03 に変更
(4)	スーパーの ID-POS データ	7651	2256	18	リフト

3.2 類型による消費行動の説明力を比較

ここで説明力とは 2.1 で説明した潜在クラス分析をクラスタリング手法と見たときの性能という意味で用いる．まず異なる類型の間で消費行動の特性が異なるということと，同じ類型の中では揃っているという観点で Calinski-Harabasz 指標[式(1)]による比較を行う．

$$\mathrm{CH} = \frac{N-K}{K-1} \times \frac{BGSS}{WGSS} \quad \text{式(1)}$$

式 1 の分子 BGSS: Between Group Sum of Square は類型間の分散を表し，分母の WGSS: Within Group Sum of Square は類型内の分散であるため，CH は大きいほど良い．結果を表 2 **に示す．**

表 2

dataset	Y の形式	BGSS	WGSS	CH
(1)	values	629	13,544	3.50
	pairs	1,477	12,696	8.77
	facts	124	14,049	0.97
(2)	values	171	20,271	2.86
	pairs	2,814	17,627	54.03
	facts	3,177	17,265	62.27
(3)	values	11	2,034	0.97
	pairs	212	1,833	21.04
	facts	235	1,810	23.61
(4)	values	5,624	1,307,416	1.93
	pairs	24,236	1,288,804	8.44
	facts	3,922	1,309,118	1.76

3.3 価値観による類型の予測精度を比較

潜在クラス分析の結果を正解ラベルとし，価値観だけの情報でどれほどの予測精度が得られるかを確認する．ランダムフォレストを分類器に用い，学習データとテストデータを半数に分けた時のラベルの一致率を表 3 に示す．

表 3

	Y の形式		
dataset	facts	pairs	values
(1)	0.301	0.600	0.476
(2)	0.303	0.450	0.610
(3)	0.294	0.338	0.730
(4)	0.251	0.311	0.596

4. まとめ

二つの実験 3.2 と 3.3 からはいずれも仮説を裏付ける結果が得られた．結果について考察し，今後の課題を述べる．

4.1 考察

表 2 では WGSS の差は小さいが BGSS が values<pairs なため CH において values<pairs となっている．消費行動において類型の特徴がはっきり出ていると解釈できる．(1)と(4)では BGSS と CH が facts<pairs となっているのは意外な結果であるが，K が最適ではなかったからかも知れない．

表 3 からは全てのデータセットで facts<pairs，つまり価値観による類型の予測精度が消費行動のみを対象とした潜在クラス分析を上回ることを確認できた．ただし(3)と(4)の予測精度はあまり高くない．

2.2 でも述べたが，今回取り上げたデータセットでは消費行動のみから潜在クラス分析を行う際に AIC 基準で K を選択できなかった．そのようなデータセットであっても価値観と組み合わせ

ることで K が決められたということは提案手法の副次的な利点と言える.

4.2 今後の課題

2.3(3)で述べた組み合わせの数を絞り込む手順において, 相関係数やリフトに対するしきい値の選び方には明確な根拠が無いので, 統計的な有意性あるいは類型の定量評価を最大化できるような基準を検討すべきであろう. またこの手法から得られる類型の有用性については今後実務における評価を重ねていきたい.

参考文献

[谷田 2013] 谷田 泰郎, 馬場 彩子, 西尾 義英: Societas と社会知ネットワーク, ヒューマンインタフェースシンポジウム 2013, 2013

[齋藤 2016] 齋藤 有紀子, 木虎 直樹, 谷田 泰郎: 複数データのマッピングによるシニア価値観分析の試み, 2016 年度人工知能学会全国大会 (第 30 回) , 2016

[石垣 2011] 石垣司, 竹中毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム: 実サービス支援のためのカテゴリマイニング技術, 人工知能学会論文誌, Vol.26, No.6, pp.670-681, 2011.

[Hofmann 1999] Hofmann, T.: Probabilistic latent semantic analysis, Proc. of Uncertainty in Artificial Intelligence, pp.289-296, 1999.

[1] 23 種の価値観変数: Societas の発表時から価値観変数の見直しを行った. 現在の変数定義は[齋藤 2016]表 2 を参照のこと.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

訪日外国人の価値観と再訪意向に関する考察

A Study of the Values and the Revisit Intention of Inbound Tourists in Japan

鎌田　裕美[*1]
Hiromi Kamata

[*1] 一橋大学
Hitotsubashi University #1

The purpose of this study is to analyze inbound tourists come to Japan. Segmenting the tourists by their values and comparing the satisfaction and the revisit intention among the segments. There are some useful studies of tourist segmentation by the motivation, however, the motivation items were defined by each study. The motivation might be changed by the time or the destinations. The values can be considered that are more stable than the motivation which will change by the situation. This study will attempt the values to inbound tourists' segmentation analysis. Additionally, satisfaction and revisit intention are compared by the segments. Based on these analysis, repeater segment's characteristics will be induced.

1. はじめに

2003 年の観光立国宣言を契機として、とくに東アジアを中心とした国々から、日本に多くの観光客が訪れている。最近では、何度も日本に訪れる、いわゆるリピーターが増えており、いかにリピーターを増やすかという点に関心が寄せられている。東アジア諸国は日本へのアクセスが容易かつ短時間であり、国内旅行のような感覚で訪れる人も多いだろう。今後、リピーターの創造を進めるためには、こうした実際のリピーターを調査分析し、なぜ何度も日本に来るのか、なぜある観光地を選ぶのか、を明らかにすることが重要である。

観光客の集客やリピーターの創造のためには、観光客の客層を把握してターゲットを定め、ターゲットである「顧客」の満足度に影響をおよぼす要因、再訪意向につながる要因を明らかにすることが必要である。つまり、STP(Segmentation, Targeting, Positioning)戦略などのマーケティング戦略が必要である。本研究では、セグメンテーションに焦点をあて、訪日外国人観光客の価値観を軸にセグメントを抽出し、セグメント間の満足度と再訪意向について分析する。

2. 先行研究

セグメンテーションの基準は、個人属性、地理的属性など無数にあるといわれている。観光客のセグメンテーションについては、年齢や性別、居住地などを用いることが多い。しかしながら、観光客が「なぜ、観光するのか」、「なぜ、その観光地に行くのか」という視点が重要であり、個人属性等によるセグメンテーションではこの部分に迫ることができない。

観光分野では、マーケティングのベネフィット・セグメンテーションに基づき、旅行動機を基準としたセグメンテーションの研究が進められている。研究対象は特定の観光客（スキー、キャンプ、フェスティバルなど）や外国人観光客など、さまざまである。旅行動機は、「push」と「pull」の 2 種類がある。「push」は、観光客の内発的な動機であり、「pull」は外的動機、いわゆる観光地の魅力である。これら旅行動機は，調査対象や時代によって異なることが一般的である。[Uysal et al. 2008] は、先行研究が取り上げた旅行動機をまとめたうえで、「調査対象に応じて、旅行動機を検討する必要がある」と指摘している。こうした旅行動機によるセグメンテーション研究は多くの蓄積があり、一定の研究成果を得ているものの、調査対象に応じて旅行動機項目が異なり、時期や対象によって結果が異なる。

満足度やリピート意向は、観光分野においても研究は盛んである。[Yoon and Uysal 2005] は、旅行動機、満足度、再訪意向の関係について分析した。再訪意向については、再訪意向や他者への推薦を含めて「観光地ロイヤリティ」ととらえている。[Bigne et al. 2009] は、観光地への再訪行動について、短期と長期では再訪行動に与える要因が異なることを明らかにした。短期または長期の再訪行動に及ぼす要因として、バラエティ・シーキング、知覚価値、観光地イメージ、満足度、スイッチング・コスト、過去のスイッチング行動の 6 つをあげて、それぞれが及ぼす影響を計測、その違いを示した。その結果、満足度が影響を与えるのは、長期の再訪行動であることを明らかにした。[Baloglu et al. 2008] は、セグメントごとに、旅行動機、総合満足度、再訪意向の関係を重回帰分析により分析し、満足度や再訪意向の違いを明らかにした。

本研究では、旅行動機よりは長期に安定して変化がないと思われる観光客の価値観に着目して、セグメンテーション分析を行う。また、セグメント間の満足度と再訪意向について分析し、検討を行う。

3. 調査分析

3.1 価値観

価値観にはさまざまな指標があるが、本研究では本研究では、[三菱総合研究所 2014]が設定した価値観の項目を用いて、調査分析を行う。調査項目は表 1 に示すとおりである。これらの項目は、世界価値観調査の項目に基づくものである。

3.2 データ

3.1 で述べた価値観について、アンケート調査によりデータを収集した。調査内容は、価値観要素の 22 項目、直近の訪日旅行の満足度、再訪意向である。再訪意向は、短期「次の休暇に訪れたい」、長期「いつかまた訪れたい」の 2 種類を設定した。いずれも、5 尺度（5:そう思う、4:ややそう思う、3:どちらでもない、2:あまりそう思わない、1:そう思わない）で回答を依頼した。

連絡先：鎌田裕美，一橋大学，〒186-8601 東京都国立市中 2-1，042-580-8871，hiromi.kamata@r.hit-u.ac.jp

アンケート調査はウェブ調査により実施した。調査の対象者は、訪日経験のある台湾人である。調査の結果、515 名から回答を得た。

3.3 分析と結果

(1) 価値観とセグメント

分析は、次の 2 段階から成る。まず、価値観を抽出するため、アンケート調査で得た 22 項目の回答結果を用いて因子分析を行う。次に、価値観に基づくセグメントに分類するため、因子分析の結果を用いてクラスター分析を行う。

分析結果は、表 1 のとおりである。因子分析では、アンケート調査のうち 22 項目の回答結果を主因子法、プロマックス回転により分析した。固有値が 1 以上かつ因子負荷量が 0.4 以上の因子のみとした結果、4 因子が抽出された。因子の名前は、各因子を構成する要素から名付けた。

表 1 因子分析の結果

健康・安全
気ままな生活をしたい
病気にかからず、健康でいたい
人生を楽しみたい
安全・安心なところに住みたい
自分で自分のやることを決めていきたい
家族との信頼関係やふれあいを大切にしたい
人生でいろいろな経験をしたい
他者とのつながり
周囲の人を助けたい、面倒をみたい
他人が必要としていることに対応したい
自然や地球環境を大事にしたい
これまでの慣習に従うことが大事だ
近所の人とのふれあいを大切にしたい
誰も見ていないときも、いつもルールに従うべきだ
人を立てる控えめな態度や精神性が大事だ
自分ひとりの幸せよりもみんなの幸せを考えたい
いつも他人に礼儀正しく接するべきだ
権力・誇示
責任者となり、他人を指導したい
人生の勝ち組になりたい
金持ちになり、高級品を持ちたい
実力を発揮し、人から賞賛されたい
アグレッシブ
新たなアイデアを考えたい
危険を冒したい、冒険したい

表 2 価値観とセグメンテーションの結果

	CL1	CL2	CL3	CL4		
人数	193	90	229	3		
%	37.5	17.5	44.5	0.6	F 値	有意
健康・安全	4.5	3.8	4.9	2.2	277.59	0.00
他者とのつながり	3.9	3.4	4.7	2.0	216.43	0.00
権力・誇示	3.6	3.5	4.5	2.1	188.52	0.00
アグレッシブ	3.5	3.1	4.5	1.6	144.17	0.00

※因子の数値は、各クラスターの回答の平均値である。

次に、因子分析の結果に基づきクラスター分析を用いて、セグメンテーションを行った（K-means 法）。クラスター数を 3～7 に設定してそれぞれ分析した結果、クラスターの割合、各クラスターの特徴から 4 を採用した。なお、クラスター4 は 3 名のみが該当するが、クラスター数を変更しても 3 名は 1 つのクラスターを形成した。

(2) 満足度と再訪意向

セグメントごとの満足度、再訪意向は、表 3 のとおりである。数値は、平均値である。

表 3 満足度、再訪意向の平均値

	CL1	CL2	CL3	CL4
満足度	4.5	4.2	4.6	3.7
再訪意向:短期	4.3	4.1	4.5	3.7
再訪意向:長期	4.4	4.2	4.6	3.3

観光の場合、満足度や再訪意向は全体的に高くなる傾向にある。もっとも高いのは CL3 であり、価値観に基づくセグメンテーションでも、高い傾向にある。CL4 は、価値観のいずれ項目にも「そう思わない」、「あまりそう思わない」と回答しているが、満足度や再訪意向も他のセグメントと比較して低い傾向にある。詳細は、学会当日に報告したい。

4. まとめと今後の課題

本研究では、セグメンテーションに焦点をあて、訪日外国人観光客の価値観を軸にセグメントを抽出し、セグメント間の満足度と再訪意向について検討した。価値観のいずれの項目にも同意する人ほど、満足度や再訪意向が高くなる傾向があった。

今後の課題は、価値観に関する先行研究をベースに、演繹的に分析を行うことである。分析を重ね、差が国や文化、社会背景によるものなのか、その要因を検討したい。

参考文献

[加納 2015] 加納史子 谷田泰郎: 世界価値観データベースに基づく世界消費者の把握、The 29th Annual Conference of the Japanese Society for Artificial Intelligence、人工知能学会、2015.

[三菱総合研究所 2014] 三菱総合研究所: 日本人はこんなタイプに分けられる！、mif 生活者市場予測システム、2014.

[Baloglu 2008] Baloglu, S., Pekcan, S. Chen and J. Santos: The Relationship between Destination Performance, Overall Satisfaction, and Behavioral Intention for Distinct Segments, Journal of Quality Assurance in Hospitality & Tourism, Vol.4, No.3-4, pp.149-165, 2008.

[Bigne 2009] Bigne, J. E., I. Sanchez and L. Andreu: The Role of Variety Seeking in Short and Long Run Revisit Intentions in Holiday Destinations, International Journal of Culture, Tourism and Hospitality Research, Vol.3, No.2, pp.103-115, 2009.

[Uysal 2008] Uysal, M., X. Li, and E. Turk: Push-Pull Dynamics in Travel Decisions, In Oh, H. and A. Pizam (ed.). The Handbook of Hospitality & Tourism Marketing, Elsevier, pp.412-438, 2008.

[Yoon 2005] Yoon, Y. and M. Uysal: An Examination of the Effects of Motivation and Satisfaction on Destination Loyalty: a Structural Model, Tourism Management, Vol.26, pp.45-56, 2005.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

インバウンド観光マーケティングモデル「グリーンAIアカデミー」

Inbound tourism Marketing Model 「Green AI Academy (GAIA) 」

猪澤也寸志
Izawa Yasushi

エコツーラボ合同会社
EcoTourLab

1. モデル開発の背景

2014年、COP12に国連生物多様性条約事務局が提出した報告書によると，「大気中の二酸化炭素（CO2）濃度が急増して起きる海洋酸性化による経済損失は徐々に増え，2100年までに年１兆ドル（約107兆円）を超える恐れがある」．水産資源や観光資源を供給し，約４億人の生活を支えているサンゴ礁が大打撃を受けるためである．また，報告書は「サンゴ礁以外の海岸などでの被害を加えれば損失額はさらに大きくなる」と指摘した．海洋酸性化は大気中のCO2が海水に溶け込むことが原因で起こるが，今のペースでCO2排出量が増えると．今世紀末にはさらにｐＨが0.3下がる可能性があり，主に炭酸カルシウムでできている貝殻やサンゴの骨格ができにくくなることが室内実験で数多く報告されているとした．

2016年、宮古島や石垣島においてサンゴ大白化現象が生じた．これも温暖化に伴う海水温上昇が影響している．この白化と酸性化に対する予知予防対策が国際的な課題となっている．

2018年，宮古島の年間観光客はここ3年で倍増し，90万人を突破した．その最大の増加要因は4割（35万人）を占めている外国人観光客である．今後，2019年春には那覇空港と同じ3,000m滑走路を有する下地島空港が開港し，2020年にメガクルーズ船埠頭の運用が開始されれば，5年後の2022年には150万人（うち外国人100万人）を突破するものと推測される．この外国人観光客のほとんどが中華系であり，サンゴ礁生態系への遊泳影響が懸念されている．筆者の経営するエコガイドカフェにも昨年は3,000名の中華系外国人がスノーケリングに参加している．この参加者にノータッチ遊泳マナーを徹底指導しているが，遊泳中いっさいサンゴ礁に接触しないECOスノーケリングには強い共感を得ている．結果的にグリーンリピーターも増加しつつある．くしくも本2018年は国際サンゴ礁年であり，十年に一度，世界が連携してサンゴ礁保全を推進する年である．ちなみに，筆者経営のエコツーラボ合同会社（屋号：エコガイドカフェ）は，去った１月末に環境大臣より，国際サンゴ礁年2018オフィシャルサポーター企業として任命を受け，当社のサンゴ礁保全活動を環境省が広報支援する協定書を締結した．

今後，人工知能リテラシーを高めることが老若男女を問わず，次世代を生き抜く力を育む教育課題となっている．しかし，その学習機会の整備は始まったばかりで，中でも自然生態系保全に資するグリーンAIに関しては，今のところ，その学習機会は未整備と察する．よって，自然環境と人工知能双方のリテラシーを高めることができる本提案が，時代の求めている教育機会の一助になれば幸いである．

2. モデル開発の目的

インバウンド観光マーケティングモデル「グリーンAIアカデミー（以降、GAIAと称する）」開発の目的は，①人工知能を活用して，サンゴ礁の健全度評価および白化酸性化等の予知予防を行い，その保全活動に資すること．②人工知能と自然環境リテラシー双方を高める実践的な学習機会を提供し，それを多言語提供することで外国人グリーンファンを形成し，その啓発効果で遊泳負荷等を抑制すること．③GAIAを垂直展開し，人工知能および生態系分野の国際知見を結集すること．④GAIAを水平展開し，他地域のサンゴ礁保全，観光振興，人材育成，他生態系の保全に資することである．

3. モデル開発の目標

GAIAの人工知能は，①サンゴ礁の生息状況（健全・白化・酸性化・回復など）を判別評価する「学習部」，②今後の白化や酸性化を予知予防する「予測部」，③予知と連動した予防ロボティクスなどを行う「対処部」，④各部の開発運用知見を抽出して、人工知能と環境リテラシ双方を高める学習プログラムを自動生成する「教育部」の４部で構成される．

GAIAの開発運用フィールドは，①宮古島に設置したサテライトラボ，②宮古島と橋で繋がった伊良部島に設置するサンゴ養殖場，③サンゴ礁に恵まれた宮古島の天然海域で構成される．

以下に，GAIA開発目標を列記する．

3.1. サンゴ礁の回復力評価（学習目標）

サンゴ礁を健全に維持するために何が必要なのか？その筆頭は，サンゴ礁の回復力（Coral Reef Resilience）を評価する仕組みである．対象海域の画像や動画データから，健全なのか，白化しているのか，回復しているのかなどを判別する推論エンジン開発が「学習部」の目標となる．

3.2. サンゴ白化等の予知予防（予測目標）

再学習で精度を高めることが期待できる前述の推論エンジンと，周辺海域のセンサデータ（水温，光子量，紫外線量，潮力，濁度など）を総合解析して，サンゴ白化等の予知予防を実現することが「予測部」の目標となる．

3.3. 予知連動の予防ロボティクス（対処目標）

天然海域でのロボティクスは困難だが，筆者が代表監事の伊良部漁協のサンゴ養殖場であれば，例えば，真夏の強烈な太陽光が白化を促進すると予知された場合，その強光を遮る遮光ロボティクスが「対処部」の目標となる．

3.4. リテラシプログラムソースの自動生成

人工知能と自然環境双方のリテラシを高めるプログラムソースとしては，①CaffeなどGAIA学習モデル、②学習＆正解データライブラリ，③適合課題や再現課題の抽出プロセス、④正解ラベル設定に必要な自然科学知見，⑤センサネットワーク構築プロセス，⑥野外における海中撮影などデータ収集ノウハウ，⑦サンゴ白化や今後深刻になる酸性化の最新知見などを抽出して，自動的にプログラムを生成することが「教育部」の目標となる．

4. GAIAトライアルフェーズに関して

GAIAは，学習部，予測部，対処部，教育部で構成される統合システムであるが，本書では，今後予定している「学習部」に関してのトライアルフェーズを述べる．

4.1. 業務フローの検討

人工知能が行う部分，人が行う部分，人工知能の役割，業務範囲の規定などを想定し，目標精度に応じて実務フローとの相関から以下のような３パターンを検討する．
①人工知能が推奨する結果を参考に，実行に関する最終意思決定を人が行う．②人工知能が自動的に実行する．人工知能が判断の自信度を判定し，自信度が低いときには人に意思決定を委ねる．③人工知能が出力したルールを人が確認し，確認されたルールに基づいて自動的に実行する．

なお，「教育部」の目標であるで学習プログラムの自動生成だが，当面は手作業にて，学習＆正解データライブラリなどを活用し，外国人観光客でも楽しめる体験学習プログラムを作成し，グリーンファン形成のためのインバウンド観光マーケティングとして本2018秋より開始する．

4.2. データ収集

保有データとして，2016年の大白化以降，現在までの野外白化観察データ（2Tb/100日）がある．他にも，日本初のサンゴ礁公害事件となった宮古島市海中公園工事被害状況の海中撮影データ（0.5tb/20日），オニヒトデ大発生時の食害観察データ（0.5tb/100日間）などがある．また，過去数年間の健全なサンゴ野外撮影データも少なからず保存している．

随時観測可能な天然海域から今後とも収集するデータとして，健全サンゴデータ疲弊サンゴデータ（白化，工事，食害），回復サンゴデータなどがある．

通年定点観測が可能な養殖場（今春設置予定）から今後収集するデータとして，健全サンゴデータ疲弊サンゴデータ（白化，工事，食害），回復サンゴデータなどに加え，養殖場に設置を予定している各種センサデータ（水温、光子量、紫外線量、潮力、濁度など）などがある．

4.3. 学習モデル設計

学習アルゴリズムは，必要に応じて，教師なしディープラーニング（クラスタリング，主成分分析等）を行い，教師ありディープラーニングの目的変数や説明変数の加工に活用することも検討する．

学習モデル候補としては，画像認識に強いCaffe等を主たるフレームワークとして検討する．生態学的な最新知見と専門家の暗黙知を反映させた正解ラベル等でサンゴ礁回復評価エンジンの開発を目指す．

開発した学習済みモデルを「Caffe Model Zoo」などを通じて段階的に提供する．原則，MacOSで構築し，当社研究員や観光客らにとっても身近な学習利用環境を整備する．なお，研究開発環境は必要に応じてLinuxOSを活用する．

4.4. データ加工（前処理）

過去データに関しては，疲弊サンゴ観察を継続実施して来た筆者が行う．具体的には，疲弊要因の判定（白化、工事，食害など），白化状態の判定（進行/回復），健全回復度の判定などであるが，これら経験知などの暗黙知を正解データ分類基準などに置換する作業は筆者が行う．

今後の正解データ整理に関しては，筆者が主体となって当社研究員も担当するが，アノテータ要員としては当社研究員以外にも学生インターンや観光客も参加貢献できるような教育体制を構築するものとする．

ちなみに，今現在，サンゴ白化に関連する数万点の撮影写真（3000pix程度）があり，これを健全に増量したり、４K動画から切り出したりすることで数十万点規模の学習データセットが確保できる．今後は，エコガイドカフェスタッフによる定点観察データ（月間１万点規模）に加え，これらデータ収集ノウハウの一部を観光客に移植し，その観光客によるデータ収集も併せて実施する．

データ加工としては，目的変数の加工（グループ化・ラベル化），説明変数の加工（自己回帰変数の作成，平滑化，移動平均法，ローパスフィルタ，ハイパスフィルタ，ラベルデータのグループ化，複数説明変数の合成・次元圧縮，異常値処理，データー数加工（リサンプリング，アンダーサンプリング，オーバーサンプリング），画像データ増幅・データオーグメンテーション（反転/回転，明度彩度の変更，拡大/縮小，画像合成，部分画像抽出）等を考えている．

4.5. 結果の評価

本書では，評価指標としての「精度」と「解釈性」に関して述べる．

精度における適合率は，白化が進行しているのか，回復しているのかの判定である．この適合率を上げるためには，以下に資する正解ラベル整理が必要となる．①海域からサンゴ群体部分の抽出，②サンゴ群体の形状や数量，③サンゴ群体の属判定、④褐虫藻密度，⑤骨格形成パターン，⑥骨格侵食パターン、⑦骨密度や骨格量など．

精度における復元率は，その対象範囲，例えば，サンゴ群集（m単位），群体（cm単位），個体（mm単位）などで精度は左右されると考えられるが，予測部が予防予知の判断材料とする上で，どの感度が適正なのかなども合わせて検証していかなければならないものと考える．

今後、白化進行と回復を区別した上で対象海域の健全度（回復力）を評価をする上でも，適合率と再現率の調和平均を取ったF値（総合的な精度指標）向上が求めらると考えている．なお，F値を向上させるため予測先時間などの分析内容定義を見直すことも重要と考えているが，これらは試行錯誤の上，走りながら調整したいとも考えている．

解釈性としては，ディープラーニングのモデルはブラックボックスであるため，さまざまな説明変数に対する出力を見比べてみて理解せざるを得ないが，白化や酸性化のメカニズム，または，その回復メカニズムなど，サンゴ礁のレジリエンスを解明する仮説検証のツールとして有効であるものと考えている．例えば，白化プロセスやその回復プロセスの骨生成メカニズムの鍵を，説明変数を加工しながら出力し，模索することも有効と考えている．

5. インバウンド観光マーケティングに関して

5.1. グリーンファンを育むGAIA

筆者の目標「観光客が来れば来るほど美しくなる宮古島」を実現するインバウンド観光マーケティングは３年前より既に始動している．繰り返しになるが，昨年3,000名の中華系外国人が筆者主催の体験学習に参加したが，遊泳中いっさいサンゴ礁に接触しないECOスノーケラへとほぼ全員を育むことができた．この結果，複数リピータ（グリーンファン）も着実に増加傾向にある．

このような外国人観光客によるサンゴ礁保全をムーブメントとなすため，GAIAを観光資源として活用する予定である．観光客が楽しみながら「人工知能と自然環境双方のリテラシ」を育める体験学習プログラムを併せて開発提供する．

今秋から新教育旅行コンテンツとして，小中学生の体験学習向け，修学旅行向け，大学ゼミ向け，企業研修向け，インターン研修者向け，来春からは，研究者向けなどに，それぞれ多言語（日本語・中国語・英語）のプログラムを開発する予定である．

5.2. モデルGAIAの水平展開

水平展開としては，湿原や森林など多生態系への水平展開も視野に入れている．情報管理方法としては、改ざんが困難な分散台帳としてブロックチェーン化し，共有する方法なども検討する．

5.3. モデルGAIAのファウンディング

将来的な開発運用の資金源としては，少額寄付（観光客などの寄付）に対応するため「エコトークン（GAIAトークン）」を発行することも検討する．このトークン価値は，GAIAが保全した生態系資産の価値として，前述の分散台帳に保全の足跡を残す．多地域多生態系の保全貢献が進めば，このGAIAトークンの価値は小さくはないものとなりうる．今後，海洋酸性化の緊急対策などに，地球規模でGAIAの垂直水平展開が求められた場合，GAIAトークンをICOして対処することも視野に入れている．

6. 2022年ビジョン（５年計画）

6.1. 一年目・・・2018年

学習部トライアルフェーズのPDCAサイクルを回し，試行錯誤を繰り返しながら開発実用性への目処を立てる．GAIA体験学習プログラム（多言語）を開発し試験運用を開始する．

6.2. 二年目・・・2019年

通年定点観測した試験養殖フィールドのサンゴ群体データにて学習させる．生息確認された正解ラベルや海中のセンサネットワークから収集したメタデータも複合的に学習させる．サンゴ礁回復力評価において，人為判定の「補助」として実用的な達成目標精度を目指す．研究者向けのGAIA体験学習プログラム（多言語）を開発し，試験運用を開始する．

6.3. 三年目・・・2020年

前年と同様だが、人為判定の「鍵」として実用的な達成目標精度を目指す．垂直展開，水平展開の検討を行う．

6.4. 四年目・・・2021年

白化予知に伴う白化予防策として，センサネットワークと連動した遮光などの予防ロボティクスの実地検証を行う．また，水中３Ｄスキャナーなどの最先端ツールを活用して，正解用データなどに三次元データの導入を行う．

なお，２年目で人為判定の「鍵」となりうる実用的な達成目標精度に，二次元データでは至らないと判断した場合は，二年目に３次元データ収集を前倒すものとする．試験的に水平展開を行う．

6.5. 五年目・・・2022年

白化に加え酸性化の予知予防を適合＆再現課題として加える．白化回復の指標は骨格変化になりうることから，３次元スキャナー等で収集した骨格変化データで学習を進めたグリーンAIは，酸性化に伴う骨格変化（骨生成阻害や骨粗鬆）に対しても，適切な学習データや正解ラベルを付与すれば，移転可能と考えられる．なお，天然海域および野外養殖フィールドにおいて酸性化の影響らしきものを確認できた時点でこの酸性化課題も前倒しにするものとする．

7. 筆者（エコガイドカフェ）の足跡

1985年　宮古島創業
2008年　国際サンゴ礁年2008登録イベント活動
　　ECOスノーケラー養成教室
2009年　カーボンオフセット認証取得
2010年　環境省主催エコツーリズム大賞（特別賞）
2010年　沖縄県サンゴ礁保全推進協議会理事
　　（現在退任）
2010年　伊良部漁業協同組合　代表監事
2014年　日本政府観光局特別賞
　　沖縄MICEコンテンツコンテスト
2011年　沖縄県オニヒトデ対策事業
2012年　沖縄県オニヒトデ対策事業 ・ 35,000個体
2013年　沖縄県オニヒトデ対策事業 ・ 22,000個体
2014年　沖縄県オニヒトデ対策事業 ・　1,300個体
2017年　日本サンゴ礁学会口頭発表 ・ 過去10年連続
2017年　国際サンゴ礁年2018登録イベント活動
　　ECOスノーケラー養成教室：実績５万名突破
2018年　国際サンゴ礁年オフィシャルサポーター企業

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Stability and similarity of clusters under reduced response data

Marisciel Litong-Palima[*1], Kristoffer Jon Albers[*2] and Fumiko Kano Glückstad[*1]

[*1] Copenhagen Business School, DK-2000 Frederiksberg, Denmark [*2] Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

This study presents a validated recommendation on how to shorten the surveys while still obtaining segmentation-based insights that are consistent with the analysis of the full length version of the same survey. We use latent class analysis to cluster respondents based on their responses to a survey on human values. We first define the clustering performance based on stability and similarity measures for ten random subsamples relative to the complete set. We find foremost that the use of true binary scale can potentially reduce survey completion time while still providing sufficient response information to derive clusters with characteristics that resemble those obtained with the full Likert scale version. The main motivation for this study is to provide a baseline performance of a standard clustering tool for cases when it is preferable or necessary to limit survey scope, in consideration of issues like respondent fatigue or resource constraints.

1. Introduction and motivation

Consumer segmentation research based on survey data has benefited from the availability of more efficient and economical means to collect survey responses, like the proliferation of survey hosting platforms, like Qualtrics [https://www.qualtrics.com/], SurveyMonkey [https://www.surveymonkey.com/], Google Surveys [https://surveys.google.com/] and SurveyXact [https://www.surveyxact.dk/] as well as crowdsourcing online platforms where the surveys can be quickly deployed, like the Amazon Mechanical Turk [https://www.mturk.com/] and Prolific Academic [https://www.prolific.ac/]. As the use of these tools are likely to increase, it also becomes likely that more and more surveys will be deployed which poses a challenge for recruiting respondents and for maintaining the quality of the responses. Hence, there is a timely need to optimize the design of surveys to counterbalance these effects.

This study aims to address an aspect of survey design related to the presentation of questions in order to minimize the survey length. The intention is to maintain consistency and reliability in the gathered responses in order to provide better input to computational tools for segmentation. While having more questions in a survey intuitively provides more data for extracting insight, this study considers how the questions can be presented in order to mitigate issues like respondent fatigue, wherein the response rates and quality decreases with survey length [Burchell 1992, Rolstad 2011]. The goal is to have more data of higher quality to support the computational tools in segmentation because these algorithms perform better when data is substantial and of good quality.

In this article, we describe a procedure for comparing the relative baseline performance of a standard clustering method when applied to survey data that has been reduced or limited as compared to the performance when clustering is applied to the original and full data. There are two interpretations of data reduction considered in this study: (1) having fewer question items, i.e. 21 to 10 questions; and (2) having fewer response levels, i.e. Likert to binary. The basic motivation for this study is to provide a recommendation for designing surveys that can be completed by respondents in less time and still provide data for clustering solutions of expected stability and similarity to the full data solution, and more importantly with comparable mean responses or item profile. This work follows closely the motivations and recommendations discussed in [Dolnicar 2007, Dolnicar 2011].

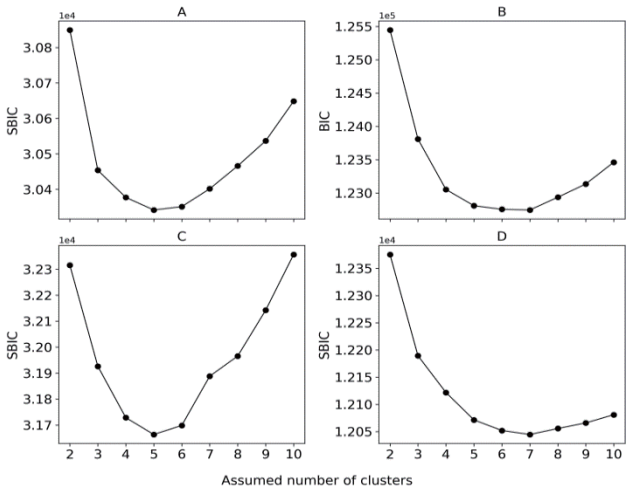


Figure 1 Based on the best, i.e. least, Bayesian information criterion (BIC) or its sample-adjusted version (SBIC), the LCA model with five clusters is a reasonable model to assume for all the cases of data sets (A) WVS Likert 10Q, (B) ESS Likert 21Q (C) PRF Likert 21Q and (D) PRF True Binary 21Q.

2. Human value survey data

As listed in Table 1, we use three samples of human value survey data: Sets 1 and 2 are secondary sources from the World Values Survey [http://www.worldvaluessurvey.org/] and the European Social Survey [http://www.europeansocialsurvey.org] respectively while Set 3 is a primary source that was disseminated through SurveyXact and Prolific Academic

Contact: Marisciel Litong-Palima, Copenhagen Business School, DK-2000, Frederiksberg, Denmark, Phone: +45 3815 3238, Email: mpa.msc@cbs.dk .

platforms. All respondents are from the United Kingdom. The questionnaires used in the three sets have 10 human value questions in common. The ESS and the PRF have 11 human value questions in addition.

Table 1 Description of data sets used. N is the number of questions in the survey set.

Set	Source	N	Sample size	Remarks
1	WVS	10	989	Secondary/Interview-based
2	ESS	21	2005	Secondary/Interview-based
3	PRF	21	516	Primary/Online

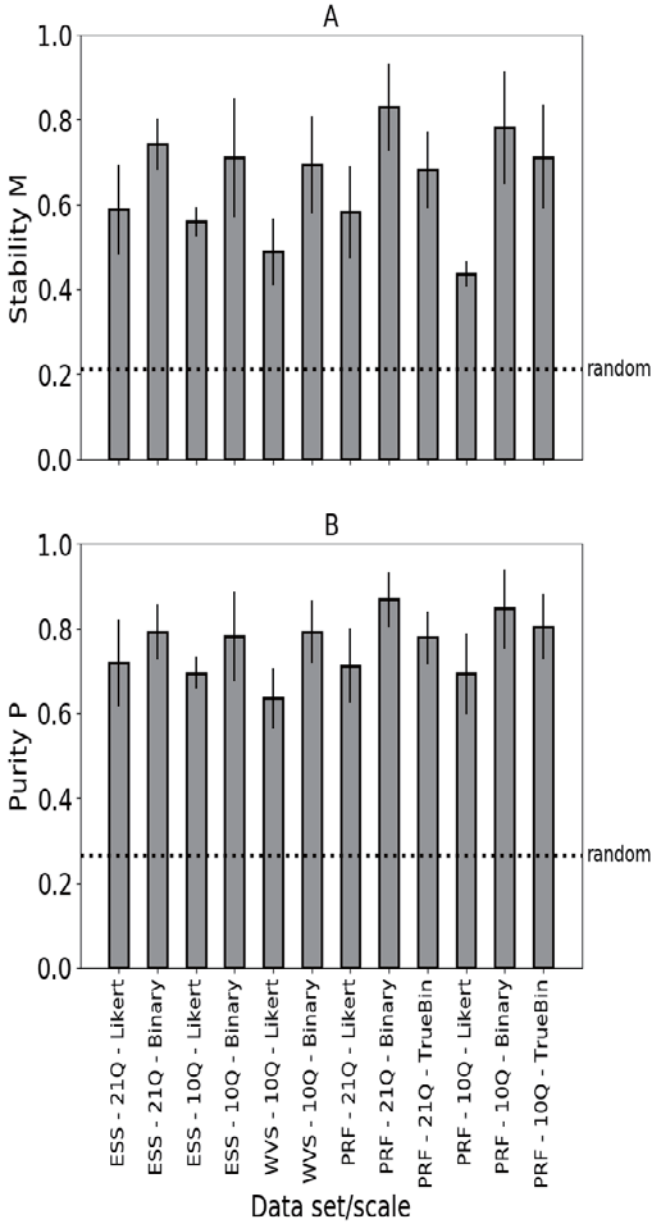


Figure 2 (A) Stability M and (B) similarity/purity P for the ESS, WVS and PRF data cases considered. The labels reflect the data case type. For example, PRF-10Q-TrueBin means the data is from the online survey Set 3 with 10 human value using the multiple-choice/tick box, i.e. true binary response scale.

In this study, we define the case of full data to correspond to a data set with 21 questions on a 6-scale Likert response scale, which consists of two negative ("Not at all like me", "Not like me") and four positive ("A little like me", "Somewhat like me", "Like me", "Very much like me") response levels, e.g. ESS. The WVS data set has at most ten value questions in common with the ESS data set and is thus an example of a data set with reduced item number. The PRF data set has the same 21 questions as the ESS but for some questions, response is limited to either a positive or a negative, i.e. reduced response levels.

Note that the clustering solutions for each of the different data sets cannot be directly compared. In order to make a meaningful comparison across the different data sets, we use a relative measure of performance that compares the clustering solution obtained with ten random subsamples to that obtained with the complete sample. To quantify performance, we use the relative measures of stability M and similarity measure P, also known as purity score.

$\mathbf{M} = \ll \delta_{\mathbf{Trc}, \mathbf{T}^{(\mu)}_{\mathbf{rc}}} \gg_{\mathbf{m}}$, where the values of $\mathbf{T_{rc}}$ and $\mathbf{T^{(\mu)}_{rc}}$ are the connectivity matrices for the complete and the resampled data, and $\ll\gg_{\mathbf{m}}$ is a two-step averaging over the neighbors in the complete set and then over the random samples, as introduced in [Levine 2001].

$\mathbf{P} = \Sigma_{\mathbf{r}}\, \mathbf{p_r}\, (\mathbf{max_c}\, (\mathbf{p_{rc}}\, /\mathbf{p_r}))$, where the values of $\mathbf{p_r}$ and $\mathbf{p_c}$ are probabilities of membership to clusters $\mathbf{r}$ and $\mathbf{c}$ obtained using the reduced and the complete data sets respectively and $\mathbf{p_{rc}}$ is the probability for membership to both.

We then compare **M** and **P** across the different cases of reduced data. Cases are labelled to indicate the following:

- 21Q – complete 21 questions on human values
- 10Q – reduced 10 questions on human values
- Likert – complete scale with six values consisting of 4 positive and 2 negative levels
- Binary – reduced scale with two levels converted from the negative and positive levels of the Likert scale
- True Bin – reduced scale, asked directly to the respondents as a multiple choice-tick box type (i.e. not converted)

3. Method to estimate relative baseline of clustering performance and results

We choose latent class analysis (LCA) [Lazarsfeld 1968] as implemented in the MPlus software [https://www.statmodel.com/] as our standard clustering method. Similar clustering approaches are used for exploring questionnaire data and examining human behavior and value priorities [Szakolczai1998, Moors2009, Magun2015]. LCA is implemented in MPlus based on a mixture model of multivariate distributions for categorical variables. In doing clustering analysis, one assumes that the respondents belong to significantly different groups but that this grouping is unknown or cannot be observed directly and can only be inferred from analyzing the data. LCA is a method to unmix the respondents into these

distinct but unobserved clusters based on the over-all structure of the entire response data.

We choose the Bayes information criterion (BIC) and its sample-adjusted version SBIC to decide how many latent clusters there are in the data set. The BIC has been shown to be a better and consistent criterion in inferring the unknown number of clusters for a given sample, even when the item profile or structure in the data is complex and even when the cluster sizes are significantly different [Nylund 2007]. For this study, we found that the related criterion SBIC gives a clearer indication of the optimal values for most of the data set cases. The use of either BIC or SBIC is enough for the purposes of this study. Results based on the BIC and SBIC, as shown in Figure 1 indicate minimum values around five to six clusters across all the different data sets considered. For this study, we choose models with five clusters.

Figure 2 shows the stability and similarity measures across the full and reduced data cases. All solutions have stability and similarity measures that are at least twice as much as random solutions, whose corresponding measures are **M**~0.21 and **P**~0.27 on average.

The clustering solutions obtained using binarized or true binary scales consistently perform better when compared against their Likert-based counterparts. This suggests that converting the scale to binary or using a binary scale designed into the survey can give clustering solutions that are more robust with **M**~0.7 and highly similar with **P**~0.8.

When it comes to the reduction of questions, performance with just ten questions are in general less stable and with less similarity, but nevertheless stability has value **M**~0.5. Note that stability is least for the PRF-10Q-Likert case, even though the response scale is Likert, and hence not reduced.

The PRF-10Q-True Bin has performance measures that are nearly the same as the PRF-21Q-True Bin cases for both **M** and **P**. This is a useful result as it supports the question style where respondents are asked just once to account which among 21 human values they relate to, instead of being asked a list of 21 questions. This supports a shorter survey design.

We recognize that the process of converting the scale to binary or the use of a true binary scale, by definition, reduces the variations in the possible responses and reduces the information available for the LCA algorithm. Nevertheless, what these stability and similarity results suggest is that despite the reduction in the information due to a limited scale, the clustering solutions obtained are still descriptive and can still reveal the clusters that resemble those obtained in the full length or scale version. To illustrate this claim, Figure 3 shows the matching in the cluster profile for the PRF-21Q-Likert case compared with the PRF-21Q-True Bin case. Pairing the clusters from the two solutions is easily done using a visual check. The purity score computed directly between the two solutions gives **P**~0.46 and supports well the correspondence of the structure of the item profiles for both solutions. The characteristics of the clusters suggested by the solution from PRF-21Q-Likert case are well-retained in the solutions from the corresponding true binary case.

4. Summary and Conclusion

The tendencies with regards to the clustering performance of latent class analysis on random subsamples relative to the complete sample and across the cases of reduced response data are summarized as follows:

- The average **M** values across different reduced data cases is 0.55 for the Likert cases and 0.75 for the binary cases. **M** is 0.21 for a purely random solution.
- The average **P** values across different reduced data cases is 0.65 for the Likert cases and 0.81 for the binary cases. **P** is 0.27 for a purely random solution.
- The reduction of the number of questions have less impact on the decrease in **M** and **P**, although can account for more variability.
- Even under a reduction in the response levels, cluster characteristics are well maintained. In the case of the PRF-21Q-Likert and PRF-21Q-True Bin, the radars plots can even be easily matched and the purity score is at 0.46.

This study recommends to reduce the amount of response data from human values survey designs primarily by using true binary response scales, particularly when the goal is to cluster the respondents. This can be particularly useful in studies where the value questions are included only as part of a more comprehensive questionnaire, and hence reducing parts of the supporting response data may reduce the overall cost of the study and limit respondent fatigue.

Though the present study is based on LCA on human value questions, the presented approach can easily be extended to other domains and clustering models as well.

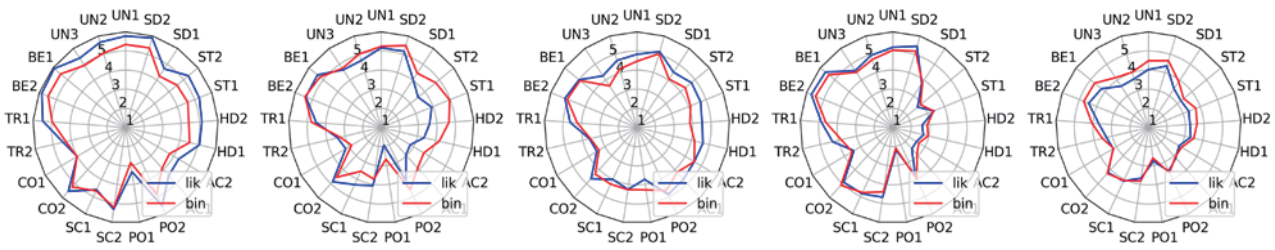


Figure 3 Comparison of the mean responses for each of the five cluster for the cases (A) PRF-21Q-Likert and (B) PRF-21Q-True Bin.

Acknowledgement

This work has been conducted as part of “UMAMI: Understanding Mindsets Across Markets, Internationally” No. 61579-00001A funded by Innovation Fund Denmark.

References

[Burchell 1992] Burchell, Brendan, and Catherine Marsh. "The effect of questionnaire length on survey response." *Quality and quantity* 26.3: 233-244, Springer, 1992.

[Dolnicar 2007] Dolnicar, Sara, and Bettina Grün. "How constrained a response: A comparison of binary, ordinal and metric answer formats." Journal of Retailing and Consumer Services 14.2: 108-122, Elsevier, 2007.

[Dolnicar 2011] Dolnicar, Sara, Bettina Grün, and Friedrich Leisch. "Quick, simple and reliable: Forced binary survey questions." International Journal of Market Research 53.2: 231, Elsevier, 2011.

[Lazarsfeld 1968] Lazarsfeld, P.F. & Henry. N.W. “Latent structure analysis.” New York: Houghton Mifflin, 1968.

[Levine 2001] Levine, Erel, and Eytan Domany. "Resampling method for unsupervised estimation of cluster validity." Neural computation 13.11: 2573-2593, MIT Press, 2001.

[Magun 2015] Magun, Vladimir, Maksim Rudnev, and Peter Schmidt. "Within-and between-country value diversity in Europe: A typological approach." European Sociological Review 32.2 (2015): 189-202.

[Moors 2007] Moors, Guy, and Jeroen Vermunt. "Heterogeneity in post-materialist value priorities. Evidence from a latent class discrete choice approach." European Sociological Review 23.5 (2007): 631-648.

[Nylund 2007] Nylund, Karen L., Tihomir Asparouhov, and Bengt O. Muthén. "Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study." Structural equation modeling 14.4 (2007): 535-569.

[Rolstad 2011] Rolstad, Sindre, John Adler, and Anna Rydén. "Response burden and questionnaire length: is shorter better? A review and meta-analysis." *Value in Health* 14.8: 1101-1108, Elsevier, 2011.

[Szakolczai1998] Szakolczai, Arpád, and László Füstös. "Value Systems in Axial Moments A Comparative Analysis of 24 European Countries." European Sociological Review 14.3 (1998): 211-229.

Oral presentation | General Session | [General Session] 13. AI Application

[2J1] [General Session] 13. AI Application

座長:星野 崇宏（慶応大）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room J (2F Royal Garden B)

[2J1-01] Effect of Increasing of Horizontal ShareHolding by Index Funds to Competitions and Market Prices -- Investigation by Agent-Based Model --

○Takanobu Mizuta[1] （1. SPARX Asset Management Co.,Ltd.）

9:00 AM - 9:20 AM

[2J1-02] Indexation of sentiment of Bank of Japan using text of Financial report and macroeconomic index

○Kyoto Yono[1], Kiyoshi Izumi[1], Hiroki Sakaji[1] （1. Graduate School of Engineering,The University of Tokyo）

9:20 AM - 9:40 AM

[2J1-03] How Does the Rule for Investment Diversification Affect Market Fluidity, Efficiency, and Investor's Performance?

○shunya maruyama[1], takanobu mizuta[2], isao yagi[1] （1. Kanagawa Institute of Technology, 2. SPARX Asset Management CO.,Ltd.）

9:40 AM - 10:00 AM

[2J1-04] Analyzing Mergers and Acquisitions (M&A) in Japan Using AI methods

○Bohua Shao[1], Kimitaka Asatani[1], Ichiro Sakata[1] （1. School of Engineering, the University of Tokyo）

10:00 AM - 10:20 AM

[2J1-05] Consideration on substitution by artificial intelligence and adaptation of organization structure

○Shunei NORIKUMO[1] （1. Osaka University of Commerce）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

水平株式保有するパッシブファンドの増加が企業間競争と市場価格へ与える影響–人工市場によるシミュレーション分析–

Effect of Increasing of Horizontal ShareHolding by Index Funds to Competitions and Market Prices – Investigation by Agent-Based Model –

水田孝信 *1
Takanobu Mizuta

*1 スパークス・アセット・マネジメント株式会社
SPARX Asset Management Co., Ltd.

Recently, some empirical studies argue that “horizontal shareholding” (or “common ownership”) lessen competition among companies and prevent growth of an industry. Especially, horizontal shareholding by index funds occupy more weights of lists of shareholders and became a subject of greater debate. In this study, I built an artificial market mode, a kind of agent based model, and investigated an effect of increasing of horizontal shareholding by index funds to competitions and market prices. Our result shows that even when holding ratio of index funds is not so much they lessen competition. Moreover, when value of a company beating competition grow, its market price grow more than the company value (overshoot) and become over valued, then shareholders who encourage competition decrease and the company lose competition power. On the other hand, when value of a company losing competition drop, its market price fall deeper than the company value (overshoot) and become under valued, then shareholders who encourage competition increase and the company gains competition power. My simulation result indicated such mechanism balances competition powers among corporates. Growing index funds may possibly weaken the balancing mechanize.

1. はじめに

近年，投資ファンドがある業界のすべての企業の大株主となる “水平株式保有”(horizontal shareholding)(または “共同保有”(common ownership) ともよばれる) が，公正な企業間の競争を阻害し，産業の発展を妨げているという主張が増えてきた [Azar 14, Elhauge 16, Fichtner 17]*1．通常投資家は，保有している企業が競争に勝ち企業価値が上昇することが自身の利益につながるため，企業の経営者に競争を促す．一方，業界のすべての株式を保有している投資家は保有する企業間が競争することによって，たとえある企業が競争に勝っても競争に負けた方の企業も必ず保有しており，その企業の価値下落によって損失もこうむるので競争を促す動機がなく，むしろ商品の販売価格を維持する方が利益になることすらある．

水平株式保有はさまざまな投資戦略のファンドで起こりえるが，日経平均株価などの指数（インデックス）と同じ収益を得られるようにインデックスを構成する銘柄と同じ銘柄を保有する “パッシブファンド” はほとんどの場合水平株式保有を行ううえ，近年急速に投資資金が増えているため，特にパッシブファンドによる水平株式保有が大きな割合となっており，大きな議論となっている [Fichtner 17].

米国においてはパッシブファンドの運用会社は上位 3 社による寡占が進んでいるうえ，全上場企業の 4 割以上の企業の筆頭株主が実質的にパッシブファンドの運用会社で，パッシブファンドが全体の約 15%を保有し，多くの企業の上位株主が重複しているという状況になっている [Fichtner 17]．このため [Azar 14] は，米国航空業界では上位株主の多くが重複しており，この水平株式保有による企業間競争の阻害の効果で航空運賃が 3%から 7% ほど高くなっていると見積もった．[Fichtner 17] は議決権行使の状況を分析し，パッシブファンドは株主総会のような公開の場ではなく，経営者との非公式ミーティングの場で経営者へ圧力をかけていると主張した．

[Elhauge 16] は，このような状況はすでに反競争法 (日本でいう独占禁止法) に違反している状況である恐れがあると述べ，[Piketty 13] が主張する所得格差の拡大はパッシブファンドが企業間競争を阻害していることが原因であって *2，これ以上の所得格差の拡大を食い止めるためには，株式に投資するファンドに対して 1 業界につき 1 企業のみに保有を制限すべきと主張した．

しかしながら，水平株式保有が企業の経営戦略や市場価格にどのように影響を与えるかは非常に複雑なメカニズムが存在すると考えられるため，これらの実証分析を中心とした先行研究では結論は全くでていない *3．特に図 1 に示すように，株式の売買が株主構成に変化を与え，それが企業の経営戦略および企業が市場価格に関わらず本源的にもつ価値 (ファンダメンタル価格) を変化させ，それが株価を変化させ，それが株式の売買の判断を変えるという，ポジティブ・フィードバック過程が内在していることが分析を難しくしていると考えられる．実証研究のみではこのような，ミクロ・マクロ相互作用を含むメカニズムを分析することは困難である．また，取引参加者に占めるパッシブファンドの割合が現在ほど多かったことは過去ないため，これ以上パッシブファンドが増えた場合の議論を実証研究だけで行うのは困難である．

このような実際の金融市場におけるミクロ・マクロ相互作用を含むメカニズムや，まだおきていない状況の変化の純粋な影響を議論するのにすぐれた手法として，コンピュータ上で仮想的にその状況を作り出し検証する，エージェントベースドモデルの一種である人工市場モデルを用いたシミュレーションがある *4．これまでの伝統的な経済学で使われてきた手法には

連絡先: 水田 孝信，mizutata@gmail.com
http://www.mizutatakanobu.com/
当日の発表スライドは以下にあります
http://www.mizutatakanobu.com/201806.pdf

*1 これらの論文のレビューもある [水田 18].

*2 [Piketty 13] が主張した所得格差拡大の原因はこれとは異なる．

*3 実際，反論 [O’Brien 17, Rock 17] も多く，それらの反論に対する反論 [Elhauge 17] もある．

*4 優れたレビューとして，[LeBaron 06, Chen 12, 水田 14, Mizuta

ない強みがあるとして，Nature と Science に人工市場モデルに期待を寄せる論考が掲載されている [Farmer 09, Battiston 16]．そして人工市場モデルを用いたシミュレーション研究は，現実の金融市場の規制・制度変更の議論に貢献 *5 したり，バブルや金融危機の発生メカニズムの解明に貢献したりした．

人工市場モデルを用いてパッシブファンドを議論した研究もいくつかある．[水田 17a, Mizuta 17b] は，パッシブファンドとは逆に値上がりが期待できる銘柄を選別しそれらに投資する “アクティブファンド” は，売買の頻度が低かったとしても，まれに起きる市場が不安定になったときに比較的多く売買することにより市場の安定化・効率化に寄与していることを示し，パッシブファンドの増加は市場の不安定化・非効率化につながる恐れを指摘した．また，[高橋 11, Braun-Munzinger 16] は人工市場モデルを用いてパッシブファンドが市場価格へ与える影響を議論した．しかしながら，パッシブファンドによる水平株式保有が市場価格に与える影響を人工市場モデルを用いて議論した研究はない．

そこで本研究では，[水田 17a, Mizuta 17b] が構築した人工市場モデルを 2 銘柄に拡張し，株式水平保有が経営戦略を変更させ企業間競争を阻害するモデルを加え，パッシブファンドの増加が企業間競争と市場価格へ与える影響を分析した．

2. 人工市場モデル

本研究では，ファンダメンタル価格にもとづいて少ない機会にのみ売買を行う，現実市場で重要な投資家のモデル化に成功した [水田 17a, Mizuta 17b] の人工市場モデルを 2 銘柄に拡張したモデルを用いる．

2.1 エージェント

全エージェント数を N とする．初め，株式 1 と株式 2 の両方を 1 株づつ保有，株式 1 のみ 1 株保有，株式 2 のみ 1 株保有，いづれの株式を保有しないエージェントが $N/4$ づつ存在する．各エージェントは，1 株保有している株式に対しては常に 1 株の売り注文を，保有していない株式に対しては常に 1 株の買い注文を出す．そのため，2 株以上の保有，空売り (マイナスの保有株数) は発生しない．両銘柄の相互作用は後に述べる株主構成によるファンダメンタル価格の変化のみによって起き，各エージェントの注文価格の決定は各銘柄ごとに完全に独立に行われる．

ファンダメンタルエージェントは $N_{\mathbf{P}}$ 体存在する．時刻 t，エージェント j の注文価格 $P_o^{t,j}$ は，$P_o^{t,j} = P_f \exp(d\sigma^j \pm m(\mu^j + 1))$ とする．ここで，d および m は定数であり，σ^j は j ごとに異なる実数を出力する正規分布乱数，μ^j は j ごとに異なる実数を出力する 0 から 1 までの一様乱数である．$\pm$ は買い注文のときは $-$，売り注文のときは $+$ をとる．

直近 (時刻 $t-1$) の市場価格 P^{t-1} に依存せず，株式がもつ本源的な価値（ファンダメンタル価格 P_f）を元に注文価格 $P_o^{t,j}$ を決める．各エージェントはファンダメンタル価格 P_f を知らないが推定を試みている．そして，推定ファンダメンタル価格から十分安い価格で買おうとし，十分高い価格で売ろうとする傾向があるといわれ，この十分な価格差のことを安全マージン (Margin of Safety) とよぶ [Graham 03]．$m(\mu^j + 1)$ は安全マージンの推定ファンダメンタル価格に対する比である．

テクニカルエージェントは $N_{\mathbf{T}}$ 体存在する．彼らは順張り戦略を採る．時刻 t，エージェント j の注文価格 $P_o^{t,j}$ は，$P_o^{t,j} = P^t(P^t/P^{t-tm^j})$ とする．現実の金融市場の価格変動を再現するためにテクニカルエージェントが必要であることが知られている *6．

ノイズエージェントは $N_{\mathbf{N}}$ 体存在し，時刻 t，エージェント j の注文価格 $P_o^{t,j}$ は，$P_o^{t,j} = P^t \exp(\eta\sigma^{t,j})$ とする．ここで η は定数，$\sigma^{t,j}$ は t および j ごとに異なる実数を出力する正規分布乱数である．本研究では常に十分な量の取引が行われている株式を取り扱う．これまでに述べたエージェントだけだと注文価格が特定の価格付近に偏り売買があまり成立しないことが多く発生するので，ノイズエージェントを導入した．なお，実際の金融市場においてもこのような，流動性 (多くの待機している注文がもたらす売買の成立のしやすさ) を供給する市場参加者が多く存在する．

2.2 価格決定メカニズム

時刻 t のすべてのエージェントの注文価格が決定されたのち，板寄せ方式 (call auction) [東証 15] で取引を成立させ各銘柄の市場価格 P^t を決定する．すなわち，買い注文は高い注文から，売り注文は安い注文から順番につき合わせていき，売買の注文価格が同じになったところを P^t とする．

2.3 ファンダメンタル価格の変更

Δt ごとに株主構成によっては両銘柄のファンダメンタル価格が変更になる．この変更が生じることを以後，“競争が起きた” とよぶ．t が Δt で割り切れる時刻のみ競争が起きる．$t - \Delta t$ から t までの各時刻において，株式 1 のみを保有するファンダメンタルエージェント数 n_1，株式 2 のみを保有するファンダメンタルエージェント数 n_2，両方を保有するファンダメンタルエージェント数 n_b を数え，n_1, n_2, n_b のうち n_1 が最も大きい時刻が最も多ければ株式 1 のファンダメンタル価格を δP_f 増やし，株式 2 のそれを δP_f 減らす．n_2 が最も大きい時刻が最も多ければその逆を行う．n_b が最も大きい時刻が最も多ければファンダメンタル価格は変更されない（競争は起きない）．

このモデルは実際の以下の現象をモデル化したものである．片方の株式のみを保有する投資家は両企業が競争し保有する企業が競争に勝ちファンダメンタル価値が上昇することが自身の利益につながる．そのため，保有しないほうの企業に競争で勝つように保有している企業に働きかける．一方，両方の株式を保有している投資家は両企業が競争することによって，たとえ競争に勝った企業のファンダメンタル価格が上昇しても，競争に負けた方の企業も必ず保有しておりファンダメンタル価値の下落によって損失もこうむるので競争を促す動機がない．このような現象が近年大きな議論になっていることは “はじめに” で述べた通りである．

ファンダメンタルエージェントのみを対象としたのは，ファンダメンタルにもとづいた投資を行う投資家やパッシブファンド以外の投資家 (投機家ともよばれる) は，企業と対話したり何かしらの経営戦略を働きかける（いわいるエンゲージメント）を行わないことが多いことをモデル化したためである．

2.4 パッシブエージェント

ファンダメンタルエージェントの初期に両銘柄を保有している $N_{\mathbf{F}}/4$ 体の内，N_{Fp} 体はまったく取引を行わない．この取引を行わないファンダメンタルエージェントをパッシブエージェントとよぶ．“はじめに” で述べたとおり，近年，良い銘

16a, Todd 16, 和泉 17a, 和泉 17b] がある．

*5 人工市場モデルの金融市場の規制・制度の議論への貢献は [水田 14, Mizuta 16a, 和泉 17a] が詳しい．また，東京証券取引所の持ち株会社，日本取引所グループは人工市場モデルによる研究を “JPX ワーキングペーパー”(http://www.jpx.co.jp/corporate/research-study/working-paper/) として多く公表している．

*6 例えば，[Chen 12]．[水田 17a, Mizuta 17b] の付録も参照．

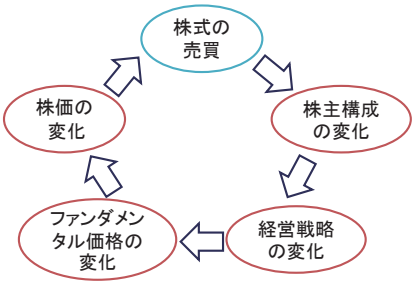


図 1: 水平株式保有におけるポジティブ・フィードバック過程

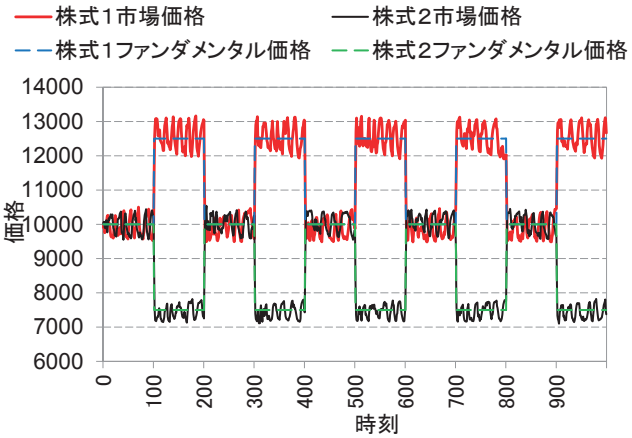


図 2: $N_{\mathbf{Fp}} = 0$ のときの各株式の市場価格 P^t およびファンダメンタル価格 P_f の時系列

柄を組み入れるための売買を全く行わないパッシブファンドが増えており，これをモデル化した．

3. シミュレーション結果

各種パラメータは，$N_{\mathbf{F}} = 400, N_{\mathbf{T}} = 100, N_{\mathbf{N}} = 1000, P_f = 10000, d = 0.05, m = 0.02, tm_{\mathbf{max}} = 100, \eta = 0.5, \Delta t = 100, \delta P_f = 2500$ とした．また，$N_{\mathbf{Fp}} = 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 = N_{\mathbf{F}}/4$ に対してシミュレーションを行い，各シミュレーションは $t = t_e = 1000$ まで行った．これらのパラメータの妥当性については [水田 17a, Mizuta 17b] の付録で示されている．また，本モデルは他のモデルに比べパラメータが少なく，恣意性が入りにくいのが特徴である．

図 2 は $N_{\mathbf{Fp}} = 0$ のときの各株式の市場価格 P^t およびファンダメンタル価格 P_f の時系列である．ファンダメンタル価格の変更が頻繁に起きており，競争が頻繁に起きていることが分かる．しかも，2 つの企業は交互に競争に勝っており，企業間競争のバランスをとるメカニズムが存在する可能性を示している．図 3 は $N_{\mathbf{Fp}}/N_{\mathbf{F}}$ ごとの競争が起きた回数である．$N_{\mathbf{Fp}}/N_{\mathbf{F}} > 12.5\%$ で競争が全く起きていない．パッシブファンドの割合がさほど大きくなくても，競争を阻害する可能性を示している．

図 4 は $N_{\mathbf{Fp}} = 0$ のときの各株式の市場価格とファンダメンタル価格の差 $P^t - P_f$ および単独保有ファンダメンタルエージェント数 n_1, n_2 の時系列 (時刻 $t = 100$ から 110 を拡大) で

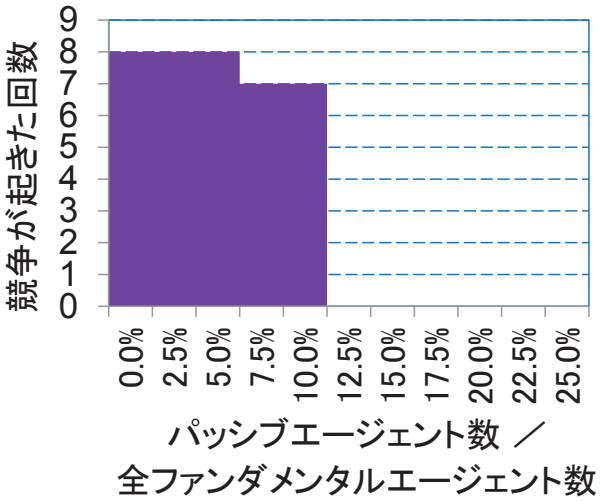


図 3: $N_{\mathbf{Fp}}/N_{\mathbf{F}}$ ごとの競争が起きた回数

ある．時刻 $t = 100$ で競争が発生し，株式 1 のファンダメンタル価格が上昇し，株式 2 のそれは下落した．ファンダメンタル価格が上昇した株式 1 では市場価格がそこに収束するまで買いが入り保有数が増えている．$t = 102$ までに新しいファンダメンタル価格へ収束しているが，それ以後はオーバーシュート [*7] してファンダメンタルより高い市場価格となっている．そのためこの期間は，株式 1 を保有しているファンダメンタルエージェント数の方が株式 2 のそれより少なくなっている．

つまり，現実においても，競争に勝った企業の市場価格が増加したファンダメンタル価格以上に上昇して割高となり競争を促す株主が離れて競争力を弱くする一方，競争に負けた企業の市場価格が減少したファンダメンタル価格よりさらに下落して割安となり競争を促す株主が増え競争力を強くして，企業間競争のバランスをとるメカニズムが存在する可能性があると考えられる．パッシブファンドの増加はこのようなメカニズムを弱める恐れがあると考えられる．

4. まとめ

本研究では，[水田 17a, Mizuta 17b] が構築した人工市場モデルを 2 銘柄に拡張し，株式水平保有が経営戦略を変更させ企業間競争を阻害するモデルを加え，パッシブファンドの増加が企業間競争と市場価格へ与える影響を分析した．

その結果，パッシブファンドの割合がさほど大きくなくても，競争を阻害する可能性を示した．また，競争に勝った企業の市場価格が増加したファンダメンタル価格以上に上昇して割高となり競争を促す株主が離れて競争力を弱くする一方，競争に負けた企業の市場価格が減少したファンダメンタル価格よりさらに下落して割安となり競争を促す株主が増え競争力を強くして，企業間競争のバランスをとるメカニズムが存在する可能性があることを示した．パッシブファンドの増加はこのようなメカニズムを弱める恐れがあると考えられる．

留意事項

本論文はスパークス・アセット・マネジメント株式会社の公式見解を表すものではありません．すべては個人的見解であります．

*7 オーバーシュートが実際に多く発生していることは実証研究で知られており，そのメカニズムに関しても人工市場モデルを用いて議論されている [Yagi 12b, 八木 12a, 水田 13, Mizuta 16b].

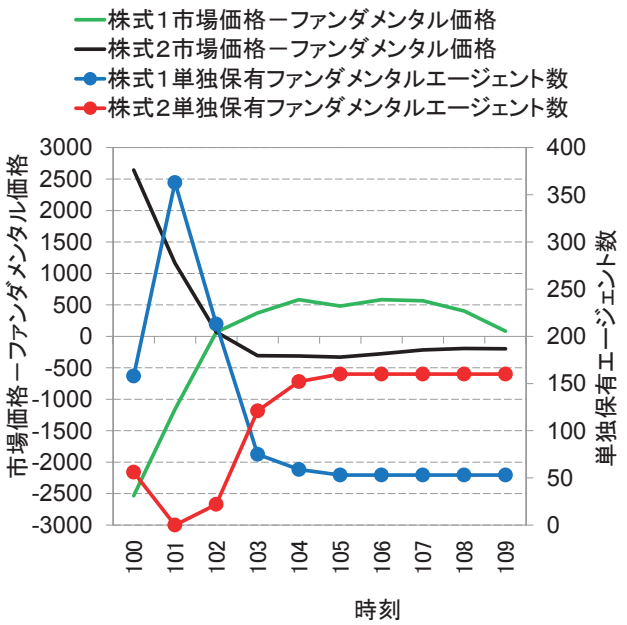


図 4: $N_{\mathbf{Fp}} = 0$ のときの各株式の市場価格とファンダメンタル価格の差 $P^t - P_f$ および単独保有ファンダメンタルエージェント数 n_1, n_2 の時系列 (時刻 $t = 100$ から 110 を拡大)

参考文献

[Azar 14] Azar, J., Schmalz, M. C., and Tecu, I.: Anti-Competitive Effects of Common Ownership, *SSRN Working Paper Series* (2014),
https://ssrn.com/abstract=2427345

[Battiston 16] Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., and Scheffer, M.: Complexity theory and financial regulation, *Science*, Vol. 351, No. 6275, pp. 818–819 (2016),
http://science.sciencemag.org/content/351/6275/818

[Braun-Munzinger 16] Braun-Munzinger, K., Liu, Z., and Turrell, A.: Staff Working Paper No. 592 An agent-based model of dynamics in corporate bond trading, *Bank of England, Staff Working Papers* (2016),
http://www.bankofengland.co.uk/research/Pages/workingpapers/2016/swp592.aspx

[Chen 12] Chen, S.-H., Chang, C.-L., and Du, Y.-R.: Agent-based economic models and econometrics, *Knowledge Engineering Review*, Vol. 27, No. 2, pp. 187–219 (2012),
http://dx.doi.org/10.1017/S0269888912000136

[Elhauge 16] Elhauge, E.: Horizontal Shareholding, *Harvard Law Review*, Vol. 129, No. 5, p. 1267 (2016),
https://harvardlawreview.org/?p=4185

[Elhauge 17] Elhauge, E.: The Growing Problem of Horizontal Shareholding, *SSRN Working Paper Series* (2017),
http://ssrn.com/abstract=2988281

[Farmer 09] Farmer, J. D. and Foley, D.: The economy needs agent-based modelling, *Nature*, Vol. 460, No. 7256, pp. 685–686 (2009),
https://www.nature.com/articles/460685a

[Fichtner 17] Fichtner, J., Heemskerk, E. M., and Garcia-Bernardo, J.: Hidden power of the Big Three? Passive index funds, re-concentration of corporate ownership, and new financial risk, *Business and Politics*, Vol. 19, No. 2, p. 298326 (2017),
https://doi.org/10.1017/bap.2017.6

[Graham 03] Graham, B. and Zweig, J.: *The Intelligent Investor: The Definitive Book on Value Investing*, HarperCollins (2003)

[和泉 17a] 和泉 潔, 川久保 佐記, 米納 弘渡：第 5 章 強靭な金融システム, 古田 一雄（編）, レジリエンス工学入門, 日科技連出版社 (2017),
http://www.juse-p.co.jp/cgi-bin/html.pl5?i=ISBN978-4-8171-9624-8

[和泉 17b] 和泉 潔：第 6 章 可能世界ブラウザとしてのエージェントシミュレーション, マルチエージェントのためのデータ解析 (マルチエージェントシリーズ), コロナ社 (2017),
http://www.coronasha.co.jp/np/isbn/9784339028126/

[LeBaron 06] LeBaron, B.: Agent-based computational finance, *Handbook of computational economics*, Vol. 2, pp. 1187–1233 (2006),
http://dx.doi.org/10.1016/S1574-0021(05)02024-1

[水田 13] 水田 孝信, 和泉 潔, 八木 勲, 吉村 忍：人工市場を用いた値幅制限・空売り規制・アップティックルールの検証と最適な制度の設計, 電気学会論文誌C（電子・情報・システム部門誌）, Vol. 133, No. 9, pp. 1694–1700 (2013),
http://doi.org/10.1541/ieejeiss.133.1694

[水田 14] 水田 孝信：人工市場シミュレーションを用いた金融市場の規制・制度の分析, 博士論文, 東京大大学大学院工学系研究科 (2014),
http://hdl.handle.net/2261/59875

[Mizuta 16a] Mizuta, T.: A Brief Review of Recent Artificial Market Simulation (Agent-Based Model) Studies for Financial Market Regulations and/or Rules, *SSRN Working Paper Series* (2016),
http://ssrn.com/abstract=2710495

[Mizuta 16b] Mizuta, T., Kosugi, S., Kusumoto, T., Matsumoto, W., Izumi, K., Yagi, I., and Yoshimura, S.: Effects of Price Regulations and Dark Pools on Financial Market Stability: An Investigation by Multiagent Simulations, *Intelligent Systems in Accounting, Finance and Management*, Vol. 23, No. 1-2, pp. 97–120 (2016),
http://dx.doi.org/10.1002/isaf.1374

[水田 17a] 水田 孝信, 堀江 貞之：忍耐強い (Patient) アクティブ投資は市場を効率的にするのか？—人工市場によるシミュレーション分析—, 第 19 回金融情報学研究会, Vol. 19, (2017),
http://sigfin.org/019-01/

[Mizuta 17b] Mizuta, T. and Horie, S.: Why do Active Funds that Trade Infrequently Make a Market more Efficient? - Investigation using Agent-Based Model, in *Computational Intelligence for Financial Engineering Economics (CIFEr), 2017 IEEE Symposium Series on Computational Intelligence on* (2017),
https://doi.org/10.1109/SSCI.2017.8280798

[水田 18] 水田 孝信：パッシブファンドの新たなる論点「水平株式保有」, スパークス・アセット・マネジメント (2018),
https://www.sparx.co.jp/report/special/2126.html

[O'Brien 17] O'Brien, D. and Waehrer, K.: AThe Competitive Effects of Common Ownership: We Know Less than We Think, *SSRN Working Paper Series* (2017),
http://ssrn.com/abstract=2922677

[Piketty 13] Piketty, T.: *Le Capital au XXI^e siècle*, Éditions du Seuil (2013), (邦訳: 山形浩生, 守岡桜, 森本正史: 21 世紀の資本, みすず書房 (2014))
https://www.msz.co.jp/book/detail/07876.html

[Rock 17] Rock, E. and Rubinfeld, D.: Defusing the Antitrust Threat to Institutional Investor Involvement in Corporate Governance, *SSRN Working Paper Series* (2017),
http://ssrn.com/abstract=2925855

[高橋 11] 高橋 大志：社会シミュレーションによる金融市場分析, 横幹連合コンファレンス予稿集, Vol. 2011, pp. 69–69 (2011),
http://doi.org/10.11487/oukan.2011.0.69.0

[Todd 16] Todd, A., Beling, P., Scherer, W., and Yang, S. Y.: Agent-based financial markets: A review of the methodology and domain, in *Computational Intelligence for Financial Engineering Economics (CIFEr), 2016 IEEE Symposium Series on Computational Intelligence on* (2016),
https://doi.org/10.1109/SSCI.2016.7850016

[東証 15] 東証：東証公式 株式サポーター 株式取引編, 東京証券取引所 (2015),
http://www.jpx.co.jp/learning/tour/books-brochures/detail/08.html

[八木 12a] 八木 勲, 水田 孝信, 和泉 潔：人工市場を用いた市場暴落後における反発メカニズムの分析, 情報処理学会論文誌, Vol. 53, No. 11, pp. 2388–2398 (2012),
http://id.nii.ac.jp/1001/00087035/

[Yagi 12b] Yagi, I., Mizuta, T., and Izumi, K.: A study on the reversal mechanism for large stock price declines using artificial markets, in *2012 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr)*, pp. 1–7 (2012),
http://dx.doi.org/10.1109/CIFEr.2012.6327791

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

金融レポート、およびマクロ経済指数による日銀センチメント指数の構築

Indexation of sentiment of Bank of Japan using text of Financial report and macroeconomic index

余野　京登[1]　和泉　潔[1]　坂地　泰紀[1]

Kyoto Yono[1], Kiyoshi Izumi[1], Hiroki Sakaji[1]

[1]東京大学大学院

[1]Graduate School of Engineering,The University of Tokyo

１　はじめに

金融市場におけるテキストデータは、投資家にとって分析対象の一つであり、その重要性は日に日に増している。決算短信をはじめとする企業の業績について書かれたドキュメント、証券会社のアナリストが書いた個別企業についてのアナリストレポート、更にはツイッター等のSNSで発信している個人投資家のつぶやき等、多種多様なテキストデータが存在する。定性的な投資判断の材料になるとともに、これらのテキストを用いて、モデルを構築し、定量的な分析を行うこともある。

本研究では、中央銀行が発行する議事録やステートメントなどの公的な文章を対象に、定量的な数値化を目的としている。具体的には、日本銀行の発行する金融政策決定会合の議事要旨から物価、生産、雇用等の各トピックに対するセンチメント指数の構築を目指す。

1.1 センチメント付与に関する問題意識

公的な文章からセンチメントを導き出す研究は、各国の中央銀行を対象に行われてきた。山本らは、日本銀行のテキストを対象に、ニューラルネットワークを用いた深層学習でスコア付を行い、指数化を行った[1]。また、D. Wu らは、FOMCのテキストを対象に、LDAによりトピックわけを行い、各センテンスに関しては、極生語辞書を用いたスコア付を行った[2]。これらの研究においては、まず一文単位でスコアを付与した後、１ドキュメントに対しては、各文の平均スコアをその時点のセンチメントとしている。その際の課題点を３つ上げる。１点目は、あるドキュメントにおいて、文章数が少なく、かつ、一つの文のスコアが極値を取った場合、ドキュメント全体がその特定のスコアに大きく影響されてしまう。２点目は、ドキュメントには時系列性を有しているはずだが、それを特に考慮していない。３点目としては、モデルにより数値化しているが、それはドキュメント単体のみを評価しているに過ぎず、中央銀行の実際に行っている「マクロ経済指標等の各データを観察→各議員が議論→その結論として、ドキュメントが発行される」という一連の構造を捉えていない。

本研究では、これらの問題意識を踏まえ、マクロ経済指標も用いた生成モデルを用いて中央銀行のトピック別のセンチメント指数の構築を試みる。

２　モデル構築

本章では、実際の日本銀行における金融政策運営の概略について説明した後、その枠組ベースにした構築したモデルの概要について述べる。

2.1 日本銀行の金融政策運営

日本銀行の公式サイト[3]によると、金融政策運営は以下の通り、行われている。

金融政策運営の基本方針は、日本銀行政策委員会の「金融政策決定会合」とよばれる会合で決定します。会合では、金融経済情勢に関する検討を行うとともに、金融市場調節方針や当面の金融政策の運営方針を決定し、決定した内容は直ちに公表しています。（中略）金融政策決定会合では、年に8回、2日間かけて集中的に審議を行い、金融政策の方針を決定しています。議決は9名の政策委員（総裁、2名の副総裁、6名の審議委員）による多数決によって行います。

つまり、金融経済情勢に関して、各９名の審議委員が議論を行う。そして、その結果として議論の内容が議事要旨として公表される。

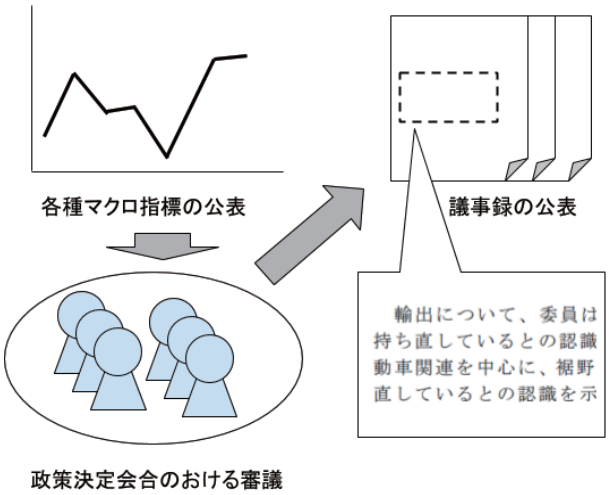


図１: 日本銀行における金融政策運営の概観

2.2 分析対象テキスト

日本銀行のトピック別のセンチメントを構築するために、対象とするテキストは金融政策決定会合の議事要旨とする。さらに、議事要旨内の「金融経済情勢に関する委員会の検討の概要」セクションのみを対象とする。対象のテキストの特徴としては、１点目としては、トピックについての記述が「～～について、」というように明示的に文頭に書かれる。２点目としては、「委員は、～～を共有した」や「一人の委員は、～～と述べた」というようにどの程度の人数の委員によって議論や意見がなされたかが文章内に必ず記述される形となっている。

わが国の景気について、委員は、所得から支出への前向きの循環メカニズムが働くもとで、緩やかに拡大しているとの見方で一致した。委員は、企業部門の動きについて、輸出は増加基調にあるほか、設備投資も、収益が過去最高水準を更新する中、緩やかな増加基調にあるとの認識を共有した。また、家計部門についても、委員は、個人消費

図２: 議事要旨の一例

2.3 スコア付

先行研究[1]で用いられた景気ウォッチャー調査を学習データに LSTMで構築したモデルでのセンチメントを各文に対して付与を行う。また、各文において、委員の意見の一致度により、そのスコアの重み付けを行う。

2.4 生成モデル

実際の日本銀行の金融政策運営を考慮し、以下のマルコフ性を考慮した生成モデルを構築する。

まず、各文のスコアSはその時点における日本銀行のセンチメントを示す潜在変数Zに基づき生成させる。その潜在変数Zはマクロ指数Mに依存しており、かつ１時点前の自身にも依存する時系列性を有している。

より詳細なモデルについては、研究会当日にて発表する予定である。

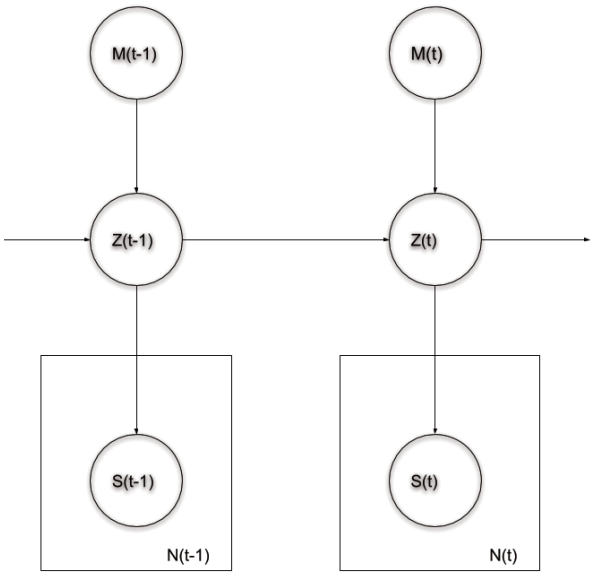


図３: 本モデルにおけるグラフィカルモデル

３　結果、およびまとめ

前章で述べたモデルを用いたシミュレーション結果、および考察を研究会当日にて発表する予定である。

参考文献

[1] 山本 裕樹,松尾 豊:景気ウォッチャー調査の深層学習を用いた金融レポートの指数化,第30回人工知能学会全国大会,(2016)

[2] Jegadeesh, N., and D. Wu: Deciphering Fedspeak:The Information Content of FOMC Meetings,AFA 2016 San Francisco Meeting Paper（2015）

[3] 金融政策決定会合の運営 ： 日本銀行 Bank of Japan, www.boj.or.jp/mopo/mpmsche_minu

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

分散投資規制は市場流動性や効率性，運用成績にどのような影響を与えるか？

How Does the Rule for Investment Diversification Affect Market Fluidity, Efficiency, and Investor's Performance?

丸山 隼矢[*1] Shunya Maruyama
水田 孝信[*2] Takanobu Mizuta
八木 勲[*1] Isao Yagi

[*1] 神奈川工科大学 Kanagawa Institute of Technology
[*2] スパークス・アセット・マネジメント株式会社 SPARX Asset Management Co., Ltd.

As financial products have grown in complexity and level of risk compounding in recent years, investors have come to find it difficult to assess investment risk. Furthermore, companies managing investment funds are increasingly expected to perform risk control and thus prevent assumption of unforeseen risk by investors. A related revision to the investment fund legal system in Japan led to establishing what is known as the ”rule for investment diversification” in December 2014, without a clear discussion of what its expected effects on market price formation having taken place. In this paper, we therefore used an artificial market to investigate how the rule for investment diversification affects market fluidity, efficiency, and investor’s performance. The results clearly demonstrate that implementing the rule for investment diversification tends to decrease trading volume, maintain market efficiency, and restrain the risk of holding assets.

1. はじめに

近年投資信託(以下，投信と略す．)のリスクが把握しにくくなっている．なぜならば，投信の組み入れ対象となる金融商品の仕組みが複雑になるとともにリスクの大きさも把握しにくくなってきたからである．その結果，投信が組み入れ保有している特定の資産にリスクが集中する状態や，投信の運用成績が特定の保有資産の運用成績に大きく依存する状態になっていることが考えられる．しかし，このような状況を投資家がすべて把握することは不可能である．そのため，投資家が想定外のリスクを負うのを未然に防ぐため，投信の運用会社側でリスクをコントロールすることが求められるようになり，2013 年には投資信託法制が改正され，2014 年 12 月より分散投資規制という規制が設けられることとなった[蒲谷 14, 杉田 14].

分散投資規制は投信の各構成資産に対して保有上限を設ける規制である．分散投資そのものの是非については，主に運用成績の良し悪しの観点から多くの実証研究で議論されている．Cremer ら[Cremers 09]は，高いアクティブシェアの投信は，ベンチマークインデックスの投信より運用成績が総じてよいのに対し，低いアクティブシェアの投信は運用成績がよくない事を明らかにした．

しかし，実証研究では分散投資に関する運用成績についての議論は盛んに行われているが，分散投資に何らかの制約(分散投資規制等)が設けられたときの市場価格形成についての議論は行われていない．なぜなら，現実の市場には，さまざまな投資家が取引に参加しているため,分散投資の制約による取引のみが価格形成に与える影響を抽出することは困難だからである．特定の取引が市場に与える影響を分析するような課題，つまり，実証研究では議論が困難な課題を分析する方法の 1 つとして，人工市場を用いる方法がある．人工市場とは，コンピュータ上に仮想的に構築されたマルチエージェントシステムの金融市場である[Chen 12, Mizuta 14]．市場側に何らかの制約(市場安定化や効率化のための制度等)をモデル化し組み込むことで，投資家がどのように振る舞うか，さらに，彼らの振る舞いによって市場にどのような影響が現れるかを検討することが可能となる．

連絡先：丸山隼矢，神奈川工科大学情報学部
神奈川県厚木市下荻野 1030
s1421007@cce.kanagawa-it.ac.jp

最近では市場規制制度が金融市場に与える本質的な影響を分析する研究が盛んに行われている[和泉 03,大井 13] が，分散投資規制に関しては八木ら[Yagi 17]が取り組んでいるのみである．八木ら[Yagi 17] は，2 資産人工市場を用いて，一方の資産価値が急落したときに他方の資産の価格形成に与える影響を調査し，分散投資規制が適用されている場合は他方の資産価格も連動して下落することを明らかにしている．このように，分散投資規制が市場の価格形成に与える影響についての議論は人工市場を用いて行われているが，分散投資規制が運用成績に与える影響についての議論はまだ行われていない．さらに市場流動性や効率性についての議論は実証研究を含めても行われていない．そこで本研究では，人工市場を用いて分散投資規制が市場流動性や効率性，さらに投信(を運用しているエージェント)の運用成績にどのような影響を与えるかを，分散投資規制の強弱も考慮して調査した．

2. 人工市場モデル

本研究では，水田らの人工市場モデル[水田 13]を基にして，人工市場を 2 資産市場へと拡張し，分散投資規制モデルを構築した．

2.1 注文プロセス

本モデルは 2 つのリスク資産(以下，単に「資産」と呼ぶ)のみを取引対象とし，価格決定メカニズムは連続ダブルオークション方式(ザラバ方式)とした．本市場には n 体のエージェントが取引に参加する．ただし，そのうち半数のエージェントが分散投資規制対象とする．エージェントはエージェント番号 $j = 1$ から順番に $j = 2,3,4,\cdots$と注文を出す．最後のエージェント $j = n$ が注文を出すと，次の時刻にはまた初めの エージェント $j = 1$ から注文を出し繰り返される．時 刻 t は 1 体のエージェントが注文を出すごとに 1 増える．つまり，注文をしただけで取引が成立しない場合も 1 ステップ進む．エージェント j は注文価格，売り買いの別を以下のように決める．時刻 t にエージェント j が予想する価格の変化率(予想リターン) $r_{e_{j,k}}^{t}$は，以下の式 (1) から得られる．

$$r_{e_{j,k}}^{t} = \left(w_{1,j,k} \log \frac{P_{f,k}}{P_k^{t-1}} + w_{2,j,k} r_h^{t-1}{}_{j,k} + w_{3,j,k} \epsilon_{j,k}^{t} \right) / \sum_{i=1}^{3} w_{i,j,k} \quad (1)$$

ここで, $w_{i,j,k}$は時刻 t, エージェント j, 資産 k の i 項目の重みであり, シミュレーション開始時に, それぞれ 0 から $w_{i,max}$ まで一様乱数で決める. この値は後述する学習過程により変化する. $u_{j,k}$ は時刻 t, エー ジェント j, 資産 k の 3 項目の重みであり, シミュレーション開始時に, それぞれ 0 から u_{max} まで一様乱数で決められ, シミュレーション中に変化することはない. $P_{f,k}$ は時間によらず一定の資産 k のファンダメンタル価格である. ファンダメンタル価格とは, 金融資産を発行する企業自体が持っている実態の価値に基づいた価格を指す. $P_{t,k}$ は資産 k の時刻 t での取引価格取引されなかった時刻では一番最近に取引された価格であり, 時刻 $t = 0$ では $P_{t,k} = P_{f,k}$ とする), $\epsilon_{j,k}^{t}$は 時刻 t, エージェント j, 資産 k の乱数項で, 平均 0, 標準偏差 σ_{ϵ} の正規分布乱数である. $r_{e_{j,k}}^{t}$ は資産 k, 時刻 t にエージェント j が計測した過去リターンで, $r_{h_{j,k}}^{t} = \log(P_k^{t-1}/P_k^{t-\tau_j})$ である. ここで, τ_j は 1 から τ_{max} まで一様乱数でエージェントごとに決める. 式 (1) の右辺の括弧内の第 1 項目は, ファンダメンタル価格と比較して, 安ければプラスの予想リターンを, 高ければマイナスの予想リターンを示すファンダメンタル投資家の成分である. なお, ファンダメンタル投資家は通常ファンダメンタル価値を参照して投資判断を行う投資家を指す. 第 2 項目は過去のリターンがプラス (マイナス) ならプラス (マイナス) の予想リターンを示す, テクニカル投資家の成分である. テクニカル投資家は過去の価格推移を参照して投資判断を行う投資家を指す. 第 3 項目はノイズを表している. 予想価格 $P_{e_{j,k}}^{t}$ は, 予想リターン $r_{e_{j,k}}^{t}$,k に基づいて 式 (2) から求められる.

$$P_{e_{j,k}}^{t} = P_k^{t-1}\exp(r_{e_{j,k}}^{t-1}) \qquad (2)$$

注文価格 $P_{0_{j,k}}^{t}$ は, 平均 $P_{e_{j,k}}^{t}$, 標準偏差P_dの正規分布に従う乱数で決まる. 価格の変化幅の最小単位は P_d とし, 注文の売り買いは予想価格 $P_{e_{j,k}}^{t}$ と注文価格 $P_{0_{j,k}}^{t}$の大小関係で決まる.

$$\begin{aligned} P_{ej,k}^{t} > P_{oj,k}^{t} \quad &\text{なら 1 単位の買い} \\ P_{ej,k}^{t} < P_{oj,k}^{t} \quad &\text{なら 1 単位の売り} \end{aligned} \qquad (3)$$

注文数は常に 1 と一定とする. 全てのエージェントは, 注文価格や注文数を決定後, レバレッジ制約に違反していないか判定を行う. さらに分散投資規制対象エージェントはその判定も行う. 本モデルの価格決定メカニズムはザラバ方式であるため, 買い (売り) 注文は, その注文価格より安い (高い) 売り (買い) 注文が市場に既に存在すれば, その中で最も安い (高い) 売り (買い) 注文と即座に取引が成立する. しかし, 相対する注文が市場になければ, 今回の注文は市場に残す. 市場に残した注文は, キャンセル時間 t_c だけ経過しても取引が成立しなかったときキャンセルされる.

2.2 学習プロセス

本研究では, 実証研究が存在を示している「投資戦略の切り替え」に焦点をあて, そのメカニズムの分析を行いやすいように比較的シンプルでパラメータの少ないモデルの構築を行った. 戦略を状況に応じて切り替えるという学習プロセス を以下のようにモデル化した. すなわち, 価格変化の方向を当てている戦略のウエイトを引き上げ, 外れている戦略のウエイトを引き下げるようにした. また, 資産 k の学習期間のリターン$r_{l_k}^{t}$ をかけることにより, 小さい価格変動を当てたり外したりしても大きくウエイトが増減しないようにした. 学習はエージェントごとに, 各エージェントが注文を出す直前に行われる. ファンダメンタル投資家成分だけの場合の予想リターンを$r_{e_{1,j,k}}^{t} = \log(P_{f,k}/P_k^{t})$, テクニカル投資家成分だけの場合の予想リターンを $r_{e_{2,j,k}}^{t} = r_{h_{j,k}}^{t}$とする ($k$ = 1,2). これらの予想リターン $r_{e_{i,j,k}}^{t}$ が資産 k の 学習期間のリターン $r_{l_k}^{t} = \log(P_k^{t}/P_k^{t-t_l})$ と比べ,

$$\begin{aligned} &\text{同符号なら,} \quad w_{i,j,k} \leftarrow w_{i,j,k} + k_l r_l^{t} p_j^{t}(w_{i,max} - w_{i,j,k}) \\ &\text{異符号なら,} \quad w_{i,j,k} \leftarrow w_{i,j,k} - k_l r_l^{t} p_j^{t} w_{i,j,k} \end{aligned} \qquad (4)$$

のように, $w_{i,j,k}$ を書き換える. ここで, k_l は定数, p_j^{t} は時刻 t, エージェント j に与えられる 0 から 1 まで の一様乱数である. 上記式(4)での学習プロセスの他に, 小さい確率 m で $w_{i,j,k}$ を再設定する. つまり, 0 から $w_{i.max}$ までの 一様乱数で決めなおす. これはランダム学習を意味しており, 実績からの学習と組み合わせることにより, エージェントが試行錯誤的により良い戦略のウエイトを求める姿を客観的にモデル化している.

2.3 制度モデル

本モデルは, 水田ら[水田 13]では実装されていなかった資産の総保有量を制限するレバレッジ制約と, 各資産の保有量を制限する分散投資規制から構成される. 既述のようにレバレッジ制約はすべてのエージェントが対象であるが, 分散投資規制は該当するエージェントのみが対象となる. 本モデルにおいて, レバレッジ制約は, 各エージェントの資産の取引高の総額を, 純資産の v_{upper} 倍以下に規制するようにモデル化した. レバレッジ制約は以下の式で表される.

$$\sum_{k=1}^{2}\left|P_k^{t} S_{j,k}^{t}\right| \le v_{upper} NAV_i^{t} \qquad (5)$$

ここで, $S_{j,k}^{t}$ は, 時刻 t におけるエージェント j の 資産 k の保有量であり, $S_{j,k}^{t} > 0$ なら買い保有している状態, $S_{j,k}^{t} < 0$ なら空売りしている状態を表している. v_{upper}はレバレッジ比率であるが, 分散投資規制が適用される投信では通常レバレッジを利かせた取引をすることはないため, 今回はv_{upper} = 1.0 とした. なお, 時刻 t におけるエージェント j のキャッシュ の保有量を C_j^{t} とすると, エージェント j の純資産 NAV_i^{t} は次のように定義される.

$$NAV_i^{t} = \sum_{k=1}^{2} P_k^{t} S_{j,k}^{t} + C_j^{t} \qquad (6)$$

分散投資規制は次のとおり定義される.

$$\frac{|P_k^{t} S_{j,k}^{t}|}{NAV_j^{t}} \le w_{dir} \qquad (7)$$

w_{dir} は, 分散投資規制の規制比率 (以降, 規制比率と略す) であり, 各エージェントの純資産に対する各資産の保有高の上限率である. 何らかの理由 (例えば, P_k^{t} が P_k^{t-1} より大きく上昇した等) で式 (7) が満たさなくなったとき, エージェントは式 (7) を満たすようになるまで資産 k の注文を出しつづけるようにしている.また, 予めそのような事態になることを防ぐため, 分散投資規制が適用されたエージェントは, 注文を発注する際にその注文が規制に違反しそうな場合 (上記の式 (7) を満たさない場合) は, それをキャンセルするよう振る舞うようにした.

3. シミュレーション結果と考察

本研究では, 表 1 のようにパラメータを設定し, 実験を行った. 実験の回数は, 各条件において 30 回試行し, 結果はその平均をもって議論する.

本研究では, 1)分散投資規制対象エージェントがいない市場 (以下, 規制なし市場と呼ぶ), 2) 全体の半数(500 体)のエージェントが弱い分散投資規制 (w_{dir}= 2/3)の対象となっている市場(以下, 弱い規制市場と呼ぶ), 3)全体の半数(500 体)のエージェントが強い分散投資規制(w_{dir} = 1/3)の対象となっている市場(以下, 強い規制市場と呼ぶ)の下で, 人工市場モデルの妥

当性を検証した後，シミュレーション結果から，出来高（市場流動性），市場非効率性，運用成績についての議論を行う．エージェントの初期キャッシュ量は，200,000，100,000，50,000 の 3 とおり用意する．この理由は，各エージェントの純資産の多寡（投信の資産規模の違い）で，分散投資規制が市場に影響が異なるのかどうかを検証するためである．なお，シミュレーション下における資産 1 と資産 2 の条件は同じであるため，結果もほぼ同じ傾向となった．よって，資産 2 の結果と考察については必要なときのみ議論し，基本的には資産 1 の結果を基に考察を行うものとする．

表 1　パラメータ

Parameters	Value
n	1,000
$w_{1,max}$	1
$w_{2,max}$	10
u_{max}	1
τ_{max}	10,000
σ_ϵ	0.03
P_d	1,000
t_c	10,000
t_t	10,000
k_l	4
m	0.01
δP	1

3.1 人工市場モデルの妥当性

現実の市場ではファット・テイルとボラティリティ・クラスタリングが現れる，と多くの実証研究で指摘されている[Cont 01]．よって，本モデルもこれらが再現できるようパラメータ設定を行った．ファット・テイルは，市場価格の騰落率の分布が正規分布ではなく裾が厚い，すなわち，尖度が正であることである．また，ボラティリティ・クラスタリングは市場価格の騰落率の 2 乗が，ラグが増えても自己相関係数が有意に正であり，長期記憶性を持つことである．表 2 は各シミュレーション条件下でのスタイライズド・ファクトを示している．いずれのシミュレーションにおいても，尖度はプラスでファット・テイルとなっており，騰落率の 2 乗の自己相関はラグが増えてもプラスでボラティリティ・クラスタリングが再現されている．

表 2　各条件下におえるスタイライズ ド・ファクト

		資産 1			資産 2		
分散投資規制の規制比率		規制なし	2/3	1/3	規制なし	2/3	1/3
尖度		2.42	2.42	2.51	2.41	2.41	2.53
	ラグ						
価格騰落	1	0.15	0.15	0.14	0.15	0.15	0.15
率の	2	0.11	0.11	0.11	0.11	0.11	0.11
2 乗の自	3	0.08	0.09	0.08	0.08	0.09	0.08
己相関	4	0.07	0.07	0.06	0.06	0.07	0.06
	5	0.06	0.06	0.05	0.05	0.05	0.05

3.2 出来高

表 3 は資産 1 の出来高の平均である．この表から，規制が強くなるほど初期キャッシュ量の大小に関係なく，出来高が小さくなる傾向があることがわかる．しかし一部例外が発生している箇所も見受けられた（初期キャッシュ量 200,000 のとき，17,038（規制なし市場）→ 17,120（弱い規制市場））．一方，規制比率が同じ場合でも，各エージェントの初期キャッシュ量が減少すると，出来高も減少している．

このような結果になった理由は次のように考えられる．まず，w_{dir}が小さくなるに従い，出来高が減少する場合について考える．w_{dir}が小さくなると，純資産が同じでも資産ごとの保有高の上限は小さくなる．そのため，特定の資産をある程度保有するとすぐに保有上限に達してしまい，それ以上保有しようとする注文がキャンセルされ，発注できないケースが増加する．キャンセル数については，表 4 に示す．

その結果，取引が成立する機会が減少し，出来高も減少すると考えられる．例えば，強い規制市場（w_{dir}が 1/3 のとき）では，規制なし市場や弱い規制市場（w_{dir}が 2/3 のとき）に比べてキャンセルされる注文が多いため出来高が減少しているが，弱い規制市場では，規制なし市場とそれほどキャンセルされる注文が変わらないため，出来高も規制なし市場とあまり違いが出ていない．そのため上記のような例外が発生したものと考えられる．

よって，分散投資規制がないとき，もしくは規制が弱いときは出来高に違いは見られないが，規制が強くなると出来高は減少傾向にあるといえる．

次に，初期キャッシュ量の値が小さくなると出来高の値も減少する場合について考える．初期キャッシュ量が小さくなると，w_{dir}の値が同じ市場でも，純資産が小さくなるため相対的に資産ごとの保有高も小さくなる．その結果，w_{dir}が小さくなる場合と同じ理由により，出来高が減少すると考えられる．

以上より，規制が弱い間は，分散投資規制が出来高に与える影響は限定的だが，規制が強くなってくると影響が出てくる．それと同時に，運用資産規模（初期キャッシュ量）が小さくなっても影響が出てくると考えられる．

表 3　資産 1 の出来高の平均

		分散投資規制の規制比率		
		規制なし	2/3	1/3
初期キャッシュ	200,000	17,038	17,120	16,770
	100,000	16,742	16,599	15,932
	50,000	15,545	15,410	13,580

表 4　資産 1 の注文キャンセル数の平均

		分散投資規制の規制比率		
		規制なし	2/3	1/3
初期キャッシュ	200,000	1,893	2,795	19,201
	100,000	33,534	37,032	78,438
	50,000	110,395	114,535	213,010

3.3 市場非効率性

金融市場には価格発見機能と流動性供給機能があり，これらによって社会に価値を提供している．金融資産は市場で取引されている価格とは別にファンダメンタル価格が存在すると考えられている．金融市場は，市場参加者たちの取引を通じて実態価値であるファンダメンタル価格を発見する，価格発見機能を持つことが期待されており，ファンダメンタル価格から大きく乖離し

ていない価格で取引されることが期待されている. それが達成された状態を効率性が高いという. 分散投資規制は投資家の各資産の保有高を規制し, 投資家の保有リスクを過剰にしないことを目的とするもので, 市場を効率化するための規制ではない. しかし, 分散投資規制の導入によって市場が非効率になることは望ましくない. そこで, 分散投資規制を導入した場合に, 市場が非効率にならないかを検証する. 市場の効率性を直接測定する指標として, 実験市場研究でしばしば用いられる市場非効率性を用いることにした. 資産 k の市場非効率性 M_{iek}は以下の式で求められる[水田 13].

$$M_{iek} = \frac{1}{t_e}\sum_{t=1}^{t_e} \frac{\left|P_k^t - P_{f_k}\right|}{P_{f_k}} \qquad (8)$$

ここで, t_e は計測期間を示し, 今回は 1,000,000 とした. M_{iek} は 0 以上の値を取り, 0 なら完全に効率的, 大きくなればなるほど非効率であることを示す. 表 4 は各条件における資産 1 の市場非効率性を示している. 分散投資規制の強弱による値の差は認められないが, 初期キャッシュ量が少なくなるほど非効率になる傾向がある.

表 5　資産 1 の市場の非効率性

	分散投資規制比率		
初期キャッシュ	規制なし	2/3	1/3
200,000	0.11%	0.12%	0.12%
100,000	0.13%	0.13%	0.12%
50,000	0.16%	0.15%	0.15%

3.4 運 用 成 績

本節では, 分散投資規制が適用されたエージェントとそうでないエージェントの運用成績の違いについて確認する. ただし, エージェントの運用成績はシミュレーション開始時の純資産に対する終了時の純資産の増減率で表す.

表 6 は分散投資規制が適用されたエージェントとそうでないエージェントの運用成績の平均である.

分散投資規制の強弱による差は見受けられないが, 初期キャッシュ量が小さくなるほど分散投資規制が適用されたエージェントの運用成績の方がよくなる傾向がある. ただし, 運用成績の差は最大でも高々0.014%(初期キャッシュ量が 50,000, 分散投資規制比率が 1/3 のとき)ほどなので, 分散投資規制が運用成績に与える影響は限定的であると思われる.

表 6　運用成績の増減率の平均

初期キャッシュ	エージェント	分散投資規制比率	
		2/3	1/3
200,000	規制あり	0.000023%	0.000003%
	規制なし	−0.000023%	−0.000003%
100,000	規制あり	0.0025%	0.0024%
	規制なし	−0.0025%	−0.0024%
50,000	規制あり	0.0052%	0.0069%
	規制なし	−0.0052%	−0.0069%

4. まとめと今後の課題

本研究では, 人工市場を用いて分散投資規制が 1)市場流動性, 2)市場効率性, 3)運用成績にどのような影響を与えるかを, 分散投資規制の強弱を考慮して調査した. その結果, ファンダメンタル価格が一定の場合は分散投資規制が強くなると出来高が減少する, つまり, 市場流動性が低くなる傾向があることがわかった. 一方で, 運用資産規模(初期キャッシュ量)が小さくなると, 間接的に規制の影響を受けて市場流動性だけでなく効率性も低下する傾向が見られた. なお, 運用成績においては規制の適用有無による明確な違いは出なかった. 今後の課題は以下の通りである. 現在の人工市場モデルでは, 注文量を常に 1 としているが, 現実ではもっと多くの資産が存在し, 1 単位以上の注文を出すことができる. そのため, 現在のモデルをより現実に近いものとするために, 2 単位以上の注文が出せるように拡張を行い, 本結果と比較することが挙げられる. さらに今回はすべての資産のファンダメンタル価格が一定であるため, 市場効率性や運用成績に明確な方向性が出なかった可能性がある. よって, 片方の資産価格のみが急落する場合など特定の状況では非効率化することや運用成績に差が出ることが考えられるため, それらの検証を行うことが挙げられる

参考文献

[蒲谷 14] 蒲谷 俊介:改正投信法が投信業界に与える影響, 金融 IT フォーカス, 野村総合研究所, pp.12-13, (2014)

[杉田 14] 杉田 浩治:投資信託の 14 年改革と今後の課題, 資本市場研究会, No. 347, pp. 4-11, (2014)

[Cremers 09] M. Cremers and A. Petajisto: How active is your fund manager? A new measure that predicts performance, Review of Financial Studies, Vol. 22, No. 9, pp.3329–3365, (2009).

[Chen 12] S.-H. Chen, C.-L. Chang, and Y.-R. Du: Agent-based economic models and econometrics, Knowledge Engineering Review, Vol. 27, No. 2, pp. 187–219, (2012)

[和泉 03] 和泉 潔: 人工市場: 市場分析の複雑系アプローチ, 森北出版, (2003)

[Mizuta 14] T. Mizuta, K. Izumi, I. Yagi and S. Yoshimura: Regulations' effectiveness for market turbulence by large erroneous orders using multi agent simulation, In Proceedings of IEEE Computational Intelligence for Financial Engineering and Economics 2014(CIFEr2014), pp.138–143, (2014)

[大井 13] 大井朋子:エージェントシミュレーションを用いた「価格規制」と「ネイキッド・ショート・セリングの禁止」の有効性 の検証, 金融庁金融研究センター FSA リサーチ・レビュー, Vol. 7, (2013)

[Yagi 17] I. Yagi, A. Nozaki, and T. Mizuta: Investigation of the rule for investment diversification at the time of a market crash using an artificial market simulation, (in press), (2017).

[水田 13] 水田 孝信, 和泉 潔, 八木 勲, 吉村 忍: 人工市場を用いた値幅制限・空売り規制・アップティックルールの検証と最適な制度の設計, 電気学会論文誌論文誌 C, Vol. 133, No. 9, pp.1694-1700, (2013)

[Cont 01] R. Cont: Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, Vol. 1, pp.223–236, (2001)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Analyzing Mergers and Acquisitions (M&A) in Japan Using AI methods

Bohua Shao[*1] Kimitaka Asatani[*1] Ichiro Sakata[*1]

[*1] School of Engineering, the University of Tokyo

If potential M&A cases can be detected automatically, this technology will improve the efficiency of M&A target recommendation and effectiveness of in-process M&A cases. However, in the past, M&A recommendation was impossible due to insufficient data and complexity of M&A. In this research, we provided a clustering method with cash flow features and company relationship features. From M&A clustering, we observed that M&A tend to concentrate in specific clusters. In order to improve the precision of M&A recommendation, we also analyzed the relationships between features from financial items and we extracted important features for identifying company relationships. The result of this research shows feasibility of recommending M&A from big data. In the future, we will design and select more features for analyzing M&A and we will associate results from AI with Management Science.

1. Introduction

Mergers and acquisitions (M&A), as we all know, are the transactions of the ownership of organizations. Historically, M&A in Japan showed a strong countercyclical trend. That is when the economy was prosperous, M&A were few while when the economy was not going well, M&A cases were prevailing. [Mehrotra 2011]

However, the "Financial Big Bang of Japan" from 1997 to 2006 changed this situation dramatically. Recently, according to the report of METI (Ministry of Economy, Trade and Industry) on 2018's tax reform [METI 2017], the shareholders of acquirees can defer the tax on revenue during M&A. This reform helps promote M&A in Japan in the future.

Currently, M&A business is still conducted by huge human labor and scientific approaches are not prevailing for selecting suitable M&A partners. Actually, many M&A deals tend to be failures. In the financial engineering field, there are abundant AI approach research fruits. Hence, it is also possible to have M&A recommendations if sufficient data are provided.

This research is a part of the research project "Research on the feasibility of the M&A target recommendation for practical use". The ultimate goal of this project is to provide companies with potential M&A target lists by an AI system. This work, aiming at understanding current M&A patterns and features in Japan, serves as a fundamental research for the project.

Previous literature about M&A preferred case studies and econometric methods. In this research, we used AI method, which was seldom used in M&A research field. By AI approach, in the future, we will be able to save labor work on M&A analysis a lot in the M&A business field. In this research, we also provided a new M&A clustering method, which is especially suitable for M&A in Japan in the past fifteen years.

2. Previous literature

Dickson argues about the relationship between the cash flow patterns and the firm life cycle. There are several stages in the firm life cycle and there are also corresponding cash flow patterns in each stage. [Dickinson 2011]

Shibayama *et al.* argued that M&A between firms of equal size is likely to result in difficulty in integration of knowledge base. [Shibayama 2006] Cloodt *et al.* had following statements. First, in order to increase innovative performance through M&A, companies have to target firms with moderately related knowledge base. Second, the author's results clearly demonstrate that in high-tech industries, the M&A of a large absolute knowledge base only contribute to improved innovative performance during the first couple years after integration. Third, non-technology M&A in high-tech industries contribute less or negatively. [Cloodt 2006]

"Mastering the merger", published by Bain & Company, Inc. in 2004, said M&A is a paradox for companies. Successful companies learn from M&A and authors recommend that companies should start from small M&A deals and then have large deals. [Harding and Rovit 2004]

In 2012, Mori *et al.* published a paper with the topic of predicting business partners by AI approach. In this paper, there are 3 feature types: suppliers' features, customer's features and their relationships' features. The main method of this research is the support vector machine (SVM), which is a kind of supervised machine learning approach. In this paper, there are two experiments for predicting the business relationship. The latter experiment, which is designed to predict the reciprocal business relationships, is the elaborated version of the former one, which only predicts business relationships. The final achievement of this paper is developing a web system for recommending business partners. [Mori 2012]

3. Data

In this experiment, we are using the databases from UZABASE, Inc. In these databases, we have access to M&A information, company information and financial information.

In the M&A database, companies participating in M&A are divided into 4 roles. They are acquirer, ultimate acquirer, seller and target. The difference between acquirers and ultimate acquirers is that acquirers are ordinary companies while ultimate acquirers are mainly Private Equity funds and venture capital firms. The difference between targets and sellers is that targets are organizations for sale while sellers are the owners or

Contact: Bohua Shao, School of Engineering, the University of Tokyo, shaobohua6@gmail.com

shareholders of targets. In this paper, we only focused on "Acquirer" and "Target" because they are typical M&A roles.

There are 7 M&A types in the databases. They are "Merger of Equals", "Fund Buy-out", "Acquisition", "Minority Stake", "Joint Venture", "MBO" and "Demerger".

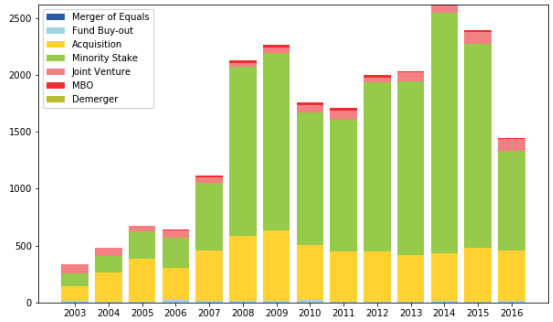


Figure 1. General information of M&A in Japan (2003-2016)

Figure 1 shows the number of M&A in each year in Japan, according to their types. From this figure, we find "Minority Stake", "Joint Venture" and "Acquisition" are prevailing in each year. In this research, we only focused on "Merger of Equals" and "Acquisition" because they are traditional understanding of M&A (M&A is the abbreviation of "Mergers and Acquisitions").

We have access to company information. In this paper, we focused only on 3692 Tokyo Stock Listed companies' information. They are labelled with the SPEEDA industrial classification codes. The SPEEDA code consists of 3 parts and 9 digits in total. The first 3 digits show the general classification, the second 3 digits show the medium classification and the last 3 digits show the detailed classification.

4. Experiments

4.1 Methods

In this research, we found M&A are very complex. In order to understand M&A patterns and phenomena, we used K-means [Shalev-Shwartz 2014] for our first trial.

Since we designed several features for analyzing M&A and we were interested in knowing which features are effective, we used Principal Component Analysis (PCA) [Shalev-Shwartz 2014].

4.2 Cash flow features

Before investigating M&A, we were interested in company behaviors. As mentioned in Chapter 2, the cash flow is a proxy for understanding company behaviors. We conducted such an investigation so as to have several basic knowledge for further research.

In this section, we used the following formula to calculate free cash flow:

$$\text{Free Cash Flow} = \text{Cash flow from operating activities} + \text{Cash flow from investing activities} \quad (1)$$

We designed 8 features shown in Table 1 for analyzing cash flow.

Table 1: 8 features

Name	Financial items
cfoa	Cash flow from operating activities
cfia	Cash flow from investing activities
cffa	Cash flow from financing activities
freecf	Free cash flow
plnetsales	Net sales
oprofit	Ordinary profit
cfb	Cash & Cash Equivalent - Beginning
cftd	Changes in cash flow

In this section, we focused on companies of all industries in Japan and we only focused on yearly financial reports. The cash flow data can be fetched from 1989, so we analyzed cash flow data from 1989 to now. We designed a vector with a length of 8, 8 dimensions. The 8 elements in a vector represent the financial data of 8 financial items of a company of a fiscal year. We also deleted vectors with data missing and finally we succeeded in generating about 70,000 vectors.

We used PCA for understanding the importance of each feature and we found that "Net sales" and "Cash & Cash Equivalent- Beginning" have high weights. Hence, these two features are important for cash flow analysis and will be fundamental knowledge for designing features and analyzing M&A further.

4.3 Features

In order to have M&A analysis, we will design 3 feature types: acquirers' features, targets' features and relationships' features. We plan to adopt the results from 4.2 and design features about financial items. As a first step, in this paper, we preferred to firstly investigate relationships' features.

We extracted M&A deal information from the M&A database. In this paper, we only focused on M&A between 3692 listed companies. According to the following statements, we made 3 features:

Dis3: If the acquirer and the target share the same general industrial classification (the first 3 digits in SPEEDA codes), the value is 1; otherwise: 0.

Hist3y: Value: 1: the acquirer had M&A experience in the period of three years in succession before the M&A announcement; 0: others.

Histbool: Value: 1: the acquirer had M&A experience before the M&A announcement; 0: others.

4.4 Evaluation

In order to have further understanding of M&A, we generated random samples for evaluation. For a fair evaluation, we again used the 3792 listed companies. We randomly make pairs from them and generated about 7,000,000 pairs by combination.

Next, we randomly selected 1,000 pairs from the nearly 7,000,000 pairs as random samples in this paper.

We deleted any data flaw and finally made 778 random sample vectors and 109 M&A case vectors. Then, we mixed up 778 random sample vectors and M&A vectors and applied K-means for clustering.

5. Results

We have three features and they are all valuing 0 or 1. Obviously, we should have 8 situations. However, there are 2 situations with "Hist3y" valuing 1 and "Histbool" valuing 0 are impossible. Hence, we decided to separate into 6 groups.

We had 6 clusters. The table below shows the number and proportion of M&A cases and random samples in each cluster. The very right column shows the ratio of M&A to random samples in each cluster.

Table 2: The number of M&A and random samples in clusters

Cluster id	M&A cases (%)	Random Samples (%)	Ratio of M&A to Random Samples
0	13 (11.93%)	399 (51.28%)	3.26%
1	7 (6.42%)	153 (19.67%)	4.58%
2	20 (18.35%)	59 (7.58%)	33.90%
3	37 (33.94%)	10 (1.29%)	370%
4	19 (17.43%)	128 (16.45%)	14.84%
5	13 (11.93%)	29 (3.73%)	44.83%
Total	109 (100.00%)	778 (100.00%)	Avg. 14.01%

From this Table 2, we find Cluster 0 has the lowest ratio of M&A cases to random samples while Cluster 3 does the highest. Hence, by only 3 features, we are able to find that M&A concentrate in a cluster. This phenomenon is interesting and we are interested in adopting other features in the future.

6. Summary

In this research, we attempted to analyze M&A with AI approaches. Results in Chapter 5 will help us design features for analyzing M&A in the future.

In the section 4.2, we analyzed cash flow. However, cash is strongly related with corporate governance and several reforms were conducted in Japan in the past years. In the future, we will associate the cash flow analysis with corporate governance.

We had only 109 M&A cases for Chapter 5 and it is far from enough. We will expand our research boundary from Japan to the whole world for our future research topic. We had only 3 features in this paper and in the future, we will adopt the results from cash flow analysis and design more features for further analysis.

Acknowledge

Here, we show our sincere gratitude to UZABAZE, Inc. for providing databases.

References

[Dickinson 2011] Dickinson V.: Cash Flow Patterns as a Proxy for Firms Life Cycle, Account Rev., 2011.

[Mehrotra 2011] Mehrotra V., et al.: Creditor-Focused Corporate Governance: Evidence from Mergers and Acquisitions in Japan, Journal of Financial and Quantitative Analysis, 2011.

[METI 2017] Ministry of Economy, Trade and Industry, 経済産業関係　税制改正について, Available at: http://www.meti.go.jp/main/yosan/yosan_fy2018/pdf/zeiseikaisei.pdf, 2017.

[Shibayama 2006] Shibayama S., *et al.*: Effect of mergers and acquisitions on drug discovery: perspective from a case study of a Japanese pharmaceutical company, Drug Discov. Today, 2008.

[Cloodt 2006] Cloodt M. *et al.*: Mergers and Acquisitions: Their effect on the innovative performance of companies in high-tech industries, Research Policy, 2006.

[Harding and Rovit 2004] Harding D., Rovit S.: Mastering the merger: four critical decisions that make or break the deal, Harvard Business Press, 2004.

[Mori 2012] Mori J., *et al.*: Machine learning approach for finding business partners and building reciprocal relationships, Expert Syst.Appl, 2012.

[Shalev-Shwartz S 2014] Shalev-Schwartz S., Ben-David S.: Understanding machine learning: from theory to algorithms, New York: Cambridge University Press, 2014.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

人工知能による雇用の代替と組織構造の適応に関する考察

-日本の経営機械化史から-

Consideration on substitution by artificial intelligence and adaptation of organization structure - From the history of management information systems theory -

法雲 俊栄[*1]
Shunei Norikumo

[*1] 大阪商業大学
Osaka University of Commerce

Although research on artificial intelligence has been carried out for nearly 70 years, expectations are rising in the world, more than anything, booms of recent years have the potential to greatly change the working environment and private life of many people. Under such circumstances, studies that threaten the anxiety that human beings lose their jobs due to the substitution of artificial intelligence and that certain job types can lead to unemployment have also been announced. From the viewpoint of management information theory, until now, when new information technology has appeared, mankind has established an organizational structure in line with the times by flexibly adapting to the technology, such as the contents of business and the flow of work. Based on the above background, I would like to look back on the history of management information theory and consider how human beings have changed from the computer age to change the organization structure and how the next generation organization structure should be.

1. はじめに

2013 年、英オックスフォード大学で人工知能（AI：Artificial Intelligence）の研究をおこなうマイケル・A・オズボーン准教授とカール・ベネディクト・フレイ研究員の共同研究による論文『The Future of Employment：雇用の未来―コンピュータ化によって仕事は失われるのか』により、702 種の職がコンピュータに取って代わられる確率を詳細に試算した内容が公表され、世間に衝撃を与えた。

人工知能に関する歴史は、1950 年のチューリングによる『Computing Machinery and Intelligence』、機械は知的かどうかの問に始まり、研究はJ・マッカーシー主催によるダートマス会議後の 1958 年、ローゼンブラットパーの『The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain』、セプトロン、などによって、現在の機械学習及びデータマイニングに繋がる基盤研究がスタートした。1960 年代から、人工知能の推論や探索が始まり、現在は3度目のブームを迎えていると言われているが、何より近年の人工知能ブームが以前と違うところは、多くの人の労働環境や私生活を含む近未来を大きく変える可能性があり、世間では良くも悪くも期待が高まりつつある。そのような中、国内の機関紙でも人工知能の代替により人間は仕事を失い、特定の職種は失業する可能性があるといった不安を脅かす情報が広がりつつある。

未来を予測し、将来への期待、最悪のシナリオ、さまざまな方面から日々研究と議論がおこなわれているが、どれが正しく、誤っているということは、まだ誰にも分からない。明らかなことは、今、人工知能の技術によって、社会が大きく変わろうとしている。そのことに対する期待値も大きいが夢見ているだけでは、また失敗に終わってしまう可能性もある。「また」というのは、周知の通り、人工知能は、科学技術の偉人たちによる労も虚しく、一度、二度と様々な問題に適応できなかった問題があり、今回、第 3 波のブームを迎えていることに起因する。

ビジネスへの実用化、とりわけ経営情報論的な視点に立てば、これまで人類は、新たな情報技術が登場したとき、業務内容や業務の流れなど、人や組織が技術に柔軟に適応することで時代に即した組織構造を確立し、革命の時代を乗り越えてきた。雇用が無くなると聞けば、誰もがAI時代到来の社会に対して批判的な意見や、どことなく不安な面を抱き、明るい未来はイメージできないが、AIは人間よりも優れている側面が沢山あると考えられる一方で、AI は苦手・不得意とする分野も多く存在することから、演繹的に人間と機械が共存できる社会を提案したい。

以上のような学術的な背景と世情から、日本の経営機械化から歴史を遡り、人はどのように変わり、組織構造の適応を遂げてきたのか。組織では、どのような人材を育成し、未来を託した若者に教育をおこなってきたのかを探り、次世代における組織構造の適応について模索したい。人工知能の導入の可能性について、明るい未来が見出せるよう、ロボットには代替が出来ない、人間に託された本質的な役割を含め見出したい。

本論の展開は、2 章は、第 3 次ブームにおける人工知能の特性と過去の情報化の時代におけるコンピュータの特性について。3 章は、主に 1970 年代の経営情報論史にみる補完的の意味と、経済的な意味の代替の差異について明らかにする。4 章は、1940 年代から始まる日本の経営機械化史から、現在の ICT 導入までの組織構造の変遷について探る。5 章は、考察。6 章は、まとめとする。

連絡先：氏名：法雲俊栄，所属：大阪商業大学 総合経営学部，
住所：〒577-8505 大阪府東大阪市御厨栄町 4-1-10，
Mail：norikumo@daishodai.ac.jp

2. 人工知能の特性と過去の情報化の時代

今まで多くの人工知能に関する研究成果がある中で、近年の人工知能の大きな特徴についてみたい。

コンピュータ自らが膨大な情報源から大量のデータを瞬時に分析し、学習、推論、考察して、質問に回答したり、人間の意思決定を支援したりする、コグニティブ・コンピューティング・システムがある[佐藤 16]。このコグニティブと呼ばれる「経験的知識に基づいた」、「認知力が優れた」という意味で、IBM の AI : Watson は、現代の人工知能の代表的なシステムであるといえる。

2011 年 2 月、米国の人気クイズ番組「ジェパディ！」再出演した AI:Watson は、2 日間にわたって人間との対戦に挑んだ。引き分けだったがトータルの正答率で完勝して、人間を越える実力を示しただけでなく、賞金 100 万ドルも獲得して話題になった。その当時から飛躍的に進歩した AI: Watson は、2016 年 8 月 4 日、東京大学医科学研究所で、それによって人間が診断できなかった 60 代の女性患者の難解な病名をわずか 10 分で白血病と正確に割り出し、適切な治療法によって患者の命を救ったと発表した。[IBM 11] [佐藤 16] 他にも、プロの棋士に勝利し、弁護士が苦労する多数の判例や証拠を瞬時に検索して事件を解決するなど、人工知能の基盤技術が様々なシーンで活躍している。

しかし、誰もが人工知能として認めたい AI: Watson であるが、IBM は、「Watson は AI の要素技術をふんだんに活用しているが、あくまで人をサポートすることが目的であり、この目的こそが AI との違いである」と説明している[吉崎 16]。つまり、人工知能とは位置づけていないのである。このことは、経済面で仕事の代替は可能であっても、実際のビジネスシーンである経営面で仕事を代替する事とは異なることに関係し、実際にビジネスで溶け込むかどうか、今後のビジネスへの導入に欠かせない考えであるため、次章で解説したい。

IT の進歩に伴い、徐々に人間の担っている業務も人工知能等を有するコンピュータで処理することが増加していくことは当然の事であるが、その事実が、「コンピュータに仕事を奪われるのではないか」、「コンピュータ乃至はロボットに人間が使われるのではないか」と危惧する言葉もあり、このような情報化に対する疑問は、過去の機械化、コンピュータ化の歴史の過去にもあり、労使間でも問題になった。

近年の人工知能の特性と従来のコンピュータの違いを比べてみると、従来のコンピュータが得意とする処理は、定型的な予めプログラム化された命令をルーチンワークとして実行される。工場等の生産現場では工作機械やロボットが、オフィスでは商品管理や顧客管理による販売促進、財務・会計処理による資産管理など、現場には OA 機械やコンピュータなどの機器が投入されてきた。私たち人が日々おこなっている仕事は、これまで数々の情報技術の導入と長い情報化の歴史背景の中、変容してきたのである。

3. 経済的な代替と、経営的な補完の意味と差異

巷では、人工知能が職種によって私たちの現在の労働を代替する可能性について議論されているが、経営的な視点と経済的な視点では捉え方が大きく異なる。現在の情報通信技術と経営組織の関係においては、「補完的」利用が基本となっている。

これは、1970 年代の MIS ブーム時に遡るが、かつての組織は、環境との関係をさほど意識せずに、構成員の職務・権限・責任を明確に規定し、作業の方法や手続きを明確にすれば、人は合理的に機械のように受動的に行動し、その成果は確定的に予測できると考えられていた。しかし、経営の現場では、既存の業務や管理活動をそのままに、伝統的に事務部門の仕事として明示的に位置づけられた情報活動だけを自己完結的に独立させて ICT に代替させても、結果的には、その効率化・合理化レベルの効果しか実現できない。[遠山・村田・岸 03]。と示されている所以にある。

また、経済学的な見解として井上氏（2016）は、「ある産業でコンピュータ化と雇用の増大が同時に起こっていて、情報技術と労働者が補完的関係にあるように見えたとしても、経済全体での技術・機械と労働者の関係は完全に代替的ではなく、程々に代替的で、新しい技術は、その職業の雇用を完全に無くすのではなく、一定程度減少させる効果を持っている。」とされている。[井上 16]

つまり、私たちの日常において、意識している競合他社の領域とは別の違った域から、突如として、需要が奪われる可能性がある。これは、昨今の破壊的イノベーション域においても、その余波として、今後、人工知能による代替的な提案が進めば、大いに影響が出てくるものと思われる。今後、政府などの機関は、新たな技術の導入によって、波及的に他業種の雇用が奪われる可能性があることを念頭におき、それが極端に特定する職業の雇用を奪ってしまう恐れがあり、大きな経済的影響を与える可能性がある場合、法規制をかける等の手段で経済をコントロールして、結果的に人の雇用を守り、経済的混乱を避け、経済的影響を抑えた導入方法について考える必要がある。

前章でも述べた、IBM は Watson を人工知能として位置づけていないことは、人工知能が人間の知性を超える境界を「シンギュラリティ: 特異点」と呼ぶが、この特異点を越えたとき、またビジネスへの導入に際し、過去の情報化の歴史からしても、非常に重要な考え方であると思われる。

このように、日常における経営的側面と、視野をマクロに置き換えた経済全体の視点とでは、捉え方が大きく異なる。経済的側面から私たちの価値観を大きく覆され、仕事が代替される可能性はあると思われるが、経営的側面で完全な代替になることは、過去の学術的背景からしても無いと思われる。今後、人工知能の技術進歩が進んでいけば、それに伴う、隣接学問の質的発展が益々求められる。

4. 日本の経営機械化史より

本稿では、国内の経営機械化史にフォーカスを当て、世代別特徴と、情報技術と組織（人々）の間にどのような歴史的な変遷があったのかを取り上げて、本論の手掛かりとしたい。

明治維新以降の 20 世紀初頭において、経済的、政治的にも解き放たれた文明開化時に、日本は、欧米をはじめとする先進国の文物や制度を積極的に導入し、近代化に向かおうとする勢いのある流れがあった。文明開化というと、明るく、社会の解き放たれた明るいイメージがあるが、コンピュータ前史においての実情は異なる。1923 年パワーズ式統計機が米国から横浜に入港するも、翌日に関東大震災に見舞われたため、神戸港で降ろされた。また唯一の国産統計機として開発中だった川口式電気集計機は完成間近で計画を中止している [米花 75] 。

1945 年には終戦を迎えることになるが、後に紹介する国産統計機の開発プロジェクトが戦時中という、経済的にも厳しい状況の最中にあることを考えると、いかに過酷な情勢下で、ホレリス式、パワーズ式の統計機が導入され、日本のビジネスを機械化へ導こうとしたか、先達の苦労が伝わってくる。

4.1 日本の経営機械化

日本の経営機械化前史において活躍した機械は、PCS、金銭登録機、EDPS がある。事務の効率化・合理化に繋がると、統計局、鉄道省、税関、保険業、証券会社、銀行を中心に試行され、導入に至った。とりわけ、国内でパンチ・カード・システム（PCS）として親しまれ導入されたホレリス式統計機（現 IBM）は、後の組織にも「統計課」、「IBM 室」が一般的に加わるほど、情報技術史上、最も組織に大きなインパクトを与えた端緒である。

PCS が国内で初めて紹介されたのは、1892 年のこと、東京統計協会編の「統計集誌」第 129 号において、高橋二郎訳によりホレリス式統計機が「人口調査電気機械の発明」として紹介された。明治政府は、米国の国勢調査に見習い、現地の視察、国産統計機の開発に着手、国内の国勢調査法の制定をおこなった。1905 年川口市太郎技師が「川口式電気集計機」を試作したが、震災により計画は中止。その後は、商社からの輸入に頼り、日本における経営機械化の一歩が始まったとされる。

1923 年「パワーズ式統計機」（Reminton Rand 社、現 Unisys）が三井物産（吉澤 審三郎）によって統計局、鉄道省（現 JR）、横浜税関に、1925 年に保険業（日本生命と第一生命）に納入された。また同年、ホレリス式統計会計機が森村組（現、森村商事）水品浩によって日本陶器（現ノリタケ）に納入された。後には、1954 年にトヨタ自動車、1958 年に三井銀行（現、三井住友銀行) に統計機が導入され、各業界で急速に事務の効率化の流れが進んだ。これら日本を代表する組織は、どのように新しい情報技術に適応して変化してきたのか、その変遷について探る。

4.2 経営機械化史の長い大企業

PCS の導入は、20 世紀に入って進展するが、戦前は、導入の初期段階で十分に活用している組織は少なかった。戦後になると、輸入規制も無くなり、商社による輸入が進み、急速に各界で利用される。大幅な事務効率の成果を出すまでには至らなかった。この頃は、試験導入の時期といえる。大幅に事務の効率化が認められるようになったのは、ようやく 1950 年代の頃からである。以下は、1920 年代から 50 年代にかけて、積極的に事務の機械化に乗り出した企業群である。

■日本生命

1897 年と早期から業務上、保険数理や加入者の計算が必要だったため、英製テートス計算機を導入している。本格的に事務の処理に使われ始めたのは、1926 年パワーズ式統計機からで、主に逐次主計、外務、医療部門の実務面に使用された。1932 年にホレリス式統計機、1934 年にバロース統計機を導入し、保険料払込案内の作成を開始している。

組織は、創業以来、約 30 年間務めた片岡直温が政界入りを表明した 1919 年に、職制を改革して、主計課、財務課、会計課、経理課、を設置している。1929 年、弘世助太郎の 3 代目社長就任と同時に 14 課を 26 課に拡大し、経理局、財務局、主計局と配置している。1939 年には、経理課を 7 係にして充実させ、1944 年には、統計課が設置されている。1959 年には、IBM7070 電子計算機の受入れ。また、1957 年に外務教育制度の確立、1951 年から管理者教育、1953 年から内務職員教育を拡充している。[長谷川 62]

■日本陶器

1925 年にホレリス式統計機を、また 1962 年に事務用途の小型コンピュータ HITAC-201、電子（パラメトロン式）会計機 NEAC-1201 を導入している。後、1966 年には、EDP が導入された。1967 年に大型汎用コンピュータ HITAC-8300 を導入し、全社で EDP 教育を開始、1969 年に受注・出荷総合システムが完成した。後、経理関連システム KM-1 を導入している。

1920 年の早期から、営業部、会計係が設置されており、1964 年に経理部があり、1968 年には、コンピュータ部、1979 年にコンピュータ室、1984 年に情報システム部の流れで設置している。

日本でもいち早く導入された統計機の活用は、従来の会計・経理で活用されていたものと思われるが、社史によると大きく組織を見直されたのは、1966 年のコンピュータ利用に関する基本指針「EDPS を中心とする総合システム確立方針」[柴垣 95]からで、経営管理体制を含めた 5 年の長期計画を、3 年の組織内教育と研究開発期間で平行して進めている。

■豊田自動車

1954 年に IBM 統計機を導入し、実績計算、材料原価計算、減価償却費計算、工数計算、原価部品費計算を、1955 年に給与計算、1957 年に自販経理部機械計算課を設置している。1960 年に NCR31 会計機、IBM650 を導入、1963 年に IBM7074・1401 を導入、1966 年に IBM1440 を導入し工場電算機に利用し、1969 年に IBM360/40 に変更している。

組織は、豊田喜一郎時代、1939 年に経理部経理課に経理掛、別会計掛。1950 年に経理部に変更。1960 年に自販機械計算部、自工調査部機械計算課を設置。1968 年に自工電算部発足、1978 年に自販システム部発足、1987 年に情報処理システム部門を設置して、傘下に第一・第二情報システム部を設置している。1974 年電算センター完成により電算部と技術電算室が統合している。販売店の経営効率化を支援するため 1980 年に「C80 委員会」を設置し、翌年からオフィスコンピューターを活用した販売店業務の統一システム「C80」の開発を始めた。営業所のサービス部門業務から着手し、整備料金の自動計算や請求書の作成、サービスカルテなどをシステム化している。1990 年に各ネットワーク部署を集結させてシステム企画部が設置された。[トヨタ自動車 12]

■三井銀行

都市銀行の中でもいち早く、機械化に乗り出し、技術環境に即した組織構築をおこなっている。1958 年に統計資料の作成に PCS（IBM 製）を導入、1961 年に IBM1401 を導入し管理を外為と経理に拡大、1965 年に IBM 1410、IBM 1440 を導入し、日本初の銀行勘定系（普通預金）、本店・支店間の第 1 次オンラインシステム化に成功。1970 年頃に IBM S/370 を導入して、第 2 次総合オンラインシステム化し、総合口座に対応している。

組織の名称変更は、1958 年に統計室を設置、1961 年に統計課に変更。1967 年に機械計算課、1968 年、1969 年に東京と大阪に大規模な事務センターを設立し、システム開発課と電子計算課に変更。1975 年に大規模なシステム開発部、1989 年にシステム部、1994 年にシステム部と電算センターへと名称を変更している。[渡部 00]。

5. 考察 -機械化による組織構造の変化-

前章までに、経営の機械化を初期から始めた企業の歴史と組織構成の変遷についてみてきた。組織変遷の流れをまとめると、大きく 6 つ段階と改編に分けられる。どの企業も機械化までに長期の試験的利用、実験、研究をおこなっており、組織の目指すところを年代毎にスローガンを掲げるなどして明確にしている。また技術の理解に対する組織内の教育、人材育成が大規模でおこなわれている。これらのことは、企業によって導入の時期が約 10 年～30 年の違いはあるにせよ、流れは同じである。

組織変遷は、まず、事務部門に統計課や計算課が加わった後、さらに分かれ電子計算課などの部門が設置されている。後、

企業の事業拡大、経済的環境需要、とともに各部門に電子計算課が分散し、再度、電算センター部、情報システム開発部などの従来の計算課を統合する組織が生まれ、自社でプログラムや情報システムを開発する流れが見られる。表 1 は組織変遷の特徴と主要な情報技術をまとめたものである。この表から集中と分散を繰り返し、EDP 処理から組織が連動してきたことがわかる。

年代	分類	機械・技術の活用	組織構成の再編	経営環境	人材育成	組織形態
1950年以前	導入期	**事務**作業の機械化 パワーズ式、ホレリス式統計機、金銭登録機、NCR ■PCS、EDP	部内に統計課などを設置 （事務部門）	個別業務	組織開発	事業部制組織
1960年	分散期	**業務**の機械化 IBM1410、1440によるオンライン化、■EDPS、TSS	各基幹部門に設置 （各部門に電子計算課）	基幹業務	能力開発	機能別組織
1970年	分裂期	**経営**の自動化×Mis-MIS 経営の支援○ 意思決定の支援 ■MIS、DSS	自社開発組織を設置 （プログラム、情報システム）	企業間連携 多角化	人材開発	-
1980年	統合期	**意思決定**の支援 戦略的な支援 ■DSS、SIS 企業間ネットワーク	大型電算センター設置	多国籍化	異文化教育	フラット組織
1990年	連携期	業務**改革**、 ■BPR、インターネット	別部門・子会社化 （○○システム会社）	全社企業	キャリア開発	マトリックス組織
2000年以降	融合期	企業間の**最適化** ■ERP、SCM、DCM、DWH	アウトソーシング、外注 （Webサービス提供）	グローバル化	再組織開発 知識創出	再フラット組織 orプロジェクト・チーム型組織

表 1　経営の機械化史と組織構成の変遷

6. まとめ

本論では、人工知能による雇用の代替と組織構造の適応に関する考察として、日本の経営機械化史から組織改編についてみてきた。

経営の機械化史は、半世紀以上前のことであり、人工知能が及ぼす影響と比較し、同様の対応手段をとっても解決にはならないと思う。しかし、当時もまた今の私たちと同じように、新たな技術の受け入れの過程に、戸惑いと試行錯誤があって、技術を組織と人々で受け入れてきたと思うと、過去の技術導入と組織の変革の時代を理解しておくことは、必要なことではないかと思われる。

また 1970 年代には、コンピュータの導入により、上手く組織と連携できず、営業不振や倒産の防止に活用できない会社も存在し、MIS の概念が Mis-MIS と指摘され、情報システムと組織（人々）のあり方が見直された時期がある。人工知能の技術革新だけが先走り、法規制や倫理、応用技術の組織内教育、人材教育が行き届かないようであれば、両者が融合できる社会は育たないと思われる。

日本の経営の機械化の前史において、戦前は PCS 導入、そしてコンピュータの登場によって、無くなった職業もあれば、新たに創出された職業も存在する。経済の社会構造はもちろん、経営システムとしてどう受け入れるか。実用化される可能性のある技術に対していち早く、基礎知識、利用、導入の方法について組織内で周知しておく必要がある。

今後、人工知能の導入により、人件費を抑え、低価格を追及したサービスや商品がたくさん登場すると思われるが、そのサービスを受けることで、どのような経済的な波及効果が起こるかを考え、一部分のサービスをあえて受けないという信念も、また人間性を重視して調整が可能な秩序ある社会を維持するための策かもしれない。

昨今の学術的な流れの中で、AI は前例の無いことやデータが無い状況、人間のひらめきや直感といった能力が苦手で、その分野を AI は認識できないとも言われている。具体的には、AI は、データに基づいた分析・統計で仮説を立て、仕組み化された構造的な問題を解決することを得意とするが、感性的・直感的な非構造的な問題解決は不得意とされている。

今回の研究、機械化の導入と人工知能の導入を同義と位置づけることは難しいことであるが、戦後の経済的に低迷していた日本が、世界でも主要な IT 立国になった先達から、情報技術の発達の適応について学ぶべきことは、多いと思われる。本研究によって、AI の未来を受身としてではなく、AI と組織の個々人とが協働することで次世代における新たな組織構造の構築に貢献できればと考えている。

謝辞

本研究は、日本経営協会より委託を受けた「経営科学研究奨励」研究の一部である。日頃の研究にご理解とご支援をいただき、深く感謝と敬意を表する。学術的な領域人のみならず、世の人々にも安心して機械と共存できる社会をイメージしてもらえるよう貢献に務めたい。

参考文献

[アンソニー 68]　R.N.アンソニー(著), 高橋 吉之助 (訳): 経営管理システムの基礎, ダイヤモンド社, 1968 年.

[岸本 59]　岸本英八郎: 経営と技術革新, 日本経済新聞社, 1959 年.

[米花 75]　米花稔: 日本経営機械化史, 日本経営出版会, 1975 年.

[遠山 15]　遠山暁、村田潔、岸眞理子: 経営情報論, 有斐閣, 2015 年.

[島田 07]　島田 達巳, 高原 康彦: 経営情報システム, 日科技連出版社, 2007 年.

[定道 82]　定道宏, 布上康夫, 松本良治: TSS のための情報処理要論, 東洋経済新報社, 1982 年.

[情報処理学会 10]　情報処理学会歴史特別委員会: 日本のコンピュータ発達史, オーム社, 2010 年.

[渡部 00]　渡部 栄: 情報管理概論, 白桃書房, 2000 年.

[柴垣 95]　柴垣和夫、伊藤正直、長谷川信、橘川武郎:日本ガイシ 75 年史, 日本ガイシ株式会社, 1995 年.

[日本 IBM 02]　日本 IBM, ノリタケと日本 IBM の歩み http://www-06.ibm.com/jp/press/2002/07041.html, 2002 年.

[トヨタ自動車 12]　トヨタ自動車 75 年史, トヨタ自動車株式会社, 人事・情報システム, http://www.toyota.co.jp/, 2012 年.

[岡本 07]　岡本行二: 日本の情報システムとコンピュータ利用, 情報システム学会誌, Vol. 2, No. 1, 2007 年.

[長谷川 62]　日本生命七十年史, 日本生命保険相互会社, 1962 年.

[経営情報学会 10]　経営情報学会情報システム発展史特設研究部会, 明日の IT システムの発展史 総合編, 専修大学出版局, 2010 年.

[桑田 10]　桑田 耕太郎, 田尾 雅夫: 組織論[補訂版]　有斐閣アルマ, 2010 年.

[桐谷 11]　桐谷 恵介, 宮下 毅: クラウド時代の情報システム部, 中央経済社 2011 年.

[井上 16]　井上智洋, 人工知能の未来, 文藝春秋, 2016 年.

[IBM 11]　IBM, クイズ番組に挑戦した IBM Watson, https://www.ibm.com/watson/jp-ja/quiz/index.html, 2011 年

[吉崎 16]　吉崎敏文, IBM が Watson を「AI」と呼ばない本当の理由, http://www.atmarkit.co.jp/ait/articles/1601/29/news022.html, ITmedia, 2016 年.

[佐藤 16]　佐藤博:　人工知能（AI）が医療を変える！ わずか 10 分で白血病を見抜き患者を救った「IBM Watson」の底力, http://healthpress.jp/2016/08/ai-10ibm-watson.html , healthpress, 2016 年.

[2J2] [General Session] 13. AI Application

座長:水田 孝信（ スパークス・アセット・マネジメント株式会社）

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room J (2F Royal Garden B)

[2J2-01] Improvement of Prediction Accuracy in Predicting Market Trends by Newspaper Article Analysis Using Deep Learning

○Kazuki Matsumoto[1], Tohgoroh Matsui[1] （1. Chubu University）

1:20 PM - 1:40 PM

[2J2-02] A Novel Approach of Purchase Prediction Considering the Purchase on Competing Company by Using ID-POS Combined with Geofencing Location Data

○Junichiro NIIMI[1], Takahiro HOSHINO[2,1] （1. RIKEN Center for Advanced Intelligence Project, 2. Faculty of Economics, Keio University）

1:40 PM - 2:00 PM

[2J2-03] Classification by Time-Series Gradient Boosting Tree

○Kei Nakagawa[1,2], Mitsuyoshi Imamura[1,3], Kenichi Yoshida[2] （1. Nomura Asset Management Co.,Ltd., 2. University of Tsukuba Graduate School of Business Sciences, 3. University of Tsukuba Graduate School of Systems and Information Engineering）

2:00 PM - 2:20 PM

[2J2-04] Construction of Strategy using Binary Decision Diagram in Stock Investment using Technical Indicators

○Ohki Kato[1], Hazime Anada[1] （1. Tokyo City University）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習を用いた新聞記事分析による市場動向予測における予測精度の改善

Improvement of Prediction Accuracy in Predicting Market Trends by Newspaper Article Analysis Using Deep Learning

松本 一樹 *1　　松井 藤五郎 *2*3
Kazuki Matsumoto　　Tohgoroh Matsui

*1中部大学 大学院 工学研究科 情報工学専攻
Department of Computer Science, Graduate School of Engineering, Chubu University

*2中部大学 生命健康科学部 臨床工学科
Department of Clinical Engineering, College of Life and Health Sciences, Chubu University

*3中部大学 工学部 情報工学科
Department of Computer Science, College of Engineering, Chubu University

In this paper, we analyze newspaper articles using deep learning to forecast market trends. We have proposed a method to forecast market trends based on time-series text analysis using deep learning. This method works very well for forecasting TOPIX from The Nikkei (Nihon Keizai Shinbun) between 2008 and 2014, but the prediction accuracy falls after 2015. In this paper, we propose to reduce the duration of the training data in order to improve the prediction accuracy after 2015. As a result of the period of training data over the past three years, the prediction accuracy has been improved by 12.2%, from 55.1% to 67.3%.

1. はじめに

金融市場の動向を予測する研究は古くから行われている。これまで、金融市場の動きを予測する研究として、過去の取引実績の時系列パターンから株価を予測するテクニカル分析や、財務諸表などから株価を予測するファンダメンタル分析が主流であった。これらの手法はいずれも定量的データを用いた分析である。しかし、定量的データを用いた分析だけでは予測できない市場の動きがある。例えば、ある企業の不正会計が発覚したとき、その企業の株価は急落する。このような株価の動きは定量的データの分析だけで予測することはできない。そこで、近年、テキスト分析を用いて金融市場の動きを予測する研究が盛んに行われている [和泉 11, 藏本 13]。

テキスト分析による金融市場の動向予測において、株価や金利といった市場データである目的変数が時系列を成すように、金融市場の動きを説明するのに用いられるテキストも時系列をなしている。そのため、新聞などの定期的に発行されるテキスト・データは、一種の時系列データであると考えることができる。これを利用して、前日の新聞記事と当日の新聞記事を比較し、その差分に基づいて特徴語の出現パターンを抽出し、SVM を用いて東証株価指数 (TOPIX) が上昇するか下落するかを予測する手法が提案されている [松井 11]。

また、近年、パターン認識の分野において、近年深層学習が注目を集めている。深層学習は、画像認識や音声認識において高い認識精度を発揮しており、テキスト分析においても従来手法と比べ高い性能であることが確認されている [Socher 13]。そこで我々は、SVM の替わりに深層学習を用いたところ、時系列テキスト分析を用いた市場動向予測においても、深層学習を用いた予測は、2008 年から 2014 年までの予測正解率が 91.3%であった [松本 17]。しかし、2015 年、2016 年を対象とした実験では、2015 年は 59%、2016 年は 55.1%と予測正解率がそれ以前の予測正解率から大幅に低下してしまった（図 1）。

連絡先: 松本一樹，中部大学 大学院 工学研究科 情報工学専攻，
tp17028-2903@sti.chubu.ac.jp

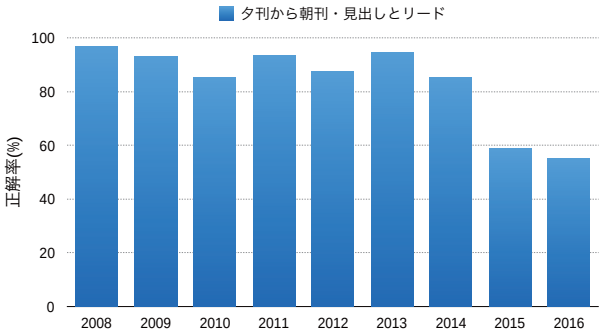


図 1: 予測正解率 [松本 17]

そこで本論文では、新聞記事を時系列データと捉えてテキストの差分に着目した分析を行い、深層学習を用いて株価が上昇するか下落するかを予測する手法において、2015 年以降において予測正解率が低下した原因について調査し、予測精度向上を図る。

2. 深層学習を用いた時系列テキスト分析による市場動向予測

まずはじめに、新聞記事を対象とした、深層学習を用いた時系列テキスト分析によって市場動向を予測する手法について説明する。

対象とする新聞記事は、日本経済新聞の見出しとリードとし、市場動向に影響を与えるテキストであると考えられる予測対象日の前営業日の夕刊から当日の朝刊を結合したテキストを用いる。

テキスト分析に用いる深層学習は、ニューラルネットワークを多層化したものであり、画像認識や音声認識で高い認識精度を発揮している。テキスト分析においても従来手法と比べて高

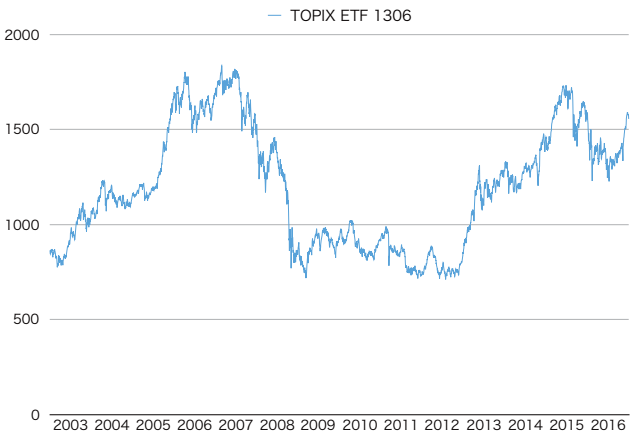


図 2: TOPIX ETF 1306 の株価推移

い性能であることが確認されている。深層学習への入力は、先行研究 [松井 11] で作成された特徴量ベクトルのデータを CSV 形式に変換したものを用いる。深層学習の出力は、予測対象日の日中の利益率が正または 0 のとき $(0,1)$、そうでないときは $(1,0)$ とする。新たに予測を行うときは、テキストを特徴量ベクトルに変換してニューラルネットワークへ入力し、その出力 (o_1, o_2) を求め、$o_1 < o_2$ なら利益率は正になると予測し、そうでないなら負になると予測する。

3. 提案手法

本論文では、従来手法において訓練データの期間を直近過去 5 年間であったのを、直近過去 3 年間として、特徴量ベクトルの作成と深層学習の学習を行う手法を提案する。

従来手法の問題点として、2015 年、2016 年を対象とした実験において、予測正解率が 2015 年以前と比べて大幅に低下したことがあげられる。2015 年、2016 年を対象とした実験では、2015 年は 59%、2016 年は 55.1%であった [松本 17]。2016 年を予測する際、訓練データの期間を直近過去 5 年間とすると、訓練データに 2013 年以前のデータが含まれる。しかし、2012 年にアベノミクスが始まり、2013 年頃からアベノミクスの効果が現れ始めたため、2013 年以前と以降では市場の動きが異なると考えられる（図 2)。

そこで、訓練データの期間を過去 5 年間より短くすることで、2016 年の予測における訓練データに 2013 年以前のデータが含まれないようにする。このように訓練データの期間を短くすることによって、2015 年以降における予測精度の向上が期待できる。

4. 実験結果と考察

提案手法の有効性を確認する為に検証実験を行った。日本経済新聞を対象として、予測対象日の前営業日の夕刊から予測対象日の営業日の朝刊までを 1 つのテキストとし、その見出しとリードを用いた。予測対象は 2016 年の東証株価指数 (TOPIX) 連動型上場投資信託 (ETF) とし、予測対象日の寄りから引けにかけて TOPIX ETF の取引価格が上昇するか下落するかを予測した。訓練データの期間を予測対象日の直近過去 3 年間とした。

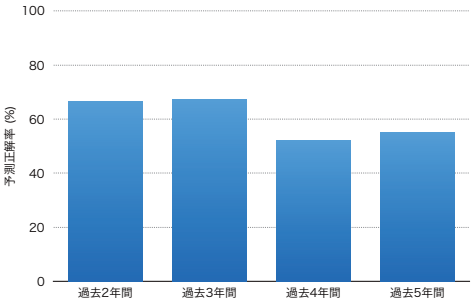


図 3: 予測正解率 (2016 年)

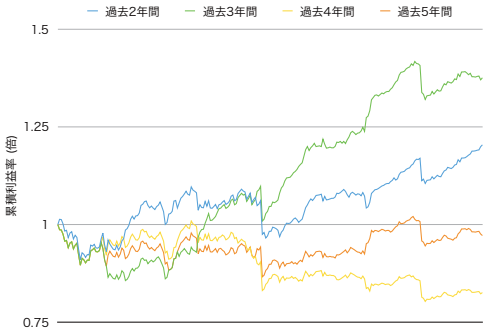


図 4: 累積利益率 (2016 年)

特徴語はテキストの差分に基づいて新出、続出、消滅の 3 つの出現パターンに分類する。分類した特徴語を出現パターンごとに出現回数をカウントし、30 回以上出現したパターンのみを対象とし、パターンが出現した日に上昇した割合が 0.55 以上または 0.45 以下のパターンを特徴として用いた。

深層学習のツールは TensorFlow を使用した。深層学習のユニット数は、入力層を特徴語数分のユニット、出力層を 2 ユニット、隠れ層は 2 層とし、1 層のユニット数を 64 ユニットとした。また、活性化関数には ReLU を用いた。

比較として、訓練データの期間を過去 5 年間、4 年間、2 年間にしたものを用いた。この時、それぞれの特徴量ベクトルは、特徴として用いる出現パターンの出現回数を、過去 2 年間の場合 20 回以上、過去 4 年間の場合 40 回以上、過去 5 年間の場合 50 回以上のパターンを対象として特徴量ベクトルを作成した。

実験結果を図 3、図 4 に示す。 実験の結果、従来手法である訓練データの期間が直近過去 5 年間では予測正解率が 55.1%であったのに対して、訓練データの期間を直近過去 3 年間にした場合では 67.3%となり、従来手法と比べて予測精度が 12.2%上昇した。また、過去 4 年間では 52.2%、直近過去 2 年間では 66.5%となった。

訓練データの期間を直近過去 4 年間にした場合と直近過去 2 年間にした場合において、直近過去 3 年間にした場合と比べて予測正解率が低下した原因として、直近過去 4 年間の場合では、アベノミクスの影響を受ける前である 2012 年の新聞記事が訓練データに含まれているためであると考えられる。

また、直近過去 2 年間の場合では、2013 年以前の新聞記事が訓練データに含まれていないものの、同じく 2013 年以前の

新聞記事が訓練データに含まれていない直近過去 3 年間の場合に比べて、訓練データが減っているために、直近過去 3 年間の場合に比べて予測正解率が低くなっていると考えられる。

本手法による予測に基づいて 1 年間の取引を行った場合の累積利益率を図 4 に示す。TOPIX ETF の価格が上昇すると予測された場合には、寄り付きに始値で買い、引けに終値で売る、下落すると予測された場合には、寄り付きで始値で信用売り、引けに終値で買い戻すことを毎日続けた場合のシミュレーションである。ただし、取引手数料は考慮していない。

1 年後の累積利益率は、直近過去 5 年間を訓練データの期間とした場合は約 0.97 倍、直近過去 3 年間を訓練データの期間とした場合では約 1.38 倍となった。このため、提案手法は有効であると考えられる。

5. まとめ

本論文では、深層学習を用いた新聞記事分析による市場動向予測と、2015 年以降において予測正解率が低下した原因について調査し、訓練データの期間を短くすることによって予測精度の向上を図った。日本経済新聞を対象として、夕刊から朝刊までを 1 つのテキストとし、その見出しとリードに対して時系列テキスト分析を行い、TOPIX ETF の取引価格が上昇するか下落するかを予測した。実験の結果、提案手法である訓練データの期間を直近過去 3 年間にした場合では予測正解率が 67.3%となり、訓練データの期間が直近過去 5 年間だった従来手法と比べて予測精度が 12.2%上昇した。よって、提案手法は有用であると考えられる。

今後は、2016 年以外の予測期間においても訓練データを短くすることが有効であるか確認したい。

参考文献

[和泉 11] 和泉 潔, 後藤 卓, 松井 藤五郎: 経済テキスト情報を用いた長期的な市場動向推定, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3309–3315 (2011)

[藏本 13] 藏本 貴久, 和泉 潔, 吉村 忍, 石田 智也, 中嶋 啓浩, 松井 藤五郎, 吉田 稔, 中川裕志: 新聞記事のテキストマイニングによる長期市場動向の分析, 人工知能学会論文誌, Vol. 28, No. 3, pp. 291–296 (2013)

[松井 11] 松井 藤五郎, 石田 智也, 中嶋 啓浩, 和泉 潔, 吉田 稔, 中川 祐志: 新聞記事を対象とした時系列テキスト分析による市場予測, 第 7 回人工知能学会ファイナンスにおける人工知能応用研究会 (SIG-FIN), pp. 44–47 (2011).

[Socher 13] Socher,R., et al : Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP 2013*, pp. 1631–1642 (2013).

[松本 17] 松本 一樹, 松井 藤五郎: 深層学習を用いた新聞記事分析による市場動向予測. 第 79 回情報処理学会全国大会, 6N-08 (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

位置情報データによる競合店舗の利用状況の多様性を用いた購買予測手法の提案

A Novel Approach of Purchase Prediction Considering the Purchase on Competing Company by Using ID-POS Combined with Geofencing Location Data

新美 潤一郎 *1　　星野崇宏 *2
Junichiro NIIMI　　Takahiro HOSHINO

*1理化学研究所革新知能統合研究センター
RIKEN Center for Advanced Intelligence Project

*2慶應義塾大学経済学部, 理化学研究所革新知能統合研究センター
Faculty of Economics, Keio University, RIKEN Center for Advanced Intelligence Project

Nowadays various kinds of data about customers behavior on online websites or applications can be collected for the marketing analysis. However, especially for the physical retailing stores, it is still difficult to acquire their behaviors on competing firms. In this research, we develop a novel approach of considering the usage of competitors We present the way to predict customers future purchase on retailing stores with ID-POS data combined with the third-party large-scale GPS locational data collected from the smartphone apps.

1. はじめに

近年では EC（電子商取引）サイトの発達やネットスーパーの普及などの要因により実店舗での小売販売も競争が激化している。マーケティングの戦略策定にあたり自社で保有する顧客について正しく情報を把握することは、正確なターゲティングやプロモーション展開にあたって必須である。中でも自社ロイヤルティの推定にあたっては消費者のシェア・オブ・ウォレット（SOW）*1 を考慮することが重要となる。

このような SOW の予測にあたっては、EC 市場においては Clickstream Data をはじめとしたアクセス履歴データの収集によるウェブサイト横断的なインターネットの閲覧履歴データを用いることで、消費者の競合企業での行動を考慮した解析が行われている。しかし実店舗企業での SOW は、企業横断的に利用履歴を得ることが困難であることから未だ事例が少ないというのが現状である。実店舗での購買を把握する代表的な手法として、これまでにもアンケートを用いた市場調査や、家計簿サービスにおける購買後のレシートのスキャンなどの方法で消費者の支出を把握することが試みられてきたが、これらの手法では認知バイアスによる回答の歪みやレシートスキャンにあたって調査対象者への負担が大きいといった問題があり、正確なデータを大規模に収集することは困難であった。

近年では広く普及したスマートフォン端末で、複数のアプリケーションを用いて特定の消費者の位置情報を継続的に収集する試みが大規模に行われている。しかし収集された位置情報は企業の小売データ等と共通の ID で連携しているわけではなく、マーケティング分析への活用にはまだ課題が残っている。そこで本研究では、大規模に収集された位置情報データと特定の小売企業で収集された ID-POS データを紐づけることで、実店舗での競合他社を考慮したマーケティング解析の手法を提示する。

連絡先: 慶應義塾大学経済学部星野崇宏 (e-mail: hoshino@econ.keio.ac.jp)

*1　顧客のある市場全体への支出に対する自社への支出額の割合

2. 先行研究

2.1 機械学習を用いた購買予測

機械学習、とりわけ深層学習はその急速な発展とともに様々な分野に応用されているが、マーケティング分野としての消費者の購買予測に応用した事例は画像認識等に比べると少数である。代表的なものとしてオンラインストアでの購買を扱ったものには、消費者の EC サイト上での行動についての Clickstream Data から、ある EC サイトへの訪問ごとの購買有無の予測とその際に購買された商品を予測し、その精度をロジスティック回帰分析、ランダムフォレスト、Deep Belief Network（DBN）で比較した [Vieira 15] がある。一方で実店舗での購買データを外部データと紐付けて解析したものとしては、購買履歴データ、アクセス履歴データ、アプリ操作履歴データから実店舗・オンラインストアでの購買を横断的に予測し、Feed-Forward Neural Network（FFNN）[Svozil 97] と一般化線形モデル（GLM）で結果を比較した [新美 17] などがある。

2.2 行動の多様性の考慮

本研究では行動の多様性にも焦点を当てている。行動の多様性とは、消費者の行動の幅広さについて特定要素への依存度を測ることで考慮したものである。消費者の行動の多様性を購買予測に考慮することについては、[新美 15] においてオンラインストアで購買された商品カテゴリや自社店舗への流入経路（リファラ情報）の幅広さを多様性指標として投入す

表 1: ある消費者 A,B が利用した店舗数の例

	消費者 A	消費者 B
利用総数	50 回	50 回
T 社	10 回	10 回
競合 A 社	10 回	37 回
競合 B 社	10 回	1 回
競合 C 社	10 回	1 回
競合 D 社	10 回	1 回
T 社利用の割合	20%	20%
特定ブランドへの依存度	低	高
HHI	2000	5888

ることで購買予測の精度が改善することが示されている。多様性指標の定式化には、経済指標としてある市場の寡占の度合いを表すハフィンダール・ハーシュマン指数（HHI）[Davis 09] を用いている。HHI は市場に参入する企業 i の市場シェア s_i の二乗和として次の式 1 のように表される。

$$HHI = \sum_{i=1}^{k} s_i^2 \quad (1)$$

HHI は一要素が完全に独占している場合に最大値 $100^2 = 10000$ を取る非負の値であるが、本研究では解析で扱いやすいよう 10000 で除し 0 から 1 の値をとる連続変数として投入している。

位置情報から得られる行動の多様性の例として、次の表 1 にある 2 人の消費者 A、B が期間中に訪問した店舗についての集計を示すと、消費者 A、B はともに自社店舗の利用の割合は同一であるものの、全体として特定ブランドへの依存度が大きく異なっている。市場全体の利用回数、自社利用回数、自社 SOW が同一な場合でも HHI を用いることで購買行動の異質性を把握することができる。消費者 A が特定のブランドにロイヤルティを持たずに様々なスーパーマーケットを利用している一方で、消費者 B の場合には競合 A 社を中心に利用しつつ T 社を補助的に利用しているような状況が推察できることから、多様性変数を用いることで顧客のターゲティングをより最適化することが可能である。

また本研究と同様にマーケティングデータに対して深層学習手法を適用する場合には、企業規模が増大するにつれてカテゴリカル変数（例として「最もよく利用した時間帯」「最もよく購買された商品カテゴリ」等）の増加により予測モデルの大規模化に繋がりうる。これは結果としてデータ解析の構築コストの増大や意思決定の遅延をもたらす可能性もある。[新美 17] ではこのような大規模なカテゴリカル変数の代替として多様性変数を投入することで、予測器の次元数を大きく削減した上で高い予測精度を維持した効率的な解析が実現できることが示されている。

3. 解析手法

3.1 使用データと分析対象

本研究では、実店舗における購買行動の予測を目的として九州地方を中心に展開する大手スーパーマーケットチェーン T 社の消費者を分析対象としている。本研究では 2 つのデータを組み合わせて解析に用いているが、まずは T 社から提供いただいた 2017 年 8,9,10 月の ID-POS データを用いて、8,9 月の購買から 10 月の購買を予測する解析を行う。したがって 2017 年 8,9 月に店舗で購買を行った消費者を T 社における自社顧客と定義している。ただし位置情報とのマッチングのため、8 から 10 月の期間全体にわたって最低でも 3 回の購買が行われていないものについては対象から除外している。

2 つ目のデータとして本研究では位置情報データを活用している。この位置情報データは一般に消費者がスマートフォンで特定の複数のアプリケーションを利用した際にバックグラウンドで取得された位置情報がサーバ上で集計されるものである。ただし GPS に基づく正確な位置情報ではなく、地域を特定の距離で区切った仮想の区画であるジオフェンス内への進入・退出を記録したものである。これを用いて T 社の各店舗および競合すると考えられる近隣の複数の店舗のジオフェンスを特定し、収集した位置情報から消費者の店舗訪問の履歴を得られることを確認した。最後に T 社店舗での購買履歴とその時間帯に該当店舗に訪問していることを条件としたマッチングを実施し、ID-POS と位置情報の紐付けを行った。マッチングの方法については 3.2 節で詳述。

3.2 購買履歴と位置情報のマッチング

本研究で用いている購買履歴データと位置情報データは異なる消費者を対象として異なる手法で収集されたものであり、必然的に異なる ID を用いて集計されている。したがって解析に先立ちこれらデータの紐付けが必要となるため、ここで本研究でのマッチング方法について述べる。位置情報データでは特定ジオフェンスへの進入・退出の時間が、購買履歴データでは ID で継続的に特定の消費者の購買日時が追跡されていることから、決済を行った日時の前後 30 分以内に滞在情報が得られていることをもって同一購買とみなしている。次に位置情報と購買履歴の各 ID について、期間中に少なくとも 3 回以上購買を行っている顧客に対して、期間中全体の購買回数に対するこの一致回数の割合の多いものから順にマッチングさせた。結果としてマッチングデータは 1,000 程度が得られ、そこからさらに購買条件に従い顧客を抽出した。

3.3 予測手法

予測には研究課題として近い手法である [新美 17] と同様に FFNN による 3 層の一般的な深層学習を用いた。複数の変数群を用意し、使用する特徴量を差し替えることによる予測精度の変化について確認し、位置情報の多様性を投入して予測精度が改善することをもって位置情報の多様性の有効性を示す。

解析結果については当日示す。

参考文献

[Davis 09] Davis, P. and Garcés, E.: *Quantitative techniques for competition and antitrust analysis*, Princeton University Press (2009)

[新美 15] 新美 潤一郎, 星野 崇宏: ユーザ別アクセス・パターン情報の多様性を用いた顧客行動の予測とモデリング, 応用統計学, Vol. 44, No. 3 (2015)

[Svozil 97] Svozil, D., Kvasnicka, V., and Pospichal, J.: Introduction to multi-layer feed-forward neural networks, *Chemometrics and intelligent laboratory systems*, Vol. 39, No. 1, pp. 43–62 (1997)

[Vieira 15] Vieira, A.: Predicting online user behaviour using deep learning algorithms, *CoRR*, Vol. abs/1511.06247, (2015)

[新美 17] 新美 潤一郎, 星野 崇宏: 顧客行動の多様性変数を利用した購買行動の予測, 人工知能学会論文誌, Vol. 32, No. 2, pp. B–G63_1–9 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

時系列勾配ブースティング木による分類学習 金融時系列予測への応用

Classification by Time-Series Gradient Boosting Tree

中川 慧 *1*2 Kei Nakagawa　今村 光良 *1*3 Mitsuyoshi Imamura　吉田 健一 *2 Kenichi Yoshida

*1野村アセットマネジメント株式会社 Nomura Asset Management Co., Ltd

*2筑波大学大学院 ビジネス科学研究科 University of Tsukuba Graduate School of Business Sciences

*3筑波大学大学院 システム情報工学研究科 University of Tsukuba Graduate School of Systems and Information Engineering

We propose a time-series gradient boosting decision tree for a data set with time-series and cross-sectional attributes. Our time-series gradient boosting tree has weak learners with time-series and cross-sectional attribute in its internal node, and split examples based on dissimilarity between a pair of time-series or impurity between cross-sectional attributes. Dissimilarity between a pair of time-series is defined by dynamic time warping method or in financial time-seires by indexing dynamic time warping method. Experimental results with stock price prediction confirm that our method constructs interpretable and accurate decision trees.

1. はじめに

株価を予測するために様々な方法が提案されている。主に時系列解析に基づくもの、クロスセクショナルなデータを特徴量として使用した、いわゆる機械学習に基づくものの 2 種類が代表的である。時系列分析手法は、データ自身の過去の変動パターンを線形方程式でモデル化するという意味で恣意性の低い方法である。一方で、機械学習による予測は時系列解析と異なり、特徴量をうまく選択する必要がある。従来のファイナンス分野で、株価を予測するための特徴量としてはバリューやモメンタム等がある [Fama 92, Fama 93, Asness 13]。最近ではアナリストレポートやニュースといったテキストデータを、特徴量として用いる試みが模索されている [Takahashi 06, 工藤 17]。

本論文では、時系列およびクロスセクションの属性を持つデータセットに対する時系列勾配ブースティング決定木を提案する。勾配ブースティング木などのツリーモデルは一般に可読性に優れており、説明責任という観点から実務への適用がしやすい。勾配ブースティング木は決定木を多段に組み合わせることで確率的な予測値を導出する手法であり、非常に高い精度を実現できる。しかしながら従来の決定木学習手法は時系列属性を想定していないため、株価のような時系列データを含むデータ集合に適用する場合、データの前処理が必要となる。最も単純な前処理の方法として、時系列データを計測値の平均値や標準偏差などの時系列データのモーメントで置き換える方法が考えられる。ただし、この方法は時系列データの構造、すなわち形状を無視しており、例えば形が大きく異なる時系列データを同一視してしまう欠点がある。

そこで、勾配ブースティング・アルゴリズムに用いる弱学習器として、以上の問題点を克服した時系列データの形を陽に扱い時系列データ全体を対象とする [Yamada 03] の基準例分割テストによる分割を用いた時系列決定木を使用する。さらに時系列属性以外の通常の属性も取り込んだ時系列決定木を弱学習器として使用することを提案する。時系列決定木は相違度基準として動的時間伸縮法 (DTW) を用いているが、我々は金融時系列の分割の際には、時系列データの類似度を計測する手法として、[Nakagawa 17] で提案された Indexing Dynamic Time Warping(IDTW) を使用する。時系列ペアに対し距離が定義されていることから、[Nakagawa 17] のように決定木学習手法ではなく、最近傍法 (k-NN や k^*-NN[Anava 16]) を用いることも考えられる。ただし、最近傍法は怠惰学習 (lazy learning) であるために分類モデルが存在せず、学習結果が分かりにくいという欠点がある。以上の時系列決定木を弱学習器として用いた時系列勾配ブースティング木を用いて、日本株式を対象に実証分析を行う。また金融時系列以外の時系列データセットについての実験評価については [今村 18] で議論し、本論文では手法の提案と金融時系列への応用を報告する。

連絡先: 中川 慧，野村アセットマネジメント株式会社，〒 103-8260 東京都中央区日本橋一丁目 12 番 1 号

2. 先行研究

勾配ブースティングは、ブースティング・アルゴリズムの一種である。ブースティングとは、集団学習の枠組みの一つで、複数の弱学習器を統合して全体の学習器を構成する手法である。弱学習器としては、決定木が使われることが多い。ブースティングにおける弱学習器として決定木を用いることは、データの外れ値に強い、数値変数と離散変数や欠損値などを扱いやすいなどの利点がある。本研究では弱学習器として、[Yamada 03] の時系列決定木に、時系列以外の通常の特徴量も追加した決定木を使用することを提案する。そこで、以下では時系列決定木および勾配ブースティング木についてレビューする。

2.1 時系列決定木

時系列決定木は、内部ノードに基準となる属性時系列を持つ決定木であり、基準例分割テストまたはクラスター分割テストによってサンプルを分割していく [Yamada 03]。[山田 03] の実験からクラスター分割テストより基準例分割テストの結果が良好であることから、本稿では基準例分割テストを分割手法として用いる。

データセット D は、n 個のサンプル $\{x_i, i = 1, ..., n\}$ を含み、各サンプル x_i はそれぞれ m 個の属性 $\{a_j, j = 1, ..., n\}$ とクラスラベルまたは目的変数 $\{y_i, i = 1, ..., n\}$ を持つ。各属性 a_j は、時間順に値を並べた時系列データ又は系列を持たない通常

の連続値、名目属性を持つ。ここで $x(a_j)$ は一つのサンプルであり、基準例と呼ぶ。基準例分割テストは、サンプル x_i の属性 a_j に関する時系列データを $x_i(a_j)$ で表すと、データセット全体を、$DTW(x(a_j), ex_i(a)) < \theta_i$ を満たすデータセットから構成される集合 $L(x, a_j, \theta_i) = \{(y, x)|DTW(x(a_j), x_i(a_j)) < \theta_i\}$ とそれ以外の集合 $R(x, a_j, \theta_i) = \{(y, x)|DTW(x(a_j), x_i(a_j)) \geq \theta_i\}$ に分割する。ここで $DTW(x, y)$ は、動的時間伸縮法 (DTW) に基づいた類似度を表す。DTW は時系列データにおける一点のデータをもう片方の時系列データにおける複数点のデータに対応づけられるため、時間方向の非線形な伸縮を許容する。このため、長さが異なる時系列データのペアに適応できる上に、結果が人間の直観により合致する。具体的なアルゴリズムは、Algorithm1 の通り。

基準例分割テストはサンプルが基準例 $x(a_j)$ と DTW で計測して、類似しているか、あるいは類似していないかという基準で閾値 θ を変えながら分割する。当該分割の評価基準としては、利得比基準など通常の決定木で用いられている方法が使える。

Algorithm 1 DTW distance

```
procedure DTW(x, y)
                                   ▷ Initialize matrix D
    Var D[N, M]
    D[1, 1] = 0
    for i = 2 to N do
        for j = 2 to M do
            D[i, j] = ∞
        end for
    end for
                                   ▷ Calculate DTW distance
    for i = 2 to N do
        for j = 2 to M do
            D[i, j] = d(x[i − 1], y[j − 1])
                      +min(D[i, j − 1], D[i − 1, j], D[i − 1, j − 1])
        end for
    end for
    return D[N, M]
end procedure
```

2.2 勾配ブースティング木

[Friedman 01] は、ブースティングの各ステップで構築するモデルに回帰ツリーを用いる勾配ブースティング木（gradient boosting tree）という手法を考案した。これは特徴量 (入力) を x、ラベルを y として、弱学習器 $f_m(x), m = 0, ..., M$ に対して全体の学習器 $F(x)$

$$F(x) = f_0(x) + f_1(x) + ... + f_M(x) \quad (1)$$

を損失関数 $L(y, F(x))$ が最小になるように弱学習器 $f_m(x)$ を逐次的に学習し統合する。つまり、最初に学習器 $F_0(x) = f_0(x)$ が与えられるとし、m ステップ目の学習では m 個の学習器からなる全体の学習器を損失関数 $L(y, F(x))$ が最小になるように弱学習器を決める。

$$F_m(x) = F_{m-1}(x) + f_m(x) \quad (2)$$

具体的なアルゴリズムは以下の通りである。勾配ブースティング木の特徴は、その時点でのモデルの予測値と結果の残差を目的変数とした回帰ツリーを構築し、そのツリーモデルの予測値による値を加算することにある。

Algorithm 2 Gradient Boosting Tree

```
procedure Gradient Boosting Tree(y, x)
                                   ▷ Initialize F_0 with a constant
    F_0(x) = Σ_{i=1}^{N} arg min_c (y_i, c)
    for m = 1 to M do
        r_im = −[ L(y_i, F(x_i)) / F(x_i) ]_{F(x)=F_{m−1}(x_{i−1})}
                                   ▷ Fit a decision tree to predict targets r_im
        f(x) = FitTree(r_im, x)
        ρ_m = arg min_ρ Σ_{i=1}^{N} L(y_i, F_{m−1}(x) + ρf(x))
        F_m(x) = F_{m−1}(x) + γρ_m f(x)
    end for
    return F(x) = Σ_{m=1}^{M} F_m(x)
end procedure
```

3. 提案手法 - 時系列勾配ブースティング木

前章で確認した [Yamada 03] の基準例分割テストによる時系列決定木において、時系列属性以外の属性、通常の連続値、名目属性も追加する。属性が時系列データの場合には基準例分割テストによって分割を行い、時系列以外の場合には、通常の決定木と同様の基準 (不純度) で分割を行う。具体的なアルゴリズム (StandardExSplit with Cross-Sectional Data) は、Algorithm3 の通り。$H(\bullet)$ は不純度を表す関数で、本稿では、よく使用される Gini 係数を用いる。

Algorithm 3 StandardExSplit with Cross-Sectional Data

```
procedure Split({y_1, ..., y_n}, {x_1, ..., x_n})
    for each samples x_i do
        for each attribute a_j do
            if attribute a_j is time-series then
                for each samples x_k do
                    L(x, a_j, θ_i)
                    = {(y, x)|DTW(x_i(a_j), x_k(a_j)) < θ_k}
                    R(x, a_j, θ_i)
                    = {(y, x)|DTW(x_i(a_j), x_k(a_j)) ≥ θ_k}
                    G(L, R, θ_i)
                    = |L|/|L+R| H(L) + |R|/|L+R| H(R)
                end for
            else
                L(x, a_j, θ_i) = {(y, x)|x(a_j) < θ_i}
                R(x, a_j, θ_i) = {(y, x)|x(a_j) ≥ θ_i}
            end if
            G(L, R, θ_i) = |L|/|L+R| H(L) + |R|/|L+R| H(R)
        end for
    end for
    θ* = arg min_θ G(L, R, θ_i)
    return best split θ*
end procedure
```

この分割アルゴリズムを用いた時系列決定木を勾配ブースティング木の弱学習器として利用する。その際に金融時系列が時系列属性として与えられた場合、時系列データのペアに関する相違度計算法には [Nakagawa 17] で提案された、株価予測

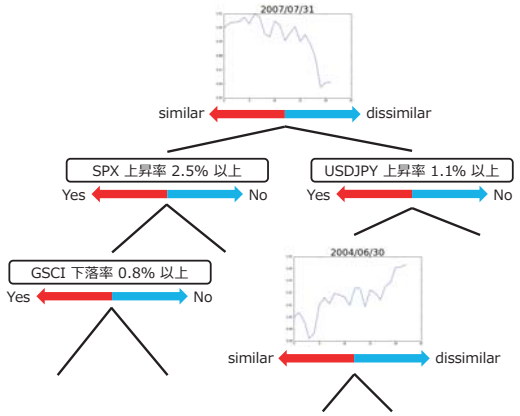


図 1: Time-Series Tree with Cross-sectinal Data

に有効な Indexing Dynamic Time Warping(IDTW) を用いる。IDTW は金融実務の観点から、月末の値を 1 として、日々の価格推移を月末との比を使って表した系列に対して DTW を適用する (Algorithm4)。

Algorithm 4 IDTW distance

procedure IDTW(x, y)
▷ Scaling data
  Var Ix, Iy ▷ Initialize Ix,Iy
  $Ix[1] = 1, Iy[1] = 1$
  for $i = 2$ to N **do**
    $Ix[i] = Ix[i-1]\frac{x[i]}{x[i-1]}$
  end for
  for $j = 2$ to M **do**
    $Iy[j] = Iy[j-1]\frac{y[j]}{y[j-1]}$
  end for
▷ Apply DTW
  return $DTW(Ix, Iy)$
end procedure

構築される時系列決定木のイメージが図 1 である。時系列に対して、通常の特徴量による分類だけでなく、時系列が自身の過去の形状と類似しているか、類似していないか、という形状に基づいた分類が行われている。そのため、どのような分類が行われるのか直感的な理解が可能となる。

4. 実証分析

前章で提案した時系列勾配ブースティング木 (TS-GBTree) を用いて、日本株を対象に実証分析を行う。使用するデータは TOPIX 指数およびクロスセクションの情報として、指数外の情報である為替レートのドル円、米国株式指数 SP500、商品総合指数である SPGCSI 指数の月間収益率を用いる。使用したデータの期間は、1985 年 1 月から 2018 年 1 月までであり、学習期間は始点を 1985 年 1 月に固定し、これより予測を実施する直近の月までとして、2005 年 1 月から 2018 年 1 月までを検証期間とした。分析にあたっては、時系列勾配ブースティング木を用いて、毎月末に予測を行い、上昇予測の場合は月末に 1 単位 TOPIX を買い、下落予測の場合には 1 単位 TOPIX を売り、これを 1 か月毎に繰り返し予測を行った。比較対象は時系列決定木 (TS-Tree)、先行研究の IDTW+k^*-NN(k^*-NN) を用いた。

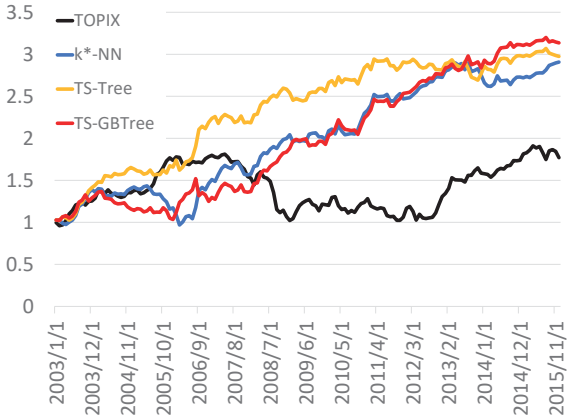


図 2: Total Return of TOPIX,k*-NN,Time-Series Decision Tree and Time-Series Gradient Boosting Tree.

表 1: Precision, Recall and F-Score of k*-NN

	precision	recall	f1-score	support
Down	0.62	0.52	0.57	65
Up	0.69	0.77	0.73	91
avg/total	0.69	0.77	0.73	156

表 2: Precision, Recall and F-Score of Time-Series Decision Tree(with CS)

	precision	recall	f1-score	support
Down	0.49	0.65	0.56	65
Up	0.67	0.52	0.58	91
avg/total	0.60	0.57	0.57	156

表 3: Precision, Recall and F-Score of Time-Series Gradient Boosting Decision Tree(with CS)

	precision	recall	f1-score	support
Down	0.55	0.65	0.59	65
Up	0.71	0.62	0.66	91
avg/total	0.64	0.63	0.63	156

表 4: Accuracy of Each Method

	k*-NN	TSTree	TSGBTree
TS	66.67%	54.49%	66.02%
TS+CS	-	57.05%	62.82%

時系列勾配ブースティング木と比較対象である k^*-NN、時系列決定木と TOPIX の累積収益率の推移が図 2 である。時系列勾配ブースティング木は検証期間前半は TOPIX 指数を下回るものの、リーマンショック以降の 2009 年から安定して TOPIX を上回っている。収益性については、時系列勾配ブースティング木は k^*-NN に対しては 20%程度、時系列決定木に対して 100%程度上回った。一方で、表 1 が k^*-NN の、表 2 が時系列決定木の、表 3 が時系列勾配ブースティング木の Precision、Recall および F スコアである。表から勾配ブースティング木と時系列勾配ブースティング木で、Precision は 0.04%、Recall は 0.06%、F スコアは 0.06%の性能差があった。一方で、勾配ブースティング木は k^*-NN との比較では、Precision、Recall、F スコア全て下回った。

表 4 は時系列決定木に対して、勾配ブースティングを用いること、およびクロスセクションのデータ (CS) を時系列決定木に組み込むことによる正答率の改善を表している。勾配ブースティングにより、精度が約 11%改善し、クロスセクションのデータを組み込むことで時系列決定木が約 2%改善している。しかしながら勾配ブースティング、クロスセクションのデータともに組み込むと、約 8%の改善にとどまった。当然ながら、TOPIX 指数に対して説明力を持つであろう特徴量は上記以外に多数あると考えられ、それらを組み込むことでさらなる精度改善が見込まれる。

5. まとめ

本論文では、時系列およびクロスセクションの属性を持つデータセットに対する時系列勾配ブースティング決定木を提案した。勾配ブースティング・アルゴリズムに用いる弱学習器として、時系列データの形を陽に扱い時系列データ全体を対象とする [Yamada 03] の基準例分割テストによる分割を用いた時系列決定木を使用した。加えて、時系列属性以外の通常の属性も取り込み、金融時系列データの類似度の計測に [Nakagawa 17] の IDTW を用いた。TOPIX 指数を対象とした実証分析により次のような結果が得られた。

- 時系列決定木に対して勾配ブースティングを用いることで予測精度が 11%程度改善。
- 時系列決定木に対してクロスセクションのデータを組み込むことで 2%程度予測精度が改善。
- 収益性については比較手法をすべて上回った。

一方で、クロスセクションのデータを時系列勾配ブースティング木に組み込むと、約 8%の改善にとどまった。当然ながら、TOPIX 指数に対して説明力を持つであろう特徴量は上記以外に多数あると考えられ、それらを組み込むことでさらなる精度改善が見込まれる。また、TOPIX 以外の株式指数や個別銘柄、あるいは債券や為替などの他のアセットへの適用が今後の課題である。

参考文献

[Anava 16] Anava, O. and Levy, K.: k*-nearest neighbors: From global to local, in *Advances in Neural Information Processing Systems*, pp. 4916–4924 (2016)

[Asness 13] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H.: Value and momentum everywhere, *The Journal of Finance*, Vol. 68, No. 3, pp. 929–985 (2013)

[Fama 92] Fama, E. F. and French, K. R.: The cross-section of expected stock returns, *the Journal of Finance*, Vol. 47, No. 2, pp. 427–465 (1992)

[Fama 93] Fama, E. F. and French, K. R.: Common risk factors in the returns on stocks and bonds, *Journal of financial economics*, Vol. 33, No. 1, pp. 3–56 (1993)

[Friedman 01] Friedman, J. H.: Greedy function approximation: a gradient boosting machine, *Annals of statistics*, pp. 1189–1232 (2001)

[Nakagawa 17] Nakagawa, K., Imamura, M., and Yoshida, K.: Stock Price Prediction using k*-Nearest Neighbors and Indexing Dynamic Time Warping, in *Artificial Intelligence of and for Business (AI-Biz 2017)* (2017)

[Takahashi 06] Takahashi, S., Takahashi, M., Takahashi, H., and Tsuda, K.: Analysis of stock price return using textual data and numerical data through text mining, in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 310–316Springer (2006)

[Yamada 03] Yamada, Y., Suzuki, E., Yokoi, H., and Takabayashi, K.: Decision-tree induction from time-series data based on a standard-example split test, in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 840–847 (2003)

[工藤 17] 工藤秀明, 永島淳, 宮崎義弘：自然言語処理技術を用いたアナリストレポートの実証分析: センチメントの変化と株式市場の反応について, 証券アナリストジャーナル = Securities analysts journal, Vol. 55, No. 9, pp. 66–77 (2017)

[今村 18] 今村光良, 中川慧, 吉田健一：ベンチマークデータを用いた時系列勾配ブースティング木の実験評価, 人工知能学会全国大会論文集 2018 年度人工知能学会全国大会 (第 32 回) 論文集社団法人 人工知能学会 (2018)

[山田 03] 山田悠, 鈴木英之進, 横井英人, 高林克日己 他：動的時間伸縮法に基づく時系列データからの決定木学習, 情報処理学会研究報告知能と複雑系 (ICS), Vol. 2003, No. 30 (2002-ICS-132), pp. 141–146 (2003)

テクニカル指標による株式投資の戦略木構築

Construction of Strategy using Binary Decision Diagram in Stock Investment using Technical Indicators

加藤 旺樹[*1]
Ohki Kato

穴田 一[*1]
Hajime Anada

[*1] 東京都市大学
Tokyo city university #1

In recent year, many researchers pay attention to stock trading using technical analysis. In the investment using technical analysis, we predict future price movements from patterns of past price movements and trade by using technical indicators to judge turning point and trends of the market. However, it is necessary to have expert knowledge and there is a problem that it is difficult to make a profit. Therefore, we construct investment strategy that can make a good profit using technical indicators.

1. はじめに

近年，テクニカル分析を用いた株式売買に関する研究が多く行われている．テクニカル分析を用いた投資では，相場のトレンドや転換点を判断するテクニカル指標を用いることで，過去の値動きのパターンから将来の値動きを予測し売買を行う．しかし，そのためには専門的な知識を必要とする上，利益を上げにくいという問題がある．

そこで，本研究ではテクニカル指標を用いた高確率で利益を生み出す投資戦略の構築を目的とする．

2. 既存研究

松村らは，戦略木としてテクニカル指標を用いた投資戦略を木構造で表現し，それぞれの個体がその戦略木に従い株式売買を行うモデルを構築した[松村 06] [松村 08]．遺伝的操作を用いることでより利益が高くなるように戦略木を進化させ，その戦略木の特徴を分析した．

2.1 テクニカル指標

テクニカル指標は株式の売買タイミングを判断するために使われる指標で，トレンド系，オシレータ系，出来高系の 3 つがある．トレンド系は株価の推移からトレンドの上昇や下降を判断する指標，オシレータ系は株価の推移からトレンドの転換点を判断する指標，出来高系は売買が成立した株数の推移からトレンドの転換点を判断する指標である[藤本 04][吉見 03][山中 15]．

2.2 戦略木

戦略木は，図 1 の例のように複数の非終端ノードと終端ノードからツリー状に構成される．それぞれの数字がテクニカル指標や売買行動を表し，各ノードから 1 番上のノードまでのノード間を結ぶ枝の数を各ノードの深さとする．非終端ノードでは，テクニカル指標に基づく判定処理によって左右どちらかのノードを実行する．終端ノードでは，株式の売買行動を実行する．非終端ノードによる分岐を終端ノードにたどり着くまで繰り返すことで行動を決定する．表 1 に非終端ノードの例と終端ノードを挙げる．

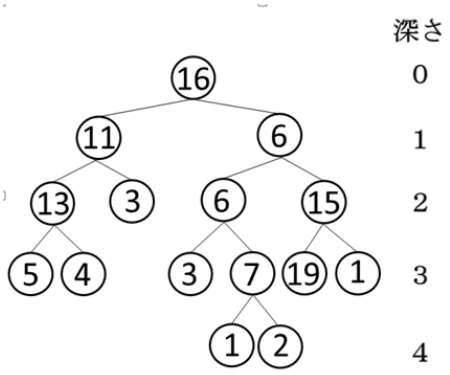


図 1 戦略木の例

個体は，戦略木に従って当日の行動を決定する．一番上のノードからテクニカル指標に基づく判定処理により，左右どちらのノードを実行するのか決定していく．この操作を繰り返し，到達した終端ノードの行動を実行する．判定に使用するデータは，各指標に必要な日数分の始値，高値，安値，終値，出来高の日足データである．

表 1 ノードの例

種類	番号	内容
終端ノード	1	1 株購入，資金がないときは待機
	2	1 株売却，株式がないときは待機
	3	待機
	4	保有している株式をすべて売却
	5	資金が許す限り株式を購入
非終端ノード	6	前日の終値が前々日の終値を上回れば左のノードを実行，それ以外であれば右のノードを実行
	⋮	⋮
	34	前日の総資産が前々日の総資産を上回れば左のノードを実行，それ以外であれば右のノードを実行

2.3 シミュレーション

戦略木を個体の遺伝子として，交叉や突然変異，淘汰の操作

連絡先：加藤旺樹，東京都市大学，〒158-8557
東京都世田谷区玉堤 1-28-1

でより高い利益を生み出す個体を作り出していく.

2.3.1 初期個体生成

表 1 のノードの中からランダムに 1 つを選択する. そのノードが終端ノードであれば, 戦略木の生成を終了する. 非終端ノードであれば, その非終端ノードの左右にノードをランダムに選択して付け加える. この作業を, 全ての非終端ノードの左右に終端ノードが付け加えられるまで繰り返す. 初期個体の生成は, 初めからノード数の多い木になることを防ぐため, 深さ 4 には終端ノードを付け加える.

2.3.2 評価

個体を評価した値である適応度 F は総資産とする. 総資産とは, 取引後の個体の現金残高と保有株式の時価総額の全銘柄の総和である.

2.3.3 遺伝的操作

以下の操作(ⅰ)～(ⅲ)を順に行うことで, より適応度の高い個体を作り出す.

(ⅰ) 複製

現世代において適応度の高い上位 20%の個体を, そのまま次世代に残す.

(ⅱ) 交叉

現世代の全個体から適応度を用いたルーレット選択により親となる 2 個体を選択する. それぞれの個体からランダムに 1 つずつノードを選び, そのノード以下の部分木を交換することで子を 2 個体作り出す. これらの操作を繰り返すことで次世代に残す N 個体の 80%を作り出す.

(ⅲ) 突然変異

次世代に残す全ての個体について適応度の順位に基づいて突然変異を施すか決定する. 個体 i の突然変異率 P_i は次式で定義される.

$$P_i = (Rank_i - 1) \times 0.1 \quad (\%) \qquad (1)$$

ここで, $Rank_i$は個体 i の適応度の順位である. これにより, 順位の高い個体は突然変異する確率が低くなる. 突然変異を施す場合ランダムにノードを 1 つ選び, 終端ノードなら他の終端ノード, 非終端ノードなら他の非終端ノードにランダムに変化させる. これらの個体を次世代に残す.

2.3.4 シミュレーションの流れ

個体を N 個体用意する. 個体はそれぞれ戦略木と現金残高, 保有株式数の情報を持ち, 1 日 1 回行動する. 各個体は独立して, 戦略木に従い複数銘柄でそれぞれ取引を行う. 1 銘柄につき元金を 10 万円とし 2 年間の取引を行う. 取引は過去のデータを用いて実市場と同じ値動きを再現した市場で行い, 株式の注文は注文を出した日の終値で成立させる. シミュレーションは以下の流れで行う.

Ⅰ)初期個体生成
Ⅱ)評価
Ⅲ)複製
Ⅳ)交叉
Ⅴ)評価
Ⅵ)突然変異
Ⅱ)～Ⅵ)を, 終了条件を満たすまで繰り返す. Ⅴ)の評価は突然変異を行う際に個体の順位を用いているため, 複製した個体と交叉によって作り出された個体の順位を求めるために行う.

3. 提案手法

3.1 既存手法の問題点

既存手法には大きく 3 つの問題点がある.

1 つ目の問題点は, 交叉で親を選び出す際の総資産によるルーレット選択では, 総資産の増えた戦略木と減った戦略木で選択確率に差が出にくい点である.

2 つ目の問題点は, 戦略木によって行動を決定する際に同じ数字のノードを 2 回目以降実行する場合, そのノードでは左右どちらに分岐するのか決まっているため, 絶対に到達できないノードが生まれてしまう点である. 例えば, 図 1 では, 深さ 1 と深さ 2 に 6 番のノードがある. 深さ 1 にある 6 番で左のノードを実行する場合, 深さ 2 の 6 番も左のノードを実行することになり, その逆側の深さ 3 の 7 番のノード以下には到達できない. これらのノードは売買行動に関係なく適応度に影響を与えないが, ノード数を制限していないためこのようなノードが際限なく増えていく. これでは, 交叉の際, 適応度に影響を与える部分が選ばれにくくなるため, 進化が停滞しやすく効率的な学習が妨げられる.

3 つ目の問題点は, 交叉によってできた個体をそのまま次世代に残しているため, 適応度の低い個体が次世代に多く残る点である.

3.2 提案

本研究では, 1 つ目と 2 つ目の問題点を解決するため個体 i の適応度 F_i を次式のように定義した.

$$\mathrm{F}_i = \mathrm{Profit}_i - \alpha \times \mathrm{NodeNumber}_i \qquad (2)$$

ここで, Profit_iは個体 i の利益, $\mathrm{NodeNumber}_i$は個体 i の戦略木のノード数, α は戦略木のノード数に関する重みを表す.

適応度に利益を用いることで, 既存手法の総資産によるルーレット選択より戦略木ごとの選択確率に差が出るようにしている. また, 戦略木のノード数を引くことでノード数の多い戦略木の適応度を小さくし, ノード数の多い戦略木を生き残りにくくしている. しかし, このままでは適応度が負になることがありルーレット選択の選択確率を求めることができない. そこで, 適応度が負の場合, 戦略木の中で最も小さい適応度の絶対値を全ての戦略木の適応度に足すことで補正する.

さらに 3 つ目の問題点を解決するため次世代に残す個体を現世代の個体と交叉によってできた個体から適応度の順位を用いたルーレット選択によって選び出す. こうすることで, 個体の多様性を保ちつつ順位の高い個体を次世代に残りやすくする. 本研究では交叉によって N 個体を作り出し, そこに現世代の N 個体を合わせた 2N 個体から次世代の N 個体を決定する.

3.3 提案手法の流れ

提案手法によるシミュレーションは以下の流れで行う.

Ⅰ)初期個体生成
Ⅱ)評価
Ⅲ)交叉
Ⅳ)評価
Ⅴ)次世代に残す個体の選択
Ⅵ)突然変異
Ⅱ)～Ⅵ)を, 終了条件を満たすまで繰り返す. Ⅳ)の評価は次世代に残す個体を決定する際に適応度の順位を用いたルーレット選択を用いているため, 現世代の個体と交叉によって出来た個体の順位を求めるために行う.

4. 結果

本研究では，取引に東京証券取引所 1 部に上場している銘柄を使用した．使用したのは，酒造関連のアサヒ，キリン，サッポロ，IT 関連のヤフー，オービック，KDDI，消費関連のイオン，イズミの 8 銘柄である．シミュレーションを行う条件として，個体数 N =100，世代数 5000，試行回数 50 とし，α は 10 ～ 150 を 10 刻みで，戦略木の学習を行う学習期間とテストを行うテスト期間を表 2 のように設け，シミュレーションを行った．

表 2 取引期間

ケース	学習期間	テスト期間
1	2013 年 1 月～2014 年 12 月	2015 年 1 月～2016 年 12 月
2	2011 年 1 月～2012 年 12 月	2013 年 1 月～2014 年 12 月
3	2009 年 1 月～2010 年 12 月	2011 年 1 月～2012 年 12 月

図 2 にケース 1 におけるテスト期間の全個体の中で最も高い利益(best)を示す．この図は α = 50 において，50 試行を平均した結果である．縦軸は利益(万)，横軸は世代数を表し，オレンジ線は既存手法の best，青線は提案手法の best を表す．

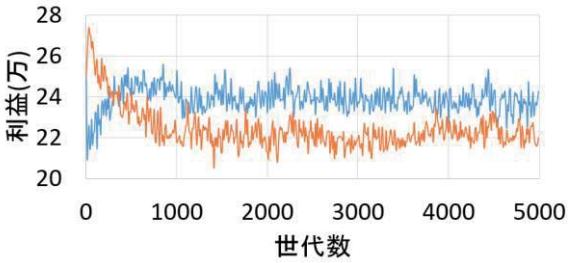


図 2 ケース 1 テスト期間の best

学習期間において提案手法ではどのケースにおいても best で既存手法を上回った．さらに，テスト期間において既存手法では世代数が進むにつれて best が下がっているが，提案手法では世代数が 0～1000 の間は上がっており既存手法を上回っている．ケース 1，3 において既存手法による戦略木の学習では世代が進むにつれてテスト期間の best の上昇を確認することができなかったが，提案手法では確認することができ，学習によって資産を増やすことができる戦略を獲得できたと考えている．これは提案手法の戦略木では，戦略木のノード数を制限したことで交叉によって適応度に影響を与える部分が交換されやすくなっている．その結果，効率的な学習ができ，より適応度の高い戦略木ができやすくなったからだと考えられる．

図 3 にケース 2 におけるテスト期間の全個体の中で最も高い利益(best)を示す．この図は α = 50 において，50 試行を平均した結果である．縦軸は利益(万)，横軸は世代数を表し，オレンジ線は既存手法の best，青線は提案手法の best を表す．

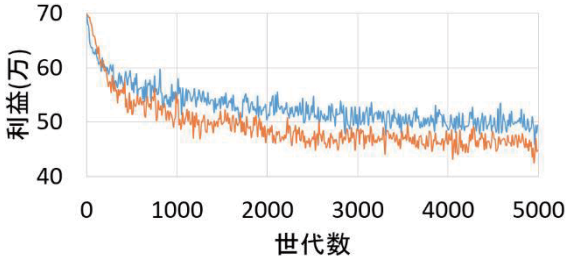


図 3 ケース 1 テスト期間の best

ケース 2 のテスト期間では提案手法，既存手法のどちらにおいても世代が進むにつれて best が下降してしまった．これは学習期間に値上がり幅の大きい上昇トレンドの銘柄が存在するため，上昇トレンドのみで資産が増えるように戦略木が進化してしまいテスト期間の株価変動に対応できなくなったからだと考えられる．

図 4 にケース 1 におけるその世代の戦略木の平均ノード数を示す．この図は α = 50 において，50 試行を平均した結果である．縦軸は平均ノード数，横軸は世代数を表し，オレンジ線は既存手法による戦略木の平均ノード数，青線は提案手法による戦略木の平均ノード数を表す．

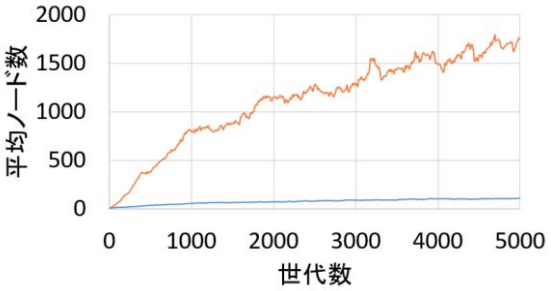


図 4 ケース 1 平均ノード数

最終世代において，既存手法による戦略木の平均ノード数は約 1700 であるが，提案手法による戦略木の平均ノード数は約 100 であり，提案手法では既存手法による戦略木よりはるかに少ないノードで戦略木を構築することができている．提案手法では，ケース 2，3 においても同様に既存手法による戦略木より少ないノードで戦略木を構築することができた．

表 3 にケース 1 の学習期間で最も総資産が高かった戦略木のノード数と無効なノードの割合を表 3 に示す．

表 3 無効なノード

	既存手法	提案手法
ノード数	787	47
無効な割合	76%	28%

既存手法による戦略木はノード数 787 に対して無効な割合が 76% と戦略木を構築するノードの約 4 分の 3 が無効なノードとなっているが，提案手法による戦略木はノード数 47 に対して無効な割合が 28% と既存手法より無効なノードを大幅に減らすことができている．

5. 今後の課題

既存手法，提案手法のどちらにおいても学習期間に偏った変動の銘柄が存在すると，その変動時のみでしか稼げないように戦略木が進化してしまい，他の株価変動に対応できなくなっている．様々なパターンの株価変動に対応する戦略木を構築するため，学習期間において様々な変動で利益をあげている戦略木を評価する方法の検討が必要である．

提案手法では，既存手法に比べ戦略木の無効なノードを大幅に減らすことができたが，無効なノードが残っており完全に削除することができていない．戦略木から全ての無効なノードを削除する方法の検討が必要である．

これらを検討した結果についても発表を行う．

参考文献

[松村 06] 松村幸輝，国屋美敬，木村周平:“遺伝的プログラミング手法に基づくエージェントベーストレーダモデル”情報処理学会論文誌，Vol.47,No9,pp.2869-2886，2006

[松村 08] 松村幸輝:“進化計算手法にクラスタリングを応用した株式投資の戦略木最適化モデル”情報処理学会論文誌, Vol.49,No1,pp.457-475, 2008

[藤本 04] 藤本壱:株価チャート分析大全, 自由国民社, 2004.

[吉見 03] 吉見俊彦:決定版チャート分析の真実, 日経ラジオ社, 2003.

[山中 15] 山中康司:相場で勝つための基本と実践 テクニカル指標の読み方・使い方, 2015.

[Koza 92] Koza, J:Genetic Programming, MIT Press, 1992

[2J3] [General Session] 13. AI Application

座長:小澤 順（産業技術総合研究所）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room J (2F Royal Garden B)

[2J3-01] Objective evaluation for the cystoscopic diagnosis of bladder cancer using transfer learning

○Yutaro Hoshino[1,2], Hidenori Sakanashi[2], Masahiro Murakawa[2], Nagatsugu Yamanouchi[1], Hirokazu Nosato[1,2] （1. Graduate School of Toho University, 2. Artificial Intelligence Research Center, National Institute of Adanced Industrial Science and Technology）

3:20 PM - 3:40 PM

[2J3-02] Estimation of Depressive Tendency Based on Lifestyle and Constitution

○Shiori Yamaguchi[1], Hiroki Tanaka[1], Hayato Maki[1], Shigehiko Kanaya [1,2], Nobutaka Suzuki[3], Satoshi Nakamura [1,2] （1. Nara Institute of Science and Technology, Graduate School of Science and Technology, 2. Nara Institute of Science and Technology, Data Science Center, 3. Kanazawa University, Graduate School of Medical Sciences）

3:40 PM - 4:00 PM

[2J3-03] Study on Measurement of Pelvic Angle from Point Cloud

○Emiri Sakiyama[1], Atsushi Hiyama[2,3], Sohei Wakisaka[3], Atsushi Izumihara[3], Masahiko Inami[3] （1. The Faculty of Engineering, The University of Tokyo, 2. Center for Advanced Intelligence Project, RIKEN, 3. Reseach Center for Advanced Science and Technology, The University of Tokyo）

4:00 PM - 4:20 PM

[2J3-04] Denoising autoencoder-based modification method for RRI data with premature ventricular contraction

○Shota Miyatani[1], Koichi Fujiwara[1], Manabu Kano[1] （1. Kyoto University）

4:20 PM - 4:40 PM

[2J3-05] Consideration of relations sit-to-stand movement and strength of lower-limbs

○Takahiro Hiyama[1], Yoshikuni Sato[1], Jun Ozawa[2] （1. Panasonic, 2. Advanced Industrial and Technology）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

転移学習による膀胱癌の内視鏡診断の客観的評価

Objective evaluation for the cystoscopic diagnosis of bladder cancer using transfer learning

星野勇太朗[*1,2] Yutaro Hoshino　坂無英徳[*2] Hidenori Sakanashi　村川正宏[*2] Masahiro Murakawa

山内長承[*1] Nagatsugu Yamanouchi　野里博和[*1,2] Hirokazu Nosato

[*1] 東邦大学大学院理学研究科
Graduate School of Science, Toho University

[*2] 産業技術総合研究所人工知能研究センター
Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology

This paper proposes an objective evaluation method for cystoscopic diagnosis of bladder cancer based on transfer learning using pre-trained DCNN (Deep Convolutional Neural Network) model. In the proposed method, lesion detection with comparatively fewer cystoscope images was realized by using the DCNN model pre-trained with a large amount of general images as fixed feature extractor and by learning the extracted feature vectors with subsequent classifiers. In this paper, in order to verify the effectiveness of the proposed method, experiments using actual bladder cystoscopic images were performed. As a result of the experiment, the proposed method achieved 95.7% in sensitivity and 93.3% in specificity in the two-class classification of normal and flat lesions which are difficult to distinguish, and showed the effectiveness for the cystoscopic diagnosis of bladder cancer.

1. はじめに

現在，膀胱に発生する癌は，一般的に膀胱内視鏡によって検査が行われる．膀胱内視鏡検査は，尿道から内視鏡を挿入することで尿道や膀胱を直接観察し，内部に異常がないかを判断する検査である．この検査において膀胱内壁に腫瘍が発見された場合，経尿道的膀胱腫瘍切除術(TUR-BT: Transurethral Resection of Bladder Tumor)によって取り除かれるが，術後 2 年以内の再発率は 50%以上といわれている．医師の知識や経験の差による検査精度のばらつきが生じており，最新の狭帯域光観察(NBI: Narrow Band Imaging)や光力学診断(PDD: Photodynamic Diagnosis)を用いた検査においても，感度 80%~100%，特異度 60%[1]とされている．特に小径の腫瘍や平坦型腫瘍，隆起病変に付随する平坦病変の広がりを同定することが困難で，新しい技術による診断精度の向上が期待されている．

近年，人工知能応用の流れは医療分野にも広がっており，診断精度を向上させる技術として，医療画像内の病変検出を目的とした機械学習を用いた診断支援技術の研究開発が行われてきた[2)3)]．しかし，人工知能を医療分野に応用する場合，複雑な医療画像の状態を捉える特徴量抽出手法の設計や，深層学習を適用する場合の十分な学習データ量の確保が難しい問題がある．特に癌の症例を異常データとして網羅的に大量に収集することは難しく，とりわけ日本では，病変画像の収集方法の整備が遅れており，学習用の対象画像を十分に集めることが困難である．

そこで本稿では，既存の一般物体用認識モデルを特徴抽出器として転用する客観的評価手法を提案する．提案手法では，複雑な膀胱内視鏡画像の状態を捉えるために，大量の一般画像で学習済みの認識モデルを特徴抽出器として使うことで，腫瘍を見分けるため分類器のみ学習を行う．これにより，分類器の構造がシンプルな場合，比較的少量の学習データでの学習により検出が可能になる．本稿では，提案手法による腫瘍検出の可能性を検証するため，実際の膀胱内視鏡検査画像を用いて，腫瘍病変画像と平坦病変画像に対し，正常画像との分類実験を行った．

連絡先：星野　勇太朗，　東邦大学大学院　理学研究科，
6517007h@st.toho-u.ac.jp

2. 提案手法

本研究では，複雑な膀胱内視鏡画像の状態を捉える特徴抽出手法として，学習済み深層学習モデルである Inception-v3[4)]を用いた．Inception-v3 は，Google 社が開発した画像認識用の大規模な CNN（畳み込みニューラルネットワーク）で，ImageNet の画像（一般画像 120 万枚）を用いて 1000 クラス分類を行えるように学習されている．TensorFlow 等の人工知能ライブラリにも実装されており，誰でも気軽に利用可能な学習済みモデルである．

提案手法では，入力された画像が Inception-v3 のネットワークを通して情報が最も圧縮される部分である Average Pooling 層の出力（2048 次元ベクトル）を特徴量として取り出し，後段の分類器に渡す．これにより，前段には一般画像で学習済みの CNN モデルをそのまま特徴抽出器として用いることで，後段の分類器のみ学習すればよく，シンプルな構造の分類器を用いる場合，比較的少ない量の対象画像（膀胱内視鏡画像）でも学習が可能となる．

本稿では，膀胱内視鏡画像からの病変検出に対する提案手法の有効性を確認することを目的として，後段の分類器には，非線形分類器として代表的なサポートベクターマシン(SVM: Support Vector Machine)と多層パーセプトロン(MLP: Multi-Layer Perceptron)の 2 つを用いて検証した．

3. 検証実験

3.1 実験の概要及び目的

本稿では，実際の膀胱内視鏡画像に対し，Inception-v3 ベースの特徴抽出器を用いて取り出した特徴量を，2 つの分類器(SVM, MLP)において学習・テストを行い，その結果を評価した．実験では，正常画像（図1a）に対し，比較的見分け易い腫瘍画像（図1b）の検出実験と，見分け難い平坦病変画像（図1c）の検出実験により検証した．

3.2 対象画像

学習・テスト用画像として，筑波大学附属病院で施行された膀胱内視鏡画像から 400x300pixels で切り出した正常 135 枚，腫瘍 177 枚，平坦病変 48 枚を用いた．腫瘍と平坦の病変画像は，病変部位を 80%以上含むように切り出されている．

3.3 評価方法

提案手法の有効性を検証するために各2クラス分類実験における感度（異常を異常として検出する割合）と特異度（正常を正常として検出する割合）を算出し，ROC(Receiver Operating Characteristic)曲線から性能評価指標である AUC(Area Under the ROC Curve）値を求めて比較した．また，性能指標として Youden 指数（感度+特異度-1）が最大となる閾値での感度，特異度を比較した．なお，各実験では，学習とテスト画像を入れ替えて 5 試行する 5-fold Cross-Validation を行った．表 1 に 1 試行での学習・テストに用いた画像枚数を示す．

4. 結果

表 2 および図2に各実験結果を示す．腫瘍／正常分類では，SVM で AUC=0.986, MLP で AUC=0.99 を達成した．また，平坦／正常分類でも SVM で AUC=0.968，MLP で AUC=0.967 を達成し，SVM, MLP のいずれの分類器でも，膀胱内視鏡画像の病変検出に対する提案手法の有効性を確認した．

5. 結論

本稿では，一般画像を用いて学習した深層学習モデルを特徴量抽出器として用いることで，比較的少ない画像の学習でも膀胱内視鏡画像から病変を検出可能な客観的評価手法を提案し，腫瘍画像と平坦画像の検出実験によりその有効性を確認した．これにより，これまで人工知能を応用した診断支援が実現していない膀胱癌の内視鏡診断においても，学習用画像データを大量に収集することなく，診断と治療の質を高める可能性が示唆された．今後は腫瘍病変のバリエーションや，データ量を増やして実験を行うことにより，より実用的な病変検出システムについての研究を進めていく．

表1　各実験に用いた画像枚数

	腫瘍検出		平坦病変検出	
	学習	テスト	学習	テスト
病変画像（腫瘍/平坦）	143	34	38	10
正常画像	107	28	48	12

表2　実験結果

	腫瘍検出		平坦病変検出	
	SVM	MLP	SVM	MLP
AUC	0.986	0.99	0.968	0.967
Youden 指数最大	0.886	0.919	0.891	0.874
感度	0.931	0.949	0.957	0.957
特異度	0.955	0.97	0.933	0.917

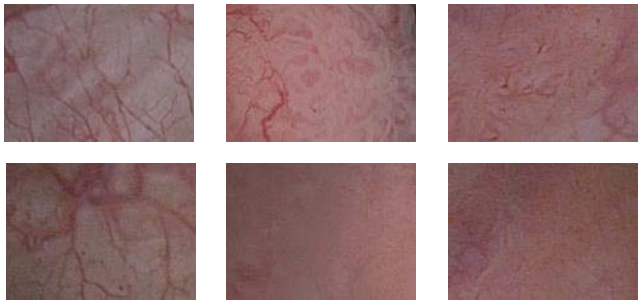


図1　膀胱内視鏡画像のサンプル

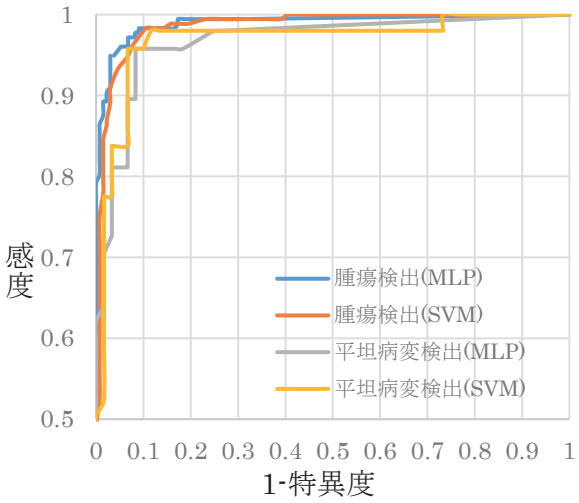


図2　実験結果の ROC 曲線

謝辞

本研究を進めるにあたり，研究にご協力いただき，貴重な膀胱内視鏡画像データをご提供いただきました，筑波大学医学医療系腎泌尿器外科西山博之教授，池田篤史病院講師に深く感謝いたします．なお，本研究の一部は JSPS 科研費 JP16678973, JP17K16775 の助成を受けたものです．

参考文献

1) [Kausch 10] Kausch I, Sommerauer M, Montorsi F, Stenzl A, Jacqmin D, Jichlinski P, Jocham D, Ziegler A, Vonthein R: Photodynamic diagnosis in non-muscle-invasive bladder cancer: a systematic review and cumulative analysis of prospective studies, Eur Urol., Vol. 57, No. 4, pp. 595-606, 2010.

2) [Nosato 14] H. Nosato, H. Sakanashi, E. Takahashi, M. Murakawa: An objective evaluation method of ulcerative colitis with optical colonoscopy images based on higher order local auto-correlation features, IEEE 11th International Symposium on Biomedical Imaging ISBI, pp. 89-92, 2014.

3) [Nuno 18] Nuno R Freitas, Pedro M Vieira, Estevão Lima, Carlos S Lima: Automatic T1 bladder tumor detection by using wavelet analysis in cystoscopy images, Physics in Medicine & Biology. Feb 2018, Vol. 63, No. 3: 035031, 2018.

4) [Christian 16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna: Rethink-ing the inception architecture for computer vision, CVPR, pp. 2818–2826, 2016.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

生活習慣や体質のデータを用いたうつ傾向の予測

Estimation of Depressive Tendency Based on Lifestyle and Constitution

山口 栞 *1 Shiori Yamaguchi
田中 宏季 *1 Hiroki Tanaka
真木 勇人 *1 Hayato Maki
金谷 重彦 *1*2 Shigehiko Kanaya
鈴木 信孝 *3 Nobutaka Suzuki
中村 哲 *1*2 Satoshi Nakamura

*1奈良先端科学技術大学院大学　先端科学技術研究科
Nara Institute of Science and Technology, Graduate School of Science and Technology

*2奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター
Nara Institute of Science and Technology, Data Science Center

*3金沢大学　医薬保健学総合研究科
Kanazawa University, Graduate School of Medical Sciences

This paper investigates the possibility of predicting depressive tendency by lifestyle data. Previous studies analyzed few aspects of lifestyle, whereas the current study utilizes multivariate analysis to examine multiple perspectives of lifestyle, such as social, sleeping, and dietary habits. We created a questionnaire including depressive tendency score (K6) as well as lifestyle, and recruited 987 participants, who answered it using a crowdsourcing service. Classification models were obtained using machine learning to classify the participants as high or low depressive tendency. Random forest classifier achieved 0.97 accuracy. Particularly effective features were chosen from Chinese medicine, and personality describing neuroticism.

1. はじめに

うつ病とは，抑うつ気分や興味の消失を特徴とする気分障害の一種である．世界中で約 3 億人が発症しており，毎年 80 万人の自殺の原因となっている [WHO 17]．日本でも約 100 万人の患者がおり深刻な社会問題になっている [MHLW]．

うつ病は，生理学的な指標を用いた診断方法は確立されていないが，北米心理学会が策定した診断ガイドラインである DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [Black 14] や，HAM-D (Hamilton's Rating Scale for Depression [Hamilton 80] などの質問票を用いて医師が診断を行う．DSM-5 は，うつ病の特徴を抑うつ気分，関心・歓心の低下，体重の著しい増減，不眠・過眠，精神運動の激越・遅滞，疲労，無価値・罪悪感，思考力・集中力・決断力の低下，死に関する反復思考などとしている．抑うつ気分または関心・歓心の低下が認められ，かつ 5 つ以上の特徴が 2 週間以上認められる場合にうつ病と診断される．

うつ病の早期発見のため，K6 [Kessler 02] や QIDS-SR [Rush 03] などの質問紙が存在し，医師によらずうつ傾向の自己チェックができる．K6 は，過去 30 日間における抑うつ気分や無価値であるという感覚ついて尋ねる 6 問から成る．0–24 点で評価され，点数が高いほどうつ傾向が高いとされる．13 点を閾値としてうつ傾向があるとみなすと，DSM-4 による診断と高い相関が見られる [Kessler 02]．

しかし，K6 は上述の通り抑うつ気分などについて明示的に尋ねる内容であるため汎用性に欠ける．また，うつ傾向のある人が，自らの状態について自覚的であるとは限らない．そのため，生活習慣など汎用的かつ客観的なデータを用いてうつ傾向を予測する技術が求められている．

うつ傾向と生活習慣との関係について調査した先行研究として，例えば人間関係の満足度 [Davidson 16] や，高齢者の孤食傾向とうつ傾向との関係が指摘されている [Kimura 12]．

これらの先行研究は，生活習慣のごく限られた側面のみに焦点を当ててうつ病との相関関係を調べている．また，うつ傾向の予測を目的としていないため，予測モデルの学習・開発・テストの手続きについて詳細に報告されておらず，未知のデータに対する汎化性能については議論されていない [Davidson 16] [Kimura 12]．

そこで本研究では，より多面的な生活習慣からうつ傾向を予測することを目的とする．そのため，本稿ではまずうつ傾向の予測に有効な特徴について調査した．

連絡先: 山口　栞, 奈良先端科学技術大学院大学, 奈良県生駒市高山町 8916-5 情報棟 B712, 電話：0743-72-5264 電子メール：yamaguchi.shiori.yk7@is.naist.jp

2. 質問紙の構成，クラウドソーシングによる生活習慣データの収集，および解析

2.1 質問紙の構成

質問紙は，うつ傾向と判定するための質問と，生活習慣等を尋ねる質問の 2 部からなる．うつ傾向判定のための質問として，K6 [Kessler 02] を用いた．

生活習慣等に関する質問は，中医学 [Zhu 08]，ストレス，認知機能検査，性格 (Big5) [Oshio 12]，コミュニケーション，基本データ (年齢，性別，職業など)，喫煙習慣，運動習慣，健康状態， 病歴， 睡眠 [Doi 00]，食生活の 12 カテゴリーから成る計 290 問を使用した．

これらの質問は，既存の質問集の中から，過去の文献でうつと関連性があるものを参考にし，また，医師や栄養士などの医療識者と相談した上で選択した．今回の研究では，ストレス，認知機能検査，中医学から K6 の質問に近い 1 問（憂鬱になるか問う質問）以外の，被験者全員が回答した質問を解析に用いた．

2.1.1 質問紙の順序効果の検証

質問紙を使用するにあたって，妥当性，信頼性，順序効果について確認する必要がある．今回は，既存の質問紙の組み合わ

表 1: ランダムフォレストの混同行列

	予測されたクラス 低うつ傾向	予測されたクラス 高うつ傾向
実際のクラス 低うつ傾向	36	1
実際のクラス 高うつ傾向	1	36

せにより質問紙を構築したから，妥当性についてはすでに確かめられていると考えられるので検証を省略した．信頼性については，今回は検証していない．順序効果の有無を検証するために，大学院生 3 名に質問紙に答えてもらい，1 週間後に同じ質問紙で順序を変えたものに回答してもらった．結果，いずれの被験者についてもカッパ係数 [Cohen 60] が 0.9 以上となったため，順序効果は無視できると判断した．

2.2 実験協力者

日本語で読み書き可能であることを条件として，実験協力者をクラウドソーシングを通じて募集した結果，987 名から協力を得た．

質問数が多く回答時間が長くなってしまうため，疲労などの影響により実験協力者が真剣に回答しない可能性を考慮して，同じ質問を 3 問尋ねた．それらの回答のうち，2 問以上異なる回答をした被験者 5 名を除いた．また，調査に 2 度以上参加した 43 名の被験者については，初回のデータのみを利用した．年齢に関して，異常値は中央値で置き換えた．その結果，男性 320 人，女性 615 人，合計 935 人の実験協力者が解析の対象となった．年齢の平均は 35.3 歳で，標準偏差は，9.2 であった．この実験協力者（935 人）において，K6 の点数の平均は 8.15，標準偏差は 5.3 だった．

2.3 解析手法

生活習慣等を特徴量として，うつ傾向が高い群と低い群に分類するモデルを学習させた．K6 スコアが 13 点以上の実験協力者を高うつ傾向群，その他を低うつ傾向群として判定した．高うつ傾向群は 186 人に対し，低うつ傾向群は 749 人であった．低うつ傾向群の実験協力者を無作為に選択し，高うつ傾向群と人数を揃えた．

この新たな実験協力者群（男性 123 人，女性 249 人）の年齢の平均は 34.6 歳，標準偏差は 9.3 だったので，元の実験協力者群に近い分布を得た．実験協力者の K6 スコアの平均は 11.3，標準偏差は 6.0 だった．

分類手法としてロジスティック回帰とランダムフォレストを使用した．全実験協力者を 7:1:2 の割合に分割し，それぞれ学習，開発，テストのために用いた．グリッドサーチを利用し，ハイパーパラメータを探索した．開発データにおける正解率が最大となるハイパーパラメータを用いて，テスト性能を評価した．

また，ランダムフォレストについては，予測に対する質問の重要性をジニ係数を用いて計算した．ロジスティック回帰については，予測に対する質問の重要性を重みの絶対値を用いて計算した．

3. 結果

うつ傾向の有無を予測するモデルで，ロジスティック回帰の正解率は，0.95 だったのに対し，ランダムフォレストの正解率は，0.97 だった（chance level = 0.5）． 表 1 は，ランダムフォレストにおいての混同行列を示している．適合率，再現率，F 値全てで 0.97 を示した．偽陽性・偽陰性の誤分類がなされた被験者がそれぞれ 1 名ずつおり，それらの実験協力者の K6 スコアはそれぞれ 10 と 19 であった．

正解率が高かったランダムフォレストによる予測への貢献が高い質問上位 10 問を表 2 に示す．続いて正解率が高かったロジスティック回帰による予測への貢献が高い質問上位 10 問を表 3 に示す．

カテゴリーとして，ランダムフォレスト，ロジスティック回帰ともに，中医学からの質問が，予測に重要な質問として多く選ばれた．ランダムフォレストでは，神経質の度合いを問う性格の質問も予測に有効であることが分かり，一方，ロジステック回帰では，食生活に関する質問も予測への貢献が高いことが分かった．ランダムフォレストにおいて 1 番貢献度が高かった「怖がりやすく、不安感がありますか。」という質問では，高うつ傾向群は，低うつ傾向群と比べると恐怖や不安感を多頻度で感じると回答した．2 番目に貢献度が高かった「元気がありますか。」という質問では，高うつ傾向群は，より元気がないと回答する傾向が見られた．また，3 番目の質問「疲れやすいですか。」には，高うつ傾向群は低うつ傾向群と比べて疲れやすい頻度が高いと回答した．

4. 考察

ランダムフォレストとロジスティック回帰では，予測への貢献が高い質問が異なる結果となった．今回は実験協力者の男女比が偏っており，その影響が予測に有効な質問に影響した可能性もあるので，今後の研究として，各性別で予測への貢献度が高い質問を調べたい．また，今回の実験では 296 問と多くの質問を用いたため実験協力者への負担が大きかったと考えられるので，今後の研究では少ない質問数でうつ傾向を予測したい．

5. まとめ

今回の研究では，うつ傾向の予測に有効な生活習慣等を調べた．ランダムフォレストを用いたうつを予測するモデルは，正解率 0.97 を示した．また，予測により有効な質問は，中医学のカテゴリーに属する体質について問うものと性格のカテゴリーに属していた．

参考文献

[Black 14] Black, D. W. and Grant, J. E.: *DSM-5 guidebook : the essential companion to the diagnostic and statistical manual of mental disorders, fifth edition.*, Washington, District of Columbia ; London, England : American Psychiatric Publishing (2014)

[Cohen 60] Cohen, J.: A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* (1960)

[Davidson 16] Davidson, S. K., Dowrick, C. F., and Gunn, J. M.: Impact of functional and structural social relationships on two year depression outcomes: A multivariate analysis., *Jornal of Affective Disorders* (2016)

[Doi 00] Doi, Y., Minowa, M., Uchiyama, M., Okawa, M., Kim, K., Shibui, K., and Kamei, Y.: Psychometric assessment of subjective sleep quality using the Japanese

表 2: ランダムフォレストによる予測への貢献上位 10 問

カテゴリー	質問
中医学	怖がりやすく、不安感がありますか。
中医学	元気がありますか。
中医学	疲れやすいですか。
中医学	精神が緊張したり、イライラしますか。
中医学	無意識によくため息をつきますか。
性格	私は自分自身のことを、冷静で、気分が安定していると思う。
性格	私は自分自身のことを、心配性で、うろたえやすいと思う。
健康	あなたの現在の健康状態はいかがですか。
病歴	あなたはこれまで、不妊症で病院に通いましたか。
中医学	体や手足が重だるく感じますか。

表 3: ロジスティック回帰による予測への貢献上位 10 問

カテゴリー	質問
中医学	元気がありますか。
中医学	どこか体に痛みがありますか。
食生活	コーヒー・紅茶に 1 回に入れる砂糖の量
中医学	怖がりやすく、不安感がありますか。
中医学	無意識によくため息をつきますか。
中医学	寒さに弱いですか。（冬の寒さ、夏のエアコンなど）
病歴	あなたはこれまで、不妊症で病院に通いましたか。
中医学	胸が重苦しかったり、腹が張ったりしますか。
食生活	夕、麺類は 1 週間に○杯
食生活	酢の物・和え物 1 週間に○回

version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects., *Psychiatry Research* (2000)

[Hamilton 80] Hamilton, M.: Rating depressive patients., *The Journal of Clinical Psychiatry* (1980)

[Kessler 02] Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., Walters, E. E., and Zaslavsky, A. M.: Short screening scales to monitor population prevalences and trends in non-specific psychological distress., *Psychological Medicine* (2002)

[Kimura 12] Kimura, Y., Wada, T., Okumiya, K., Ishimoto, Y., Fukutomi, E., Kasahara, Y., Chen, W., Sakamoto, R., Fujisawa, M., Otsuka, K., and Matsubayashi, K.: Eating alone among community-dwelling Japanese elderly: Association with depression and food diversity., *Journal of Nutrition, Health and Aging* (2012)

[MHLW] MHLW, : `http://www.mhlw.go.jp/kokoro/speciality/detail_depressive.html`

[Oshio 12] Oshio, A., Abe, S., and Cutrone, P.: 日本語版 Ten Item Personality Inventory （TIPI-J）作成の試み, パーソナリティ研究 (2012)

[Rush 03] Rush, A., Madhukar, H. T., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., and Keller, M. B.: The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression, *Biological Psychiatry* (2003)

[WHO 17] WHO, : `http://www.who.int/mediacentre/factsheets/fs369/en/` (2017)

[Zhu 08] Zhu, Y., Origasa, H., Uebaba, K., Xu, F., and Wang, Q.: Development and validation of the Japanese version of the Constitution in Chinese Medicine Questionnaire (CCMQ), *Kampo Medicine* (2008)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ポイントクラウドからの骨盤角度計測に関する研究

Study on Measurement of Pelvic Angle from Point Cloud

崎山恵美理 *1 Emiri Sakiyama　檜山敦 *2*3 Atsushi Hiyama　脇坂崇平 *3 Sohei Wakisaka　泉原厚史 *3 Atsushi Izumihara　稲見昌彦 *3 Masahiko Inami

*1東京大学工学部 The Faculty of Engineering, The University of Tokyo

*2理化学研究所　革新知能統合研究センター Center for Advanced Intelligence Project, RIKEN

*3東京大学先端科学技術研究センター Reseach Center for Advanced Science and Technology, The University of Tokyo

Pelvic angle is one of the important indicators to assess standing or walking posture. Conventional method to measure pelvic angle are using motion capture or inertial sensor that require to put markers or sensors on the body. This may impose a burden to subjects. In this research, we proposed a contactless pelvic angle measurement method using point cloud obtained from depth camera. We applied ICP algorithm to extracted point cloud of pelvis region from depth camera, and estimated the angular displacement caused by motion. As a result, proposed method was effective in measuring pelvis angle in upright position. Although, there was not much correlation with the true value in the results of estimation while walking, it was suggested that this method could capture the feature of walking postures in different pelvic angles.

1. はじめに

人間の姿勢変化の要因には加齢や疾患等, 様々なものが考えられる. 人の立位・歩行時の姿勢評価に関して, 骨盤の前後の傾きは一つの重要な指標となる [1].

従来の骨盤姿勢計測手法として主要なものにモーションキャプチャや慣性センサを用いた計測がある. これらの手法ではマーカーやセンサ自体を測定対象に装着する必要性が生じ, 計測の際の負担となりうる. そこで, 非接触で低コストな骨盤姿勢計測を実現しうるものとして, 3 次元カメラから得られるポイントクラウドに注目した. Microsoft 社開発の Kinect v2 センサのように人物の関節位置を推定し姿勢推定を行う SDK を持つ 3 次元カメラも, それ自体では骨盤の前傾・後傾等の細かな姿勢変化の検出は行えない. その一方で, ポイントクラウドを処理する技術に関しては, ICP アルゴリズム等を用いた剛体レジストレーションの研究が盛んに行われている.

これらを踏まえ, 本研究では, 「人体は非剛体であるが, 骨盤まわりのポイントクラウドを抽出し剛体とみなしてレジストレーションを行うことによって立位, 歩行時の骨盤姿勢変位を計測できる」という仮説を立てた. ポイントクラウドによる骨盤角度計測の精度や最適なパラメータについて, 計測対象の服装の種類によって比較し, ポイントクラウドによる立位・歩行時の骨盤姿勢を評価する可能性を検証した.

2. 提案手法

2.1 ポイントクラウドデータ記録部

システム全体の概略図を図 1 に示す.

データ記録部では, 複数台の Kinect からポイントクラウドデータを統合し記録するために, Kowalski らの LiveScan3D [2] を用いた. また次に, LiveScan3D によって得た人体のポイントクラウドデータに対し, ICP アルゴリズムによるレジストレーションの実行のために骨盤周りのポイントクラウドの抽出, 座標系の原点補正, 外れ値除去の 3 つの前処理を行う.

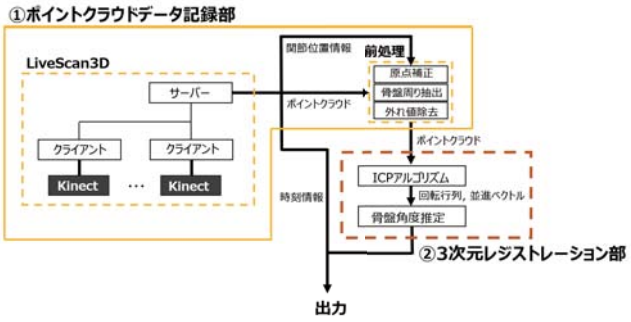


図 1: ポイントクラウドからの骨盤角度計測システム概略図

特に, 骨盤周りのポイントクラウドの抽出では, 記録した人体部分のデータからレジストレーションに必要な骨盤周りの部分のみを抽出する. 床に水平な面に関しては, およそ仙骨部位置から股関節位置までを含むように高さの上限 max・下限 min を設定してその範囲のみを切り出し, 矢状面においては, 前額面から見て Kinect で得た骨盤中心位置を中心に幅 $2w$ におさまる範囲のみ残るように切り出す (図 2). 後者は, 動作時の手の動き等によってポイントクラウドデータにノイズの乗りやすい部分を除去する役割を果たす. 抽出される範囲をに示した.

レジストレーション結果の誤差軽減のために外れ値除去を行う. 原点補正では Kinect の姿勢推定によって得られた骨盤中心が原点となるようにデータの座標系を変換する. これは後述の ICP アルゴリズムの初期位置合わせを行うため, そして得られた回転行列からの骨盤角度の解釈を身体の位置・方向に合わせた座標系設定によって単純化するためという目的を持つ.

2.2 3 次元レジストレーション部

3 次元レジストレーション部では, ICP アルゴリズムによって人体骨盤部のデータ同士の位置を合わせ, その姿勢変換の回

連絡先: 崎山恵美理, 東京大学稲見・檜山研究室, sakiyama@star.rcast.u-tokyo.ac.jp

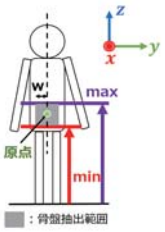


図 2: レジストレーションのための前処理のパラメータ

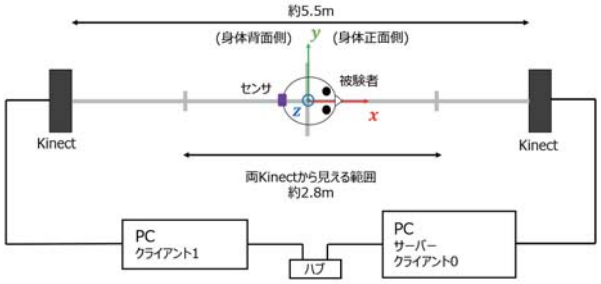


図 3: 実験 A：実験システム概略図



服装の分類		骨盤下端シルエット	
		反映する	反映しない
仙骨部シルエット	反映する	パンツ	スカート
	反映しない	—	ワンピース

図 4: 実験 A で用いた服装及びその選択基準

転行列から骨盤角度を推定する.

本手法で用いる ICP(Iterative Closest Point) アルゴリズムは剛体レジストレーションの代表的手法の 1 つで, 対応の与えられていない 2 つの点集合 $X = \{\mathbf{x_i}\}_{\mathbf{i=1}}^{\mathbf{n}}, \mathbf{Y} = \{\mathbf{y_i}\}_{\mathbf{i=1}}^{\mathbf{n}} \subset \mathbb{R}^{\mathbf{3}}$ に関して仮の対応を与えることによってレジストレーションを行う手法である. 基本的な手順としては, まず (1)X の各点について, 最近傍となる Y の点 (closest point) を対応付けることによって X, Y の仮の対応が与えられる. そして, (2)X の各点を Y の対応点へと変換するパラメータ (回転行列 R 及び並進ベクトル $\mathbf{t}$) を目的関数 $\min_{R,\mathbf{t}} \sum_{i=1}^{n} \|\mathbf{y_i} - (\mathbf{Rx_i} + \mathbf{t})\|^{\mathbf{2}}$ の最小化によって推定する. 点集合 $RX + \mathbf{t}$ と Y に関しても (1), (2) の手順を行い, 以下同様に反復 (iterate) して最終的なレジストレーション結果を算出する.

最近傍点同士で仮の対応づけを行うため, 2 つのデータ間の距離を離した場合において ICP アルゴリズムを実行すると全く意図しない局所解が推定されてしまう可能性がある. よってデータの初期位置を適切に設定する必要があることに注意しなければならない. このため本提案手法では前処理として原点補正による初期位置合わせを行う.

本手法では, 基準となるフレームと変位を測定したいフレームとの間で直接ポイントクラウドのレジストレーションを行い, 得られた回転行列からピッチ角を計算してこれを骨盤角度の推定結果として出力する. 本手法で直立時の骨盤角度変動を計測した結果, その変動範囲は約 0.8 度であった.

3. 実験

以上の提案手法の精度を, 実験 A では立位状態での骨盤運動時について, 実験 B では歩行時について評価した. いずれの実験でも骨盤角度を本提案手法及び仙骨部に装着したセンサによって計測し, センサによる結果を真値として提案手法結果と比較することで精度を評価した. センサには iPhoneX を使用し, 机上に静置した際の角度変動の大きさは 0.016 度であった. さらに実験 C では提案手法によって異なる歩行姿勢の識別が可能であるかを検証した.

3.1 実験 A：立位状態での骨盤運動における骨盤角度計測

実験 A では, まず立位状態での骨盤運動時において評価実験を行った. 図 3 に示す実験システムにおいて被験者 (健常女性 1 名) にメトロノームの示すリズムに合わせて立位状態で骨盤を前傾から後傾へゆっくりと動かす運動を行ってもらった. 図 4 に示す, 身体と衣服とのシルエットの差異に基づく服装 5 種 (レギンスパンツ, ワイドパンツ, スカート (ギャザーなし), スカート (ギャザーあり)) について同様の計測を行い, 服装による精度への影響や, 各服装における本手法の最適な測定条件を測定領域 (身体正面, 背面, 両面), 骨盤部抽出パラメータ min の値について考察した. この時基準フレームは計測開始時の直立状態のフレームとした.

その結果, 図 5 に示すようにいずれの服装についても提案手法のパラメータ最適条件下でセンサによる真値との間に非常に強い相関を示し, 本提案手法の有効性が確認された. 提案手法によって推定された変動は, 真値のそれよりも小さく表れる傾向にあった. また服装について, 表面形状の凹凸変化が大きいものほど精度が低くなることが示唆された. これらの結果について回帰分析を行い, 提案手法による推定結果から真値の推定を行う予測モデルができた. これを次の実験 B で使用した.

3.2 実験 B：歩行時の姿勢評価のための骨盤角度計測

本実験は, 姿勢が立位に近い膝関節が揃った瞬間における骨盤角度を歩行姿勢の評価に利用するために, その瞬間のデータに関して提案手法の精度の評価を行った (図 6). 実験 A と同じ実験環境で, 2 台の Kinect 間を結ぶ歩行路上を被験者 (実験 A と同じ) に歩いてもらった. この時の基準フレームは直立静止時のフレームとした. 結果として, 立位状態での骨盤運動時

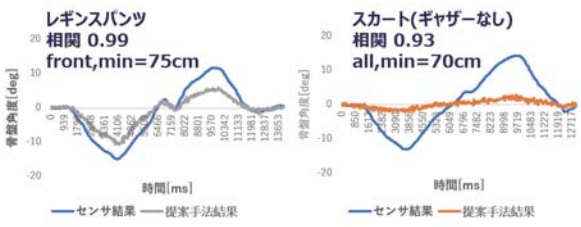


図 5: 実験 A：最適条件下での提案手法による骨盤角度推定結果と真値との比較. 相関の最も強いレギンスパンツと最も弱いスカート (ギャザーなし) の結果. 骨盤角度はプラス側が後傾, マイナス側が前傾を表す.



図 6: 実験 B で計測した, レギンスパンツでの歩行時の様子 (左) と膝関節が揃った瞬間 (右)

のような真値との強い相関は見られず, また提案手法によって推定された変動は真値のそれよりも同程度あるいは大きく表れる傾向にあり, 実験 A で得た予測モデルを適用するとその誤差は増大し, 歩行時の計測では立位時とは異なる手法の工夫・モデルの作成が必要となることが示唆された.

そこで, 基準フレームを直立状態から歩行時の膝関節が揃った瞬間のフレームに変更し, 歩行周期における同位相のデータ間のみ (軸足も同じ側のみ) でレジストレーションを行うように修正した結果, 相関関数はスカート (ギャザーなし) は −0.33 から 0.81 に, スカート (ギャザーなし) は 0.58 から 0.69 に, ワンピースは 0.59 から 0.88 へと変化し真値との相関が強まり, 本手法による骨盤角度推定の精度が向上したと考えられる.

3.3 実験 C:提案手法による歩行姿勢識別可能性の検証

提案手法による膝関節が揃った瞬間における骨盤角度推定結果から歩行姿勢の評価が可能であるか検証した. 具体的には, 正常姿勢による歩行, 骨盤の前傾した歩行, 骨盤の後傾した歩行の 3 種類の歩行について識別可能であるか分析した.

本実験では, 実験 B と同様の歩行路上を実験 A, B とは異なる被験者 (健常女性 1 名) に歩いてもらった. また各歩行姿勢の判定については, 歩行インストラクター 1 名の監修のもと行った. 5 種類の服装に関して同様の実験を繰り返した. この時の測定条件はパラメータ min, max, w の値を固定し front 条件, すなわち身体正面側のデータのみを用いた. 本実験は被験者の歩行姿勢が未評価である状況での提案手法の使用を想定し, 正しい立位姿勢を 統一的な基準として用いた.

提案手法結果について, いずれの服装に関しても推定結果の平均値は (b) 前傾 <(a) 正常 <(c) 後傾であった. また, 検定の結果, レギンスパンツ, ワイドパンツ, スカート (ギャザーあり) について, 3 群間の平均値の有意差が認められた. 残り 2 種についても (a)-(b) 間では非有意であったが, (a)-(c) 間, (b)-(c) 間では有意差が得られた.

これらの結果から, 提案手法によって各歩行姿勢の特徴を捉えられたと言える. 各歩行姿勢の識別に関しては, 全 5 種の服装のうち 3 種について 3 群間全てに平均値の有意差が認められた. 残りの 2 種についても非有意なのは (a) 正常・(b) 前傾間のみで, 提案手法による各歩行姿勢識別の可能性が示唆された.

4. おわりに

本研究では, 人体のポイントクラウドデータから, ICP アルゴリズムによるレジストレーションを行い非接触に骨盤角度を推定する手法を提案した.

評価実験の結果, 立位状態での骨盤運動時において提案手法による骨盤角度推定の有効性が確認され, 表面形状の凹凸変化の大きい服装ほど精度が下がることが示唆された. また歩行時の骨盤角度について, 立位状態での予測モデルは適用できず歩行時のための手法・モデルを要することが示唆された. このため手法の改良案について考察し、基準フレームの変更により真値に対する精度の向上が見られた.

また, 正常姿勢による歩行, 骨盤の前傾した歩行, 骨盤の後傾した歩行の 3 種類の歩行について計測を行った結果, 提案手法による各歩行姿勢の特徴捕捉及びそれらの識別の可能性が示唆された.

今後はデータを増やしていくことで計測および識別制度を向上させていきたい.

表 1: 1:レギンスパンツ, 2:ワイドパンツ, 3:スカート (ギャザーなし), 4:スカート (ギャザーあり), 5:ワンピースの各服装に関する (a) 正常姿勢による歩行, (b) 骨盤の前傾した歩行, (c) 骨盤の後傾した歩行の 3 群の提案手法による骨盤角度推定結果の分布の比較. ただし p 値は Bonferroni 法によって補正済とする.

服装	歩行	平均	分散	t 検定 (a)-(b)	t 検定 (a)-(c)	t 検定 (b)-(c)
1	(a) 正常	3.92	2.87	***	***	***
	(b) 前傾	0.761	3.17			
	(c) 後傾	9.98	8.00			
2	(a) 正常	0.893	0.874	*	***	***
	(b) 前傾	-0.193	5.86			
	(c) 後傾	7.49	3.08			
3	(a) 正常	-1.77	3.51	n.s.	***	***
	(b) 前傾	-2.24	5.54			
	(c) 後傾	0.995	5.76			
4	(a) 正常	-0.712	1.11	***	***	***
	(b) 前傾	-2.48	1.65			
	(c) 後傾	1.55	1.03			
5	(a) 正常	-2.71	6.39	n.s.	***	***
	(b) 前傾	-3.24	4.26			
	(c) 後傾	0.840	3.36			

n.s.: 非有意, *: p<0.05, **: p<0.01, ***: p<0.001

謝辞

この成果の一部は, 国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務の結果得られたものです. また, 実験に協力していただいた eikowalk 主宰の谷英子氏および講師の中野奈美氏に感謝いたします.

参考文献

[1] Kirsten Gotz-Neumann, 月城 慶一, 山本 澄子. 観察による歩行分析–筋活動とバイオメカニクス (特集 歩行分析の臨床). 総合リハビリテーション, Vol. 34, No. 2, pp. 107–116, feb 2006.

[2] Marek Kowalski, Jacek Naruniec, and Micha Daniluk. Live Scan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors. In *Proceedings - 2015 International Conference on 3D Vision, 3DV 2015*, 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Denoising autoencoder に基づく 心室期外収縮を含むRR間隔データ補正

Denoising autoencoder-based modification method for RRI data with premature ventricular contraction

宮谷将太 *1 Shota MIYATANI　藤原幸一 *1 Koichi FUJIWARA　加納学 *1 Manabu KANO

*1京都大学 Kyoto University

The fluctuation of an RR interval (RRI) on an electrocardiogram (ECG) is called heart rate variability (HRV). Since HRV reflects the activities of the autonomous nervous system, HRV has been used for many kinds of health monitoring systems. However, HRV is easily influenced by arrhythmia, which prevents the precise health monitoring. The present work focuses on premature ventricular contraction (PVC) which is common arrhythmia. To modify RRI data with PVC, the present work proposes a new method based on denoising autoencoder (DAE), referred to as DAE-based RRI modification (DAE-RM). The performance of DAE-RM was evaluated by its application to clinical RRI data which contains artificial PVC (PVC-RRI). The root mean squared error (RMSE) of modified RRI was improved by 83.5 % from PVC-RRI. The result showed that DAE-RM could modify PVC-RRI data appropriately. The proposed DAE-RM has potential for realizing precise health monitoring systems which use HRV analysis.

1. はじめに

心電図 (electrocardiogram; ECG) 波形の最も高いピークを R 波と呼び，隣接する R 波と R 波の間隔を RR 間隔 (RR interval; RRI) と呼ぶ．RRI が時間的に変動する現象を心拍変動 (heart rate variability; HRV) と呼び，HRV は自律神経活動を反映するとされる．そのため，HRV はストレスや眠気を評価する指標として用いられてきた．また，HRV は循環器疾患でも変化するため，循環器疾患の診断にも用いられている [1]．近年，容易に RRI を測定できるウェアラブルデバイスが開発された [2]．このようなウェアラブルデバイスに使用を前提としたてんかん発作の事前予知 [3] や，運転時の眠気検出 [4] などのヘルスモニタリングシステムが提案されている．

RRI データにアーチファクトが混入すると，HRV 指標が大きく変化してヘルスモニタリング性能が低下するため，HRV 解析を行う前にアーチファクトを適切に補正する必要がある．RRI データに混入するアーチファクトの原因は，R 波検出抜けエラーと不整脈の 2 つに大別される．本研究では後者に注目し，健常者でも起こりうる不整脈である心室期外収縮 (premature ventricular contraction; PVC) を対象とした [5]．

不整脈が含まれる RRI データを用いた HRV 解析における周波数領域指標を補正するアルゴリズムは提案されているが [6]，時間領域指標については触れられていない．本研究では PVC を含む RRI データを補正する手法を提案する．提案手法はノイズ除去を目的とするニューラルネットである denoising autoencoder (DAE) を用いており，本手法を DAE-based RRI modification (DAE-RM) と呼ぶ．本稿ではオープンデータとして公開されている RRI データに人工的な PVC アーチファクトを混入させ，DAE-RM を適用して補正性能を評価する．

連絡先: 藤原幸一，京都大学大学院情報学研究科システム科学専攻，〒 606-8501 京都市左京区吉田本町，fujiwara.koichi @ i.kyoto-u.ac.jp

2. 心拍変動解析と心室期外収縮

HRV 解析にはさまざまな指標が提案されているが，本節では広く用いられている時間領域指標と周波数領域指標について説明する [1]．また，PVC を含む RRI データを用いたことで生じる HRV への影響について述べる．

2.1 時間領域指標

時間領域指標は，被験者から取得した RRI データから直接計算できる．

- **meanNN :** RRI の平均値．
- **SDNN :** RRI の標準偏差．
- **Total Power (TP) :** RRI の分散．
- **RMSSD:** 隣接する RRI の差の 2 乗平均平方根．
- **NN50 :** 隣接する RRI の差が 50 ms を超えた回数．

2.2 周波数領域指標

RRI データは等間隔にサンプリングされていないため，周波数領域指標を計算する前にリサンプリングする必要がある．パワースペクトル密度 (power spectrum density; PSD) はリサンプリング後の RRI データより自己回帰 (auto regression; AR) モデルもしくは Fourier 変換を用いて計算できる．

- **LF :** PSD の低周波 (0.04-0.15 Hz) のパワー．
- **HF :** PSD の高周波 (0.15-0.40 Hz) のパワー．
- **LF/HF :** HF に対する LF の比．

2.3 心室期外収縮

心臓の正常な拍動リズムは洞結節が興奮することで生み出されるが，洞結節以外で興奮が生じると，本来の収縮に先行して心臓が興奮することがある．この不整脈を期外収縮と呼び，特に心室で興奮が生じる期外収縮を PVC と呼ぶ [5]．PVC を

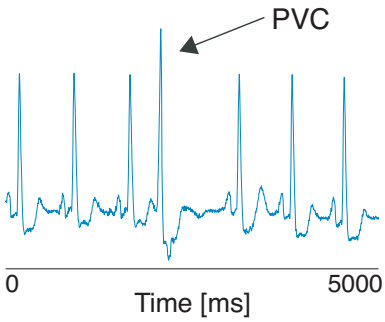


図 1: PVC を含む ECG 波形

含む ECG 波形を図 1 に示す．この図から，PVC が生じたことによって RRI が変化していることがわかる．

Kostis らは，心臓病を持たない健常者 101 名の ECG を 24 時間測定したところ，39 名が少なくとも 1 度は PVC を起こしていたと報告している [7]．また，年齢によって発生率が異なるという報告がされており，PVC を起こした人のうち 11 歳以下の割合は 1%以下である一方，75 歳以上の割合は 69%を占めている [8]．

PVC が生じると RRI データに大きな変動が生じ，RRI から計算される HRV も影響される．図 2 に PVC アーチファクトを混入させた RRI データ (PVC-RRI) から計算された SDNN と LF/HF を示す．青線と赤線はそれぞれ，元の RRI データから計算された指標と PVC-RRI から計算された指標であることを表し，緑の点線は PVC アーチファクトが混入したタイミングを表す．この図から，PVC アーチファクトを加えた直後から HRV 指標に変化が生じていることがわかる．PVC によって歪められた HRV 指標をその後の解析に用いると，誤った結果を導き出す危険性がある．

3. RRI 補正

本節では，PVC-RRI を補正するアルゴリズムについて述べる．提案するアルゴリズムでは，ノイズ除去を目的とするニューラルネットの denoising autoencoder (DAE) を用いる．

3.1 Denoising autoencoder

Autoencoder (AE) はニューラルネットを用いた次元圧縮，特徴抽出手法である [9]．AE は出力を入力とできるだけ等しくするため，入力と出力の再構築誤差を最も小さくするように学習を行う．このとき，中間層と出力層において活性化関数を恒等写像とすると，AE は代表的な次元圧縮手法である主成分分析 (principal component analysis; PCA) と一致する．次元圧縮された特徴を得るため，隠れ層のユニット数は入力変数の

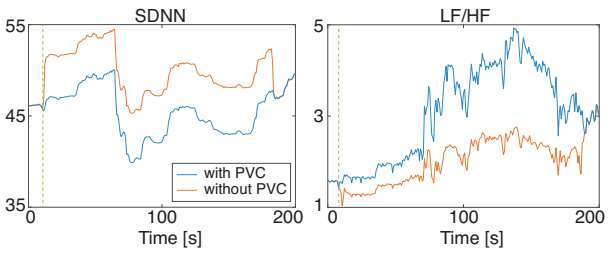


図 2: PVC が HRV 指標に与える影響: SDNN(左図), LF/HF(右図)

Algorithm 1 DAE-RM

while do
  Collect the newly measured tth RRI x_t.
  Store x_t to the buffer in the FIFO manner.
  if PVC detection **then**
    Wait the next $t+1$th to $t+T+P-1$th RRI $x_{t+1}, \cdots, x_{t+T-1}, \cdots, x_{t+T+P-1}$.
    Extract the previous P RRI $x_{t-P}, \cdots, x_{t-1}$ from the buffer.
    Construct the RRI subsequence to be modified: $\boldsymbol{x}_s = [x_{t-P}, \cdots, x_t, \cdots, x_{t+T-1}, \cdots, x_{t+T+P-1}]$.
    Calculate the mean of $\boldsymbol{x}_s$, $\bar{\boldsymbol{x}}_s$.
    $\boldsymbol{x}_s = \boldsymbol{x}_s - \bar{\boldsymbol{x}}_s$.
    Get the modified RRI subsequence $\hat{\boldsymbol{x}}_s$ by inputting $\boldsymbol{x}_s$ to the trained DAE.
    $\hat{\boldsymbol{x}}_s = \hat{\boldsymbol{x}}_s + \bar{\boldsymbol{x}}_s$.
    $d = \sum \hat{\boldsymbol{x}}_s - \sum \boldsymbol{x}_s$.
    Take the last element of $\boldsymbol{x}_s$, $\hat{x}^e_s$.
    $\hat{x}^e_s = \hat{x}^e_s - d$.
    Replace $\boldsymbol{x}_s$ to $\hat{\boldsymbol{x}}_s$
  else
    Wait until the next RRI data x_{t+1} is measured.
  end if
end while

数よりも小さくするべきであるが，正則化項を導入することで隠れ層のユニット数を入力変数の数より大きくすることができる．この手法を sparse AE (SAE) と呼ぶ [10]．

DAE は AE と同じ構造を持つニューラルネットであるが，学習時の入力にノイズを加え，ノイズを加える前の入力と同様の出力を得るよう学習を行う [11]．ノイズを加えて学習を行うことから，DAE はノイズ除去能力を有すると考えられる．本研究では，SAE と DAE を組み合わせて用いた．

3.2 DAE-RM

PVC が混入したことによる RRI データの変化はアーチファクトとみなすことができるので，DAE によって PVC アーチファクトを除去できる可能性がある．この DAE に基づく手法を DAE-based RRI modification (DAE-RM) と呼ぶ．

提案する DAE-RM のアルゴリズムを Algorithm 1 に示す．ここで，DAE は RRI 補正を行う前にすでに学習済みであるとする．ステップ 3 で first-in-first-out (FIFO) にて，新たに測定された RRI をバッファに記録する．RRI データに PVC が検出された場合，ステップ 5-15 で DAE による補正を行なう．ステップ 5-7 にて，DAE によって補正する RRI の部分データ $\boldsymbol{x}_s$ を構築する．ここで T は PVC によって変化した RRI の要素の数を表しており，$T = 2$ のときは PVC は単発性である．また，P は DAE による補正幅を表すパラメータであり，T 拍の前後 P 拍を DAE によって補正する．これは，DAE で補正を行うために PVC 前後の RRI データが必要なためである．ステップ 9 で DAE 適用前に $\boldsymbol{x}_s$ の中心化を行い，ステップ 11 で復元している．$\boldsymbol{x}_s$ の時間合計が補正の前後で変化しないようにするため，補正による時間合計の変化を補う必要がある．そこでステップ 12-14 にて，$\boldsymbol{x}_s$ の最後の要素で補正前後の変化を調整している．

4. ケーススタディ

本節では，PVC アーチファクトを加えた実際の RRI データに対し，DAE-RM を適用した結果を示す．

4.1 データセット

ケーススタディには PhysioNet の正常洞調律データベース (normal sinus rhythm database; NSRDB) を用いた [12]．NSRDB には成人 18 名の ECG と RRI のデータが含まれており，これらを被験者 A - R とする．被験者は 26 歳から 45 歳までの 5 人の男性（平均 33.8，標準偏差 7.7）と，20 歳から 50 歳までの 13 名の女性（平均 35.8，標準偏差 7.7）で構成され，その全員が不整脈でないと診断されている．データベースにある RRI データのいくつかには明らかなアーチファクトが含まれており，これらのアーチファクトは DAE-RM の性能を評価する妨げとなるため，これらを除去した．その結果，18 名の被験者から RRI データのエピソードが 166 個作成され，エピソードの合計時間は 375 時間となった．PVC アーチファクトはエピソードのランダムな位置に加え，本ケーススタディではその位置を既知であるものとした．

4.2 DAE の学習

DAE の学習には被験者 A から得られたエピソードを用いた．中間層と出力層の活性化関数にはそれぞれ，シグモイド関数と恒等写像を用いた．DAE のパラメータは被験者 B-F から得られたエピソードを用いて決定し，DAE によって補正する要素の数が 4，中間層のユニット数が 20，最大試行回数が 2000 となった．つまり，Algorithm 1 において $T = 2$，$P = 1$ となった．また，パラメータ決定時に L^2 正則化項を用いてスパース性を導入した．そのため，ユニットを結合するほぼすべての重みが 0 となっている．

4.3 RRI 補正

被験者 G-R から得られたすべてのエピソードを用いて DAE-RM の性能を検証した．なお，PVC-RRI データの作成と DAE-RM による補正は，加えたアーチファクトの位置に補正性能が依存することを防ぐため，5 回の試行を行ってその平均値を結果としている．図 3 は DAE-RM を被験者 M のあるエピソードに対して適用した結果である．青線，赤線，緑線はそれぞれ，元の RRI データ，PVC-RRI データ，PVC-RRI データを DAE-RM によって補正した RRI データを表す．この図から，DAE-RM による補正後の RRI データが元の RRI データと非常に近い値をとっていることがわかる．元の RRI データと PVC-RRI データの平方平均二乗誤差 (root mean squared error; RMSE) に対する，元の RRI データと補正後の RRI データの RMSE の改善率は 83.5 %であった．

また，図 3 に示したそれぞれの RRI データから HRV 指標を計算した結果を図 4 に示す．これらの指標における RMSE の改善率は，meanNN: 81.6%, SDNN: 98.3%, Total Power: 98.4%，RMSSD: 97.5%，NN50: 68.3%，LF: 86.8%，HF: 95.1%, LF/HF: 91.7%となった．この結果から，PVC が HRV 解析に与える影響を DAE-RM によって抑えられていることがわかる．

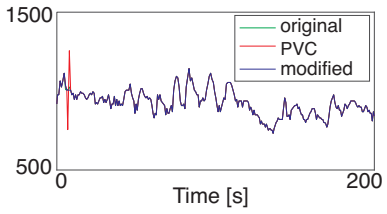


図 3: 被験者 M の補正前後の RRI データ

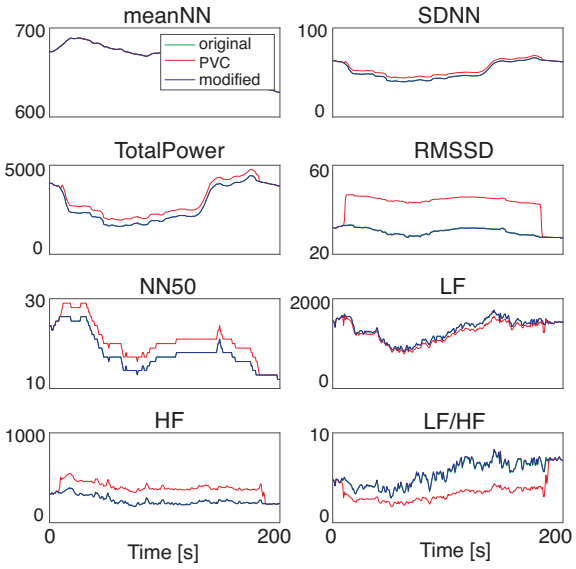


図 4: 被験者 M の補正前後の HRV 指標

4.4 考察

DAE による RRI 補正の性能を他の回帰手法と比較する．今回は，partial least squares (PLS) と，locally weighted PLS (LW-PLS) を比較の対象とした．PLS は広く使われている線形回帰手法であり，入力変数より少ない潜在変数を用いることで多重共線性の問題を回避できる [13]．LW-PLS は PLS を拡張した手法で，すでに得られているデータセットのサンプルとクエリの類似度を重みとして，クエリにおける入出力関係を最もよく表現するサンプルを用いて局所的な PLS モデルを構築する [14]．

DAE-RM，PLS，LW-PLS を用いて RRI 補正を行った結果を図 5, 6 に示す．縦軸には RRI と HRV の RMSE の改善率を用いた．これらの箱ひげ図から提案する DAE-RM が 2 つの手法と比較して，最も良い性能であることが示された，

以上の結果から，提案手法の DAE-RM は PVC が混入した RRI データの補正に有用であることが示された．また，DAE-RM によって HRV を用いたヘルスモニタリングをより正確に行える可能性が示された．

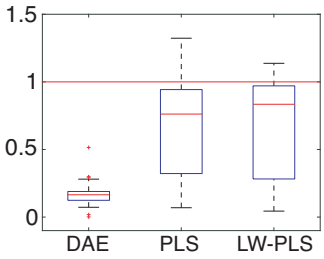


図 5: RRI の RMSE 改善率

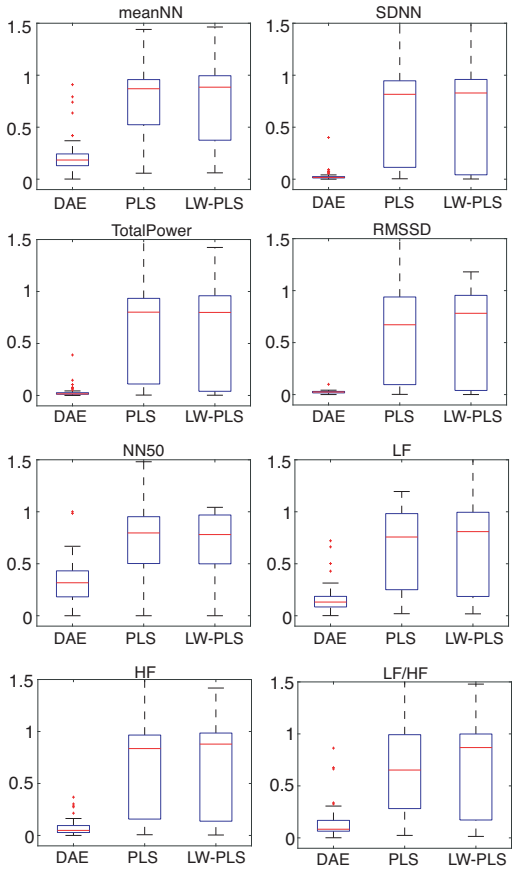


図 6: HRV 指標の RMSE 改善率

5. まとめ

本研究では，DAE を用いた PVC を含む RRI データの補正手法を提案した．PLS や LW-PLS などと比較して，提案する DAE-RM が PVC-RRI の補正に最も適していることが示された．

ケーススタディでは PVC が生じたタイミングが既知であるとしていたが，実用するためには PVC が生じたタイミングを検出しなければならない．今後の課題として，RRI データに含まれる PVC を検出する手法を開発する必要がある．また，RRI データに加わりうるもうひとつのアーチファクトである R 波検出抜けエラーについても，提案する DAE-RM が適用できるかどうか検証する．

参考文献

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,” *Circulation*, vol. 80, pp. 1043–1065, 1996.

[2] T. Yamakawa *et al.*, “Development of a wearable hrv telemetry system to be operated by non-experts in daily life,” in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, oct 2013, pp. 1–4.

[3] K. Fujiwara *et al.*, “Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1321–1332, jun 2016.

[4] E. Abe *et al.*, “Development of drowsiness detection method by integrating heart rate variability analysis and multivariate statistical process control,” *SICE Journal of Control, Measurement, and System Integration*, vol. 9, no. 1, pp. 10–17, 2016.

[5] F. C. Messineo, “Ventricular ectopic activity: Prevalence and risk,” *The American Journal of Cardiology*, vol. 64, no. 20, pp. J53 – J56, 1989.

[6] J. Mateo and P. Laguna, “Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 334–343, March 2003.

[7] J. B. Kostis *et al.*, “Premature ventricular complexes in the absence of identifiable heart disease.” *Circulation*, vol. 63, no. 6, pp. 1351–1356, 1981.

[8] Y. M. Cha *et al.*, “Premature ventricular contraction-induced cardiomyopathy,” *Circ. Arrhythm. Electrophysiol.*, vol. 5, no. 1, pp. 229–236, 2012.

[9] P. Baldi and K. Hornik, “Neural networks and principal component analysis: Learning from examples without local minima,” *Neural Networks*, vol. 2, no. 1, pp. 53 – 58, 1989.

[10] W. Sun *et al.*, “A sparse auto-encoder-based deep neural network approach for induction motor faults classification,” *Measurement*, vol. 89, no. 7, pp. 171–178, 2016.

[11] P. Vincent *et al.*, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML ’08. New York, NY, USA: ACM, 2008, pp. 1096–1103.

[12] M. Costa *et al.*, “Physionet: an nih research resource for complex signals,” *Journal of Electrocardiology*, vol. 36, pp. 139 – 144, 2003.

[13] P. Geladi and B. R. Kowalski, “Partial least-squares regression: a tutorial,” *Analytica Chimica Acta*, vol. 185, pp. 1 – 17, 1986.

[14] S. Kim *et al.*, “Development of soft-sensor using locally weighted pls with adaptive similarity measure,” *Chemometrics and Intelligent Laboratory Systems*, vol. 124, pp. 43 – 49, 2013.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

立ち上がり動作と下肢筋力との因果関係に関する考察

Consideration of relations sit-to-stand movement and strength of lower-limbs

樋山 貴洋[*1] 佐藤 佳州[*1] 小澤 順[*2]
Hiyama Takahiro Sato Yoshikuni Ozawa Jun
[*1] パナソニック株式会社 Panasonic Corporation
[*2] 産業技術総合研究所 Advanced Industrial and Technology

This research considered the relation sit-to-stand movement and strength of lower-limbs. We measured time of sit-to-stand without cushion and with cushion for five times sit-to-stand on seven subjects. As a result, the time of sit-to-stand with cushion is obtained higher mutual correction between strength of lower-limbs than without cushion. The soft environment of bearing surface is effective for exact estimation of strength of lower-limbs.

1. はじめに

高齢化に伴う運動機能の低下は運動の量と質が低下していていき，虚弱化が進行するとともに、日常生活動作能力（ADL）も低下していくため、運動機能低下の予防対策は重要である[1]. 運動機能の指標には筋力が知られているが、筋力を定量的に計測するためには大規模な装置を使用する必要があり、簡易でかつ日常生活動作から運動機能の指標となる筋力を計測することが望まれている. 2015 年 4 月の制定された改正介護保険法では，特別養護老人ホームへの入所は要介護 3 以上とされた[2]. それにより，長く在宅での生活を続けていることが求められていることから，在宅生活において要介護者の健康状態を見守ることへの需要が高まることが予想される.

本研究では日常動作の中で幾度となく行われている立ち上がり動作に着目する. これまで立ち上がり動作の文献には下肢関節モーメントを測定し，立ち上がり動作で活動する主要な筋について報告した研究[3]や，高さの異なる椅子からの立ち上がり動作において体幹の筋活動を測定し，椅子の高さが低いほど筋活動が有意に大きくなる報告[4]がある. また，5 回立ち上がりテストでの立ち上がり時間が筋力と相関関係を持つ報告もある[5]. 一方で，日常での立ち上がり動作は座面の高さだけではなく標準的な椅子やベッド、車いすといった座面環境や安定性の異なる条件で立ち上がりが必要な場面があるが、座面環境と立ち上がり動作について検討した文献は見られない. また，「できるだけ速く立ち上がる」といった条件下ではなく，立ち上がり方に条件を加えず日常の立ち上がり動作と下肢筋力との関係について示した文献は見られない. そこで，本論文では日常的に簡便に下肢筋力を推定することを目的に，異なる 2 つの座面条件での立ち上がり動作と下肢筋力の関係について考察する.

2. 立ち上がり動作と下肢筋力

2.1 従来までの下肢筋力の指標

下肢筋力は簡便性の点から，椅子に座った座位姿勢からの立ち上がり（Stand-to-stand:以降 STS と記述）動作を規定回数の成就時間により評価されている. 特に，5 回素早く立ち座り動作を行う FTSS(Five-times-sit-to-stand test)が少ない回数で定量的に下肢筋力を評価できるため，下肢筋力の評価指標に用いられやすい[5]. 先行研究では FTSS 時間が 14.4 秒を超える場合に，60% でバランス維持機能障害であることや[6]，FTSS 時間が 15.0 秒を超える場合に転倒リスクが 74%高まることが報告されている[7]. しかしながら，それらの研究は姿勢動揺計による立ち上がり方の明確な基準や座面環境を決めておらず安定した計測がなされているとはいえない. また，高齢者の場合，FTSS テストにおいても下肢に負担をかけるため、日常生活で行われる普段の STS から下肢筋力を推定できることが望ましい.

Caruthers らは，STS をいくつかのフェーズに分けて筋電位を用いて下肢筋力の活動量を調べた結果，臀部が座面から離れて体幹を後傾させる前に下肢の主要な筋に最も負担がかかることを報告している[8]. つまり，STS 動作は前傾姿勢により重心が前方方向に移動したときに最も下肢に負担がかかり，下肢への負担は動作速度にも影響しやすい. そこで本研究では，スタート時から体幹を伸展させる前の立ち上がり時間に着目して，下肢筋力との関係性について調べた.

2.2 運動環境変化によるロバスト性

歩行や立ち上がりといった日常動作は様々な環境下で行われている. 歩行の場合には環境変化に伴う予期せぬ転倒が起こる場合がある. 歩行中の転倒を評価するためには，遊脚期におけるつま先と地面との距離であるつま先クリアランスの最小値MMTC（Minimum Toe Clearance）を適切な評価することで表すことができる. Khandoker らは加齢の進行や転倒リスクが高まることで MTC のばらつきが大きくなることを報告した[9]. また，小林らは視覚障害者誘導用ブロックに健常な若年者と高齢者に歩行をさせた結果，若年者に見られていた股関節で大腿部を大きく振り出す動きが高齢者に見られなかった報告した[10]. つまり，筋力が低下していくと，運動環境変化に適応する動作が困難になることが考えられる.

本研究では立ち上がり動作に着目しているが，座面環境が影響して，立ち上がり方が困難になるといった考察をした文献は見られない. 通常，日常的に使用する椅子は長時間の座位姿勢や臀部への負担を避けるため柔軟性のあるクッションを用いる場合が多く，座面環境の違いで STS 動作と下肢筋力の因果関係を調べることに大きな意味を持つ. そこで，本実験では異なる環境下の条件（柔軟性のあるクッションの有無）での立ち上がり動作を 6 軸のセンサを用いて分析した.

〒571−8501 大阪市門真市大字門真 1006 番地 TI本部西門真地区
パナソニック株式会社 ビジネスイノベーション本部
樋山 貴洋 Email: hiyama.takahiro@jp.panasonic.com

3. 座面環境が異なる条件での立ち上がり動作実験

3.1 実験環境

(1) 被験者

被験者は下肢の障害の無い健康な男性 7 名（年齢：39.1±11.2, 身長：:174.8±6.0, 体重：68, 1±6.64）であった. 被験者には実験の目的・趣旨を十分に説明し同意を得た後に, 事前に設定したクッションの有無の条件にて STS 動作を実施した.

(2) 実験装置

下肢筋力測定には小型の筋力測定器（ミュータス F1:アニマ株式会社性製）を用いた. STS 動作の計測には 6 軸センサ(小型無線多機能センサ TSND121:ATR-Promotions 製)を用いた. そのセンサをベルトにより被験者の腹部にセンサを固定した. 6 軸センサからの信号はサンプリング 100Hz とし, センサ内のメモリによりデータを蓄積した. STS 用の椅子は椅子高 46 ㎝（脚付きマットレスベッド用：エムール（株）社製）を使用した. クッションは直径 40 ㎝充鎮物ポリエステル 100%のもの(モチモチクッション 40%ジェノアレタード T:NITORI 製)を使用した.

(3) 実験手順

被験者には STS 動作前に, 測定時の座位姿勢および動作様式について十分に説明した. 下肢筋力は, 固定用ベルトを用いて被験者の足首と椅子の脚を固定し, 座位・体幹 90 度屈曲位とした姿位から右足・左足ともに 2 回ずつ 5 秒間の等尺性筋力を計測し, 2 回の内の大きい方を取った.両足座位姿勢は, 背筋をまっすぐ伸ばした状態で膝関節を 90° に保持した姿勢とした.

STS 動作はソファに手をついて立ち上がらないように, 手は膝の上か胸の前で腕を組むように指示をした. STS 動作様式は座位姿勢の状態から, 検者の合図により立ち上がり, 立位状態で 1 秒程度保持しまた座位状態に戻るように指示をした. また, STS 動作はクッション有無のそれぞれの条件で 5 試行ずつ行った.

(4) 解析手順

STS 動作において 6 軸の内, 体幹前傾・後傾で変動する角速度データを用いて分析を行った. 計測データはカットオフ周波数 0.2Hz の HF と 2Hz の LF により平滑化した. 図 1 上図は STS 動作による体幹角の変化と図 1 下図は STS 動作における体幹前傾・後傾変化での正規化（-1~1）した 3 秒間の角速度データである. 座位状態から 1 秒付近の時間から体幹を前傾させていくと, 体幹角が減少し始める. 体幹角が最も小さくなるところでは, 角速度が 0 となる. その後, 体幹を後傾させ, 起立状態へと近づいていく.

解析区間の決め方には角速度データを用いた. 着席時（0 秒~1 秒まで）の最大値から STS 動作時の最小値の大きさを 0.8 倍した振幅値を閾値として, その区間の STS 時間を求めた. 体幹角の変化で示した場合, 体幹が前傾し始め最も前傾角が大きくなる区間となる. なお, 最大下肢筋力は両足それぞれの最大筋力の平均値とする. 求めた STS 時間と下肢筋力から相関係数を算出し, クッション有無で変化を考察する. また, 下肢筋力とクッション有無による座面環境の変化についても考察するため, 7 名の被験者それぞれの立ち上がり 5 試行分の標準偏差を求めて, その標準偏差と下肢筋力の関係を分析した.

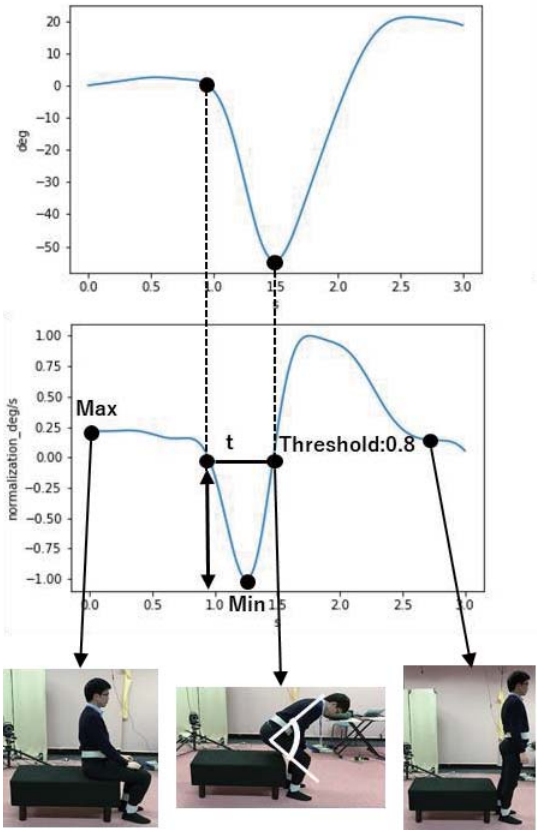


図 1. 立ち上がり動作区間の取り方

3.2 柔軟な環境での立ち上がり動作時間の分析結果

図 2 に被験者 7 名に対する下肢筋力と 5 試行分の立ち上がり平均時間との関係を示した. それぞれ 2 つの直線はクッション有無（青線：無, 赤線：有）, 点から上下に伸びる線は 5 試行分の標準偏差を示している. ともに筋力が大きいほど, 立ち上がり時間が速いことが示された. 表 1 に立ち上がり平均時間と筋力との相関係数を示した. クッション有りはクッション無しに比べて, 高い相関係数（-0.81）が得られた.

図 3 に被験者 7 名に対する下肢筋力と 5 試行分の立ち上がり時間における標準偏差の関係を示した. それぞれ 2 つの直線（青線：無, 赤線：有）はクッション有無を示している. クッション有りの方が下肢筋力との大きくなるほどばらつきが小さく, さらに相関係数も高い（-0.725）ことが示された（表 2）.

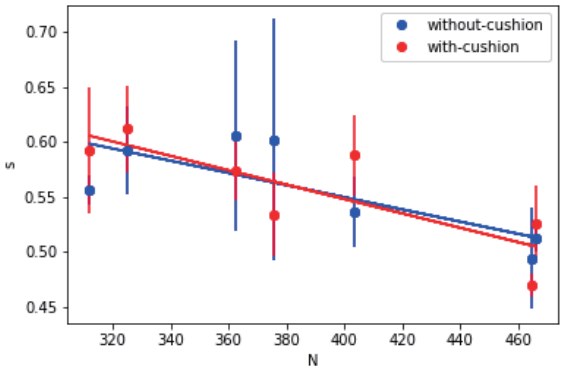


図 2. 立ち上がり平均時間と筋力の関係

表 1. 立ち上がり平均時間と筋力の相関係数

	相関係数
クッション有り	-0.817
クッション無し	-0.760

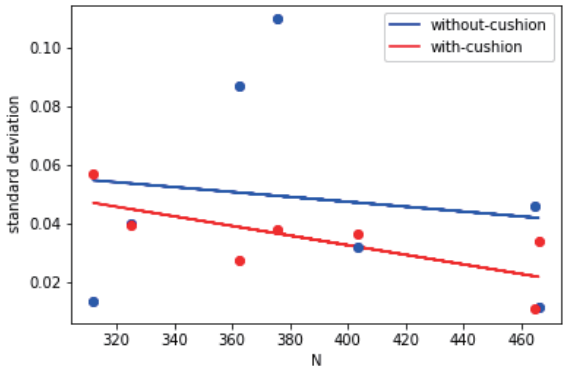


図 3. STS 時間の標準偏差と筋力の関係

表 2. STS 時間の標準偏差と筋力の相関係数

	相関係数
クッション有り	-0.725
クッション無し	-0.139

3.3 考察

本研究では異なる座面環境での立ち上がり動作と下肢筋力の関係を考察するため，被験者 7 名に対し クッション有無の条件で STS 動作を 5 試行実施し，STS 動作と下肢筋力の相関関係を分析した．図 2 の結果から，STS 時間と下肢筋力に負の相関がみられた．通常，FTSS テストにおいても STS 動作が速いほど筋力が強い傾向がみられることから，本実験でも同様な結果が得られたと考える．なお，クッションが有ることで座面が高くなり，立ち上がり時間が速くなるといった傾向は被験者間で異なることから座面高の高さの影響が少ないと考える．また，日常の STS 動作から下肢筋力を推定可能であると示唆される．

図 3 の結果から，5 試行中の STS 時間の標準偏差と筋力の関係性からクッション有りで筋力が大きい人ほどばらつきが少ない傾向が見られている．つまり，座面の材質によって STS 動作に違いが見られ，柔軟な座面ほど下肢筋力推定に有効なことが示唆される．

日常生活動作と運動機能との関係において歩行動作に着目した研究は多くある．歩行速度が速い人ほど Activities of Daily Living(ADL)能力が高かったとの報告[11]や，歩行周期のどの時期に各筋が活動するのか検討した研究[12]があるが，在宅生活中での自然な「見守り」を行う場合には住居の広さや環境が影響して歩行動作のみでは適切に運動機能を推定できるとはいえない．一方で STS 動作は，座面環境や高さといった環境の違いはあるものの場所を取らずに運動可能である．また，毎日の行動に離床や排泄といった STS が必要不可欠な動作があり，日常生活から運動機能を推定するためには STS 動作に着目することは重要であるといえる．

4. おわりに

本研究では座面環境の異なる 2 つの条件で立ち上がり動作を実施し，立ち上がり時間と下肢筋力の関係に着目し考察した．被験者 7 名に対して，クッション有無でそれぞれ 5 試行分の立ち上がり動作を実施し，下肢筋力との関係性について調べた．その結果，クッション有りの方が下肢筋力との相関が高いことが示された．また，STS 動作 5 試行の標準偏差と下肢筋力の関係においてもクッション有りの方が高い相関が得られた．つまり，座面環境によって下肢筋力を推定しやすい条件があり，特にクッション性の高い柔軟な素材が立ち上がりの安定性を高めることを示唆している．

本実験では，椅子の高さを揃え，さらにクッションのバリエーションも一つに設定している．しかしながら，下肢の筋力は椅子の高さに依存するため[3]，今後は椅子の高さを変更した場合や，クッションの柔軟性を増やした場合について評価する必要がある．また，本実験では若年者を対象に行った研究であり，高齢者で同様の結果が得られるのかは明らかになっていない．今後は高齢者を対象に検討し，さらに被験者数を増やすことが課題である．

参考文献

[1] 大杉紘徳，村田伸，堀江淳・他：地域在住高齢者の各歩行パラメータに関連する要因分析．日本ヘルスプロモーション理学療法学会，4（1），pp.31 - 35, 2014.

[2] 村井順子, 光田恵, 板倉朋世, 毛利志保, 棚村壽三:在宅介護の実態と介護環境の評価，日本建築学会環境系論文集，第 81 巻　第 725 号, Vol.81, pp.599-606, 2016.

[3] 浅井葉子, 金子誠喜, 大津慶子, 椅子からの立ち上がり動作における体幹前傾角度と下肢関節モーメントとの関係，日保学志, Vol.8 NO.1, 2005.

[4] 染谷富士子，三秋泰一，椅子の高さの違いが立ち上がり動作の下肢・体幹金の筋活動に与える影響，金大医保つるま保険学会，Vol.29, 101-104, 2005.

[5] M. Csuka, D. J. McCarty, “Simple method for measurement of lower extremity muscle strength”, American Journal of Medicine, Vol.78, pp.77–81, 1985.

[6] S. Whitney, D. Wrisley and G. Marchetti, et al, “Clinical measurement of sit-to-stand performance in people with balance disorders: validity of data for the Five-Times-Sit-to-Stand Test”, Physical Therapy, Vol.85, pp.1034-1045, 2005.

[7] S. Buatois, D. Miljkovic and P.Manckoundia, et al, “Five times sit to stand test is a predictor of recurrent falls in healthy commu-nity-living subjects aged 65 and older”, Journal of the American Geriatrics Society, Vol56, pp.1575-1577, 2008.

[8] E. J. Caruthers, J. A. Thompson and A. W. Chaudhari, et al, “Muscle Forces and Their Contribution and Horizontal Acceleration of the Center of Mass During Sit-to-Stand Transfer in Young, Healthy Adults”, Journal of Applied Biomechanics, Vol. 32, pp.487-503, 2016.

[9] A. H. Khandoker, S. B. Taylor and C. K. Karmakar,

"Investigating Scale Invariant Dynamics in Minimum Toe Clearance Variability of the Young and Elderly During Treadmill Walking", IEEE Transactions on neural systems and rehabilitation engineering, Vol. 16, pp.380-389, 2008.

[10] 小林吉之, 嶺也守寛, 高嶋孝倫, 藤本浩志, 視覚障害者誘導用ブロックが高齢晴眼者の歩行に与える影響に関する研究, 日本機械学会, Vol. 76,pp.2574-2579, 2006,

[11] 甲斐義浩，村田伸，堀江淳・他：要介護高齢者の 5m最速歩行速度と ADL 能力との関連．総合リハビリテーション，Vol. 39, pp. 795-799, 2011.

[12] R. Shiavi, N. Green, B. McFadyen, et al: "Normative childhood EMG gait patterns", J Orthop Res, Vol. 5, pp.283-295, 1987

[2J4] [General Session] 13. AI Application

座長:小澤 順（産業技術総合研究所）

Wed. Jun 6, 2018 5:20 PM - 6:40 PM Room J (2F Royal Garden B)

[2J4-01] Improvement of CT image segmentation by Deep Residual 3D U-Nets and 3D-CNNs

○Keita Ninomiya[1], Yoshinobu Furuyama[2], Joji Ota[2], Hiroki Suyari[1] （1. Chiba University, 2. Department of Radiology, Chiba University Hospital）

5:20 PM - 5:40 PM

[2J4-02] Feature selection and over-adaptation prevention in neural networks using the auxiliary weight method

○Akira NODA[1] （1. Technology Research Laboratory, Shimadzu Corp.）

5:40 PM - 6:00 PM

[2J4-03] Mental Disorder Diagnosis using fMRI Images by Structured Deep Generative Model

○Tetsuo Tashiro[1], Takashi Matsubara[1], Kuniaki Uehara[1] （1. Kobe University）

6:00 PM - 6:20 PM

[2J4-04] Segmentation of anatomical structures in chest X-ray pictures using U-Net

○Kenji Kondo[1,2], Jun Ozawa[1], Masaki Kiyono[2,3], Shinichi Fujimoto[3], Masato Tanaka[3], Toshiki Adachi[3], Harumi Ito[3], Hirohiko Kimura[3] （1. Advanced Industrial Science and Technology, 2. Panasonic Corporation, 3. University of Fukui）

6:20 PM - 6:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Deep Residual 3D U-Net と 3DCNN による CT 画像セグメンテーションの改善

Improvement of CT image segmentation by Deep Residual 3D U-Nets and 3D-CNNs

二宮 啓太 *1 Keita Ninomiya　古山 良延 *2 Yoshinobu Furuyama　太田 丞二 *2 Joji Ota　須鎗 弘樹 *3 Hiroki Suyari

*1千葉大学 工学部 情報画像学科
Department of Informatics and Imaging Systems, Faculty of Engineering, Chiba University

*2千葉大学医学部附属病院 放射線部
Department of Radiology, Chiba University Hospital

*3千葉大学 大学院工学研究院
Graduate School of Engineering, Chiba University

Segmentation of medical images with high precision and speed is an important task in many medical scenes. One such method for this task is GraphCut based on energy minimization problem. However, in GraphCut, it is difficult to perform segmentation completely and automatically if adjacent pixel values are similar. There are many methods for this problem, but most of them are not suitable in speed. In deep learning methods, automatic segmentation is possible because of its capability of capturing complicated features. In this research, we propose a model incorporating 3D U-Net extended with Residual Unit and 3DCNN for correcting segmentation results.

1. はじめに

医療現場において高精度で高速な医療画像セグメンテーションは重要な課題である．特に一刻を争う救急医療においてその需要が高い．現在はグラフカットなどの人の手を借りた作業が一般的であり，これらの手法では多くの時間を必要とするのが現状である．時間短縮のために，最近では深層学習の Fully Convolutional Networks[Jonathan 15] や U-Net[Olaf 15] による手法が多く研究されている．しかし，これらのモデルでは臓器の種類によっては，セグメンテーションが困難なことがある．その原因として，臓器が高い可変性を有していることが挙げられる．特に本研究で対象とした膵臓は，臓器のセグメンテーションの中でも困難な課題として知られている．本研究では膵臓のセグメンテーションの精度向上のために，Residual Unit を組み込んだ 3D U-Net と，セグメンテーション結果の訂正を行う 3DCNN を用いたモデルを提案する．

2. 従来モデル

2.1 3D U-Net

[O. Cicek 16] では U-Net を三次元に拡張した 3D U-Net が提案されている．3D U-Net は U-Net の畳み込み層を三次元の畳み込み層に置き換えたもので，三次元の特徴量を捉えることができるネットワークである．二次元の U-Net で三次元画像を処理するには，二次元スライス画像にしなければならない．この際に，スライスした方向の連続性を情報量として失っており，U-Net が三方向の全ての軸の連続性を利用するには，U-Net を三つ用意する等の工夫が必要である．それに対して，3D U-Net は入力に三次元画像が使用可能で，複数のネットワークを準備する必要がない．この点で U-Net より優れており，セグメンテーションの精度も二次元の U-Net と比較して向上したことが報告されている [O. Cicek 16].

連絡先: 二宮啓太，千葉大学 工学部 情報画像学科，263-8522 千葉市稲毛区弥生町 1-33，email : nino-keita1727@chiba-u.jp

2.2 Deep Residual U-Net

[Zhengxin Zhang 17] では U-Net に Residual Unit を組み込んだ Deep Residual U-Net を提案している．Residual Unit とは [Kaiming He 15] に用いられた手法で，ネットワークを深くした際に起こる学習が進まない現象を緩和する働きを持つ．その構造は単純で図 1 のように出力に対して入力を加算するだけである．[Zhengxin Zhang 17] では，他の手法と比較して高い精度を出したことが報告されている．

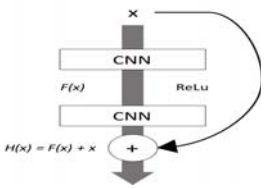


図 1: Residual Unit の構造の一例

3. 提案手法

3.1 提案モデル

本論文では 2.1 節に Residual Unit を組み込んだモデルと，3DCNN の二つを組み合わせたモデルを提案する．図 2 は提案ネットワークの概略図である．Deep Residual 3D U-Net と 3DCNN の詳細は 3.2 節と 3.3 節にて説明する．

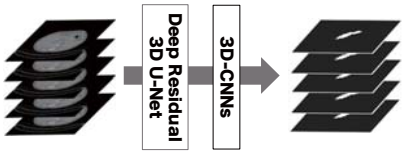


図 2: 提案モデル

3.2 Deep Residual 3D U-Net

図 3 に Residual Unit を組み込み，より層を深くした 3D U-Net を示す．Encoder 層及び Decoder 層の詳細は図 4 と図 6 で

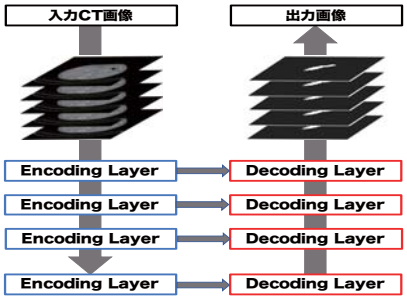
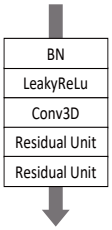
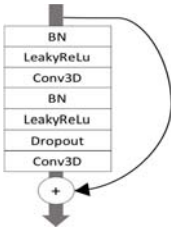


図 3: 提案する Deep Residual 3D U-Net





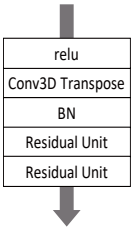


図 4: Encoder Layer　図 5: Residual Unit　図 6: Decoder Layer

ある．二次元の Deep Residual U-Net は 2.2 節で紹介したが，提案する Deep Residual 3D U-Net は更に Residual Unit を加えることで層を深くした．本モデルでは畳み込み層のフィルタサイズを 3 × 3 × 3，Conv3DTranspose 層は 2 × 2 × 2 とした．

3.3 3DCNN

3DCNN は三次元の畳み込み層を三つ重ねたものとした．フィルタサイズは初めの二層を 4 × 4 × 4，三層目を 1 × 1 × 1 とした．3DCNN では入力に学習済みモデルによるセグメンテーション結果と CT 画像を入力として与える．本モデルを提案した目的は学習済みの手法に適応することでもとのセグメンテーション結果を高精度化する役割を期待している．

4. 評価実験

4.1 データ

データセットは [Holger R. Roth 16] の Pancreas-CT を使用した．このデータセットには 82 人分の CT 画像があり，学習毎に学習データとして 50 人分，学習中の検証データとして 10 人分，評価用のテストデータとして 20 人分をランダムに選択する．この CT 画像は体の横断面を写した二次元画像であるが，それを層のように積み重ねることによって，三次元画像として捉えることができる．本実験では，膵臓が写った二次元 CT 画像を層のように重ねることで三次元画像とし，この三次元画像より膵臓を含む直方体を抜き出し，縦と横を 64 にリサイズを行なったものをデータとして用いた．また，CT 値は水を 0，空気を-1000 とした値であり，組織ごとにおおよその値が決まっている．本実験では 0〜300 を膵臓のセグメンテーションを行うのに重要なデータと考え，範囲外のものは値を 0 とした．その後，-1〜1 の範囲に正規化を行い入力画像とした．CT 画像の一枚を図 7 に示す．

4.2 損失関数

損失関数にはセグメンテーション問題において，一般的に利用されるダイス係数を 1 から減算したものを用いた．ダイス係数を式 (1) に，損失関数を式 (2) に示す．ダイス係数は画像の類似度を計算できる関数であり，0〜1 の範囲を持ち，1 に近

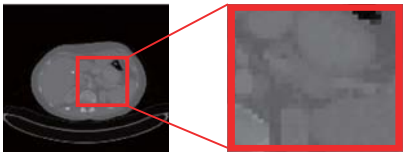


図 7: 入力三次元 CT 画像のうち，横断面 CT 画像の一枚



図 8: データセットの例

いほど正解に近いことを示す．A は教師データ，B はモデルによる予測結果とする．

$$\text{ダイス係数} = \frac{2 \times |A \cdot B|}{|A| + |B|} \tag{1}$$

$$\text{損失関数} = 1 - \text{ダイス係数} \tag{2}$$

4.3 実験環境

今回の実験で筆者が使用した環境を表 1 に示す．

表 1: 実験環境

OS	CentOS7
メモリ	60GB
プロセッサ	Intel(R) core(TM) i7-5820K CPU @ 3.30GHz
GPU	NVIDIA GeForce GTX TITAN X 12GB

4.4 評価方法

提案モデルによる出力結果は入力と同じサイズのボリュームデータであり，各ボクセルには膵臓であるかの確率が格納されている．今回は 2 クラス分類問題であるため，50%以上である場合は膵臓，50%未満である場合は膵臓でないとする．予測結果の評価には，式 (1) に示すダイス係数を用いた．

4.5 実験結果・考察

実験結果を表 2 に示す．表 2 は各手法 5 回ずつ実験を行なった結果である．これより，ダイス係数の平均，最小，最大の全てにおいて，提案手法である Deep Residual 3D U-Net + 3DCNN が最も高い値を記録していることがわかる．図 9 は各手法の平均をグラフ化したものである．このグラフを確認すると，図 4 と図 6 のような Encoding Layer，Decoding Layer を実装した図 3 の Deep Residual 3D U-Net は，平均で 5.2%の精度向上を果たしている．更に図 2 のように 3DCNN を通すことで，1.0%の精度向上が見られた．

次に 3D U-Net と Deep Residual 3D U-Net + 3DCNN の精度のばらつきに関して確認する．図 10 は二つの手法をそれぞれ 5 回ずつの実験を行い，その結果を箱ひげ図にしたものである．比較を行うと，Deep Residual 3D U-Net + 3DCNN は精度の向上だけでなく，ばらつきが小さくなっていることがわかる．

最後に実行速度について確認する．救急医療の現場で用いられること等を想定する以上，実用的な実行時間でなければな

表 2: 3D U-Net と提案手法の比較

ダイス係数	3D U-Net	Deep Residual 3D U-Net(提案)	Deep Residual 3D U-Net + 3DCNN(提案)
平均	0.688	0.740	0.750
最小	0.336	0.325	0.361
最大	0.855	0.879	0.882

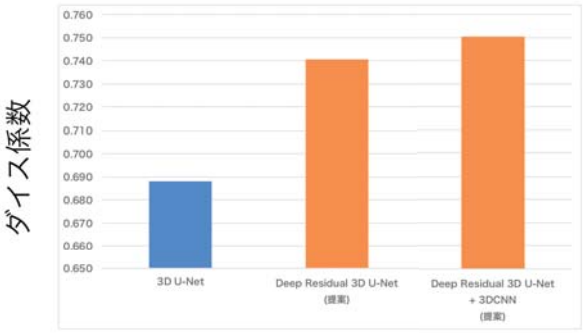


図 9: 各手法の平均ダイス係数の比較

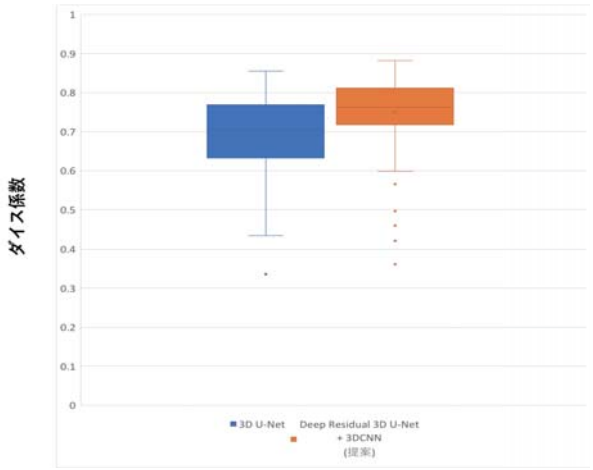


図 10: U-Net と Deep Residual 3D U-Net + 3DCNN の箱ひげ図

らない．従来手法及び，提案手法の実行時間を表 3 に示す．この時の実行環境は表 1 と同じである．比較を行うと，従来手法より実行時間は長くなっているが，0.38s と実用上問題のない実行時間となった．

表 3: 実行時間

	3D U-Net	Deep Residual 3D U-Net + 3DCNN(提案)
1 人あたり (s)	0.18	0.38

5. おわりに

本研究では 3D U-Net に Residual Unit を組み込んだ Deep Residual 3D U-Net と，セグメンテーション結果をより改善するように学習する 3DCNN の二つを組み込んだモデルを提案した．どちらも膵臓のセグメンテーションの精度向上に繋がった．しかし，欠点が存在する．例えば，今回は膵臓が写った二次元 CT 画像のみを利用して，三次元に再構築した三次元画像をデータセットとして利用したが，これは GPU の主記憶メモリが不足していたためである．より広い範囲からのセグメンテーションを行うには，この問題を解決するためにアルゴリズムの改良を検討しなければならない．

また，本実験では対象を膵臓に限定したが，実際の医療現場では他の臓器も対象である．更に対象が一つとは限らず，同時にいくつもの臓器のセグメンテーションを要求されることがある．今回提案したネットワークがこれらのことに対応できるかを今後検証する必要がある．

参考文献

[Jonathan 15] Jonathan Long, Evan Shelhamer, and Trevor Darrell: Fully Convolutional Networks for Semantic Segmentaion, CVPR 2015 pp 3431-3440

[Olaf 15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 pp 234-241

[O. Cicek 16] O. Cicek , A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation, arXiv preprint arXiv:1606.06650, 2016

[Zhengxin Zhang 17] Zhengxin Zhang, Qingjie Liu, et la.: Road Extraction by Deep Residual U-Net, arXiv:1711.10684v1 [cs.CV] 29 Nov 2017

[Kaiming He 15] Xiangyu Xhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition arXiv:1521.03385 10 Dec 2015

[Holger R. Roth 16] Holger R. Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. (2016): Data From Pancreas-CT. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU

[Roth HR 15] Roth HR, Lu L, Farag A, Shin H-C, Liu J, Turkbey EB, Summers RM. DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation. N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 556 - 564, 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

補助重み法によるニューラルネット上での特徴選択と過適応防止

Feature selection and over-adaptation prevention in neural networks using the auxiliary weight method

野田 陽 *1
NODA Akira

*1島津製作所 基盤技術研究所
Technology Research Laboratory, Shimadzu Corporation

In this study, we propose a method called auxiliary weight (AW) for neural networks in which each input value is weighted according to its contribution to the input dimension. AW is similar to Lasso regularization in the sense that it can extract features; however, AW is faster than Lasso in processing data that contains a several contributing dimensions and massive non-contributing dimensions, such as the data of medical mass spectrometry.

1. Introduction

In applications of discrimination using a chemical analyzer, such as in disease diagnosis using a mass spectrometer (MS), the input data generally contains a large number of dimensions and a high level of independence. Therefore, feature extraction is often performed by the partial differential of a neural network (NN) [**?**] or by Lasso regularization. In the field of medicinal chemistry, feature extraction is applied for biomarker candidate discovery. It is important not only to understand the reactions that occur in disease but also to prevent over-adaptation. In this study, we propose a novel method, which is called the auxiliary weight (AW*1) method in which each input dimension is multiplied by a weight based on its degree of contribution to the output of NN.

2. About Input Data

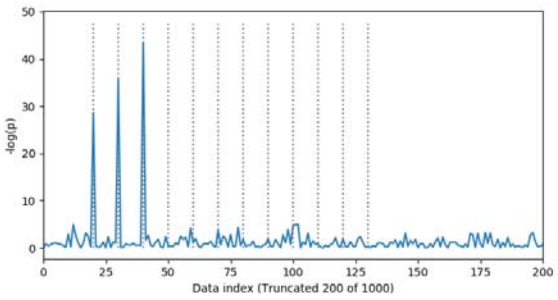
MS is an analytical device that is used to output an intensity value that correlates with the concentration of molecules having a certain mass-to-charge ratio (m/z). *2 Thus, in the obtained data, the concentration information regarding one or few molecular species is provided in a single dimension. Therefore, the independence of each dimension is observed to be high. The MS spectrum contains approximately 1000-100000 dimensions and one to several tens of these dimensions are observed to contribute to the diagnosis of a disease. In the diagnosis of several diseases, the discriminant function is nonlinear. Hence, discrimination is often performed using SVM(Support Vector Machine), and feature selection is often performed using SVMRFE(SVM recursive feature elimination) [**?**].

2.1 Simulation data

In this study, we use relatively simple simulated data. The input data is a 1000-dimensional vector of random numbers that depict a normal distribution. The training labels are provided on the basis of an equation, 1. Twelve of these dimensions of simulation data are contributing to the training label.

$$\begin{aligned} & Label = (Boolean)(y < 0) \\ y \quad = \quad & (x_1 + 0.5)(x_2 - 0.4)(x_3 + 0.3) \\ & +(x_4 + 0.25)(x_5 - 0.15)(x_6 + 0.05) \\ & +(x_7 + 0.3)(x_8 - 0.2)(x_9 + 0.1) \\ & +(x_{10} - 0.2)(x_{11} + 0.1)x_{12} \end{aligned} \tag{1}$$

Fig. 1 depicts the logarithmic p-value using the t-test for each dimension. Although twelve dimensions ($20, 30, \cdots, 130$ on the horizontal axis) contribute to the discrimination process, it can be observed that only three of them (indices 20, 30, and 40) can be identified using the t-test.



The dashed line indicates the contributing dimension.

Fig. 1: Log p-value of each dimension

3. Method

3.1 AW method

In the AW method, weights are applied to each dimension of the input data. The outputs of the AW layer are inputted to the NN, $f(x)$ (Fig.2). The AW method is similar to the Lasso method such that the weights of the dimensions that provide no contributions are zero; however,

連絡先: a-noda@shimadzu.co.jp

∗1 Code: `https://bitbucket.org/akira_you/awexperiment`.

∗2 MS-data is expressed as a vector in which the length covers the m/z range, whereas the elements denote the signal intensities at each value of m/z.

the AW method contains an additional loop to update each weight based on the degree of contribution. The degree of contribution, g_i, of each dimension is calculated using the norm of $\partial f(x_i)/\partial x_i$, which is obtained using a mini-batch. Further, g_i is normalized such that $|G| = 1$. Each weighting coefficient, a_i, of the AW layer is updated by $a_i* = g_i^u$, where u is the update rate parameter and is usually set to approximately 0.1. The AW layer and $f(x)$ are updated alternately. For the dimensions that depict large contributions, the input value is increased during the AW-update. Further, in the next batch, an NN-update of $f(x)$ is performed so that the contribution degree of the dimension becomes smaller. Finally, the contribution degrees of all the dimensions become equal to one. Thus, the AW-weights of the dimensions with no contributions become zero. AW can also be considered to be dynamic normalization. In several situations, medical MS data is normalized according to the variance, which may be caused by a disease (the target of interest in this technique) or may be caused due to the natural differences between various patients. Before conducting learning on the network, there is no information that which variance is caused by disease. In AW, data is normalized to the contribution degrees rather than to the variances. The contribution degree is defined using the NN and the learned parameters. Thus, the weights and NN must be updated alternately.

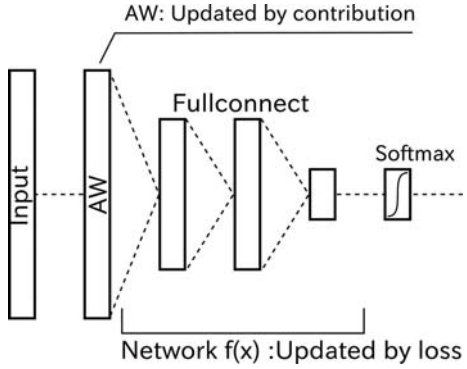


Fig. 2: The AW network model

3.2 Stochastic invalidator

The contributing dimensions can be extracted using the AW method. However, if there are massive random-number dimensions in the input, it is difficult to discern their contributions due to over-adaptation and contribution due to true discrimination. Several dimensions are required for over-adaptation by random values, the number of which is correlated with the numbers of the training data. However, in case of disease diagnosis and similar applications, only a few dimensions contribute because a disease may be caused due to a partially abnormal metabolism or because only a few chemical substances are affected by the condition.

Therefore, we intended to limit the dimensions that were contributed by the entanglement of multiple dimensions by implementing stochastic invalidation (SI) in the AW method. SI stochastically gates the output of the AW layer according to the value of a_i for each mini-batch. In this study, the value of a_i is compared with a random number, $N(0, 2.5\sigma_i)$, which was generated according to a normal distribution, where σ_i is the average of all the values of a_i. Further, a_i is replaced by zero if it is smaller than the random number. For example, consider a set of dimensions, $d_1, d_2, ..d_N$, with over-adaptation and a gate-open probability of $p_1, p_2, ..p_N$. The joint probability, $\prod p_i$, decreases exponentially with respect to the required number of dimensions, thereby sharply decreasing the learning opportunities of these dimensions.

Algorithm 1 AW update during the back-propagation of NN

while all epochs are incomplete **do**
  $G \leftarrow 0$;
  for i **in** range(B_a/B_n)
    //Normal back-propagation
    $x \leftarrow A \circ D[i]$;
    $y \leftarrow forward(x)$;
    $grad \leftarrow backword(\partial|Softmax(y) - L[i]|/\partial y))$;
    [update the NN parameters by grad]

    //Summing the degrees of contribution
    $G+ = |\partial f(x)/\partial x|$;
  end for
  $G/ = |G|$;
  $a_i* = g_i^u$;
end while

Where: B_a is the batch size for the AW, B_n is the batch size for the NN, D is the input data batch array, L contains the labels, a_i is the weight of the AW layer, A is is a vector of a_i u is the update rate, $\circ$ is the element-wise product, and grad represents the gradient for all variables.
Note:$\partial f(x)/\partial x$ can be obtained by $backword(1)$

Algorithm 2 Stochastic invalidator

$P \leftarrow 1$;
while all epochs are incomplete **do**
  for i **in** range(B_a/B_n)
    $x \leftarrow P \circ A \circ D[i]$;
    [NN update]
    [Summing the degrees of contribution]
  end for
  [Update A]
  $ave \leftarrow \sum_i a_i/N$
  $h_i \leftarrow N(0, ave * 2.5)$
  $p_i \leftarrow 0 \quad if(a_i < h_i) \quad else \quad 1$
end while

where P is the 0/1 filter for each input dimension, A is the weight for the AW, D is the input data batch array, and N is a random vector that is generated according to a normal distribution.

4. Experiment

Three analytical methods - no AW, with AW, and with AW and SI - are compared on fully connected NNs that depict the following grid-searched hyper-parameters:

- Number of middle layers: $(5, 6)$
- Dimension of the middle layer: $(100, 300)$
- Dimension of the first layer of the middle layers: $(1000, 2000)$
- Lasso: $(0, 1e-5, 1e-6)$

- Lasso (AW) :$(0, 1e-3, 1e-4, 1e-5)$
- Batch size: 800 / AW batch size: 2400

Lasso (AW) was applied to weights of AW when AW was valid and was applied to the first layer of the middle layers for the 'No AW' case. The number of training data was 27000, and the number of test data was 3000. Fig.3 depicts the accuracy that was observed during the three repeats of the experiment.

'AW+SI' provides high accuracy during the early epoch. Fig.4 depicts the results of feature extraction. These are the AW weights (or the weights of the L2 norms of the first layers). These observations illustrate that the use of AW alone or AW with SI results in the successful extraction of all twelve features, whereas the use of Lasso alone, without using AW, results in the extraction of only six features.

5. Conclusion and Future work

AW can be used with or without SI to perform feature extraction and prevent over-adaptation more effectively than the performance of Lasso when there are massive non-contributing dimensions. This technique will be useful for searching the biomarkers and for conducting diagnostics using medical MS.

In this study, only the simulated data was analyzed. However, in actual medical data, the number of data will be low. Therefore, the extracted features will not be very accurate. It is necessary to consider these differences while observing the extraction results of each cross-validation and conduct verification process based on the knowledge of human physiology.

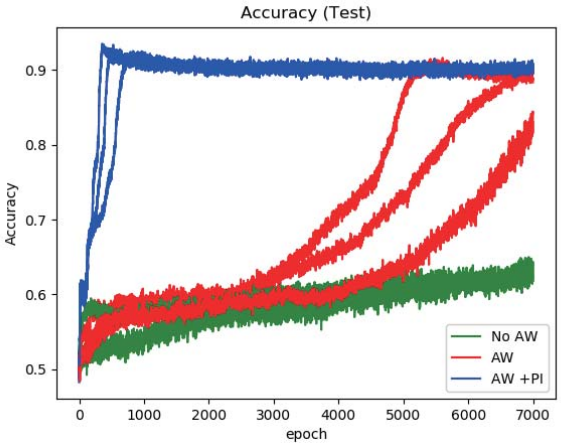


Fig. 3: Accuracy (Test)
Parameters (#of Mid, Mid.Order, 1stMid.Order, Lasso, LassoAW) are
No AW: (5, 300, 2000, 1e-6, 1e-5), (5, 100, 2000, 1e-5, 1e-3), (6, 100, 2000, 1e-5, 1e-4)
AW: (5, 300, 2000, 1e-6, 1e-4), (6, 100, 1000, 1e-5, 1e-4), (5, 300, 2000, 1e-6, 0)
AW+SI: (6, 300, 2000, 1e-5, 1e-3), (6, 300, 1000, 1e-5, 1e-3), (6, 300, 2000, 1e-5, 1e-3)

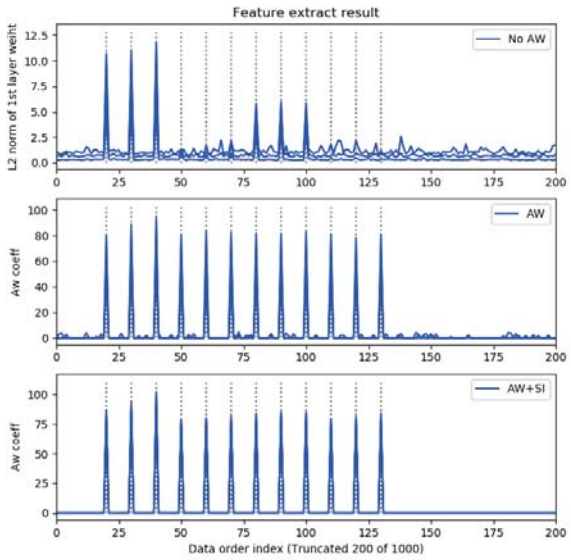


Fig. 4: Feature extraction result

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

構造化深層生成モデルによるfMRI画像を用いた精神疾患診断

Mental Disorder Diagnosis using fMRI Images by Structured Deep Generative Model

田代哲生 *1 Tashiro Tetsuo　松原崇 *1 Matsubara Takashi　上原邦昭 *1 Uehara Kuniaki

*1神戸大学 大学院システム情報学研究科 計算科学専攻
Graduate School of System Infomatics, Kobe University

In mental disorder diagnosis based on fMRI images, conventional studies perform preprocessing such as feature-extraction using correlation analysis since the fMRI dataset is composed of a small number of high-dimensional samples. However, this preprocessing could miss features necessary for diagnosis. On the other hand, deep generative models achieved good accuracy even with a small dataset with limited preprocessing. In this paper, we model fMRI brain images using a deep generative model with a subject-wise variable. The proposed model explicitly separates individual differences from the mental disorder and noise in fMRI images. The proposed model achieves accuracy higher than conventional methods.

1. はじめに

機能的磁気共鳴画像 (fMRI) のような神経画像データセットを使った機械学習技術に関する多くの研究が行われている．これらの技術は統合失調症などの精神的障害が持つと考えられる特定のバイオマーカーを発見することが期待されている．加えて，適切な治療の機会を提供し，潜在的な治療の有効性を評価することもできる．しかしこれらの神経画像データセットは，そのほとんどが他の機械学習向けのデータセットと比較して，限られた数のデータによって構成されている．したがって，機械学習技術を適用するためには，相関解析や特徴抽出，次元削減を行うことが一般的である．これは，fMRI 脳画像内に存在する撮影時の雑音や，被験者ごとの個人差 (脳の大きさや形の違い) を抑制することが目的でもある．

こうして，特徴抽出で得られた特徴を使って，機械学習を行う手法は多様なものが提案されている．具体的には，特徴の時系列に生成モデルを適用する研究などがある．Suk らは，深層学習による特徴抽出の後に，隠れマルコフモデル (HMM) によって時間的変化をモデル化し，精神疾患の診断を行い精度の向上を達成した [Suk 16]．Yahata らは相関解析による次元削減の後に，Sparse Logistic 回帰 (SLR) を使って精神疾患の診断を行った [Yahata 16]．しかしながら，これらの手法においても，前処理では取り除ききれない被験者ごとの個人差や，雑音などがある程度残ってしまう．このような精神疾患の診断に関連しない情報を強く学習してしまう事が，診断精度の向上の妨げとなっていると考えられる．

一方で，近年，Deep Neural Network (DNN) が柔軟な機械学習フレームワークとして注目されている [Schmidhuber 15]．DNN は任意の関数を近似することで，データセットが持つ高度な特徴を自動的に学習することができる．DNN は教師あり分類器 (Multi-layer Perceptron; MLP) [Castro 16] と，教師なし特徴抽出器 (Auto-encoder; AE) として用いられている [Castro 16, Suk 16]．これらに限らず，深層生成モデル (DGM) と呼ばれる Neural Network は，ネットワーク構造上に複数の要因の関係性を記述する生成モデルを実装している [Tashiro 17]．深層生成モデルの研究では，fMRI 画像と，精神疾患であるかどうかのクラスラベル，各時間のスキャンごとに起こる変動の 3 つの因子関係をモデル化し，分離する事で，他の比較手法よりも良い精度を達成した．これは，データの要因の関係性を正しく記述できていれば，生成モデルが小規模データセットにおいて識別モデルよりも良い精度が得られることによる [Raina 03]．よって，fMRI データセットの事前知識に基づいて，深層生成モデルを構築することで，診断精度が改善する可能性がある．以上を踏まえて，被験者の個人差の特徴を要因として明示的にモデル化する Structured subject-wise DGM (SDGM) を提案する．SDGM はそれぞれの被験者が，その fMRI 画像全てに共有する個々の変動 (すなわち，被験者特有の特徴) を要因として加える．この推論によって，データセットに対して，より妥当なモデル化を行う．本研究では，統合失調症と双極性障害の resting-state fMRI データセットを用いて実験を行い，SDGM を評価した．この実験では提案手法が，従来手法として行った SVM (Support Vector Machine) [Pereira 09], LSTM (Long Short-Term Memory) [Dvornek 17], DGN [Tashiro 17], AE+HMM [Suk 16], fMRI 脳画像解析でよく用いられるピアソン相関行列 (PCC) 用いて抽出された機能的接続性に基づいた手法 [Shen 10, Yahata 16] よりも良い精度の診断を達成できること示す．

連絡先: 連絡先: 田代哲生，神戸大学大学院システム情報学研究科計算科学専攻，tashiro@ai.cs.kobe-u.ac.jp

2. Structured subject-wise DGM

Structuerd subject-wise Deep Generative Model (SDGM) は，Auxiliary Deep Generative Models [Maaløe 15] の導出過程を応用し，DGM を fMRI 解析のために拡張したものである．まず，SDGM を構築するための各要因のモデル化を行う．N を被験者数，i を被験者の番号を示すとし，fMRI 画像を $\boldsymbol{x}_i$，クラスラベルを y_i とすると，データセットは $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$ となる．各被験者 i は健常者 ($y_i = 0$) か，精神疾患者 ($y_i = 1$) であり，それぞれ一人につき，T_i 枚の fMRI スキャン画像 $\boldsymbol{x}_i = \{x_{i,t}\}_{t=1}^{T_i}$ を持つ．また本手法では被験者 i は事前分布 $p(s)$ に従う特徴量 s_i (subject-wise feature) を持つものとする．この s_i は前処理によって取り除ききれなかった特徴を表し，各被験者が持つ，全スキャン画像に共通する個人の特徴 (脳の形など) である．ここで，fMRI 画像 $x_{i,t}$ は，クラスラベル y_i，被験者個人の特徴 s_i，時間的変化を含む潜在変数 $z_{i,t}$ に関連しているとする．なお，この潜在変数 $z_{i,t}$ は事前分布 $p(z)$

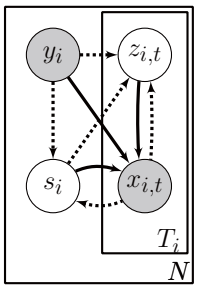


Fig. 1: Graphical model of our proposed generative model.

に従うものとする．z_i は時間変化によって起こる脳活動の変化や頭の物理的な動きなどを表す．ここまでの仮定を用いて，fMRI 画像 x_i を生成する，θ をパラメータとして持つモデル p_θ を構成する．この生成モデルを図で表したものを図. 1 に，式で表したものを以下に示す．

$$p_\theta(\boldsymbol{x}_i|y_i) = \prod_{t=1}^{T_i} p_\theta(x_{i,t}|y_i)$$
$$= \prod_{t=1}^{T_i} \int_{s_i}\int_{z_{i,t}} p_\theta(x_{i,t}|z_{i,t}, y_i, s_i)p(z_{i,t})p(s_i)$$

この式を変分法 [Kingma 13] によって変形することで，モデルエビデンス $\log p_\theta(\boldsymbol{x}_i|y_i)$ は，ϕ をパラメータとする推論モデル q_ϕ を用いて次のように表される．

$$\begin{aligned}
&\log p_\theta(\boldsymbol{x}_i|y_i) \geq \mathbb{E}_{q_\phi(s_i,\boldsymbol{z}_i|\boldsymbol{x}_i,y_i)}\left[\log \frac{p_\theta(\boldsymbol{x}_i, s_i, \boldsymbol{z}_i|y_i)}{q_\phi(s_i, \boldsymbol{z}_i|\boldsymbol{x}_i, y_i)}\right] \\
&= -D_{KL}(q_\phi(s_i|\boldsymbol{x}_i, y_i)||p(s_i)) \\
&-\mathbb{E}_{q_\phi(s_i|\boldsymbol{x}_i,y_i)}\left[\textstyle\sum_{t=1}^{T_i} D_{KL}(q_\phi(z_{i,t}|x_{i,t}, y_i, s_i)||p(z_{i,t}))\right] \\
&+\mathbb{E}_{q_\phi(s_i|\boldsymbol{x}_i,y_i)}\left[\textstyle\sum_{t=1}^{T_i} \mathbb{E}_{q_\phi(z_{i,t}|x_{i,t},y_i,s_i)}\left[\log p_\theta(x_{i,t}|z_{i,t}, y_i, s_i)\right]\right] \\
&=: \mathcal{L}_g(\boldsymbol{x}_i, y_i)
\end{aligned} \tag{1}$$

それぞれ，$D_{KL}(\cdot||\cdot)$ はカルバック・ライブラー情報量を，$\mathcal{L}_g(\boldsymbol{x}_i; y_i)$ はモデルエビデンスの下界を表す．この $\mathcal{L}_g(\boldsymbol{x}_i; y_i)$ が，条件付き生成モデル p_θ と条件付き生成モデル q_ϕ のパラメータである θ と ϕ の最適化のための目的関数となる．

すなわち，$\mathcal{L}_g(\boldsymbol{x}_i; y_i)$ をモデルエビデンス $\log p_\theta(\boldsymbol{x}_i|y_i)$ の近似として使用する．加えて，本研究では生成モデルを識別に応用するために，識別的な目的関数を定義する．クラスラベル y_i の事後分布 $p(y|\boldsymbol{x}_i)$ をベイズの公式を用いて計算すると以下のように変形できる．

$$\begin{aligned}
p_\theta(y|\boldsymbol{x}_i) &= \frac{p(y)p_\theta(\boldsymbol{x}_i|y)}{\sum_{y'\in\{0,1\}} p(y')p_\theta(\boldsymbol{x}_i|y')} \\
&\approx \frac{p(y)\exp\mathcal{L}_g(\boldsymbol{x}_i, y)}{\sum_{y'\in\{0,1\}} p(y')\exp\mathcal{L}_g(\boldsymbol{x}_i, y')} \\
&=: \exp\mathcal{L}_d(\boldsymbol{x}_i, y)
\end{aligned}$$

この式変形では，クラスラベル y の事前分布 $p(y)$ が $p(y=0) = p(y=1) = 0.5$ としている．したがって，$\mathcal{L}_g(\boldsymbol{x}_i, y=1)$ は大きな値を持つ場合，被験者 i はより精神疾患である可能性が高くなる．加えて，クラスラベルの対数尤度である $L_d(\boldsymbol{x}_i, y_i)$

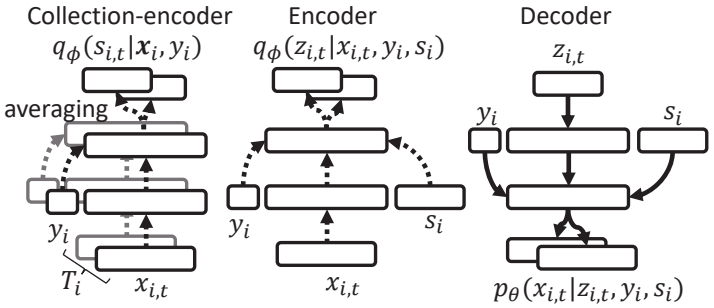


Fig. 2: Our proposed generative model on the deep neural networks.

は，これを目的関数とすることで，識別的な学習を可能とする．本研究ではこの二つの目的関数の調整を行うハイパーパラメータ $\omega \in [0, 1]$ を用いて，以下の式を目的関数とする．

$$\mathcal{L}(\boldsymbol{x}_i, y_i) = \omega\mathcal{L}_g(\boldsymbol{x}_i, y_i) + (1-\omega)\mathcal{L}_d(\boldsymbol{x}_i, y_i)$$

ここまでで得られた生成モデル p_θ と推論モデル q_ϕ を，Neural Network を用いて構築する．確率モデルを出力する Neural Network の構築には，Kingma らと同様の手法によって行う [Kingma 13]．前処理済みの fMRI 信号 $x_{i,t}$，被験者個人の特徴 s_i，スキャンごとの変動 $z_{i,t}$ をそれぞれ次元が n_x，n_s，n_z のベクトルとする．推論モデル $q_\phi(z_{i,t}|x_{i,t}, y_i, s_i)$ と生成モデル $p_\theta(x_{i,t}|y_i, s_i, z_{i,t})$ を多変量ガウス分布とする．それぞれのモデルのパラメータを出力する Neural Network を構築する．推論モデルの実装では，$x_{i,t}$，y_i，s_i を入力として，$q_\phi(z_{i,t}|x_{i,t}, y_i, s_i)$ のパラメータを出力する encoder を構成する．生成モデルの実装には $z_{i,t}$，y_i，s_i を入力として，$p_\theta(x_{i,t}|y_i, s_i, z_{i,t})$ のパラメータを出力とする decoder を構成する．ここで，データとして存在する $x_{i,t}$，y_i と異なり，s_i は事前に得られない．そこで x_i，y_i を使って，s_i を推論するモデル $q_\phi(s_i|\boldsymbol{x}_i, y_i)$ を Neural Network を用いて構成する必要がある．ここで，s_i が，各被験者に対して，全ての t において共通する特徴であることを考え，各被験者 i の全ての t における $x_{i,t}$ と，y_i を入力とし，$q_\phi(s_i|\boldsymbol{x}_i, y_i)$ のパラメータを出力とする encoder を構成する．この s_i を推論するモデルは，T_i 枚存在する $x_{i,t}$ と，y_i をそれぞれ別々の Neural Network に入力する．その各 Neural Network の出力の平均値を入力とする Neural Network によってこれを構成する．以上のようにして，Neural Network で構成した提案手法のネットワーク構造を図. 2 に示す．

このモデルを用いて診断を行う場合には，y_i が未知のデータに対して，精神疾患の尤度の近似として学習している $\mathcal{L}_d(\boldsymbol{x}_i, y_i = 0)$ と $\mathcal{L}_d(\boldsymbol{x}_i, y_i = 1)$ の値を比較することで，精神疾患であるか，健常者であるかを診断する．

3. 実験

本研究では統合失調症患者，双極性障害患者，健常者の resting state fMRI (rs-fMRI) 脳画像のデータセットを使用した．データセットは OpenfMRI データベースの ds000030 (`https://openfmri.org/dataset/ds000030/`) より得られる．まず，このデータに SPM12 (`http:///www.fil.ion.ucl.ac.uk/`) を使用して以下の前処理を施した．最初に fMRI 撮影開始時の磁気の乱れを取り除くために，各被験者の撮影した fMRI 画像の最初の 10 スキャンを除去した．その後，撮影時の時間的ズレを補正するために Slice Timing 処理を施した．続いて，撮

Table 1: Diagnostic Accuracies.

	Schizophrenia			Bipolar		
Model	**BACC**	**SPEC**	**SEN**	**BACC**	**SPEC**	**SEN**
PCC+Kendall+PCA+c-means[Shen 10]	0.640	0.635	0.645	0.602	0.565	0.640
PCC+SCCA+SLR[Yahata 16]	0.639	0.779	0.500	0.607	0.735	0.480
LSVM[Pereira 09]	0.505	0.788	0.223	0.512	0.855	0.169
LSTM[Dvornek 17]	0.661	0.854	0.467	0.571	0.802	0.340
DGM[Tashiro 17]	0.722	0.920	0.524	0.619	0.650	0.587
AE+HMM[Suk 16]	0.618	0.554	0.682	0.616	0.490	0.742
SDGM (proposed)	**0.767**	0.812	0.722	**0.622**	0.844	0.401

影中の被験者の頭や脳の揺れを除去するために，Realigment 処理を施し，各被験者の画像の位置ズレを抑制した．次に脳画像の個人差や，MRI 撮影時の環境差をある程度抑制するために，Normalization 処理行った．結果として，ボクセルサイズ $3 \times 3 \times 3$mm の MNI 空間に正規化される．ここまでの処理で，SPM12 によって正規化された画像を，AAL (Automated Anatomical Labeling) [Tzourio-Mazoyer 02] による 116 部位に分割した．この各分割した領域のボクセルの平均値を計算する．結果として，rs-fMRI 画像はベクトル化され，[116 部位 × スキャン枚数] の形となる．ここで，rs-fMRI では，0.025 Hz から 0.06 Hz ，または 0.07 Hz の間の周波数の帯域に，有用な特徴が存在することが知られている [Malinen 10]．そこで，本研究では，$0.01 \leq \mathrm{f} \leq 0.1$ Hz の周波数帯域が残るようにバンドパスフィルタを使用した．最後に，データのボクセル値が，時間方向において平均 0，分散 1 になるように，正規化を施した．なお，scrubbing として，スキャン全体での平均画像との，脳の位置のずれが 1.5mm を超えるスキャン画像と，任意の方向に 1.5 度を超える角度のズレが存在するスキャン画像を除去している．また，被験者ごとのスキャン枚数のバランスを保つために，scrubbing によって残ったスキャン画像が 100 枚以下となった被験者については学習するデータセットから取り除いた．よって，最終的に，健常者 113 人，統合失調症患者 44 人，双極性障害患者 45 人を使用した．

4. 結果

提案手法の効果を検証するために，統合失調症の診断と，双極性障害の診断の 2 つを行った．具体的には，患者の前処理済みデータと，健常者の前処理済みデータを学習し，識別を行った．なお，10 fold cross-validation を 5 回行った平均値を用いて精度を検証した．また精度の評価において，データに偏りが存在することを考慮して，Accuracy ではなく，Sensitivity と Specificity の平均値を使用する．これはデータの偏りに依らない精度となり，これを Balanced Accuracy として使用する．また学習がデータの偏りに影響されないよう，oversampling を行って学習した．SDGM のハイパーパラメータは探索範囲を狭めるために，encoder，decoder，s の推論器ともに，3 層のニューラルネットワークで構築し，入力層と隠れ層は同じ unit 数で，$\{50, 100, 200, 400\}$ の範囲とした．また encoder，s の推論器の出力層は $\{5, 10, 20, 50, 100\}$ の範囲から探索した．また，モデルの入力層には dropout を行った．ここでの dropout 率は $\{0.0, 0.5\}$ の 2 つのパラメータで探索した．最後に学習には Adam アルゴリズム [Kingma 14b] を使用した．Adam は初期パラメータを使用し，学習率は 10^{-4} とした．

比較手法としては，まず脳画像解析によく用いられる手法である，脳の機能的接続性を考慮したピアソン相関係数 (PCC) を計算する手法を用いる．ここでは PCC の後にケンドールの順位相関係数 (Kendall) を使って特徴選択を行い，その後主成分分析による特徴抽出の後に，C-means によって識別を行う手法 (以後これを PCC+Kendall+PCA+c-means とする) [Shen 10] と，Sparse canonical correlation analysis (SCCA) によって特徴選択を行った後に Sparse Logistic 回帰 (SLR) によって識別を行う手法 (PCC+SCCA+SLR) [Yahata 16] を行う．fMRI の時間的変化を学習し，識別する手法として，LSTM (Long Short-Term Memory) [Dvornek 17] と Auto-enoder による特徴抽出の後に隠れマルコフモデルによって識別を行う手法 (AE+HMM) [Suk 16] を行った．他の手法としては，一般的な機械学習の手法として Liner Support Vector Machine (LSVM) [Pereira 09] と，同様に深層生成モデルを使った手法 (DGM) [Tashiro 17] による識別とも比較した．各種手法のハイパーパラメータは，Kendall，SCCA による特徴選択数は $\{50, 100, 200, 400, 600\}$，PCA による特徴抽出の次元数は $\{2, 5, 10, 20, 50\}$，LSVM の Cost パラメータは $\{0.1, 0.2, 0.5, 1, 2, 5, 10\}$ の範囲から探索した．LSTM，Auto-encoder，DGM は，提案手法である SDGM と同様の条件で探索を行った．また HMM においては $\{2, 3, 4, 5, 6, 7\}$ の範囲から探索した．

この実験結果を表. 1 に示す．統合失調症，及び双極性障害両方の診断において，提案手法である SDGM が最も良い精度を達成した．特に統合失調症において大きな改善が見られた．SDGM と，PCC+Kendall+PCA+c-means，PCC+SCCA+SLR との比較によって，本手法が PCC のような相関分析を行わずとも，深層学習によって，一定の精度を得られることを示している．これは SDGM のネットワーク内で，各入力部位間での相関性や，機能的接続性を PCC と同様，あるいはそれ以上に，より良く学習していると考えることができる．また，DGM と提案手法である SDGM との比較によって，被験者固有の特徴として推論した s を分離する事によって，精度の向上が行われたことが分かる．一方で，双極性障害の診断においては，精度の改善こそあったものの，大きな改善とはならなかった．ここで，AE+HMM の精度が統合失調症の場合に比べて，相対的に高い精度を示している事が分かる．この事から，双極性障害の診断に置いては，統合失調症の場合よりも，脳の時間的変化に診断に必要な情報が存在すると考えられる．よって，SDGM が DGM に対して大きな精度の向上を見込めなかったのも，今回の提案手法で行った，subject-wise feature を加えるというモデルの拡張が，時間的変化を捉えられるようにするものではなかったからであると考えられる．

Table 1: Top 5 Weights of AAL for Diagnosis.

Schizophrenia		**Bipolar**	
ROI	weight	ROI	weight
Cerebelum_6_L	0.0555	Cingulum_Ant_R	0.0132
Postcentral_L	0.0532	Frontal_Inf_Orb_L	0.0121
Cingulum_Mid_L	0.0531	Cerebelum_7b_R	0.0116
Lingual_R	0.0529	ParaHippocampal_L	0.0114
Lingual_L	0.0526	Temproal_Mid_L	0.0106

さらに定性的な評価として，提案手法である SDGM を用いて，診断に特に寄与した部位の特定を行った．これは入力が，前処理において，AAL で部位ごとに分割して計算した平均値となっているために，116 次元の入力の中で，特に誤差が大きかった入力が，識別に関連したと判断できる．具体的には，SDGM のクラスラベルを変化させた時に，もっとも誤差の変化 (weight) が大きい部位を特定する事によって行う．具体的な計算式は，DGM における方法 [Tashiro 17] と同様である．結果は表. 1 の通り．各部位において，ある程度疾患との関連性がある事を示している．統合失調症においては，小脳 (Cerebelum) の関連性 [Yeganeh-Doost 11] や，中心後回との関連性 [Ferro 14] などを言及した研究も存在する．一方で，双極性障害においては，提案手法自体の診断精度が特別に高いわけではないため，今回の実験で得られた部位が，特別に強く双極性障害に関連すると判断するのは難しい．しかしながら，それぞれ，帯状回 (Cingulum) との関連性 [Fei 08] や，海馬傍回 (ParaHippocampal) との関連性 [Almeida 09] などを示す研究も存在する．

5. 結論

本研究では，fMRI 脳画像と機械学習を用いた精神疾患の診断に有効な深層学習モデルである SDGM を提案した．SDGM では，fMRI 画像を生成モデルとしてモデル化した．加えて，機械学習における診断の障害になる考えられる，被験者ごとの個人差を明確にモデル化し，分離する事によって，診断精度の向上を試みた．この診断精度を検証するために，本研究では，統合失調症と双極性障害のデータセットを用いて実験を行った．結果として，SDGM は，両方の精神疾患の実験において，従来手法よりも良い精度を達成した．これにより，SDGM が小規模データの場合の fMRI 画像を用いた診断に有効である事示した．また，DGM と同様に，入力と出力の誤差から，脳の部位と精神疾患との関連性を確認する事ができた．これにより，精神疾患の持つバイオマーカーの発見に役立つ可能性を示している．

本研究は科研費 (16K12487) の支援，総務省 SCOPE(受付番号 172107101) の委託を受けて行われた．

参考文献

[Schmidhuber 15] Schmidhuber, J. (2016) Neur. Netw., vol.61 pp.85-117

[Suk 16] Suk, H.I., et al. (2016), NeuroImage, vol.129, pp.292-307

[Castro 16] Castro, E., et al. (2016) IEEE Trans. Med. Imaging, vol.35(7), pp.1729-1740

[Raina 03] Raina et al. (2003), Advances in Neural Information Processing Systems 16, pp.545-552

[Pereira 09] Pereira, F., et al. (2009) NeuroImage, vol.45 pp.S199-S209

[Dvornek 17] Dvornek, N.C., et al. (2017) In Proc. of Machine Learning for Multimodal Interaction, pp.362-370

[Tashiro 17] Tashiro, T., et al. (2017) In Proc. of International Symposium on Nonlinear Theory and its Applications.

[Shen 10] Shen, H., et al. (2019), NeuroImage, vol.49(4), pp.3110-3121

[Yahata 16] Yahata, N., et al. (2016), Nature Communications, vol.7(7), 11254

[Maaløe 15] Maaløe, et al. (2015) In Proc. of International Conference on Machine Learning vol.48, pp.1445-1453

[Yeganeh-Doost 11] Yeganeh-Doost et al. (2011), Clinics, vol.66(1), pp.71-77

[Ferro 14] Ferro et al. (2014), European Psychiatry, vol.29(1), pp.1

[Fei 08] Fei et al. (2008), The British journal of psychiatry: the journal of mental science, vol.193(2), pp.126-129

[Almeida 09] Almeida et al. (2009), Psychiatry research, vol.174(3), pp.195-201

[Kingma 13] Kingma et al. (2013), In Proc. of International Conference on Learning Representations

[Tzourio-Mazoyer 02] Tzourio-Mazoyer et al. (2002), NeuroImage, vol.15, pp.273-289

[Malinen 10] Malinen et al. (2010), Proceedings of the National Academy of Sciences of the United States of America, vol.107, pp.6493-6497

[Kingma 14a] Kingma et al. (2014), Advances in Neural Information Processing Systems 27, pp.3581-3589

[Kingma 14b] Kingma et al. (2014) In Proc. of International Conference on Learning Representation, p.1-15

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

U-Netを用いた胸部X線画像からの解剖学的構造の領域抽出

Segmentation of anatomical structures in chest X-ray pictures using U-Net

近藤 堅司[*1,*2] 小澤 順[*1] 清野 正樹[*2,*3]
Kenji Kondo Jun Ozawa Masaki Kiyono

藤本 真一[*3] 田中 雅人[*3] 安達 登志樹[*3] 伊藤 春海[*3] 木村 浩彦[*3]
Shinichi Fujimoto Masato Tanaka Toshiki Adachi Harumi Ito Hirohiko Kimura

[*1] 産業技術総合研究所 [*2] パナソニック株式会社 [*3] 福井大学
Advanced Industrial Science and Technology Panasonic Corporation University of Fukui

We report a segmentation process of multiple anatomical structures in chest X-ray pictures, which is a key element of computer-aided diagnosis (CAD) systems, and its evaluation results. The segmentation process utilizes U-Net, that is, a type of fully convolutional network. The segmentation targets are a small region such as a first thoracic vertebra and a line structure which is a boundary between anatomical structures. In the evaluation, we achieve Dice index 0.91 as the segmentation accuracy for the first thoracic vertebra, and Dice index 0.71 through 0.81 as the segmentation accuracy for the line structures.

1. はじめに

近年、医用画像を解析処理することで異常検出や所見疾患推定を行う装置／ソフトウェアが開発されている。それらを用いる診断はコンピュータ支援診断(Computer-Aided Diagnosis: CAD)と呼ばれ、医師の読影精度向上および負担軽減が期待されている。医用画像の中でも胸部X線画像は豊富な情報を含み、その撮影装置は安価で普及率も高いため、胸部X線撮影は胸部疾患診断の第一選択方法になっている。

胸部X線画像の CAD 技術としては、事前に機械学習した病変を検出するものが多く提案されている[Rajpurkar 17] [Wang 17]。しかしながら、胸部X線画像では奥行き方向に複数の解剖学的構造物が重なって描写され、さらに病変がそれらの解剖学的構造物と重なった場合は病変検出が困難になる。このような場合に、解剖学的構造の正常状態を基に、疾患による解剖学的構造の変化を検出することで異常検知を行う胸部X線画像 CAD システムが必要となる。この CAD の実現には、胸部 X 線画像から解剖学的構造の領域を正確に抽出する必要がある。

一方、深層学習を用いて医用画像から領域抽出を行う U-Net が提案されている[Ronneberger 15]。U-Net は encoder-decoder 型の Fully Convolutional Network (FCN)の一種であり、encoder 部で画像に対し convolution と downsampling を繰り返し適用して高次の特徴抽出を行い、decoder 部で convolution と upsampling を繰り返し適用して特徴マップの解像度を向上させ、目的とする領域抽出結果を得る。この時、同一解像度の特徴マップを encoder から decoder に受け渡すことで正確な位置精度を実現している。

本研究では、U-Net を用いて胸部 X 線画像から解剖学的構造の領域抽出を行い、正常胸部 X 線画像 697 症例を用いてその性能を評価した。

2. 関連研究

胸部X線画像から解剖学的構造の領域を抽出する先行研究としては、[Ginneken 06] や[Novikov 17]がある。

連絡先: 近藤堅司, 産業技術総合研究所人工知能研究センター, 〒305-8560 茨城県つくば市梅園 1-1-1 中央第 1, kondo.kenji@aist.go.jp

[Ginneken 06]は、アプリケーションとして心胸比の自動計算を想定し、左右肺野、心臓、左右鎖骨の5つの解剖学的構造について active shape model (ASM), active appearance model (AAM), 及び pixel classification 法などの画像処理手法により領域抽出を行っている。肺野の領域抽出精度については、人（human observer）による抽出結果と同等、心臓については人より精度が低く、鎖骨については十分な精度を得ていない。実験に用いた胸部X線画像は、日本放射線技術学会(Japanese Society of Radiological Technology: JSRT)が公開した胸部X線画像データベース(JSRT database [Shiraishi 00])であり、この JSRT database に対し、評価に用いる領域マスク画像を手動作成し、Segmentation of Chest Radiograph (SCR) database として公開している。

[Novikov 17]は、JSRT database および SCR database に対し、U-Net [Ronneberger 15]を適用し、肺野、心臓、鎖骨領域に対する multi-class segmentation 手法を提案した。[Novikov 17]の領域抽出精度として、肺野と心臓については人（[Ginneken 06]での human observer）を上回り、鎖骨については人には及ばなかった。鎖骨のような小さい部位については、正則化手法を工夫したり、より多くの高レベル特徴を抽出することが必要であると結論付けた。

このように、従来までの胸部 X 線画像からの領域抽出においては、鎖骨のような小さい部位に対して十分な精度が実現されていない。

一方、胸部X線画像の異常検知に関する新たな取組として、[黒滝 17]らは、生成モデルの一種である変分オートエンコーダ(variational autoencoder: VAE)を用いて正常胸部X線画像をモデル化し異常検知を行う手法を提案している。具体的には、入力画像に関する対数尤度の変分下界を利用して異常検知を行っているが、十分な定量評価については未報告である。

3. 解剖学的構造の正常状態に基づく胸部X線画像 CAD システム

解剖学的構造の正常状態に基づく胸部 X 線画像 CAD システムでは、局所的な解剖学的構造の正常状態をモデル化し、疾患による解剖学的構造の変化を検出することで、胸部 X 線画像の所見推定を行う（図1）。

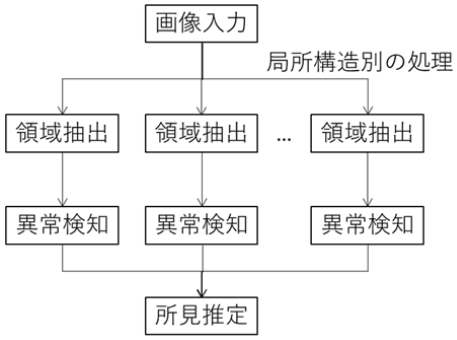


図 1　解剖学的構造の正常状態に基づく胸部X線 CAD システム

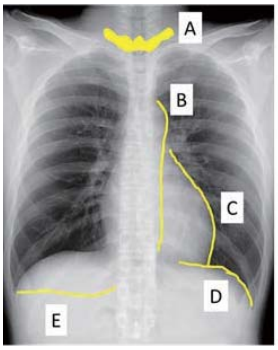
図1の CAD では、入力された胸部X線画像から、予め定めた複数の局所解剖学的構造の領域を抽出し、局所構造毎に事前に構築した正常モデルを基に異常検知を行い、最後に、それらの結果を総合して異常所見を推定する。次節以降で、U-Net を用いて実装した、図1における領域抽出部の詳細と、その評価について述べる。

4. U-Net を用いた解剖学的構造領域抽出

[Novikov 17]では、鎖骨のような小さい解剖学的構造に対して、人と同等な領域抽出精度は実現できていない。図1で示す CAD システムでは線状やより小面積の解剖学的構造を扱う必要がある。本節では、それらの解剖学的構造に関する領域抽出の取組結果を報告する。

4.1 評価対象の解剖学的構造

本稿では、図2に示す5個の解剖学的構造に対して、U-Net を用いた領域抽出の評価を行った。



	名称
A	第一胸椎
B	下行大動脈陰影
C	左室陰影
D	左横隔膜ドーム陰影
E	右背側横隔膜陰影

図 2　対象とする解剖学的構造

A は解剖学的構造そのものであるが、B,C,D,E については、正確には、X線吸収率が互いに異なる複数の解剖学的構造物によって胸部X線画像上に描出される、解剖学的構造間の境界線である。本稿では、この境界線を線構造と名付け、線構造も含めて解剖学的構造と呼ぶこととする。

4.2 U-Net を用いた解剖学的構造領域抽出

[Novikov 17]では、U-Net を胸部X線画像の解剖学的構造の領域抽出に適用し、肺野、心臓、鎖骨領域に対する multi-class segmentation 手法を提案している。オリジナルの U-Net に対し、①convolution 層の後ろに dropout 層を設けることで汎化性能を高める　②高解像層ほどフィルタ数（＝特徴マップ数）を多く配置し、最適化すべきパラメータの総数を減らす、等の改良を加えることにより、胸部X線画像の領域分割においてオリジナルの U-Net よりも高い性能を得ている。特に、鎖骨のような位置変動が少なく、かつ、サイズが小さい構造では、高次特徴をそれ程必要としないため、特に②の効果により、オリジナルの U-Net を上回る抽出性能を得ている。

我々の今回の評価実験においては、①②の改善を取り入れた U-Net を single-class segmentation タスクで用いる。領域抽出処理のデータフローを図3に示す。処理対象の胸部X線画像を共通の入力とし、5 個の U-Net が、図2における A～E の各解剖学的構造物の領域確率画像をそれぞれ出力する。

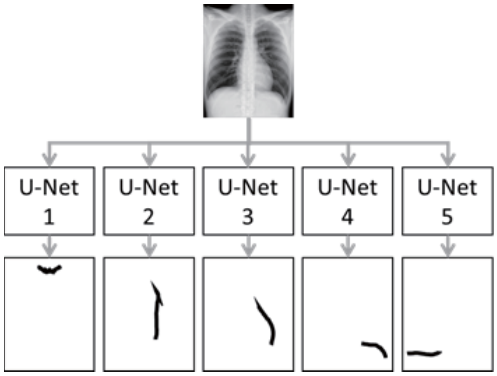


図 3 領域抽出処理のデータフロー

5. 評価実験

5.1 実験データ

実験データとして、福井大学病院の正常胸部X線画像 697 症例（DICOM format）を用いた。図2の 5 つの解剖学的構造の領域マスク画像については、胸部X線画像読影に関する教育を受けた作業者が手作業で作成し、診療放射線技師が最終確認・修正を行った上で使用した。なお、本実験は、福井大学倫理審査委員会の承認を得て実施している。

U-Net へ入力する胸部X線画像は次のように加工して用いた。

- 画像の幅と高さが異なる DICOM 画像について余白を付加して正方形状に変換後、256×256 画素、または、512×512 画素に縮小
- 輝度値に関しては、DICOM header の WW,WL を利用して正規化を行った後、float に変換
- 領域マスク画像も胸部X線画像と同様の方法で縮小

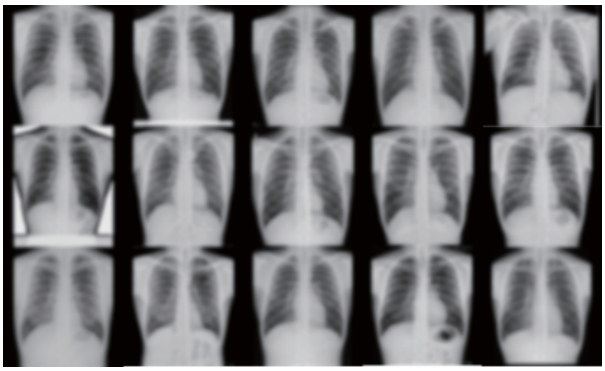


図 4　実験に用いた胸部X線画像の一部

訓練用として 697 症例のうちの 9 割(627 症例)を無作為に選定して用い、評価用として残りの 70 症例を用いた。

図4に、今回の実験に用いた胸部X線画像の一部を示す。図4では、個人情報保護の観点から、意図的に平滑化を適用した画像を掲載している。実験で用いた画像は、平滑化を適用していないため、より高周波成分を含んでいる。

5.2 U-Net の構造

実験に用いた U-Net の構造を図5に示す。この構造は、[Ronneberger 15]の構造、及び、我々が実験に用いた GPU のメモリ量、計算時間を考慮して決定した。

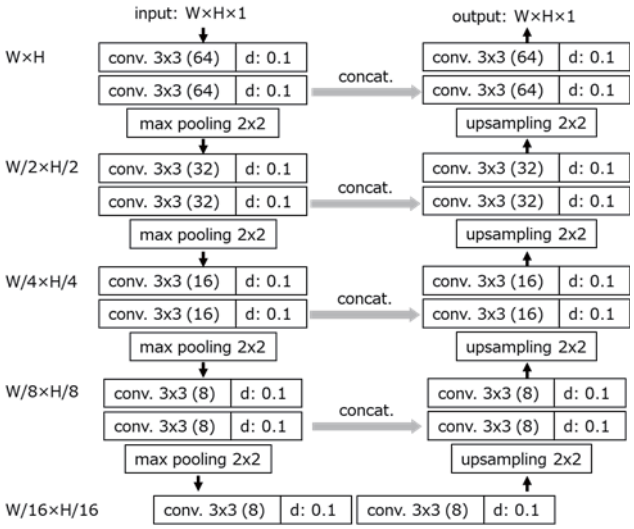


図 5 U-Net の構造

図5の記載法について補足する。”conv. 3×3(64)” はサイズ 3×3(×前層の特徴マップ数)のフィルタを 64 個用いる convolution 層を、”d: 0.1” は確率 0.1 の dropout 層を、左端の ”W×H” 等は入出力の画像サイズが W×H 画素である時の特徴マップの画像サイズを表している。図5の U-Net では 2×2 の max pooling を 4 回行うことで、最も低解像の特徴マップは入力画像の 1/16 のサイズとなっている。

5.3 実装および計算機

領域抽出の評価指標を、教師データとしての領域マスク画像とネットワーク出力としての領域マスク画像間の Dice 係数とし、学習時の損失関数を、Dice 係数×(-1)とした。

U-Net の実装には、python2.7 と keras を用いた。最適化には Adam を用い、学習率を 10^{-5} とした。計算機のスペックは、OS: Ubuntsu14.04, GPU: GeForceGTX1080Ti 11GB, CPU: Core i7 6850K, memory: 64GB である。

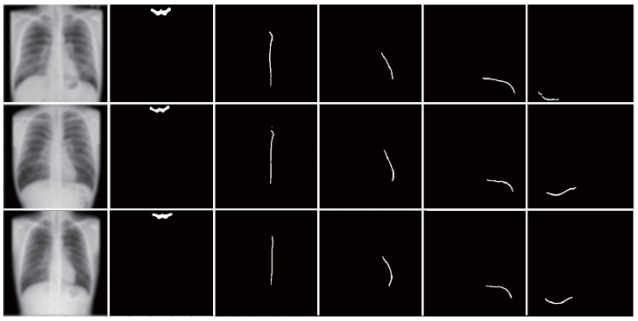
5.4 評価結果と考察

表1に Dice 係数での評価結果を、図6に評価データから無作為に選んだ3症例に対する領域抽出結果を示す。表中の数字は、入力画像解像度が 256×256 画素および 512×512 画素である場合それぞれの、学習終了後の訓練／テストデータに対する Dice 係数の平均値である。

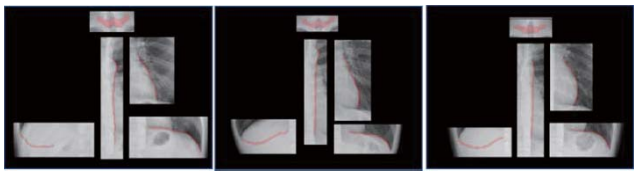
解剖学的構造そのものである第一胸椎については、両解像度ともテストデータに対する Dice 係数が約 0.91 となった。[Novikov 17]においては 256×256 画素の解像度時に、鎖骨に対する Dice 係数が 0.902 であった。データセットやデータ数が異なるため厳密な比較はできないが、今回の我々の実験では、鎖骨よりも小さく形状が微細な第一胸椎に対する Dice 係数が 256×256 画素の解像度時で 0.909 であり、良好な性能が確認できた。

表 1 評価結果（Dice 係数）

画像サイズ		第一胸椎	下行大動脈	左室	左横隔膜ドーム	右背側横隔膜
256	訓練データ	0.932	0.833	0.821	0.821	0.775
	テストデータ	0.909	0.796	0.763	0.769	0.711
512	訓練データ	0.929	0.827	0.810	0.816	0.760
	テストデータ	0.908	0.807	0.709	0.770	0.722



(a) 入力画像と出力画像



(b) 入力画像と抽出結果の重畳表示

図 6 評価データ(3 例)に対する領域抽出結果

一方、下行大動脈、左室、左横隔膜ドーム、右背側横隔膜という線構造については Dice 係数が 0.71～0.81 となった。但し、目視で領域抽出結果を見る限り、ほぼ正しく領域抽出できていた。Dice 係数が低くなった理由としては、Dice 係数が本来領域間のオーバラップを評価する指標であり、線構造の場合、位置が正しくても線の太さが異なれば値が低下するためと考えられる。今後は、線構造に対する損失関数及び評価指標を改良する必要がある。

線構造のうち、特に、右背側横隔膜陰影について Dice 係数が他よりも低い値となったが、この右背側横隔膜陰影は、濃淡差が微妙な陰影であることに依ると考えられる。

入力画像解像度 256×256 画素の場合と 512×512 画素の場合との比較においては、左室を除き、Dice 係数に差は見られなかった。左室については、入力画像解像度が 512×512 画素の場合に 256×256 画素の場合と比べ Dice 係数が約 0.05 低下していた。抽出結果を目視した結果、512×512 画素の入力画像の場合のみで、真の左室陰影以外の位置に過抽出が生じているものがテストデータ70症例のうち6割程度あり、さらにその半分が図7のように下行大動脈陰影の一部を含めた状態での過抽出であった。その理由を以下のように考えた。

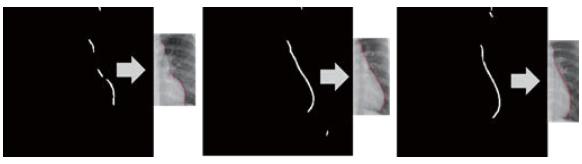


図 7 左室陰影に関する過抽出の例

図5の U-Net のネットワーク構造を、仮想的な1つの畳み込みフィルタとして考えた場合、その受容野の大きさを計算すると 204×204 画素になる。入力画像解像度が 256×256 画素の場合、理論上は、辺の長さで入力画像の約8割の正方領域を基に判断を下せることになる。しかし、入力画像解像度が 512×512 画素の場合には、辺の長さで入力画像の約4割の正方領域しか考慮することができず、そのため図7のような過抽出が生じたと考えられる。今後は、領域抽出対象の解剖学的領域のサイズ、および、その位置変動範囲等に応じたネットワーク構造の最適化を行う。

フィルタ受容野の問題を除けば、今回用いた5個の解剖学的構造の領域抽出では、解像度が 256×256 画素で十分と考えられる。

6. 結び

局所的な解剖学的構造の正常状態をモデル化し、疾患による解剖学的構造の変化を検出することで異常検知を行う、胸部X線画像 CAD の概要について述べた。さらに、その構成要素である、胸部X線画像からの解剖学的構造の領域抽出に対して、U-Net を用いて実装し、その評価結果について述べた。領域抽出対象として、第一胸椎のような小面積の解剖学的構造や、下行大動脈陰影などの解剖学的構造間の境界線である線構造を扱った。第一胸椎の領域抽出精度として、Dice 係数 0.91、線構造の領域抽出精度として Dice 係数 0.71～0.81 を確認した。

今後は、線構造に適した損失関数の設計、及び、領域抽出対象に応じたネットワーク構造の最適化などにより、解剖学的構造領域抽出の更なる改善を行い、さらに、異常検知手法の開発に取り組む。

参考文献

[Ginneken 06] B. Ginneken, M. Stegmann, and M. Loog, “Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database,” Medical Image Analysis, vol. 10, pp. 19–40, 2006.

[Novikov 17] A. Novikov, D. Lenis, D. Major, J. Hladuvka, M. Wimmer, and K. Buhler, “Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs”, arXiv:1701.08816v3, 2017

[Rajpurkar 17] P. Rajpukar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, R. Ball, C. Langlotz, K. Shpanskaya, M. Lungren, A. Ng, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning”, arXiv:1711.05225v3, 2017.

[Ronneberger 15] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation”. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 234–241. Springer, 2015.

[Shiraishi 00] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, “Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules,” American Journal of Roentgenology, vol. 174, no. 1, pp. 71–74, 2000.

[Wang 17] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. Summers, “ChestX-ray8: Hospital-scale ChestX-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases”, CVPR2017 , 2017.

[黒滝 17] 黒滝紘生、中山浩太郎、上原雅俊、山口亮平、河添悦昌、大江和彦、松尾豊、”深層学習による胸部X線写真からの診断補助”, 第 31 回人工知能学会全国大会, 2017

[2K2] [General Session] 12. HI / Education Aid

座長:瀬田 和久（大阪府立大学）

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room K (3F Azisai Mokuren)

[2K2-01] Pulley Learning Support with Feedback based on Haptic and Pseudo-Haptic

○Takanori Kono[1], Yukihiro Matsubara[1], Masaru Okamoto[1] （1. Graduate School of Information Sciences, Hiroshima City University）

1:20 PM - 1:40 PM

[2K2-02] Virtual Antique Folding Screen with Optical Effects of Sacred Light

○Toshiyuki Tsuno[1], Mariko Inamoto[2], Akihiko Konagaya[1] （1. Tokyo Institute of Technology, 2. Keisen University）

1:40 PM - 2:00 PM

[2K2-03] Analysis of Gaze Behaviors in Virtual Environments for Cooperative Pattern Formation

○Norifumi Watanabe[1], Kota Itoda[2] （1. Advanced Institute of Industrial Technology, 2. Keio University）

2:00 PM - 2:20 PM

[2K2-04] Development and Evaluation of a AR Drawing System suitable for real time gripping movement by using Kinect

○Hiroki Inatome[1], Masato Soga[1] （1. Wakayama University）

2:20 PM - 2:40 PM

[2K2-05] An Effect of User's Degree of Intimacy on Agent that Expresses Requital of a Favor

○Mitsuhiro Hayase[1], Taro Nakamura[2], Masayoshi Kanoh[2] （1. Toyohashi Sozo University, 2. Chukyo University）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

力覚，擬似力覚提示機能を持つ滑車学習支援システムに関する研究

Pulley Learning Support with Feedback based on Haptic and Pseudo-Haptic

河野 貴範[*1] 松原 行宏[*1] 岡本 勝[*1]
Takanori KONO Yukihiro MATSUBARA Masaru OKAMOTO

[*1] 広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

Learning systems with haptic or Pseudo-haptic feedback has developed. However, a study have not been conducted about effect of haptic and Pseudo-haptic feedback in learning system. Therefore in order to investigating effect of haptic and Pseudo feedback in learning system, we developed Learning system with feedback based on haptic and Pseudo-haptic. This paper describes the experiment using Kanji learning support system and the experiment using Pulley learning support system.

1. はじめに

ユーザに力のフィードバックを行える力覚提示デバイスを用いた学習支援システムの研究が行われている．学習支援システムにおいて力覚提示は，動作訓練の指導や仮想オブジェクトの重さ，動作，接触，衝突などの表現や教示的なフィードバックなど，様々な用途で用いられる．[檜谷 15]では力覚提示デバイスである SPIDAR-tablet を用いて滑車を学習題材とした力覚を伴う仮想実験環境が開発された．このシステムは滑車の組み合わせにより錘を持ち上げる際に必要な力の変化を力覚提示デバイスから体感することができる．

また視覚情報から力を錯覚させる擬似力覚と呼ばれる手法がある．擬似力覚は視覚情報と身体動作の間に不整合が生じた際に起きる．身体動作や身体動作を反映するポインタ等の移動量を増減させることで力を感じさせる手法である．擬似力覚提示は，ユーザの動きを制限しない，また視覚情報のみで行えるため力覚提示の代替方法としての利用可能性がある．この擬似力覚提示を用いた学習支援システムの研究も行われている．[柏原 15]では概念マップ作成に擬似力覚提示を用いた概念マップ作成支援システムを開発，擬似力覚提示効果の評価が行われた．この研究では概念マップ作成時に擬似力覚提示を用いることで重要な概念，概念間の関係を示すことを目的とした．

力覚提示，擬似力覚提示を用いた学習支援システムに関する研究は行われているが，それら各提示方法を学習支援システムに用いた場合の学習の比較を行った研究はされていない．そこで，[河野 17]では筆順学習支援システムに力覚提示，擬似力覚提示を用いた際の学習効果や学習時の比較を行った．その結果から間接指導で学習を進める題材の場合は力覚提示でも擬似力覚提示でも同様の学習結果が得られること，また静止状態での提示に関して各提示方法による学習の差が生まれる可能性を確認できた．そこで著者らは提示を静止状態と移動状態の両方で行う直接指導で進める題材として滑車実験を選択した．今後滑車学習支援システムを製作し，システムを用いて実験を行い提示時の状態に着目して学習結果を確認する予定である．本稿では 2 節で既に行った筆順学習支援システムを用いた実験について，3 節では 2 節の実験を踏まえた題材の検討について，4 節では今後開発予定のシステムとそれを用いた実験について記述する．

連絡先：河野貴範，広島市立大学大学院情報科学研究科，広島県広島市安佐南区大塚 3-4-1，TEL 082-830-1500，lkono@lake.info.hiroshima-cu.ac.jp



図 1 筆順学習支援システムの外観

2. 筆順学習支援システム

2.1 筆順学習支援システムの概要

図 1 に筆順学習支援システム[河野 17]の外観を示す．このシステムは学習者に正しい筆順を認識させる学習支援システムである．システムは事前に選択された力覚提示，擬似力覚提示，音声提示のいずれかにより学習者の誤った知識を修正する．学習者はタブレット端末の画面上に表示された漢字をなぞることで，漢字の筆順入力を行い，学習者が誤った筆順を入力した場合のみ各提示方法により誤りの指摘を行う．力覚提示は図 1 内の SPIDAR-tablet と呼ばれる力覚提示デバイスを用いる．擬似力覚提示はタブレット端末の画面上に表示されたポインタを誤った入力が行われた場合のみ実際の指より遅れて追従させることにより行う．音声提示は単一のブザー音により行う．

本システムでは筆順の誤りを誤り(1)「画の順番を間違う」，誤り(2)「画の途中で離す」，誤り(3)「次の画を続けて書く」に分類し，学習者が誤りを判別できるよう，誤りごとに異なる提示の使用方法で指摘を行う．誤り(1)，(3)は移動状態，誤り(2)は静止状態での提示である．音声提示は単一のブザー音から行われるが，提示されるタイミングにより誤りの判別を行うことができる．

2.2 筆順学習支援システムを用いた実験

システムを用いて，提示方法の違いによる学習効果の差，それぞれの特性についての調査を行った．被験者は大学生，大学院生 9 名とし，事前テストとして漢字 24 字を被験者に回答させた．その後システムを用いて，各提示方法により 8 字ずつ学習を行った．事前に誤り指摘が各提示方法により行われると周知した．学習後事後テストにより学習結果の確認を行った．

実験結果からいずれの提示方法でも同様な学習結果が確認できた．実験中における各誤りをした際の正しい画が書けるまで

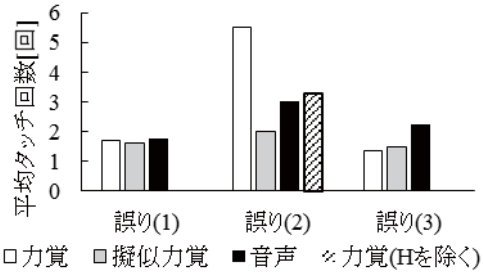


図 2 筆順学習支援システムを用いた実験における各誤りをした際の正しい画がかけるまでの平均タッチ回数

の平均タッチ回数を図 2 に示す. 図 2 から誤り(1)に関しては提示方法による違いは見られなかったが, 誤り(2), (3)に関しては提示方法による違いが見られた. 力覚提示と擬似力覚提示に着目して比較を行うと, 静止状態での提示(誤り(2))に関してのみタッチ回数に差が見られ, ここから静止状態での提示に関して提示方法による学習の差が生じる可能性があると考えられる.

3. 題材の検討

2 節で説明したシステムの学習題材である漢字の筆順では, 各提示方法による学習結果に差は見られなかった[河野 17]. これは筆順学習支援システムが学習者の知識を修正するものであり, こうした誤った知識の修正を行う場合, いずれの提示方法でも学習者は誤りを認識し, 修正することで学習直後は同様の学習結果が得られたと考えられる. 今回のような学習者が間違えた際に指摘を行う間接指導ではなく直接指導で学習を支援する題材の場合, 提示方法による学習効果の差が確認できる可能性がある. そのため学習題材の再検討が必要である.

力覚提示を用いて直接指導を行う題材としてスキル学習や物理実験などが考えられるが, スキル学習では力覚提示, 擬似力覚提示の与え方が学習効果に大きな影響を与えるため, 提示方法の与え方について十分に考察が必要で有り, 困難である. そのため本研究では物理実験を題材とする. 物理実験を題材としたシステムでは仮想物体の重さや動作を力覚提示, または擬似力覚提示で表現する事で学習の支援を行う.

2 節の実験から静止状態での提示に関して各提示方法に差が生じる可能性を確認できた. 2 節のシステムでは力覚提示, 擬似力覚提示ともに移動時の提示と静止時の提示が存在するが, 1 つの問題に対してその 2 つが混在しており, それら 2 つの提示による学習効果の調査を行うのは難しい. そこで新たな題材として移動時, 静止時の提示が並存し, 2 つの提示が独立した題材を選択する必要があると考えられる. そこで本研究では滑車実験を選択した. 滑車を題材とした際の問題を図 3 に示す. 図 3(a)は糸を引く際の力 A を求める問題, 図 3(b)は滑車の構成図の点 B にはたらく力を求める問題である. これらをシステム上で体感させる場合, (a)は移動時の提示, (b)は静止時の提示

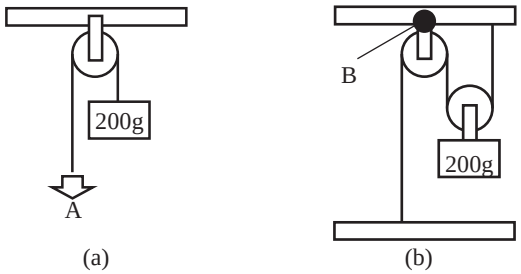


図 3 滑車の問題例

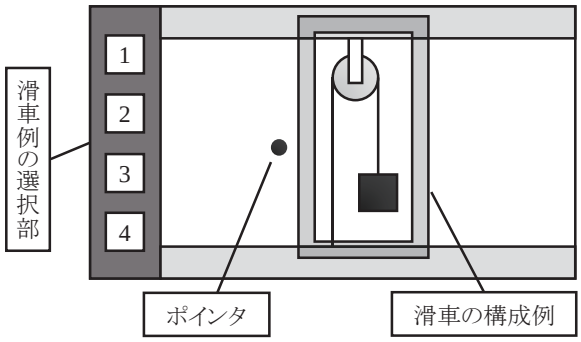


図 4 滑車学習支援システムの画面例

となる. このように静止時の提示と移動時の提示が存在するが 1 つの問題に対して 1 つの提示方法で学習を行い, 静止時の提示と移動時の提示は独立している. システムを用いた実験では, これら二つの問題の被験者の回答, 実験時の様子を中心に学習の差の検証を行う.

4. 滑車学習支援システム

本システムでは滑車の組み合わせによる糸を引く際の重さの変化や, 滑車の構成物にかかる力の大きさを力覚提示, または擬似力覚提示によって体感することが出来る. 図 4 に滑車学習支援システムの画面例を示す. 力覚提示は Phantom-omni を用いて行い, 擬似力覚提示は Phantom-omni のペン先の位置座標のみを用い, 画面内のポインタの移動量や位置を制御することにより行う. システムでは滑車の組み合わせの例を複数用意し, 学習者は用意された滑車例の構成物を選択することで, 構成物のある点ではたらく力の大小(静止状態での提示), 滑車の組み合わせによる糸を引く際の力の変化(移動状態での提示)を力覚提示, 擬似力覚提示を通して体感することが出来る.

本システムを用いた被験者実験は, 被験者を力覚提示で学ぶグループと擬似力覚提示で学ぶグループの 2 つのグループに分割し行う. 実験は事前テスト, システムを用いた学習, 事後テストの順で行う. 事前テスト, 事後テストはペーパーテストで行い, 事後テストはシステムを用いた学習の直後と, 一定期間後に再度行う. 提示方法による学習の差異は学習結果や, 提示回数, 学習時間などの学習時の様子から得る.

5. まとめ

本研究では力覚提示, 擬似力覚提示を学習に用いた際の各提示方法による学習の差異の調査を目的としている. 本稿では間接指導を力覚提示, 擬似力覚提示で行う学習支援システムを用いた実験と, 実験結果を踏まえた今後作成予定であるシステムとそのシステムを用いた実験について述べた.

参考文献

[柏原 15 年] 柏原 昭博, 塩田 剛: 擬似力覚呈示による知識構築支援, 電子情報通信学会論文誌, Vol. J98-D, No. 1, pp. 104-116 (2015)

[河野 17 年] 河野 貴範, 松原 行宏, 岡本 勝: 力覚及び擬似力覚提示機能を持つ漢字学習支援システムに関する研究, 日本教育工学会論文誌, Vol.41, Supplement, pp. 77-80 (2017)

[檜谷 15 年] 檜谷 直樹, 松原 行宏, 岡本 勝: タブレット PC とポータブルな反力デバイスを用いた滑車の仮想実験環境, 教育システム情報学会誌, Vol. 32, No. 3, pp. 220-225 (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

仮想空間上の灯明光源効果を用いた時代屏風の再現

Virtual Antique Folding Screen with Optical Effects of Sacred Light

津野 駿幸[*1] 稲本万里子[*2] 小長谷明彦[*1]
Toshiyuki Tsuno Mariko Inamoto Akihiko Konagaya

[*1] 東京工業大学情報理工学院
School of Computing, Tokyo Institute of Technology

[*2] 恵泉女学園大学
Keisen University

Sacred light is commonly used in a residence in Heian period, however it is not appropriate to use in museum now, because of fire prevention. Virtual antique folding screen with optical effects of sacred light enables us to appreciate the antique folding screen as appreciated by Heian noble people.

1. はじめに

近年，IT 技術と芸術を融合した新しい研究領域の試みとして，日本美術の代表となる源氏絵のオントロジーを用いた研究[1]および深層学習を用いた源氏絵の分析[2]などの研究が進められている．本研究では仮想空間技術を利用して，平安貴族が鑑賞したと思われる時代屏風の再現を試みる．時代屏風は背景に金箔が貼られており，平安時代には照明器具として灯明が用いられていた．一般に，芸術作品を灯明のような燃焼を伴う灯りで鑑賞することは安全性の観点から困難である．この問題を仮想空間上で灯明の灯りと金箔からの光源反射を忠実に再現することで解決する．

2. 時代屏風

2.1 作品の概要

本研究で用いる小長谷研究室所蔵の屏風は，縦 155.5 センチメートル，横 53.5 センチメートル(第一扇)，59.0 センチメートル(第二～五扇)，53.0 センチメートル(第六扇)の金地着色の六曲一隻屏風である．後述するように，屏風の主題が不明であるため，入手時の名称をそのまま用いて時代屏風と称す．

2.2 図様と表現

屏風の図様は，向かって右の第一扇から記述するのを通例とするが，この時代屏風は主要なモチーフが左に描かれているため，第五・六扇から述べる．第五・六扇下方には，檜皮葺の寝殿造の建物が大きく描かれている．巻き上げられた御簾の奥にいる人物は，繧繝縁の畳に坐り，顔を見せないところから，帝のようである．簀子と階の下には，黒の袍，青，緑，橙の直衣姿の男性貴族が控え，向かって右手には，鳥と蝶に扮した人物が向かい合っている．『源氏物語』胡蝶の場面を彷彿とさせるが，桜と山吹の持物がなく，龍頭鷁首の船も描かれていないため，主題は不明である．第四扇から第二扇にかけて幔幕が張られ，奥には楽人と雅楽で用いられる大太鼓，松と紅葉が描かれ，季節は秋であることがわかる．幔幕の手前には土坡が広がり，画面上方にはなだらかな山並みが見える．第一扇は，画面中央に塀，画面下方の土坡の奥に牛車が配される．

屏風全体に金箔が貼られ，金雲はない．人物の着衣には色鮮やかな絵具が用いられているが，人物の顔貌は土佐派や狩野派などの正系絵師のものとは異なり，松や紅葉も頼りなげである．画面の大きさに比べてモチーフが少なく余白が多いが，灯明を照らすとモチーフがシルエットになり，金箔が反射する．

2.3 絵師と制作年代

人物や樹木の表現から町絵師の手によるものであり，制作年代は江戸時代中期から後期と考えられる．

図 1 時代屏風（小長谷研究室所蔵）



表 1 課題と実現状況

再現内容・目標	実現
Unity 上で炎オブジェクトを作る	○
炎の大きさ・形・色・輝度を自由に変えられる	○
炎に動き(揺らぎ)を付けることができる	○
より現実に忠実な炎の再現のための評価指標がある	×
炎の再現の評価指標に添って実装をしている	×
Unity 上で時代屏風オブジェクトを作る	○
時代屏風の金箔部分に金属反射がある	○
金属反射の具合の評価指標がある	×
金属反射の評価指標に添って実装している	×
屏風の表面に立体感・微小な凹凸が見られる	×

連絡先：小長谷明彦，東京工業大学情報理工学院，
kona@c.titech.ac.jp

3. 時代屏風の仮想化

VR 技術を用いた屏風の再現例として，凸版印刷株式会社による”風神雷神図屏風”が知られている[3]. これは，金属光沢や表面の凹凸，照明環境や鑑賞する向きによって見え方を変化させる，といったことを再現した展示作品である. さらには，ロウソクの光による照明・月明かりによる場合の見え方など，実際の文化財鑑賞においては実現性に乏しい条件下での鑑賞を再現している. 本研究では時代屏風の鑑賞における「灯明の光源効果」に特に注目していることから，①照明器具である「灯明」の忠実な再現，②時代屏風の光反射効果，に着目した. 本研究の目標と現状を表1にまとめる.

3.1 実現法

仮想空間として 3D ゲーム開発用のプラットフォームである Unity を使用した. VR 用のデバイスとしては HTC 社の HTC Vive を用いた. まだ未検討である光源となる炎の明るさと金属反射の指標は輝度分布分析[4]を，炎の動き・揺らぎはモーション解析[5][6]をこれから用いるつもりである.

3.2 時代屏風の再現

本研究で用いる時代屏風は本研究室所有のものであり，高解像度カメラで時代屏風を撮り，イメージファイルを得た. 仮想空間上では屏風の型となるようなオブジェクトを作り，上述のイメージファイルを張り付けることにより，時代屏風オブジェクトを作成した. また，本研究の鍵となる時代屏風の背景の金箔部分を再現するため，Unity の機能の 1 つである material 機能を用いて金を表現した. さらに，背景となる金箔部分だけに金の material を反映させるため，屏風にレンダリングする 1 枚目は金の material を全体に反映させたイメージファイルを，そして 2 枚目としてその上に金箔部分を透明にした通常のイメージファイルを被せることで，金箔部分を再現した.

3.3 灯明の再現

時代屏風の照明器具となる灯明を再現するため，Unity 上で炎オブジェクトを再現する必要がある. そのため，Unity の機能の 1 つである Particle System を用いて「炎」を再現した. Particle System では，多数の Particle を発生させることができる. そして，その Particle の 1 つ 1 つに炎を表現するためのイメージをレンダリングすることにより，「炎」を再現している. また，Particle System の多彩な機能により，炎の大きさ・形・色・明るさなどを自由に再現することができる.

3.4 光源効果の再現

Unity には光源オブジェクトが用意されており，これにより仮想空間上で光源を再現することができる. 本研究で作成した仮想空間上では灯明の炎を構成する Particle の 1 つ 1 つに光源オブジェクトを付けているため，この灯明の炎によって様々な光源効果が得られる. 具体的には(1)発せられる光の色，(2)時代屏風オブジェクトの反射具合(特に金箔部分の金属反射)，(3)風オブジェクトによる炎の揺らぎ，といったような光源効果が再現可能である.

4. まとめと今後の課題

本研究は着手したばかりであり，時代屏風の再現にあたっては様々な課題が残されている. 重要な課題の一つは，「灯明の炎をより現実らしく振舞う」ことである. そのためには，実際の灯明が灯っている動画を解析し，灯明の炎のダイナミクス，すなわち，風・炎・反射に関する数理モデルの構築とパラメータ解析が必要となる. また，動画では得られなかった灯明の振る舞い，つまり起こりうる炎の振る舞いをどう予測するかについても十分な検討が必要である. さらに，時代屏風の表面の絵に現れる微小な凹凸を仮想空間上でどのように実現するかも重要な課題の1つである.

謝辞

本研究を推進するにあたり，貴重なご意見と議論を頂いた小長谷研究室ならびに源氏絵 DB 研究会の皆様に感謝を申し上げます. また，本研究の一部は科研研究課題 17H02295「オントロジーに基づく源氏絵データベースを共有・活用した源氏絵の総合研究」に支援されている.

参考文献

[1] 稲本万里子:オントロジーに基づく源氏絵データベースを共有・活用した源氏絵の総合研究, 科研費 17H02295

[2]加藤,稲本,小長谷:深層学習法による源氏絵の流派推定, 人工知能学会全国大会(2018)

[3]凸版印刷株式会社 “風神雷神図のウラ -夏秋草図に秘めた想い- ”, http://www.toppan-vr.jp/mt/showing/, (2018)

[4]鎌田慎也“デジタルデータ分布分析 – デジタル画像からの輝度分布分析”, UNISYS TECHNOLOGY REVIEW 第 87 号, (Nov. 2005)

[5]James Davis, Gary Bradski, “Real-time Motion Template Gradients using Intel CVLib”, IEEE ICCV Workshop on Framerate Vision, (1999)

[6]James Davis, Aaron Bobick, “The Representation and Recognition of Action Using Temporal Templates”, CVPR, (1997)

図２　灯明の光源効果と仮想空間上での再現例





The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

集団の協調パターン形成を目的とした仮想環境での視線行動の分析

Analysis of Gaze Behaviors in Virtual Environments for Cooperative Pattern Formation

渡邊紀文 *1　　糸田孝太 *2
Norifumi Watanabe　　Kota Itoda

*1産業技術大学院大学　　*2慶應義塾大学
Advanced Institute of Industrial Technology　　Keio University

In a goal type ball game such as soccer and handball, a plurality of players who can pass are searched for, and each player's intention is estimated and a player who can pass is selected. Furthermore, it confirms the position and behavior of enemies that exist around passable players, estimates their intentions, and determines teammate players who pass the pass most successfully. In order to realize instantaneous intention estimation and judgment subject to strong temporal and spatial constraints, cooperative patterns shared within the group are considered to exist. Therefore, in this research, we presented to subjects a first-person perspective of professional soccer players by using virtual environment, and analyzed the gaze behaviors pre- and post-training. Based on the results, we discuss that subjects switch their behavior by estimating the intention of otherplayers who pass and receive to subject by presenting the visual information based on the first-person perspective.

1. はじめに

我々は日常生活において他者の行動からその意図を推定しその結果に基づいて自己の行動戦略を決定する．日常生活の多くの場合においては，1 対 1 のコミュニケーションが主となり，自己の内部モデルと照らし合わせて他者の意図を推定する．しかし意図を推定すべき対象が複数存在する集団行動においては，着目すべき他者を決定し，その意図が推定できた段階で更に別の他者の意図を推定するといった連鎖的な意図の推定を行う必要がある．具体的にはサッカーやハンドボールといったゴール型ボールゲームにおいては，パス行動時にパスが可能な味方を複数探索し，それぞれの意図を推定してパスが可能な選手を選択する．更にパス可能な選手の周囲に存在する敵の位置や行動を確認してその意図を推定し，最もパスが成功する味方選手を決定する．このような時間的・空間的に強い制約を受けた瞬時の意図推定および判断を実現するためには，集団内部で共有される協調パターンが存在すると考えられる．そこで本研究では，このような人間の協調パターンをモデル化したエージェントの作成を目指し，サッカーにおいて協調パターンを形成することが重要であるパスシーンに着目した実験を行い，被験者が新たにパス行動を獲得する際の視線行動を分析して，その意図推定の過程を明らかにする．

実験では仮想環境を利用してプロのサッカー選手の一人称視点を訓練データとして与え，学習前後の被験者の視線行動を分析する．選手のポジション・トラッキングデータを元に協調行動を仮想環境で再現し，特定の場面を繰り返し提示する事でパス行動におけるボールホルダー及びレシーバーの選択行動を獲得し，その前後における視線行動の変化の分析とその評価を行う．その結果を基に，一人称視点による視覚情報の提示により，被験者がパスを出す選手およびパスを受ける選手の意図を推定して視線行動を切り替えることができるのか議論する．

連絡先: 渡邊紀文，産業技術大学院大学産業技術研究科，東京都品川区東大井 1-10-40，watanabe@aiit.ac.jp

2. 先行研究

協調行動を実現したエージェントモデルとして，ロボカップサッカーシミュレーションを利用した研究が行われている．秋山らはアクション連鎖探索フレームワークという探索木を作成し，オンラインで有効なアクションの連鎖をエージェントに実装したすることで，サッカー戦術の連動性について研究を行った [4]．また人間の視線行動から，団子状態を解消したチームワーク形成において必要なパラメータを抽出し，新たなサッカープレイヤーのエージェントモデルを構築することで，チームワーク形成の違いによる試合への影響を分析した研究が行われている [5]．このようにエージェントなどを利用したモデル研究では，データを元に繰り返し実験を行い，人間と同様な行動を再現することはできるが，協調パターン形成における選手の意図を推定することは困難である．

一方人の行動に関する協調パターンの研究では，試行間での行動決定過程を分析する必要があるため，協調パターンを獲得している選手および未獲得の選手それぞれの行動を繰り返し実施する必要がある．従来のスポーツ行動分析においては，実際のフィールドでチームメンバーによってミニゲーム形式で場面を再現する形式 [1] がとられ，参加するメンバーや天気によって状況は変化してしまい再現性が低い実験であった．ディスプレイを用いて表示する形式 [2, 3] も行われていたが，再現性は高い一方で，被験者の視線行動を基本として状況を切り替えるには 360 度に近い曲面ディスプレイ等の大掛かりな装置が必要である事や，敵や味方が自分の目の前にいるといった臨場感に乏しい事が問題となっていた．

このような刺激の再現性及び人間の意図推定の問題に対し，近年一人称視点で広視野角の仮想環境を提示可能なヘッドマウントディスプレイが安価で開発され，多くの認知訓練で利用されている [6]．これらのデバイスを利用する事で，選手の訓練に仮想環境が使われ，その有効性も示されている [7, 8]．このように集団行動における行動実験や分析において，仮想環境を用いた研究が広く行われるようになってきている．本研究においても，仮想環境を用いて一人称視点の入力による集団行動を被験者に提示することで再現性及び臨場感の問題を解決し，さらに繰り返し実験可能な仮想エージェントを利用することで

実際の場面に近い環境での行動変化を分析する．

3. HMD と仮想環境を用いた実験環境

実験では，頭部に装着する事で仮想環境を呈示する事が可能なヘッドマウントディスプレイ (FOVE 社製 FOVE 0）を利用した．HMD の性能は視野角が約 100 ° であり，ヘッドトラッキング機能に加えてアイトラッキングシステムが搭載されている．これにより，仮想環境を呈示しながら被験者の視線移動を分析することが可能となる．なお仮想環境については，Unity Technology 社製のゲームエンジン”Unity3D”を用いて構築した．

仮想環境において選手の役割を果たすエージェントはチーム毎に赤と白で色分けされ，フィールドを含むパラメータは実際の選手や試合の規定を基に作成を行った [11](図 2)．

実験は頭部と HMD の位置を揃え，被験者は選手の中の 1 体と視野を共有する事で行った．これによりフィールドを見下ろすような俯瞰映像ではなく，一人称視点で行動を決定する，より実際の場面での判断に近い形での行動を分析する事が可能である．また，被験者エージェントの身体を透過処理する事により自らの身体で周囲の物体が隠れてしまう事を防いでいる．仮想環境と実験の様子を図 1 に，また仮想環境における主なパラメータを表 1 に示す．

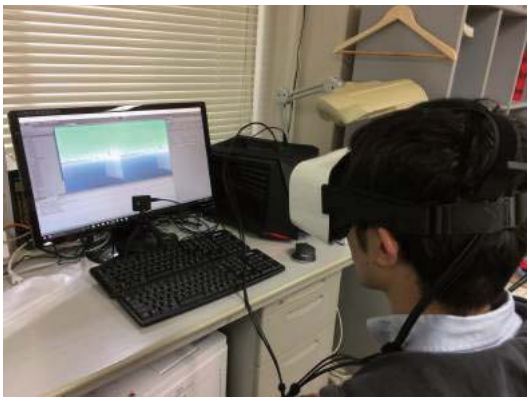


図 1: FOVE 0 を装着した被験者と実験環境 (被験者の目の前のディスプレイに表示されている画面と同じものが HMD に提示される)

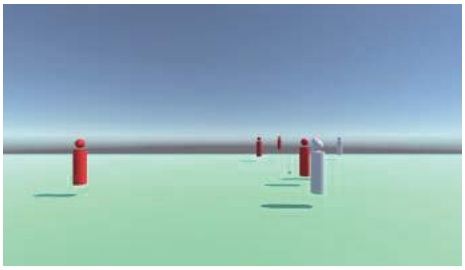


図 2: 被験者の一人称視点での仮想環境

表 1: 仮想環境の主なパラメータ設定（m）

ボールの直径	0.2	フィールドの横	105.0
選手の高さ	1.7	フィールドの縦	68.0
選手の幅	0.4	ゴールの幅	7.3
選手の頭の直径	0.3	ゴールの高さ	2.2

4. 視線行動に関する協調パターン獲得実験

4.1 実験に用いる場面

被験者に提示する場面は，FIFA Confederations Cup 2013 年のスペイン対イタリア戦で，前半において中盤のボール確保から味方内でのパスを含みシュートに至るまでの約一分間の場面である．仮想環境で再現するため，10fps で選手の位置を記録したポジション・トラッキングデータ及び動画データを用い，データには視線方向は含まれていないため，動画データを元に選手の頭部方向を用いて選手の視線方向を近似した．近年の研究では人の頭部方向から視線や注意の方向を近似できる事が示唆されている [12]．

被験者が視野を共有する選手は，周囲の選手の意図を推定し行動を切り替えるミッドフィルダー（MF）とし，攻撃の序盤から終盤へと繋がる各段階での視線行動を分析する．今回分析に用いたスペインチームはボールを保持し続ける戦術を持つ，パス回しが非常に優れているチームであり [13]，精度の良いパスを実現するため選手同士が意図を共有し，それに基づいた行動が瞬時に現れていると考えられる．

4.2 実験手順

実験では，仮想環境で作成した場面を繰り返し提示する事で被験者はミッドフィルダーのパス選択行動を獲得する．仮想空間におけるエージェントはトラッキングデータを基に自動で移動し，被験者は実験で指定したミッドフィルダーに対応するエージェントとして視線行動を取る．被験者は HMD を頭部に装着した状態で首を振る事で仮想環境における周囲の状況を自由に確認することができる．なお実験は被験者の安全のため着席状態で行った．

実験は図 3 に示すように,『事前フェーズ』,『学習フェーズ』,『事後フェーズ』の三つのフェーズに分かれている.『事前フェーズ』では被験者は頭部を動かし，対応の場面での状況を自由に確認する事ができる.『学習フェーズ』では被験者の視野範囲を固定し，実際の選手が行った視線行動を被験者に提示する．最後の『事後フェーズ』では，被験者は学習フェーズで選手が行っていた行動を基に頭部を動かす．被験者には，場面について事前に説明し，その場面にいた場合にどこに注意するのかを判断するよう指示し，全てのフェーズを三試行ずつ実施する．実験の後には，ボールゲーム経験や視線行動を取る時にどこに着目していたかといった簡単なアンケートを行った．被験者は二十代の大学生及び大学院生 3 名（被験者 1〜3 で行った．なお，VR 酔いを防ぐため各試行の間で最長 1 分程度の休憩時間を取っている．

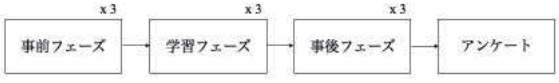


図 3: 実験手順

5. 事前・事後フェーズにおける視線移動の変化

FOVE のアイトラッキングの機能を用いて，被験者の視線の角度変化の差分について，各 3 回の試行における分散を出力した結果を図 4 に示す．角度の分散は，一周回るとまた同じ値として戻ってくるという特殊な性質を持つため通常のように観測値の総和を用いて平均，分散を求める事ができない．そこで式 1, 2 のように三角関数を用いて表現する方法が用いられる．

$$(R\cos\bar{\theta}, R\sin\bar{\theta}) = \frac{1}{N}(\sum\cos\theta, \sum\sin\theta) \quad (1)$$

$$V = 1 - R \quad (2)$$

ここで，$\bar{\theta}$ は平均，V は分散を，N はデータ数を表し，角度が同方向に揃う程動径 R が単位ベクトルに近づき，揃わない程動径 R は 0 に近づく事から，式 2 のように 0 から 1 の間の値として分散の大きさを表現できる．

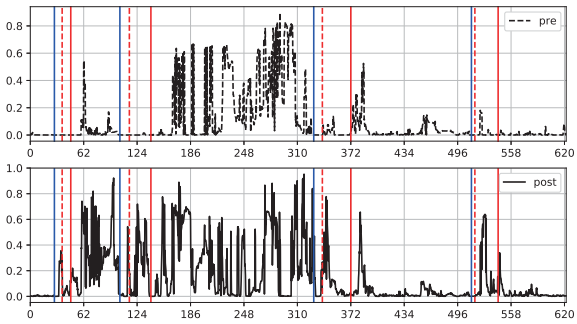


図 4: 事前（上）・事後（下）での被験者 1 の視線の角度変化の分散．味方選手が被験者に対してパスを出すタイミングを青色実線，被験者のボールタッチのタイミングを赤色点線，被験者がパスを出すタイミングを赤色実線で示す．

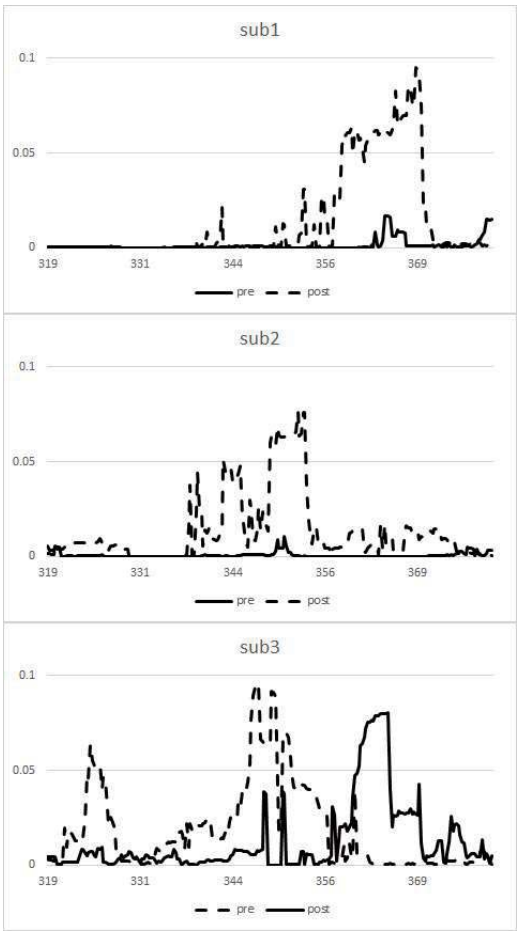


図 5: 事前（pre）・事後（post）での被験者 1，2，3 の視線の角度の分散．呈示した映像中で味方とのパスの受け渡しのある 319-379frame の値を示す．

図 4 より，事前フェーズでは異なる注視点を連続して見続けていたのに対し，事後フェーズでは短い時間ではあるが分散が小さくなる回数が多くなり，同じ注視点を確認している傾向が見られる．これにより非常に狭い範囲で視線を移動して自分の周囲を確認していると考えられる．味方選手がパスを出す前，被験者がボールを受け取る前，ボールを味方に出す前それぞれの 1 秒間の分散の試行間平均については，事前フェーズと事後フェーズで有意な差はみられなかった．

次に各 3 回の試行における被験者 3 名の視線の角度の分散を図 5 に示す．なおここでは味方選手が被験者に対してパスを出し，被験者が更に別の選手にパスを出す 319〜379fram までの値を抽出して表示する．グラフより，事前フェーズでは各 3 回においてほぼ同じ部位を注視しているが，事後フェーズにおいては注視する部位が異なっている．よって事後フェーズでは，狭い範囲で細かく視線を移動する探索的行動をとっていると考えられる．

6. 考察

事前・事後フェーズにおいて試行間の角度変化による分散の違いは見られないが，視線が一致するタイミングが増加し，狭い範囲で注視していると考えられる．本結果は事前フェーズでは頭部を移動して視野を大きく変化させているため，そこで注視する部位も変化し分散が大きくなるが，事後フェーズでは MF の視野範囲と共通しているためそこで注視する部位が明確となり，分散が小さくなるタイミングが増加したと考えられる．これにより，特定の注視点から得られる情報が増加し，他の選手行動の分析及びその意図の推定が可能になったと考えられる．

味方選手が被験者に対してパスを出し，被験者が別の選手にパスを出すタイミングにおいては，事後フェーズでは各試行での分散が大きくなった．これは事前フェーズにおいては被験者は視覚的注意を行わず周辺視などを利用して空間全体の状況を確認しているのに対し，事後フェーズではパスを出す味方の選手を探索し，その選手の行動を中心視によって確認して細かく視線を移動しているためであると考えられる．

そこで次に被験者が実際に注視していた選手を分析する．図 6 はボールを受け取った被験者が更に別の選手にパスを出す 319〜379fram までの値を抽出しており，このタイミングでは被験者は次に味方の FW 選手にパスを出している．被験者の視線からの FW 選手 2 名の相対角度を確認すると，340frame 前後までは FW1 および FW に両名をほぼ同じ角度で見ているが，340frame からそれぞれ視線を細かく移動し，350frame 以降は FW1 の選手の相対角度が 0 に近づき中心視で判断している傾向が見られる．このように学習フェーズを経ることで，

被験者自身が協調パターンを獲得し，パス行動における能動的な探索行為を行っていると考えられる．



図 6: 被験者 1 の視線からの他の FW 選手（1,2）の相対角度（rad）．呈示した映像中で味方とのパスの受け渡しのある 319-379frame の値を示す．

7. おわりに

本研究では，仮想環境を利用した一人称視点での呈示により，サッカー集団行動での視線移動を分析し，協調パターン形成に関する意図推定の過程を明らかにした．実験結果より，プロのサッカー選手の視線行動を共有することで，探索すべき視野範囲を限定し，意図を推定すべき選手を限定して視線を向けていた．更に探索すべき視野範囲の中で細かく視線を移動し，実際に複数の FW の選手を中心視で確認し，その意図を推定しようと試みたと考えられる．これらの行為を実践することができた理由としては，学習フェーズにおいて本実験で呈示したパス行動の協調パターンを獲得し，それに基づいて被験者が能動的な探索を行ったためであると考えられる．

本研究の視線移動に関しては，被験者数がおよび分析するシーンが限定されているため，今後実験を追加して分析する必要がある．また実際の集団行動においては，自己の行動の変化により他者がその意図を推定して行動を変化する．本実験では他者が自己の視線行動に基づいて行動を変化させる状況を実現することはできていないため，今後協調パターンに基づいて行動を変化させるエージェントモデルを構築し，被験者との相互作用を分析する必要があると考えている．

8. 謝辞

本研究は科学研究費助成事業若手研究（B）（課題番号 16K21442）の助成を受けたものである．

参考文献

[1] 中川昭: ボールゲームにおけるゲーム状況の認知に関するフィールド実験：ラグビーの静的ゲーム状況について, 体育學研究, Vol. 27, No. 1, pp. 17-26, 1982

[2] 中本浩揮, 杉原隆, 及川研: 知覚トレーニングが初級打者の予測とパフォーマンスに与える効果, 体育學研究, Vol. 50, No. 5, pp. 581-591, 2005

[3] W. Lee, Tomohiko Tsuzuki, Masato Otake, and Osamitsu Saijo: The effectiveness of training for attack in soccer from the perspective of cognitive recognition during feedback of video analysis of matches, Football Science, Vol. 7, No. pp. 1-8, 2010

[4] 秋山英久 ：アクション連鎖探索によるオンライン戦術プランニング， 人工知能学会研究 資料，SIG Challenge B101-6，2012

[5] 矢作拓也，渡邊紀文：人間行動データに基づいたロボカップサッカーシミュレーションにおけるチームワークの形成，2010 年度人工知能学会全国大会（第 24 回），ROMBUNNO.1C5-1，2010

[6] 近藤敏之, 長嶺伸, 大村優慈, 矢野史朗: 没入型ヘッドマウントディスプレイの認知心理学実験への活用事例，日本神経回路学会誌，Vol. 23, No. 3 pp. 87-97, 2016

[7] Benoit Bideau, Richard Kulpa, Nicolas Vignais, Sebastien Brault, Franck Multon, and Cathy Craig: Using virtual reality to analyze sports performance, Computer graphics and applications, IEEE, Vol. 30, No. 2, pp. 14-21, 2010

[8] Helen C. Miles, Serban R. Pop, Simon J. Watt, Gavin P. Lawrence, and Nigel W. John: A review of virtual environments for training in ball sports, Computers & Graphics, Vol. 36, No. 6, pp. 714-726, 2012

[9] Oculus Rift DK2. https://www.oculus.com/en-us/dk2. Accessed: 2017-07-04

[10] FOVE 0. https://www.getfove.com/. Accessed: 2017-07-04

[11] Federation Internationale de Football Association, et al. Laws of the game. FIFA., 2015

[12] Ryoichi Nakashima and Satoshi Shioiri: Facilitation of visual perception in head direction: Visual attention modulation based on head direction, PloS one, Vol. 10, No. 4, pp. e0124367, 2015

[13] Ian Ladyman. World cup 2010: Beat spain? it's hard enough to get the ball back, say defeated germany. http://www.dailymail.co.uk/sport/worldcup2010/article-1293239/WORLD-CUP-2010-Beat-Spain-Its-hard-ball-say-Germany.html, 7 2010. Accessed: 2017-07-04.

[14] Alexandre Hervieu, Patrick Bouthemy, and Jean Pierre Le Cadre: Trajectory-based handball video understanding, In Proceedings of the ACM International Conference on Image and Video Retrieval, pp. 43. ACM, 2009

[15] Kota Itoda, Norifumi Watanabe, and Yoshiyasu Takefuji: Model-based behavioral causality analysis of handball with delayed transfer entropy. Procedia Computer Science, Vol. 71, pp. 85-91, 2015

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

実物体の把持移動に適合する AR 描画システムの構築と評価

Development and Evaluation of a AR Drawing System suitable for real time gripping movement by using Kinect

稲留 広貴
Hiroki Inatome

曽我 真人
Masato Soga

和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University

Virtual human figures have been used in the conventional human figure sketch learning support system. A learner selects one of the virtual figures for a motif. However, the learner could not intuitively manipulate the virtual human figure in real time to change the posture of the figure for learner's favorite pose. Therefore, in this research, we propose a method to change the posture of the virtual figure model intuitively by using the drawing doll as a tangible interface. Specifically, it uses PCL which can acquire point cloud data from RGB-D camera with KINECT, then it acquires and tracks the three-dimensional coordinates of the real object and superimposes and displays the virtual human figure on the real drawing doll. In the verification experiment tracking accuracy was verified and the improvement of the system was discussed.

1. はじめに

1.1 研究背景

近年でのスキル学習支援に関する研究は，技能継承だけでなく災害学習やリハビリテーションなどの分野においても広く行なわれている．また，現実空間を仮想物体によって拡張する AR(Augmented Reality)の技術が一般化するにつれて，AR 分野での研究も行われるようになってきている．そこで，AR との親和性の高い学習支援として，人物画スケッチ学習支援システムが挙げられる．しかし，従来の人型の仮想物体を利用した人物画スケッチ学習支援システムでは，事前に用意された人型モチーフを選びスケッチを行うというもので，モチーフに対しリアルタイムに自分で好きなポージングをさせることが出来なかった．この問題を解決することによって，より直感的な操作が可能なモチーフ姿勢決定システムが実現できる．

1.2 先行研究

従来の人物画スケッチ学習支援システムに関する研究として，山田卓らが 2011 年に発表した「視点を自由に設定できる人物画の輪郭線スケッチ学習支援環境の構築」が挙げられる(1)-(3)．この研究は，本研究と同じく人物画スケッチ学習支援環境の構築を目的としている．この研究では，3DCG の人型モチーフを自由な視点から構図を決定し，人物画スケッチの骨格と輪郭線の診断を可能にした．しかし，このシステムでは事前に用意された複数のモデルの中から，好きなポージングのモデルを選ぶというもので，自分で好きなポージングを作り出すことが出来ないという，直感的なモデル操作における課題があった．

2013 年に Dong Woo Seo らが発表したシステムでは，マーカレスで AR 空間を作成し，Kinect の深度カメラと点群データを利用して手の部分を抽出することで仮想物体への接触判定を実現した(4)．しかし，仮想物体に対する直感的な把持移動には対応していなかった．

連絡先：稲留 広貴,和歌山大学システム工学部
s195003@center.wakayama-u.ac.jp

1.3 先行研究

本研究では，1.2 節で挙げた先行研究の問題点を踏まえて，学習者がデッサン人形をタンジブルインタフェース(5)として用い，それをリアルタイムに好みのポーズに変化させ，仮想のモデルの姿勢を変化させ，そのモデルを AR で表示する手法を提案する。そして，その試作システムの構築と評価を行う．

構築するシステムは，試作段階として 3D モデルの位置・姿勢に着目して構築する．人物モデルのポージングを操作するシステムを構築していく上で，操作後の 3D モデルの位置と姿勢の評価が必要となるため，試作段階として採用した．

実物体をトラッキングし，3D モデルを重畳表示することによって直感的なモデル操作が可能な AR 描画システムを構築することを構築することを目標とする．

2. システム設計

2.1 提案手法

前章 1.3 節で挙げた目標を達成するために，以下のことを行う．

- 実物体と，それに近い造形の 3D モデルを用意する
- 実物体を対象物体として認識し，トラッキングさせる
- 対象物体の参照点群に 3D モデルを重畳表示する

把持動作に適応させるために，実物体と実物体に近い形状の 3D モデルを使用する．また，3D モデルだけではなく実物体も使用することで，3D モデルの位置や姿勢を重畳表示先の実物体の姿勢から計測することが可能となる．これにより，前章で挙げた問題点にあった，既存モチーフの視点決定を PC 上で行う必要がなく，マーカレス AR で正しい位置・姿勢を推定することが出来る．

2.2 システム概要

本システムでは人物画スケッチ学習支援への応用を鑑みて，実物体にデッサン人形を使用している．3D モデルを実物体の位置・姿勢に追従させるためには，対象物体の位置情報や法線の向きを取得しなければならない．今回は物体の把持移動による位置・姿勢の変化に対応させるために，深度センサを搭載

した Kinect を使用する. このデバイスによって現実空間の映像を, 深度を含めた3次元データとして取得することができる. 本システムの物理構成は図1である. 対象物体認識の精度を考慮し, Kinect と対象物体は1m程離し何もない平面上に配置している. 実物体と仮想物体を適応させるために RGB-D カメラの映像を点群データとして取得することが出来るライブラリ PCL を用いる(6)(7). これにより対象物体を認識し, 対象物体の点群データを取得する. そして, 対象物体の点群の位置と法線ベクトルを取得することによって実物体の位置と姿勢を仮想物体と対応させる.

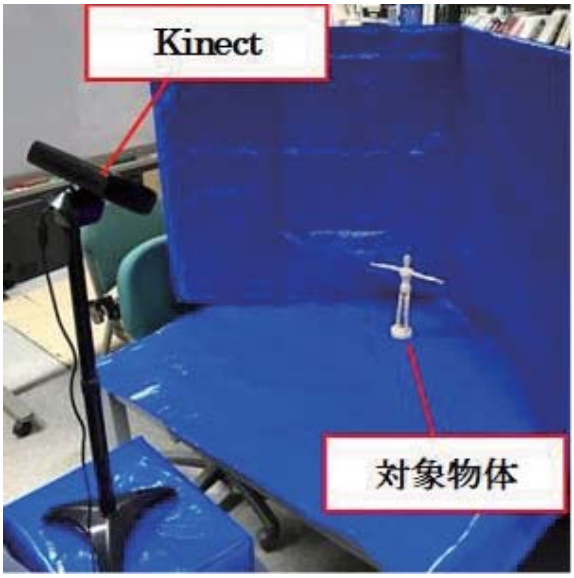


図 1　物理構成

位置と法線ベクトルを取得した対象物体に, 3D モデルを重畳表示するためには 3D モデルの点群データと法線ベクトルも取得・計測する必要がある. 本研究では 3D モデルを用意するにあたり, 3 次元点群処理が可能な 3D データ編集ソフトの MeshLab を使用した(8). 図 2 は本研究で対象物体に頂上表示する 3D モデルであり, 図 3 は MeshLab を用い法線ベクトルを可視化したものである.

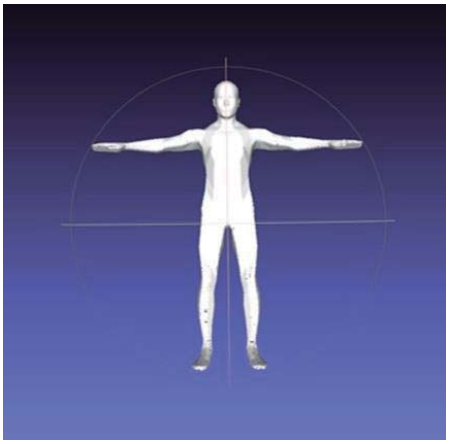


図 2　人型モチーフ



図 3　算出した法線

2.3 システム実装

本システムでは大別して以下の 2 種類の処理を実装した.

① 対象物体のリアルタイムな物体追跡
② 対象物体への 3D モデル重畳表示

対象物体を追跡するには, まず対象物体を分割化されたモデルとして Kinect から取得する必要がある. 取得したモデルの点群データをトラッカーとして設定し, 対象物体の追跡を行う. 物体追跡の処理の流れは, まず以前の粒子の位置と回転の情報を利用して次のフレームでそれぞれの位置と回転を予測する. 次に, これらの粒子の重さを下記の尤度式で計算する(9).

$$L_j = L_{distance}(\times L_{color})$$

$$w = \sum L_j$$

最後に, 深度センサから取得した実際の点群データを予測粒子と比較し, 粒子を再サンプリングする. この処理の際に得られる, トラッカーとして設定した分割化されたモデルの参照点群にテンプレートとして用意した人型 3D モデルの点群を重畳表示する. 図 4 は分割化されたモデルの参照点群であり, 図 5 は参照点群に 3D モデルの点群を重畳表示したものである. 本システムでは, 対象物体に追跡する仮想物体の情報のみを必要とするため, 重畳先のモデルの参照点群は描画していない.

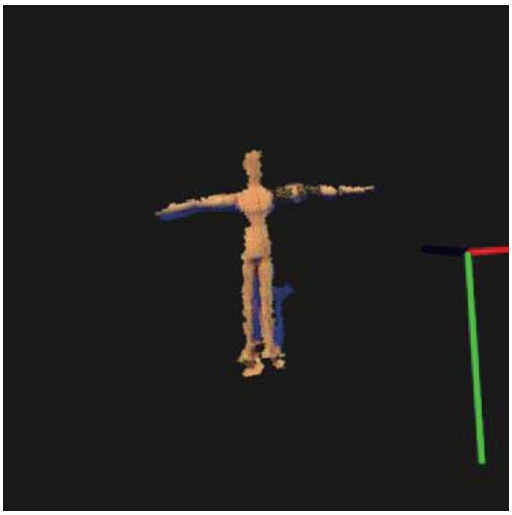


図 4　分割化されたモデルの参照点群

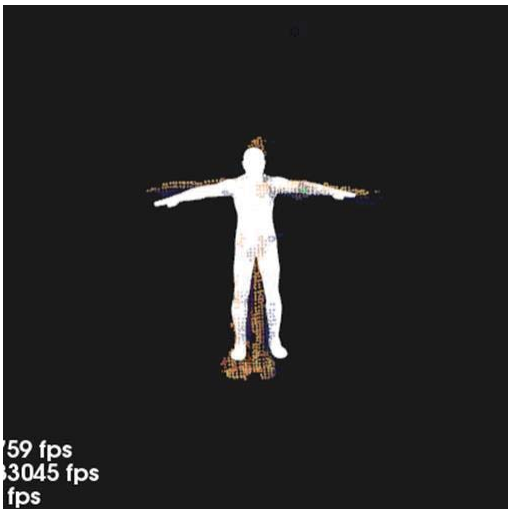


図 5　重畳表示された 3D モデル

3.　検証実験

検証実験は以下の項目を実施した.

① 対象物体の認識限度の検証
② 物体移動追跡の精度検証
③ 物体姿勢追跡の精度検証

3.1 予備実験

本システムでは人物画スケッチ学習システムへの応用のため, 対象物体にデッサン人形を使用した. しかし, 形状の異なる対象物体を用いた場合の挙動を考慮し, 対象にデッサン人形を用いることの妥当性を検証するために予備実験を行った. 予備実験で行う動作は本実験の動作と同様である. 予備実験では, デッサン人形の他に, 立方体と球を用意した. また, 立方体と球に近しい形状の 3D モデルとしてウサギ型のモデルを使用した.

予備実験の結果, デッサン人形を使用した場合は3次元的な移動と回転の両方に適合していることが分かった. しかし, 球と立方体は移動には適合していたものの, 回転には適合し難いということが分かった. これは, 表面の法線ベクトルが一定であり, 面の変化に乏しい形状であるために, 粒子の回転の追跡を誤認してしまうことが原因であると考えられる.

3.2 検証の内容

本システムは物体の把持移動に適合する必要性があったため, 認識限度の検証実験を行った. この実験では対象物体の一部を隠す動作を繰り返し, その時の仮想物体が正しく追従できている認識の限度を検証する.

②, ③の精度検証では, 条件を合わすために, 対象物体の配置地点を揃えた. また, 計測をシステム起動した後対象動作を行うことを1セットとしてそれを 20 回ずつ行った.

物体移動追跡の精度検証では, 対象物体を3次元的に移動させた時に, 仮想物体が対象を見失うことなく正しい位置で追跡できているかを検証する. 図 6 は位置精度検証実験中のシステム画面である.

物体姿勢追跡の精度検証では, 物体を開始地点から動かさず 360 度回転させた時に, 仮想物体が対象の回転に正しく追跡できているかを検証する. 図 7 は姿勢精度検証実験中のシステム画面である.

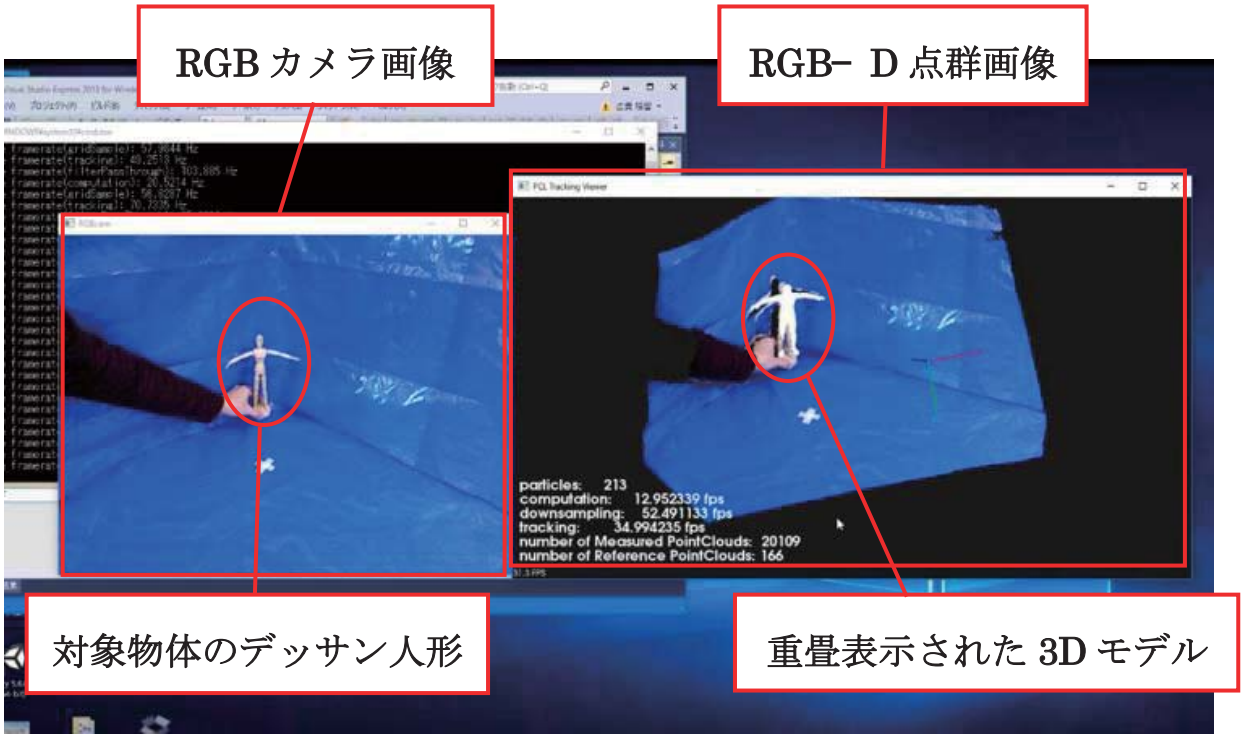


図 6　位置精度検証時システム画面

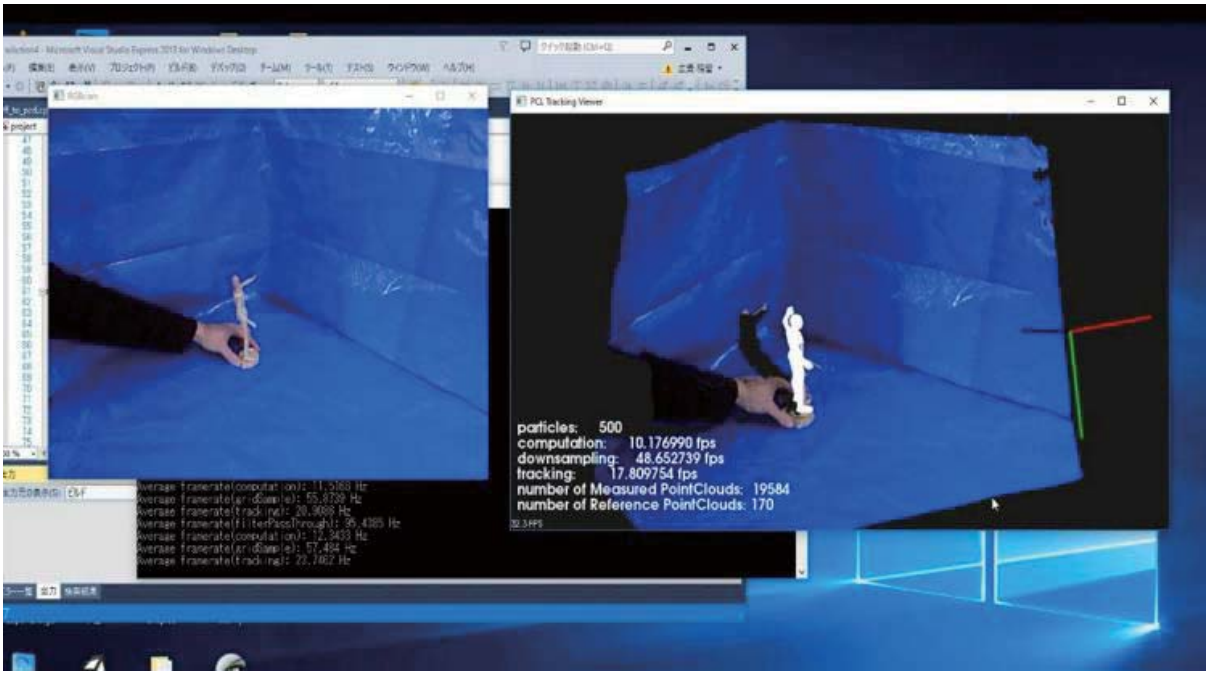


図 7　姿勢精度検証時システム画面

4. まとめ

従来の人型の仮想物体を利用した人物画スケッチ学習支援システムでは，学習者がリアルタイムで人型モチーフを直感的に操作して好みのポーズに決めることは出来なかった．本研究では，従来の人物画スケッチ学習システムの問題を解決するために，対象となる実物体をタンジブルインタフェースとして操作することによって仮想物体を操作する手法を提案し，その試作システムの構築と評価を行った．試作段階として対象動作を把持移動時のリアルタイムな位置，姿勢としてシステムを構築した．実物体を追跡対象として認識するために，RGB-D カメラである Kinect と点群処理用ライブラリ PCL を用いた．

システム完成後，予備実験と検証実験を行った．予備実験では，把持動作の認識精度と物体の形状による挙動の差について検証し，検証実験でデッサン人形を用いる妥当性を確認した．検証実験では，位置精度検証と姿勢精度検証の 2 種類を行い，本システムが位置と姿勢の追跡に適合していることを確認した．また，予備実験と検証実験を通して本システムの改善点を考察した．

今後は，本システムを人物画スケッチ学習支援システムへ発展させていく予定である．

参考文献

(1) 山田卓,曽我真人,瀧寛和,”視点を自由に設定できる人物画の輪郭線スケッチ学習支援環境の構築”,人工知能学会全国大会論文集, 2011,
https://kaigi.org/jsai/webprogram/2011/pdf/285.pdf

(2) 山田卓,曽我真人,瀧寛和,”骨格と輪郭線を診断する人物画の学習支援環境の構築”，情報処理学会 インタラクション 2012,
http://www.interaction-ipsj.org/archives/paper2012/data/Interaction2012/interactive/data/pdf/1EXB-36.pdf

(3) Masato Soga, Suguru Yamada, Hirokazu Taki, Development of a Learning Environment for Human Body Drawing by Giving Error Awareness for Bones and Contours 12th International Conference on Intelligent Tutoring Systems, ITS2014, pp.640-643

(4) Dong Woo Seo, and Jae Yeol Lee. “Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences”, Expert Systems with Applications, Volume 40, pages 3784 – 3793, 2013

(5) Ishii, H. and Ullmer, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, United States, March 22 - 27, 1997). S. Pemberton, Ed. CHI ‘97. ACM, New York, NY, 234-241,
DOI: http://doi.acm.org/10.1145/258549.258715

(6) Tsukasa Sugiura, Point Cloud Library 1.8.0 has been released, http://unanancyowen.com/pcl18/

(7) Radu Bogdan Rusu, Steve Cousins,” 3d is here: Point cloud library (pcl)”, Robotics and automation (ICRA), 2011 IEEE International Conference on, 1-4

(8) MeshLab, http://www.meshlab.net/

(9) Point Cloud Library, Tracking object in real time, http://pointclouds.org/documentation/tutorials/tracking.php

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

好意の返報性を表出するエージェントがユーザの親密度に与える効果

An Effect of User's Degree of Intimacy on Agent that Expresses Requital of a Favor

早瀬光浩 *1 Mitsuhiro Hayase　中村太郎 *2 Taro Nakamura　加納政芳 *2 Masayoshi Kanoh

*1豊橋創造大学 Toyohashi Sozo University　*2中京大学 Chukyo University

In this study, we verify an effect of user's degree of intimacy on agent that expresses requital of a favor. We implement an agent with non-verbal behaviors that gives favor like a smile or a movement to tuck hair behind the ear. These actions are performed while the agent is speaking. We take a questionnaire survey on the user's degree of intimacy both cases with and without these non-verbal behaviors.

1. はじめに

近年，スマートフォンが急速に普及し，世帯保有率が 2016 年には 71.8%と高い値を示し，特に，20〜40 歳代の世帯では，90%を超える保有率となっている [総務省 16]．スマートフォンには，人の操作を支援するエージェントが搭載されている．エージェントがパートナーとしてユーザと親密になることは，ユーザエクスペリエンスを高めることとなり，より高度な情報提供や支援が可能となる．このことから，エージェントがユーザから好意を得ることは重要と言える．

心理学の分野では，人からの好意は，自己開示，単純接触，同調行動，アイコンタクト，好意の返報性などの行動から得られると示されている [齊藤 05]．これまでのパートナー型ロボットやエージェントによる研究では，自己開示 [小川 05]，単純接触 [Robert 68]，同調行動 [村中 04] やアイコンタクト [中沢 05] などの手法が取り入れられている．しかし，好意の返報性に着目した研究は見当たらない．

好意の返報性とは，相手から好意を寄せられると自分も相手に好意を寄せようとする心理のことを指す．好意を与える非言語行動は，好意の返報性により会話後の親密化願望に対して増加に影響を与えることが示されている [仲嶺 13]．このことから，好意を与える非言語行動により，自身の好意を相手に伝えることで，返報性の原理により，相手の好意を向上させることが可能と考えられる．

本研究では，好意の返報性を表出するエージェントの行動モデルがユーザの親密度に与える効果を検証する．

2. エージェントの行動モデル

エージェントの開発には，MMDAgent[MMDAgent] を使用する．

男性は，会話前に会話相手から外見的魅力を感じるほど，会話後の親密化願望が高くなることから，エージェントは女性モデル [MMD 16] を使用する．

[仲嶺 13] の研究により笑顔は好意の返報性によって好意を向上させることが明らかになっている．このことから，好意の返報性を表出する行動として笑顔（3 種類）を，さらに，相手に興味があるという心理を示す行動として髪を耳にかける行動を実装する．また，お辞儀，うなずき，首をかしげる好意とは関係がない行動も実装する [中村 18]．図 1 に好意のある非言語行動を示す．



図 1: 好意を与えるとされる非言語行動 [中村 18]

3. 実験

3.1 実験設定

非言語行動は，エージェントが会話中に行い，行動と会話内容が不自然とならないようなタイミングで表出させる必要がある．エージェントが行う会話は，エージェントの自己紹介とする．会話シナリオでは，好意を伝えるような内容は排除する．会話の中では，エージェントが対話者に対し質問を行う．質問は，クローズドクエスチョンとし，対話者が「はい」か「いいえ」で答えられるようになっている．会話は，エージェントが質問を行い，対話者が回答し，その回答に応じてエージェントが返答を行うことで進んでいく．笑顔などのリアクションが取り入れやすいという点から，エージェントが返答を行う際に非言語行動を行う．具体的には，喜び，尊敬の気持ちをエージェントが感じていると推測されるときは笑顔のいずれか 1 種類を表出し，エージェント自身のこと，もしくは相手を気遣う発言をしているときは髪を耳にかける行動を表出させる．図 2 に，実験環境を示す．被験者は，マイク付きヘッドホンを装着し，モニターに映るエージェントと会話をしてもらう．

連絡先: 早瀬光浩，豊橋創造大学，愛知県豊橋市牛川町松下 20-1，0532-54-2111，m-hayase@sozo.ac.jp

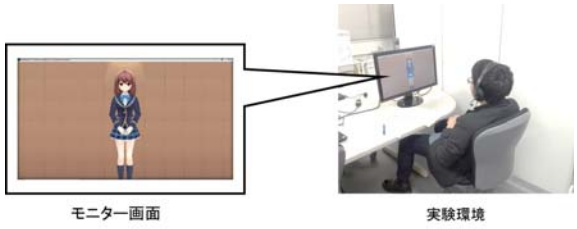


図 2: 実験環境

- 親密性
  - 彼女とうまくコミュニケーションがとれた
  - 彼女といて心地よかった
- 情熱
  - 彼女は理想的に見えた
  - 彼女は個人的にとても魅力的にみえた
- コミットメント
  - 彼女を大事にしようと思えた
  - もう一度彼女と話してみたいと思った

図 3: アンケート項目

3.2 被験者と評価方法

被験者には，会話中に好意を与える非言語行動を行うエージェントと行わないエージェントの 2 つを体験してもらう．

被験者は，男子学生 20 名で行った．カウンターバランスを考慮し，好意を与える非言語行動を行うエージェントを体験した後に行わないエージェントを体験する学生を 10 名，逆の順番で体験する学生を 10 名と分けた．

評価は，スタンバーグの愛のトライアングル理論 [吉田 07] の測定尺度を使用した．図 3 にアンケートの内容を示す．アンケート項目は，内容がエージェントに関係ないものを除外し，各要素（親密性，情熱，コミットメント）から 2 つずつの計 6 項目とした．アンケートへの回答は，1 つ体験するごとに行ってもらう．各項目は，7 段階のリッカート尺度で評価してもらう．

4. 結果

アンケート項目の各要素で平均を算出し，評価を行った．図 4 に実験結果を示す．図中の「好意アリ」が好意を与える非言語行動を行うエージェントを,「好意ナシ」が非言語行動を行わないエージェントを体験した場合の評価を示している．全ての要素で,「好意アリ」の方が「好意ナシ」よりも高い値を得ることができた．各要素について対応あり t 検定を行った結果，情熱の要素で有意傾向が確認できた．

情熱の要素で有意傾向が確認できたことは，恋愛関係の中で強い情動に繋がる動因，身体的魅力や性的魅力が影響力を持つと考えられる．また，髪を耳にかける仕草が関係すると推測される．

5. おわりに

本研究は，好意の返報性を表出するエージェントの行動モデルがユーザの親密度に与える効果の検証を行った．実験結果より，エージェントが好意を与える非言語行動を行うことで、ユーザの親密度を高められる可能性が示唆された．今後，髪をかける動作の有無で実験を行いたい．

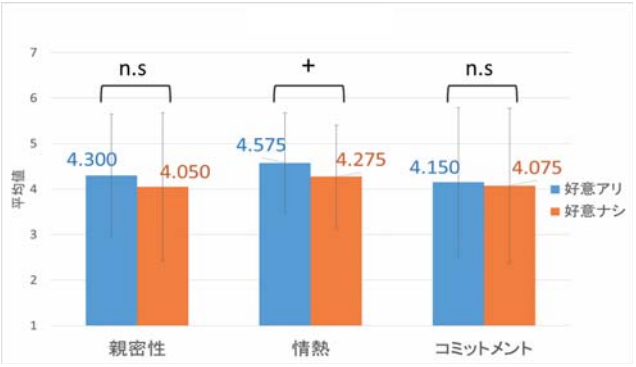


図 4: 実験結果

参考文献

[総務省 16] 総務省: 平成 28 年通信利用動向調査報告書（世帯編） (2016).

[齊藤 05] 齊藤勇: 恋愛心理学, ナツメ社 (2005).

[小川 05] 小川義人, 菊池英明: 自己開示する情報がエージェントのパーソナリティ印象に与える影響, Human-Agent Interaction Symposium 2011, I-1A-3 (2011).

[Robert 68] Robert. B. Zajonc: Attitudinal Effects of Mere Exposure, Journal of Personality and Social Psychology, 9(2, Pt.2), pp.127 (1968).

[村中 04] 村中彬人, 竹内勇剛: 複数エージェントとのインタラクションにおけるユーザの同調行動, 電子情報通信学会技術研究報告, HCS ヒューマンコミュニケーション基礎, 103(742), 6570 (2004).

[中沢 05] 中沢正幸, 西本卓也, 嵯峨山茂樹: 視線制御モデルによる擬人化音声対話エージェントの制御, 人工知能学会全国大会, 3B2-07 (2005).

[仲嶺 13] 仲嶺真, 大坊郁夫, 松井豊: 初対面異性間における対人魅力と会話行動が親密化願望に及ぼす影響, 筑波大学心理学研究, 46, 4956 (2013).

[MMDAgent] MMDAgent, http://www.mmdagent.jp/

[MMD 16] MikuMikuDance スターターパック DX, 晋遊舎 (2016).

[中村 18] 中村太郎, 早瀬光浩, 加納政芳: 好意の返報性を表出するエージェントの行動モデル, 第 44 回東海ファジィ研究会, P1-03 (2018).

[吉田 07] 堀洋道（監）, 吉田冨二雄（編）: 心理測定尺度集 II 人間と社会のつながりをとらえる＜対人関係・価値観＞, サイエンス社 (2007).

[2K1] [General Session] 13. AI Application

座長:高橋 大志（慶応大 ）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room K (3F Azisai Mokuren)

[2K1-01] The customyzation of neural machine translation with user's dictionary

○Shota Yamauchi[1], Yohei Murakami[2] ,Takao Nakaguchi[3], Toru Ishida[4] （1. Kyoto University, 2.Unit of Design, Kyoto University, 3.Department of Web Business Technology, Kyoto College of Graduate Studies for Informatics, 4.Department of Social Informatics, Kyoto University）

9:00 AM - 9:20 AM

[2K1-02] Dialog Summarization using Phrase Structure Analysis and Text Clustering

○Ayako Hoshino[1], Itaru Hosomi[1] （1. NEC Corp.）

9:20 AM - 9:40 AM

[2K1-03] Language generation system for sightseeing guidance using neural language model

Kazuya Ikuta[1], Seitaro Shinagawa[1], ○Koichiro Yoshino[1,3], Yu Suzuki[1,2], Satoshi Nakamura[1,2] （1. Graduate School of Information Science, Nara Institute of Science and Technology, 2. Data science center, Nara Institute of Science and Technology, 3. PRESTO, Japan Science and Technology Agency）

9:40 AM - 10:00 AM

[2K1-04] Analysis of News and Stock Returns in Korean Markets Using High Frequency Data: Exploring Conditions of Classification Algorithm

Sungjae Yoon[1], ○Aiko Suge[1], Hiroshi Takahashi[1] （1. Keio University）

10:00 AM - 10:20 AM

[2K1-05] A Consideration of Estimating Spoiler Pages in Comics

○Yoshiki Maki[1], Yuji Shiratori[1], Kenta Sato[1], Satoshi Nakamura[1] （1. Meiji University）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ユーザ辞書を用いたニューラル翻訳のカスタマイゼーション

The customyzation of neural machine translation with user's dictionary

山内 翔大 *1　村上 陽平 *2　中口 孝雄 *3　石田 亨 *4
Shota Yamauchi　Yohei Murakami　Takao Nakaguchi　Toru Ishida

*1京都大学工学部情報学科
Undergraduate School of Informatics and Mathematical Science, Kyoto University

*2京都大学デザイン学ユニット
Unit of Design, Kyoto University

*3京都情報大学院大学ウェブビジネス技術専攻
Department of Web Business Technology, Kyoto College of Graduate Studies for Informatics

*4京都大学社会情報学専攻
Department of Social Informatics, Kyoto University

Neural machine translation (NMT) has significantly improved quality over traditional statistical-based machine translation (SMT). However, it is known that it is difficult for NMT to translate sentences containing rare terms. Therefore, in this research, we propose a method of replacing rare terms with synonyms and translating it, and replacing the translated synonyms with translations of the rare terms in the bilingual dictionary. This approach has two technical issues: acquisition of synonyms from small scale corpus, and selection of the synonyms. The proposed method shows the better result than the result of the existing technics in sign test.

1. はじめに

ニューラルベース機械翻訳 (NMT) は，従来の統計ベース機械翻訳 (SMT) よりも大幅に品質を向上させている．例えば，Google が開発した GNMT は，エラーの少なさや正確性などにおいて，PBMT よりも良い結果を示している [Johnson 17]．また，TOEIC の点数で換算すると 900 点相当の英文作成能力があるニューラル翻訳機 *1 も登場している．SMT では，入力文の各単語をマッピングテーブルを用いて翻訳し，言語モデルを用いて適切に並び替えることによって翻訳結果を生成し出力する．一方，NMT では，原文はエンコーダと呼ばれるニューラルネットワークによってベクトル表現に変換され，その後，デコーダと呼ばれる別のニューラルネットワークが翻訳文を生成する．したがって，SMT が単語を最小単位としたマッピングテーブルを用いて翻訳するのに対し，NMT では，入力文そのものが最小単位になり，入力文に唯一に対応する翻訳結果を出力する．

しかしながら，NMT には，普段使われることが少ないような専門用語を含む文を翻訳することが困難であるということが知られている．また，従来の SMT では，マッピングテーブルに専門用語を追加することで，辞書を用いた翻訳に対応することができた．しかし，NMT にはマッピングテーブルがないため，辞書を用いて専門語を含む文章を翻訳することが難しい．したがって，ユーザが NMT をカスタマイズすることも困難であると言える．

そこで，本研究では，専門用語を類義語で置き換えて翻訳し，翻訳結果の中に含まれる類義語の訳語を対訳辞書から得られる専門用語の訳語でおきかえるという手法を提案する．この手法により，ユーザは任意の NMT に対して外付けでユーザ辞書を連携させることができ，専門用語を含む文の翻訳が可能になる．このアプローチを実現する上で解決するべき技術課題として，以下の二つが挙げられる．

1. 小規模コーパスからの類義語の獲得

   専門用語は大規模コーパスの中にさえほとんど出現しないため，モデルを学習するにはデータが足りないということがある．その場合，適切な類義語を獲得することが非常に困難である．したがって，小規模コーパスから適切な類義語を獲得する手法が必要である．

2. 翻訳機に適用する類義語の選択

   類義語の中のどの単語を使用すれば良い翻訳結果を出力できるかは，使用する翻訳機に依存する．したがって，翻訳機に適用する類義語の選択手法が必要である．

*1 https://www.miraitranslate.com/index.php#txt-t

2. 未知語の翻訳

NMT には，専門用語を含む文を翻訳することが困難であるという欠点がある．計算の複雑さを制御するために，NMT は小さな語彙を使用しなければならず，語彙外の大量の専門用語はすべて単一の記号 (以下，unk 記号) に置き換えられるからである．このアプローチにより，NMT モデルが専門用語の翻訳を学ぶことができないという問題が生じている．したがって，入力文中のある単語や，その単語の翻訳結果がその頻出単語群に含まれていない場合，入力文の単語と出力文の単語の対応関係が取れず，モデルはその単語の適切な翻訳を生成することができない．また，最悪の場合には，一部の語やフレーズが翻訳結果から欠落してしまう可能性がある [Li 16]．この問題は，語彙外の単語という意味の out-of-vocabulary word の頭文字をとって OOV 問題として知られている．この問題に対して，これまで様々なアプローチから研究がなされてきた．これらのアプローチは，大きく分けて 2 種類に分類することができる．一つは，ニューラル翻訳モデル学習時にカスタマイゼーションを施す手法であり，もう一つは，翻訳時にカスタマイゼーションを施す手法である．

2.1 学習時のカスタマイゼーション

[Li 16] では，学習段階においては，まずモノリンガルコーパスから単語の分散表現を学習する．得られたモデルは単語間

の類似度を評価するために用いられる．それと同時に，バイリンガルコーパスを用いて両言語の文章対の中の単語レベルでの対応を学習する．さらに，副産物として，整列されたバイリンガルコーパスから語彙の翻訳テーブルが導出される．このテーブルと類似度モデルを用いて，バイリンガルコーパス内の専門用語の単語対は，類似語に置き換えられる．最終的に，得られた新たなバイリンガルコーパスを用いて，NMT モデルを作成する．

[Jean 15] では，専門用語を含む文章を NMT でうまく翻訳できない問題に対して，NMT モデル作成時に学習させる語彙の量を増やすというアプローチを根幹においた研究を行なっている．そもそも，専門用語が学習できない理由は，語彙サイズに比例して大きくなる計算量を制御するために学習させる語彙数を制限し，それ以外の単語は単一の unk 記号に置き換えて学習させているからである．この研究では，その問題を解決するため，語彙全体をクラスタリングすることで，非常に大きな目標語彙を有するモデルに対してデコードを効率的に行うというアプローチをとっている．

[Luong 15] では，学習時に unk 記号に置き換えられた専門用語がどんな語だったかを追跡することにより，専門用語を含む文章を NMT でうまく翻訳できない問題を解決しようとしてる．具体的には，まず，両言語の対となる文中に含まれる単語の対応関係を情報として含んだデータを用いて訓練し，翻訳後の文中の各 unk 記号に対応する入力文中の語の位置を出力することができるような NMT モデルを作成する．このモデルを用いて翻訳することにより，翻訳後の文に含まれる unk 記号に対応する専門用語の正体を突き止めることができ，最後に対訳辞書等を使って置き換えを行うことで最終的な翻訳結果を出力することができる．

2.2 翻訳時のカスタマイゼーション

[Li 16] は，学習時だけでなく，翻訳時にもカスタマイゼーションが施される．翻訳段階においては，まず，入力文中の専門用語を学習段階で得られた類似度モデルを用いて類義語に置き換える．そして，作成した NMT モデルを用いて翻訳し，翻訳結果の中の類義語の訳語を語彙の翻訳テーブルを用いて専門用語の対訳に置き換えることで，最終的な翻訳結果を獲得する．この手法では，NMT の学習時に使用した類似度モデルや語彙の翻訳テーブルを翻訳の際にも使用しているため，高い翻訳精度を示していると思われる．そして，提案手法は外付けでユーザ辞書を NMT に連携させるアプローチであるため，翻訳時のカスタマイゼーションに分類される．提案手法と [Li 16] との大きな差異は，[Li 16] が自ら作成した NMT モデルを用いて翻訳を行うのに対し，提案手法はユーザが翻訳に使用する任意の NMT に対して適用できるという点である．

3. 専門用語の類義語の獲得

本研究では，word2vec [Lilleberg 15] を用いて獲得された特徴ベクトル間のコサイン類似度が大きいものを類義語と定義する．しかしながら，大規模コーパス中にすら登場回数が少ないような専門用語は，学習を十分に行うことができず，適切な単語ベクトルを生成できないため，類似していない単語が類義語として獲得されてしまう．この問題を解決するために，本研究では word2vec を用いたコーパスの自動生成を提案する．この章では，コーパスの自動生成を用いた専門用語の適切な類義語を獲得する手法について説明する．この手法は，1 回目の類義語学習プロセス，コーパスの自動生成プロセス，2 回目の類義語学習プロセスの 3 つのプロセスから構成される．

3.1 1 回目の類義語学習プロセス

まずはじめに，大規模コーパスと，専門用語が使用されている小規模コーパスを統合し，1 回目の学習用コーパスを作成する．次に，学習用コーパスに対し word2vec による学習を行うことにより，学習用コーパス内の用語の単語ベクトルを学習したモデルを獲得する．最後に，この学習モデルから学習用コーパス内の用語の単語ベクトルを獲得する．ここで得られた各用語の単語ベクトルの類似度を求めることにより，各用語の類義語を獲得することができる．

3.2 コーパスの自動生成プロセス

図 1 は，コーパス自動生成の手順を表したものである．1 回目の学習で獲得された word2vec モデルを用いて，専門用語が含まれた文の中の専門用語以外の名詞を類義語に置き換えることで多様な文を獲得する．

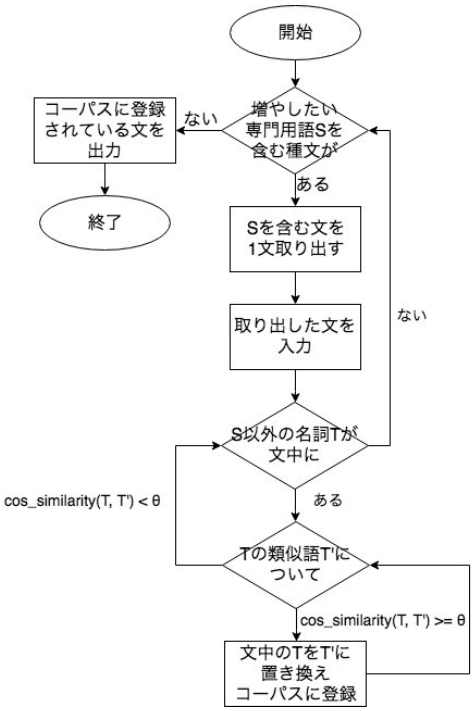


図 1: word2vec を用いたコーパスの自動生成のフローチャート

コーパス自動生成の手順は以下の通りである．まず，増やしたい専門用語 S を含む文（種文とする）を一つ選択する．種文中に S 以外の名詞 T があれば，名詞 T の類義語の中で，T とのコサイン類似度が 1 番高い T’ を選ぶ．名詞 T がなければ，別の種文を選択し直す．T と T’ のコサイン類似度が，閾値 θ 以上ならば種文の中の T を T’ に置き換える．そうでないならば新たな T を選択し直す．T を T’ に置き換えた文を獲得する．その後，T’ を再び選択し同様の処理を繰り返す.（この際，T’ は前回選んだものの次に類似度が高いものを選ぶ．）

3.3 2 回目の類義語学習プロセス

まず，1 回目の学習用コーパスと，自動生成したコーパスを統合し，2 回目の学習用コーパスを作成する．次に，2 回目の学習用コーパスに対し word2vec による学習を行うことにより，コーパス内の用語の単語ベクトルを学習したモデルを獲得する．最後に，この学習モデルから専門用語の単語ベクトルを獲得する．ここで得られた各専門用語の単語ベクトルと，その他の用語との類似度を求めることにより，各専門用語の類義語を獲得することができる．

4. 適切な類義語の選択

この章では，適切な類義語の選択手法として，類似度積を用いた，最終候補となる類義語の選定と，折り返し翻訳及び BLEU 値計算 [Papineni 02] を用いた，最終候補からの類義語の選択を提案する．

4.1 類似度積を用いた類義語選定プロセス

まず，類似度積について説明する．類似度積とは，翻訳対象文中に登場する専門用語の単語ベクトルと，各専門用語に対応する類義語の単語ベクトルとのコサイン類似度の積のことである．まず，提案手法の翻訳を行うための必要十分条件は，類義語で専門用語を置き換えた文の翻訳結果の中に類義語単体の翻訳結果が含まれていることである．したがって，翻訳対象文中に含まれる各専門用語の類義語のうち，この条件を満たし，なおかつ類似度積が高いような類義語の組を最終候補として選定する．

4.2 折り返し翻訳及び BLEU 値計算を用いた類義語の選択プロセス

図 2 は，類義語選択の手順を表したものである．

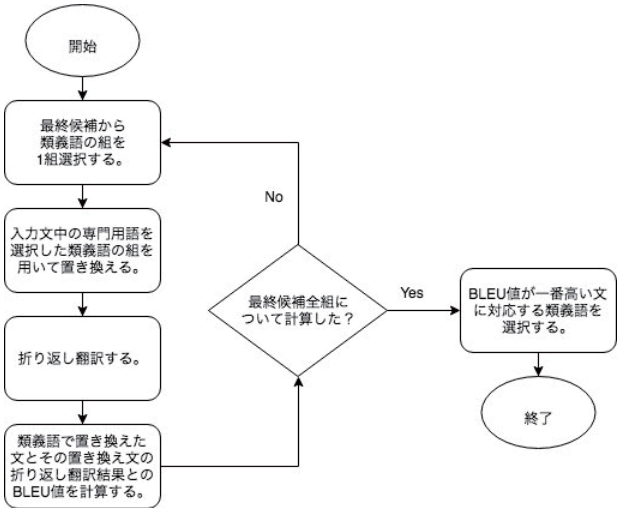


図 2: 折り返し BLEU 値を用いた類義語の選択プロセス

まず，類似度積を用いた類義語選定プロセスによって得られた各類義語の n-組で，入力文中の専門用語を置き換え，置き換え文を作成する．次に，置き換え文を折り返し翻訳する．最後に，置き換え文と，折り返し翻訳結果との BLEU 値を計算し，一番高い BLEU 値を示した際に用いた類義語を選択する．このプロセスにより，最終的に翻訳に使用する類義語が獲得される．

5. 提案手法の評価

ここでは，100 文を対象にして，提案手法の翻訳結果と，専門用語の置き換え語として類義語ではなくランダムな文字列を使用する既存手法 [Murakami 11] を用いた場合の翻訳結果を評価，比較する．今回の実験には，大規模コーパスとして wikipedia コーパス [*1] を，word2vec ライブラリとして gensim[*2] を使用した．翻訳対象の 100 文及び辞書は，京都の観光分野のものを作成した．

∗1 https://dumps.wikimedia.org/jawiki/latest/
∗2 https://radimrehurek.com/gensim/

5.1 平均点による比較

評価方法は，翻訳結果の 5 段階絶対評価であり，評価者は 3 人である．表 1 は，3 人の評価者 A, B, C が二つの手法で実施した 5 段階評価の 100 文の平均点である．

表 1: 3 人の 100 文に対する 5 段階絶対評価の平均値

	A	B	C
提案手法	3.98	3.82	4.10
既存手法	3.86	3.63	4.09

実験の結果，3 人とも提案手法の評価が既存手法の評価を若干上回ったが，t 検定によって統計的に有意な差が得られなかった．

5.2 相対評価による多数決を用いた符号検定結果の比較

次に，100 文に対して，3 人の相対評価による多数決を用いた片側符号検定を行う．この相対評価では，5 段階評価で同点だったものについても強いていえばどちらが良い結果を示したかを答えてもらい，両者に優劣をつけてもらった．どうしても優劣がつけれないものに関しては同点としてもらった．帰無仮説として，「翻訳結果に差がない」という仮説を設定する．また，有意水準 α は 5%とした．

本実験においては，100 文の，2 種類の翻訳手法における 3 人の相対評価の多数決は，表 2 のように分布した．

表 2: 3 人の 100 文に対する相対評価の多数決

提案手法	既存手法	同点
43 文	25 文	32 文

結果，$p = 0.019 < 0.05$ となり，帰無仮説を棄却され，提案手法の統計的優位性が示された．

5.3 翻訳結果の考察

次に，提案手法が，定性的にどのような文に対して効果的な翻訳を行えたのかを考察していく．表 3 の上 3 行は，「仁和寺は平安時代に建てられました.」という文章の二つの手法での翻訳結果である．なお，この文章では専門用語は「仁和寺」「平安時代」であり，提案手法において類義語として選ばれたのはそれぞれ「円城寺」「中世」である．

表 3: 翻訳結果例

翻訳対象	仁和寺は平安時代に建てられました.
提案手法	Ninnaji Temple was built in Heian Period.
既存手法	Ninnaji Temple was built on Heian Period.
翻訳対象	私の父は白雲寺の住職を知っています.
提案手法	My father knows the priesthood of Hakuunji Temple.
既存手法	My father knows the job title of Hakuunji Temple.

二つの翻訳結果の大きな違いは，提案手法では「in Heian Period」と訳されており，辞書連携翻訳では，「on Muromachi period」と訳されている．これは，前者において，「平安時代」が「中世」という時代を司る語に置き換わった結果，適切な前置詞が選択されたと思われる．一方，後者では，「平安時代」が無意味な文字列に置き換えられた結果，その文字列の前に適切な前置詞を配置できなかったものだと思われる．このように，

適切な前置詞を求められる翻訳に対しては，提案手法は有効に働くことが考察される．

次に，表 3 の下 3 行は，「私の父は白雲寺の住職を知っています.」という文章の二つの手法による翻訳結果である．なお，この文章では専門用語は「白雲寺」であり，提案手法において類義語として選ばれたのはそれぞれ「高千穂峰」である．

この翻訳結果においては，提案手法では適切に翻訳されているのに対し，既存手法では無意味な文字列の影響が文全体に伝播し，「住職」の翻訳が不適切なものになってしまったのだと思われる．このように，提案手法においては，既存手法を用いる際に発生した誤訳の伝播を抑える効果が期待される．

6. おわりに

本研究では，専門用語を含んだ文の新たな翻訳手法として，入力文中の専門用語を類義語で置き換え，翻訳した後に翻訳結果に含まれる類義語の翻訳結果を専門用語の対訳と置き換えるという手法を考案した．本研究における技術課題は，以下の二つであった．

1. 専門用語の適切な類義語群の獲得

   出現頻度の低い専門用語の類義語を獲得することは困難であった．したがって，少ないデータから適切な類義語群を獲得する手法が必要であった．解決策として，本研究では word2vec を用いたコーパスの自動生成を考案した．

2. 類義語群からの適切な類義語の選択

   類義語群の中から，正しい翻訳結果を与える類義語を選択する手法が必要であった．解決策として，本研究では，類似度積を用いた類義語選定と，折り返し翻訳及び BLEU 値計算を用いた類義語選択を考案した．

また，提案手法を京都の観光分野のテキストに用いた結果，翻訳精度の向上が示され，既存手法に対する提案手法の優位性が示された．

謝辞

本研究は，日本学術振興会科学研究費基盤研究 (A) (17H00759, 平成 29 年度～32 年度) および日本学術振興会科学研究費若手研究 (A)(17H04706, 平成 29 年度～32 年度) の支援を受けた．

参考文献

[Jean 15] Jean, S., Cho, K., Memisevic, R., and Bengio, Y.: On Using Very Large Target Vocabulary for Neural Machine Translation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pp. 1–10 (2015)

[Johnson 17] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F. B., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J.: Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation, *TACL*, Vol. 5, pp. 339–351 (2017)

[Li 16] Li, X., Zhang, J., and Zong, C.: Towards Zero Unknown Word in Neural Machine Translation, in Kambhampati, S. ed., *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pp. 2852–2858, IJCAI/AAAI Press (2016)

[Lilleberg 15] Lilleberg, J., Zhu, Y., and Zhang, Y.: Support vector machines and Word2vec for text classification with semantic features, in Ge, N., Lu, J., Wang, Y., Howard, N., Chen, P., Tao, X., Zhang, B., and Zadeh, L. A. eds., *14th IEEE International Conference on Cognitive Informatics & Cognitive Computing, ICCI*CC 2015, Beijing, China, July 6-8, 2015*, pp. 136–140, IEEE Computer Society (2015)

[Luong 15] Luong, T., Sutskever, I., Le, Q. V., Vinyals, O., and Zaremba, W.: Addressing the Rare Word Problem in Neural Machine Translation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pp. 11–19 (2015)

[Murakami 11] Murakami, Y., Lin, D., Tanaka, M., Nakaguchi, T., and Ishida, T.: Service Grid Architecture, in Ishida, T. ed., *The Language Grid - Service-Oriented Collective Intelligence for Language Resource Interoperability*, Cognitive Technologies, pp. 19–34, Springer (2011)

[Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, in *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318Association for Computational Linguistics (2002)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

句構造解析とクラスタリングを用いた会話履歴の要約

Dialog Summarization using Phrase Structure Analysis and Text Clustering

星野　綾子　　細見　格
Ayako Hoshino　　Itaru Hosomi

NEC セキュリティ研究所
NEC Security Research Laboratories

Contact centers have massive amount of dialog records to be used to improve the quality of their service. This paper describes a method of summarizing a large number of dialogs on the same topic into a tree structure. The method consists of two steps: 1) summarization of each dialog with phrase structure rules, and 2) organizing dialogs into a tree structure using text clustering. A common oddness was observed among summarized dialogs, and this problem was mitigated by applying Multi-Sequence Alignment (MSA). With the proposed method, we were able to summarize real-life dialogs into a reasonably small tree with only two hours of rule writing labor. Also, applying MSA helped to reduce the number of nodes and led to a higher purity score.

1. 概要

コンタクトセンタには，日々大量に蓄積される会話履歴を活用し，オペレータの応対品質を上げるという課題がある．そのためには，多数の会話履歴を会話の流れが把握できる形で，一覧できるよう要約することが求められている．本稿では，大量の同一トピックに関する会話履歴を木構造のグラフに要約する手法を提案する．

手法は２段階から成り，１）各会話を構造解析の後，２）会話集合をクラスタリング，を行う．各会話の構造解析においては，文脈自由文法を用いて会話の構造を解析し，ラベルを付与する．その後に，同一層のノード群に対し，ラベル及び出現単語に基づきクラスタリングを行う．

提案手法により，会話の流れを保持した形で大量の会話履歴を要約することが出来る．また，本稿では，この手法により起こる会話要約上の課題を提示し，多重整列（Multi-Sequence Alignment）により問題を軽減する．

提案手法により，実データを２時間程度のルール作成の手間により，充分小さいサイズの木に要約することができた．また，多重整列により，全体ノード数の削減とクラスタリング結果の評価指標である純度の向上が見られた．

2. はじめに

コンタクトセンタ [*1] では，同一トピックに関する会話が多数行われ，記録されている．この会話履歴データを生かし，オペレータの応対品質を上げることは，コンタクトセンタ業者にとって重要な課題である．しかし，単純な文書要約では，セールストークや質問の出現有無や順番等，会話の流れを把握することができず，オペレータにとって具体的な改善策が得られない．大量の会話履歴を，会話の流れが分かるように集約する手法が必要である．

例えば，会話履歴中の連続する発話を，「本人確認」「注文受付」等，業務の単位にラベル付けし，複数会話の流れを木構造で把握できるようにすることは，会話履歴集合の要約として有用である．

連絡先: 星野　綾子，NEC セキュリティ研究所，〒 211-8666
神奈川県川崎市中原区下沼部 1753，TEL 044-455-8652，
a-hoshino@cj.jp.nec.com

*1　電話のみを扱うコールセンタに対し，メールやチャットによる顧客接点をコンタクトセンタと呼ぶ．

3. 先行研究

会話履歴を文書と見なすと，大量の会話履歴の要約には，複数文書分割技術の利用が考えられる．複数文書分割技術とは，文書内の要素同士の対応関係を考慮に入れ，複数の類似内容の文書を分割する技術である．Jeoug と Titov は，複数文書トピック分割技術を提案し，教師なしベイジアンモデルと単純な split-merge アルゴリズムを用いて，文書を全文書に共通のセグメントと各文書に独自のセグメントとに分割した [Jeong 10]．

また，コンタクトセンタを対象分野とした会話履歴を木構造に構造化する技術として，Bangalore と Stent の研究ある [Bangalore 09]．Bangalore と Stent の手法による木構造は，Dialog を根とし，Task，Topic/Subtask, Dialog Act, Pred-Args, Utterance, Clause を順に各層に配置している．

また，Luis と Cohen はオンラインフォーラム上の会話履歴から木構造を生成した [Louis 15]．彼らは PCFG と LCFRS の２つの異なる文法規則による構造解析を行い，スレッドをサブトピックに分割し，特に LCFRS により隣接しない同一サブトピックに関する投稿をリンクさせ，クラスタリングを含む３種の手法により，会話履歴を幅方向に時系列順にサブトピックが並ぶ木構造に整理した．

Bangalore と Stent，Luis と Cohen のいずれの手法も会話の流れを深さ方向に保持した木構造を作ることは出来ない．つまり，木構造を根から葉まで辿ることにより，ある会話を開始から終了まで再現する木構造を生成できない．そこで，本稿では，トピックのラベルを与えることができ，尚且つ，根から葉を辿ることにより会話を辿ることができる木構造を生成する手法を提案する．

4. 手法

提案手法のステップ１ (句構造解析による会話要約; PRS)，ステップ２ (テキストクラスタリングによる会話履歴要約; CLS) を説明する．また，提案手法により生じた課題に対し，多重配列 (MSA) を用いた緩和策を説明する．

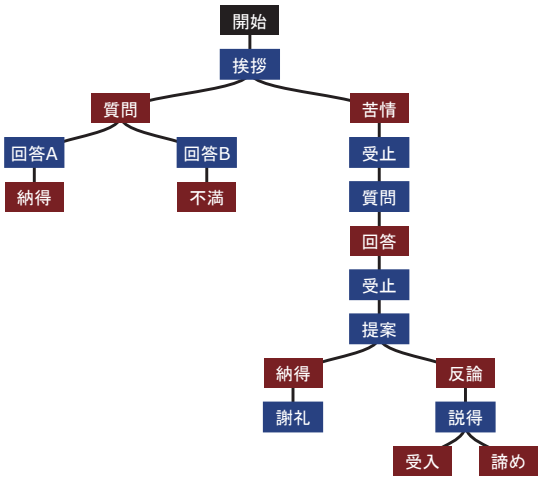


図 1: 会話構造化結果の例

4.1 句構造解析による会話要約

先行研究が示すように会話履歴データを纏め上げるのに句構造解析を用いることが適している．本稿では，解釈可能な要約結果を得るために，人手作成の句構造規則を利用し会話を句構造解析する方式を採用した．まず，文に対し,「質問」「回答」等の，発話行為に相当するラベルを付与する（非終端記号→終端記号規則）．実装には，正規表現パタンを用いた．次に，ラベル列をさらに要約するような規則を作成する（非終端記号→非終端記号規則）．これにより，複数文にまたがる事象を 1 ノードに要約する．実装には，独自のパタンマッチングプログラムを用いた．

規則作成の際には，コンタクトセンタにおける業務の目的（セールス，質問回答，等）を考慮し，目的に関連する事象をなるべく漏らさず捉えるようにする．以下に，作成した非終端記号→非終端記号規則の例を挙げる．O はオペレータの発話，C は顧客の発話を示す．

質問回答 → C:質問, O:回答
注文要求 → C:商品・数量指定
注文要求 → C:商品指定, C:数量指定

また，いずれの規則にもマッチしない文には「その他」ラベルを付与した．その他ラベルは，スペースの要求に応じて，連続するその他ラベルを 1 ノードに纏めたり，直前のその他以外のノードにマージすることもできる．

4.2 テキストクラスタリングによる会話履歴要約

ステップ 1 の結果，会話開始を根とし，深さ方向に会話の流れを保持する木を得た．さらに幅方向に複数ノードの纏め上げを行うことで，図 1 のようなラベル付の木を得る．

大量の会話履歴を要約するため，テキストクラスタリングを用い，同一親を持つ同一層における類似の内容のノードを統合する処理を，根から開始し再帰的に行った．尚，ノードは，そのノードに属する全ての文に含まれる一定頻度以上の形態素を保持している．アルゴリズムは，1) k-means または 2) 階層クラスタリング SLINK を採用し，実装には ELKI[*2] ツールキットを用いた．特徴には，形態素原形とラベルの BOW (Bag-Of-Words) を用い，統合されるノード群中の最頻のラベルを統合後のノードのラベルとした．ベースラインとして，同一親ノードを持つ同一ラベルのノードを纏める手法 (Simple) と比較した．

*2 https://elki-project.github.io/

表 1: 実データ処理時の木サイズ

	深さ	ノード数
before PRS	285.4	35,966
after PRS	20.9	2,407
after PRS-CLS	20.9	362

4.3 提案手法における課題と多重整列を用いた緩和

提案手法では，同一層のノード群を統合処理の対象としたため,「挨拶-注文要求-注文要求」等，同一のラベルが連続する木が多く見られた．これは，層 1 で 1 クラスタをなしたノード群のラベルとして「挨拶」が選ばれ，層 2，層 3 において 1 クラスタをなしたノード群のラベルとして「注文要求」が選ばれたためである．近接する層で同一のラベルが発生することが多いため，不自然に連続するラベル列が生じやすい．この課題を解決するために，ステップ 1 と 2 の間に，近接する層の同一ラベルノードが同一層となるようノード群を整列する多重整列の処理を加えた．多重配列を行うツールは，Clustal Omega[*3] を使用した．会話履歴群から得たラベル列をアミノ酸配列を表す記号に変換し，多重整列させた．整列の結果，挿入された空白に該当する個所に，空ノードを挿入した．

5. 評価

コンタクトセンタにおける実データについて，文法規則の作成時間，要約結果の木のサイズ，プログラム実行時間を評価した．また，多重整列を用いた要約の向上については，コーパスデータ [*4] により，木サイズと純度を計測した．

5.1 文法規則の作成時間

コンタクトセンタにおける実データ 126 件から，4 件程度のランダムサンプルを抽出し，2 時間程度で，非終端記号→終端記号規則 16 件，非終端記号→非終端記号規則 16 件，合計 32 件の規則を作成することができた．

5.2 要約結果の木サイズ，プログラム実行時間

ステップ 1，2 前後の木サイズを表 1 中に示す．要約前に比して，要約後はノード数がステップ 1 により 6.7%に削減され，ステップ 2 によりさらに 15%に削減された．実データで計測したプログラムの実行速度は，ステップ 1 は 2 分で処理した．また，ステップ 2 は 307 回クラスタリングを繰り返し，処理時間は 30 秒であった．要約結果は，末端が 45 本の木となり，紙やディスプレイ上で一覧可能な大きさとなった．

5.3 多重整列を用いた要約の向上

コーパスデータ処理時の木サイズ（ノード数）を表 2 中に示す．データセットに合わせて作成した，計 22 件の規則を用いた．木サイズを計測する際には，終端の空ノード及び空ノードのみを子孫として持つ空ノードは除外した．純度は，あるノードに統合されたノード群における，ノードに選出されたラベルと一致するラベルを持つノードの割合を木全体に渡って算出したものとする．

ステップ 2 に Simple を採用した場合, 多重整列によりクラスタリング後の木サイズは 117 になり，これは多重整列を行わない場合に比して 1.27%となった．尚，Simple を用いた場

*3 https://www.ebi.ac.uk/Tools/msa/clustalo/
*4 NTT 雑談対話コーパス [東中 14]

表 2: ノード数と純度

	Simple		Clustering	
	ノード数	純度	ノード数	純度
PRS	17,738	-	17,738	-
PRS-CLS	9,248	1.00	314	0.57
PRS-MSA-CLS	117	1.00	55	0.97

合には，純度は必ず 1.0 となる．また，多重整列を行うことにより木サイズが 17.5%に削減され，且つ，純度が 0.97 に向上した．

6. まとめ

本稿では，会話履歴を木構造に要約する手法として，句構造規則とテキストクラスタリングを用いた要約手法を提案した．実データに対し，句構造規則を作成することができ，コンタクトセンタの実データから，充分小さい木を生成できることを確認した．また，提案手法による課題の解決策として，多重整列を用いた改良を行い，ノード数の減少と純度の向上を達成した．

参考文献

[Bangalore 09] Bangalore, S. and Stent, A.: Incremental Parsing Models for Dialog Task Structure, in *EACL* (2009)

[Jeong 10] Jeong, M. and Titov, I.: Multi-document Topic Segmentation, in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pp. 1119–1128, New York, NY, USA (2010), ACM

[Louis 15] Louis, A. and Cohen, S. B.: Conversation Trees: A Grammar Model for Topic Structure in Forums., in Marquez, L., Callison-Burch, C., Su, J., Pighin, D., and Marton, Y. eds., *EMNLP*, pp. 1543–1553, The Association for Computational Linguistics (2015)

[東中 14] 東中 竜一郎, 孝太郎 船越 : Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション, *SIG-SLUD*, Vol. B4, No. 02, pp. 45–50 (2014)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

観光案内におけるニューラル言語モデルを用いた説明文の生成

Language generation system for sightseeing guidance using neural language model

生田　和也 *1　品川 政太朗 *1　吉野 幸一郎 *1　鈴木 優 *1*2　中村 哲 *1*2
Kazuya Ikuta　Seitaro Shinagawa　Koichiro Yoshino　Yu Suzuki　Satoshi Nakamura

*1奈良先端科学技術大学院大学 情報科学研究科　*2データ駆動型サイエンス創造センター
NAIST Graduate school of information science　NAIST Data science center

In sightseeing information navigation systems, the information presented by natural language has a potential to improve usability. Several systems tried to embed the informing contents in a prepared template for generating sentences which are useful for tourists, which is called a slot filling based method. However, it is difficult for the systems to generate diverse expressions and unseen patterns. To solve this problem, we propose a neural network based sentence generation method instead of using a slot filling based method. In this research, we construct the contents as a one-hot vector representation and construct the neural network based language generator and the one-hot content vectors for generating natural and understandable sentences. We collected a tourist information corpus via crowdsourcing. Existing language generation systems used word classes. However, these systems often connect words unnaturally. In this research, we also proposed a re-ranking system based on a neural language model to solve the problem. In our experiments, we confirmed the naturalness and validity of the sightseeing guidance sentences generated by our proposed method.

1. はじめに

旅行者が観光プランを決定する上で必要以上の情報が溢れ，逆に意思決定を困難とする問題を「情報過負荷状態」[Barta 09] と呼ぶ．この問題に対し，観光案内システムの開発では位置情報や天候・混雑情報といったコンテキストを用いて提供する情報を絞り込むのが一般的である [Abowd 99]．この際，どのような情報形式で提案を行うのかが重要となる．[Venkataiah 08] では，Web サービスなどで提供される様々なモーダルによる断片的な情報を「Discrete Visualization」と定義し，それらを単一モーダルとして集約した「Continuous Visualisation」と比較することで，単一モーダルによる情報提示のほうが理解を得やすいという結果が得られたと論じている．単一モーダルによる情報提示のひとつとして，言語によるコンテンツの集約が考えられる．例えば観光者がルート案内システム [Tumas 09] により移動する時，臨時休業や運行ダイヤの変更によるルート変更はしばしば起こり得る．この際，ルートのみによる情報提供において，何故変更がなされるのかについての補足を行うのは難しい．これに対し，言語による情報提示では理由も含めて提示を行う事ができ，ユーザの疑問を解消することができる．また情報提供を行うシチュエーションは多様に存在するのに対し，モバイルインタフェースでは表示可能な画面の大きさやアクセシビリティが限られている。こうした状況に対し、言語による情報提示では，Push 通知や音声ガイダンスなどの情報提供が可能である．本研究では，言語による情報集約を用いた観光案内システムを想定し，伝達したいコンテンツが断片的に定まっている際の文生成を行うことで，その有効性についての検証を行う．

2. 文生成手法

文生成（Natural Language Genaration : NLG）は，非言語情報や中間表現を基いて人間が理解可能な文を生成することであり，文書要約 [Conroy 01] や機械翻訳 [Cho 14], 詩の生成 [Zhang 14]，株価動向の記述 [Murakami 17][Aoki 16] など幅広いタスクにおいて用いられる．

2.1 ルールによる文生成

NLG において出力は文であるが入力としては様々なものをとりうる。これをルールによって生成するシステムの例として，地理情報システム (GIS) の構造化データを入力としてメッセージを生成してルート案内を行うシステム [Dale 03] や，一定時間ごとに観測した時系列データを入力とし，海上予報文を生成する研究 [Sripada 03] などがある．これらでは文を生成する際の言語ルールを人手により設計しているため高コストである．

2.2 テンプレートマッチングによる文生成

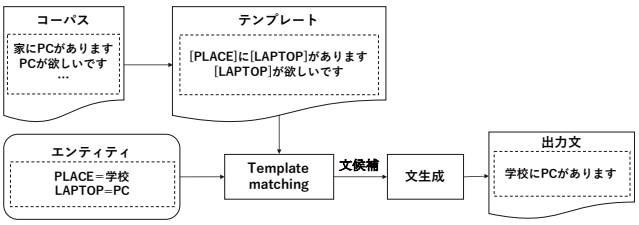


図 1: テンプレートマッチングによる文生成

言語ルールを自動構築する研究として [Kondadadi 13][山崎 16] では，NLG をテンプレート抽出およびテンプレートマッチングの問題として扱うことで自然な文を生成することに成功している (**図 1**)．テンプレートとは固有名称をクラス表現タグとして置き換えた文の骨組みである．この手法では言語ルールの設計コストを少量に抑えて最適化を行うことができる反面，表現できる文の幅がテンプレートに依存するといった問題があり，クラス表現の組み合わせによっては候補となるテンプレートを見つけ出すことが難しい．

2.3 言語モデルによる文生成

文生成の異なるアプローチとして，言語モデルを用いるものがある．言語モデルとは文 $s = x_{1:n}$ において「単語 x_t の生起確率はそれまで出現した単語 $x_{1:t-1}$ に従って決まる」という仮定をおいて文をモデル化したものであり，予測される単語の生起確率に従ってサンプリングを行う事で生成が行わ

表 1: コンテンツと埋め込み単語の対応

PLACE	SHOP	TREE	EVENT	TREE-STATE	SALE
四条	松栄堂	桜	講演会	見頃	限定販売
河原町	舞扇堂	紅葉	祭り	散る直前	割引セール
…	…	…	…	…	…
BEAUTIFUL	GOOD-VALUE	CONGRSTION	BEGIN	STAY	END
True/False	True/False	True/False	True/False	True/False	True/False

れる．モデル構築は再帰結合を有するニューラルネットワーク（RNN）が用いられる事が多く，これを**ニューラル言語モデル**と呼ぶ．RNN において内部にメモリセルを持つものを Long-Short Term Memory（LSTM）と呼び，学習において保持する情報，忘却する情報，出力する情報を最適化する事で学習が行われる．これにより，時刻 t において生成される単語が k である確率は単語予測分布 y を用いて下記で表される．

$$y_{t=k} \simeq P(x_t = k|x_{1:t-1}) \tag{1}$$

特に，コンテキストを用いて生成文を条件付けるモデルを Encoder-Decoder モデルと呼び，機械翻訳の研究 [Cho 14] などで用いられる．[Cho 14] では翻訳元言語をコンテキストとして，翻訳先言語と対応付ける形で LSTM 言語モデルを構築することで翻訳文を生成する．コンテキストを d とすると，Encoder-Decoder モデルにおける予測分布 y は下記で表される．

$$y_{t=k} \simeq P(x_t = k|x_{1:t-1}, d) \tag{2}$$

このアーキテクチャをそのまま本研究に適用しようとした場合，コンテキストと説明文の対応関係が膨大となる．そのため、コンテキストをある程度汎化して表現することができる枠組みが必要となる。

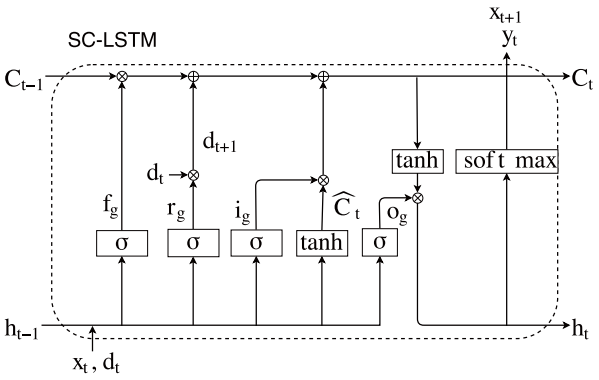


図 2: SC-LSTM

対話システムの研究において Wen らが提案した Semantically conditioned LSTM (SC-LSTM）[Wen 15] は，LSTM に新たな読み込みゲートとよばれる機構を加えて応答文の生成を行っている (図 2)．発話するクラス表現を「1」，発話しないクラス表現を「0」としたバイナリベクトルをコンテキスト d として与え，読み込みゲートから得られるパラメータを減衰係数としてコンテキストを収束させることで同時最適化を行い言語モデルを構築する. これによりコンテキストの状態に応じた柔軟な生成が可能となり，予測分布 y は下記になる．

$$y_{t=k} \simeq P(x_t = k|x_{1:t-1}, d_{1:t-1}) \tag{3}$$

本研究では観光案内システムのコンテンツは伝達内容を 0/1 のバイナリで表現したコンテンツベクトルとして生成することができると仮定し，その出力をコンテキストとして用いる事でテンプレートに依存しない観光案内文生成を行う．生成にあたり，Wen らのアーキテクチャをもとに実装を行う．

3. 観光案内コーパスの構築

文生成を行うにあたり，まず観光案内を行う範囲を定め，対応したコンテンツの設計を行う．次に，設計したコンテンツにしたがってコーパスの収集を行う．

3.1 コンテンツの設計

Wen らが提案する生成システムのデザインと観光案内で必要とされる案内内容から、**表 1** に示す 12 種類のコンテンツベクトルをデザインした。この際 Wen らの提案システムでは、コンテンツベクトルは状態の時系列的な変化を考慮していないが、観光案内ではこうした状態にも対応を行う必要がある。そこで本研究では，観測状態の開始や終了，継続を表現可能な「BEGIN」「STAY」「END」といったコンテンツの定義を新たに加えた．今回のコーパス収集では，観光案内を行う地域を京都とした．

設計したコンテンツは,「Place =四条」といった埋め込み単語として得られるもの,「Beautiful = True」といった抽象的な表現として表現されるものがある．この表のコンテンツを組み合わせてコンテンツパターンの設計を行う．そのためにまず，あらかじめ文として表現可能なコンテンツパターン 69 通りを設計し，コーパスの収集を行った．

3.2 コーパスの収集

学習に用いる文の収集にあたってはクラウドソーシング [*1] を用いた．収集作業を行うにあたり，まずコンテンツパターンからランダムに 1 つパターンを選択する．パターンに含まれるコンテンツは対応したランダムな埋め込み単語あるいは True/False フラグに置き換えられ，作業を行うワーカに指示される．ワーカには，その指示を満たす形で観光案内を行う説明文を **1 文で**記入してもらった (**表 2**)．これにより構築されたコーパスを用いてモデルの構築を行い，観光案内文を生成する．

表 2: ワーカによる記入例

コンテンツパターン	もみじ, 見頃,BEGIN=True
ワーカによる記入	もみじは見頃を迎え始めています
コンテンツパターン	松栄堂, タイムセール,GOODVALUE=True
ワーカによる記入記入	松栄堂のタイムセールはお得です

4. 観光案内文の生成

観光案内文を生成するために，SC-LSTM を用いて文の候補を列挙する生成フェーズを行う．その後，文候補から適当な文を選択するリランキングフェーズを経て最終的な観光案内文を出力する. 生成文の有効性を検証するにあたり，関連研究で述べたテンプレートマッチングによる手法をベースラインとした．以下で両手法の生成プロセスについて述べる．

*1 クラウドワークス：https://crowdworks.jp/

4.1 生成フェーズ

生成フェーズでは，コンテンツパターンをベクトル表現としたコンテンツベクトルを入力として与え，SC-LSTM ジェネレータを用いて文候補の生成を行う．

テンプレートマッチングでは，入力となるコンテンツパターンと一致するテンプレートを探索し，文候補とする．一致するテンプレートが見つからない場合，レーベンシュタイン距離を用いて最も近いパターンを持つテンプレートを用いる．

4.2 リランキングフェーズ

リランキングを行う際，埋め込み単語によって言い回しが変化する可能性がある．例えば「講演会」というイベントを文に埋め込む場合,「開演された」という言い回しは自然であるが,「祭り」というイベントを埋め込む場合,「開演された」といった表現は不自然になる．リランキングを行った後に埋め込み単語を当てはめる場合，この言い回しの変化を考慮することができない．よって本論文における観光案内文生成では，SC-LSTM が生成した文のクラス表現 **(Entity)** に埋め込み単語 **(Phrase)** を当てはめた言語モデルによるリランキングを行う事で，埋め込み単語を考慮した文生成を行った (**Phrase-reranker**)．また，言い回しを踏まえたリランキングを行うことによる自然性・妥当性の変化について比較するため，クラス表現を埋め込み単語に置き換えずにリランキングを行う場合においても同様に生成を行った (**Entity-reranker**). リランキングに当たっては，生成フェーズで得られた SC-LSTM エントロピに加えて，LSTM エントロピ，生成文に含まれるべきコンテンツの不足数・剰余数を総コンテンツ数で割った $SERR$ を用いて最適化を行った．

テンプレートマッチングでは，Phrase-reranker の LSTM 言語モデルによるコストに従ってリランキングを行う．

5. 評価実験

5.1 実験設定

モデルの学習には，クラウドソーシングにより収集したコーパス 9645 文とそれに付随するコンテンツパターン 69 種を用いる．69 種のうち無作為に抽出した 10 種のコンテンツパターンとそれに対応する観光案内文 1405 文を未知の案内文として除外し，残った 8240 文をトレーニング：バリデーション：テスト＝ 7:1.5:1.5 となるようにコーパスを分割し，トレーニングコーパスを用いて学習を行った．

生成文の評価は, テンプレートを用いた言語生成の先行研究 [山崎 16] に習い，BLEU スコア及び主観評価により行う．BLEU は n-gram 一致率により参照文とどれだけ近しいかを測るものであり，人間が作成した参照文との機械的な比較を行うことで、言語生成が適切に行われているかについてのいくつかの指標と相関を持つことが指摘されている。今回は、参照文はワーカにより記入された案内文を用いる．主観評価では,「自然性」と「妥当性」について「1〜5」の 5 段階評価を行うとともに，各生成手法における「−2〜2」の 5 段階の嗜好テストを行った．

実験にあたって,「テストコーパスにおける説明文を参照文としたもの」と「予め除外した案内文を参照文としたもの」の 2 種類について文生成を行い，上述の評価を行った．以降ではこれらをそれぞれ「既知のコンテンツパターンによる文生成」「未知のコンテンツパターンによる文生成」と呼ぶ．これにより、学習データに存在しないパターンに対する頑健性を考察する．

5.2 実験結果-BLEU スコア

BLEU スコアにおける評価結果を**表 3** に示す．

評価結果について，以下のような傾向が見られた．

表 3: BLEU スコア

既知のコンテンツパターンによる文生成	
手法	BLEU スコア
Template matchng	<u>47.70</u>
Entity-reranker	46.55
Phrase-reranker	46.72
未知のコンテンツパターンによる文生成	
Template matching	37.99
Entity-reranker	<u>41.09</u>
Phrase-reranker	41.07

(*i*) 未知のコンテンツパターンによる BLEU スコアは，既知のコンテンツパターンに比べて全手法が低下する結果となった．Template matching では BLEU スコア大きく低下したのに対し，他の手法のスコア減少は小さかった．

(*ii*) 既知のコンテンツパターンにおいては Template matching による手法が最も高スコアであるが，他の手法と比較してスコアの差は少なかった．

(*iii*) 未知のコンテンツパターンについては，Template matching の BLEU スコアが最も低くなった．Entity-reranker が最も高いスコアであるが，Phrase-reranker のスコアとほぼ同程度であった．

(*i*)〜(*iii*) より，未知のコンテンツパターンが与えられた場合，提案手法はベースラインに比べて柔軟な文生成が行われていると考えられる．

5.3 実験結果-主観評価

主観評価における結果を**表 4**，**表 5** に示す．

表 4: 主観評価-自然性・妥当性

既知のコンテンツパターンによる文生成		
手法	自然性	妥当性
Template matchng	<u>4.01</u>	4.20
Entity-reranker	4.00	4.22
Phrase-reranker	3.96	<u>4.24</u>
未知のコンテンツパターンによる文生成		
Template matchng	3.98	3.10
Entity-reranker	<u>4.00</u>	<u>4.10</u>
Phrase-reranker	3.75	3.94

表 4 は，生成文を「自然性」及び「妥当性」について「1〜5」で 5 段階評価されたものである．**表 5** は，3 つの生成手法のうちそれぞれ 2 つについて，どちらが好ましいかを「-2〜2」で 5 段階で評価したものである．主観評価結果から，提案手法による文生成の自然性・妥当性は 3.75〜4.24 と高い結果となり，概ね適切な文が生成できていると言える．

既知のコンテンツパターンによる文生成においては，各手法の自然性・妥当性及び嗜好性全てにおいて BLEU スコアによる評価と同一の結果が得られた．

未知のコンテンツパターンによる文生成では，自然性・妥当性及び嗜好性において Phrase-reranker による手法が Entity-reranker による手法に比べて劣る結果となった．Phrase-reranker で用いられる LSTM 言語モデルは埋め込み単語が含まれるため学習難易度が相対的に高く，リランキングが適切に行われない場合があった事が要因だと考えられる．実際にいくつかの出力文では，同じ内容を複数回繰り返すといった不適切な文がリランキングされている例が確認された．

表 5: 主観評価-嗜好

既知のコンテンツパターンによる文生成			
Pref	Template matchng	Entity reranker	Phrase reranker
Template matchng	-	0.026	0.041
Entity reranker	-0.026	-	0.013
Phrase reranker	-0.041	-0.013	-

未知のコンテンツパターンによる文生成			
Pref	Template matchng	Entity reranker	Phrase reranker
Template matchng	-	-0.561	-0.332
Entity reranker	0.561	-	0.232
Phrase reranker	0.332	-0.232	-

表 6: 未知のコンテンツパターンによる生成結果

パターン (A)	EVENT=祭り, CONGESTION, START
Template matchng	平等院では祭りが行われ、混雑しはじめています。
Entity reranker	祭りで混雑が始まっています。
Phrase reranker	祭りは混雑し始めています。

パターン (B)	PLACE=上賀茂神社, EVENT=特別展
Template matchng	上賀茂神社では特別展が行われていて混雑しています。
Entity reranker	上賀茂神社ではお特別展が行われています。
Phrase reranker	上賀茂神社では特別展が開催されています。

5.4 生成例

未知のコンテンツパターンによる生成例を**表 6** に示す．

Template matching による手法では, 指示したコンテンツパターンを満たさない文生成が行われている．例えば (A) では，指示していない「平等院」という埋め込み単語が現れている．(B) では，指示していない「CONGESTION」(混雑) という表現が含まれている．Entity-reranker，Phrase-reranker の両手法では，多くのコンテンツパターンにおいて未知のパターンに対応した文生成が行われており，コーパスに存在しない観光案内文の表現が可能となっている．また Phrase-reranker による手法では，いくつかの文において埋め込み単語による言い回しの変化に対応した生成が行われている事を確認した．例えば**表 6 パターン (B)** では，Entity-reranker で「お祭り」という表現に対して「特別展」を誤って埋め込んだ「お特別展」といった出力がされている．Phrase-reranker ではこのような生成は見受けられなかった．

5.5 結言

本論文では，観光案内システムにおいて文による情報提示を行うシステムを提案し，伝えたいコンテンツが得られている時の文生成について検証を行った．従来手法では，未知のクラス表現の組み合わせを考慮した文生成が難しいという課題があったが，本研究では SC-LSTM アーキテクチャをベースに，LSTM 言語モデルを用いてリランキングを行うことで，未知のコンテンツの組み合わせを考慮した文生成において提案手法の有効性を確認した．

今後の課題として，以下のようなものが考えられる．1 つ目に，コンテンツベクトルの設計をより具体化する必要がある．実際に観光案内システムを開発するにあたって，より広域のドメインにおいて観光案内を行うことが想定され，そのため観測するデータ及びそこから推定できるコンテンツについてより具体化を行う必要がある．2 つ目に，コンテンツの前後関係や重複を考慮した文生成について検討する必要がある．コンテンツの種類によっては生成する順番により意味合いが変化する場合があるため，位置関係や記述回数を考慮した文生成が可能となる枠組みが重要であると考えている．

謝辞

本研究成果は、国立研究開発法人情報通信研究機構の委託研究により得られたものです。

参考文献

[Abowd 99] Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M., and Steggles, P.: Towards a better understanding of context and context-awareness, in *Handheld and ubiquitous computing*, pp. 304–307Springer (1999)

[Aoki 16] Aoki, K. and Kobayashi, I.: Linguistic summarization using a weighted N-gram language model based on the similarity of time-series data, in *Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on*, pp. 595–601IEEE (2016)

[Barta 09] Barta, R., Feilmayr, C., Pröll, B., Grün, C., and Werthner, H.: Covering the semantic space of tourism: an approach based on modularized ontologies, in *Proceedings of the 1st Workshop on Context, Information and Ontologies*, p. 1ACM (2009)

[Cho 14] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078* (2014)

[Conroy 01] Conroy, J. M. and O'leary, D. P.: Text summarization via hidden markov models, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 406–407ACM (2001)

[Dale 03] Dale, R., Geldof, S., and Prost, J.-P.: CORAL: Using Natural Language Generation for Navigational Assistance, in *Proceedings of the 26th Australasian Computer Science Conference - Volume 16*, ACSC '03, pp. 35–44, Darlinghurst, Australia, Australia (2003), Australian Computer Society, Inc.

[Kondadadi 13] Kondadadi, R., Howald, B., and Schilder, F.: A Statistical NLG Framework for Aggregated Planning and Realization., in *ACL (1)*, pp. 1406–1415 (2013)

[Murakami 17] Murakami, S., Watanabe, A., Miyazawa, A., Goshima, K., Yanase, T., Takamura, H., and Miyao, Y.: Learning to Generate Market Comments from Stock Prices, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1, pp. 1374–1384 (2017)

[Sripada 03] Sripada, S., Reiter, E., and Davy, I.: SumTime-Mousam: Configurable marine weather forecast generator, *Expert Update*, Vol. 6, No. 3, pp. 4–10 (2003)

[Tumas 09] Tumas, G. and Ricci, F.: Personalized mobile city transport advisory system, *Information and Communication Technologies in Tourism 2009*, pp. 173–183 (2009)

[Venkataiah 08] Venkataiah, S., Sharda, N., and Ponnada, M.: *A Comparative Study of Continuous and Discrete Visualisation of Tourism Information*, pp. 12–23, Springer Vienna, Vienna (2008)

[Wen 15] Wen, T.-H., Gasic, M., Mrksic, N., Su, P.-H., Vandyke, D., and Young, S.: Semantically conditioned lstm-based natural language generation for spoken dialogue systems, *arXiv preprint arXiv:1508.01745* (2015)

[Zhang 14] Zhang, X. and Lapata, M.: Chinese Poetry Generation with Recurrent Neural Networks., in *EMNLP*, pp. 670–680 (2014)

[山崎 16] 山崎健史, 吉野幸一郎, 前田浩邦, 笹田鉄郎, 橋本敦史, 舩冨卓哉, 山肩洋子, 森信介 他：フローグラフからの手順書の生成, 情報処理学会論文誌, Vol. 57, No. 3, pp. 849–862 (2016)

高頻度データを用いたニュースと韓国株式の分析:ニュース分類の条件探索

Analysis of News and Stock Returns in Korean Markets Using High Frequency Data: Exploring Conditions of Classification Algorithm

尹聖在[*1] Sungjae Yoon　菅愛子[*1] Aiko Suge　高橋大志[*1] Hiroshi Takahashi

[*1] 慶應義塾大学大学院経営管理研究科
Graduate School of Business Administration, Keio University

News articles play an important role in financial markets. This study analyzes the relationship between news articles and stock price fluctuations using high frequency trading data in Korean stock markets. Especially, we analyze differences in market reactions according to languages of news articles. In order to understand the influences of news articles, this study explores conditions of Long Short Term Memory (LSTM) models that classify news articles.

1. はじめに

フィンテックなど金融分野における技術の進展を背景とし，資産運用への関心が高まっている．資産運用分野において，新たな情報技術を用いた分析などが盛んに報告される中，金融市場において発信されるニュースは，資産価格に関する重要な情報の一つに挙げられる．効率的な市場において，情報は資産価格に速やかに反映されるという理論背景において，日々発信されるニュースと株式市場の関連性について，これまで数多くの実証研究がなされてきた．近年では，ニュース記事の内容を定量化・指標化する試みがなされ，それらと投資家行動に及ぼす要素とを関連付ける研究もされている．

我々は以前，様々な言語のニュースが株価に与える影響を韓国株式市場において分析した（図 1）[Yoon, 2017] [Yoon, 2018]．更にその中で，Long Short Term Memory (LSTM)ブロック [Hochreiter & Schmidhuber, 1995] [Hochreiter & Schmidhuber, 1997] を用いた深層学習モデルにて，英語ニュースによる韓国株式のリターン分類を報告した．そのモデル作成時のハイパーパラメータ探索の試みを本稿で報告する．

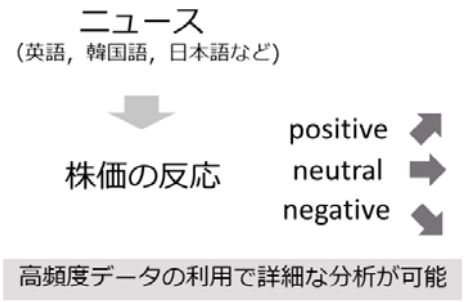


図 1:ニュースガ株価に与える影響

2. 目的

機械学習モデルの学習において，良いモデルを作るためにハイパーパラメータ（外部から与えるモデルのパラメータ）のチューニングが必要となる．良いモデルの一例として，精度の高いモデルが挙げられる．目的の精度をもとめてモデル学習を繰返す際，選択するアルゴリズム，データ量，計算環境などによって，一回の試行に時間を要することも多い．そこで効率よく探索を行うため，様々な試みがなされており，ランダムサーチ，グリッドサーチ，ベイズ最適化，などが知られている．

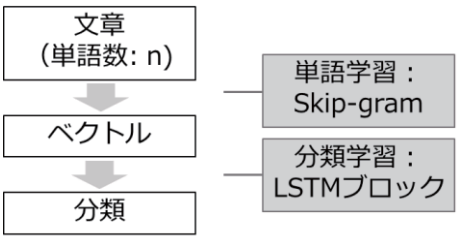


図 2:モデルの概要

深層学習では，多数のパラメータが存在する．例えば筆者ら用いたモデルは図 2 のように要約されるが，そこで用いるハイパーパラメータとして例えば，学習係数，隠れ層のユニット数，層の段数，epoch 数，dropout，batch size などが挙げられる．

本稿ではパラメータ探索の始めとして，どの段階におけるパラメータの影響が大きいのか，ベクトル（入力段階），分類学習の段階，で大きく分けパラメータの変更を行い，分類結果（精度）の変化を観察した．

3. データ

3.1 ニュースデータ

ニュースデータは韓国株式市場に上場している時価総額上位 5 社（2012 年末時点）について Thomson Reuters から抽出した．2013 年ニュースをモデル作成用（学習用）に用い，2014 年ニュースを予測に用いた．ニュース発信言語は英語の他，韓国語，日本語，中国語など多岐に渡るが，本稿では英語ニュースを扱う．

3.2 株価データ

Thomson Reuters の高頻度データから，上記で抽出したニュース発信前後の株価を 15 分ごとに取得した．ニュース発信 30 分前に比較し，ニュース発信 30 分後に株式が 0.3%以上値上がりしたものをポジティブ，0.3%以上値下がりしたものをネガティブ，それら以外をニュートラルなニュースであると分類した．結果，表 1 のようなデータセットが得られた．

連絡先:菅愛子，慶應義塾大学大学院経営管理研究科，aikosuge@keio.jp

表 1:ニュースデータの分布

	ネガティブ	ニュートラル	ポジティブ	計
2013 年	15.9%	63.5%	20.6%	908
2014 年	19.3%	62.8%	17.9%	643

4. 方法

基本モデルを設定し, その中のハイパーパラメータを変化させ, 精度の変化を検証した.

4.1 基本モデルの設定

図 2 に示すようにニュース分類をセットした. すなわち, 文章(単語)のベクトル化には skip-gram 法 [Mikolov, 13] を, 株価文章分類の学習には LSTM ブロックを用いた. 2013 年ニュースデータ(908 サンプル)を訓練データとテストデータに分け(9:1), モデル学習には訓練データを用いた.

4.2 パラメータの変更

入力段階においては, 文章をベクトル化する際に用いている skip-gram のパラメータ, すなわち学習に用いる単語数やベクトルの次元, 学習効率(initial learning rate: initial LR, epoch 数)を変化させた. また, ニュース分類段階では LSTM ブロックの学習効率(initial LR, epoch 数)を変化させた.

4.3 結果の検証

訓練データとテストデータの, 精度の比較を行った. また, いくつかのモデルにおいて, 2014 年ニュースデータを用い, 株価リターン分析を行い, 分類予測を行った. なお, 結果は 20 回の学習を行い, その精度を平均して示している.

5. 結果

表 2 にパラメータ探索の結果を一部示す. なお, 結果は 20 回の学習を行い, その精度を平均して示している.

表 2:条件による精度の変動

	入力側	LSTM ブロック		検証	
条件	単語数: n	Initial LR	Epoch 数	訓練精度	検証精度
1	100	0.01	100	63.6%	63.3%
2	100	0.01	200	63.4%	64.4%
3	100	0.01	300	63.4%	64.4%
4	100	0.1	200	70.7%	63.3%
5	200	0.01	200	63.3%	63.6%

今回の結果では, 入力する単語数 n が精度に与える影響は小さいことが分かった. しかしながら, 今回用いたテキストデータの最長は 900 単語を超えるものもあり, 今後, 入力条件について詳細に分析していきたい. また, 分類学習における initial LR を 0.1 以上にすると, 訓練データが示すモデル精度と検証データが示す精度で乖離が見られ, 過学習が起きていることが分かった. 精度が 60%程度にとどまっている原因として, サンプル数の少なさや, データの不均衡分布が考えられる. 今後はその点を調整して学習モデルの改善に取り組む予定である.

最後に表 2 の条件 2 について, 2014 年のニュースデータを用い分類を行い, 実際の株価リターン(平均)を算出した. その分類結果を表 3 に示す. この時の予測精度は 55.7%となった.

表 3:学習モデルによる 2014 年ニュース分類(条件 2)

予測	ネガティブ	ニュートラル	ポジティブ	計
分布	17.0%	73.7%	9.3%	643

今回はデータが限られており, 別の期間などのデータを用いて, モデルや精度の検証をする必要がある.

6. まとめ・課題

ニュースによる株価分類を深層学習にてモデル化した際のハイパーパラメータについて探索した. 数多くのパラメータが存在するが, 今回は入力単語数と, モデル学習における initial RL と epoch 数について変動させ, モデル精度に与える影響を観察した. 過学習が起きない一定の条件を見出したものの, 全体の精度は 60%程度にとどまった. 今後, 更なる精度向上のためにモデルやパラメータ設定が課題である.

また深層学習以外の分類法との比較も望まれる.

参考文献

[Hochreiter & Schmidhuber, 95] Sepp Hochreiter and J¨urgen Schmidhuber: Long Short Term Memory, Technical Report FKI-207-95, Technische Universit¨at M¨unchen, M¨unchen, August 1995.

[Hochreiter & Schmidhuber, 97] Sepp Hochreiter and J¨urgen Schmidhuber: Long Short Term Memory, Neural Computation, 9(8):1735–1780, November 1997.

[Mikolov, 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, ICLR, 2013.

[Yoon, 17] Yoon, S.J., Suge, A., and Takahashi, H.: Analyzing the Influence of News Articles on the Korean Stock Market with High Frequency Trading Data, JSAI International Symposia on AI, Workshop 3: Artificial Intelligence of and for Business (AI-Biz 2017), 2017.

[Yoon, 18] Yoon, S.J., Suge, A., and Takahashi, H.: The Analysis on Korean Stock Market by High Frequency Data and Text Classification Using LSTM, JSAI Special Interest Group on Business Informatics (SIG-BI #8), 2018.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

コミックのネタバレページ推定に関する検討

A Consideration of Estimating Spoiler Pages in Comics

牧　良樹[*1]
Yoshiki Maki

白鳥　裕士[*1]
Yuji Shiratori

佐藤　剣太[*1]
Kenta Sato

中村　聡史[*1]
Satoshi Nakamura

[*1] 明治大学大学院先端数理科学研究科先端メディアサイエンス専攻
Frontier Media Science Program, Graduate School of Advanced Mathematical Science, Meiji University

Previous studies suggested that spoilers might increase the enjoyment of novels. However, these studies have not clarified the problem of spoilers sufficiently. Therefore, we focus on the spoilers of comics and consider the influence of comic spoilers depending on the reading progress and immediately after spoiler. In this work, we selected eight comics and investigated the spoilers' influence by changing the spoiler timing for readers. As a result, we revealed that spoilers reduced the interest in continue of comics. And we consider how to determine spoiler pages automatically using image processing and character detection.

1. はじめに

小説やコミック，ドラマなどのストーリーコンテンツは、昔ながらに楽しまれてきたコンテンツである．その中でもコミックは，描かれている物語に絵やセリフを通して読者が入り込むことにより，読者が興奮や感動，笑いなどを得ることが可能なものである．最近では，コミックは情報科学の分野でも注目されており，様々な研究が行われている[松下 13a][松下 13b].

コミックは人気コンテンツであり，情報分野でも注目されているが，コミックコンテンツで頻繁に扱われる問題としてネタバレが挙げられる．ネタバレとは，「主人公が勝つ」や「登場人物が死ぬ」といった，コンテンツを楽しみにしているひとがその部分を読んだり，観賞したりする前に触れてしまうことで，本来コンテンツから得られたはずの興奮や感動などを低下させる要因となってしまうものである．そのため，ネタバレはコンテンツがもっている面白さを 100%楽しみたい読者や視聴者には忌み嫌われるものであると考えられてきた．

ここで Leavitt ら[Leavitt 11]は，小説のネタバレに関する研究を行っており，ネタバレは問題ないということを明らかにしている．また，Levine ら[Levine 16]は，ネタバレタイミングの変化による影響の変化について研究しており，小説を読んでいる最中にネタバレをしても問題ないことを明らかにしている．しかし，対象コンテンツが小説であり，コミックにどこまで適用可能なものかは不明である．また，こうした研究で提示されているネタバレは日常的に受けるネタバレではないため，本質的なネタバレの影響を調査したものといえなかった．コミックに関するネタバレの研究では，Hassoun[Hassoun 13]がコミックのコマ割りによるネタバレの影響について調査をしている．しかし，Hassoun の研究ではコミックの 1 ページ内のネタバレに対して研究しているものであり，物語全体のネタバレの影響については研究されていなかった．

こうした点を踏まえ，我々は以前の研究[牧 17][Maki 17]において，コミック内での読書進度に応じたネタバレの影響に関する実験を行い，面白さと面白いと感じるシーン数が反比例することや，ネタバレが許容されるコミックと許容されないコミックが存在することを明らかにした．しかし，前回の研究では，読み終わりの面白さについての比較を行っているだけであり，ネタバレの本質的な解明を行うことができていなかった．また，そもそもどういったページをひとはネタバレと判断するのかについても分析が不十分であった．

そこで本研究では，まずコミックにおけるネタバレの問題について明らかにするため，ネタバレタイミング毎にネタバレを提示した直後の影響を調査する．また，ひとはどのページをネタバレであると判断するのかについて，過去の実験結果を踏まえて分析を行う．さらに，ネタバレページ推定を行うための実験手法およびデータセット構築手法についての検討を行う．

連絡先：牧良樹，明治大学総合数理学部先端メディアサイエンス学科，東京都中野区中野 4 丁目 21-1，03-5343-8334，tekkanomaki01@gmail.com

2. ネタバレの影響に関する分析

2.1 ネタバレ影響実験

我々はこれまでの研究[Maki 17]において，読書進度によってネタバレの影響度を測る実験を行った．ここでは，コミックを巻数によって 3 つのパートに分け実験を行った．また，ある作品について，ネタバレを提示するタイミング毎に 4 つのグループ（図 1）に分け，1 作品内のグループ毎にそれぞれ 5 人の協力者に実験に取り組んでもらった．なお，グループ 4 についてはネタバレを行っていない．

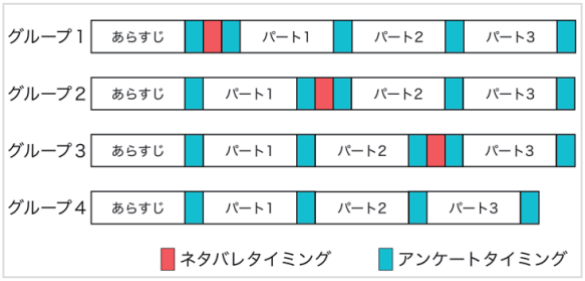


図 1　ネタバレとアンケートのタイミング

2.2 分析結果

パート 3 後の面白さ項目のアンケートをネタバレタイミング別に分けた結果を図 2 に示す. 図 2 より, パート 2 後で値が最小となったが, 有意差は見られなかった. そこで, ネタバレ直前および直後のアンケートを比較することによってネタバレされた直後の影響について傾向を分析する.

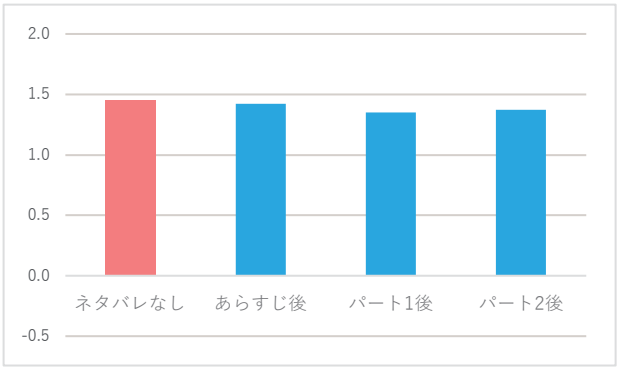


図 2 パート 3 後の面白さ比較

それぞれのアンケート項目について, 二要因混合計画(独立変数 1:ネタバレ要因(実験協力者内配置, 2 水準), 独立変数 2:ネタバレタイミング要因(実験協力者間配置, 3 水準), 従属変数:各アンケート項目における値)の分散分析を行った. また, 多重比較には Holm 法を用いた. 結果を図 3, 4 に示す.

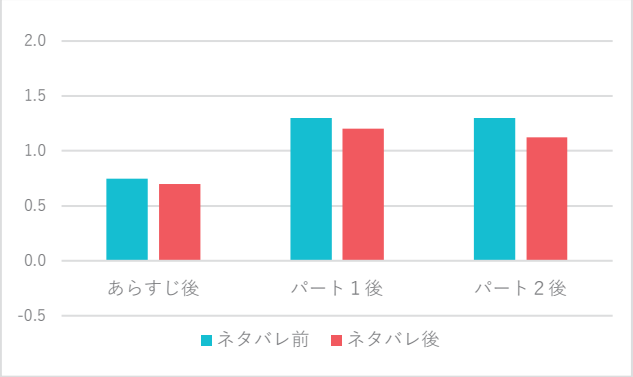


図 3 面白さ項目における分散分析結果

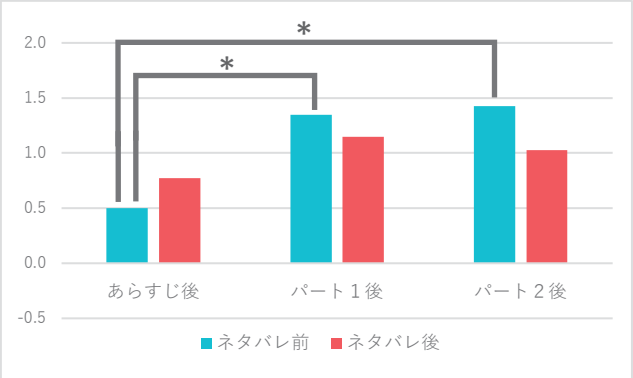


図 4 続きへの関心項目における分散分析結果

図 3 より, 面白さ項目でネタバレタイミング要因について主効果が観測された(F[2,117]=7.59, p<.01). 多重比較を行った結果, あらすじ後とパート 1 後間とあらすじ後とパート 2 後間に有意差が確認された(あらすじ後パート 1 間:p<.05, あらすじ後パート 2 間:p<.05). しかし, 交互作用が観測されなかった.

図 4 より, 続きへの関心項目でネタバレタイミング要因について主効果が観測された(F[2,117]=9.62, p<.01). 多重比較を行った結果, あらすじ後とパート 1 後間, あらすじ後とパート 2 後間に有意差が確認された(あらすじ後パート 1 間:p<.05, あらすじ後パート 2 間:p<.05). さらに, ネタバレタイミング要因とネタバレ要因の間に交互作用が観測され(F[2,117]=8.04, p<.001), あらすじ後内とパート 2 後内のネタバレ要因について有意差(あらすじ後:p<.05, パート 2 後:p<.01)が確認された.

また, 続きへの関心項目について, ネタバレなしには, ネタバレなしグループで一回でも減少した人数と割合を, ネタバレありにはネタバレをされたグループにおいてネタバレした直後に減少した人数と割合を表 1 に示す. この表より, ネタバレが提示されずとも 10%の実験協力者が続きへの関心を失うこと, またネタバレされると 37%の実験協力者が続きへの関心を失っていることが分かる. つまり, ネタバレなしに比べ, ネタバレありの方が, 4 倍程度の読者が続きへの関心を失っているといえる.

表 1 続きへの関心の減少割合

	実験協力者	減少人数	減少割合
ネタバレなし	40	4	10%
ネタバレあり	120	44	37%

2.3 考察

ネタバレ全体の面白さについては, どのタイミングにおいても有意な差が見られなかったことから, ネタバレはコミックの最終的な面白さには影響がないことがわかった. これは, ネタバレである結果を知っていたとしても他のシーンを楽しむことによって面白さが変わらなかったことや, ネタバレをされても最後までコミックを読みきることでネタバレされてない状態と同じくらいの面白さを感じられると考えることができる.

ネタバレ前後の結果から, 交互作用が確認され, ネタバレ前後に変化があったのは続きへの関心のみであった. 面白さの項目でネタバレ前後に差が確認されなかったことから, ネタバレをされても面白さは変わらないことがわかった. これは, 今回使用したネタバレページが見開き 1 ページのみであり, 内容が少なかったことから作品自体の面白さが損なわれることがなかったためであると考えられる.

また続きへの関心については, あらすじ後とパート 2 後に有意差が確認された. あらすじ後にネタバレをされると続きへの興味が増えたことから, 序盤の概要を知った状態でネタバレをされるとさらに読みたい気持ちが増え, 購買意欲が増すことが考えられる. しかし, パート 2 後にネタバレを受けると続きへの興味度合いが減ることより, ある程度読み進めた状態でネタバレを受けると続きへの興味を失ってしまうことがわかった. これによって, 気になっていたところ部分を知ってしまうことによって, 作品の続きを読むことを断念してしまい, 購買意欲が減ってしまう可能性があることがわかった. また, そもそも表 1 からネタバレをしなかったグループに比べてネタバレをしたグループの方が続きへの

減少割合が高いこともからも，ネタバレによって続きを読むのを断念してしまうことがわかる．

今回の実験では，表 1 で割合が提示されているとおり，続きへの関心が失われても，最後まで読むことを要求されている．途中で興味を失ったとしても，最後まで読めば面白いということは多々あるため，最終的な面白さは大差がないという結果につながっているといえる．以上のことからも，ネタバレにより購買意欲が減少してしまう可能性があるため，ネタバレは問題であるといえる．

3. ネタバレページ

コミックにおいてネタバレは問題があるという結果を踏まえ，今後はネタバレページを自動推定する技術を実現していく．ここで，ネタバレの自動推定においては，まずはネタバレページに関するデータセットをどのように構築していくかという点を検討する必要がある．

そこで，本章ではまずこれまで構築したネタバレデータセットについて分析を行い，どういった方針で今後構築するべきかといった点について議論を行う．

3.1 ネタバレページの特徴分析

我々はこれまでの研究[Maki 17]において，4 人の実験協力者に 4 ジャンル 2 作品ずつの，合計 8 作品のコミックに対して，各作品から 3 つの見開きページをネタバレとして選定してもらうことによって，ネタバレデータセットを構築してきた．

ページ毎のネタバレとして選ばれた回数を図 5 に示す．図の横軸は各作品の最終巻のページ番号であり（ネタバレは最終巻のみから選定された），縦軸はネタバレとして選ばれた回数となっている．

まず，サスペンスジャンル（予告犯，魍魎の匣）では，ネタバレとして選ばれたページが分散していることがわかる．これは，サスペンスジャンルでは複数の謎がそれぞれの場所で解決がされるため，実験協力者が各自の興味を持った謎が解かれている場所を選んだためだと考えられる．また，その他の作品においては，右肩あがりの作品（幽☆遊☆白書，All You Need Is Kill，GIANT KILLING）と右肩下がりの作品（ピンポン，ReRe ハロ，赤髪の白雪姫）の 2 種類の作品があった．ここで，我々は以前の研究[Maki 17]において，ネタバレ OK 作品とネタバレ NG 作品があることを明らかにしており，右肩あがりの作品はネタバレ NG 作品，右肩あがりの作品はネタバレ OK 作品と一致していることがわかっている．ここから，コミックの作品がネタバレ NG，OK に分けられる原因としてネタバレとなるシーンの位置や作品の盛り上げ方が関係する可能性があると考えられる．

しかし，今回の実験では，選択可能なページ数を 3 ページで固定してしまったため，選べる数が少なかった．また，図 5 からわかるように，隣接したページが選ばれていることも多いことから，1 つのネタバレが複数のページに渡って描かれている可能性がある．そこで，コミックにおけるネタバレをどのように教示するかについて再検討を行う必要がある．

ネタバレとして選定されていたページは，バトルジャンルでは，セリフや説明に「勝ったぜ」や「最期」など，勝敗がわかる言葉が出現していた．また，ネタバレページと周辺のページで他のページよりも大きなコマが使われていた．スポーツジャンルでは，表彰台のシーンや結果となるスコアなど文字情報として勝敗がわかるものが多かった．スポーツジャンルが，バトルジャンルと異なるのは，言葉ではなく，数字などで記載されてることが多かった点である．例えば，GIANT KILLING では，1，2 位ネタバレの周辺で文字情報が極端に少なくなっており，ピンポンでも 1 位ネタバレページで文字情報が極端に少なくなっていた．さらに，周辺で大きなコマが使われていたという特徴もあった．一方，サスペンスジャンルでは，文字情報として特定の文字が使われているということはなかったが，ネタバレページやその前後で大きなコマが使われているということが多かった．特に，魍魎の匣における 2〜4 位のネタバレページは，文字情報が他のページに比べて極端に少なくなっていた．予告犯でも 2〜4 位のネタバレページで文字数が少なくなっていた．ロマンスジャンルでは，ネタバレページでは「好き」や「すき」など，愛を伝える言葉が使われており，周辺のコマで大きなコマが連続的に使われていた．以上のことから，ネタバレページには，そのページ内の言葉とコマ

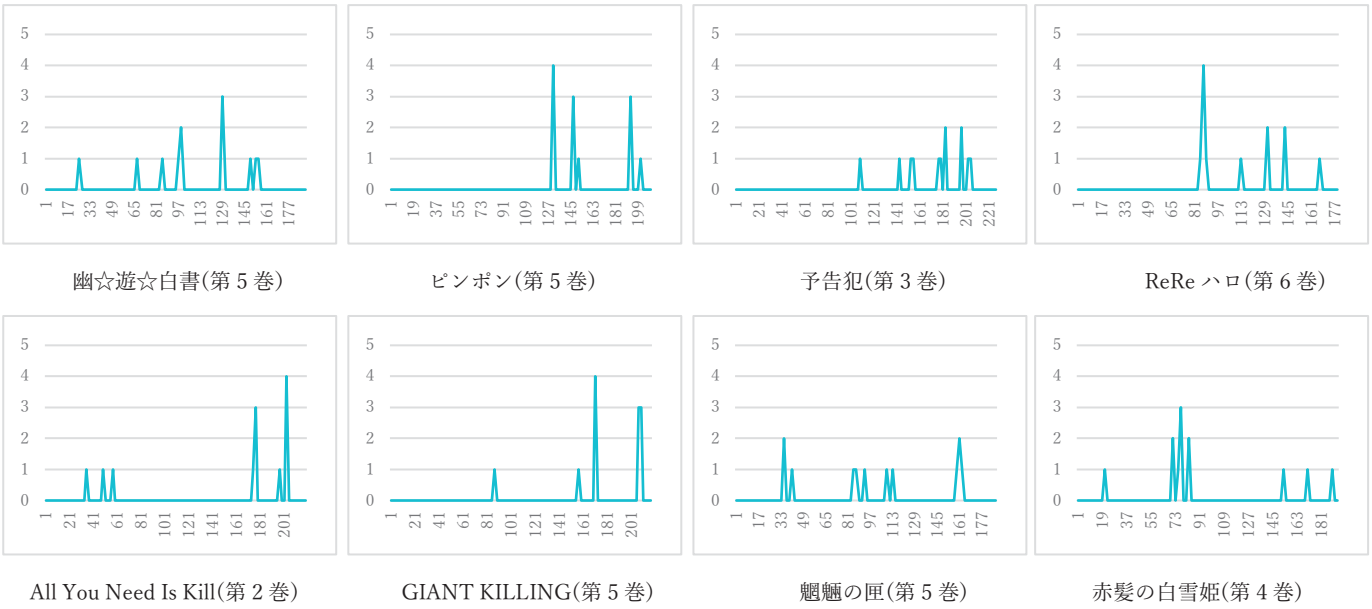


図5 各コミックにおけるネタバレページ帯とポイント

の大きさに特徴があり，さらに周辺のコマでも大きいものが使われるといった特徴があることがわかった．

3.2 ネタバレページデータセットの構築指針

実験結果より，ネタバレはどこまで読んだのかというコンテキストに依存すると考えられる．例えば，シリーズ物の 1 巻を読んでいる人に対して，第 20 巻のネタバレを提示しても，それをネタバレと認識することはできないと考えられる．そこで，ネタバレのデータセットを構築するうえでは，随時状況を設定し，そのシーンにおけるネタバレを選定してもらう必要がある．

現在構築しているデータセットでは，主観でネタバレであると感じるページを 3 ページのみ選定してもらい，さらに順位を付与してもらっていた．しかし，このネタバレであるという判断にぶれが大きくあり，結果的に適切なネタバレページを選定できていなかった可能性がある．また，ページに限りがあるために，限定されてしまった可能性がある．

そこでまず，ネタバレページかどうかの判定には，「そのコミックの N 話までを読んでいる他者に対し，N+1 話の中にあるページ p を提示したときに嫌な思いをされるもの」という基準を設定し，データセット構築者に対して教示を行う．これにより，評価指標が他者との関係性になり，明確化されると期待される．また，読み進めるに従い状況が変化するという点を考慮することが可能となる．

また，それぞれのページについてネタバレ度合いに応じて 3 段階のタグを付与してもらうことで，複数ページにわたるネタバレを選定していく．さらに，実験は作品を読んだことのない人に作品を読み進めてもらいつつ，ネタバレページであったと考えられるページに対してこのタグを付与してもらうことを検討している．これは，コミックを読んだことがあり事前知識がある人は，読んだことのない人と各ページに対する印象が異なってしまうことが考えられるためである．

3.3 ネタバレページの自動推定

現時点でネタバレページにはどのような特徴が存在するのかは明らかにされていない．特徴については様々なものが考えられ，その 1 つがコマの大きさである．コマは，一般的なシーンでは小さなコマ，盛り上がるシーンでは大きなコマが使われているが，盛り上がるシーンではバトルの結果や事件の真相など物語の核心をつく場面が多くなり，その分そのページがネタバレである可能性も高くなることが考えられる．つまり，コマの大きさを使ってネタバレページの推定を行うことが可能であると考えられる．コミックにおけるコマの推定手法は，石井ら[石井 07][石井 12]の濃度勾配の方向を生かして推定する手法や，野中ら[野中 12]のコマの形を利用して推定する手法が提案されており，これらの手法を用いることでコマの大きさを推定することを検討している．また，ネタバレページの特徴としては，ページ内のセリフや説明文の内容なども挙げられる．コミックでは，絵を用いて結果を表示することの他に，説明やセリフによって結果を見せることも多い．以前の実験でも，言葉で勝敗や結末を伝えているコマが選出されていた．そこで，それぞれのページで文字認識を行うことによってページ内の文字を抽出し，推定する手法についても行っていく．文字認識手法については，Google 社の Vision API[Google 18]を用いて行うことを考えている．さらに，コマやセリフ内容以外にも，顔検出を行うことや，吹き出し検出などを用いることも考えている．上記の手法を用いて検出された画像情報や文字情報の特徴と，コミック内のページそれぞれに付与されたネタバレのタグ情報を用いることにより，機械学習にて推定を行っていくことを考えている．

4. まとめ

本研究では，コミックにおける読書進行度によるネタバレタイミングの変化がネタバレの影響度にどう作用するのかを実験によって検証した．

その結果，全体の面白さの結果から，ネタバレを受けたとしても最後まで読み切ることによってネタバレされていない状態と変わらない面白さを得ることができることがわかった．続きへの関心の結果から，ネタバレを受けると続きへの興味度合いの減少割合がネタバレを受けなかった場合に比べて 4 倍程度高くなっていたことから，ネタバレによって購買意欲が減ってしまう可能性があることもわかった．

また，ネタバレページの特徴分析から，作品のジャンルによってネタバレページの特徴は異なっていた．しかし，ネタバレページには前後のコマの大きさやセリフに特徴があるという点では共通していた．この結果を元に，ネタバレページ推定を行うための実験手法について提案した．具体的には，まずデータセット構築において，作品未知の人に対し，対象コミックのページごとにネタバレ度合いに応じた 3 段階のタグを付与してもらうことを検討している．また，判定手法では，コマの大きさやセリフ内容などを検出し，機械学習を用いることによって判定する手法を考えている．

謝辞

本研究の一部は，JST ACCEL（グラント番号 JPMJAC1602）の支援を受けたものである．

参考文献

[松下 13a] 松下光範: コミック工学の可能性，第 2 回ウェブインテリジェンスとインタラクション研究会，2013

[松下 13b] 松下光範: コミック工学のこれまでとこれから，人工知能学会インタラクティブ情報アクセスと可視化マイニング研会，2013.

[Leavitt 11] Leavitt. J. D, Christenfeld. N. J. S：Story Spoilers Don’t Spoil Stories，Psychological Science，2011.

[Levine 16] Levine. William. H, Betzner. Michelle, Kevin S : The Effect of Spoilers on the Enjoyment of Short Stories，Journal of Discourse Processes, 2016.

[Hassoun 13] Hassoun. D : Sequential outliers. The role of spoilers? Need for cognition, need for affect, and narrative selection and enjoyment. Psychology of Popular Media Culture, 2016.

[牧 17] 牧良樹，中村聡史:コミック読者のネタバレ遭遇タイミングによる興味度合い変化，第31回人工知能学会全国大会(JSAI2017), 2017.

[Maki 17] Yoshiki Maki, Satoshi Nakamura: Do Manga Spoilers Spoil Manga?. The Sixth Asian Conference on Information Systems (ACIS 2017), 2017.

[石井 07] 石井大祐, 河村圭, 渡辺裕: コミックのコマ分割処理に関する一検討, 電子情報通信学会論文誌, 2007.

[石井 12] 石井大祐, 渡辺裕: マンガからの自動キャラクター位置検出に関する検討, 情報処理学会研究報告, 2012.

[野中 12] 野中俊一郎, 沢野拓也, 羽田典久: コミックスキャン画像からの自動コマ検出を可能とする画像処理技術「GT-Scan」の開発, FUJIFILM RESERCH & DEVELOPMENT, 2012.

[Google 18] Google Cloud Platform: Cloud Vision API, 2018.

[2K3-NFC-3a] [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

座長:西村 拓一（産業技術総合研究所）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room K (3F Azisai Mokuren)

[2K3-NFC-3a-01] Consideration of application cases of structured manual and its utilization

○Satoshi Nishimura[1], Ken Fukuda[1], Takuichi Nishimura[1] （1. National Institute of Advanced Industrial Science and Technology）

3:20 PM - 3:40 PM

[2K3-NFC-3a-02] Knowledge Sharing and Acquisition among Teachers using Guitar Rendition Ontology

○Nami Iino[2,3], Satoshi Nishimura[2], Takuichi Nishimura[2], Ken Fukuda[2], Hideaki Takeda[1,3] （1. National Institute of Informatics, 2. National Institute of Advanced Institute Science and Technology , 3. Graduate University for Advanced Studies）

3:40 PM - 4:00 PM

[2K3-NFC-3a-03] Development of "Tailor-made Brain Training" program suitable for environment

○Chiaki Oshiyama[1], Hiroyasu Miwa[2], Chihiro Sutou[3], Eiji Shimizu[3], Takuichi Nishimura[2] （1. Osaka University, 2. National Institute of Advanced Industrial Science and Technology, 3. Chiba University）

4:00 PM - 4:20 PM

[2K3-NFC-3a-04] Design of Activity Program for expression :project for expression on caring mind

○maiko kobayakawa[1], takeshi sunaga[2], sunao maruyama[2], tomoki hirano[1], kousuke yamada[4], takuichi nishimura[3], kentaro watanabe[3], sachiko fujimitsu[5] （1. Tokyo University of the arts, 2. Tokyo University of the Arts, 3. National Institute of Advanced Industrial Science and Technology, 4. Keio University , 5. Saga University hospital）

4:20 PM - 4:40 PM

[2K3-NFC-3a-05] Treatment strategy of Septicemia based on DPC data analysis by using pLSA method

○Kazuya Yamashita[1], Yuichiro Sakamoto[2], Eiichi Sakurai[1,2], Yoichi Motomura[1,2] （1. National Institute of Advanced Industrial Science and Technology Artificial Intelligence Research Center, 2. Saga University）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

構造化マニュアルの適用事例の分析と利用に向けた考察

Consideration of application cases of structured manual and its utilization

西村悟史[*1] Satoshi Nishimura　福田賢一郎[*1] Ken Fukuda　西村拓一[*1] Takuichi Nishimura

[*1] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

The authors have proposed the methodology, which is called knowledge explication, to construct qualitative knowledge base (KB). The authors focus on employee-driven approach to construct KB. Knowledge explication method is one of the approach. The constructed knowledge is call structured manual in this context. This manuscript provides clarification of the cases which the authors applied knowledge explication method or structured manuals. The authors also classify the contents of structured manuals which was constructed in each case. The last part of this manuscript shows developing systems for knowledge explication as future prospects.

1. 序論

本稿では，説明できる AI (Artificial Intelligence)システム[DARPA]の実現に向けて，西村らの提案した知識発現[西村 2017]がどのように貢献し得るかについて論じる．

本研究では，説明を「人間が分かる形で情報を表現したもの」と捉える．具体例としては，移乗介助のやり方を対話文で表現しロボットが発話した音や，楽譜の画像データとそれに重畳するような位置情報を持つように「スラ―する」というアノテーション文字列を表現しディスプレイ上に表示した光，などが挙げられる．ここでは，そのような説明を生成することを説明能力と呼ぶ．AI システムが説明能力を持つためには，説明したい内容を持つ必要がある．ここでの内容とは，先の例での「移乗介助のやり方」や「スラ―する」に相当するものである．筆者らは，このような内容を専門家から引き出すために知識発現と呼ぶ方法論について研究を進めてきた[西村 2017]．

本稿では，知識発現とその中で用いられる構造化マニュアルの事例を整理する．まず，2 章で現状の知識発現について述べる．続く 3 章にて内容を構造化し，説明を生成するまでに関連する作用について整理する．4 章ではこれまでの知識発現の応用事例を整理する．5 章にて，2 章で整理した作用の計算機による支援について考察し，今後の展望を述べる．

2. 知識発現とは

西村らは，プロセス知識共有のために，図 1 に示す知識発現と呼ぶ方法を提案し，介護現場での検証を行っている[西村 2017]．ここでのプロセスとは，介護業務のための従業員の行為や使用する道具の作用を総称した概念である．この方法は，プロセス知識の共通部分（以下，共通プロセス知識と呼ぶ）を基盤として，現場固有のプロセス知識（以下，固有プロセス知識と呼ぶ）を従業員が主体的に記述することに特徴がある．具体的には，2 人以上の従業員が集まり，共通プロセス知識を参照し話し合いながら，自分たちの持つ現場固有の知識を追加したり，必要ない知識を削除したりする．このイベントをワークショップと呼ぶ．現場の従業員が主体的に知識を記述する方法は，従来のインタビューを主体とした知識獲得[Schreiber 2000]や，大量のデータからの知識発見[河野 1997]とは異なる概念であるため，“知識発現”と呼んでいる．

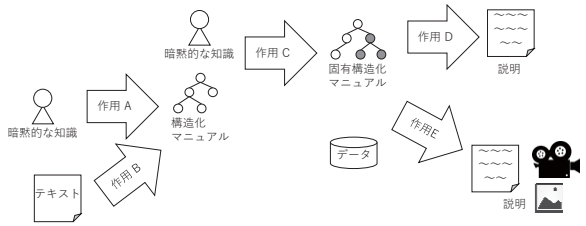
また，知識モデルとして CHARM (Convincing Human Action Rationalized Model)[Nishimura 2013]を用いることで，目的指向でプロセス知識を構造化するという特徴も備える．ここでの構造化とは，知識を構成要素毎に分け，要素間の関係性を表現することである．構造化の結果として大きな目的を達成するために詳細な行為が分解されるという階層性を持つ構造化知識が得られる．ここではそれを構造化マニュアルと呼ぶ．

大きな目的レベル（移乗するや入浴介助する等）では提供サービスに寄らず共通性がある．目的指向でプロセス知識を記述することで，目的レベルでの共通性を利用し，施設や提供サービスごとの相違を吸収することが期待できる．

3. 作用の整理

3.1 概要

図 1 に知識発現の前後で発生する作用(プロセス)を整理した模式図を示す．ここでは，作用を物の状態変化として捉えているが，それぞれの作用を分けている要因は，状態変化前後で，説明内容を持つ物への注目である．すなわち，作用 A は，人の頭の中などにある暗黙的な知識から構造化マニュアルへの変化であり，作用 B は自然言語文から構造化マニュアルへの変



	Type	Formalism	Description
Knowledge source	Implicit knowledge	Low	The knowledge is usually stored in human brain. It is difficult for others to refer the knowledge. This type of knowledge is shared by oral instruction in OJT.
	NL text Natural Language Text	Medium	The knowledge is usually stored in the form of text-manual. It is easier to refer the knowledge rather than implicit knowledge. However, others need to interpret the knowledge to understand it. For the information system, it is necessary to pre-process the text to utilize the content.
	Structured manual	High	The knowledge is not usually used in practical service fields. It is easy to refer the knowledge if the knowledge content is well-structured in the consistence manner. However, it is sometimes difficult to construct the knowledge in such structured manner.

図 1 知識発現に関連する作用

連絡先：西村悟史，産業技術総合研究所，東京都江東区青海 2-4-7，03-3599-8315，satoshi.nishimura@aist.go.jp

化であり，作用 D は構造化マニュアルから説明への変化である．作用 C は作用 A の特殊概念であり，作用 E は作用 D の特殊概念である．ここで，暗黙的な知識，自然言語文，構造化マニュアルは形式化の度合いで分類しており，それぞれ低い，中程度，高いとした．

3.2 作用 A

作用 A は，人間の中にある暗黙的な知識を構造化マニュアルとして記述することを指す．知識工学の分野では，これはインタビューなどの方法で行われてきた．その作用発揮主体は知識工学者が主であった．

これまでの知識発現の応用においては，作用 A は行われてこなかった．前述の通り，知識発現では共通プロセス知識を構築し，それをもとに現場の専門家が知識を記述する．これまでの応用範囲では，自然言語文で共通プロセス知識が記述されている場合のみであったため，作用 A は必要なかった．今後，共通プロセス知識が明示的でない場合は，従来の知識工学の方法に則って，作用 A を実施することになる．

3.3 作用 B

作用 B は，自然言語文で明示的に記述された内容を構造化マニュアルとして記述することを指す．知識工学の分野では，作用 A 同様に知識工学者が作用発揮主体として行う方法は過去に研究されている．関連研究として，Stanford Open Information Extraction[Angeli 2015]のように，自然言語で書かれた文章から重要語と重要語間の関係を自動的に抽出し，三つ組みに構造化するシステムが開発されている．

一方で，CHARM のように目的指向で手続き的知識を記述することを考えると，自然言語文には明示的に含まれていない中間目的やリスク等の情報が必要となる．計算機システムが作用 B を実施する場合は，このような足りない情報をどのように取得するかが課題となる．

3.4 作用 C

作用 C は，構造化マニュアルと暗黙的な知識を持った人間を入力として，暗黙的な知識を構造化マニュアルに追記することを指す．これは，作用の実施前に既に記述された構造化マニュアルが存在する必要がある点で，作用 A より特殊化された概念と言える．

一般に知識工学で作用 A を実施する過程でこの作用 C は表れる．一度に知識表現が出来上がるわけではなく，計算機が解釈しやすいように明示的に記述された知識表現を見ながらドメインの専門家からフィードバックを得て，知識が整理された知識表現が得られる．その過程は作用 C の一種として位置づけられる．

知識発現も作用 C の一種として位置づけられる．特に，従来研究との違いは，作用発揮主体が現場の従業員であるということである．これにより，継続的に知識表現の改訂の実施が期待できる．

3.5 作用 D

作用 D は，構造化マニュアルとして記述された知識をもとに，形式変換や部分抽出等の加工を通して人間向けの説明を生成することを指す．例えば，構造化マニュアルをリスト形式に表示方法を変えて提示することや，介護サービス利用者の行う行為のように条件に合致する内容のみを抽出して提示することなどが挙げられる．また，一度に情報を提示するだけではなく，インタラクティブな部分抽出を行うことで，チャットや音声対話で説明を生成するということもユースケースとして想定できる．5.2 節にて展望を述べる．

3.6 作用 E

作用 E は，構造化マニュアルに加えて現場の保有するデータから，説明を生成することを指す．これは，作用の実施前に現場の保有するデータが存在する必要があるという点で，作用 D より特殊化された概念と言える．

説明として，データが加わることにより，説明される人の理解促進が期待できる．すなわち，現場の保有するデータを構造化マニュアルの”根拠”として提示することで，説得性が増すことが期待される．

関連研究として，画像からキャプションを生成する研究がある[Ushiku 2015]．画像はデータの一例として捉えられるので，これはデータのみからの説明生成と言える．この手法では，構造化された他の言語資源を組み合わせた説明生成は考えられていないが，本研究でこのような手法を援用するためには，どのように構造化マニュアルと関連するデータを計算機が意味処理可能に表現するかが課題となる．

3.7 考察

本章では，知識発現とその前後で現れる作用について整理した．作用 A, B, C は知識工学や知識管理といった分野に属する．これらの分野では，どのように知識を作りだし，明示的に記述するかに焦点が当てられている．一方，作用 D, E は情報抽出やユーザインタフェース，自然言語生成などの分野に属する．これらの作用が示すところは，記述された知識をどのように利用するかに焦点が当てられている．

4. 知識発現と構造化マニュアルの適用範囲

本章では，これまでに行ってきた知識発現と構造化マニュアルの適用分野についてまとめる．

4.1 産業分類に則った分類

表 1 にこれまでに知識発現と構造化マニュアルを適用した事例を示す．産業分類の列には，日本標準産業分類[総務省 2013]の要素を示している．8 種類中，2 種類は業として行うものではないため分類には当てはまらなかった．

日本標準産業分類に当てはめるとこれまでに適用していない分野が多く存在する一方で，医療，福祉業や教育，学習支援業のようなサービス産業だけではなく，建設業や製造業での需要があることも見受けられる．そして，どのような目的で知識発現を用いてきたのかを外観すると，何かの知識を共有，継承することが目的とされていることが分かる．このことから，産業分類に則った分類のみでは，知識発現と構造化マニュアルの適用範囲を明確化することは難しいと考え，記述内容に則った分類についても考察した．

4.2 記述内容に則った分類

記述された構造化マニュアルの内容に基づいた分類を試みる．分類にあたり，[住田 2012]の基底作用，上位作用の考え方を参考にする．基底作用とは，「作用要求者が直接望む作用」であり，上位作用とは，「それを可能にする別の作用」のことである．

(1) 介護

介護分野では，基底作用は，介護サービス利用者（以下，単に利用者と呼ぶ）が生活する行為であると捉える．利用者は生

表 1 日本標準産業分類に則った適用範囲の分類

No.	分野	説明	産業分類
1	介護	マニュアル構築と統一	大分類P 医療，福祉
2	大学教育	アクティブラーニング時の振り返り支援	大分類O 教育，学習支援業
3	地域活性化支援	地域の介護予防リーダー育成支援	−
4	共想法	共想法に関する各種ノウハウの整理と記録の関連付け	−
5	建設	施工管理および各種情報の関連付け	大分類D 建設業
6	ごみ処理施設	プラントオペレーションの整理	大分類R サービス業（他に分類されないもの）
7	自動運転	運転行為と法令情報の関連付け	大分類E 製造業
8	ギター演奏	ギター演奏の訓練支援	大分類O 教育，学習支援業

活するという行為を実施したいが，老化等が原因となり，それを単独で実現することが出来ない．そこで介護士等が，生活する行為を可能にするように介助することが介護行為の主目的であり，上位作用と言える．特にこの上位作用は，「作用主体能力準備上位サービス」に分類できる．この分野では，上位作用の記述に主眼を置いてきた．

(2) 大学教育

大学教育分野では，基底作用は，学生が行う振り返る行為であると捉える．それを可能にする行為として，教員が学習教材を提供する行為があり，それが上位作用として認定できる．つまり，学習教材という基底作用で使用されるものを利用可能にする上位作用として捉え，「永久使用権提供上位サービス」に分類できる．

(3) 地域活性化支援

地域活性化支援分野では，基底作用は，地域の介護予防リーダーが活動する行為であると捉える．介護予防リーダーが活動することは，介護予防リーダーにより構成される組織とその管理行為が必要となる．そのような管理行為がここでは上位作用として認定できる．実際には，介護予防リーダーが組織管理を含めて活動を実施できることが望ましいが，ノウハウの不足等によりできないことが多い．そのため，地域の支援者が介護予防リーダーを支援することが行われる．そのような支援行為は，もう一つ上の上位作用として認定される．介護予防リーダーの組織管理行為と支援者の支援行為は，どちらも介護同様に，「作用主体能力準備上位サービス」に分類できる．この分野では，2 つの上位作用の記述に主眼を置いている．

(4) 共想法（認知症予防）

共想法分野では，基底作用は，共想法参加者が自分の認知能力を向上する行為であると捉える．上位作用は，共想法に参加する行為となる．つまり，共想法参加者は，加齢とともに認知能力が減衰するという基底作用を満たせないリスクを排除するために，共想法参加行為という上位作用を実施していると整理することが出来る．この点で，上位作用は，「作用主体能力準備上位サービス」に分類される．また，3 の地域活性化分野と同様に，共想法実施者による共想法イベントを成立させるという行為は，共想法参加者が共想法に参加する行為を可能にするという点で，上位作用とみなせる．共想法実施者の行為は，共想法イベントにおいて，共想法参加者の認知能力を向上させる主体を準備していると捉えられるので，「作用主体存在準備上位サービス」に分類できる．この分野では，共想法参加者，実施者それぞれの実施する上位作用に主眼を置いている．

(5) 建設

建設分野では，基底作用は，建設現場における重大事故を減らす行為であると捉える．上位作用は，それを可能にするための観察と判断行為である．すなわち，観察と判断行為により，ミスが予め発見，取り除かれることで，基底作用が実行可能になる．これは，4 の共想法参加者の参加行為と似ているが，作用主体の能力を準備しているというよりは，作用発揮を阻害する要因を排除することで，基底作用を可能にしていると認定できる．そのため，[住田 2012]の上位サービスの分類に当てはめるならば，「作用発揮関与物準備上位サービス」に分類される．この分野では，上位作用の記述に主眼を置いた．

(6) ごみ処理

ごみ処理施設分野では，基底作用は，プラントを操作する行為と捉える．ここでは，プラントを操作するという作用要求者の要求する作用がそのまま実施されているので，上位作用は考えない．この分野では，基底作用の記述に主眼を置いた．

(7) 自動運転

自動運転分野では，基底作用は，A 地点から B 地点へ移動することであり，上位作用は考えない．6 のごみ処理分野同様に基底作用の記述に主眼を置いた．

(8) ギターの演奏

健康増進分野のギターの演奏行為[Iino 2017]については，基底作用が，ギタリストによる演奏行為と捉える．これは，6, 7 と同様に上位作用は考えない．これら 3 つの分野は，行為やサービスの記述というよりも機械的な操作や機能の記述を行ったものとして分類できる．この分野では，基底作用の記述に主眼が置かれている．

4.3 考察

まず，日本標準産業分類に従って適用分野の分類を試みたが，産業として認定されない分野も見られるため，分類しきれない事例があった．また，分野をまたがっていたとしても，記述内容には共通性があるように思われた．そこで，次に，[住田 2012]で述べられている，基底作用，上位作用の分離に基づいて，記述内容から分類を試みた．結果として，事例(3), (4)に見られるように，基底作用，上位作用の 1 階層のみの整理ではなく，多段に階層を考慮できる内容もあった．これは，被支援者と支援者がそれぞれ実行主体となる作用間の相互作用を表していると考えられる．特に，被支援者の行為を阻害するリスク要因を支援者が取り除いている構図があることが想定される．

5. 知識発現の応用システム

3 章で述べたように，知識発現の前後で大きく分けて 5 つの作用がある．それらを支援するためのシステム開発について展望を述べる．特に，作用 A, B, C と作用 D に着目する．

5.1 知識発現支援システム

作用 A, B, C を支援することを目的として，kNowledge eXplication augmentaR (kNeXaR)と呼ぶシステム[Nishimura 2016]を開発中である．現状では，簡易なオントロジー構築機能とそこで定義された用語を用いて手続き的知識を構造化マニュアルとして構築する機能を実装している．スロットフィリングによって，入力可能な情報に制約を与えることで構造化マニュアルの構築が可能となっている．将来的には，発想支援や関連情報の推薦，オントロジーの情報を用いた無矛盾性の検証等の拡張を考えている．

5.2 知識ベースト対話システム

作用 D を実施するシステムとして，対話ロボットを用いたシステムについても考察を進めている．以下に，[Jokinen 2018]に示された対話例の和訳を載せる．

U: こんにちは，ゆき！
N: こんにちは，Kristiina! 何か手伝えることはありますか？
U: ええと，体位変換について手伝ってもらいたいです．
N: はい，体位変換ですね．どのように体位変換したいですか？
U: 座位に変えたいです．
N: はい，利用者さんの身体状態は確認しましたか？
U: はい，意識ははっきりしていて，ベッドで横になっておられます．
N: はい，では，私の指示に従ってください．準備が良ければ，おっしゃってください．
U: 準備できています．
N: はい，まず，水平に利用者さんを移動します．それが終わったら，次とおっしゃってください．
U: 次を教えてください．
N: はい，利用者さんを抱えて，水平に押しながら移動してください．その際に頭を支えてください．次に，利用者さんが適性に座れているかと気分が良いか確認してください．
U: 少し待ってください．
N: はい，ごゆっくりどうぞ．
U: ご利用者さまがきちんと座れました．
N: はい，気分はよさそうですか？
U: はい，気分もよさそうです．
N: はい，素晴らしいです．これで体位変換は終わりです．

ここで，U は対話ロボットと対話をしているユーザ，N は対話ロボットを指す．対話で用いる介助に関する知識は構造化マニュアルとして記述されている．このような対話を実現するためには，構造化マニュアルに記述された文字列を抽出するだけではなく，文脈に応じて語形を変換することや，ユーザの発話からキーワードを抽出し，それに基づいて適切な文字列を返却するような検索モジュールの開発が必要となる．

6. 結論

本稿では，西村らが提案した知識発現[西村 2017]の取り組みについて整理した．まず，知識発現を中心として，どのような作用がなされるのかを整理した．そして，これまでの適用範囲について，それが属する産業分野と，記述された内容から分類を試みた．最後に，今後の展望としてのシステム開発について述べた．

今後の展開として，知識発現を円滑に進めるためのノウハウを整理し，知識発現支援システムへの実装を進める．また，構造化マニュアルを OWL (Web Ontology Language)[1]等の標準言語を用いて形式化することで，既存の推論機構を援用し，無矛盾性検証など，構築支援についても検討を進める．同時に，形式化の議論は応用を伴って進めるべきものであるため，対話システムへの利用のしやすさの観点で議論を進めたい．

謝辞

本研究の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO）の委託業務および，JSPS 科研費 16K16160 の助成を受けたものです．

参考文献

[DARPA] David Gunning: Explainable Artificial Intelligence (XAI), https://www.darpa.mil/program/explainable-artificial-intelligence. (accessed: 2018.02.20)

[西村 2017] 西村悟史, 大谷博, 畠山直人, 長谷部希恵子, 福田賢一郎, 來村徳信, 溝口理一郎, 西村拓一: 現場主体の“知識発現”方法の提案, 人工知能学会論文誌, Vol. 32, No. 4, p. C-G95_1-15, 2017.

[Schereiber 2000] Schreiber, G., Akkermans, H., Anjewierden, A., Hoog, R. D., Shadbolt, N. R., Van de Velde, W., and Wielinga, B. J.: Knowledge Engineering and Management the CommonKADS Methodology，MIT Press 2000.

[河野 1997] 河野浩之: データベースからの知識発見の現状と動向 (< 特集> 大規模データベースからの知識獲得), 人工知能学会誌, Vol. 12, No. 4, pp. 497-504, 1997.

[Nishimura 2013] Nishimura, S., Kitamura, Y., Sasajima, M., Williamson, A., Kinoshita, C., Hirao, A., Hattori, K., and Mizoguchi, R.: CHARM as activity model to share knowledge and transmit procedural knowledge and its application to nursing guidelines integration, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 17, No. 2, pp. 208-220, 2013.

[Angeli 2015] Gabor Angeli, Melvin Johnson Premkumar, and Christopher D. Manning: Leveraging Linguistic Structure For Open Domain Information Extraction. In Proceedings of the Association of Computational Linguistics (ACL), 2015.

[Ushiku 2015] Yoshitaka Ushiku, Masataka Yamaguchi, Yusuke Mukuta, and Tatsuya Harada. Common subspace for model and similarity: Phrase learning for caption generation from images. IEEE International Conference on Computer Vision (ICCV), pp.2668-2676, 2015.

[総務省 2013] 日本標準産業分類, http://www.soumu.go.jp/toukei_toukatsu/index/seido/sangyo/02toukatsu01_03000023.html. 2013. (accessed: 2018.02.20)

[住田 2012] 住田光平, 來村徳信, 笹嶋宗彦, 高藤淳, 溝口理一郎: オントロジー工学に基づくサービスの本質的性質の考察, 人工知能学会論文誌, Vol. 27, No. 3, pp. 176-192, 2012.

[Iino 2017] Nami Iino, Satoshi Nishimura, Ken Fukuda, Kentaro Watanabe, Jokinen Kristina, Takuichi Nishimura: Development and use of an Activity Model based on Structured Knowledge - A music teaching support system -, The 5th International Workshop on the Market of Data, pp. 1-6, 2017.

[Nishimura 2017] Nishimura, S., Fukuda, K., and Nishimura, T.: Knowledge Explication: Current situation and future prospects, IJCAI 2017 Workshop on: Cognition and Artificial Intelligence for Human-Centered Design, pp. 1-7, 2017.

[Jokinen 2018] Jokinen, K., Nishimura, S., Watanabe, K., Nishimura, T.: Human-Robot Dialogues for Explaining Activities, International Workshop on Spoken Dialog System Technology, pp. 1-10, 2018.

[1] https://www.w3.org/TR/owl2-overview/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ギター奏法オントロジーを用いた指導者間の知識共有・獲得

Knowledge Sharing and Acquisition among Teachers using Guitar Rendition Ontology

飯野 なみ[*1,2] 西村 悟史[*1] 西村 拓一[*1] 福田 賢一郎[*1] 武田 英明[*2,3]
Nami Iino Satoshi Nishimura Takuichi Nishimura Ken Fukuda Hideaki Takeda

[*1] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

[*2] 国立情報学研究所
Graduate University for Advanced Studies

[*3] 国立情報学研究所
National Institute of Informatics

In music field, the knowledge of players and teachers is important, and it can be learning and teaching support by describing and providing them correctly. But the knowledge of classical guitar is not standardized now. Therefore, common teaching method has not been established even for fundamental skill. In this paper, we developed a Guitar Rendition Ontology of classical guitar for knowledge sharing/acquiring among teachers. We structured knowledge for expanding the goal-oriented knowledge that we constructed on our previous research. Through hearing survey on the ontology we developed, we considered the usefulness of knowledge sharing/acquisition and future work for application to teaching site.

1. はじめに

近年の情報社会に伴い，適切な知識の使用が重要視されている．膨大な知識を扱うためには，目的に応じた情報の収集，選択，統合化を行う必要がある．特に知識処理の分野においては，オントロジーに関する様々な研究が行われている．オントロジーとは，哲学用語で「存在論」のことであり，人工知能の用語では「概念の明示的な仕様」と定義される．知識処理システムが対象とする世界のオントロジーを構築することで，対象となる概念の合意による知識の共有や再利用への貢献が期待できる[伊藤 2000].

音楽分野においても，演奏家や指導者が持つ知識は重要な資産であり，それらを正しく記述し提供することで学習や指導を支援できる．中でも，クラシックギターは演奏に関する知識や学習プロセスの標準化がなされておらず，演奏者ごとの技能の差が大きい．指導現場においては，指導者間の知識共有がなされず，基礎的な技術でさえ共通の教授法が確立されていない．

そこで本研究では，指導者間の知識共有・獲得を目的としてクラシックギターの奏法オントロジーを構築する．先行研究で取り組んだ目的指向知識を一次情報として，オントロジーに基づいた構造化を行う．そして構築したオントロジーに対するヒアリング調査から，知識共有・獲得の有用性と，指導現場への応用に向けた今後の展望について述べる．

2. 関連研究

2.1 オントロジーの活用

オントロジーの活用としてまず，語彙の標準化が挙げられる．農業現場では，農作業語彙の多様性と基準となる情報がないためにそれぞれの農業 IT システムが独自の用語や表記で運用されており，データ連携が難しい．そこで，共通農業語彙として農作業基本オントロジー（Agriculture Activity Ontology）が構築され，農作業における標準語彙として適用されている[朱 2015].

連絡先：飯野なみ，産業技術総合研究所，東京都江東区青海 2-3-26，090-7276-5103，nami-iino@aist.go.jp

また医療分野では，多様な患者状態を表現する方法として，臨床医学オントロジーが構築されており，病名用語の意味関係を持たせることで疾患の体系化を図っている[大江 2009].

クラシックギターにおいても，指導者によって表現が異なることから，語彙体系が求められる．また，エレキギターやフラメンコギターといった他ジャンルで使用される用語との差別化も必要である．オントロジーを用いることで，ギター界の語彙の標準化だけでなく，複数の表現を持つ奏法との関係を示すことができる．

2.2 現場主体の知識共有・獲得

知識発現（Knowledge Explication）は，サービス現場における知識の共有を目的として提案された表現枠組みである [西村 2017]. まず，記述モデル CHARM（Convincing Human Action Rationalized Model）を用いて，基盤となる手続き的知識（プロセス知識）を構造化する．その後，基盤の知識に対して従業員が主体となって現場固有のプロセス知識を記述していく．既に介護現場での使用が試みられており，その有用性が示されている．実施方法はワークショップ形式で，紙に印刷した共通プロセス知識に対して，従業員らが議論を通して付箋に記述した事例を貼り付けることで固有プロセス知識を得るという方法を取っている[西村 2016].

3. ギター奏法オントロジーの構築

3.1 既存知識の収集

筆者らの先行研究では，2.2 節で述べた知識発現に基づき．クラシックギターの目的指向知識の構築を行なった[飯野 2017]. クラシックギター指導支援を目的として，クラシックギターの奏法に関する知識と習得プロセスの共通理解を図ることを試みた．既存の教本などから知識を収集した後，(1)研究者による既存知識の収集とグループ化，(2)プロギタリスト兼指導者による知識の具体化と構造化を行った．(2)はワークショップ形式で実施し，習得プロセスの分類や知識の修正，追加などを行った．

記述形式は手続き的知識であり，「押弦する」，「撥弦する」，「消音する」，「音色を変える」といった 20 あまりの目的行為を上位として取り上げ，それらを達成するための手順を構造化した．し

かし，クラシックギターの知識の基盤としてこの記述方法には大きな問題がある．まず，使用する語彙が概念化されていないことから，宣言的知識と手続き的知識が混同している．また，目的行為の手順における言語表現の曖昧性も指摘されることから，この構造化知識をそのまま用いることは困難である．

3.2 奏法概念の構造化

図 1 に，オントロジー構築ソフトである protégé を用いて記述した概念（owl:Class）と属性（owl:ObjectProperty）を示す．本研究では，クラシックギター奏法を概念として，その目的別に is-a 関係（上位・下位関係）による階層化を行った．具体的には，音に関する目的を持つ 7 つの奏法：弦を使用せずに他楽器を模倣した「効果音奏法」，和音の弾き方に関連する「和音変化奏法」，記譜された音に装飾的に音を加えるなどの「装飾奏法」，音の長さを変化させる「音価変化奏法」音と音との繋ぎ方を変える「音節変化奏法」，弦を使用して音色を変化させる「音色変化奏法」，重音などを用いて響きを変化させる「音響変化奏法」と，身体に関する目的を持つ奏法：押弦や撥弦といった基本的な行為に関する「基本奏法」に大別した．さらに基本奏法は，右手を安定して動かすための「固定奏法」，押弦時のフォーム（型）をまとめた「押弦奏法」，撥弦の仕方である「撥弦奏法」に分類した．

3.3 属性と値

奏法の属性として，「目的音」，「表現形態，「実行行為」をメインに構造化した．「目的音」は理想とする音を示し，「ギター音」クラスの「打楽器音」や「滑走音」，「倍音」といった 22 の値を定義した．ただし，基本奏法のように目的が“安定させる”などの場合は適用されないため，拡充が必要である．

「表現形態」は楽譜に書かれている記号を示す．値は，「演奏記号」クラスの，楽器に依存しない音の扱い方を示す「アーティキュレーション記号」，バロック時代に頻繁に見られる即興的に音を追加する「装飾記号」，楽器独自の特殊奏法を表す「楽器固有記号」とした．

3.4 実行行為の記述方法

前節で述べた 3 つ目の属性である「実行行為」では，奏法の持つ複数の行為の組み合わせや順序性，並列性を考慮して記述した．本稿では，特に重要である身体の表現，行為の順序・並列について述べる．

クラシックギターは，身体に楽器を固定した状態で演奏するため，奏法の記述に必要な身体部位は左右の手指の使用に限定することができる．ただし，利き手を考慮して「指板側の手指」と「胴板側の手指」と分類することとした．また，実行行為の値は「弦を弾く（はじく）」，「弦を押す」，「弦を叩く」などと動詞で表している．そこで，身体の表現には「弦を弾く（胴）」，「弦を押す（指）」という記述方式をとった．

順序性や並列性を持つ奏法は，各行為を分解して定義した．まず，順序を持つ奏法は，実行行為 1，実行行為 2 と属性を分けることで表現した．例えば，撥弦奏法の下位概念である「アポヤンド」は，実行行為1：「弦を弾く（胴）」，実行行為2：「隣の弦に持たれかかる（胴）」，と定義される．ただし，これらの数値に順序的意味は持たない．順序の定義は Protégé では限界があるため，OWL で公理を作成する必要がある．次に，「B をしながら A をする」と言った並列行為については，A を主行為，B を条件行為とし，ブランクノードを介すことで実行行為プロパティに各プロパティが属することを表した．例えば，音色変化奏法の下位概念である「ピッチカート」は，主行為：「弦を弾く（胴）」，条件行為：「弦に触れる（胴）」と記述した（図 2）．

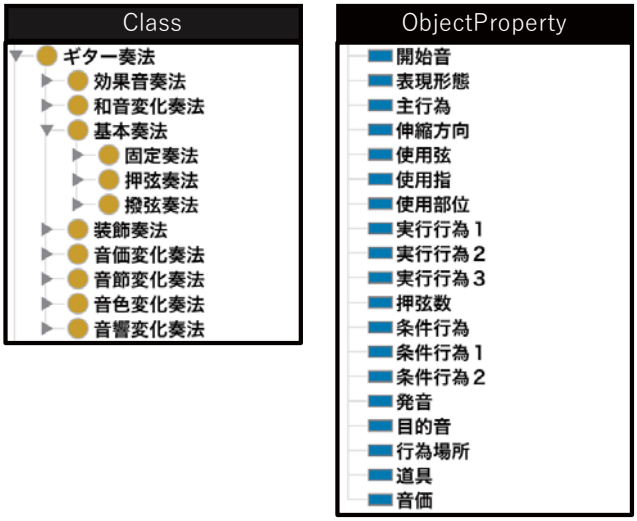


図 1　ギター奏法オントロジーの概念と属性



図 2　実行行為の記述形式

4. オントロジーを用いた知識共有・獲得

4.1 拡張性の検討

構築したギター奏法オントロジーの標準化に向けて，ヒアリング調査による知識の十分性や問題点を確認した．調査対象はクラシックギターの演奏経験を持つ者とし，ギター学習者 3 名と指導者兼演奏家 2 名それぞれ個別に，ギター奏法オントロジーに対する意見を求めた結果，以下の有益な知見が得られた．

●**概念**：知識の十分性を確認したところ，記述されていない知識が得られた．それらは主に近・現代に生み出された奏法であり，より多くの作曲家や演奏者との知識共有が必要であることが分かった．また，「ギターの構え方や爪の削り方，磨き方も奏法ではないか」，「消音も奏法ではないか」「トリルやモルデントなどの装飾奏法はスラー（奏法）を用いて実現するので奏法ではない」といった意見から，ドメインを明確にすることも課題であることが分かった．

●**属性**：属性の記述には課題が見られ，属性や値の修正，追加が行われた．特に実行行為の動詞表現については主観的要素が強いことから．標準化に向けた語彙の統一が求められた．例えば，和音奏法の下位概念であるカッティングの行為である「かき鳴らす」は連続した撥弦に感じるため，「弾き下ろす」と表現する，といった意見が得られた．

以上のように，研究で構築したギターオントロジーは知識の獲得や知識共有に有用であり，拡張できることが示された．さらに

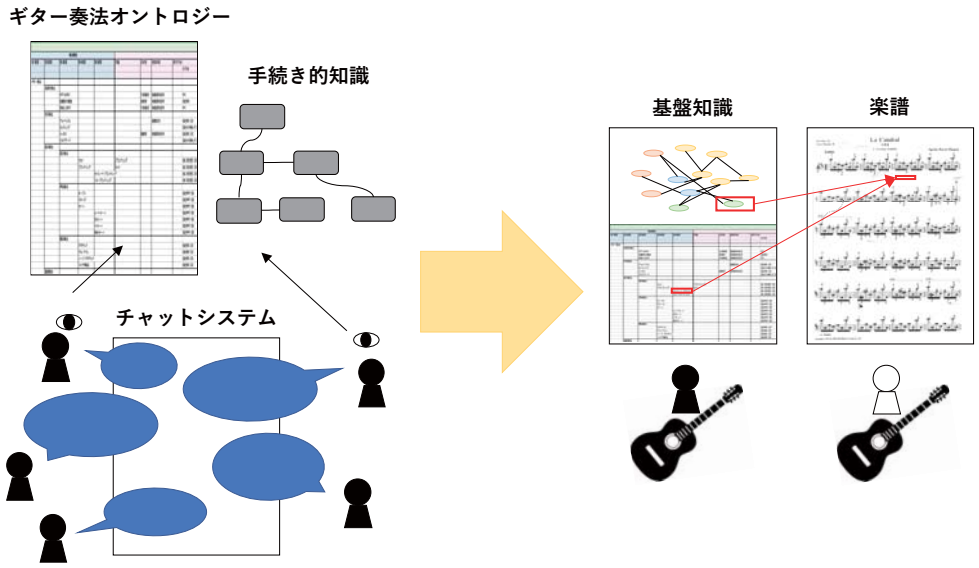


図 3　基盤知識の構築と利用

有識者の意見を反映させることで，より確かな標準的なオントロジーの構築が可能となる．

また，ギター奏法オントロジーの利用の可能性について学習者と指導者それぞれの立場から意見を求めたところ，前者は「従来の教本以上に体系化されていて分かり易いが，より具体的な記述が欲しい」との意見があった．後者では，「このオントロジーを通して議論を行うことで知識共有と指導者自身の学習に繋がる」，「自身の理解の曖昧さに気づける」，「オントロジーによって適切な指導ができる」，といった肯定的な意見が多く得られた．また，両者共通の意見として，実際に学習・指導現場で使用する際には，理解しやすいオントロジーデザイン，イラスト・動画データとの連携が挙げられた．

4.2　指導現場への応用

本節では，今後の展望として，構築したギター奏法オントロジーの指導現場への応用について述べる．

図 3 は，ギター奏法オントロジーと手続き的知識を用いた基盤知識の構築と利用の構図である．前節で示したように，ギター奏法オントロジーは知識共有・獲得には有用であるが，指導現場では，より具体的な説明や情報提供が必要である．そのためには，人が理解しやすい方法として知識発現のような手続き的知識が適切だと考える．そこで，オントロジーと手続き的知識の連携によって指導基盤となる知識の構築を試みる．実装としては，連携された知識に対して，チャットシステムによる指導者間の知識の共有・獲得を行う．そして，構築された基盤知識を学習者の楽譜にリンクすることで，指導の記録や学習過程の可視化を可能にする．

5.　おわりに

本研究では，指導者間の知識共有・獲得を目的としてギター奏法オントロジーを構築した．まず，クラシックギター奏法を概念として，その目的別に分類することで階層化を図った．次に，属性を目的や表記，行為に着目して定義した．構築したオントロジーに対するヒアリング調査では，知識共有・獲得の有用性を示した，そして今後の展望として，指導支援に向けたオントロジーの展開について考察した．

今後，より確かな標準的なオントロジーの構築を目指し，知識の拡充とシステム開発を進めていく方針である．

参考文献

[伊藤　2000]　伊藤英毅:オントロジーを利用した知識の共有/再利用，Unisys Technology Review，第 64 号 (2000).

[飯野　2017]　Nami Iino, Satoshi Nishimura, Ken Fukuda, Kentaro Watanabe, Jokinen Kristina, Takuichi Nishimura: Development and use of an Activity Model based on Structured Knowledge ‐ A music teaching support system ‐, IEEE International Conference on Data Mining Workshop, November 18, 2017.

[朱　2015]朱成敏，武田英明，法隆大輔，竹崎あかね，吉田智一，農業 IT システム間データ連携のための農作業 基本オントロジーの構築，電子情報通信学会知的 環境とセンサネットワーク研究会，ASN2015-75 (2015).

[大江　2009]　Kazuhiro Ohe: Standardization of disease names and development of an advanced clinical ontology, Journal of Information Processing and Management, olume 52, Issue 12, pp.701-709 (2009).

[西村　2017]　西村悟史，大谷博，畠山直人，長谷部希恵子，福田賢一郎，來 村徳信，溝口理一郎，西村拓一: “現場主体の“知識発現”方法の提 案”，人工知能学会論文誌，Vol. 32, No. 4C (2017)

[西村　2016]　西村悟史，大谷博，畠山直人，長谷部希恵子，福田賢一郎，來村徳 信，溝口理一郎，西村拓一:現場主体で介護業務知識を作るための知識モデルの検討，人工知能学会研究会資料，SIG-SWO-039-03.

http://protege.stanford.edu/

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

環境に適したテーラーメイドの「脳トレ」プログラム開発のための効果検証
～Development of "Tailor-made Brain Training" program suitable for environment～

押山千秋*1　三輪洋靖*2　須藤千尋*3　清水栄司*3　西村拓一*2
Chiaki OSHIYAMA　Hiroyasu MIWA　Chihiro SUTOU　Eiji SHIMIZU　Takuichi NISHIMUA

*1 大阪大学 Osaka University #1
*2 産業技術総合研究所 National Institute of Advanced Industrial Science and Technology #2
*2 千葉大学 Chiba Universiy#3

The problem of child mental health has been increasing. In recent years, it were reported that prevention of school education were important. Cognitive function affects mental health. The effect of cognitive function training "brain training" on children has also been verified, and it was reported that "brain training" have improved "cognitive function" and "intelligence factor". But there is no brain training program aimed at children's mental health. Also, there is no program that incorporates it into the context of school life. We have developed a brain training program which can be done in a short time at school and to verify effect. First, we used “mental rotation task” that activates a wide range of brain parts. We would like to report on the result of training for 61 university students for 2 weeks and the result of training at one class of elementary school 4th.The purpose of this research is to verify the effect to lead to tailor-made program construction by AI

1. はじめに

(1)学校での予防教育と脳トレ　神経科学研究や実験心理学研究において、様々な課題が用いられてきた。精神医学においてはコンピューター認知課題プログラムによる訓練を取り入れた認知機能リハビリテーションが行われ、認知機能が向上したという検証結果が報告されている。それらの認知課題を組み合わせた「脳トレ」と言われるものが、様々に作られている。大人を対象にした「脳トレ」の効果は近年、その効果が実証されたが、子どもに対する「脳トレ」の効果も、500 人以上のサンプルで検証され[Wexler et al., 2017]、「認知機能」や「知能指数」が向上したと報告されている。子どものメンタルヘルスの問題は顕著になってきている。近年、子どもたちがその発達の中で多くの時間を過ごす学校教育現場での予防的取り組みの重要性が指摘されている。認知機能は心の状態にも影響するが、認知機能向上のための脳トレの効果が近年報告されているが、子どもの精神状態を健全にすることを目的にした脳トレプログラムはまだない。また、予防教育が行われる学校生活の文脈に取り入れる形での検証は行われていない。さらに、どのプログラムが、どのような効果をもたらすのかについて、詳細に検討されていない。子どもの精神疾患予防の視点から、学校において短時間で行える、タブレット型コンピューターを用いた、脳トレーニングプログラムを開発し、その効果を検証することを本研究の目的とした。

(2)メンタルローテーション課題　数多くある脳トレのプログラムのなかから、まずは、1000 報を超える先行研究がある認知課題であるメンタルローテーションでの効果検証を行い、個別のテーラーメイドのシステム構築へと繋ぐ。「メンタルローテーション課題」は、心の中に思い浮かべたイメージ（心的イメージ）を回転変換する認知的機能を使った課題である[Kosslyn et al.,1998]。メンタルローテーション機能のおかげで、私たちは、実生活で、図や文字を回転した角度から見てもその文脈を判断することが出来る。例えば、違う角度から地図をみて、適切な方向を判断したり、レストランなどでメニューを逆さから見て内容を判断する。自らの身体を移動することなくそれらが可能になるのもメンタルローテーション機能のおかげである。

メンタルローテーションで賦活する神経回路は、うつや不安に強靭な神経回路と重なっている。毎朝のトレーニングで、うつや不安に強い健康な脳への発達を促す可能性がある。メンタルローテーションを遂行するとき、広範囲の脳領域が活性化する[Tomashino et al., 2016]。中でも、回転に関連した後部頭頂葉という脳領域は、認知的柔軟性に関わっている部分でもある[Cohen et al.,1997]。認知的柔軟性が高いとうつや不安が低いということが明らかになっている[Oshiro et al., 2016]。メンタルローテーションの正答率は神経ネットワークを表現しており[Semrud-Clikeman et al., 2012]、反応時間は精神的な運動状態を反映している[Rogers et al., 2002]。メンタルローテーションの正答率が有意に高い男性は、うつや不安障害の罹患率が女性の 2 分の 1 である[Rosenfield, 1999]。

上記から、以下の 3 つも予測される。①「メンタルローテーション課題」を授業前に短時間行うことで後部頭頂葉を含む広範な神経ネットワークが鍛えられ、うつや不安に強い脳と心の発達を促進する②授業前に様々なメンタルローテーショントレーニングを積み重ねていくことで、高い脳の可塑性を持つ子どもの認知機能が向上する。③朝の学活時に毎日行うことで、子どもが学校生活へと気持ちを切り替えられ、適応的行動が身につく。それらは副次的に、計算や言葉の能力向上に波及し、効果を及ぼすと考えられる。

2.1 研究1　大学生を対象とした検証

(1)目的: 青年期の対象者に対するメンタルローテーショントレーニングの効果を検証する。

(2)方法: ＜対象＞大学生 61 名（男性 23 名, 女性 38 名, 平均年齢 22.4 歳, *SD*=3.69, 範囲 19-28）

(3)手続き:無作為に 2 群に割り付け、メンタルローテーショントレーニング期間とメンタルローテーショントレーニングを行わない対照期間のクロスオーバー試験を行った(Figure1)。 メンタルローテーショントレーニング期間には, 1 日 2 回, 朝と夜, iPad での MR_tr. を 2 週間試行してもらった。メンタルローテーショントレーニング期間および対照期間前後で PHQ-9（うつ）, GAD-7（不安）, CFS（認知的柔軟性）および正答率と反応時間を評価した。

メンタルローテーション課題は、十字のフィックスに続いて 300msec のブランク、その後、手の画像 1000msec、反応があるか、画像 1000msec とその後のブランクの合計 4000msec で次の十字フィックスへと繋がる。使用された手の画像は肢位 4 方向（手掌、手甲、親指側回内、小指側回外）×回転角度 4（0°、90°、180°、270°）×右手左手 2 の 32 枚である。

(4)結果：グループ×時期の 2×3 の 2 要因反復測定分散分析の結果、主効果が見られたのは、メンタルローテーション正答率（$F(1,59) = 15.20$, $p < .0001$, $\eta_p^2 = .205$），メンタルローテーション反応時間（$F(1,59) = 35.59$, $p < .0001$, $\eta_p^2 = .376$），PHQ-9（$F(1,59) = 13.054$, $p = .001$, $\eta_p^2 = .181$）であった。グループ×時期での有意な交互作用は、メンタルローテーション反応時間で見られた（$F(1,59) = 8.73$, $p = .004$, $\eta_p^2 = .129$）。反応時間におけるグループの単純主効果はグループ A において($F(2.118)=20.17$, $p<.0001$, $\eta_p^2 = .221$) ,グループ B において ($F(5.56)=25.67$, $p<.0001$, $\eta_p^2 = .252$)であった。次に、トレーニング期の介入前後での多重比較を行った結果、グループ A の反応時間で（$F(1,60) = 39.14$, $p < .0001$, Hedge's $g = -.966$）グループ B の反応時間で、（$F(1,60) = 4.61$, $p = .036$, Hedge's $g = -.319$），グループ A の反応時間で（$F(1,60) = 11.16$, $p = .001$, Hedge's $g = .489$）グループ B の反応時間で（$F(1,60) = 4.68$, $p = .035$, Hedge's $g = .401$）、グループ A の PHQ-9 で（$F(1,60) = 8.18$, $p = .006$, Hedge's $g = -.539$）と、有意な結果となった。

(5)考察：メンタルローテーショントレーニングによって、正答率は向上するが、正答率は2，3回のメンタルローテーション課題を経験することによる回帰効果でも向上することが示唆された。反応時間はトレーニングによって有意に短縮することが示された。対象を大学生とした場合、トレーニングによって、正答率が向上し、トレーニング効果は反応時間により反映されると考えられた。また、大学生にメンタルローテーショントレーニングをうつ傾向が改善する可能性があることが示唆された。認知的柔軟性は統計的に有意な向上は認められなかった。

2.2 研究 2　小学校での検証

(1)目的：小学校の朝の学活時に一斉にメンタルローテーショントレーニングが行えるかどうかを試行し、トレーニングの効果を検証する。

(2)方法：＜対象＞関東にある小学校 4 年 X 組に所属する小学生 34 名（1 名はお休みにより、介入前のデータが欠落したため、解析対象からは除外したが、トレーニングには参加してもらった）

(3)手続き:20XX 年 10 月 23 日から 1 か月間、毎朝の学活時にアプリの入った iPad を配り、一斉にメンタルローテーショントレーニングを行った。また、トレーニング後に児童と教員にアンケート調査を行った。メンタルローテーション課題は、はじめの 2 週間は十字のフィックスに続いて 300msec のブランク、その後、手の画像 1000msec、反応があるか、画像 1000msec とその後のブランクの合計 4000msec で次の十字フィックスへと繋がる。使用された手の画像は肢位 4 方向（手掌、手甲、親指側回内、小指側回外）×回転角度 4（0°、90°、180°、270°）×右手左手 2 の 32 枚である。また、児童の個人差を考え、2 週目までの結果から、正答率が 90%以上を高群、正答率 80%未満の児童を低群、それ以外を中群とした 3 群に分けた。中群にはそのまま、十字のフィックスに続いて 300msec のブランク、その後、手の画像 1000msec、反応があるか、画像 1000msec とその後のブランクの合計 4000msec で次の十字フィックスへ、高群には十字のフィックスに続いて 300msec のブランク、その後、手の画像 800msec、反応があるか、画像 800msec とその後のブランクの合計 3000msec で次の十字フィックスへ、低群には十字のフィックスに続いて 300msec のブランク、その後、手の画像 2000msec、反応があるか、画像 2000msec とその後のブランクの合計 4000msec で次の十字フィックスへとした。アンケートは、毎朝 1 か月間のトレーニングを終えて、①とても面白かった②まあまあ面白かった③特に面白くも面白くなくもなかった④面白くなかった⑤全く面白くなかった、の 5 段階で評価をしてもらった。

(4)結果：メンタルローテーションの正答率と反応時間を、トレーニング前後で反復測定分散分析による比較を行った。メンタルローテーションの正答率は($F(1.33)=89.53$, $p<.0001$, $\eta_p^2 = .731$)、メンタルローテーションの反応時間は($F(1.33)=62.22$, $p<.0001$, $\eta_p^2 = .653$)。児童 35 名のうち、①とても面白かったと答えた児童は 32 名、②まあまあ面白かったと答えた児童は 3 名で、③特に面白くも面白くなくもなかった④面白くなかった⑤全く面白くなかったと答えた児童はいなかった。また、もっとやりたい、とか、インターネットでメンタルローテーション課題について調べたので他のメンタルローテーション課題もみんなでやってみたいという感想もあった。教員からは、タブレットを配り始めてから終了まで 4 分であったことから、学校生活への負担はないと考える、とのコメントをいただいた。

(5)考察：朝の学活試行に無理のないシステムであったと考える。

参考文献

[Wexler et al., 2017]Bruce E. Wexler, Markus Iseli, Seth Leon, William Zaggle, Cynthia Rush, Annette Goodman, A. Esat Imal & Emily Bo. Cognitive Priming and Cognitive Training: Immediate and Far Transfer to Academic Skills in Children.Scientific Reports, 6:32859,.2016.

[Kosslyn et al.,1998]Kosslyn, S.M., DiGirolamo, G.J., Thompson,W.L., Alpert, N.M. Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. Psychophysiology. 35, 151–161.1998.

[Tomashino et al., 2016]Tomasino, B., Gremese, M., Effects of stimulus type and strategy on mental rotation network: an activation likelihood estimation meta-analysis. Front. Hum. Neurosci. 9, 693. 2016.

[Cohen et al.,1997]Cohen, M.S., Kosslyn, S.M., Breiter, H.C., DiGirolamo, G.J., Thompson, W.L., Anderson, A. K., Brookheimer, S.Y., Rosen, B.R., Belliveau, J.W., Changes in cortical activity during mental rotation. A mapping study using functional MRI. Brain. 119, 89–100. 1996.

[Oshiro et al., 2016] Oshiro K, Nagaoka S, Shimizu E. Development and validation of the Japanese version of cognitive flexibility scale. BMC Res Notes. ;9:275. 2016.

[Semrud-Clikeman et al., 2012]Semrud-Clikeman, M., Fine, J.G., Bledsoe, J., Zhu, D.C., Gender differences in brain activation on a mental rotation task. Int. J. Neurosci. 122, 590–597. 2012.

[Rogers et al., 2002] Rogers, M.A., Bradshaw, J.L., Philips, J.G., Chiu, E., Mileshkin, C., Vaddadi, K., Mental rotation in unipolar major depression. J. Clin. Exp. Neuropsychol. 24, 101–106. 2002.

[Rosenfield, 1999] Rosenfield, S., Gender and mental health: do female have more psychopathology, men more, or both the same (and why), in: Horwitz, A.V., Schneid, T.L. (Eds.), A Handbook for the Study of Mental Health: Social Contexts, Theories, and Systems. Cambridge University Press, New York, pp. 348–360. 1999.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

表現のための活動プログラムのデザイン:看護の心を表現すること

Design of Activity Program for Expression :project for expression on caring mind

小早川真衣子*1 Maiko Kobayakawa
須永剛司*2 Takeshi Sunaga
丸山素直*2 Sunao Maruyama
平野友規*1 Tomoki Hirano
山田クリス孝介*4 Kosuke C Yamada
西村拓一*3 Takuichi Nishimura
渡辺健太郎*3 Kentaro Watanabe
藤満幸子*5 Sachiko Fujimitsu

*1 東京藝術大学大学院 Tokyo University of the Arts Graduate School
*2 東京藝術大学 Tokyo University of the Arts
*3 産業技術総合研究所 National Institute of Advanced Industrial Science and Technology
*4 慶應義塾大学 Keio University
*5 佐賀大学附属病院看護部 Saga University Hospital

I am researching on designing "a new activity for nursing work" at a hospital in collaboration with nurses. In 2012, when this project began, we recognized the problem of design was to redesign the information system terminal in the hospital to such a thing. And now, we have practiced a "workshop to expression on nursing mind" with nurses. In this design project that include research, in order to clarify the formation of that project, we have reflected the design process on “Activity program” and “tools”. As a result, we found some characteristics attitude of Co-design.

1. 概要

病院での「看護の仕事の新しいあり方」を看護師たちと共同でデザインすることに取り組んでいる. このプロジェクトがはじまった 2012 年, 私たちは院内の情報システム端末を如何に改善するのかということをデザインの問題として認識していた. それから 6 年が経った今, 私たちは看護師と共に「看護を語る場」をつくっている. そのねらいは, 看護の心を表現することである.

このプロジェクトにおける共同デザインは、どのように成り立っているのだろうか？本稿では, 私たちが出会った予想外で驚きの出来事に注意を向け, 生成される状況とどのように関わったのか. また, 「看護を語る場」のための「活動プログラム」と「道具」のデザインをどのように志向していったのかを私たちの認識とその変化に焦点を当て, ふり返る. そして, そうした状況と対話しながらデザインしている実態を考察する.

2. 本研究の位置づけ

2.1 着想:活動プログラムのデザイン

これまでに見たことのない新しい非物質の道具が, 人に受け入れられない, 人の営みに埋め込まれないという問題はしばしば起こりうる. そこで, 私たちは非物質の道具と一緒に最初にデザインされる企図としての「活動プログラム」という対象に着眼している[小早川, 2009]. 活動プログラムは, 「活動」そのものではなく, 道具の使用を含む活動を方向づける企図でありシナリオである. 実際の活動は人間がさまざまなリソースとの相互作用によって生成するものであるという理解を実践ごとに深めていきながら, それらリソースを組み立てるための活動プログラムがデザインされることが必要と考えている. 物質であれ非物質であれ, 道具はリソースのひとつであり, 生成する活動によって意味づけられるものだという着想である. しかし, 活動プログラムをデザインする方法は未だ明らかではない.

連絡先:小早川(小原)真衣子, 東京藝術大学大学院美術研究科, 東京都台東区上野公園 12-8, koba@asu.aasa.ac.jp

2.2 探究:どのようにデザインしているのか？

ここで探究している活動プログラムのデザインの「方法」は, デザインした活動プログラムが計画的に正しく実行され完成するためのやり方を意味するものではないことを注意しておかなければならない. なぜならば, 実際の活動は, 何が起きるかわからないからだ. 例えば, 私たちが一種の“ドローイング”として実施する「表現ワークショップ」は一般の人たちが参加する体験型の活動であるが, その日集まる表現者の個性や, そこにどのような反応が起きるかは分からない. さらに, そこで起きたことと同じことは二度とは起こせない. 不確実で個別的な状況がそこにある. この観点をふまえ, ふり返る際は, 想定外の出来事や私たちが驚いたことに注意を向ける必要があると考える. 次章では, 実際にそうした特徴をもつ出来事を事例に挙げ, その状況を改めて記述し, 解釈していきたい.

3. 共同デザインの実践:MED プロジェクト

3.1 プロジェクトの概要

病院での看護の仕事の新しいあり方をデザインするために, 現場の実践者である看護師たちとの協働に取り組んでいる. 大学病院の看護部の看護師長数名(以下, 看護 G)と心理学系の研究者, サービス工学を専門とする研究者, デザイン系の研究者からなる研究グループ(以下, 研究 G)が参加しており, 学際共同研究「MED プロジェクト」と呼んでいる[藤満 2013, 須永 2013]. このプロジェクトでテーマとなっている看護の仕事の新しいあり方のデザインとは, 病院に次々に導入される情報システムの利用によって失われつつある「看護の心」を表現することによって外化し共有する仕組みと文化をつくることである. ここでいう「看護の心」とは, 看護師たちが通常業務で患者に対面・対話し

看護処置にあたる際に欠かせない，患者やその家族，そして仲間を想う気持ちである．

このプロジェクトがはじまった 2012 年，研究 G も看護 G も病院内の情報システム端末を如何に改善するのかという課題を認識していた．それから 6 年が経った今，研究 G は看護師と共に「看護を語る場」をつくっている．その場が

3.2 現時点の成果：新聞ワークショップ

「看護を語る場」としてデザインした「新聞ワークショップ」の概要を示す．看護師 20 名程が集まり 90 分間で行う下記 6 つの行為系列（S01～S06）からなる活動プログラム「表現ワークショップ」と道具立てとしての「新聞ワークショップ・キット」をデザインした．当初は研究 G が運営主体となっていたが，2017 年以降，年に数回，看護 G によって自立的に運営がなされている[宮之下 2017]．（写真 1）

S01 導入：他者の表現を見る
S02 個人表現　：体験を思い出し作文を書く
S03 個人表現を味わう：一人ずつ朗読する／傾聴する
S04 共同表現：グループで新聞をつくる
S05 共同表現を味わう：発表する
S06 ふり返り：表現体験をふり返る

各行為でどのような体験ができるのか，概要を示す．部屋に入ってきた看護師（表現者）は，まず過去に他者が創作した「新聞」をみる．その意図は，自分がこれから行うことのゴールを知り，これなら「私にもできるかも」と思ってもらうことにある．全員が集まり着席したら，まずは個人で表現する．自分の仕事をふり返り，看護の仕事で「心に残る体験」を思い出し，一人一人が作文する（S02）．「あの出来事は，こんなふうに私の心に残っていたんだ」と書きながら気付いていく．次に，それを一人ずつ朗読する．書いた内容を説明するのではなく，書いた作文をただ読み上げる．他の人はそれを傾聴し味わう（S03）．各人が読み終わる毎に拍手する．全員が読み終わったら，次に，4 人 1 グループになり，協働で新聞づくりを行う．各自の作文を貼り合わせ，共通した思いや違いに着目し，対話しながら新聞のタイトルを決めていく（S04）．写真などを用いて装飾しながら，自分たちの表現をさらに味わう．出来上がったら，新聞を見せ合い，タイトルを発表する（S05）．最後に，この表現体験をふり返り，気づきや感想を語る（S06）．なお，S04 で創作された新聞は，別の WS で S01 の対象となる．

表現者として参加した看護師の一人は「気持ちを出す場がなかったので，皆で気持ちをわけることのできるいいワークショップだと思った」という感想を述べていた．また，今では主催者側と写



真 1. 新聞ワークショップの様子（2017）

なった看護 G の一人は次のように「新聞ワークショップ」に看護の継承という価値を付与している．「看護の心を表出すると，見える，読める，残せる．その仕組みとかたちをつくりたい．表現していくと残っていく．自分たちの当たり前を残す試み．」．現在も，活動プログラムと道具が共に細かな改良が続いている．

3.3 ふり返り：予想外の出来事

こうした共同デザインの成果にいきつくまでに，私たち研究 G はその場で生成される状況とどのように関わりながら活動プログラムと道具を改善・考案していったのか？私（たち）の認識とその変化に焦点を当てて，その状況を解釈していきたい．次に挙げるのは，この出来事がなかったら，今の共同デザインの状態はなかっただろうと直感する予想外であり驚きのあった特徴的な出来事である．

(1) 事例① S 氏の周辺参加と「Zuzie Poetry」

2013 年 6 月 27 日，私たちは看護 G と一緒にデザインした心象分析ツールと活動プログラム「Zuzie Poetry」を看護師 20 数名に向けて発表する機会をもった．「Zuzie Poetry」は，看護体験を複数人で対話しながらふり返り，可視化して共有することをねらいとしている[渡辺 2015]．発表会では，「活動プログラム」に則った未来のストーリー（以前看護師の一人が表現ワークショップで書いた「ある患者さんとの出来事」という作文が題材）を創作し，実際のツールを使う場面を演じてみせるアクティングアウトの方法を用いることとした．会議室に看護師が集まり，前方に研究 G と看護 G が「Zuzie Poetry」の入った PC と共に座った．その後，看護 G の内 2 名が，ストーリー（台詞）が描かれた紙を片手に，PC を操作しながら，「Zuzie Poetry」を使った対話のかたちを演じて見せた（写真 2）（図 1）．

発表直後，発表を見ていた看護師たちに混ざり座っていた S 氏（研究 G でも看護 G でも看護師たちでもない，周辺的に参加した人）が，私たちに向かって「今のは出来事を説明するものになっている．大事なのはそこでの気づきだと思う．」と発言した．また，「みなさんがやっておられるのは電子カルテで失われたものを取り戻すためでしょ」と発言．S 氏は研究 G に同行してきたにも関わらず，アクティングアウトで現れた活動が，本来実現したかった表現になっていないという問題点を的確に指摘した．

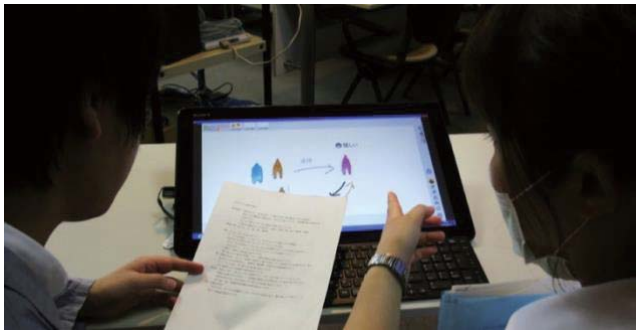


写真2.「Zuzie Poetry」を使ったアクティングアウト（2013）



図 1.「Zuzie Poetry」のストーリーボード（2013）

解散後，看護師の一人（T さん）が私に近寄り「思いを伝える機会をつくるのにすごい労力が必要なんです」と言った．

その後，研究 G はツール機能の改善点を抽出しつつもその改良を棚上げし，心象を外化するという活動プログラムを問い直し始めた．同年，9 月と 11 月に「Zuzie Poetry」を使った表現活動

を看護 G に参加してもらいやってみながら，道具と活動プログラムの妥当性を検討した．翌年 2014 年 3 月，今後の進め方を話し合うために病院を訪れ，「Zuzie Poetry」を使った実験ワークショップをふり返った．そこで，H 師長が「いいけど，これを使う時間はない」と発言した．

(2) 事例② Y 師長と「新聞ワークショップ」

2016 年 7 月，初めてとなる「新聞ワークショップ」を院内で行った．この時，表現のテーマとワークショップ中の進行は研究 G が行い，参加する看護師たちとグループ分けは看護 G が決めた．10 月にも参加者を変えて，同様の役割分担で実施．その頃，研究 G は先走って「新聞ワークショップ」が自立するため道具立て（キット）を作りたいと考え，その協働の仕方を模索していた．2017 年 3 月，顔を合わせて行ったミーティングの中で，看護 G から病院内の MED メンバーを増やし，こうした活動を草の根的に広げたいとの発言があった．そして，早速に Y 師長の部署を対象にした「新聞ワークショップ」を実施することが表明された．そのミーティングの終わりかけに Y 師長が席を立ちながら，「キットつくっているんだよね．よろしく．」と研究 G に言った．研究 G はキットをつくりながら，看護の「実践知」を共有するための活動プログラムと道具のデザインを新たに始めた．

2017 年 6 月，Y 師長の部署のスタッフを対象とした第 1 回目の「新聞ワークショップ」が実施され，Y 師長は研究 G がつくったキットを使い運営した．行為の詳細が描かれた「新聞ワークショップの流れ」を片手にもっての進行．そこに参加した研究 G（2 名）は写真等の記録を行った．終了後，看護 G よりキットの改善要望が多く挙がった．研究 G はそれを持ち帰り，第 2 回目となる 10 月に向けて改良を行った．10 月の終了後にもまた改善要望があがったため，研究 G はその次の回向けた改良を行うことになった．第 3 回目（2018 年 3 月），別の部署にはなったが，やはり司会進行はキットを使い Y 師長が行った．そして，そのワークショップの最後，「私はワークショップをやる人間として，いろんな部署に出向いていく」と，参加者全員に向かい発言し，協力を促した．

3.4 考察

前項に挙げた 2 つの事例は，いずれも 2 年程の期間の中で想定外の出来事が連続して起きている事態である．研究 G は，そのような予想外に生成される状況とどのように関わりながら，活動プログラムと道具を改善・考案していったのだろうか？私（たち）の認識とその変化について考察する．

(1) 関係性の変化によって起きた共同デザイン

この関わり合いの中で興味深いのは，偶然に居合わせた S 氏の周辺的な参加が，あの場に居合わせたそれぞれの立場の人の関係性に変化をもたらしたと解釈できる点にある．あの部屋に入ってきた時，S 氏，研究 G，看護 G，看護師たちは，上下に序列化された組織のようであったと考えられる．なぜなら，S 氏は研究 G にとって師のような存在であり，また看護師長で構成されている看護 G は看護師たちからすると上司だからである．また，研究 G と看護 G の関係は明確ではないが，看護 G は研究 G に対して，ものを言い難い状況にあったのではないかと推測する．ある種の誤解とも言えるその上下関係は，S 氏が看護師たちの立場に立って研究 G に意義を唱えたことをきっかけに変化した．研究 G が何か正解を知っている人で，看護師たちが知らない人ではない．そのような気づきが，思いを伝える労力について教えてくれた看護師の態度や，看護 G がこの直後に「先生たちも正解を知っているわけじゃないんですね」と発言したことにもみてとれる．プロジェクトの当初，「ゴールを教えて欲しい」と訴えていた看護 G は，それ以後はそういったことは全く発言しなくなっていた．

看護 G と研究 G の関係性は，その後のデザインの展開にも変化をもたらしている．あの時，私たちは，S 氏の本質をついた発言や，その発言に触発された看護師の反応に対して，耳と傾けるという関わり方をした．そして，それを受け入れ，起こしたかった活動の本質に立ち戻るという選択をした．心象を表現するのはいい．しかし，それを外化して共有した時に，自分たちが立ち向かおうとしていた記録業務と同じような営みをつくり出してしまっていることに全員が気づくことができた．その気づきは，道具のデザインを一度立ち止まらせ，活動プログラムの再検討という新たなデザインの展開の起因としてはたらいた．また，最後に示した H 看護師の発言も加わって，新たな活動プログラムというのは，研究 G が看護 G に協力をしてもらってデザインするというデザイン方法の認識は，看護 G と共に行わなければならないという認識へ変化した．

(2) 役割の移行と新たなデザインの始まり

この関わり合いの中で興味深いのは，新聞ワークショップの実施主体が研究 G から看護 G に完全に移行している起点に，2017 年 3 月のミーティングでの看護 G からの突然の表明があるという点である．おそらく，それまで研究 G の振る舞いをそばでみていた看護 G が，「その行為は自分でもできそう」と思ったこと，さらに院内に活動を広めていく道筋を見通したことによって，その状況がつくられたのだろう．当該ミーティングの内容を予め知ることはなかった研究 G は，そこに立ち現れた看護 G の表明に耳を傾け，受け入れていった．その時点で，研究 G は自分たちが何かを計画し，それに参加してもらうという研究 G 主体のかたちではなく，自分の意思にかかわらず「巻き込まれる」としか言いようのない状況の中にいる．

プロジェクトが始まって4年は，研究 G が表現ワークショップを運営しなければならないと考えていた．しかし，ここに来てはじめて，看護 G がやれるのだという気づきを得た．また，巻き込まれ，それまでの役割を手放したことは，私たちは今ここで何をすればよいのかを考え実践する契機となった．そして，そこにキットをデザインするという役割，そして，新聞ワークショップで起きた語りを継続的に起こしという新たなデザイン問題を見出している．私たちのデザインは，

新たな道具のデザインは，活動プログラムを共同でデザインするということは，現場の埋め込みと同時に行われることであるという認識を新たにもつことができた．

4. まとめ

共同デザインにおける想定外かつ驚きの出来事を中心に省察することから得た気づきは，その成り立ちを下記のように特徴づける．すなわち，1）周辺的参加の状況をつくりだし，関わる人たちの関係性の変化を起こす．2）目的を固定化せず，目的がないことにも耐えながら，発見的に前進する．3）巻き込まれることによって，自分の意思に関わらず，その状況の中にいることを選ぶ．このような行き当たりばったりに見えるデザインの思考と行為について，ティム・インゴルドは「応答」という概念で説明している．それは今生じていることにしたがって順々に反応すると

いうことを意味している[Ingold 2013]. また, ドナルド・ショーンは, そのプロセス全体を専門的実践家の専門性として「行為の中の省察」という概念で意味付けている[Schön 2007]. 私たちのデザインを成り立たせている上記の1）～3）の特徴は, 活動プログラムのデザインがいつも不確実で不安定な状況に置かれているためであり, 私たちはそれをあえて選び, 状況と対話することをデザインの動力にしていると言える. 今後, こうした記述が次なるデザインの道標になるよう, まとめていきたい.

＜謝辞＞
本プロジェクトに参加する佐賀大学医学部附属病院看護部の皆様に感謝申し上げます. また, 事例に登場する成果物のデザインには多摩美術大学の学生が参加しています. なお, 本研究の一部は JSPS 科研費 15K16174 の助成を受けたものです.

参考文献

[小早川, 2009]小早川真衣子, 敦賀雄大, 高見知里, 永井由美子, 須永剛司, ミュージアム学習のための表現活動プログラム設計の枠組み, デザイン学研究 2008 度作品集, 14 号, pp.50‐55, 日本デザイン学会, 2009 年 3 月

[藤満 2013] 藤満幸子, 山口真由美, 原田由美子, 椛島久美子, 宮之下さとみ, 南里美貴, 百武朋美, 山田クリス孝介, 須永剛司, 小早川真衣子, 新野佑樹, 渡辺健太郎, 西村拓一: 医美工連携による看護情報システムの開発を目指したデザイン・プロジェクト」, 第 33 回医療情報学連合会（第 14 回日本医療情報学会学術大会, 2013

[宮之下 2017]宮之下 さとみ, 藤満 幸子, 原田 由美子, 山口 真由美, 椛島 久美子, 山田 クリス孝介, 須永 剛司, 小早川 真衣子, 丸山 素直, 看護体験を表現する作文・新聞ワークショップの検証, 日本デザイン学会 第 64 回春季研究発表大会, 日本デザイン学会, 2017 年

[渡辺 2015] 渡辺 健太郎, 藤満 幸子, 原田 由美子, 山田 クリス孝介, 須永 剛司, 小早川 真衣子, 新野 佑樹, 阪本 雄一郎, 西村 拓一, 本村 陽一: 看護現場における業務経験の表現・共有支援システムの開発, 情報処理学会論文誌, 56(1), 137-147, 2015

[須永 2013] 須永 剛司, 小早川 真衣子, 山田 クリス孝介, 渡辺 健太郎, 新野 佑樹, 西村 拓一:Co-design プロジェクトが自発的に回ること：社会を形づくるデザインに向けて, 人工知能学会誌, 28(6) 886-892 , 2013

[Ingold 2013]ティム・インゴルド, 金子遊訳: メイキング 人類学・考古学・芸術・建築, 2017, 左右社（Tim Ingold : Making: Anthropology, Archaeology, Art and Architecture, Routledge, 2013）

[Schön 2007] Donald A. Schön 著, 佐藤学他訳, 2007,「省察的実践とは何か―プロフェッショナルの行為と思考」, 鳳書房

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

医師の知見と経験を蓄積・伝承する判断支援システム実現に向けた pLSA による診療クラスタの時間遷移パターンの解析

Treatment strategy of Septicemia based on DPC data analysis by using pLSA method

山下和也[*1] 阪本雄一郎[*2] 櫻井瑛一[*1*2] 本村陽一[*1*2]
Kazuya Yamashita Yuichiro Sakamoto Eiichi Sakurai Yoichi Motomura

[*1]産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

[*2]佐賀大学
Saga University

This paper shows our study about treatment strategy of septicemia based on DPC data by using pLSA, probabilistic latent semantic analysis, which find hidden segments in data. Our aim of using this method is modeling patients' change along with time. In this study, we estimates segments from the data that consists of times when each patients had some medical treatment and name of treatments. We compared each segments from the points of death rate, medical cost and patients' hospitalization. We analyze patients' movement pattern between segments. We also pay attention to that these movements is related to medical decision.

１．はじめに

医療の質を向上させながら増え続ける医療費の圧縮が求められているなか、DPC(Diagnosis Procedure Combination=診断群包括分類)データの蓄積が進み、これら医療ビッグデータの重要性が益々認識されその利活用に大きな期待が寄せられている。DPC データを用いた集中治療関連の報告は散見されるが、治療法の効果などを検証した結果が多く、治療戦略への応用などの報告は認められないのが現状である。本研究では患者にとって重要な「１．予後（特に死亡率）」、医療制度を支える国にとって重要な「２．医療費」、安定的に患者を受け入れる為に病院にとって大事になる「３．在院日数」の３つの変数に注目し、他の変数２つが同じならもう１つの変数はより「生存率が高く」、「安価で」、「短い在院日数」となるような治療戦略を持つことを目的とした上で、多彩な治療薬が存在する敗血症に対して DPC データで得られた治療薬を pLSA(Probabilistic Latent Semantic Analysis)手法[Hofmann,1999]によりクラスタ分類するとともに同一患者におけるクラスタの時間遷移を解析する事で、DPC 入力データが敗血症治療ナビゲーション（医療判断支援システム）として治療戦略に有益である可能性を検証している。

pLSA を用いた解析の実社会への応用としては、[井手,2016]が高齢者の大規模アンケートを用いた高齢者生活行動モデルを構築し、[井手,2017], [Ide,2017]では個人の代わりに空間（地域）についての潜在クラスを抽出する確率的潜在空間意味解析を用いて地域の自治体等の政策決定の提言に活用出来る事を示している。またスーパーの購買履歴である ID-POS データを用いた顧客理解・来店人数予測[山下, 2017a] ,[石垣,2011]、需要予測[本村,2011]などマーケティング分野への例がある。更に顧客 ID と月を組み合わせた変数により購買行動について時間軸上での変化のモデリングを行った研究[原田, 2016]がある。同様の手法を医療分野に適応した初の例として[山下, 2017b]は敗血病患者と注射薬剤を同時に分類し、これまでの年齢や重症度による分類とは異なる、診療行為（この場合は注射薬剤）によって情報量最大となる患者分類を行っている。[山下, 2017b]では「入院初日から 3 日まで」、「4～30 日まで」、「31 日目以上」と 3 期間における患者の時間変化も考慮した解析であり、[山下,2018]では上記のように固定した期間に離散化する代わりに 1 日毎の時間変化を解析し、患者の変化の変わり目をデータから検出出来るようにしている（動的離散化）。また診療行為についても注射だけでなく、投薬、処置、手術等も含めた診療行為としている。[山下,2018]では同一患者の時間変化に伴うクラスタ遷移パターンについて幾つかの代表例が示されたが、本研究では全体像の解析を行った。

２．時間変化も考慮した pLSA によるモデリング手法

連絡先:山下和也, 産業技術総合研究所・人工知能研究センター, 〒135-0064 東京都江東区青海 2-4-7, yamashita-kazuya@aist.go.jp

pLSA は、2 つの変数 x,y の背後に共通の特性、潜在変数 z ∈ Z = {z1,...,zk} があると仮定し、x,y が各 z に関係する確率をデータから期待尤度最大化法によって推定し、その確率値が最大となる潜在変数 z に x と y を分類する。本研究では、DPC データを用い上記の x,y のうち x を患者 ID と診療行為実施日（入院初日から 1 日毎）を組み合わせた変数“患者 ID_実施日”とした。y を注射、投薬、処置、手術等の各診療行為名称（注射や投薬の場合は各薬剤名、処置の場合は各処置名等）とした。上記の変数 x,y を入力として産業技術総合研究所の知財ソフトウェア「PLASMA」を用いて pLSA を行い、同じような診療行為を受けている患者及び実施日と、同時に行われやすい診療行為を同時に分類した。使用した患者 ID 数は 458(患者 ID_実施日の総数は 19658)、診療行為数は 4406 である。得られた潜在変数 z を診療クラスタと呼ぶこととし、クラスタ数 9 が AIC 最小となった。このクラスタ毎に、各クラスタに所属する患者 ID について死亡率（死亡の有無の平均）を集計して死亡率の低い順に aC1,aC2,..,aC9 とクラスタに名前を付けた。各クラスタの詳細は[山下,2018]に示されている。本研究では前日から当日へのクラスタ遷移の事例数や遷移時期を、全体、生存・死亡の各ケース、在院日数の長短別、入院初期かそれ以降かによる別に分けて集計を行う事でクラスタ遷移パターンを調べた。

３．同一患者の時間変化に伴うクラスタ遷移パターン結果

３．１．前日から当日へのクラスタ遷移数

２章で示した方法で pLSA を行い死亡率の順に番号を付けた診療クラスタ aC1,aC2,..,aC9 について、前日から当日へのクラスタ遷移の事例数を図１に示す。全体の傾向として遷移なし（同一クラスタへの遷移）が多い事が分かる。クラスタ aC3（以後単にクラスタ 3 と示す）に留まっている事例が最も多い事が分かる。また入院初日はクラスタ 6 で以後はクラスタ 6 への遷移は少なく、クラスタ 6 は入院最初期の特徴を持っている事が分かった。

図２に、結果的に(左)生存したケース、(右)死亡したケースそれぞれに分けたクラスタ遷移パターンを示す。生存した患者は特にクラスタ 3 に留まっていることが分かる。全体として死亡率が低いクラスタを遷移している。クラスタ 4 とクラスタ 2 の間の両方向の遷移が多い特徴が見てとれる。一方死亡ケースでは、死亡率の高いクラスタに留まっている特徴が見られる。クラスタ 8 にある日数が最も多い。ここまでをまとめると、全体として死亡率の高いクラスタ間の遷移は少なく、それに対して、死亡率の低い 4 や 2 等のクラスタ間では遷移が多いことが分かる。

図３では生存した患者について、在院日数が短かった患者（在院 28 日以下）と長かった患者（在院 29 日以上）に分けてそれぞれの遷移を示した。なお 28 日は在院日数の中央値である。長かった患者については、入院初期（入院初日から 28 日以前）とそれ以降（入院初日から 29 日以降）に分けて示している。在院日数が短い患者はクラスタ 2 の日が多い。またクラスタ 2 と 4 の双方向の遷移も見られる。一方在院日数が長い患者の入院初期でも上記と同様の傾向が見られるがそれに加えてクラスタ 3 状態の日も多く、クラスタ 4 と 3 の遷移も見られる。またクラスタ 8 の日も比較的多い。入院 29 日目以降ではクラスタ 3 に留まる事がより多くなっている。以上まとめると生存した患者は入院初期にクラスタ 2 と 4 を遷移し、そのまま早期に退院する患者と、その後、クラスタ 3 に移行し滞留した後に退院する患者がいる。

図４では死亡した患者について、図３と同様の図を示した。早期に死亡する患者は死亡までクラスタ 8 の状態が多い。長期入院後死亡する場合、入院初期にはクラスタ 3 や 5 の状態が続いたり他のクラスタの合間にクラスタ 4 にいる状態を経過した後、クラスタ 7 か 9 に移行して亡くなる様子が見られる。

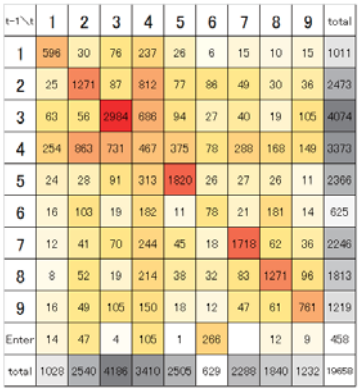
３．２．クラスタ遷移日

図５に各クラスタ遷移した日（入院初日からの日数）の平均を示した。生存・死亡別の図をみると、死亡では入院後期に死亡率の高いクラスタにあることが分かる。

図６は前日からクラスタ遷移したか否かを示した図である。遷移した割合を示していて、全体として入院後期程、遷移はせず同じクラスタに留まる傾向が分かる。長期入院してかつ死亡したケースでは入院初期からクラスタ遷移が比較的少ない事が分かる。

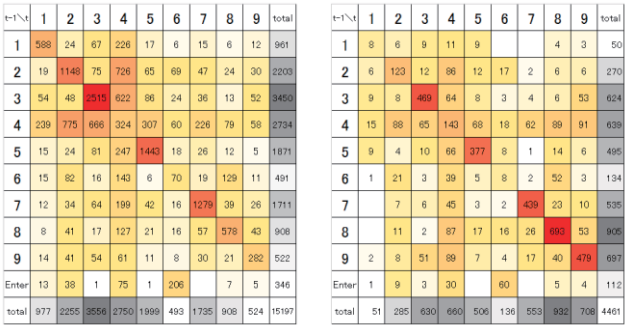
４．まとめ

診療クラスタの時間遷移には予後や在院日数毎にそれぞれのパターンがある事が分かった。これらクラスタ遷移は医師の何らかの医学的判断と結びついていて、その判断により診療行為を変えた事を意味する。また同じクラスタが継続されているのも医学的判断の結果である。医学的判断には「その際の患者の状態における標準的な診療を行っている」や「集学的治療を行った」、「ある種の挑戦的な診療を行った」、または「with draw」や「with hold」などがあり、今後これら医学的判断とクラスタ遷移、また検査値を紐づけ、さらにそれらの結果として、目的変数としての予後、在院日数、医療費の症例データを蓄積する事で、医療現場の医師の判断も含めた現象を、ベイジアンネットワーク等により計算空間にモデル化する事が出来る。そのモデルは診療判断システムとして実際の集中治療室で活用する事が出来、さらにその活用結果は新たな事例の蓄積として医療現場にフィードバックが可能である。このように医療現場と計算空間でのループを回していく事で、経験を積んだ医師の知見の蓄積と伝承が可能となり、全体の医師の知見とモデルの両方が共に成長していく事が可能となる。本研究ではそのようなシステム実現に向けて診療クラスタ遷移の全体像を把握した。



t-1\t	1	2	3	4	5	6	7	8	9	total
1	596	30	76	237	26	6	15	10	15	1011
2	25	1271	87	812	77	86	49	30	36	2473
3	63	56	2984	686	94	27	40	19	105	4074
4	254	863	731	467	375	78	288	168	149	3373
5	24	28	91	313	1820	26	27	26	11	2366
6	16	103	19	182	11	78	21	181	14	625
7	12	41	70	244	45	18	1718	62	36	2246
8	8	52	19	214	38	32	83	1271	96	1813
9	16	49	105	150	18	12	47	61	761	1219
Enter	14	47	4	105	1	266		12	9	458
total	1029	2540	4186	3410	2505	629	2288	1840	1232	19659

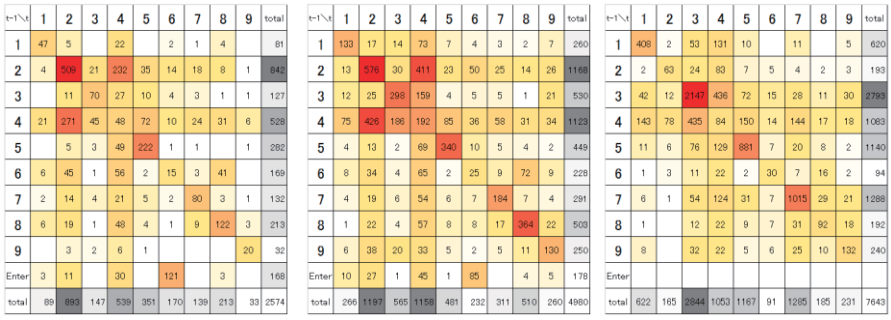
図１．クラスタ遷移数（全ケース）
縦軸が前日のクラスタ番号（Enter は入院）
横軸が当日のクラスタ番号



t-1\t	1	2	3	4	5	6	7	8	9	total
1	588	24	67	226	17	6	15	6	12	961
2	19	1148	75	726	65	69	47	24	30	2203
3	54	48	2515	622	86	24	36	13	52	3450
4	239	775	666	324	307	60	226	79	58	2734
5	15	24	81	247	1443	18	26	12	5	1871
6	15	82	16	143	6	70	19	129	11	491
7	12	34	64	199	42	16	1279	39	26	1711
8	8	41	17	127	21	16	57	578	43	908
9	14	41	54	61	11	8	30	21	282	522
Enter	13	38	1	75	1	206		7	5	346
total	977	2255	3556	2750	1999	493	1735	908	524	15197

t-1\t	1	2	3	4	5	6	7	8	9	total
1	8	6	9	11	9			4	3	50
2	6	123	12	86	12	17	2	6	6	270
3	9	8	469	64	8	3	4	6	53	624
4	15	88	65	143	68	18	62	89	91	639
5	9	4	10	66	377	8	1	14	6	495
6	1	21	3	39	5	8	2	52	3	134
7		7	6	45	3	2	439	23	10	535
8		11	2	87	17	16	26	693	53	905
9	2	8	51	89	7	4	17	40	479	697
Enter	1	9	3	30		60		5	4	112
total	51	285	630	660	506	136	553	932	708	4461

図２．クラスタ遷移数（左：生存、右：死亡）

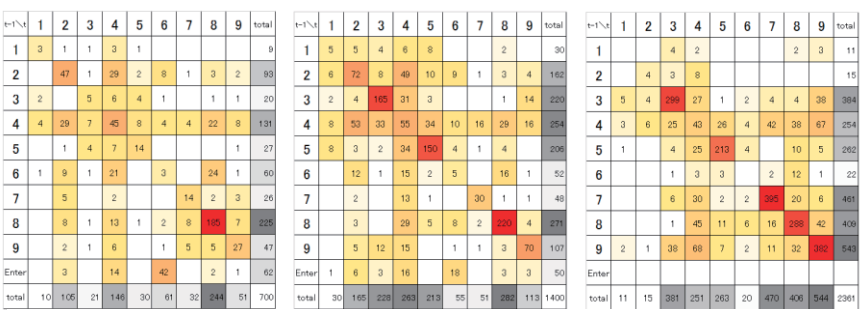


t-1\t	1	2	3	4	5	6	7	8	9	total
1	47	5		22		2	1	4		81
2	4	509	21	232	35	14	18	8	1	842
3		11	70	27	10	4	3	1	1	127
4	21	271	45	48	72	10	24	31	6	528
5		5	3	49	222	1	1		1	282
6	6	45	1	56	2	15	3	41		169
7	2	14	4	21	5	2	80	3	1	132
8	6	19	1	48	4	1	9	122	3	213
9		3	2	6	1				20	32
Enter	3	11		30		121		3		168
total	89	893	147	539	351	170	139	213	33	2574

t-1\t	1	2	3	4	5	6	7	8	9	total
1	133	17	14	73	7	4	3	2	7	260
2	13	576	30	411	23	50	25	14	26	1168
3	12	25	298	159	4	5	5	1	21	530
4	75	426	186	192	85	36	58	31	34	1129
5	4	13	2	69	340	10	5	4	2	449
6	8	34	4	65	2	25	9	72	9	228
7	4	19	6	54	6	7	184	7	4	291
8	1	22	4	57	8	8	17	364	22	503
9	6	38	20	33	5	2	5	11	130	250
Enter	10	27	1	45	1	85		4	5	178
total	266	1197	565	1158	481	232	311	510	260	4980

t-1\t	1	2	3	4	5	6	7	8	9	total
1	408	2	53	131	10		11		5	620
2	2	63	24	83	7	5	4	2	3	193
3	42	12	2147	436	72	15	28	11	30	2793
4	143	78	435	84	150	14	144	17	18	1083
5	11	6	76	129	881	7	20	8	2	1140
6	1	3	11	22	2	30	7	16	2	94
7	6	1	54	124	31	7	1015	29	21	1288
8	1		12	22	9	7	31	92	18	192
9	8		32	22	5	6	25	10	132	240
Enter										
total	622	165	2844	1053	1167	91	1285	185	231	7643

図３．生存したケースのクラスタ遷移数（左：在院短期、中：在院長期の入院初期、右：在院長期の入院 29 日目以降）



t-1\t	1	2	3	4	5	6	7	8	9	total
1	3	1	1	3	1					9
2		47	1	29	2	8	1	3	2	93
3	2		5	6	4	1		1	1	20
4	4	29	7	45	8	4	4	22	8	131
5		1	4	7	14				1	27
6	1	9	1	21		3		24	1	60
7		5		2			14	2	3	26
8		8	1	13	1	2	8	185	7	225
9		2	1	6		1	5	5	27	47
Enter		3		14		42		2	1	62
total	10	105	21	146	30	61	32	244	51	700

t-1\t	1	2	3	4	5	6	7	8	9	total
1	5	5	4	6	8			2		30
2	6	72	8	49	10	9	1	3	4	162
3	2	4	165	31	3			1	14	220
4	8	53	33	55	34	10	16	29	16	254
5	8	3	2	34	150	4	1	4		206
6		12	1	15	2	5		16	1	52
7		2		13	1		30	1	1	48
8		3		29	5	8	2	220	4	271
9		5	12	15		1	1	3	70	107
Enter	1	6	3	16		18		3	3	50
total	30	165	228	263	213	55	51	282	113	1400

t-1\t	1	2	3	4	5	6	7	8	9	total
1			4	2				2	3	11
2		4	3	8						15
3	5	4	298	27	1	2	4	4	38	384
4	3	6	25	43	26	4	42	38	67	254
5	1		4	25	213	4		10	5	262
6			1	3	3		2	12	1	22
7			6	30	2	2	395	20	6	461
8			1	45	11	6	16	288	42	409
9	2	1	38	68	7	2	11	32	382	543
Enter										
total	11	15	381	251	263	20	470	406	544	2361

図４．死亡したケースのクラスタ遷移数（左：在院短期、中：在院長期の入院初期、右：在院長期の入院 29 日目以降）



t-1\t	1	2	3	4	5	6	7	8	9	average
1	72	13	58	62	32	13	35	17	45	64
2	15	13	23	16	14	10	19	12	16	15
3	65	25	60	47	63	47	56	87	58	58
4	60	15	46	29	38	23	48	33	63	36
5	35	23	64	40	52	28	50	48	41	50
6	5	7	34	14	28	23	34	15	14	16
7	34	17	52	51	43	37	69	57	56	64
8	10	8	88	30	41	42	44	26	50	29
9	41	13	59	67	44	39	51	58	72	65
Enter	1	1	1	1	1	1		1	1	1
average	63	14	57	35	48	15	63	28	64	44

t-1\t	1	2	3	4	5	6	7	8	9	average
1	72.8	14	61.3	64.2	42.4	12.7	35.3	5.67	29.8	66
2	17.7	13.3	23	16.5	15.2	10.2	19.4	13.1	16.6	14.9
3	68.8	24.4	64	48.9	66.9	45.8	56.6	111	43.6	60.5
4	62.4	15.2	47.7	26.7	39.3	22.6	44.6	23.7	28.4	34.4
5	47	24.8	68.6	42.2	55.6	24.1	51.2	31.3	18.8	53.3
6	5.6	7.51	38.6	14.5	30.2	25.2	31.4	12.4	13.7	15.7
7	34.2	17.5	52	47.9	42.8	34.7	59.6	56.5	39	55.9
8	10.3	7.71	85	18.6	27.4	47.3	44.8	19.3	26.1	22.6
9	33.6	12.8	43.2	30.4	27.7	39.5	39.3	29.5	33.7	32.8
Enter	1	1	1	1	1	1		1	1	1
average	65	13.9	59.6	33.4	51.3	14.7	55	21.6	31.8	42.14

t-1\t	1	2	3	4	5	6	7	8	9	average
1	9.75	6.67	32.8	12.9	12.6			34	108	22.5
2	5.33	10.8	21.2	13.1	8	9.47	6.5	8.83	14	11.7
3	42.9	29.6	41.5	29	21.9	52.7	54	34.5	71.8	42.5
4	16.8	11.9	26	35.6	30.7	23	60.9	41.8	84.4	40.3
5	15.6	11.3	25.7	32.8	37.3	35.9	23	62.7	59.2	36.8
6	2	5.24	12	11.3	25.8	6.75	59.5	20.6	16	15
7		15.3	57.5	63.5	48.7	58.5	97.5	56.8	98.8	90.9
8		10.9	31.5	46.1	59.8	36.1	41.2	31.7	70.1	35.9
9	94	16.1	76.5	92.2	69.6	39	71.4	72.3	94.2	89.4
Enter	1	1	1	1		1		1	1	1
average	21.2	11.1	41.8	39.3	36.2	14.6	89	34.6	87.7	49.4

図５．クラスタ遷移した日（入院初日からの日数）の平均（上：全体、下左：生存、下右：死亡）

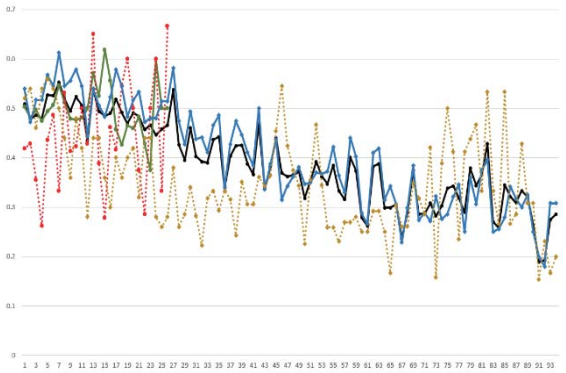


図６．クラスタ遷移の有無　横軸は入院初日からの日数、縦軸は遷移率の平均（29 日以降も伸びている線の実線＆丸点：全体、実線＆♦点：在院長期かつ生存、点線＆♦点：在院長期かつ死亡、28 日までしかない線の実線＆丸点：在院短期かつ生存、点線＆丸点：在院短期かつ死亡）

謝辞

本研究は NEDO 委託事業「人間と相互理解できる次世代人工知能技術の研究開発」の支援を受けて行いました。

参考文献

[Hofmann,1999] Hofmann, Thomas. ”Probabilistic latent semantic analysis.” Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., (1999)

[Ide,2017]Ayae Ide,Kazuya Yamashita,Yoichi Motomura and Takao Terano,”Analyzing Regional Characteristics of LivingActivities of Elderly People from Large Survey Data with Probabilistic Latent Spatial Semantic Structure Modeling”, IEEE BigData2017 Workshop,(2017)

[石垣,2010]石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, 56, 2, (2010)

[石垣,2011] 石垣司, 竹中毅, 本村陽一: 潜在クラスモデルによる流通量販店舗の来店人数予測の精度改善の評価, 人工知能学会全国大会, (2011)

[井手,2016] 井手絢絵, 本村陽一:高齢者の生活行動推論のため の大規模アンケートデータを用いた確率的潜在意味構造 モデル構築, 人工知能学会「社会における AI」研究会第 27 回研究会,(2016)

[井手,2017] 井手 絢絵,川本 達郎,山下 和也,本村 陽一:ミクロアグリゲーションを用いた匿名化による確率的潜在空間意味解析, 人工知能学会全国大会,(2017)

[近藤,2017] 近藤那央,竹内理人,山下和也,櫻井瑛一,本村陽一, 大規模イベントにおける人の行動履歴情報収集とそれを利用したインタラクティブシステムの開発に向けて,人工知能学会全国大会,(2017)

[阪本,2018] 阪本雄一郎,山下和也,櫻井瑛一,本村陽一:統計技術による DPC データの敗血症治療戦略への活用, 第 45 回日本集中治療医学会学術集会,(2018)

[原田, 2016] 原田奈弥, 山下和也, 本村陽一: ID 付 POS データによる購買行動の季節変化の分析と視覚化, 人工知能学会社会における AI 研究会 27 回研究会,(2016)

[本村,2011] 本村陽一, 竹中毅, 石垣司: 条件付層別差分モデルによる需要予測の高精度化, 人工知能学会全国大会, (2011)

[山下, 2017a] 山下和也,原田奈弥,黄冬陽,吉開朋弘,本村陽一: ID-POS データによる来店行動・購買行動の潜在的時空間意味構造分析, 人工知能学会全国大会,(2017)

[山下, 2017b] 山下和也,阪本雄一郎,櫻井瑛一,本村陽一: 時間変化も考慮した pLSA 手法による敗血症治療戦略への応用, 人工知能学会 社会における AI 研究会 27 回研究会,(2017)

[山下,2018] 山下和也,阪本雄一郎,櫻井瑛一,本村陽一:診療クラスタ遷移と医学的判断の紐づけによる敗血症治療ナビゲーションシステム実現可能性の検証,第 5 回「JAMI 医用知能情報学研究会-JSAI 医用人工知能研究会」合同研究会,(2018)

[2K4-NFC-3b] [NFC Session] NFC-3 PADSED (Product and Activity Design Support based on Event Database)

座長:西村 拓一（産業技術総合研究所）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room K (3F Azisai Mokuren)

[2K4-NFC-3b-01] Analysis and utilization of the action history of the event participants

○Tomoyuki Owada[1], Kazuya Yamashita[1], Tomotaka Omae[1], Yoichi Motomura[1] （1. National Institute of Advanced Inductrial Science and Technology）

5:20 PM - 5:40 PM

[2K4-NFC-3b-02] Reflecting COTO Database Research: toward Implementation in the Society

○Kentaro Watanabe[1], Hiroyasu Miwa[1], Satoshi Nishimura[1], Ken Fukuda[1], Takuichi Nishimura[1] （1. National Institute of Advanced Industrial Science and Technology）

5:40 PM - 6:00 PM

[2K4-NFC-3b-03] Product and Activity Design Support based on Event Database

○Takuichi Nishimura[1], Kentarou Watanabe[1], Ken Fukuda[1], Satoshi Nishimura[1], Yoichi Motomura[1] （1. Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)）

6:00 PM - 6:20 PM

[2K4-NFC-3b-a] (Panel discussion) Data and Knowledge Construction, Sharing and Utilization in Real World

Satoshi Nishimura[1], Takuichi Nishimura[1], Ken Fukuda[1], Kentaro Watanabe[1], Kristina Jokinen[1], Yasuyuki Yoshida[1], Nami Iino[1], Lihua Zhao[1] （1. National Institute of Advanced Industrial Science and Technology）

6:20 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

イベント参加者の属性情報および行動履歴データの分析と活用

Analysis and utilization of the action history of the event participants

大和田智之[*1] 山下和也[*1] 大前智嵩[*1] 本村陽一[*1]
Tomoyuki Owada Kazuya Yamashita Tomotaka Omae Yoichi Motomura

[*1] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

There's more application in the field of data collection and data analysis recently. POS data for marketing, search or access log in web, image data with GPS to track human action history, are the remarkable examples. In this research, we collected the action history of the event participants by using ID card. We report the result of clustering using pLSA , the investigation for the causal relationship between participants' movement and booths arrangement, and also consider more application in the future of this system.

1. はじめに

実社会におけるビッグデータの収集，例えば購買行動に関する人の ID と商品 ID，場所(店舗)や時刻などのデータ(ID-POS データ)収集と，これを用いたデータ分析はサービスの改善への応用事例も多い[関 03，石垣 11]. こうしたマーケティングの分野における商品の購買履歴(POS データ)だけでなく，web における検索履歴やアクセス履歴の活用，さらに実世界の位置情報を付与した GPS データ付き画像データなどを用いた人の行動分析など，その対象はさらに広がってきている.

本研究では新たに大規模なイベントにおいて，来場者全員に ID 付きカードを付与することで来場者の属性(年齢や性別など)を登録し，さらに来場者が訪問した展示ブースでカードを端末にタッチすることで得られる回遊データの分析を行った結果を報告する. また，このイベント支援システムの今後やさらなる活用方法についても検討する.

2. データ収集

2.1 収集方法

人の行動履歴を収集する場合，まず検討しなければならないのがデータの収集方法である. 方法として，GPS を利用する方法，カメラによる人認識を利用する方法，ID カード等の IC タグを利用する方法などが提案されている.

GPS を利用した場合，同じ建物内などの細かい移動には十分な精度を発揮しづらいという問題がある. カメラを利用する場合，プライバシーの問題があり，また，技術的にも設備的にも高コストである.

本実験ではシール型の IC タグをイベント来場者に配布し，各ブースでタッチしてもらうことによりデータ収集を行った. これにより，行動履歴の収集と同時に各ブースの満足度を来場者に入力してもらうことが可能になった.

2.2 イベント概要

実験は 2017 年 11 月に行われたサイエンスアゴラ 2017 というイベントで行った. サイエンスアゴラは 2006 年から毎年行われている科学と社会をつなぐ日本最大級のオープンフォーラムで，例年の出展団体数は 100~200 団体，参加者数は 4000 人~1 万人に及ぶ. サイエンスアゴラ 2017 では、ブースは 1 階，3～5 階の 4 つのフロアで実施され，フロアごとでそれぞれテーマが分かれていた. セッションは 8 階と 20 階で実施された. このイベントにおいてデータ収集を兼ねたタッチラリー企画を行い，参加者に各ブースやセッションを回ってもらい，参加賞として飲料を配布した. 3 日間の開催のうち 1 日分，分析に利用可能な約 1000 人の来場者データについて分析を行った.

2.3 タッチラリーの流れ

イベント来場者はまず1階の受付でシール型 IC タグを貼付したカードを受け取り，受付横にあるサイネージ(図 1)で年齢や性別などの簡単な属性情報を入力する. その後来場者は，各ブースを見学した後，設置された端末に IC タグをタッチし，満足度を入力する. 各ブースに置かれた端末の様子を図 2 に示す.

最後に，5 つ以上のブースを見学した参加者は，出口に設置されたサイネージ一体型の自販機(図 3)にICタグをタッチすることにより，ジュースを受け取ることができる. また同時に，参加者がイベント中に回ったブースが画面上に経路として表示される.

また企画提供者は，イベント開催中からイベント開催後に出展したブースの年齢別，男女別の来場者数や，平均評価を見ることができる. これにより，各々の提供者が企画の質の向上に役立てることができる.



図 1: 受付に設置されたサイネージ端末

連絡先：大和田智之，産業技術総合研究所人工知能研究センター，〒135-0064 東京都江東区青海 2-4-7，t.owada@aist.go.jp

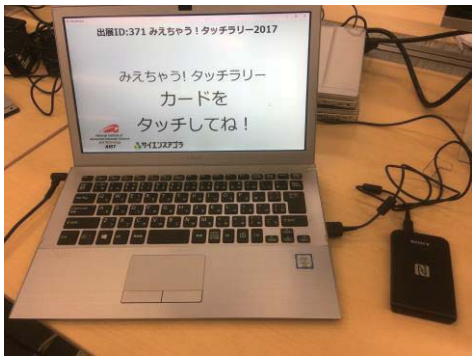


図 2: 各ブースに設置された端末



図 3: サイネージ一体型の自販機

3. 分析手法

3.1 pLSA を用いたクラスタリング

pLSA[Hofmann 99, Hofmann 01]は, 2 つの変数 x,y の間に共通の潜在変数 $z \in Z=\{z_1,z_2,\ldots z_k\}$があると仮定し, x,y に関係する確率が最も高いような z を選ぶことにより x,y を同時にクラスタリングする手法である.

pLSA は, 元々は自然言語処理などの分野で活用されてきた手法であるが, 最近では本研究に関連した分野も含む多くの分野で応用されている. pLSA を用いた POS データの活用事例として, 石垣らは, 顧客と商品の同時クラスタリングを行っている[石垣 10].

前章で示した方法により得られたデータを用い, 来場者を x, 各ブースを y として pLSA によるクラスタリングを行った. 使用したデータにおける来場者の総数は 976 人, ブースの総数は 51 個, 来場者によるブース訪問数の合計は 6440 件であった.

3.2 ベイジアンネットワークの構築

ベイジアンネットワークは確率変数をノードとし, 変数同士の依存関係を確率的なネットワークとしてモデル化したものである. ベイジアンネットワークによって構築したモデルはグラフとして可視化することが容易であり, 変数の意味を視覚的にもとらえやすいという利点がある. また, ある変数 x とその変数に影響を及ぼす変数 $z_1, z_2,\ldots z_k$ の関係は条件付確率 $P(x| z_1, z_2,\ldots z_k)$として得られるため, 因果関係の分析が容易である.

4.2 節では, 参加者の属性情報, 行動履歴(各ブースの滞在時間, 評価等), および pLSA による同時クラスタリングの結果を利用したベイジアンネットワークによるモデルの構築結果とその考察を示す.

4. 結果と考察

4.1 ブースのクラスタリング結果と階層の関係

pLSA を用いて来場者とブースを4クラスタに分けた場合のクラスタリング結果を, ブースごとにまとめたものを表1に示す. 表の各列はそれぞれのクラスタに属する確率を示し, 各行が各ブースに対応する. 表から, 各クラスタが各ブースの位置するフロアとかなりの精度で対応していることが分かる. 来場者は同じ階のブースを優先して回る傾向があり, 逆に言えばブースの位置関係は来場者の行動傾向に極めて大きい影響を与えているといえる.

また, 2 クラスタに分けた場合では, 1 階と 3 階のブースがクラスタ 1, 4 階と 5 階のブースがクラスタ 2 に属する傾向があり, 3 クラスタに分けた場合では 4 階と 5 階のブースがさらに 2 つに分かれることが分かった. いずれも表 1 に示した結果と同様, クラスタと位置するフロアは高い相関を示した.

4.2 参加者の属性情報と行動履歴の関係

参加者の属性情報, 行動履歴, および前節の結果から得られたクラスタをベイジアンネットワークによりモデル化し, グラフを可視化した結果を図 4 に示す. ここで, 前節の考察に基づき, クラスタ名には各フロアを回っている確率が高いという意味で, それぞれ floor1like, floor3like, floor4like, floor5like と名付けた.
ここで Discharge は出口で飲み物を受け取ったかどうか, また受け取ったならばその時間帯, TotalResidence はイベント滞在時間, AveResidence はブース一つあたりの平均滞在時間, AveEvaluation はブース一つあたりの平均評価を表す.

得られたモデルから, floor1like に属する参加者はブース訪問数が少なくブース一つあたりの滞在時間が長い傾向にあることが分かった. このため, 1 階を回って満足したもしくは疲れてほかの階を回らなかったことが考えられる. また floor5like に属する参加者は比較的早い時間からきて長く滞在している傾向にあった.



図 4: 属性情報, 行動履歴から構築したベイジアンネットワーク

表 1: ブースの位置するフロアとクラスタの関係

floor	Z1	Z2	Z3	Z4
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	1.00	0.00	0.00	0.00
3	0.30	0.54	0.00	0.16
3	0.00	0.81	0.19	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
4	0.18	0.00	0.82	0.00
4	0.06	0.00	0.94	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.19	0.00	0.81	0.00
4	0.00	0.00	1.00	0.00
4	0.15	0.00	0.85	0.00
4	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.26	0.74
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.29	0.71
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
8	1.00	0.00	0.00	0.00
20	0.00	0.00	0.00	1.00

4.3 ブースの位置と参加者の興味に関する考察

ブースのクラスタリング結果およびユーザーの行動を分析した結果はいずれもブースの位置が参加者の行動に強い影響を与えていることを示している. 2.2 節で記したように分析対象としたイベントではフロアごとのテーマ分けがなされており, ブースの位置が及ぼす影響とテーマ分けが及ぼした影響を完全に切り分けることはできない. しかしながら, ベイジアンネットワークの結果から, 近いフロアから順に回っていき, 時間の都合やイベントへの満足度から遠くのブースまで回るかを考える参加者が多いと考えられる.

今後のブース配置を検討するにあたっては, 時間がない人でも楽しめるように低層階に様々なジャンルのコンテンツをちりばめる一方で, 各階にも目玉となるようなブースを用意することで参加者に各フロアを回ってもらうといったことが考えられる.

5. まとめ

本稿で示したデータ収集のシステム, データ解析のプロセスはイベントの価値を高めるにあたって一定の成果を出せたといえる. 今後は, 参加者の回遊情報からリアルタイムにブースを提案するシステムや, AI を用いたインタラクティブなイベント支援など, さらなる応用が期待される.

謝辞

本研究は NEDO 委託事業「人間と相互理解できる 次世代人工知能技術の研究開発」の支援による. 国立研究開発法人科学技術振興機構の主催するサイエンスアゴラ 2017 において, 公式イベントとして実験を実施した. サイネージ一体型自動販売機により来場者に提供した飲料は, キリン株式会社様より提供を受けた.

参考文献

[関 03] 関 庸一: ID 付き POS データからの顧客行動パタンの抽出, オペレーションズ・リサーチ : 経営の科学, Vol.48, No.2, pp.75-82, 2003.

[石垣 11] 石垣司, 竹中毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌, Vol.26, No.6, pp.670-681, 2011.

[Hofmann 99] T.Hofmann and J.Puzicha, "Latent class models for collaborative filtering", *Proc. 16th international joint conference on Artificial intelligence*, 1999.

[Hofmann 01] T.Hofmann, "Unsupervised Learning by Probabillistic Latent Semantic Analysis", *Machine Learning*. Vol.42, No.1-2. pp.177-196, 2001.

[石垣 10] 石垣司, 竹中毅, 本村陽一: 確率的潜在意味解析を用いた大規模 ID-POS と顧客アンケートの統合利用による顧客-商品の同時カテゴリ分類, 信学技報, Vol.109, No.461, NC2009-160, pp.425-430, 2010.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

コト・データベース研究の振り返り：社会実装に向けて

Reflecting COTO Database Research: toward Implementation in the Society

渡辺 健太郎[*1] 三輪 洋靖[*2] 西村 悟史[*1] 福田 賢一郎[*1] 西村 拓一[*1]
Kentaro Watanabe Hiroyasu Miwa Satoshi Nisimura Ken Fukuda Takuichi Nishimura

[*1] 産業技術総合研究所 人工知能研究センター
Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology

[*2] 産業技術総合研究所 人間情報研究部門
Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology

In this paper, we reflect the research on COTO Database and discuss the future challenges toward the implementation of COTO Database in the society.

1. はじめに

サービス分野，特に労働集約性の高い人起点のサービス現場の改善を目的とし，提案されたのが，近未来チャレンジセッション「コト・データベースによるモノ・コトづくり支援」[西村 2014]である．ここ数年で，労働人口の減少等に伴う，介護，飲食，運輸，建設等のサービス現場における労働力不足の問題は益々深刻化しており[日本経済新聞 2017]，本セッションが設定したミッションは，今まさに喫緊の社会問題として立ち表れてきている．

本稿では，改めてコト・データベース（以降，コト DB）のコンセプトを整理し，その意義を再考すると共に，最新の技術システムのプロトタイプの実装状況を紹介する．本開発を元に，コト DB の社会実装に向けた更なる課題と取り組みについて論ずる．

2. コト・データベース

2.1 コンセプト

まず，コト DB のコンセプトについて再掲する．コト DB は，大きく下記の 3 つのコンセプトで構成される．

(1) 現場参加に基づくデータベース構築と活用

西村ら[西村 2014]は，サービス現場の生産性向上のために，現場関係者のコミュニティ形成，並びに関係者の参加に基づく業務プロセス（コト）とその支援技術（モノ）の一体的な開発の重要性を提示している．コト DB は本開発過程において業務プロセスの情報を収集し，利活用する枠組みとして提案されている．一方，コト DB は端に情報を格納する技術システムとしてのデータベースを指すだけではなく，現場参加に基づき情報収集・利活用する取り組みも含めた概念として提唱されている．本概念は，サービスの生産性に強く影響を与える，個々の関係者の主観性や状況依存性を適切に扱う上で，人の判断や思考過程を捉える過程をデータベースと切り離して考えることはできない，という考えに基づく．これは，Computer-Supported Cooperative Work (CSCW)研究における，Common Information Space (CIS) [Schmidt 1992]とも共通する考え方である．具体的なコト DB の構成要素として，業務プロセス情報を収集，分析，活用するためのデータモデル，DB 構造，情報収集システム，分析・可視化システム等の技術システムに加え，情報収集のためのワークショップ等の方法論も研究対象となっている[西村 2014]．

(2) 個別目的性を有する多種システムのデータ統合

コト DB は，現場のデータ収集において，多様なシステムからのデータを統合することで，サービス現場の業務プロセスを多面的に表現する[西村 2014, 渡辺 2014]．そのためのデータ収集技術として，モバイル端末や IoT，ヒューマンセンシング，業務支援ロボット等が考えられるが，渡辺らは個々の技術が現場の課題に対し，単独でも有益な効果を発揮する，個別目的性の重要性を主張した [渡辺 2014]．これは，現場関係者にとって短期的にも有効性を感じさせるシステムでないと，現場に定着しないという問題意識に基づくものであり，個々のシステムの個別目的性と，そこから得られたデータの統合・分析を通じた長期的な効果を合わせて追求することが，コト DB 研究の1つの方針となっている．

異なるシステムで生成される異種データの統合にあたり，Linked Data が採用されており[渡辺 2015a]，相互にデータ関連づけを行うにあたり，個々のデータ間で使われる概念の共通化を目的としたコト・オントロジーの開発が行われている[Nishimura 2016]．本オントロジーを介することで，異なるシステムが提供する情報を重ね合わせ，サービス現場を多面的に表すことができる．

(3) 主観，状況の記録

前述のとおり，人起点のサービスの業務プロセスは，個々の関係者の主観性，状況依存性に特徴づけられる．コト DB は，関係者の主観やサービスが行われる状況を記録・利活用することで，個々の関係者とその置かれている状況に，より即したサービスの開発や支援を目指している[西村 2014]．これを実現するためには，外部から観測可能な情報に留まらず，業務プロセスで発生する関係者の内部状態（気づきや感情）の記録が必要となる[渡辺 2015a]．これを実現するために，関係者の内省を促すワークショップや現場で起こった状況の記録を進めるための方法論，技術システムの開発が行われている．例えば，看護師同士の討議を通じ，個々の観点を共同記述し，サービスの業務デザインに活用するシステム Zuzie Poetry[渡辺 2015b]や，介護現場の業務情報共有システム内データの意図[中島 2015]や主観 [福田 2017]を分類する試みは，主観や状況に配慮し

連絡先：渡辺健太郎，産業技術総合研究所人工知能研究センター，〒135-0064 東京都江東区青海 2-4-7，kentaro.watanabe@aist.go.jp

た業務プロセスや支援システム開発をデータ駆動で支援するコトDB研究の一種と考えることができる.

2.2 最新プロトタイプ実装

コト DB の技術システムのプロトタイプは段階的に拡張され, 対応システム, 並びにオントロジーの拡張が行われてきた[渡辺 2014, 渡辺 2015a]. 現段階の最新のプロトタイプは下記のシステムのデータを相互に連結し, Web インタフェースを介し, 所定のクエリーに基づく情報可視化, 関連検索, 任意のクエリー検索を行うことができる. 表 1 に連携可能なシステムのリストと主な取り扱い情報を示す.

表1. コトDBプロトタイプ連携システム

システム名	機能・データ
DANCE [福原 2013]	業務情報共有のためのメッセージングシステム, テキスト・写真・動画・絵文字等
DRAW [渡辺 2015c]	業務プロセス・支援機器デザイン支援, テキスト・写真・手書きの組み合わせデータ
Quality Study [三輪 2014]	タイムスタディ支援, 業務行動データ
ドローンカメラ [Srisamosorn 2016]	バルーン型ドローンのカメラ映像
センサ付ぬいぐるみ [米澤 2017]	接触センサによる利用者との接触情報の記録
据付カメラ	静止画・動画(固定, 360 度)記録

コト DB 上での, 上述のシステム間データの相互連携を確認するために, 本研究では介護施設におけるセラピーセッションの仮想シナリオを設定した. 本シナリオに基づき, 入居者の事前状態の記録(DANCE)と介入プラン(DRAW), セッション当日の様子の映像記録(ドローンカメラ)と, 入居者とぬいぐるみの接触情報(センサ付ぬいぐるみ), さらにはその後の経過観察(DANCE)を, コト DB 上で一連の関連情報として提示することができることを確認した. 図1はコト DB のスクリーンショットである. 画面下部には上部に表示されたデータの関連情報が提示される. サービス現場への実装はまだこれからだが, 現場の状況や職員の気づきの記録も含め, 多面的に介護サービスにおける業務プロセスを表すことができることを確認した.



図1. コトDBプロトタイプ動作画面

3. コト・データベースの社会実装とその課題

コト DB 研究において, これまでにサービス現場の業務プロセスの記録と活用を, 現場参加の活動と支援システム両面で実装, 分析した例は既に複数実現されている. 例えば, DANCE のデータの共起ネットワークを用いた, 介護職員参加による業務改善ワークショップ[Watanabe 2015]の例が挙げられる.

一方, 前述のような複数システムを連携したコト DB の社会実装はまだ道半ばである. 本稿では, 今後広く社会実装を進めていく上での課題と対応する取り組みを下記に列挙する.

● プラットフォーム構築と社会実装の方法論化

単体の支援システムのサービス現場での開発・導入については, 近年複数の研究事例や実績が挙げられている[西村 2014]が, コト DB は複数のシステムの導入と現場内連携を想定しており, その社会実装はより難易度が高い. 例えば欧州では高齢者支援技術システムのプラットフォームとして UniversAAL [Hanke 2011]が挙げられるが, 複数のシステムを取りまとめるプラットフォーム開発は, コト DB においても重要な技術開発テーマとなる. また, 支援システムの導入プロセス自体の方法論化も重要なテーマである.

● インパクトの評価

コトDBの導入効果の評価も重要なトピックである. コトDBの効果を一義的なコスト換算で評価するだけでは十分ではなく, モチベーションやチームの連携度等を表す, 多様な尺度を導入しつつ, 総合的な評価を確立することが重要と考えられる.

● データ管理

サービス現場の従業員や生活者の生活・業務を多側面から捉えるコト DB 技術については, データ管理の方法や, データの分析方法に関する倫理的側面等についての検討が今後さらに求められる.

4. さいごに

本稿では, コトDB研究の変遷と最新のプロトタイプ・システムを紹介すると共に, 今後の社会実装に向けた課題について提言を行った. 今後もコト DB のコンセプトを更に発展させ, 社会により根ざした研究として取り組んでいきたい.

謝辞

本研究の一部は JSPS 科研費 15K16174, 15H01698, 16H02916 の助成を受けたものです.

参考文献

[Hanke 2011] Hanke, S. et al.: universAAL–an open and consolidated AAL platform. In: Ambient assisted living, (pp. 127-140). Springer, Berlin, Heidelberg, 2011.

[Nishimura 2016] Nishimura, S., Fukuda, K., Watanabe, K., Miwa, H., Nishimura, T.: Ontology Development for Interoperable Database to Share Data in Service Fields. In: Qi G., Kozaki K., Pan J., Yu S. (eds) Semantic Technology. JIST 2015. Lecture Notes in Computer Science, vol 9544. Springer, Cham, 2016.

[Schmidt 1992] Schmidt, K., Bannon, L.: Taking CSCW seriously: supporting articulation work. Computer Supported Cooperative Work, 1(1), 7–40, 1992.

[Srisamosorn 2016] Srisamosorn,V., Kuwahara ,N., Yamashita, A., Ogata,T., Ota, J.: Design of face tracking, System Using Environmental Cameras and Flying Robot for Evaluation of Health Care. In V. G. Duffy (Ed.), Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. DHM/HCII 2016, LNCS 9745, (pp. 264-273). Switzerland, Springer, 2016.

[Watanabe 2015] Watanabe, K., Fukuda, K., Nishimura, T.: A Technology-Assisted Design Methodology for Employee-Driven Innovation in Services, Technology Innovation Management Review, 5(2), 6-14, 2015.

[中島 2015] 中島正人:介護施設におけるスタッフ間連携のための情報共有, 筑波大学博士論文, 2015.

システムの社会実装に関する研究

[西村 2014] 西村拓一, 渡辺健太郎, 福田賢一郎, 本村陽一:コト・データベースによるモノ・コトづくり支援, 第 28 回人工知能学会全国大会講演論文集, 2014.

[福田 2017] 福田賢一郎, 西村悟史, 渡辺健太郎, 三輪洋靖, 西村拓一:介護施設における情報共有によるモチベーションデザインを目指して, 第 31 回人工知能学会全国大会講演論文集, 2017.

[福原 2013] 福原知宏, 中島正人, 三輪洋靖, 濱崎雅弘, 西村拓一: 情報推薦を用いた高齢者介護施設向け申し送り業務支援システム, 人工知能学会論文誌, 28(6), 468-479, 2013.

[三輪 2014] 三輪洋靖, 渡辺健太郎, 長尾知香, 福原知宏, 堀田美晴, 西村拓一: 介護サービスにおける感性スタディの提案, ロボティクス・メカトロニクス講演会 2014 講演論文集, 日本機械学会, 3P2-J01, 2014.

[米澤 2017] 米澤朋子, 山添大丈:触れ合いとふれあい〜相互接触へ向けた接触方法による感情生成, ヒューマンインタフェースシンポジウム 2017, 5Td3-11, 2017.

[渡辺 2014] 渡辺健太郎, 西村拓一, 本村陽一:モノ・コトづくり支援システム「UPAD toolkit」の開発, 第 28 回人工知能学会全国大会講演論文集, 2014.

[渡辺 2015a] 渡辺健太郎, 福田賢一郎, 西村拓一, 本村陽一:コト・データベースのシステム開発:その基本構造, 第 29 回人工知能学会全国大会講演論文集, 2015.

[渡辺 2015b] 渡辺健太郎他:看護現場における業務経験の表現・共有支援システムの開発, 情報処理学会論文誌, 56(1), 137-147, 2015.

[渡辺 2015c] 渡辺健太郎, 西村拓一:サービス従業員による業務設計とその分析, 日本機械学会第 25 回設計工学・システム部門講演会講演論文集, 2015.

[日本経済新聞 2017] 日本経済新聞: 人手不足, 業種格差くっきり 銀行にはリストラ余地 (2017/12/18), 2017.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

コト・データベースによるモノ・コトづくり支援

ー構造化知識データ融合によるサービスインテリジェンスを目指してー

Product and Activity Design Support based on Event Database

西村 拓一 Takuichi Nishimura　渡辺 健太郎 Kentarou Watanabe　福田 賢一郎 Ken Fukuda　西村 悟史 Satoshi Nishimura　本村 陽一 Yoichi Motomura

産業技術総合研究所 人工知能研究センター
Artificial Intelligence Research Center, AIST

For human-oriented service fields, we have been developing technologies to enhance the intelligence (observation, judgment, cooperation) of people by Artificial I intelligence(AI). We will increase their observation ability by information sharing technologies of awareness and action measurement technologies of human behavior and body motion. We will raise their judgment ability by knowledge structuring technology that helps organize the knowledge of activities. This technology is aimed at reducing risk, improving productivity, and improving quality of service. As such community, we target on not only nursing care, nursing, health promotion but also education and manufacturing industry. In order to enhance human intelligence, we are building structured knowledge as an example of events database. AI and robot will use structured knowledge and modify through interaction with employees. It is possible to innovate involving various stakeholders because they can clearly grasp the current situation of the community by the structured knowledge.

1. はじめに

生活者の行動様式やニーズの多様化，少子高齢化や環境問題等の社会問題を背景とし，様々な利害関係者の要求を満たす手段としてのサービスの改善は社会，産業上の主要課題となっている．特に人と人の相互作用により価値創出を行う，人起点のサービス現場の改善には，業務プロセス(コト)の正確な理解と改良と，業務に用いる技術システム(モノ)を一体的に開発し，サービス現場の能力を拡張することが求められる．そのような研究領域をサービスインテリジェンスと呼び，新しいサービス設計技術，及び方法について研究開発が進めている．

そこで，本稿では，人々が主体的・共創的にインテリジェンス(観察，判断，行動力)を高める方法論とそれを効率的に実現する人工知能技術，現場での取り組みを概観する．具体的には，モノだけでなく人々の活動とその意味，感情，知識を「コト」としてデータベース化，モデル化し，新たなコトとモノを設計することを支援する「モノ・コトづくり支援技術」および，介護，看護，健康増進，保育，教育，理美容，製造業など，人々の能力を拡張することを支援する技術を必要としている現場での取り組みを紹介する．

医療，介護，教育などチームワークで顧客ニーズを深堀しカスタマイズするサービス現場や健康増進や地域などお互いに能力や時間を提供し合うコミュニティ活動の改善には，業務プロセス(コト)の正確な理解と改良に加え，業務に用いる技術システム(モノ)の一体的な開発が必要である．しかし，各従業員がどのような主観世界のもと，どのようなプロセスでサービス要素を構築し顧客に価値を提供しているか理解し改善することは困難である．さらに，サービス現場の改善の鍵となる技術システムを開発する関係者が正確な要求仕様を決定することも困難である．

このために，現場関係者が主体的にコミュニティを形成し，外部専門家とともに業務プロセスの把握や改善，技術システム開発を推進する現場参加型開発[本村 2013][須永 2013]が注目されている．これは，介護・看護現場や教育現場などチームワーク型やマニュアル化できない都度対応型の業務プロセスの改善に特に有効である．このような方向の成功例として，IT ベンダーで技術もコミュニティデザインも把握した人材を現場に送り，新たな業務フローを構築しつつ，その業務フローに最適なシステムを開発するフィールド・イノベーションというアプローチが知られている．[有馬 2013]

しかし，主観性や状況依存性，観点の多様さのために業務プロセス(コト)の形式化と分析が困難であった．このため，過去の活動の成果を記録・再活用できず，システム開発者などコミュニティ外のメンバーとの知見の共有も困難となり非効率で「人」依存となっていた．また，現場依存性も高く水平展開が困難となっていた．[西村 2013]

そこで，業務プロセスを蓄積・共有するコト・データベース(コト DB)により，効率的な現場参加型のモノ・コトづくり支援技術が役立つと考えている．コト DB では，主観性や状況依存性を含む現場のコト情報を収集，分類，検索する技術(ワークショップ手法やコトを記述するモデル化手法，DB 構造)など，モノ・コトづくり支援では，従業員，システム開発者等，それぞれの視点で業務プロセスの分析，可視化技術などが課題となる．具体例として，お互いに切磋琢磨しつつ身体活動を楽しむ健康増進コミュニティ支援[西村 2015]を提案している．

本稿では，コト DB の具現化である構造化知識データ融合によるサービスインテリジェンスの可能性を考える．

2. サービス現場のインテリジェンス

2.1 背景

介護・介護予防・健康増進における専門職の能力を AI により拡張し，例えば通常は AI が在宅介護者を支援しその様子を見守る専門家が必要に応じて支援するなど，専門家の能力を普及することが求められている．保育，教育でも同様な支援技術が有効であり，製造業でも少量多品種の現場では人の経験や勘を共有し従来の QC 活動やそれを支える QC 七つ道

連絡先：西村拓一，産業技術総合研究所サービス工学研究センター，〒135-0064 東京都江東区青海 2-3-26 ，takuichi.nishimura@aist.go.jp

具だけでは困難となっている．これらのサービス現場におけるインテリジェンスを支援するためには，現場関係者の主体的で積極的な参加や行動変容が重要である．そこで，各種技術の開発及び導入を現場主体で進める，現場参加型開発が重要である．健康増進コミュニティや地域コミュニティなどの無償のサービス現場も含む各種現場の活動と知識を無理なく形式化し共有しブラッシュアップすることで，実践知化し，各種現場や分野を超えた知識連携を実現する．

2.2 現場参加型開発

現場で活用できるシステムやロボットを構築するためには，業務プロセスの改善や関係者の考え方の変化も必要である．そこで，製造業における QC 活動のように現場主体の改善が重要となる．現場コミュニティが協働し自発的・創造的に改善活動を行うことで，現場力の向上が可能となる．また，このような現場では，品質が主観や状況依存となるため計測が困難となり，QC 七つ道具などの工学的な改善支援技術だけでなく，活動や運動の質を可視化する技術が必要となる．このために，コミュニティ形成のためのワークショップや現場の現状を可視化することで，経験と勘の表出と相互理解が進む．ミッション，役割，業務プロセスの合意形成だけでなく，既存のルールを破壊することから新プロセスを創出できる場合もある．関係者が主体的になるため，現状把握能力が向上し，業務プロセスの改良能力が向上することも期待できる．

2.3 コミュニティ支援技術

コミュニティ支援に役立つ技術として日々の申し送りを通して現場の気付きや暗黙知を記録・共有するためのツール DANCE (Dynamic Action and kNowledge assistant for Collaborative sErvice fields)および従来のタイムスタディを拡張しサービス品質の記録と分析を支援するツール QualityStudy を開発して来た．さらに，複数の従業員が自身の持つサービス現場の情報や知識を表現・共有し，よりよいサービスをデザインするためのツール DRAW (Design Representation tool for Autonomous Work systems)[渡辺 2015]を開発している．

2.4 活動や運動の質可視化技術

理学療法士や教師が運動指導時に運動の質を高めるために活用できる技術も開発している．身体の動きや傾きを可視化する Axis Visualizer を開発している．[西村 2016] これは，体幹の意識を高めるために，特定の基本動作をすることで，体幹をぶらさず効率的な動きをしているか計測できるものである．日常生活や様々なスポーツで用いられているモーションキャプチャ技術も有用である．Axis Visualizer はこのモーションキャプチャと併用することで，通常は Axis Visualizer を使用し，より詳細な知見を得る場合はモーションキャプチャで評価する．

3. コト DB の実例：目的と根拠を「説明できる AI」

前節のサービスインテリジェンスの実現のために，本節では，さらに現場の専門家の知識を構造化し，状況のデータと融合したコト DB を構築することで目的と根拠を「説明できる AI」を実現することを提案する．この「説明できる AI」により，現場のノウハウの共有による業務効率化やリスクの共有による事故の低減，業務品質の標準化など，現場の人々の能力を拡張することが可能となる．

介護業務を例に説明すると，介護サービスの現状は，少子老齢化に伴う介護コストの低減と質の向上が急務となっており，日本の介護サービスの輸出も期待されている．このために，業務マニュアルを構築することで，業務効率化（教育コスト低減，リスクの低減）と業務の標準化と質の向上を実現し，輸出可能（成功事例：学習塾やコンビニなど）とすることが期待されている．

しかし，介護業務の知識化は困難である．まず，被介護者は多様であり介護者の身体能力や知識も異なり施設など状況に応じて，業務手順は多様となるため，すべての場合の手順を書きつくすことが困難である．また，一度マニュアルを作っても，その後活用せず，内容も古くなることが多い．さらに，マニュアルに対する指導者の懸念として，マニュアル手順の丸覚えとなり各作業の目的を理解せず，異なる状況へ適応困難な人材が育つのではということがある．また，マニュアルを作っても，その手順に納得せず従わないのではとの懸念もあった．

そのため，図1のように，手順の目的と根拠を「説明できる AI」の実現を目指している．これは，介護の一つである，褥瘡予防に関して目的指向で知識を構造化したものである．「褥瘡を予防する」ためには「血行を良くする」を含む 3 種類の行為があり，さらに「血行を良くする」ためには，利用者が「寝返りをする」か寝返りできない場合は，介護者が「利用者に声をかけ」た後に「体位を変える」．対位を変える頻度は 2 時間に1回であり，持ち上げるとリスクがあることが書かれている．図1の右上に書いた記法で構造化されている．

通常のマニュアルは，第4階層のように手順のみが書かれており，その上位の中間目的が書かれていないことが多い．この各行為の中間目的を把握していることにより，状況に応じた行為を思いつくなどの応用力を高めることになる．多様な被介護者に必要な介護行為すべてをかけないため，このような状況適応能力を高めることは重要である．

また，各行為には，IoT，センサ情報や事故ヒヤリハット情報，被介護者個人のケアプラン，介護実施結果などの現場の活動を捉えたデータをリンクさせていく．通常の介護施設では，これら様々なデータがつながることなく存在し，利活用されにくい状況である．このことは，他のサービス現場や製造現場でも頻発している．そこで，この目的指向構造化知識の各行為に対して，現場の様々活動データを自然にリンクする仕組みを構築している．

各行為へのデータ化が集まってくれば，図1右下のように，機械学習により，AI が予測して注意したり，新たな手法を示唆することが可能になる．例えば，指導者が腰痛になりにくい移乗介護動作を新人に指導する際に，腰に加速度センサを用いるスマホアプリを装着していたとする．良い動作の場合と危険な動作の場合の教師付きデータが集まることで，機械学習し，以降は新人が腰にスマホアプリをつけていると指導者と同じ指導内容を AI が行うことが可能となる．

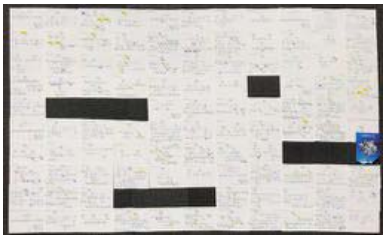
特に，これらの知識構築とデータ融合は，外部専門家が実施するのではなく，介護現場の介護師らその現場を詳細に把握しているメンバーがそれぞれの視点と知識を用いて日々の活動の中で無理なく構築できるよう「知識発現」手法を提案している[西村 2017]．これは，図1右上のようにわかりやすい記法を開発することで実現した．さらに現場共通の知識を事前に構築し，それを各現場へ配布し，各現場のメンバーは，その現場固有の業務プロセスへ修正するだけで済むようにすることで実現した．

今まで，図2のように複数の教科書から介護業務の基本構造化知識を人手で構築してきた．この際，様々な介護士に指導いただいている．特に，認知症の被介護者のみのグループホームや要介護度が高い老健など業態によって介護業務の内容が異なるため，様々な業態で活用できるものとしている．今後，この基本知識をもとに様々な現場の固有知識を構築していくなかで，

基本知識自体の改良，各業態の基本知識の構築を進めていく予定である．

図3に，目的と根拠の知識をもとに「説明できる AI」の将来像を示した．将来は，既存のマニュアルから半自動で固有知識を構築し，不明な中間目的などを AI のチャットボットやロボットがインタラクティブに従業員から収集する．また，基本的な初心者への指導は，構造化知識から実施し，この知識に含まれない事象が起きた場合は，指導者へ問い合わせる．指導者から新たな知識を獲得し，構造化知識を改良した上で指導を進める．また，始動時に収集された各行為のセンサ情報と教師データから機械学習を行い，各行為のリスクを予測する．危険がある場合は本人と指導者に通告し，さらに指導者の知識を学習していく．

将来像としては，このような現場のインテリジェンスを拡張する「説明できる AI」，「説得できる AI」を目指しているが，文章構造の自動分析が困難，中間目的が書かれていないなど，既存のマニュアルから構造化知識を自動で生成することも多くの研究課題が存在する．特に指導者も中間目的を意識することなく業務を行なっていたり指導をしている場合があり，その場合は，暗黙知を引き出す高度な対話制御技術を必要となる．



介護業務の種類	行為数
排泄介助	277
入浴介助	253
移乗介助	247
移動介助	236
更衣介助	226
体位変換	197
口腔ケア	170
食事介助	140

図2　人手で構築した介護業務の構造化知識

4.　身体動作の専門知識とデータ融合例

本節では，前節の技術を身体動作に対しして実施した例を示す．図4のように「歩く」動作などを構造化したところ「体幹をひねる」動作が重要であることがわかった．様々な日常動作，スポーツ動作は体幹と四肢の協調運動とすることで障害予防，上達促進が可能となる．体幹は前後左右へのスライドや曲げの動きも重要であるが，最も日常的に使用される動きはひねり動作となる．歩行時は，体幹のひねり振動と四肢が連携している．

そこで，体幹をひねる動作に関してモーションキャプチャや床反力計に座位で座り，胸郭を回転する動きを計測した．この際

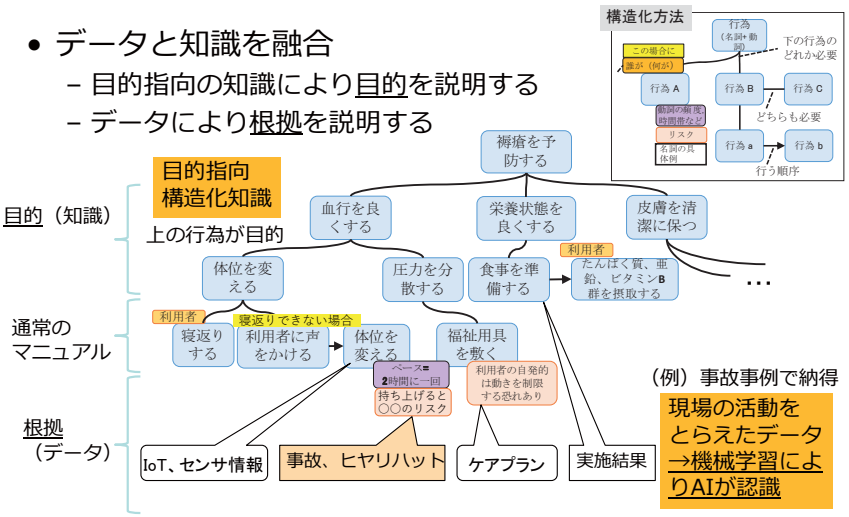


図1 「説明できる AI」のデータ構造化知識

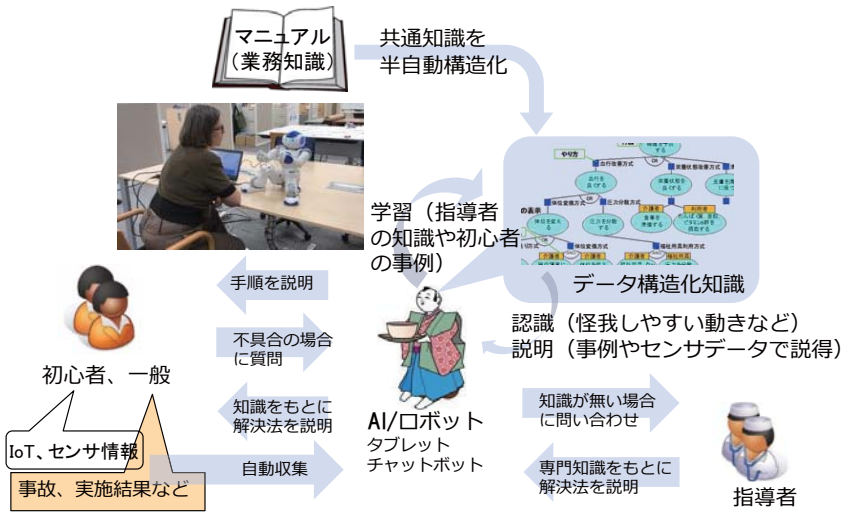


図3 「説明できる AI」のイメージ

モーションキャプチャシステムにより胸郭および骨盤の回転角度を計測した．このデータを分析したところ，バネモデルで記述できることを突き止めた．そこで，スマートフォン内蔵の加速度センサを用い，ユーザがスマートフォンを胸部に当てて体幹ひねり動作を行うことで体幹ひねりの品質を計測する簡易アプリを構築した．[吉田 2018]

この例は，データ知識融合の一例である．つまり，図4のように構造化知識から発見した重要な要素動作を，実験室で分析し，バネモデルを構築した．このモデルを用いて，日常空間で計測可能な簡易アプリを開発し，現場でのデータ収集と簡易診断を行うことで，AI により専門家の指導能力を拡張したと言える．

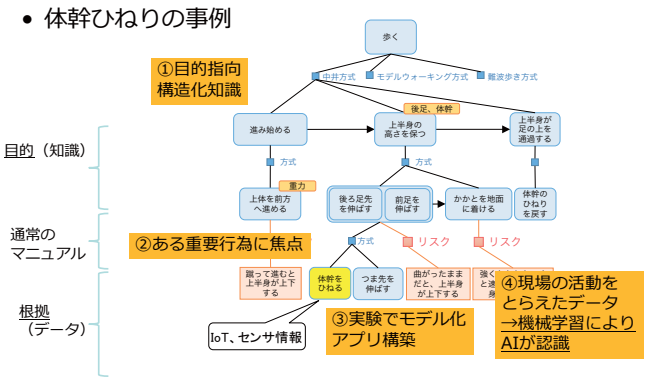


図4　データ知識融合の例

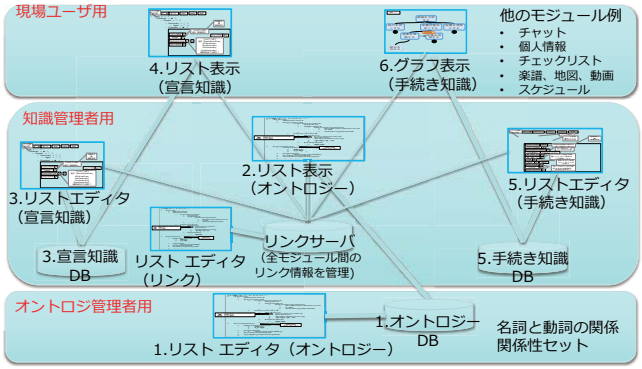


図5　リンクのイメージ

図5には，データ知識融合を実現するための様々なツールのイメージを示した．構造化知識は，宣言知識と手続き知識を有し，それらはリンクされている．例えば，手続き知識の一つのノード「褥瘡を予防する」の「褥瘡」は宣言知識の「褥瘡」とリンクされ，その詳細説明を参照できる．また，当該現場では「褥瘡」ではなく，「とこずれ」という表現をしている場合は，現場ユーザ様の構造化知識では，「とこずれ」という用語を用いる．その場合，現場の知識管理者は，これらの二つの用語のリンクを作成し，全国共通のオントロジーのなかの「褥瘡」とリンクする．これにより現場間の比較や知識の共有も容易になる．

また，他に従業員間でチャットを用いて手続き知識の改良を議論していたら，そのスレッドと当該要素知識の間にリンクを貼る．これにより，手続き知識が変更された理由や過去の議論がすぐに参照可能となる．他に，図5右上のような様々なデータと知識をリンクすることで，組織内に存在する，センサ情報，リスク情報，実施記録，事例，やりとりなどを知識ベースで把握し，経緯や根拠を把握できる．

5. まとめ

本項では，サービス現場の人々の能力を拡張する技術を導入するために現場の状況を把握する重要性を述べ，さらに現場の状況だけでなく，暗黙知を形式化する知識構造化技術を述べた．また，このデータ知識融合によるサービスインテリジェンス実現の可能性を考えた．今後は，サービス現場の能力を拡張する「説明できる AI」の構築を進めるとともに，データ構造化知識をもとに新たなサービスをデザインする研究も進めたい．

謝辞　本研究の一部は経済産業省ロボット介護機器開発・導入促進事業，平成 23 年度経済産業省委託事業 次世代高信頼・省エネ型IT基盤技術開発・実証事業（サービス工学研究開発分野）「本格研究による人起点のサービス工学基盤技術開発」，国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務「次世代ロボット中核技術研究開発」並びに科研費 (課題番号 24500676, 25730190) の助成の下，実施されました．各種研究にご協力いただいた介護老人保健施設・和光苑，はなまる会，伸こう福祉会福祉会のみなさまに御礼を申し上げます．

参考文献

[本村 2013] 本村 陽一，西村 拓一，西田 佳史，佐藤 洋，大山 潤爾，介護・医療における現場参加型アプローチの課題と展望 ―持続的・自律的サービスシステムの実現に向けて―，人工知能学会誌，Vol.28 No.6, p8.924-929, 2013.

[須永 2013] 須永　剛司・小早川　真衣子・山田クリス孝介・渡辺 健太郎・新野　佑樹・西村　拓一，Co-design プロジェクトが自発的に回ること ―社会を形づくるデザインに向けて―，人工知能学会誌，Vol.28 No.6, p8.886-892, 2013.

[西村 2013] 西村 拓一，渡辺 健太郎，本村 陽一，コト・データベースによるモノ・コトづくり支援，2013 年人工知能学会全国大会，3A1-NFC-03-3, 2013.

[有馬 2013] 有馬 淳，千葉 広隆，中川 肇，フィールド・イノベーション：現場参加型の業務改革，人工知能学会誌，Vol.28 No.6, p8.880-885, 2013.

[長尾 2014] 長尾知香，西村拓一，中高年者の健康増進活動支援における新たなサービス体系の構築-介護予防と地域活性化の支援-,第2回サービス学会国内大会，P1−6, 2014.

[渡辺 2015] 渡辺健太郎,藤満幸子,原田由美子,山田クリス 孝介,須永剛司,小早川真衣子,新野佑樹,阪本雄一郎，西村拓一,本村陽一: 看護現場における業務経験の表現共有支援システムの開発,情報処理学会論文誌，Vol.56, No.1, pp. 137-147, 2015.

[西村 2015] 西村 拓一，渡辺 健太郎，福田 賢一郎，本村 陽一，コト・データベースによるモノ・コトづくり支援ー現場の気付き能力と設計能力の拡張を目指してー，2015 年人工知能学会全国大会，1K5-NFC-05b-1, 2015.

[Iino 2017] Nami Iino, Satoshi Nishimura, Ken Fukuda, Kentaro Watanabe, Jokinen Kristina, Takuichi Nishimura (2017), Development and use of an Activity Model based on Structured Knowledge – A music teaching support system –, IEEE International Conference on Data Mining Workshop, November 18.

[吉田 2018] 吉田 康行，梁 滋路，西村 悟史，鴻巣 久枝，長尾 知香，西村 拓一（2018）運動指導のための簡易評価技術の開発 --携帯端末による体幹捻り運動評価情報処理学会論文誌 59(2) 1-11.

6:20 PM - 7:00 PM (Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room K)

[2K4-NFC-3b-a] (Panel discussion) Data and Knowledge Construction, Sharing and Utilization in Real World

Satoshi Nishimura[1], Takuichi Nishimura[1], Ken Fukuda[1], Kentaro Watanabe[1], Kristina Jokinen[1], Yasuyuki Yoshida[1], Nami Iino[1], Lihua Zhao[1] （1. National Institute of Advanced Industrial Science and Technology）

©The Japanese Society for Artificial Intelligence

[2L1] [General Session] 9. NLP / IR

座長:柳瀬 利彦（株式会社 日立製作所）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room L (3F Sapphire Hall Asuka)

[2L1-01] Location identification using local context and relevant documents

○Ryo Seki[1], Takashi Inui[1] （1. University of Tsukuba）

9:00 AM - 9:20 AM

[2L1-02] Convolutional Neural Network for Event Knowledge Extraction

○Kai Ishikawa[1], Hiroya Takamura[2], Manabu Okumura[2] （1. NEC Corporation, 2. Tokyo Institute of Technology）

9:20 AM - 9:40 AM

[2L1-03] Search for New Application of Invention based on Extraction of Causality from Patents

○Takahisa Ota[1], Takuya Minami[1], Yusuke Yamazaki[1], Yoshishige Okuno[1], Chinatsu Tanabe[1], Hiroyuki Sakai[2], Hiroki Sakaji[3] （1. SHOWA DENKO K.K., 2. Seikei Univ., 3. The Univ. of Tokyo）

9:40 AM - 10:00 AM

[2L1-04] Mathematical Expression Retrieval in PDF documents from Web using Mathematical Terms as Queries

○Kuniko Yamada[1], Harumi Murakami[1] （1. Osaka City University）

10:00 AM - 10:20 AM

[2L1-05] Politician classification based on political probrems extracted from Twitter Data

○Mayumi Nomura[1], Tomonobu Ozaki[1] （1. College of Humanities and Sciences）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

局所文脈と関連文書を用いた地名に対する地理的位置の同定

Location identification using local context and relevant documents

関 龍[*1] 乾 孝司[*2]
Ryo Seki Takashi Inui

[*1]筑波大学情報学群
School of Informatics,University of Tsukuba

[*2]筑波大学大学院システム情報工学研究科
Graduate School of System and Information Engineering,University of Tsukuba

In this paper, we discuss the task of matching the location name appearing in the document to the geographical location on the real world. In this task, it often happens that one location name points to more than one geographical location. In order to resolve this situation, many previous methods have been proposed for selecting the best candidate from the candidates. For example, a method of outputting a candidate having the largest population was proposed. In this research, we propose a method to select the best candidate by using local context and relevant documents which are easily obtained and easily preprocessed than the external data necessary for the previous research. Through experiments, the proposed method achieves about 70% accuracy, and further improvement of accuracy is expected by handling more relevant documents.

1. はじめに

近年，SNS などの様々なサービスが普及し，人々が気軽に情報を発信できる時代になってきている．これら発信された文書内にある，場所を示す表現と実世界上での地理的位置を結びつけたいというニーズが存在する．SNS の代表とも言える Twitter では投稿時の位置情報をつぶやきに付与することが可能であるが，Middleton ら [1] によればこの機能を利用しているものは全体の 1%にも及ばない．文書内の表現に対して実世界上でのその地理的位置が同定できれば，特定区画に対象を限定したテキストベースの社会分析などがこれまで以上に高精度・高被覆で実現できるようになる．

文書内の場所を示す単語と実世界での地理的位置を結びつけるには様々な問題があるが，その中でも場所を示す単語の曖昧性の問題がある [2]．曖昧性とは例えば,「中央区」という単語があった場合に，日本に数ある中央区のうちのどの位置を実際に参照しているのかわからない状況のことである．実際に日本には数多くの中央区が存在するが，簡単のため「大阪市中央区」と「東京都中央区」に絞った例を図 1 に示す．この地理的位置推定の研究は 10 年ほど前から行われており，どのように場所を示す単語の曖昧性を解消するかの方法には様々な手法が提案されてきた．文書レベルでの地理的位置推定や単語レベルでの地理的位置推定，教師データの有無などによって手法は異なるが，本研究では単語レベルかつ教師データを用いない地理的位置推定における手法を扱う．その代表的な例が Speriosu[3] の人口データを利用する「POPULATION」である．この手法では人口データが必要となる．また一般に，人口が集計・管理される粒度は市区町村のレベルに留まっており，それより狭い区画を指す大字（おおあざ）レベルの人口データを全国規模で入手することは困難である．

本研究では先行研究で必要な外部データを使わず，それに準じかつ入手や扱いが容易な局所文脈や関連文書を用いた手法を提案する．文書中の局所文脈の情報を利用して曖昧性を解消する文脈参照法と関連文書を用いて曖昧性を解消する MENTION_COUNT の二つの手法である．

本論文では文書中に表れる実世界上のどこかを指す単語を「地名」と呼び，その地名に対応する可能性のある実世界上の住所を「候補地」あるいは単に「候補」と呼ぶ．



図 1: 中央区に対する曖昧性

2. 関連研究

POPULATION[3] とは人口データ（以降，人口 DB）を使用する方法であり，候補のうち一番人口が多い候補を出力とする．主に文書で言及される都市は有名，つまり人口が多い都市であることが多いという前提のもとに考えられた方法である．例えば「ロンドン」と文書中にあった場合は主にイギリスのロンドンを指すことが多いがカナダのオンタリオ州にもロンドンが存在する．しかし，人口はイギリスのロンドンの方が多いためイギリスのロンドンを出力とし，カナダのロンドンを避けることができる．

文書中の地名を t，t が指す候補地（地理的位置をあらわす住所）の集合を $C^t = \{c_i^t\}$ とすると，地名 t に対する地理的位置同定の問題は次式で書くことができる．

$$\hat{c^t} = \arg\max_{c_i^t \in C^t} score(c_i^t) \quad (1)$$

POPULATION は上式のスコア関数 $score(c_i^t)$ として候補地 c_i^t の人口を返す関数を想定した手法である．

しかしこの方法はターゲットに対応した人口 DB が必要であり，対象とする地名のレベルを市区町村でなく大字レベルにまで細分化させた場合，人口 DB の量も非常に大きくなる．また，大字レベルの人口 DB を日本全国の範囲で入手しようとすると困難である．

表 1: 住所 DB の形式

地名	候補数	候補 1	候補 2	•••
松屋町	8	岐阜県_岐阜市_松屋町	京都府_京都市_下京区_松屋町	京都府_京都市_上京区_松屋町
		京都府_京都市_中京区_松屋町	京都府_京都市_伏見区_松屋町	大阪府_寝屋川市_松屋町
		大阪府_大阪市_中央区_松屋町	香川県_丸亀市_松屋町	
本田	7	山形県_鶴岡市_本田	埼玉県_深谷市_本田	埼玉県_南埼玉郡_宮代町_本田
		新潟県_新発田市_本田	富山県_射水市_本田	岐阜県_瑞穂市_本田
		大阪府_大阪市_西区_本田		
亀田	3	福島県_郡山市_亀田	千葉県_富津市_亀田	新潟県_新潟市_江南区_亀田
旭区	2	神奈川県_横浜市_旭区	大阪府_大阪市_旭区	
パルプ町	1	北海道_旭川市_パルプ町		
芝浦	1	東京都_港区_芝浦		

3. 住所データベース

本研究では候補の情報を得るために住所データベース（以降，住所 DB）を参照する．住所 DB とは日本中全ての住所を含んだデータベースであり，実世界上の住所と一致していなくてはならない．しかし，住所は合併などにより流動性をもったものであるから必ずしも住所 DB と実際の住所が合致しているとは限らない．本論文では住所 DB のおおもとのデータとして国土交通省の発行している「街区レベル位置参照情報」*1を利用した．今回使用するこのデータのサンプリングは 2016 年である．

本論文では一つの地名 t に対してどのような候補が存在しているのかを知りたいため，上記データに前処理を施し，表 1 の形式のデータを得る．例えば，地名「パルプ町」の場合，この名前をもつ地理的位置は北海道に 1 地点存在するだけであるので曖昧性がない．一方,「旭区」は 2 地点存在するため曖昧性の解消が必要である．なお，候補は都道府県名からの住所を含んでいるため候補間に曖昧性はない．住所 DB のエントリ数は 152,937 である．この内，候補数が 2 以上になる（曖昧性がある）エントリの割合は約 5%であり，これらエントリにおける平均候補数は 4.49 である．

4. 提案手法

4.1 文脈参照法

文脈参照法は非常に直感的な方法であり，ターゲットとなる地名の前後 k 文字を参照し，そこにターゲット以外の地名があった場合にその情報を用いて地理的位置を決定する方法である．例えばターゲットとして「中央区」があり，候補として「東京都中央区」と「大阪府大阪市中央区」があるとする．このときに「中央区」の前後 k 文字中に「東京都」があった場合出力として「東京都中央区」を選択する．前後 k 文字中に複数の地名が出現することもあるが，この場合は次のような規則に従う．

- （優先度 1）後方文脈よりも前方文脈を優先させる
- （優先度 2）ターゲットに近いものを優先させる

これは文書において地名は曖昧性が無いように書かれることが多く，例えば「東京都の中央区」というように地名の前にその地名の属する地名を書くことが多いため前方文脈を優先している．また，ターゲットに近いものを優先するのは単純にその方が遠いものよりもターゲットの説明をしている場合が多いと考えたからである．この方法では前後 k 文字内にターゲット以外の地名がない場合は曖昧性は解消されない．

文脈参照法の具体的な手続きを以下に示す．まず，準備として，以下の処理をおこなう．

- 文書中の地名 t に対して，その前後文脈の文字列を取得する．
- 前節でみたように，各候補は都道府県から始まる住所である．そこで，候補を都道府県，市区町村，大字の 3 つの要素に分解する．

以上の準備のあと，各候補の要素（都道府県／市区町村／大字）が t の前後文脈に出現しているか否かを文字列マッチングによって確認し，出現していた候補を出力する．複数の候補がマッチした場合は上述の優先順位に従って出力となる候補を決定する．また，どの候補ともマッチングが取れない場合は出力なしとなる．

4.2 MENTION_COUNT

MENTION_COUNT は式 (1) のスコア関数 $score(c_i^t)$ として，関連文書集合 D 中での候補地 c_i^t の言及回数を返す関数を用いる手法である．事前に各候補地の言及回数を保存した言及回数データベース（言及回数 DB）を作成しておき，地理的位置同定の際はこの言及回数 DB を参照することでスコア値を得る．

4.3 言及回数 DB

言及回数 DB とは候補とその言及回数の対を保存したデータベースである．生コーパスを使用し，文脈参照法を一記事の範囲に対して行うことによって作成した．ここで使用するコーパスは固有表現抽出などの処理が加えられておらず，真にテキストのみのデータであるから非常に入手が簡単である．本論文で用いる外部データはこのコーパスのみであるが，先行研究で用いるような人口 DB や座標データに比べると入手は簡単であり，前処理もシンプルであるため使いやすいと言える．

候補地の言及回数を数えるには，関連文書中において「大阪府大阪市中央区」のように，候補間で曖昧性を持たないように記述されていることが理想であるが，そのような記述は地名の言及の一部にすぎない．そこで，都道府県レベルに注目し，また，文脈を一つの記事全体とした文脈参照法を適用することで近似的に各候補地の言及回数を求める．手順としてはまず，関連記事に出現する都道府県名を記事毎にリスト化しておく．次に住所 DB から地名をひいていき，全ての記事に対して検索をかける．ある記事にその地名があった場合，先の都道府県リストを見に行き，リスト中の都道府県に属するターゲットの

*1 http://nlftp.mlit.go.jp/isj/index.html

候補があったらその候補をインクリメントする．全ての記事についてこれを行ったらひとつの地名に対しての処理は終わりである．これを全ての地名について行えば言及回数 DB が完成する．

5. 評価実験

5.1 実験の設定

5.1.1 実験条件

先行手法である POPULATION，提案手法である文脈参照法と MENTION_COUNT，文脈参照法を使用して出力なしとなった場合に POPULATION を使う手法 (文脈参照法+POPULATION)，文脈参照法を使用して出力なしとなった場合に MENTION_COUNT を使う手法 (文脈参照法+MENTION_COUNT) の 5 つを比較する実験を行なう．文脈参照法において参照する文字数 k を指定する必要がある．今回はコーパスの 1 記事あたりの平均文書長 $L = 1,657$ のとき，前方文脈を k_1，後方文脈を k_2 とすると $k_1 = k_2 = \frac{L}{2} = 829$ とした．また今回の設定ではコーパス内にあるターゲットは入力として与えられる．

5.1.2 データセット

評価を行なうコーパスは拡張固有表現タグ付きコーパス [4] の新聞ジャンルを用いる．候補データ生成に用いる住所データベースは 3 節に示したものを利用する．POPULATION に用いる人口 DB は政府が発行している人口統計データ「e-stat 統計で見る日本」*2 の調査年 2015 年のデータを用いる．言及回数 DB を作成する際に使用する生コーパスとしては毎日新聞社のコーパス [5] 1 年分を用いる．

正解データは人手で作成した．拡張固有表現タグ付きコーパスに付与された City タグが付く地名のうち，住所 DB を参照して複数の候補が得られるものについて，その近隣文脈を読むことで正解候補を選択した．今回は新聞ジャンルのみの実験であるから，新聞ジャンル中の候補が複数ある City タグ全 234 個について正解を付与した．

5.1.3 評価尺度

評価尺度には正解率を用いる．これは手法により選択した候補と正解ファイル中の候補とを比較し一致していた場合に正解，そうでない場合に不正解としたときの候補が複数ある全 City のうちの正解の割合である．

5.2 実験結果

5 つの手法それぞれの正解率を表 2 に示す．この結果に対して，ターゲットとなる地名が大字レベルのものとそれ以外*3に分割して再評価した結果を表 3 に示す．ここで，不正解を誤選択，正解なし，出力なしの 3 つに分けた．誤選択とは正解候補と出力候補がどちらも存在するが一致しない場合であり，提案手法によって正しい候補を選ぶことができなかったことをあらわす．正解なしとは正解候補が存在しない場合であり，コーパス中の地名と住所 DB の地名とが一致せずに正解候補を正しく選択できなかったものである．これは合併などによる原因が考えられる．出力なしとは出力候補が存在しない場合であり，提案手法によって候補選択が行えなかった状態である．また正解なしの場合は出力候補の有無はドント・ケアであり，出力がなくても出力なしにはカウントされない．

表 2 より，先行手法の POPULATION よりも提案手法の文脈参照法や MENTION_COUNT の方が正解率が高いことがわかる．また，文脈参照法+POPULATION や文脈参照法+MENTION_COUNT は POPULATION や MENTION_COUNT 単体よりも大きく正解率が上がっている．つまり，文脈参照法に他の手法を組み合わせることで性能が向上すると言える．文脈参照法+POPULATION と文脈参照法+MENTION_COUNT では正解率がほぼ同じであることもわかる．表 3 より，大字レベルでは POPULATION の正解数 0 に対して文脈参照法と MENTION_COUNT は表 2 とくらべてあまり正解率が変わっていない．これは人口 DB が大字レベルのものはなく，POPULATION では市区町村レベルの曖昧性解消しかできないことを示してる．

以上の結果より，POPULATION，文脈参照法，MENTION_COUNT の 3 つでは文脈参照法が単体では一番性能が良く，文脈参照法では出力がない場合に他の手法を適用する組み合わせ手法を用いればより性能が上がるとわかった．組み合わせる手法として POPULATION と MENTION_COUNT を用いたが，どちらも同等の性能であり，人口 DB を用意できない場合に言及回数 DB を代わりに用いれば同程度の精度が出せるということがわかった．また，曖昧性を解消する地名のレベル（市区町村，大字）によって有効な手法が違うということもわかった．

文脈参照法+MENTION_COUNT の結果の例を次に示していく．コーパスは簡単のためターゲットを含む一文のみを示す．

まず，正解の例を示す．上がコーパス，下が住所 DB のクエリで，太字になっている候補が出力である．この例では文脈参照法により，ターゲットである「内幸町」の直前にある「東京」を用いて「東京都_千代田区_内幸町」を出力することができ，正解となる．

> （古沢　由紀子）池田弘子さん １７日にＮＩＥ講習会 中学校教員を対象にＮＩＥ（教育に新聞を）活動の実践例を報告し、活用に役立てる「第７回ＮＩＥ講習会」が １７日午後２時３０分から、東京・内幸町の日本プレスセンターで開催される。

> 内幸町 **東京都_千代田区_内幸町** 富山県_富山市_内幸町

誤選択の例を次に示す．この例ではターゲットである「宇都宮」の前にある「新潟」を用いて「新潟県_十日町市_宇都宮」を選択してしまい，誤選択となる．なお，$k_1 < 9$ であれば MENTION_COUNT を使用するため正解となる（言及回数 DB では宇都宮に対し栃木県_宇都宮市が一番言及回数が多い）．

> ７月に秋田、新潟、盛岡、９月に宇都宮、東京でも開く。

> 宇都宮 **新潟県_十日町市_宇都宮** 栃木県_宇都宮市

MENTION_COUNT による正解の例を次に示す．この例ではターゲットである「横浜」の前後に曖昧性を解消できる地名が全くないため文脈参照法では解決できずに MENTION_COUNT を使用する．言及回数 DB では「横浜」に対し「神奈川県_横浜市」が一番言及回数が多いためそれが出力となり，正解となる．

*2 https://www.e-stat.go.jp/regional-statistics/ssdsview

*3 都道府県レベルでは曖昧性が生じないため，ここに該当する事例はすべて市区町村レベルの地名である

表 2: 5 手法の正解率

	正解	誤選択	正解なし	出力なし	正解率
POPULATION	109	51	32	81	39.9
文脈参照法	178	39	32	24	65.2
MENTION_COUNT	137	104	32	0	50.2
文脈参照法+POPULATION	196	42	32	3	71.8
文脈参照法+MENTION_COUNT	194	47	32	0	71.1

表 3: 5 手法の正解率（大字レベル・市区町村レベル）

	大字レベル					市区町村レベル				
	正解	誤選択	正解なし	出力なし	正解率	正解	誤選択	正解なし	出力なし	正解率
POPULATION	0	5	0	53	0	109	46	32	28	50.7
文脈参照法	42	13	0	3	72.4	136	26	32	21	63.3
MENTION_COUNT	30	28	0	0	51.7	107	76	32	0	49.8
文脈参照法+POPULATION	42	13	0	3	72.4	154	29	32	0	71.6
文脈参照法+MENTION_COUNT	43	15	0	0	74.1	151	32	32	0	70.2

日本では昨年２月に大阪市で初めて誕生、以後神戸、横浜、北九州、金沢、富山県高岡、長野県上田の各市と東京都で設立された。

横浜 高知県_高知市_横浜 福岡県_福岡市_西区_横浜 **神奈川県_横浜市**

以下に，言及回数 DB の横浜に対する候補を記述する．神奈川県_横浜市が言及回数が一番多いため選ばれたとわかる．

神奈川県_横浜市 175
福岡県_福岡市_西区_横浜 21
高知県_高知市_横浜 11

6. おわりに

本論文では地名に対する地理的位置の同定を行う際に発生する地名の曖昧性問題について，文脈参照法と MENTION_COUNT の提案を行った．先行研究の手法では地名に対応する人口 DB が必要であったが，提案手法ではこれらよりも入手が簡単で前処理も簡単な関連文書，具体的には新聞の生コーパスを用いることで曖昧性の解消を行った．実験により先行手法 POPULATION の正解率が 40%程度に対し提案手法である文脈参照法では正解率が 65%，文脈参照法+MENTION_COUNT では精度が 71% 程度になることがわかった．

今後の課題と検討として，次のようなものが挙げられる．

- 対象とするコーパスの拡大
- 言及回数 DB 作成において，都道府県レベルを市区町村レベルにまで範囲を小さくしたり扱うデータ（生コーパス）量をより大きくすることで言及回数 DB の精度を高める
- コーパスと住所 DB のサンプル年度の整合をとる
- 他ジャンルのコーパスでの実験

参考文献

[1] Stuart Middleton,Lee Middleton,and Stefano Modafferi.Real-time crisis mapping of natural disasters using social media.2014.

[2] 北本朝展.G 空間 オープンな地名情報システム GeoNLP.THE JOURNAL OF SURVEY 測量,pp6-10,2014.

[3] Michael Adrian Speriosu.Methods and Applications of Text-Driven Toponym Resolution with Indirect Supervision.THE UNIVERSITY OF TEXAS AT AUSTIN,2013.

[4] 橋本 泰一, 乾 孝司, 村上 浩司. 拡張固有表現タグ付きコーパスの構築. 情報処理学会研究報告自然言語処理,pp113-120,2008.

[5] 毎日新聞社.CD-毎日新聞 2004 データ集.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

イベント知識抽出のための畳み込みニューラルネットワーク

Convolutional Neural Network for Event Knowledge Extraction

石川 開[*1]
Kai Ishikawa

高村 大也[*2]
Hiroya Takamura

奥村 学[*2]
Manabu Okumura

[*1] NEC データサイエンス研究所
Data Science Research Laboratories, NEC Corporation

[*2] 東京工業大学
Tokyo Institute of Technology

1. はじめに

今日、高度な知識処理を備えた対話システムは、スマートフォンやスマートスピーカーなどの日常的なデバイスを通じて我々の生活に浸透し、身近なツールとなりつつある。また、社会や産業の高度化に伴って、ますます増大しつつある知識労働者の負担を、知識処理システムで支援するという試みも、医療、法務などの様々な領域で活発化している。このように、AI による社会変革への関心の高まりと共に、実社会課題への適用が進み始めた知識処理システムではあるが、その実現には、システムの適用領域や目的に応じた膨大な知識が必要となる。この知識の開発・メンテナンスコストが、これまで知識処理システムを実用化する上での大きな障壁となってきた。知識処理システムの社会適用を本格的化する上で、知識開発を自動化するブレークスルーが必要となる。それにより、例えば、知識処理システム自体が、運用の過程でシステムに不足する知識を自己診断し、自律的に文書群や専門家などの情報源にアクセス、システムの知識の網羅性、信頼性、鮮度を維持する、といったアプローチが現実的なものとなる。

本研究では、このような知識開発の自動化の導入として、シェアードタスク NIST TAC KBP2016 Event Nugget Detection [Mitamura 2016] を対象に、ベースラインとなる知識抽出システムを構築し、改善方法を提案する。同タスクでは、課題となるイベントに対し、ラベル付きの事例が与えられるが、事例に含まれない語彙や表現も正しく抽出しなければならない。このため、教師データの限られた語彙や表現から、汎用性の高いモデルを構築する必要がある。そこで我々は、ベースラインに文分類のための畳み込みニューラルネットワーク [Kim 2014] を選択する。分散表現 [Mikolov 2013] の適用は、大規模離散的な語彙空間を低次元の連続的な意味空間に写像するため、事例の表層的な違いを吸収する効果が期待できる。また、畳み込みニューラルワーク(CNN)については、学習データに含まれない単語列の意味を、分散表現の畳み込みで構成、推定できる様な効果を期待している。並列処理向きで GPU による高速化の効果が大きいという利点もある。一方、CNN の固定長入力、近接する入力要素の階層的な畳み込みといった特性が、知識抽出タスクでは次のような問題を生じる。

(1) 短いフレーズでイベントが確定する事例では、入力中の周辺語がノイズとなって精度が低下する。一方、入力を短くすると、文脈を考慮する必要のある事例が正しく検出できない。

(2) 近接する入力の畳み込みを階層的に繰り返す CNN では、構成要素の離れた複単語表現(MWE)の学習が難しい。

以降、上記の問題に対する改善方法を提案、評価実験でベースラインと比較、実証する。また、NIST TAC KBP2016 の参加システムの公式評価の比較で、提案手法が KBP2016 の全ての参加ステムの性能を上回る事を示す。

2. イベント知識抽出システムの概要

本研究では、NIST TAC KBP2016 の Event Nugget Detection に対する評価システムを構築し、提案手法の評価実験を行う。Event Nugget Detection では、システムは定義された全 18 種類のイベントタイプのいずれかに該当するトークン列をテキスト中から検出する。この際、トークン列のテキスト中の位置や事実性の分類を含めたシステム全体での精度が評価される。

本研究の知識抽出システムの処理フローを図1に示す。本システムでは、入力文書に簡易なトークナイザー [Kim 2014] を適用してトークン列を生成する。

次に、イベントトークン検出では、各文書のトークン列から、全 18 種類のイベントタイプのいずれかに該当するトークンを検出する。さらに、検出されたイベントトークンに対し、イベントタイプ分類と、事実性分類が適用される。イベントタイプ分類は、検出されたイベントトークンのタイプを一意に特定し、事実性分類は、イベントトークンの事実性の状態が、事実(actual)、一般(general)、その他(others)のいずれに該当するかを特定する。.

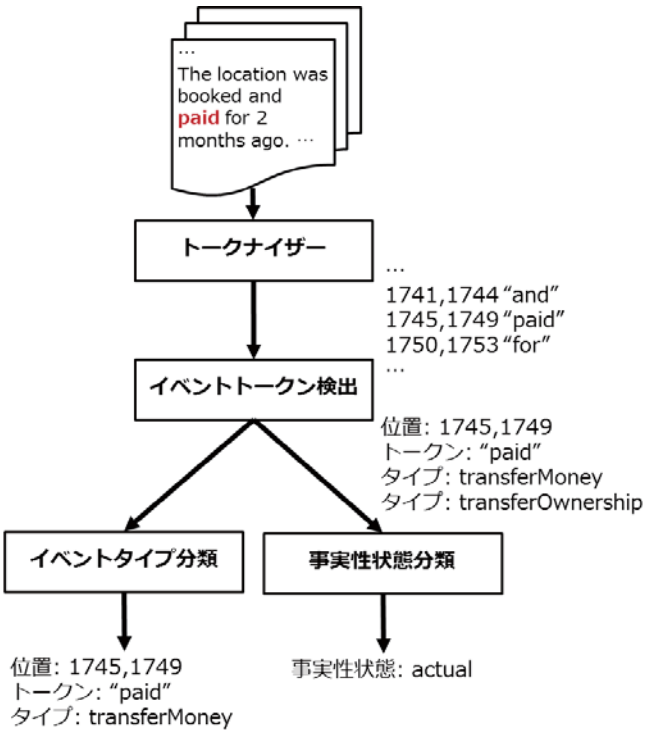


図１： 本研究のイベント知識抽出システムの処理フロー

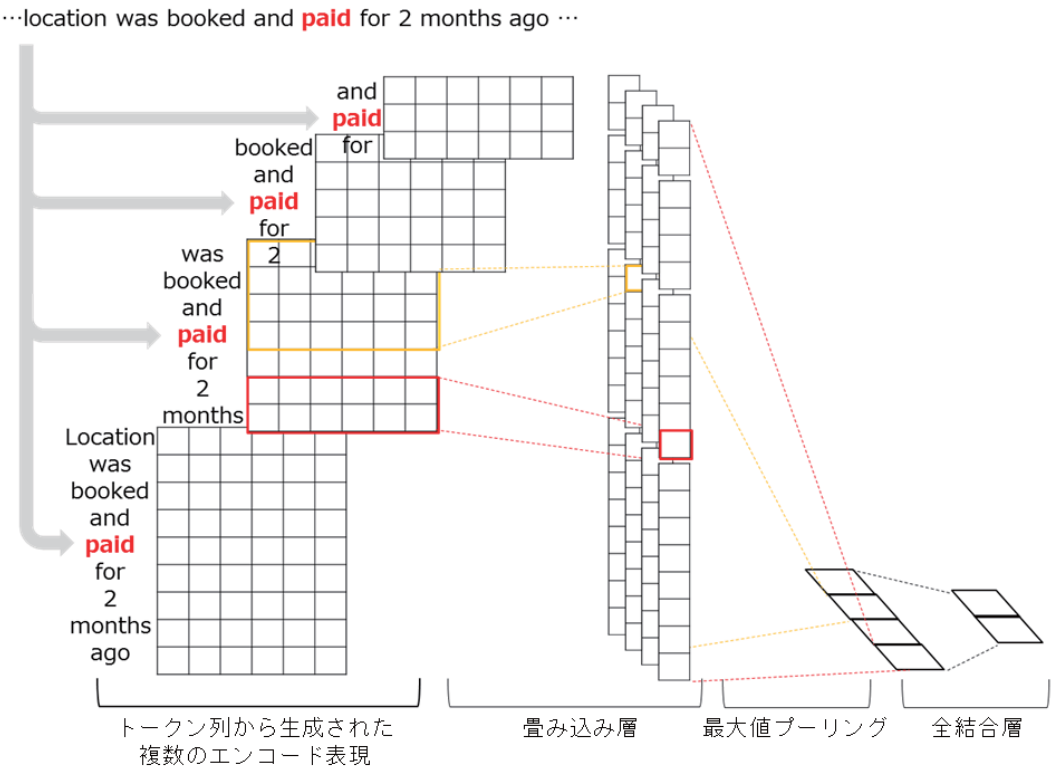


図2： MW 法におけるトークン列からの入力表現の生成例とニューラルネットワークの構成

3. イベントトークン検出モデル

3.1 ベースライン： Single Window (SW) 法

前述のように、イベント検知では、各文書から生成されたトークン列中の各トークンに対し、全 18 種類の各イベントタイプに該当するかどうかの二値分類を行う。本研究では、この二値分類のベースラインとして、Collobert 等のモデル [Collobert 2011] の改良版である文分類のための畳み込みニューラルネットワーク [Kim 2014] を用いる。

この分類手法では、各トークンのエンコーディングに、の word2vec [Mikolov 2013] によって得られる 300 次元の単語ベクトル表現を用いる。そして、分類対象のコンテキストの長さを制限するために、固定長の窓を導入する。入力文がこの窓よりも長い場合は窓の長さにトリミングし、この窓の長さよりも短い場合は、パディング（padding）により入力を窓の長さに一致させる。この文分類の手法を元に、窓の中心にあるトークンを検出の対象トークン、窓内の周辺トークンを、対象トークンの文脈と読み替えれば、イベントトークンの検出に適用することが出来る。以降、この手法をベースライン手法とし、Single Window (SW) 法と呼ぶ。

SW 法において、ある対象トークンのニューラルネットワークへの入力は、対象トークンと、前後に長さ l のトークン列を加えた、式(1)に示すような長さ 2l+1 のトークン列に対する単語ベクトル表現である。

$$x_{-l:l} = x_{-l} \oplus \ldots \oplus x_0 \oplus \ldots \oplus x_l \qquad (1)$$

ここで、x_iは窓中の i 番目の単語に対する k 次元の単語ベクトル、⊕は結合（concatenation）オペレータである。

SW 法は、原理的には窓を大きく設定することで長い文脈を考慮できるが、実際には冒頭で述べた通り、次のような問題が存在する。

・問題点1： 窓中の各トークンが同じ重みでネットワークに入力されるため、窓を大きくすると、表現によってはイベント検知と直接関係の無い入力情報が増え、窓の拡大によって全体的な検知精度が低下する。

・問題点2： 隣接トークン対に比べて離れたトークン対の意味の構成性は CNN では学習が難しく、非隣接の出現パタンで、イベント検知に有用な複単語表現(MWE)などが上手く学習されない可能性が生じる。

以降、これらの問題に対する改善手法の提案を行う。

3.2 提案1： Multiple Window (MW) 法

第1の提案は、ベースラインの問題点1の解消を目的とした Multi Window (MW) 法の導入である。MW 法は、対象トークンを共通の中心とし、大きさが段階的に異なる窓を複数用意し、各窓に対するエンコードを生成、多チャネルでネットワークに入力する。具体的には、式(2)に示すような、長さ 2l+1, 2l-1, …, 1 の l + 1 個の窓に対するエンコードの集合として表される。

$$\{x_{-l:l}, x_{-l+1:l-1}, \ldots, x_{-1:1}, x_0\}. \qquad (2)$$

図2は、MW 法におけるトークン列からエンコード表現を生成し、ネットワークに入力する全体構成を示している。図に示すように、式(2)で得られた l+1 個のエンコードを並列でネットワークの畳み込み層（入力層）に入力する。畳み込み層の出力は、最大値プーリングで圧縮後、全結合層に入力される。ソフトマックスにより、最終的に、統合された予測結果（二値分類の予測結果と予測確率）が得られる。

3.3 提案2: Pseudo Multiword Expression (PME) 法

第2の提案は、ベースラインの問題点2の解消を目的とした Pseudo Multiword Expression (PME) 法の導入である。PME 法は、対象トークンと、そこから一定距離内にある全てのトークンとの対を、擬似的に複数語表現(multiword expression)の構成要素とみなし、各トークン対に対するエンコードを生成、多チャネルでネットワークに入力する。対象トークンから距離 l 以内のトークンから得られるトークン対を用いて、対象トークンに対して式(3)のような 2l 個のエンコードが得られる。

$$\{x_{-l} \oplus x_0, \dots, x_{-1} \oplus x_0, \dots, x_0 \oplus x_l\}. \quad (3)$$

この 2l 個のエンコードを、MW 法で用いたのと同じネットワークに並列で入力することによって、最終的に、統合された予測結果(二値分類の予測結果と予測確率)が得られる。

3.4 イベント検知モデルの動的選択

イベント検知モデルは、窓サイズや用いるエンコード表現(SW 法、MW 法、PME 法)によって検知特性が異なる。システムの全体性能を向上するためには、対象トークンごとに相性の良いモデルが選択され、全体としてシステムの検知精度が向上することが望ましい。このような考えから、本研究では、CNN から出力される予測確率を確信度のように見なすことで、対象トークンごとに動的にモデルを選択する方法用いる。具体的には、式(4)に示すように、モデル(model)がトークン w をイベントタイプ(event type)として検知する際、予測確率 $P_{model}(event\ type|w)$ が最大となるモデルを選択し、その予測結果を、トークン w に対する複数モデルの統合的な予測結果として用いる。

$$\begin{aligned} &model \\ &= \underset{model}{\operatorname{argmax}}\{\max[P_{model}(event\ type|w), \\ &1 - P_{model}(event\ type|w)]\} \end{aligned} \quad (4)$$

4. 分類モデル

4.1 イベントタイプ分類

イベントタイプ分類では、前段の各イベントタイプ毎の検出結果の和集合に対し、複数のイベントタイプに判定されたトークンの再分類を行う事で曖昧性を解消し、タイプを一意に特定する。分類モデルには前述の SW 法、MW 法、PME 法を用いる。学習するデータセットは、イベントトークン検出においては、いずれのイベントタイプにも該当しないトークンを負例として学習するが、イベントタイプ分類においては、イベントとして検出されたトークンの分類が目的なので、学習対象から外す。得られた複数モデルの各イベントタイプに対する予測結果、予測確率を元に、式(5)を用いてイベントタイプの推定を行う。

$$\begin{aligned} &event\ type \\ &= \underset{event\ type}{\operatorname{argmax}} \left\{ \max_{model} P_{model}(event\ type|w) \right\} \end{aligned} \quad (5)$$

4.2 事実性状態(realis status)分類

事実性状態分類においても、イベントタイプ分類と同様、前段のイベントトークン検出の結果を元に、事実性状態を一意に特定する。分類モデルはここでも SW 法、MW 法、PME 法を用いる。ここでも、イベントとして検出されたトークンの分類が目的なので、訓練セットから、いずれのイベントタイプにも該当しないトークンを外して学習する。得られた複数モデルの各事実性状態に対する予測結果、予測確率を元に、式(6)を用いて事実性状態の推定を行う。

$$\begin{aligned} &realis \\ &= \underset{realis}{\operatorname{argmax}} \left\{ \max_{model} P_{model}(realis|w) \right\} \end{aligned} \quad (6)$$

5. 評価実験

5.1 データセットと学習条件

本評価実験では、NIST TAC KBP2016 の Event Nugget Detection (English)におけるシステムの公式の評価データセット・評価ツールを用いて評価セットを行った。また、訓練データには、KBP の訓練評価データをまとめた"TAC KBP Event Nugget Detection and Coreference - Comprehensive Training and Evaluation Data (LDC2017E02)"から KBP2016 の evaluation data を除いたデータと、"DEFT Rich ERE English Training Annotation (LDC2016E31, LDC2015E68, LDC2015E29)"を用いた。

畳み込みニューラルネットワークのハイパーパラメーターはベースライン、提案手法の全てのモデルで共通のものを用いた。具体的には、Zhang によるパラメータ設定の分析[Zhang 2017]を参考に、フィルタサイズを[2, 3, 4, 5]、特徴マップの次元数を 100、ドロップアウト率を 0.5、ミニバッチサイズを 50 とした。学習には、Zeiler の Adadelta による更新ルールとミニバッチのシャッフルを加えた確率的最急降下法[Zeiler 2012]を適用、10fold の交差検定による F1 値に基づいてアーリーストッピングを実施した。

単語分散表現モデルには、Google News に対する公開モデルである GoogleNews-vectors-negative300 を用い[Mikolov 2013]、このモデルでカバーされない語についてはランダムな初期値を与え、逆伝播による word2vec のチューニングを実施した[Kim 2014]。

5.2 イベントトークン検出の評価

イベントトークン検出における、ベースラインと提案手法の各モデルの適合率、再現率、F1(マクロ平均)を表1に示す。以降の表では、手法とパラメータ l(式(1)、(2)、(3)で定義)を、次のように表記する。SW(1)は、SW 法で l=1 としたモデル、SW(1,3,5)は、SW 法で l=1, 3, 5 とした 3 つのモデルを動的モデル選択で統合したモデルを表す。PME(3,5)+MW(0,3,5)は、PME 法で l=3, 5、MW 法で l=0, 3, 5 とした計5つのモデルを動的モデル選択で統合したモデルを表す。

表1: ベースラインと提案手法のイベントトークン検出の性能

モデル(窓サイズ)	適合率	再現率	F1
SW(0)	49.5	50.0	49.8
SW(3)	43.8	33.6	38.1
SW(5)	16.1	13.3	14.6
SW(0,3,5)	49.9	38.1	43.2
MW(0)	49.5	50.0	49.8
MW(3)	53.4	45.5	49.1
MW(5)	51.2	47.1	49.1
PME(3)	52.0	48.7	50.3
PME(5)	50.4	51.1	50.8
PME(3,5)+MW(0,3,5)	**53.0**	**53.8**	**53.4**

表 1 から、SW 法では、周辺語を考慮しない SW(0)が最も F1 が高いことから、問題点1が実際に確認できる。PME 法では、最

も離れた周辺語を考慮する PME(5)が最も F1 が高く、SW(0)を上回ることから、PME 法が離れた周辺語の考慮で精度を向上している事が確認できる。MW 法は適合率で優れており、PM 法と MW 法のモデルの統合が、全モデル中最良の F1 を得た。

5.3 イベントタイプ分類の評価結果

続いて、イベントタイプ分類におけるベースラインと提案手法の各モデルの適合率、再現率、F1 を表2に示す。SW 法の入力には SW(0)で検出されたトークン、MW 法と PME 法の入力には、PME(3,5)+MW(0,3,5)で検出されたトークを用いた。ベースライン中、提案手法中で、モデル間の性能差が少ないが、これは、トークン検出で生じたイベントタイプの曖昧性が少ない事を意味する。わずかであるが、PME 法の優位性が確認できる。

表2: ベースラインと提案手法のイベントタイプ分類の性能

モデル(窓サイズ)	適合率	再現率	F1
SW(3)	**41.7**	**41.3**	**41.5**
SW(5)	41.6	41.2	41.4
SW(3,5)	41.7	41.3	41.5
MW(3)	45.2	45.8	45.5
MW(5)	45.4	46.0	45.7
PME(3)	45.5	46.1	45.8
PME(5)	**45.5**	**46.1**	**45.8**
PME(3,5)+MW(3,5)	45.5	46.1	45.8

5.4 事実性状態分類の評価結果

事実性状態分類におけるベースラインと提案手法の各モデルの適合率、再現率、F1 を表3に示す。SW 法の入力には SW(0)で検出されたトークン、MW 法と PME 法の入力には、PME(3,5)+MW(0,3,5)で検出されたトークを用いた。ベースライン中では SW(3,5)が、提案手法中では PME(5)+MW(3,5)が最も F1 が高く、いずれも単一のモデルよりも、動的モデル選択による統合モデルが優れた結果を得た。このことから、異なるモデルが同等の性能水準である場合に、動的モデル選択による統合が有効に働くと考えられる。イベントトークン検出(表1)において、SW(0,3,5)が SW(0)を下回り、動的モデル選択が上手く機能しなかったのは、SW(0)と SW(3)、SW(5)の間の性能差が大きく開いていたためと考えられる。

表3: ベースラインと提案手法の事実性状態分類の性能

モデル(窓サイズ)	適合率	再現率	F1
SW(3)	37.0	37.5	37.2
SW(5)	36.0	36.6	36.3
SW(3,5)	**37.5**	**38.1**	**37.8**
MW(3)	39.9	40.4	40.1
MW(5)	40.5	40.8	40.7
PME(3)	40.4	40.9	40.6
PME(5)	40.7	41.2	41.0
PME(5)+MW(3,5)	**41.7**	**42.2**	**41.9**

5.5 タスク全体性能の評価結果

最後に、提案手法、ベースラインのタスク全体での性能水準を見るため、NIST TAC KBP2016 の Event Nugget Detection (English)の公式評価結の上位 10 チームの性能と比較した(表4)。その結果、MW 法と PME 法の動的モデル選択による提案手法 MW+PME が最も高い F1 値 35.85%となり、KBP2016 でトップの NYU1[Nguyen 2016]を上回ることが分かった。この事から、提案手法は、知識抽出システムとして現時点で世界トップレベルの性能水準にある事が示された。

表4: ベースライン、提案手法、KBP2016 参加システム上位10チームの Event Nugget Detection タスクの全体性能

参加システム/提案手法	適合率	再現率	F1
MW+PME (proposed)	**35.67**	**36.03**	**35.85**
NYU1	33.47	37.21	35.24
UTD1	34.05	33.12	33.58
wip3	38.38	29.1	33.1
SoochowNLP3	37.26	28.97	32.59
SW (baseline)	**31.13**	**30.85**	**30.99**
RPI-BLENDER2	31.92	28.59	30.16
LTI-CMU1	40.19	22.76	29.06
SYDNEY1	35.93	23.45	28.38
aipheshd-t161	29.94	23.81	26.53
Washington1	30.71	21.42	25.24
UMBC2	25.64	18.76	21.67

6. まとめ

本稿では、文書からの知識抽出の方式として、畳み込みニューラルネットワークを用いた知識抽出システムを開発し、NIST TAC KBP2016 Event Nugget Detection への適用を試みた。文書分類向けに開発されたベースライン SW 法における、固定長単一窓が不要な周辺語によるノイズを招く問題、構成要素が離れて出現する複単語表現(MWE)と畳み込みの相性問題への対処として、それぞれ MW 法、PME 法を提案した。KBP2016 の公式の評価セット・評価ツールを用いて評価を行った結果、提案手法(MW 法と PME 法の動的モデル選択による統合方式)は、NIST TAC KBP2016 の全参加チームを上回る F1 値 35.85%を得た。

今後、正例が極端に少ないイベントに対しても安定的に高品質なモデルを獲得することを課題に、さらなる改善を進めていく。

参考文献

[Kim 2014] Yoon Kim: Convolutional Neural Net-work for Sentence Classification, In Proceedings of EMNLP, 2014.

[Zhang 2017] Ye Zhang, Byron Wallace: A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification, IJCNLP, 2017.

[Mikolov 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean: Efficient estimation of word representations in vector space, In ICLR, 2013.

[Mitamura 2016] Teruko Mitamura, Zhengzhong Liu, and Eduard Hovy: Overview of TAC-KBP2016 Event Nugget Track, TAC, 2016.

[Nguyen 2016] Thien Huu Nguyen, Adam Meyers, and Ralph Grishman: New York University 2016 System for KBP Event Nugget: A Deep Learning Approach, TAC, 2016.

[Zeiler 2012] M. Zeiler: Adadelta: An adaptive learning rate method, CoRR, abs/1212.5701, 2012.

[Collobert 2011] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuglu, P. Kuksa: Natural Lan-guage Processing (Almost) from Scratch , Journal of Machine Learning Research12, 2011.

特許文書を対象とした因果関係抽出に基づく発明の新規用途探索

Search for New Application of Invention based on Extraction of Causality from Patents

太田 貴久[*1] Takahisa Ota　南 拓也[*1] Takuya Minami　山崎 祐介[*1] Yusuke Yamazaki　奥野 好成[*1] Yoshishige Okuno

田辺 千夏[*1] Chinatsu Tanabe　酒井 浩之[*2] Hiroyuki Sakai　坂地 泰紀[*3] Hiroki Sakaji

[*1] 昭和電工株式会社 Showa Denko K.K.　[*2] 成蹊大学 Seikei Univ.　[*3] 東京大学 The Univ. of Tokyo

In this research, we propose a method for searching for new application of invention. The proposed method consists of two steps. The first step is to extract technical features and their effects from patents. The second step is to search for patents with similar technical features and dissimilar effect to a specific query. This method was demonstrated to be useful in searching for new application of invention.

1. はじめに

特許は，発明の公開を代償に，発明者へ独占排他権を付与する産業上重要な権利である．その際に公開される特許文書には，権利の範囲だけでなく，具体的な実現方法等の様々な技術的な情報が含まれている．特許法の目的でもある産業の発達を目指すためには，これらの情報を積極的に活用する必要がある．

一方，近年，知財を積極的に活用する取り組みの 1 つに「IP ランドスケープ」と呼ばれる概念がある[特許庁 2017]．IP ランドスケープとは，従来の事業保護を目的とした特許分析とは異なり，知財による事業拡大を目的とした分析，すなわち，将来を見据えた事業に資する知財分析を意味する．

このような背景のもと，本研究では，最も典型的な知財であり，かつ大量の情報を含む特許を用いて，事業拡大を目的とした分析を支援するための手法を提案する．具体的には，既存技術の新たな展開先（すなわち，既存発明の新しい用途）を探索する手法を提案する．

2. 従来研究

従来，特許文書から手段（原因）と効果（結果）の対を含む因果関係文を，コンピュータを用いて抽出する技術が知られている[日本特許 2011]．しかしながら，本特許[日本特許 2011] は，特許文書から手段と効果を抽出することにとどまっており，抽出した情報（因果関係文）の活用方法については触れられていない．

また，前記特許の技術を利用した研究として Nonaka らが提案するパテントマップ自動生成システムがある[Nonaka 2010]．本研究は，特許文書から抽出した手段と効果を用いて，技術動向を把握するための図（パテントマップ）を自動的に生成する研究である．しかしながら，本研究によって得られるパテントマップは，指定した技術分野（特許集合）の範囲で生成されるものである．このようなパテントマップは，「ある分野における既知の効果に対して，未知の手段を把握する」ことには役立つが，技術分野の枠を超えた，事業拡大を目指した分析を行えるものではない．

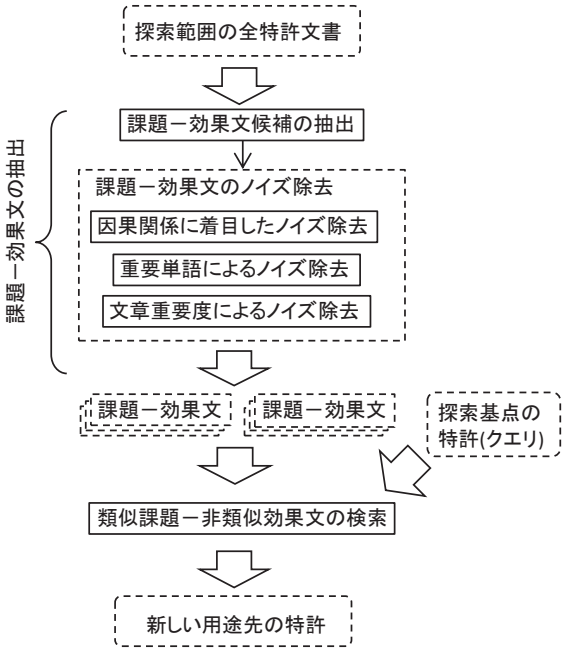


図 1: 提案手法の概要

また，特許文書から手段と効果の因果関係を抽出し，手段が類似の因果関係は同様の効果を有するであろうという推測をコンピュータで行う研究が知られている[石川 2007]．しかしながら，石川らの研究[石川 2007]は，過去の特許文書に記載されている手段と効果の因果関係の範囲で，技術者の発想支援を行うものであり，事業に資する提案するようなものではない．

これらの研究に対して，本研究では，既存の技術が利用できる新たな技術分野（用途先）の発見を支援する．これにより，技術分野の枠を超えた事業に資する知財分析を提供することができる．

3. 提案手法

図 1 に本研究で提案する用途探索手法の概要を示す．

連絡先：太田貴久，昭和電工株式会社 法務・知的財産部，東京都港区芝大門 1-13-9，(TEL) 03-5470-3256，(FAX) 03-3433-0960，ota.takahisa.xiuec@showadenko.com

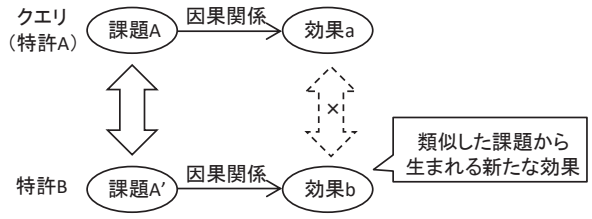


図 2: 類似課題－非類似効果

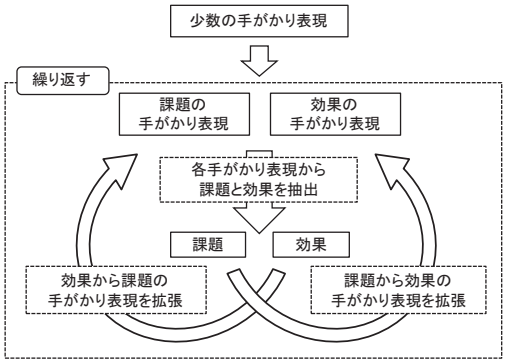


図 3: 坂地らの手法 [坂地 2010] の概要

図 1 のように，提案手法は，大きく分けると，「課題－効果文の抽出」と「類似課題－非類似効果文探索」の 2 つのステップに分けられる．

提案手法の最初のステップは，探索範囲のすべての特許文書（例えば，特定年代のすべての公開特許公報）から，課題－効果文の抽出を行う．ここで，課題－効果文とは，発明における最終的な因果関係を表す文であり，発明が達成する効果（以下，単に効果）と，その効果を達成するために解決する必要のあった技術課題（以下，単に課題）の組で表される情報である（従来研究では、課題を「手段」と呼んでいる）．例えば，「○○層を形成するため，熱伝導率を低減できる」という文が課題－効果文である．ここで，「○○層を形成するため」が課題を表し，「熱伝導率を低減できる」が効果を表す．効果の原因となる課題を抽出することで，単なる発明の構成要素ではなく，その発明の技術的な特徴を捉えることが可能となる．

課題－効果文の抽出ステップでは，課題－効果文の候補を抽出した後，3 種類のノイズ除去手法を適用することで，精度（適合率）の高い課題－効果文を抽出する．本研究は，新規用途探索を行う際に，分析者に新たな気付きを与えることを目的としており，技術を俯瞰することが目的ではない．そのため，すべての特許から漏れなく課題－効果文を抽出できなくても問題ない．すなわち，本研究の課題－効果文抽出においては，再現率より適合率（精度）の方が重要である．そこで，提案手法では，再現率を犠牲にしてもノイズとなるような課題－効果文を除去するようにしている．

提案手法の 2 つ目のステップは，類似課題－非類似効果文探索ステップである．本ステップは，探索の基点とする既存の課題－効果（以下，クエリと呼ぶ）をもとに，探索範囲の特許集合から新しい用途先（の候補）となる特許を探索するステップである．なお，クエリとする課題－効果は，課題と効果を表すようなテキストであればどのような指定を行ってもよい（必ずしも特許の文章を用いなくとも良い）．

ここで，探索するべき用途先の特許について考える．例えば，自社特許の新しい用途先を探索する場合，低コストで参入可能な用途先とは，自社の技術的特徴をそのまま応用できる技術分野である．すなわち，前ステップで抽出した課題－効果文において，課題が類似している技術分野を探索できることが好ましい．一方，誰もが容易に想到する分野は好ましくない（特許文書の場合，公開までにタイムラグがあるため，すでに他社に取り組まれている可能性がある）．すなわち，前ステップで抽出した課題－効果文において，類似しない（意外性の高い）効果を持つ技術分野が好ましい（図 2）．

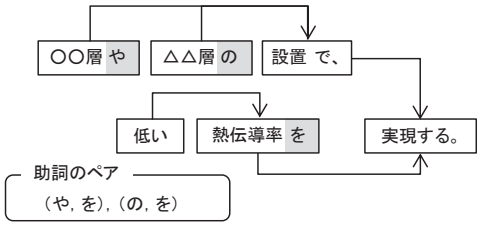


図 4: 坂地らの手法 [坂地 2011] の概要

以上の理由により，提案手法の 2 つ目のステップでは，クエリと「類似した課題」を持ち，かつ「類似していない効果」を持つ特許を，探索範囲の特許集合から検索する．

提案手法は，以上の 2 ステップによって，類似している課題を持つ特許において，課題と対応している効果が類似していない特許を抽出する．これにより，クエリとした技術の新しい用途先を探索することが可能となる．以下では，各ステップの詳細について説明する．

3.1 課題－効果文候補の抽出

課題－効果文の抽出ステップの最初の処理では，探索範囲のすべての特許文書から，課題－効果文の候補を抽出する．本研究では，課題－効果文候補の抽出手法として，坂地らの手法（Cross-Bootstrapping）[坂地 2010] を用いた（以下，本手法を，坂地の手法(2010)と呼ぶ）．

坂地らの手法(2010)では，特許明細書の【発明の効果】は，課題と効果が同じ文に出現することが多いことを利用した手法である．坂地らの手法(2010)では，「手がかり表現」と呼ぶ，課題および効果を現す表現の周辺に出現する特徴的な表現に着目する．例えば，「本発明は、無効スペースとなる冷凍サイクル関連機器の配置構成を工夫することで、使用者が使いやすい位置の収納容積や収納性を高めることができる。」という文では，前の下線部が課題を，後の下線部が効果を現している．本例では，課題の直後の「ことで、」と，効果の直後の「ことができる。」が手がかり表現にあたる．坂地らの手法(2010)では，このような手がかり表現から課題－効果文を抽出する．簡単に手法を説明すると，坂地らの手法(2010)では，「手がかり表現から課題－効果を抽出」するステップと「課題－効果から手がかり表現を拡張」するステップを交互に繰り返すブートストラップ手法である（図 3）．

3.2 因果関係に着目したノイズ除去

前の処理で抽出した課題－効果文の候補には多くのノイズが含まれる．例えば，坂地らの手法 (2010) では，手がかり表現として「（課題）であり、（効果）である。」というパターンを用いて課題－効果文が抽出されることがある（この手がかり表現は，一見，完全な誤りに見えるが，「本発明の特徴は、○○層を設けたことであり、これにより，熱伝導率を低減できた点である。」という文も存在するため，完全な誤りとは言えない）．提案手法では，このようなノイズとなる課題－効果文を 3 つの方法で除去する．

1 つ目のノイズ除去の方法は，助詞のペアを用いたノイズ除去である．具体的には，坂地らの手法[坂地 2011]を用いる（以下，本手法を，坂地の手法 (2011) と呼ぶ）．

坂地らの手法(2011)では，課題末尾の文節と効果末尾の文節のそれぞれに直接係る文節の助詞を抽出し，そのペアを基に因果関係を含む文か否かを判定する手法である．例えば，図 4 のように「○○層や△△層の設置で、低い熱伝導率を実現する。」という文の場合，課題末尾の文節である「設置で」に係る「○○層や」と「△△層の」と，効果末尾の文節である「実現する。」に係る「熱伝導率を」において，（課題側の文節の助詞，効果側の文節の助詞）のすべてのペアを用いて因果関係を判定する．この場合，（や，を）と（の，を）が因果関係の判定に用いられる．坂地らの手法(2011)では，事前に因果関係を含むと人手で判定した文の助詞のペアを教師データとして，SVM 用いて因果関係の判定を行っている．

3.3 重要単語に基づくノイズ除去

2 つ目のノイズ除去の方法は，重要単語に基づくノイズ除去である．最初のノイズ除去では，因果関係の有無に着目してノイズ除去を行った．しかしながら，因果関係の有無のみではノイズ除去としては不十分である．本研究において，理想的な課題－効果文とは，課題が，特別な技術的特徴 (STF: Special Technical Feature) を表している文である．ここで，特別な技術的特徴とは，特許の核となる「発明の先行技術に対する貢献を明示する技術的特徴」[特許庁 2016] を表す．

このような特別な技術的特徴を含む課題を抽出するために，2 つ目のノイズ除去では，特徴的な単語（キーワード）を課題に含まない課題－効果文候補はノイズとして除去する．なお，本研究において，特徴的な単語とは，しきい値以上の TF-IDF を持つ単語である．ここで，IDF は，提案手法の入力である探索範囲の全特許文書より算出する．

3.4 文章重要度に基づくノイズ除去

最後のノイズ除去の方法は，文章重要度に基づくノイズ除去である．本ノイズ除去では，一般的な表現や具体性に欠ける表現等を候補から取り除くことを目的とする．

例えば，典型的な効果の手がかり表現である「できる。」を考える．多くの場合，「○○できる。」は正しく発明の効果を表すものであるが，「特徴づけることができる。」といった表現も抽出される．このような表現は特許文書において一般的に用いられる表現であるが，具体的な効果を表すものではない．また，課題においても，「このような材料を用いることにより、」といった照応表現を用いて具体的な技術的内容を含まない不適切な課題が抽出されることがある．

以上のような効果や課題の共通の特徴は，ともに一般的な語で表現が構成される点にある．そこで，最後のノイズ除去では，課題－効果文の候補のうち，一般的な語のみで構成されるような文を除去する．具体的には，課題－効果文候補に含まれる語の TF-IDF の合計値を文章重要度とし，文章重要度がしきい値以下の場合に，当該文はノイズであると判定する．

上記のように，課題－効果文の抽出ステップでは，課題－効果文の候補を抽出した後，さらに 3 種類のノイズ除去手法を適用することで，探索範囲の各特許から課題－効果文を抽出する．

3.5 類似課題－非類似効果文の検索

次に，実際に新しい用途先となる特許を探す，類似課題－非類似効果文探索ステップについて説明する．本ステップは，前ステップによって抽出された課題－効果文から，クエリ（探索基準の課題－効果）と比較して，課題が類似し，かつ効果が類似していない課題－効果文を検索する処理である．

ここで，検索の対象となる課題－効果文は，元の特許文書と比較すると非常に短い．そのため，課題や効果の類似性を単純な one-hot 表現を用いて求めようとすると，多くの場合で類似していないと判定されてしまい，正しく類似課題－非類似効果文を検索することができない．

そこで，本研究では，課題－効果文を構成する語の分散表現を求め，課題（および効果）を構成する語の分散表現の和（以下，表現ベクトルと呼ぶ）を比較することで課題（および効果）の類似／非類似を判定する．具体的には，本研究では，word2vec[1] によって語の分散表現を求める．また，表現ベクトルの類似性判定は，コサイン類似度がしきい値を超えたか否かによって判定を行う．

なお，本研究では，コサイン類似度による比較を行う前に，高 TF-IDF のキーワードによるフィルタリングを行い，類似度を算出する課題－効果文を絞り込んでいる．

以上の方法によって，提案手法は，探索範囲の課題－効果文から，クエリと類似課題－非類似効果を持つ文を抽出し，当該文を含む特許を出力する．すなわち，類似している課題を持つ特許において，課題と対応している効果が類似していない特許を抽出する．

4. 実験

本研究では提案手法について，2 つの実験を行った．1 つ目の実験は，実際に新しい発明の用途が探索可能であるか否かを確認する実験（以下，新規用途探索実験と呼ぶ）である．2 つ目の実験は，過去，実際に他の用途へ展開した発明をクエリとして，提案手法によって，その展開した用途が検出できるか否かを確認する実験（以下，他用途展開再現実験と呼ぶ）である．

なお，今回の実験では，探索対象の特許集合や各種パラメータは同一のものを用い，クエリ（探索基点の課題－効果）のみ別のものを用いた．両実験で共通のデータやパラメータを表 1 に示す．表 1 において，初期手がかり表現とは，坂地らの手法(2010)で最初に与える手がかり表現である．また，助詞 SVM の教師データとは，坂地らの手法(2011)で因果関係の有無を判定する SVM を学習するために用いた文である．また，重要単語のしきい値とは，重要単語によるノイズ除去において，単語がキーワードであるか否かを判定するためのしきい値である．また，文章重要度のしきい値とは，文章重要度によるノイズ除去において，課題－効果文がノイズか否かを判定する合計 TF-IDF のしきい値である．なお，本研究では，形態素解析器は MeCab[2]を（辞書は NEologd[3]を使用），構文解析器は CaboCha[4]を用いた．

以下，各実験の詳細について述べる．

4.1 新規用途探索実験

最初の実験では，ある特定の用途の素材に関する特許 A についての課題－効果文をクエリとし，当該課題－効果と類似課題－非類似効果を持つ特許を出力し，その出力にどの程度他の用途先が含まれるかを確認した．

[1] https://code.google.com/archive/p/word2vec/
[2] http://taku910.github.io/mecab/
[3] https://github.com/neologd/mecab-ipadic-neologd
[4] http://taku910.github.io/cabocha/

表 1: 実験条件

探索対象	2013 年～2015 年の すべての公開特許公報
初期手がかり表現	[坂地 2010]と同じ
助詞 SVM の 教師データ	上記探索対象から人手で 判定した 6,300 文
重要単語のしきい値	3
文章重要度のしきい値	10
(word2vec) 学習アルゴリズム	CBOW
(word2vec) 次元数	300
課題類義語しきい値	0.7

実験の結果, 特許 A の課題－効果文から, 306 件の特許が抽出された. 実際に306件のすべての特許を読み込み, 他の用途先となっているかを人手で判定した結果, 20 個の用途先を確認できた.

4.2 他用途展開再現実験

次の実験では, 過去, 実際に起こった用途展開例を用いて, その用途展開を再現できるか否かを検証した. 本実験では, 分野 Y の製品へ展開した分野 X の特許について, 特許の課題－効果文から分野 Y を発見することができるか否かを確認した.

実験では, 分野 X の特許 4 件の課題－効果文から, 200 件の特許を抽出した. それぞれの特許の技術分野を判定した結果, 図 5 のような結果となった. 図 5 のように, 分野 Y の特許を 6 件 (5%) 確認した. このように、用途展開した実例を再現することに成功した.

5. おわりに

本研究では, 積極的に知財を活用するために, 既存発明の技術的特徴を用いて, 新しい効果を生み出すような用途先(展開先)を発見するための手法を提案した.

提案手法は, はじめに, 大量の特許文書から発明の効果とその効果を達成するために必要であった技術的要素を表す文(課題－効果文)を抽出する. その後, 提案手法は, 抽出した課題－効果文の中から, 探索基点とする特許(クエリ)の課題と類似し, かつ効果が類似していない文を抽出する. このようにすることで, 類似した技術的特徴が生み出す, 新しい効果を探索することが可能となる.

さらに, 実際の特許文書を用いて提案手法を検証した結果, 新しい用途先が探索可能であることを確認し, さらに, 過去の用途展開事例を再現できることも確認した. これより, 本研究で提案する用途探索手法は, 有用であると考える.

なお, 著者らは, 自動的に新規用途先特許を探索する本提案手法を実装したシステムを開発している(図 6).

また, 提案手法は, 多少の変更で, さらなる応用が可能である. 例えば, 提案手法の類似課題－非類似効果を探すステップを変更し, クエリと同一の技術分野の特許から類似した効果を持ち, かつ類似していない課題を持つ文(非類似課題－類似効果文)を抽出することで, クエリと同じ目的を, 別の手段で解決するような特許を発見できる. このようにすることで, 開発者は, 発明を改良するための参考資料を容易に発見することができる.

以上のように, 提案手法は, 知財業務の様々な場面で活用が可能な手法であるが, 取り組むべき課題も存在する. 今後の課題を以下に示す.

- より大規模かつ詳細な実験

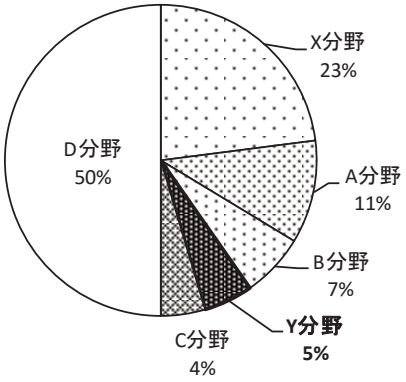


図 5: 他用途展開再現実験の実験結果

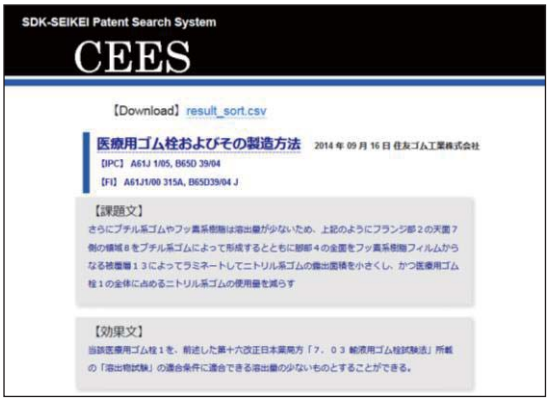


図 6: 用途先探索システム（CEES）

- 提案手法の再現率向上

今後, これらの課題に対する取り組みを行う予定である.

参考文献

[Nonaka 2010] Hirofumi Nonaka, Yusuke Suzuki, Hiroki Sakaji, Hiroyuki Sakai, Shigeru Masuyama: Extracting EFFECT Expressions Form Patent Documents, Proc. of the 25th International Technical Conference of Circuits/Systems, Computers and Communications (ITC-CSCC 2010), pp.729-731, 2010.

[石川 2007] 石川 大介, 石塚 英弘, 藤原 譲: 特許文献における因果関係を用いた類推による仮説の生成と検証 -ライフサイエンス分野を対象として-, 情報知識学会誌, Vol. 17, No. 3, pp. 164-181, 2007.

[坂地 2010] 坂地 泰紀, 野中 尋史, 酒井 浩之, 増山 繁: Cross-Bootstrapping：特許文書からの課題・効果表現対の自動抽出手法, 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 6, pp. 742-755, 2010.

[坂地 2011] 坂地 泰紀, 増山 繁: 新聞記事からの因果関係を含む文の抽出手法, 電子情報通信学会論文誌. D, 情報・システム, Vol. 94, No. 8, pp. 1496-1506, 2011.

[特許庁 2016] 特許庁: 特許・実用新案審査基準 第 II 部 明細書及び特許請求の範囲 第 3 章 発明の単一性(特許法第 37 条), http://www.jpo.go.jp/shiryou/kijun/kijun2/pdf/tukujitu_kijun/02_0300.pdf, 2016 (2018 年 2 月 28 日アクセス).

[特許庁 2017] 特許庁: 知財人材スキル標準(version 2.0), https://www.jpo.go.jp/sesaku/kigyo_chizai/chizai_skill_ver_2_0.htm, 2017 (2018 年 2 月 28 日アクセス).

[日本特許 2011] 日本特許, 公開特許公報, 情報出力装置、情報蓄積装置、およびプログラム, 特開 2011-22668, 2011.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

数学用語をクエリとするWeb上のPDF文書を対象とした数式検索

Mathematical Expression Retrieval in PDF documents from Web using Mathematical Terms as Queries

山田 奉子　　村上 晴美
Kuniko Yamada　　Harumi Murakami

大阪市立大学大学院創造都市研究科
Graduate School for Creative Cities, Osaka City University

Since mathematical expressions on the web are not annotated with natural language, searching for expressions by conventional search engines is difficult. Our proposed method performs web searches using a mathematical term as a query and extracts expressions related to it from the obtained PDF documents. We convert PDF to TeX, create images from the mathematical descriptions in TeX and obtain image feature quantities. The expressions related to the query are discriminated by SVM using the feature quantities. Our experimental results showed that MRR's performance is the best when using both PDF and HTML.

1. はじめに

Web 上の数式を，通常の検索システムで効率よく検索することは難しい．以前，数学用語をクエリとしてキーワード検索を行い，得られた HTML 形式の Web ページから数式画像を抽出し，画像周辺情報と共に上位 10 件の数式画像を提示する研究を行い，一定の成果を得られた[Yamada 17]．その後，新しいデータセットを用いて再度実験を行った．今回，その際に得られた PDF 文書を用いて，HTML との比較実験を行う．PDF 文書には画像がないため，PDF を TeX に変換後，数式記述部分から画像を作成することによって行う．クエリと数式との関連度は，数式が「独立行にある」「近傍にクエリを持つ」「適切な画像特徴量を持つ」「文書の最初の方に出現する」という観点より測る．

2. 提案手法

図 1 に沿って説明する．

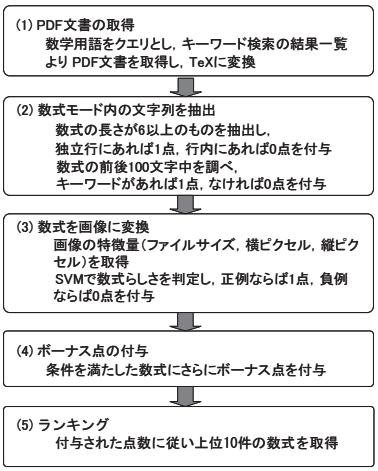


図 1：提案手法の概要

2.1 PDF 文書の取得

数学用語をクエリとしてキーワード検索を行い，上位 100 件の結果一覧から PDF 文書を取得し，InftyReader (http://www.sciaccess.net/jp/InftyReader/) を用いて TeX に変換する．

連絡先：山田 奉子，大阪市立大学創造都市研究科，
大阪市住吉区杉本 3-3-138，k16uz90z10@st.osaka-cu.ac.jp

2.2 数式モード内の文字列を取得

変数などの数式の断片を除くために，数式モード内の文字列の長さを定義する．TeX のコマンドの中で，関数の一部として用いるものや文字（例えば，\frac, \partial, \alpha, x）は 1 点，それ以外のもの（例えば，\mathrm \overline）は 0 点として，合計点を出し長さとする．$y = f(x), E = mc^2$（TeX の表記では E=mc^2）などが長さ 6 となるので，6 以上を数式として抽出する．ディスプレイ数式モードなら独立行として 1 点，インライン数式モードなら行内として 0 点を付与する．

抽出した数式の前後 100 文字中にキーワードがあれば 1 点，なければ 0 点を付与するが，キーワードを持つ数式が得られない場合は，代替キーワードを用いて再度調べ点数を付与する．ここで，代替キーワードとはキーワード検索の結果一覧のスニペット中の太文字をさす．

2.3 数式を画像に変換

数式を python のライブラリを用いて png 形式の画像に変換し，その特徴量を用いて SVM で数式らしさを判定する．使用する特徴量はファイルサイズ・横ピクセル数・縦ピクセル数・密度（式 (1)）・横縦比（式 (2)）である．

$$\text{密度} = \frac{\text{ファイルサイズ}}{\text{横ピクセル数} \times \text{縦ピクセル数}} \quad (1)$$

$$\text{横縦比} = \frac{\text{横ピクセル数}}{\text{縦ピクセル数}} \quad (2)$$

Libsvm を用いて判定し，結果が正例なら 1 点，負例なら 0 点を付与する．

2.4 ボーナス点の付与

この時点で各画像は検索エンジンのランク順かつ PDF 文書の出現順で並んでいる．点数を付与して降順に並び替えても，元々下位にあったものはなかなか上位に上がれない．そこで「重要な事柄は文書内の最初の方に出現する」として，ボーナス点を導入する．これは同一 PDF 内で最初に出現した満点 3 点の数式に，追加 1 点を与えるものである．これにより下位にある正解数式を上位に上げることができる．以上より，各数式 i_k に式 (3) により得点 $score(i_k)$ を与える．

$$score(i_k) = x_{line} + x_{key} + x_{svm} + x_{bo} \quad (3)$$

ここで，x_{line} は独立行か否か，x_{key} はキーワードを持つか否か，x_{svm} は SVM の判定が正例か負例か，x_{bo} はボーナス点を持つか否かである．

2.5 ランキング

score(i_k) に従ってランキングをし，上位 10 件を取得する．

3. 実験

3.1 データセット

表 1:データセット

	PDF			HTML		
	正解数	全画像数	PDF 数	正解数	全画像数	HTML 数
合計	398	29,708	297	181	6,030	554
正解割合	1.34%			3.00%		

ビショップの『パターン認識と機械学習』上下巻の索引から無作為にキーワードを選びクエリとして Web 検索を行い，各キーワードについて上位 100 件の Web ページを得た．先の研究で 30 キーワード分の HTML 中の画像データを用いた分類器を作成しているので，今回これを使用する．別に 10 キーワード分の PDF 文書をテストデータとした．キーワードは「正定値行列」「多峰性」「等式制約」「ニュートン-ラフソン法」「平均二乗平方根誤差」「ラグランジュ乗数」「確率の加法定理」「カーネルトリック」「一様サンプリング」「運動エネルギー」である．得られた数式の正解判定は人手により 2 値で行った．表 1 の PDF 数は TeX に変換できた数であり，HTML 文書に比べると 1 文書が多くの画像を持つが，その中で正解の占める割合が少ないことがわかる．

3.2 結果と考察

精度（式(4)）・再現率（式(5)）・F 値（式(6)）・平均逆順位（MRR）（式(7)）・平均精度の平均（MAP）を用いて，出力結果の上位 10 件を評価する．MAP は平均精度（AP）（式(8)）を計算し，さらにそのマクロ平均を求めることで得る．

$$\text{Precision@}n = \frac{r}{n} \tag{4}$$

$$\text{Recall@n} = \frac{r}{c} \tag{5}$$

$$\text{F-measure@}n = \frac{2 \cdot \text{Precision@}n \cdot \text{Recall@}n}{\text{Precision@}n + \text{Recall@}n} \tag{6}$$

$$\text{MRR} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{r'_k} \tag{7}$$

$$\text{AP@n} = \frac{1}{\min(n,c)}\sum_{s} I(s)\text{Prec}(s) \tag{8}$$

ここで，r は上位 n 件の正解数式数，c は全正解数式数，r'_k は k 番目のキーワードにおける最上位にある正解の順位，$I(s)$ は第 s 位の数式が正解か否かを表すフラグ，Prec(s)は第 s 位の精度である．

表 2:実験の結果（精度・再現率）

	精度			再現率		
	PDF	HTML	MIX	PDF	HTML	MIX
上位 1 件	0.20	0.50	**0. 60**	0.02	**0. 06**	**0. 06**
上位 2 件	0.40	**0. 75**	0.65	0.08	**0. 20**	0.13
上位 3 件	0.33	**0. 63**	0.60	0.10	**0. 24**	0.18
上位 4 件	0.28	**0. 63**	0.58	0.11	**0. 32**	0.23
上位 5 件	0.30	**0. 54**	0.48	0.15	**0. 37**	0.24
上位 6 件	0.28	**0. 47**	**0. 47**	0.17	**0. 38**	0.28
上位 7 件	0.31	0.44	**0. 46**	0.22	**0. 41**	0.32
上位 8 件	0.30	**0. 43**	**0. 43**	0.24	**0. 44**	0.34
上位 9 件	0.33	**0. 42**	**0. 42**	0.30	**0. 49**	0.38
上位 10 件	0.33	0.40	**0. 41**	0.33	**0. 52**	0.41

表 3:実験の結果（MRR・MAP）

	PDF	HTML	MIX
MRR	0.50	0.75	**0. 77**
MAP	0.17	**0. 34**	0.28

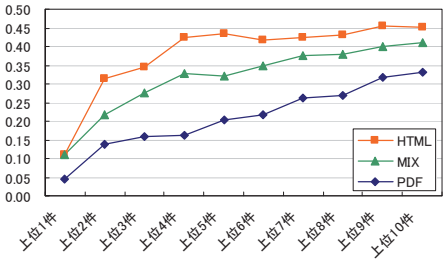


図 2:実験の結果（F 値）

表 2・表 3・図 2 中の，PDF は PDF 文書から得られた数式，HTML は HTML 文書から得られたもの，MIX はこれらを元の検索結果の順位に戻して改めてランキングしたものである．概ね，HTML の方が結果が良かったが，精度の一部及び MRR では，両方を用いた方が良かった．PDF 文書中 22.6%はスライドを PDF にしたものであり精度が悪く，上位 10 件で正解は 23 個中 2 個しか得られなかった．今後の課題は，スライド由来の PDF を排除する方策を考えること，新しい分類器で精度を上げること，長さ以外の前処理を行うことなどである．今後，テストデータの件数を増やし，新しい機械学習の方法を試してみることも必要と考える．

4. 関連研究

数式検索とは，通常類似数式検索を指すことが多いが，多くの検索システムがキーワードと数式の両方もしくは片方で検索可能である．しかしその検索範囲は特定のデータベース内である．例えば，MathWebseaerch は the arXMLiv corpus より構築したデータセット内を検索している [Hambasan 14]．また，LaTeX 形式の数式及びそれらを画像に変換したものを用いて類似数式検索を行っているものに [Zanibbi 11] の手法がある．この手法も arXiv よりデータセットを構築しており，Web 検索の結果得られたデータよりデータセットを構築しているものは見当たらなかった．

5. おわりに

数学用語をクエリとしてキーワード検索を行い，得られた PDF 文書を TeX に変換後，その数式部分の記述から数式画像を作成しその特徴量を利用することによって，キーワードに関連する数式を抽出することができた．HTML 文書を用いたものと比較した結果，MRR については両方を用いた方が良いことがわかった．

PDF は HTML より多くの正解数式を持っているので，上で述べた今後の課題について取り組み，精度を上げて行きたい．

参考文献

[Hambassn 14] Hambassa, R. et al.: Mathwebsearch at NTCIR-11, in *NTCIR-11*, pp. 114-119 (2014)

[Yamada 17] Yamada, K. et al.: Presenting Mathematical Expression Images on Web to Support Mathematics Understanding, in *IEA/AIE 2017*, Springer, pp40-46 (2017)

[Zanibbi 11] Zanibbi, R. and Yuan, B.: Keyword and Image-Based Retrieval for Mathematical Expressions, in *DRR XVIII*, pp. 011-019 (2011)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Twitterデータを用いた政治的論点の抽出と政治家分類

Politician classification based on political problems extracted from Twitter Data

野村 真由美　　尾崎 知伸
Mayumi Nomura　Tomonobu Ozaki

日本大学 文理学部
College of Humanities and Sciences, Nihon University

In order to assist the decision making for political elections, in this paper, we propose a simple method to identify political problems from Tweets by politicians. Furthermore, using cluster analysis, we discover hidden communities among politicians by characterizing each politicians based on their attitudes on detected political problems.

1. はじめに

現在，日本の政治場面において，論点や立場が判然としない政党や団体の増加が散見される．また，政策や公約，政治活動のすべてが政治広報や選挙公報に記載されるわけではなく，各政治家の状態や思想が一覧できる資料を見つけることは非常に困難である．

これらのことを背景に，これまでに，ネット上の情報を用いて種々の論点を抽出し分析する技術が提案されている．テーマを問わないシステムとして，Web 上の意見や感想といった主観情報を用いて意思決定支援を行う OpinionReader[1] が挙げられる．OpinionReader では，賛否の両立するテーマに関する記事を入力し，テーマの何が問題となるのかという論点を抽出した上で，論点ごとの賛否を可視化する．また鷹栖ら [2] は，テーマに関する意見ツイートに対して係り受け解析を適用することで論点を抽出し，クラスタリングを通じて意見文を分類する手法を提案している．同様に Jiang ら [3] は，ユーザの関連ツイートを考慮して意見文を肯定，否定，中立に分類する手法を，Kim ら [4] は文書を賛成反対の程度に応じて多値に分類する手法をそれぞれ提案している．

一方，論点抽出や意見分析の政治分野への応用も数多く試みられている．例えば，橋本ら [5] や東ら [6] は，Web サイトのデータを用いた議員分類を行っている．具体的には，Yahoo!JAPAN が運営する「みんなの政治」内に存在した「みんなの議員評価」と呼ばれるサービスを用い，一般ユーザが議員に対して行ったレビュー記事データから，各政党のイデオロギーの分析や議員の分類を行っている．また東ら [7] は，特定の政治テーマに対して明るい知識人のツイートを収集して賛成反対を分析し，その結果を基に著名議員のツイートから議員の立場を明らかにする手法を提案している．大南ら [8] は政治家ではなく一般のツイッターユーザに焦点を当て，政治に関係のない情報の混在する中から特定の政治テーマに関するツイートを集め，ユーザを分類している．また，暴力的なツイートを発するユーザ層のイデオロギーの相関についても検証を行っている．

これら既存研究の多くは，特定のテーマや論点を与えることを前提としている場合も少なくない．しかし，政治に詳しくない利用者にとっては，与えるべき論点を自主的に見つけること自体が難しいといった問題が挙げられる．そこで本研究では，政治家の日常の発信である Twitter[*1] に焦点を当てる．議員のツイートから論点となりそうな単語を抽出するとともに，それらを整理し，また論点に対する立場から議員を分類する．

連絡先: 尾崎 知伸，日本大学 文理学部 情報科学科，〒 156-8550
東京都世田谷区桜上水 3-25-40，tozaki@chs.nihon-u.ac.jp

2. データセット

ツイート集合収集時（2017 年 6 月 30 日）に国会議員を擁した政党 8 党に関し，各政党の公式アカウントがフォローしているアカウントからプロフィール等を参考に国会議員および関係者を推定し，分析の対象とした．分析対象のアカウント数は 413，総ツイート数は 753,841 件である．政党別の公式アカウントおよび分析対象アカウント数を表 1 に示す．

表 1: 分析対象政党とアカウント数

政党名	公式アカウント	対象アカウント数
自由民主党	jimin_koho	165
民進党	MinshintoNews	101
公明党	komei_koho	47
日本共産党	jcp_cc	30
自由党	seikatsu1pr	26
日本のこころ	nipponkokoro	23
日本維新の会	osaka_ishin	21
社会民主党	SDPjapan	20

3. 政治的論点に関する賛否行列の構築

政治的論点を抽出するため，賛成語類語辞典と反対語類語辞典からそれぞれ賛成語 297 語，反対語 1,087 語を特定し，是非を問う単語群として利用する．賛成語・反対語の例を表 2 に示す．次に，各ツイートに対して係り受け解析ツール CaboCha[*2] [9, 10] を適用して賛成語・反対語に係る名詞を特定し，これを政治的論点の候補とした．これらの候補から使用アカウント数が 5 未満の単語を排除し，最終的に 1,933 の政治的論点を特定した．特定された政治的論点の例を表 3 に示す．

各議員による各論点に対する態度を表現する，行数 413（＝対象議員数），列数 1,933（＝政治的論点数）の賛否行列 M

*1 https://twitter.com/
*2 https://taku910.github.io/cabocha/

表 2: 賛成語・反対語の例

賛成語		反対語	
賛成	受け入れ	反対	拒否
満足の意	好都合	不可	禁止
イエス	肯定	ノー	否定
異議なし	同意	異議あり	受け入れがたい
後押しする	承認	水を差す	抗議

表 3: 抽出された政治的論点の例

新た	安倍政権	法案	賛成
活動	議員	時	一
最後	向け	多く	出席
様々	考え	必要	対応
声	今後	皆様	選挙
思い	党	議論	何
全く	何	応援	問題
政治	参加	意見	政策
方	国会	中	質問
国	国民	今	私
岡田代表	選出	疑惑	続投
見極め	明記	国会招致	展望
2 日	歴史認識	乱暴	事実上
ケア	母校	政権運営	おり
石川	抱え	期間	いじめ
留学生	第一	撤去	いくつか
sns	先行	細野豪志	現時点
再稼動	防衛	修正案	二重国籍
国会内	18 時	岩手	全国的
内閣委員会	不足	静岡	出産
国会質問	決着	動向	南シナ海
4 項目	イスラエル	お答え	配備
命令	保育所	天候	サポーター
氏	ガイドライン	半ば	見つけ
自己	明後日	請願	神奈川
遠く	四	北海道 5 区	入国
了承	金融	一時	未来のために
進展	内閣支持率	19 日	私学
動い	おかげ	懇親	インターネット
大勢	リアル	中途半端	供給
後半	熊本県	教育委員会	発議

を考える．賛否行列 M の第 i 行，第 j 列の要素 $M_{i,j}$ は，議員 i による政治的論点 j に対する態度であり，議員 i が論点 j に賛成した回数を $A_{i,j}$，反対した回数を $D_{i,j}$ としたとき，$M_{i,j} = A_{i,j} - D_{i,j}$ と計算する．

一方比較のため，収集したツイートに含まれる名詞に着目した発言行列 U を考える．発言行列 U は，行数 413（=対象議員数），列数 1,621（=名詞数）の行列であり，議員 i が単語 j を発信した回数を要素 $U_{i,j}$ として持つ．なお，名詞の抽出には解体素解析器 MeCab [*3] [11] を用い，またすべてのアカウントで 100 回未満しか発信されなかった名詞は排除している．

賛否行列 M と発言行列 U に対し，共通して現れる名詞（列）数は 408 であった．また M にのみ現れる名詞数は 1,525，U にのみ現れる名詞数は 1,213 であった．これより，両者の Jaccard 類似度は 0.130 となり，両者で捉えている特徴が大きく異なることが分かる．

*3 http://taku910.github.io/mecab/

表 4: 政治的論点・発言クラスタにおけるクラスタ中心

クラスタ数 8		クラスタ数 4	
賛否表現	発言表現	賛否表現	発言表現
態勢	こと	態勢	こと
ため	一	ため	一
特定	会	抗議	会
5 兆円	日本共産党	安保	反対
義務付け	疑惑		
安倍政権	辺野古		
中国	瀬谷		
辺野古	行使		

4. 政治的論点の分類

賛否行列 M 及び発言行列 U の各列に対し，コサイン類似度を用いた K-medois 法を適用し，政治的論点及び利用される名詞をそれぞれグループ化する．得られたクラスタの中心（Medoid）を表 4 にまとめる．

賛否行列に対するクラスタ中心の一つである ⟨5 兆円 ⟩ は，防衛省の国家予算と年金運用にあたる損失額の両方を指す語である．また，⟨ 安保 ⟩ は安全保障関連法案の略，⟨ 特定 ⟩ は秘密法案にかかる語である．その他，⟨ 安倍政権 ⟩ ⟨ 中国 ⟩ ⟨ 辺野古 ⟩ など，全体として論点として妥当な単語がクラスタ中心として選択されていることが分かる．

5. 政治家の分類

賛否行列 M 及び発言行列 U の各行に対してクラスタリング技術を適用し，対象となる議員をグループ化する．なお，クラスタリングにはコサイン類似度を用い，政党数である K=8-medoids 法を用いた．得られたクラスタ中心（議員）の所属政党と政党別のクラスタ所属率を表 5 に示す．

賛否表現に対するクラスタリング結果は，発言表現による結果と比べ，議員の所属政党のクラスタに関してばらつきが存在している．なお，議員の本来の所属を正解とした場合との RAND 尺度は，賛否表現で 0.71，発言表現で 0.76 であり，必ずしも政党内で各論点に対する賛否が分かれているわけではないこと，個々人の発言に沿った語句によりクラスタが形成されていることが伺える．

次に，賛否表現による政治家分類の結果を考察する．日本共産党，社会民主党所属議員の多くはクラスタ 3 に，公明党所属議員の多くはクラスタ 4 に，日本維新の会所属議員はクラスタ 5 に，それぞれ集まっていることが確認された．その一方で，政党内で最も多く分類されたクラスタと，そのクラスタの medoid の議員の政党が一致しているのはクラスタ 4 の公明党のみであり，政党という意味では，必ずしもクラスタ中心（の議員）がそのグループを代表しているわけではないという結果となった．

またクラスタリング結果から，自由民主党と民進党の議員が各クラスタに分散していることが確認できる．これは，両党ともに議員数が多いため，党の内部において同一の論点の発信を行えないことを考えると妥当であると言える．クラスタ 7 に関しては，与党である自由民主党と日本のこころの議員が多く集まっているが，野党議員は非常に少ないことが分かる．この傾向は，同じく与党である公明党の議員が集まるクラスタ 4 においても確認できる．これらの結果から，与党と野党の論点

表 5: 各政党別のクラスタ所属率

ID	クラスタ中心	自由民主党	民進党	公明党	日本共産党	自由党	日本のこころ	日本維新の会	社会民主党
					賛否表現				
1	自由民主党	0.07	0.03	0.07	0.00	0.22	0.06	0.00	0.00
2	民進党	0.14	0.15	0.14	0.04	0.17	0.00	0.16	0.08
3	自由民主党	0.07	0.13	0.00	0.63	0.22	0.19	0.00	0.54
4	公明党	0.14	0.09	0.41	0.04	0.00	0.06	0.16	0.00
5	民進党	0.18	0.19	0.18	0.11	0.11	0.13	0.32	0.15
6	日本共産党	0.12	0.06	0.07	0.04	0.17	0.13	0.26	0.15
7	自由党	0.20	0.20	0.11	0.07	0.06	0.31	0.05	0.08
8	自由民主党	0.09	0.13	0.02	0.07	0.06	0.13	0.05	0.00
					発言表現				
1	民進党	0.15	0.37	0.00	0.00	0.35	0.25	0.05	0.23
2	日本共産党	0.01	0.03	0.00	1.00	0.20	0.00	0.00	0.23
3	民進党	0.15	0.15	0.09	0.00	0.05	0.00	0.00	0.00
4	自由民主党	0.38	0.08	0.05	0.00	0.15	0.00	0.00	0.00
5	日本維新の会	0.00	0.00	0.00	0.00	0.00	0.25	0.84	0.00
6	自由民主党	0.01	0.24	0.00	0.00	0.05	0.25	0.00	0.38
7	公明党	0.10	0.02	0.82	0.00	0.05	0.13	0.00	0.15
8	自由民主党	0.20	0.10	0.05	0.00	0.15	0.13	0.11	0.00

発信の方向性が大きく異なることが考えられる．

各政党の少数派によって形成されているクラスタも確認できる．これは，論点の数が政党数に満たない，または複数の政党で論点を共有している結果とも考えられる．また，論点クラスタの中心（表 3）に，政治的ではない一般名詞が含まれており，政治家分類においても，（本当の意味での）論点だけでなく，（係り受け解析により抽出された）一般名詞が大きく関係していると考えられる．

6. おわりに

本研究では，議員および政治家関係者の発信するツイートを対象に論点を抽出し，得られた論点を用いて政治家の分類を行った．その結果，与野党間の差をある程度反映しながらも，政党には寄らない分類結果が得られることが示された．

今後の課題としては，政治的論点抽出の精度向上を目的に，政治分野特有の固有名詞を対象とした固有表現抽出技術の開発や，一般名詞の排除などが挙げられる．また，二重否定や反語表現など，単純な係り受け解析だけでは特定することが困難な論点への態度の決定についても，更なる考察を行う予定である．

参考文献

[1] 佐々木 千晴，藤井 敦，石川 鉄也： 意思決定支援のための主観情報マイニング，言語処理学会第 *12* 回年次大会発表論文集，pp.77–80 (2006)

[2] 鷹栖 弘明，内海 彰： 文脈を考慮した観点に基づく意見ツイートクラスタリング，言語処理学会第 *21* 回年次大会発表論文集， pp.433–436 (2015)

[3] L. Jiang, M. Yu, Mi. Zhou, X. Liu, and T. Zhao : Target-dependent Twitter Sentiment Classification, *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.151–160 (2011)

[4] S.-M. Kim and E. Hovy : Determining the sentiment of opinions, *Proc. of the 20th international conference on Computational Linguistics*, Article No. 1367 (2004)

[5] 橋本 悠，掛谷 英紀： Web 上のレビュー記事のイデオロギー分析とその応用， 言語処理学会第 *16* 回年次大会発表論文集，pp.740–743 (2010)

[6] 東 宏一，橋本 悠，掛谷 英紀： WEB 上の言語資源に基づく国会議員の分類， 言語処理学会第 *17* 回年次大会発表論文集，pp.476–479 (2011)

[7] 東 宏一，掛谷 英紀： 国会議員のツイッター分類とその応用， 言語処理学会第 *18* 回年次大会発表論文集，pp.559–562 (2012)

[8] 大南 勝，掛谷 英紀： 政治的立場の異なるツイッターユーザ群の特徴分析，言語処理学会第 *22* 回年次大会発表論文集, pp.537–540 (2016)

[9] 工藤 拓，松本 裕治： チャンキングの段階適用による日本語係り受け解析， 情報処理学会論文誌，Vol.43, No.6, pp.1834–1842 (2002)

[10] Taku Kudo and Yuji Matsumoto : Japanese Dependency Analysis using Cascaded Chunking, *Proc. of the 6th Conference on Natural Language Learning 2002*, pp.63–69 (2002)

[11] T. Kudo, K. Yamamoto and Y. Matsumoto : Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp.230–237 (2004)

[2L4] [General Session] 9. NLP / IR

座長:角森 唯子（NTTドコモ）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room L (3F Sapphire Hall Asuka)

[2L4-01] Measuring Beginner Friendliness of Web Pages explaining Academic Concepts by Integrating Heterogeneous Features

○Shiokawa Hayato[1], Takahide Kasuga[1], Kota Kawaguchi[1], Bingcai Han[1], Takehito Utsuro[1], Yasuhide Kawada[2] （1. University of Tsukuba, 2. Logworks Co., Ltd.）

5:20 PM - 5:40 PM

[2L4-02] The Poc on automatic detection of erroneous telop using AI technology

○Makoto Nakano[1] （1. Nippon Television Network Corporation）

5:40 PM - 6:00 PM

[2L4-03] A Simple Reproducible Judgment Method for Lexical Simplification based on Word Difficulty

○Riku Takada[1], Tomoyuki Kajiwara[2], Noriyuki Okumura[1] （1. National Institute of Technology, Akashi College, 2. Osaka University）

6:00 PM - 6:20 PM

[2L4-04] Country Adaptation in Neural Machine Transliteration of Person Names

○Dawoon Jung[1], Satoshi Sato[1] （1. Nagoya University）

6:20 PM - 6:40 PM

[2L4-05] Scoring support system using sequence alignment

○Kohei Takai[1], Kengo Taketani[1], Yasukuni Mori[2], Hiroki Suyari[2] （1. Department of Informatics and Imaging Systems, Faculty of Engineering, Chiba University, 2. Graduate School of Engineering, Chiba University）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

異種の素性を併用した 学術用語解説ウェブページの分かり易さの自動評定

Measuring Beginner Friendliness of Web Pages explaining Academic Concepts by Integrating Heterogeneous Features

塩川 隼人[*1] Hayato Shiokawa　春日 孝秀[*1] Takahide Kasuga　川口 輝太[*1] Kota Kawaguchi　韓 炳材[*1] Bingcai Han　宇津呂 武仁[*2] Takehito Utsuro　河田容英[*3] Yasuhide Kawada

[*1]筑波大学大学院システム情報工学研究科 Grad. Sch. Sys. & Inf. Eng, Univ. of Tsukuba　[*2]筑波大学システム情報系 Fclty. Eng, Inf. & Sys, Univ. of Tsukuba　[*4](株) ログワークス Logworks Co., Ltd.

Search engine is an important tool of modern academic study, but the results are lack of measurement of beginner friendliness. In order to improve the efficiency of using search engine for academic study, it is necessary to find a method of measuring the beginner friendliness of a Web page explaining academic concepts and to build an automatic measurement system. This paper studies how to integrate heterogeneous features including those generated from the image of the Web page by a variant of CNN as well as those extracted from the body text of the HTML file of the Web page. Integration is performed through the framework of the SVM classifier learning. Evaluation results show that heterogeneous features perform better than each individual type of features.

1. はじめに

インターネットが普及した現在，ウェブ上には多くの学術関連のコンテンツが存在する．このコンテンツを使って学術用語を学ぶことは，学校で使用される教科書・参考書による学習とは別に，学習手段の一つとしてたびたび用いられている．インターネットで学術用語を学ぶ際，ウェブ検索に頼れば関連ページは容易に見つかる．しかし,「分かり易い用語解説」を見つけるには，検索上位ページを一つずつ見比べ読み進める非効率な作業が必要となる．これは，ウェブ検索では，分かり易さをベースにした解説ページ群の体系化ができないことが原因である．そこで，本研究では，検索上位ページの中で分かり易さを充足する必要十分な数の用語解説ページを見つけ体系化することを目的とする．さらに，この研究においては，最終的に，用語解説を提示してウェブ学習を促進する学習教材読解支援システムを実現する．

以上の目的をふまえて，本論文では，図 1 に示すように，6 個の個別因子を手がかりとして，分かり易さの全体評定を行う手法を構築する．特に，本論文では，6 個の個別因子のうち「レイアウトの見易さ」についての因子を対象とし，深層学習を用いてこの因子を自動評定する方式 [塩川 18] によって得られる特徴量，および，用語解説ウェブページの HTML ソーステキスト中のテキスト情報の特徴量 [春日 18] を併用して，分類器として SVM を用いて全体評定の自動判定を行う手法 [春日 18] を適用し，異種の特徴量を効果的に併用可能であることを示す．特に，理工系学術用語を検索クエリとして収集した用語解説ウェブページを対象として行った評価結果をふまえて，提案手法の有効性を示す．

2. 用語解説ウェブページの「分かり易さ」評定の因子

本節では，用語解説ウェブページの個別因子および全体評定の評定基準について述べる．これらの評定においては，本論文の著者が二値での評定を行った．

連絡先: 塩川 隼人，筑波大学大学院システム情報工学研究科，〒 305-8573 茨城県つくば市天王台 1-1-1, 029-853-5427

2.1 個別因子

(a) 用語の定義がある

必要十分な用語の定義を述べている．例えば，統計分野の用語を検索クエリとしたとき，エクセル操作の解説ページがあり，学術用語に関する説明がされず，応用的な内容のみであれば定義無しとする．

(b) 図がある

用語解説や例示の為に使用されている図を一つ以上含む．文字のみの表などは対象としない．

(c) 数式がある

一般的に数式と呼べるものを含む．アルファベット一文字等は含めない．

(d) 例がある

例題や,「例えば，～ のように」などの例示を含む．例示については，用語を説明するための短すぎない例示に絞る．一つでも例題，例示と判断できるものがあれば良いものとした．

(e) レイアウトが見易い

ページのレイアウトの見易さ．学術用語解説ウェブページの見易さに特化する，というよりは，一般的なウェブページの見易さに準ずる．例えば，シンプルさ，改行や章題による構成の工夫，色遣い等が挙げられる．

(f) 文章が分かり易い

用語解説ページの文章の分かり易さ．そのページで扱う用語と，対象とするページ閲覧者層を考慮し，総合的な評価で行う．具体的には，例えば，専門用語と平易な用語のバランス，情報量，文章量等を基準に判断を行った．

2.2 全体評定

6 個の個別因子を元に，ウェブページ全体の評価付けを行う．全体評定の評価は，個別因子の評定がどの程度満たされているかを基準として判断しているが，個別因子全てが充足されている必要はなく，ある程度充足されていれば，全体評定も充足すると評定している．例えば,「図あり」や「式あり」などの因子は必ずしも必要ではなく，これらの因子が欠けていたとしても

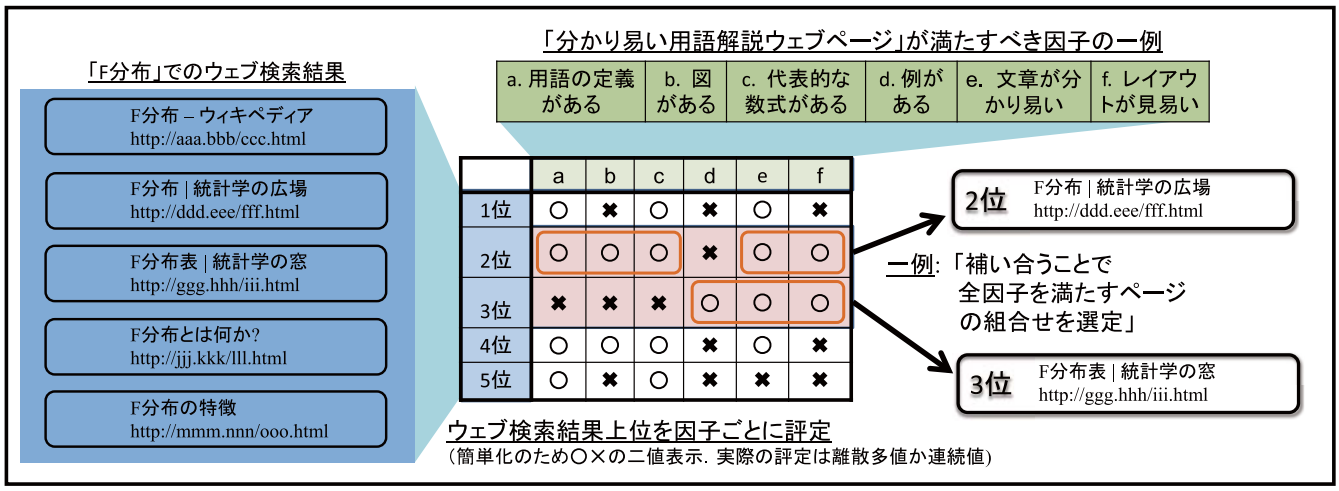


図 1: 用語解説ウェブページの「分かり易さ」評定の因子

表 2: 対象学術分野ごとの正例数及び負例数

	対象学術分野	正例数	負例数	合計
訓練事例	線形代数	36	54	89
	物理	26	101	127
	生物	17	47	64
	プログラミング	70	72	142
	IT	39	50	89
	小計	188	323	511
評価事例	統計	66	84	150
	化学	52	98	150
	合計	340	620	960

全体評定が充足されないわけではない．6 個の因子の中で特に重要なのは「レイアウトが見易い」と「文章が分かり易い」のため，この両方が欠けている場合，あるいはどちらか片方が欠けていて，かつ，その他の個別因子が複数欠けている場合は，全体評定を「充足しない」と評定している．

3. 参照用学術用語解説ウェブページ集合の作成

3.1 対象学術分野および用語

事前調査を行い,「レイアウトの見易さ」,「文章の分かり易さ」,「全体評定」の評定基準が似ている理工系学術分野である統計・線形代数・物理・生物・化学・プログラミング・IT 分野を本研究での対象分野として選んだ．次に，各分野の用語の内，主に高校 3 年生から大学の学校教育レベルのものの中から，表 1 に示すように，統計・物理・化学・プログラミング・IT は 15 語，線形代数のみ 14 語，生物のみ 7 語を対象学術用語として選んだ．

3.2 作成手順および結果

前節で収集した学術用語に対して用語解説ウェブページを収集し，2. 節で述べた個別因子および全体評定を人手で判定した事例を蓄積する．

まず，ウェブページの収集には検索エンジンを用い，各用語を検索クエリとした場合の検索結果上位 10 件のウェブページを集めた．その際，HTML ファイル収集プログラムではアクセスできないページは参照用ページ集合に加えない．そして，2.2 節で述べた基準に基づき全体評定の判定を人手で行い，その判定結果を付与した事例を蓄積する．以上の手順によって，全体評定が「充足する」と判定された事例を正例,「充足しない」と判定された事例を負例として，各分野ごとに表 2 に示す数の正例・負例を収集した．全 7 分野のうち 5 分野を訓練用分野，残り 2 分野を評価用分野とした．

4. 深層学習による学術用語解説ウェブページの見易さの自動評定結果の付与

4.1 VGG16 [Simonyan15]

近年，深層学習は多様な分野のタスクにおいて，その有用性を示している．特に画像認識の分野では，畳み込みニューラルネットワーク (CNN) と ImageNet などの大規模なデータセットの登場により，様々なタスクにおいて驚異的な性能を発揮している．また，ImageNet のような大規模なデータセットを用いて学習を行った CNN のパラメータは，その汎用性の高さから高性能な特徴抽出器として異なるドメインのタスクにおいても活用できることが知られている．

そこで，本論文では，用語解説ウェブページの「レイアウトの見易さ」を評定するタスクにおいて，VGG16 モデル [Simonyan15] を基盤の特徴抽出器として用いる．VGG16 は ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 [Russakovsky14] の「画像分類タスク (Image classification)」で準優勝,「位置特定タスク (Single-object localization)」で優勝したモデルであり，このモデルは 13 層の畳み込み層，その途中に挟まれている 5 層の最大プーリング層，3 層の全結合層 (1,000 値分類用)，および，出力層から成り立っている．訓練済みの VGG16 としては，ImageNet2014 のデータセットによって 1,000 値分類用に訓練済みのモデルが一般に公開されており，この訓練済みモデルは他のタスクにも広く転用可能な

表 1: 学術分野ごとの対象学術用語一覧

分野	対象とする学術用語数	対象とする学術用語
線形代数	14	階数，共役勾配，行列式，クラメルの公式，クロネッカーのデルタ，三角行列，正規直交基底，対角化，直交行列，特性多項式，二次形式，ノルム，メネラウスの定理，ヤコビ行列 (計 14 語)
物理	15	電気力線，張力，慣性の法則，遠心力，電波，電流，万有引力，交流，音波，ホイートストンブリッジ，反発係数，相互誘導，正電荷，速度，変圧器 (計 15 語)
生物	7	DNA，ショウジョウバエ，原核生物，減数分裂，光合成，細胞，葉緑体 (計 7 語)
プログラミング	15	C 言語，Java，エスケープシーケンス，コマンドライン引数，スコープ，フィールド値，ポインタ，メソッド，繰り返し処理，構造体，算術演算子，条件分岐，配列変数，文字列，論理演算 (計 15 語)
IT	15	API，DBMS，HTML，IP アドレス，JDBC，RDB，SDK，SQL，Unicode，URL，スコープマネジメント，ステークホルダーマネジメント，タイムマネジメント，ナレッジマネジメント，リスクマネジメント（計 15 語）
統計	15	事後分布，確率，相関係数，信頼区間，k 平均法，回帰分析，十分統計量，確率密度関数，ガンマ分布，事前分布，主成分分析，コーシー分布，f 分布，自己回帰モデル，自己回帰 (計 15 語)
化学	15	イオン結合，エステル，カルボン酸，ケトン，化学反応式，化学平衡，共有結合，合成高分子，酸化還元，遷移元素，典型元素，天然高分子，燃料電池，物質の三態，芳香族有機化合物 (計 15 語)
合計	96	—

ことが知られている．本論文では，Python の深層学習ライブラリ Keras[*1] 用に公開されているモデル[*2] を利用して評価実験を行った．

4.2 見易さの自動評定結果の付与手順

本論文では，ImageNet データセットを用いて訓練済みの VGG16 モデルを基盤に用いて用語解説ウェブページの「レイアウトの見易さ」の判定の評価実験を行った．参照用学術用語解説ウェブページデータとして，3. 節で作成した用語解説ウェブページ事例を用い，各ウェブページのトップページを画像化したものを実際の事例として用いる．この訓練・評価事例集合の正例・負例数は，表 2 の通りである．

本論文において VGG16 を利用して「レイアウトの見易さ」を評定するモデルを訓練する際には，まず，VGG16 の 1,000 値分類用の 3 層の全結合層と出力層を取り払い，2 クラス分類用の全結合層 3 層と出力層を接続し，以下のモデルを用意する．

> 「fine-tuning-VGG16 ありのモデル」
>
> VGG16 の畳み込み層 10 層，および，最大プーリング層第 4 層目までのパラメータを VGG16 の訓練済みのパラメータに固定し，それ以降の畳み込み層 3 層，最大プーリング層 1 層，全結合層 3 層に対して，本論文の「レイアウト見易さ」判定タスク用の訓練事例を用いて，再度パラメータの訓練を行う．

対象学術分野のうち，統計および化学分野に属する学術用語をクエリとして収集した参照用学術用解説ウェブページ集合を評価事例とし，それ以外の事例を訓練事例とする．そして，以下の手順で，訓練・評価事例に対して深層学習による見易さの自動評定結果を付与する．

訓練事例 全 5 分野のうちの 4 分野で訓練したモデルを用いて，残りの一分野の事例に対して見易さの自動評定結果およびその信頼度を付与する．

評価事例 訓練事例の全 5 分野のうちの 4 分野で訓練したモデルを全 5 通り作成し，各評価事例に対して 5 つのモデルのうちの一つを無作為に適用し，見易さの自動評定結果およびその信頼度を付与する．

5. 深層学習による見易さ自動評定結果およびテキスト特徴量を素性とする分類器学習による全体評定の判定

5.1 素性

5.1.1 深層学習による見易さ自動評定結果の素性

4.2 節で付与された見易さの自動評定結果およびその信頼度の情報は，[0,1] の区間の数値情報として付与されており，信頼度=0.5 の場合を中間点として，値が 1 に近いほど，「自動評定=見易い」の信頼度が高く，値が 0 に近いほど，「自動評定=見易くない」の信頼度が高くなる．この数値を一つの素性として用いる．

5.1.2 テキスト素性

大別して，以下の三種類，合計して全 10 種類の素性を用いる．

文字種ごとの文字数 漢字，平仮名，片仮名の各文字種の文字数 (全三種類).

特定の語・文字列の有無または頻度 特定の語・文字列の頻度として，「利用規約」，「相談」，「ノウハウ」，「法」[*3]，「困」[*4]の各々の語・文字列の頻度を素性として用いる (全五種類)．同様に，「例」および「Q0」，…，「Q9」の頻度の合計を素性として用いる (全一種類).

HTML タグ中の文字列パターンの頻度 HTML タグに囲まれた文字列パターンの頻度として，“>” と “<” に囲まれ

*1 https://keras.io/ja/
*2 https://github.com/keras-team/keras/blob/master/keras/applications/vgg16.py

*3 「方法」，「手法」等の語の頻度を集計することを想定し導入している．
*4 「こんな困ったことがありませんか?」等のフレーズの頻度を集計することを想定し導入している．

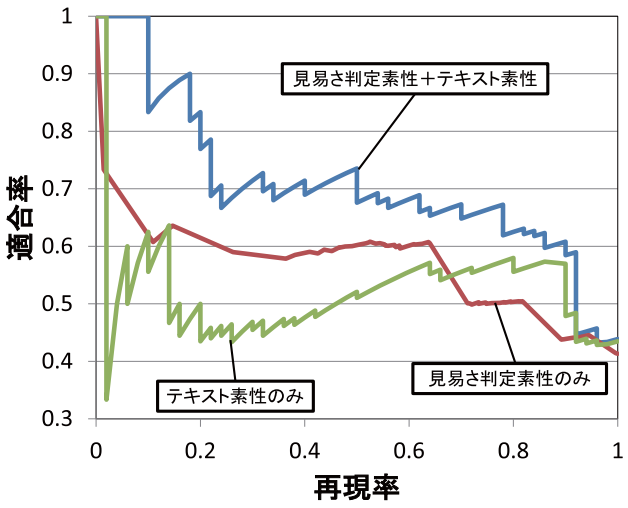


図 2: 評価結果

た一文字以上の数字列の頻度を素性として用いる (全一種類)*5.

5.2 評価手順

分類器としては，scikit-learn [Pedregosa11] パッケージにおける SVM(sklearn.SVM.SVC ツール) を用いる．3 値以上の多値をとる素性については，素性のとる値の分布に応じて，重複のない多数の範囲をとる二値素性，および，重複のある多数の範囲をとる二値素性等，様々な二値素性のとり方を評価した．その結果，左端を固定し，右端が異なる 20~40 個の範囲をとる二値素性を用いた場合に最も高い性能となったため，この素性設定を用いた．SVM のカーネル関数としては RBF カーネルを用い，最適化の目的関数を再現率・適合率の ROC 曲線の面積として，グリッドサーチによりコストパラメータ C(1 および 10) および RBF カーネルのパラメータ γ(0.01，0.001，および，0.0001) を最適化する．そして，全体評定の値をクラスとして SVM の訓練を行う．

5.3 評価結果

素性として，(a) 深層学習による見易さ自動評定結果の素性，および，テキスト素性を併用した場合 (図 2 中の「見易さ判定素性+テキスト素性」のプロット)，(b) 深層学習による見易さ自動評定結果の素性のみを用いた場合 (図 2 中の「見易さ判定素性のみ」のプロット)，(c) テキスト素性のみを用いた場合 (図 2 中の「テキスト素性のみ」のプロット) の三通りを用いた場合を比較した結果を図 2 に示す．この結果から分かるように，深層学習によりレイアウトの見易さを自動評定した結果の素性，および，ウェブページの HTML ソーステキスト中のテキスト情報の素性という異種の素性を併用することにより，それぞれの素性を単独で用いた場合よりも全体評定の性能を改善できた．

6. 関連研究

分かりやすさの全体評定を行うのに必要な個別因子のうち，見易さ以外の因子の自動評定については，本論文の関連タスクとして，[春日 18, Han18] において，HTML 構造上の特徴を利用した見易さ以外の個別因子及び，全体評定の自動評定の手法を提案している．この研究と本稿で述べた手法を組み合わせることによって，個別因子全ての自動評定が可能になり，本研究の目的である用語解説ページの体系化を行うことができる．

この学術用語解説ウェブページの分かり易さ自動評定タスクに関連して，コミュニティ型質問応答の分野においては，回答者による回答の良質さを自動評定する手法についての研究が行われている [石川 11, Sakai11]．また，特に，本論文における学術用語解説ウェブページの見易さの自動評定に関する関連タスクとして，[大山 17] においては，本論文とほぼ同様の深層学習手法によって，プレゼンテーションスライドの画像情報に対する印象を予測する手法を提案している．

7. おわりに

本論文では,「レイアウトの見易さ」についての因子を対象とし，深層学習を用いてこの因子を自動評定する方式 [塩川 18] によって得られる特徴量，および，用語解説ウェブページの HTML ソーステキスト中のテキスト情報の特徴量 [春日 18] を併用して，分類器として SVM を用いて全体評定の自動判定を行う手法 [春日 18] を適用し，異種の特徴量を効果的に併用可能であることを示した．

参考文献

[Han18] Han, B., Shiokawa, H., Kawaguchi, K., Utsuro, T. and Kawada, Y.: Measuring Beginner Friendliness of Chinese Web Pages explaining Academic Concepts using HTML Structures, 第 32 回人工知能学会全国大会論文集 (2018).

[石川 11] 石川大介, 酒井哲也, 関洋平, 栗山和子, 神門典子：コミュニティ QA における良質回答の自動予測, 情報知識学会誌, Vol. 21, No. 3, pp. 362–382 (2011).

[春日 18] 春日孝秀, 塩川隼人, 韓炳材, 宇津呂武仁, 河田容英：HTML 構造上の特徴を利用した学術用語解説ウェブページの分かり易さの自動評定, 第 10 回 DEIM フォーラム論文集 (2018).

[大山 17] 大山真司, 山崎俊彦, 相澤清晴：プレゼンテーションスライドの客観評価と印象予測, 第 16 回 FIT 講演論文集, 第 3 巻, pp. 45–52 (2017).

[Pedregosa11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830 (2011).

[Russakovsky14] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., Berg, A. C. and Li, F.: ImageNet Large Scale Visual Recognition Challenge, *CoRR*, Vol. abs/1409.0575, (2014).

[Sakai11] Sakai, T., Ishikawa, D., Kando, N., Seki, Y., Kuriyama, K. and Lin, C.-Y.: Using Graded-Relevance Metrics for Evaluating Community QA Answer Selection, *Proc. 4th WSDM*, pp. 187–196 (2011).

[塩川 18] 塩川隼人, 春日孝秀, 韓炳材, 宇津呂武仁, 河田容英：深層学習を用いた学術用語解説ウェブページの見易さの自動評定, 第 10 回 DEIM フォーラム論文集 (2018).

[Simonyan15] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *Proc. 3rd ICLR* (2015).

*5 複数のページから構成される一連のページ群の間で任意のページに遷移するためのリンク群におけるページ数を集計することを想定し導入している．

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

AI 技術を使った誤テロップ自動検出に関する技術検証

The Poc on automatic detection of erroneous telop using AI technology

中野　信[*1]
Makoto Nakano

[*1] 日本テレビ放送網株式会社
Nippon Television Network Corporation

Keyword. Television, Telop, Broadcast,TVprogram

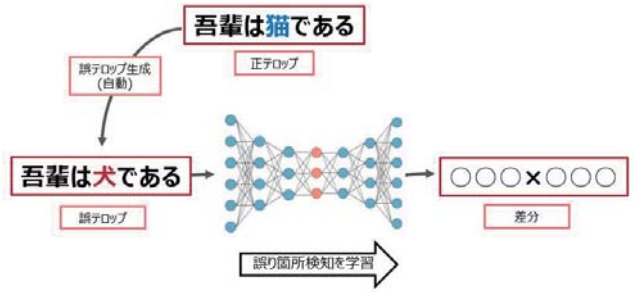
1. 背景

放送局で制作・放送しているテレビ番組では、画面上にいくつもの「文字テロップ」が、次々に切り替わりながら表示されている。これは、放送している内容を視聴者の方にスムーズに分かっていただくためであったり、演出上の工夫であったりする。しかし、時に、このテロップの文字に誤りがあり、不体裁となる場合が少なからず発生している。放送している画面にテロップが表示されるまでには、人間による校閲を経ているのではあるが、どうしても見逃しが出てしまっている。これは、熟練した校閲者であっても人手による確認では限界があること、また、テロップは毎日大量に作られており、チェック作業量は膨大である、などもあり、誤テロップを完全になくすことは困難である。そこで、このチェック作業を補助することを目的とした自動チェックの仕組みが、AI 関連技術を使ってできないか、技術検証を行った。

なお、検証に当たっては、富士通株式会社様に技術協力を頂いた。

2. 概要

基本的なアプローチは、まず、正しいテロップ＜正テロップ＞から、過去の校閲事例を参考にして誤りを埋め込み生成した＜誤テロップ＞を生成、この誤テロップを RNN (Recurrent Neural Network)へ入力し、テロップの「誤り箇所検出の学習」をさせる、という手法とした。



誤テロップ生成の方式は、実際の校閲事例を参考に、以下8つの誤りパターンに分類した。

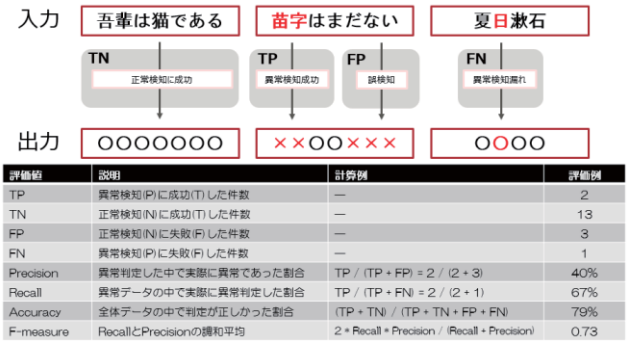
・同じ読みの漢字で置換（漢字 → 幹事）
・似た平仮名/片仮名を置換（づくし → ずくし）
・文字削減（Congratulations → Congratulation）
・文字追加（ドラゴンズ → ドランゴンズ）
・文字入れ替え（ノビレチン → ノビチレン）
・ら抜き言葉（見られる → 見れる）
・近い意味の言葉で置換（猫 → 犬）
・数字を置換（1998 年 → 1996 年）

また、この誤りパターンごとに Neural Network を分けて学習させるという「モデル分割」という手法も取り入れ、分割有り・無しでの比較も行った。

なお、RNN へ入力する学習データであるが、実際に放送で利用されたテロップのテキストデータの十分な量の入手が困難であるという事情のため、Wikipedia から抽出した文章もあわせて使い、以下のようなデータセットで検証を行った。

	学習データ	モデル分割	学習データ数	検証データ	検証データ数
A	Wikipedia	有り	640万件（×5モデル）	Wikipedia	3.2万件（×5モデル）
B		無し	3200万件		16万件
C				校閲事例集	2668件
D	校閲事例集		2500件		168件

また評価指標は以下とした



評価値	説明	計算例	評価例
TP	異常検知(P)に成功(T)した件数	—	2
TN	正常検知(N)に成功(T)した件数	—	13
FP	正常検知(N)に失敗(F)した件数	—	3
FN	異常検知(P)に失敗(F)した件数	—	1
Precision	異常判定した中で実際に異常であった割合	TP / (TP + FP) = 2 / (2 + 3)	40%
Recall	異常データの中で実際に異常判定した割合	TP / (TP + FN) = 2 / (2 + 1)	67%
Accuracy	全体データの中で判定が正しかった割合	(TP + TN) / (TP + TN + FP + FN)	79%
F-measure	RecallとPrecisionの調和平均	2 * Recall * Precision / (Recall + Precision)	0.73

3. 検証

技術検証を進めて行った結果、以下のような課題や新たなアイデアが出てきた。

●誤テロップ補正の学習が収束しない
■推定原因
・学習データ生成時の自動誤テロップ生成に問題がある。
■改善対策案

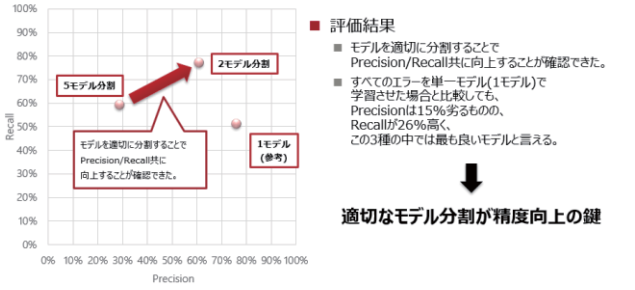
連絡先：日本テレビ技術統括局技術開発部　中野　信
m-nakano@ntv.co.jp

＜誤テロップ生成条件を制約する＞
・数字誤りは同一桁数の置換に限定
・前後の文字置換は同一文字種別の交換に限定
・過剰文字は同一文字種別に挟まれた箇所への挿入に限定
＜誤変換生成辞書の拡充＞
・ipadic 辞書（36 万件）⇒ipadic 辞書+SKK 辞書（52 万件）
＜誤りパターンの整理＞
・前述 8 パターンを、5 パターンへ集約

●学習済みモデルの分割について

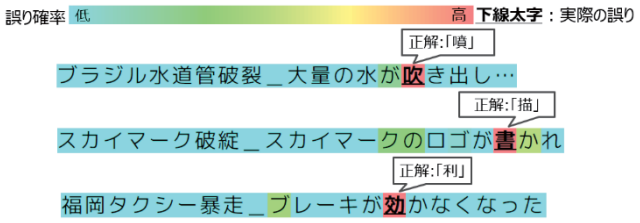
前述した「モデル分割」の手法であるが、誤りパターンごとにモデル分割した際に、個別のモデルでは高い精度が出るものの、単一文章に対してモデルの重ね掛けを行うと精度が非常に低い、ということがわかった。そこで、さらにモデルの分割数を減らす（前述の 5 パターンを、言語的誤り検出、意味的誤り検出の 2 モデルへ集約）ということを試みたのでその結果を以下に示す。

モデル数	TP	TN	FP	FN	Precision	Recall	Accuracy	F-measure
2モデル	637,852	23,508,900	416,401	192,442	61%	77%	98%	0.68
5モデル	154,875	7,800,621	370,237	106,841	29%	59%	94%	0.39
（1モデル）	132,246	8,130,080	40,653	129,470	76%	51%	98%	0.61



●誤り検出結果の可視化

今回の取り組みでは、実際にテロップ制作や、校正・校閲作業を担っている現場担当者に対するヒアリングも行った。その際に「検出した誤りを、校閲者に分かりやすく提示できないか」という意見が出て、それを実現するために「どれくらいの確率で誤りである、と判断したか含め分かりやすく可視化する」ということに取り組み、以下のように、誤り推定確率を色の違いで表現する、という手法を考案した。



4. まとめ

今回、DeepLearning を誤り検出に応用した事例がほとんど無い中で技術検証を行い、条件に制約はあるものの、技術的に可能であることを確認できた。

また、大量の正しいテキストに誤りを埋め込んで教師データ化し、学習を行う手法の有効性を実証した。これは、テロップの実データに限りがある中、今後の取り組みに大いに役に立つものと思われる。一方、実データを使っての学習においては、校閲事例の蓄積フォーマットが、学習用データとするには整っておらず、クレンジングの手間が多く掛かる結果となった。そのため、今後の取り組みに備えるためにも、蓄積フォーマットを整備する、という現場運用の課題も見出せた。

また、校閲事例から作った学習用データの数を増やせば増やすほど、自動誤テロップ検出精度が向上することも確認でき、前述のフォーマットの整備とあわせて、データ整備の必要性を認識できた。

最後に、現場で使う場合を想定した手法、Neural Network による誤り推定確率を色で表現し、校閲者に分かりやすい形で出力する、ということもできることが分かり、作業量削減のための現場での適用という目的に沿う技術であるという目処も得られた。

5. 今後に向けて

① 実データ(テロップデータ及び校閲事例集)の蓄積・管理

DeepLearning を使った誤テロップ検出の精度改善には実データの教師データ化が非常に重要であることが実証できたのでデータを効率よく蓄積、管理する仕組みを整備する。

② 誤テロップ検出モデルの改善

実際に放送に使用したテロップに対しての誤り箇所検出精度が Precision 45%、Recall 34%程度であり、誤テロップ検出モデルのさらなる改善の余地がある。

③固有表現のモデル分離

固有名詞や数詞の誤りは、校閲事例を振り返っても約 5 割を占めていてぜひ自動チェックしたいところであるが、パターンで判断できるものではなく、現状のモデルには不適。ここを実現できるモデルをぜひ見つけたい。

③ 事実を参照できる仕組み

上記③にも書いたが、一般的な Neural Network はパターンを判断するのみで、情報を記憶しない。しかし、時々刻々変化する人の年齢や、スポーツにおける成績（勝率など）は、人間でもチェックが困難で時間が掛かる。これが自動化できるとかなりの校閲作業省力化が見込まれるため、ぜひ実現したい。

最後ではあるが、実験検証実施、本論文作成に当たり多大なご協力を頂いた富士通株式会社様へ深謝いたします。

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

再現が容易な単語の平易化判定手法

A Simple Reproducible Judgment Method for Lexical Simplification based on Word Difficulty

高田　理功 *1　　梶原　智之 *2　　奥村　紀之 *1
Takada Riku　　Kajiwara Tomoyuki　　Okumura Noriyuki

*1明石工業高等専門学校　　*2大阪大学
National Institute of Technology, Akashi College　　Osaka University

A lexical simplification is one of the important techniques for leaners of second languages. In this paper, we propose a simple reproducible judgment method for lexical simplification without using probabilistic model. In addition, our proposed method can simplify vocabularies to easier vocabularies at a significant advantage to previous researches.

1. はじめに

グローバル化が進む中で，日本語学習者が日本人だけでなく外国人を含むようになっている．しかし，日本語学習者のうち，日本語初学者である子供や外国人は難解な単語の意味を理解することができない．そのため，日本語初学者は，文が平易な単語で構成されていると読みやすい．それを実現する技術として，自動的に難解な単語を平易な単語に言い換える語彙平易化という技術がある．語彙平易化は言い換え元と言い換え先の単語のどちらが簡単か判定する平易化判定を行う必要がある．そのため，本稿は平易化の基盤である単語の平易化判定を行う．

先行研究として，梶原ら [梶原 17] がある．梶原らの手法は素性に単語分散表現の 1 つである Word2Vec[Mikolov 13] を用いている．しかし，単語分散表現は学習結果が初期値の乱数に依存するため再現が困難であり，かつ平易化判定時にどの素性が有用か判断することができない．

そこで我々は Word2Vec がどの程度初期値の乱数に依存するか調査する．また，文字数，文字種，頻度のように結果の解釈が容易な素性のみを用いる平易化判定手法を構築する．

本稿の結果，我々の平易化判定手法は，先行研究よりも accuracy が 6.5 ポイント改善し，かつ再現が容易な手法と分かった．また素性として，日本語 Wikipedia 中の文字単位の頻度と BCCWJ 中の単語単位の頻度が有用であることが分かった．

2. 関連研究

語彙平易化では，入出力の単語が以下の 2 つを満たすように変換される．

[**同義性**] 言い換え先の単語が言い換え元の単語の意味を含有する
[**平易化**] 言い換え先の単語が言い換え元の単語より平易である

同義性に関する研究として，PPDB[Ganitkevitch 13] がある．PPDB は言い換え辞書を自動生成する研究であり，英語をある言語に翻訳した後，再度翻訳して英語に戻すことで，言い換え元の意味と同義なフレーズを取得できるとし構築された．また日本語で同様に自動作成された言い換え辞書として PPDB: Japanese*1[Mizukami 14] がある．PPDB: Japanese は行ごとに言い換え元，言い換え先，言い換え確率などが書かれている言い換え辞書である．PPDB: Japanese は言い換え確率の閾値が異なる 5 つのサイズが公開されており，本稿ではその中で最も言い換え対が多く収録されている 10best を利用している．

平易性に関する研究として，Simple PPDB[Pavlick 16] がある．Pavlick らは，PPDB を改良した PPDB: 2.0[Pavlick 15] を元に平易化判定を行い，平易化用の大規模な言い換え辞書 Simple PPDB を構築した．Pavlick らの平易化判定は，フレーズから抽出できる素性を利用した，ロジスティック回帰モデルを用いるモデルである．

これと同様に日本語の平易化用の大規模な言い換え辞書を構築した例として，Simple PPDB: Japanese*2[梶原 17] がある．梶原らは以下の 2 つの手順で平易化判定を実現した．

[**難易度推定**] 言い換え元の単語と言い換え先の単語の難易度を推定する
[**平易化判定**] 言い換え元の単語と言い換え先の単語の難易度差を算出する

難易度推定には RBF カーネルを採用した SVM を用いており，素性には日本語 Wikipedia*3 中に出現する単語の頻度, 単語長, 文字種 (ひらがな，カタカナ，漢字), 日本語 Wikipedia 上の Word2Vec[Mikolov 13] による単語分散表現を用いている．最も高い accuracy を示した Word2Vec は，100 次元の skip-gram with negative sampling モデルであった．

3. 単語分散表現の再現の容易性

2 節のように，梶原らは単語分散表現を用いている．しかし単語分散表現は，学習結果が初期値の乱数に依存するという問題が存在する．どの程度学習結果に初期値の乱数が依存するか調査するため，以下の手順で実験を行った．

本稿では，梶原らと同様に，MeCab*4 と mecab-ipadic-NEologd[佐藤 17] によって単語分割された日本語 Wikipedia の本文を用いて，100 次元の skip-gram with negative sampling モデルを学習した．同一のテキストデータから構築され

連絡先: 高田 理功，明石工業高等専門学校，兵庫県明石市魚住町西岡 679-3,e1319@s.akashi.ac.jp

*1 http://ahclab-server.naist.jp/resource/jppdb/
*2 https://github.com/tmu-nlp/simple-jppdb
*3 https://dumps.wikimedia.org/jawiki/20161001/
*4 http://taku910.github.io/mecab/

る複数の単語分散表現間の差について調査するため，初期値以外のパラメータを統一して単語分散表現の学習を二度（単語分散表現 A および B）行った．

3.1 コサイン類似度

単語分散表現 A と単語分散表現 B がどの程度一致するかを確認するため，日本語 Wikipedia から無作為抽出した 1 万種類の単語について，単語分散表現 A-B 間のコサイン類似度を測定した．単語分散表現の学習およびコサイン類似度の測定を 5 回実施し，その平均値を表 1 に示す．

表 1: 再現による各単語のコサイン類似度

実験回数	コサイン類似度
1 回目	-0.004
2 回目	-0.001
3 回目	-0.005
4 回目	-0.008
5 回目	-0.002
平均	-0.004

実験の結果，コサイン類似度は平均-0.004 であり，ベクトルがほとんど一致していないことがわかる．そのため，初期値の乱数を指定しない場合，単語分散表現を再現することは困難であると言える．

3.2 梶原らの再現実験

梶原らの手法を単語分散表現 A，単語分散表現 B それぞれを用いて再現した．単語分散表現 A および B それぞれを用いて再現実験を行うことで，分散表現によりどの程度予測に差が生じるのか確認することができる．単語分散表現 A および B を用いて再現実験した結果の accuracy を表 2 に示す．なお，単語の難易度推定の評価は日本語教育語彙表 [*5] のうち，単語分散表現 A および B に存在する 16,421 単語でテストを行っている．また，平易化判定の評価は PPDB: Japanese の 10best うち上記の難易度予測が可能であった単語 1 語ずつで構成されている 47,423 単語対でテストした．

表 2: 梶原らの再現実験の accuracy

手法	難易度推定	平易化判定
梶原ら [梶原 17]	0.709	0.607
単語分散表現 A	0.652	0.649
単語分散表現 B	0.651	0.651

再現実験によると，難易度推定，平易化判定ともに梶原らの手法とは 5 ポイント程度 accuracy が異なる．また，単語分散表現 A-B 間は accuracy に差がない．

次に，単語分散表現 A と単語分散表現 B の，難易度推定を行った予測単語のうち，予測が一致しない単語数を調査した．結果，単語分散表現 A-B 間で難易度予測が異なった単語数は，715 単語存在した．これは，16,421 単語中の約 4.4%に該当する．平易化判定も同様に予測が一致しない単語対数を調査したところ，判定結果が異なる対は 1,954 単語対存在した．これは，47,423 単語中の約 4.1%に該当する．よって accuracy に差が生じるとは限らないが，実験ごとに単語の予測難易度の一部が変化することがわかる．

*5 http://jhlee.sakura.ne.jp/JEV.html

4. 提案手法

先行研究である梶原らの研究では素性の 1 つに Word2Vec が用いられている．また，Word2Vec は 3 節で述べたように，学習の初期値に乱数を利用するため，実験の再現が困難である．そのため本稿では，実験の再現が容易なモデルを提案する．

本稿で提案するモデルは梶原らの手法 [梶原 17] を踏襲し，以下の手順で単語の平易化判定を行う．

[**難易度推定**] 言い換え元の単語と言い換え先の単語の難易度を推定する

[**平易化判定**] 言い換え元の単語と言い換え先の単語の難易度差を算出する

本稿では推定する難易度は 3 段階と 6 段階の 2 種類を用意した．3 段階難易度は初級，中級，上級であり，6 段階難易度の場合は 3 段階の各難易度を前半と後半に分ける．難易度推定の際は数値に対応付けるため，平易なほど小さい値をとるようにし，難解なほど大きい値を取るようにする．3 段階難易度の場合，難易度が初級のものを難易度 1 とし，上級のものを難易度 3 とした．6 段階難易度の場合，難易度が初級前半のものを難易度 1 とし，上級後半のものを難易度 6 とした．単語の平易化判定では，言い換え元の単語の難易度が，言い換え先の単語の難易度より大きい場合を平易，同じ場合を同等，小さい場合を難解とする．

単語の難易度推定モデルは，梶原らと同じ RBF カーネルによる SVM を用いる．また多クラス分類への拡張は one-versus-rest 法で行った．SVM は sklearn(0.19.1) の SVC[*6] を用いた．ハイパーパラメータである C と gamma はグリッドサーチを用いて最適解を求めた．

4.1 素性

素性は単語長，文字種 (ひらがな，カタカナ，漢字)，日本語 Wikipedia 中の単語単位の頻度，日本語 Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度を用いた．以下で各素性について説明する．

単語長とは単語の文字列としての長さであり，1 次元の素性である．単語長は，Flesch Reading Ease[Flesch 48] に単語長の平均が用いられているように，可読性を測る有用な素性になると考えられる．

文字種は以下の 2 パターンを用意した．どちらの場合も，素性の次元数は 3 次元である．

[**文字種の有無**] 単語に含まれるひらがな，カタカナ，漢字の有無 (1，0)

[**文字種の文字数**] 単語に含まれるひらがな，カタカナ，漢字の文字数

日本語 Wikipedia 中の単語単位の頻度とは，MeCab と mecab-ipadic-NEologd によって日本語 Wikipedia を単語分割した際の各単語の対数頻度で，素性の次元数は 1 次元である．頻度情報は，SemEval-2012 の English Lexical Simplification タスクで Baseline として用いられたが，優れたスコアを示したことなどから，有用な素性になると考えられる．Pavlick らの研究 [Pavlick 16]，梶原らの研究 [梶原 17] においても Wikipedia 中の単語単位の頻度を素性として利用している．なお，対数頻度において出現回数が 0 回のとき，対数頻度を 0 としている．

日本語 Wikipedia 中の文字単位の頻度は，単語に含まれる各文字単位の日本語 Wikipedia 中の頻度である．文字単位の

*6 http://scikit-learn.org/0.19/modules/generated/sklearn.svm.SVC.html

頻度は，平易な単語に用いられやすいひらがなやカタカナは頻度が大きくなると仮定する．一方漢字の頻度は低くなり，特に難解語に用いられる漢字ほど頻度が低くなると仮定する．

文字単位の頻度は，「子犬」という単語に対して，「子」「犬」それぞれに頻度が存在する．このように既出の各素性と異なり，単語長に依存して素性の数が変化する．しかし SVM の入力素性は固定長にしか対応しない．素性の数を単語長に依存させないようにするため，素性は文字単位の頻度の合計，平均，最大，最小の 4 パターンを用意し，対数化した後素性として加えた．よって素性の次元数は 4 次元である．

BCCWJ 中の単語単位の頻度は，現代日本語書き言葉均衡コーパス (BCCWJ)[*7][Maekawa 10] 中の単語単位の対数頻度である．BCCWJ は書籍全般，雑誌全般，ブログなどの様々なジャンルにまたがって無作為にサンプリングされた均衡コーパスである．BCCWJ 自体は限定公開であるため，BCCWJ に含まれる各ジャンルごとの頻度や全体における頻度などをまとめた『現代日本語書き言葉均衡コーパス』語彙表 ver1.0[*8] を用いて頻度を取得した．語彙表は 2 種類の単語区切りが公開されており，本稿では固有表現に対応する長単位を選択した．

語彙表に掲載されている情報のうち素性として用いるのは，ジャンルごとの頻度の対数とした．可変長と固定長があるものは合計された頻度のものを用いる．BCCWJ 中の単語単位の頻度は，13 種類のジャンルごとの頻度と，全体における頻度の合計 14 次元の素性となる．

4.2 素性の組み合わせ

本稿で用いる素性は単語長，文字種の有無と文字数，日本語 Wikipedia 中の単語単位の頻度，日本語 Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度である．しかし，これらの全てが有用な素性であるとは限らない．そこで SVM に入力する素性の構成を複数パターン用意した．素性の組み合わせとして用意した baseline1 および 2 と，提案手法 1 から提案手法 8 までを以下に箇条書きで紹介する．なお基本素性 1 は Wikipedia 中の頻度，単語長，文字種の有無で，基本素性 2 は Wikipedia 中の頻度，単語長，文字種の文字数である．

[**baseline1**] Wikipedia 中の頻度
[**baseline2**] 基本素性 1
[**提案手法 1**] 基本素性 2
[**提案手法 2**] 基本素性 1，文字単位の頻度
[**提案手法 3**] 基本素性 2，文字単位の頻度
[**提案手法 4**] Wikipedia 中の頻度，BCCWJ 中の頻度
[**提案手法 5**] 基本素性 1，BCCWJ 中の頻度
[**提案手法 6**] 基本素性 2，BCCWJ 中の頻度
[**提案手法 7**] 基本素性 1，文字単位の頻度，BCCWJ 中の頻度
[**提案手法 8**] 基本素性 2，文字単位の頻度，BCCWJ 中の頻度

baseline1 は単語の対数頻度のみを素性として用いた．基本素性 1 は梶原らの論文 [梶原 17] で提案されている基本素性と同じ素性である．梶原らの論文で有用な素性として発表されているため，これを基本素性 1 とした．また，基本素性 1 のみを素性とする手法を baseline2 とした．

基本素性 2 は梶原らの論文で用いられている文字種の真理値の代わりに文字種の文字数を素性としたものである．文字種の文字数を素性とすることで，単語長の情報を含んだ素性となり，accuracy の向上を期待する．

*7 http://pj.ninjal.ac.jp/corpus_center/bccwj/index.html
*8 http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html

基本素性 1 と基本素性 2 に対して，Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度をそれぞれ加えたものと加えないものを用意した．この素性の構成パターンにより，基本素性 1，基本素性 2，日本語 Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度それぞれが有用な素性か判断することができる．

5. 実験

提案手法で述べたモデルを用いて，実際に単語の難易度推定と平易化判定を行い，提案手法を評価する．また，再現が容易か調査した．

5.1 単語の難易度推定

4.2 節に示した各素性の組み合わせで，SVM を用いた単語の難易度推定を行う．本稿では教師データに日本語教育語彙表を用いた．日本語教育語彙表は日本語教育用の語彙である 17,920 単語の表記，読み，語彙の難易度などの情報を収録した表である．各語彙の難易度は 5 人の匿名教師により初級前半，初級後半，中級前半，中級後半，上級前半，上級後半の 6 段階難易度で設定されている．

推定する単語の難易度は 6 段階難易度と初級，中級，上級の 3 段階難易度の 2 種類である．評価は 10 分割交差検定を用いて行った．

日本語教育語彙表を用いた baseline1 および 2 と提案手法 1 から提案手法 8 の単語の難易度推定を行った．10 分割交差検定での難易度推定の accuracy を表 3 に示す．

表 3: 難易度推定の accuracy

手法	3 段階難易度	6 段階難易度
梶原ら [梶原 17]	**0.709**	-
baseline1	0.557	0.411
baseline2	0.572	0.420
提案手法 1	0.575	0.419
提案手法 2	0.587	0.432
提案手法 3	0.585	0.432
提案手法 4	0.624	0.458
提案手法 5	0.633	0.465
提案手法 6	0.636	0.470
提案手法 7	0.626	0.415
提案手法 8	**0.644**	**0.478**

表 3 からわかるように，日本語 Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度を素性として用いるのは有用である．特に提案手法 3 と提案手法 8 に着目すると，BCCWJ 中の単語単位の頻度を用いることで，BCCWJ 中の単語単位の頻度を含まない同様の手法より 5.9 ポイント accuracy が上がることを確認できる．提案手法 1 から提案手法 6 までを見ると基本素性 1 と基本素性 2 ではどちらを用いても accuracy が大きくは変わらないが，提案手法 7 と提案手法 8 には accuracy に大きな差が生じた．

最も優れた accuracy を示した手法は提案手法 8 で，3 段階難易度の場合 baseline1 より 8.7 ポイント，6 段階難易度の場合 baseline1 より 6.7 ポイント高い accuracy となっている．先行研究の梶原らの手法と比べると提案手法 8 は accuracy が 6.5 ポイント低い．しかし，Wikipedia 中に含まれない単語でも対数頻度を 0 として予測するため，より予測できる単語の総数は増加している．

5.2 単語の平易化判定

5.1 節の難易度推定の結果を利用して単語の平易化判定を行った．本稿では平易化判定のための単語対を PPDB: Japanese の 10best データから取得した．PPDB: Japanese の 10best データのうち，言い換え先と言い換え元が日本語教育語彙表に載っていて，かつ単語 1 語で構成されている単語対を評価に利用する．以上の手順で取得できた単語対は 48,646 対存在する．平易化が可能か判定するため，単語の難易度差を言い換え先より平易，同等，難解の 3 段階で算出した．なお難易度は 3 段階難易度により平易化判定したもの，6 段階難易度により平易化判定したものの 2 種類を用意した．評価データは日本語教育語彙表による難易度差を利用する．表 4 に，以上の方法で単語の平易化判定を評価した結果を示す．

表 4: 平易化判定の accuracy

手法	3 段階難易度	6 段階難易度
梶原ら [梶原 17]	**0.607**	-
baseline1	0.614	0.597
baseline2	0.625	0.600
提案手法 1	0.624	0.597
提案手法 2	0.625	0.578
提案手法 3	0.625	0.576
提案手法 4	0.661	0.652
提案手法 5	0.668	0.658
提案手法 6	0.668	0.658
提案手法 7	0.670	0.625
提案手法 8	**0.672**	**0.662**

単語の難易度推定と同じく，最も優れた accuracy が得られたのは提案手法 8 であった．3 段階難易度の場合 baseline1 より 5.6 ポイント，6 段階難易度の場合 6.5 ポイント向上したことがわかる．梶原らの手法は 3 段階難易度による平易化判定で accuracy が 0.607 であるため，提案手法 8 は梶原らの手法と比べて accuracy が 6.5 ポイント改善していることがわかる．ただし本稿では，対数頻度 0 の単語の難易度を予測することでより多くの単語対に難易度を付与している．そのため，baseline を含め先行研究と単純には比較できない．

また，BCCWJ 中の頻度を素性とした判定が，BCCWJ 中の頻度を用いない手法より 4 ポイント程度高い accuracy を示すことがわかる．基本素性 1 と基本素性 2 では最大でも 0.3 ポイントしか accuracy に差が生じないため，文字種の文字数を素性としても accuracy は向上しない．

5.3 再現の容易性の評価

我々は accuracy が高いだけでなく，再現性が高いモデルの構築を目指している．そのため，最も accuracy の高かった提案手法 8 を用いて，全く同じ実験を 5 回繰り返し再現の容易性を確かめた．

結果，5 回とも難易度推定において全ての単語の予測難易度が一致した．また，平易化判定においても難易度推定と同様に判定が一致している．本稿で用いた素性は一切の確率的な変動が存在しないため，容易に再現が可能であることがわかる．

6. おわりに

本稿では難易度推定と平易化判定共に，baseline より accuracy を 5.8 ポイント以上向上させ，かつ再現が容易な手法を提案した．また素性として，日本語 Wikipedia 中の文字単位の頻度，BCCWJ 中の単語単位の頻度が有用であることを示した．先行研究と比較すると，平易化判定は梶原らの手法 [梶原 17] より accuracy を 6.5 ポイント改善した．梶原らの構築した平易化用の辞書の改善など，本研究が今後活躍することを期待する．

参考文献

[Flesch 48] Flesch, R.: A new readability yardstick, *Journal of Applied Psychology*, Vol. 32, pp. 221–233 (1948)

[Ganitkevitch 13] Ganitkevitch, J., Durme, B. V., and Callison-Burch, C.: PPDB: The Paraphrase Database, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 758–764 (2013)

[Maekawa 10] Maekawa, K., Yamazaki, M., Maruyama, T., Yamaguchi, M., Ogura, H., Kashino, W., Ogiso, T., Koiso, H., and Den, Y.: Design, Compilation, and Preliminary Analyses of Balanced Corpus of Contemporary Written Japanese, in *Proceedings of the Seventh conference on International Language Resources and Evaluation*, pp. 1483–1486 (2010)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Advances in neural information processing systems*, pp. 3111–3119 (2013)

[Mizukami 14] Mizukami, M., Neubig, G., Sakti, S., Toda, T., and Nakamura, S.: Building a Free, General-Domain Paraphrase Database for Japanese, in *Proceedings of the 17th Oriental COCOSDA Conference*, pp. 129–133 (2014)

[Pavlick 15] Pavlick, E., Rastogi, P., Ganitkevitch, J., Van Durme, B., and Callison-Burch, C.: PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Vol. 2, pp. 425–430 (2015)

[Pavlick 16] Pavlick, E. and Callison-Burch, C.: Simple PPDB: A paraphrase database for simplification, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2, pp. 143–148 (2016)

[梶原 17] 梶原 智之, 小町 守：Simple PPDB: Japanese, 言語処理学会第 23 回年次大会, pp. 529–532 (2017)

[佐藤 17] 佐藤 敏紀, 橋本 泰一, 奥村 学：単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討, 言語処理学会第 23 回年次大会, pp. 875–878 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Country Adaptation in Neural Machine Transliteration of Person Names

Dawoon JUNG Satoshi SATO

Graduate School of Engineering, Nagoya University

We have developed a neural machine transliteration system that can transliterate person names in many countries into Japanese. This system is constructed by two-step training. First, a base model is trained on a large country-unknown person-name dataset. Then, country-specific models are constructed from the base model by applying transfer learning with a set of small country-spcific datasets. An experiment shows that our system outperforms the previous CRF-based system.

1. Introduction

Machine transliteration is the task of automatically transforming the script of a word from a source language to a target language, while preserving the original pronunciation [1]. Typically, transliteration is used for translation of proper nouns, such as person names, place names, and company names.

The 2020 Olympic & Paralympic Games will be held in Tokyo, Japan. Since the alphabet of Japanese language is different from Latin alphabet, the local committee has to transliterate the official participant list written in Latin alphabet into the Japanese one written in Katakana. This transliteration task is heavy since the expected participants are more than 10,000 people from more than 200 countries. Thus, machine transliteration aid is needed to reduce the human effort.

A number of machine learning based methods for machine transliteration have been developed in the past. These methods require a large amount of training data for each language pair. This requirement becomes a problem when we have to deal with many language pairs. To overcome this problem, Yasue et al. [2] have proposed a two-step construction of transliterators, where they first construct a base transliterator by using a large country-unknown dataset and then, for each country, construct a country-specific transliterator by re-training the base one using a small country-specific dataset.

Recently in natural language processing, neural networks have been widely used because of their performance and simplicity, including machine transliteration [3]. In this paper, we propose a neural machine transliteration system using the transfer learning technique, which can transliterate person names of many countries into Japanese for the compilation of participant lists of Olympic games written in Japanese. We focus on 'countries' instead of 'languages', because of two reasons: (1) nationalities of all participants are provided by IOC (International Olympic Committee), and (2) it is difficult to determine the language from one single name (e.g. the language of the name 'Peter' can be English, German, or Italian).

Contact: Dawoon JUNG, Dept. of Information and Communication Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, jung.dawoon@c.mbox.nagoya-u.ac.jp

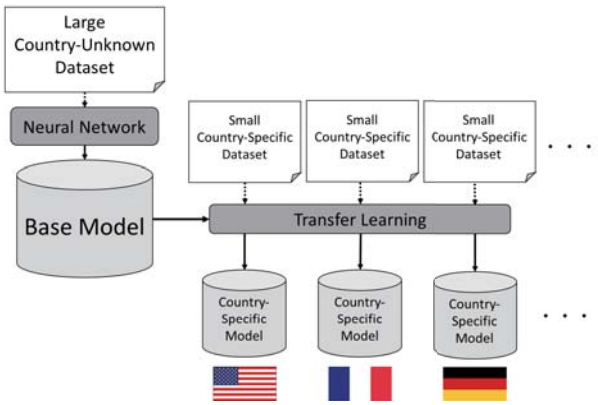


Figure 1: An overview of proposed system

2. Proposed System

Figure 1 shows an overview of the proposed system. The system is constructed by two-step training.

(1) The base model is trained on a large country-unknown person-name dataset.

(2) For each country, a country-specific model is constructed from the base model by applying transfer learning with a small country-specific dataset.

2.1 Base Model

We build our base model based on a Seq2Seq framework. Our model takes a person name (e.g., "michael") as a Latin character sequence and produces the transliterated one (e.g., マイケル) as a Katakana character sequence.

Tacotron [4] is an end-to-end neural text-to-speech synthesis model which takes characters as input and outputs the corresponding raw spectrogram. We use Tacotron since both text-to-speech synthesis and machine transliteration deal with pronunciation. We adapt Tacotron to machine transliteration with two modificatons. First, the output layer is modified to produce a character sequence as output, according to [5]. Second, CBHG module is simplified by omitting highway network, because the length of an input sequence is much shorter than that of the original Tacotron.

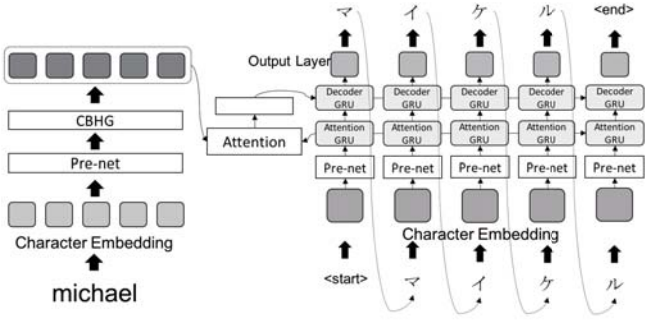


Figure 2: The architecture of the base model

Figure 2 shows the proposed model. It can be divided into the encoder and the decoder. The encoder consists of input-character embedding, pre-net, and CBHG module. The decoder consists of output-character embedding, pre-net, attentional GRU, decoder GRU, and output layer.

2.2 Country-Specific Model

A country-specific model is a fine-tuned model for a specific country. We bulid a country-specific model by applying transfer learning to the base model. Transfer learning enables us to leverage the pre-learned knowledge from related tasks, which leads to the reduction of training data [6]. For example, Zoph et al. [7] demonstrated that training only certain parameters of a pre-trained model can improve the performance of low-resource machine translation. Similarly, we construct a country-specific model by re-training only certain parameters of the base model. In Seq2Seq framework, the encoder produces a vector representation from an input sequence and the decoder produces an output sequence from the representation. In our case, since the pronunciation varies from country to country while the target language is fixed as Japanese, the encoder would be especially involved in learning the pronounciation of input sequences. Therefore, certain parameters of the encoder should be re-trained.

3. Postprocessing

To improve the quality of the final outputs of the transliteration system, we further apply a set of heuristic rules for standardizing Katakana person names in the postprocessing stage. Examples are shown below.

1. Non-standard Katakana characters are standardized; e.g., (ヲ → オ), (ヅ → ズ) (ヂ → ジ)
2. Non-standard use of small characters (小書き文字) is standardized; e.g., (レァ → レア), (シィ → シー), (デェ → デー), (タィ → タイ)
3. Unnecessry consecutive characters are omitted; e.g., (ーー → ー), (ンン → ン)

Table 1: Evaluation of base model with country-unknown dataset

	Top5 Accuracy	MRR
Proposed Model	**76.41**%	**0.537**
CRF	74.49%	0.511
Seq2Seq	71.56%	0.506

(*:tested with 13,611 examples of a country-unknown dataset)

Table 2: Evaluation of base model with a set of country-specific datasets

	Top5 Accuracy	MRR	The Number of Countries Over 85%
Proposed Model	**81.23**%	0.659	**106**/202
CRF	80.89%	**0.668**	103/202
Seq2Seq	76.56%	0.506	80/202

(*:tested with 31,629 examples of a set of country-specific datasets)

4. Experiments

4.1 Datasets

We assess our proposed system on Latin alphabet to Katakana transliteration. We used two types of datasets. First one is a large country-unknown dataset, which contains 136,107 person names. This dataset was collected by web-crawling method [8]. Second one is a set of small country-specific datasets that created manually by human translators, which are provided by Jiji Press Ltd. These datasets contain 31,629 person names in 202 countries; the number of examples in a country-specific dataset varies from less than 50 to more than 1000.

4.2 Evaluation of Base Model

First, we evaluated the base model. For the construction, we used 80% of the country-unknown dataset for training, and 10% for development. For evaluation, we used the rest 10%. As the baseline systems, we built a CRF-based model [8] and a simple Seq2Seq model [9].

As evaluation metrics, we used top-5 accuracy and MRR (Mean Reciprocal Rank) since a single person name can have mutiple transliterations in general. For example, 'Murray' is ambiguous as 'マリー (ma-ri-i)' and 'マレー (ma-re-e)', both of which are widely used in Japan. We also set the development goal of the performance to 85% Top-5 accuracy for each country.

The evaluation result is shown in Table 1. The proposed base model achieved 76.41% top-5 accuracy and outperformed the two baseline models. An additional advantage of the proposed base model is that the construction of the model does not require the local alignment of transliteration pairs, which is required by the CRF-based model.

We also examined the performance of the base model for a set of country-specific datasets. Table 2 shows the result. The proposed base model achieved 81.23% top-5 accuracy and outperformed the two baseline models for these

Table 3: Settings of re-training parameters

Parameter			Model (1)	(2)	(3)	(4)	(5)
Encoder	Input Character Embedding		✓	✓	✓	✓	✓
	Encoder Pre-net		✓	✓	✓	✓	
	CBHG Module	Parameters Except Encoder GRU	✓	✓	✓		
		Encoder GRU	✓	✓			
Decoder	Output Character Embedding		✓				
	Decoder Pre-net		✓				
	Attentional GRU		✓				
	Decoder GRU		✓				
	Output Layer		✓				

Table 4: Evaluation of country-specific models; Model (1) ∼ (5) are re-trained models that follow the settings shown in Table 3.

Model	Top5 Accuracy	MRR	The Number of Countries Over 85%
CRF	87.11%	0.739	56/83
Model (1)	83.95%	0.702	39/83
Model (2)	87.51%	**0.744**	53/83
Model (3)	**87.99**%	**0.744**	**58**/83
Model (4)	85.32%	0.692	39/83
Model (5)	85.17%	0.693	44/83

(*:tested with 29,482 examples of a set of country-specific datasets that contain more than 50 examples.)

datasets. The higher accuracy (81.23%) than the result in Table 1 (76.41%) is due to the overlap between these country-specific data and the training data.

4.3 Evaluation of Country-Specific Models

Second, we evaluated country-specific models. Applying transfer learning to the base model constructed in the previous subsection, we built 83 country-specific models using country-specific datasets, each of which contains more than 50 examples. We examined five settings of the re-training parameters, shown in Table 3. The performance measured by 5-fold cross-validation and the baseline is the re-trained CRF model.

Table 4 shows the overall performance of each setting and the baseline; Table 5 shows the performance of each country, where country name is presented by its IOC code. Model (3) achieved the best performance of 87.9% top-5 accuracy, which outperformed the re-trained CRF model. This model also achieved over 85% top-5 accuracy in 58 countries among 83 countries.

Table 4 shows that it is better to re-train certain parameters of the encoder (Model (2) and (3)) than to re-train all the parameters (Model (1)). This result implies that transfer learning method can avoid overfitting by restricting the

Table 5: Evaluation of country-specific models; The column '#' shows the number of examples.

	#	Top5 Acc.		#	Top5 Acc.
PHI	52	96.15	BRN	89	88.76
KGZ	97	95.88	HUN	362	88.40
BUL	228	94.74	MAS	136	88.24
PER	54	94.44	SLO	305	88.20
ESP	632	94.30	CAN	1167	88.00
COL	278	94.24	BRA	628	87.90
JAM	262	93.89	KSA	74	87.84
ARG	287	93.73	SVK	218	87.61
AUS	982	93.58	SIN	113	87.61
ITA	922	93.49	AUT	505	87.13
VEN	197	93.40	UKR	701	86.73
NGR	175	93.14	SUI	583	86.28
MEX	318	93.08	GER	1408	86.22
DOM	85	92.94	FIN	362	86.19
UZB	297	92.93	MAR	181	86.19
TJK	70	92.86	MDA	94	86.17
ISR	181	92.82	RSA	330	84.85
IND	318	92.45	EGY	171	84.80
NZL	421	91.92	FRA	1220	84.75
GBR	1000	91.70	GRE	307	84.36
USA	2503	91.33	SWE	453	83.66
POR	184	91.30	ARM	108	83.33
AZE	218	91.28	IRL	168	82.74
PUR	68	91.18	NED	640	82.50
EST	135	91.11	BLR	472	81.99
ROU	392	91.07	TUN	116	81.90
KAZ	491	91.04	CZE	424	81.84
CRO	212	91.04	PRK	82	81.71
RUS	1558	90.76	ALG	103	81.55
KEN	511	90.61	LTU	178	81.46
CUB	350	90.57	IRI	251	80.48
INA	125	90.40	NOR	476	79.83
BAH	83	90.36	VIE	69	79.71
CHI	90	90.00	LAT	226	79.65
TTO	79	89.87	POL	583	79.42
GEO	167	89.82	CMR	63	79.37
ETH	330	89.70	MGL	192	78.13
GUA	58	89.66	DEN	260	76.92
QAT	113	89.38	TUR	331	75.23
ECU	93	89.25	BEL	272	70.96
UGA	55	89.09	THA	180	68.89
SRB	210	89.05	TOTAL	29482	87.99

(*:tested with 29,482 examples of a set of country-specific datasets that contain more than 50 examples.)

Table 6: The effect of transfer learning

Name	Katakana Name	Predicted Name by Base Model
altansu**kh**	アルタンスフ	アルタンスク
mun**kh**	ムンフ	マンク
narmanda**kh**	ナルマンダフ	ナーマンダク
nazmanda**kh**	ナズマンダフ	ナズマンダク
tuvshinbaya**r**	ツブシンバヤル	タブシンバヤー
uuganbaata**r**	ウーガンバートル	ウーガンバーター
tsogtbaata**r**	ツォグトバートル	ツォグバーター
tuvshinbaya**r**	ツフシンバヤル	タブシンバヤー

number of re-training parameters. However, the effect of transfer lerning is limited if the number of re-training parameters is too small (Model (4) and (5)).

From Table 5, you can see that the performance varies with the country. The country with the best performance is Philippines (PHI) and the country with the worst performance is Thailand (THA).

4.4 Discussion

First, we examined the effect of transfer learning by comparing some names predicted by Model (3) with those by the base model.

Table 6 shows some examples that Model (3) can produce the correct transliterations, whereas the base model cannot. These examples are person names of Mongol (MGL), which showed the largest improvment of performance by transfer learning. You can see that incorrect substrings such as (kh →ク) and (r →ー) have been corrected as (kh →フ) and (r →ル). This implies that country adaptation has been achieved through transfer learning.

Second, we examined the countries that could not obtain good performances even if we built country-specific models by transfer learning. Table 7 shows the performances of five countries with lowest performance. These countries can be classified into three groups.

1. Countries that use alphabets different from Latin alphabet: Thailand (THA) and Mongol (MGL). Some important hints for pronunciation were disappeared in the process of converting the original spelling into the Latin-alphabet spelling.

2. Countries that use a language that has pronunciation far from well-known English-based pronunciation: Turkey (TUR) and Denmark (DEN).

3. Multilingual countries: Belgium (BEL). It is difficult to predict pronunciation by our approach that uses a country instead of a source language.

The performance of Group 1 and 2 may be improved by addtional country-specific traning examples. However, multilingual countries cannot be handled by our approach.

Table 7: Results of five low rank countries

	Top5 Accuracy	Accuracy Improvement
THA	68.89%	+7.23%
BEL	70.96%	+0.74%
TUR	75.23%	+9.07%
DEN	76.92%	+7.31%
MGL	78.13%	+23.97%

5. Conclusions

In this paper, we proposed a neural person-name transliteration system that can transliterate person names of many countries into Japanese. Experiments showed that the proposed model outperformed the baseline system. For future work, it is required to improve the performance of multilingual countries and extremely low-resource countries that could not apply transfer learning.

References

[1] Sarvnaz Karimi, Falk Scholer, and Andrew Turpin. Machine Transliteration Survey. ACM Computing Surveys, Vol. 43, No. 3, 2011.

[2] 安江祐貴, 佐藤理史. 外国人名カタカナ表記自動推定における各国適応. 人工知能学会第 30 回年次大会論文集, 2016.

[3] Amir H. Jadidinejad. Neural Machine Transliteration: Preliminary Results. arXiv, 2016.

[4] Yuxuan Wang, RJ Skerry-Ryan, and Daisy Stanton et al. Tacotron: Towards End-to-End Speech Synthesis. arXiv, 2017.

[5] Kyubyong Park. Neural Japanese Transliteration —can you do better than SwiftKey™ Keyboard?. `https://github.com/Kyubyong/neural_japanese_transliterator`, 2017.

[6] Lisa Torrey and Jude Shavlik. Transfer learning. IGI Global, 2009.

[7] Barret Zoph, Deniz Yuret Jonathan May, and Kevin Knight. Transfer Learning for Low-Resource Neural Machine Translation. ACL, 2016.

[8] Satoshi Sato. Crawling English-Japanese Person-Name Transliterations from the Web. WWW, 2009.

[9] Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. ICLR, 2014.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

シーケンスアライメントを用いた記述式問題の採点支援システムの提案

Scoring support system using sequence alignment

高井 浩平 *1 Kohei Takai　竹谷 謙吾 *1 Kengo Taketani　森 康久仁 *2 Yasukuni Mori　須鎗 弘樹 *2 Hiroki Suyari

*1千葉大学 工学部 情報画像学科
Department of Informatics and Imaging Systems, Faculty of Engineering, Chiba University

*2千葉大学 大学院工学研究院
Graduate School of Engineering, Chiba University

In the university unified entrance examinations from 2020, a new kind of question to require answers in description style will be introduced. In order to reduce the cost of scoring these answers, automatic scoring system is expected to be realized. In this paper, we propose a scoring support system using sequence alignment with high precision.

1. はじめに

2020 年度より大学入試センター試験に短答記述式問題が導入される．記述問題の採点には既存のマークシートタイプの自動採点システムは使えないため，膨大な数（2017 年度の受験者数は 55 万人弱）の解答文を人の手で採点する必要がある．採点にかかる時間や人件費といったコストは計り知れない．そのため，記述式問題に対する採点をどのように行うかがよく議論されている [中川，平井 16]．

機械学習を用いた採点システムの提案もされているが，機械学習を用いた採点システムは「採点前に学習のためのデータセットを揃えるのが困難である」,「機械学習による判断は，たとえ 90%の高精度でも，母集団の大きさや受験生の進路に関わる精度としては不十分である」といった問題点が多い．

そこで本論文では，採点のコスト削減を目的として，シーケンスアライメントを用いた採点支援システムを提案する．単語の並びを考慮した条件キーフレーズを採点前に設定し，そのキーフレーズに基づいてシーケンスアライメントを用いることで採点の支援を行う．大学入学共通テストでは，国語と数学に記述式問題を導入することになっているが，本手法では，まずは社会科の記述式問題を対象とする．実験では 4 つの解答データに対して採点を行い，採点の精度・自動採点率を評価する．また，採点支援システムの精度を向上させるための手法も提案する．

2. 関連研究

筆記試験の採点システム作成を試みた関連研究としては，Visual Basic で作成した採点システムの試作 [五月女 17] などがある．[五月女 17] は生徒の解答を採点者に提示し，点数ではなく正誤の判断をさせるものであった．手作業よりも作業時間の短縮が可能になっているが，あくまで採点は手動で行うもので自動採点ではない．

採点を正解・不正解の 2 値分類問題と捉え，ニューラルネットワークを用いた研究 [寺田，久保，柴田，黒橋，大久保 16] や，採点基準に基づく採点と機械学習による採点予測を組み合わせた採点支援システム [亀田，石岡，劉 17] なども研究されている．しかし，正解・不正解の教師データを人の採点により与える必要があり，より良い精度を得るためには，大量の採点済の教師データを与える必要がある．

連絡先: 高井浩平，千葉大学 工学部 情報画像学科，263-8522
千葉市稲毛区弥生町 1-33，e-mail : aena2593@chiba-u.jp

3. 提案手法

3.1 採点支援システム

通常，記述式問題の採点には，採点基準があり，文字数制限，使うべきキーワードの指定など，いくつかの条件が与えられていることがほとんどである．それを踏まえて，提案する採点支援システムの採点フローを 図 1 に示す．出力は正解・不正解の 2 値とし，自動採点が行えなかったものに対しては手動採点を行う．

3.1.1 【初期設定】解答条件・条件キーフレーズ

解答条件として，必須単語，文字数制限を設定する．条件キーフレーズとは，正解判断にとって重要な単語を含む短いフレーズのことである．条件キーフレーズによる採点を行うことで，重要な単語とその並び，自然な文かどうかを判断する．また，これにより，部分点の付与も可能になる．採点前にこの条件キーフレーズを，採点者が複数設定する（表 1）．

表 1: 条件キーフレーズの設定例

条件 1	条件 2
家柄関係なく	才能や功績のある人
家柄にとらわれず	功績や才能のある人
家柄によらず	才能ある人や功績ある人
家柄を気にせず	功績ある人や才能ある人
家柄を重視せず	才能や功績がある優秀な人

3.1.2 【照合】正解/不正解リスト

採点が済んだ解答を採点後に正解/不正解リストに蓄積していく．新たに採点を行う解答文と一致するものがリストに存在していれば，即座に正解/不正解を判断し採点を終了する．同一文に対する処理を省略することで採点時間を短縮している．

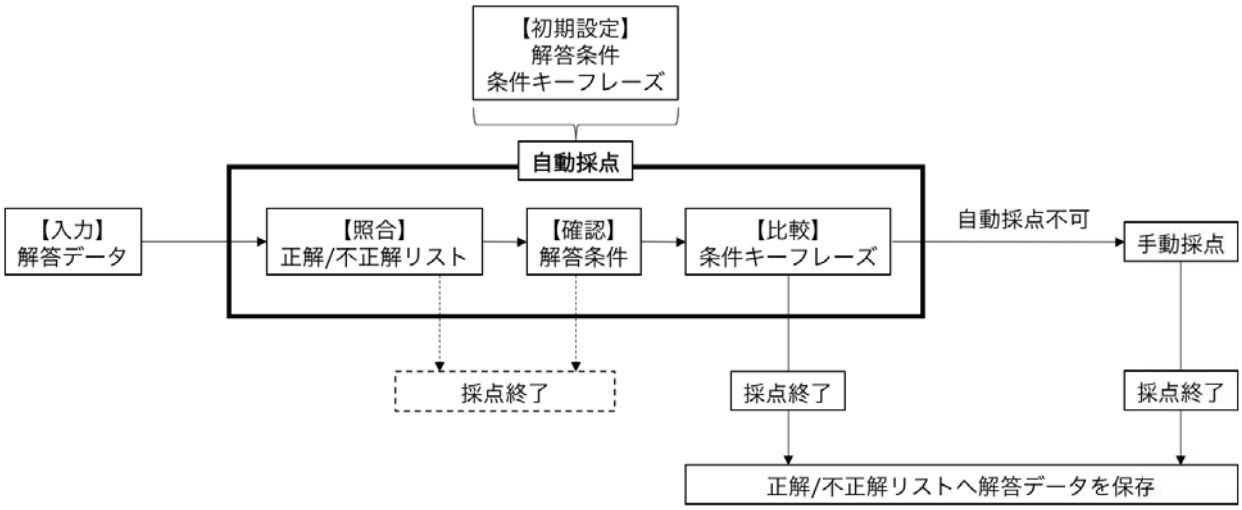


図 1: 採点支援システムの採点フロー

3.1.3 【確認】解答条件

必須単語が含まれているか，文字数制限を超えていないかを確認する．この解答条件が満たされていない場合，即座に不正解と判断し採点を終了する．

3.1.4 【比較】条件キーフレーズ

解答文と条件キーフレーズを比較する．解答文は形態素解析エンジン MeCab[*1] を使用して形態素解析を行い，文書を単語に分かち書きする．この際，名詞，助詞，形容詞のみを抽出する．重要な意味を持つ単語のみを抽出することにより，文書の持つ意味を大きく損なうことなく情報量を削減することができる．比較手法にはシーケンスアライメントを用いた．

3.1.5 シーケンスアライメント

シーケンスアラインメントとは，主にバイオインフォマティクスの分野で利用され，2 つ以上の DNA や RNA などの配列において類似した領域を特定できるように並び替える技術である．この技術を分かち書きされたテキストデータに応用することで，2 つのテキストの比較を行った．

本手法では，シーケンスアライメントアルゴリズムの 1 つである Needleman-Wunsch Algorithm [Needleman, Wunsch 70] を用いる．形態素解析を行なった文章に対してこのアルゴリズムを適用することで，解答文と条件キーフレーズとの共通部分・非共通部分を抽出することができ，単語の並びに着目して解答文と条件キーフレーズとの一致率を計算することができる（図 2）．これにより，設定されているいくつかの条件キーフレーズの中から，最も類似したものを選択することができる．表 1 の場合，条件 1，条件 2 の両方の条件で類似する条件キーフレーズが選択された場合，正解と判断する．

また，文字変換プログラム/辞書 KAKASI[*2] と，形態素解析エンジン MeCab を用いて，シーケンスアライメントにより抜き出された非共通部分の表記揺れの補正を行う．単純な文字列の比較では，同一の単語であっても漢字とひらがなのように表記が異なる場合や，形容詞・動詞の活用形により表記が異なる場合について正しく採点を行うことができない．KAKASI による読み方に着目したローマ字文への変換と，MeCab による基本形への変換により，表記方法に依存せずに単語の意味による採点を行うことができる．

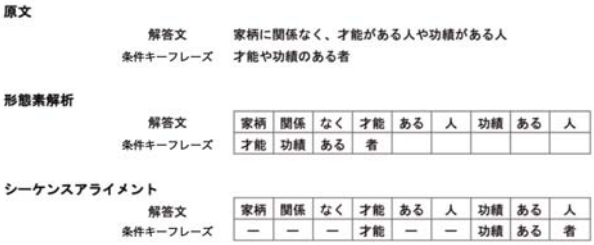


図 2: シーケンスアライメントによる比較

さらに，概念辞書 WordNet[*3] を使用して，シーケンスアライメントにより抜き出された非共通部分の意味を比較する．Wordnet は単語を類義関係のセット (synset) でグループ化しており，一つの synset が一つの概念に対応している．解答文の条件キーフレーズとの非共通部分について，それぞれの単語の synset を比較し，同じ概念の単語であると判断した場合に解答文の該当する単語を条件キーフレーズの該当する単語で置換し，再度比較を行う．単語の意味に着目して比較を行うため，少ない数の条件キーフレーズで広範囲の解答を採点することができる．

3.2 条件キーフレーズ生成による精度向上の提案

提案する採点支援システムは採点前に設定される条件キーフレーズに大きく依存しており，設定される条件キーフレーズが不十分だと自動での採点が上手く行えないという問題がある．

そこで，生徒の正解文を集めた正解文リストからクラスタリング・アライメントを用いて，正解文の共通・非共通部分を抽出し，採点者が採点前に想定することができなかった条件キーフレーズの抽出を行う．

∗1 http://taku910.github.io/mecab/
∗2 http://kakasi.namazu.org
∗3 http://compling.hss.ntu.edu.sg/wnja/

正解文リストに対してクラスタリングを行い, 文章の類似度によっていくつかのクラスタに分類する（図 3）. クラスタリングの際，文書のベクトル化は Bag-of-Words モデルを用いた.

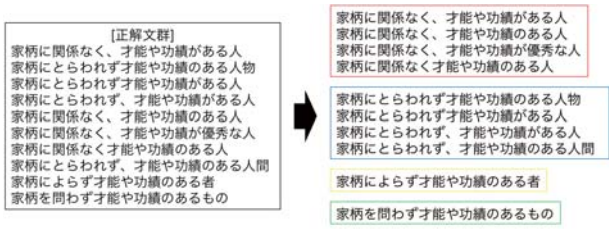


図 3: 正解文群のクラスタリングの例

そして, クラスタ内の文章にアライメントを行うことで, 文の共通部分･非共通部分を抜き出し, 新しく出現したキーフレーズを出力する（図 4）.

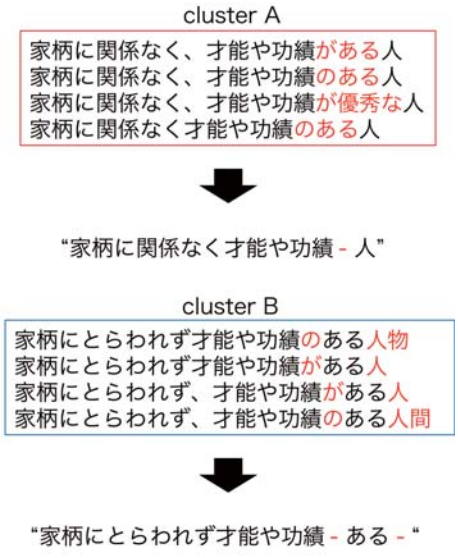


図 4: 正解文クラスタに対するアライメントの例

非共通部分が抽出された場合 WordNet を使用し，入力されたワードの類義語を入力することで複数のキーフレーズを生成する（図 5）.

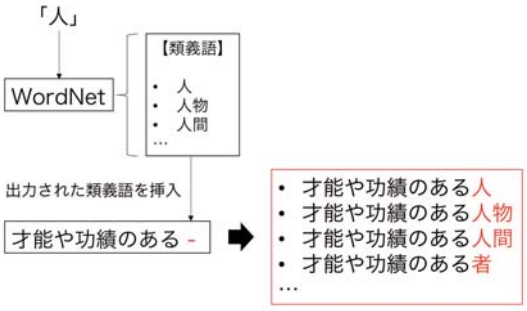


図 5: WordNet を用いた生成の例

生成された条件キーフレーズを次回採点時に初期条件に追加することで，正解文に対する自動採点率が向上し，手動採点の割合が減少すると予想している（図 6）.

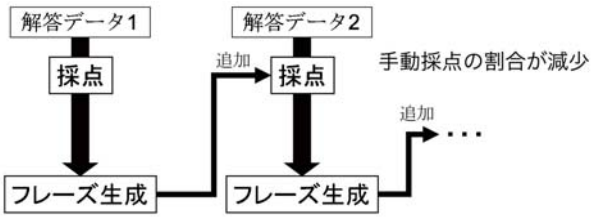


図 6: 精度向上の予想

4. 評価実験

4.1 実験用データ

大学生を中心に 68 名に「千葉の公立高校における社会の記述式問題」を 4 題解いて貰い, 実験用データとして使用した (図 7) . 解答は決められた単語を使用することと, 約 20 から 30 字程度の文字数制限が課される.

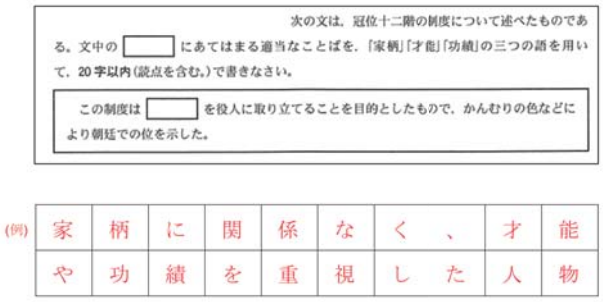


図 7: 問題と解答の例

4.2 結果・考察

各問題に対する採点支援システムの採点結果を 表 2 に示す.

表 2: 採点結果

採点指標 (%)	問題 1	問題 2	問題 3	問題 4
精度	100.0	100.0	100.0	100.0
自動採点率	30.0	38.0	39.0	43.0
正解文自動採点率	47.5	76.2	41.1	38.9
不正解文自動採点率	7.1	21.3	33.3	61.5

全ての問題に対して誤採点のない 100%の精度で採点が行えた．非常に信頼性の高い採点を行えていることが分かる．採点精度が 100%であれば一度採点した問題を再度採点する必要がないため，採点コストの削減に繋がることになる.

正解文に対する自動採点率が高く，不正解文に対する自動採点率が低いものは，正解については判定基準として正答を用意しているが，不正解文については “正解ではないと断定できる” という判断基準しか存在しないためである．不正解を自動で判別できる適切な条件キーフレーズを設定できれば，自動採点率は大幅に向上すると考えられる.

問題によって自動採点率にばらつきが生じているのは，正答の設定に問題があることや，問題の難易度による正解/不正

解数のばらつきが原因であると思われる．人手で設定した条件キーフレーズの数が少なかったことや，用意した解答データが 68 人分と少なかったことを考えると，設定する条件キーフレーズの数を増加させることや，より多くの解答データを用意することにより自動採点率は大きく向上すると考えられる．

また，3.2 節で述べた，条件キーフレーズ生成手法を行った結果を 図 8 に示す．

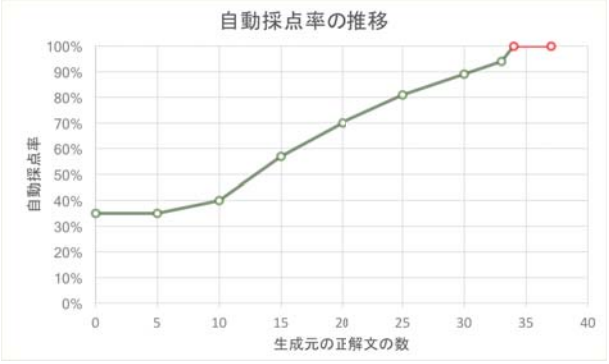


図 8: 問題 1 の正解文自動採点率の変化

問題 1 の解答文から条件キーフレーズ生成を行い，初期条件に生成された条件キーフレーズを追加した．その後，再度，問題 1 の解答データに対して採点を行った結果である．

横軸は “生成の元となった正解文の総数”, 縦軸は “正解文をシステムが自動で採点した割合を示す自動採点率 (%) ” ．結果から，生成元の正解文数が増える毎に採点率が向上していることがわかる．正解文のうち 34 文を使用した段階で，自動採点率が 100%に達した．

生成に使用した解答データに対して再度採点を行っているので，正解自動採点率が 100%になるのは当然のことではあるが，抽出/生成したキーフレーズが, 正答判断にとって重要なキーフレーズであったと言える．また，全文からの生成でなくても精度の向上が見込めることが示せた．さらに大規模な解答データに対しても検証する必要がある．

5. おわりに

大学入学共通テストに記述式問題が導入されることを背景として，シーケンスアライメント等を用いた文字列比較をベースとする自動採点を行うことによる採点支援手法を提案した．従来の自動採点及び採点支援の手法と比較すると，精度が非常に高く，人間が再度採点を行う必要が生じないため，採点の信頼性においては優位性があると考えられる．また，機械学習を利用しない手法であるため，大学入学共通テストのような大規模なテストだけでなく，比較的小規模な解答データに対する採点支援システムとしての利用も見込める．

課題として，さらなる自動採点率の向上や，より大規模なデータでの評価が挙げられる．今回実験に使用したデータが 68 名分と少ないため，解答が数百から数千集まるような実際の採点でどのような結果になるかが確認できなかった．また，精度向上が期待できる提案手法に関しても，データが少ないため満足な検証が行えていない．さらに解答データを集める必要がある．

実際の採点を考慮すると，精度 100%でないと実用化は困難であると考えているので，精度を落とすことなく自動採点率を上げることが重要である．

参考文献

[中川，平井 16] 中川 正樹，平井 佑樹 ：記述式解答の採点支援・自動採点に向けて -手書き認識からの挑戦-，情報処理学会，情報処理 57 巻 9 号，pp.920-924，2016

[五月女 17] 五月女 仁子 : Visual Basic で作成した採点システムの試作，日本女子体育大学紀要，第 47 巻，2017

[寺田，久保，柴田，黒橋，大久保 16] 寺田凛太郎，久保顕大，柴田知秀，黒橋禎夫，大久保智哉 : ニューラルネットワークを用いた記述式問題の自動採点，言語処理学会，第 22 回年次大会 発表論文集，pp.370-373，2016

[亀田，石岡，劉 17] 亀田雅之，石岡恒憲，劉東岳 : 短答記述式問題解答文の採点支援システム JS4 の試作，言語処理学会，第 23 回年次大会 発表論文集，pp.1137-1140，2017

[Erkan, Radev 04] Gunes Erkan, Dragomir R. Radev : LexRank: Graph-based Lexical Centrality as Salience in Text Summarization, Journal of Artificial Intelligence Research 22，pp.457-479，2004

[Yogatama, Liu, Smith 15] Dani Yogatama, Fei Liu, Noah A. Smith : Extractive Summarization by Maximizing Semantic Volume, The 2015 Conference on Empirical Methods on Natural Language Processing(EMNLP 2015)，pp.1962-1966，2015

[Needleman, Wunsch 70] Needleman, Saul B. and Wunsch, Christian D : A general method applicable to the search for similarities in the amino acid sequence of two proteins., Journal of Molecular Biology, 48, pp.443-453, 1970

Oral presentation | Organized Session | [Organized Session] OS-6

[2L2-OS-6a] [Organized Session] OS-6

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room L (3F Sapphire Hall Asuka)

[2L2-OS-6a-01] Emerging Interactions

○Ichiro Tsuda[1] （1. Chubu University）

1:20 PM - 1:40 PM

[2L2-OS-6a-02] Spatial Concept Formation-based Semantic Mapping

○Yuki Katsumata[1], Akira Taniguchi[1], Yoshinobu Hagiwara[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）

1:40 PM - 2:00 PM

[2L2-OS-6a-03] Learning actions using Generative Adversarial Imitation Learning added task achievement reward

○Akira Kinose[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）

2:00 PM - 2:20 PM

[2L2-OS-6a-04] Analysis of Action Decision Algorithm Using Episodic Memory and Value Association

○Shintaro Kayanuma[1], Sana Kawazoe[1], Yuna Tsutsumi[1], Masahiro Miyata[2], Takashi Omori[1] （1. College of Engineering, Tamagawa University, 2. Graduate School of Engineering, Tamagawa University）

2:20 PM - 2:40 PM

[2L2-OS-6a-05] Reward learning through interaction between self-organization of a spiking neural network and a bodily behavior

○Yuji Kawai[1], Tomohiro Takimoto[1], Jihoon Park[1], Minoru Asada[1] （1. Graduate School of Engineering, Osaka University）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

創発インタラクションの可能性：脳科学に基づく創発型人工知能を目指して
Emerging Interactions: Towards the construction of artificial intelligence with a basis of computational neuroscience

津田一郎[*1]
Ichiro Tsuda

[*1] 中部大学創発学術院
Chubu University Academy of Emerging Sciences

We propose emerging interactions in order to unify the developing two directions of research: neural networks with a basis of brain sciences and program-driven artificial intelligence. We aim to construct an agent which is able to adapt to complex environments by establishing functional differentiations. In this lecture, we show our approach to this issue.

1. 概要

学習するニューラルネットとプログラム駆動型の人工知能の融合は1980年代から始まり今日の爆発的な進展へと至っている。その本質的な融合において創発インタラクションの果たす役割は大きいと考えられる。我々は特に脳の進化的及び機能的結合を介した機能分化に着目し、複雑な環境に対して機能分化を通じて即時適応する機構の解明とそれを実現するエージェントの提案を目指している。講演ではその一端をお示ししたい。

参考文献

[Tsuda 2016] Ichiro Tsuda, Yutaka Yamaguti and Hiroshi Watanabe: Self-Organization with Constraints―A Mathematical Model for Functional Differentiation, Entropy vol.18, 74 (2016); doi:10.3390/e18030074

[Tsuda 2015] Ichiro Tsuda: Chaotic itinerancy and its roles in cognitive neurodynamics, Current Opinion in Neurobiology, Vol.31, pp.67-71 (2015), DOI:10.1016/j.conb.2014.08.011

[Yamaguti 2015] Yutaka Yamaguti, Ichiro Tsuda: Mathematical Modeling for Evolution of Heterogeneous Modules in the Brain, Neural Networks, vol.62, 3-10 (2015) DOI:10.1016/j.neunet.2014.07.013

連絡先：津田一郎、中部大学創発学術院、愛知県春日井市松本町 1200 番地、0568-51-9520, tsuda@isc.chubu.ac.jp

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

場所概念形成に基づいた空間のSemantic Mapping

Spatial Concept Formation-based Semantic Mapping

勝又勇貴 *1 Yuki Katsumata　谷口彰 *1 Akira Taniguchi　萩原良信 *1 Yoshinobu Hagiwara　谷口忠大 *1 Tadahiro Taniguchi

*1立命館大学
Ritsumeikan university

This paper propose a new statistical semantic mapping method called *SpCoMapping*, which integrates probabilistic spatial concept acquisition and simultaneous localization and mapping via Markov random field to yield semantic information. Using a simulation environment, we compared in an experiment SpCoMapping with other previously proposed methods. We show that our method can generate better semantic maps than CNN-based semantic mapping and spatial concept formation.

1. はじめに

ロボットは，人間が実環境で使用する場所・物体・行動・人の名前などの語彙を理解して，ユーザが求めるタスクを実行する必要がある．例として，ロボットがユーザから与えられるコマンド「キッチンを清掃して」や「ジョセフの部屋のゴミを拾って」などを実行するためには，ロボットがその環境の地図上において「キッチン」や「ジョセフの部屋」の領域を知る必要がある．空間の Semantic Mapping は，ロボットが人間とコミュニケーションをとり，要求されるタスクを適切に実行するために，環境の地図に意味情報を与えるタスクである [Kostavelis 15]．

日常生活において語彙は自宅やオフィスといった環境に依存し，記号システム自体が動的であるため，未知の語彙が含まれている [T.Taniguchi 16]．従来の Semantic Mapping に関する研究では，事前知識として場所を表す語彙のラベルを用いた研究が多く行われてきた．しかし，これらの手法では各環境に存在する未知の語彙を扱うことが困難である．様々な環境と語彙に対応し，人間との円滑なコミュニケーションを実現するために，家庭環境において動作するロボットにとって，人間の発話に含まれる未知の語彙に対応できる Semantic Mapping は重要である．

Semantic Mapping は，Simultaneous Localization and Mapping (SLAM) を用いて得られた環境の地図を場所を表す語彙のラベルを持つように拡張する研究である．Sunderhauf らは，RGB データを語彙のラベルに変換する Convolutional Neural Network (CNN) を用いた Semantic Mapping を提案した [Sunderhauf 16]．この手法では，Semantic Mapping の語彙のラベルとして CNN による画像認識の結果を使用している．しかし，この画像認識をベースとした手法には 3 つの課題がある．

1 つ目は過去のサイクルで地図に塗られたラベルを別のラベルで上書きしてしまう問題である．例えば，ロボットの位置が同じであっても，廊下から部屋に入るときと部屋を出るときに得られる画像認識の結果が異なることで，先行研究では部屋に入るときに生成されたラベルを部屋から出るときの新しいラベルで上書きしてしまうという問題が生じる．Semantic Map 上のラベルは情報を単純に上書きするのではなく，統計的に保存すべきであり，提案手法は，地図上の各セルのラベルを確率変数としてモデル化することで，この問題を解決する．

2 つ目の問題は，Semantic Map が画像情報のみに基づいている点である．人間は視覚情報のみに基づいて場所のカテゴリを形成するわけではない．環境内で形成される Semantic Map 上のカテゴリの領域や種類は，視覚情報だけでなく，位置情報，ユーザーの発話などの他の情報によっても影響されることが望ましい．提案手法では，ユーザの発話などのマルチモーダル情報を扱う確率的生成モデルの一部としてマルチモーダルカテゴリ分類を用いることでこの問題を解決する．

3 つ目の問題は，既存の Semantic Mapping では，各環境の未知の語彙を扱うことができない点である．実環境において考えられる全ての場所を表す語彙とその特徴量を訓練データとしてロボットに与えることは困難である．これに対し提案手法では，未知語を扱う階層的ベイズモデルに基づく教師なし学習でこの問題を解決する．

ロボットが場所の語彙を獲得するのと同時に，カテゴリと場所の領域を推定する手法として場所概念獲得に関する研究がある [A.Taniguchi 16,Ishibushi 15]．Taniguchi らは，ロボットが未知語を段階的に獲得することを可能にする non-parametric Bayesian spatial concept acquisition method (SpCoA) を提案した [A.Taniguchi 16]．これらの研究では，マルチモーダル情報を扱うことができ，また未知の語彙を獲得することが可能である．しかし既存の場所概念獲得に関する研究では，位置分布がガウス分布によってモデル化されているため，環境の地図への適切な Semantic Mapping を行うことができない．そのため，形成された Semantic Map 上の領域において，環境の様々な形状，例えば壁や障害物による領域の分断などが考慮されていないという問題がある．

本稿では，マルコフ確率場 (Markov Random Field: MRF) と場所概念獲得の手法 [Ishibushi 15, A.Taniguchi 16] を統合した Semantic Mapping 手法である SpCoMapping を提案する．SpCoMapping には，上記の問題を解決する以下の特徴がある．

1. Semantic Map の各セルが確率変数を持ち，ラベルを上書きせず確率的に求めることができる．
2. MRF によって，Semantic Map 上の各領域は環境の形状を考慮できる．
3. 視覚情報だけでなく，人間とロボットとの対話によって得られる語彙などの情報，つまりマルチモーダル情報に基づいて Semantic Map が形成される．

連絡先: 勝又勇貴，立命館大学情報理工学研究科，滋賀県草津市野路東 1-1-1，yuki.katsumata@em.ci.ritsumei.ac.jp

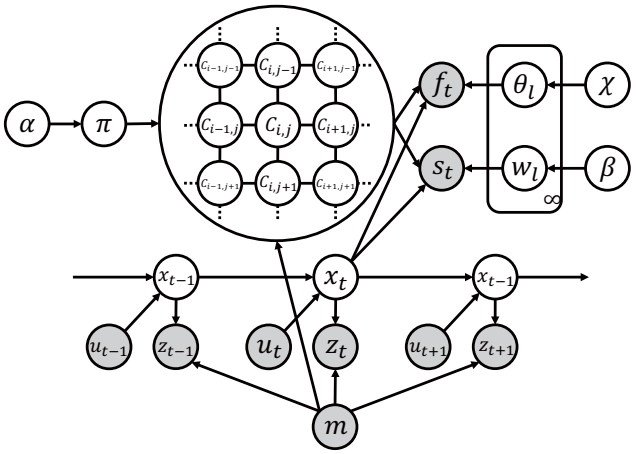


図 1: SpCoMapping のグラフィカルモデル

表 1: SpCoMapping のグラフィカルモデルの変数

m	環境の地図
x_t	ロボットの自己位置
u_t	制御情報
z_t	深度情報
$C_{i,j}$	セル (i,j) の場所概念のインデックス
f_t	画像特徴量
s_t	語彙の特徴量 (bag-of-words)
π	$C_{i,j}$ の多項分布のパラメータ
θ_l	f_t の多項分布のパラメータ
w_l	s_t の多項分布のパラメータ
α, β, χ	事前分布のハイパーパラメータ

4. 音声認識器の辞書に含まれる語彙であれば、新たな場所の名前として学習することができる．
5. Dirichlet process (DP) [Teh 05] をカテゴリ数の事前分布として使用し，場所概念のカテゴリ数を推定できる．

本稿では，SpCoMapping[*1] によって，上記の問題を解決できることをシミュレーション環境で示す．

2. 提案手法

2.1 概要

SpCoMapping のグラフィカルモデルを図 1，各変数の説明を表 1 に示す．まず，ロボットは自己位置推定を用いて環境を移動し，RGB データを取得する．ロボットが移動する間，人間は場所を表す語彙をロボットに与える．SpCoMapping では，Sunderhauf ら [Sunderhauf 16] と同様に事前に訓練された CNN を用い，画像特徴量を得る．また，人間が発話した音声を音声認識システムで認識し，その認識結果の Bag-of-Words (BoW) を語彙の特徴量として用いる．次に，ロボットはマルチモーダル情報を統合することによって場所概念を学習し，確率的生成モデルを用いて Semantic Map を形成する．

2.2 MRF

本手法では従来の場所概念学習においてガウス分布で表現されていた位置分布を，MRF によって表現する．環境の地図の各セルを MRF のノードとして用いることで，ロボットは環境の地図の形状（壁や障害物など）を考慮できるようになる．マルコフ確率場は以下の式で表される．

$$p(C_{i,j}|C_{\partial i,j};\gamma) \quad \propto \quad \exp\left\{\sum_{r\in\partial i,j}\gamma\,\delta(C_{i,j},C_r)\right\} \qquad (1)$$

ここで，$C_{\partial i,j}$ は $C_{i,j}$ の隣接するノード，γ は相互作用係数，$\delta()$ はデルタ関数を示す．

3. 実験

SpCoMapping の Semantic Mapping の性能を評価し，既存の方法と比較する実験を行った．定量評価を行うため，SIGVerse[*2] を用い，複数の家庭環境を模したシミュレーション環境において実験を行った．本実験において SpCoMapping に使用したデータセットは Github に公開した [*3]．

3.1 実験条件

画像特徴量を得るための CNN の実装には，CNN のフレームワークである Caffe [Jia 14] を使用した．また，CNN の訓練データとして，Places205 dataset [Zhou 14] を使用した．CNN は AlexNet [Krizhevsky 12] を使用した．語彙情報をロボットに与える際，Semantic Mapping を評価することに焦点を当てるため，音声認識を使用せずに，ロボットにテキストデータを直接与えた．実験で比較したのは以下の 5 つの方法である．

(A) カテゴリ数が未知の場合の SpCoMapping (DP)

(B) カテゴリ数が既知の場合の SpCoMapping (Dir)

(C) Spatial concept formation に語彙の学習を追加した手法 [Ishibushi 15]

(D) Sunderhauf らの CNN-based semantic mapping [Sunderhauf 16]

(E) Nearest neighbor

手法 (A) では，DP に week-limit 近似 [Fox 11] を用い，場所概念の上限数を 120 と設定した．手法 (B) では，場所概念の数を正解ラベルのカテゴリ数と同じ値に設定し，ディリクレ分布から場所概念の多項分布のパラメータのサンプリングを行った．手法 (E) は最近傍法である．この手法は位置的に最も近い語彙のラベルと同じラベルを獲得する手法である．Ishibushi らの手法 [Ishibushi 15] は，画像特徴量と自己位置情報のみを使用しているが，実験では語彙情報を扱うモデルを手法 (C) として使用した．手法 (D) は，占有格子地図上のフリースペースのセル全てに対してはカテゴリ分類を行わない．そこで，全てのピクセルにラベルを与えるために手法 (D) を用いて Semantic Map を形成した．図 2 は，各手法によって生成された Semantic Map を示している．

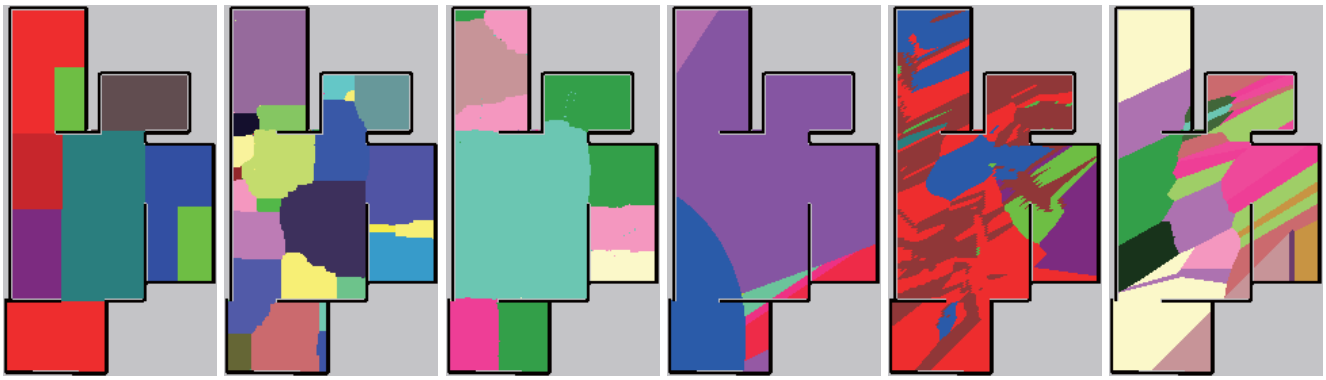
正解ラベルは，任意に選ばれた参加者が 3D シミュレータからの情報と 2D マップを参照して，各環境の Semantic Map を描画することで作成した．実験では，各環境で使用されると考えられる場所を表す語彙をロボットに与えた．用いた語彙のリストを，表 2 に示す．語彙のリスト内で下線を引いた語彙は，CNN のラベルには存在しないものを表している．また，これらの語彙には正解ラベルで用いられていない語彙も含まれる．

SpCoMapping は場所概念のカテゴリ数が未知の場合と既知の場合，いずれも 5000 イテレーション実行する．カテゴリ数が

∗1 Source code on Github: https://github.com/EmergentSystemLabStudent/spco_mapping.git

∗2 SIGVerse: http://www.sigverse.org/wiki/en/

∗3 Dataset on Github:https://github.com/EmergentSystemLabStudent/Dataset_of_SpCoMapping.git



(a) 正解ラベル (b) Proposed method (DP) (c) Proposed method (Dir) (d) The spatial concept [Ishibushi 15] (e) CNN-based semantic mapping [Sunderhauf 16] (f) Nearest neighbor

図 2: Room2ldk4 において形成された Semantic Map の例: (a) は人間が生成した正解ラベル，(b) は SpCoMapping の場所概念のカテゴリ数が未知の場合，(c) は SpCoMapping の場所概念のカテゴリ数が既知の場合，(d) は既存の場所概念獲得の Semantic Mapping 結果，(e) は CNN ベースの Semantic Mapping の結果，(f) は Nearest Neighbor の結果である．それぞれの Semantic Map の色の違いはそれぞれの Map におけるカテゴリの違いを示している．

表 2: シミュレーション環境で使用した語彙のリスト

bedroom	shower	kitchen	entrance	living
dining	closet	corridor	parlor	
window	tree	shelf	chair	sofa
washroom	toilet	refrigerator	TV	wall
bed	table	bath	door	

未知の場合のハイパーパラメータは $\alpha = 1.0 \times 10^8$，$\beta = 500$，$\chi = 1.0 \times 10^4$，$\gamma = 4.0$ とする．カテゴリ数が既知の場合のハイパーパラメータは $\alpha = 5.0 \times 10^7$，$\beta = 100$，$\chi = 1.0 \times 10^4$，$\gamma = 3.0$ とする．使用したハードウェアは，Ubuntu 14.04 LTS 64-bit，ROS Indigo with 31.4 GB memory，Intel Core i7-4770K CPU @ 3.50 GHz × 8，Gallium 0.4 on NVE4 である．

3.2 結果

3.2.1 クラスタリング

各手法に対して Adjusted Rand Index (ARI) [Hubert 85] を計算した．ARI は，2 つのクラスタ間の類似性の尺度である．2 つのクラスタが同じ場合は ARI は 1 となり，ランダムである場合は ARI は 0 となる．Semantic Mapping は，地図上のピクセルをクラスタリングするタスクと見なすことができるため，ARI を用いて提案手法と既存手法の性能を比較した．

結果を表 3 に示す．SpCoMapping は，他の手法と比較して各環境で高い値を示した．このことから，SpCoMapping は 1 章で述べた上書きの問題の解決と環境に合わせた任意の形状を持つ領域を生成することができることを示している．つまり，SpCoMapping によって生成された Semantic Map のカテゴリは，人によって生成された Semantic Map のカテゴリに近いと言える．

3.2.2 単語から場所の予測

ナビゲーション，部屋の掃除，探しものなどのユーザとのコミュニケーションが必要なタスクをロボットが行う場合，ロボットは場所を表す語彙からユーザの指示する場所を推定する必要がある．そこで，各手法で推定した場所の正解率を比較した．正解率の計算は次のように計算する．

$$R_{match} = \sum_{s \in V}^{|V|} \frac{\sum_{i=1}^{M} \delta(s_i, L_i)}{M} \tag{2}$$

ここで V は正解ラベルに含まれる語彙の数，M は占有格子地図中のフリースペースの数，s_i は Semantic Map のピクセル i に与えられた語彙，L_i は正解ラベルのピクセル i に与えられた語彙を表す．

SpCoMapping において，語彙 s を与えられたときの場所概念のインデックスの推定は以下の式で行う．

$$C_s = \underset{C}{\operatorname{argmax}}\, p(C \mid s_t = s, \pi, w) \tag{3}$$

ここで s_i は Semantic Map 中のピクセル i における $C_i = C_s$ のときの語彙を表す．

結果は表 4 に示す．手法 (B) は，他の方法よりも高い値を示した．手法 (A) の値は，Room1dk5 と Room1dk4 において 5 つの手法の中で最も良い結果が得られた．しかし，ピクセル数の大きな環境（Room1dk6 と Room2ldk4）の結果は他の手法に劣っている．この 2 つの環境で SpCoMapping の性能が低下した原因はカテゴリ数の推定が間違っていたためである．カテゴリ分類のノンパラメトリックベイズ推定は，不安定であるため，SpCoMapping (Dir) は SpCoMapping (DP) よりも安定した結果が得られることがわかる．

4. 結論

本稿では，MRF を用いて場所概念獲得手法を拡張し，SpCoMapping と呼ばれる Semantic Mapping について述べた．シミュレーション環境における実験により，SpCoMapping は既存の Semantic Mapping の問題に対処できることが示された．さらに実験において，SpCoMapping によって生成された Semantic Map は，人間によって生成された Semantic Map と，既存の手法よりも一致している結果が得られた．また，SpCoMapping により生成された Semantic Map は，単語入力からの場所推定，すなわち人間とロボットとのコミュニケーションの観点から既存の方法よりも優れていることを示した．

表 3: ARI の結果

Method	Room1dk5	Room1dk6	Room1ldk4	Room1ldk5	Room2ldk4
(A) SpCoMapping (DP)	**0.5356**	0.4548	**0.5877**	**0.4925**	0.4623
(B) SpCoMapping (Dir)	0.4471	**0.5355**	0.4393	0.3354	**0.4963**
(C) Spatial concept formation [Ishibushi 15]	0.3524	0.4692	0.2657	0.3740	0.2883
(D) CNN-based semantic mapping [Sunderhauf 16]	0.3102	0.2742	0.2505	0.3067	0.2559
(E) Nearest neighbor	0.3371	0.2870	0.4830	0.3337	0.3968

表 4: 領域の正解率

Method	Room1dk5	Room1dk6	Room1ldk4	Room1ldk5	Room2ldk4
(A) SpCoMapping (DP)	**0.2612**	0.0268	**0.2185**	0.1338	0.0001
(B) SpCoMapping (Dir)	0.2071	**0.6368**	0.1461	0.1119	0.1780
(C) Spatial concept formation [Ishibushi 15]	0.0463	0.2587	0.2154	**0.3200**	**0.1876**
(D) CNN-based semantic mapping [Sunderhauf 16]	0.1174	0.0761	0.0949	0.0766	0.1245
(E) Nearest neighbor	0.1260	0.1607	0.1729	0.1252	0.1292

今後の研究では，SpCoMapping を人間とのコミュニケーションを必要とするタスク，例えば「私の部屋をきれいにしてください」と言ったようなタスクなどに適用することを考えている．また，シミュレーション実験で明らかになった SpCoMapping (DP) の安定性を向上させることも課題の一部である．さらに，SpCoMapping はバッチ学習を採用しているため，SpCoMapping をオンライン学習が行えるように改善し，環境の地図が未知の環境において動作するために SLAM と統合することも今後の課題として挙げられる．

いくつかの課題は残っているが，提案手法は教師なし学習に基づく空間の Semantic Mapping の性能を大幅に改善し，実環境における人間の発話を Semantic Mapping に利用できることを示した．本研究は，近い将来，家庭環境において能動的に学習するロボットと人のコミュニケーションを円滑化することに貢献すると考えられる．

参考文献

[Kostavelis 15] Ioannis Kostavelis, Antonios Gasteratos, "Semantic mapping for mobile robotics tasks: A survey", Robotics and Autonomous Systems, 2015, pp.86-103

[T.Taniguchi 16] Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, "Symbol Emergence in Robotics: A Survey, Advanced Robotics", Advanced Robotics vol.30 no.11-12, 2016, pp.706-728

[Sunderhauf 16] Niko Sunderhauf, Feras Dayoub, Sean McMahon, Ben Talbot, Ruth Schulz, Peter Corke, Gordon Wyeth, Ben Upcroft, and Michael Milford, "Place Categorization and Semantic Mapping on a Mobile Robot", International Conference on Robotics and Automation (ICRA), 2016, pp.5729-5736

[A.Taniguchi 16] Akira Taniguchi, Tadahiro Taniguchi and Tetsunari Inamura, "Spatial Concept Acquisition for a Mobile Robot That Integrates Self-Localization and Unsupervised Word Discovery From Spoken Sentences", Transactions on Cognitive and Developmental System, 2016, pp.285-297

[Ishibushi 15] Satoshi Ishibushi, Akira Taniguchi, Toshiaki Takano, Yoshinobu Hagiwara and Tadahiro Taniguchi, "Statistical Localization Exploiting Convolutional Neural Network for an Autonomous Vehicle", IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, 2015, pp.1369-1375

[Teh 05] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal and David M. Blei, "Sharing clusters among related groups: Hierarchical dirichlet processes", Advances in neural information processing systems (NIPS), 2005, pp.1385-1392

[Jia 14] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding", Proceedings of the 22Nd ACM International Conference on Multimedia, 2014, pp.675-678

[Zhou 14] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, "Learning Deep Features for Scene Recognition using Places Database", Advances in Neural Information Processing Systems 27, 2014, pp.487-495

[Krizhevsky 12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems 25, 2012, pp.1097-1105

[Fox 11] Emily B Fox, Eric B Sudderth, Michael I Jordan and Alan S Willsky, "A sticky HDP-HMM with application to speaker diarization", Institute of Mathematical Statistics, 2011, pp.1020-1056

[Hubert 85] Lawrence Hubert, "Comparing Partitions", Journal of Classification Volume 2, Issue 1, 1985, pp.193-218

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Generative Adversarial Imitation Learning に タスク達成報酬を付加した動作の学習

Learning actions using
Generative Adversarial Imitation Learning added task achievement reward

黄瀬 輝 *1　　谷口 忠大 *1
Akira Kinose　　Tadahiro Taniguchi

*1 立命館大学
Ritsumeikan University

In manipulation tasks, it is extremely costly for a human to manually create a program such as the complex motion like a human arm. On the other hand, learning manipulation tasks using machine learning has a huge range of actions, so it is difficult for robots to learn without demonstrations by a human. For these reasons, it is required for robots to learn motions smoothly using human demonstrations. In this research, based on Generative Adversarial Imitation Learning (GAIL) which is a representative method of Learning from Demonstrations (LfD) in recent years, we propose the method that can learn manipulation actions in the simulation environment by newly using task achievement rewards. As a result, the proposed method succeeded in learning the policy of manipulation task which was difficult to learn in the existing models.

1. はじめに

近年，ロボットは研究の発展とともに利用価値が高まり，日常生活環境や災害現場などで人の代わりに作業を行うことが期待されている．ロボットアームの制御によって物体の把持，移動を行うマニピュレーションは，ロボットが作業を行う上で最も基本的な技術の一つである．しかし，人間の腕の動作のような，複雑な制御を伴うプログラムを手作業で作成することは，非常に手間がかかる．一方，機械学習を用いたマニピュレーション制御は行動の取りうる範囲が膨大であるため，ロボットが人間による教示無しで学習を行うことは困難である．このような理由から，人間の教示によってロボットの動作を獲得することが必要である．本研究では，これらの問題に対して，タスクのデモンストレーションをもとに動作の学習を行う Learning from Demonstration (LfD) に焦点を当てる．

近年の LfD に代表される手法として，Ho らによる Generative Adversarial Imitation Learning (GAIL)[Ho 16] が挙げられる．GAIL は環境のダイナミクスと人間による特徴量の設計が不要かつ，高次元環境でのモデルフリーの優れた LfD の手法である．従来，逆強化学習によって推定した報酬関数に基づいて強化学習を行うことで，タスクのエキスパートの動作を学習していたのに対し，GAIL は Generative Adversarial Networks (GAN) [Goodfellow 14] を適用させることでこれらを統一的に扱っている．GAIL の特徴として，環境からの報酬を受け取らず，エキスパートの軌道から推論した報酬関数のみを用いて学習を行っていることが挙げられる．しかし，本研究で想定するような，タスクのゴールが存在するマニピュレーションタスクでは，タスク達成の要因となるオブジェクトやゴールの初期位置が変化するため，環境からの報酬を用いずに学習することは困難であると考えられる．

本研究では，タスクのゴールが存在するマニピュレーションタスクにおいて，LfD の代表的手法である GAIL のモデルにタスク達成報酬を導入することで，パフォーマンスの向上を図る．また，その実験として，MuJoCo シミュレーション環境 [Todorov 12] のマニピュレーションタスクにおいて，GAIL と提案手法の性能評価を行う．

連絡先: 黄瀬輝, 立命館大学情報理工学研究科, 滋賀県草津市野路東 1 - 1 - 1 ,akira.kinose@em.ci.ritsumei.ac.jp



(a) テレビゲーム Atari2600[Bellemare 13]



(b) ロボットによるマニピュレーション

図 1: タスクの例

2. GAIL へのタスク達成報酬の導入

2.1 強化学習と逆強化学習

一般的な強化学習では，エージェントが環境から現在の状態における行動の報酬を得ることによって最適方策を学習する．近年の強化学習の代表例として Deep Q-network [Mnih 13, Mnih 15] があり，図 1(a) のようなテレビゲームでは，ゲームのスコアを報酬に設定することで学習を行っている．また，碁やチェス，将棋などのボードゲームでは，ゲームの勝利を報酬に設定することで学習を行う．

しかし，現実のタスクにおいて，この報酬を定義することが困難な場合がある．例として，図 1(b) のようなロボットによる「ボトルの飲料をコップに移す」といったマニピュレーションタスクを考える．現実の環境では行動空間が膨大になるため，タスク達成報酬のみでは学習が難しく，タスク達成に繋がる要素を考慮した報酬が必要になる．深層強化学習によるマニピュレーションの学習に成功している Gu らの研究においても，学習にはタスク達成に繋がる要素を考慮した報酬を必要とし，タスク達成のみを報酬にした場合，学習は困難になることが述べられている [Gu 17]．タスク達成に繋がる要素とは，マニピュレーションタスクにおいて「ロボットの腕をどの位置

へ移動させ，何度傾けるのが好ましいか」などにあたる．しかし，「ロボットの腕をどの位置へ移動させ，何度傾けるのが好ましいか」を考慮した報酬を人間が明確に定義することは困難である．

このような問題に対して，逆強化学習は任意のタスクにおける最適な方策を持つエージェントであるエキスパートの行動から，人間によって定義が困難な報酬関数を推定する．そして，逆強化学習によって推定した報酬関数に従って強化学習を行うことで，報酬関数の定義が困難なタスクの最適方策を求めることができる．この LfD の手法は一般に人間が設計する特徴量が必要であり，高次元連続環境では計算量が膨大になるといった問題がある．

2.2 GAIL

Ho らが提案した GAIL は，課題点であったダイナミクスと人間による特徴量の設計が不要かつ，高次元環境でのモデルフリーの LfD の手法である．Ho らは同論文で逆強化学習に強化学習を用いた 2 段階最適化問題を 1 段階最適化問題に定式化できることを示している．その後，定式化された関数において，Goodfellow らによって提案された GAN の目的関数に一致するように正則化関数を定義することで，報酬関数を推定することなく，方策を直接推定ことができる GAIL を提案している．GAIL は式 (1) を解くことによって学習を行う．

$$\min_{\pi}\max_{D} \quad \mathbb{E}_{\pi_\theta} \quad [\log(D(s,a))] \\ + \quad \mathbb{E}_{\pi_E} \quad [\log(1-D(s,a))] - \lambda H(\pi) \qquad (1)$$

$D(s,a)$ はその状態行動のペア (s,a) がエキスパートから生成されたものである確率を示す．

GAIL モデルにおいて，Discriminator の役割は Generator によって生成された方策 π_θ とエキスパート方策 π_E を識別することである．Discriminator は勾配ベクトル:

$$\mathbb{E}_{\tau_i}[\nabla_w \log(D_w(s,a))] + \mathbb{E}_{\tau_E}[\nabla_w \log(1-D_w(s,a))]$$

に対して勾配降下法を行い，エキスパートから生成されたものか正しく識別できるように学習する．w は Discriminator のパラメータを示す．D は $D(s,a) \in [0,1]$ の値を取り，擬似的なコスト関数の役割をする．

Generator の役割はエキスパート方策に似るように方策を生成することである．Generator は勾配ベクトル:

$$\mathbb{E}_{\tau_i}[\nabla_\theta \log \pi_\theta(a|s) Q(s,a)] - \lambda \nabla_\theta H(\pi_\theta)$$
$$\text{where } Q(s,a) = \mathbb{E}_{\tau_i}[\log(D_{w_{i+1}}(s,a)|s_0 = s, a_0 = a)]$$

に対して方策勾配法を行い，Discriminator によって定義されるコストを最小化するような方策を学習する．θ は方策のパラメータを示す．この方策更新は強化学習手法 TRPO [Schulman 15] で勾配計算を行う．

2.3 タスク達成報酬を導入した GAIL

本研究では，タスク達成条件を満たした状態行動対に対して，従来手法では利用していなかったタスク達成報酬を導入することでパフォーマンスの向上を図る．

GAIL の学習則の特徴として，エキスパート軌道から推論した報酬関数のみを用いて強化学習を行うため，環境からの報酬を受け取らず，エキスパート軌道のみから学習を行っていることが挙げられる．この性質は，エージェントが自分の行った動作がタスクの達成にどのような影響を及ぼしたかがわからないという問題を引き起こす．よって，マニピュレーションタスクのように，タスクのゴールやオブジェクトの位置座標が変化するような環境下において，GAIL による学習は困難である．そこで，エキスパートだけでなく環境からも報酬を受け取ることができれば，GAIL の学習性能を上げることが期待できる．しかし，GAIL がエキスパートからの推論によって得られる報酬のみを利用している理由として，学習のための報酬関数の設計は非常に手間が掛かり，技術や知識を要することが挙げられる．以上のことから，本手法では設計に手間がかからないような報酬関数を利用することで学習を行う GAIL のモデルを考える．

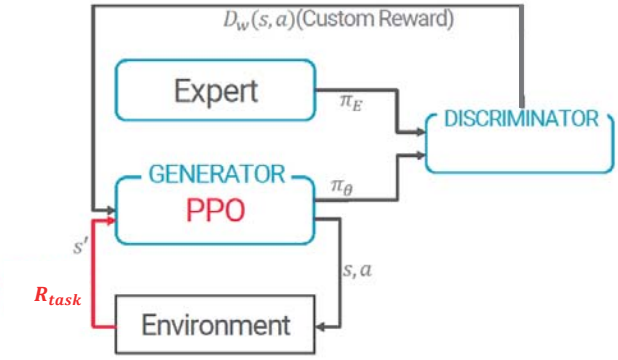


図 2: TRGAIL のモデル図

学習を行うタスクとして，本論文の研究対象としているマニピュレーションタスクを想定する．物体を指定した場所へ移動させることを目指すように学習を行うマニピュレーションタスクでの学習では，タスクのゴールが明確に存在するため，タスク達成条件を明示的に定義することが可能である．

そこで，本手法では，状態と行動を入力として，タスク達成条件を達成したかというバイナリ値を出力とする環境の報酬関数として定義する．その後，エキスパート軌道から推論される報酬とタスク達成バイナリ報酬の両方の報酬を用いることでタスク達成が可能な方策の学習を行う．ここで，バイナリ報酬とは，一般的に連続環境の複雑なタスクで用いられるような学習を円滑化させる工夫を用いた複雑な報酬とは異なり，ある状態行動対がタスクを達成しているかが 0 または 1 の値で与えられるものであるとする．

GAIL にタスク達成報酬を利用したモデルを本稿では Task Reward oriented Generative Adversarial Imitation Learning (TRGAIL) と呼称する．図 2 に TRGAIL のモデル図を示す．TRGAIL において定義される報酬関数の式を以下に示す．

$$Q(s,a) = \mathbb{E}_{\tau_i}[\alpha \log(D_{w_{i+1}}(s,a) + (1-\alpha) R_{task}(s,a)] \qquad (2)$$

ここで，α は報酬関数においてタスク達成報酬が報酬関数を占める割合を決めるパラメータである．本稿では $\alpha = 0.5$ を採用している．$R_{task}(s,a)$ はバイナリ値で与えられるタスク達成報酬を示す．GAIL において $D(s,a)$ は報酬ではないため，ここにタスク達成報酬を加えることはヒューリスティクスなアプローチであると言える．また，TRGAIL では Generator に強化学習手法である PPO[Schulman 17] を用いる．TRGAIL のアルゴリズムを Algotithm 1 に示す．

TRGAIL によるエキスパート軌道の学習では，タスク達成に近づくが完全なタスク達成までに至らない場合，タスク達成

Algorithm 1 TRGAIL

1: **Input:** エキスパート軌道 $\tau_E \sim \pi_E$, 方策と Discriminator の初期パラメータ θ_0, w_0
2: **for** $i = 0, 1, 2, \ldots$ **do**
3: 学習軌道をサンプリング $\tau_i \sim \pi_{\theta_i}$
4: Discriminator の勾配ベクトル:

$$\mathbb{E}_{\tau_i}[\nabla_w \log(D_w(s,a))] + \mathbb{E}_{\tau_E}[\nabla_w \log(1 - D_w(s,a))]$$

に対して勾配降下法を行うことでパラメータを w_i から w_{i+1} へ更新
5: Generator の勾配ベクトル:

$$\mathbb{E}_{\tau_i}[\nabla_\theta \log(\min(r_t(\theta), \mathrm{clip}(r_t(\theta), 1+\epsilon, 1-\epsilon))\, Q(s,a))] - \lambda \nabla_\theta H(\pi_\theta)$$

$$Q(s,a) = \mathbb{E}_{\tau_i}[\alpha \log(D_{w_{i+1}}(s,a) + (1-\alpha) R_{task}(s,a)]$$

に対して PPO の方策勾配法を行うことでパラメータを θ_i から θ_{i+1} へ更新
6: **end for**

報酬を環境から得ることによって，エージェントがエキスパート軌道の学習とタスクの達成軌道の学習を両立できると考えられる．

このアプローチを強化学習の観点から考えると，タスクのエキスパートによる手本を得ることで，膨大な探索範囲を絞り込めるという見方をすることができる．また，タスクのエキスパートとタスク達成報酬を準備する代わりに，従来の強化学習の課題点であったタスク達成に繋がる要素を考慮した複雑な報酬関数の設計が不要になるという利点が考えられる．

3. 実験

3.1 実験目的

本実験では，エキスパートの軌道からマニピュレーション動作の獲得をする実験を行う．本実験の目的は，既存の GAIL モデルとタスク報酬を導入した GAIL のモデルのタスクのパフォーマンスを比較し，その妥当性を検証することである．比較手法として，既存の GAIL，提案手法，LfD 手法のベースラインとして Behavior Cloning を用いる．

3.2 実験条件

実験環境は物理エンジンソフトウェアである MuJoCo [Todorov 12] を用いたシミュレーション環境で実験を行う．各タスクは OpenAI Gym [Brockman 16] によって定義されたモデルの状態空間，行動空間を利用して行う．本実験では，Pusher，Striker，Thrower の 3 タスクについてマニピュレーション動作獲得の実験を行う．それぞれのタスクの図を図 3，図 4，図 5 に示す．

各タスクにおいて，1 回の試行を合計 100 タイムステップとする．イテレーション数は 3000 であり，各イテレーションのタイムステップ数が 2400 になるようにバッチをサンプリングした．ベースラインとして，同タスクで Behavior Cloning で 10000 イテレーションの学習を行ったものを利用した．実験は 3 タスク，3 手法に対してシード値を変えてそれぞれ 3 試行ずつ行った．エキスパート軌道は実験条件の簡易化のため，連続環境における state-of-the-arts の手法の一つである強化学習手法の PPO によって学習を行った最適方策による軌道を用いた．PPO による強化学習は人による複雑なタスク達成に繋がる要素を考慮した報酬関数の設計を行い，最適軌道から 500 軌道をサンプリングした．

それぞれのタスクにおいて，行動空間は 7 次元の連続値でアームの各 7 関節に与えるトルクの値を示す．状態空間は 23 次元の連続値で 7 関節の角度角速度，オブジェクトゴール手先の xyz 座標の値を示す．以下に各タスクの設定について述べる．

Pusher

Pusher タスクは，エージェントの手先が U の字の型をしたアームを用いて白い円柱のオブジェクトをゴールである赤い円内へ押し込むように移動させるタスクである．環境の初期状態はエージェントの姿勢とゴールの座標が固定値で，オブジェクトの座標がランダム値で決まる．タスク達成条件はオブジェクトの中心の xy 座標がゴール範囲内に入ることとする．

Striker

Striker タスクは，エージェントの手先が T の字の型をしたアームを用いて，白いボールをテーブル上のゴールである離れた白い柵付きの円内へ打ち込むタスクである．環境の初期状態はエージェントの姿勢とオブジェクトの座標が固定値で，ゴールの座標がランダム値で決まる．タスク達成条件はオブジェクトの中心の xy 座標がゴール範囲内に入ることとする．

Thrower

Thrower タスクは，エージェントのアームを用いて白いボールを離れたテーブル上の四角の箱のゴールへ投げ入れるタスクである．環境の初期状態はエージェントの姿勢とオブジェクトの座標が固定値で，ゴールの座標がランダム値で決まる．タスク達成条件はオブジェクトの中心の xy 座標がゴール範囲内に入り，ボールが箱の底面に接触していることとする．

3.3 評価指標

軌道スコア

各試行の最終タイムステップにおける状態に対する評価を行い，スコアを算出する．この評価指標は時間的要因を考慮しないため，学習した軌道の試行における最終的な状態に対して評価を行う．よって，この評価値からは学習した方策が各試行においてタスク達成に近づいたかがわかる．Pusher タスクは最終タイムステップのオブジェクトの中心座標とゴール範囲の距離，Striker タスクは最終タイムステップのオブジェクトの中心座標とゴール範囲の距離，Thrower は最終タイムステップまでに最も近づいたオブジェクトの中心座標とゴール範囲の距離を軌道スコアと定義する．

タスク成功率

この評価指標は，各試行の最終タイムステップにおける状態を参照し，その時点でタスクの達成条件を満たしているかの判定を行い，そのイテレーションにおけるタスク成功率を示す．よって，この評価値が高い場合，タスクの達成条件を正確に満たす方策を学習できたということがわかる．

3.4 実験結果

実験結果を表 1 に示す．TRGAIL の実験結果について，この手法による評価指標は全手法において総合的に最も良いパフォーマンスを示した．実験結果の軌道と成功率を確認すると，従来までの手法と比較して，エキスパート軌道の学習に加えてタスクのゴールを目指すような方策を学習していることが確認できた．また，Pusher，Thrower の軌道スコア，成功率において，エキスパートのスコアと同等のパフォーマンスを示していることがわかる．

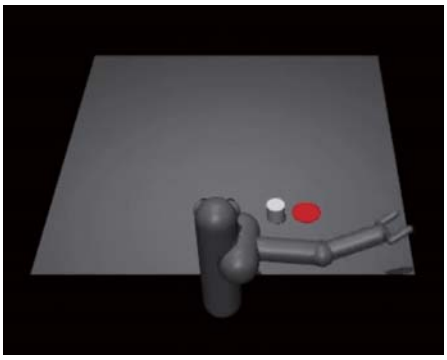


図 3: Pusher タスク

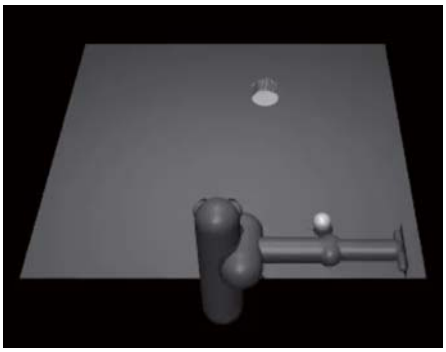


図 4: Striker タスク

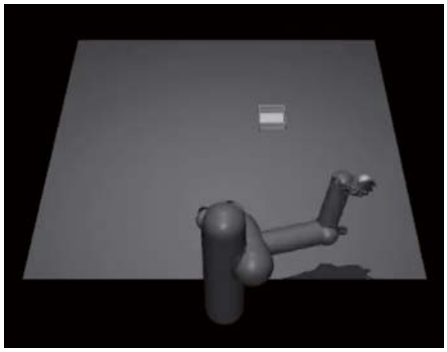


図 5: Thrower タスク

表 1: 実験結果

	Pusher		Striker		Thrower	
手法	軌道スコア	成功率	軌道スコア	成功率	軌道スコア	成功率
Expert Policy	-0.0006	93.2	-0.0095	87.1	-0.0077	79.3
Random Policy	-0.1675	0.5	-0.6365	0.0	-0.8586	0.2
Behavior Cloning	-0.1418	25.0	-0.1346	33.0	-0.0092	81.0
GAIL	-0.1975	9.4	-0.6466	6.9	-0.0535	28.0
TRGAIL	**-0.0004**	**92.3**	**-0.0270**	**79.7**	**-0.0082**	**91.6**

4. おわりに

本稿では，GAIL にタスク達成報酬の導入を行い，タスクのゴールが存在するシミュレーション上のマニピュレーションタスクにおいて，エキスパート軌道からタスク達成動作の獲得を行う TRGAIL を提案した．実験結果より，提案手法がタスクのゴールが存在するマニピュレーションにおいて高いパフォーマンスを示す方策を学習したことが示された．

GAIL はエキスパートとなる軌道と似たような軌道を生成する方策を学習するという問題設定だが，タスク達成までの過程がより複雑になるようなタスクの学習は困難であると考えられる．本研究の今後の課題として，階層型強化学習の理論を適応させることで，タスクを階層化し，階層的な報酬関数の推定を行うことで，より複雑なタスクの学習を行うことが期待される．

参考文献

[Bellemare 13] Bellemare, Marc G., et al. "The Arcade Learning Environment: An evaluation platform for general agents." J. Artif. Intell. Res.(JAIR) 47 (2013): 253-279.

[Brockman 16] Brockman, Greg, et al. "Openai gym." arXiv preprint arXiv:1606.01540 (2016).

[Goodfellow 14] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.

[Gu 17] Gu, Shixiang, et al. "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates." Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, 2017.

[Ho 16] Ho, Jonathan, and Stefano Ermon. "Generative adversarial imitation learning." Advances in Neural Information Processing Systems. 2016.

[Mnih 13] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." in Deep Learning, Neural Information Processing Systems Workshop, 2013.

[Mnih 15] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529.

[Schulman 15] Schulman, John, et al. "Trust region policy optimization." International Conference on Machine Learning. 2015.

[Schulman 17] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).

[Todorov 12] Todorov, Emanuel, Tom Erez, and Yuval Tassa. "Mujoco: A physics engine for model-based control." Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, 2012.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

エピソード記憶と価値の連合した行動決定アルゴリズムの評価

Analysis of Action Decision Algorithm Using Episodic Memory and Value Association

栢沼 晋太郎[*1] 川添 紗奈[*1] 堤 優奈[*1] 宮田 真宏[*2] 大森 隆司[*1]
Shintaro Kayanuma Sana Kawazoe Yuna Tsutsumi Masahiro Miyata Takashi Omori

[*1] 玉川大学 工学部 [*2] 玉川大学大学院 工学研究科
College of Engineering, Tamagawa University. Graduate School of Engineering, Tamagawa University.

Episodic memory is the important function of hippocampus, and we can't lack it for our wholesome life. But its theory and role in decision making is not clear yet. In this study, we analyze and discuss on its model that we proposed at the WBAI hackathon. As the result, an association of the episodes and the value enabled rapid action learning with small number of experiences, and takes complementary role with the strong but slow feature of Deep Learning.

1. はじめに

エピソード記憶は，人間の重要な情報処理の一つであり，知識やその人物の生活史を支える記憶体系の一部である．脳科学ではその機能の責任部位として海馬が知られており，その損傷は健忘症を引き起こす．人間の知能という側面からみると，エピソードは間違いなく重要であろうと思われるが，実際場面におけるその役割は現時点では明確ではない．

海馬の理解がそういう状態であるときに，全脳アーキテクチャ・イニシアティブ主催のハッカソン「目覚めよ海馬」が開催された[WBAI 2017a]．我々はこれに参加し，エピソード記憶に価値を連合させてオンライン学習させるアルゴリズムを提案・実装した結果，準優勝であった．ハッカソンでは，1 次元の仮想迷路として 8 個のタスクが用意され，そこで行動するエージェントのプロトタイプが与えられて，それを改造してタスクを解くことが求められた．結果として，我々の海馬モデルは 1 次元仮想迷路課題 8 タスクのうち 7 タスクを単一エージェントで連続して解決する成績が出せた．しかしその内部で何が起きていたのか，エピソードに価値を連合されたことは有効だったのか，など不明な点が多く残るままにハッカソンは終わった．

そこで本研究は，我々が提案した行動決定アルゴリズムの内部過程を分析し，なぜこれがタスク解決，特に徐々に難しくなるタスク群のすばやい解決に有効であったのかを明らかにすることを試みる．そして，従来から知られている行動決定アルゴリズムとの関係を議論することで，機械学習におけるエピソード記憶や価値の連合の効果について議論する．

2. エピソードと価値を連合させる行動決定モデル

2.1 ハッカソン

2017 年 9 月，NPO 法人全脳アーキテクチャ・イニシアティブ主催で，第 3 回全脳アーキテクチャ・ハッカソン「目覚めよ海馬！ 汎用人工知能プロトタイプに向けた海馬モデルの組み込み」が開催され，全 8 チームがマウスの行動実験における迷路のシミュレーション課題に取り組んだ．

課題は，全長 24m，幅 1m の仮想迷路でエージェントが各タスクで事前に決められた条件をクリアすると報酬が得られ，解決できたトライアル数が一定条件を満たすと次のタスクへ進むものであった（Fig.1）．基本的にはエージェントが緑色のブロックを取るとそのトライアルは成功となるが，同時にエージェントは常に負の報酬を受け続け，累積報酬がある値を下回った時点でそのトライアルが失敗となる．あるタスクで，成功数が失敗数よりも 23 回多くなったなら，それまでの学習結果を維持したまま次のタスクに進むことができる．1 次元の迷路を使うタスク 1〜7 の特徴を表 1 に挙げる．

連絡先：栢沼 晋太郎，玉川大学 工学部，東京都町田市 玉川学園 6-1-1，kynms5is@engs.tamagawa.ac.jp



Fig.1 1 次元の仮想迷路，エージェント(水色の玉)がゴール(緑色のブロック)に到達すると報酬が与えられ，トライアルが終了する．迷路の途中に壁に色のついた部分があり，タスクごとにその場所でのタスクの要請が変化する．)

表 1 1 次元仮想迷路でのタスク一覧

タスク 1	無条件で，ゴール地点で報酬が与えられる．
タスク 2	緑の壁の通過時に(すぐにその場で) 報酬が与えられる．その後，ゴールでも報酬が与えられる．
タスク 3	緑の壁の地点で 2 秒間待機すると，報酬が得られ，さらにゴールでも報酬が得られる．緑の壁の地点で待機せずにゴールに到達すると報酬は得られない．
タスク 4	報酬が得られる壁が赤の壁の地点に変わる．報酬が与えられる条件はタスク 3 と同じ．
タスク 5	報酬が得られる壁が青の壁の地点に変わる．報酬が与えられる条件はタスク 3 と同じ．
タスク 6	「タスク 4→タスク 5→タスク 3→タスク 4・・・」と報酬が与えられる地点が毎回変化する．
タスク 7	スタート時に幾何学模様(△，○，□)をエージェントに提示される．△ならタスク 3 と同じように報酬が与えられる．○ならばタスク 4 と同じ，□ならタスク 5 と同じように報酬が与えられる．

エージェントには RGB カメラと深度カメラが搭載されており，行動ステップごとに画像を取得する．ハッカソン参加者はエージェントのプログラムを作り，この画像から何らかの処理を用いて行動する．エージェントの行動は前進(1m/step)，右回転(+10 度回転/step)，左回転(-10 度回転/step)，停止(±0/step)の 4 種類であり，結果としてエージェントはその 2 次元的位置と回転方向の状態量(x, y, θ)を持つ[WBAI 2017b].

2.2 エピソード記憶と価値による意思決定モデル

我々は，海馬のエピソード記憶と場所細胞の機能に注目し，行動学習モデルを提案した．エピソード記憶は，視覚・聴覚・触覚などの感覚情報の処理結果の集合体である特徴ベクトルとした．さらに，エピソードには多くの場合に感情が付随すると考える．例えば，道を歩いている際にかわいい犬に会って嬉しかった場合は快の感情(正の価値)がエピソードに付随し，犬に吠えられて恐かった場合は不快な感情(負の価値)がエピソードに含まれる．これより，我々はエピソード記憶には価値が紐づけられていると考え，我々の提案モデルではエピソード記憶には現在の感覚情報の特徴ベクトルに価値情報が付随し，それがエピソードに基づく意思決定を可能にしている．例えば，分かれ道で左に行けばかわいい犬がいて，右に行けば吠える犬がいるという記憶があれば，私たちは左に行くことが多いだろう．これは，過去のエピソードを想起したときにその価値が同時に想起され，過去の経験を現状に当てはめた結果としての未来の価値が予測されて行動選択がなされた[Miyata 2017]，と説明できる．

さらに，このタスクでは場所ごとに適切な行動が対応する．そのため，エージェントが自己位置を特定できるならば，その場での行動の価値を過去のエピソードから高精度で予測できると期待できる．そこで，海馬にある場所細胞の機能として視覚情報の時系列中で変化の少ない部分を抽出して自己位置を特定する機能をもつ Slow Feature Analysis(SFA) を実装した[Schoenfeld 2013]．以上より，本稿では上記のエピソード記憶のみで行動決定するモデル A と，エピソード記憶と場所細胞の組み合わせで行動決定するモデル B の2つのモデルを検討した．

Fig.2 は本モデルの基本的な構造である[堤 2018]．モデルは認識系から視覚情報の特徴ベクトルXを受け取り，エピソード記憶として溜め込まれている過去の特徴ベクトル群 Y^t より，特徴ベクトルの近いエピソード群 $Z = \{Z^k : \| X - Y^t \| < T\}$ を選ぶ．選ばれたエピソード群Zを行動ごとに平均化して行動に対する平均価値を求めることで，入力された特徴ベクトルに対する行動価値を計算し，SoftMax 法で行動を選択する．同時に，現在の特徴ベクトルと環境から与えられた価値をエピソード記憶として保存する．行動した結果に報酬を得た場合はそのトライアル内のエピソード記憶の価値情報を過去に遡って更新することで，最終結果に応じた価値をエピソードに付与する．エピソード記憶は過去 10,000 ステップのエピソードを保持できるとしているが，エピソード記憶の保持数に関しては検討が必要である．

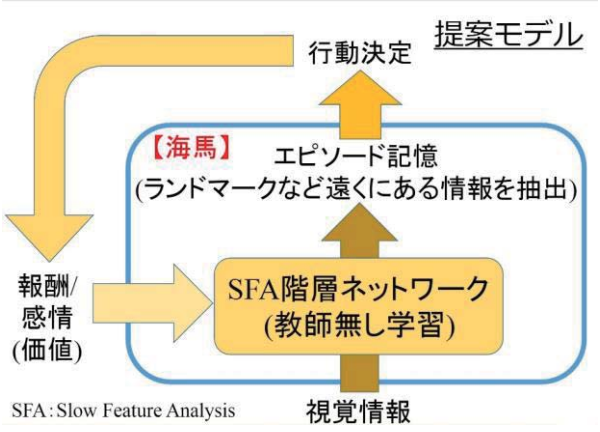


Fig.2 エピソード記憶による意思決定モデルの構造

2.3 学習のパフォーマンスと場所細胞の導入の効果

本ハッカソンでサンプルとして提供された行動学習モデルでは，経験した状態，行動，報酬などをメモリに蓄積し，ランダムサンプリングする Experience Replay を用いて行動決定している[Bendor 2012]．この方法では経験の時系列を学習に利用できないためか，一定の場所で待機するという動作の学習が成功せず，タスク 3 で行き詰った．しかし，我々が提案したエピソード記憶から行動決定するモデルでは，一定の場所で待機することの学習に成功し，サンプルでは解けなかったタスク 3 を始めタスク 7 までをクリアできた．これには，エピソード記憶に価値を紐づけて未来の価値を予測する方式が効果を持った可能性がある．

一方で問題となったのが，エージェントが迷路内で不必要に回転してフィールド内を無駄に往復することで起こる，タスクをクリアするまでのステップ数の多さである．これに対して我々は，エージェントが迷路を不必要に往復するのは Fig.1 の黒い壁の方をエージェントが向いている時の視覚情報からでは，今いる場所が正確に分からないのではないかと仮説を立てた．このことから，場所を特定する場所細胞の機能を SFA で再現し，実装した．場所細胞の効果は大きく，タスク 1 から始めてタスク 7 をクリアするまでにかかったステップ数は，SFA を実装しないモデルAの約 40,000 ステップに対し，SFA を実装したモデルBは 23,000 ステップ強であった．Fig.3 より各タスクのクリアまでの効率が格段に向上していることがわかる．

では，場所細胞を実装したモデルAと実装しなかったモデルBの振る舞いはどう違うのか．次節ではそれぞれのモデルの内部でなにが起きていたか，エージェントの動きを分析することで明らかにする．

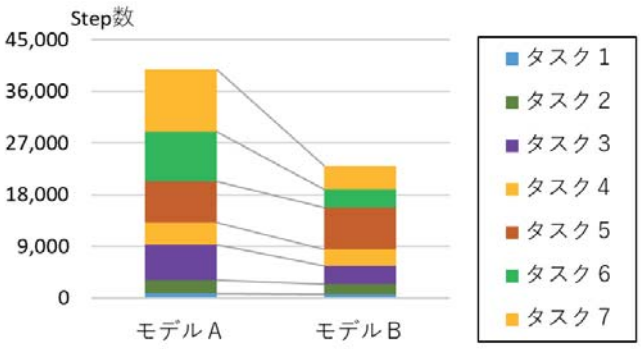


Fig.3 場所細胞がなしのモデルAとありのモデルBでのタスク 7 終了までのステップ数の差

3. モデルの内部過程の分析

3.1 エージェントの軌跡からいえること

我々は，提案モデルの解析を行っていくうえで，タスク 3 に着目した．タスク 1 とタスク2は直進をするだけで報酬が得られたのに対して，タスク3は直進だけでなく，回転や停止などの行動により緑の壁の前で一定時間留まってはじめて報酬が得られる．つまり，直進だけでなく，場面に応じて回転・停止などの行動の選択が必要なタスクであった．この行動選択をどのように学習してタスクをクリアしたかを明らかにすることで，我々の提案したエピソードと価値を紐づけた行動選択モデルの特性が明らかになると期待する．

まずモデルAがタスク 3 をクリアする過程を分析した．価値を含んだエピソードの検索は，過去の類似した場面での価値を用いて，未来の価値を予測する機能を持つ．そのため，このモデ

ルの学習（適切なエピソードの蓄積）によりエージェントの行動がどう変化したか確認した．具体的には，個々の場所に対する行動確率を可視化し，学習によるその変化の傾向を見た．

可視化に用いるエピソードは，タスク 3 で最初に報酬を得たグループと最後に報酬を得たグループから，なるべくステップ数が近いものを選んだ．これを用いて，各ステップにおける行動を進行方向(z 軸)に対応させて Fig.4, Fig.5 に示した．また，両方の図で一定時間留まる必要がある壁の緑色の部分を緑の背景で示した．

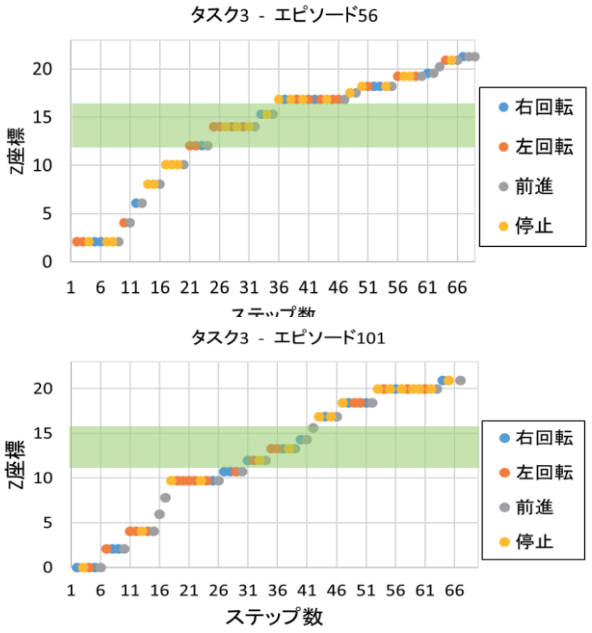


Fig.4 タスク 3 の序盤(a: 上段)と終盤(b: 下段)から抽出したエピソードの進行状況

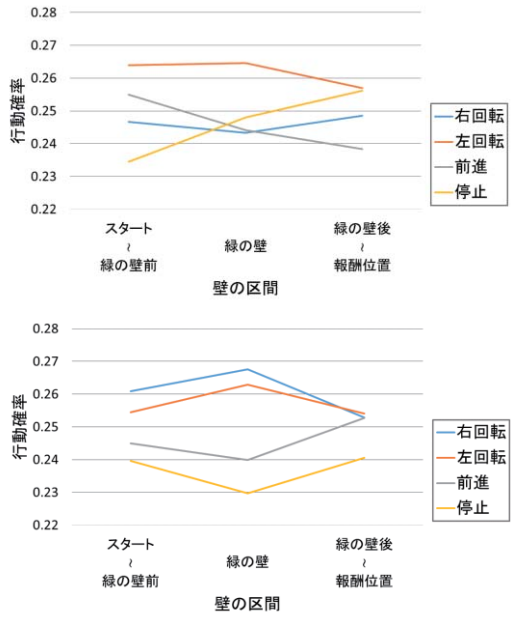


Fig.5 序盤エピソード(a: 上段)と終盤エピソード(b: 下段)の行動確率の平均値

Fig.4(a：上段)では，緑の壁付近で左回転と停止の頻度が高く，壁の前で待てていた．しかし，特定の位置で長く停止していることから，偶然に直進行動が出なかった可能性が高い．それに対し Fig.4(b：下段)では可能な行動がバランスよく選択されて緑の壁の範囲で停止していることから，その場にとどまりやすい行動確率が学習できているようにみえる．

そこで，価値の高いエピソードの蓄積により行動確率がどう変化したか分析した．そのために，タスク 3 全体をスタート位置から緑の壁の前まで，緑の壁の範囲，緑の壁の後から報酬が得られる位置までの 3 区間に分け行動確率の平均値を計算した．Fig.5(a)(b)より，価値の高いエピソードの蓄積により行動全体で右回転と左回転の確率がバランスして進行方向がゴール方向からぶれる可能性が減り，さらに緑の壁の範囲では左右の回転の確率が高くなって前進の確率が下がり，結果としてこの区間の滞在時間が長くなるように変化していることがわかる．

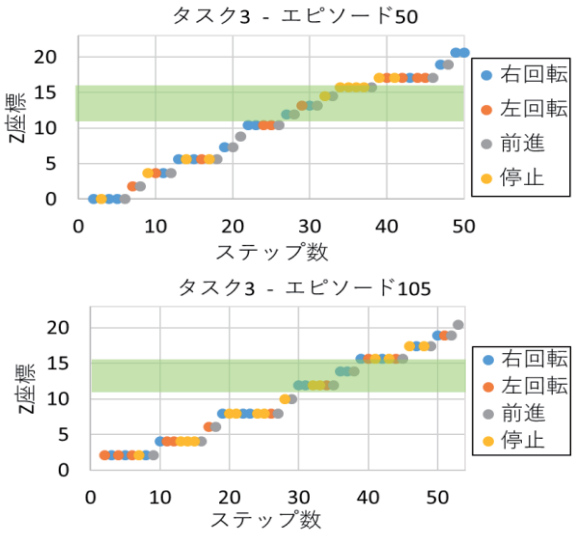


Fig.6 序盤付近のエピソードの経路(a: 上段)と終盤付近のエピソードの経路(b: 下段)

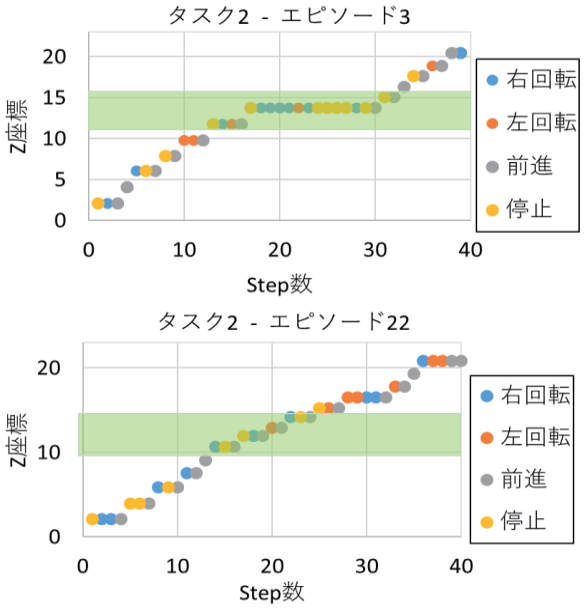


Fig.7 タスク 2 序盤のエピソードの経路(a: 上段)とタスク 2 終盤のエピソードの経路(b: 下段)

以上より，モデルAはエピソードの蓄積により行動学習ができることが確認できた．また，場所に対応して実際に行った行動の

割合だけでなく，行動確率も場所に準じて変動し，タスク 3 の間に効率の良い行動が選ばれやすくなっていることが分かった．

3.2 場所細胞の効果の分析

次に，モデルAに SFA を実装したモデルBで，効率が格段に向上したメカニズムを分析した．そのため，前節と同様の可視化を行った(Fig.6 (a),(b))．これを比較すると，モデルA (Fig.4)とは異なり，エピソードの序盤と終盤での行動の変化が少ないように見える．これより，タスク 2 で報酬を獲得したエピソードにおいてすでに緑の壁の範囲内で停止・回転の行動をとっている可能性を考え，同様の可視化をタスク 2 でも行った(Fig.7 (a),(b))．

Fig.7 (a),(b)から，モデルBではタスク 2 の時点で，緑の壁の前でたまたま長く滞在した場合に報酬が得られ，さらに前進してゴールで報酬を得た，というトライアルがあったと推測される．そういう成功トライアルがあると，以降はその行動を再現するのが本モデルである．すなわち，タスク 2 の段階でタスク 3 に通じる行動学習を行ったことが，タスク群を通じて効率があがった一因であったと推測できる．

では，なぜ SFA の導入が学習を加速したのだろうか？一つの仮説は，位置を特定する精度がモデルB(SFA あり)の場合に高くなったということである．それを検証するため，現在のエピソードに近いとして抽出された過去のエピソードの位置関係を可視化した(Fig.8)．図中の黄色の星が現在のエピソードの位置で，その他の丸が抽出されたエピソードの位置である．上段では，現在の位置に対して遠い位置に類似エピソードが出ているが，下段では近い位置・角度のエピソードだけが抽出された．また，類似とされた個数が上段では 38 個，下段では 102 個と，蓄積されたエピソードの影響もあるが，SFA ありの方が明らかに近い場所のエピソードを検出していた．



Fig.8 SFA を実装しない場合(上段)と実装した場合(下段)の類似エピソードの位置

4. 考察

エピソード記憶に価値を紐づけることによってタスク 1 から始めてタスク 7 までクリアできた．またエージェントの内部計算では，過去の経験から未来の価値がある行動を選択できた．しかし，ハッカソンの課題で与えられた感覚入力のままのエピソード情報では現在の位置(事象)を正確に判別できず，適切なエピソードが選べず，結果として未来の価値予測が狙い通りには働かなかったと思われる．そしてこの問題は，SFA による場所細胞を用いることで大きく改善された．これは，エージェントが現在の位置に関係するエピソードを記憶から的確に選ぶことで現状にマッチした行動－価値関係の計算が可能となり，未来の価値が高いエピソードを精度よく現在に当てはめることができるようになったことによると考えられる．

本ハッカソンのナビゲーションという課題は，場所に依存して行動価値が決まるものであり，エピソード記憶の「どこ」という情報の精度よい認識が成功の鍵となった．同ハッカソンで1位となったチームはその学習に強化学習を用いており，その中の特徴抽出部が本研究の SFA と同じ機能を実現したと考えられる．一方でタスク 1 からタスク 7 までを連続して過去の学習結果を再利用しながら高速に解いていくという効率性からは，場所細胞とエピソード記憶を用いる本モデルは極めて効率的であると考えられる．実際，タスク 3 については 48 エピソード(試行)で完了しており，より困難な課題があるタスク 1 からタスク 7 までの全体でも 361 エピソードで完了している．これは，極端な場合には 1 回の成功エピソードで次からは適切な行動がとりうる本モデルのメリットである．

本分析でわかったことは，エピソードベースの意思決定は，現在の状況判断が重要である，ということであろう．過去のエピソードのうちどれが現在の状況に近いのか？それを的確に選択できたなら，エピソードの再利用は極めて速い学習を可能とする．げっ歯類においては位置の認識は極めて重要であり，それがゆえに場所細胞や環境地図がエピソードの結果として海馬に保存されているとも言えよう．これがより多様な状況のなかで生活する人の場合には，より強力な状況認識の能力が求められる．それが海馬の周辺および大脳皮質に求められる情報処理ともいえよう．

5. まとめ

以上，エピソード記憶に価値を紐づけした行動決定モデルのメカニズムについて，全脳アーキテクチャ・イニシアティブのハッカソンの課題を例に分析した．この課題の特性は，タスク 1 からタスク 7 まで順次に，それまでの学習結果を活かしたまま次の課題に取り組むという，知識の蓄積に相当する機能を要求するものであった．このような課題は現実世界では珍しくはないが，DQN を代表とする Deep Learning による行動学習では学習に時間がかかることが予想され，ここに Deep Learning の一つの限界があるよう思える．それに対してエピソード記憶を用いる方法は，状況変化や新規の経験にすばやい解決を与える意味で，効果的であろう．

大量のデータに基づき特徴抽出から行動決定までを統合的に学習する Deep Learning は，時間をかけてよいのであれば強力である．逆にエピソードによる方法は事例の揺らぎに強く依存し，最適性を求めるのは難しい．両者を統合して，データや経験が増えるにしたがってエピソードベースから階層ネットワークベースに移行していく方式の開発が求められよう．これについては，今後の課題としたい．

参考文献

[WBAI 2017a] 第 3 回全脳アーキテクチャ・ハッカソン，https://wba-initiative.org/2391/

[WBAI 2017b] 第 3 回全脳アーキテクチャ・ハッカソン 迷路課題仕様，https://github.com/wbap/hackathon-2017-sample/wiki/

[Miyata 2017] Masahiro Miyata, Takashi Omori : Modeling emotion and inference as a value calculation system , BICA2017, 2017.

[Schoenfeld 2013] Fabian Schoenfeld, Laurenz Wiskott: RatLab: an easy to use tool for place code simulations , frontiers in Computational Neuroscience, 2013

[堤 2018] 堤 優奈, 栢沼 晋太郎, 川添 紗奈, 宮田 真宏, 大森 隆司: エピソード記憶と価値を紐づけた海馬モデルによる行動学習の分析，第 8 回人工知能学会 汎用人工知能研究会，2018

[Bendor 2012] Bendor, D. & Wilson, M. A.: Biasing the content of hippocampal replay during sleep, Nature Neuroscience, 2012

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

スパイクニューラルネットワークの自己組織化と身体運動の相互作用を通した報酬学習

Reward learning through interaction between self-organization of a spiking neural network and a bodily behavior

河合 祐司 Yuji Kawai　瀧本 友弘 Tomohiro Takimoto　朴 志勲 Jihoon Park　浅田 稔 Minoru Asada

大阪大学 大学院工学研究科
Graduate School of Engineering, Osaka University

Interaction between the brain network and the body is important for behavioral and cognitive development. Behaviors cause sensory input to the brain, and the neural network self-organizes under the sensory stimulus, which leads to the acquisition of new behaviors. We aim to figure out how this brain-body interaction develops from a viewpoint of a constructive approach. We construct a spiking neural network (SNN) model for the reward learning of canonical babbling, i.e., combination of a vowel and consonant. Vocalization is created by the output of the SNN and is fed back into the SNN. Synaptic weights in the SNN are modulated using STDP. Our experiment showed that STDP under auditory feedback enabled the model to rapidly acquire babbling. We found that neurons in the SNN of the model activated much more when vocalizing a consonant. This body representation in the SNN seems to accelerate the acquisition of babbling.

1. はじめに

脳と身体は相互作用しながら機能を発現させる．身体から脳への感覚信号が脳神経ネットワーク活動を誘発し，活動依存的な可塑性によって神経ネットワーク構造が変化する．その自己組織化した神経ネットワークの活動が身体が駆動し，その感覚信号が脳神経にフィードバックされる．このような脳と身体の動的な相互作用は，特に乳幼児期に脳の可塑性が大きく，身体の成長が著しいことから，発達科学において重要視されている [Byrge 14]．しかし，神経ネットワークと身体という異なる性質を持つシステムが相互作用する構成論的なメカニズムは明らかでない．また，このような相互作用において，報酬学習といった課題を与えたときのモデルの学習性能や神経活動の振る舞いは，人工システムへの応用という点でも興味深い．

従来の脳神経モデル研究の多くは身体性を考慮していないが，以下に挙げる研究では，脳神経ネットワークと身体の関係を検討している．Yamada et al. は胎児の脳神経，筋骨格，および，胎内環境の精緻なシミュレーションを構築し，体性感覚信号が可塑性を有するスパイクニューラルネットワーク (spiking neural network: SNN) に入力されることで，ある身体部位の運動時に統計的有意に活動が多くなるニューロンが現れる（身体表現が獲得される）ことを示した [Yamada 16]．しかし，このモデルでは，SNN が身体を駆動することはなく，獲得された身体表現が身体運動へ与える影響は検討されていない．また，Warlaumont et al. は，SNN の出力が発声シミュレータの運動指令となるモデルを提案し，SNN からの出力のシナプス結合を報酬に基づいて更新することで発声の学習が可能であることを報告した [Warlaumont 16]．しかし，このモデルは聴覚などの感覚器を想定しておらず，また，可塑性のない SNN であるため，脳と身体の相互作用的な発展は考慮されていない．Park et al. は神経振動子ネットワークからの運動指令と，筋骨格系からの感覚情報の相互作用によって，身体運動パターンが創発されることを報告したが，神経の可塑性や学習系は考慮されていない [Park 17]．

そこで本研究では，SNN を用いた発声学習の課題 [Warlaumont 16] において，自身の発声の聴覚情報を SNN へフィードバックし，SNN を神経可塑性に基づいて自己組織化させるモデルを提案することで，SNN の身体表現獲得と学習性能の関係について構成論的に検討することを目的とする．従来研究において，音声の顕著性（スペクトログラムの時間変化量の平均）を報酬とした学習により，乳児様の規準喃語，すなわち，母音と子音の両方を含む断続的な発声が獲得されることが示されている [Warlaumont 16]．この従来研究と同様に，SNN の一部のニューロンから出力シナプスを介した出力信号を発声シミュレーションの運動指令とするが，その出力シナプス荷重のみを顕著性の報酬に基づいて更新する．それに加えて本モデルでは，発声音をスペクトル分解し，逐一，SNN に入力し，同時に，SNN 内のシナプス荷重を spiking-timing-dependent plasticity (STDP) 則により更新する．ここで問題となるのは，SNN の活動と発声シミュレータの活動の性質の違い，特に，活動の時間スケールの違いである．このため，SNN のスパイク活動が規準喃語様の発声活動に反映されることは自明ではなく，課題に適した神経活動を SNN が獲得することが望ましい．聴覚信号の時間スケールは発声のものと同じであるため，その聴覚フィードバック下での STDP による発火タイミング学習によって，発声関連の身体表現が SNN の活動に反映されることが期待される．そのような課題に適した神経活動によって，報酬学習が促進され，さらに規準喃語様の発声が増加し，SNN の自己組織化が進む，という身体と脳の相互作用を通した効率的な規準喃語獲得がなされると考えられる．シミュレーションにより SNN の活動と学習性能を解析することによって，上記の仮説を検証する．

2. モデル概要

図 1 に規準喃語獲得モデルの概要を示す．本モデルは主に，SNN，出力層，および，発声シミュレータから構成される．SNN は 1,000 個の Izhikevich スパイクニューロンで構成されており，それらのパラメータは従来研究 [Izhikevich 03] に従う．1 つのニューロンはランダムに選ばれた他の 100 個のニューロン

連絡先: 河合祐司，大阪大学大学院工学研究科知能・機能創成工学専攻，〒 565-0871 大阪府吹田市山田丘 2-1，kawai@ams.eng.osaka-u.ac.jp

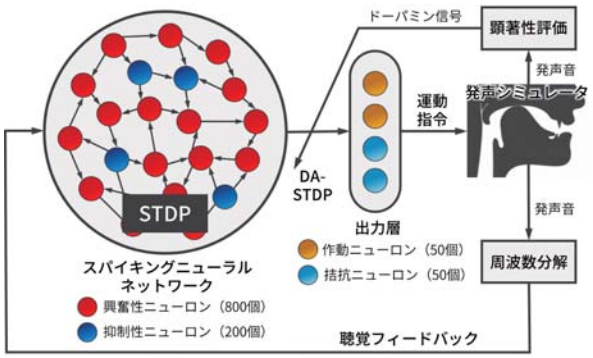


図 1: スパイクニュラルネットワークを用いた規準喃語学習モデルの概要.

とシナプス結合している．出力層は 100 個のスパイクニューロンで構成され，SNN 中のランダムに選ばれた 100 個の興奮性ニューロンから投射を受ける．出力層には，50 個の作動ニューロンと 50 個の拮抗ニューロンがあり，次式により，それぞれの発火数（時刻 t において $S_{\mathrm{e}}(t)$ と $S_{\mathrm{i}}(t)$）の差の履歴に基づいて運動指令 $m(t)$ が決まる．

$$m(t) = g \sum_{i=0}^{100} \left\{ (S_{\mathrm{e}}(t-i) - S_{\mathrm{i}}(t-i)) \left(1 - \exp\left(\frac{-100+i}{\tau}\right)\right)\right\} \tag{1}$$

ここで，運動指令の係数 $g = 0.05$，減衰定数 $\tau = 20$ ms である．この一次元運動指令を発声シミュレータ Praat [Boersma 15] の咬筋と口輪筋に入力することで，口を開閉しながら発声する．その発声音を周波数スペクトル分解し，100 個の帯域（1–20 Hz, 21–40Hz, ⋯, 1,981–2,000Hz）ごとの周波数強度に応じた電流を SNN に入力する．SNN に入力する i 番目の入力 I_i は i 番目の周波数帯域の周波数強度 E_i に比例する．

$$I_i = bE_i \quad i = 1, 2, \cdots, 100 \tag{2}$$

ここで，入力係数 b を 13 とする．それぞれの I_i が SNN 内のランダムに選ばれた 2 つのニューロンに入力される．この周波数帯域と入力されるニューロンの対応関係はシミュレーション中変化せず，SNN の 200 個のニューロンが一貫した帯域の聴覚フィードバックを受ける．

SNN 内の興奮性ニューロンのシナプス荷重を STDP 則と intrinsic plasticity (IP) 則により更新する．STDP は結合関係にある二つのニューロン間のスパイクの時間差に基づいて，それらの間のシナプス荷重を強（弱）化する可塑性である [Izhikevich 06]．この可塑性によって，これらのニューロンの結合が，両者の発火活動の相関関係を反映するようになる．IP 則は，ニューロンの発火率を適切なレベルに維持する可塑性である．一つのニューロンの連続したスパイク間の時間差が許容範囲外であるときに，ニューロンモデルの敏感性に関わるパラメータを更新する [Li 18]．この可塑性により，ランダムな外部入力がなくても，定常外部入力のみにより SNN の自発的な発火を維持することができる．ランダム入力がある場合，表現すべき外部入力が SNN の活動に反映されること（表現獲得）を，そのランダム性が阻害する可能性があるため，IP 則を用いた恒常的な自発発火が有用である．これらの可塑則のパラメータは従来研究 [Li 18] に従う．

SNN から出力層へのシナプス荷重を dopamine modulated STDP (DA-STDP) により，報酬に基づいて更新する [Izhikevich 07]．これは，報酬が与えられたときにのみ STDP を適用する手法であり，これにより，出力層は報酬が得られる発火パターンを生成するようになる．報酬は従来研究 [Warlaumont 16] と同様に，発声音の顕著性が閾値を超えたときに与えられるものとする．顕著性は音声の周波数スペクトルの時間的変化量で定義され [Coath 09]，規準喃語のような母音と子音を多く含む音は大きな顕著性を有する．1,000 ms 間の発声ごとに顕著性を評価し，DA-STDP により出力層へのシナプス荷重を更新する．

3. 実験

3.1 実験条件

本モデルにおける聴覚フィードバックの効果を調査するために，以下の三つのフィードバック入力条件を設けた．

聴覚フィードバック条件 前節で述べたフィードバック方法であり，発声音に基づいた聴覚信号を SNN に入力する．

サロゲートフィードバック条件 現在の試行での発声音ではない，別の試行での聴覚フィードバック条件で得られた聴覚信号を SNN に入力する．

フィードバックなし条件 聴覚入力のない条件であり，IP 則に伴う定常入力と SNN 内で結合している他のニューロンからのシナプスを介した入力によってのみニューロンが駆動する．

さらに，STDP による SNN の自己組織化の効果を調べるために，STDP を適用しないモデルについても実験した．以上の，三種類のフィードバック条件 × STDP の有無の合計六条件で実験した．各条件においてシミュレーション時間 2,000 s の実験を 10 回実行し，それらの平均値で結果を示す．

3.2 ネットワーク活動の解析

2,000 s のシミュレーション後に，シナプス荷重やニューロンパラメータを固定して，さらに 200 s のシミュレーションを実施した．聴覚フィードバックは前節で述べた通りに行った．初めの 100 s は神経活動の過渡状態として除去し，後の 100 s の神経活動を定常状態として解析対象とした．

自己組織化した SNN の神経活動の複雑性を主成分分析の次元数を用いて評価した．この指標は，100 s の神経活動履歴，すなわち，1,000 個のニューロン（1,000 次元）の時系列に対して，主成分分析を施し，累積寄与率が 80%を超える主成分の数を求めたものである．この値は，SNN の神経ダイナミクスを表現するために必要な線形次元数を示しており，元々の 1,000 次元からどれだけ次元を圧縮できるかでその活動の複雑性を表す．

発声運動（運動指令 $m(t)$）に対する SNN の個々のニューロンの振る舞いを，発火率の t 検定により評価した．t 検定により，特定の運動指令の間に発火率が，それ以外の間の発火率に比べて，統計的有意に大きくなったニューロンの個数を求めた．運動指令は -1 から 1 の値をとり，この値が大きくなるほど口がすぼみ，子音を発声し，この値が小さくなるほど母音を発声する．したがって，今回，子音的な発声時（$m(t) \geq 0.5$），母音的な発声時（$m(t) \leq -0.5$），および，それ以外の発声時（$-0.5 < m < 0.5$）に区分して，それぞれの場合について評価した．

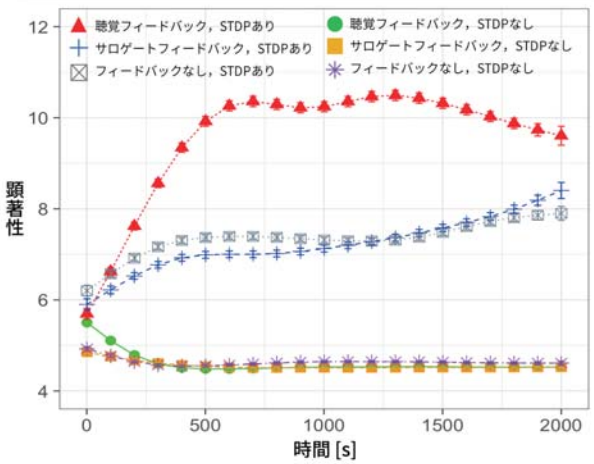


図 2: 顕著性の推移.

運動指令は出力層における作動ニューロンと拮抗ニューロンの差から求められ，これらのニューロンの活動が大きく学習性能に寄与する．そこで，これらのニューロンの平均発火率を評価し，学習性能との関連を検討する．

3.3 実験結果

3.3.1 顕著性の推移

図 2 に発声音の顕著性の推移を示す．図中の曲線は 10 回の試行の平均値に一般化加法モデルをフィッティングしたものであり，エラーバーは 95%信頼度区間を示す．この図より，聴覚フィードバック条件で STDP を適用したモデル（赤色の三角形マーカー）が最も早く学習できたことがわかる．およそ 600 s で顕著性の値が 10 を超え，学習が収束している．また，サロゲートフィードバック条件とフィードバックなし条件で STDP を適用したモデル（それぞれ，青色の十字マーカーと灰色の四角形マーカー）でも，顕著性が緩やかに増加しており，学習が可能であることがわかる．一方で，STDP を適用しないモデル（緑色の丸マーカー，黄色の四角形マーカー，および，紫色の米印マーカー）は学習ができていない．これらの結果より，規準喃語獲得には STDP が必要であり，さらに，聴覚フィードバックが学習を促進させているといえる．

3.3.2 神経ネットワーク活動の複雑性

図 3 に神経ネットワーク活動の複雑性を示す．それぞれの棒の色は図 2 に対応しており，エラーバーは標準偏差を示す．この図より，STDP のあるモデル（赤，青，および，灰色の棒）では，複雑性が大きく，STDP のないモデル（緑，黄，および，紫の棒）では，複雑性が小さくなったことがわかる．したがって，STDP が神経ダイナミクスの複雑性の向上をもたらしたといえるが，聴覚フィードバック条件とサロゲートフィードバック条件の複雑性にはほとんど差がなく，複雑性では図 2 における聴覚フィードバック条件での学習促進を説明できない．

3.3.3 発声を表現するニューロン

図 4 に STDP あり条件において，子音，母音，または，それ以外の発声時に統計的有意に発火率が上昇するニューロン数を示す．聴覚フィードバック条件では，サロゲートフィードバック条件とフィードバックなし条件に比べて，子音発声時に発火率が上昇するニューロンがより多く存在することがわかる．一方で，子音以外の発声時に発火率が上昇するニューロンは，サロゲートフィードバック条件とフィードバックなし条件

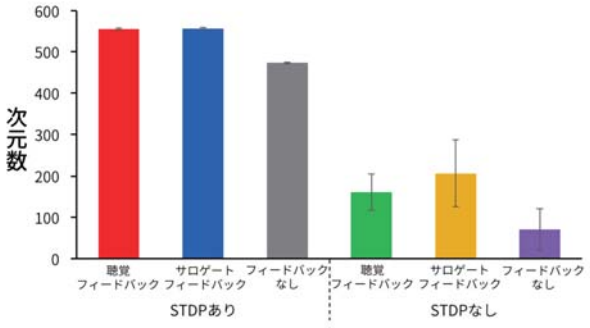


図 3: 神経ネットワーク活動の複雑性.

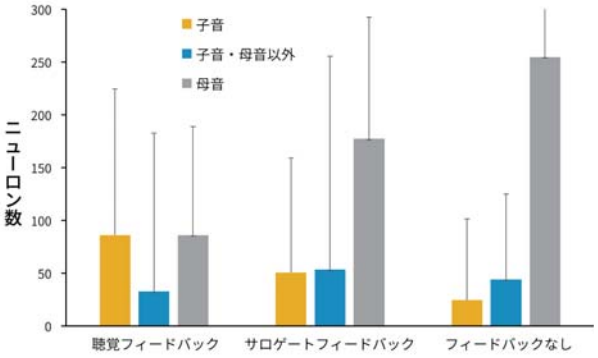


図 4: 子音，母音，または，それ以外の発声時に統計的有意に発火率が上昇するニューロン数．エラーバーは標準偏差を示す．

で多くなっている．大きな顕著性には，多くの子音発声が必要であるため，SNN 内に子音発声に関連するニューロンが存在することは，学習の効率化につながる．聴覚フィードバック条件では，SNN の自己組織化の結果，学習課題上で重要な子音に関する表現が獲得されたことによって，最も早い学習が可能となったと推測される．

3.3.4 出力層ニューロンの活動

図 5 に STDP あり条件において，出力層での作動ニューロンと拮抗ニューロンの 1 s あたりの平均発火率を示す．聴覚フィードバック条件では，サロゲートフィードバック条件とフィードバックなし条件に比べて，出力層ニューロンの発火率がわずかに低下していることがわかる．すなわち，この条件では，出力層における比較的小さな活動コストで，学習が達成されたことを示す．また，聴覚フィードバック条件では，作動ニューロンの活動が，拮抗ニューロンの活動よりも大きくなっている．運動指令はこれらの活動の差で定義されているため，作動ニューロンの活動が大きいことは，口をすぼめる動作が多くなり，子音が多く発声されることを示す．図 4 で示した SNN における子音発声の表現の下での報酬学習によって，出力層ニューロンが大きく活動することなく，子音発声の運動指令を生成できたものと推測される．一方で，サロゲートフィードバック条件とフィードバックなし条件では，SNN の子音の発声とはあまり関係のない複雑な活動から，報酬学習により子音発声に必要なダイナミクスを取り出す必要があるため，学習に時間を要したと考えられる．

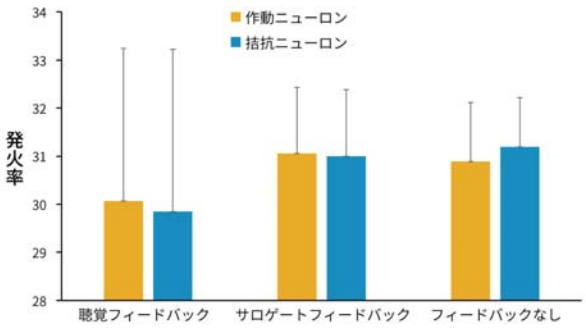


図 5: 拮抗・作動ニューロンの平均発火率．エラーバーは標準偏差を示す．

4. 議論と結論

本研究では，聴覚フィードバック下での STDP による SNN の自己組織化と，DA-STDP による SNN と出力層間の報酬学習によって，規準喃語を獲得するモデルを提案した．実験結果から，身体と脳の相互作用を通した自己組織化が，報酬学習を促進することが示された（図 2）．その一つの要因は，STDP により SNN の活動がより複雑になったことが挙げられる（図 3）．STDP は関連のあるニューロン間の結合を強化し，不必要な結合を除去するため，神経ネットワーク結合を疎にする．このような疎な神経ネットワーク活動は複雑になることが示されている [Friston 96]．複雑な神経ダイナミクスには，様々なパターンが含まれているといえ，DA-STDP により，その中から顕著性の大きな発声に必要なパターンを抽出するように学習したと考えられる．一方で，Laser et al. は，単純なバイナリニューロンのネットワークに STDP を適用することで，活動の複雑性が低下することを示している [Lazar 09]．したがって，STDP により複雑性が向上する結果は，スパイクニューロンのネットワークに特有である可能性がある．

もう一つの要因は，聴覚フィードバック下での STDP により，ニューロンが課題（子音発声）関連の活動を獲得したことである（図 4）．聴覚フィードバック条件において，子音発声時に大きく活動するニューロンが増加したことで，SNN の神経ダイナミクスが子音発声に適したパターンに変化したと考えられる．Park et al. の研究においても，体性感覚信号が神経振動子ネットワークに入力されることで，ネットワーク活動が身体運動に適したものとなり，身体運動のパターン性が現れることが示唆されている [Park 17]．本研究ではさらに，そのような身体表現を持つ神経活動を用いることで，報酬学習が促進されることを示す結果を得た．また，この身体表現により，出力層ニューロンの活動が小さくても，学習が可能となる（図 5）．より少ない神経活動で高い性能を発揮できたことは，代謝の点で優れているといえる．

以上の結果から，聴覚フィードバック条件での SNN は，自己組織化の結果,「複雑さ」と「表現」という，直感的には相反する性質を獲得したといえる．表現は既知の課題や入力に対して，適切に応答するために必要である一方，複雑さは多様な課題や入力に即時的に適応するために必要である．これらを両立した今回の結果は，様々な課題に素早く適応可能な人工システムにおける SNN の有用性を示唆する．今後，学習後の本モデルに新しい学習課題を与えたときの学習性能によって，本モデルの新奇課題に対する適応性を評価したい．

謝辞

本研究は，JST，CREST，JPMJCR17A4 の支援を受けたものである．

参考文献

[Boersma 15] Boersma, P. and Weenink, D.: Praat: Doing phonetics by computer (2015), `http://www.praat.org`

[Byrge 14] Byrge, L., Sporns, O., and Smith, L. B.: Developmental process emerges from extended brain–body–behavior networks, *Trends in Cognitive Sciences*, Vol. 18, No. 8, pp. 395–403 (2014)

[Coath 09] Coath, M., Denham, S. L., Smith, L. M., Honing, H., Hazan, A., Holonowicz, P., and Purwins, H.: Model cortical responses for the detection of perceptual onsets and beat tracking in singing, *Connection Science*, Vol. 21, No. 2-3, pp. 193–205 (2009)

[Friston 96] Friston, K. J.: Theoretical neurobiology and schizophrenia, *British Medical Bulletin*, Vol. 52, No. 3, pp. 644–655 (1996)

[Izhikevich 03] Izhikevich, E. M.: Simple model of spiking neurons, *IEEE Transactions on Neural Networks*, Vol. 14, No. 6, pp. 1569–1572 (2003)

[Izhikevich 06] Izhikevich, E. M.: Polychronization: computation with spikes, *Neural Computation*, Vol. 18, No. 2, pp. 245–282 (2006)

[Izhikevich 07] Izhikevich, E. M.: Solving the distal reward problem through linkage of STDP and dopamine signaling, *Cerebral Cortex*, Vol. 17, No. 10, pp. 2443–2452 (2007)

[Lazar 09] Lazar, A., Pipa, G., and Triesch, J.: SORN: a self-organizing recurrent neural network, *Frontiers in Computational Neuroscience*, Vol. 3, No. 23 (2009)

[Li 18] Li, X., Wang, W., Xue, F., and Song, Y.: Computational modeling of spiking neural network with learning rules from STDP and intrinsic plasticity, *Physica A: Statistical Mechanics and its Applications*, Vol. 491, pp. 716–728 (2018)

[Park 17] Park, J., Mori, H., Okuyama, Y., and Asada, M.: Chaotic itinerancy within the coupled dynamics between a physical body and neural oscillator networks, *PLOS ONE*, Vol. 12, No. 8, p. e0182518 (2017)

[Warlaumont 16] Warlaumont, A. S. and Finnegan, M. K.: Learning to produce syllabic speech sounds via reward-modulated neural plasticity, *PLOS ONE*, Vol. 11, No. 1, p. e0145096 (2016)

[Yamada 16] Yamada, Y., Kanazawa, H., Iwasaki, S., Tsukahara, Y., Iwata, O., Yamada, S., and Kuniyoshi, Y.: An embodied brain model of the human foetus, *Scientific Reports*, Vol. 6, p. 27893 (2016)

Oral presentation | Organized Session | [Organized Session] OS-6

[2L3-OS-6b] [Organized Session] OS-6

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room L (3F Sapphire Hall Asuka)

[2L3-OS-6b-01] Symbol Emergence in Robotics towards Adaptive Autonomous AI Architecture

○Tadahiro Taniguchi[1] （1. Ritsumeikan University）

3:20 PM - 3:40 PM

[2L3-OS-6b-02] Validation of symbolic inference based on a value driven associative memory

○Masahiro Miyata[1], Takashi Omori[2] （1. Graduate School of Engineering, Tamagawa University, 2. College of Engineering, Tamagawa University）

3:40 PM - 4:00 PM

[2L3-OS-6b-03] Active SLAM and Spatial Concepts Formation for a Mobile Robot in Home Environment

○Yoshiki Tabuchi[1], Akira Taniguchi[1], Yoshinobu Hagiwara[1], Tadahiro Taniguchi[1] （1. Ritsumeikan University）

4:00 PM - 4:20 PM

[2L3-OS-6b-04] Stories as Living Representation

○Taisuke Akimoto[1] （1. Kyushu Institute of Technology）

4:20 PM - 4:40 PM

[2L3-OS-6b-05] A Capability Map for Guiding AGI Developments

Masahiko Osawa[2], ○Takashi Omori[1], Koichi Takahashi[3], Naoya Arakawa[4], Naoyuki Sakai[4], Michihiko Ueno[4], Michita Imai[2], Hiroshi Yamakawa[4] （1. Tamagawa University, 2. Keio University, 3. RIKEN, 4. Dwango Artificial Intelligence Labo.）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

3:20 PM - 3:40 PM (Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room L)

[2L3-OS-6b-01] Symbol Emergence in Robotics towards Adaptive Autonomous AI Architecture

○Tadahiro Taniguchi[1] （1. Ritsumeikan University）

Keywords: Symbol emergence, Cognitive architecture, Deep learning, Probabilistic generative models

汎用 AIは本来，地球上の生命がそうであるように，自律的に環境で活動し
その感覚・運動情報から適応的かつ発達的に知能・機能を形成すべきである．
記号創発ロボティクスは感覚運動情報から言語利用までをボトムアップに
繋げ，トップダウンな社会からの制約も考慮に加えて記号創発システムの理解に
構成論的にアプローチする学問領域である．
階層ベイズモデルや深層学習をさらに統合し，より統合的な汎用 AIに
向かっていくための記号創発ロボティクスが目指す自律適応型 AIアーキテクチャ
について議論したい．

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

価値に駆動される連想記憶によるシンボル的推論の検証

Validation of symbolic inference based on a value driven associative memory

宮田 真宏[*1] 大森 隆司[*2]
Masahiro Miyata Takashi Omori

[*1] 玉川大学大学院 工学研究科 [*2] 玉川大学 工学部
Graduate School of Engineering, Tamagawa University College of Engineering, Tamagawa University

Intuitive inference and logical inference are modeled separately as the two different processes in brain. However, a site for the logical inference has not been found in the brain, and it is necessary to model a process that realizes the logical inference in a distributed neural network. In this study, we propose an integrated model in which the intuitive inference is represented as a search process of in a continuous and distributed associative memory, and is switched to a symbolic inference mode that biases an associative gain when it found values during the intuitive inference search. In this study, we show the computational model of symbolic inference by associative memory, and we verify its practicality by solving a maze task. As a result, we observed the intuitive one can be represented as the Bayesian probabilistic inference, and the symbolic one based on value.

1. はじめに

現在，多くの人工知能(AI)を搭載したロボットなどの行動決定には強化学習が用いられている．しかし強化学習を用いた行動学習には多くの試行錯誤が必要であり，学習されるまでに非常に時間がかかることが問題である．それに対して我々人間は，その場の状況にすぐに対応するだけでなく，その場の状況を受け，次に起こる状況を予測する推論をする．従来の AI 研究ではその行動を Tree 探索を典型とする論理的な手法によりモデル化してきた．

さらに我々の推論には，「考える前に体が動いた」などの表現があるように，意識的で論理的な推論とは異なる素早い直観的な推論もある．これら二つの推論に共通することは，推論の後に意思決定またはそれに伴う行動が行われている点である．

では，推論と意思決定にはどのような関係があるのだろうか．我々が意思決定をする際，その状況すべてが既知とは限らない．未知の情報があっても我々は推論をすることができ，意思決定することができる[服部 2015]．これは，推論時には過去に例のない新奇の場面においても，その先に起こる事象を複数かつ並列的に予測し，過去の経験に完全に合致せずともその情報を想起でき，さらにその経験に基づいて見出された情報に価値を割り振り，その価値が最大となる行動を選択する，という一連の内部過程があるからであろうと考えられる[Miyata 2017]．これらより，推論とは広い意味での価値探索であり，我々はその結果を受けて行動決定している，と考えることができる．

そこで本稿では，価値評価に駆動された 2 つの推論過程を，連想記憶モデルを用いて確率的な推論，および論理的な推論により実現できることを示し，簡単なモデルによる迷路探索問題を解くことで，その検証を試みる．

2. 人の推論

認知科学では，人の推論には直観的推論と論理的推論の二種類ある[Evans 2009]とされ，それぞれを別のシステムとしてモデル化してきた(表1)．それに対して，従来の AI における推論の基本モデルには Tree 探索があげられる．Tree 探索は個々の離散状態の予測，およびその評価を行う意識的かつシンボル的な方式である[Russel 2008]．また，我々はシンボル的推論とは別に，何かを知覚するとその影響の予測と評価を素早く行う直観的な推論過程も持っている．これは，感覚刺激からの自動的な連想による無意識的な予測と評価によると考えられる．これを踏まえると，予測は局所的で多方向に並列的に進み，それを価値計算システムが部分的に評価したものが，直観と呼ばれていると考える[大森 2017]．

表1 推論の二重過程と二重システム仮説

直観的推論	論理的推論
作業記憶は不要	作業記憶が必要
無意識的，自律的	意識的， メンタルシミュレーション
推論が速い	推論が遅い
バイアスに影響されやすい	規範的，公平
文脈依存	抽象的
確率的，分散的	論理的，シンボル的
暗黙知（経験的確率）を利用	明示的な知識を利用
推論が浅い	深い推論が可能
進化的に古い	進化的に新しい

3. 連想記憶モデル

3.1 連想記憶とは

連想記憶とは，記憶パターンを貯蔵し，部分的な記憶情報を基に必要な記憶を読み出す機能である．神経回路による分散型の連想記憶モデルの例として，記憶事項を一つのベクトルで表したアソシアトロン[Nakano 1972]や時間的な連想を考慮したモデルがある[Sompolinsky 1986]．これらのモデルでは，複数個の記憶事項の記銘はそれらの相関行列の和(記憶行列)で表し，想起用の入力ベクトルと記憶行列の積を計算することで想起を再現できる．本稿ではこの連想記憶モデルを用いて，人の二種類の推論のモデル化およびその検証を行う．

3.2 直観的推論と論理的推論の実現

先述したように推論は2種類あるとされ，それぞれが別の処理システムとしてモデル化されている．実際に人の推論システムが二種類あるとすると，状況に応じてどちらかを選択して 2 つの推論を制御する上位システムが必要となろう．しかし，そのような上

連絡先：宮田真宏，玉川大学工学研究科，194-0041 町田市玉川学園 6-1-1，mytma4re@engs.tamagawa.ac.jp

位システムは知られておらず，そもそも論理的推論の脳部位も知られていない．そこで本研究では，これら 2 つの推論は 1 つの処理システムで実装され，その動作モードのスイッチングにより切り替えられると考えた．その過程を連想記憶モデルの連想と仮定して表現すると，その計算式は以下になる．

$$y^{t+1} = \alpha \left(\sum_j Pr\left(y_j^{t+1} | y_i^t\right) W_{ij}^e \right) y^t + (1-\alpha) \sum_j W_{ij}^S y^t \cdots (1)$$

(1)式右辺の第一項は，直観的推論を実現する相互想起の項である．ここでは過去の経験を表現する記憶ベクトルからイベントごとの連想行列(W^e_{ij})を作成し，入力ベクトル(y_i)から想起される記憶ベクトルを連想的に探索する．なお，記憶ベクトル群は直交していると想定する．これより想起ベクトルは，入力ベクトルに関係があるほど連想行列から強い強度で想起される．そして想起されたベクトルは過去の経験頻度を表す条件付き確率に基づき強さが決まる．この計算の反復により，短時間で広い範囲に対する連想的な探索が可能となる．そして探索中に価値のある部分を発見した際には，その価値に従い意思決定する．

(1)式右辺の第二項は，第一項で見出された価値に焦点を当て，その価値を最大化させる記憶想起の反復計算を表す．すなわち，入力ベクトル中の価値のある成分を強化して，想起によって特定の価値が支配的になるまで自己想起型の連想計算を反復し，それに対応する行動を選択する [宮田 2018] (図1) .

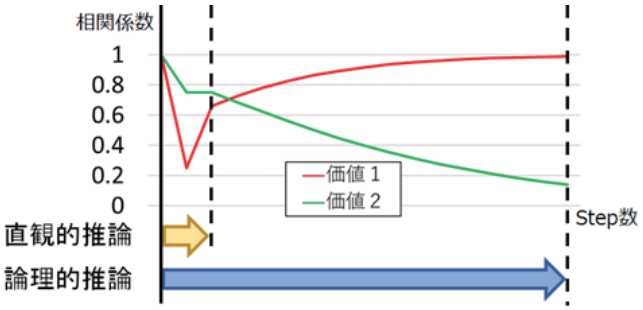


図1 完全直交ベクトルを用いた連想記憶シミュレーション

4. 計算機シミュレーション

本モデルを用いて，直観的推論と論理的推論とを組み合わせによる迷路探索課題の解決シミュレーションを行った(図2)．状態ベクトルは長さ5000の±1(興奮/抑制)からなるランダムベクトルとして，地図上の各位置に対応させた．迷路中の水色の領域は正の価値，濃い紫の領域は負の価値に結び付き，黄緑色の領域は正と負の価値が競合する領域である．エージェントのタスクは，地図上の現在位置から行動を開始し青のキューブ(報酬)を取ることである．

エージェントの推論過程を横軸に想起強度，縦軸に時刻を取りグラフ化したものが図3(左)である．このグラフのように，時間の変化に伴い，はじめに示した過去の経験のみに影響される直観的推論の直後に，その後に得られる価値を予測してその最大化を行う論理的推論の過程が続き，その反復による順次的な推論の振る舞いを実現できた．またこの結果をTree探索風に示したものが図3(右)である．この結果から，Tree 探索の深さ優先探索のような推論をエージェントが行っていることが示せた．

5. まとめ

本稿では，連想記憶モデルを用いた直観的推論と論理的推論のアーキテクチャを提案し，簡単な迷路探索課題の解決過程

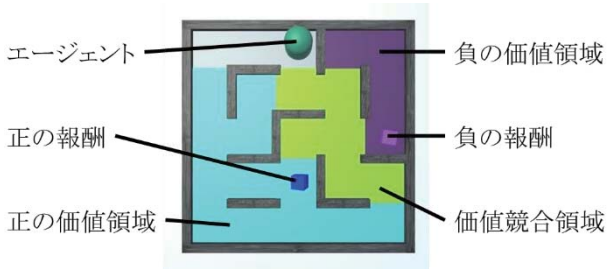


図2 迷路探索用の地図，および価値配置

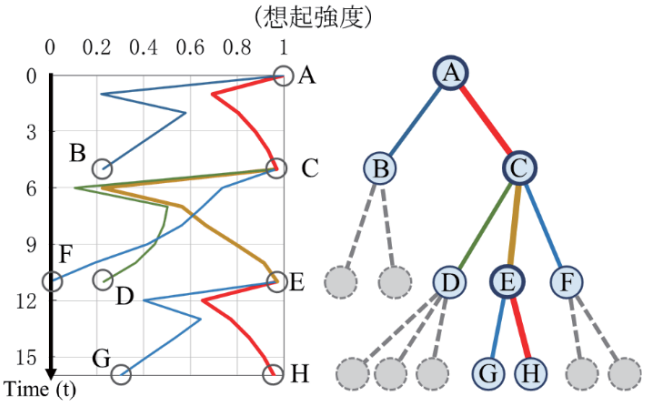


図3 ステップごとのエージェントの推論結果

を示した．しかし，直観的な推論から得られる場所を示すベクトルには経路情報が含まれていない．ランダムベクトルに途中の経路や行動の情報を付加する方法は検討が必要である．

その解決のための一つの方法が粒子モデルであろう．粒子モデルは個々の粒子に経路情報などを付加することが実装上容易である．粒子モデルを用いてもベイズ的な挙動を示す直観的推論が実現できることは既に確認されている[Miyata 2017]．粒子モデルを用いた論理的推論の手法について検討することが今後の課題である．本研究は文部科学省科研費 15H01622 の助成を受けた．支援に感謝する．

参考文献

[服部 2015] 服部雅史: 思考と推論: 理性・判断・意思決定の心理学，北大路書房，2015

[Miyata 2017] Masahiro Miyata, Takashi Omori : Modeling emotion and inference as a value calculation system, BICA2017, 2017

[Evans 2009] Jonathan St. B. T. Evans et al.: How many dual-process theories do we need? One, two, or many? , Oxford Scholarship Online, 2009

[Russel 2008] Russel, Norvig 他:エージェントアプローチ人工知能第 2 版，共立出版, 2008

[大森 2017] 大森隆司，宮田真宏:粒子モデルと価値評価系による直観的推論の計算アーキテクチャ，p.55-56,日本神経回路学会全国大会講演論文集，2017

[Nakano 1972] Kaoru Nakano : Associatron-A Model of Associative Memory, IEEE, 1972

[Sompolinsky 1986] Haim Sompolinsky: Temporal Association in Asymmetric Neural Networks, Physical review letters, The American Physical Society, 1986

[宮田 2018] 宮田真宏，大森隆司: 連想記憶モデルに基づく人のシンボル的推論のモデル化，第 8 回 人工知能学会 汎用人工知能研究会，2018

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

家庭環境における移動ロボットの能動的地図生成と場所概念形成

Active SLAM and Spatial Concepts Formation for a Mobile Robot in Home Environment

田渕 義基 *1 Yoshiki Tabuchi　谷口 彰 *1 Akira Taniguchi　萩原 良信 *1 Yoshinobu Hagiwara　谷口 忠大 *1 Tadahiro Taniguchi

*1立命館大学
Ritsumeikan University

In order for robots to commnucate with users in daily environment, it is important for robots to learn spatial concepts. Although passive learning of spatial concepts has been studied, these researches have a problem that the time burden on users is large for the work that the robot learn spatial concepts. Therefore, we propose a method to reduce the time burden on users by actively operating the robot. In the experiment, we show that our method is effective for reducing the time burden on users.

1. はじめに

近年，ユーザとコミュニケーションを取りながら共同で生活する生活支援ロボットの需要が増加し，開発されている．そのようなロボットがマルチモーダルな観測情報から場所のカテゴリ分類を行う場所概念形成は，ユーザとコミュニケーションを取るうえで重要である．そして，ユーザによるロボットの操作を介してロボットに受動的に場所概念を形成させる研究が行われている [Kostavelis 13, Taniguchi 16, Taniguchi 17]．場所概念を形成することによって，例えばユーザの「キッチンからコップを取ってきて」という指示に対し，ロボットは「キッチン」という語彙からその場所を推測し，「キッチン」に移動することが可能となる．

Taniguchi らは，ユーザがロボットを操作しながら環境を移動し，地図生成及び場所概念形成を行う手法を提案している [Taniguchi 17]．この手法において，ロボットの操作に慣れていないユーザへの時間的負担について議論されていないことが課題として挙げられる．例えば，ユーザが生活する家庭環境においてロボットが地図生成と場所概念形成を行うためには，ユーザがロボットを操作しながら共に環境内を移動する必要がある．しかし，この作業はユーザへの時間的負担が大きいという課題がある．

そこで，本稿ではロボットが家庭環境を能動的に移動しながら地図生成と場所概念形成を同時に行う手法を提案し，ユーザへの時間的負担の軽減における提案手法の有用性を検証する．

2. 提案手法

提案手法の概要図を図 1 に示す．提案手法は，Stachniss らの能動的地図生成手法 [Stachniss 05] と Taniguchi らの場所概念形成手法 SpCoSLAM[Taniguchi 17] を統合したものである．これらの手法は，Grid-based FastSLAM2.0[Grisetti 07] を拡張している手法であるため，容易に統合することができる．これによりロボットは能動的に環境を移動しながら地図生成を行い，ユーザが場所の名前を発話すると，そのときの位置情報・画像情報から逐次的に場所概念の形成を行う．次節以降で統合する 2 つの手法について詳しく述べる．

連絡先: 田渕義基，立命館大学情報理工学研究科，yoshiki.tabuchi@em.ci.ritsumei.ac.jp

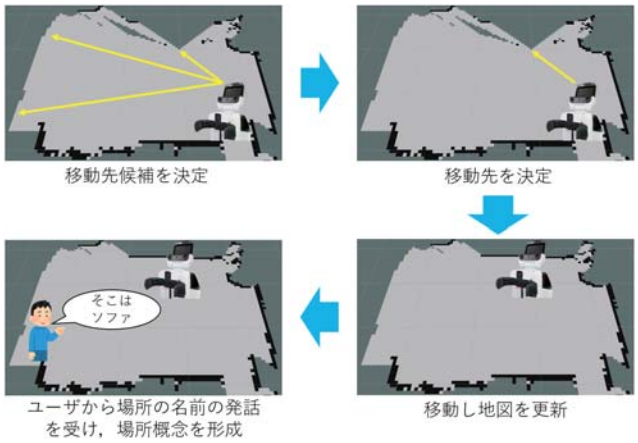


図 1: 提案手法概要図

2.1 能動的地図生成手法

Stachniss らは，ロボットが能動的に環境内を移動して環境の地図を生成する手法として，自己位置と地図の事後確率についてのエントロピーを最小化するようにロボットの動作を決定する手法を提案した．この手法では，Grisetti らの Simultaneous Localization and Mapping (SLAM) 手法 Grid-based FastSLAM2.0 を拡張している．

2.1.1 Grid-based FastSLAM2.0

Grid-based FastSLAM2.0 では，Rao-Blackwellized パーティクルフィルタとして知られるパーティクルフィルタを用いている [Doucet 00]．これにより SLAM における目標である，時刻 t での自己位置と地図の事後確率を以下のように定式化し，因子分解することができる．

$$p(x_{1:t}, m \mid z_{1:t}, u_{0:t-1}) = p(m \mid x_{1:t}, z_{1:t}) p(x_{1:t} \mid z_{1:t}, u_{0:t-1}) \quad (1)$$

ここで式 (1) において，x はロボットの自己位置，m は地図，z は観測値，そして u は制御値を表す．因子分解した式中の $p(m \mid x_{1:t}, z_{1:t})$ は，$x_{1:t}$ と $z_{1:t}$ が求まれば解析的に求めることができる．また，$p(x_{1:t} \mid z_{1:t}, u_{0:t-1})$ を求めるためには，Rao-Blackwellized パーティクルフィルタを使用する．このパーティクルフィルタは，各パーティクルがロボットの自己位置を表現

しており，また地図を持っている．各パーティクルが持っている地図は，対応するパーティクルによって表現される観測値 $z_{1:t}$ と自己位置 $x_{1:t}$ を与えられることによって構築される．そして，各パーティクルは $p(z_t \mid m, x_t)$ に比例した重み ω でリサンプリングされる．$p(z_t \mid m, x_t)$ は，各パーティクルに関連付けられた地図 m と自己位置 x_t が与えられたときの観測値 z_t の尤度である．

2.1.2 エントロピーの求め方

Stachniss らが提案している能動的地図生成の目標は，ロボットの自己位置と地図の不確実性，つまり式 (1) のエントロピー $H(p(x_{1:t}, m \mid z_{1:t}, u_{0:t}))$ を最小限に抑えることである．ここで，$z_{1:t}, u_{0:t-1}$ を d_t とおくと，このエントロピーは以下のように 2 つの成分に分解することができる．

$$H(p(x_{1:t}, m \mid d_t) = H(p(x_{1:t} \mid d_t)) + \int_{x_{1:t}} p(x_{1:t} \mid d_t) H(p(m \mid x_{1:t}, d_t)) dx_{1:t} \tag{2}$$

そして重みづけされたパーティクルを用いることで以下のように近似される．

$$H(p(m, x_{1:t} \mid d_t) \approx H(p(x_{1:t} \mid d_t)) + \sum_{i=1}^{\#particles} \omega_t^{[i]} H(p(m^{[i]} \mid x_{1:t}^{[i]}, d_t)) dx_{1:t} \tag{3}$$

2.1.3 情報利得の期待値の求め方

式 (3) に示したエントロピーの近似式は，ロボットの現在地におけるエントロピーである．能動的地図生成を実現するためには，ロボットがある動作をした後のエントロピーが最小となるような動作，つまりエントロピーの変化量が最大となる動作を実行して自律移動する必要がある．エントロピーの変化量を表す情報利得は以下の式で示される．ここで，予測動作を $a_t = \hat{u}_{1:T-1}$，予測自己位置を $\hat{x} = \hat{x}_{t+1:T}$，そして予測観測を $\hat{z} = \hat{z}_{t+1:T}$ とおく．

$$I(\hat{z}, a_t) = H(p(m, x_{1:t} \mid d_t)) - H(p(m, x_{1:t}, \hat{x} \mid d_t, a_t, \hat{z})) \tag{4}$$

しかし，予測観測 $\hat{z}$ は実際に次の動作を実行しなければ分からないため，式 (4) について $\hat{z}$ で積分する必要がある．すると，情報利得の期待値は，次のように表される．

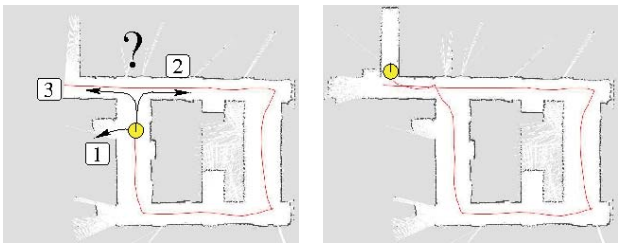
$$E[I(a_t)] = \int_{\hat{z}} p(\hat{z} \mid a_t, d_t) \cdot I(\hat{z}, a_t) d\hat{z} \tag{5}$$

そして，式 (5) は次のように近似される．

$$p(\hat{z} \mid a_t, d_t) \approx \sum_{i=1}^{\#particles} p(\hat{z} \mid a_t, m^{[i]}, x_{1:t}^{[i]}, d_t) \cdot \omega_t^{[i]} p(m^{[i]} \mid x_{1:t}^{[i]}, d_t) \tag{6}$$

近似式 (6) を式 (5) に代入し，情報利得の期待値を計算することは可能ではあるが，大量の計算資源を必要とする．そこで，パーティクルの集合から重みに比例する確率に従ってサンプリングしたパーティクル v によって予測観測 $\hat{z}$ を近似する．すると以下のように情報利得の期待値を表すことができる．これにより計算の複雑さは軽減され，リアルタイムで能動的地図生成を行うことができる．

$$E[I(a_t)] \quad \approx \quad I(\hat{z}(v), a_t) \tag{7}$$



(a) 3 つの動作候補ある場合，それぞれに対して式 (9) により評価する　(b) 式 (9) により評価した結果，3 番の動作が最適な動作として選択される

図 2: 能動的地図生成における動作候補及び最適な動作 [Stachniss 05]

2.1.4 最適な動作の決定方法

ロボットが最適な動作を選択する際に，考慮しなければならないのは情報利得の期待値だけではない．考慮しなければならないもう一つの点は，移動コストである．動作 a_t を実行したときの移動コストを $cost(a_t)$，α を任意の係数とすると，動作 a_t の有用性は以下のように定義される．

$$E[U(a_t)] = E[I(a_t)] - \alpha \cdot cost(a_t) \tag{8}$$

このように定義することで，情報利得の期待値と移動コストをトレードオフに扱うことができる．そして式 (8) で表される有用性が最大のものが最適な動作として選択される．それは以下のように表される．

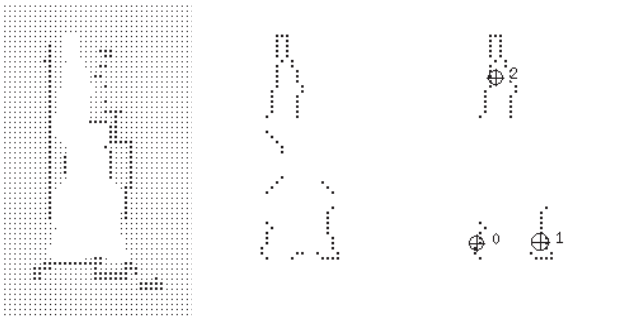
$$a_t^* = \underset{a_t}{\operatorname{argmax}} E[U(a_t)] \tag{9}$$

2.1.5 最適動作候補の決定方法

ここまで最適動作の決定方法を説明してきたが，環境内で実行することのできる動作の数は膨大であり，それらすべてについて，式 (9) を評価することは難しい．そこで，最適な動作を決定するために，ロボットの動作を探索動作と再訪問動作の 2 種類に分ける．探索動作は，地図の事後確率のエントロピーを低減する動作であり，Yamauchi らのフロンティアアプローチを適用する [Yamauchi 98]．再訪問動作は，ロボットが既知の場所を再び訪れることで，自己位置の事後確率エントロピーを低減する動作である．これらの動作を候補とすることで，効率良く最適な動作を決定することができる．その例が図 2 である．図 2(a) に示す動作 1 および 3 は探査動作であり，動作 2 は再訪問動作である．式 (9) により評価した結果，最適な動作は動作 3 となりロボットは図 2(b) に示されるように自律移動をする．

2.1.6 フロンティアアプローチ

Yamauchi らのフロンティアアプローチは，能動的地図生成において生成途中の占有格子地図のオープンスペースと未知領域の境界を動作候補とする手法である [Yamauchi 98]．その概要図を図 3 に示す．図 3(a) の白色の格子で表される部分がオープンスペースと定義され，障害物がなくロボットが移動可能なスペースである．次に，黒色の格子で表される部分が壁や障害物に占有されており，ロボットが移動不可能なスペースである．そして，灰色の格子で表される部分が未知領域と定義され，ロボットが未観測のスペースである．図 3(a) の占有格子地図からオープンスペースと未知領域の境界を検出したものをフロンティアエッジと定義し，示したものが図 3(b) である．そして隣接するフロンティアエッジを一つのフロンティ



(a) 占有格子地図：白色の格子がオープンスペース，黒色の格子が占有されているスペース，灰色の格子が未知領域

(b) 占有格子地図から検出したオープンスペースと未知領域の境界（フロンティアエッジ）

(c) 隣接するフロンティアエッジを一つのグループ（フロンティアグループ）とし，ロボットの大きさ以下のものを排除

図 3: フロンティアアプローチの概要図 [Yamauchi 98]

アグループとし，ロボットの大きさ以下のものを排除して示したものが図 3(c) である．ロボットの大きさ以下のフロンティアグループを排除する理由は，その場所にロボットが移動することは不可能だからである．このようにして図 3(c) の例では，3 つのフロンティアグループがロボットの探索動作の候補となる．

2.2 場所概念形成手法

Taniguchi らは，ロボットが地図生成を行っている既知環境において場所の名前と位置情報から場所概念を形成する non-parametric Bayesian spatial concept acquisition method (SpCoA) を提案した [Taniguchi 16]．また，SpCoA と Grid-based FastSLAM2.0 をベイズ生成モデルの枠組みで統合することでで，ロボットが地図生成を行っていない未知環境においてユーザがロボットを操作して行う地図生成，および場所の名前・位置情報・画像情報から場所概念の形成を逐次的に行う Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping (SpCoSLAM) を提案した [Taniguchi 17]．SpCoSLAM では，場所のカテゴリゼーションと教師なし語彙獲得，および地図生成を 1 つのモデルとして表現することにより，互いの情報の不確実性を相互に補完することが可能となっている．

3. 実験

3.1 実験目的

本実験では，図 4 に示す実験環境において 3 つの手法による地図生成と場所概念形成を行う．3 つの手法とは，提案手法，ユーザの操作によりロボットが移動する SpCoSLAM (User Operating)，そしてロボットが不規則に自律移動する SpCoSLAM (Random Movement) である．3 つの手法における被験者の作業時間を被験者への時間的負担の指標として比較し，提案手法のユーザへの時間的負担軽減への妥当性を検証する．

3.2 実験条件

実験には，図 5 に示すロボットを用いた．カメラはロボットの前方の画像情報のみを取得するように取り付けた．PC の OS は Ubuntu 14.04 であり，Robot Operation System (ROS) indigo によりロボットを制御した．能動的地図生成は，Grid-based FastSLAM2.0[Grisetti 07] が実装されている ROS の *gmapping* パッケージと *turtlebot_navigation* パッケージを拡張することで実装し，レーザレンジファインダの最大測定距離は 2.6 m とした．実験環境は，図 4(a) に実験環境において形成された占有格子地図を示す．実験環境の大きさは 6.7 m × 4.0 m である．被験者数は 4 名であり，この実験環境において被験者は，ロボットが図 4(b) に示す 3 つの場所に位置する時に場所の名前の発話を各々1 回行う．本実験における被験者の作業には，ロボットの操作，ロボットへの場所の名前の発話が含まれる．実験終了条件は，3 つの場所の名前の発話を完了し，かつ実験環境内の地図生成が完了した場合である．



(a) 実験環境で形成された地図

(b) 各マーカーと場所の名前の関係

図 4: 実験環境で形成された地図と場所の名前の関係



図 5: 実験に用いたロボット

3.3 実験結果

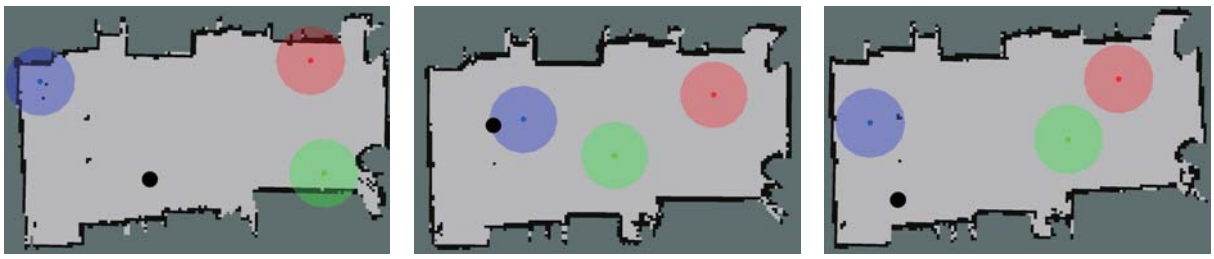
表 1 に，各被験者が 3 つの手法によりロボットに場所概念形成を行わせた際に要した被験者の平均作業時間とロボットの平均稼働時間を示す．また，図 6 に 3 つの手法により形成した場所概念の一例を示す．

表 1 より提案手法と SpCoSLAM (Random Movement) は，被験者の平均作業時間が 3 秒台であるが，SpCoSLAM (User Operating) は被験者の平均作業時間が 1 分を超えていることが分かる．この要因は，提案手法と SpCoSLAM (Random Movement) には被験者がロボットを操作する必要がないためである．また，提案手法のロボットの平均稼働時間は，SpCoSLAM (User Operating) の 220 %，SpCoSLAM (Random Movement) の 70 % であることが分かる．提案手法のロボットの平均稼働時間が SpCoSLAM (User Operating) より長くなった要因としては，ロボットが自律移動する際に各時刻において移動先までの経路生成に多くの時間を要したと考えられる．そして，提案手法のロボットの平均稼働時間が SpCoSLAM (Random Movement) より短くなった要因としては，第 2.1 節で説明した手法によりロボットが環境内を自律移動することで，ロボットが環境内を不規則に自律移動するよりも効率的に地図生成を行うことができたと考えられる．

図 6(a) より SpCoSLAM (User Operating) では，実験環

表 1: 被験者の平均作業時間とロボットの平均稼働時間

	SpCoSLAM (User Operating)	SpCoSLAM (Random Movement)	提案手法
被験者の平均作業時間	1 分 12 秒 6	3 秒 8	3 秒 6
ロボットの平均稼働時間	1 分 12 秒 6	3 分 39 秒 2	2 分 43 秒 3



(a) SpCoSLAM (User Operating) (b) SpCoSLAM (Random Movement) (c) 提案手法

図 6: 各手法による場所概念形成結果

境に近い場所概念を形成させることができた．しかし図 6(b) と図 6(c) より SpCoSLAM (Random Movement) と提案手法では，実験環境に近い場所概念を形成すること困難であった．SpCoSLAM (User Operating) では被験者がロボットを操作するため任意の位置で場所の名前を発話することができるが，提案手法と SpCoSLAM (Random Movement) では，ロボットが自律移動するため任意の位置で場所の名前を発話することができないからであると考えられる．

4. おわりに

本稿では，能動的地図生成手法と場所概念形成手法を統合することで，ロボットが能動的に環境を移動しながら地図生成を行い，ユーザから場所の名前を発話されるとそのときの位置情報・画像情報から逐次的に場所概念の形成を行う手法を提案した．次に，ユーザがロボットに場所概念形成を行わせる作業におけるユーザの時間的負担について従来手法と比較することで，ユーザによるロボットの操作が不必要になったことによる時間的負担軽減への妥当性を示した．一方，従来手法ではユーザがロボットを操作するためユーザの意思通りにロボットに場所概念形成を行わせることができるが，提案手法ではロボットは能動的地図生成手法に基づいて環境内を自律移動するため，ユーザの意思通りにロボットに場所概念形成を行わせることができなかった．

今後の展望として，ロボットに場所概念形成を行わせる作業におけるユーザの時間的負担をさらに軽減していくことが求められる．具体的には，ロボットが場所概念を形成すべき場所を能動的に知覚し，ユーザに場所の名前の発話を促すことで，ユーザが常にロボットの位置を把握しておく必要がなくなり，ユーザの時間的負担軽減につながると考えられる．また，ユーザの意思通りにロボットに場所概念形成を行わせるために，ロボットの最適な動作の決定方法を場所概念形成手法に適応させていくことも求められる．

参考文献

[Kostavelis 13] I. Kostavelis and A. Gasteratos, ”Learning spatially semantic representations for cognitive robot navigation,” Robotics and Autonomous Systems, vol. 61, issue. 12, 2013, pp. 1460-1475

[Taniguchi 16] A. Taniguchi, T. Taniguchi and T. Inamura, ”Spatial Concept Acquisition for a Mobile Robot That Integrates Self-Localization and Unsupervised Word Discovery From Spoken Sentences,” Transactions on Cognitive and Developmental Systems, vol. 8, no. 4, 2017, pp. 285-297

[Taniguchi 17] A. Taniguchi, Y. Hagiwara, T. Taniguchi and T. Inamura, ”Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping,” Intelligent Robots and Systems (IROS), 2017, pp. 811-818

[Stachniss 05] C. Stachniss, G. Grisetti and W. Burgard, ”Information gain-based exploration using Rao-Blackwellized pa rticle filters,” RSS, 2005, pp. 65-72

[Grisetti 07] G. Grisetti, C. Stachniss and W. Burgard, ”Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters,” Transactions on Robotics, vol. 23, no. 1, 2007, pp. 34-46

[Doucet 00] A. Doucet, N. de Freitas, K. Murphy and S. Russell, ”Rao-Blackwellised particle filtering for dynamic Bayesian networks,” in Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence, pp. 176-183, Morgan Kaufmann Publishers Inc. (2000)

[Yamauchi 98] B. Yamauchi, ”Frontier-based Exploration Using Multiple Robots,” Proceedings of the Second International Conference on Autonomous Agents, 1998, pp. 47-53

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

生きた表象としてのストーリー —創発的な世界構築の計算モデルのための考察—

Stories as Living Representation: A Discussion Toward a Computational Model of Emergent World Making

秋元 泰介
AKIMOTO Taisuke

九州工業大学大学院情報工学研究院知能情報工学研究系
Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

One of the essential natures of the human intelligence is to mentally organize the world in the form of narrative or story, through the interaction with physical or social environments. Based on this thought, we are tackling the systematization of essential mechanism of a narrative-centered cognitive/social architecture for autonomous agents. The focus of this paper is on the cognitive mechanism of generating a story as a mental representation of the external world. We sketch the concept of an emergent story which develops from the internal generative actions of each representational element in the story.

1. はじめに

人間の知能の本質の一つは，外界（諸々の物理的・社会的環境）との相互作用を通じて，心の中で，世界を物語的に組織化していくことである．このような考えを，人工エージェントのための自律的な知能の枠組みとしての認知／社会アーキテクチャとして具現化することが，本研究の大局的な目標である．

本研究では，物語の生成を，相互に関連し合う以下の三つの水準に分けて捉える．

- 一つ目，社会的水準．複数の個体間の共創的な相互作用を通じて多様な物語が生成していく仕組み[秋元 16]．
- 二つ目，心的水準（多）．ある個体の心の中で，様々な環境との相互作用を通じて，多数のストーリー（心的表象）が生成・組織化されていく仕組み[Akimoto 18]．
- 三つ目，心的水準（一）．二つ目を背景として，その中で一つのストーリー（これもまた複雑な構造物である）が生成する仕組み．

これらの仕組みを全体として体系化することが当面の課題であるが，本稿では，三つ目の水準（一つのストーリーの生成）に焦点を合わせてこの問題に取り組む．これは二つ目の水準（多数のストーリーの組織化）を背景とするため，これについても少し述べる．一つ目の社会的水準には本稿では立ち入らないこととする．

2. 世界の表象としてのストーリー

本節では，ストーリーは世界の表象であり，心の中に多種多様なストーリーを生成・組織化していくことは，個体にとっての世界を構築し，物理的・社会的な環境に適応していくことである，という考えについて述べる．

2.1 物語，ストーリー，物語言説

本研究では，「物語」（narrative）という用語を，現実または虚構の世界における一つ以上の事象を筋立てたものないし筋立てること，という意味で用いる．さらに，心的表象としての物語と，言語（またはその他の表現媒体）により表現された物語とを，「ストーリー」（story）と「物語言説」（discourse）というように呼び分ける．これらは物語論[Genette 72; Prince 03]に由来する用語であり，その本来の意味において，ストーリーは物語の内容面（何を語るか），物語言説は物語の表現面（如何に語るか）を指すが，本研究ではこれらを心的表象（representation）と表現（expression）という関係に置き換えている．

以上のような意味でのストーリーと物語言説の関係を，送り手-受け手間のコミュニケーションに当てはめると，次のようになる．送り手にとってのストーリーは話の内容に相当する心的表象（例えば過去の記憶）であり，それを言語またはその他の表現媒体により表現したものが物語言説である．受け手側では，物語言説を解釈・理解することによって，心的表象としてのストーリーが形成される．

2.2 ストーリーの性質

ストーリーとしての世界は，事を中心とする，時間的な広がりのある，主観的に構成された世界である．ある個体にとって，一つ一つのストーリーは，一つ以上の事象を結び付けた一塊の小さな世界であり，過去・未来・仮想・虚構を成す多数のストーリーが組織化されたものが，個体にとっての世界（全体）である，というように考える．ストーリーの性質を，「時間」と「意味」という二つの観点から述べる．

時間: 人間は，過去のことを思い出したり，未来のことを思い描いたりしながら生活している．人間の現在進行的な知覚や行動は，常に，様々な水準の時間的文脈（過去や未来との結び付き）の中で生じる．物・者の意味には，時間との関わりのもとに生じる面がある（自分の人生，他者に関する記憶，物の用途，場所の歴史や記憶，等）．事象は，動き続ける世界を時間的に分節化することによって形作られる．世界をストーリーとして構築することは，以上のような，人間的な時間を作り出すことである．

意味: 人間は，同じような状況・事態においても，多かれ少なかれ，人それぞれが異なる経験をする．例えば，ボクシングの試合が行われている会場において，選手と観客の経験は当然異なるだろうが，観客どうしの間でも，ボクシングに関する知識の多少や，どちらの選手に肩入れするか等によって，同じ試合に対して各々が異なるストーリーを見出すだろう．但し，ストーリーの作り方には文化的に共有される面があるため，人それぞれが全く異なる世界（ストーリー）を持つわけではない．これは，本研

連絡先：秋元 泰介，akimoto@ai.kyutech.ac.jp

究の枠組みにおいては，社会的水準と心的水準の相互浸透的な関係として位置付けられる．

2.3 人工エージェントの心的世界

本研究では，人工エージェントの内部に形成される，多数のストーリーの集積としての世界のことを，「心的世界」と呼ぶ．その全体像を[Akimoto 18]が示している．ここに含まれ得るストーリーの種類は多岐に渡る．例えば，物理的世界（外界）における経験の記憶，物語のコミュニケーション（ニュース，文学作品，歴史叙述，世間話，他）を通じて形成されたストーリー，未来の予想・計画・展望，仮想的または虚構的な想像といったものが挙げられる．ストーリーそのものは，事象や実体（人物・物・場所）の形式的な関係構造を表し，その意味的側面は，概念・スキーマ・心像という 3 種類の心的要素（表象）との結び付きによって表現される（図 1）．これら各要素の基本概念を簡単に説明する．

ストーリー: ストーリーは，一つ以上の事象が時間的・因果的・意味的等の関係によって結び付けられたものである．加えて，ストーリーに関連する実体（人物・物・場所）の関係構造を表わす「ストーリー世界」が，ストーリーに包含される形で付随する．

概念: 概念は，動詞的に表現される事象や，名詞的に表現される実体の抽象的水準の意味を形作る心的要素である．本研究では概念を，一般的な単語（主に名詞・動詞・形容詞・副詞）に相当する「一般的概念」と，個別的な存在者・物，及び言語的に分節化された時間・空間（場所）を指し示す「存在的概念」の 2 種類に大別する．

スキーマ: スキーマは，物語のある構造的側面（主として，事象間の関係，実体間の関係，語り方）に対応する抽象的かつ一般的な構造の枠ないしパターンに相当する知識構造である．その構成要素は複数の概念またはスキーマである．スキーマは，主として，物語のトップダウン的な認知や，抽象的水準の認知の基盤となる．

心像: 心像は，言語・視覚・聴覚・嗅覚・味覚・触覚等の感覚・知覚情報に対応する心的なイメージであり，ストーリーの部分（事象・場面・実体）に紐付けられる．

2.4 ストーリー生成の諸側面

心的世界におけるストーリーの生成がどのような状況の中で生じるのかを，三つの観点から述べる．

行動・知覚: エージェントが諸々の環境の中で行動することは，自身を主人公とする行動計画のストーリー（未来）を生成しながら，それを実行していくことに対応する．そして，一連の行動の結果や知覚を統合することによって，過去のストーリー（エピソード記憶に相当）が作られる．

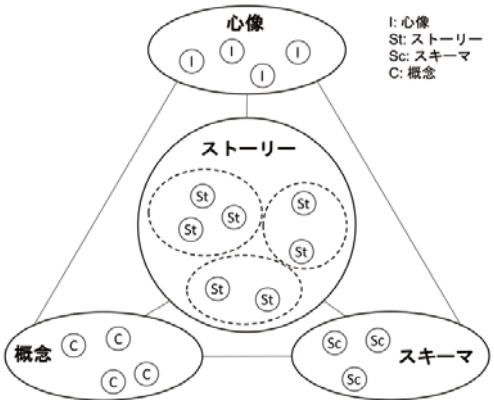


図 1 心的世界の全体像

コミュニケーション: 様々な社会的環境において，他者との間で行われる物語の生成や受容を通じて，ストーリーが生成する．現代の人々は，現実世界を構成する情報の多くをニュース，歴史叙述，噂話等といった物語から得ているだろうし，文学・ドラマ・映画・漫画のような虚構の物語からも膨大な量のストーリーを得ている．また，虚構物語の受容経験によって，現実世界の見方（ストーリーの作り方）が変化することもある．

想像: 実際に見聞きしたことの他に，仮想や虚構のストーリーを作り出すことも，人間的な知能における重要な性質である．例えば，人間は，山道を歩いているときに茂みから熊が現れるかも知れないと警戒し，土曜の早朝に街中で寝そべっている人を見て昨晩飲みすぎたのだろうと思い，風呂に浸かりながら日本の将来を憂慮し，喫茶店で珈琲を飲みながら人を楽しませるための作り話を考えることができる．

2.5 環境への適応としてのストーリー生成

心の中で多種多様なストーリーを生成・組織化（忘却や変化を含む）していくことは，エージェントが物理的・社会的な環境に適応していくことである（この点に関する主な背景として，Schank らによる一連の人工知能研究を挙げる[Schank 77, 82, 90]）．エージェントが，様々な環境の中で活動するためのストーリーを動的に生成していくこと（生成できるようになっていくこと）は，自律的な知能の基盤となる．例えば，日常的な経験を記憶し物語ること，社会的な出来事に関する会話をすること，公共交通機関や飲食店を利用すること，仕事の手順を覚えること，様々な危険を想定しながら安全に街を移動すること，様々な社会的文脈の中で他者が心の中に持つストーリーを想像すること，こうした活動に共通して，ストーリーを作り出す能力が求められる．

人間が，様々な経験を積み重ねながら少しずつ多様な環境に適応していくように，自律的な知能は，発達的に学習する仕組みを持たなければならない．この点に関しては，[Piaget 48]や[Schank 82]の知能観を参考にして，次のように考える．心的世界における既存のストーリー・スキーマが，新しいストーリー・スキーマを作るための知識・資源になる．慣れ親しんだ環境に対しては既知のストーリー・スキーマを再利用し，未知の環境には，既知のストーリーやスキーマを再構成することによって適応しようとする．このような形で，単純な構造を作ることからはじまり，徐々により複雑な構造を作れるようになっていく．このような原理を探っていく必要がある．この問題に対する現時点での考えは，概略として次のようになる．

心的世界の内部に，各種の環境に対応するストーリー生成の場として，「心的空間」が形成されていく．心的空間におけるストーリー生成は，一般的水準の構造であるスキーマによって統制される．つまり，ストーリーの一般化が，心的世界の組織化における一つの鍵となる．また，新しい心的空間は，既存のストーリー・スキーマの再構成によって作られるというように考えるが，それは，既存の心的空間では対応できない環境に直面した際に，それに適応するために，既存の心的空間から逸脱しようとする力によって生じる．あるいは，芸術的創造のように，新しい何かを作り出そうとする内発的な力によって生じる場合もある．以上の概念の模式図を図 2 に示す（破線の枠は心的空間を表す）．

3. 生きた表象としてのストーリー: 概念的な整理

以上の枠組みを背景として，その中で一つ一つのストーリーが生成する仕組みについて検討する．

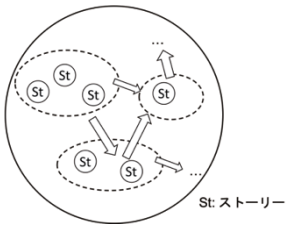


図 2 心的世界を構築していく過程の模式図

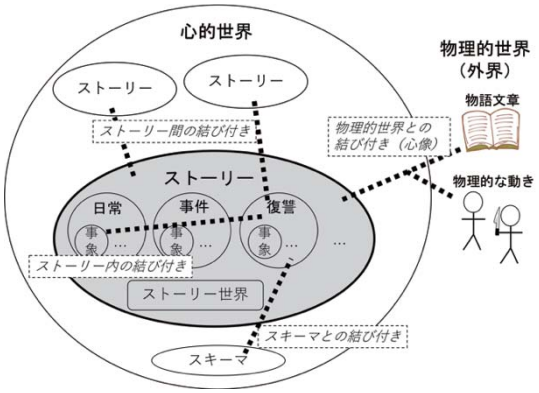


図 3 心的世界におけるストーリー

3.1 ストーリーの複雑性

はじめに，ストーリーの構造的な複雑性について述べる．一つのストーリーの構造は，当のストーリー内の構成要素間の関係構造に当たる「ストーリー内的構造」と，当のストーリーとは別の表象との関係ないし結び付きとしての「間表象的構造」という，二つの観点から捉えられる．図 3 にその全体像を示す．

(1) ストーリー内的構造

[Akimoto 17]は，計算論的な観点から物語の構造を体系化している．これを心的表象としてのストーリーの内的構造に応用してみる．

構造の多次元性: [Akimoto 17]は，テクスト（表層的表現）としての物語に対応する深層的な構造を，表現の構造（語り方）である物語言説（discourse），そこで語られる内容としての事象群の関係構造であるストーリー（story），ストーリーに関与する実体群（人・物・場所）の関係構造であるストーリー世界（story world）という，3 つの構造的次元に分けている．これらは，“(物語言説 (ストーリー (ストーリー世界)))”という形で包含的に統合される．すなわち，物語言説はストーリーについての語りであり，ストーリーはストーリー世界に含まれる実体群の動きに対応するというように，内側の次元が外側の次元の前提的な要素になる．同時に，ストーリー世界はストーリーに関連する実体情報のまとまりであり，ストーリーは語られることによって構成される事象情報のまとまりであるというように，外側の次元が内側の次元を限定するという関係もある．以上の枠組みを心的表象としてのストーリーに応用すると，一つのストーリーは“(ストーリー (ストーリー世界))”という構造になる．

階層と関係: ストーリーとストーリー世界は，それぞれが，複数の部分が集まってより大きな部分を形作るという階層的な構造と，部分どうしが様々な関係で結び付いたネットワーク状の構造とを併せ持つものと考える．[Akimoto 17]は，それを階層グラフ構造として整理している．その基本的な考え方は概略次の通りである．ストーリーにおける部分の最小単位は事象であり，それが集まって場面やより大きな意味的まとまりを形作る．部分間は，時間的・因果的・意味的等の関係によって相互に結び付けられる．同様に，ストーリー世界における部分の最小単位は一つの実体（人・物・場所）であり，複数の実体がまとまってより大きな構造的要素（例えば組織）を成す．また，実体やそのまとまりどうしは様々な関係で結び付けられる（例えば人間関係）．なお，ストーリーやストーリ世界の構造には，全体の構造・意味がそれを構成する部分によって形作られるのと同時に，部分の構造・意味が全体的な文脈によって規定されるというように，全体と部分の間に相互依存的な関係があると考える必要があるだろう．

(2) 間表象的構造（ストーリー外の表象との関係）

文学理論の用語として，「間テクスト性」[Kristeva 69]という，あらゆるテクストが常に他のテクストとの関係のもとに成り立つという考え方があるが，これと同様の性質が心的表象としてのストーリーにも当てはまると考える．間表象的構造は，あるストーリーが，他のストーリーやその他の心的要素（スキーマ，概念，心像）との関係のもとに成り立つという性質に対応する．この性質は，2.5 節で述べた，新しいストーリーが常に既存の表象をもとにして構成されるという考え方にも対応している．

3.2 生きた表象による創発的なストーリー生成

以上のような見方をすると，ストーリーの構造は非常に複雑である．このような複雑性を回避するような問題設定を探るというアプローチもあるかも知れないが，この複雑性は恐らく，物語の認知の本質として重要である．このような複雑な構造を生成するのは，従来の中央制御的な記号処理（例えばストーリーグラマーのような文法型知識，プランニング，事例ベース推論）の延長では難しいように思われる．そこで本研究では，表象を外から操作するという形ではなく，表象の内的な力によって，創発的にストーリーが生成するような仕組みを探ってみることにする．

ストーリーの創発といっても，登場人物達の行動のシミュレーションのような類の問題設定ではない．本研究が考えるのは，ストーリーの構造における各水準・各部分が，自らの構造を作り出そうとする力を持ち，それらの相互作用を通じて，全体としてのストーリーが創発するような仕組みである．（ストーリーの生成を全面的に内発的な動きに委ねると荒唐無稽なストーリーが生成しそうなので，生成を外側から制御しようとする力（意識的な思考のようなもの）もあると考えておく．）

3.3 表象の力

以上の考えを実現する具体的な方法はまだ無いが，この問題の予備的な整理を行ってみたい．心的表象としてのストーリーが，創発的に構造を作り出す仕組みを検討するに当って，次のような考え方をする．3.1 節で述べたように，ストーリーは，複数の部分が集まってより大きな部分を成すという，階層的な構造を持つものと仮定する．この階層構造における各層の各部分（事象，場面，あるいはより大きなまとまり）が，自律性を備えた一つの表象となる．各部分は，自らの内部構造を規定するトップダウン的な力を持つ．同時に，部分の内部では，そこに寄せ集められる複数の表象群が相互に結び付いていくことによって，ボトムアップ的に構造を作り出していく．このボトムアップ的に形成された内部構造によって，当の部分の性質（トップダウン的な力の性質）が変化することや，新しい部分（上位構造）を形成したりすることもある．以上のような仕組みによって，各部分の内部で生じるトップダウン的な力とボトムアップ的な力との双方向的な動きを通じて，ストーリー全体の構造が生成する．

以上の概念を一つの例を用いて説明する．ストーリー中のある部分が，「太郎が花子に復讐をする」という意味を持つとする．この部分の内部構造を，上記の意味に対応するような内容に大筋で規定するのが，トップダウン的な力である．一方，この大筋を具体化するためのアイディアないし資源となるような表象が，

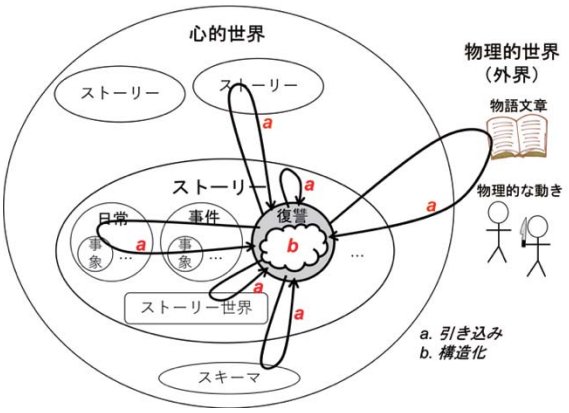


図 4 表象による引き込み-構造化

心的世界の各所から寄せ集められ，それらが相互に結び付くことによって，この部分のより詳しい内容（どのような形で復讐するのか）を形成する．これがボトムアップ的な力である．このボトムアップな構造化を通じて，例えば「復讐」が「太郎が花子に恋をする」という意味に変化するような形で，部分（が持つトップダウン的な力）の性質が変化する場合もある．

(1) 表象の力の基本分類について

ストーリーにおける一つ一つの部分が自らの構造を作り出そうとする力を，「引き込み」と「構造化」の 2 種類に分ける．引き込みは，部分が，自らの構造を作るための資源（表象）を，ストーリー内外から引き込む働きである．構造化は，部分の構造が，内部の表象どうしの相互作用を通じて，作り出される働きである．

図 4 にこれらの概念を表す模式図を示す．この図は，一つのストーリーが生成する過程における，一つの部分（復讐）に焦点を合わせている．

(2) 引き込み-構造化の原理について

引き込みは，部分が自らの内部構造を作るための資源（表象）を各所から寄せ集める働きであるため，引き込みと構造化は密接に関連し合う．これらの主要な原理として考えられるものを以下に挙げる．

アナロジー: 表象間の構造的な類似性に基づく引き込み-構造化．アナロジーは，表面的に異なる事物（表象）どうしの間に，特に関係構造上の類似性を見出す力であり，人間の認知の本質の一つとされる[Gentner 83; Holyoak 95]．新奇な事物（表象・ターゲット）を，構造的に類似する既知の事物（表象・ベース）に対応付けて構成する仕組みとして，物語の認知においても重要な役割を担うと考えられる[Schank 82]．事例ベース推論[Riesbeck 89]も，アナロジーに基づく方法の一つとして位置付けられる．本研究の枠組みに当てはめると，アナロジーは，ストーリー（の部分）が生成する際に，構造的な類似性に基づいて別のストーリー（の部分）の引き込みと構造化を行う原理となる．また，アナロジーは，複数の表象間の構造的な共通性に基づく一般化の基礎にもなる．

空所・予期: 物事を筋立てるために欠如している情報や，後続する情報に対する予期を充足しようとする力による引き込み-構造化．当該部分が持つトップダウン的な構造化の力，スキーマ，または他のストーリーとの類似性に基づいて生じる．例えば，スクリプト[Schank 77, 82]のようなスキーマに基づいて話の行間を補うことや，事例（過去の経験のストーリー）に基づく予期によって知覚が駆動されることに対応する．

背景・文脈: 当該部分に含まれる各種の存在者・物に関する背景的・文脈的な情報（例えばある場面における人物の属性情報やその場面に至るまでの経緯）の編集的な引き込みと，それに基づく構造化の限定や調整．

抽象: 抽象も知能の基本的な性質の一つである[Zucker 03]．部分の内部構造から上位構造としての部分を形成する，あるいは部分の性質を変化させる，ボトムアップ的な構造化に相当する．これに伴い，内部構造に対するトップダウン的な力の性質も変化する．（本研究では，ストーリー内の上位構造を作る働きを抽象（abstraction），複数のストーリーに通じる一般的な構造であるスキーマを形成することを一般化（generalization）と呼ぶ．）

その他: 表象間の類似性や関連性に基づく引き込みや，表象の分解・合成・再結合・調整等による構造化．

4. むすび

本稿では，物語を生成することが人間的な知能の本質の一つであるという考えのもとに，心的水準において，世界の表象としてのストーリーが創発的に生成する仕組みに関する概念的な整理を行った．この考えを実現する具体的な方法にはまだ至っていないが，今後取り組むべき主要な問題は幾つか見出された．ストーリーの創発的な構造化の原理となる，アナロジー，空所・予期，背景・文脈，抽象等の仕組みを，心的世界という背景のもとに，全体として体系化してくことが，ここでの主な課題である．

謝辞

本研究の一部は電気通信普及財団研究調査助成の支援を受けた．

参考文献

[秋元 16] 秋元 泰介, 小方 孝: 共創型物語生成システムに向けて, 人工知能学会論文誌, Vol. 31, No. 6, AI30-O_1-8 (2016)

[Akimoto 17] Akimoto, T.: Computational modeling of narrative structure: A hierarchical graph model for multidimensional narrative structure, *International Journal of Computational Linguistics Research*, Vol. 8, No. 3, pp. 92-108 (2017)

[Akimoto 18] Akimoto, T.: Narratives of an artificial agent: Mental world and narrating, Ogata, T. and Asakawa, S. Eds., *Content Generation Through Narrative Communication and Simulation*, pp. 241-264, IGI Global (2018)

[Genette 72] Genette, G.: *Discours du Récit, Essai de Méthode, Figures III*, Seuil (1972) (花輪 光, 和泉 涼一 訳, 物語のディスクール, 水声社 (1985))

[Gentner 83] Gentner, D.: Structural-mapping: A theoretical framework for analogy, *Cognitive Science*, Vol. 7, pp. 155-170 (1983)

[Holyoak 95] Holyoak, K. J. and Thagard, P.: *Mental Leaps: Analogy in Creative Thought*, MIT Press (1995) (鈴木 宏昭, 河原 哲雄 監訳: アナロジーの力—認知科学の新しい探求, 新曜社 (1998))

[Kristeva 69] Kristeva, J.: 記号の解体学—セメイオチケ 1, 原田 邦夫 訳, せりか書房 (1983) (原著 1969)

[Piaget 48] Piaget, J.: *La Naissance de L'intelligence chez L'enfant, 2e ed.*, Delachaux et Niestlé (1948) (谷村 覚, 浜田 寿美男 訳: 知能の誕生, ミネルヴァ書房 (1978))

[Prince 03] Prince, G.: *A Dictionary of Narratology, Revised Edition*, University of Nebraska Press (2003) (遠藤 健一 訳, 改訂 物語論辞典, 松柏社 (2015))

[Riesbeck 89] Riesbeck, C. K. and Schank, R. C.: *Inside Case-Based Reasoning*, Lawrence Erlbaum (1989)

[Schank 77] Schank. R. C. and Abelson, R. P.: *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum (1977)

[Schank 82] Schank, R. C.: *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge University Press (1982) (黒川 利明, 黒川 容子 訳, ダイナミック・メモリ—認知科学的アプローチ, 近代科学社 (1988))

[Schank 90] Schank, R. C.: *Tell Me a Story: Narrative and Intelligence*, Northwestern University Press (1990) (長尾 確, 長尾 加寿恵 訳, 人はなぜ話すのか, 白揚社 (1996))

[Zucker 03] Zucker, J.-D.: A grounded theory of abstraction in artificial intelligence, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1435), pp. 1293-1309 (2003)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

AGI開発をナビゲートするための能力マップ

A Capability Map for Guiding AGI Developments

大澤正彦 *1*2 Masahiko Osawa
大森隆司 *3*6 Takashi Omori
高橋恒一 *4*6 Koichi Takahashi
荒川直哉 *5*6 Naoya Arakawa
坂井尚行 *5*6 Naoyuki Sakai
上野道彦 *5*6 Michihiko Ueno
今井倫太 *1 Michita Imai
山川宏 *5*6 Hiroshi Yamakawa

*1 慶應義塾大学大学院理工学研究科 Keio University, Graduate School of Science and Technology
*2 日本学術振興会特別研究員 (DC1) Japan Society for the Promotion Science, Research Fellow (DC1)
*3 玉川大学 工学部 Tamagawa University, College of Engineering
*4 理化学研究所　生命機能科学研究センター RIKEN Center for Biosystems Dynamics Research
*5 ドワンゴ人工知能研究所 Dwango Artificial Intelligence Laboratory
*6 全脳アーキテクチャ・イニシアティブ The Whole Brain Architecture Initiative

本研究では，人レベル AGI の開発を効率的に行うためのガイドとしての能力マップを新たに提案する．能力マップは，人レベル AGI に必要とされる能力を階層的に整理し研究の現状を客観的に評価することで，重点的な研究開発が必要な能力を明確化することができる．さらに，能力マップにおいて階層化した個々の能力に対して，それを実現する内部実装，評価する工学的タスク，そして対応する脳領域をそれぞれ割り当てる．特に能力に脳領域を結びつけることは，同じ脳領域に対応づけられた複数の能力を統一的なメカニズムで実現する可能性を示唆し，AGI において重要な汎用性の向上を促進させると期待する．能力マップを AGI 研究開発コミュニティの開発サイクルの中心として位置付けることで，研究活性化が可能と考えられる．

1. はじめに

深層学習の成功をきっかけに，人間レベルの汎用人工知能 (AGI) の実現への期待が高まっている．人間レベル AGI が実現すれば，多くの産業分野で発展を加速できると期待される．一方で，人間レベル AGI は汎用的であるがゆえに，それを実現するために工学的に実現する必要がある能力 (認知機能) は多岐にわたり，その全体像を把握することは容易ではない．ここで，能力とは，工学的に実現することが期待されるあらゆる粒度の実行機能とする．

したがって今後は，開発済みの AI 技術でカバーできる部分とできない部分を明確化し，体系的にまとめながら技術開発を効率的に進めていく必要がある．

類似のものとして，DARPA が提供している認知機能のリストがある [DARPA 05]．しかし，リスト化だけでは直近の課題の明確化はむずかしい．実装する必要がある能力 (認知機能) を要素に分解・階層化して，研究の現状と結び付けて提供することが重要であろう．

そこで本研究では，人間レベル AGI 実現に向けて実装すべき能力を分解・階層化した能力マップの構築を提案する．能力マップを作成する主な目的は以下の 4 つである．

- 人間レベルの AGI の全体像を踏まえた，開発に必要な構成要素の洗い出し
- 直近で着手可能な，あるいは着手すべき技術課題の可視化
- 現在の研究進捗状況や，過去からの技術発展からの差分を定量化
- 技術的な現状や見通しを踏まえた，ロードマップの材料として利用

能力マップはノードとエッジからなるグラフとして表現される．1 つのノードは，能力，内部実装，タスクリスト，対応する脳領域がセットになっており，エッジは能力の階層関係を表現する．能力と内部実装が対応づくことで，工学的な手法と AGI に組み込まれた場合の位置付けを示すことができる．マップを構成する個々の能力は，それを含むエージェントが解くべき工学的タスクと対応づけることで，能力の実現度を定量化することができる．そして，各能力に主に寄与する脳領域や神経回路を対応づけることで，同一のメカニズムでクリア可能な，あるいは関係があると考えられるタスク群の可能性を絞ることや，モジュールの再利用可能性が示唆できる．

連絡先: 大澤正彦，慶大/学振 (DC1)，〒 223-8522 神奈川県横浜市港北区日吉 3-14-1，045-566-1749，mosawa@ailab.ics.keio.ac.jp

2. 能力マップ

2.1 能力マップの概要

提案する能力マップとは，人間レベル AGI の構築を戦略的に行うために，AGI 開発コミュニティで知識共有するためのツールである．図 2 にナビゲーション能力を最上位ノードにおいた場合の能力マップの一例を示す．能力マップは知的な能力をノードとし，それらの依存関係をエッジとした有向グラフで表現される．

能力マップは能力の他に 3 つの要素を 1 組にしてノードとする．

内部実装 対応する能力を実現するための工学的なメカニズム．

タスクリスト 達成することが求められる工学的タスクのリスト．タスクリストを能力と対応づけることで，その能力がどの程度達成されたかを定量的に評価することができる．さらにある能力の要素への分解をタスクから裏打ちすることで，能力の階層性の妥当性を向上させる．

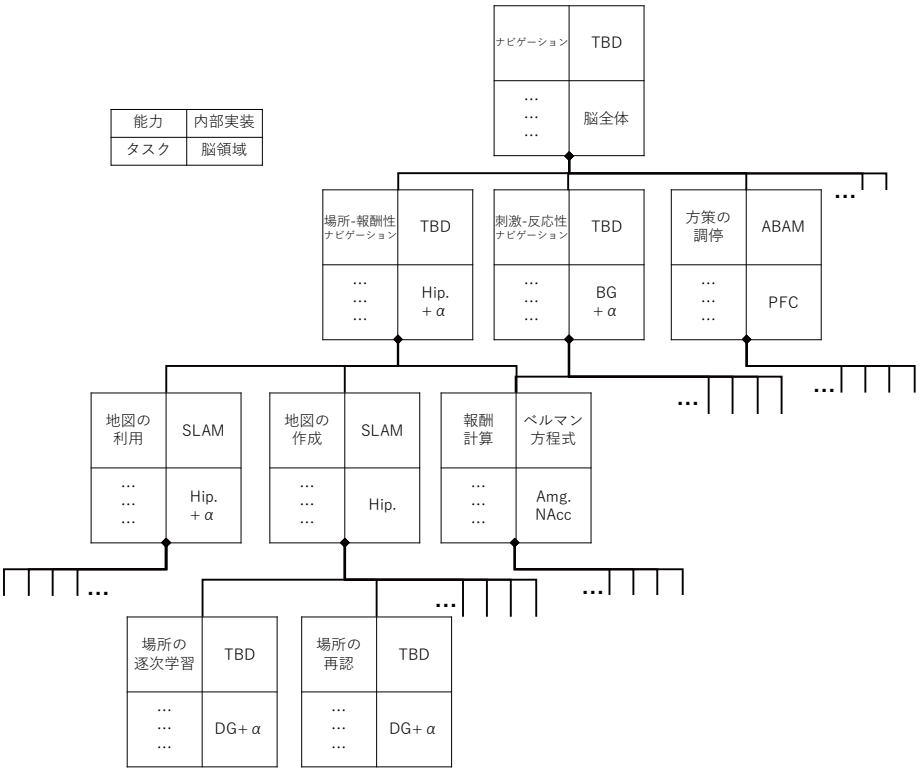


図 1: 能力マップの例．最上位のノードにナビゲーションをおいて構築した例．1 つのノードは複数の上位もしくは下位ノードを持ちうる．内部実装，タスクリスト，脳領域はいずれも完全に埋めることは困難で，これら 3 つを埋めることで AGI 研究が発展すると考えられる．

対応する脳領域 能力を実現するために中心的な役割を担っている脳領域や特定の神経回路を対応づけることで，最終的に目指す人工知能の汎用性を向上させる．詳細は 2.3 で説明する．

2.2 能力マップの利用例

先に述べたように，能力マップの利用は人間レベルの AGI を長期的に研究開発していく一助となると考えている．本節では現在想定している具体的な利用方法について列挙する．

タスク設定とタスク解決のサイクル

これまでの AI 研究では，タスクを設定して定式化することは新たな機能の実現に寄与して来た．能力マップにおいては 1 つの能力に対して 1 つ以上のタスクを割り当てるが，適切なタスク設定も AGI 開発コミュニティの中で提案・精緻化されることを想定している．能力マップ上でタスク設定とタスク解決が同時に取り組まれながら，能力マップ自体も精緻化されていくというサイクルを実現することで，人工知能の未解決問題に取り組んでいけると考えられる．

神経科学的な知見の体系的整理

各能力は，それ自体またはその下位階層の全ての能力に神経科学的な対応づけがされることを目指す．それにより，神経科学的知見と認知機能の対応づけに寄与できる可能性がある．さらに，認知機能として広く認められながらも，その神経基盤が明らかでない機能を明確化することにも貢献しよう．

知能研究に体系性を与える参考資料として

人間レベル AGI を開発する目的で開発した能力マップが多くの関係者の合意を取れるなら，知能研究に体系性を与える参考資料となろう．特に，タスクリストを用いて能力の体系性の裏付けをするという優位性がある．

研究の活性化

能力マップを利用することで，研究があまり行われていない領域を明確化することがで能力マップを分析すると，研究が進んでいない，あるいは少ない知能の領域を明確化できよう．その場合には，活性化が必要と考えられる領域に対して研究プロジェクトを発足させることで，活性化させられる可能性がある．また，能力を詳細に分割してタスクを定義することで，解決するべき要件が明確化できるため，多くの人材が AGI 開発コミュニティに参画しやすくなると期待する．

ロードマップの作成

能力マップに工数推定とリソース計画を組み込むなら，ロードマップとしても利用することができる．それは，技術的な課題を体系的にまとめたロードマップとなることから，抽象的な予測から作るロードマップよりも高い信頼性があるであろう．

2.3 脳型制約による汎用性の向上

能力マップでは全ての能力が少なくとも間接的に脳器官に接地していることを目標とする．能力を脳領域と対応づけることは，能力マップが表現する知的システムの汎用性を 2 つの方法で向上させるきっかけを作ることができる．

能力の一本化

別の能力のラベルがはられたノードが、同一の脳領域を割り当てられていた場合に，実は同じ能力について記述している可能性を示唆する．これによりグラフのノードが一本化されるため，グラフが小さくなる．能力マップ自体が洗練されることで、汎用性に寄与する．

内部実装の一本化

別の内部実装によって実現されていた能力ノードが，同一の脳領域を割り当てられていた場合に，実は同じ内部実装で実

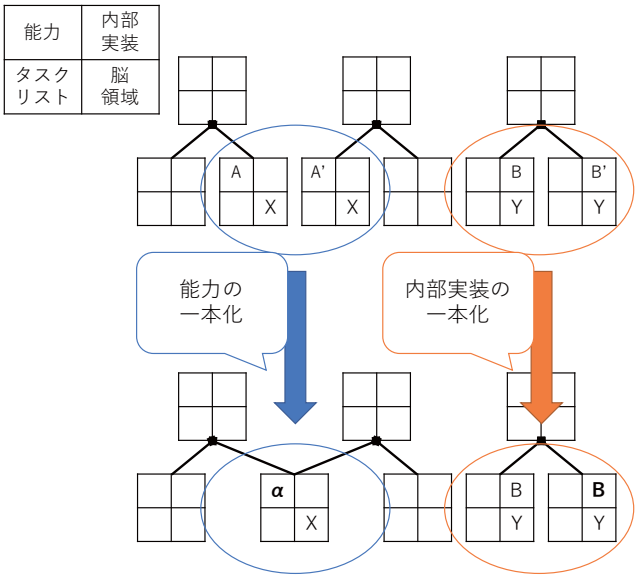


図 2: 脳型制約による汎用性の向上

現可能もしくは実現すべきである可能性を示唆する．能力マップは能力の階層構造によってグラフが形成されるため，内部実装の変更はグラフ構造を直接変化させない．一方でグラフ全体を実現するために利用される技術の数が削減されるため，汎用性・再利用性に寄与する．

2.4 全脳アーキテクチャプロジェクトにおける能力マップの位置付け

全脳アーキテクチャプロジェクト [Yamakawa 16] では，脳全体を AI 技術モジュールの組み合わせとして捉え，モジュールを組み合わせることで人間レベル AGI の実現を目指している．またプロジェクトで参考にする全脳のモデルは最初から人間を想定するよりも，人間と脳のアーキテクチャの共通性が高くありながらも，構造がよりシンプルで神経科学分野における知見の多い動物をマイルストーンとするべきだと考える．具体的には，参考にする脳構造をげっ歯類から，類人猿，そして人間へと推移していく予定である．その過程で動物間に共通する，あるいは異なる認知機能を明確化させることで，げっ歯類から類人猿，類人猿から人間へとプロジェクトの対象をシフトしていく際に，円滑な技術の再利用ができるようにする必要がある．

本研究で提案している能力マップは，段階的に扱う能力を拡張していく全脳アーキテクチャプロジェクトと相性のいい設計になっている．能力マップは目標とする全体像を還元論的に分解して記述する．げっ歯類レベルの認知機能として体系化された能力マップは，類人猿や人間に拡張した際にほとんど同一の階層構造のまま転用できる可能性が高い．

3. おわりに

本研究では，AGI 開発をナビゲートするための能力マップを提案した．能力マップは人間レベル AGI に必要な能力が階層的に記述されている．特に汎用性を向上させるために脳との対応づけを行うことで，一本化すべきノードの発見や，共通の内部実装で解決すべきタスク群の候補を選出することができると考えられる．

AGI 開発は今後，長期的なスパンでの研究開発が必要な領域であり，効率化を図るための方法論は今後さらに需要が大きくなるであろう．今後のアクションプランとしては，提案した能力マップをさらに精緻化して，実際に AGI 研究開発コミュニティにおけるサイクルのかなめとして運用して行くことを目指す．特に能力マップは多くの立場から合意を得られることが重要であるため，ワークショップ形式で議論を進めることを次なるアクションプランとして想定している．

参考文献

[DARPA 05] DARPA, ”BICA, Biologically-inspired cognitive architectures, Proposer Information Pamphlet (PIP) for Broad Agency Announcement 05—18.,” Arlington, VA: Defense Advanced Research Projects Agency, Information Processing Technology Office, (2005)

[Yamakawa 16] Hiroshi Yamakawa, Masahiko Osawa and Yutaka Matsuo: Whole Brain Architecture Approach Is a Feasible Way Toward an Artificial General Intelligence, The 23rd International Conference on Neural Information Processing, pp. 275-281, Kyoto, Japan, (2016).

[2M2] [General Session] 13. AI Application

座長:大西 貴士（NEC/産総研）

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room M (2F Amethyst Hall Hoo)

[2M2-01] Utilizing deception information for dialog management of doctor-patient conversations.

○Tung The Nguyen[1], Koichiro Yoshino[1], Sakriani Sakti[1], Satoshi Nakamura[1] （1. Graduate School of Information Science, Nara Institute of Science and Technology）

1:20 PM - 1:40 PM

[2M2-02] Cost reduction using stochastic programming in supply chain management

○TAKESHI KONNO[1], Ken Kobayashi[1], Shuji Awai[2], Go Yamamoto[2] （1. FUJITSU LABORATORIES LTD., 2. FUJITSU SOCIAL SCIENCE LABORATORY LTD.）

1:40 PM - 2:00 PM

[2M2-03] AI utilization for membrane structure design and manufacturing

○Kazuo Yokobori[1], Tomo Miura[1] （1. Taiyo Kogyo Corporation）

2:00 PM - 2:20 PM

[2M2-04] Optimization for the Migration Units in Information Systems based on CRUD Relations with Semi-Supervised Graph Neural Networks

○Kugatsu Sadamitsu[1] （1. Future Corporation）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Utilizing deception information for dialog management of doctor-patient conversations.

Nguyen The Tung*[1], Koichiro Yoshino, Sakriani Sakti, Satoshi Nakamura

Graduate School of Information Science, Nara Institute of Science and Technology

Almost all of existing negotiation systems assume that their interlocutors (the user) are telling the truth. However, in negotiations, participants can tell lies to earn a profit. In this research, we proposed a negotiation dialog management system that detects user's lies and designed a dialog behavior on how should the system react with. As a typical case, we built a dialog model of doctor-patient conversation on living habits domain. We showed that we can use partially observable Markov decision process (POMDP) to model this conversation and use reinforcement learning to train the system's policy.

1. Introduction

Recently, negotiation dialog has attracted the attention of many researchers around the world. Other than improving the efficiency of dialog management, the focus of these studies have switched from passive role systems (eg. restaurant, tourist information provider) to become more actively systems that can influence the user's decision (persuasive technology) [5, 10]. Most of current researches mainly deal with cooperative dialog, in which the system and the user work together to reach a balanced point . However, there are situations that the mutual goal cannot be reached (non-cooperative dialog) [11, 3].

Existing works [3] did not cover the situation in which participants use "false promises" tactics - which can directly impact the outcome of negotiation. For example, someone pretends to give up unhealthy custom (eg. smoking/drinking) after receiving advices from friends or family but he actually does not have intention to follow the advice. [12] put a learning agent that can tell lies against an adversary who consider lies with contradiction in simulation.

In this paper, we focus on a typical example of negotiation, the doctor-patient conversation on living habits domain. In this type of dialog, the patient has their own perspective and opinions about their health condition. The doctor needs to consider these opinions when making a treatment plan (recommend a new habit) and negotiates with users to reach a plan that satisfy both user's demand and requirements of the treatment. Patients sometimes tell lies because they do not want to change their habit, thus considering deceptive information will improve the negotiation strategy of the doctor.

2. Scenario and Modeling

2.1 Dialog scenario

This work considers a dialog scenario between a system (doctor) and a user (patient). They discuss about user's living habit, which can be about: sleeping, food, working/studying, exercise, social media usage and leisure activities. The system tries to convince the user that they need to change to a more healthy living habit. The system persuades the user by giving them information about the new habit (system's recommendation), health benefits of the new habit and negative effects of user's current habit. This action is denoted as Framing in this research. On the other hand, the user wants to continue the current habit and gives reasons to show that it is too difficult for him to change. The system behaves cooperatively; if user's reason is honest, system will give an easier recommendation.

To make the conversation simpler, only the system can propose recommendations, the user cannot suggest what habit they should change to. However, the users are allowed to use dishonest reasons to make the system offer an easier recommendation. The user can also pretend to accept the system's offer while they actually do not intend to change their current habit.

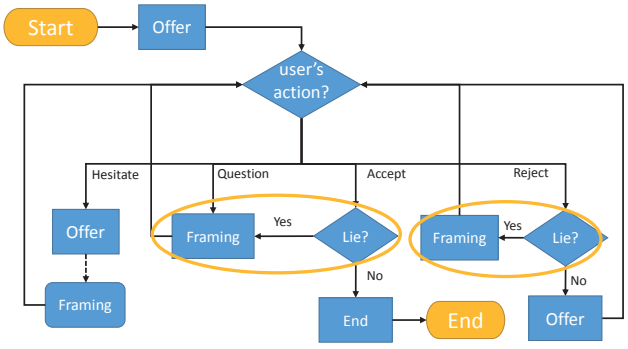


Figure 1: *Flow chart of the conversation.*

Figure 1 describes the proposed dialog behavior, which considers deceptions of the user. In this flowchart, rectangles indicate system actions. The set of dialog acts for the system includes:

Offer: the system offers a recommendation to the user, which means the system suggests the user should change to a new habit.

Framing: the system provides information about the new living habit, the necessity and health benefits of the habit and consequences if the user keep the current habit.

End: the system ends the conversation.

Contact: Nara Institute of Science and Technology;
8916-5 Takayama, Ikoma, Nara;
{nguyen.tung.np5,koichiro,ssakti,s-nakamura}@is.naist.jp

For persuasive dialog, framing technique was shown to be a very effective way to convince users to accept the system's goal ([5]). On the other hand, the user can react with different actions as described below:

Accept: the user accepts the system's recommendation (offer). The user can pretend that s/he accept the system's offer.

Reject: the user rejects the recommendation given by the system and gives reason why s/he rejects it. In order to make the system to give an easier habit, the user can reject with a deceptive reason.

Hesitate: the user says he/she is unsure about whether to accept the offer or not.

Question: the user asks the system about the benefits of new habit or asks the system to give a clear explanation it.

According to a study by [7], when the patient is lying, the doctor should tell the patient about the necessity and benefits of the treatment plan and consequences if patient refuses to follow it. Applying to the "living habit" scenario, the most logical reaction when the user is telling lie (uses fake reasons or pretend to agree) is Framing. In contrast, a conventional negotiation system that does not consider user's deception always offer a new recommendation when the user rejects and end the conversation when the user agrees regardless whether they are telling the truth or not.

2.2 Policy Management using POMDP

To find the best strategy of dialog system against the user deception, considering errors of dialog act classification and deception detection for user utterance is necessary, because these model do not have 100% accuracy. Partially observable Markov decision process (POMDP) is widely used to learn the best strategy of dialog systems for such error containing cases [13].

The equation for updating belief state of a POMDP can be written as:

$$b^{t+1} \propto P(o^{t+1}|s^{t+1}) \sum_{s_i} P(s^{t+1}|s^t, \hat{a}^t) b^t \tag{1}$$

Apart from user's action s^t, the proposed dialog system also uses deception information for dialog management. To solve this problem, we used method in a similar work for user focus by [13]. By extending Equation (1) with deception information of the current turn d^t and next turn d^{t+1}, we have the belief update of the proposed system:

$$b^{t+1}_{s,d} \propto \quad P(o^{t+1}_s, o^{t+1}_d | s^{t+1}, d^{t+1}) \sum_s \sum_d P(s^{t+1}, d^{t+1} | s^t, d^t, \hat{a}^t) b^t_{s,d} \tag{2}$$

With the observation result come from SLU and Deception Detection modules being denoted as o_s and o_d respectively.

In this research, we use Q-learning, a popular method to train the optimal policy π^*. We utilize Grid-based Value Iteration method proposed by [2] . The belief is calculated by:

$$b_{s_i} = \begin{cases} \mu & \text{if } s = o \\ \frac{1-\mu}{|S|-1} & \text{otherwise} \end{cases} \tag{3}$$

μ represent the rounded probability for every 0.1 that the observation comes from SLU and Deception Detection equal to actual user state and deception information.

Dialog state		Rewards		
User DA (s)	d	Offer	Framing	End
Accept	0	-10	-10	+100
	1	-10	+10	-100
Reject	0	+10	+10	-100
	1	-10	+10	-100
Question	0	-10	+10	-100
Hesitate	0	+10	+10	-100

Table 1: Rewards in each turn.

In table 1, we show the reward received for each turn.

2.3 Deception Detection using multi-modal approach

There are various clues that can help us to detect lies, which include lexical modal, acoustic modal, gestures and facial expressions. As a result, multi-modal approach that combines those modalities was proved to be very efficient in detecting deception [8]. Thus, we also utilized multi-modal approach with acoustic and facial features to build the deception detection module.

To extract facial features, we used OpenFace toolkit developed by [1]. From this tool, we were able to extract 14 face AU (Action Unit) regression and 6 AU classification values as well as head position, head direction parameters. All these values are then normalized and discretized into 5 different levels of intensity.

Acoustic features are extracted from audio files using OpenSMILE tool [4]. The acoustic feature template was taken from the work by [6]. From the pitch and loudness values, we calculated maximum (*max*), minimum (*min*), mean (*mean*) and standard deviation (*std*). The duration-related features include: percentage of frame with voice, percentage of frame with lower pitch than previous frame (falling pitch), percentage of frame with higher pitch than previous frame (raising pitch).

The classification model that we used for deception detection is Multi Layer Perceptron with hierarchical structure to combine acoustic and facial features. This method was proposed by [9] for emotion detection task. In particular, for our research, we put facial features directly into the first layer of the network while acoustic features are incorporated at the second hidden layer.

To test the effectiveness of the deception detection module, we used data from recorded conversations between 2 participants. One of them plays the role of a doctor and the other one's role is a patient. In this conversation, the "patient" will try to get a prescription from the "doctor" by telling lies about his health condition. Deception label for each utterance was manually annotated by themselves. The total number of utterances in this data is 146. We took 34 of them (17 honest, 17 deceptive) to use as test data. Results of the experiment are shown in Table 2.

	Features combine	Decision combine	Hierarchical
Precision	61.77%	67.65%	64.71%
Recall	35.29%	41.17%	58.23%

Table 2: Deception detection accuracy.

3. Experiments of dialog management system

3.1 Data

Due to the difficulties of recruiting actual doctors and patients to collect the data, the participants who took part in data collection are students who have good level of English fluency. All participants are working at the same academic environment at the time of data collection.

The data used for the experiments are recorded using the "living habits" dialog scenario in Wizard-of-Oz (WoZ) setup. Each recording session is carried out by 2 participants who play the role of the system and the user respectively. Each session consists of 6 dialogs for each of the living habit topic. The participants who play the role of the user (patient) are given payment as reward for the outcome of the conversation. If they pretend to agree with the system's offer, they will receive lower payment. On the other hand, if they choose to truly agree with the system's offer they will get higher payment with a condition that they will need to change to the new habit for one week. The payment is to create the situation where the user has to choose between an easy activity (continue current habit) with low reward and an activity that is difficult but has higher reward in return (change to new habit) to observe more lies.

There are total of 7 participants who take part in the recordings. 4 of them played the role of the system and 6 played the role of the user (participants who played the role of the user can switch their role to doctor in later recordings.) In the end, the recorded training data are about 3 hours 20 minutes long with a total of 883 utterances. Labels of DA are annotated by one expert and labels of deception are provided by the participant who made the deception.

With test corpus, recordings are done as direct conversations between participants. The recording setup is similar to WoZ scenario but without the help of TTS since the participants are now talking directly with each other. This type of evaluation data contains 936 utterances and about 2 hours 35 minutes in length. Table 3 shows statistics of the collected data.

Data	System DA			User DA				% lie in user's utterances
	End	Framing	Offer	Hesitate	Question	Accept	Reject	
Train	14.43%	43.30%	42.27%	21.69%	3.61%	51.81%	22.89%	18.07%
Test	17.54%	36.26%	46.20%	17.64%	9.86%	19.72%	52.82%	14.08%

Table 3: Deception and dialog acts statistic.

3.2 Results

Dialog policy	Success rate	Avg. offer
Conventional	21.83%	2.472
Proposed policy	**29.82%**	2.447

Table 4: Success rate and average offer per succeeded dialogs.

First, we test the negotiation efficiency of the learned policy by comparing it with a baseline negotiation policy (denoted as Conventional in Table 4) that treats user's utterances as honest. With this baseline policy, the system receives reward similar to Table 1 but only for the case when user is being honest ($d = 0$); because with this policy, deception is not taken into account. We let the system interacted with a simulated user created from test data for 100,000 dialogs. The simulator is created with the same method as described in [13]. The transition probabilities are given by:

$$P(s^{t+1},d^{t+1}|s^t,d^t,\hat{a}^t) = \underbrace{P(s^{t+1}|d^{t+1},s^t,d^t,\hat{a}^t)}_{\textbf{intention model}}\underbrace{P(d^{t+1}|s^t,d^t,\hat{a}^t)}_{\textbf{deception model}} \quad (4)$$

Probabilities for **intention model** and **deception model** are calculated from test data using maximum likelihood.

Performance is evaluated with success rate and average offer per succeeded dialog. Success rate is the percentage of dialogs in which user truly accepts the system's offer. For the second evaluation metric, average offer; since every time the system makes an Offer action, the new habit will be easier but gives less health benefit, thus it is less favorable for the system. Therefore, using fewer offers to successfully persuade the user is better. From the results shown in Table 4, it is clear that our proposed system outperforms the baseline.

Dialog system	DA accuracy	Deception handling
Conventional	68.15%	35.00%
Proposed policy + gold-label deception	**80.45%**	**80.00%**
Proposed policy + predicted deception	**79.32%**	**55.00%**

Table 5: Dialog selection acts accuracy.

Next, we evaluate the performance of the system's dialog acts decision. The accuracy are measured with 2 metrics. DA accuracy refers to the accuracy of the system's chosen dialog acts against reference actions that were chosen by human. Deception handling indicates the accuracy of dialog acts decision when user is lying. From the results of Table 5, we can see that our proposed system outperformed the baseline again.

4. Conclusions

In this paper, we present a negotiation strategy that counter deception in doctor-patient conversation. We also propose a dialog system that utilize this strategy and showed the construction and performance of each module. Experiment results proved that the proposed strategy outperformed normal negotiation strategy significantly, beating it by more than 8% in chance of successful persuasion. DA accuracy experiments indicated that the learned policy achieved a good level of naturalness when compared to human behavior in the same scenario. In future, we would like to combine the modules and conduct experiments of the system with human.

References

[1] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, pages 1–10.

[2] Blai Bonet. 2002. An e-optimal grid-based algorithm for partially observable markov decision processes. In Proc. of the 19th Int. Conf. on Machine Learning (ICML-02).

[3] Ioannis Efstathiou and Oliver Lemon. 2014a. Learning non-cooperative dialogue behaviours. In Prooceedings of the

15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). Association for Computational Linguistics, Philadelphia, PA, U.S.A., pages 60–68. http://www.aclweb.org/anthology/W14-4308.

[4] Florian Eyben, Martin Wollmer, and Bjorn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia. ACM, pages 1459–1462.

[5] Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2014. Reinforcement learning of cooperative persuasive dialogue policies using framing. In COLING. pages 1706–1717.

[6] Julia Hirschberg, Stefan Benus, JasonMBrenier, Frank Enos, Sarah Friedman, Sarah Gilman, Cynthia Girand, Martin Graciarena, Andreas Kathol, Laura Michaelis, et al. 2005. Distinguishing deceptive from non-deceptive speech. In Interspeech. pages 1833–1836.

[7] Karin I Kjellgren, Johan Ahlner, and Roger Säljö. Taking antihypertensive medication controlling or co-operating with patients? International journal of cardiology, 47(3):257–268, 1995.

https://doi.org/10.1023/A:1005719122519.

[8] Verónica Pérez-Rosas, Mohamed Abouelenien, Rada Mihalcea, and Mihai Burzo. 2015a. Deception detection using real-life trial data. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, pages 59–66.

[9] Leimin Tian, Johanna Moore, and Catherine Lai. 2016. Recognizing emotions in spoken dialogue with hierarchically fused acoustic and lexical features. In Spoken Language Technology Workshop (SLT), 2016 IEEE. IEEE, pages 565–572.

[10] Kristian Torning and Harri Oinas-Kukkonen. 2009. Persuasive system design: state of the art and future directions. In Proceedings of the 4th international conference on persuasive technology. ACM, page 30.

[11] David R Traum. 2008. Computational models of noncooperative dialogue .

[12] Aimilios Vourliotakis, Ioannis Efstathiou, and Verena Rieser. 2014. Detecting deception in noncooperative dialogue: A smarter adversary cannot be fooled that easily. In Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (SemDIAL). Edinburgh, Scotland, U.K., pages 252–254.

[13] Koichiro Yoshino and Tatsuya Kawahara. 2015. Conversational system for information navigation based on pomdp with user focus tracking. Computer Speech & Language 34(1):275–291.

[14] https://www.wsj.com/articles/SB10001424127887323478004578306510461212692

[15] https://www.zocdoc.com/about/news/new-zocdoc-study-reveals-women-are-more-likely-than-men-to-lie-to-doctors/

[16] http://articles.latimes.com/2009/jun/08/health/he-lying8

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

SCM における確率計画法を用いたコスト削減の検証

Cost reduction using stochastic programming in supply chain management

紺野 剛史[*1] 小林 健[*1] 粟井 修司[*2] 山本 剛[*2]
Takeshi Konno Ken Kobayashi Shuji Awai Go Yamamoto

[*1] 富士通研究所
FUJITSU LABORATORIES LTD.

[*2] 富士通ソーシアルサイエンスラボラトリ
FUJITSU SOCIAL SCIENCE LABORATORY LTD.

Production of a given product at a manufacturing site is usually planned on the basis of predicted demand for the product in question. However, actual demand is uncertain, and it is important to take this uncertainty into consideration during production planning. In this paper, we present a solution to problems for production planning in component plants that takes the uncertainty of demand into account, and a verification of this solution using real data.

1. はじめに

製造の現場における製品の生産計画は，その製品の需要予測に基づいて立案されることが多い．しかし需要量は不確実性を伴うものであり，生産計画立案においてはその不確実性を考慮することが望ましい．第 4 次産業革命(インダストリー4.0)に向けて需要のある製品を直ちに必要量だけ生産する変種変量生産へと生産方式が移行する[正田 16]．需要量の激しい変動を考慮して製造コストを削減する生産計画を立案することは今後重要な課題となる．

そこで本稿では需要の不確実性を考慮した生産計問題として在庫費用とバックオーダー費用を最小化する問題に着目し，この問題に対する確率計画法を用いた解法について検討する．そして実データを用いた数値実験により確率計画法の有用性を検証する．

2. 生産計画問題の定式化

この節では，本稿で扱う生産計画問題の定式化について述べる．まず 2.1 節で需要量の不確実性を考慮しないモデル(基本モデル)について述べ，続いて 2.2 節で不確実性を考慮した確率計画問題としての定式化について述べる．

2.1 基本モデルの定式化

ここではT期にわたってI種類の製品の生産計画を決定するモデル(基本モデル)について考える．なお，定式化は久保らの在庫方策決定モデルを参考にした[久保 14]．まず，今回は 1 つの製造ラインで複数品目の製品を生産することを想定し，各期$t \in \mathcal{T} = \{1, \dots, T\}$において，製品全体の生産量の総和の上限を$M_t$とする．工場は在庫を持つことができ，各製品$i \in \mathcal{I} = \{1, \dots, I\}$は，それぞれ製品 1 個/期あたり$h_{it}$の在庫費用がかかる．また，製品$i$の期 tにおける在庫量は需要量 d_{it}だけ減少する．ただし，需要量に対して在庫が不足して品切れが生じた場合，その不足分に対し，製品 1 個/期あたりd_{it}の追加費用(バックオーダー費用)を支払うことで顧客が再び商品が到着するまで待つものとする．本来欠品時はバックオーダーのほか需要が消滅する機会損失も考慮する必要があるが，今回は簡単のため欠品時の需要はすべてバックオーダーされると仮定する．

以上の仮定のもと，計画期間全体にわたる在庫費用とバックオーダー費用を最小化する生産計画を決定するモデルは，t期における製品iの生産量を表す変数をx_{it}，t期における製品iの在庫費用とバックオーダー費用を表す変数をy_{it}とすると，以下のような線形計画問題として定式化できる：

$$\min \quad \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} y_{it} \tag{1}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}} x_{it} \le M_t \quad (t \in \mathcal{T}), \tag{2}$$

$$y_{it} \ge h_{it} \sum_{t'=1}^{t} (x_{it'} - d_{it'}) \quad (t \in \mathcal{T}, i \in \mathcal{I}), \tag{3}$$

$$y_{it} \ge b_{it} \sum_{t'=1}^{t} (d_{it'} - x_{it'}) \quad (t \in \mathcal{T}, i \in \mathcal{I}). \tag{4}$$

ここで式(2)は各期の製品生産量の上限を表す制約である．また式(3), (4)は

$$y_{it} \ge \max\left\{ h_{it} \sum_{t'=1}^{t} (x_{it'} - d_{it'}), b_{it} \sum_{t'=1}^{t} (d_{it'} - x_{it'}) \right\}$$

と等価で，最適解においてy_{it}は製品iのt期における在庫費用とバックオーダー費用どちらか発生した費用に一致する．

2.2 確率計画法としての定式化

実際の現場で基本モデル(1) - (4)を用いて将来の生産計計画を定めることを想定した場合，予測需要量d_{it}のデータとしては過去の実績値から何らかの方法で推定した予測値(あるいは予測分布の平均値など)を代用することになる．しかし，予測需要量d_{it}は不確実性をもつため，実際の現場ではその不確実性も考慮した生産計画を定めることが望ましい．予測需要量の不確実性を考慮して生産計画を定める最も簡便な方法としては，事前に過去の実績値から何らかの方法で各期の各製品の予測需要量の同時分布を推定し，その同時分布から生成した離散的なシナリオデータを用いて所望の計画を求める最適化問題を解くことである．

まず，式(1) - (4)に対して，需要の不確実性を考慮する場合の定式化について考える．シナリオ数を S とし，それぞれのシナリオ$s \in \mathcal{S} = \{1, \dots, S\}$の生成確率を$p^{(s)}$とする．ここでシナリオ$s$における製品$i$の期$t$での需要量を$d_{it}^{(s)}$とし，シナリオ$s$における製品$i$の期$t$での在庫費用とバックオーダー費用を表す変数$y_{it}^{(s)}$を新たに導入すると，与えられたシナリオのもと在庫費用とバッ

連絡先：紺野剛史，konno.takeshi@jp.fujitsu.com

クオーダー費用の期待値を最小化する問題は以下の線形計画問題として定式化できる:

$$\min \quad \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} p^{(s)} y_{it}^{(s)} \tag{5}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{I}} x_{it} \le M_t \quad (t\in\mathcal{T}), \tag{6}$$

$$y_{it}^{(s)} \ge h_{it}\sum_{t'=1}^{t}\left(x_{it'} - d_{it'}^{(s)}\right) \quad (t\in\mathcal{T}, i\in\mathcal{I}, s\in\mathcal{S}), \tag{7}$$

$$y_{it}^{(s)} \ge b_{it}\sum_{t'=1}^{t}\left(d_{it'}^{(s)} - x_{it'}\right) \quad (t\in\mathcal{T}, i\in\mathcal{I}, s\in\mathcal{S}). \tag{8}$$

3. 数値実験(期待値)

この節では, 需要の不確実性を考慮しないモデル(1) - (4)と不確実性を考慮したモデル(5) - (8)の性能を比較するため数値実験を行う.

3.1 ケーススタディ

図 1 に数値実験を対象とするサプライチェインネットワークを示す. 本稿では, 部品工場と部品工場の貸倉庫を対象とする. 部品工場は見込み生産をしており, 組立工場に納品する部品を作っている. 部品工場は倉庫を持たないため, 完成品は毎日, 組立工場近くの貸倉庫に運ばれ保管される. 組立工場から発注があった分だけ, 貸倉庫にある在庫を納品する. 部品工場は見込み生産のため, 需要予測が外れると過剰在庫や欠品が発生する.

数値実験では, 部品工場の在庫費用とバックオーダー費用を最小化するための生産計画について評価を行う.

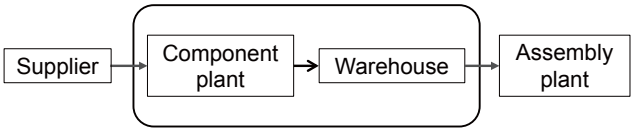


図 1:サプライチェインネットワーク

3.2 計算条件

以下の設定で作成した問題を解いた.

- 製品数: $I = 10$
- 計画期間(日):$T = 5$
- 各シナリオ数:$S = 100, 1000, 10000$
- シナリオの生成確率: $p^{(s)} = 1/S\ (s\in\mathcal{S})$
- 各製品の需要量:
  - 従来手法(基本モデル):実績データから各製品iの需要の平均ベクトル$\overline{d_i}$
  - 提案手法(不確実を考慮したモデル):実績データから各製品iの需要の平均ベクトル$\overline{d_i}$と分散共分散行列Σ_iを推定し, 各シナリオ s に対応する多変量正規分布$N(\overline{d_i}, \Sigma_i)$に従う乱数を生成した$d_i^{(s)}$
- 生産量の上限:

$$M_t = \sum_{i\in\mathcal{I}} \frac{\overline{d_i}}{2}$$

製品$i\in\mathcal{I}$の期$t\in\mathcal{T}$の在庫管理費h_{it}およびバックオーダー費用b_{it}は実際の倉庫単価および製品価格を参考にして設定した. 数値実験では線形計画問題の求解に Cbc version 2.9.0 を使用し, 計算機としては CPU: Intel(R) Xeon E3-1280 V2 3.60GHz, RAM: 32GB, OS: Windows 8 の計算機を使用した.

3.3 計算結果

提案手法の求解に使用したシナリオ集合とは異なるシナリオ集合($S = 10000$)を各製品の需要量として用意し, 提案手法と従来手法で得られた生産計画の総費用を比較した. 表 1 に提案手法による期待値の改善率を示す. なお, 本稿では改善率を以下のように定義した.

$$\text{改善率} = \frac{\text{従来手法の総費用} - \text{提案手法の総費用}}{\text{従来手法の総費用}} \times 100 \tag{9}$$

表 1 では各シナリオ数で改善率の平均値は正の値をとり, 不確実性を考慮することによって総費用が削減できていることが分かる. また, シナリオ数を増加することで総費用の最小値は改善したが, 平均値と最大値に変化はみられなかった.

表 1:各シナリオにおける改善率

シナリオ数	100	1000	10000
平均値	9.1	9.1	9.1
最小値	4.1	5.4	5.5
最大値	4.6	4.6	4.6
計算時間	1秒未満	1分未満	約25分

図 2 に提案手法($S = 10000$)と従来手法における総費用のヒストグラムを示す. 提案手法の分布の方が従来手法より期待値が改善していることがわかる.

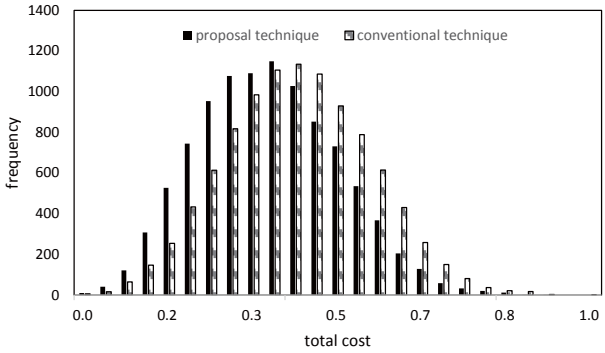


図 2:総費用のヒストグラム

4. 数値実験(実データ)

この節では, 実データにおいて需要の不確実性を考慮しないモデルと不確実性を考慮したモデルの性能を比較するため数値実験を行う.

4.1 総費用の比較

3.3 節で作成した生産計画について, 需要量を 2016/08/01 から 2016/08/05 に組立工場へ実際に検収した数量に置き換えて, 計画期間全体で生じた総費用を計算した. 図 3 に提案手法($S = 10000$)と従来手法で得た生産計画を実データによる評価

結果を示す. 提案手法は従来手法より約 50%総費用が改善する結果が得られた.

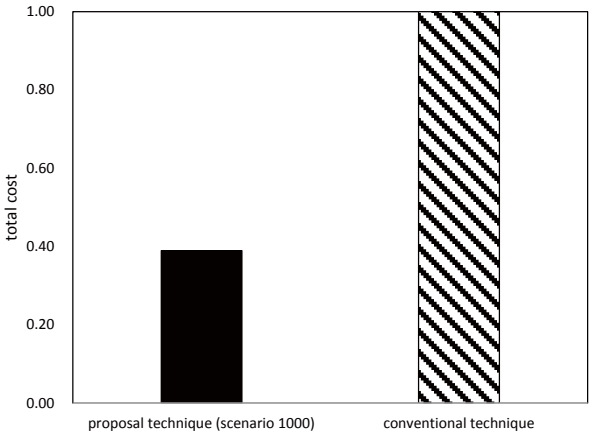


図 3:手法毎の総費用

図 4 に製品毎で生じた総費用の差を示す. 製品 2 から製品 4 の総費用の差が大きくなっている. これらは全商品の中でも需要量が大きい製品になっていることから欠品が発生している. 提案手法では欠品が改善されていることから総費用削減につながっている. 製品 9 については, 需要量は少ないが単価が高い製品となっている. これに関しても提案手法では総費用削減につながっている.

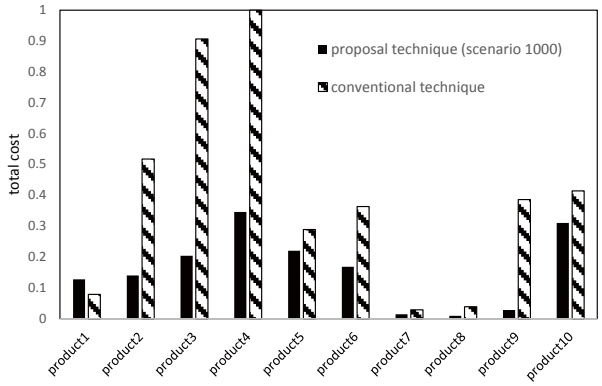


図 4:製品毎の総費用

4.2 生産量の上限を変化させた場合の総費用の比較

部品工場では作業員の調整によって生産量の上限を変化させることが可能である. ここでは, 3 パターンの生産量の上限について従来手法と提案手法で生産計画を立案し, 総費用の改善率を評価した. なお, 需要量は 4.1 節と同じ検収量を用いた.

以下の設定で作成した問題を解いた.

- 製品数: $I = 10$
- 計画期間(日):$T = 5$
- 各シナリオ数:$S = 1000$
- シナリオの生成確率: $p^{(s)} = 1/S\ (s \in \mathcal{S})$
- 生産量の上限:
  - パターン 1:
$$M_t = \sum_{i \in \mathcal{I}} \frac{\overline{d_i}}{2}$$
  - パターン 2:
$$M_t = \sum_{i \in \mathcal{I}} \frac{3 \times \overline{d_i}}{4}$$
  - パターン 3:
$$M_t = \sum_{i \in \mathcal{I}} 1.5 \times \overline{d_i}$$

図 5 に式(9)により求めたパターン毎の総費用の改善率を示す. パターン 1 と 2 では総費用の改善率が小さく, パターン 3 から総費用の改善率が大きい. 生産量の上限が小さい場合は, 生産計画を行う製品の優先順位が限定され差がでないためである. しかし, 上限を大きくすることで様々な組み合わせを考慮できるため総費用の差は大きくなる. なお, 今回の評価では, 生産量の上限を増やすための作業員増加の費用は考慮していないため実際の総費用の差は小さくなることに注意する.

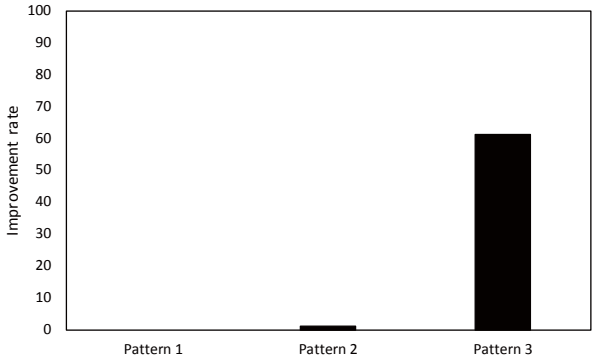


図 5:生産量の上限を変化させた場合の総費用の改善率

5. おわりに

本稿では, 不確実性を考慮して生産計画を定める問題に対し, 確率計画法を用いたアプローチとその有用性の検証を行った. 本稿で扱った生産計画問題は総費用の期待値を最小化するシンプルな問題であるが, 今後はロバスト最適化問題や CVaR などのリスク尺度を用いた生産計画問題などへモデルを拡張させることが考えられる. また本稿で扱った生産計画問題は単一のラインで複数製品を製造する場合の問題であり, 解くべき最適化問題を製品ごとに分解できないという特徴がある. したがってより多品種の生産計画を策定するような大規模問題に対しては, 問題に応じた効率的な解法の設計も必要となる.

参考文献

[正田 16] 正田 聡:IoT社会における製造業の方向性, 経済産業省調べ (2016).

[久保 14] 久保 幹雄, 小林 和博, 武田 朗子, 田中 未来, 村松 正和:サプライ・チェイン最適化における2次錐最適化の応用, オペレーションズ・リサーチ59, Vol.12, pp.739-747 (2014)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

膜構造の設計・製造への AI 利用

AI utilization for membrane structure design and manufacturing

横堀　一雄[*1]
Kazuo Yokobori

三浦 智[*2]
Tomo Miura

[*1] 太陽工業株式会社　設計本部
Taiyo Kogyo Corporation Design Department

[*2] 太陽工業株式会社 情報システム部
Taiyo Kogyo Corporation Information System Department

Membrane structure is the space structure that is composed from clothes, cables and steel frames and so on. In this study, we are developing a management system that is automatically estimation and manufacturing process for membrane structures, then do verification with real job solutions. Though membrane surface has 3D curvature, it is composed from flat (2D) cloth parts which are cut out from the roll of cloth. The roll width is about 1 – 4 meters width, it varies depending on each cloth material. The process of 3D->2D has many complex judgements and it is difficult to achieve good accuracy as estimation system. Therefore AI algorithms are tried to use to estimation system for accuracy improvement.

1. はじめに

膜構造とは東京ドームや東京駅八重洲口グランルーフ（写真1）の屋根のように織物生地やフィルムなどの膜材、ケーブル、鋼材の柱梁で構成される恒久建築物である。



写真 1　東京駅八重洲口グランルーフ(膜屋根部施工：太陽工業)

膜構造は構造力学的に張力で釣合う形状となっており、最小限の部材から構成される軽量な構造物である。1960 年代頃は模型によるデザイン検討により設計され、コンピューターによる解析が可能になって以降はデザイナーと構造解析技術者が解析とデザイン検討を繰り返し、そのデザインと部材構成を決定するという手間のかかる建築物である。 曲面形状の膜屋根の製造段階においても屋根を構成するパーツを1～4m 幅の生地ロール（反物）から切出して製造するため、人手では非常に困難な事から長らく自社では製造工程のデジタル化及び自動化に取り組んできた。 そして近年では ERP（Enterprise Resource Planning：企業資源計画）システムの導入により、設計から製造までのコストがコンピューター上で管理できるようになってきた。
筆者らは、これと設計・製造方法を照らし合わせることで、

（1）設計・製造工程の自動化に役立てる
（2）ERP システムに蓄積された膨大な情報から規則性を見つけ設計・製造の効率化に役立てる

以上を目的に AI 利用の開発を行っている。

2. 設計・製造のプレ段階としての AI 積算

AI 利用という観点から言うと、即設計・製造に利用するという段階には残念ながら至っていない。仮に製造した後で求められた要求事項が適切でなかった場合、その修正による損害は巨額になる事から AI 利用には慎重ならざるを得ない。

しかし、建築物は多くの場合が 1 物件限りのオーダーメイドであるため、「この建築物の建設費用は幾らか？」という積算作業が繰り返し行われ、この積算作業は設計・製造過程を簡易試行する事で行われている。筆者らは、この積算作業に AI 技術の導入を試みた。

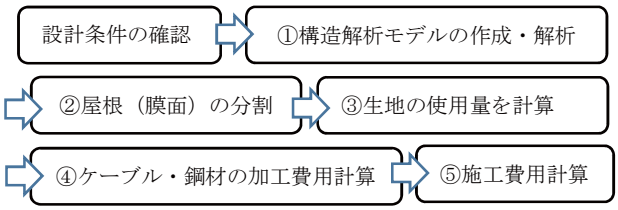


図 1.　フロー　積算作業の流れ

設計条件から自動的にデザインして構造解析モデルの検討①を行う事は現状では難しく将来的な課題と考えている。
しかし、デザインが決まった後の②と③については、

(a)　②でどのように分割すると③での使用量を抑えられるか
(b)　③で生地上にどのようにパーツを配置するか

という問題があり、これらに対してルールなり判断する仕組みが明らかにできれば、②、③を自動化する事ができ、その結果、製造を効率化できる。この問題解決に AI 技術を導入した。

④、⑤については標準仕様と単価が決まっていれば ERP システムと連動させて簡単に計算できる。

AI 利用により積算されたうち何件かは実際に建設され、そのコスト情報（実際の設計と製造結果に基づく）が ERP システム上に記録された。 本研究では、サポートベクトルマシンと確率的最適化手法（文献 1、2）を利用して②、③を行い生地の使用量を計算し、実際に施工された物件の使用生地量との比較を行った。精度の良い方法が見つかれば、実際の積算業務の精度向上に直ぐに役立ち、将来的には実際の設計・製造の自動化にも使用出来るようになる。

連絡先：太陽工業株式会社　〒532-0012 大阪市淀川区木川東 4-8-4　　Fax：06-6306-3168
＊1 e-mail: yk002500@mb.taiyokogyo.co.jp　　*2 e-mail: mt003685@mb.taiyokogyo.co.jp

2.1 使用生地量の予測プログラム(人をまねる)

図 2 は膜屋根モデルの鳥瞰図と構成パーツ状態の一例を示している。このようなモデルに対して単に必要生地量を求めるだけで無く、人による作業手順を調べ、それを再現する予測プログラムを作成し実績(ERP 上の使用量)との比較を行う。

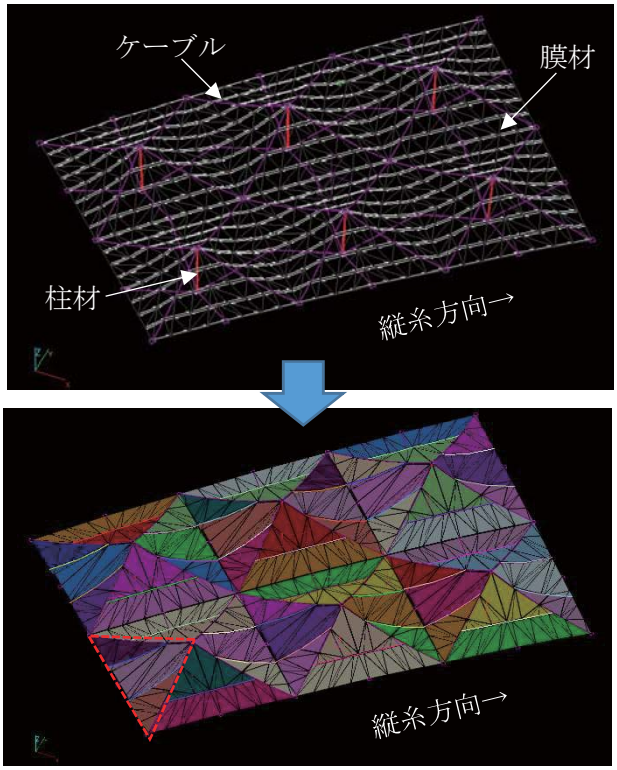


図 2. 生地幅を考慮し膜面を自動分割

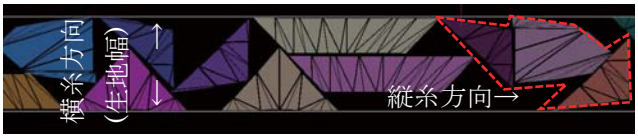
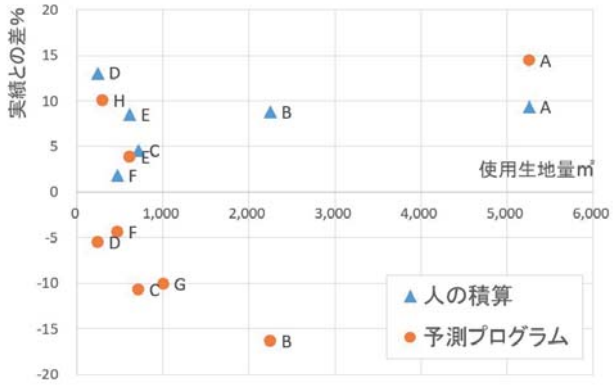


図 3. 生地上に配置する(プログラムによる自動配置)

織物は縦糸(ロール巻方向)と横糸方向で伸びが異なり、製作上これを考慮する必要がある。パーツ分割の際にも生地上でこの方向が一致するように分割・配置をする。図 3 に予測プログラムによる自動配置の結果を示す。図 2、3 中の赤い点線部分は同じ部位を示している。2 分割だと生地幅に収まらない為、3 分割となっている。この様にして求めた使用生地量を実績値、人の積算値(参考)と比較した結果をグラフ 1 に示す。グラフ中の



グラフ 1.　実績と人・予測プログラムの比較

A～H は各物件を示し、実績値との誤差を0%として、それからの差を表示している。膜材の種類により生地幅が異なるため、横軸は“使用生地長×生地幅＝使用生地量㎡”としている。但し、物件 G,H については人が作業した積算結果がない。

2.2 使用生地量の予測プログラム(AI 利用)

グラフ 1 において実績と差が生じている原因を以下で考察する。

(1)人の判断による処理の分岐

グラフ 1 の結果は、予測プログラムが人間に劣る場合があることを示している。この原因を調べた結果の一例を示す。図 2、3 の例について工場の加工データ(自動裁断機をコントロールするデータ)を調べると、図 4 で示すような配置になっていた。

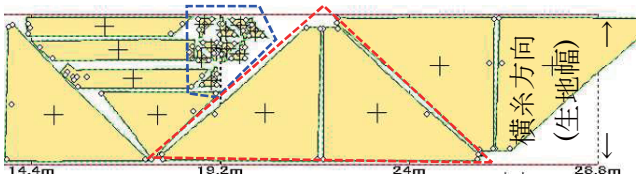
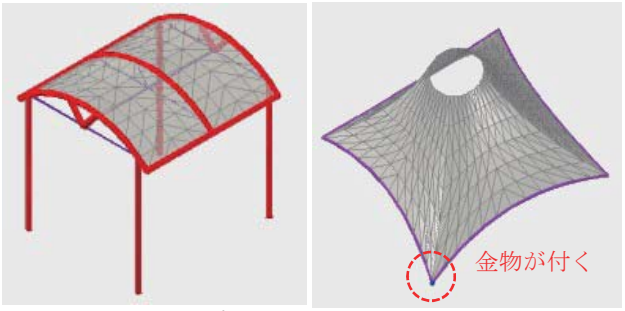


図 4. 工場の加工データ(人による配置)

図 2、3 および図 4 の赤い点線の領域は同じ分割領域を示している。 図 2 の自動分割が 3 パーツに分割しているのに対して、人による分割(図 4)では 2 パーツに分割されている。 よく見ると三角形のパーツの角部が別パーツ(青の点線部分)に分割されている事が分かる。これにより、生地の空隙(歩留まりが良くなりロールの使用長さ)が減り, これが原因である。さらに, 図 5 に示す代表的な膜構造物についても述べる。



フレームタイプ　　サスペンションタイプ

図 5. 膜構造の代表的な形式

フレームタイプではこのような角部だけ別パーツという事はあまり行わないが、サスペンションタイプでは角部に引き込む為の金物(固定の為の金属パーツ)が付くため(図 5 赤い点線の○部分)、デザイン上別パーツにしても違和感が無く、このような分割が行われる場合がある。

これ以外にも、以下図 6 のようにデザイン上、分割ラインの位置が揃う、等分割の方が良いと言った人の判断もある。

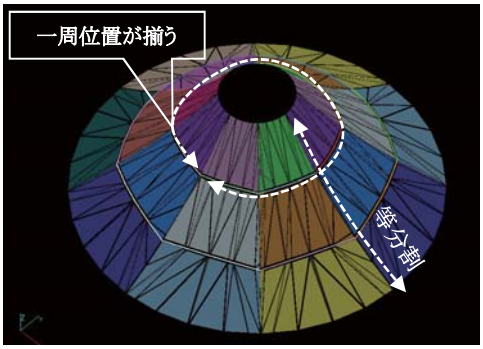


図 6. デザイン上の好ましい判断

このような人の判断等により予測プログラムとの差が生じる。これらの人の判断は、しきい値や判断条件を明確にすることが難しく、条件分岐で処理を切り替えてゆく従来のプログラムでは、曖昧なケースでも妥当と思われる判断をして結果を得ることは難しいように思える。本研究では、これらの判断について、デザイン性の要求度合い、膜材-ケーブル-鋼材の比率を元にサポートベクトルマシンの手法を用いて個々の判断を行い、使用生地長を計算した。

(2)初期段階で未定の項目についての予測

積算段階(顧客の要請に応じて建設費を計算する段階)において未定な項目がある事が原因で実績が変動する。

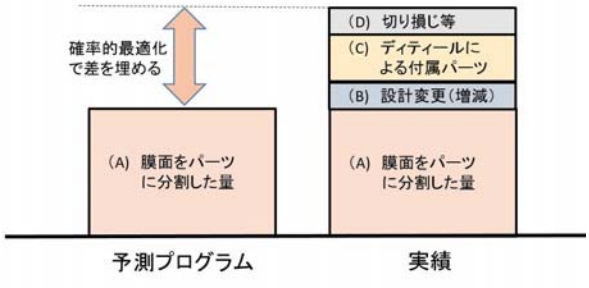


図 7. 積算時に未知の量

実績(工場で実際に使用された量が ERP システムに入力されたもの)の中には、積算時点で未知の量がある(図 7、実績)。以下で未知量(B)～(D)について述べる。

(B)設計変更による増減

実際に建設される際に設計変更が行われ、寸法が変更になる可能性がある。

(C)ディティールの付属パーツ

詳細な設計が進むと膜面を取りつける部位(ディティール)の寸法が決まり、付属する膜材のパーツの量が計算できるが、積算段階では未定な為、予測プログラムでは計算していない。

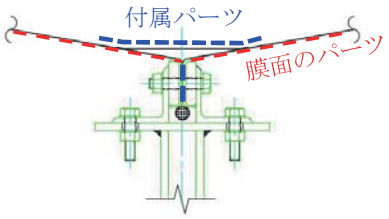


図 8. ディティールにおける付属パーツ

(D)切り損じなど

実際に生地からパーツを切出す段階で、自動裁断装置のトラブルなどで切り損じが生じ、計画より使用量が長くなってしまう事がある。

(B)～(D)の量は積算段階では考慮するのが難しい数量だが、積算結果に考慮したい量である。

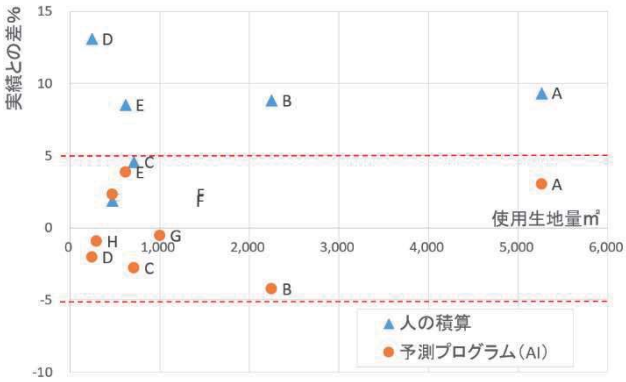
そこで本研究では、確率的最適化手法に基づき、前節(1)で計算した生地使用量を入力x_i、実績の使用長さをy_iとして、損失関数を

$$\sum_{i=1}^{n} l(y_i, f(x_i)) \quad 、 \quad f(x_i) = (1 + (A \log x_i + B))x_i$$

として、過去の実績値からこれが最小となる A、B を求め、$f(x_i)$ を予測値 Y_i としている。式中の$A \log x_i + B$ はグラフ 1 における横軸の生地使用量 x に対する実績との差を示している。

現時点では、既存の一定数サンプルよりA、B を計算しているが、ERP システム上で新たな物件データが得られる毎にA、B を更新して行く逐次学習方式を予定している。

(1)、(2)の処理を行った生地使用量の予測値Y_i を実績値、人の積算値と比較した結果をグラフ 2 に示す。



グラフ 2. 実績と人・予測プログラム(AI)の比較

予測される使用生地量が実績の±5%以内に収まるように改善された。

3. まとめ

本研究では作業者の処理手順をまねるプログラム(明確な判断基準に従う)では難しい、

(1) 作業者の判断基準が明確でない判断
(2) 想定していないケース判断

についてサポートベクトルマシンを利用し判断を行い、

(3) 予測する段階で未知の数量の結果への影響

について、ERP システム上の過去の実績値から確率的最適化手法により推測する事で予測プログラムの精度を上げることが出来た。今後は、

・ERP システム上に蓄積されてゆく実績データにより推測がどのように推移して行くか
・設計の自動化に向けて「(1)作業者の判断基準が明確でない判断」を AI 技術の利用で最適な設計に導けるか

という観点で開発を進めて行きたい。

参考文献

1) 竹内一郎、烏山 昌幸: サポートベクトルマシン (機械学習プロフェッショナルシリーズ)、講談社 、2015 年 8 月 7 日.
2) 鈴木大慈: 確率的最適化(機械学習プロフェッショナルシリーズ)、講談社 、2015 年 8 月 7 日.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

半教師有りグラフニューラルネットワークを用いたCRUD関係に基づくシステム移行単位の最適化

Optimization for the Migration Units in Information Systems based on CRUD Relations with Semi-Supervised Graph Neural Networks

貞光 九月 *1
Kugatsu Sadamitsu

*1フューチャー株式会社 Strategic AI Group
SAIG, Future Corporation

Migrating large-scale information system as a whole is impossible due to its size, thus it is necessary to divide the system into manageable migration units. Typically, optimizing the migration units can only be done by experienced engineers, and takes large amount of time. In this paper, we propose a method finding optimal migration units. To find optimal migration units, we interpret CRUD relations between functions and tables as a graph structure, and adapt semi-supervised graph neural network, Diffusion Convolutional Neural Networks, with a graph embedding method. Furthermore, we unify CRUD relation with file arrangement information and attributes of each node. In experiments, the proposed methods find practical migration units that can be applied to real examples.

1. はじめに

産業の基幹となる大規模な情報システムにおいて，外部環境変化への対応や業務の最適化等の目的から，システム刷新の要求は必然的に生じる [木原他 2015]．大規模な情報システムの刷新を行う際，一度にすべての機能を新システムに切り替えることはリスクが高く非現実的であり，実際の開発では，一部の機能やテーブルの集合を部分的に移行単位として切り出し，段階的に移行するという安全策が採用されることが多い．この時の移行単位の設計は，業務とプログラムに精通する技術者が自身の経験則に基づいて決定することが多く，設計自体に多大な工数を要するという課題や，対象システムに対する知識と経験を持つ一部の技術者にスキルが偏在してしまうという属人性の課題があった．我々はこれらの課題を解決するために，大規模システムの段階移行開発における，移行単位の自動最適化手法を提案する．

本稿では，移行単位最適化の手がかりとして，システム設計時に用いられる，機能とテーブルの操作関係を記述したCRUD表に着目する．CRUDとは，テーブルに対する代表的な操作である，Create, Read, Update, Deleteの頭文字をとったものであり，機能からテーブルへの操作がどのように行われるか記述したものである．CRUD表は人手による移行単位設計の際にも参照されることが多く，自動化においても有益なリソースとして活用できることが見込まれる．また，2. 節で述べる通り，CRUD表は一種のグラフ構造とみなすことができ，そのためシステム移行単位の最適化問題は，CRUD関係を表すグラフにおける各機能ノード及び各テーブルノードの分類問題と定義できる．さらに本稿ではCRUD関係に加え，機能のソースコードやテーブルデータのコンピュータ内のディレクトリ配置情報や，機能やテーブル自体が持つ属性情報を含めた統一的な1つのグラフとして取り扱うことで，より高精度な解析を可能とする．

グラフ中のノード分類手法はこれまでに多くの手法が提案されており，本稿では半教師有りグラフノード分類手法の一種で，近年高い精度が報告されているグラフニューラルネットワークの一種，Diffusion Convolutional Neural Network [Atwood and Towsley 2016] を用い，さらにグラフ埋め込み (graph embedding) 法の一種であるnode2vec [Grover and Leskovec 2016] と Poincaré embedding [Nickel and Kiela 2017] をそれぞれ併用することで，比較的少量の教師データのみを用いた場合でも，高い精度で移行単位の推定が可能となることを示す．

連絡先: 貞光九月，フューチャー株式会社，東京都品川区大崎 1 － 2 － 2 アートヴィレッジ大崎セントラルタワー，k.sadamitsu.ic@future.co.jp

2. 移行単位最適化のためのCRUD関係の利用

2.1 CRUD関係とCRUD表

本節では，移行単位の最適化に必要な手がかりとして，CRUD関係およびCRUD表を用いることを考える．システム開発において，機能とテーブルの関係は極めて重要である．システムの設計や開発の各フェーズにおいて，機能とテーブルの操作関係を漏れなく設計・開発に落とし込むために広く用いられるものがCRUD表である．図1 (a) にCRUD表の例を，図1 (b) にCRUD表に対応するシステム中のCRUD関係を表す．例えば，図1(a)において，機能ノード “Func2” は，テーブルノード “Table2” に対し “Create” の関係を持ち，“Table3” に対し “Read” の関係を持つことを表す．CRUD関係を持つノード同士は比較的強い結びつきを持つと考えられることから，本稿ではこのCRUD表に着目した移行単位の最適化を行う．

本稿で扱うタスク概要についても同図を用いて表す．まず図1 (a) 中，“Unit class1”，“Unit class2” と記した囲いが，人手で解析した移行単位の理想的な分類結果であるとする．本稿で扱う移行単位最適化は，一部の機能またはテーブルに対応する移行単位クラスを教師信号として与え (図 (b) 中 “seed” と記した “Func2” と “Func5”)，その他のノードの移行単位クラスを正しく分類することが目的となる．

さらにCRUD表は，CRUD関係をグラフとみなした場合の隣接行列とみなすことが可能であり，実際に，図1 (b) のCRUD関係に対応するグラフ構造を，図1(a)のCRUD表という隣接行列に等価に変換することができる．システム操作関係をグラフとして解釈することにより，3. 節で述べるように，グラフ解析のための機械学習を直接的に適用できるという利点が生じる．

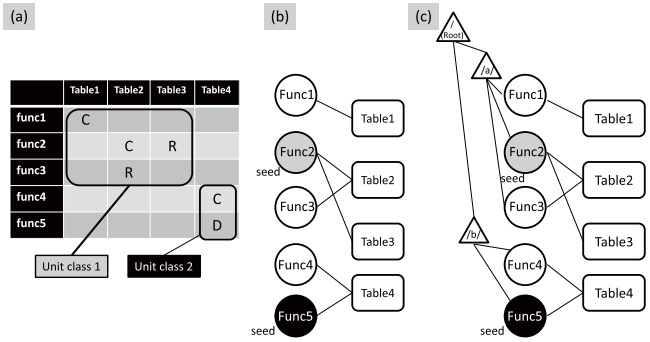


図 1: (a)CRUD 表と人手付与した正解移行単位 (Unit class 1, 2)，(b) CRUD 表に対応する機能とテーブル間の関係図．“seed” と記す “Func2” と “Func5” が教師データで，残る全ノードがクラス推定対象ノード．(c) 機能とテーブル間の関係図に，ファイル配置情報を重畳したもの．Func1 と Func2 が同じクラスに属す可能性を示している．

2.2 ファイル配置情報とノード属性の利用

CRUD 関係は機能とテーブルの関係を表現する上で重要な手がかりとなりえるが，本節では CRUD 関係以外で追加できる手がかり情報を検討する．

1 つ目の手がかりは，各機能のソースコードやテーブルデータのコンピュータ上でのファイル配置情報である．システム開発を行う上で，同じディレクトリに配置される機能やテーブルは，1 つの単位としてまとめて移行する方が適切であるという仮説に基づいて導入する．ファイル配置情報の利用方法にはいくつか方策が考えられるが，本稿では直接 CRUD 関係のグラフ構造に重畳する方法を提案する．具体的には，元となる CRUD 関係を表すグラフ構造に対し，OS のディレクトリに相当する仮想ノードを設け，そのディレクトリ内に配置される機能のソースコード・テーブルデータ，あるいは親・子ディレクトリに対し，CRUD 関係と同様にエッジを張ることとする．実際に重畳した具体例を図 1 の (c) に示す．このような重畳を行うことで，図 1 の (b) における “Func1” と “Table1” のように他ノードから孤立したサブグラフが，(c) ではファイル配置情報を考慮することで “Func2” 等と同じディレクトリ “/a/” 以下の配置であることが分かり，“Func2” と同じ移行単位とすべきと判断する根拠とすることができる．なおグラフ中のエッジを定義する際に，CRUD 関係の種別や，ファイル配置情報か否かといったエッジの種類を複数分けて想定したり，エッジの重みとして考慮することも考えられるが，本稿では全てのエッジを同一種，同一重みのエッジとして扱うこととした．

2 つ目の手がかりは，ノード自体に付与される属性情報である．CRUD 関係やファイル配置情報は機能・テーブルノード間のエッジに関する情報であるが，ノード自体にも情報が付属する場合がある．例えば，テーブルの場合「マスタ」「トランザクション」といったテーブルの種類を表す属性が付与されている場合が多くある．これらの属性を考慮することで，重要な属性値を持つ機能ノードとテーブルノード間に CRUD 関係がある場合，他の CRUD 関係よりも優先してそれらノードを同じクラスにまとめるべきである，といった指針を得ることが可能となる．次節で述べるグラフノード分類法の中には，各ノードの属性を直接取り入れることが可能な手法が多く存在し、これら属性情報を導入することは容易である．

3. 半教師有りグラフニューラルネットワークを用いた移行単位最適化

3.1 半教師有りグラフニューラルネットワーク

前節において，CRUD 関係を一種のグラフ構造とみなせることを示し，これにより移行単位の最適化をグラフ構造におけるノードのクラス分類問題と捉えることができた．

グラフ中のノードのクラス分類手法としてはこれまでに数多くの手法が提案されている．まず，教師無しのグラフノードクラスタリング法として，ラプラシアン行列から得られるグラフスペクトルを用いた手法や，スピングラス法 [Reichardt and Bornholdt 2006] 等の手法が知られている．教師無しのアプローチは，教師データを要しないという点でコスト面で優位であるものの，人が意図する移行単位とは異なる分類がなされる可能性が高く，実用に適さない場合もある．

一方，少量の教師データを用意することで，残るノードを人の意図に沿った形で分類できる半教師有り学習法も多く提案されており，ラベル伝搬法やランダムウォークに基づく手法 [Talukdar et al. 2008] が知られている．本稿でもシステム開発者の意図に沿った分類を行うという目的に則り，半教師有りのアプローチを採用することとする．

近年ではさらに，グラフに対して適用可能なニューラルネットワークが提案されており，グラフノードの分類や複数のグラフにおける各グラフの分類が可能となっている [Atwood and Towsley 2016, Kipf and Welling 2017]. Kipf らは，グラフスペクトルに基づく Graph Convolutional Network による半教師有り学習法を提案している [Kipf and Welling 2017]．一方，Atwood らは，diffusion process を用いた Diffusion-Convolutional Neural Network (DCNN) を提案している [Atwood and Towsley 2016]．本稿では高精度なグラフノードの分類結果を得るため，CRUD 表によって定義されるグラフ構造に対し DCNN を適用する．

DCNN は以下の式で定義される．

$$Z = f(W^c \odot P^* X) \tag{1}$$

$$Y = \arg\max(f(W^d \odot Z)) \tag{2}$$

ここで W^c, W^d は $H \times F$ のモデルパラメータ，X は $N \times F$ の全ノードに関する素性行列，Z は diffusion-convolutional activation と呼ばれ，$N \times H \times N$ のテンソルを表す．N は全ノード数，F は全素性数，H はホップ数をそれぞれ表す．P^* は 隣接行列を P としたときの冪級数であり，$P^*_{ijl} = P^j_{il}$ であらわされる．i は元ノード，l は行き先ノード，j はホップ数を表す．DCNN における diffusion とは，あるノードの素性値がホップ数と冪級数に基づいて拡散することを指し，これにより周辺ノードの影響を加味した上でグラフ中の各ノードの特徴を捉えることが可能となる．

3.2 グラフノードの分散表現

前節で述べた半教師有りグラフノード分類法を移行単位の最適化に適用した場合，各ノードの属性情報に基づく元の表現能力が十分でないために，学習が進まないケースが実験的に多く見られた．本節ではグラフ埋め込み法に基づく各ノードの分散表現を用いることでこの課題を解決する．具体的なグラフ埋め込み法としては node2vec [Grover and Leskovec 2016] と Poincaré embedding [Nickel and Kiela 2017] を導入する．本稿では，これら手法で得られた各ノードの分散表現を新たな追加の属性情報として扱う．

node2vec は，自然言語処理の分野で用いられる word2vec [Mikolov et al. 2013] をグラフ構造へ適用するよう拡張した手法で，word2vec の推定に用いられる Negative Sampling を用いた skip-gram 法の枠組みを援用し，任意のノードの近傍を得ることで自ノードに対する分散表現を得る手法である．一列のシーケンス構造で構成されるテキスト中の近傍とは異なり，グラフにおける近傍の定義は様々考えられるが，node2vec ではランダムウォークによって近傍を得る．

一方 Poincaré embedding では双対空間の一種である Poincaré 空間においてグラフ埋め込みを行う手法で，ユークリッド空間に比べ，空間を効率的に利用することができる．特に階層構造における表現能力の高さが知られており，wordnet 等への適用において高い精度が報告されている [Nickel and Kiela 2017] ．本稿で扱うシステム移行単位の解析においても，2.2 節で導入したファイル配置情報がまさに階層構造を取ることから，Poincaré embedding の効果が期待される．

4. 実験

4.1 実験条件

本節では提案手法の効果を確認するため，実際の段階的移行に基づくシステム開発において，人手で設計された 8 クラスの移行単位を教師信号，および正解とみなすことで実験を行った．機能ノード数は 5108，テーブルノード数は 6953，CRUD 関係によるエッジは約 86000，ファイル配置情報によるエッジは約 15000 である．CRUD 関係の種類は，通常の C,R,U,D に加え “Truncate” (T) という操作関係を含め計 5 種を扱う．

各ノードには事前に定義された 32 種類の属性の有無をそれぞれ付与した．人手で設計した移行単位クラスに含まれるノード数は，それぞれ 1106, 1542, 743, 2019, 2806, 800, 481, 2564 で，ノード数最大のクラスを常に推定結果とした場合の正解率，チャンスレートは，21.3% となる．

また，初期に与える教師データ（シード）は，0.1% , 1%, 10%, 30%, 50%, 70%, 90% とデータ量を変化させつつ，全機能・テーブルノードからランダムに選択した．さらに DCNN においては，教師データのうち，1 割を validation のために使用した．

本実験で取り扱う比較対象として，以下 3 つの条件の組み合わせを行う．

- グラフ構造: CRUD 関係を単独で用いる場合 (“CRUD”) と，CRUD 表にファイル配置情報を重畳させた場合 (“CRUD+filepath”) の 2 種類を比較する．
- 半教師有りノード分類手法: 半教師有りグラフノード分類法の一種であるラベル伝搬法 (LP) と，DCNN の 2 種類を比較する．DCNN の epoch 数は 100 とし，validation の誤差が最小となった時点のモデルをテストに用いた．hop 数は 3，dcnn layer および dense layer 数はそれぞれ 5 層とした．
- ノードの分散表現: グラフノード分散表現を用いないベースラインに加え，node2vec, Poincaré embedding を 128, 256, 512 次元で試したうち，LP ではテスト時最良となった node2vec の 128 次元をそのまま用いることとし，DCNN では validation のエラー率が最小となった各手法の 512 次元を用いた．

評価は，各手法が推定したクラスが，人手で作成した移行単位と一致するか否かの正解率に基づいて評価した．

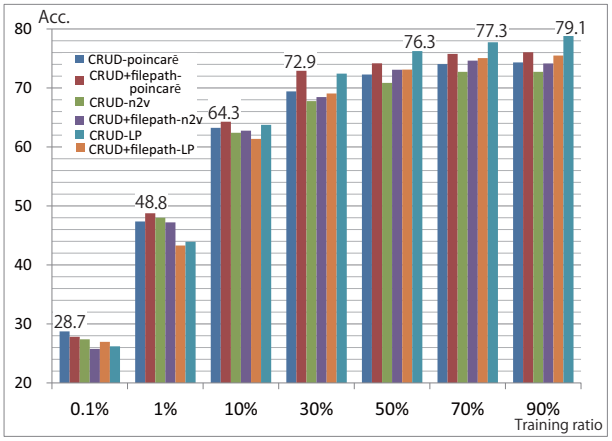


図 2: 推定した移行単位と人手で設計した移行単位との一致率．横軸は学習量割合を示す．グラフ中の数値は各学習量における最高値を表している．

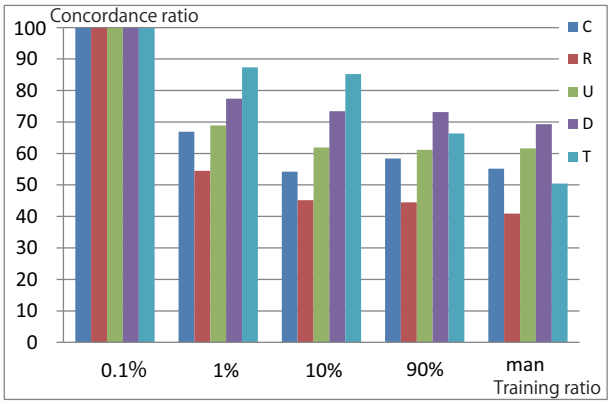


図 3: CRUD 関係を持つ 2 つのノードが同じ移行単位クラスに属する割合を学習量毎に比較した結果．“man” は人手で設計した移行単位の結果を示す．

4.2 実験結果

各比較条件における精度を教師データ量毎に比較した結果を図 2 に示す．ただし，グラフノードの分散表現を追加しない場合については，CRUD 表とファイル配置情報を併用した場合 (“CRUD+filepath”) でも，教師データが 0.1% の時に約 23%，教師データが 90% の時に約 51% と，分散表現を追加する場合と比べ精度が大きく劣ったため，見やすさのため図中から除外している．Poincaré embedding と node2vec の比較では，DCNN の全ての条件下において，Poincaré embedding の方が高い精度を示した．階層構造を持つファイル配置情報を用いる場合だけでなく，通常の CRUD 表に基づく場合でも Poincaré embedding の効果を得られた．

DCNN と LP を比較した場合，教師データ量の少ない条件下 (0.1%～30%) で比較的高い精度が得られた．一方 LP では，教師データ量が多い条件下 (50% 以上) において DCNN を用いた場合よりも高い精度を示した．実利用の面において，本提案法が特に効果を発揮するのは，教師データ量が少ない初期導入での利用と想定され，初期に人手で用意可能な教師データの数を 10～100 個程度と考えると，初期導入においては DCNN

を用いた手法が優位と言える．その後，人手のクリーニング等により移行単位の整理を進めた後，再度 LP で分類，チェックを行うことで，より実用的な効果を得られることが示唆される．

次にファイル配置情報の効果について，“CRUD” と “CRUD+filepath” の結果を比較すると，ほとんどのケースにおいてファイル配置情報を併用した場合に精度向上の効果をもたらしている．しかし，教師データ量 0.1% の DCNN においてのみ，ファイル配置情報を用いない方が精度が高い傾向が見られた．0.1% は 10 個のデータしかなく，教師データの選択方法に精度が敏感に反応した可能性が考えられるが，詳細な検証は今後の課題である

最後に提案手法によって得られた移行単位分類結果をもとに，CRUD 関係にのみ着目した最適性を評価する．図 3 は，各 CRUD 関係 (“C”,“R”,“U”,“D”,“T”) で結ばれた機能とテーブルの対が，同じ移行単位クラスに属する割合を示す．まず，人手で作成した正解となる移行単位の場合 (“man”)，機能とテーブルのクラス一致割合は全体平均で 47.8% であった．人は CRUD 関係以外の要因も加味した上で移行単位を決定するため，CRUD 関係における最適性を満たさなくなったものと考えられる．これに比べ提案手法では，教師データ量が少ない場合，比較的高い一致率を示し，特に学習量が 0.1% においてはほぼ 100% 一致している．そして教師データ量が増えるに従い，人手で作成した場合の一致率に近づく結果 (学習量が 90% において，全体平均で 51.0% の一致率) となった．このことから，提案手法は CRUD 関係における最適化を指向しつつ，人が設計した移行単位に近い分類となるように学習が進んでいると言える．

5. 結論

本稿では，大規模なシステム開発において必要となる移行単位の最適化を目的に，システムを構成する機能とテーブルの関係を表す CRUD 表に着目し，ソースコードのディレクトリ構造や機能・テーブルの属性情報を加味した一元的なグラフとみなした上で，半教師有りグラフノード分類法，特に半教師有りグラフニューラルネットワークの一種である DCNN を用いることで，比較的少量の教師データ (0.1%) でも高精度な移行単位の推定を可能とした．

今後の課題としては以下の三点が挙げられる．第一に，実際の移行単位設計時には人が考慮しているが，現在まだ扱えていない情報を，推定器に組み入れることが挙げられる．例えば，対象システムを用いた実際の業務がどのような手順で行われているかを知ることで，よりリスクの低い，人の分類により近い移行単位の設計が可能となると考えられる．第二に，本稿では既知として与えた移行単位の数を推定することで，これを解決するためには教師無しグラフクラスタリングとの併用アプローチが考えられる．第三に，本稿で各ノードの情報を分散表現から得ている箇所について，本来的にはグラフニューラルネットワークの中で一括して得られることが望ましく，グラフニューラルネットワーク自体及び適用法の改良が必要となる．これらの課題を解決することで，より高精度かつ利便性の高い移行単位の最適化が可能となり，ひいては情報システムの頑健性を一層高められると考える．

参考文献

[木原他 2015] 木原 純平, 鈴木 邦彦, 小倉 孝昭, 浜崎 俊行, レガシーマイグレーションプロジェクトにおける提供品質確保の取り組み, 2015, プロジェクトマネジメント学会誌 17(5), pages 9-14.

[Atwood and Towsley 2016] James Atwood and Don Towsley, Diffusion-Convolutional Neural Networks, 2016, In *Proceedings of the Advances in Neural Information Processing Systems 29 (NIPS 2016)*, pages1993–2001.

[Kipf and Welling 2017] Thomas N. Kipf and Max Welling, Semi-Supervised Classification With Graph Convolutional Networks, 2017, In *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)*.

[Grover and Leskovec 2016] Aditya Grover and Jure Leskovec, node2vec: Scalable Feature Learning for Networks, 2016, In *Proceedings of the 22nd ACM SIGKDD (KDD 2016)*, pages 855-864.

[Reichardt and Bornholdt 2006] Jörg Reichardt and Stefan Bornholdt, Statistical mechanics of community detection, 2006, *Physical Review*, 74. 1

[Talukdar et al. 2008] Partha P. Talukdar, Joseph Reisinger, Marius Paşca, Deepak Ravichandran, Rahul Bhagat and Fernando Pereira, Weakly-supervised acquisition of labeled class instances using graph random walks, 2008, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, pages 582-590.

[Nickel and Kiela 2017] Maximilian Nickel and Douwe Kiela, Poincaré Embeddings for Learning Hierarchical Representations, 2017, In *Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 6338–6347.

[Mikolov et al. 2013] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, Efficient estimation of word representations in vector space, 2013, *arXiv preprint, arXiv:1301.3781.*

[2M3] [General Session] 13. AI Application

座長:荒井 幸代（千葉大学）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room M (2F Amethyst Hall Hoo)

[2M3-01] Drone Flying Area Estimation Method based on Deep Learning II

○Masatoshi Hamanaka[1] （1. RIKEN）

3:20 PM - 3:40 PM

[2M3-02] Toward Building the Evaluation Dataset of Non-Signal Crossing Images

○Shinichiro Goto[1], Yuki Murata[1], Masayasu Atsumi[1] （1. Graduate school of Eng., Soka University）

3:40 PM - 4:00 PM

[2M3-03] Large Scale Driver Identification Using Driving Data

○Daiki Tanaka[1,2], Yukino Baba[1,2], Hisashi Kashima[1,2], Tomoya Saito[3,2], Yuta Okubo[4,2] （1. Kyoto University, 2. RIKEN Center for AIP, 3. SOMPO RISK MANAGEMENT &HEALTH CARE, 4. Sompo Japan Nipponkoa Insurance Inc.）

4:00 PM - 4:20 PM

[2M3-04] Application of individual behavior model in machine learning based cafeteria visitor prediction model and its effectiveness analysis

○Kiyotaka Matsue[1], Hideki Noda[1], Koichi Kondo[1] （1. Toshiba Infrastructure Systems &Solutions Corporation）

4:20 PM - 4:40 PM

[2M3-05] A Preliminary Implementation of Followee Recommendation Mechanism with Self-evaluation Maintenance Model on Mastodon

○Takuma Matsushita[1], Naoki Fukuta[1] （1. Shizuoka University）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ディープラーニングに基づくドローン飛行エリアの推定 II

Drone Flying Area Estimation Method based on Deep Learning II

浜中 雅俊[*1]
Masatoshi Hamanaka

[*1] 理化学研究所
RIKEN

This paper describes a method for estimating the flight area of drones based on deep learning. The position of a drone can be detected by using the global positioning system (GPS). However, GPS sometimes has problems capturing signals from satellites that are shielded by mountains and/or buildings. Moreover, GPS signals are very weak and subject to a variety of disturbances. Such problems increase the chance of a crash when flying a GPS controlled drone. As a solution to this problem, we propose a flight area estimation method using a 3D map created on the basis of deep learning. Our method could estimate the flight area with 98.4 percent accuracy in a field experiment.

1. はじめに

自動運転中のドローンは，衝突を回避するために物体を発見すると，速度を低下あるいは停止して回避する．しかし，ホバリング中や低速移動中でもドローンはプロペラを回転し続けているため，回避を行うほどエネルギーを消費することになる．そこで我々は，各ドローンの位置を把握し管制するシステムを構築してドローンのエネルギー効率を向上させること目指している．

従来のドローンでは GPS（グローバル・ポジショニング・システム）による位置検出が行われていたが，ビルの谷間や山の沢や谷では，衛星からの GPS 信号の遮蔽や反射が生じ，正確な位置が求まらない場合があった．また，GPS 信号は微弱であるため，妨害により受信困難となる場合もあった[Kerns 2014, Urbina 2016] ．

そこで我々は，衛星 3 次元地形図を用いてディープラーニング[LeCun 2015, Amari 2018]に基づき GPS 信号が受信できない場合でもドローンが飛行しているエリアを推定することを可能とする手法を提案し，シミュレーション実験により手法の基本性能の確認を行ってきた[浜中 2016]． 推定手法の特徴は，以下の 3 点である．

- **2 次元 LIDER の使用**： 飛行エリアの推定は，衛星や航空機が取得した 3 次元地形図と，ドローンが取得する地表面の形状のパターンマッチングにより実現できる．このとき，航空レーザ測量に用いられる 3 次元 LiDAR を用いて 3 次元形状どうしのパターンマッチングを行ったほうが簡単な問題となるが，重量が重くドローンに搭載することは難しい．そこで，本研究では，近年自動車の自動運転などに使用させる小型軽量の 2 次元 LiDAR をドローンに搭載し，2 次元の地表断面形状を取得する（図 1）．ドローンによる地表形状ベクトルの取得については 2.1 で述べる．

- **エリア分割**： 衛星や航空機が取得した 3 次元地形図とドローンが取得した 2 次元の地表断面形状のパターンマッチングは，マッチング回数が膨大となり，リアルタイムで処理することが困難である．そこで，我々は推定を行う飛行エリアを分割し，分割された各エリアにラベルを付けることで，位置推定の問題を 2 次元の地表断面形状からラベルを推定する機械学習の問題へ置き換える．これにより計算量を大幅に削減することが可能となる．仮に，ドローンの位置をピンポイントで知りたい場合には，分割されたエリア内で 3 次元地形図と断面形状のパターンマッチングを行い位置の確定を行えばよい．エリア分割については，2.3 で述べる．

- **ディープラーニングの導入**： 機械学習するのはドローン飛行したときに取得する 2 次元の地表断面形状とドローンが飛行しているエリアのラベルであるが，それらは 3 次元地形図から作成される．3 次元地形図から取得できるデータの量は膨大であるために地表断面形状とそのエリアのラベルとの関係をサポートベクターマシン(SVM) [Vapnik 1995]のような手法で求めようとすると，多大な計算時間が必要となるだけでなく，データがメモリに載らず，計算することが困難となる．そこで我々は，機械学習手法としてディープラーニングを導入する．ディープラーニングは，学習データをミニバッチと呼ばれるグループに分割して学習することが可能であるため，多くのデータを用いて学習することが，SVM に比べて容易である．ディープラーニングの一種である MLP（multi-layer perceptron）[Amari 2018]については，3 節で述べる．

本稿では，屋外で実際にドローンを飛行させた場合のエリア推定精度の評価を行う．実験の結果，98.4%の精度でエリア推定が可能であることを確認した．

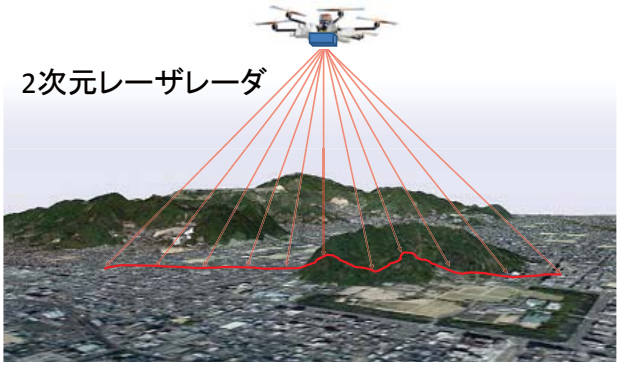


図 1: 地表面の断面形状の取得

masatoshi.hamanaka@riken.jp
http://gttm.jp/hamanaka/

2. 学習・評価データの作成

ドローンの飛行エリアの推定は 2 つのステップからなる. まず, ドローンに 2 次元 LiDAR を搭載し, 地表面の断面形状を取得する. 次に, 3 次元地形図を用いて学習したネットワークに, LiDAR から取得した断面形状を入力することでエリアを推定する. 本節では, 学習・評価用データの作成について述べる.

2.1 地表面形状の取得

SICK 社の LD-MRS をドローンの鉛直下向きに搭載し地表断面の形状を取得する. ドローンは 50 メートル以上の高度を飛行することができるが, ドローンに搭載可能な小型軽量の LiDAR の測定範囲は, 地表の反射率によって異なるがおおよそ 50 メートルである[SICK 2018]. そこで, ドローンの飛行高度は 50 メータ以下とした.

2.2 で述べる 3 次元地形図の分解能は 1 メートルであるため, 地表の高度変化を測定する点は, 1 メートル間隔で左右 25 点ずつ, 合計 50 点の計測点を用意し, その 50 点とドローンとの高度差を 50 次元の地表形状ベクトルとして取得する.

2.2 3 次元地形図の作成

群馬県昭和村における北緯 36.6217 から 36.619900 まで, 東経 139.1090 から 139.111240 の 200 メートル四方に LiDAR (SICK 社 LD-MRS)を搭載したドローン(DJI 社 Matrice600)を 10 回飛行させ, 3 次元地形図を作成した(図 2). Matrice600 に搭載の GPS は 1 メートル程度の誤差があるから, 解像度を 1 メートルとして, 複数回の平均を求めた[DJI 2018](図 3).



図 2: ドローン(Dji 社 Matrice600)と LiDAR (SICK 社 LD-MRS)

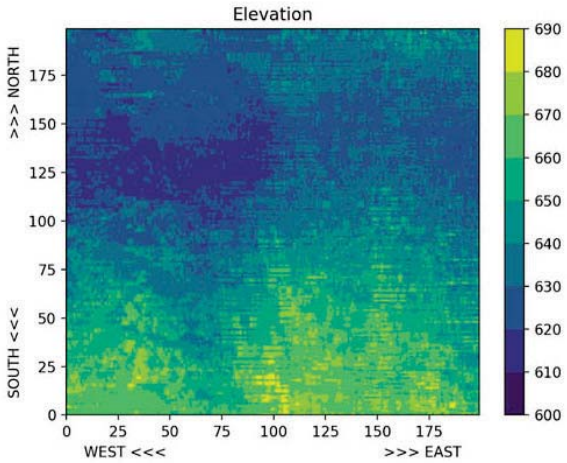


図 3: 3 次元地形図

2.3 学習データの作成

200 メートル四方の飛行エリアを縦横 5 等分に分割し, 分割で生じた 40 メートルの正方形にそれぞれ異なるラベルを付与する.

次に, ドローンの位置からその時の地表形状ベクトルを 3 次元地形図から算出したものと, ドローンがいるエリアのラベルの組を学習データとする. 学習データが大きすぎて GPGPU 上のメモリに乗らない場合には, 計算時間が多大となることから, 本稿では, ドローンの向きを北向きに, 高度を 3 次元地形図の平均高度から 35 メートル上空に固定することにした. 向きや高度を固定しない場合でも推定可能であることは確認済である[浜中 2016].

ドローンは 200 メートル四方の範囲を飛行し, LiDAR のスキャン幅は 50 メートルである. また, 3 次元地形図の解像度は 1 メートルである. したがって, 地表形状ベクトルとラベルのセットは 3 万件(=200x150)作成した(図 4).

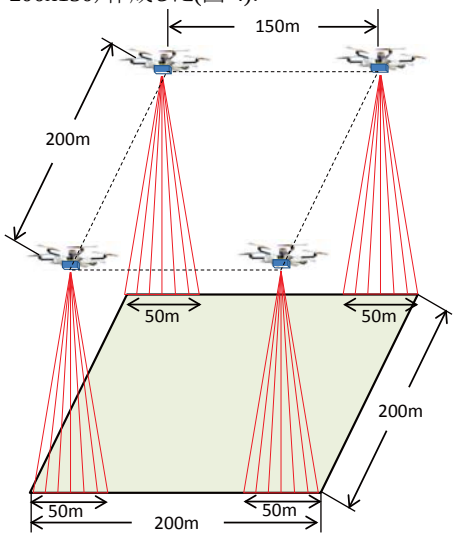


図 4: 飛行範囲

2.4 評価データの作成

ドローンを高度 35 メートルで 20 分間飛行させ, 2 秒ごとに取得した地表形状ベクトルと, そのときの GPS ポイントから得たラベルの組を, 評価データとした. 評価データの作成は, ドローンを 5 回飛行させて行ったためデータ件数は 3000 件(5 回x20 分 x60 秒/2 秒)である.

オープンな条件での評価とするため, 3 万件の学習データから評価データ 3000 件の各ポイントから最も近いデータを除き, 学習には 2 万 7 千件を用いることにした.

3. 実験結果

位置推定精度は, MLP の隠れ層の層数およびユニット数によって変化する(図 5). 層数およびユニット数をランダムに変化させて精度を比較したところ, 深層に向かって徐々にユニット数が減少していくネットワークで精度が向上する傾向があることがわかった.

適切な層数およびユニット数を求めるため, 深層に向かって一定の割合でユニット数が減少するネットワークに限定して探索を行う. 層数は 1 から 7 まで, ユニット数の減少率は, 0.3 から 1.0 まで 0.1 刻みで全探索した. ユニット数の最大は 3000 である. たとえば, 3 層で減少率 0.5 のネットワークでは, 1 層目が

1500(3000x0.5)ユニット, 2 層目が 750(1500x0.5)ユニット, 3 層目が 375(750x0.5)ユニットである.

深層学習には多大な時間がかかることから, 段階的に探索エリアを絞っていくことにした. 表 1 は, 2000 エポックまでの学習でのエリアが正しく推定できた精度である. 精度が 0.5 以下のセルの背景を灰色としている. 精度が低いのは, 以下の 3 つのいずれかの場合である.

・層数およびユニット数の両方が大きい
・層数が 1 である
・出力層の前の層のユニット数が非常に少ない

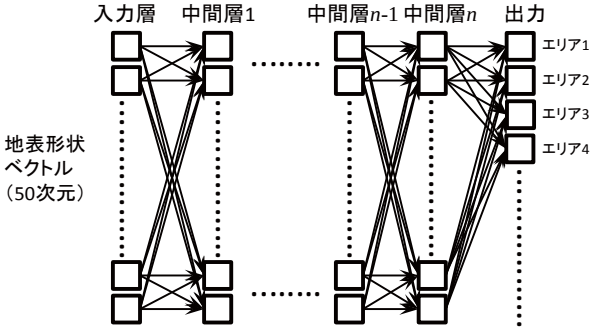


図 5: 位置推定のための MLP

表 1: 2000 エポックまでの結果

	ユニット数の減少率							
層数	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1	0.51	0.50	0.50	0.49	0.49	0.46	0.49	0.50
2	0.67	0.71	0.75	0.76	0.74	0.74	0.73	0.66
3	0.61	0.67	0.66	0.69	0.70	0.70	0.64	0.63
4	0.57	0.62	0.61	0.61	0.59	0.66	0.30	0.05
5	0.55	0.60	0.62	0.58	0.61	0.05	0.05	0.06
6	0.36	0.60	0.61	0.66	0.58	0.05	0.05	0.05
7	0.19	0.49	0.63	0.66	0.05	0.05	0.05	0.05
8	0.14	0.23	0.05	0.05	0.05	0.05	0.05	0.06

表 2 は, 表 1 で精度 0.4 以上だったものについて 1.5 万エポックまでの学習した結果である.

表 2: 15000 エポックまでの結果

	ユニット数の減少率							
層数	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1	0.65	0.65	0.64	0.65	0.65	0.64	0.64	0.65
2	0.83	0.87	0.85	0.85	0.87	0.87	0.87	0.84
3	0.83	0.90	0.93	0.95	0.95	0.92	0.93	0.89
4	0.80	0.91	0.94	0.94	0.93	0.91		
5	0.77	0.91	0.95	0.93	0.88			
6	0.53	0.91	0.94	0.93	0.85			
7		0.81	0.95	0.94				

表 3 は, 表 2 で精度 0.94 以上だったものに対して 3 万エポックまで学習した結果である. 図 6 は, その学習曲線である. 3 層で減少率 0.6 のネットワークでは, 20509 エポックで精度 0.9762, 7 層で減少率 0.5 のネットワークでは, 29677 エポックで精度 0.98757 となった.

表 3: 30000 エポックまでの結果

	ユニット数の減少率		
層数	0.5	0.6	0.7
3		0.9762	0.9705
4	0.9615	0.9607	
5	0.9670		
6	0.9756		
7	0.9759	0.9669	

ディープラーニングでは, データオーギュメンテーションにより認識性能が向上することが知られている[Lv 2017]. 我々は, 地表断面ベクトルの各要素に一様乱数を付加することで学習データを拡張することを試みた.

拡張した学習データは, オリジナルの学習データ 2.7 万件に, ランダムを付加した 2.7 万件の合計 5.4 万件である. 付加するランダムノイズ値が±0.25m, ±0.50m, ±1.00m となる 3 種類の拡張学習データを作成した. 学習には, 表 3 で高性能だった, 3 層で減少率 0.6 のネットワークおよび 7 層で減少率 0.5 のネットワークの 2 種類を試みた.

その結果, 7 層で減少率 0.5 のネットワークでノイズレベルが±0.50m の場合に 22119 エポックで精度 0.9838 となり最も高性能となった(表 4). 学習時間は, GPGPU (NVIDIA GTX 1080Ti)を用いて 7 時間であった.

表 4: ランダムノイズを加えて拡張学習データでの精度

	ノイズレベル		
層数および減少率	±0.25m	±0.50m	±1.00m
3 層で減少率 0.6	0.9730	0.9721	0.9712
7 層で減少率 0.5	0.9782	0.9838	0.9813

4. おわりに

我々はこれまで, ディープラーニングを用いてドローンに搭載した小型軽量の 2 次元 LiDAR が取得した地表断面ベクトルからドローンが飛行しているエリアを推定する手法について提案し, シミュレーション実験により手法の基本性能を確認してきた. 本稿において達成したのは以下の 3 点である.

- **手法の実世界における性能評価**: 群馬県昭和村における 200 メートル四方を実験場所として, ドローン搭載の LiDAR を用いて 3 次元地形図を作成した. そして, その 3 次元地形図を用いて 2 万 7 千件の学習データを作成し, MLP で学習を行った結果, ドローンが取得した地表面ベクトルからエリア推定が可能であることを確認した.

- **ネットワーク構成の検討**: MLP によるエリア推定精度はネットワーク構成により変化する. MLP の層数およびユニット数をランダムで変化させて試行を繰り返した結果, 深層に向かって徐々にユニット数が減少する構成で性能が高いことが分かった. そこで, 深層に向かって一定の割合でユニット数が減少する構成に限定して適切なネットワークを探索したところ, 深層に向かってユニット数が 1800, 1080, 648 と変化する 3 層で減少率 0.6 のネットワークおよび, ユニット数が 1500, 750, 375, 187, 93, 46, 23 と変化する 7 層で減少率 0.5 のネットワークで精度 0.976 となり高い性能を示した.

- **拡張データによる評価**: ディープラーニングでは，データオーギュメンテーションにより認識性能が向上することから，地表断面ベクトルの各要素にランダムにノイズを付加したデータを作成し，性能を評価した．その結果，オリジナルのデータ 2.7 万件に，ノイズのサイズが±0.50m になるようノイズを負荷したデータ 2.7 万件を加えた 5.4 万件のデータを用いて 7 層で減少率 0.5 のネットワークを学習させたところ，精度が 0.984 となり高い性能を示した．

今後，陸域観測技術衛星だいち 2 号（ALOS-2）の PALSAR 2 (Phased Array type L-band Synthetic Aperture Radar 2)から得られるデータを用いて，広範囲のエリアを MLP に学習させて性能評価を行っていきたい．

謝辞

本研究の一部は，JSPS 科研費 17K19972 の助成を受けたものです．

参考文献

[Kerns 2014] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, “Unmanned Aircraft Capture and Control Via GPS Spoofing,” Journal of Field Robotics, Vol. 31, Issue 4, pp. 617-636, 2014.

[Urbina 2016] D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, H. Sandberg, “Limiting the Impact of Stealthy Attacks on Industrial Control Systems,” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 1092-1105, 2016.

[LeCun 2015] Y. LeCun, Y. Bengio, G. E. Hinton, “Deep Learning,” Nature, Vol. 521, Issue 7553, pp. 436-444, 2015.

[Amari 2018] Amari S, Ozeki T, Karakida R, Yoshida Y, Okada M, Dynamics of Learning in MLP: Natural Gradient and Singularity Revisited, Neural Comput, Vol. 30, Issue 1, pp. 1-33, 2018.

[浜中 2016] 浜中雅俊, 塩見英樹, 小田雄一: ディープラーニングに基づくドローン飛行エリアの推定，人工知能学会全国大会, 3E4-5, 2016.

[Vapnik 1995] V.Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.

[SICK 2018] SICK, The LD-MRS family – compact laser scaners for outdoor use, from: https://www.comtrol.com/elements/uploads/fckeditor/file/LD-MRS_4001021.pdf, 2018.

[DJI 2018] Dji, Matrice600, from https://www.dji.com/matrice600, 2018.

[Lv 2017] JJ. Lv, XH. Shao, JS. Huang, XD. Zhou, and X. Zhou, “Data augmentation for face recognition, Journal of Neurocomputing," Vol. 230, Issue C, pp. 184-196, 2017.

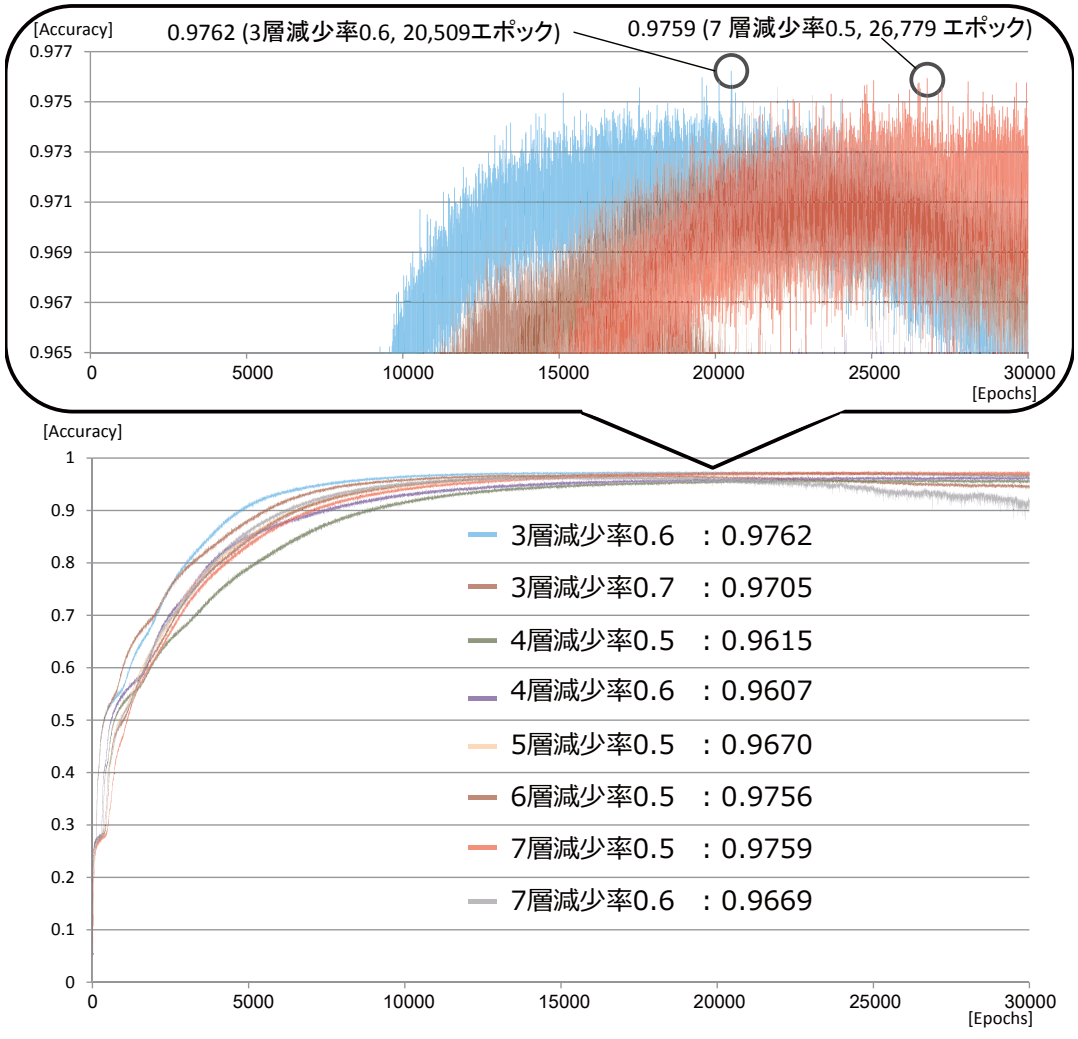


図 6：30000 エポックまでの学習曲線

無信号交差点画像の評価用データセット構築に向けて

Toward Building the Evaluation Dataset of Non-Signal Crossing Images

後藤紳一郎[*1] Shinichiro Goto　村田祐樹[*1] Yuki Murata　渥美雅保[*1] Masayasu Atsumi

[*1] 創価大学大学院工学研究科情報システム工学専攻

Information Systems Sci., Graduate school of Eng., Soka Univ.

We are considering a system capable of providing a route of minimum driving risk to reduce accidents of elderly drivers. In this system, mapping the driving risk is an important issue so that it is necessary to estimate the risk level of non-signal crossings where accidents occur frequently. To estimate the risk level of intersections, the viewing environment, especially the viewing angle, has a big influence. Thus, estimating the viewing angle from the dash-cam image was selected as a research topic. In order to construct an estimation model, it is necessary to prepare a data set for machine learning, and as a first step, a data set of images and viewing angles was built as a prototype by using Google Street View.

1. はじめに

近年，日本では高齢ドライバーによる交通事故が増加し社会問題となっている．2017 年の交通事故死者は 3,694 人と昭和 23 年以降の統計で最少となったが，半数以上が高齢者であり増加傾向である[警察庁 18]．2015 年の統計によると3,392 万人が 65歳以上の高齢者である[内閣府 16]．事故類型別の事故件数割合を年齢別に比較すると，65 歳以上の高齢ドライバーは，出会い頭事故が相対的に高くなっている．その中でも，信号機なし交差点の件数が 65%と高い．つまり，高齢ドライバーの事故は無信号交差点が特徴的であると言える．[ITARDA 16]

我々は，高齢ドライバーの事故低減のために運転危険度最小ルート提供が可能なシステムを検討している[後藤 2017]．そこでは，図 1 に示すように，運転危険度をマッピングする事が課題となる.特に，高齢ドライバー事故が多発している無信号交差点の危険度を推定する必要がある．

交差点の危険度の推定には視環境，特に視野角が大きな影響を持つ．そこで，視野角をドライブレコーダー画像から推定する事を研究課題とした．

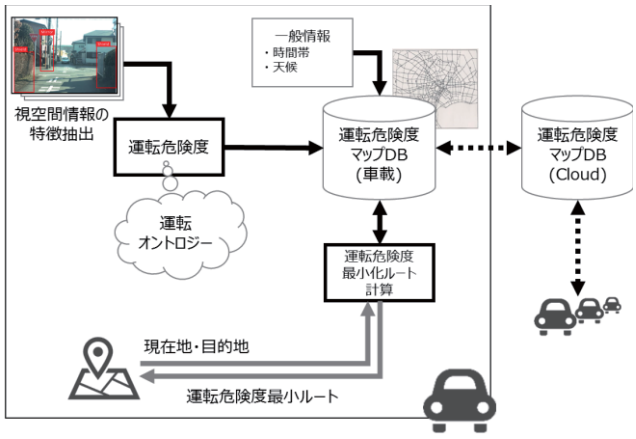


図 1　提案システム

連絡先：後藤紳一郎，shinichiro.goto.510@gmail.com

注 1: https://www.google.co.jp/intl/ja/streetview/

推定モデル構築の為に機械学習用データセットを準備する必要がある．その要件を明確化するために，第一段階としGoogle ストリートビュー(注 1)から得た画像と視野角のデータセット試作を行い，今後のドライブレコーダー映像からデータセットを生成するための課題整理を行う．

2. 関連研究

無信号交差点の環境要素を考慮した出会い頭衝突発生危険度の研究[小竹 16]では，事故に寄与する環境要素として「見通し」を上げている．その計測手法として，レーザ計測車両で交差点実環境の点群・位置情報を取得し，ドライバーのアイポイントから見通せる路面までの距離を算出している．

この研究は正確な見通しの数値化が期待できる一方で，交差点環境の取得範囲の拡大や時系列変化への対応などにデータ取得コストの増大が予想される．また，レーザ計測では判定が難しいフェンスや透過性のある材料が使われた遮蔽物の取り扱いも課題になると言える．

本研究では，取得コストが比較的小さなドライブレコーダーを用い，その画像から視野角を推定し，交差点の運転危険度付けを目指す．画像による推定には，機械学習を用い，データセットから教師付き学習をした推定モデルを構築する手法が一般的であるが，その学習用データセットを準備する方法が課題となる．

3. 試作データセットの準備

現在，交差点の視野角がアノテーションされたドライブレコーダー画像の公開データセットは存在しない．

構築には全て自前で準備する方法も考えられるが，コストが安く，効率的に推進する方法の検討が必要である．そのため，まずはオープンデータを用いたデータセットの試作を行う事とした．もっとも身近な交差点画像といえば Google ストリートビューで，生活道路も含め，多くのパノラマ画像が準備されている．これらのデータを使用し，交差点画像と視野角からなるデータセットを試作することとした．以下に試作した手順を紹介する．

図 2 に示すように，信号無し交差点のデータ取得を行うには，衛星画像から信号無し交差点を探す為に「止まれ」の文字を検出し，その位置座標の格納を行う．今回は手動操作で 1,000 地点のデータを取得した．次に，取得した信号無し交差点における視環境を確認するため，衛星画像から拡大し，ストリートビューに

切り替わる機能を活用した. 交差点入り口に最も近い場所の画像にし, その結果として出力される URL を記録する.

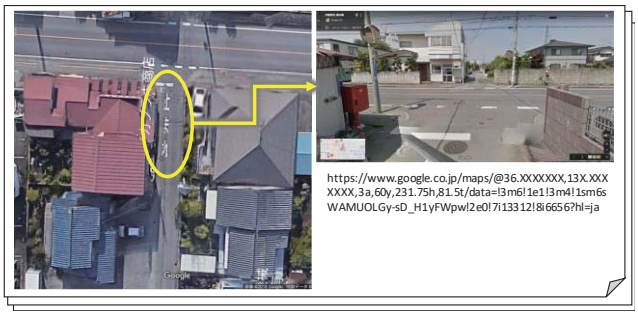


図 2 信号無し交差点画像の取得例

続いて, 視野角に関しては Google ストリートビューの 360 度カメラの映像を用いて, 運転視野の範囲を探った. その方法を図 3 に示す. ストリートビュー画像の特徴として, 方向は動的に URL 表示されている. そこで, 運転視野の左端の方向を中心になるようにし, 表示された URL を記録し, 同様に右端に関しても行い, 左右の方向の差を取れば, 視野角は算出される. こうして, 交差点 1,000 箇所のストリートビュー画像と視野角のデータの準備を行った.



図 3 視野角情報の抽出例

そこで発見された取り扱いに配慮すべき点を表 1 に示す.
上記の問題点の対応のため, 以下の対応を行った.

- 視野角が 120 度以上のデータを省く
- 交差点からの距離が遠い画像を省く
- 高さを考慮して, 視野角の再アノテーションを行う

これらを実施することで, 1,000 データ中の半分近くである 600 件程度のデータに集約された.

4. 試作データセットで得られたこと

視野角を決定する遮蔽物が何であるかの確認を行った. その結果を図 4 に示す.

建物の壁が 36%と最も多い対象物であった. 次に塀(17%), 植栽(16%), 駐車車両(15%)となった. 駐車車両のような移動可能な対象物が 4 番目に多い結果であった事が新たな発見であった.

その他, 数は少ないものの, 自動販売機(2%)なども遮蔽物となっていた. また, 透過性のある半遮蔽対象物として, メッシュ状等の透過性のあるフェンス(9%)が含まれていた. これらより, 交差点画像中でこれらの対象物の認識を行うことが, 視野角を判定する事に大きな影響がある.

表 1 試作データセットの配慮すべき点

No.	問題点	画像例
1	視野角が 120 度を超えると画像の歪みが大きくデータとして適さない.	
2	交差点からの距離が不定であり, 同じ視野角でも見え方に差がある	交差点のかなり手前(63deg)  交差点直前(63deg) 
3	カメラの取り付け位置が実際の運転者の目線より大分高いため, 実際よりも広い範囲の視野角が得られている	

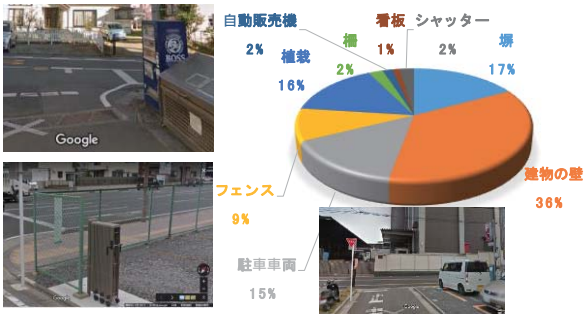


図 4 遮蔽対象物

これらの対象物について, 実際に Faster R-CNN[Ren 15], SegNet [Vijay 15], YOLO[Redmon 15]の機械学習を用いた物体認識技術を活用すると, 図 5 に示すように駐車車両, 建物の壁, 道路標識等は認識可能である. これらの技術を組み合わせることで, 進行方向における遮蔽物の位置を推定し, 視野角の自動認識を可能にすれば, アノテーション作業の効率化に繋がる.

今後は次項に示すような実際のドライブレコーダー画像からアノテーションされたデータセットを十分な量で効率よく構築し, 交差点の危険度推定に向けた視野角推定モデルの試作へと進む. また,データセットの充実のために, 都内と地方, 北海道と九州, のような地域別の道路形状の違いによるバリエーションを準備し, 推定モデルの頑健性を検証する.



図 5 物体認識事例

5. ドライブレコーダーデータセット作成に向けて

上記の試みにより得られた知見から, ドライブレコーダー映像からのデータセットとしては以下の要件で構築を行う事とする.

- 水平画角 120 度程度
- 取り付け位置は普通乗用車程度の 1200～1500mm
- 位置座標と車速が記録され, 交差点からの距離が確認可能なこと

また、上記データセットを構築するためのツールの要件は以下の通りとなる.

.

- ドライブレコーダー映像から無信号交差点映像を自動で抽出する
  - ➢ 一時停止標識を認識し, 非優先から優先道路への進入画像を自動抽出する
  - ➢ 視野角 120 度を越える画像は「見通しの良い視野角」クラスとして自動分類する
- 自動化も見据えながら, 効率よく視野角のアノテーションを行える

以上の要件を満たすデータの取得と, ツールの構築, アノテーション作業を行う事で, 交差点の視野角推定器を開発し, 危険度推定研究を進める.

6. おわりに

高齢者ドライバーに対する, 運転危険度の最小化を目的としたルート提案システム実現に向け, その技術要素となる交差点の視野角推定を行う上で必要なデータセット試作に取り組んだ. 今後は, ドライブレコーダー映像も対象にしてデータセットを構築し, 危険度推定の研究を進める. 高齢ドライバーの運転負荷が低減し, 運転寿命が延びる事で,「すべての人が心から安心して, どこへも自由に移動できる社会」を目指していきたい.

参考文献

[警察庁 18] 警察庁交通局交通企画課: 平成 29 年中の交通事故死者数について, 警察庁ホームページ報道発表資料 (2018)

[内閣府 6] 内閣府: 平成 28 年版高齢社会白書(概要版), 内閣府ホーページ, 第 1 章 高齢化の状況(2016)

[後藤 17] 高齢ドライバーの運転負荷低減を目指す支援システムの提案, 2017 年度 人工知能学会全国大会（第 31 回）

[小竹 16] 無信号交差点の環境要素を考慮した出会い頭衝突発生危険度の一提案, 2016 年秋季大会 学術講演会

[Ren 15] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[Vijay 15] Vijay Badrinarayanan, et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling." arXiv preprint arXiv:1505.07293, 2015.

[Redmon 15] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." arXiv preprint arXiv:1506.02640 (2015).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

運転データによる大規模ドライバー識別

Large Scale Driver Identification Using Driving Data

田中 大貴 *1*2 Daiki Tanaka　馬場 雪乃 *3*2 Yukino Baba　鹿島 久嗣 *1*2 Hisashi Kashima　齋藤朋也 *4*2 Tomoya Saito　大久保雄太 *5*2 Yuta Okubo

*1京都大学 Kyoto University　*2理化学研究所 革新知能統合研究センター RIKEN Center for AIP　*3筑波大学 University of Tsukuba

*4SOMPO リスケアマネジメント株式会社 SOMPO RISK MANAGEMENT & HEALTH CARE　*5損害保険ジャパン日本興亜株式会社 Sompo Japan Nipponkoa Insurance Inc.

We address a large-scale driver identification problem whose goal is to predict the driver who drives a car from driving data collected using GPS devices. In contrast with the prior research considering at most a few dozen of drivers, we try to identify a huge number of drivers up to 10,000 drivers. The experiment show our method can identify drivers more precisely than baseline method. We also show that temporal features are quite effective in the large scale driver identification and that speed and acceleration features also contribute to driver identification.

1. はじめに

ドライバー識別問題とは、自動車を運転しているドライバーに関する様々な情報をもとに、ドライバー個人を識別する問題である。その応用としては、複数人が同じ車を運転する可能性がある場合に、各ドライバーに合ったミラー角度の設定、運転席の高さの設定、エアコンの温度の設定などの適切な車両設定の自動化や、ドライバーの好みに合わせた情報推薦、ドライバーの運転経歴の自動作成、車両盗難の検知など多岐にわたる。

近年、機械学習をはじめとするデータ解析技術は様々な分野においてその利用が検討されており、ドライバー識別もその例外ではない。機械学習を用いてドライバー識別に取り組む研究は過去にも試みられてきたものの [Fung 17, Hallac 16, Van Ly 13, Miyajima 07]、データの大規模収集の困難さから、識別対象とするドライバー数は高々十数人の規模であった。一方、GPS を搭載したスマートフォン等の普及に伴い、これらを用いたアプリケーションを車上で用いることによって、運転行動に関連するデータが比較的簡単に大規模に収集できるようになってきた現在、大規模なドライバー識別の実現可能性が高まってきた。

本研究ではスマートフォン上で動作する運転診断アプリケーションを用いて収集された運転データをもとに、数千人規模のドライバー識別の実現可能性について検証する。最大 1 万人を対象としたドライバー識別実験では、ベースライン手法と比較して高い精度でドライバーを識別できることが示された。また、位置や時間に関する特徴量、特に位置による特徴量が大規模なドライバー識別において極めて有効であることを確認するとともに、速度や加速度情報もドライバー識別に一定の寄与があることを示した。

本研究の貢献は以下のように纏められる：

- 実データに対して数千人規模という大規模なドライバー識別を行なった。
- ドライバー識別に有効な特徴量について検討した。

連絡先: 田中 大貴, daiki_yosky@ml.ist.i.kyoto-u.ac.jp

2. 関連研究

これまでの研究においても、ドライバー識別問題は比較的少人数を対象とした設定で取り組まれてきた。Fung らは、14 人の熟練ドライバーの加速時、減速時の速度、加速度、躍度、車体角度から特徴抽出を行った [Fung 17]。運転データに含まれる加速、減速それぞれの識別結果を多数決を用いて統合し、識別器には多クラス線形判別分析を用いて、2 人でのドライバー識別において約 80%、14 人でのドライバー識別において約 50%の精度を得た。

Hallac らは、車両がコーナーを曲がる数秒の短い区間の情報のみからドライバー識別に取り組んだ [Hallac 16]。ハンドル角度、ステアリング速度・加速度、車体速度・加速度、車体の向き、エンジン回転数、アクセルペダル・ブレーキペダルの位置、トルク、スロットル位置から特徴量を抽出し、分類を行なった。ウェーブレット変換による特徴生成も行った。識別器にランダムフォレストを使い、2 人でのドライバー識別において 76.9%、5 人でのドライバー識別において 50.1%の精度を得た。

Van Ly らは、加速、減速、コーナリングという 3 つのイベントを用いて 2 人ドライバー識別に取り組んだ [Van Ly 13]。彼らは運転データからこれら 3 種のイベントを抽出し、イベント毎に特徴抽出を行い、ドライバー識別を行なった。彼らは CAN バスから得られる加速度とジャイロ情報から特徴抽出を行なった。識別器には SVM と k-means クラスタリングを用いて、約 60%の精度を得た。

Miyajima らは、ガウス混合モデルを用いてアクセルペダルとブレーキペダルの踏み具合をモデル化することで、2 人ドライバー識別に取り組んだ [Miyajima 07]。時系列データに対してケプストラム解析を行い、その有用性について示した。精度は 76.8%を得た。

表 1 に纏めるように、過去にもドライバー識別に取り組んだ研究はいくつかあるが、いずれも高々数十人の識別を対象としている。本研究では、これらの研究が取り組んでいない数千人規模でのドライバーの識別に取り組んだ。

表 1: ドライバー識別の関連研究の概要

著者	年	分類器	ドライバー数
Fung et al. [Fung 17]	2017	多クラス線形判別	14
Hallac et al. [Hallac 16]	2016	ランダムフォレスト	5
Van et al. [Van Ly 13]	2013	SVM, k-means	2
Miyajima et al. [Miyajima 07]	2007	ガウス混合モデル	2

3. データセット

本研究で使用するデータセットは、損害保険ジャパン日本興亜株式会社によって収集された約 36,000 人分の運転データである。なお、位置情報を含む運転データについては、個人情報の削除ならびに変換を行い、データから個人特定が困難であるように加工したのちに用いた。

3.1 データ取得

本研究で用いたデータセットは、自動車内にドライバーが持ち込んだスマートフォン上で動作する運転診断アプリケーションから収集されたものである。ドライバーはアプリケーションを起動させ、運転データの送信を開始し、運転を開始する。運転が終了したら運転データの送信を終了する。運転データのサンプリング頻度は 1Hz である。

3.2 運転データ

運転データは、ドライバーが行なう運転に関するデータを提供する。運転データはスマートフォン ID に紐づけられている。運転データには 1Hz 毎に記録された時間 (タイムスタンプ)、緯度、経度、速度、前後加速度、左右加速度が含まれている。運転データに含まれる総走行時間は約 70 万時間で、欠損データを除いた総走行時間は約 58 万時間である。総運転データ数は 1,575,841 であり、36,725 人の運転データが含まれている。また、運転データの時系列長はデータにより様々である。本研究で行なう実験にはこの運転データのサブセットを用いる。

4. 問題設定

本研究で扱うドライバー識別問題とは、運転中のタイムスタンプ、経度、緯度、前後加速度、左右加速度、速度のデータから、その運転を行なったドライバーを識別することである。

入力として、各ドライバーの各運転に対して、運転データが与えられる。なお、各ドライバーに対して、1 つ以上の運転データが存在する。運転データとは、1 回のアプリケーションの開始から終了までの走行時の情報を記録したものであり、具体的にはタイムスタンプ、位置情報、加速度データ、速度データが時系列で記録されている。すなわち、i 番めの走行データは、タイムスタンプ $\mathbf{time}_i$ と緯度 $\mathbf{lat}_i$ と経度 $\mathbf{lon}_i$ と前後加速度データ $\mathbf{acc}_i^{\mathrm{FB}}$ と左右加速度データ $\mathbf{acc}_i^{\mathrm{LR}}$ と速度データ $\mathbf{v}_i$ で与えられる。なお、$\mathbf{time}_i$, $\mathbf{lat}_i$, $\mathbf{lon}_i$, $\mathbf{acc}_i^{\mathrm{FB}}$, $\mathbf{acc}_i^{\mathrm{LR}}$, $\mathbf{v}_i$ は時系列データである。時系列データ $\mathbf{time}_i$, $\mathbf{lat}_i$, $\mathbf{lon}_i$, $\mathbf{acc}_i^{\mathrm{FB}}$, $\mathbf{acc}_i^{\mathrm{LR}}$, $\mathbf{v}_i$ から i 番目の走行に対する特徴ベクトル $\boldsymbol{x}_i = (x_i^1, x_i^2, ..., x_i^n)$ を構成し、その特徴ベクトル集合ならびにドライバーの識別子から分類器を訓練する。

なお、本研究では 3 種類の特徴ベクトルを設計し、それらからそれぞれドライバー識別を行う。特徴量と特徴ベクトルの構成については次の章で述べる。

表 2: 実験で用いる特徴量の種類

データ	特徴量の種類
速度、加速度	前後加速度の最大値、最小値、平均値、分散
	左右加速度の最大値、最小値
	速度の最大値、平均値、分散
位置、時間	アプリ開始・終了時の緯度、緯度の平均・分散
	アプリ開始・終了時の経度、経度の平均・分散
	運転開始時の曜日、時間

5. 特徴量の設計

本研究でドライバー識別に用いる特徴量は、速度と加速度に関する特徴量、位置に関する特徴量、時間に関する特徴量の 3 種類である。それぞれの特徴量について、以下で説明するとともに表 2 にまとめる。

5.1 速度と加速度に関する特徴量

運転データに含まれる速度データと加速度データから、9 つの特徴量を設計する。速度については、最大値、平均値、分散を、加速度については、前後加速度の最大値、最小値、平均値、分散ならびに左右加速度の最大値、最小値を用いる。

なお、信号待ちなどによる停車中のデータや徒歩によるデータを除外するため、速度に関する特徴量を抽出する際には時速 7km/h 以下のデータを排除し、加速度に関する特徴量を設計する際には −0.001G 以上 0.001G 以下のデータを排除する。

5.2 位置に関する特徴量

運転データに含まれる緯度データと経度データから、アプリケーション開始時と終了時の緯度経度ならびに緯度経度の平均値と分散の 8 つの特徴量を用いる。

5.3 時間に関する特徴量

運転データに含まれるタイムスタンプから、アプリケーションが起動された時間と その日の曜日の 2 つの特徴量を設計する。なお、曜日は月曜日から日曜日まで曜日ごとに、0 から 6 までの異なる数を割り当て、それらを one-hot-vector として特徴量とする。得られた one-hot-vector を他の特徴量から構成された特徴ベクトルに連結する。

6. 評価実験

前述のデータを用いて、多人数でのドライバー識別の実現可能性を検討する。まずは大規模ドライバー識別で用いる識別器を決定するための小規模での予備実験を行った後、全データを用いた本実験を行った。

6.1 評価指標

数千人を超える多数のドライバー識別を確実に行うのは困難である。識別器はより本人であると思われる順位を出力するため、上位に予測されるドライバーに本人が含まれていれば良

いとする予測精度指標として、情報検索の精度評価にしばしば用いられる Recall@k と平均逆順位を用いる。

Recall@k 上位 k 件の候補中に含まれる正解ラベル数を全正解ラベル数で割ったものの平均値である。本研究では正解ラベルが運転をしたドライバーのラベル 1 件であるので、Recall@k は以下の式で定義される：

$$\text{Recall@k} = \frac{1}{|M|} \sum_{m \in M} t_{m,k} \tag{1}$$

ただし、$t_{m,k}$ はサンプル $m \in M$ の予測結果の上位 k 位に正解ラベルが含まれていれば 1、含まれていなければ 0 をとる。実験では $k = 1, 3, 5, 10$ の場合について Recall@k を求めた。

平均逆順位 正解ドライバーの順位の逆数の平均値であり、以下の式で定義される：

$$\text{MRR} = \frac{1}{|M|} \sum_{m \in M} \frac{1}{r_m} \tag{2}$$

ただし、r_m はサンプル $m \in M$ の予測結果中の正解ラベルの順位である。

6.2 予備実験

大規模ドライバー識別問題で用いる識別器を決定するため、まずは予備実験として小規模なドライバー識別問題によって各識別器の性能を確認する。

識別対象とするドライバーは、ドライバー集合からランダムに選択された 30 人とする。用いる運転データ数は合計 6,248 件で、80%を訓練データ、20%をテストデータとして使用する。なお、識別に用いる特徴ベクトルは全ての特徴量で構成する。識別器として、SVM、ニューラルネットワーク、ランダムフォレストの 3 種を用いる。各識別器に対して 5-交差検定によるグリッドサーチによってハイパーパラメータチューニングを行う。SVM のハイパーパラメータの探索空間は、誤識別数に対するペナルティC の値域は $\{0.1, 1.0, 10, 30, 50\}$、カーネル関数は {RBF カーネル, 線形カーネル, 多項式カーネル, シグモイドカーネル } の 4 種の中から選択し、多クラス識別方式は {One-vs-One 方式, One-vs-Rest 方式 } の 2 種から選択する。ニューラルネットワークについては隠れ層の層数 {1、2、3} の中から、各層のユニット数 {50、100} の中から瀬選択する。ランダムフォレストについては決定木の最大の深さを $\{10, 15, 20, 25, 30\}$ から、決定木の本数を $\{100, 500, 1,000, 1,500\}$ から選択する。

チューニングの結果、SVM は誤識別数に対するペナルティ C は 10、カーネル関数は線形カーネルが、One-vs-One 方式のものが選択され、ニューラルネットワークは中間層が 3 層のもので、各中間層のユニット数はそれぞれ 1000 のものが選ばれた。ランダムフォレストの決定木の最大深さは 20、決定木の本数は 1,000 のものが選択された。

識別器ごとの予測精度を表 3 に示す。ランダムフォレストによる識別が、全ての指標において最も良い精度となったため、大規模ドライバー識別問題ではランダムフォレストを用いる。

表 3: 30 人ドライバー識別での各識別器の性能

識別器	SVM	NN	ランダムフォレスト
Recall@1	0.726	0.788	**0.885**
Recall@3	0.939	0.955	**0.981**
Recall@5	0.984	0.985	**0.997**
Recall@10	0.999	0.999	**0.999**
MRR	0.842	0.875	**0.934**

6.3 本実験の設定

本実験では、(1) 時間に関する特徴量と位置に関する特徴量のみを使って予測を行なう場合、(2) 速度と加速度に関する特徴量のみを使って予測を行なう場合、(3) 全ての特徴量を使って予測する場合の 3 つの予測モデルに (4) ベースライン手法を加えた 4 種類のモデルを比較する。識別器としては予備実験で決定されたランダムフォレストを使用する。ランダムフォレストのハイパーパラメータとして、森の中の決定木の本数、決定木の最大深さは改めて調整を行う。ベースライン手法としては、単純にデータ数の多いドライバー順にランキングして、これを予測結果とする方式を用いる。すなわち、識別対象とするドライバー数を M として、使用したデータ数で降順ソートしたドライバーを $y_1, \ldots, y_M$ とする時、ベースライン手法による予測結果は、上位から $y_1, \ldots, y_M$ となる。

ドライバーによって運転データの数は異なるため、100 個以上のデータが存在するドライバーはおよそ 100 個のデータを、100 個未満しかデータがないドライバーについてはそのユーザのデータ全てを使用する。このようにして得られた全データ中の 80%を分類器の学習に使い、20%を性能評価用のテストデータとして使う。なお、実験に用いたドライバーはデータ数に関して降順ソートした順に選択する。例えば 10 人識別であれば、ドライバー集合からデータ数の多い順に 10 人のドライバーを選択する。したがって、識別人数を増やすに従い 1 人当たりのデータ数が減っていくことに注意されたい。

識別対象とするドライバーの数は 10、100、1,000、5,000、10,000 人とした。評価指標には 6.1 節で述べた Recall@k ($k = 1, 3, 5, 10$)、平均逆順位を用いて評価を行なう。用いるデータ数は、訓練データとテストデータを合わせて、10 人識別では 1,189 件、100 人識別では 11,358 件、1,000 人識別では 107,135 件、5,000 人識別では 430,844 件、10,000 人識別では 609,279 件である。

6.4 本実験の結果

各特徴ベクトルの設計方法について、Recall@k ($k = 1, 3, 5, 10$)、平均逆順位を識別人数ごとにプロットしたグラフを示す。なお、横軸の識別人数は対数プロットしている。図 1 が Recall@1、図 2 が Recall@3、図 3 が Recall@5、図 4 が Recall@10、図 5 が MRR を示したものである。All が全ての特徴量で識別する場合、Loc and Time が位置と時間に関する特徴量で識別する場合、Speed and Acc が速度と加速度に関する特徴量で識別する場合、Baseline がベースライン手法でドライバーを識別をする場合の結果を示している。

識別人数が 10、100、1,000、5,000、10,000 人の全ての場合において、設計した特徴ベクトルを用いてドライバー識別を行なった方がベースラインよりも良い結果となり、本研究で構成した特徴量がドライバー識別に有効であることが分かった。10,000 人という大人数でのドライバー識別においても、全ての特徴量を用いて識別を行った場合平均逆順位が 0.33 という比較的高い精度を得ることができた。全ての特徴量を用いて識別を行なった場合と、位置と時間による特徴量のみを用いて識別を行なった場合には大きな精度の差はなかったが、ほぼすべての評価指標において、速度と加速度による特徴量を用いることでのわずかな精度向上が見られた。また、速度や加速度によ

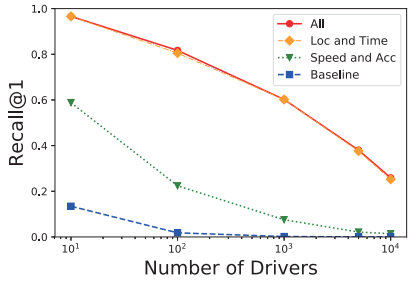


図 1: Recall@1 による各手法の比較

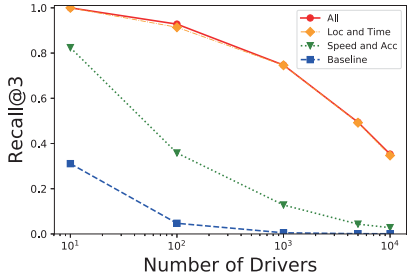


図 2: Recall@3 による各手法の比較

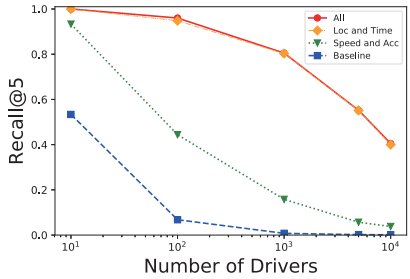


図 3: Recall@5 による各手法の比較

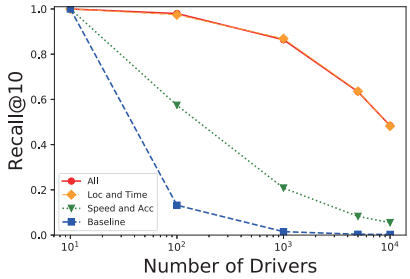


図 4: Recall@10 による各手法の比較

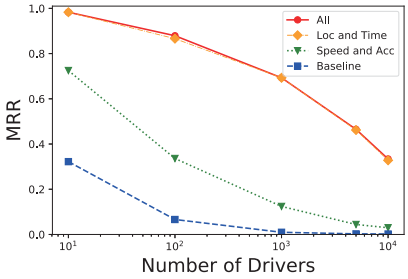


図 5: 平均逆順位による各手法の比較

る特徴量のみから識別を行なった場合に、ベースライン手法よりも良い精度を得たことからも、速度や加速度の情報は大規模ドライバー識別問題において一定の役割をもつと考えられる。とりわけ、様々な理由により位置情報の使用が制限される場合には、相対的にその重要度は高まると考えられる。また、識別対象の人数が増えるに従い、識別問題の難易度が上がるため、精度が落ちていくことも確認できる。

7. 結論

本研究では実運転データを用いた大規模ドライバー識別に取り組んだ。1 万人のドライバー識別実験では、ベースライン手法と比較して高い精度でドライバーを識別できることが示された。また、位置や時間に関する特徴量、特に位置による特徴量が大規模なドライバー識別において極めて有効であることを確認するとともに、速度や加速度もドライバー識別に一定の寄与があることを示した。

なお、本研究で使用したデータセットは、様々な車種の自動車に持ち込まれた携帯電話から収集されたデータをもとにしたものであり、速度と加速度といった特徴量は、加速性能やブレーキ性能、車両重量、トランスミッション方式などの、車種によるスペックの違いも大きく受けていると推測される。そのため、特に速度と加速度による特徴量のみによるドライバー識別問題は、車両識別問題とドライバー識別問題が混在した問題になっていると考えられる。今後はこれらを区別したより詳細な実験設定を行っていくことを予定している。

参考文献

[Fung 17] Fung, N. C., Wallace, B., Chan, A. D., Goubran, R., Porter, M. M., Marshall, S., and Knoefel, F.: Driver identification using vehicle acceleration and deceleration events from naturalistic driving of older drivers, in *Proceedings of the 12th Annual IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 33–38 (2017)

[Hallac 16] Hallac, D., Sharang, A., Stahlmann, R., Lamprecht, A., Huber, M., Roehder, M., Sosic, R., and Leskovec, J.: Driver identification using automobile sensor data from a single turn, in *Proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 953–958 (2016)

[Miyajima 07] Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., and Itakura, F.: Driver Modeling Based on Driving Behavior and Its Evaluation in Driver Identification, *Proceedings of the IEEE*, Vol. 95, No. 2, pp. 427–437 (2007)

[Van Ly 13] Van Ly, M., Martin, S., and Trivedi, M. M.: Driver classification and driving style recognition using inertial sensors, in *Proceedings of the 24th IEEE Intelligent Vehicles Symposium (IV)*, pp. 1040–1045 (2013)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

機械学習による食堂食事数予測問題への個人行動モデルの適用とその効果分析

Application of individual behavior model in machine learning based cafeteria visitor prediction model and its effectiveness analysis

松江 清高[*1] 野田 英樹[*1] 近藤 浩一[*1]
Kiyotaka Matsue Hideki Noda Koichi Kondo

[*1] 東芝インフラシステムズ株式会社
Toshiba Infrastructure Systems & Solutions Corporation

Various data about individual behavior are becoming more easily available based on IoT related technologies. In this study, a machine learning based cafeteria visitor prediction model using individual behavior models has been developed. Each individual behavior model is built up from personal information in our office building. The effectiveness analysis of the model has been studied by comparing to another prediction model using conventional Random Forest algorithm. Authors clarified that RMSE of the conventional model and the developed model were 47.0 and 44.3, respectively. These results were better than the error in the cafeteria's prediction.

1. はじめに

近年 IoT (Internet of Things) の普及に伴い，さまざまなデータが取得できるようになってきている．たとえば外食産業への適用事例では，飲食店の売上予測を行うために，飲食店舗の隣接催事場でのイベントの予想平均入場者数，天候，曜日などの情報を用いて状態空間モデルから売上を予測する手法[山口 04]などが発表されている．我々が入居するラゾーナ川崎東芝ビルは近隣の飲食店街に加え同ビル内での弁当販売やコンビニ利用が可能なことから，日々5000 名程度で変動している同ビルの在館者のうち同ビルの食堂利用者は半数程度，人間の経験に基づく食堂食事数の予測誤差は 100 食程度であり，食堂事業者より予測精度向上のニーズがあった．一方，同ビルは IoT 基盤を導入しており個人データを活用できる環境にある．個人データを食堂食事数予測に適用した事例は無いことから，筆者らは個人の食堂利用有無を推定するモデル（以降，個人行動モデルという）を適用した食堂食事数予測モデル（以降，検討モデルという）を開発した．本稿は検討モデルと従来の食堂食事数予測モデル（以降，ベースモデルという）との比較分析結果をまとめたものである．

2. 使用データと食事数の定義

ラゾーナ川崎東芝ビルの利用者はセキュリティゲートを通らないと入館，退館できない仕組みになっている．セキュリティゲート通過時に読み取る個人 ID データを利用することで入館，退館を追跡でき，任意の時刻の在館者数を把握できる．また，同ビル内の食堂には POS (Point of sale) 端末が設置され，入館者は現金に加えて個人 ID カードでの支払いも可能である．そのため POS データを利用して個人の食堂利用回数や日々の食堂食事数を取得できる．これらのデータに加えて天気情報も利用してモデルを作成する．これは個人行動が天気や気温などに依存する可能性を考慮するためである．

連絡先：松江 清高，東芝インフラシステムズ株式会社 技術企画部，川崎市幸区堀川町 72 番地 34，044-331-1365，kiyotaka.matsue@toshiba.co.jp

また，食事数とは「昼間（11 時 00 分～14 時 00 分）の食堂での支払い回数」と定義する．支払い回数は支払い金額や食事メニューに依存しないものとする．

3. 予測モデル

3.1 ベースモデル

個人行動を加味せずマクロな視点で予測するモデル（以降，ベースモデルという）を作成する．ベースモデルの作成にはランダムフォレスト（以降，RF という）[Breiman 01]を用いて，変数選択とモデル作成を繰り返すアルゴリズム[Diaz-Uriarte 06] を採用する．全説明変数を用いて RF モデルを作成した後に説明変数の重要度を算出し，その重要度の上位 80%の変数を用いて再度 RF モデルを作成する，という処理を繰り返し実施し，作成した複数のモデルから最適なモデルを選択するアルゴリズムである．これにより，直近 90 日間の食堂食事数やビル内の在館者数などの実績値から食堂利用率を算出するモデルを作成する．食堂食事数は当日の在館者数に食堂利用率を乗算して算出する．

3.2 個人行動モデルを用いた検討モデル

個人行動モデルを用いた検討モデルでは各個人を表 1 に示す 4 つのいずれかに分類して食堂食事数を予測する．支払い方法が現金の個人はベースモデルと同一手法で作成した RF 回帰モデルを作成する（分類 1）．支払い方法が個人 ID カードの個人は各個人の食堂利用状況から食堂利用有無を推定する個人行動モデル（分類 2-1～2-3）を作成する．これら分類 1 の予測値と分類 2-1～2-3 の食堂利用有りと推定した個人の人数を足し合わせることで食堂食事数を予測するモデル（検討モデル）を作成する．

分類 2-1 は食堂利用回数の少ない個人を対象にしている．データ数が少ないため，個人ごとにモデルを作成せず，個人 ID カードを利用した利用者全員のデータを使用して RF 分類モデルを作成する．同様に，分類 2-2 は食堂利用率に偏りがある個人を対象にしており，食堂利用回数または未利用回数のどちらか一方のデータが極端に少ない又は存在しない個人を対象

にしている. 分類 2-1 同様, データ数が少なく特徴を捉えたモデル作成が難しいため食堂利用率から食堂利用有無を推定するモデルを個人ごとに作成する. 分類 2-3 は上記以外の個人を対象にしたモデルであり個人ごとに RF 分類モデルを作成する.

モデル 2〜4 を分類している表 1 の数値について, 食堂利用回数 10 回とは 10 営業日に相当し, 個人行動が曜日に依存すると想定した場合に祝日等を考慮するとモデルを作成する上での必要最低限のデータ数と考えている. また, 食堂利用率 0.2 や 0.8 は, 一ヶ月間(営業日換算で 20 日間)を想定した場合, 食堂利用率 0.8 超は食堂未利用回数が 3 日間以下を意味し, 週 5 日間, 食堂を利用するケースが必ず存在することを意味する. 食堂未利用データが少ない上に行動に偏りのある個人は特徴抽出が難しい個人と考えられる.

表 1 検討モデルの詳細

分類	支払い方法	訓練データ	食堂利用回数 N	食堂利用率 α
1. 現金払い (RF 回帰モデル)	現金	現金 全員	−	−
2-1. 訓練データ少 (RF 分類モデル)	個人 ID	個人 ID 全員	N＜10	−
2-2. 訓練データ偏り (食堂利用率モデル)	個人 ID	個人 ID	10≦N	α＜0.2, 0.8＜α
2-3. 上記以外 (RF 分類モデル)	個人 ID	個人 ID	10≦N	0.2≦α≦0.8

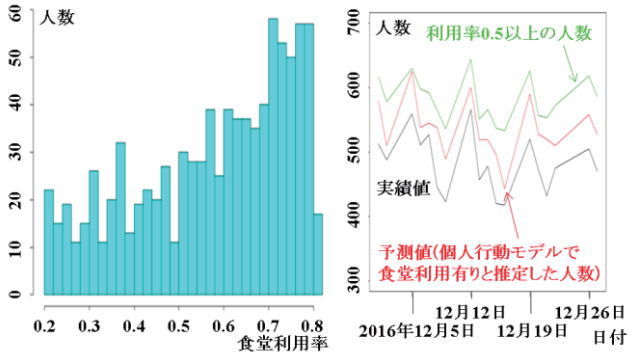
4. 評価結果と効果分析

ベースモデルおよび検討モデルを R 言語の汎用ライブラリ ranger[Wright 15]を用いて作成し食堂食事数の予測結果を評価した. 両モデルの RMSE (Root Mean Square Error) と MAPE (Mean Absolute Percentage Error) の値を表 2 に示す. ベースモデルは RMSE で 47.0 となり, 人間の経験に基づく予測値の 100 食と比較して良好な結果が得られた.

表 2 各モデルの RMSE と MAPE

	RMSE	MAPE
ベースモデル	47.0	1.5
検討モデル	67.2	2.2

しかし検討モデルはベースモデルと比較して RMSE で 20.2, MAPE で 0.7 悪化する結果となった. そこで精度悪化の理由を探るため個人の動きを調べた. 図 1 に 2016 年 12 月 1 日の個人行動モデル適用者の食堂利用率のヒストグラムと個人行動モデルの予測値, 実績値, 利用率 0.5 以上の人数の推移を示す.



(a)食堂利用率ヒストグラム (b)利用人数推移(予測と実績)

図 1 個人行動モデル適用者の分析結果

図 1(a)の食堂利用率ヒストグラムより食堂利用率 0.5 以上の人数のほうが 0.5 未満の人数より多い. また図 1(b)を見ると全日程において予測値が実績値を上回っており, さらに利用率 0.5 以上の人数が予測値を上回っている. つまり食堂利用率の高い個人が多いために, 食堂利用有りと推定する個人が増えてしまい, 結果として全体の予測食事数を押し上げていると考えられる. そこで, この推定値と実績値の差分を学習し予測モデルを補正することで予測精度の向上を試みた. 一ヶ月間の平均残差を個人行動モデルに加算して補正したモデルの評価結果を図 2 と表 3 に示す. その結果, RMSE で 22.92, MAPE で 0.8 改善し, ベースモデルの予測精度を若干上回った.

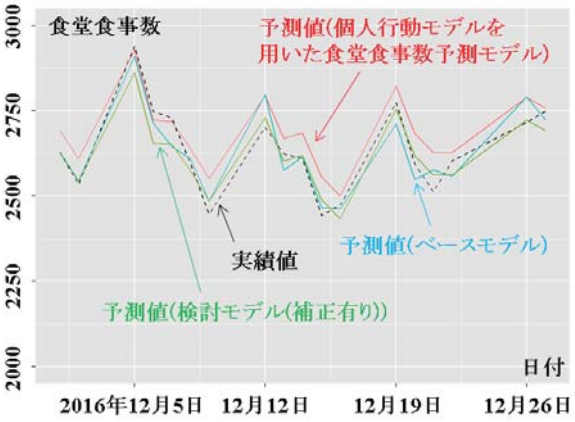


図 2 食堂食事数の推移(予測値と実績値)

表 3 補正を適用した場合の RMSE と MAPE

	RMSE	MAPE
検討モデル(補正有り)	44.3	1.4

5. おわりに

食堂食事数予測に個人行動モデルを適用し, 以下の 3 点を明らかにした.

- マクロな視点で予測したベースモデルによる予測値は人間の経験に基づく予測結果より良好な結果であること
- 食堂利用率ヒストグラムが 0.5 を中心に正規分布していない場合には個人行動モデルに補正項が必要なこと
- 個人行動モデルに補正項を付加することでベースモデルの予測精度を若干上回る結果が得られること

更なる精度向上を目指し, 個人行動モデル作成時の数値最適化(食堂利用回数および食堂利用率によるモデル分類時の最適化)や食堂利用率ヒストグラム形状に基づいた個人行動モデルの補正方法の検討などの課題を解決していく.

参考文献

[山口 04] 山口類, 土屋映子, 樋口和之: 状態空間モデルを用いた飲食店売上の要因分解, 日本オペレーションズ・リサーチ学会, 2004.

[Breiman 01] Breiman, L.,: Random Forest, Machine Learning 45, 2001.

[Diaz-Uriarte 06] Diaz-Uriarte, R., and Andres, S. A., Gene selection and classification of microarray data using random forest, BMC Bioinformatics, January 6, 2006.

[Wright 15] Wright, M. N. and Ziegler, A., ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, arXiv: 1508.04409v1, 18 Aug 2015.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Mastodonにおける自己評価維持モデルに基づいたフォロー先推薦機構の試作

A Preliminary Implementation of Followee Recommendation Mechanism with Self-evaluation Maintenance Model on Mastodon

松下拓真 Takuma Matsushita　福田直樹 Naoki Fukuta

静岡大学情報学部 Faculty of Informatics, Shizuoka University
静岡大学学術院情報学領域 College of Informatics, Academic Institute, Shizuoka University

Cyber-bullying and suicide are one of the most serious issues on the SNS-based communication. Mastodon is a destributed SNS and it doesn't have own protecting system for these issues. We use self-evaluation to measure that our approach is working or not. In this paper, we propose preliminary idea of a recommending algorithm to prevent decreasing of self-evaluation and give a simulation-based comparison with another recommending algorithm which has been used on a major microblogging service.

1. はじめに

SNS におけるいじめや自殺問題への対策は重要な課題であり，対策方法の検討がされている．中央管理型の SNS である Twitter では，ユーザ間のトラブルを防止する機能として，危険な単語を排除する機構を備えている他，自傷行為をしているアカウントを報告するなどの機能が存在する．分散型 SNS である Mastodon ではサーバが個人によって管理され，一部のボランティアなどにより Twitter のような対策を行なっている．このような対策には管理が行き届かない可能性のほか，対策そのものが終了する可能性も考えられる．このような対処方法では，コミュニケーションから現実世界にまで踏み込んだような事件が起こる可能性については対処されていないという問題がある．Facebook のような匿名性の低い SNS と比べ，Mastodon は Twitter のような比較的匿名性の高い SNS である．本研究では，周りとの繋がりを維持，向上するための支援を行うため，自己評価維持モデルに基づいた支援を行う．社会心理学で提唱された自己評価維持モデルでは，他人との関係が自己評価に大きな影響を与えるとされており，agent-based model による自己評価計算モデルで自己評価が低下する状況では自殺が増加することが指摘されている [Liu 17]．本研究では分散型 SNS のひとつである Mastodon に注目し，自己評価の低下を防ぐためのユーザ支援機構を試作する．

2. 背景

自己評価維持モデルとは，人間が自己評価を維持もしくは向上することを動機づけられていること及び他人との関係が自己評価に大きな影響を与えるということを述べたモデルである [Tesser 88]．自己評価とは主観的な評価であり，他者から見た自分ではなく，自分から見た自分を評価したものである．また他人との関係の喪失が自殺に影響を与えることが，Durkheim らによって述べられている [Durkheim 52]．この自己評価維持モデルと Durkheim の法則に基づき，gent-based model を用いて自己評価を計算する機構が提案されている [Liu 17]．Liu らのモデルでは繋がりの強さを用いて，自己評価を計算しており，SNS での繋がりの強さを求めるための研究は Gilbert らが行なっている [Gilbert 12]．Gilbert らのモデルでは，繋がりの強さとしてコミュニケーションを取っている期間や TF-IDF による特徴語の類似度，感情を意味する単語などから繋がりの強さを求めている．SNS における自殺問題への対策は様々なアプローチで行われている．[Abboute 14] では Twitter の投稿を分類し，自殺の危険を自動的に検知することを行なっており，[Won 13] では Twitter での自殺率とその特性を線形回帰によって分析している．また推薦手法において，Daly らは Friend-of-Friend, Content Matching, Content-plus-Link, SONAR の 4 つの異なる推薦手法を比較し，人口を増やす目的では Friend-of-Friend，長い期間において繋がりを維持するのであれば Content Matching を使用するべきだとしている [Daly 10]．また Friend-of-Friend では'rich getting richer' と呼ばれる，推薦手法によって，近傍するユーザを多く持つユーザほど多くのユーザを獲得することがわかっており，ネットワークの価値が低下すると述べている．Twitter では Who-to-Follow システムの推薦手法として Friend-of-Friend が使用されている [Su 16]．

2.1 自己評価維持モデル

Liu らのモデルではソーシャルネットワーク $G = (V, E, \omega)$ の V を agent，E をエージェント間の繋がりとすることで以下のように表現する [Liu 17]．

$$E \subseteq V^2 \backslash \{(u, u) | u \in V\}$$

ω はそれぞれの E に対する繋がりの重み $E \to [0, 1]$ であり，この ω が高ければ高いほど強い繋がりを持つ．例えば $\omega(u, v) = 1$ はとても強い繋がりを意味し，$\omega(u, v) = 0$ はとても弱い繋がりを意味する．このようなネットワークから自己評価を求めたいノードを ego とし，そのノードに対する ego-network を設定する．ego-network では，あるノードと隣接するそれぞれのノード及びエージェント間のエッジ E と ω を考慮し，そこから自己評価 ϵ を求める．このときの自己評価 ϵ は ego-network を隣接行列で表現した際の ego に対する固有ベクトル中心性である．ここで固有ベクトル中心性を求める際の ego が持つ self-loop は，ego の先天性のストレス耐性とする．またネットワーク G が変化した際，自己損失 γ を以下のように定義する．

$$\gamma(v) = 1 - \frac{\epsilon'(v)}{\epsilon(v)}$$

連絡先: 松下拓真，静岡大学情報学部情報科学科，
〒 432-8011 静岡県浜松市中区城北 3-5-1，
takuma.matsushita.14@shizuoka.ac.jp

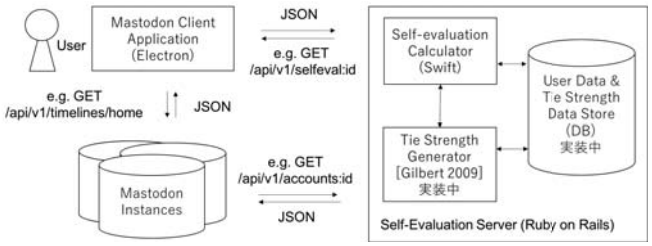


図 1: 試作機構の構成図

この自己損失は自分がある期間で，どの程度自己評価が低下したのかを表す指標となる．

2.2 Mastodon

Mastodon は分散型 SNS と呼ばれる SNS の一つである *1．分散型 SNS には複数の実装が存在する．近年実装されたもので Mastodon に近いものとしては GNU Social と呼ばれる実装が存在する *2．Mastodon では，個人がインスタンスと呼ばれるサーバを所有することができ，管理することが可能である．インスタンスにはそれぞれ管理者が目的を設置したりすることで，それぞれが特色を持ち，そこに所属している人たちがコミュニケーションをとっている．ユーザはインスタンスに所属し，インスタンス内のユーザを Twitter や Instagram などに見られるフォロー機能を活用することで，インスタンス内のユーザの投稿した文章を読むことができる．また Twitter とは異なり，インスタンス内の全ての投稿を観測できるストリームが存在する．さらに連合と呼ばれるストリームも存在し，インスタンスの管理者が協定を結んでいる場合，連合に参加している他のインスタンスの文章を観測できる．本稿執筆時点における Mastodon のインスタンス数は，1390 であるとされる *3．この API から得られるインスタンスの情報は，Mastodon 開発者のインスタンスと連合を結んでいるインスタンスのみであり，実際にはさらに多くのインスタンスが運営されていると考えられる．Mastodon には文字制限が 500 文字となっており，240 文字である Twitter とは大きく異なり，ユーザの投稿に表現の幅が広いと考えられる．またセンシティブな内容には，フィルタリングがかけられており，表示するかどうかは投稿ごとにユーザに委ねられる．また電話帳などのインポート機能がせず，現実世界での繋がりよりもインターネット上での繋がりが多いことが予想される．

3. Mastodon への適用

図 1 に本機構の構成図を示す．Mastodon インスタンスに改良を加え防止機構を備えるには，全てのインスタンスに新たな機構を挿入しなければならないため，本機構ではクライアントサイドでの提案を行う設計にしている．ただしユーザの情報やユーザに何らかの提案を行う機能に関しては，サーバサイドで計算することが望ましいため，サーバサイドとクライアントサイド両方の実装で実現する．サーバサイドで自己評価を計算することには，プライバシーの観点，およびサーバサイドに計算結果をストアすることで，計算にかかるコストを削減できるためである．特にプライバシーに関しては，他人にユーザの自己評価が直接見えてしまうことは大きな問題であり，自己評価の計算結果ではなく，自己評価によって得られた結果を返す仕様とする．サーバサイドではユーザの繋がりの重み計算機構と自己評価計算機構及び，あるユーザをフォローした際の将来的な自己評価の変化を計算する機構を備える．クライアントサイドでは基本的な Mastodon クライアントアプリケーションとしての機能のために Mastodon API と通信が可能であり，前述した自己評価管理サーバと通信することで支援を行う．自己評価を計算する機構を使用するためには，前述したように自己評価を計算したいユーザとその周りの人々との繋がりの重みが必要となる．SNS の情報からユーザ間の繋がりの重みを計算する手法については Gilbert らによって提案されており [Gilbert 12]，この計算手法によって Mastodon 上での自己評価の計算が実現可能となる．これらの機能は後から標準の機能としても組み込めるよう，既存のライブラリと親和性が高い設計にしている．図 2 にユーザインタフェースを示す．

*1 https://github.com/tootsuite/mastodon
*2 http://www.gnu.org/software/social/
*3 https://instances.social



図 2: ユーザインタフェースの一部

4. 推薦手法

自己評価を向上させるには他のユーザとの新たな繋がりを増やすということが考えられ，その際に全体としての自己評価が向上するようなこと以外に，極度に自己評価が低いユーザが少ないことが望ましいと考えられる．本研究が支援を対象とするユーザは，被験者実験への参加を求めることが適切でないと考えられるような具体的かつ切迫した問題を持った立場となっている場合もあり，もし被験者実験を行った場合には，本機構のなんらかのアクションの副作用によって致命的な帰結をもたらす可能性がある．そのような可能性を持つ被験者実験には倫理的な面から避けるべきである．そのため，ここでは理想的な推薦手法をシミュレーションによって検討する．Mastodon におけるネットワーク構造は判明していないため，ランダムネットワークとスケールフリーネットワークにおけるシミュレーションを行い，その上で推薦手法を適用し，自己評価を計算することで，アルゴリズムの選定を行う．

推薦手法の 1 つである Friend-of-Friend には”rich get richer” と呼ばれる，多くのフォロワーを持っているユーザがこの推薦手法によって，より多くのフォロワーを多く獲得するという傾向がある．Twitter はこの傾向に対して，フォロワーが少ない人もこの推薦手法によってフォロワーを獲得していると述べている [Su 16]．

先述したように，Mastodon にはいじめを防止する機構が存在しないため，フォロワーが少ない人の自己評価が低下し

Algorithm 1 Week Helper

$min \leftarrow 0$
for $(u, v) \in V^2, u = v$ **do**
  for $(v, s) \in V^2, v = s$ **do**
    Save s, if s's self-evaluation $< min$
    $min \leftarrow s$'s self-evaluation
  end for
end for
Add directed edge (u, s) with probability
Add directed edge (s, u) with probability

てしまうことを避けるために，推薦手法として自己評価の低い人の自己評価が向上するようなものであることが望ましい．Mastodon ではインスタンス内にいる全てのユーザを API によって取得することはできず，自己評価が低い人を全体の中から探し出すことはできない．本研究ではユーザのフォロワーを遡って推薦ユーザを決定を行う．具体的には Friend-of-Friend アルゴリズムにおけるユーザの選定過程に自己評価を考慮することで，対処を行う．Friend-of-Friend は 2 ステップのランダムウォークを行い，フォロワーのフォロワーから対象者を選ぶアルゴリズムである．ここで提案アルゴリズムは，フォロワーのフォロワーから自己評価が低い人を選び，推薦を行う．提案アルゴリズムが自己評価を利用しているのは，”rich get richer” に対して，フォロワーが少ないユーザのみを推薦対象に選ぶことが自己評価の観点から危険性が考えられるためである．例えばフォロワーが多くてもフォロワーとの繋がりが弱く，自己評価が低いユーザを推薦することでこのようなユーザを推薦対象にすることができる．以下にアルゴリズムを示す．

5. シミュレーション

まず，スケールフリーネットワークである Barabasi-Albert モデル [Albert-L á szl ó 16] と，ランダムネットワークである Erdos-Renyi モデル [Watts 98] を用い，無向グラフを生成する．スケールフリーネットワークでは，べき乗則に倣ったネットワークが生成され，リンクの数が少ないノードが大量に生成されるが，リンクを多く持つようなノードがわずかに生成される．ランダムネットワークでは，ポアソン分布にしたがったネットワークが生成され，ほとんどのノードが同じ数のリンクを持ち，繋がりの数が大きなノードがほとんど生成されない．このような違いを持つ 2 つのネットワークから，Liu らのアルゴリズムを用いて有向グラフを生成した [Liu 17]．その後，生成された有向グラフに対して，Friend-of-Friend と提案アルゴリズムである Week Helper を適用し，比較を行なった．Friend-of-Friend では，あるノードに対して向かっているエッジを 2 回遡って取得したノードをランダムに取得する．Week Helper では，Friend-of-Friend のように 2 回遡って取得したノードから，自己評価が最も低いノードを選択する．このように選択されたノードに対して，エッジを生成する．この際エッジの重みは 0.5 とした．

表 1，2 にシミュレーションによって得られた，各手法の適用後の自己評価値の平均，分散，および最小値を示す．表中の FoF は Friend-of-Friend アルゴリズムによる推薦によって新たなエッジを接続したものであり，WH は Friend-of-Friend アルゴリズムによって見つかった人たちの中でも自己評価の低いノードを選択したものである．計算結果は 10 回試行したものを平均したものである．平均値においてはどちらの BA，

表 1: p = 1 のときの Erdos-Renyi model におけるシミュレーション結果

	Friend-of-Friend	Week Helper
中央値	0.031592551	0.031534616
平均	0.031586206	0.031566056
最小	0.03020614	0.030151507
最大	0.033002222	0.035785367
標準偏差	0.000479481	0.000593782
分散	2.30E-07	3.53E-07

表 2: p = 1 のときの Barabasi-Albert model におけるシミュレーション結果

	Friend-of-Friend	Week Helper
中央値	0.015172684	0.015458042
平均値	0.01945937	0.019557755
最小	0.004880366	0.006467888
最大	0.351715418	0.372655329
標準偏差	0.019872288	0.020100777
分散	0.000394908	0.000404041

ER ネットワークの推薦後は僅かな変化であった．また BA ネットワークでは分散を抑えつつ最小値を向上させることができたことが読み取れる．ER ネットワークでは最小値をわずかに向上させたが，分散が僅かに大きくなった．このときの Barabasi-Albert モデルにそれぞれの推薦手法を適用した際の，自己評価の値をある範囲で区切った際とその人数は図 3 のようになる．Friend-of-Friend よりも提案アルゴリズムのほうが，自己評価が極端に低いノードの数が減っていることが読み取れる．

さらに，この推薦手法がネットワーク中で，ユーザの自己評価の低下を防ぐことを示すためのシミュレーションを行なった．各ステップごとに，自己評価が低くなったユーザとランダムに選択されたノードを取り除き，Friend-of-Friend と提案アルゴリズムを適用する．ステップは 100 回繰り返し行なった．図 4 に結果を示す．どちらの推薦手法も，自己評価が低いノードがある程度減った後に，ノード数の減少が少なくなっている．しかし，Friend-of-Friend アルゴリズムの方がノード数の減少が激しく，減少の変化が大きいステップが多いことが観測できる．

6. 結論と今後の課題

本研究では Mastodon における自己評価維持モデルを利用した推薦手法を提案した．また自己評価計算モデルの Mastodon への適用を行い，繋がりの重みの計算などで課題が残るものの適用可能性を示した．使用される推薦手法はシミュレーションによって確認し，推薦手法のパラメータを変更した上で、どの程度提案手法が自己評価に影響を及ぼすのかを確認した．また提案手法は長期的な適用を考慮した際に効果が発揮されることが確認された．今後の課題として，繋がりの重みおよび自己評価計算モデルの信頼性に関わる問題と，過去の研究で明らかにされてきた社会心理学の事実を確認することがあげられる．しかし倫理的な観点からこれらはシミュレーションによって十分に確認されることが必要となる．推薦手法の改良は今後の課題であり，より複雑な状況でのシミュレーションが必要である．また自己評価が低下しないような状況をメカニズムデザイ

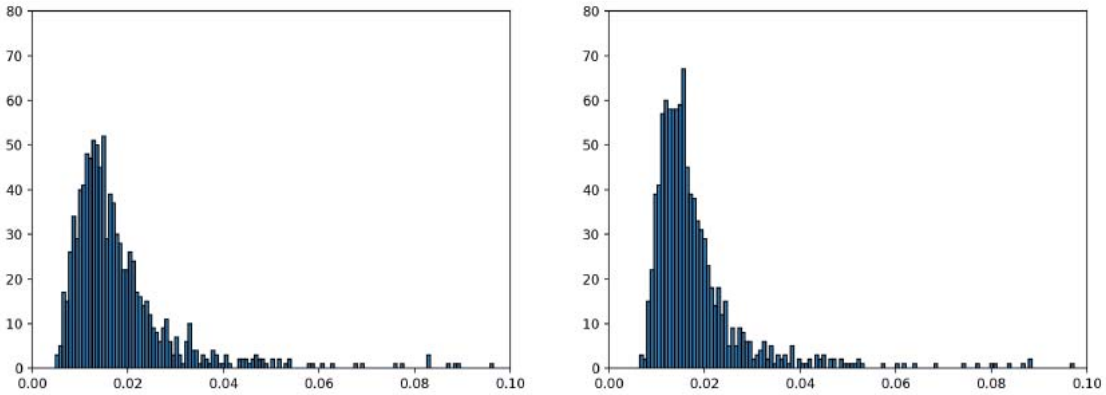


図 3: Barabasi-Albert モデルにおける推薦手法の比較．左側が Friend-of-Friend．右側が提案アルゴリズム．

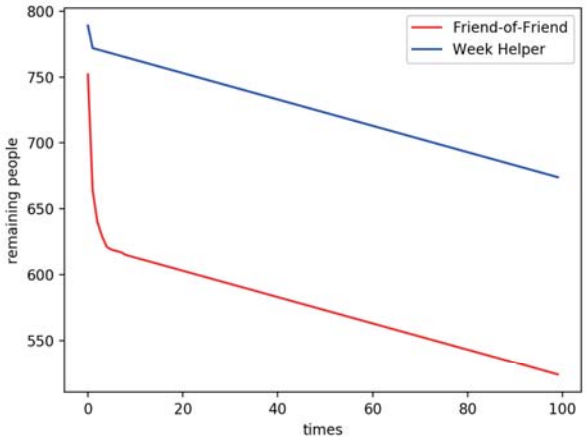


図 4: 100 ステップまでノードを取り除く処理を実行した結果

ンなどの手法から明らかにすることができないか模索することも重要な課題である．さらにインスタンスが管理者によって削除されたり，過疎化してしまうことへの対策も課題として挙げられる．

参考文献

[Abboute 14] Abboute, A., Boudjeriou, Y., Entringer, G., Azé, J., Bringay, S., and Poncelet, P.: Mining Twitter for Suicide Prevention, in Métais, E., Roche, M., and Teisseire, M. eds., *Natural Language Processing and Information Systems*, pp. 250–253, Cham (2014), Springer International Publishing

[Albert-L á szl ó 16] Albert-L á szl ó Barab á si, M á rton P ó sfaiF*Network science*, Cambridge University Press, Cambridge (2016)

[Daly 10] Daly, E. M., Geyer, W., and Millen, D. R.: The Network Effects of Recommending Social Connections, in *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pp. 301–304, New York, NY, USA (2010), ACM

[Durkheim 52] Durkheim, E., Simpson, G., and SPAULDING, J. A.: *Suicide. A Study in Sociology... Translated by John A. Spaulding and George Simpson. Edited with an Introduction by George Simpson*, London (1952)

[Gilbert 12] Gilbert, E.: Predicting Tie Strength in a New Medium, in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pp. 1047–1056, New York, NY, USA (2012), ACM

[Liu 17] Liu, J., Li, L., and Russell, K.: What Becomes of the Broken Hearted?: An Agent-Based Approach to Self-Evaluation, Interpersonal Loss, and Suicide Ideation, in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '17, pp. 436–445, Richland, SC (2017), International Foundation for Autonomous Agents and Multiagent Systems

[Su 16] Su, J., Sharma, A., and Goel, S.: The Effect of Recommendations on Network Structure, in *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pp. 1157–1167, Republic and Canton of Geneva, Switzerland (2016), International World Wide Web Conferences Steering Committee

[Tesser 88] Tesser, A.: Toward a Self-Evaluation Maintenance Model of Social Behavior, Vol. 21 of *Advances in Experimental Social Psychology*, pp. 181 – 227, Academic Press (1988)

[Watts 98] Watts, D. J. and Strogatz, S. H.: Collective dynamics of 'small-world' networks, *Nature*, Vol. 393, No. 6684, pp. 440–442 (1998)

[Won 13] Won, H.-H., Myung, W., Song, G.-Y., Lee, W.-H., Kim, J.-W., Carroll, B. J., and Kim, D. K.: Predicting National Suicide Numbers with Social Media Data, *PLOS ONE*, Vol. 8, No. 4, pp. 1–6 (2013)

[2N1] [General Session] 10. Vision / Speech

座長:辻川 剛範（NEC）

Wed. Jun 6, 2018 9:00 AM - 10:20 AM Room N (2F Sakurajima)

[2N1-01] Expressive Speech Synthesis through modeling the variety of expressions by Variational Autoencoder

○Kei Akuzawa[1], Iwasawa Yusuke[1], Matsuo Yutaka[1] （1. University of Tokyo）

9:00 AM - 9:20 AM

[2N1-02] Response Timing Estimation Using Dialogue Situations for Spoken Dialogue Systems

○Genki Akai[1], Ryu Takeda[1], Kazunori Komatani[1] （1. The Institute of Scientific and Industrial Research, Osaka University）

9:20 AM - 9:40 AM

[2N1-03] Incremental estimation of user's turn-taking state

○Shinya Fujie[1,2], Katsuya Yokoyama[2], Tetsunori Kobayashi[2] （1. Chiba Institute of Technology, 2. Waseda University）

9:40 AM - 10:00 AM

[2N1-04] Semantic-Usage Classification of Mimetic Words using Sound Symbolism

○Natsumi Takeuchi[1], Daiki Urata[1], Tsuyoshi Nakamura[1], Masayoshi Kanoh[2], Koji Yamada[3] （1. Nagoya Institute of Technology, 2. Chukyo University, 3. Institute of Advanced Media Arts and Sciences）

10:00 AM - 10:20 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

変分自己符号化器を用いた表現の多様性のモデル化による表現豊かな音声合成

Expressive Speech Synthesis through modeling the variety of expressions by Variational Autoencoder

阿久澤 圭 *1 Kei Akuzawa　岩澤 有祐 *1 Yusuke Iwasawa　松尾 豊 *1 Yutaka Matsuo

*1東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

Recent advancements in the deep autoregressive generative modeling improve the performance of speech synthesis (SS). However, how to equip the expressiveness into the deep autoregressive based SS-system is an open issue due to the lack of ability to model the global characteristics of speech (such as speaker individualities or speaking styles). In this paper, we propose a model called VAE-Loop, which integrates variational autoencoder (VAE) with VoiceLoop: one of the autoregressive based speech synthesis models. Unlike the traditional SS with autoregressive modeling, the proposed method explicitly model the global characteristic of speech by VAE, enabling control of the expressiveness of the synthesized speech. Experiments on VCTK and Blizzard2012 showed that VAE helps VoiceLoop to generate higher quality speech and control expressions through learning the global characteristics.

1. はじめに

我々は自然な会話の中で様々な感情，話し方，話者の年齢や性別 (以下話者性と呼ぶ) などによって特徴付けられた表現豊かな音声を発する．表現の豊かさは，近年深層学習を基盤とした音声合成モデルが向上させた自然さと明瞭さに加えて，音声合成システムが備えるべき性質の一つとされている (例えば [Eyben 12])．表現豊かな音声合成を実現するための典型的な手法は感情や話し方，話者性のラベルを用いることである (例えば [Luong 17]) が，本研究はそうした教師ラベルなしでの表現豊かな音声合成に取り組む．なぜなら，表現豊かな音声は動画投稿サイトやオーディオブックからの入手が容易であるが，その音声に対応するラベルを得るにはコストがかかるし，ヒューリスティックに定めた感情や話し方のカテゴリを信頼できない場合もある．しかし，ラベルなしでの表現豊かな音声合成は一つのテキストに対して数多くの正解音声が存在するという困難な逆問題に直面する．

近年音声合成全般及びラベルなしでの表現豊かな音声合成において，自己回帰生成モデルが重要な役割を果たしている．例えば [Oord 16] は，既存の統計的または波形接続型の音声合成システムから音声の自然さと明瞭さを向上させた．また，自己回帰モデルを用いた sequence-to-sequence モデル (seq2seq モデル) の VoiceLoop[Taigman 18] は，youtube から入手した多様な韻律を含む音声を用いて音声合成モデルを訓練できることを初めて示した．

しかしラベルなしでの表現豊かな音声合成に自己回帰生成モデルを用いることは 2 つの問題を抱えている．第一に，自己回帰生成モデルはラベルなしでデータの大域的な特徴を学習するのが不得手である (例えば [Gulrajani 17])．話者性や感情などが発話を大域的に特徴づけていることを考えれば，自己回帰生成モデルも前述の逆問題に十分に対抗できない可能性がある．第二に，自己回帰生成モデルは変分自己符号化器 (Variational Autoencoder, VAE)[Kingma 14] といった他の深層生成モデルと違って潜在変数を持たないので，生成する音声の特徴をコントロールすることが難しい．

そこで本研究では前述の自己回帰生成モデルの問題を克服し，自然さ，明瞭さに加えて表現の豊かさを備えた音声合成をラベルなしで実現するために，VoiceLoop を VAE に組み込んだモデル VAE-Loop を提案する．VAE は話者性といった音声データの大域的な特徴の学習に有効であることが近年示されてきた (例えば [Hsu 17]) が，それを音声合成に適用することを提案したのは，我々の知る限り本研究が初めてである．VAE-Loop では，VAE を用いて獲得した潜在変数を自己回帰生成モデルへの入力とすることで，データの大域的特徴を組み入れた生成が行える．また話者性や話し方の潜在表現を学習することで，生成する音声の特徴を柔軟にコントロールできる．

実験ではまず，VoiceLoop が VAE によって獲得した大域的な特徴を利用することで困難な逆問題を緩和し，より高品質な音声を生成できることを示す．加えて，獲得した潜在変数を用いて VAE-Loop が話者性や話し方をコントロールする能力を持つことを示す．

本研究の貢献は，自己回帰生成モデルをラベルなしでの表現豊かな音声合成に適用する際の問題点を克服するために，VAE が獲得した大域的な特徴を利用することを提案したことである．この手法により、高コストなラベルありデータが得られないより現実的な問題設定においても，高品質で多様な表現を持つ音声の生成が可能になる．

2. 関連研究

深層学習を基盤にした seq2seq モデルによる音声合成モデルは，テキストから音響特徴量を直接予測する単純な構造を持ち，また自然で明瞭な音声を生成することが示されている．(例えば [Wang 17b][Taigman 18]) 特に VoiceLoop[Taigman 18] は多様な韻律を含む youtube から入手した音声を用いて訓練できることが示されているので，ラベルなしでの表現豊かな音声合成に有効であると考えて提案モデルの一部に組み入れた．

また [Taigman 18] や [Wang 17a] は，話し方を教師なしで学習しコントロールできる seq2seq モデルを提案した．これらの手法と違い，本研究が用いる VAE は背後に潜在表現の明確な定式化があり，話し方の潜在表現を確率分布として陽に学習することができる．その結果，簡単で柔軟な話し方のコント

連絡先: 阿久澤圭，東京大学大学院工学系研究科技術経営戦略学専攻，〒 113-8656 東京都文京区本郷 7-3-1，akuzawa-kei@weblab.t.u-tokyo.ac.jp

ロールが可能になる．

seq2seq モデルに限定しないラベルなしでの表現豊かな音声合成の研究には [Eyben 12] などがある．こうした研究では，クラスタリングなどの教師なし学習の手法で獲得した感情の表現を訓練に用いることが多い．これらの手法では表現の学習と音声合成モデルの訓練を別々に行う必要があるが，提案モデルは一括で訓練を行うことができる．

VAE を音声の生成に利用した研究には [Hsu 17] などがある．これらの研究は VAE によって音声データが持つ話者性などの特徴を潜在変数として獲得できることを示した．これらの研究と異なり，本研究では音声合成を目的としてテキストによる条件付けを行う．また音声以外の領域では，本研究と同様に VAE と自己回帰生成モデルを組み合わせる研究が多くある (例えば [Gulrajani 17])．

3. モデル

本節では，まず提案モデルの基盤となる VAE と VoiceLoop について解説し，その後提案モデルについて述べていく．

3.1 Variational Autoencoder

VAE は Kingma らによって提案された深層生成モデルである [Kingma 14]．VAE では観測変数 $\boldsymbol{x}$ と潜在変数 $\boldsymbol{z}$ のデータ生成過程を，確率密度分布のモデルパラメータ $\boldsymbol{\theta}$ を用いて $\boldsymbol{z} \sim p(\boldsymbol{z}), \boldsymbol{x} \sim p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})$ のように定める．すると，真の事後分布 $p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x})$ を近似した分布 $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$($\boldsymbol{\phi}$ はモデルパラメータ) を用いて，周辺尤度の下界 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x})$ が以下のように求まる．

$$\begin{aligned}\log p_{\boldsymbol{\theta}}(\boldsymbol{x}) &= D_{KL}(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})||p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x})) + \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x}) \\ &\geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x}) \\ &= -D_{KL}(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})) + \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})]\end{aligned}$$

そして訓練では，下界 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x})$ が最大になるようにパラメータ $\boldsymbol{\theta}, \boldsymbol{\phi}$ について最適化を行う．本稿では $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$ をエンコーダー，$p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})$ をデコーダーと呼ぶ．

3.2 VoiceLoop

VoiceLoop は，Taigman らによって提案された seq2seq モデルである [Taigman 18]．$\boldsymbol{x}_1, ..., \boldsymbol{x}_T$ を予測したい音響特徴量の可変長の系列とする．VoiceLoop は条件付き独立性

$$p_{\boldsymbol{\xi}}(\boldsymbol{x}|\boldsymbol{c}) = \prod_{t=1}^{T} p_{\boldsymbol{\xi}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{c}) \quad (1)$$

$$p_{\boldsymbol{\xi}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{c}) \sim \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_{\boldsymbol{\xi}}(\boldsymbol{x}_{1:t-1}), \boldsymbol{I}) \quad (2)$$

を利用して，各タイムステップ $t \in T$ ごとに $p_{\boldsymbol{\xi}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{c})$ を逐次的に推定する．ここで (2) 式は，モデルパラメータ $\boldsymbol{\xi}$ とタイムステップ t 以前の音響特徴量 $\boldsymbol{x}_{1:t-1}$ によって決まる変数 $\boldsymbol{\mu}_{\boldsymbol{\xi}}$ を平均，単位行列を分散とするガウス分布に確率変数 $\boldsymbol{x}_t$ が従うことを仮定している．

$p_{\boldsymbol{\xi}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{c})$ の推定は以下の手順に従う．VoiceLoop はバッファと呼ばれる潜在変数 $S \in \mathbb{R}^{d \times k}$ を持つ．S はそれぞれが d 次元の変数である $S[1]...S[k]$ から成り立っている．それぞれのタイムステップ t で，バッファは次のように変化する．

$$\begin{aligned}S_t[i+1] &= S_{t-1}[i] \quad if \;\; 1 \leq i < k \\ S_t[1] &= u\end{aligned}$$

ここで u は，テキスト情報 $\boldsymbol{c}_t$，音響特徴量 $\boldsymbol{x}_{t-1}$，話者の埋め込みベクトル s，およびバッファ自身を用いて，以下のように決まる変数である．

$$\begin{aligned}C_t &= [\boldsymbol{c}_t + tanh(F_u(s)), \boldsymbol{x}_{t-1}] \\ u &= N_u([S_{t-1}, C_t])\end{aligned} \quad (3)$$

そして，VoiceLoop はバッファ S_t と話者の埋め込みベクトル s を用いて音響特徴量の予測値 $\hat{\boldsymbol{x}}_t$ を以下のように出力する．

$$\hat{\boldsymbol{x}}_t = N_o(S_t + F_o(s)) \quad (4)$$

ただし，F_u, N_u, F_o, N_o はそれぞれ全結合層，または全結合層と活性化関数を多層重ねたものである．また，$\hat{\boldsymbol{x}}_t$ は (2) 式の $\boldsymbol{\mu}_{\boldsymbol{\xi}}(\boldsymbol{x}_{1:t-1})$ と同値である．

3.3 提案モデル: VAE-Loop

VAE-Loop は，VAE を補助特徴量 $\boldsymbol{c}$ で条件づけたモデル $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{c})$ を出発点とする．本研究においては $\boldsymbol{x}$ が音響特徴量，$\boldsymbol{c}$ が音素に対応する．そして，その際の尤度の下界は以下のように書ける．

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x}, \boldsymbol{c}) = &-D_{KL}(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})) \\ &+ \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{c})]\end{aligned} \quad (5)$$

ここで，本研究では簡単化のために $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c}) = q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$ と仮定した．そして，事前分布 $p(\boldsymbol{z})$，近似事後分布 $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$ はそれぞれ以下の正規分布に従うと仮定する．

$$p(\boldsymbol{z}) \sim \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I}), \;\; q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_{\boldsymbol{\phi}}, \boldsymbol{\sigma}^2_{\boldsymbol{\phi}}\boldsymbol{I})$$

ここで，$\boldsymbol{\mu}_{\boldsymbol{\phi}}, \boldsymbol{\sigma}_{\boldsymbol{\phi}}$ はパラメータ $\boldsymbol{\phi}$ とエンコーダーの入力 $\boldsymbol{x}$ によって決まる変数である．

VAE-Loop では，VoiceLoop によるテキスト $\boldsymbol{c}$ から音響特徴量 $\boldsymbol{x}$ の生成過程を，VAE におけるデコーダー $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{c})$ と見なす．VoiceLoop をデコーダーとして用いるために，(1),(2),(3) 式を以下のように置き換える．

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{c}) = \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{z}, \boldsymbol{c}) \quad (6)$$

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{z}, \boldsymbol{c}) \sim \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:t-1}, \boldsymbol{z}), \boldsymbol{I}) \quad (7)$$

$$u = N_u([S_{t-1}, C_t, \boldsymbol{z}]) \quad (8)$$

(8) 式では，VoiceLoop による音声の生成過程に潜在変数 $\boldsymbol{z}$ を組み入れている．$\boldsymbol{z}$ は S_{t-1}, C_t と違いタイムステップ t に依存しない大域的な特徴であり，自己回帰生成モデルの欠点を補っている．

$\boldsymbol{z}$ は訓練時にエンコーダー $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$ からサンプリングされ，音声の大域的な特徴を学習する．エンコーダーは通常の VAE と同様に多層のニューラルネットワーク (Deep Neural Networks, DNN) でモデル化する．一方推論時には，$\boldsymbol{z}$ は単純な事前分布 $p(\boldsymbol{z})$ を介してサンプリングされる．VAE-Loop の概念図は図 1 で確認できる．

提案モデルの目的関数は，(5), (6) 式を用いて以下のように書ける．

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x}, \boldsymbol{c}) = &-D_{KL}(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})) + \\ &\mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\sum_{t=1}^{T} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{z}, \boldsymbol{c})]\end{aligned} \quad (9)$$

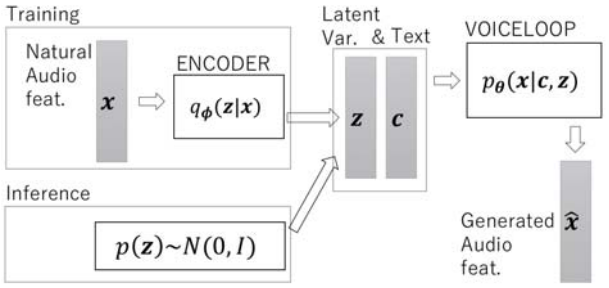


図 1: 提案モデル VAE-Loop の概念図．訓練時はエンコーダーから，推論時は事前分布から潜在変数のサンプリングを行う．

ここで (9) 式の第一項は正則化項，第二項は再構成誤差である，第二項は VoiceLoop が推定した $\hat{\boldsymbol{x}}_t$ と真の $\boldsymbol{x}_t$ の間で平均二乗誤差をとることで推定できる．つまり，$\boldsymbol{z}$ を用いる点を除いて通常の VoiceLoop の目的関数と同様である．

4. 実験

4.1 データセット

本研究では，複数話者のデータセットと，様々な感情や話し方を含むデータセットの二つを訓練に用いた．複数話者のデータセットとして，109 人の英語話者の音声を集めた VCTK Corpus[Veaux 17](以下 VCTK と呼ぶ) を用いた．[Taigman 18] とデータの前処理などの条件を揃えるために，VCTK に含まれるすべてのデータではなく，VoiceLoop のソースコードページ *1 から入手できる男性 4 人女性 17 人計 21 人の北米の話者による合計約 5 時間の音声データを用いた．それぞれの音声の長さは 5 秒以下である．

二つ目のデータセットとして，Blizzard Challenge2012[King 12] で使用された，4 編のオーディオブックからなるデータセット (以下 Blizzard2012 と呼ぶ) を用いた [Braunschweiler 10]．オーディオブックは多様な感情や話し方を含むことから，表現豊かな音声合成の研究によく利用される．Blizzard2012 は VCTK と違って全て同一の男性話者によって朗読されている．VCTK と条件を揃えるため，また勾配爆発を避けるため長さが 5 秒以下の音声のみを用いたところ，全体で約 10 時間の音声となった．

二つのデータセットに共通の前処理として冒頭の無音期間の削除を行った．また，全体の 9 割を訓練データとして，残りの 1 割を検証データと 50 サンプルのテストデータに分けた．実験に用いる特徴量は [Taigman 18] と同様に，モデルへの入力を音素，モデルが出力する音響特徴量を 60 次元のメルケプストラム，基本周波数 (F0)，非周期性成分，有声・無声フラグとした．

4.2 実験設定

実験では VoiceLoop と提案モデル VAE-Loop について比較を行った．提案モデルのエンコーダーには，時間方向への畳み込みを基盤とした DNN を用いた．具体的には，ストライド幅 2 の畳み込み，ドロップアウト，バッチ正規化，ReLU を 5 層繰り返した後に，可変長の音響特徴量を扱うための時間方向の MaxPooling，全結合層を重ねた構造を用いた．一方，提案モデルのデコーダーおよびベースラインに用いた VoiceLoop のハイパーパラメータは，[Taigman 18] での著者らによる実装と同様とした．

訓練時には最適化アルゴリズムを Adam，エポック数を 150 とした．学習率は [1e-3, 1e-4, 5e-5, 1e-5] の中から検証誤差が最小になるものを選んだところ，VCTK では 1e-4，Blizzard2012 では 5e-5 となった．

*1 https://github.com/facebookresearch/loop/

表 1: MOS (Mean±CI)

Method	VCTK	Blizzard2012
Ground Truth(1)	4.07 ± 0.23	3.94 ± 0.30
VoiceLoop(w/o labels)	2.51 ± 0.34	2.23 ± 0.24
VoiceLoop(w/ labels)	3.24 ± 0.27	N/A
VAE-Loop($\sigma = 1$)	**3.25 ± 0.29**	2.47 ± 0.32
VAE-Loop($\sigma = 0.7$)	N/A	**2.89 ± 0.32**
VAE-Loop($\sigma = 0$)	N/A	**3.03 ± 0.32**

4.3 MOS による評価

本節では大域的な特徴のラベルが得られないときも VAE-Loop が逆問題を緩和して高品質な音声を生成する能力を持つと示すために，crowdMOS toolkit[Ribeiro 11] と Amazon Mechanical Turk を用いた Mean Opinion Score(MOS) による評価を行った．MOS とは，評価者がアルゴリズムによって生成された音声を聞いて「人間の音声のように自然に聞こえるか」に基づいて 5 段階評価でつける点数である．評価者はそれぞれのモデルごとに 50 個の音声サンプルを聞いて評価した．評価者はアメリカ在住者に限定し，二つのデータセットに対してそれぞれ 15 人以上の評価者が参加した．表 1 は，MOS とその 95%信頼区間 (Confidence Interval, CI) を表している．ただし，表 1(1) は Vocoder によって再構成した後の音声サンプルを，MOS の上限として用いている．

表 1 より，VCTK において提案モデルの評価がベースラインの評価を上回ることが観測された．さらに，VAE-Loop は話者ラベルを用いていないにも関わらず，話者ラベルを用いる場合の VoiceLoop と同レベルの評価を得ていることが観測された．また我々の非公式な試聴では，VAE-Loop がベースラインに比べて明瞭な音声の生成に失敗する（例えば，吐息やある特定の音素を数秒間生成し続ける）ことが少ないと観測された．したがってこの MOS の差は，VAE-Loop が話者性という音声の大域的な特徴を学習することで困難な逆問題を軽減したことを示唆していると考えられる．

Blizzard2012 では，テストサンプルの生成に使う $p(\boldsymbol{z})$ が高い分散を持つとき，VAE-Loop も VoiceLoop と同様に明瞭な音声の生成に失敗する現象が度々確認された．そこで，$p(\boldsymbol{z}) \sim \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \sigma^2\boldsymbol{I})$ として，様々な σ を用いて $\boldsymbol{z}$ のサンプリングを行った．ただし，$\sigma = 0$ とは常に $\boldsymbol{z} = 0$ をサンプリングすることを指す．その結果，$p(\boldsymbol{z})$ の分散を抑えると提案モデルの MOS が向上し，ベースラインを上回ることが観測された．ここで，σ を小さくすると毎回似たような $\boldsymbol{z}$ がサンプリングされるので，σ を変えることは安定した音声の生成と音声の多様性の間でトレードオフを取ることだと捉えることができる．

4.4 潜在変数を用いた音声の特徴のコントロール

本節では，提案モデルが潜在変数を用いて生成する音声の特徴をコントロールできることを示す．図 2 は，VCTK を用いて訓練した VAE-Loop が生成した音声の F0 の軌跡を表している．左右の図では，別々の発話文を同一の潜在変数 $\boldsymbol{z}$ を用いて生成している．図 2 で $\boldsymbol{z}$ の値に応じて F0 の高さが決まることから，提案モデルが女性らしい高い声，男性らしい低い声といった特徴を潜在変数を用いてコントロールできることが確認できた．また，二つの $\boldsymbol{z}$ の中間の値を用いて音声を生成すると，F0 の値も平均化されることが確認できた．図 3 は，Blizzard2012 を訓練に用いたときの F0 の軌跡を表している．VCTK と同様に，潜在変数の値が F0 の軌跡を特徴付けてい

ることが確認できた．音声の例は次の URL で試聴できる．: https://akuzeee.github.io/VAELoopDemo/

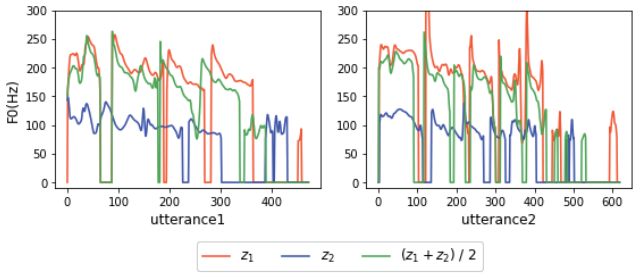


図 2: VCTK で訓練した VAE-Loop を用いた，二つの発話文に対する F0 の軌跡．z_1 は高い声 (我々の試聴によれば女性の声)，z_2 は低い声 (男性の声) を生成する．また二つの潜在変数の平均をとることで，F0 の値も平均化される．

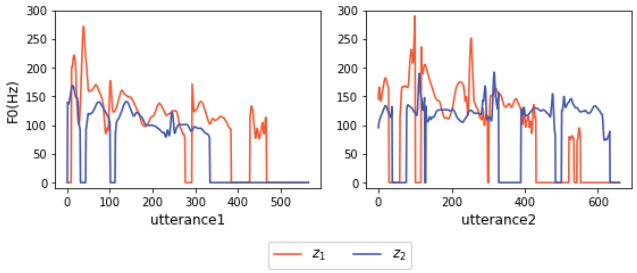


図 3: Blizzard2012 で訓練した VAE-Loop を用いた，二つの発話文に対する F0 の軌跡．z_1 は高低差の大きい声 (我々の試聴によれば芝居掛かった声)，z_2 は高低差の少ない声 (落ち着いた声) を生成する．

5. 議論と結論

本研究では，既存の自己回帰生成モデルが抱える問題を克服し，自然さ，明瞭さに加えて表現の豊かさを持つ音声合成をラベルなしで実現するために，VAE を用いて多様な表現をモデル化することを提案した．実験では，ラベルなしでの表現豊かな音声合成という困難な逆問題において，VoiceLoop が VAE によって獲得した大域的特徴を利用することでより高品質な音声を生成すること，そして音声の話者性や話し方を操作できることを示した．

本研究の限界に，Truncated Back Propagation Through Time を含む十分なハイパーパラメータ探索を実験コストの問題から行えなかったことがある．しかし 4.2 節で述べた実験設定に限れば二つのデータセットで提案モデルの有効性を確かめることができた．また VAE が困難な逆問題を緩和し高品質な音声の生成に役立つことは示せたが，「表現の豊かさ」を直接評価することは課題として残された．

将来の研究の方向性としては，少量のラベルありデータを用いた半教師あり学習への拡張が考えられる．また，VAE のエンコーダーの構造を改良するなどして，より系列データに特化した潜在表現を獲得することも重要だと考える．

参考文献

[Braunschweiler 10] Braunschweiler, N., Gales, M. J. F., and Buchholz, S.: Lightly supervised recognition for automatic alignment of large coherent speech recordings, in *Proc. interspeech 2010*, pp. 2222–2225 (2010)

[Eyben 12] Eyben, F., Buchholz, S., and Braunschweiler, N.: Unsupervised clustering of emotion and voice styles for expressive TTS, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4009–4012 (2012)

[Gulrajani 17] Gulrajani, I., Kumar, K., Ahmed, F., Taiga, A. A., Visin, F., Vazquez, D., and Courville, A.: PixelVAE: A Latent Variable Model for Natural Images, in *Proc. 5th International Conference on Learning Representations* (2017)

[Hsu 17] Hsu, W.-N., Zhang, Y., and Glass, J.: Learning Latent Representations for Speech Generation and Transformation, in *Proc. Interspeech 2017*, pp. 1273–1277 (2017)

[King 12] King, S. and Karaiskos, V.: The Blizzard Challenge 2012, in *Proc. Blizzard Challenge workshop* (2012)

[Kingma 14] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes., in *Proc. 2nd International Conference on Learning Representations* (2014)

[Luong 17] Luong, H. T., Takaki, S., Henter, G. E., and Yamagishi, J.: Adapting and controlling DNN-based speech synthesis using input codes, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4905–4909 (2017)

[Oord 16] Oord, van den A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. W., and Kavukcuoglu, K.: WaveNet: A Generative Model for Raw Audio, *CoRR*, Vol. abs/1609.03499, (2016)

[Ribeiro 11] Ribeiro, F., Florncio, D., Zhang, C., and Seltzer, M.: CROWDMOS: An approach for crowdsourcing mean opinion score studies, in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2416–2419 (2011)

[Taigman 18] Taigman, Y., Wolf, L., Polyak, A., and Nachmani, E.: VoiceLoop: Voice Fitting and Synthesis via a Phonological Loop, in *Proc. 6th International Conference on Learning Representations* (2018)

[Veaux 17] Veaux, C., Yamagishi, J., and MacDonald, K.: CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit (2017)

[Wang 17a] Wang, Y., Skerry-Ryan, R. J., Xiao, Y., Stanton, D., Shor, J., Battenberg, E., Clark, R., and Saurous, R. A.: Uncovering Latent Style Factors for Expressive Speech Synthesis, *CoRR*, Vol. abs/1711.00520, (2017)

[Wang 17b] Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyrgiannakis, Y., Clark, R., and Saurous, R. A.: Tacotron: Towards End-to-End Speech Synthesis, in *Proc. Interspeech 2017*, pp. 4006–4010 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

音声対話システムにおける 対話の状況を利用した応答タイミング推定

Response Timing Estimation Using Dialogue Situations for Spoken Dialogue Systems

赤井元紀 武田龍 駒谷和範
Genki Akai Ryu Takeda Kazunori Komatani

大阪大学産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

Responding at appropriate timing is important in spoken dialogue systems. In conventional systems, response timing is determined regardless of situations. We newly incorporate dialogue situations and estimate appropriate response timing by taking them into consideration. We design two processing units corresponding to a current situation and specific response timing. Prosodic features in the two different units are used for estimating the situation and response timing, respectively. We collected data and trained two models: discriminative and regression models. Experimental results showed the regression model outperformed that without the dialogue situations by about 9 points.

1. はじめに

近年，雑談など対話そのものを目的とした対話システムの必要性が高まっている．そのような対話システムでは，応答の内容だけでなく，そのタイミングも重要である．つまり、ユーザの発話に対して適切な応答のタイミングを推定できるようにする必要性がある．

適切なタイミングでの応答に関する概念として話者交替（ターンテイキング）がある．話者交替とは，対話の参加者が順番に発話を行う現象のことである．話者交替の分析の際は，ターンが基本単位となる．ターンは，対話の参加者のうち一人が話し続けている時間軸上の区間として定義される [中野 15]．1 つのターンは，1 つ以上の発話と発話間の休止時間からなる．ただし，休止時間や発話に並行して他の話者による相槌がなされたとしても，話者交替が起きたとはみなされないことが多い．本研究においても，相槌はメイナード [メイナード 93] にならい「休止時間や相手の発話に並行してなされる，ターンを取得しない（話者交替が行われない）発話」として考える．

相槌も適切なタイミングでなされるべき応答の種類の 1 つである．相槌はタイミングだけでなく頻度も重要となる．例えば，現在の話題が盛り上がっていれば，高い頻度で相槌が打たれる方が望ましいはずである．以降，単に応答と述べたときは話者交替を伴う応答と相槌の両方を示すものとする．これらの応答を適切なタイミングで行うことで，ユーザのシステムに対する満足度を高めることが期待できる．これまでにもユーザからの音声入力に対して，応答タイミングを推定する研究がなされている [大須賀 06, 西村 09, 小川 14, Lala 17]．

このような従来研究の問題として，どのような対話の状況に対しても同様に応答する点がある．例えば現在の話題が盛り上がっているか等の状況に応じて，適切な応答のタイミングや相槌の頻度は異なる．つまり，応答タイミングの推定は対話の状況に依存すると考えられる．

本研究では，図 1 のような対話システムの実現を目指す．図の左側の例のように，「駅前のラーメン屋がー・・・」と話が盛り上がって興奮した様子で話す男の子に対しても，従来の対話システムでは，通常と同じ頻度で「うんうん」と応答する．これに対して本研究では，図の右側の例のように「駅前のラーメン屋がー」「うんうん」・・・ というように，男の子の盛り上がり具合に合わせて高い頻度で応答するようなシステムを目指す．



図 1: 理想のシステム

本論文では，上記のシステムの実現のために，複数の処理単位を用いた応答タイミングの推定を行う．複数の処理単位を用いて現在の対話の状況と応答タイミングのそれぞれに特化した推定を行うことで，状況に応じた適切な応答タイミングを推定する．システムの行動推定の評価に使用するデータとして，ユーザとシステムとのインタビュー形式の対話における 7 名のユーザの回答音声を収集した．収集したデータを用いて，システムの行動推定における対話の状況の利用の有無による性能の変化を評価した．

2. 対話の状況の設定

適切なシステムの応答タイミングや相槌の頻度は一意に定まるものではなく，ユーザの状態や対話の内容などの大域的情報（以降，対話の状況と呼ぶ）によって変化すると考えられる．本研究では，対話の状況を応答タイミングの推定に利用することで，より適切な応答タイミングの推定を目指す．

対話の状況として利用できる情報として考えられるのは，(1) ユーザの状態（現在の話題が盛り上がっている，考え中や言いよどみ），(2) ユーザの特性（年齢，性別，話すのが好きかどうか），(3) 対話のコンテキスト（明るい話，暗い話，対話のトピック）の 3 つである．例えば，(1) の場合，現在ユーザが話している話題が盛り上がっているならば，応答のタイミングは早めで相槌の頻度が高い方が望ましいはずである．

本研究では，対話の状況として，これらの 3 つの情報のうちまず (1) の導入を試みる．特に，(1) の中でも，現在の話題

連絡先: 赤井元紀，大阪大学産業科学研究所，大阪府茨木市美穂ケ丘 8-1，06-6879-8416，akai@ei.sanken.osaka-u.ac.jp

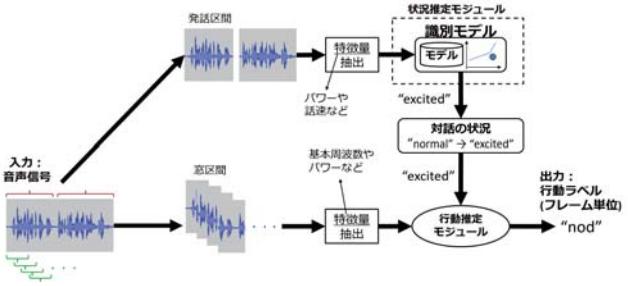


図 2: 応答タイミング推定モジュールの全体像

が盛り上がっているかどうかに着目する．以降，対話の状況として「話が盛り上がっている」と「普通」の 2 種類を考え，それぞれ excited と normal というラベル（以降，状況ラベルと呼ぶ）を割り当てる．

3. 複数の処理単位を用いた応答タイミングの推定

3.1 応答タイミング推定モジュールの全体像

システム実行時における対話の状況に応じた応答タイミングの推定について説明する．本研究では，100 ミリ秒を 1 フレームとし，システムの応答タイミングを推定する問題を，1 フレームごとにシステムが取るべき行動を推定する問題とする．図 2 に本対話システムの全体像を示す．

本モジュールを使用する対話システムでは，ユーザとシステムとの 1 対 1 でのインタビュー形式での対話を想定している．具体的には，(1) システムはインタビュワー役であり，ユーザに対して質問を行う，(2) ユーザはシステムの質問に対して，何らかの返答を返す，(3) システムは適切に相槌を打ちながら，ターン取得可能と判断した時点で次の質問に移る，といった対話の流れを想定している．

入力音声信号を発話区間および窓区間に分割することで，対話の状況推定およびシステムの行動推定に利用する音声区間を得る．発話区間は無音で区切られた音声区間（*Inter Pausal Units:* IPU）として定義する．また，窓区間は入力音声に対してシフト幅 s ミリ秒ずつずらしながら窓幅 w ミリ秒ごとに区間を切り出すという処理により切り取られた音声区間とする．本研究ではシフト幅 s は 100 ミリ秒，窓幅 w は 3000 ミリ秒として固定した．

状況推定モジュールは，発話区間ごとに対話の状況を推定し，現在の対話の状況を更新する．行動推定モジュールは，窓区間から抽出した特徴量と現在の対話の状況を入力としてシステムが取るべき行動を出力する．本研究では，複数の処理単位を使い分けることで，対話の状況に応じた応答タイミングの推定を実現する．

3.2 行動ラベルの設計

システムが取れる行動である「何も話さない」,「相槌を打つ」,「ターンを取得する」に対して，それぞれ silent，nod，talk というラベル（以降，行動ラベルと呼ぶ）を割り当てる．まず，インタビュー形式での対話を行うために最低限必要な行動として,「何も話さない」と「ターン取得を行う」の 2 つが挙げられる．ユーザの返答の最中は何も話さず，話者交替のタイミングが来た時点でターン取得を行い次の質問に移行することで，インタビュー形式の対話を最低限実行できる．さらに，本研究ではこれらの 2 つの行動に加えて「相槌」を考える．相槌のタイミング（や頻度）は，話者交替のタイミングに比べて対話の状況によって変化しやすいと考えられる．つまり，相槌のタイミングや頻度は，提案手法を用いることで最も改善される部分であると考えられる．以上より，本提案手法の性能を評価するためにシステムの行動として「相槌」を扱う．

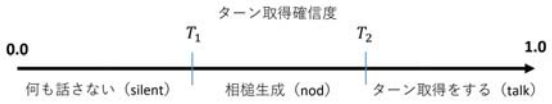


図 3: ターン取得確信度

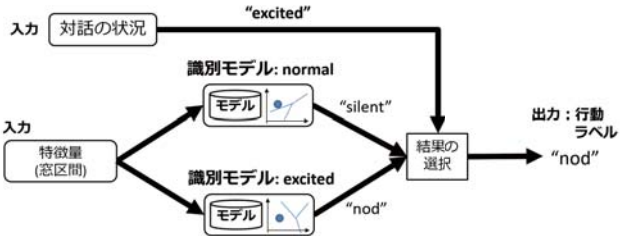


図 4: 行動推定モジュール（識別モデル）

システムの行動推定を行うためにターン取得確信度という概念を導入する．ターン取得確信度は，システムがターン取得すべきかどうかの度合いを表す 0-1 の連続値として定義する．ターン取得確信度 c に応じて 2 つのしきい値 T_1, T_2 に基づき，図 3 のように，$c \leq T_1$ ならば silent，$T_1 < c \leq T_2$ ならば nod，$c > T_2$ ならば talk というようにシステムの行動を決定することでシステムの行動推定を行う．しきい値 T_1, T_2 を変えることでシステムの挙動が変化するため，対話の状況に応じてしきい値を変えることで，システムの行動推定が可能となる．

3.3 システムの行動推定

窓区間から抽出した特徴量と現在の対話の状況をもとにシステムの行動を推定する方法について説明する．本論文では，識別モデルと回帰モデルのそれぞれを用いた場合について対話の状況を利用した応答タイミングの推定方法を提案する．

識別モデルの場合，図 4 のように対話の状況ごとにモデルを用意して使い分けることで応答タイミングを推定する．識別モデルは窓区間から抽出した特徴量を入力として取るべきシステムの行動を出力する．

回帰モデルの場合，図 5 のようにしきい値 T_1, T_2 を対話の状況ごとに変えることで応答タイミングを推定する．回帰モデルは窓区間から抽出した特徴量を入力としてターン取得確信度を出力する．ターン取得確信度を推定する回帰モデルを構築し，モデルの出力に応じて 2 つのしきい値 T_1, T_2 に基づき，図 3 のようにシステムの行動を決定する．

先行研究でよく用いられている識別モデル以外に回帰モデルを用いるのは，回帰モデルは 1 つのモデルで出力を柔軟に変化させられるため，必要な学習データの量が少なくなるからである．識別モデルにおいて対話の状況を利用する場合，対話の状況ごとにモデルを用意するために複数のモデルが必要となり，対話の状況を利用しない場合に比べてより多くの学習データが必要となる．一方，回帰モデルの場合は対話の状況を利用する場合でも，使用するモデルは 1 つなので，モデルの学習に必要なデータの量は対話の状況を利用しない場合と同じである．

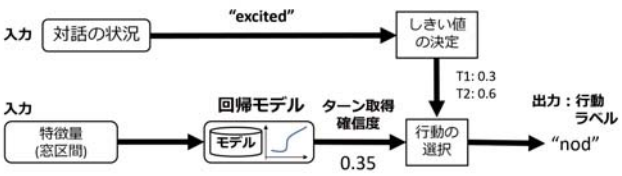


図 5: 行動推定モジュール（回帰モデル）

表 1: 作成したデータセット

種類	状況ラベル		行動ラベル		
	excited	normal	silent	nod	talk
ラベル数	587	1097	52909	12537	3864

3.4 システムの行動推定に用いる特徴量

本研究では，システムの行動推定には特徴量として韻律情報を用いる．特徴量には INTERSPEECH 2009 Emotion Challenge [Schuller 09] で使用されたもの（以降，IS09 と呼ぶ）を用いる．IS09 は音声区間に対して計算される 384 次元の韻律に関する特徴量である．IS09 は基本周波数やパワーなどの特徴量と最大値や平均などの統計量の組み合わせにより計算される．

3.5 対話の状況の推定に用いる特徴量

対話の状況推定にはシステムの行動推定に用いた特徴量とは異なる特徴量を用いる．これは，対話の状況推定に特化したモデルの構築を行うためである．対話の状況を明示的に切り離して扱うことにより，システムの行動推定と異なる特徴量を利用する．

本研究では，excited に相当する音声と normal に相当する音声を比較した結果，excited は normal に比べて，(1) 声が大きくなる，(2) 話す時間が増える，(3) 話すスピードが速くなるといった傾向がある．これらの傾向をもとに (a)IS09 における pcm_RMSenegrgy およびその変化量に対する統計量（24 次元），(b) 発話区間の長さ（秒），(c) 話速（モーラ / 秒）の 3 種類の特徴量が有用だと判断した．(1) と (a)，(2) と (b)，(3) と (c) がそれぞれ対応している．ここでは，「ぁぃぅぇぉゃゅょわ」以外のひらがな 1 文字を 1 モーラとみなした．

4. 評価実験

4.1 使用データ

本節では，対話の状況推定やシステムの行動推定で使用する，モデルの構築のためのデータについて述べる．まず，システムの質問に対する 7 名のユーザの回答音声を収集した．このとき，複数の話し方で収録を行うことで，各対話の状況の正解に相当する音声を収集した．次に，ユーザの回答音声を発話区間単位に分割し，各発話区間に対して正解となる状況ラベルを割り当てた．そして，ユーザの回答音声をフレーム単位に分割し，各フレームに対して正解となる行動ラベルをアノテートした．以上の手順により得られたラベルの数を表 1 に示す．状況ラベルは対話の状況推定に，行動ラベルはシステムの行動推定にそれぞれ用いた．

4.1.1 ユーザの回答音声の収集

研究室の学生 7 名にシステムとの対話音声の収録を依頼した．収録は，システムが行う 10 個の質問に対してユーザがそれぞれ回答するという手順で行った．以降，10 個の質問に対するユーザの回答をひとまとまりとして，1 セッションと呼ぶ．また，収録の際は (a) 普段通りの話し方，(b) 何を話そうか考え込む感じ，返答に困っているような話し方，(c) 自分の好きなもの，興味のあることについて「語る」ような話し方の 3 通りの話し方で答えるように指示した．これは，対話の状況の正解に相当する音声を収集するために行った．(a), (b) は normal，(c) は excited に相当する．以上より，$3 \times 7 = 21$ セッション分の音声が得られた．1 セッションあたりの収録時間は平均約 5 分だった．

4.1.2 対話の状況推定に用いる学習データの収集

対話の状況推定に用いる学習データとして，発話区間と正解となる状況ラベルのペアを収集する．excited に相当する音声から得られた発話区間に対しては excited を正解とみなし，normal に相当する音声から得られた発話区間に対しては normal を正解とみなした．これにより得られた状況ラベルの数は，excited が 587，normal が 1097 であった．

4.1.3 システムの行動推定に用いる学習データの収集

システムの行動推定に用いる学習データとして，フレームと正解となる行動ラベルのペアを収集する．まず，収録した対話音声への行動ラベルの付与を学生 6 名に依頼した．アノテート対象の発話を再生し，その発話に対して，リアルタイムでシステムが取るべき行動をアノテートするよう指示した．

次に，各アノテータごとにフレームに対して正解となる行動ラベルをアノテートした．この段階では，1 つのフレームに対して 6 つの行動ラベルがアノテートされていることになる．

さらに，nod や talk がアノテートされたフレームの近傍のフレームの正解は，1 フレームの長さが 100 ミリ秒と短いため，silent よりも nod や talk の可能性が高いと考えられる．そこで，本実験では，まず，各アノテータが付与したラベルに対して，nod や talk がアノテートされたフレームから前後 3 つのフレームのラベルを nod や talk で置き換えるという処理を行った．次に，各フレームごとにアノテータ全員のラベルから 1 つの行動ラベルを決定した．各フレームに対するラベルは，(1) そのフレームに nod や talk の付与したアノテータが一人もいなければ silent に決定，(2)nod と talk のうち付与された数が多いラベルに決定（同数なら nod）という手順で決定した．これらの手順は，アノテータが付与した nod や talk は全て正しいという仮定のもとに行った．この仮定にもとづいて，一人でも nod や talk を付与したフレームは nod か talk が正解になるとしている．以上の手順により得られた行動ラベルの数は，silent が 52909，nod が 12537，talk が 3864 であった．

4.2 対話の状況推定の評価

4.2.1 実験条件と評価方法

対話の状況 excited と normal の 2 値分類について実験した．本実験では話者 open の cross validation を行い，正解率によって評価を行った．ただし，チャンスレートが 50.0%になるように重みを設定して学習と評価を行った．

モデルには Support Vector Classification（SVC）を使用し，カーネル関数には線形カーネルと RBF カーネルの 2 種類を用いた．モデルの構築には Weka[*1] を用いた．

本実験では，(1)IS09 の全特徴量 [IS09]，(2)IS09 における pcm_RMSenegrgy およびその変化量に対する統計量（24 次元）と発話区間の長さ（秒），話速（モーラ / 秒）[custom] の 2 種類の特徴量を使用した．以降，(2) の特徴量を custom とする．特徴量の次元は，IS09 が 384 次元，custom が 26 次元である．custom における話速の計算は以下のように行った．

*1 http://www.cs.waikato.ac.nz/ml/weka/

表 2: 状況推定の正解率の平均 (%)

	線形カーネル	RBF カーネル
IS09	64.8	65.9
custom	68.4	**69.3**

まず，Google Cloud Speech API[*2] を用いて音声認識を行った．次に，認識結果を KAKASI[*3] を用いてひらがなに変換した．そして，認識結果のモーラの数を発話区間の長さで割ることで話速とした．

4.2.2 実験結果

正解率の平均を表 2 に示す．custom および RBF カーネルを用いた場合において，正解率が 69.3%と最も高くなった．以降のシステムの行動推定において，対話の状況推定は，モデルに SVC，カーネルに RBF カーネル，特徴量に custom を用いて行う．

4.3 システムの行動推定の評価

4.2 節で評価した状況推定モジュールを用いたシステムの行動推定に関して評価する．

4.3.1 実験条件と評価方法

システムの行動 silent，nod，talk の 3 値分類について実験した．本実験では話者 open の cross validation により評価を行った．ただし，チャンスレートが 33.3%になるように重みを設定して学習および評価を行った．

モデルは，識別モデルとして SVC を，回帰モデルとして Support Vector Regression（SVR）を使用した．SVR はデータの正解ラベルをターン取得確信度の数値に変換して学習を行った．本実験では，silent = 0.0，nod = 0.5，talk = 1.0 とした．SVR の学習に関しては学習セットの一部を development set としてしきい値の最適化に使用した．しきい値 T_1, T_2 は 0.0 から 1.0 までの 0.05 刻みのグリッドサーチにより決定した．これらのモデルの構築には Weka を用いた．カーネル関数には RBF カーネルを用いた．特徴量は IS09 を窓区間ごとに抽出した．

本実験では，評価指標として正解率以外に各行動ラベルごとの precision を使用する．nod や talk のタイミングを間違えると，ユーザの対話に対する満足度が低下しやすいと考えられるので，nod と talk の precision が高い方が望ましい．一方で，silent のタイミングを間違えても，ユーザの対話に対する満足度が低下しにくいと考えられるので，silent の precision は高くなくてもよい．

4.3.2 実験結果

システムの行動推定の正解率の平均と nod および talk に関する precision を表 3 に示す．対話の状況の項目の「あり（推定）」は 4.2 節で構築したモデルを用いて推定した対話の状況を行動推定モジュールに入力として与えた場合の結果である．「あり（正解）」は対話の状況の正解を行動推定モジュールに入力として与えた場合の結果である．

回帰モデルの場合，対話の状況の推定結果および正解を利用することでシステムの行動推定の正解率が約 9 ポイント上昇した．また，nod の precision が対話の状況の推定結果を利用した場合は 0.142，対話の状況の正解を利用した場合は 0.152 上昇した．一方で，talk の precision は対話の状況の推定結果および正解を利用することでともに若干減少している．しかし，

表 3: システムの行動推定の実験結果

モデル	対話の状況	正解率（%）	precision(nod)	precision(talk)
識別	なし	73.3	0.717	0.889
	あり（推定）	71.0	0.661	0.888
	あり（正解）	71.3	0.621	0.896
回帰	なし	60.4	0.426	0.928
	あり（推定）	**69.0**	**0.568**	0.913
	あり（正解）	**69.7**	**0.579**	0.912

回帰モデルを用いた場合の talk の precision はいずれも 0.9 以上と高く，talk の precision の減少幅は nod の precision の上昇幅に比べて十分小さいので，回帰モデルの場合は提案手法が有効だといえる．

一方，識別モデルの場合，対話の状況の推定結果および正解を利用するとシステムの行動推定の正解率が約 2 ポイント減少した．また，talk の precision は対話の状況の利用の有無でほぼ変化がないのに対し，nod の precision は対話の状況の推定結果を利用した場合は 0.056，対話の状況の正解を利用した場合は 0.096 減少した．これは，識別モデルにおいて対話の状況を利用する場合，対話の状況を利用しない場合に比べてより多くの学習データを必要とするためと考えられる．つまり，本実験において十分な学習データが確保できていなかったといえる．学習データを増やすことで識別モデルに関しても提案手法が有効となる可能性がある．

今後の課題としては，まず，十分な学習データを確保することが挙げられる．これは，既存のコーパスなどを利用することで解決できると考えられる．また，本研究では窓幅 w を固定したが，窓幅 w の最適化も必要である．さらに，本研究で提案した対話システムを実際にユーザに使用してもらい，アンケートなどで性能を評価する必要がある．

参考文献

[Lala 17] Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., and Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking, in *Proc. of SIGDIAL*, pp. 127–136 (2017)

[Schuller 09] Schuller, B., Steidl, S., and Batliner, A.: The INTERSPEECH 2009 Emotion Challenge, in *Proc. of INTERSPEECH* (2009)

[メイナード 93] メイナード 泉子：会話分析，くろしお出版 (1993)

[小川 14] 小川 翼, 伊藤 敏彦：リアルタイム発話継続/交替予測システムの構築, HAI シンポジウム, pp. 192–198 (2014)

[西村 09] 西村 良太, 中川 聖一：応答タイミングを考慮した音声対話システムとその評価, 研究報告音声言語情報処理 (SLP), Vol. 2009, No. 22, pp. 1–6 (2009)

[大須賀 06] 大須賀 智子, 堀内 靖雄, 西田 昌史, 市川 熹：音声対話での話者交替/継続の予測における韻律情報の有効性, 人工知能学会論文誌, Vol. 21, No. 1, pp. 1–8 (2006)

[中野 15] 中野 幹生, 駒谷 和範, 船越 孝太郎, 中野 有紀子：対話システム, 自然言語処理シリーズ, コロナ社 (2015)

∗2 https://cloud.google.com/speech/

∗3 http://kakasi.namazu.org/index.html.ja

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

音声対話システムのためのユーザの発話権維持状態の逐次推定

Incremental estimation of user's turn-taking state for a spoken dialog system

藤江 真也 *1*2 Shinya Fujie
横山 勝矢 *2 Katsuya Yokoyama
小林 哲則 *2 Tetsunori Kobayashi

*1千葉工業大学 Chiba Institute of Technology
*2早稲田大学 Waseda University

Turn-taking state estimation to determine utterance timing of a spoken dialog system is discussed. We propose the recurrent neural network based method to estimate user's turn-taking state incrementally. The proposed method utilizes acoustic feature extracted using a spectrogram autoencoder as well as linguistic feature extracted from a partial speech recognition result using a neural network based language model. The article shows an example of estimation result and discuss the performance of the proposed method.

1. はじめに

音声対話システムの発話タイミング決定のための，ユーザの発話権維持に対する状態を逐次予測する手法を提案する．

従来の音声対話システムの多くは，主に音声認識の前処理として実行される音声区間検出（VAD; Voice Activity Detection）の結果を利用し，音声信号の終了時点をユーザの発話終了として扱っている．更に，多くのシステムはその時点でユーザが発話権をシステムに渡している，あるいは放棄しているとみなし，即座にシステムの発話を開始するような単純な制御をしている．しかしながら，音声区間の途切れ目が発話終了とは限らず，ユーザは発話権を維持してそのまま発話を継続する可能性がある．また，必ずしも音声が途切れていなくともシステムが発話を開始してもよいタイミングが存在するのも事実である．これらに対し，音声区間ごとに発話の継続，終了を推定する手法がある [1, 2]．これらの研究では依然として VAD が検出する音声区間を対象としてその推定を行っている．VAD は，音声信号の終了を正確に検出するために，ハングオーバー処理（一定の長さの無音区間が続くまで音声区間が続いているとする処理）を行う．このため，VAD が検出する音声区間に対して推定を行っている限り，一定の処理遅れが生じることは避けられない．また，音声区間の途中でシステムの発話を開始することは原理的に不可能である．

そこで，逐次的に音声信号を処理し，フレームごとにユーザの発話権に対する状態（維持しようとしているか，譲渡または放棄しようとしているか）を評価することで，システムの発話タイミングを決定しやすくする方法を提案する．従来研究においては，主にシステムの相槌（back-channelling）のタイミング検出のためにこのような処理が行われている [1, 3]．その理由として，相槌生成においてはタイミングの遅れの悪影響が明らかであることが挙げられる．我々は相槌以外のシステム発話に関してもそのタイミングの遅れは好ましくないと考える．

本稿では，単純な方法で発話を認定して行った従来研究 [4] に対して，より厳密な方法で発話権の維持を認定し，予測モデルを構築した結果について報告する．

連絡先: 藤江真也，千葉工業大学 先進工学部 未来ロボティクス学科，shinya.fujie@p.chibakoudai.jp

2. 話者交替タイミングの分析

2.1 分析条件

すでに収集あるいは公開されている人対人の音声対話コーパスに関して，話者交替がどのようなタイミングで行われているかを分析する．

分析対象としては，以下の 4 つのコーパスを選んだ．

- 早稲田コーパス（W）: 早稲田大学と ATR-Trek で独自に収集したもの．レストラン検索サービスを想定した対話．391 対話，計 16 時間 30 分．
- 日本語話し言葉コーパス（CSJ）: 講演者とインタビュワーによる講演に関するインタビュー．58 対話，計 12 時間 10 分．
- RWCP 音声データベース（RWCP）: 自動車購入，海外旅行計画のスタッフと客を想定した対話．96 年度版と 97 年度版両方を対象とした．61 対話，計 8 時間 29 分．
- 重点領域研究「音声対話」対話音声コーパス（PASD）: 様々なタスクの対話で，部分的に人とシステムの対話を含む．62 対話，計 5 時間 8 分．

機械的に話者交替をカウントするために，音声区間を利用する．音声区間情報がコーパスに含まれるもの（CSJ，RWCP）は，それを利用した．その他のコーパスに関しては，公開されているツール（`https://github.com/ASTL-NICT/VAD`）を利用して区間検出を行った．音声区間の間のポーズ長が 700ms 以下の場合は，結合して 1 つの発話区間とした．

以上の処理を施した上で，一方の話者の発話区間の終了タイミングに対して，もう一方の話者の発話区間開始タイミングがどのように分布するかを調査した．

2.2 分析結果

話者交替の発生数は，W で 30,872，CSJ で 8,132，RWCP で 3,493，PASD で 4,131 であった．話者交替タイミングの累積相対度数を図 1 に示す．

タスクの違いによらず，話者交替の約 50%は先行話者の発話終了までに行われており，約 80%は先行話者の発話終了後 600ms まで，約 95%は 1,000ms までに行われていることが分かる．なお，PASD に関してはシステムとの発話が含まれており，システムの発話，あるいはシステムに対する人の発話が人同士の対話に比べて遅れる傾向にあることも分かった．

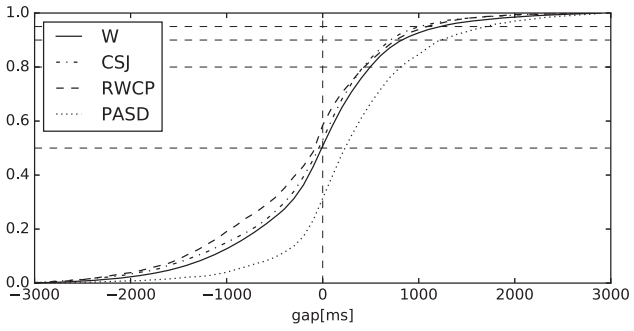


図 1: 話者交替の累積相対度数

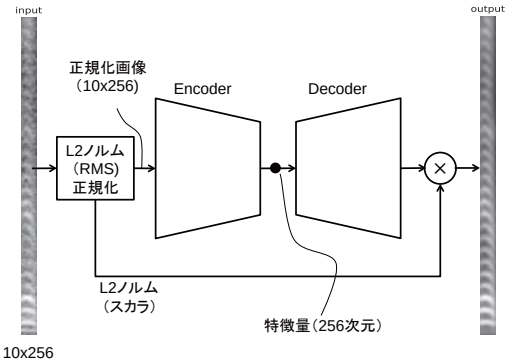


図 2: 音響特徴量抽出器（音声スペクトログラム自己符号化器）

3. 発話権維持の認定

前節の分析では 700ms 以上の無音で区切られた音声区間を発話区間とみなし，その交換を話者交替と考えた．ここでは，より正確に発話権の維持をモデル化するため，対話する二者の音声区間の相互関係をもとに発話権維持を認定する．

まず，音声区間を，それらの間の $\theta_{\text{short\_pause}}$ 以下の長さの無音を無視して 1 つに統合する．これを**発話区間**と呼ぶ．次に，$\theta_{\text{short\_utterance}}$ 以下の短い発話区間は**短い発話**と呼び，単独では発話権を維持する発話ではないと考える．

短い発話でない発話区間（$\theta_{\text{short\_utterance}}$ 以上の長さの発話区間）は，原則的に発話権を維持する発話であり，**発話権維持発話**と呼ぶ．発話権維持発話は，$\theta_{\text{keep\_turn}}$ より短い無音区間を無視して統合される．ここで，$\theta_{\text{keep\_turn}} > \theta_{\text{short\_pause}}$ である．すなわち，一度維持され始めた発話権は，ある程度長い無音区間が挟まれても維持されると考える．

例外的に，$\theta_{\text{short\_utterance}}$ 以上の長さを持つ発話であっても，先行して開始されている相手の発話権維持発話に時間的に内包される（相手の発話権維持発話より遅く始まり，早く終わる）場合は，短い発話とみなす．このルールに従うと，原理的にはどんなに長い発話でも，短い発話と認定される可能性があるが，実際には，本研究で扱うような協調的な対話においては，発話の衝突が起きた場合に発話権を獲得していない話者が喋り続けるということは起こらない．そのため，これを短い発話とみなすことは実用上問題ないと考える．

4. 発話維持状態推定モデル

4.1 問題設定

前節で認定した状態を，フレーム毎に逐次推定する．認定した状態のうち，発話権維持発話と短い発話，さらに発話していない状態の 3 状態は排他的であるため，これらは識別的に推定する．また，単に発話区間か，音声か非音声かについてもそれぞれ別々に同時に推定する．

4.2 特徴量抽出

発話終了の検出などにおいては，ピッチ変動などの韻律情報が重要であることが知られており，従来研究でも広く用いられている．しかしながら，ピッチ抽出は手法により精度が異なり，ある程度の精度を担保するためには複数フレームにより平滑化が必要となる．従って陽にピッチ抽出を行うと，全体の処理遅れが避けられなくなる．一方，音声認識で利用される MFCC などの音響特徴量を用いると，処理遅れは無くなるが，韻律情報が失われるという欠点がある．そこで本研究では，図 2 に示すような狭帯域スペクトログラムを符号化，復号化する

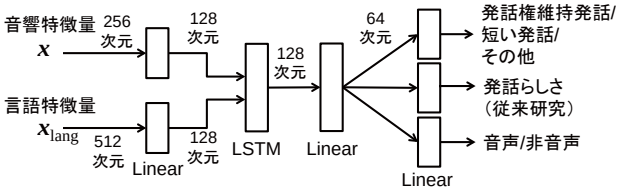


図 3: 発話権状態推定モデル

自己符号化器をニューラルネットワークにより構築し，その中間層の出力を音響特徴量とする．具体的には，10ms 毎に得られる 256 点のパワースペクトルを 10 本分並べたものを入力とし，中間層の出力 256 次元を特徴量とする．

発話権維持状態の推定においては，言語情報も有用であると考えられる．本研究では，Google Cloud Speech API (`https://cloud.google.com/speech/`) のストリーミング音声認識によって逐次得られる部分的な音声認識結果から言語特徴量を抽出する．具体的には，LSTM 言語モデルの中間出力 (512 次元) を用いる．なお，音声認識結果は毎フレーム得られるものではない．音声認識結果が得られないフレームについては，言語特徴量はゼロベクトルとする．

4.3 発話維持状態推定モデル

計算された音響特徴量，言語特徴量を入力とし，発話権維持状態を逐次推定するモデルをニューラルネットワークにより構築する．具体的な構造を図 3 に示す．

発話権維持状態に関しては，発話権維持発話，短い発話，その他（特に発話をしていない状態）を識別的に推定する．また，単に発話をしているかどうか，また，音声か非音声かを推定するネットワークを並列に入れたマルチタスク推定を行う．

5. 推定例

音響特徴量抽出に用いる自己符号化器は，日本語話し言葉コーパスの講演音声 1,000 講演からそれぞれ 10 発話ランダムに抽出した音声で学習した．発話権維持推定モデルは，早稲田コーパスの 100 対話の客役の音声で学習した．$\theta_{\text{short\_pause}}$，$\theta_{\text{short\_utterance}}$ は 1000ms，$\theta_{\text{keep\_turn}}$ は 1500ms とした．

学習に利用していないデータに対する推定例を図 4 に示す．発話権維持状態と短い発話の推定結果を見ると，発話権維持状

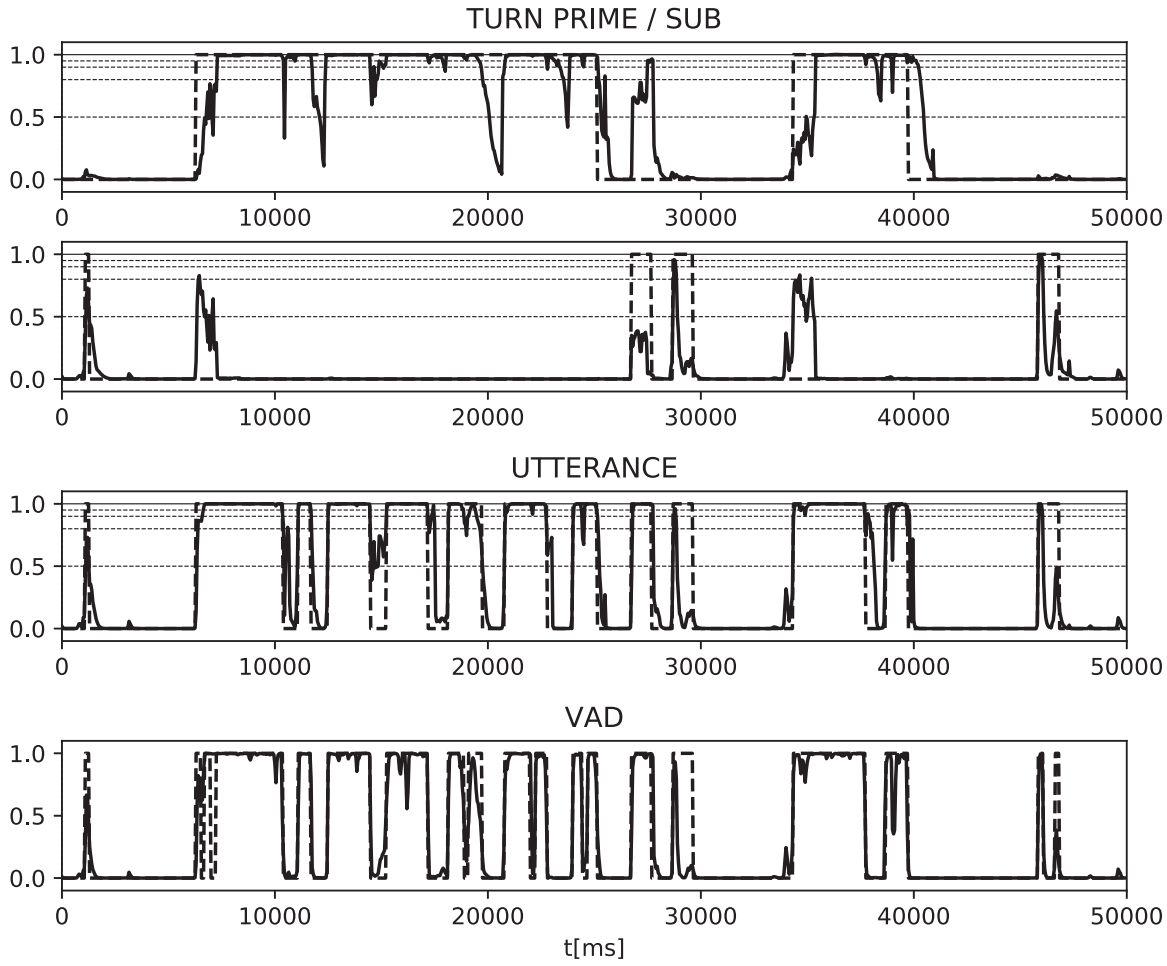


図 4: 予測結果の例．上から発話権維持状態，短い発話状態，発話らしさ [4]，VAD を表す．破線が目標値で，実線が予測結果．

態の始めは短い発話のスコアが高くなり，徐々に発話権維持状態のスコアが高くなっていくという傾向が見られる．これは，音声区間が開始した直後は，短い発話か発話権を維持するかの区別が付きづらいということの現れだと考えられる．また，一度発話権維持状態のスコアが高くなると，ある程度長い無音区間（発話区間や音声／非音声の目標値が 0 になっている区間）があっても，スコアが低くならずに発話権を維持していることを正しく推定できていることが分かる．一方，発話権維持が終了していると判断すべきところでも，しばらくスコアが高いままで，時間とともに徐々にスコアが低くなっている．これは，システムの発話開始の遅延につながるため，素早くスコアが下がるような改善を検討する必要がある．

6. むすび

音声対話システムの低遅延な発話タイミング決定を目指し，ユーザの発話権維持状態の逐次推定モデルを提案，実装した．今後の課題として，本手法で決定したシステム発話のタイミングが，VAD を用いた従来手法に比べて人同士の対話における発話タイミングに近いことを評価することや，実際の音声対話システムに組み込むことが挙げられる．

参考文献

[1] Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanishi, K., and Kawahara, T.: Attentive Listening System with Backchanneling, Response Generation and Flexible Turn-taking, SIGDIAL (2017)

[2] 藤江真也，木下裕介，小林哲則: 発話期待度／意欲度に基づく発話タイミング制御, 人工知能学会研究会資料，SIG-SLUD-B102，pp.43-48 (2011)

[3] Fuji, S., Fukushima, K., and Kobayashi, T.: Backchannel Feedback Generation using Linguistic and Nonlinguistic Information and its Application to Spoken Dialogue System, Proc. Interspeech (2005)

[4] 藤江真也，横山勝矢，小林哲則: 音声対話システムのためのユーザ発話終了タイミングの逐次予測，日本音響学会春季研究発表会講演論文集，2-Q-11 (2018)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

音象徴を用いた擬態語の意味的用法分類

Semantic-Usage Classification of Mimetic Words using Sound Symbolism

竹内なつみ *1 Natsumi Takeuchi　浦田大貴 *1 Daiki Urata　中村剛士 *1 Tsuyoshi Nakamura　加納政芳 *2 Masayoshi Kanoh　山田晃嗣 *3 Koji Yamada

*1名古屋工業大学 Nagoya Institute of Technology　*2中京大学 Chukyo University　*3情報科学芸術大学院大学 Institute of Advanced Media Arts and Sciences

Japanese language has many onomatopoeias(imitative and mimetic words). These words are used in various scenes as an useful method to communicate with other people intuitively and conveniently. Especially, some creators make and use onomatopoeias which fit with their own images to appeal to someone emotionally. However, it is difficult to make onomatopoeias because we don't have an objective evaluation criteria. To make unique and communicative onomatopoeias, it is important to estimate image of unknown onomatopoeias from existing onomatopoeias and imagine how people interpret the unknown onomatopoeias. Thus, we focus sound symbolism of onomatopoeias and attempt to classify onomatopoeias into appropriate usage by using its sound symbolism.This paper presents the results of classification of mimetic words as an investigation of onomatopoeias' semantic-usage classification.

1. はじめに

擬態語とは，オノマトペのうち音の発生しない状態を表現した言葉である．人は擬態語を用いてイメージや感覚を直接伝えることができ，初めて見聞きした擬態語に対してもそのイメージをある程度推定することができる．また，擬態語を含むオノマトペには音象徴があるとされており，オノマトペはその音から特定のイメージを喚起することができるとされる [1]．このような特徴からオノマトペはイメージや感覚を直接・簡潔に伝達することができる便利なコミュニケーション手段であると考えられており，様々な分野で活用されている．

オノマトペが活用される場の一つとして，漫画など表現のスペースが限られているものがある．言葉を尽くして状況や心理状態を説明できる小説と違い，漫画などの決まったスペースで具体性・感覚性を伝えるにはオノマトペが最適であるといわれる [2]．また，作者は自分の持つイメージに適合するオノマトペを創作し用いることで受け手の感性に訴えかけるだけでなく，他の漫画との差異化を図ることができる．しかし，オノマトペは感覚的なもので個人の感覚に依存するため，客観的な評価基準がなく新たなオノマトペの作成は困難ではないかと考えられている [3]．そこで，既存のオノマトペから新たに作成されたオノマトペのイメージを推定し受け手にどう感じられるかを判定することができれば，独創性がありかつ伝わりやすいオノマトペを作成する一助となるのではないかと考えられる．

これまでにオノマトペの用法推定について，オノマトペにあるとされる音象徴を用いる研究が試みられてきた．市岡ら [5] はオノマトペの音韻情報を用いて類似したオノマトペの自動分類を試みている．また，浦田ら [6] は音響特徴量 (MFCC) によって定量化したオノマトペの意味的用法分類の研究を行っており，未知のオノマトペに対しても既存のオノマトペと同様に適切な用法へ分類することを試みている．

浦田らの研究ではオノマトペの意味的用法分類について，擬音語での実験・考察がされているが，擬態語については未調査である．擬態語についても擬音語同様に音象徴を用いた分類が可能であるか，また，音を表現する擬音語と音のない状況を表現する擬態語の言葉の性質の違いが分類結果に影響するのではないかということが考えられる．そこで本研究では，音象徴による擬態語の意味的用法分類について調査した．

連絡先: 竹内なつみ takeuchi@ai.nitech.ac.jp

2. 擬態語の意味的用法分類実験

実験として擬態語の意味的用法分類を行った．実験ではまずオノマトペの定量化を行い，その後 10 分割の交差検証法を用いた分類器によるクラス分類を行った．オノマトペの定量化には市岡ら [5] の用いた音韻と浦田ら [6] の用いた音響特徴量を使用した．分類器については，ランダムフォレスト (RF) を用いた場合とニューラルネットワーク (NN) を用いた場合の実験を実施した．また，10 分割の交差検証法を行うことで既存の擬態語の一部をテストデータ (未知の擬態語) として分類実験を行った．実験の概要を図 1 に示す．

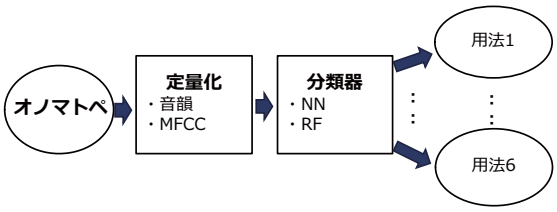


図 1: 実験の概要図

また，分類実験に使用した擬態語は，日本語オノマトペ辞典 [4] に収録されているオノマトペのうち人間の動作に関する「せき・むせる (10 語)」「飲む・酔う (52 語))」「言う・話す (68 語)」「食べる・かむ・なめる (59 語)」「騒ぐ (24 語)」「歩く・走る (50 語)」の六つの用法の擬態語計 263 語である．本実験の分類では，一つの語が一つの用法に分類されるため，二つ以上の用法に属する擬態語は除外している．それぞれの用法ごとにオノマトペに 0〜5 のラベルを付け，これを RF，NN の学習データ，テストデータに使用した．実験の評価指標には適合率・再現率・F 値を用いた．

3. 実験結果・考察

3.1 擬態語の分類実験・考察

それぞれの定量化手法および分類器を用いた場合の分類実験の結果をまとめて表 1 に示す．表 1 より，定量化手法においては，音韻を用いた場合の方が音響特徴量を用いた場合よりすべての指標の値が高いことが確認された．また，分類器については RF を用いた場合の方がより良い精度で分類された．本実験では NN を用いるにはデータ数が少なく，より多くの語が属する用法で実験する必要があると考えられる．

表 1: 擬態語実験結果

定量化手法	分類器	適合率	再現率	F 値
音韻	RF	0.65	0.65	0.64
	NN	0.56	0.56	0.55
音響特徴量	RF	0.54	0.54	0.52
	NN	0.51	0.53	0.51

定量化手法の違いによる分類結果の差について，それぞれの定量化手法を用いた場合に誤分類される語について調べたところ違いがみられた．音韻を用いた分類において誤分類された 93 語のうち 37 語が音響特徴量を用いた場合では正しく分類されていた．一方，音響特徴量を用いた分類において誤分類された 120 語のうち 64 語が音韻を用いた場合で正しく分類されていた．このことから，それぞれの定量化方法で分類しやすい語の傾向が異なるのではないかと考えられる．また，二つの定量化手法を統合して分類に利用することができれば，分類の精度を向上させることができるのではないかと考えられる．

3.2 追加実験

3.1 節の考察より，二つの定量化手法を同時に用いた分類実験を試みた．実験では，それぞれの定量化手法によって定量化された擬態語の分類結果を用いたアンサンブル学習を行った．分類器には RF を用いた．RF は複数の決定木 (弱識別器) を用いてそれぞれの結果を補い合うことで高い予測性能を得るアンサンブル学習の一種である．決定木を T 個用いる場合，入力データ $\boldsymbol{v}$ はすべての決定木に並列に入力され，この場合，オノマトペがクラスラベル c に分類される確率 $P(c|\boldsymbol{v})$ は式 (1) で表される ($p_t(c|\boldsymbol{v})$ は t 番目の決定木で得られた事後確率を表す)．

$$P(c|\boldsymbol{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\boldsymbol{v}) \tag{1}$$

RF では，この確率の値が大きいクラスが分類結果として出力される．

本実験では，まず，オノマトペの用法のクラス集合を $C = \{0, 1, 2, \cdots, 5\}$ としそれぞれのクラス $c(c \in C)$ に分類される確率について求めた．式 (1) によって音韻による定量化を行った場合の RF の確率 $P_1(c|\boldsymbol{v}_1)$ と音響特徴量による定量化を行った場合の RF の確率 $P_2(c|\boldsymbol{v}_2)$ を出力した．その後，式 (2) によって音韻を用いた場合と音響特徴量を用いた場合の結果を統合したとき最大値を取るクラス $c_{\max}$ を分類結果とした．

$$c_{\max} = \arg\max_{c \in C} P_1(c|\boldsymbol{v}_1)P_2(c|\boldsymbol{v}_2) \tag{2}$$

使用したデータおよび評価方法は先述の実験と同様である．

得られた実験結果を，表 2 に示す．表 2 では，アンサンブル学習を行った場合の適合率・再現率・F 値の他に音韻・音響特徴量のそれぞれを用いた場合の RF の分類結果を示す．表より，一つの定量化手法を用いた場合よりアンサンブル学習を行った場合の方が分類精度が向上したことが確認された．誤分類された語について調べたところ，それぞれの定量化手法でのみ誤分類されていたようなオノマトペの多くが正しく分類されていることが分かり，そのため分類精度が良くなったことが考えられる．

また，表 2 より，どの分類でも再現率 (正分類率) が 5 割以上であり，6 クラス分類のチャンスレベル (0.16) よりは高い．しかし，分類性能の議論は十分ではなく，未知のオノマトペに対して人と同等，あるいはそれ以上の精度であるかどうかの指標が必要となることが考えられる．

表 2: アンサンブル学習の結果と比較

適用手法	適合率	再現率	F 値
音韻 (RF)	0.65	0.65	0.64
音響特徴量 (RF)	0.54	0.54	0.52
アンサンブル学習	0.72	0.70	0.70

4. まとめ

本研究では，音韻と音響特徴量という二つの方法で音象徴を用いた擬態語の定量化を行い，意味的用法分類を行った．定量化方法としては音韻を用いた方が良い精度で分類ができることや定量化方法によって誤分類される語が異なることが確認できた．また，分類器に RF を用いてそれぞれの定量化手法で分類を行い，それぞれ得られた結果を用いたアンサンブル学習を行うことで，二つの定量化手法を一つの分類に用いた実験を行い分類精度の向上が確認できた．

今後は，どの程度分類できれば人と同程度の性能となるのかという指標の調査を行うことを考えている．そのため，人はどの程度オノマトペを分類できるのか調査し，機械との比較を行うことを検討していきたいと考えている．

参考文献

[1] 田守育啓：オノマトペ擬音・擬態語をたのしむ，岩波書店，2002.

[2] 西見真衣子：学生優秀卒業論文 優秀卒業論文: オノマトペの果たす役割と効果について，コミュニケーション文化，10，pp.223-240，2016.

[3] 清水祐一郎：音象徴に基づくオノマトペの印象評価システムと生成システムの設計，2016.

[4] 小野正弘 (編)：擬音語・擬態語 4500 日本語オノマトペ辞典，小学館，2007.

[5] 市岡健一，福本文代：Web 上から取得した共起頻度と音象徴によるオノマトペの自動分類，電子情報通信学会論文誌，Vol. J92-D，No. 3，pp.428-438，2009.

[6] 浦田大貴，中村剛士，加納政芳，山田晃嗣：音響特徴量を用いたオノマトペの用法分類に関する一考察，第 31 回人工知能学会全国大会，2G4-3，2017.

[2O1] [General Session] 3. Data Mining

座長:原 聡（大阪大学）

Wed. Jun 6, 2018 9:00 AM - 10:20 AM Room O (2F Kaimon)

[2O1-01] Modeling Stream Data with Fluctuated Periodicity and Periodicity Analysis Using Decision

○Koji Ohba[1], Mitsuki Kimura[1], Kazuya Watanabe[1] （1. Honda R& D Co.,Ltd）

9:00 AM - 9:20 AM

[2O1-02] Marketing data analysis by hierarchical pattern mining based on swarm intelligence

○Kazuaki Tsuboi[1], Satoshi Suga[1], Hiroaki Kanemitsu[2], Keiji Sakamoto[3], Satoshi Kurihara[1] （1. The University of Electro-communications, 2. {Trial Company, Inc., 3. Trial Holdings, Inc.）

9:20 AM - 9:40 AM

[2O1-03] Graph-based Visualization of Timeseries Information in Multiple Texts

○Hironari Kawada[1], Mina Akaishi[2], Hiroshi Hosobe[2] （1. Graduate School of Computer and Information Sciences, Hosei University, 2. Faculty of Computer and Information Sciences, Hosei University）

9:40 AM - 10:00 AM

[2O1-04] Tourist Transition Model among Sightseeing Spot based on Trajectory Data

○Takeshi Watabe[1], Hidekazu Kasahara[2], Masaaki Iiyama[2], Michihiko Minoh[2] （1. Undergraduate School of Informatics and Mathematical Science, Kyoto University, 2. Academic Center for Computing and Media Studies, Kyoto University）

10:00 AM - 10:20 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ゆらぎ周期をもつデータのモデル化と決定木を用いた周波数解析

Modeling Stream Data with Fluctuated Periodicity and Periodicity Analysis Using Decision Tree

大場 孝二 Koji Ohba　木村 光樹 Mitsuki Kimura　渡邉 和哉 Kazuya Watanabe

株式会社本田技術研究所
Honda R&D Co., Ltd.

It is important to capture the periodicity of stream data in the field of natural sciences and real world applications. However, the periodicity often has not only some fluctuations but also noises. To find out the effect of such fluctuations and noises, we model the data and analyze. In data modeling, we apply the some randomness to express the fluctuations and noises. Then, we perform a Discrete Fourier Transform to the data and analyze these power spectrums using Decision Tree.

1 はじめに

気温の変化や潮汐や天気などの自然現象だけではなく心拍数といった生理現象等周期的な現象が数多く存在する[1]. 周期を特定することができれば予測やアノマリー検出を行うことができる. しかしながら, 現実世界において事象が完全に同じ周期で起こることはなく, 一定の幅で周期がゆらぐことが多い. さらに測定装置にノイズが混入することで本来あるべき周期が歪められることもある. したがって, ゆらぎやノイズを前提とした周期分析が求められる.

本稿では, 完全に周期的とはいえないゆらぎ周期をもつ事象を, ランダムデータ列を用いてモデル化した. さらにそのデータ列にノイズを加えた. こうして生成したデータ列に対して離散フーリエ変換を行い, パワースペクトルを得ることで, ゆらぎやノイズの影響がどのような形で表れるのかを明らかにした. さらにパワースペクトルを決定木によって分類することで, ノイズの影響がパワースペクトルのどの部分に影響を与えるのかを考察した.

2 データ生成と解析手法

2.1 データ定義

0 か 1 の二値をとる時系列デジタルデータを考える. a[x]を連続する x 個のデータ列で, 0 が x 個並んだデータ列とする. この時, 完全に周期的なデータ列を, ある整数 n を用いて [a[n-1], 1] が繰り返されているものとして定義する. 次にゆらぎ周期を持つデータ列を定義する. w[x]を連続する x 個のデータ列で, x 個のうち一か所のみが連続する x 個のデータの中で 1 で, その他は 0 となるデータ列とする. ただし, 1 となる箇所は完全にランダムである. この時, ゆらぎ周期を持つデータ列を, ある整数の組(q,p,m,n)を用いて, [a[q], [a[n-m], w[m]]*p]が繰り返されているものとして定義する. ただし, [a[n-m], w[m]]*p は[a[n-m], w[m]] を p 回繰り返したものである(図 1).

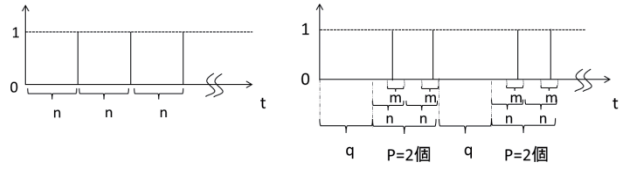


図 1 完全に周期的なデータ(左)とゆらぎ周期を持つデータで p=2 の場合(右)

連絡先:東京都港区赤坂 5 丁目 3 番地 1 号
Koji_Ohba@n.t.rd.honda.co.jp

2.2 生成するデータ列

生成するデータ列を以下のクラス 0,1,2 の 3 種類とする.

クラス 0 : 完全に周期的なデータ列, ゆらぎ周期をもつデータ列

クラス 1 : ランダムなデータ列

二項分布に従って 0,1 を各時刻にランダムに割り当てる. 各試行の成功確率を p とする.

クラス 2 : クラス 0 にクラス 1 を加えたデータ列

クラス 0 に属するデータ列に対して, クラス 1 に属するデータ列を加える. 値は各時刻のデジタル値を足して, 0 なら 0 のまま, 1 以上なら 1 とした.

2.3 分析手法

周期的な事象を判別するのに重要な特徴を分析するために以下を行った. まず, クラス 0,1,2 のデータ列を人工的に生成した. 生成したデータ列に対して, フーリエ変換を行い, パワースペクトルを得た. 得られたパワースペクトルデータを決定木にかけて分類し, 分類条件について分析した.

3 実験

本来であれば無限長のデータ列を解析しなければならないが, 今回は連続で 1460 個のデータが並んだデータ列を生成し, 分析した.

3.1 ノイズの影響によるパワースペクトルの変化

まず, クラスが 0,1,2 に属するデータ列を生成し, フーリエ変換を行い, 得られたパワースペクトルの可視化を行った. 生成したデータ列は, 以下の 5 種類である. クラス 0 のデータ列として, ①:n=28 の完全に周期的なデータ列と②:(q,p,m,n)=(24,1,4,4) のゆらぎ周期をもつデータ列の 2 種類を生成した. クラス 1 のデータ列として, ③:p=0.06 のランダムなデータ列を生成した. クラス 3 のデータ列として, ④:①に③を加えたものと⑤:②に③を加えたものを生成した. これら 5 種類のデータ列とそれらにフーリエ変換をして得られるパワースペクトル結果を図 2,3,4 に載せる.

図 2 上はデータ列①の, 図 2 下はデータ列②の結果である. 図 2 上のパワースペクトルは周期に相当する振動数(frequency) 1/28 の箇所にピークが立っている. 1/28 の整数倍にもピークが立っているが, これは 0,1 という離散データをフーリエ変換したためである. 一般に離散データに関しては, ある特定の振動数の整数倍の位置にピークが立つ. 図 2 の下図のパワースペクトルも同様に 1/28 の位置にピークが立っているが, スペクトル値の大きさが異なる. 1/28 の整数倍のピークが, 右にいくにつれて

小さくなっている．これは，ゆらぎ周期に導入したランダム列 w[x]の影響によるものである．小さな振動数帯では 28 という大きな周期を反映して 1/28 にピークが立つが，w[x]を反映して振動数が大きくなるとノイズのようなパワースペクトルとなる．

図 3 はデータ列③の結果である．パワースペクトルに特徴的なピークが見られず，特徴的な周期を持たないことがわかる．ランダムノイズでは，パワースペクトル値が全体に均等に散らばる．

図 4 上はデータ列④の，図 4 下はデータ列⑤の結果である．図 2 と比較すると同じ振動数のところにピークが立っているが，ピーク自体は小さくなっている．ランダムノイズがパワースペクトル値の値を全体的に均等にする役割を果たすからだと考えられる．

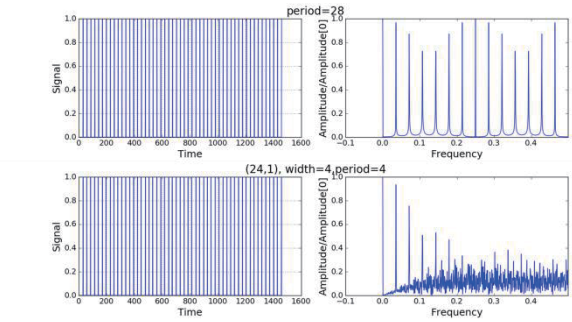


図 2 クラス 0 の 時系列データ（左）と パワースペクトル（右）．
上図はデータ列①，下図はデータ列②．

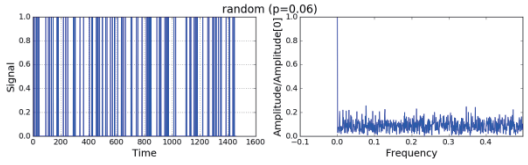


図 3 データ列③．

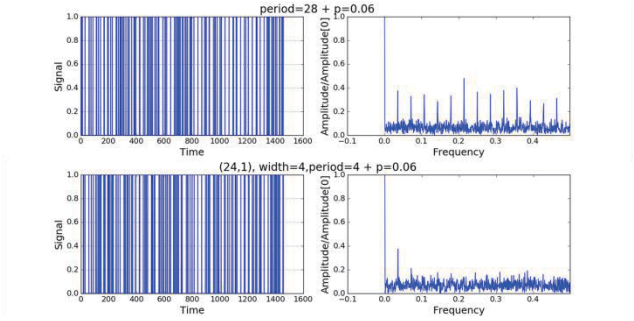


図 4　上図はデータ列④，下図はデータ列⑤．

3.2 決定木による分類結果

データを生成して，それぞれのデータのパワースペクトルを決定木にかけ，クラス分類に必要なパワースペクトル値を明らかにした．生成したデータは，クラス 0,1,2 それぞれ 1600 列である．クラス 0 は 2 種類の完全に周期的なデータ列と，3 種類のゆらぎ周期を持つデータで構成される．完全に周期的なデータは n=4, 28 で，ゆらぎ周期を持つデータは(q, p, m, n) = (8, 20, 1, 4), (20, 2, 4, 4), (24, 1, 4, 4)である．それぞれを 320 列ずつ生成した．クラス 1 は p=0.01, 0.02, …, 0.08 の 8 種類のランダムなデータ列をそれぞれ 200 列ずつ生成した．クラス 2 は，クラス 0 の 5 種類それぞれにクラス 1 の 8 種類をそれぞれ加えた 40 種類をそれぞれ 40 列ずつ生成した．以上のように生成したデータ列に対してフーリエ変換を行って得られたパワースペクトルを特徴として決定木によってデータ列のクラス分類を行った．分類の指標にはジニ係数を用いた．完全に分類できるまで決定木を動作させた結果を図 5 に載せる．

図中の各ノードについて説明する．Samples は分類対象となるデータ列数であり，value に各クラスのデータ列数が表示されており，左から順にクラス 0,1,2 のデータ列数である．gini はそのノードに含まれるデータ列のジニ係数である．ノードの分割条件は各ノードの一番上に書かれており，ルートノードでは振動数 (frequency) が 1/365.0 の時のパワースペクトル値が 0.001 以下であるかどうかが条件となっている．この条件を満たすデータ列は，図中で左下のノードに分類され，そうでないデータ列は右下のノードに分類される．

第一分岐でクラス 0 をほぼ分類できた．第一分岐では，振動数 1/365.0 でのパワースペクトル値の大小によってクラス 0 とクラス 1,2 を分類する．その振動数に注目したのは，クラス 0 は低振動数領域のパワースペクトル値が小さいが，ランダムノイズの影響によってクラス 1 のその領域のパワースペクトル値が大きなったことだと考えられる．第二分岐では，振動数 1/28.0 の振動数のパワースペクトル値の大小によってクラス 1 とクラス 2 を分類する．これは，クラス 1 が 28 という周期を持つためだと考えられる．第三分岐まで行うと，クラス 1,2,3 をほぼ完全に分類できた．

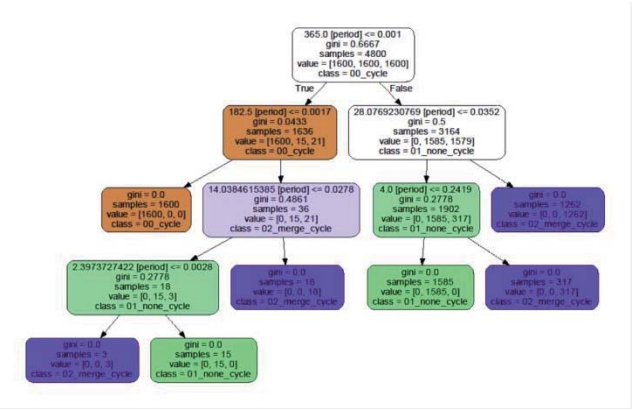


図 5: クラス 0,1,2 の決定木による分類

おわりに

本研究では，ゆらぎを持つ周期データ列をモデル化し，それらのデータ列に離散フーリエ変換を行い，結果を決定木で分類した．フーリエ変換後のパワースペクトル値はノイズの影響で減少する．パワースペクトルの値を見ることで，完全なノイズデータとノイズを加えた周期データの分類ができることを確認した．決定木の分類条件を確認することで，データ列がどのような周期を強く持っているのかを判定することができる．今後は，実データを用いた解析を行っていきたい．実際のデータには，ランダム性が成功確率 p の二項分布に従うものだけでなく，ほかの確率分布に従うものもあることが予想される．そのような状況においても，どの周期を強く持っているかの判定ができるかかの解析を行いたい．

参考文献

[1] VLACHOS, Michail; YU, Philip; CASTELLI, Vittorio. On periodicity detection and structural periodic similarity. In: *Proceedings of the 2005 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, 2005. p. 449-460.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

群知能メカニズムを用いた時系列階層型パタン抽出法による マーケティングデータ分析

Marketing data analysis by hierarchical pattern mining based on swarm intelligence

坪井一晃 *1 須賀聖 *1 金光博明 *2 坂本圭司 *3 栗原聡 *1
Kazuaki Tsuboi Satoshi Suga Hiroaki Kanemitsu Keiji Sakamoto Satoshi Kurihara

*1電気通信大学
The University of Electro-Communications

*2株式会社トライアルカンパニー
Trial Company, Inc.

*3株式会社トライアルホールディングス
Trial Holdings, Inc.

The behavior patterns in everyday life such as home, office, and commuting, and buying behavior model by day of the week, season, location have hierarchies of various temporal granularity. Generally, in usual hierarchical data analysis, basics hierarchical structure is given in advance. But it is difficult to estimate hierarchical structure beforehand for complex data. Therefore, in this study, we propose algorithm to automatically extract both hierarchical structure and pattern from time series data using swarm intelligent method. We try to extract the patterns from real marketing data.

1. はじめに

創発というキーワードは大変興味深い．創発とは，部品や要素が集合することによって，単独では見られなかった様相が発現することである．生物は創発という現象を上手く活用している．たとえば，アリやハチといった昆虫たちである．それぞれのアリやハチの単体では優れた知能を持つとは言いがたいが，一つの集団として観察すると極めて優れた知能を発揮している．アリを例に一つ具体的に説明する．アリの採餌行動においてアリ単体では餌をみつけ巣に戻るときにはフェロモンという揮発性物質を移動経路に付加することと，餌を探索するにあたりフェロモンが残る経路を好んで探索しやすいという本能的な指針によって行動するのみである．しかし，これを集団として行うことによって，餌場と巣におけるほぼ最短の経路に対して列をなし行進している．このような自然界において単体では単純ながらも集団を形成することで複雑な性質や振る舞いを誘起する様子から着想を得る，群知能の研究が盛んである．

インターネットの急激な進化やコンピュータの高性能化および低価格化を背景に，社会のいたるところに IT を活用したシステムが導入されるようになっている．導入されたシステムを利活用するたび，システムからはデータが生成され続けている．我々が意思を向けずとも稼働しているシステムも存在し，日々多種多様にわたるデータが大量に生成・蓄積されている．近年，この大量に蓄積されるデータを用いることで有用な情報を得られるのではないかという期待が高まっている．このように，蓄積されたデータから有用な知見を得るための技術としてデータマイニングに注目が集まっている．データマイニングにおける重要な技術の一つに，データベースの中から頻出するアイテムの組み合わせを抽出するパターンマイニングがある．パターンマイニングを用いることで．例えばコンビニエンスストアやスーパーマーケットなどの小売店におけるよく同時に購入される商品などを効率的に発見できるようになる．さらに，このパターンマイニングを現実世界において度々重視される時系列に着目し発展させた，シーケンシャルパターンマイニングなどの技術が提案されている．時系列性を考慮したパターンを抽出することによって，この商品が購入された次にはこの商品が購入されるはずであるという予測することができるようになる．これらのパターンマイニングでは，解析対象のデータベースはトランザクションと呼ばれる単位で分割されている．したがって，このトランザクションの中という範囲のなかでパターンと考えられるアイテムの組み合わせを抽出している，つまり，データベース中から分析した粒度というもの明確に設定し，それに合わせてトランザクションに分割している．複数の異なる粒度の分析を行うときには，それぞれの設定に応じたトランザクションを生成しその都度分析を行う．

しかし，現実世界を考えたときに，日常における自宅やオフィスなどをつなぐ通勤といった行動様式や小売店における曜日や季節などによる商品の売れ方などの中には，様々な時間的粒度の異なるパターンが含まれていることは明白である．そこで，本研究では，異なる時間的粒度を階層的にとらえ，時系列データから階層構造とパターンの両方を自動的に抽出する手法を提案する．

今回，我々は群知能分野における Ant Colony Optimization（ACO）アルゴリズムに着目している．ACO アルゴリズムは自然界におけるアリの群れとしての採餌行動をモデル化したものであり，巡回セールスマン問題を代表とする様々な最適化問題に対して良い性能を示すメタヒューリスティクスであることが知られている．

既存のパターンマイニング技術でも度々問題となる項目として計算量の課題がある．網羅的にパターンを探索する場合，入力データが増えることによって指数関数的に計算量が増加する．そこで，ACO アルゴリズムから着想を得たアルゴリズムを構築することによって，分散処理を行う等の対応を可能とすることで，有限な資源にそった準最適な解を得ることに期待できる．また，分析対象が時系列データであるということは，時間経過に伴いパターンが変化することを容易に想定できる．このような問題に対しても，ACO アルゴリズムが有する優れた適応性，頑健性を発揮できるアルゴリズムの構築を試みる．

本稿の構成を次に示す．2 章では，パターンマイニングや ACO アルゴリズムに関する関連研究について述べる．3 章で

連絡先: 坪井一晃，電気通信大学，東京都調布市調布ヶ丘 1-5-1，042-443-5660，tsuboi@ni.is.uec.ac.jp

は，提案手法となる群知能メカニズムによる時系列階層構造を有するパターンの抽出手法について述べる．4 章では，提案手法に対してマーケティングデータ用いた実験を行い消費者の購買パターンについて考察する．5 章では，まとめと今後の課題について述べる．

2. 関連研究

大量のデータから頻出するアイテムセットの組み合わせをパターンとして抽出する手法がパターンマイニングである．データベース上から頻出するアイテムセットのパターンを抽出するためのアルゴリズムとして Apriori アルゴリズム [1] が有名である．Apriori アルゴリズムは，あるアイテムセットの出現頻度が高いためにはその部分集合のアイテムセットの出現頻度が高い必要があるという考えから，小さいアイテム集合から順にデータベースにおける出現回数を数える幅優先探索を実行する．あるアイテムセットが頻出とならない場合，その上位となるアイテムセットである，そのアイテムセットは包括するより大きいアイテムセットについて探索する必要がないとみなす．このことから解の探索範囲空間を減らしている．しかし，解の候補となるアイテム集合が生成されるたびに出現回数を数えるために全ての入力データベースを調べなければならなくなる．そこで，入力データベースが膨大なものとなったときに，計算時間や計算機のメモリなどに問題が生じる．また，Apriori アルゴリズムでは，アイテムの時系列性を考慮していないといった課題が指摘されている．

そこで，データベース上の時系列性を考慮したアイテムのパターン抽出を行うために Apriori アルゴリズムを拡張した AprioriAll アルゴリズム [2] やパターン抽出にあたって時間制約などの条件をつけることで計算処理の高速化を図った GSP アルゴリズム [3] が提案されている．また，深さ優先型の探索法を採用したシーケンシャルパターンマイニング手法として PrefixSpan アルゴリズム [?] が提案されている．これらのアルゴリズムは計算処理の高速化を図ったものであり，入力データを一様に等しく参照している．しかし，マーケティングなどの現実問題に適用させることを考えると，流行や季節などの影響を受けながら消費者の行動も変化し続けている．このようなパターン自体が変化するような場合には柔軟に対応することが難しい．

そこで, 我々は群知能分における ACO アルゴリズムに着目している．ACO アルゴリズムの特徴として，どのような環境に対しても解を生成することができる適応性や，環境の変化に対しても柔軟に解を再探索できる頑健性が挙げられる．このような特徴は現実問題に対して有効であると考えられる．

そもそも，ACO アルゴリズムは自然界におけるアリの採餌行動をモデル化したものであり，最適化問題の巡回セールスマン問題に対する解法として提案された．ACO アルゴリズムにおいて，アリは通過した経路にフェロモンを残すことと，アリはフェロモンの濃度が濃い経路を好んだ経路選択を行うという二つをアリの行動の前提とする．この前提によって，アリが集団として行動するにつれて，最短経路を探索するアリの数が増加しフェロモンの濃度がいっそう濃くなる．一方，フェロモンは揮発性物質であるために時間経過にともない蒸発する．結果的に，アリが通過する頻度が低い経路のフェロモンの濃度は薄くなる．環境に残留するフェロモンが最短経路問題の解となる．結果として，ACO アルゴリズムの適応性や頑健性の高さが特徴となっている．

ACO アルゴリズムをデータマイニングのクラスタリング手法として応用した研究として Ant Miner+アルゴリズム [4] が提案されている．このアルゴリズムは分類結果がわかりやすい分類器でありながら精度が高く，データが分散した環境にも適するとされる．また，ACO アルゴリズムをパターンマイニングに応用した研究として Tamaki ら [5] の研究がある．Tamaki らは，センサが人の行動を読み取り反応することから，連続した人の行動から隣接するセンサは連続して反応するという前提によって，センサの隣接関係の推定を行っている．ACO アルゴリズムを応用したことで，複数の人による連続してセンサが反応する場合への対応や，センサ自体の故障や移動といった出来事にも柔軟に対応している．

このように，もともとは最適化技術として提案された群知能分野のアルゴリズムであるが，単純な行動ルールに基づいたエージェントの移動と環境に対するフェロモンの付加および蒸発を応用することで，柔軟なシステムの構築が達成できる，本研究においても，ACO アルゴリズムが有する適応性や頑健性に着目し，群知能メカニズムによる時系列階層構造を有するパターンの抽出を試みる．

3. 群知能メカニズムによる時系列階層構造を有するパターンの抽出手法

提案アルゴリズムでは，アイテム集合 $I = \{i_1, i_2, \cdots, i_m\}$ で構成させる一つの時系列データ $D_n = \{d_1, d_2, \cdots, d_n\}$（ただし，$d \subset I$）から，時系列にそって頻出するアイテムの組み合わせをパターンとして順序関係および出現間隔を考慮しながら抽出する．

提案するアルゴリズムでは時系列データから時刻順に入力し，それぞれの入力に合わせて嗜好を有するアリエージェントが探索しフェロモンの付加を行う．また，それぞれの入力に対するアリによるフェロモンの付加が完了するとともに，時間経過となるフェロモンの蒸発を行う．提案するアルゴリズムを擬似コード（アルゴリズム 1）に示す．

Algorithm 1

Input: 時系列データ D_n
Output: フェロモンが付加された仮想マップ M_t

```
t := 0
while t < n do
    アリエージェントの生成・追加
    アリエージェントによる探索および M_t に対するフェロモンの付加
    フェロモンの蒸発
    探索期間が終了したアリの削除
    t := t + 1
end while
```

3.1 仮想フェロモンマップ M_t の用意

まず，フェロモンを付加するための環境として仮想マップ M_t を用意する．時刻 t における仮想マップ $M_t = (P, v(P))$ は，パターン集合 $P \ni p$ およびそのパターン p に対するフェロモン量 $v(p)$ を記録した仮想マップである．パターン $p = (F, S, l)$ は，順序関係を考慮した複数のアイテムまたは既出のパターン（F）および，その複数のアイテムまたはパターンの時間間隔（S），パターンとしての階階数（l）からなる．

3.2 アリエージェントの生成

各時刻において，アリエージェントは $ANTS_{max}$ だけ追加生成する．それぞれのアリエージェントはパターンを探索する

にあたって，順序性を有する複数の嗜好 $F \subset \{I, P\}$ とそれぞれの嗜好に対しての好む間隔 S を有する．また，一匹のアリエージェントが探索できる探索期間 r が与えられる．

アリエージェントには 2 種類あり，1 つは最小のアイテムの組み合わせから嗜好を選択するタイプであり，もう 1 つは生成されたフェロモンマップから嗜好を選択するタイプである，最小のアイテムの組み合わせから嗜好を選択するタイプのアリは，そのアリエージェントが生成された時刻の入力からランダムに第一の嗜好を選択し，与えられる嗜好間のパラメータ分だけ時間経過したときの入力からランダムに第二の嗜好を選択する．

パターンからパターンを階層的に構築するために，フェロモンマップから嗜好を選択するアリエージェントのそれぞれの嗜好 f は，フェロモンマップ M におけるフェロモン量 $v(P)$ の割合を確率としてパターンから選択される．つまり，パターン p が選択される確率 $q(p)$ は式 1 に従う．

$$q(p) = \frac{v_p}{\sum v(P)} \tag{1}$$

また，アリエージェントに設定する嗜好に対する間隔 S は一様乱数によって定める．アルゴリズムを**??**に示す．

Algorithm 2 アリエージェントの生成・追加

```
ANT :=0
while ANT < ANTS_max do
  if t < T_1 | ANT < ANT_θ then
    s := random(1, s_max)
    d_t から f_1 の選択
    d_{t+s} から f_2 の選択
  else
    M_{t-1} から f_1 と f_2 の選択
    if f_1 を構成する嗜好の 2 番目のものと f_2 を構成する
    嗜好の 1 番目のものが同一 then
      s := random(0, s_max)
    else
      s := random(1, s_max)
    end if
  end if
  f_1 探索モードに設定
  ANT := ANT + 1
end while
```

3.3 アリエージェントによる探索及びフェロモンの付加

それぞれのアリエージェントは設定された嗜好および嗜好に対する間隔に基づいて入力された時刻におけるデータに対して探索を行う．アリエージェントが嗜好および嗜好に対する間隔からなるアイテムまたはパターンの組み合わせを発見することができた場合，そのアイテムまたはパターンの組み合わせを一つのパターン p とみなし，そのパターンに対するフェロモン量を 1 だけ増加させる．なお，発見したパターンが，新規登場であった場合にはパターン番号として順次固有のナンバリングを割り当てる．また，アリエージェントが嗜好およびその間隔から発見したパターンにおいて，アリエージェントが有する嗜好のうち最大の階層数に 1 加えたものをそのパターンの階層数とする．なお，初期のアイテム集合はそれぞれ階層数を 1 とする．

Algorithm 3 アリエージェントによる探索

```
ANT := 0
while ANT < ANTS_all do
  if f_1 探索モードか then
    if D_t に f_1 が出現 then
      f_2 探索モード（嗜好までのカウント s_1）に設定
    end if
  else
    if 嗜好までのカウントが 1 then
      if D_t に探索 f が含まれる then
        if f_next があるか then
          f_next 探索モード (嗜好まで s_next) に設定
        else
          発見パターンに対してフェロモンの付加
        end if
      else
        f_1 探索モードに戻る
      end if
    else
      嗜好まで (のカウント):= − − 1
    end if
  end if
  ANT := ANT + 1
end while
```

3.4 蒸発にともなうフェロモン値の減算

時間経過に伴い，残留するフェロモン量を減少させる．フェロモン量の減少に際して，フェロモン量自体は指数関数的に減少することおよび，パターンのパターンとなっているような深い階層のパターンに関しては出現率が低いと考えられるため，減少するフェロモン量は少ないことを考慮した．

$$v = (\exp(\log(v+1) - 0.00001) - 1) \times (1 - 10^{-l}) \tag{2}$$

4. マーケティングデータを用いた実験

実際のマーケティングデータに対してアルゴリズムを実行し，消費者の購買のパターンについて考察を行う．

4.1 マーケティングデータ

本研究で使用したマーケティングデータは株式会社トライアルホールディングスより提供を受けた，2014 年 1 月 1 日から 2016 年 12 月 31 日までの 3 年間の 1 消費者による商品購買データである．商品購買データにおいては，入力の粒度として商品のカテゴリ名を採用している．また，1 日における購買商品を 1 ステップとし，1095 ステップの時系列データとして利用する．実験の対象消費者は 3 年間のうちに 1000 日以上購買した経歴がある消費者から，30 代女性の消費者を対象とした．

4.2 実験

各ステップ 10000 のアリエージェントを生成する．アリエージェントに与える嗜好は 2 つとした．なお，100 ステップまでに生成されるすべてのアリエージェントは，生成された時刻において順次パラメータを設定する．また 100 ステップ以降は 100 のアリエージェントは生成された時刻において順次パラメータを設定する一方，残りの 9900 のアリエージェントはフェロモンマップ M_{t-1} から嗜好を選択する．また，2 つの嗜好に対する間隔は最大 11 最小 0 の一様乱数から選択される．

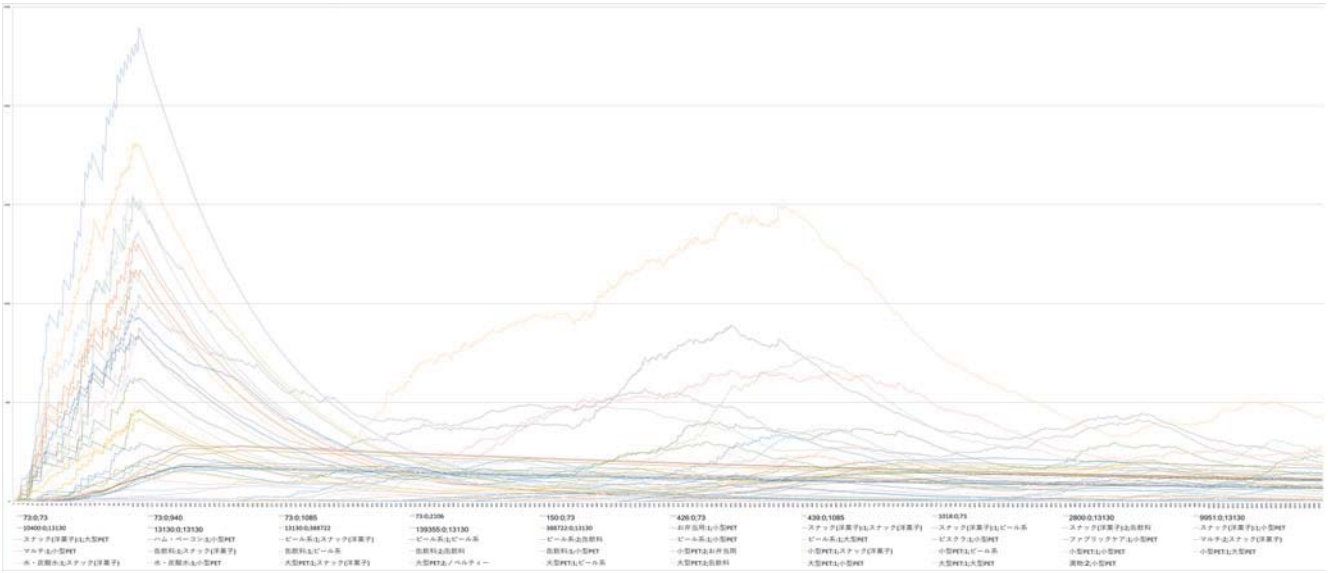


図 1: 30 代女性のパターンのフェロモン量の推移

なお，2 つの嗜好に対する間隔において 0 が選択されるのは，アリエージェントの有する 2 つの嗜好をみたときに一つ下の階層において重複が許容される場合のみである．

4.3 実験結果

各消費者の商品購買データに対して，提案アルゴリズムを実行した．300，400，500，600，700，800，900，1000，1095 ステップのそれぞれの時点におけるフェロモン量が多く残る 50 パターンについて，フェロモン量の推移をそれぞれ図 1 に示す．ただし，得られたパターンのうち嗜好間の間隔のみが違うものについては，嗜好間の間隔が小さいもののみを取り出した．

4.4 考察

実験の結果，図 1 より，前半部の 100 ステップ目を頂点に 250 ステップまでにかけて，(缶飲料，2，缶飲料）に多くのフェロモンが残っているのにたいして，それ以降から 900 ステップまでにかけては（小型 PET，1，小型 PET）に多くのフェロモンが残っていることがわかる．2014 年 3 月あたりまでは缶飲料を隔日で購入していたものが，小型 PET ボトルを連日購買するようになった．このように，本手法を用いて各消費者の購買行動を分析することで変化する消費者の購買行動に柔軟に抽出できることがわかる．

5. おわりに

本研究では，群知能アルゴリズムに代表される ACO アルゴリズムを基にしたパターン抽出手法を提案するとともに，実際のマーケティングデータを用いて消費者の購買パターンについて抽出する実験を行った．実験の結果，30 代女性の隔日で缶飲料を買っていたパターンが，連日の小型ペットボトルを購買するパターンに変化する様子が確認できた．

今後の課題として，アリエージェントに対する適切なパラメータに対する知見の獲得が考えられる．また，現在のアルゴリズムでは，一度アリエージェントにパラメータを設定したらそのパラメータに従い探索し続けるが，各アリエージェントが入力データに対して柔軟に嗜好や嗜好間の間隔などを変化されることによってより適応性を向上させられることも考えられる．

また，実際のマーケティングデータなどの時系列データに対して提案するアルゴリズムを適用させることで，消費者の購買パターンなどの変化する様子を抽出することを目指す．

参考文献

[1] Agrawal, Rakesh, and Ramakrishnan Srikant. ”Fast algorithms for mining association rules.” Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215. 1994.

[2] Agrawal, Rakesh, and Ramakrishnan Srikant. ”Mining sequential patterns.”Data Engineering, 1995. Proceedings of the Eleventh International Conference on. IEEE, 1995.

[3] Srikant, Ramakrishnan, and Rakesh Agrawal. ”Mining sequential patterns: Generalizations and performance improvements.” Springer Berlin Heidelberg, 1996.

[4] Martens David, et al . ”Classification with ant colony optimization” IEEE Transactions on Evolutionary Computation 11.5, pp.651-665,2007.

[5] Tamaki, Hiroshi, et al. ”Pheromone Approach to the Adaptive Discovery of Sensor-Network Topology.” Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 02. IEEE Computer Society, 2008.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

事象情報グラフによる複数文書中の時系列情報可視化

Graph-based Visualization of Timeseries Information in Multiple Texts

河田 裕成[*1] Hironari Kawada
赤石 美奈[*2] Mina Akaishi
細部 博史[*2] Hiroshi Hosobe

[*1] 法政大学大学院情報科学研究科 Graduate School of Computer and Information Sciences, Hosei University
[*2] 法政大学情報科学部 Faculty of Computer and Information Sciences, Hosei University

Newspapers and other media often describe particular topics over some periods by giving a series of texts that introduces new information and corrects previous information. By analyzing such texts, we can understand how a topic changed over time and how the media treated it. However, analyzing such texts manually is time-consuming. To solve this problem, we propose a method for visualizing texts that describe a topic changing over time. For this purpose, we first generate an event information graph from such texts by using an exiting method. Then we transform it into two different graphs. The first graph shows changes in the topic that we obtain by separating the initial graph and comparing the resulting time-dependent graphs. The second graph depicts a summary of the topic that we obtain by extracting high-frequency words from the initial graph. The results of our experiment show that our method can visualize changes and important information described in the texts.

1. はじめに

新聞等のメディアが報道する時事情報は，時間と共に刻々と変化する．新たな情報が次々と報じられる中で，時には過去の情報が間違いだったと判明することもある．こうした時系列的な情報変化に関する文書群を分析することで，時事情報に関する状況の推移と，メディアにおけるその扱いを知ることができる．しかし，長期間に渡って報じられた時事情報について，文書群中の情報変化を人手により分析する作業には手間と時間がかかる．

そこで，本論文では，文書中の情報を表現する有向グラフに対してグラフ操作を行うことによる，情報変化の可視化手法を提案する．具体的には，先行研究の田代らの手法[田代 2013]を用いて文書に含まれる実世界の出来事の情報をグラフ化した後，それらの時系列的な比較と重要部分の抽出を行い，それぞれを別のグラフとして可視化する．これにより，時系列情報に関する情報変化や，時系列全体を通して出現する重要な情報の分析支援を試みる．

2. 関連研究

情報の時系列的変化の可視化手法としては，伊藤らの研究[伊藤 2013]がある．伊藤らは，係り受け解析により動詞句と名詞句のペアをイベントとして抽出した．また，イベント群の時系列的な構成や頻度の変化を 3 次元的に可視化した．伊藤らの手法を用いることで，特定キーワードに関する話題の変化を可視化することができる．しかし，キーワードとその関連単語の間にある関係は明示されず，ユーザの推測に任せている．

田代ら[田代 2013]は，文書から事象を抽出し，単語間の関係をネットワークとして可視化する手法を提案した．ここで，田代らの定義する「事象」とは，実世界の出来事を表す情報のことである．具体的な手法は，まず，係り受け解析とCRFによって文書中における 5W1H 等の事象属性句を推測する．その後，人手で定義したルールによって事象属性句間の関係を推測し，事象情報として LOD により表現する．田代らはこの手法を用いて特定テーマに関する文書群中の事象情報を可視化し，メディアによる情報の差異を分析した．一方で，田代らは情報の時系列的変化は研究の対象としていない．本論文では，田代らの手法を用いて文書中の事象情報を可視化する．加えて，文書の公開時期を考慮することで，文書の題材に関する情報の推移を分析する．

連絡先：河田裕成，e-mail：17t0009@cis.k.hosei.ac.jp

3. 提案手法

本論文では，時系列的な情報変化を扱うために，特定テーマに関する，情報公開日時の明示された文書を分析対象とする．その例としてニュース記事，歴史年表，SNS への投稿文等が挙げられる．図 1 に示すような提案手法の流れを説明する．

まず，2 章で述べた田代らの手法を用いて文書から事象情報を抽出し，文書中の事象情報を有向グラフとして得る．この事象情報グラフを 2 通りのアプローチで操作することで，時系列的な情報変化と文書のテーマに関する重要情報をそれぞれ可視化する．以下，3.1 節において，事象情報グラフの概要を述べる．その後，事象情報グラフを用いた，情報変化の可視化手法を 3.2 節で，重要情報の可視化手法を 3.3 節でそれぞれ説明する．

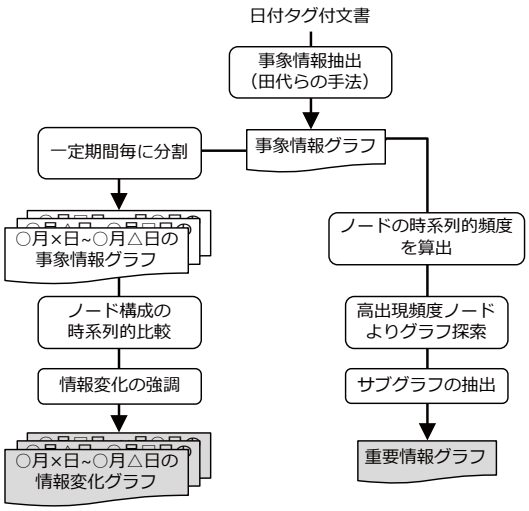


図 1 提案手法の流れ

3.1 事象情報グラフ

田代らの手法による事象情報グラフは，形態素からなる事象属性句をノードとし，「Activity」や「Status」，「Location」など、13種の関係をエッジとしている．これらのエッジは，文書内における事象属性句の係り受け関係や，それを構成する形態素の品詞情報，手がかり語を用いるルールによって設定される．なお，ノードとなる語は形態素解析等により原形化することで，活用による表現揺れを可能な限り吸収している．例えば，「会社員の男が本を買った」という文から生成される有向グラフでは，「会社員」に Modifier 関係で修飾された「男」から，「買う」に動作を表す Activity 関係でエッジが張られる．さらに「買う」は，動作対象を表す Object 関係で「本」に接続される．

3.2 情報変化グラフの生成

時間経過による新たな出来事の発生や，過去の情報に対する補足や訂正がなされることを，時系列的な情報変化とする．これを可視化した情報変化グラフは，時系列中のある期間に出現した事象情報群と，それらが直前の期間からどのように変化したかを表現する．その中で，事象情報は 3.1 節で述べたノードとエッジからなる有向グラフとして表現されるが，各ノードが情報変化の種別を表すように着色されているという点で異なる．ここでの種別とは，直前の時期に出現したノードの①削除，②継続した出現，対象時期における③新規ノードの出現の 3 つである．

ユーザは着色されたノードとそれらの接続関係を見ることで，情報変化を分析する．例えば，②継続出現のノードに対して①削除と③新規出現のノードが状態を表す「Status」関係で接続されている場合，①のノードが表す単語に関して，時間経過による状態の推移が発生したと考えることができる．図 1 左側のフローに示す，情報変化の分析と可視化手法を説明する．

まず，文書の公開日情報を用い，事象情報グラフをユーザの指定する一定期間毎に分割する．各グラフは，対応する期間に出現した事象情報群を表す．次に，グラフ群を時系列順にソートし，直前期間のグラフとのノード構成の差異を計算する．これにより，先に述べた①，②，③が得られる．可視化では，各グラフ中の①のノードを青色，②を黒色，③を赤色に着色する．

3.3 重要情報グラフの抽出

本論文での文書の重要情報とは，文書の扱うテーマを説明するものとする．例えば，自動車事故に関する文書であれば，当事者の名前や事故の起きた状況等が重要情報にあたる．事象情報グラフにおいて，重要情報は文書のテーマに強く関連する単語ノードや，それらと近い距離にあるノードからなるサブグラフとして現れると考えた．そして，文書のテーマに強く関連する単語は，そのテーマを扱う文書群において時系列に関わらず出現しやすいことに注目した．以上より，重要情報グラフは事象情報グラフから抽出したサブグラフに対して，ノードが表す単語の出現時期が判別できるように，ノードの着色処理を行うことによって可視化することとした．

ユーザは着色されたノードによるサブグラフを見ることで，高頻度で出現する単語にどのようなものがあるか知ることができる．また，高頻度語間の関係や，それらを低頻度のノードが表す単語とその出現時期を分析することができる．以下に，図 1 右側のフローで表される，重要情報の可視化手法を説明する．

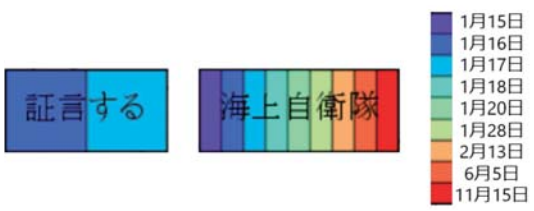


図 2　単語の出現時期を示すノードの着色例

始めに，文書の公開日情報を用いて，単語の時系列的な出現頻度を計算する．これは，時系列の各期間での単語の出現有無を 2 値で求め，それらを合計したものである．次に，出現頻度が閾値以上であるノードを求める．その後，これらのノードを始点とし，そこから指定距離以内で到達できるノードを探索する．最後に，探索によって得られたノードと始点ノードから構成されるサブグラフを抽出する．なお，ノードの出現頻度に関する閾値やグラフ探索の距離は，パラメータとして変更可能である．

可視化では，文書の公開時期ごとに色を割り当て，各ノードを出現時期に対応する色で着色している．図 2 に着色例を示す．ここでは，ノード「証言する」は時系列的に初期の 2 文書のみに出現し，ノード「海上自衛隊」は全ての文書において登場した．よって，「証言する」は青色 2 色で，「海上自衛隊」は青色から緑色，そして赤色までのグラデーションとなるように着色されている．

4. 適用実験

実際の文書群に対して本手法を適用し，情報の変化と重要情報を可視化できているか評価する．本論文では，2014 年 1 月 15 日に発生した海上自衛隊の輸送艦「おおすみ」とプレジャーボート「とびうお」の衝突事故を対象とし，実験を行った．運輸安全委員会の船舶事故調査報告書[*1] によると，この事故では，左舷前方から接近してきた「とびうお」と，それに対し右転して回避しようとした「おおすみ」が衝突したとされる．その結果，転覆した「とびうお」の乗員 2 名が死亡した．この事故に関して，2014 年 1 月から 11 月までに公開された毎日新聞社の新聞記事 9 日分 14 記事を人手で収集し，事象情報を可視化した．図 3 に田代らの手法による事象情報グラフの全体図を示す．このグラフは，431 個のノードと 885 のエッジからなる．提案した 2 つの手法による可視化結果について，以下に述べる．

4.1 情報変化グラフ例

可視化における記事の分割期間は 1 日単位とし，可視化を行った．提案手法では，時系列的に分割した期間毎に事象情報グラフが出力される．本節では，出力例として 2014 年 1 月 16 日に公開された記事に関する情報変化グラフを図 4 に示し，評価を行う．このグラフは，1 月 15 日から 16 日にかけて発生した，記事中の事象情報変化を可視化したものである．

図 4 のグラフには，124 の色付きノードが存在する．その情報変化種別ごとの内訳は，情報の削除が 43，新規出現が 64，継



図 3　田代らの手法で生成された事象情報グラフ

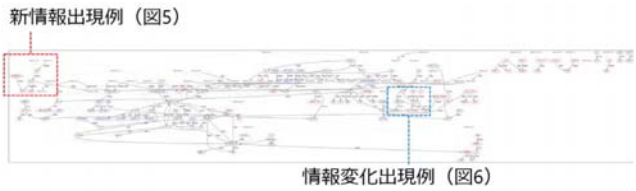


図 4　2014 年 1 月 16 日の記事に関する情報変化グラフ

*1　http://jtsb.mlit.go.jp/jtsb/ship/detail.php?id=6270

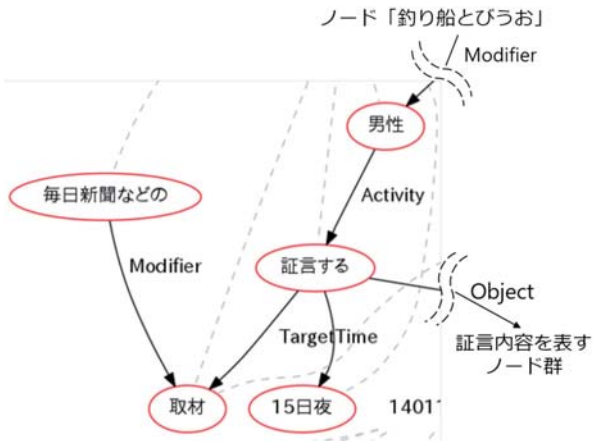


図 5 情報変化グラフ中の新規情報出現例

続出現が 17 ノードである. なお, 継続出現として可視化された 17 ノードのうち 9 つは事故の当事者名や「衝突事故」など, 事故の基本的な情報を表すもので, 図で示した 1 月 15 日から同 16 日までの期間に特有な情報は半数以下であった.

ここで, ノードの色に注目することで, より詳細な情報変化の分析を行う. 図 4 のグラフ中における情報の新規出現例を図 5 に, 特定単語に関する情報変化の例を図 6 に示す. ただし, 図 6 中のノードにおける人物名は黒塗りとして伏せてある.

図 5 では,「新聞社などの取材に対し, 男性が 15 日夜に証言した」という情報の新規出現が可視化されている. また, ノード「男性」に向かう Modifier 関係のエッジは「釣り船とびうお」から伸びており,「男性」が事故で転覆した「釣り船とびうお」の関係者であると把握できる. 同様に,「証言する」からの Object 関係のエッジを辿ることで, 証言の内容を知ることができる.

図 6 では, ノード「2 人」に関して「15 日に病院へ搬送された」から「16 日に死亡した」という時間経過による情報の推移が可視化されている. この「2 人」とは, とびうおの乗員のことを指している. しかし, グラフでは「2 人」と「釣り船とびうお」の間にエッジは存在しないため, グラフのみからそうした情報を得ることはできない. 実際の文書において, 図 6 に示した 2 つの情報はそれぞれ, 1 つの文として出現する. しかしながら,「釣り船とびうお」という単語はこれらの文中に存在しなかった. そのため, 事象情報グラフ構築時にノード間の関係設定が期待通りに行われず, 情報の不足が発生したと考えられる.

以上より, 提案手法は, 時間経過による文書中の単語の出現や削除に注目した情報変化を可視化できる. グラフにおいて, 単語を表すノードは事象内の関係を基に接続される. ユーザはエッジに付与された関係ラベルやノードの色が示す情報変化種別の組み合わせを見ることで, 情報がどのように変化したか知ることができる. 一方で, 提案手法は文中で明示的に書かれていない関係を可視化することはできない. また, 情報変化の構造や, グラフ中におけるその範囲はユーザが判断する必要がある.

4.2 重要情報グラフ例

時系列全体の文書からなるグラフに対し, 全 9 日中 8 日以上の頻度で出現した単語ノードから, 距離 2 で到達できるノードからなる重要情報グラフを抽出した. その結果を図 7 に示す. なお, 時系列的な出現頻度によるノードの着色は, 図 2 の右側に示すように, 全 9 日のそれぞれに色を割り当てることとした.

図 7 において, 閾値の 8 回を超える頻度で出現したノードは「海上自衛隊」「輸送艦おおすみ」「釣り船とびうお」「衝突事故」「瀬戸内海」の 5 つである. これらの高出現頻度ノードの描画された部分を図 8 に示す. 高頻度のノードの表す単語や, それら

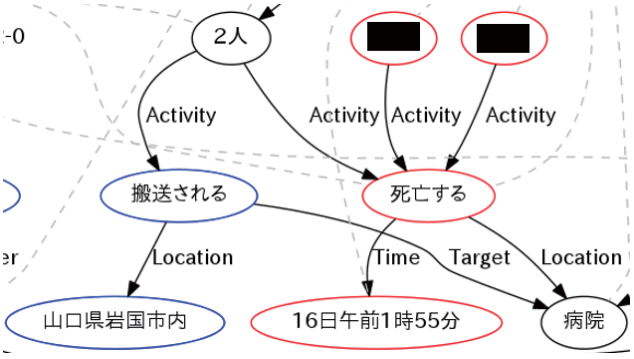


図 6 情報変化グラフ中の情報推移出現例

がどのように接続されているかを見ることで, 文書の対象とするテーマの基本的な情報を知ることができる.

まず, 高頻度のノード間の関係について分析する. 図 8 に点線四角で強調したように,「海上自衛隊」に修飾される「輸送艦おおすみ」と「釣り船とびうお」が, Activity 関係で「衝突事故」に接続されている. 加えて,「衝突事故」は「瀬戸内海」と Location 関係をもっている. 以上のことから, 分析対象の文書が扱うテーマは, 瀬戸内海で発生した海上自衛隊の輸送艦と釣り船の衝突事故であることが分かる.

次に, 高頻度ノードに接続される低頻度ノードにどのようなものがあるか分析する. 例えば,「衝突事故」は「横切る形」で修飾され, Time 関係で「1 月」や「15 日」と接続されている. また,「輸送艦おおすみ」は「右旋回している」を経由して「釣り船とびうお」や「衝突事故」と接続されている. こうした関係から, 事故が発生したのは 1 月 15 日であることや, 輸送艦は右旋回中に釣り船と衝突したという情報が出現したと分かる.

ここで, ノードの色によって表現される単語の出現時期に注目する. 図 9 では,「釣り船とびうお」に修飾された「2 人」は,「搬送される」と「死亡する」に Activity 関係のエッジをもつ. これらのノードの色によれば,「搬送される」は時系列の最初期のみに出現する. そして,「死亡する」は「搬送される」の出現直後と時系列の末期周辺に出現している. これは, 図 6 で示したような情報変化が, 図 9 での点線矢印のようなノード色の変化として現れていると解釈できる. また,「とびうおの乗員が死亡した」という情報は, 始めに出現した後はしばらく文書中で取り上げられなかったことが分かる. 実際, 分析対象の記事では, この情報が出現しなかった 1 月 17 日から 28 日までの記事では, 1 記事を除いて調査で判明した事実や関係者の証言のみを述べていた. 一方で, 例外である 1 記事には, それらに加えて「死亡した乗員 2 人のうち 1 人の通夜が営まれた」という情報が含まれていた. この記事では「2 人」という単語を用いずに以上の情報を述べていたため, ノードの出現頻度計算に誤りが生じたと考えられる.

このように, 提案手法による時系列的な出現頻度を用いたサブグラフでは, 高頻度で出現する情報と, それに付随する低頻度な情報を可視化することができる. 特に, 高頻度ノードが示す単語や, それらの関係は, 3.3 節で定義した重要情報を表すものであった. また, 低頻度ノードは, 重要情報に関連する情報が, 時系列の中でどのように出現したかを示した. ただし, 低頻度の情報にはある一時期のみ注目されたものや後に誤りだと判明したものなど, 様々な種類があることが推測される. 従って, ユーザは低頻度の情報が事実を正しく表しているとは限らないことに注意する必要がある.

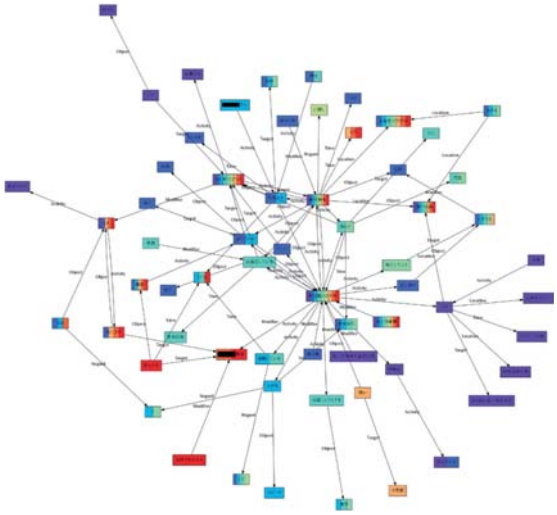


図 7 グラフ探索によって得られた重要情報グラフ

5. 課題

4 章での可視化結果より得られた，提案手法における課題について述べる．提案システムで利用する事象抽出手法に関する問題の 1 つとして，グラフ中におけるノードの多さが挙げられる．その原因としては，「こと」や「する」等，それ自体では多くの情報を持たないノードの存在がある．こうしたノードは他のノードと接続されやすく，グラフ操作による削除が難しい一方，表示が煩雑になる．もう 1 つの問題として，文中における単語の省略に起因する，エッジの設定ミスがある．先行研究と異なり，本論文ではグラフの分割を行うために 1 グラフ当たりの文書数が少なくなりやすい．そのため，グラフ中の情報欠損が比較的発生しやすく，分析の支障となる可能性がある．

これらへの対策としては，田代らの事象抽出手法を改良することによる，ノードとなる単語のフィルタリングや，照応解析による情報の補完などが考えられる．

(1) 情報変化グラフにおける課題

情報変化の可視化では，連続する 2 期間における短期的な変化を計算している．そのため，初期に錯綜していた情報が次第に訂正され，統一されていくといった，長期的視点での情報変化を可視化することができない．また，情報変化を検知する以上の処理を行わないため，情報の矛盾や人物状態の推移等，情報変化の特徴を判別することができない．

これらの改善策として，情報変化の対象や変化パターン等，長期的変化の可視化に有用な情報の利用を検討中である．

(2) 重要情報グラフにおける課題

提案した重要情報可視化手法は，ノードがどのような単語を表すかを考慮せず，出現頻度とノード間の距離によってサブグラフを抽出する．そのため，特にサブグラフの末端ノードにおいて，それがどのような事象に出現したのか解釈するのが難しい．例えば，図 7 のグラフでは「衝突事故」が「に関し」と接続されているが，「に関し」の対象となるノードは抽出できていない．

この問題の解決策として，ノード間の関係やエッジの向きを考慮することで，ノードの削除や追加取得を行うことが考えられる．

6. おわりに

本論文では，時系列的に複数作製される文書群における情報の変化と，文書の対象とするテーマに関連する重要情報の

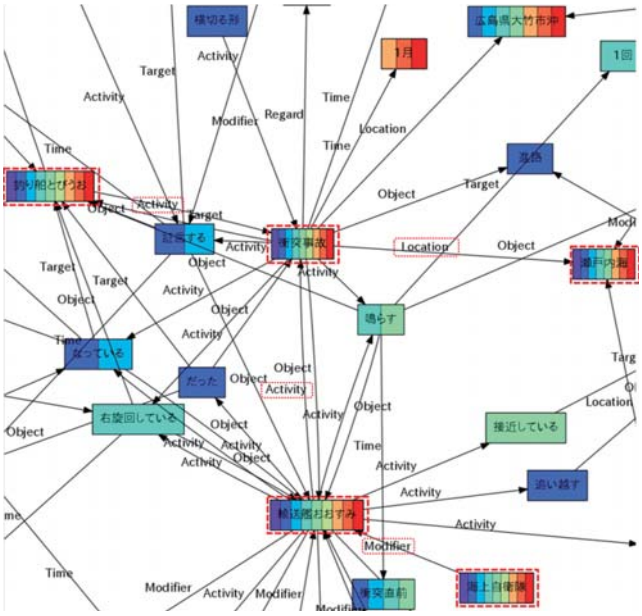


図 8 高出現頻度ノードに関する部分

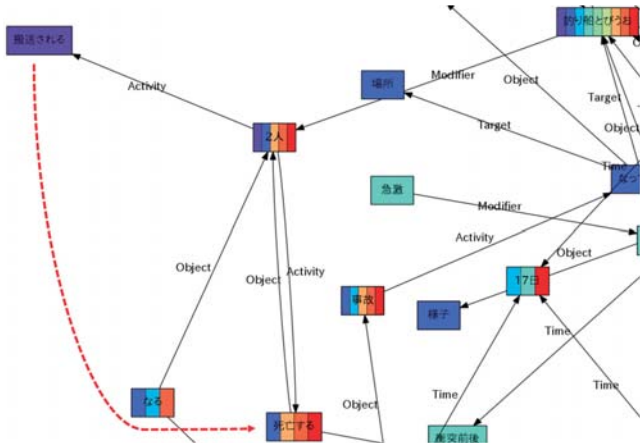


図 9 時間経過に関する重要情報例

可視化手法を提案した．その後，実際の新聞記事に対して提案手法を適用し，可視化結果を評価した．

実際の文書群に適用した結果，提案手法は短期間での情報の新規出現や時間経過による変化を可視化できることを確認した．また，重要情報として文書のテーマを表すような情報や，それに関する様々な情報を可視化できることを示した．

今後の予定としては，人物等に注目することによる長期的な情報変化の可視化や，ノード間の関係などを考慮することによる重要情報のより高精度な抽出などを考えている．

参考文献

[田代 2013] 田沼和浩，王冕，越川兼地，森田武史，長野伸一，清雄一，中川博之，田原康之，川村隆浩，大須賀昭彦: “Linked Data を用いたソーシャルメディア×マスメディアの比較実験”，人工知能学会全国大会論文集，2013.

[伊藤 2012] 伊藤正彦，吉永直樹，豊田正史，喜連川優: “係り受け解析を用いたブログユーザの行動・興味に関する時系列推移 3 次元可視化システム”，電気情報通信学会論文誌 D，Vol. J95-D，No.7，pp.1454-1466，2012.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

軌跡データのみを用いた観光スポット遷移モデルの構築

Tourist Transition Model among Sightseeing Spot based on Trajectory Data

渡部岳志 *1 Takeshi Watabe　笠原秀一 *2 Hidekazu Kasahara　飯山将晃 *2 Masaaki Iiyama　美濃導彦 *2 Michihiko Minoh

*1京都大学工学部情報学科
Undergraduate School of Informatics and Mathematical Science, Kyoto University

*2京都大学 学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University

In recent years, congestion of famous sightseeing spots and public transportation has become a problem in tourism areas. It is a way to solve it that local governments forecast congestion by a congestion simulation based on a tourist transition model and adjust the operation of public transportation. In this research, we propose a method to construct a tourist transition model among sightseeing spots based on trajectory data. We construct a tourist transition model using actual trajectory data of tourists who stayed around Kyoto city.

1. はじめに

近年観光都市では観光スポットや公共交通機関の混雑が問題となっている．それを解消する方法として行動モデルに基づいた混雑予測を元に交通機関の運行を変更するといった方法が考えられる．既存手法ではモデル構築に必要な情報を得るために Wikipedia など外部の情報を用いている．しかし，これらの手法では既知の観光スポットしかモデルに反映できない．そのため，実データに基づいて観光客が実際に訪問した観光スポット間の遷移を表す行動モデルが求められている．

本研究では観光客の軌跡データのみから構築できる観光スポット遷移モデルの構築手法を提案する．具体的には観光客の移動速度に着目して観光客が集まる点 (観光客集中点) を抽出する．その後，観光客集中点を観光スポットと移動中に立ち寄る場所 (経由地) に分類し，観光スポット間の遷移確率を求め，観光スポット遷移モデルを構築する．

2. 軌跡からの観光スポットの抽出

軌跡データから観光スポット遷移モデルを構築するには，測定誤差を除去した軌跡データから観光スポットを抽出する必要がある．観光スポットでは観光客は足を止めて観光を行うので，速度に着目すれば観光スポットを抽出できる．しかし，観光客はバス停や駅など経由地でも足を止めることがあるので，両者を区別しなければならない．観光スポットと経由地では観光客の移動方法の割合が異なり，前者では徒歩が多く，後者は非徒歩が多いとと考えられるので，この特徴を用いて両者を分類する．以下，処理の詳細について述べる．

2.1 軌跡データの前処理

本研究で用いる軌跡データは，観光客が携帯する GPS 機器により計測した観光客の位置の時系列データである．軌跡データ内の各点は，軌跡 ID(id)，緯度 (lat)，経度 (lon)，時刻 (t) を持つ．GPS の軌跡データは観光客が携帯する GPS 機器から収集できる．そのため，容易に広い範囲のデータを収集することができる．

軌跡データは GPS の測定誤差を含んでいるので除去する．軌跡データが測定誤差を含んでいると訪問していない地点を観光スポットまたは経由地として推定してしまうからである．大きな誤差を含んだ軌跡データは移動速度が急に変化すると考えられる．観光客の移動速度は移動方法によって変化するが誤差による移動速度の変化はそれよりも大きい．このため，軌跡データの各点に前後の緯度 lat と経度 lon から求めた速度 v を付与し閾値 v_e より v が大きい点を異常値として除去する．

2.2 観光客集中点の抽出

本節では，測定誤差を除去した軌跡データから，観光客が一定時間以上速度を低める点を観光集中点として抽出する．観光客集中点には観光スポットと経由地が含まれる．

観光客が一定時間以上速度を低めるパターンには，観光・乗り換えによる停止とそれらと関連のない一時的な立ち止まりの２種類がある．前者を停止点群として抽出する．一時的な立ち止まりは観光や乗り換えに比して短時間である．本研究では一時的な立ち止まりの最長時間を t_s，観光・乗り換えや一時的な立ち止まりの中の移動速度の最大値を v_s とする．移動速度 v に対し，t_s 以上の間連続して v が v_s を下回ったとき，その間の点を停止点群として抽出する．次に停止点群を mean-shift 法によってクラスタリングすることで観光客集中点を得る．こうして得られた各クラスタの重心を観光客集中点 c とする．

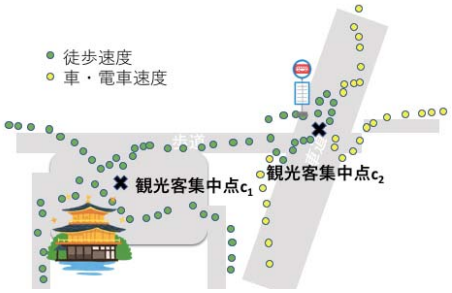
2.3 観光客集中点の観光スポットと経由地への分類

本節では，前節で抽出した観光客集中点を，観光スポットと経由地に分類する．観光客は観光スポットでは観光を行うため主に徒歩で移動し，経由地では乗り換えを行うため主に非徒歩で移動する傾向を持つ．この傾向を用いて観光客集中点を分類する．

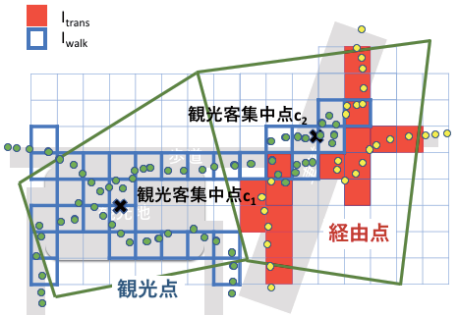
まず全ての観光客集中点にボロノイ分割を用いてモデルの対象地域を分割した空間を集中点空間 S とする．集中点空間 S において徒歩で進入可能な領域を I_{walk}，非徒歩で進入可能な領域を I_{trans},I_{walk} と I_{trans} の和集合を I_{all} とする (図 1)．観光スポットと経由地の移動方法の違いから集中点空間 S 内で領域 I_{all} のうち，領域 I_{trans} が占める面積の割合は観光ス

連絡先: 連絡先〒 606-8501 京都市左京区吉田二本松町 京都大学学術情報メディアセンター南館 美濃研究室，渡部岳志，E-mail:watabe@mm.media.kyoto-u.ac.jp

ポットでは小さく，経由地では大きいと仮定する．この仮定を用いて観光客集中点を観光スポットと経由地に分類する．



(a) 地図と観測点の図示



(b)I_{walk} および I_{trans}

図 1: I_{trans} と I_{walk} の抽出例

I_{walk} と I_{trans} の面積を得るために I_{walk} と I_{trans} を画像として得る．まず，軌跡データから速度 v_w 以下の点を徒歩，それ以上の点を非徒歩で移動しているとする．集中点空間 S 内に存在する軌跡データの内，徒歩で移動している点からヒートマップを作成し，グレースケール画像を二値化することで二値画像 I^c_{walk} を作成する．これらの二値画像において値を持つ画素に対応する領域は観光客が徒歩で進入可能であるとし，値を持つ画素の数 $M(I^c_{walk})$ をその面積とする．同様に軌跡データから非徒歩で移動している点から抽出し同様の二値画像 I^c_{trans} を作成し，$M(I^c_{trans})$ をその面積とする．I^c_{trans} と I^c_{walk} の和集合を I^c_{all} とし，$M(I^c_{all})$ をその面積とする．$M(I^c_{all})$ と $M(I^c_{trans})$ の面積比

$$R_c = \frac{M(I^c_{trans})}{M(I^c_{all})}$$

を画像の特徴量とする．特徴量 R_c が閾値 R_α より小さいものを観光スポット，大きいものを経由地として分類する．

3. 観光スポット遷移モデルの構築

3.1 観光スポットネットワークの構築

本節では，集中点ネットワークを定義し，このネットワークから得られた集中点間遷移確率を用いて観光スポットネットワークを構築する．

まず，集中点ネットワークとして，観光客集中点をノードとして持ち，観光客集中点間での遷移が軌跡データ内に存在するならばエッジを持つような有向重み付きネットワークを考える．このとき，エッジの重み $p_{i,j}$ は観光客集中点 i から行われた全遷移のうち，観光客集中点 j に遷移した割合とする．これをあるノード i から別のノード j へ直接遷移する確率とする．

次に観光スポットネットワークを構築する．遷移行列 $M = (p_{i,j})$，ノード i からノード j へのネットワーク上での最短距離を $d_{i,j}$ とする．このとき，観光客は最短経路のみを用いる訳ではない．そのためノード i とノード j の間を遷移する上で経由する可能性のあるノードの数の上限を $l_{i,j}$ とおくと求めたい $p_{i\to j}$ は以下のように求まる．

$$p_{i\to j} = \alpha(\sum_{r=d_{i,j}}^{l_{ij}} M^r)_{ij} \quad (1)$$

本研究では l_{ij} は $d_{i,j}$ の関数であると考え，$l_{i,j} = 2 \times d_{i,j}$ とした．α は正規化項であり．

$$\alpha = \frac{1}{\sum_{c\in K} p_{i\to c}} \quad (2)$$

とする．K は観光スポットからなる集合である．観光スポットをノードとして持ち $p_{i\to j}$ をエッジの重みとしてもつネットワークを観光スポットネットワークと呼ぶ．

3.2 観光スポット遷移モデルの構築

本研究では観光客の観光スポット遷移を単純マルコフモデルでモデル化する．つまり，観光客が次に遷移する観光スポットは現在訪問している観光スポットのみに依存する．観光スポットネットワーク上の対応するノード間のエッジの重みを，観光スポット間遷移確率とする．

4. 実験

4.1 軌跡データの前処理

本研究で使用したデータは，班別自由行動中の修学旅行生 (班長が端末を携帯) 及び旅行代理店スタッフの計 579 名の GPS 移動軌跡である [笠原 15]．実験データは，携帯端末を対象者に携帯したまま一日移動して貰い，その行動を携帯端末にインストールした GPS ロガーアプリケーションで収集した．GPS はおおよそ 1 秒毎に位置を計測している．対象期間は 2015 年 12 月である．実験地域としては修学旅行生が多数訪れている京都市周辺とするために緯度 34.80 度以上，35.15 度以下，経度 135.65 度以上，135.85 度以下の地域とした．また，閾値 v_e を 180(km/h) として異常値を除去した．また，GPS 以外の A-GPS や wifi で測位された点も測定精度が低いため全て除去した．実験データは 9,530,489 の観測点を含み，前処理によって実験地域外の観測点と異常値を削除した後で 5,108,676 の観測点を含んでいた．

4.2 観光客集中点の分類

本研究では実験に用いた軌跡データの特性をもとに，停止点群を抽出する際の定数として $v_s = 3.6$(km/h)， $t_s = 200$(秒) を用いた．また，mean-shift 法のカーネル関数としてガウシアンカーネルを用い，バンド幅として 0.0010 を用いた．

提案手法によって観光客集中点が 354 箇所抽出された．抽出した観光客集中点に人の手によるタグ付けを行った結果は観光スポットが 170 箇所，経由地が 171 箇所であった．またその他に渋滞や交差点など観光客集中点の誤抽出が 13 箇所含まれていた．これに対し提案手法による観光客集中点の特徴量 R_c の分布が図 2，ROC 曲線が図 3 である．図 2 の特徴量 R_c の分布から観光スポットにおいて特徴量 R_c はほとんどの場合 0.4187 より小さく，経由地において R_c は同じく 0.4187 より大きいことがわかる．そのためモデルの構築には 0.4187 を閾

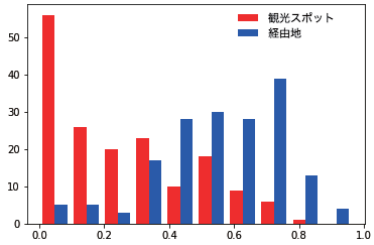


図 2: 提案手法における特徴量 R_c の分布

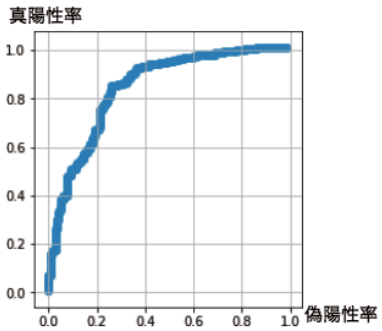


図 3: 提案手法の ROC 曲線

値として用いた．人の手によるタグ付けを正解としたとき，分類の正答率は 76.6%だった．

次に提案手法を用いて正しく分類できた例について観光スポットと経由地をそれぞれ図 4，5 に示す．

図 4 は銀閣寺に対応する観光客集中点において作成した二値画像である．この観光客集中点の面積比 R_c は 0.040 と非常に小さい．下図より銀閣寺内の道が I_{all} として抽出されていることがわかる．また，上図より非徒歩で進入可能な領域 I_{trans} として抽出されている領域がほとんどないことがわかる．この領域内では車で移動できないが上図には非徒歩で進入した領域が存在している．これは GPS の測定誤差などによる影響で速度が正しく認識できていないためと考えられる．

図 5 は，山科駅上の観光客集中点において作成した二値画像である．この観光客集中点の特徴量 R_c は 0.855 と非常に大きい．下図より非徒歩で進入可能な領域として線路が抽出されていることがわかる．また，上図より進入可能な領域は非徒歩で移動できる領域とほとんど重なっていることがわかる．

次に分類が失敗した例を図 6 に示す．

図 6 は東福寺に対応する観光客集中点において作成した二値画像である．この観光客集中点は東福寺の境内に存在している．地図から分かる通り集中点空間内には東福寺と直接の関連を持たない道路や線路が西側に存在している．このため，この観光客集中点は特徴量 R が 0.626 と大きく経由地に分類された．このように近隣に他の観光客集中点がなく，集中点空間が広く定まってしまう場合には空間内に本来観光スポットではない道路や線路が含まれてしまい，誤分類が発生することがある．このような誤分類を防ぐための方法としてボロノイ分割で空間を分割するのではなく停止点群の分布から集中点空間を得る手法を用いることなどが考えられる．



非徒歩で進入可能な領域 I_{trans}



進入可能な領域 I_{all}

図 4: 観光スポットへの分類成功例, 銀閣寺, $R_c = 0.040$
図左中央の太い道は参道であり徒歩以外では移動できない



非徒歩で進入可能な領域 I_{trans}



進入可能な領域 I_{all}

図 5: 経由地への分類成功例, 山科駅, $R_c = 0.855$

4.3 観光スポットネットワークの構築

実験データセットに対し提案手法を用い観光スポットネットワークを構築した．例として金閣寺の観光スポットからの遷移確率が高い観光スポット 10 箇所に関し，地図とともに図 7 に示す．

また，モデル全体の評価を行うために観光スポットをいくつかの地域に分け，それぞれの地域の間の遷移確率を求めた．その結果を表 1 に示す．表 1 が示す通り，ほとんどの地域において，同一地域へ遷移する確率が最も高く，それに次いで清水寺または京都駅に遷移する確率が高い．また，地理的に隣接している地域間では，遷移確率が高い傾向が見られる．この結果は，観光客は京都駅や清水寺のような人気のあるスポットには，距離にかかわらず遷移することがあるが，それ以外の場合は同一あるいは近隣の観光スポットに遷移することを示していると考えられる．

表 1: 地域間遷移確率

		目的地									
	地域	京都駅	清水	河原	嵐山	岡崎	金閣	伏見	東山	銀閣	その他
	京都駅	0.492	0.128	0.126	0.019	0.005	0.055	0.059	0.020	0.017	0.080
	清水	0.101	0.591	0.078	0.013	0.016	0.029	0.017	0.006	0.064	0.080
	河原町	0.072	0.223	0.453	0.036	0.011	0.027	0.022	0.010	0.031	0.111
出	嵐山	0.122	0.202	0.117	0.232	0.003	0.108	0.029	0.005	0.014	0.163
発	岡崎	0.101	0.237	0.068	0.013	0.224	0.029	0.012	0.004	0.167	0.140
地	金閣	0.108	0.098	0.090	0.057	0.015	0.347	0.013	0.004	0.098	0.164
	伏見	0.229	0.132	0.061	0.015	0.008	0.050	0.360	0.023	0.019	0.099
	東山	0.320	0.144	0.194	0.054	0.004	0.032	0.049	0.073	0.017	0.108
	銀閣寺	0.018	0.268	0.118	0.006	0.093	0.044	0.013	0.001	0.281	0.151
	その他	0.131	0.215	0.110	0.071	0.019	0.090	0.016	0.017	0.069	0.264

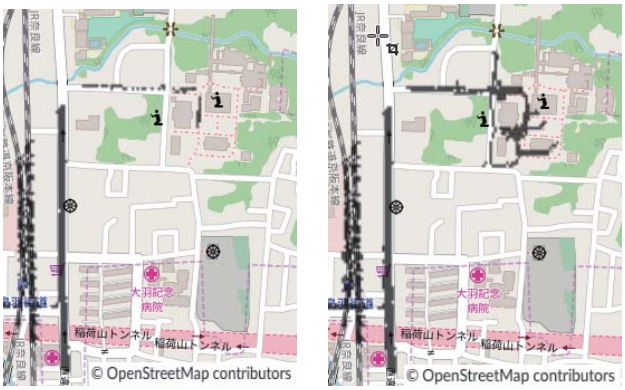


図 6: 観光スポットへの分類失敗例，東福寺， $R_c = 0.626$

5. おわりに

本研究では，軌跡データのみから観光客集中点を抽出し，観光客集中点を観光スポットと経由スポットへ分類した．そして，得られた観光スポットと経由地を用いて観光客の観光スポット間遷移をモデル化する手法を提案した．この手法では経由地を考慮することで既存手法より観光客の移動に即したネットワークを構築できる．

今後の課題として，mean-shift 法のバンド幅や集中点ネットワーク上で観光客が取りうる経路，特徴量の閾値を自動的に決定する手法の考案などが挙げられる．

謝辞

本研究は JSPS 科研費 17K00438 の助成を受けたものです．

参考文献

[笠原 15] 笠原秀一, 森幹彦, 椋木雅之, 美濃導彦.: 環境制約を用いた GPS 移動軌跡からの旅行者状態推定, 人工知能学会全国大会論文集 (CD-ROM), Vol.29, ROMBUNNO.2B3-NFC-02A-4, 2015.

	遷移先	遷移確率
A	金閣寺舎利殿	0.136
B	銀閣寺	0.072
C	龍安寺前土産物屋	0.064
D	龍安寺	0.045
E	北野天満宮	0.045
F	銀閣寺前土産物屋街	0.045
G	白峯神社	0.037
H	京都駅	0.031
I	清水寺前土産物屋街	0.030
J	京都駅バスターミナル	0.023



図 7: 金閣寺からの遷移

[2O4] [General Session] 3. Data Mining

座長:大知 正直（東京大学）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room O (2F Kaimon)

[2O4-01] Sightseeing Tweet Classification using Convolutional Neural Networks

○Shuichi Hashida[1], Keiichi Tamura[1], Tatsuhiro Sakai[1] （1. Hiroshima city university）

5:20 PM - 5:40 PM

[2O4-02] User constrained independent topic analysis

○KENTA YAMAMOTO[1], TAKAHIRO NISHIGAKI[1], TAKASHI ONODA[1] （1. AOYAMA GAKUIN UNIVERSITY）

5:40 PM - 6:00 PM

[2O4-03] Discussion on topic change detection method in Incremental Independent Topic Analysis

○Takahiro Nishigaki[1], Takashi Onoda[1] （1. Aoyama Gakuin University）

6:00 PM - 6:20 PM

[2O4-04] Extraction of Causal Knowledge from Annual Securities Report

○Fumihito Sato[1], Hiroaki Sakuma[1], Shunya Kodera[1], Yoshinori Tanaka[1], Hiroki Sakaji[2], Kiyoshi Izumi[2] （1. Nikko Research Center, Inc., 2. Graduate School of Engineering, The University of Tokyo）

6:20 PM - 6:40 PM

[2O4-05] Nursing Literature Review Matrix System based on Machine Learning

○Takanori Yamashita[1], Rieko Izukura[1], Naoki Nakashima[1], Sachio Hirokawa[1] （1. Kyushu University）

6:40 PM - 7:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

畳み込みニューラルネットワークを用いた観光ツイートの分類手法

Sightseeing Tweet Classificaiton using Convolutional Neural Networks

橋田修一 *1　田村慶一 *2　酒井達弘 *2
Shiuichi Hashida　Keiichi Tamura　Tatsuhiro Sakai

*1広島市立大学 情報科学部
Faculty of Information Sciences, Hiroshima City University

*2広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

Recently, the icreasing interest in the social media networking service (SNS) is remarkable. Many researchers work for the discovery of new information from the posted tweets. In this study, we focus on the tweets related to tourist spots for finding the spot's improvement and appeal point. However, tweets on Twitter site include not only descriptions that are related to sightseeing but also descriptions that are not related to sightseeing. As the solution, we use convolutional neural networks for classifying these tweets. Moreover, we propose the multi-channel method to improve the performance of classification. In this method, the model was passed two test information that is words and form classes. In experiments, several models are used for comparing each other, Naive Bayes, convolutional neural networks (CNN), long short-term memory, and the proposed model. The experimental results show that the proposed model can classify tweets better than the other models.

1. はじめに

近年，国内の観光産業における経済効果は目覚ましく，観光振興に大きな注目が集められている．そこで，観光地の発展を目指すため，観光地の満足点や不満点を調査し，その後の改善策や PR 計画に役立てる事業が盛んに行われている．しかし，実際に行われている観光客への調査はアンケート形式が多く *1，その結果が出るまでに時間がかかることや分析に多くの人手が必要となるようなコスト面の問題が挙げられる．そこで，ソーシャルメディア，特に Twitter 上に投稿されるツイートをリアルタイムに分析することで，観光地の満足点や不満点を明らかにすることが期待されている．

SNS ユーザの増加に伴って，投稿されたツイートから新たな情報の発見や分析を行う研究 [Jansen 09, Sakai 15] が盛んに行われており，その一つとして，観光情報に着目した研究がある [松本 14, 佐藤 15]．観光客が投稿したツイートから観光情報を抽出することで，リアルタイム性のある調査結果が得られることが期待できるが，実際には，Twitter 上に投稿されるツイートは様々なバリエーションのあるツイートが存在する．そのため，観光地の分析に使用できるツイートと使用できないツイートを分類する必要がある．

そこで，本研究では，観光地に関して投稿されたツイート（以下，観光ツイートと呼ぶ）に対する，畳み込みニューラルネットワークを用いた分類手法を提案する．具体的には，観光地に対する意見を含むツイートと含まないツイートの分類，観光地に対する満足を示すツイートと，不満を示すツイートの分類を行う．また，分類精度の向上のために，単語列だけでなく品詞列も組み合わせてモデルに入力する新しいモデルを提案する．

連絡先: 橋田修一，広島市立大学情報科学部，広島市安佐南区大塚東三丁目 4 番 1 号，082-830-1676，email : shu.hashida@gmail.com

*1　http://www.mlit.go.jp/kankocho/siryou/toukei/shouhidoukou.html

2. 観光ツイートの分類

観光ツイートとして投稿されたツイートは観光運営者に限らず，観光客が投稿したツイートも含む．そのため，さまざまなバリエーションを含んでいるツイートが多く存在する．ここでは，観光ツイートを分類するために，それぞれのツイートに対して付与される分類ラベルの定義を行う．

まず，投稿されたツイートが観光地に関する意見を含んでいるツイートであれば *Opinion_is_Included*(以下,*OI*) クラスとし，観光地に関する意見を含んでいないツイートであれば *Opinion_is_Not_Included*(以下,*ONI*) クラスと分類する．本研究における意見には観光地への評価や苦情が含まれるものとする．具体的には，*OI* クラスには観光地に対する満足を示す *Positive* クラスと不満を示す *Negative* クラスが含まれる．

次にツイートを分類する処理の流れを示す (図 1)．

1. Twitter 上に投稿されたツイートを，キーワード検索を用いて取得する．この時，キーワードには観光地名を示す単語が設定される．例として，広島県内の観光地であ

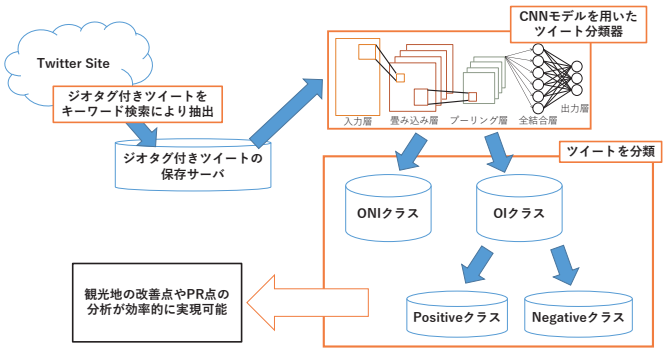


図 1: 観光ツイートの分類の流れ

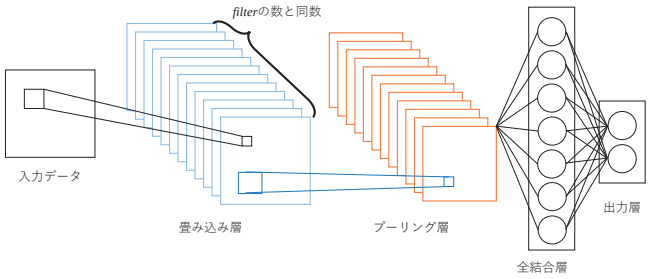


図 2: 畳み込みニューラルネットワークの概要

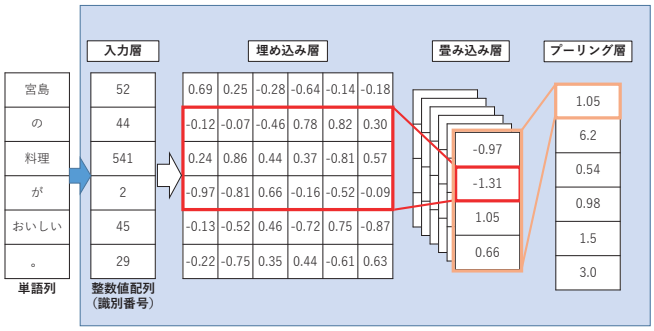


図 3: テキスト分類のための畳み込みニューラルネットワーク

る宮島に注目するときは，キーワードを “宮島” と設定する．

2. 取得したツイートを分類器で分類ラベルの予測を行う．取得したツイートは最初に $OI-ONI$ 分類器によって，OI クラスと ONI クラスに分類される．

3. 次に OI クラスに分類されたツイートに対して，$Positive-Negative$ 分類器が $Positive$ クラスと $Negative$ クラスの分類を行う．

3. 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク [Lecun 98] とは，主に画像認識で用いられている手法であり，畳み込み層とプーリング層から構成される．図 2 に畳み込みニューラルネットワークの概要を示す．まず最初に，畳み込み層では，入力された 2 次元，もしくは 3 次元行列に対してフィルタを適用し，スライドしながら対象領域の畳み込み処理を行う．出力では，入力データの特徴マップが生成され，畳み込み層内にあるフィルタの数だけ特徴マップが生成される．次に，プーリング層では畳み込まれた特徴マップのそれぞれに対して，注目領域の代表値を算出し，情報圧縮が行われる．

4. 畳み込みニューラルネットワークを用いたテキスト分類手法

本章では，畳み込みニューラルネットワークと分散表現を用いたテキスト分類の研究について Kim の行った研究 [Kim 14] を紹介する．

4.1 分散表現

畳み込みニューラルネットワークを用いたテキスト分類の学習を行うため，テキストを数値空間上に表現する必要がある．そこで一般的に分散表現 [Mikolov 13b] を利用し，テキスト中の単語を数値化し，モデルに入力として与える．分散表現とは，単語を示す記号に対してベクトルを付与する手法である．分散表現を用いることで，単語間の意味関係が数値空間上で表現され，テキストを分散表現の系列として表現できる．分散表現の獲得方法について，word2vec[Mikolov 13a] を用いて大量のテキストデータから事前学習を用いて獲得する手法や，モデル内の埋め込み層を用いた獲得方法 [Collobert 11] などが挙げられる．

4.2 モデル

Kim が提案したモデル (以下，Yoon-Kim’s モデルと呼ぶ) では，埋め込み層の事前学習を行う手法と畳み込み層で複数のフィルタサイズを設定する手法で実験が行われている．分散表現獲得方法については，事前学習ありの埋め込み層と事前学習なしの埋め込み層の両方が示されている．事前学習ありの埋め込み層を用いた手法では，Google News から得られた 1 千億単語を用いて事前学習で収得した分散表現が使用される．また，事前学習なしの埋め込み層を用いた手法では，単語ベクトルの初期値をランダムに決定し，テキスト分類の学習を通して単語ベクトルの重みを更新する．

次に，モデルの構成について，主な層の説明とその役割について示す．図 3 に Yoon-Kim’s モデルの概要を示す．埋め込み層では，入力層から得られた単語識別番号の配列 (配列長:N) を受け取り，対応する単語ベクトル (次元数:M) を埋め込まれた，$N \times M$ の 2 次元行列を出力する．次に，畳み込み層では，単語ベクトルの次元数と揃えられたフィルタが適用され，テキストの特徴を抽出している．また，プーリング層では，畳み込まれた特徴マップ全体に対してマックスプーリングが適用され，最大値が抽出される．その後，全結合層，出力層を構成することで分類ラベルの予測を行う．

実験結果として，分散表現を事前学習で獲得することで分類モデルの精度向上が行えることが確認された．しかし，事前学習により得られた分散表現は，文法的に類似する単語のベクトルが類似することが確認された．もう一方の，事前学習が行われていない埋め込み層を用いる手法では，意味的な単語関係が表現されるため，意味的に似ている単語のベクトルが類似することが確認された．

5. 提案手法

本章では，Yoon-Kim’s モデルをベースとした提案手法を示す．本研究のモデルは Yoon-Kim’s モデルでも，事前学習なしの埋め込み層を用いるモデルを用いる．従来手法では畳み込みニューラルネットワークに単語列のみを入力として与えていたが，提案手法では単語列に対応する品詞列を与える．品詞列を与えることで，学習の促進が期待でき，テキスト中の各単語が持つ情報を増やすことで分類精度の向上が期待できる．主な手順を次に示す (図 4)．

1. 訓練データ内のツイートの分かち書きを行い，そこから単語識別番号を記録する辞書を作成する．

2. 次に，分かち書きされたツイートに対して，それぞれの

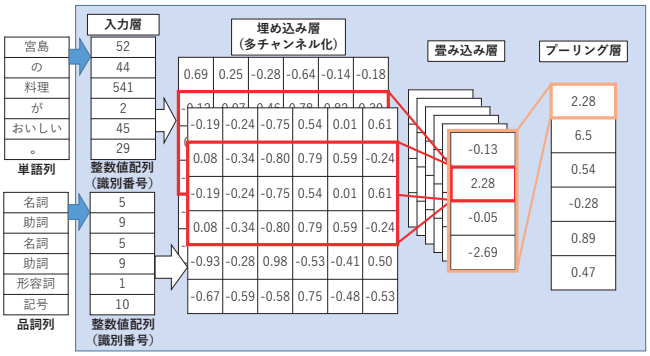


図 4: 提案モデルの概要

単語に単語識別番号を割り振る．テストデータ内において，出現した単語の単語識別番号を記録した辞書に存在しなければ，単語には空白を示す識別番号 “0” を格納し，品詞列にはその単語の品詞が格納される．これにより，単語列と品詞列の位置関係の差がなくなり，テキストの特徴を抽出することができる．

3. 整数値番号に変換された単語列と品詞列はモデルの入力層に与えられる．モデル内には 2 つの入力層と埋め込み層が存在し，それぞれの埋め込み層で単語の分散表現，品詞の分散表現を学習し，埋め込み処理を行う．その後，畳み込み層に与えるために，それぞれのチャンネルを積み重ねることで「多チャンネル化」を行う．

6. 実験

提案手法を評価するために評価実験を行った．本章では実験結果を示す．

6.1 データセット

本実験で用いるデータセットについて説明する．本研究では，実際に投稿された観光ツイートを，Twitter API を用いてキーワード検索 “宮島” により取得し，手作業でラベル付けを行った．ラベルは，OI クラスに含まれる $Positive$ クラスと $Negative$ クラス，そして ONI クラスとした．これらの 3 つのラベルはそれぞれのツイートに付与され，1 つのツイートに 1 つのラベルを付与する．表 1 にそれぞれのクラスに属するツイート数の内訳を示す．

表 1: データセット内の内訳

$Positive$	$Negative$	ONI	計
1914	826	3253	5993

本研究では比較モデルとして，次に示すモデルを用いて，10 回交差検定を行い，テストデータに対する分類精度の平均値をそのモデルの精度として評価を行う．また，Yoon-Kim’s モデルでは 3 手法を用いて，実験を行った．

- Naive Bayes
  - Naive Bayes を用いて分類を行う．
- Long short-term memory(LSTM)
  - 深層学習の手法の一つとして知られ，自然言語処理分野で多くの研究に使われているモデルである．時系列データに対して有効な手段として，長期依存の学習が行える．
- Yoon-Kim’s モデル
  - Kim の提案したモデルである．本研究では，次の 3 つの手法を用いて分類の学習を行った．
    1. 訓練データに対して word2vec を使用し，事前学習を通して得られた重みを用いて単語の埋め込みを行うモデル
    2. 訓練データとは別のデータセットに対して事前学習を行い，その単語の重みを用いて単語の埋め込みを行うモデル
    3. モデルに対する分類の学習を通して，単語の重みを学習する埋め込み層を用いて単語の埋め込みを行うモデル
- 提案手法
  - 本研究の提案モデルであり，Yoon-Kim’s モデルをベースとして入力に品詞列を追加したモデルである．

また，モデル内に Dropout を設定しているため，10 回交差検定を 20 回行い，分類精度の統計値を算出した．それぞれのモデルの精度を $OI-ONI$ 分類は表 2 に，$Negative-Positive$ 分類は表 3 に示す．

実験結果から，$OI-ONI$ 分類において，比較モデルよりも提案モデルが高い分類精度を示すことが確認できた．しかし，$Negative-Positive$ の分類では提案モデル以外のモデルで高い精度が確認された．これはデータセット内の $Negative$ クラスと $Positive$ クラスのツイート数に大きな偏りがあることが考えられる．

7. まとめ

本研究では，観光地の改善点や PR 点を分析するため Twitter 上に投稿される観光ツイートに着目し，畳み込みニューラルネットワークを用いて観光ツイートを分類する手法を提案した．また，多チャンネル化を行うことで分類精度の向上を図った．実験では，実際に投稿されたツイートに対してモデルの学習，分類を行い比較実験を行った．実験結果として，提案モデルの $OI-ONI$ 分類における分類精度が最も良いことが分かった．しかし，$Negative-Positive$ クラスの分類精度では他のモデルに劣る部分があることが確認できた．今後の課題として，分類されたツイートに対して観光地の改善や PR に関する情報の抽出が考えられる．具体的には，観光地の「どの施設が」「悪いか，もしくは良いか」，また,「それはなぜなのか」などの要素の抽出が考えられる．

謝辞

本研究の一部は，総務省 SCOPE(受付番号 : 162308002) と広島市立大学・特定研究費の支援により行われた．

参考文献

[Collobert 11] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. P.: Natural Language Processing (Almost) from Scratch., *Journal*

表 2: モデルの評価値 ($OI - ONI$ 分類)

モデル	OI クラス			ONI クラス			正解率
	再現率	適合率	F 値	再現率	適合率	F 値	
ナイーブベイズ	0.832	0.649	0.728	0.752	0.890	0.814	0.776
LSTM	**0.815**	0.777	0.795	0.821	0.851	0.835	0.817
Yoon-Kim's CNN model (1. 事前学習済み分散表現)	0.718	0.645	0.677	0.727	0.784	0.753	0.721
Yoon-Kim's CNN model (2. 事前学習済み分散表現)	0.591	0.705	0.642	0.705	0.587	0.639	0.641
Yoon-Kim's CNN model (3. 埋め込み層)	0.830	0.781	0.804	0.825	0.864	0.844	0.826
Yoon-Kim's CNN model (品詞列のみ)	0.784	0.743	0.761	0.794	0.826	0.808	0.788
提案手法 (Yoon-Kim's CNN model)	**0.832**	**0.789**	**0.809**	**0.831**	**0.865**	**0.847**	**0.830**

表 3: モデルの評価値 ($Negative - Positive$ 分類)

モデル	$Negative$ クラス			$Positive$ クラス			正解率
	再現率	適合率	F 値	再現率	適合率	F 値	
ナイーブベイズ	0.605	**0.831**	0.700	**0.914**	0.763	0.831	0.784
LSTM	0.825	0.752	0.784	0.898	0.930	0.913	0.876
Yoon-Kim's CNN model (1. 事前学習済み分散表現)	0.708	0.448	0.541	0.795	0.917	0.851	0.775
Yoon-Kim's CNN model (2. 事前学習済み分散表現)	0.618	0.205	0.299	0.733	**0.943**	0.825	0.721
Yoon-Kim's CNN model (3. 埋め込み層)	0.835	0.746	0.789	0.897	0.936	0.915	0.878
Yoon-Kim's CNN model (品詞列のみ)	0.754	0.471	0.569	0.804	0.929	0.861	0.791
提案手法 (Yoon-Kim's CNN model)	**0.840**	0.764	**0.797**	0.903	0.936	**0.918**	**0.884**

of Machine Learning Research, Vol. 12, pp. 2493–2537 (2011)

[Jansen 09] Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A.: Twitter Power: Tweets As Electronic Word of Mouth, *J. Am. Soc. Inf. Sci. Technol.*, Vol. 60, No. 11, pp. 2169–2188 (2009)

[Kim 14] Kim, Y.: Convolutional Neural Networks for Sentence Classification, in Moschitti, A., Pang, B., and Daelemans, W. eds., *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1746–1751, ACL (2014)

[Lecun 98] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, in *Proceedings of the IEEE*, pp. 2278–2324 (1998)

[Mikolov 13a] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, Vol. abs/1301.3781, (2013)

[Mikolov 13b] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J.: Distributed Representations of Words and Phrases and Their Compositionality, in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pp. 3111–3119 (2013)

[Sakai 15] Sakai, T., Tamura, K., and Kitakami, H.: Emergency Situation Awareness during Natural Disasters using Density-based Adaptive Spatiotemporal Clustering, in *Database Systems for Advanced Applications, DASFAA 2015 International Workshops, SeCoP, BDMS, and Posters, Hanoi, Vietnam, April 20-23, 2015*, Vol. 9052, pp. 155–169 (2015)

[佐藤 15] 佐藤 圭, 関 良明 : Twitter を用いた観光地の混雑状況分析手法の検討, 情報処理学会研究報告. SPT, セキュリティ心理学とトラスト, Vol. 2015, No. 15, pp. 1–6 (2015)

[松本 14] 松本 義之, 藪内 賢之 : B3-3 Twitter からの地域・観光情報収集とその有用性の検討 (B3 時系列・言語情報, 一般講演), バイオメディカル・ファジィ・システム学会大会講演論文集, Vol. 27, pp. 87–88 (2014)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ユーザ制約付き独立話題分析 -Separate Link 制約における制約の簡略化-

User constrained independent topic analysis
-Simplification of constraint on separation link constraint-

山本 健太 *1　西垣 貴央 *1　小野田 崇 *1
Kenta Yamamoto　Takahiro Nishigaki　Takashi Onoda

*1青山学院大学理工学部経営システム工学科
Department of Industrial and Systems Engineering，School of Science and Engineering，Aoyama Gakuin University

In this paper, we deals with topic extraction of text mining . Independent topic analysis is a method for extracting mutually independent topics from the text data by using the independent component analysis . User-constrained independent topic analysis is a method of extracting topics satisfying the user's request from the topic of independent topic analysis , and Merge Link constraint and Separate Link constraintare are proposed . We focus on Separate Link constraint that separates one of the topics into two topics . In Separate Link constraint, there is a problem that the burden on the user of selection of the word to be separated is large. In this paper, we propose method that can automate separation word selection and separate topics with high independence . We conducted an evaluation experiment with three benchmark data . We evaluated the independence of topics obtained by the proposed method and obtained the results showing its effectiveness .

1. はじめに

情報技術の発展により，現代は大量の文書データを容易に蓄積，配信できるようになった．また電子ニュースなど継続的に文書データが生成するサービスも盛んになり，日々新たな文書データが生成され続けている．このような利用可能な文書データが増え続けることに対し，人間の情報処理能力は変わらない．検索エンジンの発展により，わずかな入力で大量の文書データを得ることができるようになったが，全ての文書データを確認することは困難である．そのため，蓄積された文書データから有益な情報を抽出するテキストマイニングの研究が盛んに行われている．本論文ではテキストマイニングの課題の 1 つである話題抽出に着目する．ここでの話題とは，大量の文書間で複数の単語の共起によって表現される情報である．現在では，たくさんの話題抽出方法の研究が行われており，その中でもユーザの求める話題を得る方法が必要になってきている．本論文では話題間の関係に着目した話題抽出方法の 1 つで，独立性が高い話題抽出が可能な独立話題分析 [篠原 00] に着目した．

2. 制約付き独立話題分析

独立話題分析は，主に信号処理の分野で注目されている独立成分分析 (Independent Component Analysis; ICA) [Aapo05] を用いて話題を抽出する．独立成分分析とは，入力信号の統計的な性質を利用して異なる特性を持つ信号を分離，抽出する信号処理あるいは多変量解析の問題として多く定式化されている．独立話題分析では 3 つの共通変数がある．話題インデックス $t \in (1,\ldots,k)$，文書インデックス $d \in (1,\ldots,n)$，単語インデックス $w \in (1,\ldots,m)$ がある．次に独立話題分析の諸概念を記述する．$\mathbf{V}$ は $m \times k$ の行列であり，単語 w の話題 t での重要度を示す．また $\mathbf{v}_t$ は，行列 $\mathbf{V}$ の t 列目のベクトル $\mathbf{v}_t = (v_{1,t},\ldots,v_{m,t})^{\mathrm{T}}$ を表し，$\mathbf{v}_w^{\mathrm{T}}$ は，行列 $\mathbf{V}$ の w 行目のベクトル $\mathbf{v}_w = (v_{w,1},\ldots,v_{w,k})$ の転置を表す．$\mathbf{U}$ は $n \times k$ の行列であり，文書 d の話題 t での重要度を示す．同様に $\mathbf{A}$ は $n \times m$ の行列であり，文書 d 中での単語 w の頻度を示す．話題間の独立性を評価する指標には高次統計量の尖度を使用する．尖度を使用した話題の単語集中度の定義は以下のようになる．

$$\sum_{w}^{m}\left(v_{w,t}^{4}P(w)\right) - 3\left(\sum_{w}^{m} v_{w,t}^{2}P(w)\right)^{2}$$

$v_{w,t}$ は行列 $\mathbf{V}$ の w 行 t 列の要素である．$P(w)$ は単語 w の全文書中での出現確率を示し，定義は以下のようになる．

$$P(w) \equiv \frac{\sum_{d}^{n} a_{d,w}}{\sum_{d,w}^{n,m} a_{d,w}}$$

$a_{d,w}$ は行列 $\mathbf{A}$ の d 行 w 列の要素である．話題の単語集中度を用いることで，単語や文書の重要度を 0 に近づけることができる．少数の重要度の大きい単語や文書で話題を表現することができる．独立話題分析は，文書データから話題の単語集中度が最大となり，各話題の独立性が最大となる ${}^*\mathbf{V}$ を回転行列 $\mathbf{R}$ を用いて求める．また独立話題分析で求めたい話題数 k は人が与える．

独立話題分析による話題抽出は，話題間の独立性にのみ着目している．したがって，独立話題分析により得られた話題がユーザの求める話題と異なる場合がある．そのため，独立話題分析によって得られた話題に対し制約を加え，ユーザの要求を満たした話題を生成するユーザ制約付き独立話題分析 [西垣 16] の研究が行われている．ユーザ制約付き独立話題分析は，2 つの制約がある．1 つ目は 2 個の話題を統合する Merge Link 制約，2 つ目は 1 個の話題を 2 個に分離する Separate Link 制約がある．本論文では，1 個の話題を 2 個に分離する Separate Link 制約に着目する．

2.1 Separate Link 制約

Separate Link 制約を満たしてかつ独立性の高い話題を求める Separate Link 制約付き独立話題分析について考える． 各話題は単語の重要度を示す行列 ${}^*\mathbf{V}$ によって表現されている．ユーザはある 1 個の話題から ${}^*\mathbf{V}$ の各要素の絶対値を取った時に値が大きいものの中から，分離したい 2 個の単語（単語 p と単語 q）を選択する．分離したい単語を分離単語とよぶ．そして，選択された分離単語が重要単語となる独立性の高い話題を 2 個（話題 x と話題 y）生成する．具体的には，${}^*\mathbf{V}$ のベク

連絡先: 山本健太，青山学院大学理工学部
経営システム工学科，a5714104@aoyama.jp

トルである $\mathbf{v}_p^{\mathrm{T}}$ と $\mathbf{v}_q^{\mathrm{T}}$ で，要素の絶対値を考えた時，それぞれ話題 x と話題 y で最大値をとるように話題を生成する．制約とは関係のない話題は，はじめに得られた話題から可能な限り変化してほしくないため，制約を満たす話題 x を求めてから更新し，最後に話題 y を求める．Separate Link 制約付き独立話題分析はユーザが分離したい話題と 2 個の分離単語を選択し，その分離単語が重要単語となる独立性の高い話題を抽出する方法である．

2.2 研究課題と目的

ユーザ制約付き独立話題分析の Serprate Link 制約付き独立話題分析では，制約実行時に分離したい話題の選択と 2 個の分離単語の選択が必要がある．特に 2 個の分離単語の選択はユーザの確認量がとても多くなり，ユーザの負担が大きいことが課題である．以上の課題点から本論文では Separate Link 制約実行時のユーザの負担の少ない話題抽出方法を考える．提案手法としてユーザが分離単語を選択せずに，より独立性の高い話題が得られる方法が必要である．

3. 提案手法

Separate Link 制約付き独立話題分析の課題点を解決する話題抽出方法を考える．既存手法の Separate Link 制約付き独立話題析分析は分離したい話題の中でユーザが分離単語を選択し，話題を分離する方法である．提案する話題抽出方法は Separate Link 制約実行時の分離単語の選択によるユーザの負担が少ない方法を考える．また，制約に関係のない話題は制約を与える前の話題から大きく変化しないものと仮定する．本論文の提案手法はユーザの負担を減らすため分離単語選択の自動化を行う．分離単語選択の自動化は，制約実行後の話題間の独立性が高くなるように自動で単語を選択する．Separate Link 制約を満たし，ユーザが分離したい話題を選択するだけで独立性高く話題を分離可能な話題抽出方法を提案する．

3.1 アルゴリズム

提案手法を用いた Separate Link 制約付き独立話題分析のアルゴリズムを以下に述べる．アルゴリズムは 12 個のステップで構成されている．

1. 独立話題分析で任意の数 k の独立な話題を得る．ただし，独立話題分析において，$k+1$ 個の成分を抜き出し，$(k+1)\times m$ の行列 $\mathbf{X}$ を得る．また，回転行列 $\mathbf{R}$ の初期値を $(k+1)\times(k+1)$ の零行列として k 個の話題を求めた．なお，$k+1$ 番目の列は全て零となっているが，これは回転行列を正方行列にするための処置であり，独立話題分析の特性上問題ない．
2. Separate Link 制約を与える話題 z ($z \in \{1,\ldots,k\}$) をユーザが選択する．
3. 文書を話題ごとに分類する．各文書がどの話題に分類されるかは，話題に対する文書の重要度 $\mathbf{U}$ を用いる．各文書の重要度の最も大きい話題に分類する．

   $$C_{\mathrm{t}} = \{a_d | \arg\max_{1\le t\le k} u_{d,t}\},\ C_{\mathrm{t}} \in \{1,\ldots,k\}$$

   Separate Link 制約を与える話題 z に分類された以下の式を満たした文書データを $\mathbf{F}$ とする．また文書データを $\mathbf{F}$ の文書数を h とする．

   $$C_{\mathrm{z}} = \{a_d\}$$

4. Separate Link 制約を与える話題 z に分類された文書データ $\mathbf{F}$ に再び独立話題分析を行い特徴抽出を行う．独立話題数は 2 とし，2 つの独立な話題を求める．ここで求めた話題における単語の重要度を $\mathbf{V}^{\dagger}$ とする．
5. 求めた各話題における単語の重要度 $\mathbf{V}^{\dagger}$ の最も大きい単語を制約での分離単語 p, q とする．

   $$\text{分離単語 } p = \arg\max_{1\le w\le m} v^{\dagger}_{w,1}\ (p \in \{1,\ldots,m\})$$

   $$\text{分離単語 } q = \arg\max_{1\le w\le m} v^{\dagger}_{w,2}\ (q \in \{1,\ldots,m\})$$

   ここで ${\mathbf{v}^{\dagger}}_t$ は，行列 $\mathbf{V}^{\dagger}$ の t 列目のベクトル ${\mathbf{v}^{\dagger}}_t = (v^{\dagger}_{1,t},\ldots,v^{\dagger}_{m,t})^{\mathrm{T}}$ を表す．
6. 単語 p が重要単語となる話題 x を生成する回転ベクトルを次の式で設定する．

   $$\mathbf{c} = (\mathbf{u}^{\mathrm{T}}_{*\alpha}).^2, \quad \mathbf{c} = \mathbf{c}/\|\mathbf{c}\|$$

   $$\alpha = \arg\max_{1\le d\le h} \mathbf{f}_{d,p}, \quad \alpha \in \{1,\ldots,h\}$$

   ここで $\mathbf{u}^{\mathrm{T}}_{\alpha}$ は，行列 $\mathbf{U}$ の α 行目のベクトルの転置を表す．また $\alpha = \arg\max_{1\le d\le n} \mathbf{f}_{d,p}$ は，行列 $\mathbf{F}$ での単語 p を示す列ベクトル $\mathbf{f}_p$ の各要素の値が最も大きい文書番号（行番号）を α としている．$^*\alpha$ は全文書データ内での α の文書番号（行番号）を表す．
7. Separate Link 制約を満たして独立な話題の回転行列を $\ddot{\mathbf{R}}$ として，その最初の列にステップ 6 で求めた $\mathbf{c}$ を代入する．
8. ステップ 1 で得た回転行列 $\mathbf{R}$ の z 列目を抜き出した $(k-1)\times k$ の行列を独立話題分析での $\mathbf{e}_t$ の代わりとして，独立話題分析と同様に $2 \le t \le k$ までの新しい回転行列 $\ddot{\mathbf{R}}$ を求める．
9. 最後にステップ 6 と同様に，単語 q が重要単語となる話題 y を生成する回転ベクトルを次のように表す．

   $$\mathbf{g} = (\mathbf{u}^{\mathrm{T}}_{*\beta}).^2, \quad \mathbf{g} = \mathbf{g}/\|\mathbf{g}\|$$

   $$\beta = \arg\max_{1\le d\le h} \mathbf{f}_{d,q}, \quad \beta \in \{1,\ldots,h\}$$

10. ステップ 9 で得た $\mathbf{g}$ を次の回転行列化を行う．

    $$\mathbf{g} = \mathbf{g} - \ddot{\mathbf{R}}\ddot{\mathbf{R}}^{\mathrm{T}}\mathbf{g}, \quad \mathbf{g} = \mathbf{g}/\|\mathbf{g}\|$$

11. $\mathbf{g}$ を $\ddot{\mathbf{R}}$ の最後の列に代入し，その時の $\ddot{\mathbf{R}}$ が Separate Link 制約を満たす独立な話題の回転行列である．
12. 新たな $k+1$ 個の独立な話題を得る．

これらのステップ 2 からステップ 12 を複数回繰り返すことで，話題を 1 個ずつ増やしていくことが可能である．この方法で Separate Link 制約を満たし，独立性の高い話題を求める．

3.2 既存手法との比較

既存手法と提案手法の特徴の比較を記述する．既存手法はまず独立話題分析の結果からユーザが分離したい話題を選択する．次に分離したい話題の重要単語の中からユーザが 2 個の分離単語（単語 p と単語 q）を選択する．$^*\mathbf{V}$ のベクトルである $\mathbf{v}_p^{\mathrm{T}}$ と $\mathbf{v}_q^{\mathrm{T}}$ で，要素の絶対値を考えた時，それぞれの分離話題で最大値をとるような話題を抽出する話題抽出方法である．既存手法はユーザの分離したい単語を 2 個の話題に分離可能な話題抽出方法である．

対して提案手法は，まず独立話題分析の結果からユーザが分離したい話題を選択する．次に分離したい話題の重要単語の

中から 2 個の分離単語（単語 p と単語 q）を制約実行後の話題間の独立性が高くなるように分離単語を自動で抽出する．既存手法同様に ${}^*\mathbf{V}$ のベクトルである $\mathbf{v}_p^{\mathrm{T}}$ と $\mathbf{v}_q^{\mathrm{T}}$ で，要素の絶対値を考えた時，それぞれの分離話題で最大値をとるような話題を抽出する話題抽出方法である．提案手法は，分離したい話題を独立性高く 2 個の話題に分離可能な話題抽出方法である．

図 1 は既存手法の分離単語選択の一例を表している．図 2 は提案手法の分離単語選択を表している．各図では話題 1 に制約を与え，話題の分離を行うと仮定する．話題 1 を構成する単語を赤円，話題 2 を構成する単語を青円，多くの話題に共通する単語を四角で表現している．Separate Link 制約では分離単語をユーザが重要単語の中から自由に選択することができる．自由に分離単語を選択できるため，得られる話題は一意には定まらない．また図 1 のように単語間の関係性が高い単語も分離単語に選択される場合もあり，話題間の独立性も分離単語の選択で大きく変化する．既存手法はユーザが分離したい単語を選択し分離することが可能な制約である．しかしユーザは分離単語選択の負担が大きく，その分離単語選択で抽出される話題や話題間の独立性が変化する．対して，提案手法ではユーザが分離単語を選択する必要はなく，図 2 のように制約実行後の話題間の独立性が高くなるよう分離単語を自動で決定する．また提案手法では分離単語と得られる話題は一意に定まる．そして得られる話題が定まるため，話題間の独立性も一意に定まる．以上の比較を表 1 にまとめる．

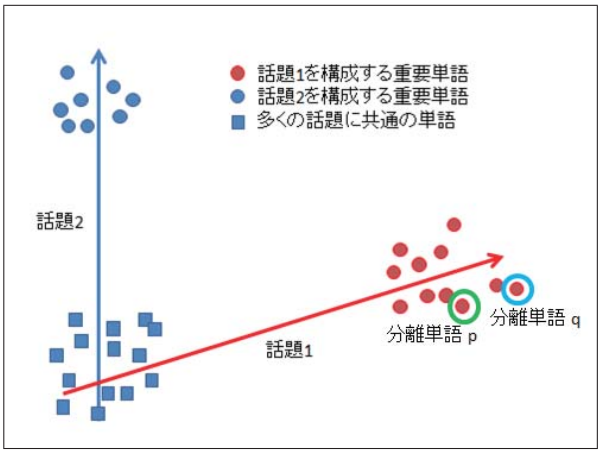


図 1: 既存手法の分離単語例

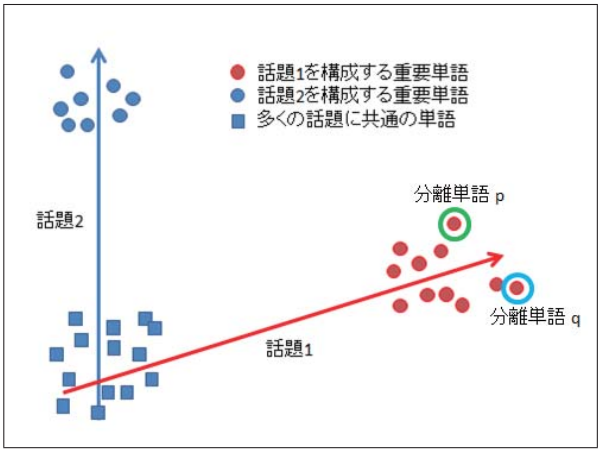


図 2: 提案手法の分離単語例

表 1: 制約の比較

手法	既存手法	提案手法
話題選択	行う	行う
単語選択	行う	行わない
話題	分離単語選択で変化	一意に定まる
独立性	分離単語選択で大きく変化	独立性高く定まる

4. 評価実験

評価実験では，Separate Link 制約を満たしているかの充足性および，提案手法によって得られた話題間の独立性の評価する．

4.1 実験条件

実験データは 3 つのベンチマークデータを用いた．

- $LosAngelsTimes(LATimes)$ の新聞データで文書数は 6279，単語数は 31472 の文書データ [la12][CLUTO]
- $DAILY\,KOSblog(KOS)$ のブログデータで文書数は 3430，単語数は 6906 の文書データ [UCIrep]
- $NeuralInformationProcessingSystems(NIPS)$ の論文データで文書数は 1500，単語数は 12419 の文書データ [UCIrep]

実験の評価は Separate Link 制約を満たしているかの充足性および，提案手法によって得られた話題間の独立性の評価する．まず Separate Link 制約の充足性の評価では，各話題における単語の重要度を示す行列 $\mathbf{V}$ を用いる．2 つの分離単語の各話題における重要度の値から，それぞれの分離話題の重要度が最も大きいことを確認する．

提案手法によって得られた話題間の独立性の評価は，制約により抽出された 2 個の分離話題間の独立性と制約に関係ない話題を含めた全話題間の独立性を評価する．独立性の評価は話題間の相互情報量を用いる．話題間の相互情報量が 0 の場合，2 つの話題は完全に独立していることを意味し，相互情報量が小さいほど話題間の独立性は高くなる．相互情報量は独立話題分析によって得られた $\mathbf{V}$ の値を用いて求める．提案手法は分離単語と得られる話題は一意に定まるが，既存手法はユーザが分離単語を選択するため，得られる話題は一意には定まらない．そこで既存手法との比較では，分離単語を重要単語の中からランダムで選択した 20 回の平均を用いて行う．

4.2 実験結果，考察

実験結果の一例として $LATimes$ に話題数 7 で独立話題分析を行った結果の話題 6 に評価実験を行った結果を示す．$LATimes$ に話題数 7 で独立話題分析を行った結果を表 2 に示す．まず，提案手法は分離したい話題の文書を抽出し，再度その文書に独立話題分析を行う．分離したい話題 6 の文書に対し，独立話題分析を行なった結果を表 3 に示す．分離したい話題 6 を構成する 2 つの話題が抽出された．提案手法での分離単語は抽出された表 3 の各話題で最も重要度の大きい bank と price に決定された．次に bank と price を分離単語として話題 6 を分離した結果を表 4 に示す．

次に提案手法により得られた話題の評価を行う．まず，Separate Link 制約の充足性の評価として各話題における分離単語 bank と price の重要度を表 5 に示す．表 5 よりそれぞれの分離話題での重要度が最も大きいことから Separate Link 制約の充足性が確認できる．次に提案手法によって得られた話題間の独立性の評価を表 6 に示す．分離話題間の独立性，全話題

間の独立性ともに提案手法により得られた話題間の独立性が高く，提案手法の有効性が確認できる．*LATimes* に異なる話題数，話題の評価実験の結果，*Nips*, *Kos* への適応結果も同様に提案手法によって得られた話題間の独立性が高く，提案手法の有効性が確認できた．

比較手法の独立性が低くなった理由としては，重要単語の中からランダムに分離単語を選択することで分離単語間の関係性が高い単語のペアが選択される場合があることが考えられる．また制約に関係のない話題の重要単語と関係性が高い単語が分離単語として選択される場合があることも理由の 1 つであると考えらえる． 単語間の関係性が高い分離単語を用いて話題を分離することで，分離話題間の相互情報量や分離話題と制約に関係のない話題間の相互情報量が高くなり独立性が低くなる．対して提案手法は分離したい話題がどのような話題で構成されているかを分析し，2 個の分離単語を決定することで独立性が高くなるように単語を選択することができる．その結果提案手法は比較手法と比べ独立性が高い話題抽出ができたと考えられる．

表 2: *LATimes* の新聞データに話題数 7 で独立話題分析による話題抽出結果

話題	重要度が高い単語				
	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	million	earn	quarter	revenu	net
2	scor	game	lead	rebound	league
3	soviet	afghanistan	israel	foreign	militari
4	macmin	aleen	art	entertain	report
5	polic	bush	counti	car	arrest
6	stock	bank	price	market	rate
7	game	team	player	coach	bowl

表 3: 独立話題分析結果表 2 の分離したい話題 6 の文書に独立話題分析実行結果

話題	重要度が高い単語				
	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	bank	insur	loan	sav	deposit
2	price	bond	index	stock	rate

表 4: 独立話題分析結果表 2 の話題 6 に bank と price を分離単語とし，提案手法による話題抽出結果

話題	重要度が高い単語				
	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	stock	polic	bank	price	market
2	million	earn	quarter	revenu	net
3	scor	game	lead	rebound	league
4	soviet	afghanistan	israel	foreign	militari
5	macmin	aleen	art	entertain	report
6	polic	bush	counti	car	arrest
7	game	team	player	coach	bowl
8	earn	quarter	stock	price	rate

5. まとめ

本論文では，独立性の高い話題抽出方法である独立話題分析にユーザ制約を加えたユーザ制約付き独立話題分析を扱った．そして，ユーザ制約付き独立話題分析の Separate Link 制約付き独立話題分析に着目した．Separate Link 制約は 1 個の話題を 2 個の話題に分離する制約である．Separate Link 制約付き独立話題分析にはユーザの分離単語選択の負担が大きい課題がある．提案手法は分離単語選択の自動化により Separate Link 制約実行時のユーザの負担を軽減し，独立性の高い話題を抽出する方法を考案した．分離単語の自動化ではまず，分離したい話題にどのような話題が含まれているかを独立話題分析により再度抽出した．そして再度独立話題分析した結果から自動で分離単語を選択する方法を用いた．実験では提案手法を 3 つのベンチマークデータに適応させた．実験結果から提案手法は，ユーザの負担を軽減した上で独立性の高く話題を分離可能な手法であると示された．今後は提案手法で得られる話題の独立性の高さの理論的証明が必要であると考えられる．また分離できる話題数を 3 個，4 個と増やすことで汎用性が高まると考えている．

表 5: 提案手法による話題抽出結果表 4 での分離単語 bank と price の各話題の重要度 $|v_{w,t}|$

	$w=$ bank	$w=$ price
$v_{w,\text{ 話題 }1}$	**34.243**	22.963
$v_{w,\text{ 話題 }2}$	1.028	12.963
$v_{w,\text{ 話題 }3}$	8.056	2.709
$v_{w,\text{ 話題 }4}$	8.208	0.973
$v_{w,\text{ 話題 }5}$	7.197	3.609
$v_{w,\text{ 話題 }6}$	3.790	10.826
$v_{w,\text{ 話題 }7}$	17.410	1.151
$v_{w,\text{ 話題 }8}$	25.03	**36.302**

表 6: 提案手法により得られた話題間の独立性の評価

	提案手法	比較手法 (ランダム)
分離話題間	**0.0383**	0.0680
全話題間	**2.1034**	2.4624

参考文献

[篠原 00] 篠原 靖志: 文書データベースの主要話題の発見と変化の追跡を行う文書閲覧支援システムの開発，電力中央研究所報告，(2000)

[Aapo05] Aapo Hyvarinen， Juha Karhunen， Erkki Oja: 詳解　独立成分分析　信号解析の新しい世界，東京電機大学出版局，(2005)

[西垣 16] 西垣 貴央， 新田 克己， 小野田 崇:制約付き独立話題分析，人工知能学会論文誌，(2016)

[la12] Shi Zhong， Joydeep Ghosh:A comparative study of generative models for document clustering，Data Mining Workshop on Clustering High Dimensional Data and Its Applications，(2003)

[CLUTO] Ying Zhao， George Karypis:Evaluation of hierarchical clustering algorithms for document datasets，Conference of Information and Knowledge Management (CIKM)，(2002)

[UCIrep] M. Lichman:UCI machine learning repository，http://archive.ics.uci.edu/ml，(2013)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

データ追加に基づく独立話題分析の話題変化検知手法の検討

Discussion on topic change detection method in Incremental Independent Topic Analysis

西垣 貴央 *1　　小野田 崇 *2
Takahiro Nishigaki　　Takashi Onoda

*1青山学院大学
Aoyama Gakuin University

We have proposed a topic extraction method that independent topic in increasing data. This algorithm extract independent topics form a small number of document data, update the independent topics when the new data comes. However, this method does not assume that the topic changes and the number of topics changes. In this paper, we propose a method to detect when the topic changes and when the number of topics changes. And we use some benchmark datasets to evaluate the proposed method. Evaluation results using benchmark data show that the proposed method is able to detect the topic changes and detect the number of topics changes.

1. はじめに

Web 上や個人所有のハードディスクドライブ (HDD) には大量の文書データが日々生成および蓄積されている．蓄積されている大量の文書データの中から，有益な知識を発見・抽出するためのテキストマイニング技術の一つである話題抽出について取り上げる．ここで話題とは， bag-of-words として与えられた大量の文書間で，複数の単語の共起によって表現される情報のことである [佐藤 15]．この話題を抽出する方法には様々な方法が存在するが，本稿では，独立性の高い話題を求める独立話題分析 [篠原 00] について考える．独立話題分析では，信号処理の分野で使用される独立成分分析 [Hyvärinen01] を用いて話題を求めている．ここで独立性が高い話題とは，話題間の相互情報量が小さい話題を示している．独立性が高い話題を求める利点として，より多くの情報量を持つ要約の作成が，容易にできる可能性が高いことが挙げられる．また，この独立話題分析を用いたシステムとして文書閲覧支援システム IT-DMS (Independent Topic-based Document Management System)[篠原 00] や，IT-DMS を改良した大量文書データに対する文書整理システム [田中 03] などがある．

逐次増加する文書データへ独立話題分析を適用できるように改良した，データ追加に基づく独立話題分析とよばれる方法 [Nishigaki17, 西垣 17] も提案されている．この方法では，初期データが抽出した独立話題を，新しい文書データが入ってくる度に更新することで，全てのデータを使用したときの話題と同等の話題を得ることが出来る．しかしこれら方法では，話題の中身が変化する場合や，話題の数が変わるということを想定していない．そこで本論文では，データ追加に基づく独立話題分析において話題の変化を検知する手法の検討を行う．

以下，2 章でデータ追加に基づく独立話題分析について簡単に紹介し，3 章で話題変化を検知する手法の検討を行う．4 章では，3 章で検討した方法をベンチマークデータに適用し，その評価を行う．最後に 5 章でまとめと今後の課題について述べる．

連絡先: 西垣 貴央，青山学院大学 理工学部 経営システム工学科，〒 252-5258 神奈川県相模原市中央区淵野辺 5-10-1 O-421c，nishigaki@ise.aoyama.ac.jp

2. Incremental Independent Topic Anlaysis

本章では，初期データのみで抽出した独立性の高い話題を，データが増加するたびに更新することで，全てのデータを用いて抽出した独立性の高い話題に近づける方法 [Nishigaki17, 西垣 17] について説明する．提案手法では，まず初期データに対して独立話題分析を適用し，独立性の高い話題を抽出する．次に，抽出した独立性の高い話題にもっとも影響を与えている文書データを抽出した独立性の高い話題の数だけ抜き出し，抽出した独立性の高い話題と抜き出した文書データ以外のデータ全てを削除する．それから，データが追加されると追加されたデータと抽出した独立性の高い話題と抜き出した文書データを合わせ一つのデータとする．その合わせたデータを用いて，抽出した独立性の高い話題を FPICA を用いて更新する．再び，更新した独立性の高い話題と抜き出した文書データのみを残して他のデータ全てを削除する．これを繰り返すことで，最終的に独立性の高い話題を得る．提案したデータ追加に基づく独立話題分析のアルゴリズムを以下に述べる．共通の変数として，話題インデックスを $t \in \{1, \ldots, k\}$，文書インデックスを $d \in \{1, \ldots, n\}$，単語インデックスを $w \in \{1, \ldots, m\}$ とする．

1. 初期データの各文書中の各単語の頻度の行列 $\mathbf{A}$ を作成し，単語数の偏りが出ないように正規化を行い $\tilde{\mathbf{A}}$ を得る．
2. $\tilde{\mathbf{A}}$ に対して特異値分解を行い，$\tilde{\mathbf{A}}$ を次のように分解する $\hat{\mathbf{U}}^{\mathrm{T}}\tilde{\mathbf{A}}\hat{\mathbf{V}} = \hat{\mathbf{S}}$．ここで，$\hat{\mathbf{S}}$ は特異値の対角行列である．
3. ステップ 2. で得た行列 $\hat{\mathbf{U}}$ と $\hat{\mathbf{S}}$，$\hat{\mathbf{V}}$ を，$\hat{\mathbf{S}}$ の値の大きい順に k 個の成分を抜き出し，行列 $\mathbf{U}$，$\mathbf{S}$，$\mathbf{V}$ を作成する．
4. k 次元空間での話題を示す $k \times m$ の行列 $\mathbf{X}$ を次の式で定義する $\mathbf{X} = \mathbf{S}^{-1/2}\mathbf{U}^{\mathrm{T}}\tilde{\mathbf{A}}$.
5. 各話題の独立性最大化：ステップ 4. で得られた話題に対して，FPICA[Hyvärinen99] に基づいて最大の独立性を与えるための回転行列 $\mathbf{R}$ を次のように決定する．
   (a) $\mathbf{R}$ の初期値を $k \times k$ の零行列とする $\mathbf{R} = \mathbf{0}$.
   (b) 単位行列 $\mathbf{I} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k)$ の $t \in \{1, \ldots, k\}$ 列目の列ベクトルを $\mathbf{e}_t$ として，回転行列 $\mathbf{R}$ の t 列目

$\mathbf{r}_t = (r_{1,t}, r_{2.t}, \ldots r_{k,t})^{\mathrm{T}}$ に代入する $\mathbf{r}_t = \mathbf{e}_t$．ここで，$\mathbf{e}_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$，$\mathbf{e}_2 = (0, 1, 0, \ldots, 0)^{\mathrm{T}}$ の $k \times 1$ の単位ベクトルである．

(c) $\mathbf{r}^{(old)}$ に $k \times 1$ の零ベクトルを代入して，$\mathbf{r}^{(old)}$ を次のように初期化する $\mathbf{r}^{(old)} = (0, 0, \ldots, 0)^{\mathrm{T}}$．

(d) $\mathbf{r}_t$ を次の式で更新する $\mathbf{r}^{(old)} = \mathbf{r}_t$，$\mathbf{r}_t = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{r}_t).^3 - 3\mathbf{r}_t$．$(\mathbf{X}^{\mathrm{T}}\mathbf{r}_t).^3$ は $\mathbf{X}^{\mathrm{T}}\mathbf{r}_t$ の行列要素の 3 乗を表す．

(e) $\mathbf{r}_t$ を次の回転行列化を行う $\mathbf{r}_t = \mathbf{r}_t - \mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{r}_t$，$\mathbf{r}_t = \mathbf{r}_t / \|\mathbf{r}_t\|$．

(f) $\|\mathbf{r}_t \pm \mathbf{r}^{(old)}\|$ が閾値以上ならば，ステップ 5d. へ．閾値より小さければステップ 5g. へ．

(g) $t < k$ ならば，t を 1 つ増やして，ステップ 5b. へ．$t = k$ ならば，その時の $\mathbf{R}$ を回転行列として，ステップ 6. へ．

6. 独立な話題中での単語の重要度 ${}^*\mathbf{V}$ と独立な話題中の文書の重要度 ${}^*\mathbf{U}$ を下記により計算する．

$$ {}^*\mathbf{V} = \mathbf{V}\mathbf{R}, \quad {}^*\mathbf{U} = \mathbf{U}\mathbf{R} $$

7. 得られた各話題に対して $\mathbf{u}_k$ の絶対値が最も大きい文書データを抜き出す．

8. ステップ (6) で得られた話題とステップ (7) のデータを除いて他の全てのデータを削除する．

9. 新しく追加されたデータに，ステップ (8) で残ったデータを連結する．

   - このとき新しく追加されるデータは，ステップ (1) での初期データに対して非常に小さいものとする．

10. データが増加した後の新たな独立な話題を求める回転行列を $\tilde{\mathbf{R}}$ とし，その初期値をステップ (6) で得られた $\mathbf{R}$ とする．

11. ステップ (9) のデータに対して，ステップ (10) の回転行列 $\tilde{\mathbf{R}}$ を独立話分析のステップ (5) と同様に FPICA に基づいて新しい回転行列を得る．

12. データが加えられた後の，新しい独立な話題が得られる．

このステップ (7) からステップ (12) をデータが増加する度に繰り返すことで，データ追加に基づいた独立性の高い話題を抽出することができる．

データ追加に基づく独立話題分析のイメージを図 1 に示す．この方法によって，増加するデータに対しても全てのデータを同時に用いることなく，独立な話題を求めることができる．

3. 話題変化を検知する手法の検討

独立話題分析では，抽出された話題中での単語の重要度 ${}^*\mathbf{V}$ と話題中の文書の重要度 ${}^*\mathbf{U}$ によって，話題が表現される．抽出された話題の $\mathbf{v}_t$ と $\mathbf{u}_t$ はそれぞれ，t 番目の話題の単語の重要度と文書の重要度を表すベクトルである．また，τ 回目に得られた t 番目の話題の単語重要度を $\mathbf{v}_t^{(\tau)}$ と表現することにする．$\mathbf{u}_t^{(\tau)}$ についても同様である．

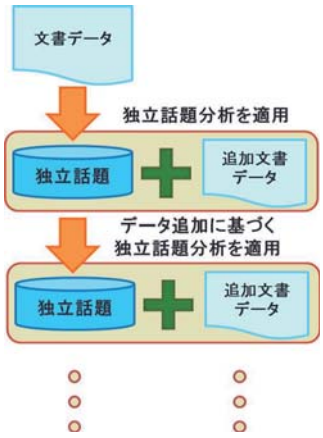


図 1: データ追加に基づく独立話題分析のイメージ

本稿では，1) τ 回目に抽出された話題と $\tau + 1$ 回目に抽出された話題が，同じ話題なのか異なる話題なのかを検知することと，2) τ 回目までは存在しなかった話題が $\tau + 1$ 回目から新たに現れたことを検知することを，目標とする．

3.1 話題の変化の検知

抽出された話題の $\mathbf{v}_t^{(\tau)}$ および $\mathbf{u}_t^{(\tau)}$ と，$\mathbf{v}_t^{(\tau+1)}$ および $\mathbf{u}_t^{(\tau+1)}$ を比較することで，話題の強弱を調べることができると考えた．比較には，各話題はすべてベクトルで表現されるので，コサイン類似度を用いて以下の式で表す．

$$ \cos(\mathbf{v}_t^{(\tau)}, \mathbf{v}_t^{(\tau+1)}) = \frac{\mathbf{v}_t^{(\tau)} \cdot \mathbf{v}_t^{(\tau+1)}}{|\mathbf{v}_t^{(\tau)}||\mathbf{v}_t^{(\tau+1)}|}, \quad t \in \{1, \cdots, k\} $$

k 個の話題それぞれで，$\mathbf{v}_t^{(\tau)}$ と $\mathbf{v}_t^{(\tau+1)}$ のコサイン類似度を計算することで，その話題を構成する単語の重みの違いが明確になる．コサイン類似度の値が 1 もしくは -1 の場合は，全く同じ強さを持った話題となり変化はなかったことを示す．一方で，コサイン類似度の値が 0 に近づくにつれて，異なる話題となり変化があったことを示す．この値の絶対値がある値 α よりも小さければ，τ 回目の話題 t と $\tau + 1$ 回目の話題 t は異なる話題であるということが言える．また，抽出された話題が変化がなかったとしても，その話題に分類される文書数の変化によって，その話題の強弱を調べることができる．話題の強弱とは，その話題が強い場合，多くの文書がその話題に分類されていることを指し，反対に弱い場合は，その話題に分類される文書数が少ないことを指す．話題への文書データの分類は，次の式を用いて行う．式中の $\mathbf{a}_i$ は i 番目の文書を表しており，$\mathbf{u}_i^{\mathrm{T}(\tau)}$ は，τ 回目における i 番目の文書の話題への影響度を表したベクトルとなっている．つまり，$\mathbf{u}_i^{\mathrm{T}(\tau)}$ の値が最大値を取るのが t 番目の話題のとき，その文書 $\mathbf{a}_i$ は話題 k に分類される．$\mathcal{T}_t$ は話題 t に分類された文書数を表す．

$$ \mathcal{T}_t = \{\mathbf{a}_i | \arg\max_i \mathbf{u}_i^{\mathrm{T}(\tau)} = t\}, \, i \in \{1, \ldots, \text{追加される文書数}\} $$

これをすべての話題で行い，各話題へ分類されている文書数の多さで話題への強弱を推定する．

3.2 新規話題の検知方法

新しい話題の出現の検知方法にも，コサイン類似度を使用する．$\tau + 1$ 回目で抽出されたある話題が，τ 回目の全ての話題と一定以上離れている場合，$\tau + 1$ 回目で抽出されたある話

表 1: LA Times に話題数 6 で独立話題分析を行った時の各話題に分類された文書数

話題番号	話題 1	話題 2	話題 3
文書数	898	704	853
話題番号	話題 4	話題 5	話題 6
文書数	432	924	2468

題は新規話題であると推測できる．

$$\cos(\mathbf{v}_t^{(\tau)}, \mathbf{v}_j^{(\tau+1)}) = \frac{\mathbf{v}_t^{(\tau)} \cdot \mathbf{v}_j^{(\tau+1)}}{|\mathbf{v}_t^{(\tau)}||\mathbf{v}_j^{(\tau+1)}|}, \quad (t, j \in \{1, \cdots, k\},\ t \neq j)$$

このコサイン類似度を計算し，この値の絶対値がある値 α よりも小さければ，その話題 j は新規話題であるということが言える．

4. 実験

4.1 実験に使用するデータについて

実験でベンチマークデータとして Los Angeles Times (LA Times) の新聞データ [Zhong03] で文書数は 6279，単語数は 31472，話題数は 6 の文書データを使用する．

その新しい文書の中に新規話題がある場合を考える．しかし，LA Times のデータは時系列データではないため，話題の変化および話題数の変化については情報がない．そこで本実験ではデータの準備として，最初に LA Times に話題数 6 で独立話題分析を行い，各文書データを各話題に分類した．分類には抽出した独立性の高い話題とのコサイン距離を用いた．その時，各話題での文書数は，表 1 となった．表 1 の数で各文書が分類され，話題 5 を除いた話題 1，話題 2，話題 3，話題 4，話題 6 の 5 個の話題に属する文書数 5355 と話題 5 の文書数 924 を別に扱う．

本実験では，データ追加に基づく独立話題分析では，初期データとして全データの 50% を用いて初期の独立話題を抽出する．その後，必要最低限の情報だけを残して，新しい文書数 200 を逐次追加していく．初期データとして全データの 50%（3139 文書）を話題 5 を除いた文書からランダムに選択し，初期の独立話題を抽出を行った．その後，200 ずつデータを追加していくため，全部で 16 回データの追加が行われた．追加されるデータの 4 回目までは，既存の話題のみが含まれるデータとして，5 回目から少しずつ新規話題（話題 5）が含まれるデータを追加することとした．

以上のデータを用いて，提案する方法で話題の変化および話題数の変化を検知することができるか検証を行う．

4.2 結果と考察

まず話題の変化の検知について結果を示す．話題の変化の検知では，同じ話題と考えられる話題同士のコサイン類似度によって決定する．表に，τ 回目と $\tau+1$ 回目での話題でのコサイン類似度を角度に変換した値を示す

表 2 では $\tau = 0$ で非常に大きな角度になっていることが分かる．これは，使用しているデータ数が初期データでは 3139 あるのに対して，追加以降は 212 になるためであると考えられる．そのため，1 回目は除外する．$\tau = 4$ 回目までは既存の話題のみを追加しているため，5 回目と 6 回目の角度に注目すると，$\tau = 5$ 回目では話題 4 と話題 6 が，6 回目では話題 5 と話題 6 のときに，10 度以上離れていることが分かる．この時，

表 2: τ 回目と $\tau+1$ 回目での話題の角度

τ	話題 1	話題 2	話題 3	話題 4	話題 5	話題 6
0	16.41	3.39	4.35	4.14	12.59	18.48
1	7.81	4.05	8.00	0.14	0.25	0.19
2	8.12	2.28	7.53	5.45	4.01	1.43
3	4.88	6.55	4.95	3.8117	4.74	0.26
4	4.96	3.85	3.98	11.38	4.62	13.24
5	3.98	3.2383	4	0.94738	11.71	11.40
6	0.02	6.24	4.19	3.89	1.946	4.50
7	5.04	5.66	2.33	4.32	1.17	2.06
8	3.8	3.84	4.97	2.09	0.22	0.137
9	4.99	3.84	3.37	3.93	6.38	1.94
10	0.16	5.23	4	2.78	1.81	4.075
11	4.35	3.31	1.81	2.13	0.42	1.21
12	4.18	4.95	3.77	0.37	1.93	1.77
13	3.77	8.03	4.41	2.2	8.08	0.31
14	4.16	8.07	3.76	3.48	8.15	3.75
15	3.76	5.26	3.99	3.76	2.15	3.47

話題が変化したのではないかと考えられる．以上のことより，α を 0.9848 とすることで，10 度以上離れたら話題が変化したとすることができる．

次に話題の強弱について，各話題でどの程度文書が含まれているかを表 3 を用いて考察する．

表 3 をみると，各話題に含まれる文書数は各回ごとにかなり異なることが分かる．話題が変化したと考えられる，5 回目の話題 4 と話題 6，6 回目の話題 5 と話題 6 を見ると，大きく数が増減していることが分かる．各話題に分類される文書数が大きく変わったから，話題の変化が起こった様にも考えられるが，2 回目の話題 6 と 3 回目の話題 6 に分類される文書数を見ると，ここも大きく減少している．しかし，表 2 の $\tau = 2$ の話題 6 を見ると，1.43 と非常に小さいことが分かる．これは話題数は大きく減っているが，話題の内容は変化していないことを示している．

最後に，提案した新規話題の検知方法について検証を行う．本実験では，$\tau = 4$ のときに新しい話題が出現している．その時の各話題での角度を表 4 に示す．表 4 では，各話題同士の角度を表しており，0 度もしくは 180 度のときは同じ話題であるということを示している．90 度のときは，独立した話題であると言える．全く同じ話題でコサイン類似度を計算した場合，ある 1 個の話題が 0 度もしくは 180 度となり，他全ての話題が 90 度となる．合計の行では，それぞれの話題が 90 度あるいは 0 度もしくは 180 度からどの程度離れているのかの合計値を示している．そこで話題が大きく変化して新規話題が出現した場合，この合計値の値が大きな値となると考えられる．表 4 の合計値を確認すると，$\tau+1$ の話題 6 は τ の全ての話題から 38 度以上離れていること示しており，ここで $\tau+1$ 回目の話題 6 が新規話題である可能性が高いことを示している．

以上の結果より

5. おわりに

本論文では，データが逐次増加する場合における独立話題分析において，データが増加した時に話題の変化および話題数の変化を検知する手法を提案した．また，LA Times のデー

表 3: τ 回目での各話題に分類された文書数

τ	話題 1	話題 2	話題 3	話題 4	話題 5	話題 6
1	6	4	48	2	67	85
2	19	15	34	22	12	110
3	72	55	30	12	41	2
4	14	120	10	46	17	5
5	32	28	24	3	37	88
6	34	71	26	73	4	4
7	41	74	39	4	5	49
8	35	11	43	8	98	17
9	109	42	11	15	12	23
10	22	47	27	82	25	9
11	39	68	41	22	5	37
12	5	29	84	52	4	38
13	28	18	65	42	49	10
14	7	20	94	53	29	9
15	7	6	18	102	40	39
16	7	17	65	43	16	4

表 4: $\tau = 4$ の時と $\tau = 5$ の時の各話題との角度

話題	$1(\tau+1)$	2	3	4	5	6
$1(\tau)$	93.43	87.85	176.02	89.93	90.12	90.25
$2(\tau)$	93.11	86.79	89.53	80.15	93.73	13.24
$3(\tau)$	175.04	90.47	86.85	89.79	90.08	94.35
$4(\tau)$	89.95	3.85	87.78	90.75	89.66	92.82
$5(\tau)$	90.41	90.09	90.13	90.54	175.38	94.79
$6(\tau)$	90.96	90.31	90.53	11.38	91.51	103.02
合計	12.93	10.04	10.47	22.80	10.40	38.79

タを使用して，提案手法によって話題の変化および話題数の変化の検知が行うことができるのか検証した．検証の結果，提案した話題の変化の検知および新規話題の検知方法の両方共 LA Times の場合はうまく検知できていると考えられる．しかし，LA Times は時系列データではないため，理想環境としてデータの追加を行っており，時系列情報が含まれるデータや他のベンチマークデータでも良い結果が得られるとは限らない．また，話題の変化や新規話題の検知で用いる角度は，どの程度を離れていると変化が起こった，あるいは新規話題であるということができるのか，LA Times の結果だけではわからないため，他のデータに対して行う必要がある．さらに，本稿ではコサイン類似度の値のみでの分析しか行っていないため，それが本当に新規話題を示しているのか，たまたま異なる話題が得られてしまったのか検証が行えていないため，話題に含まれる単語を比較して分析を行う必要がある．

謝辞

本研究の一部は，青山学院大学アーリーイーグル研究支援制度の支援によって行われた．

参考文献

[Hyvärinen99] Aapo Hyvärinen: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis, IEEE Trans. on Neural Networks, Vol.10, No.3 (1999).

[Hyvärinen01] Aapo Hyvärinen and Juha Karhunen and Erkki Oja: Independent Component Analysis, John Wiley & Sons (2001).

[Nishigaki17] Takahiro Nishigaki, Katsumi Nitta, Takahis Onoda: Incremental Learning of Independent Topic Anlysis, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol. 11, No. 2, pp.191–197 (2017).

[Lichman13] M. Lichman: UCI machine learning repository, http://archive.ics.uci.edu/ml , 2013.

[Zhong03] Shi Zhong and Joydeep Ghosh: A comparative study of generative models for document clustering, Data Mining Workshop on Clustering High Dimensional Data and Its Applications (2003).

[佐藤 15] 佐藤 一誠: トピックモデルによる統計的潜在意味分析, 自然言語処理, 第 8 巻, コロナ社 (2015).

[篠原 00] 篠原 靖志: 文書データベースの主要話題の発見と変化の追跡を行う文書閲覧支援システムの開発, 電力中央研究所報, R99036 (2000).

[田中 03] 田中 真人, 篠原 靖志: 重要話題発見のための大量文書自動整理システム, 電力中央研究所報告, R02015 (2003).

[西垣 17] 西垣 貴央, 新田 克己, 小野田 崇: 逐次データ追加がある状況下での独立話題分析とその実験的特性分析, 人工知能学会全国大会 (第 31 回) 論文集, 1E3-1, (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

有価証券報告書からの因果関係文の抽出

Extraction of Causal Knowledge from Annual Securities Report

佐藤 史仁[*1] 佐久間 洋明[*1] 小寺 俊哉[*1] 田中 良典[*1] 坂地 泰紀[*2] 和泉 潔[*2]
Fumihito Sato Hiroaki Sakuma Shunya Kodera Yoshinori Tanaka Hiroki Sakaji Kiyoshi Izumi

[*1] 日興リサーチセンター株式会社
Nikko Research Center, Inc.

[*2] 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

In annual securities report, various information such as corporate policy, risk management, R&D, and so on, is included other than business performance. Previous researches proposed the extraction methods of important sentences containing causal information from financial articles and texts but not annual financial reports. In this paper, we applied these extracting methods based on SVM discriminant model to annual securities reports in our original way. Our method indicated high performance and all evaluations, that were precision, recall and F-score, showed more than 0.8. By using our model, useful information from annual securities reports would be collected effectively, which allow us to make unique investment decisions.

1. はじめに

投資判断に利用できる情報として，財務データやマーケットデータなどの数値データの他，ニュースや新聞記事，決算短信，有価証券報告書などに含まれるテキストデータがある．これらテキストデータの特徴は，過去又は将来の業績に対する理由及び根拠や経営戦略に関する情報，進行中の施策，新商品発表情報，抱えているリスク，企業の不祥事に関する情報など，投資判断において数値データにはない重要な情報を含んでいる点が挙げられる．しかしながら，構造化することが難しく，そのほとんどは投資家が直接読まなければ投資判断に利用できない場合が多かった．近年，この問題に対し，テキストマイニングなどの人工知能分野の技術を，金融市場における分析に導入して解決を試みる研究が盛んに行われている．例えば，ある企業に関連するニュースがその株価にとってポジティブに働くかネガティブに働くか（極性）でニュースを定量化し，株式リターンとの関係を分析した研究[五島 2015] [沖本 2014]が挙げられる．これらの研究は，投資家が直接そのニュースを読まなくとも，ニュースに極性を付与し定量化することで，投資戦略に活用できることを示唆している．

テキストデータの定量化以外にも，テキストデータから投資判断等に関する重要文を抽出する手法の研究が行われている．例えば，坂地ら[坂地 2011] [坂地 2015]は，文の表現と機械学習によって，過去の業績や製品の売れ行きなどに対する因果関係文を経済新聞の記事や決算短信から抽出する手法を提案している．因果関係文とは，出来事（結果）とその理由（原因）の組から構成される文と定義される．例えば，原因「猛暑」による結果「冷房需要の盛り上がり」等の因果関係を提示することで，「猛暑」の際には「冷房需要」が高まる可能性があるという情報を得られるとした．因果関係文の抽出以外にも，決算短信から業績の要因を含む文を抽出した研究[酒井 2015b]や，業績の予測を示す文を抽出した研究[北森 2017]がある．さらに，抽出された重要文に極性を付与する研究[酒井 2015a]も行われている．このように，テキストデータからの重要文の抽出手法は，投資家が投資判断に有益な情報を効率良く把握することを可能にし，また，ある特定のテーマを持った重要文の定量化データを既存の分析や投資戦略等へ導入することで，新しい投資戦略や手法の開発に役立つだろう．しかしながら，多様なテキストデータからある特定のテーマを持った重要文についての抽出手法やその定量化が多岐にわたり研究されているが，現在のところ，それらの統一的な抽出手法や数量化モデルは存在しない．言い換えれば，抽出対象となるテキストデータや重要文によって，抽出手法や定量化手法に工夫が必要であると言える．

有価証券報告書は，上場会社が証券取引所から求められている適時開示資料である決算短信に対し，金融商品取引法により提出が定められている開示資料である．有価証券報告書は，決算短信と比較して速報性はないものの[*1]，「業績等の概要」等の業績に関する情報だけでなく，「対処すべき課題」や「事業等のリスク」など投資判断に有益と考えられる情報をより多く含む．また，ニュースや新聞記事は限られた企業に関する情報が多いが，有価証券報告書は全ての上場企業から公表されているという利点もある[*2]．投資判断に有益なテキスト情報としては，まずは業績に関する情報が考えられる．どんな事象が原因で，そのような業績結果となったのかを知るには，業績に関する項目の因果関係文を抽出すれば良い．また，何をリスク要因や経営課題として捉えて，それに対しどんな対策を講じているのかを知るには，リスクやその企業がもつ課題に関する項目の因果関係文を抽出することで把握できると考えられる．つまり，有価証券報告書の各項目から因果関係文を抽出することは，単純に業績に関する原因と結果だけでなく，その企業のリスク対策や目指している方向性などを把握するための有力情報を取得することになると言えるだろう．にもかかわらず，因果関係文の抽出に関する研究の中で，新聞記事や決算短信を対象とした研究はあるものの，有価証券報告書を対象にした報告はない．そこで，本稿では，有価証券報告書から「業績等の概要」，「対処すべき課題」，「事業等のリスク」を対象に因果関係文を抽出する判別モデルを提案する．具体的には，坂地ら[坂地 2011]の手法を応用し，

連絡先：佐藤 史仁，日興リサーチセンター株式会社，〒103-0026 東京都中央区日本橋兜町 8-1

[*1] 決算短信は，決算期末後 45 日以内の開示が適当とされ，30 日以内の開示がより望ましいとされている．一方，有価証券報告書は，やむを得ない場合を除いて事業年度末から 3ヵ月以内の公表が求められている．

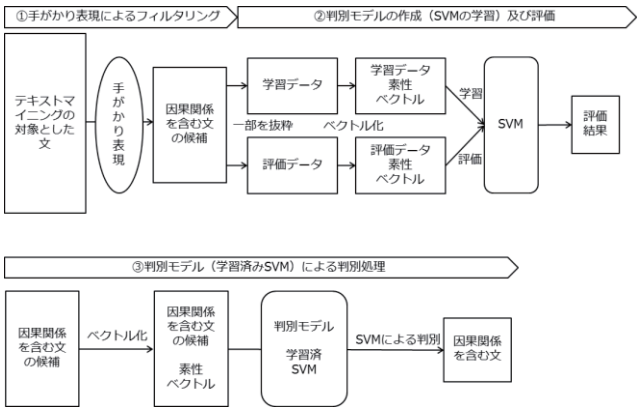
[*2] 有価証券報告書は，金融商品取引法第二十四条により提出が定められている．企業の HP の他，金融庁の EDINET などから入手可能．http://disclosure.edinet-fsa.go.jp/

有価証券報告書専用の因果関係文を抽出する機械学習を用いた判別モデルを作成した. データは TOPIX1000 を構成する企業の 2008 年から 2016 年までの有価証券報告書(本決算)に含まれるテキストデータを用いた.

2. 因果関係文の判別モデル

本稿では, 新聞記事から因果関係を含む文を抽出する研究[坂地 2011]で用いられた手法を応用した. この手法は, より広範の因果関係文を対象とできることや高い抽出性能となったことが報告されている. 本稿で行った因果関係を含む文の抽出の実際の処理として, まず, テキストマイニングの対象とした文から手がかり表現で因果関係を含む文の候補を抽出した. 次に, この因果関係を含む文の候補の一部から学習データ及び評価データを作成した. そして, 素性でベクトル化した学習データでサポートベクターマシン(以下, SVM[*3]) を学習させ判別モデルを作成した. 判別モデルの評価は, 素性でベクトル化した評価データで行った. 最後に, 因果関係を含む文の候補を素性でベクトル化し, 判別モデルで因果関係を含む文を抽出した. なお, SVM のカーネルは線形を用い, ソフトマージン SVM でのペナルティの重要度を表すコストパラメータ C は 1.0 とした. 手がかり表現と素性及び学習データと評価データについては後述する. 図表 1 は因果関係文の抽出処理の概要を示す. また, 抽出対象となる因果関係を含む文と抽出対象外である因果関係を含まない文の具体例を図表 2 に示す. ただし, 太字は手がかり表現を示している.

図表 1 因果関係文の抽出処理の概要



図表 2 因果関係を含む文と含まない文の例

因果関係を含む文	シューズ部門では、ランニングブームの継続と、フィッティングの取組みを強化したこと**により、**ランニングシューズの販売が堅調に推移いたしました。
	また、新興国を中心とする旺盛な需要や新しいエネルギー資源の開発など**を背景に、**当社グループの事業環境は好転しております。
因果関係を含まない文	また、通商、独占禁止、特許、消費者、租税、為替管制、環境・リサイクル関連の法規制**を受けております。**
	平成 17 年 7 月 1 日**から、**製造たばこの販売に際しては、これらの規定に従っております。

[*3] 本稿では, python の機械学習のオープンソースライブラリである scikit-learn を用いた.
http://scikit-learn.org/stable/index.html

2.1 手がかり表現によるフィルタリング

手がかり表現とは因果関係文を判定する上で重要な手がかりとなる表現を示す. 例えば,「猛暑日が連続したため、飲料水の売上が伸びた。」という文の「ため、」が手がかり表現となる. 本稿では, 決算短信と有価証券報告書は記載される文が類似していることから, 決算短信の手がかり表現[坂地 2015]を参考に 37 個の手がかり表現を選定した. これらの選定した手がかり表現を含む文が, 因果関係を含む文の候補となる. ただし, 2 文にまたがる因果関係や, 手がかり表現が含まれていない文は対象外とした. 選定した手がかり表現の例を図表 3 に示す.

図表 3 手がかり表現の例

から、	を背景に、	を受けております。
を反映し、	を反映して	に支えられて
によって	により	ためであります。
に伴う	に伴い、	を受け、

2.2 サポートベクターマシンによる因果関係文の抽出

(1) 学習データ及び評価データの作成方法

SVM の学習データ及び評価データは因果関係を含む文の候補から下記の手順 1-1～手順 1-4 の手順で抽出した後, 因果関係を含む場合に正例, 含まない場合に負例とするラベル付与を人手で行った. 文の抽出においては, 精度の高い判別モデルを作成するため, 学習データについてより広範の表現を抽出できるような工夫を行った. 具体的には, 有価証券報告書に記載される文が時期や業種, 項目によって特徴が異なることが考えられるため, これらが均一に抽出されるようにした.

手順1-1: 分析対象の有価証券報告書を, 有価証券報告書の発表日ベースで年ごとに振り分ける.

手順1-2: 手順 1-1 の各年の有価証券報告書から業種ごとにランダムに 3 社を抽出. 計 459(3 社×17 業種×9 年)の有価証券報告書が抽出される. 業種は東京証券取引所が定めた東証 17 業種分類を利用する.

手順1-3: 手順 1-2 で抽出した 459 の有価証券報告書から手がかり表現によるフィルタリングで因果関係を含む文の候補を抽出する.

手順1-4: 手順 1-3 で抽出した各有価証券報告書の文から,「業績等の概要」,「対処すべき課題」,「事業等のリスク」の各項目からランダムに 1 文ずつ抽出. 合計 1377(459×3)文が抽出される. なお, 1 文も抽出されなかった有価証券報告書があった場合は, 既に抽出済みの有価証券報告書以外で, 同年, 同業種内でランダムに取得された有価証券報告書を使用して同じ項目の 1 文をランダムに抽出する.

学習データ及び評価データをそれぞれ 1377 文抽出した後, 人手で正例と負例のラベルを付与した. 正例と負例の誤判定の発生を抑えるために, 少なくとも 3 人の判定が一致するような判定手順で作業を行った. 判定員は金融業務に従事する実務者が担当した. 正例と負例のラベル付与の手順を手順 2-1～手順 2-4 に示した. 学習データ及び評価データとも共通の手順となる. 得られたラベル付きデータの内訳は, 学習データが正例 782 文, 負例 595 文, 評価データが正例 733 文, 負例 644 文となった.

手順2-1: 抽出した学習データ(または評価データ)1377 文を 3 グループ(各 459 文)に分割する.

手順2-2: 判定員計 9 人を各グループに 3 人ずつ割り当てる.

手順2-3: グループ内のそれぞれの文に対して, 判定員 3 人が同じ判定だった場合はその判定を採用し正例または負例のラベルを付与する.

手順2-4: 手順 2-3 でラベルが付与されなかった文に対しては, 別の判定員 2 人が判定を行う. 5 つの判定のうち, 同じ判定が 3 つ以上となった判定を採用しラベルを付与する.

(2) 素性の作成

判別モデルとして SVM を用いる際, 学習, 評価及び学習後の判別処理において, 文の特徴を表す素性が必要となる. 本稿では坂地ら[坂地 2011] を参考に 4 つの素性を利用した. 各素性の概要を図表 4 に示す. ただし, 図表 5＜文節と係り受けの関係＞は係り受け解析を行った文節と係り受けの関係を表しており, 一塊の文字列が文節を, 矢印が係り先を示している. 因果関係を含む文の特徴を適切に捉えるために, 文に含まれる手がかり表現と構文的な素性である助詞のペアが素性として含まれている. SVM に入力される最終的なデータは, 取得した全ての素性を並べて, 文に含まれている素性を 1, 含まれていない素性を 0 としたベクトルとなる. 素性の作成においては, 形態素解析では形態素解析器 MeCab[*4][工藤 2004]を, 構文解析では係り受け解析器 CaboCha[*5][工藤 2002]を用いた.

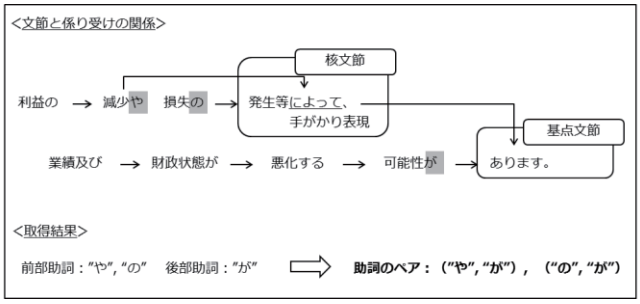
図表 4 素性の概要

素性の名前	概要
助詞のペア	核文節に係る文節に含まれる助詞を前部助詞, 基点文節に係る文節の助詞を後部助詞とし, 前部助詞と後部助詞を合わせた全ての助詞のペア(重複を除く). ただし, 前部助詞が取得できない場合は, 核文節より前の最も近い文節の助詞を前部助詞とする. 存在しない場合は欠損値. また, 後部助詞が取得できない場合は, 核文節より後で基点文節に最も近い助詞を後部助詞とする. 存在しない場合は欠損値. ※核文節は手がかり表現を含む文節. 基点文節は核文節の係り先の文節. ※助詞のペアの取得イメージは図表 5 を参照.
文に含まれる手がかり表現	図表 3 を参照.
形態素ユニグラム	因果関係を含む文の候補を形態素解析器で分解した形態素のうち, 頻度が 2 以上のものを抽出し重複を除いたもの.
形態素バイグラム	因果関係を含む文の候補を形態素解析器で分解し, 隣り合った全ての形態素ペア(重複を除く).

[*4] http://taku910.github.io/mecab/

[*5] https://taku910.github.io/cabocha/

図表 5 助詞のペアの取得イメージ



3. 判別モデルの評価結果

SVM を学習させ作成した判別モデルの評価結果を図表 6 に示す. ただし, 平均／合計の欄は, 精度, 再現率, F 値について, 因果関係を含む文と含まない文の数で加重平均した値を, データ数については合計を示している.

因果関係を含む文の場合も, 含まない文の場合も, 精度, 再現率, F 値 とも 0.8 を超える結果となった. また, 図表 7 に抽出された因果関係を含む文の例を示す.

図表 6 判別モデルの評価結果

	精度	再現率	F 値	データ数
因果関係を含む文	0.85	0.89	0.87	733
因果関係を含まない文	0.87	0.82	0.84	644
平均／合計	0.86	0.86	0.86	1377

図表 7 判別モデルの因果関係文抽出結果例

項目名	因果関係文として抽出された文
業績等の概要	投資活動によるキャッシュ・フローは、船舶の取得による支出などにより、当連結会計年度は 1,455 億 40 百万円のマイナスとなりました。
事業等のリスク	一方、9 月以降には取引先からの返品が発生するため、第 4 四半期の収益が低下いたします。
対処すべき課題	ロール事業では、事務機の構造変化や高耐久化の加速による補修品市場の縮小といった環境変化により販売の低迷が想定されます。

4. 考察

判別モデルの評価は, 因果関係を含む文の場合も, 含まない文の場合も, 精度, 再現率, F 値 とも 0.8 を超え, 良好な判別結果となったが, この要因としていくつかの理由が考えられる. 1 つは, 上手くモデルの学習が機能するような学習データが準備できたことが考えられる. 実際に, 年, 業種, 項目ごとに万遍なく文を取得していることや, 手がかり表現でフィルタリングしたことで正例候補が絞られ, 結果的に正例と負例の割合がおおよそ 6 対 4 となり, 偏りなく広範の表現の学習データが準備できた. その他, 本稿で対象とした有価証券報告書の因果関係を含む

文の構文や単語は似たものが多かったという可能性や，本稿の素性で上手く捉えられるようなシンプルな特徴を持つ文が多かった可能性も考えられる．ここで，シンプルな特徴を持つ文は，因果関係を含むか否かの人による判別が容易であると考え，SVM の性能評価の追加確認を行った．評価データとしては，手順 2-3 において，最初の判定員 3 人の判定が一致した文を判別の容易な文としたデータ，反対に，3 人の判定が一致しなかった文を判別の難しい文としたデータのそれぞれを用いた．その結果を図表 8 に示す．判別の容易な文については，判別性能の評価が高く，一方で，判別の難しい文は判別性能の評価が低い結果となった．判別の難しい文は 5 人の判定員で判別されており，単純にルール化できるような因果関係を含むか否かの基準だけではなく，判定員それぞれの異なる経験と様々な観点から判別されている．他方，SVM では，限られた学習データと決められた素性で学習しているため，ルール化できるような基準は反映できても，個々人の経験や様々な観点までは反映しきれなかったのかもしれない．このようなニュアンスと呼べる要素を加味するには，まず，判別が難しい文の学習データを増やすことや，判別の難しい文に対する因果関係が含まれるか否かの判別基準となる特徴などを，素性として新たに加える等の工夫が必要になると考えられる．

図表 8 判別モデルの追加の評価結果

判別難易	因果関係有無	精度	再現率	F 値	データ数
易	有	0.93	0.93	0.93	590
	無	0.91	0.91	0.91	454
難	有	0.58	0.74	0.65	143
	無	0.75	0.60	0.67	190

5. まとめ

本稿では，有価証券報告書から因果関係を含む文を抽出する判別モデルを提案した．作成したモデルの評価では，因果関係を含む文の場合も，含まない文の場合も，精度，再現率，F 値 とも 0.8 を超え，良好な判別結果となった．この判別モデルにより有価証券報告書独自の投資判断に有益な情報の効率的な抽出が期待できるだろう．

また本稿における SVM による判別モデルについて，判別が難しい文の学習データを増やすことや，判別の難しい文の因果関係が含まれるか否かの判別基準となる特徴などを素性として新たに加える等の工夫をすることで，判別モデルのさらなる性能の向上が見込めるだろう．

参考文献

[五島 2015] 五島圭一，高橋大志，寺野隆雄: ニュースのテキスト情報から株価を予測する，第 29 回人工知能学会全国大会 大会論文集, Vol.29, pp.1-3, (2015)

[北森 2017] 北森詩織，酒井浩之，坂地泰紀: 決算短信 PDF からの業績予測文の抽出，電子情報通信学会論文誌 (D), Vol.J100-D, No.2, pp150-161, (2017)

[工藤 2002] 工藤拓，松本裕治: チャンキングの段階適用による日本語係り受け解析，情報処理学会論文誌, Vol.43, No.6, pp1834-1842, (2002)

[工藤 2004] 工藤拓，山本薫，松本裕治: Conditional Random Fields を用いた日本語形態素解析，情報処理学会研究報告自然言語処理 (NL), Vol.2004, No.47, pp89-96, (2004)

[沖本 2014] 沖本竜義，平澤英司: ニュース指標による株式市場の予測可能性，証券アナリストジャーナル，Vol.52, No.4, pp.67-75, (2014)

[酒井 2015a] 酒井浩之，小林義和，坂地泰紀: 企業の決算短信PDFから抽出した業績要因への極性付与，第 15 回 金融情報学研究会, pp.7-12, (2015)

[酒井 2015b] 酒井浩之，西沢裕子，松並祥吾，坂地泰紀: 企業の決算短信 PDF からの業績要因の抽出，人工知能学会論文誌, Vol.30, No.1, pp.172-182, (2015)

[坂地 2011] 坂地泰紀，増山繁: 新聞記事からの因果関係を含む文の抽出手法，電子情報通信学会論文誌 (D), Vol.J94-D, No.8, pp1496-1506, (2011)

[坂地 2015] 坂地泰紀，酒井浩之，増山繁: 決算短信 PDF からの原因・結果表現の抽出，電子情報通信学会論文誌 (D), Vol.J98-D, No.5, pp811-822, (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

機械学習を利用した看護文献レビューマトリックスシステム

Nursing Literature Review Matrix System based on Machine Learning

山下 貴範 *1 Takanori Yamashita　伊豆倉 理江子 *1 Rieko Izukura　中島 直樹 *1 Naoki Nakashima　廣川 佐千男 *2 Sachio Hirokawa

*1九州大学病院メディカル・インフォメーションセンター
Medical Infomormation Center, Kyushu University Hospital

*2九州大学情報基盤研究開発センター
Research Institute for Information Technology, Kyushu University

The literature review matrix is a method commonly used in writing a summary of literature surveys in the nursing field. In the literature review matrix, the literatures and literary groups obtained by the search are arranged on the vertical axis and the analysis aspects such as purpose, data, method, result and discussion are arranged on the horizontal axis, and features of each object are described in the cell. We applied machine learning to grasp the point of view being performed manually and to extract features of each view. We constructed a system that dynamically generates a review matrix according to the search. The present paper explains the idea, the method and case studies on 215 literatures on “text mining in nursing”.

1. はじめに

多くの論文から研究分野の特徴や傾向の推移を把握するための研究は多くされてきている．IBM 社の Watson は，膨大な医療論文の非構造化データを読み込み，患者のデータと比較して CDSS（医療意思決定支援システム）を目的とした診断や診療行為を推論することは有名である．他にも医療のリアルワールドデータである患者の検査結果数値，画像，テキストなど，あらゆる構造化・非構造化データに対して機械学習を用いた解析が活発になっている．しかし，これらは医師視点の解析が多いのが現状である．一方，看護においては，健常者から患者，病気の早期から末期，新生児からお年寄り，一人の看護師から組織全体と，医師よりも対象の範囲が広く複雑であるため看護を立案・実施するにあたり，医学的な視点だけではなく，心理・社会的・宗教的・スピリチュアルといった様々な視点から対象を捉え，把握することが必要とされる．診療行為はデータ記録されるが，患者等と直接的に関わる時間や看護内容の多様性を考慮すると，患者に対するテキストデータ記録量は医師に比べ多いのが現状である．

近年， 看護研究の領域において，テキストマイニングを用いた研究が増えてきており，今後の展開が注目されている [大野 13, 西川 13]．[大野 13] らは，禁煙外来の会話から質的，量的に分析し，探索的研究を行なっている．[西川 13] らは，国際看護を対象にしてレポートなどの自由記載のデータからテキストマイニングを実施し，国際看護の教育を考察している．

そこで本研究では，「看護＋テキストマイニング」の研究動向を調査するべく，それに関連する論文概要（215 件 1,166 文）を対象にして，文の分類（背景，目的，技術，結果，議論）を元に機械学習を行い，その結果のレビューシステムを構築した．

連絡先: 山下貴範，九州大学病院メディカル・インフォメーションセンター，〒 812-8582 福岡市東区馬出 3-1-1，092-642-5971，t-yama@med.kyushu-u.ac.jp

2. 動機

看護分野では以前よりグランデット・セオリーや現象学，エスノグラフィー等の手法を用いた質的研究が行われており，複雑な人間・社会的環境，生活の文脈や信念，価値観などを全体的にとらえることに対して有用である [グレッグ 16]．これらの研究では，研究者自身が自分の手で自身の主観性・感受性に基づき対象者の内面に入り深く解析し，客観的なモデルを作成するが，結果の客観性の点で未だ論争の一つとなっている [松原 10]．客観性を高めるという点からすれば，テキストマイニングは有用であるし，最近ではテキストマイニングを組み合わせた質的研究もみられる [稲葉 16]．

看護研究におけるテキストマイニングのモチベーションとしては，前述した通り複雑な背景をもつ『人』の抽象的な事象・現象に対して具体性を持たせることに貢献できると考える．また，研究者自身のデータのコード化やカテゴリー化に対して，コンピュータの解析結果を可視化し，照合や検討することで，客観性が高まり，対象がテキストで語る現象に近づく可能性がある．テキストマイニングを用いた看護研究はまだまだ発展途上ではあるが，複雑な看護分野だからこそ対象を明確化することに意味があると考える．

3. 機械学習による文種類の識別

学術論文は文学作品と違い独自の構造に従って書かれる場合が多い．それは，IMRAD あるいは MOVE と呼ばれる [Teufel 10, Hashimoto 16]．IMRAD は Introduction, Method, Result and Discussion の略で，論文はそれらの部分的な文書から構成されることを表している．文献調査のまとめ方として，対象とする論文を縦軸にとり，このような観点を横軸としたレビューマトリックスの形式がある [Judith 12]．文献データベースで文献リストを求め詳細に読みながら，マトリックスとしてまとめることで，文献調査を整理する方法として利用されている．しかし，その作業は人手で行われている．一方，機械学習を利用することで，論文中のそれぞれの文の内容が，IMRAD のどれに合致するかを自動的に推定する研究がある．本論文では，SVM と属性選択による [Sakai 12] の手法を論文概要に適

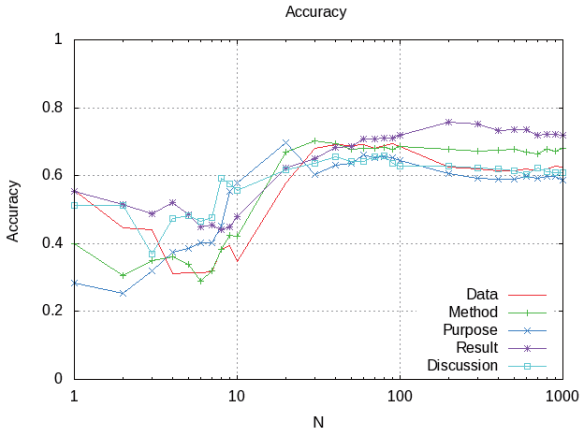


図 1: 属性選択による識別性能向上

用し，それぞれの文に目的，データ，手法，結果，議論という分類ラベルを割り当てる．これにより，自由記述の概要を意味的な半構造化文書として分析することできる．

具体的には，各ラベルに応じて SVM で得られる識別モデルを使い文のラベルを推定する．SVM で得られる識別モデルは，各単語がそのラベル推定にどれだけ有効かを表すスコアのベクトルとして表現される （表 1）．このスコアの上位 N 個の単語（スコアは正）と，下位 N 個数の単語 (スコアは負) の合計 2*N 個だけで文をベクトル化して，推定する．図 1 は，N を変化させたときの識別性能 (accuracy) をプロットしたもので，表 2 は最適な N とそのときの accuracy を示している．全ての単語でのベクトル化よりもごく少数の単語で識別性能が向上していることが分かる．

4. 文献レビューマトリックスシステム

本稿では,「看護」と「テキストマイニング」を含む 215 件の論文概要を対象とした検索エンジンを構築した．図 2 は，全論文を求める予約語「z」を検索語として与えたとき検索結果の画面である．検索窓に検索語を入力し，マトリックスの横軸の項目を「h」の行で複数してすることができる．図 2 では，Purpose, Data, Method, Result, Discussion, Author, Keyword の 7 項目を選んでいる．「h」の行の 5 項目は機械学習で識別する文種類であり，残りは，論文概要に本来含まれるメタデータである．分析対象とする項目を「v」の列の中から一つ選ぶことができる．図 2 では，研究者 (author) を選んでいる．その下の部分が検索結果の各対象について，各項目の特徴語の上位 5 個を表示するマトリックスである．一列目には，論文数の降順で研究者名が表示され，各行にその研究者の特徴を IMRAD の文種類とメタデータで示している．例えば，3 列目の Data を見ると 3 行目の「城丸瑞恵」は，他の 4 人とは違い文献情報が分析対象ということが分かる．これは，8 列目の Keyword に計量書史学があることと符号する．

図 3 は，分析対象を組織 (Organization) とした検索結果であり，研究者が 5 列目に表示され，各大学の研究グループや研究概要を把握することができる．

5. まとめと今後の課題

文献レビューマトリックス方式は，看護分野で一般的に利用されている文献調査のまとめ方である．文献レビューマトリックスでは，検索で得られる文献や文献グループを対象として縦軸に並べ，目的，データ，手法，結果，考察などの分析観点を横軸として並べ，各対象の特徴をセルに記載する．本研究は，人手で実施している観点の把握と特徴抽出に機械学習を応用することで，検索に応じて文献レビューマトリックスを動的に生成するシステムを構築した．看護におけるテキストマイニングについての文献 215 件を対象とし，分析事例を紹介した．

観点ごとに生成した機械学習のモデルの識別性能が低いので改良が必要である．また，今回は実験対象が 215 件だけなので規模を拡大し，より一般的なモデルが必要と思われる．一般の文献レビューマトリックスでは，特徴語だけでなく短い文章での記述をする．特徴語を含む文の抽出で，どれだけできるかも今後の課題である．

参考文献

[Judith 12] ジュディス ガラード: LaTeX: 看護研究のための文献レビュー マトリックス方式, 安部 陽子 訳, 医学書院, (2012)

[Hashimoto 16] Hashimoto, K., Soonklang, S., Hirokawa, S: Feature Words of Moves in Scientific Abstracts, *Proceedingsof5thIIAI − AAI*, pp. 144-149, (2016)

[Sakai 12] Sakai, T., Hirokawa, S: Feature Words that Classify Problem Sentence in Scientific Article, *Proc.iiWAS*2012, pp.360-367, (2012)

[Teufel 10] Teufel, S.: The Structure of Scientific Articles: Application to Citation Indexing and Summarization, *Stanford* : *CSLIPublications*, (2010)

[西川 13] 西川まり子, 村田直己, 小櫻愛美: 国際看護学への挑戦をテキストマイニングする, 特集看護研究におけるテキストマイニング（II） 看護研究, Vol. 46, No. 6, pp. 577586, (2013)

[大野 13] 大野佳子, 磯村毅: テキストマイニングによる成功率の高い禁煙プログラムの会話構造の特徴, 特集看護研究におけるテキストマイニング（II） 看護研究, Vol. 46, No. 6, pp. 552567, (2013)

[グレッグ 16] グレッグ 美鈴, 麻原きよみ, 横山美江: よくわかる質的研究の進め方・まとめ方第 2 版看護研究のエキスパートをめざして, 医歯薬出版, 第 2 版, (2016)

[稲葉 16] 稲葉光行, 抱井尚子: 混合研究法としてのグラウンデッドなテキストマイニング・アプローチ, 看護研究, Vol. 49, No. 1, , pp. 25-35, (2016)

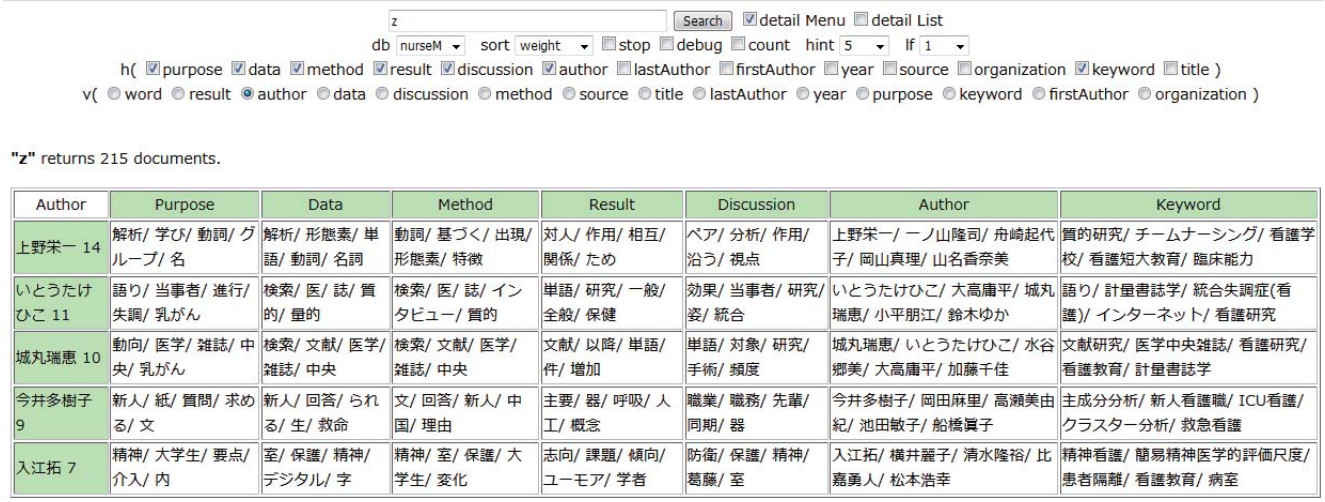
[松原 10] 松原弘子: 質的研究における客観性に関する論考—GTA 法と写真表現との比較を素材にアジア太平洋研究センター年報, (2010-2011)

表 1: 最適属性選択による識別性能

文種類	正の特徴語	負の特徴語
データ	名, 対象, もらう, 平成, カンファレンス	目的, なる, 法, いる, 離職
手法	面接, 用いる, 行う, 解析, アンケート	目的, 結果, 出産, 多い, 月
目的	目的, 明らか, 検討, ため, 個別	れる, 結果, 自分, なる, 手術
結果	結果, 多い, 抽出, 上位, 共起	示唆, 目的, 月, 日常, 考える
考察	示唆, 考える, れる, よう, 多く	目的, 分類, 名, 用いる, オノマトペ

表 2: 文種類ごとの特徴語

文種類	全単語性能	最適性能	最適 N
データ	0.6091	0.6938	90
手法	0.6716	0.7019	30
目的	0.5985	0.6962	20
結果	0.7299	0.7573	200
考察	0.6146	0.6582	80



Author	Purpose	Data	Method	Result	Discussion	Author	Keyword
上野栄一 14	解析/ 学び/ 動詞/ グループ/ 名	解析/ 形態素/ 単語/ 動詞/ 名詞	動詞/ 基づく/ 出現/ 形態素/ 特徴	対人/ 作用/ 相互/ 関係/ ため	ペア/ 分析/ 作用/ 沿う/ 視点	上野栄一/ 一ノ山隆司/ 舟崎起代子/ 岡山真理/ 山名香奈美	質的研究/ チームナーシング/ 看護学校/ 看護短大教育/ 臨床能力
いとうたけひこ 11	語り/ 当事者/ 進行/ 失調/ 乳がん	検索/ 医/ 誌/ 質的/ 量的	検索/ 医/ 誌/ インタビュー/ 質的	単語/ 研究/ 一般/ 全般/ 保健	効果/ 当事者/ 研究/ 姿/ 統合	いとうたけひこ/ 大高庸平/ 城丸瑞恵/ 小平朋江/ 鈴木ゆか	語り/ 計量書誌学/ 統合失調症(看護)/ インターネット/ 看護研究
城丸瑞恵 10	動向/ 医学/ 雑誌/ 中央/ 乳がん	検索/ 文献/ 医学/ 雑誌/ 中央	検索/ 文献/ 医学/ 雑誌/ 中央	文献/ 以降/ 単語/ 件/ 増加	単語/ 対象/ 研究/ 手術/ 頻度	城丸瑞恵/ いとうたけひこ/ 水谷郷美/ 大高庸平/ 加藤千佳	文献研究/ 医学中央雑誌/ 看護研究/ 看護教育/ 計量書誌学
今井多樹子 9	新人/ 紙/ 質問/ 求める/ 文	新人/ 回答/ られる/ 生/ 救命	文/ 回答/ 新人/ 中国/ 理由	主要/ 器/ 呼吸/ 人工/ 概念	職業/ 職務/ 先輩/ 同期/ 器	今井多樹子/ 岡田麻里/ 高瀬美由紀/ 池田敏子/ 船橋眞子	主成分分析/ 新人看護職/ ICU看護/ クラスター分析/ 救急看護
入江拓 7	精神/ 大学生/ 要点/ 介入/ 内	室/ 保護/ 精神/ デジタル/ 字	精神/ 室/ 保護/ 大学生/ 変化	志向/ 課題/ 傾向/ ユーモア/ 学者	防衛/ 保護/ 精神/ 葛藤/ 室	入江拓/ 横井麗子/ 清水隆裕/ 比嘉勇人/ 松本浩幸	精神看護/ 簡易精神医学的評価尺度/ 患者隔離/ 看護教育/ 病室

図 2: 文献レビューマトリックスシステム

Organization	Purpose	Data	Method	Author	Keyword
聖隷クリストファー大学 10	精神/ 失調/ 統合/ 変化/ 症	保護/ 受講/ 体験/ 精神/ 企画	精神/ 保護/ イメージ/ 大学生/ 変化	入江拓/ 小平朋江/ 横井麗子/ 清水隆裕/ 比嘉勇人	精神看護/ イメージ(知覚)/ 教育手法/ 統合失調症/ 簡易精神医学的評価尺度
関西福祉大学 6	発言/ もつ/ 母親/ 育児/ 内容	ベネッセコーポレーション/ 株式会社/ 発言/ 母親/ ママ	発言/ 母親/ ベネッセコーポレーション/ ママ/ 株式会社	井田歩美/ 猪下光/ 谷川和昭/ 趙敏延/ 橋本裕	ソーシャルメディア/ 母/ 育児/ 不安(予防)/ 母子保健
昭和大学 6	動向/ 点眼/ 動作/ 容器/ わが国	中央/ 医学/ 検索/ 雑誌/ 質的	中央/ 医学/ 雑誌/ 検索/ 質的	城丸瑞恵/ いとうたけひこ/ 加藤千佳/ 鈴木ゆか/ 小川鑛一	看護研究/ 医学中央雑誌/ 事故/ 人工呼吸器/ 計量書誌学
県立広島大学 5	必要/ 新人/ コミュニケーション/ スキル/ 表	新人/ 救命/ 当該/ センター/ 知識	知識/ 新人/ 回答/ 基礎/ レベル	今井多樹子/ 池田敏子/ 宮腰由紀子/ 高瀬美由紀/ 伊藤良子	ICU看護/ 新人看護職/ 主成分分析/ 病院救急医療サービス/ CCU
兵庫県立大学 4	感染/ 清掃/ 病室/ システム/ テーマ	感染/ 清掃/ 演題/ 会員/ 学会	感染/ 清掃/ 度数/ 病室/ 物品	高島真美/ 石垣恭子/ 周藤俊治/ 神崎初美/ 東ますみ	ドレナージ/ 経口投与/ 挿管法/ 療養病床/ 臨床検査技師

図 3: 分析対象を組織 (Organization) とした検索結果

Oral presentation | Organized Session | [Organized Session] OS-24

[2O2-OS-24a] [Organized Session] OS-24

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room O (2F Kaimon)

[2O2-OS-24a-01] Analysis using a Triangular Pyramid Game Simulating Restaurant Service

○Tomomi Nonaka[1], Nobutada Fujii[2], Takeshi Shimmura[1] （1. Ritsumeikan University, 2. Kobe University）

1:20 PM - 1:40 PM

[2O2-OS-24a-02] Finding cultural differences on food using dish descriptions from social media: a comparative study of culinary perceptions and preferences between Japanese and Americans

○Yukie Nakamoto[1], Masanao Ochi[1], Takeshi Sakaki[2], Junichiro Mori[3], Ichiro Sakata[1] （1. Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 2. Hotto Link Inc., 3. Graduate School of Information Science and Technology, The University of Tokyo）

1:40 PM - 2:00 PM

[2O2-OS-24a-03] An analysis of sizzle words based on co-occurrence networks

Shingo Kasai[1], ○Fumiaki Saitoh[1], Syohei Ishizu[1] （1. Aoyama Gakuin University）

2:00 PM - 2:20 PM

[2O2-OS-24a-04] Analysis of waiters know-how using a game

○Fumihide Nishimura[1], Emi Kinoshita[1], Tomomi Nonaka[1], Hajime Mizuyama[1] （1. aoyamagakuindaigaku）

2:20 PM - 2:40 PM

[2O2-OS-24a-05] Approach to improvement of restaurant management using machine language

○Tomohiro Hoshino[1], Takashi Tanizaki[1], Takeshi Shimmura[2], Takeshi Takenaka[3] （1. Graduate School of Kindai University, 2. Ritsumeikan University, 3. National Institute of Advanced Industrial Science and Technology）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

レストランサービスを模擬した三角錐ゲームを用いた分析

-サービスロボットの生産性に関する一考察-

Analysis using a Triangular Pyramid Game Simulating Restaurant Service

- A Study on Productivity of Service Robot -

野中　朋美[*1], 藤井　信忠[*2], 新村　猛[*1]
Nobutada Fujii, Tomomi Nonaka, Takeshi Shimmura

[*1] 立命館大学食マネジメント学部, [*2] 神戸大学大学院システム情報学研究科
College of Gastronomy Management, Ritsumeikan University, Graduate School of System Informatics, Kobe University

This paper discusses productivity of service robot in interpersonal service in restaurant service as a case. A subject experiment using a game that creates a triangular pyramid by paper processing simulating labor intensive work in restaurant service is performed. From the results, we analyze the behavior, productivity and quality of subjects performing autonomously diverse tasks and consider productivity of service robots.

1. はじめに

サービス産業の人手不足や労働生産性向上に向けて，サービス分野におけるロボット活用が期待されている．サービス現場の多くは労働集約的であり，省人化や効率化は喫緊の課題である．経済産業省は，ロボット分野の導入シナリオとして特定の作業を行う単機能ロボット，特定の人に自らの制御で特定の作業を行うロボット，人と周囲状況を判断して自律的に多様な作業を行うロボットと段階的な開発・普及を描いている．すでに，単機能ロボットは旅館・ホテルやレストランをはじめサービス現場への導入が進んできており[Qing-xiao 10]，生産性向上が図られている．そこで，本稿では，レストランサービスを事例に，単機能ロボット以降の2タイプのロボットに着目し，対人サービスにおけるサービスロボットの生産性を考察する．レストランサービスにおける労働集約的な作業を模擬した，紙工作で三角錐を作成するゲームを用いて被験者実験を行い，その結果から，自律的に多様な作業を行う被験者の振る舞いや生産性，品質を分析し，サービスロボットの生産性を考察する．

2. レストランサービスにおける生産性

レストランサービスは，セントラルキッチンやクックチル方式，事前調理などが導入されている場合もあるが，基本的には顧客の注文を受けてから調理を行う受注生産であり，調理・配膳と喫食において生産と消費が同時に発生する「同時性」や，消費者側と提供者側双方の「異質性」などサービスの特徴的な性質を有している．一般的に，製造業における工場生産では，ある機械が生産・加工する材（部品や製品）はその機械の性能によって一定範囲の品質に収まるが，レストランサービスにおいてはこの同時性や異質性ゆえに，知覚品質が異なる可能性がある．また，その日の需要の大小によって労働投入量に対してこなすべき仕事量の変動が大きい状況が起こりやすいため，従業員は需要変動に対して適応的に対処することが求められており，事前の準備や段取り作業を効率的に行うことが重要である．

労働生産性は，付加価値を労働投入量で除した値で計算され，労働生産性向上は，分子の付加価値を高めることと，分母の労働投入量を減らすことの両輪である．レストランサービスでは，サービスの異質性や労働集約的な作業が多いことにより，分子・分母双方のアプローチに対して難しさがあるといえる．

連絡先：野中朋美，立命館大学食マネジメント学部，滋賀県草津市野路東 1-1-1，077-561-5053，nonaka@fc.ritsumei.ac.jp

2.1 従業員満足とサービス品質・生産性

サービス研究の分野では，古くから着目されてきた顧客満足（CS）に加え，従業員満足（ES）の重要性が指摘されてきている．ES の増加は，従業員のパフォーマンスを向上させより良いサービス提供につながることによって CS を高める cycle of success モデル[Schlesinger 91]や，従業員と顧客の関係を satisfaction mirror と表し，相互に影響し合うとする「サービス・プロフィット・チェーン」[Heskett 94]が提案されている．特に従業員と顧客が直接対面する対人型のサービス現場においては，従業員の状態や気持ちがサービス品質に与える影響は大きいと考えられる．

3. サービス提供プロセスにおける従業員メンタルモデル

2015 年 5 月から 8 月に国内和食レストランチェーンの調理，接客および配膳・洗い場それぞれの職種の従業員合計 21 名を対象に従業員ヒアリング調査を行った．結果，従業員が行う作業は複数の工程の組み合わせおよび繰り返しで構成されていること．また，各従業員は厳密かつ詳細な作業スケジュールを与えられるわけではないが，事前に店長から与えられる需要予測やシフト計画（スタッフ構成）の情報を確認し，自身の頭の中で想定需要に対して準備や段取りの計画を立てていることが分かった．特に，前もって作業を見積もることができる予約客に対する準備や段取りの計画を，より詳しく立てていた．さらに，当日仕事が始まってからも店舗内の受注状況や作業場全体の仕事の進捗，あるいは顧客の様子を随時観察し，仕事の優先順位や作業時間を調整していることが確認された．

ここで，サービス現場で働く従業員を生産設備に倣えると，「インプット」や「アウトプット」，「センシング」を柔軟性および同時性をもって行っていると捉えられる．ここでいうインプットおよびアウトプットは，生産に必要な材料と生み出される生産物である．この時，従業員は環境を通じて受動的あるいは能動的にセンシングする．受動的なセンシングは，与えられる注文データや顧客からの呼び出しなどである．一方能動的なセンシングは，客席の顧客の観察や，来店状況や天候などから需要変動を感じ事前に頭の中に組み立てた自身の仕事の予定や段取りの計画すなわち初期スケジュールをリアルタイムに更新，修正しながら

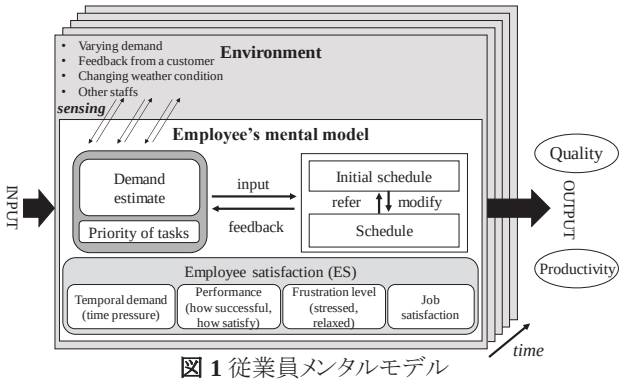


図 1 従業員メンタルモデル

仕事を行っていると言い換えられる. そこで, オーダーの変更や環境の変化をセンシングし, 環境情報を自身の中に取り込み更新することで柔軟に適応し, 結果として生産性が向上する. この際, 生産性向上や自身の計画をやりきることによる達成感を通じ ES の向上, あるいは, 計画通りに仕事が進まないことや想定需要との乖離によって焦りを感じることにより生産性や品質に影響するという仮説を構築する(図 1). 従業員メンタルモデル内の ES を構成する要素は, 左から, 余裕や焦りなどのタイムプレッシャー, 達成感や満足などの作業成績, 不安や落胆などのフラストレーション[Hart 98]に, 楽しさや面白さなどの職務満足を加えた 4 要素で構成される.

ここで, 本モデルをサービスロボットに置き換えて考えると, ロボットはあらかじめ与えられた初期スケジュールを, 環境情報のセンシングにより必要時に適宜更新し作業を行う. 受動センシングによる顧客からの要望に対しては, 対話的に対応し, スケジューリングにおいては重要度に応じて特急ジョブとして扱う. 顧客の様子を観察し, 事前期待を推定しながら優先度を設定することや, サービス提供後の顧客の反応をフィードバックするサイクルを早くまわすことで, 顧客とインタラクションしながら CS 向上や生産性向上を志向することができる. このとき, 知識獲得モデルやエキスパートシステムを利用する. 本仮説モデルを検証のために, サービスプロセスを簡素化したゲームアプローチを用いる.

4. レストランサービスを模擬した三角錐ゲーム

プレイヤは, 到着する注文に対して, はさみとセロテープを用いて紙で三角錐を作成し納品する手作業を, 高得点獲得を目指すゲーム形式で行う. 三角錐は料理に見立てられ, 受注前に作業を始めることや段取りが許される. プレイヤには,「より早く, 出来立て, 品質の良い三角錐がお客さんに喜ばれます」と説明し, 高得点を獲得するためには, 提供 LT, 劣化度や品質を考慮して生産スケジュールを工夫する必要があることを認識させる. このとき, プレイヤが需要変動に対して適応的に対処する振る舞いを分析し, 周囲の情報を把握しながらの自律的な作業スケジュール更新と, 生産性・品質への影響を分析する. また, 実験条件として, 精度が異なる需要予測情報をプレイヤに与える. 与えられる需要予測情報の精度の高低に応じて, 想定需要と実需要との差が変わり, 計画修正や変更の程度が異なる. このときの生産性や品質, ES を分析する.

注文に対する対応として, 1 個ずつ生産する逐次生産, ロット生産, 注文到着前に作り始める, 注文到着前に生産を完了し完成在庫を保持する戦略がある. 生産は, 単数, あるいは与えられる最大ロット数の制約下でロット生産が許される. 紙数枚をまとめて作業することで大幅な時間短縮を実現できるが, ずれやはみ出しなどの品質に影響する. プレイヤは, 注文量と注文到着間隔を予想しながら, 戦略を考慮して三角錐を作成する.

4.1 被験者実験

大学生 18 名を対象に, 三角錐ゲームを実施する. 需要予測情報精度の異なる 3 条件に対して, 各 10 分行う. 1 試行あたりの注文数は 9, 1 注文あたりの注文個数は平均 2, 標準偏差 1.5 の正規分布で与える. 作業に慣れることによる習熟効果の影響が出ないよう, 注文データセット, 試行順序, および需要予測情報に関する条件を実験計画法により設定する. 評価は, 納品数, 品質加点, 劣化減点, 仕掛品に対する廃棄ロスペナルティで構成されるゲーム得点に加え, 生産性, 品質, ES および CS の視点から評価する. ES に関するアンケート調査を行う.

結果, 品質得点の平均は(精度高)が 2.46 で最も得点が高かった. 提供 LT の平均は(情報なし)が最も早く 126.6 であり, 次いで 142.5(精度高), 158.8(精度低)であった. ただし(情報なし)は未納入注文個数が最も多いため, それらが提供 LT の計算に含まれていない影響により提供 LT の値が小さくなっている可能性がある. 従来提供 LT と提供品質は, 標準作業時間前後の閾値まではある程度の品質を確保できるが, 生産能力を超えた注文に対して急いで生産する場合には, 生産性と品質はトレードオフの関係にあると考えられる. 被験者は頭の中で想定需要を考えながらロット数や作業着手のタイミングを計画する. その際に, 精度の高い需要予測情報が有用に役立ち, 品質得点の高さと提供 LT の早さにつながったと考えられる.

ES と生産性や品質の関係においては, 提供 LT と「焦り」では, (情報なし)に中程度の相関がみられた. これは, 需要予測情報が無いため想定外あるいは想定以上の注文が入る状況が起こりやすく, 提供 LT が長くなっていると認識することで焦りが生まれていることが考えられる. 一方, (精度高)においては弱い相関がみられた. (精度高)の状況下では, 想定需要との乖離は起こりにくいため, 需要は分かっている上で余裕が小さい計画を立て間に合わせようと急ぐ焦りであり, 焦りを感じる要因が(情報なし)と(精度高)では異なっている可能性がある.

5. おわりに

本研究では, レストランサービスにおける労働集約的な作業を模擬した紙工作で三角錐を作成するゲームを用いて, 自律的に多様な作業を行う被験者の振る舞いや生産性, 品質を分析し, サービスロボットの生産性を考察した. 多様な条件設定における分析を進めることが今後の課題である.

謝辞

本研究の一部は JSPS 科研費 基盤研究(B)17H01827 の助成を受けたものです.

参考文献

[Hart 98] Hart, S. G. and Staveland, L. E., Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, P. A. Hancock and N. Meshkati (eds.), Human Mental Workload, North-Holland, pp.139-183 (1998).

[Heskett 94] Heskett, J.L., Jones, T.O., Loveman, G.W., Sasser Jr., W.E. and Schlesinger, L.A. : Putting the Service-Profit Chain to Work, Harv. Bus. Rev., Vol. 72, pp. 164-174 (1994).

[Qing-xiao 10] Qing-xiao, Y., Can, Y., Zhuang, F., and Yan-zheng, Z., Research of the Localization of Restaurant Service Robot, Int. J. of Advanced Robotics Systems, Vol.7, No.3, pp.227-238 (2010).

[Schlesinger 91] Schlesinger, L.A., Heskett, J.L., Breaking the Cycle of Failure in Services, Sloan Management Review, Vol. 32, No.3, pp.17-28 (1991).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

食の描写表現の相違を用いた言語圏ごとの料理の感じ方の比較手法

Finding cultural differences on food using dish descriptions from social media:
a comparative study of culinary perceptions and preferences between Japanese and Americans

中元 雪絵 *1 Yukie Nakamoto　大知 正直 *1 Masanao Ochi　榊 剛史 *2 Takeshi Sakaki　森 純一郎 *3 Junichiro Mori　坂田 一郎 *1 Ichiro Sakata

*1東京大学工学系研究科技術経営戦略学専攻
Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo

*2株式会社ホットリンク
Hotto Link Inc.

*3東京大学情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

With the progress of globalization and the overseas expansion of the food industry, the importance of understanding regional differences in food culture is increasing. However, the traditional questionnaire-based study is costly in time and money. In this study, a new method for detecting cultural differences using social media is proposed with an experimental analysis on Japanese and American attitudes toward beverages. We gathered descriptive expressions (adjectives) about beverages with both languages from Twitter, and classified them into thirty meaning-based groups. We compared the co-occurrence frequencies of these groups between Japanese and Americans. The results reveal cross-cultural differences in impression and value standards on beverages. For example, coffee is generally recognized as 'bitter' and 'sour' in Japan, while it frequently appears in 'boring' contexts in the U.S. The proposed method enables a fast specific investigation on regional food culture, which can be applied to the localization strategy of the food industry.

1. 研究背景

グローバル化に伴い, 食産業でも海外進出が増加しているが, 進出先の地域社会へのローカライゼーション戦略を打ち出し成功する事例が多く見られるようになった. ローカライゼーション戦略の肝となるのは, 地域ごとの文化的差異を理解することである. これまで文化に関する研究は, アンケート調査がほとんどであったが, 近年生活に密着した情報発信手段であるソーシャルメディアのデータに注目が集まっている. Silva らは, Foursquare というソーシャルメディアの飲食行動に関するデータを用いて, 世界の国と地域に対して文化境界を定める手法を提案した. Sajadmanesh らは, レシピデータを用い, 味や材料からみた各国食習慣の距離を算出した [Sajadmanesh 2017]. これらの研究では文化の距離を世界規模で議論することに成功している一方で, 各文化の間に具体的にどのような差異があるのか検出することは難しい.

2. 目的

本研究では, ソーシャルメディアを用いて地域ごとの対象の感じ方の違いを明確化することを課題とし, 言語的表現の違いを分析することでこれら差異を得ることが可能であるかを検証する. 具体的には, 描写表現に着目し, 描写表現の共起頻度の違いを用いて文化間差異を自動的に比較する手法を提案する. 提案手法を用いて Twitter の食に関する英日言語の感想データを比較することにより, 文化間の感じ方の差異を検出し, 提案手法の有効性を考察する.

連絡先: 中元 雪絵，東京大学工学系研究科技術経営戦略学専攻，東京都文京区本郷 7-3-1，snopict@gmail.com

3. 提案手法

3.1 着想と概要

ソーシャルメディアは私たちの生活に密着したツールである. 利用者は気軽に自分の意見や感想を投稿し, 日々膨大な量の情報がインターネット上で共有される. 気軽に主観的な考えを共有できることがソーシャルメディアの特徴であり, だからこそ多種多様な人の感性や価値観がソーシャルメディアの投稿に如実に現れる.



図 1: 英語と日本語のツイートの例

図 1 は Twitter から得られる麺についての英語と日本語のツイートの例である. 英語では麺に対して「hot」や「spicy」が共起し, 日本語では「真冬の夜中」という場面設定や「旨い」が共起する. ここから, 英語ツイートの投稿者は麺類に対して辛くてスパイシーなものを求めており, 日本語ツイートの投稿者はラーメンに対して「真冬の夜中」という場面設定や, 旨味を想起することが推測できる. もしも, さらに多くの投稿を分析した時に, 言語ごとに同じ傾向があったとしたらどうだろうか. それは英語圏と日本語圏の麺類に対する考え方や感じ方に違いがあるとは言えないだろうか.

このように, ソーシャルメディア上で人々が無意識に用いている描写表現は, 文化圏における感性や価値観を反映している可能性がある. 描写表現の使い方の差異を定量的に評価するこ

とで, 異なる文化圏において, 対象に対して一般にどのような認識があるのか, また, どのような文脈に対象が登場するのかといった違いを明確化することができる.

しかし, ここで問題となるのが, 異なる言語の描写表現をどのように統一尺度で測るかということである. 最も簡単なのは辞書を用いて翻訳することだが, 全ての描写表現を言語間で一対一対応させることは難しい. そこで本研究では, 言葉の持つ意味や概念に着目し, 近い意味を持つ言葉を言語に関わらず全て同じコンセプトとして扱うことが有効であると考えた. コンセプトを定義する際は, 表現同士の意味の近さでエッジを定義したネットワークを構築し, ネットワーク上で近い位置に集まった表現群を一つのコンセプトとみなす. これにより複数の言語の表現を各コンセプトに分類し, 言語を超えた尺度で扱うことが可能となる.

それぞれの言語圏で, 分析したい対象に対し, どのようなコンセプトを多く用いているか, その頻度を比較することで異なる言語圏における文化的差異の検出を実現できると考えた. 提案手法の概要を図 2 に示す. 以降で詳細を説明する.

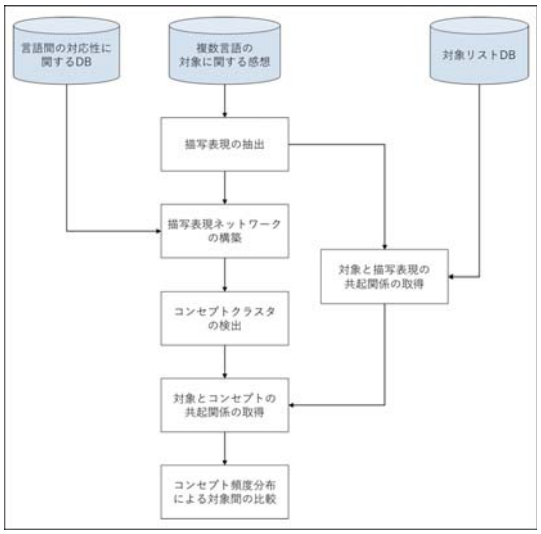


図 2: 提案手法の概要

3.2 描写表現ネットワークの構築

入力として, 分析対象のリスト, これら対象に関する複数言語における感想のデータ, 言語間の対応性に関するデータベースの 3 つを用いる. あらかじめ分析したい対象を決定し, Twitter などのマイクロブログやレビューサイトを用い, 分析対象への感想データを取得しておく.

複数言語の感想データから, それぞれの言語について描写表現を抽出する. 本研究では, 形容詞が描写表現に相当すると考え, それぞれの感想データに対して形態素解析を行うことで, 英語と日本語の形容詞群を抽出した. (**描写表現の抽出**)

続いて, 各描写表現をノードとみなし, 近い意味を持つ描写表現同士に重みを付与して繋ぐことで, ネットワークを構築する. このネットワークを描写表現ネットワークと呼ぶことにする. 本研究では, 言語間の対応性に関するデータベースとして, ConceptNet[Speer 2017] を用いて近い関係のあるペアと重みを取得した. ConceptNet は異なる言語間を繋ぐリンクの情報を持つため, 描写表現ネットワークは全言語の描写表現に対し意味の近いものを結んだネットワークとなる. 一例として, 英語と日本語の形容詞群に対して構築した描写表現ネットワークを図 3 に示す. 図は Gephi[*1] を用いて描画した. (**描写表現ネットワークの構築**)

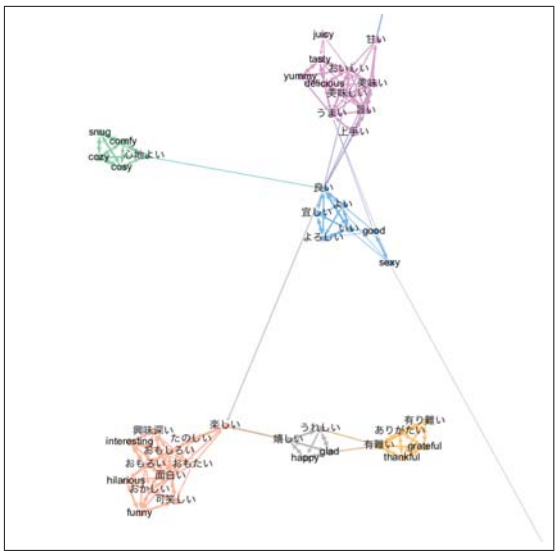


図 3: 描写表現ネットワークの例

3.3 コンセプトクラスタの検出

構築した描写表現ネットワークに対してクラスタリングを行い, 全ての表現を複数の部分集合に分割する. 近い意味をもつ表現の間にエッジを張っているため, それぞれの部分集合には, 似た意味をもつ表現が集まる. また, エッジには異なる言語間を繋ぐエッジが存在するため, 異なる言語の表現で意味が近いもの同士もまた同じ部分集合に属することとなる. クラスタリングの結果として得られる部分集合一つ一つを, 言語を超えたコンセプトを表現する「コンセプトクラスタ」と定義する. (**コンセプトクラスタの検出**)

3.4 コンセプト頻度分布による対象間の比較

ここで, 入力に用いた感想データと分析対象のリストから, 分析対象それぞれに対して, 描写表現 (形容詞) がどのように共起したかを求めておく (**対象と描写表現の共起関係の取得**). これを先で検出したコンセプトクラスタ単位で合計し, 分析対象リストと各コンセプトクラスタの間の頻度分布を計算する. これをコンセプト頻度分布と呼ぶことにする. (**対象とコンセプトの共起関係の取得**)

分析対象リストに属する二つの対象について, それぞれコンセプト頻度分布を求め, 頻度分布の JS 距離を用いて対象間の関係や近さを評価する. また, コンセプト頻度分布の差が大きいコンセプトクラスタを比較することで感じ方の差異を検出する. (**コンセプト頻度分布による対象間の比較**)

4. 実験

4.1 Twitter の食データに対する提案手法の適用

飲み物に関するクエリを英語･日本語に対して設定し, Twitter からそれぞれ感想データを取得した. 実験に用いたクエリを表 1 に示す. 英語と日本語間で, クエリに用いている単語の種類や粒度が異なるが, これは飲み物群の分類の差もまた文化的差異を表していると考え, アメリカと日本を代表するレシピサイトにおけるカテゴリ分類を元にクエリを設定したためである.

*1 https://gephi.org/

英語は allrecipes[*2] の「Drinks Recipes」カテゴリ, 日本語は楽天レシピ [*3] の「飲みもの」カテゴリの分類を参照した.

表 1: 実験に用いた英語・日本語のクエリ

英語	日本語
beer cocktail, cider, cocktail, coffee, eggnog, hot chocolate, juice, lemonade, liqueur, mocktail, mulled wine, punch, sangria, shot, shake, float, slushy, smoothie, tea	コーヒー, ココア, 紅茶, 抹茶, 豆乳, ヨーグルト, はちみつドリンク, チョコレートドリンク, しょうがドリンク, お酢ドリンク, サワードリンク, ソーダドリンク, ミックスジュース, シェイク, セーキ, スムージー, グリーンスムージー, チャイ, ビール, 焼酎, 梅酒, 甘酒, カクテル, モヒート, ジントニック, 卵酒, 健康酒

4.2 提案手法の評価

提案手法の妥当性を評価するため, Twitter から取得した飲み物の感想データを用い, 以下の三つの点から評価を行った.

全体傾向の言語間比較

描写表現から文化的差異を取得することが実際に可能であるか, また, 描写表現として形容詞を使うことが妥当であるかを検証するため, 取得した全ての英語ツイートと日本語ツイートを用い, 英語言語圏の傾向と日本語言語圏の傾向の差異を正しく抽出できているか評価する. 具体的には, 極性分析 (全ツイートの極性平均の比較), 形容詞の種類による分析 (松尾による形容詞の分類 [松尾 2014] に基づく, 各形容詞の種類の出現割合の比較), コンセプト頻度分布による分析 (英語全体, 日本語全体のコンセプト頻度分布において, 差が大きいコンセプトクラスタの検出) の三つの分析を行う.

ネットワーク図による関係性の可視化

本研究では, 分析対象のコンセプト頻度分布間の JS 距離を求めることで, 対象間の距離とし, 対象間の関係を捉えることを提案した. この手法の妥当性を確認するため, JS 距離で対象間の距離を定めたネットワーク図を作成し, 実際に対象同士の関係を捉えることが可能であるかを検証する.

同一対象への感じ方の言語間比較

本研究では, 対象のコンセプト頻度分布の差を具体的に分析することで, 異なる言語圏における感じ方の差異を得ることを提案した. この手法の妥当性を検証するため, 分析に用いたクエリのうち, 日本語と英語で同一対象を扱っているものについてコンセプト頻度分布の具体的な比較を行い, 言語圏における感じ方の差異が抽出できていることを確認する.

5. 結果

5.1 全体傾向の言語間比較

取得した全ての英語ツイートと日本語ツイートを用い, 描写表現を用いて英語圏と日本語圏の全体傾向の差異を正しく抽出できるか評価した.

ツイートの極性分析の結果, 英語の極性平均は 0.266, 日本語は 0.0137 となった. また, 形容詞の種類による分析を行った結果, 英語では対象の属性や性質を描写する形容詞が多く, 日本語では自身の感情や評価に関する形容詞が多く用いられるという結果が得られた (表 2). ここから, 英語圏でポジティブな形容詞を用いて料理を描写する投稿が多い傾向に対し, 日本語圏でネガティブな評価表現で不満を言う投稿が多いという示唆を得た. これは, 日本語圏で自分に関する言及が多く, 英語圏で食事の状況に関する投稿が多いという先行研究の結果と一致する [Vidal 2015, Acar 2013].

さらに, 英語全体と日本語全体のコンセプト頻度分布を比較することで, 色への描写が英語圏にて多いことを検出した. これはアメリカでカラフルな食品が多い事実と合致する.

以上より, 描写表現を用いて食文化に対する知見を得ることが可能であり, 描写表現, 中でも形容詞から文化的差異を検出する有効性を示している.

表 2: 形容詞の種類ごとの出現割合（%）

		英語	日本語
五感表現	味覚	1.25	6.16
	嗅覚	0	0.21
	触覚	17.75	2.95
	視覚	4.95	0.37
	聴覚	0	0
五感を限定しない表現	直喩	0.34	1.21
	擬人化	2.23	4.87
	性質	**23.83**	10.39
	状況	5.39	5.79
評価表現	一般評価	19.44	**45.26**
	感情	3.02	**9.52**
その他		21.8	13.28

5.2 ネットワーク図による関係性の可視化

二つの対象について, コンセプト頻度分布の JS 距離を用いて距離の近さを定めた. この値を用いて英語のクエリ集合, 日本語のクエリ集合それぞれについてネットワーク図を描画し, 対象間の関係を正しく評価できていることを確認した. 得られたネットワーク図を図 4 に示す.

ネットワークにおいて, 英語では juice, lemonade, tea など果物が用いられる飲み物が近い位置に存在し, 日本語ではココア, 甘酒, はちみつドリンク, 紅茶と温かい飲み物が近い位置に存在する. 全ての飲み物の関係を正しく評価できている訳ではないが, 私たちの感性に合致する結果が見えていることから, 感性の評価にあたり, コンセプト頻度分布の JS 距離を用いることは妥当であると考える.

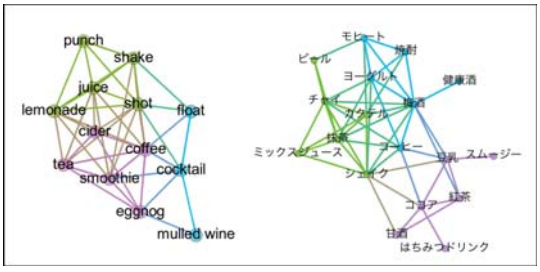


図 4: ネットワーク図 (左：英語, 右：日本語)

*2 http://allrecipes.com/
*3 https://recipe.rakuten.co.jp/

5.3 同一対象への感じ方の言語間比較

クエリのうち, 同一対象を扱っている 4 つのペア (コーヒー, 紅茶, スムージー, ジュース) について差の大きいコンセプト頻度分布を比較し, 感じ方の差異を検出できるか検証した. ここではコーヒー対する分析の結果を取り上げて説明する. 図 5 は coffee とコーヒーの頻度分布の差を表し, 値が正のクラスタは英語で共起が多く, 負のクラスタは日本語で共起が多いことを示す. この図において, 英語全体と日本語全体における差よりもさらに大きい差を持つ (左側の棒グラフが右側の棒グラフより突出している) コンセプトクラスタの内容を取り上げて議論を行った. これは, あるコンセプトクラスタで, ソーシャルメディアの用法による違いなどの全体傾向を排除してなお差があるということは, その対象への認識自体に差がある可能性が高いと考えたためである.

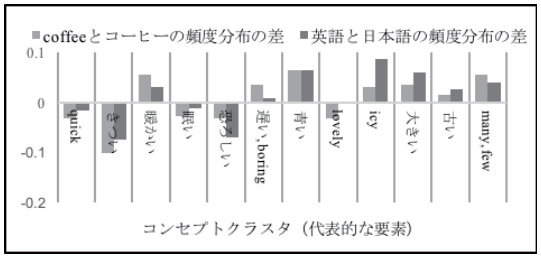


図 5: コーヒーに対するコンセプト頻度分布の差

英語ではコンセプト《暖かい, warm》や, コンセプト《遅い, boring》が日本語に比べて多く共起した. アイスコーヒーの起源は日本であり, 英語圏では冷たいコーヒーを飲む習慣があまりない背景から, コーヒーに対する「熱い」という印象がアメリカの方が強い可能性がある. また, コンセプト《遅い, boring》の共起傾向から, アメリカ人にとってコーヒーが「boring」な文脈に存在することが多く,「exciting」なコーヒーに対する欲求がある可能性が示唆される. 実際に,「boring」でないコーヒーを作る方法を紹介する記事は多く存在する.

日本語では, コンセプト《lovely, 可愛い》が多く共起した. これは日本でラテアートが独自の変化を遂げたためだと思われる. アメリカにもラテアートは存在するが, 日本のようにキャラクターを描く文化は一般的ではなく, コーヒーやラテアートに対して「可愛い」という一般認識はないと予測できる.

他のペアに対しても, 同じように文化的な感じ方や捉え方の差異を検出することが可能であった. 描写表現を分析することで得られる示唆には次のものがあげられる.

一般認識
一般にその対象が文化圏においてどのような印象で認識されているか. 日本語でのスムージーに対する不味いという印象や, コーヒーに対する苦いという印象, 英語でのジュースに対する冷たいという印象がこれにあたる.

評価基準
それぞれの文化圏が対象にどのような評価基準や選択基準を用いているか. 英語でのスムージーに対する濃さへの言及や, ジュースに対する量への言及, 紅茶に対する色への言及がこれにあたる.

文脈
対象がどのような場面で用いられているか. 英語でのコーヒーに対する「boring」という言及や, 日本語でのミックスジュースに対する「楽しい」等の言及がこれにあたる.

独自のジャンル
その文化圏のみで存在する対象のジャンル. 日本語でのラテアートや, 英語でのホットサイダーがこれにあたる.

このように, 提案手法により, アンケート調査や聞き込み調査の労力を使うことなく, 食文化の文化間差異に対する知識を自動的に得ることが可能である. 地域ごとの市場の文化や風習, 嗜好を理解し, 地域に特化した製品開発を行う際の多くの気づきを与えることができる点で, 食産業が海外進出する際のローカライゼーション戦略に対する一助となると考える.

6. 結論

本研究では, ソーシャルメディア上の描写表現を用いて人の感性に基づく文化間比較の行うことの有用性を示した. また, 複数言語の描写表現をコンセプト単位で分類した上で, 共起頻度の分布を用いて比較を行うことで, 異なる言語圏に関する文化的差異を明確化する手法の開発に成功した.

参考文献

[Silva 2014] Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2014, June). You Are What You Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food and Drink Habits in Foursquare. In Proceedings of the 8th AAAI International Conference on Weblogs and Social Media (ICWSM'14).

[Sajadmanesh 2017] Sajadmanesh, S., Jafarzadeh, S., Ossia, S. A., Rabiee, H. R., Haddadi, H., Mejova, Y., Musolesi, M., Cristofaro, E. D., and Stringhini, G. (2017, April). Kissing Cuisines: Exploring Worldwide Culinary Habits on the Web. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1013-1021). International World Wide Web Conferences Steering Committee.

[Speer 2017] Speer, R., Chin, J., and Havasi, C. (2017, February). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In AAAI (pp. 4444-4451).

[松尾 2014] 松尾章子. (2014). 食べ物のおいしさを表すことばに関する研究, 京都府立大学. Ph.D. thesis.

[Vidal 2015] Vidal, L., Ares, G., Machn, L., and Jaeger, S. R. (2015). Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations". Food Quality and Preference, 45, 58-69.

[Acar 2013] Acar, A., and Deguchi, A. (2013). Culture and social media usage: Analysis of Japanese Twitter users. International Journal of Electronic Commerce Studies, 4(1), 21.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

共起語ネットワークに基づいたシズルワードの特性分析

An analysis of sizzle words based on co-occurrence networks

笠井信吾[*1] 齊藤史哲[*1] 石津昌平[*1]
Shingo KASAI Fumiaki SAITOH Syohei ISHIZU

[*1] 青山学院大学・理工学部・経営システム工学科
Department of Industrial and Systems Engineering, Aoyama Gakuin University

In recent years, Sizzle Wards have been attracted attention as it is important in marketing. In this study, we analyze Sizzle Words based on the characteristics of co-occurring word networks, as a novel approach for conventional semantic-based research. We attempt to visualize the properties of sizzle word which could not be grasped by conventional analysis by cluster analysis using co-occurrence word network features such as mediocentricity and order centrality. It can be possible to grasp the contents by detecting the Sizzle word which is used with similar properties, despite the meaning and nuance being completely different.

1. はじめに

近年，食におけるマーケティングにおいてシズル感に注目が集まっている．シズル感とは，米国の経営コンサルタント E.Wheeler によって提唱された概念であり，消費者の五感を刺激することによって食欲や購買意欲を刺激する感覚である．これは商品のデザインや広告，製品コンセプトの策定などの要因と深く結びついており，マーケティング周辺分野を中心に重要視され始めている．

これに対して，シズル感に直結する曖昧な質感のニュアンスを端的かつ的確に表現する用語としてシズルワードがある．これは，日常会話はもとよりツイートやレビューなどといった，製品・サービスやレシピなどを対象とした評価に関わる Web 上への書き込みにおいても広く利用されている．シズルワードはオノマトペなどの感覚的な表現とも密接に結びついており，特に日本語はオノマトペの種類が豊富であることから，これらに関する知識は製品のパッケージや広告においても重要な役割を果たすことが期待されている．

このような背景から，シズルワードは様々な角度から分析がなれており，盛んに研究がなされている．シズルワードに関する既往の研究に目を向けると，シズルワードの用法や印象などに関するアンケート調査をベースとした研究[石橋 15]と，実際の用例であるテキストデータの解析をベースとした研究[渡辺 15, 森田 17, 長谷川 18]に大別できる．前者は実際に複数の被験者を募り，アンケートを通じたシズルワード（あるいはオノマトペなどの感覚的な表現）に関する結果を定量的に解析することで，その意味構造の解明を目指している．対して後者は，用例に基づくテキストデータを解析することによって，それぞれのシズルワードが持つ意味構造の解明を目指している．

これらはいずれも，アプローチや目的が異なるものの，利用者が抱く印象や感覚，利用対象などといったあいまいな言語に対する意味の類似性が単語間の類似性に影響するという前提に基づいた解析やツールの開発である．よって，これまでの研究の多くは，アンケートベース，文書解析ベースいずれにおいても，シズルワードのニュアンスや意味構造を解析に軸足を置いた研究であったといえる．曖昧さを含む言語に対する意味・ニュアンスの解析は最重要項目の一つであるが，関連研究の大多数がこのアプローチに関するもので，たとえば，共起した単語間の関連性やその傾向といった，単語の使われ方の特徴やその可能性に関する議論は充分になされていない．

そこで，我々はこれまでのアプローチとは異なる新たな視点として，各シズルワードの使われ方に着目した．ひとえにシズルワードといっても，利用範囲が多岐にわたるものや限定的なもの，など様々な切り口に基づいて分類できると考えられることから，これらに関する新たな気づきの獲得を目指し，実際の用例の文書データを対象とした解析を行った．分析においては，共起語ネットワークの特性に着目することで，各シズルワードの性質の分析を行っている．各単語をノードとしたネットワークに対して計算した中心性指標を変数とした分析により，意味解析のみでは把握しきれなかった性質について報告する．

連絡先：齊藤史哲，青山学院大学，神奈川県相模原市中央区淵野辺 1-5-10

2. 基礎事項の準備

ここでは，本研究で行う分析において基礎をなす共起語ネットワークおよびその中心性指標について概説する．

2.1 共起語ネットワーク

共起語ネットワークとは，同一文書内において共起した単語をリンクで結んだグラフである．ここでは，各単語がノードであり，リンクが同時出現の関係（共起）を表している．共起語ネットワークを用いた研究に目を向けると，文書データからのキーワード抽出に共起語ネットワークの指標を SVM の特徴量として扱う研究[山本 09]が挙げられる．本研究では，シズルワードの性質の解析においても同様に，共起語ネットワークの特徴量を素性として解析することで新たな知見が得られると考え，共起ネットワークを解析対象とした．

以下本稿中ではノードiとノードjの間のリンクに関する情報を記号a_{ij}で表し，リンクが存在する際は$a_{ij}=1$，リンクが存在しない際は$a_{ij}=0$と表示する．なお，一般的には共起語ネットワークは無向グラフとして扱われるため，$a_{ij}=a_{ji}$となり，a_{ij}をi行j列の要素とした対角行列として隣接行列Aとして表現される．nはネットワーク内に含まれる全ての頂点の数を表している．

2.2 ネットワークの中心性指標

ネットワークやその構成要素（ノード）が持つ特徴を表す指標として，中心性指標が挙げられる．中心性指標を解析することに

よってネットワークの特徴を定量的に把握することが可能になる. ここでは, 本研究にて用いる特徴量について整理する.

≪次数中心性≫

次数中心性は他のノードと繋がるリンク数が多いほど大きな値を持つ指標である. ノードiに対する次数中心性$C_d(i)$はネットワーク内でより多くの他ノードと繋がりを持つノードを高く評価されるように, 次式によって定義されている.

$$C_d(i) = \sum_{j=1}^{n} a_{ij} = \sum_{i=1}^{n} a_{ji} \quad (1)$$

これは, リンクの多さを評価しているので, 共起語ネットワークにおいては, 共起する単語のバリエーションの多さを表している. すなわち, 対象のシズルワードの利用範囲の大きさと解釈することができる.

≪媒介中心性≫

媒介中心性はネットワーク内において他の頂点間の最単経路上に位置する程度を評価する指標である. ノードiに対する媒介中心性$C_b(i)$は次式のように定義されている.

$$C_b(i) = \sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}} \quad (2)$$

ここでは, g_{jk}はノードjとノードkの間に存在する最短経路数を表しており, $g_{jk}(i)$はその内でノードi を介する経路数を表している.

これは, ノードのクラスタを介する媒介点を高く評価する傾向にある事から, レシピなどのジャンルや領域をまたいで利用することができる可能性を持つシズルワードを高く評価することができる指標と考えられる.

≪固有ベクトル中心性≫

固有ベクトル中心性は次数中心性を拡張した中心性指標であり, ネットワーク内において他の(次数)中心性が高いノードと繋がりを持つノードを高く評価する指標である. ノードiに対する固有ベクトル中心性$C_b(i)$は隣接行列Aの固有ベクトルに基づいて, 次式のように定義されている.

$$C_{ev}(i) = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij}\, C_{ev}(j) \quad (3)$$

中心性の高いノードと繋がりが多いノードを高く評価する中心性であることから, 共起語ネットワーク内では, 他の重要性の高いキーワードとの結びつきが強いシズルワードを評価していると解釈できる.

≪情報中心性≫

情報中心性は媒介中心性が最短経路を対象としてノードを評価するのに対して, ノード間の最単経路以外の経路の重みも考慮した中心性指標である. ノードiに対する情報中心性$C_{info}(i)$は経路長に関する情報量I_{ij}を導入することによって, 次式のように定義されている.

$$C_{info}(i) = \frac{n}{\sum_{j=1}^{n} \left(1 / I_{ij} \right)} \quad (4)$$

他の頂点との距離が小さいノードを高く評価する傾向にあることから, 直接的に関わるキーワードが多いシズルワードを評価できる可能性がある.

3. 分析方法について

本研究では, 上記の共起語ネットワークにおける指標およびデータセット内における単語の総出限頻度(TF 値)を全単語に対して算出し, その中からシズルワードのみを解析対象として選択する. これらを特徴量としてクラスタ分析を試みることで, シズルワードが文書表現において果たす役割の性質を調査する. このように各シズルワードのネットワーク指標を特徴量とすることで, 意味内容に捕らわれずに, それら利用傾向や利用方法に関する特徴の把握が可能になる.

これらのデータに対してクラスタ分析を用いることによって, 各シズルワードの類似性について議論を行う. クラスタリング方法としてはウォード法を採用し, 解析結果をデンドログラムとして表示することでシズルワード間の類似性を可視化する. 併せて, 特徴量間の類似性についてもデンドログラムとして表示することでクラスタリングの結果に対して解釈を与える.

4. 実データに基づいた分析

4.1 実験設定

本研究では, NII 情報学研究データリポジトリにおいて楽天株式会社より提供されている“楽天レシピ”のデータセットを解析対象としている. 楽天レシピのデータセット内においては, レシピに対するレビューである“つくったよレポート(つくレポ)”のデータおよびレシピ提供者からそれに対する返答がテキストデータとして提供されている. 本研究では, 隣接行列のサイズを抑えるために提供された約 11 万件のレビューデータの中からランダムに選択した 5000 件を用いている. 形態素解析によって抽出した品詞は名詞, 形容詞, 動詞, 副詞である.

本研究にて解析対象としたシズルワードは, [大橋 15]にて紹介されているものから単語リストを作成し, その中でもデータセット中に存在したものを解析対象のシズルワードとしている. 本研究は, レシピサイトのレビューデータを扱っていることから, シズルワードとレシピ内容の関連性が重要になる. このため, 共起は同一レビュー内に出現した単語を共起関係とみなしている. クラスタリングにおいては, 先述の各中心性をそれぞれ標準化し, 無単位化したものを特徴量として解析を行う.

4.2 解析結果

解析結果の一部を図 1, 図 2 に示す. 図 1 は中心性指標と TF 値を特徴量としたシズルワードのクラスタリング結果をデンドログラムとして表示したものであり, 図 2 はその結果に対する各特徴量の大きさを表している. 図 2 クラスタリング結果に対する各特徴量の大きさをヒートマップとして表示したものである. ここでは, 左の列から順に固有ベクトル中心性, 次数中心性, 媒介中心性, TF 値, 情報中心性の順に表示されており, その上のデンドログラムは特徴量間の類似度を表している.

また, シズルワードは大橋らによって“食感系”, “味覚系”, “情報系”に類別されている[大橋 15]. 表 1 は一部のクラスタとその関係をまとめたものを表している.

4.3 まとめ

クラスタリングの結果より, 例えば, 「甘い」と「たっぷり」は全く意味や用途は異なるものであるが, データセット内では類似した役割を持っている可能性がある. これは次数中心性や媒介中心性が高く, 情報中心性が低いことから幅広い用途で利用できるキーワードであることが見て取れる. 「さっぱり」や「香ばしい」, 「コク」「香ばしい」などは, 固有ベクトル中心性や次数中心性が高いが, 媒介中心性が低いことから, 対象領域が限定的ではあるが, 対象領域内の単語の種類が多い単語と解釈できる.

このように, 共起語ネットワークの中心性指標に着目してデータ解析を行うことで, シズルワードの使われ方やその応用範囲

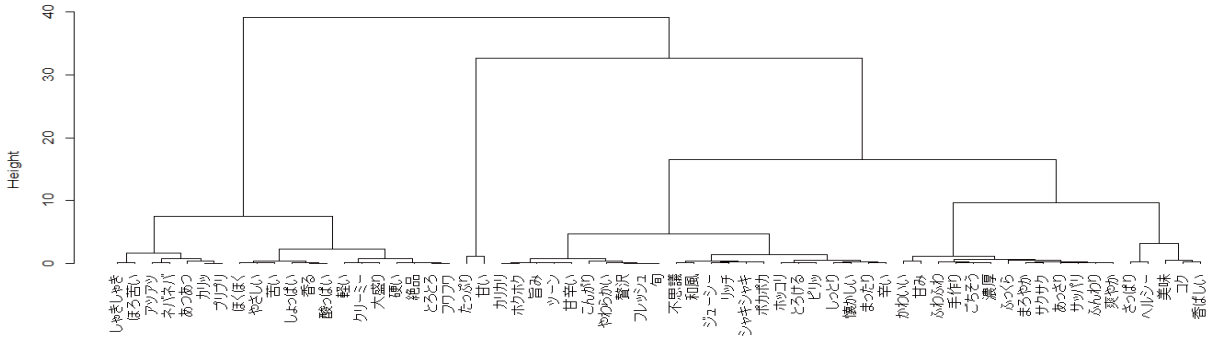


図 1: 共起語ネットワークの特徴量に基づいたシズルワードのクラスタリング結果

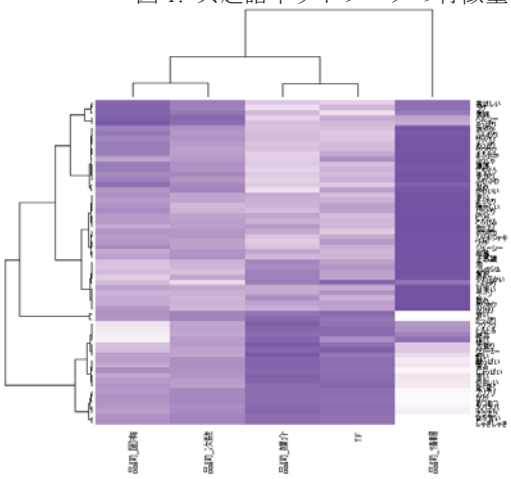


図 2: クラスタの特徴

に対して新たな解釈を与えられると考えている. また, 表 1 のように一部のクラスタでは, “食感系”や“味覚系”に偏りがあるものも確認できたことから, この類別化と用法の間には部分的に関係性が認められる可能性がある.

表 1 クラスタリングの結果とシズルワードの関係性
（食感系を青色, 味覚系を橙色, 情報系を緑色）

クラスタ A	クラスタ B	クラスタ C
しゃきしゃき	さっぱり	かわいい
ほろ苦い	ヘルシー	甘み
アツアツ	美味	ふわふわ
ネバネバ	コク	手作り
あつあつ	香ばしい	ごちそう
カリッ		濃厚
プリプリ		
ほくほく		

5. おわりに

本研究では, 意味や用途の類似性に基づいた分析を前提としている一般的なシズルワードの研究に対して, シズルワードの用法の特徴に着目した新たな分析を試みた. 方法としては, ネットワークの評価指標であるネットワーク中心性に着目し, 共起語ネットワークにおける中心性を特徴量とした分析を行った. 共起語ネットワーク内におけるシズルワードの特徴をクラスタ分析により比較することで, 意味が全く異なるにもかかわらず, シズル感を表現する上で類似した役割を果たす単語を類別化し, 新たな気づきを与えた.

今後は, さらに大規模なデータを解析することやレシピ以外の食品の評価レビューなど, 対象データの範囲を拡大する必要があると考えている. また, 学習等のより高度なデータ解析を通じて, 同様なデータからクラスタを構築する要因に関する知識の抽出を試みる必要がある. これらを通じて, 意味が異なるにもかかわらず類似な共起の特徴を持つクラスタの原因や言語構造等を明らかにしていく予定である.

謝辞

実験用のデータセットをご提供いただきました楽天株式会社様に厚く御礼申し上げます.

参考文献

[石橋 15] 石橋賢, 深瀧創, 宮田一乗: 米菓を対象としたシズル語の印象評価, 人工知能学会論文誌, Vol. 30, No. 1, pp. 229-236, (2015)

[大橋 15] 大橋正房, 光岡祐子, 須藤正彦, 北村奈津子, 米山真知子, 猿渡一兵: sizzle word シズルワードの現在「おいしいを感じる言葉」調査報告, BMFT 出版部, (2015)

[長谷川 18] 長谷川永奈, 小宮香乃, 齊藤史哲, 石津昌平: 食における言語資源に基づいたシズル感に関する因子情報の抽出, 日本感性工学会論文誌, Vol. 17, No. 2, (2018)

[森田 17] 森田真季, 荒牧英治, 灘本明代, 宮部真衣:マイクロブログにおける“おいしさ”情報分類器の適用可能性の検証, 信学技報（LOIS）Vol. 116 No. 488, pp. 229-234, (2017)

[山本 09] 山本優樹, 折原良平: 共起語ネットワーク特徴の言語・文書種非依存性に基づくキーワード抽出と見出し語の予測による性能評価, 人工知能学会論文誌, Vol. 24, No. 3, pp. 303-312, (2009)

[渡辺 15] 渡辺知恵美, 中村聡史: オノマトペロリ:味覚や触感を表すオノマトペによる料理レシピのランキング, 人工知能学会論文誌, Vol. 30, No. 1, pp. 340-352, (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ゲームを用いたホールスタッフのノウハウの分析

Analysis of waiters know-how using a game

西村 文秀 Fumihide Nishimura
木下 枝美 Emi Kinoshita
野中 朋美 Tomomi Nonaka
水山 元 Hajime Mizuyama

青山学院大学
Aoyama Gakuin University

Behavior of clerks working in a restaurant affects customer satisfaction and turnover rate. If the clerks can quickly respond to more customers, the turnover rate will rise and dissatisfaction due to waiting time can be reduced. Their tasks include seat guidance, receiving orders, serving dishes, clearing tables and accounting. Since these tasks need to be prioritized quickly and subjectively on the fly, they are difficult to be automated and often handled by humans. For this reason, the efficiency of the tasks depends on the experience and intuition of the clerks. This paper focuses on three tasks, receiving orders, serving dishes and clearing tables, and analyzes know-how of the clerks on the tasks.

1. はじめに

飲食店業界において，売上を継続的に増加させるための重要な要素として，顧客満足度の向上がある [Nonaka 14] ．顧客満足度とは，製品やサービスのパフォーマンスに対して顧客が期待する水準を満たしているかどうかを示す指標のことである．つまり，売上を継続的に増加させるためには，顧客の期待値を超えるようなサービスを提供する戦略が必要である．顧客満足度が向上すると，その分集客が見込める．したがって回転率の向上と待ち時間の短縮が求められ，顧客との接点の多いホールスタッフの効率的な作業が重要となる．そこでこれまでに，作業効率の向上やコスト削減などについてのサービス提供者視点の研究や，顧客の待ち時間などについてのサービス受給者視点の研究が行われてきた．例として，顧客とスタッフとの接触回数が多い飲食店において，ホールスタッフの作業軌跡の可視化によって，サービス作業改善を行った研究 [新村 12] がある．

ホールスタッフの作業は，顧客の要求に臨機応変に対応しなければならないため，自動化することが難しく人間によって処理されることが多い．そのため，その良し悪しが作業を行う人間の経験や勘に依存するという問題がある．本研究では，ゲームを用いることで，ホールスタッフの経験と勘に基づく「動き方」や「選択」をデータで残し，回転率が求められる飲食店においてのホールスタッフのノウハウを抽出する．そして，そのノウハウが，待ち時間や売上にどのように影響を与えているかを分析する．

2. ホールスタッフのノウハウ

ホールスタッフの主な役割として，座席案内，呼び出し，配膳，バッシング，会計がある．本研究では，待ち時間により大きな影響を与えていると考えられる，呼び出し，配膳，バッシングに着目する．ホールスタッフが一定の区間において，配膳，呼び出し，バッシングをまとめて処理する行動を「タスクまとめ」とする．顧客が呼び出すことや食事を終了することが予想できた場合，先を見越して近くまで移動することを「予測行動」とする．この「タスクまとめ」と「予測行動」に着目し，ノウハウを分析していく．

連絡先: 青山学院大学水山研究室，神奈川県相模原市中央区 5-10-1 青山学院大学相模原キャンパス O 棟

レストランゲームでは一定の区間を，図 1 に示している料理カウンターポイントを出発してから，次に料理カウンターポイントへ戻ってくるまでの間とし，これを 1 ムーブと呼ぶことにする．「タスクまとめ」の二つの行動パターンの例を挙げる．一つ目は，料理カウンターで配膳料理をとり，その料理を提供してから，そのまま料理カウンターへ戻る行動，二つ目は，料理カウンターで配膳料理を取り，その料理を提供した後，他の卓の呼び出しを受けてから，さらに他の卓のバッシングをして料理カウンターへ戻る行動である．前者は 1 ムーブで配膳一つのタスクを処理しただけだが，後者は 1 ムーブで配膳，呼び出し，バッシングの三つのタスクを処理したことになる．その分，それぞれのタスクにかかる時間は短縮されたことになる．「タスクまとめ」とは、このように，タスクが発生している位置や，タスクの待ち時間に応じて，まとめてタスクを処理することである．「予測行動」はタスクが発生する前から発生時までの従業員とタスクの位置の距離の変化を見る．その変化を「近づき距離」とする．この「近づき距離」が大きいほど，次のタスクがどこで発生しそうであるかを予測してると考え，その近づく行動が「予測行動」である．

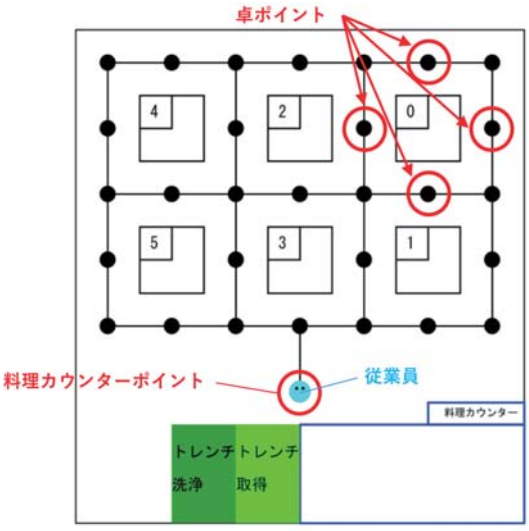


図 1: 店内の状況

3. ゲームの設計

今回のゲームでは六つのテーブルを用意した．プレイヤは従業員を操作し，呼び出し，配膳，バッシングの作業を行う．

3.1 従業員の設定

従業員はグリット上の点を移動し，各テーブルに対して顧客から要求された作業を行う．テーブルに対して作業を行うためには，図 1 に示している，そのテーブルの上下左右の卓ポイントにいなければならない．従業員は配膳，バッシングをする際はトレンチと呼ばれるお盆を持つ．トレンチに乗せることができる皿の数は 5 個までとする．顧客の呼び出しを受ける際はトレンチに上に皿がない状態にしておかなければならない．1 ムーブにおける従業員の処理の制約条件を図 2 に示す．

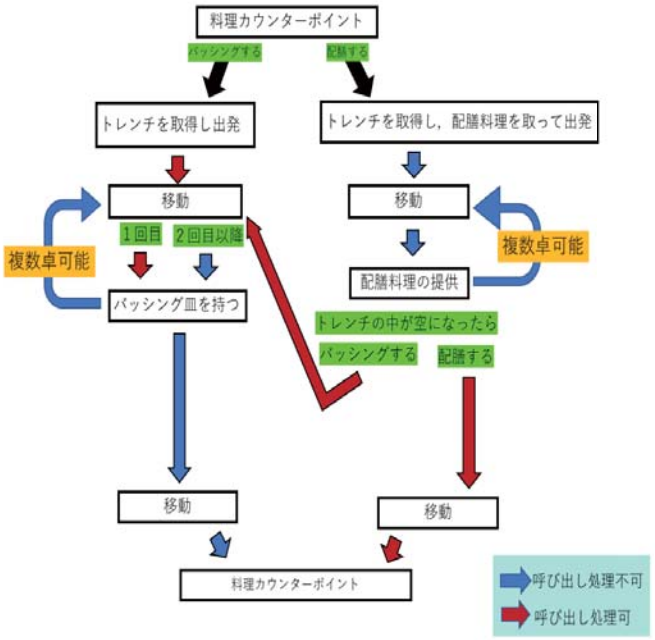


図 2: 従業員の作業の制約条件

3.2 顧客の設定

1 人ずつランダムに来店し，注文する回数は 1 回であり，注文する料理は 1 ～ 3 品のランダムである．注文した料理が届き食事時間が終了すると自動で退店する．

3.3 得点配分

ホールスタッフがすべきタスクが発生すると，そのタスクが発生してから時間に応じて白色から始まり，黄色，オレンジ色，赤色変化していく（図 3）．今回はタスクの色に応じて次のように得点を変化させる．

- タスクの色が黄色の場合は処理したタスク数 × 10 点
- タスクの色が赤色の場合は処理したタスク数 × -10 点

また，売上として注文を受けた数に応じて得点が入る．

- 注文数 × 10 点

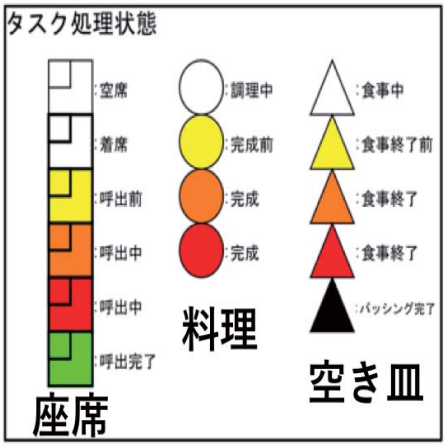


図 3: タスクの状態

4. 分析アプローチ

ノウハウを分析するためのログデータを取ることが重要である．タスクまとめは，1 ムーブごとに処理を行ったタスクの種類と場所，またその回数を記録しておく．一方，予測行動ではタスクが発生する前から，発生時までの従業員とタスクの位置の距離の変化を記録しておく．そして得られたログからノウハウを分析する．

5. 結果

1 ムーブでのタスク処理数，また，近づき距離をそれぞれ売上と待ち時間（タスクが発生してから処理されるまでの時間）に着目してグループわけをした．タスクまとめに関しては，待ち時間が少なく，売上が高いグループに着目した結果，1 ムーブでタスクを組み合わせた処理をした回数が他のグループより多く，場所を移動せずに複数のタスクを処理した回数も多かった．予測行動に関しても，待ち時間が少なく売上が高いグループに着目すると，近づき距離が他のグループより大きかった．

6. おわりに

本研究では人間が主観で行っているホールスタッフの作業ノウハウを，ゲームを用いて分析することができた．

参考文献

[新村 12] 新村猛, 竹中毅, 上村玲子, 蔵田武志, 大浦秀一: 外食産業におけるサービス工学の導入事例, 精密工学会誌, Vol78, No.3(2012)

[Nonaka 14] T. Nonaka, M. Igarashi and H. Mizuyama: A Customer Satisfaction Model for Effective Fast Fashion Store Service, Advances in Production Management Systems: Innovative and Knowledge-Based Production Management in a Global-Local World Part II, Springer, pp.587-594 (2014)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

機械学習を用いた飲食店運営の効率化へのアプローチ

Approach to improvement of restaurant management using machine language

星野智洋[*1], 谷﨑隆士[*1], 新村猛[*2], 竹中毅[*3]
Tomohiro Hoshino, Takashi Tanizaki, Takeshi Shimmura, Takeshi Takenaka

[*1] 近畿大学大学院, [*2] 立命館大学, [*3] 国立開発法人産業技術総合研究所
Graduate School of Kindai University, Ritsumeikan University, National Institute of Advanced Industrial Science and Technology

Keywords: 機械学習, 需要予測, レストラン運営

Summary

The service industry is an important industry that accounts for about 70% of Japan's GDP. However, since the labor productivity of the service industry is lower than that of the manufacturing industry, productivity improvement in the service industry is the country's most important policy issue. In order to solve such problems, we research support method for sophisticated store management based on highly accurate future prediction for face-to-face service industry. As part of it, we research prediction methods using external data existing in the ubiquitous environment such as weather, events and internal data such as POS data etc. In this paper, we describe comparison of forecasting methods and material ordering for dishes based on machine learning.

1. 背景

サービス産業は, わが国の GDP の約 7 割を占める重要な産業である. しかし, 製造業と比較してサービス産業の労働生産性は低いことから, その向上は国の重要政策課題である. 特に, 労働集約型サービス産業はサービス財の同時性により労働生産性が低い. 労働集約型サービス産業の中でも, 製販一体型サービス産業である飲食業は, セントラルキッチン他の導入により, サービス生産機能を販売から分離し, 在庫可能性を向上してきた. しかし, ディナーレストラン等の様に対面サービスの品質を重視(対面重視型サービス産業)する場合は, サービス機能を分離することはサービス品質の毀損につながるため, 上記手法の適応は難しい. このような課題を解決するために対面重視型サービス産業を対象に精度の高い将来予測(来店客数予測, 注文数量予測等)に基づき従業員勤務配置, 材料発注等を効率化し, 店舗運営の高度化支援について研究している. その一環として将来予測の精度向上のために POS データ等の内部データに天候, 催事等のユビキタスに存在する外部データを用いた予測手法について研究してきた. 今回, 機械学習を主体とした予測手法の比較と材料発注への取り組みについて報告する.

2. 予測手法

機械学習は, コンピュータ内でデータを反復的に学習することによってデータに内在する規則的なパターンを見つけ出す手法である. 学習結果に新たなデータをあてはめることにより, パターンにしたがって将来の予測が可能となる. 本研究では機械学習のツールとして Azure Machine Learning を用いる.

2.1 Bayesian Linear Regression

ベイジアンネットは, 複数の確率変数の間の条件付き依存関係をグラフ構造によって表し, 確率変数の間の依存関係を条件付き確率で表した確率モデルである[本村 03]. 確率変数, 確率変数間の条件付き依存条件, その条件付き確率の 3 つによって定義される[本村 02]. ベイジアンネットを用いることで, 一部の変数を観測したときの未観測の変数についての確率分布を求めることや, 確率値が最も大きい値をその変数の予測値として得ることができる[鈴木 18]. Bayesian Linear Regression(以下 Bayesian と記す)はベイジアンネットワークを機械学習へ応用した手法である.

2.2 Boosted Decision Tree Regression[ヨアブ・フロイント 99]

Boosted Decision Tree Regression(以下 Boosted と記す)はブースティングを用いて学習を行う手法である. ブースティングとは, N 個の学習機を用いて学習した結果, 誤って予測された事例の重みを増すことでその事例に集中して学習する手法である.

2.3 Decision Forest Regression[渡辺 02]

Decision Forest Regression(以下 Decision と記す) は Random Forest を用いた手法である. Random Forest とは複数の決定木を用いて森を構成し, 各々の決定木ごとに学習を行った結果を統合する手法である. 各決定木で利用する学習データを抽出する際に, ランダム性を取り入れることで, 各決定木の学習に極端な偏りを防ぐ. この結果, 過学習を防ぎ, 高い汎化性能を取得することが出来る.

3. 来店客数の予測

3.1 対象データ

共同研究先のレストランチェーン A から 5 店分の来客データを入手し, 上記手法にて来店客数を予測した. '14/5/1~'15/4/30 の来店実績をもとに'15/5/1~'16/4/30 の来店客数を予測し, 同期間の来店実績と比較した. 予測に用いた説明変数を表 1 に示す.

(1)式, (2)式を用いて, 予測率 α を算出した.

p_i : 1 日の来店実績値
e_i : 1 日の来店予測値
N : 予測期間
α_i : i 日の予測率

連絡先 : 谷﨑隆士, 近畿大学大学院, 〒739-2116 広島県東広島市高屋うめの辺 1, 082-434-7384,
tanizaki@hiro.kindai.ac.jp

α:N 日間の予測率の平均

$$\alpha_i = \frac{p_i - |p_i - e_i|}{p_i} \quad (1)$$

$$\alpha = \frac{\sum \alpha_i}{N} \quad (2)$$

3.2 予測結果

予測結果を表 2 に示す. 統計的手法の1つであるステップワイズ法との比較を行った. 3つの手法およびステップワイズ法はほぼ同等の結果であった.

最も予測率の高い Bayesian を用いた銀座一丁目店の来店客数の予測結果を図 1 に示す. 予測値と実測値の推移は同じ傾向を示している. 予測値平均と実績値平均の差は, 15.5 であり, 年間総数の差は総来客実績の約 5%である. 予測値と実績値の乖離が大きいのは, お盆, 年末年始, 年度初め等である. この乖離は予約客数が多い時期に大きい傾向がある. そこで「予測値と実績値の差」と「予約客数」を散布図にした(図 2). R^2 値が 0.88 以上であり高い相関がある.

表 1 予測に用いる説明変数

与える特性	説明変数名	定義	与える特性	説明変数名	定義
月	1月	1月1日～1月31日を示す.	イベント	連休最終日	3日以上の連休の最終日.
	2月	2月1日～2月28日を示す. (うるう年の場合は2月29日まで)		1月1日	1月1日.
	3月	3月1日～3月31日を示す.		1月2日	1月2日.
	4月	4月1日～4月30日を示す.		1月3日	1月3日.
	5月	5月1日～5月31日を示す.		年末	12月29日, 12月30日, 12月31日.
	6月	6月1日～6月30日を示す.		忘年会	12月の平日.
	7月	7月1日～7月31日を示す.		クリスマスイブ	12月24日.
	8月	8月1日～8月31日を示す.		成人の日	1月8日.
	9月	9月1日～9月30日を示す.		節分	2月3日.
	10月	10月1日～10月31日を示す.		盆	8月13日, 8月14日, 8月15日.
	11月	11月1日～11月30日を示す.		新年会	1月の成人式前までの平日. ただし, 正月(1月1日, 1月2日, 1月2日)を除く.
	12月	12月1日～12月31日を示す.			
曜日	月曜日	平日であり, 翌日も平日の月曜日.		送迎会	3月の平日.
	火曜日	平日であり, 翌日も平日の火曜日.		歓迎会	4月の平日.
	水曜日	平日であり, 翌日も平日の水曜日.	天候	平均風速(m/s)	1日の平均風速(秒速).
	木曜日	平日であり, 翌日も平日の木曜日.		最大風速(m/s)	1日の平均風速(秒速).
	金曜日	平日であり, 翌日も平日の金曜日.		最高気温(℃)	1日の最高気温.
	土曜日	土曜日. 祝日であっても「土曜日」とする.		最低気温(℃)	1日の最低気温.
	日曜日	翌日が平日の日曜日. 祝日であっても「日曜日」とする.		降水量の合計(mm)	1日の降水量の合計.
	連休中日曜日	翌日が休日の日曜日. 祝日であっても「日曜日」とする.		10分間降水量の最大(mm)	1日の10分間降水量の最大.
	祝日	翌日が平日の祝日. 振替休日であっても「祝日」とする.		最大瞬間風速(m/s)	1日の瞬間最大風速(秒速).
	連休中祝日	翌日が休日の祝日. 振替休日であっても「祝日」とする. 土曜日が祝日のときは「土曜日」とする.	機械学習に必要な変数	客数	1日に来店した客の総数.
	祝日前	祝日前の平日. ただし, 金曜日を除く.			

表 2 予測結果

	Bayesian	Boosted	Decision	ステップワイズ
銀座一丁目店	83.8%	81.1%	81.2%	79.0%
銀座四丁目店	82.0%	80.7%	78.4%	82.0%
上野店	76.7%	73.5%	76.8%	81.0%
川崎店	77.9%	77.2%	77.1%	79.0%
立川店	83.2%	81.7%	83.7%	85.0%

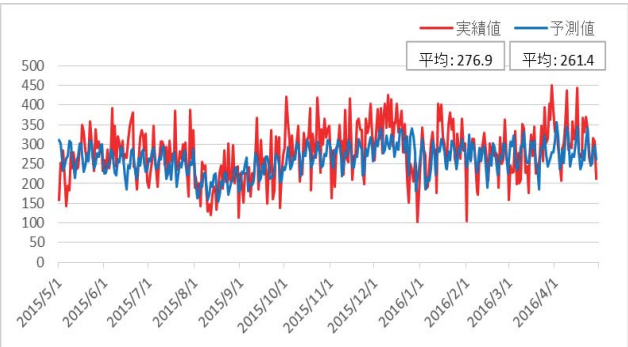


図 1 銀座一丁目店の来店客数予測

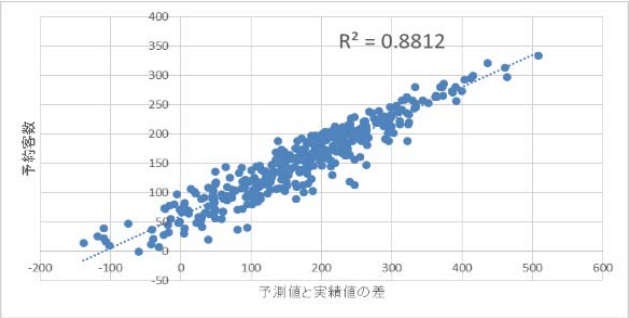


図 2 銀座一丁目店の散布図

3.3 予測手法の改善

3.2 より「予測値と実績値の差」と「予約客数」には高い相関があることが分かった. 来店予約を行う動機は「出張」「接待」等のビジネス理由, 「家族の記念日」等の私的理由に基づくことが多い. このイベントが発生する可能性が高い時期を予測することは可能であるが, 日単位で確実に予測することは困難である. 一方, 研究目標である従業員勤務配置, 材料発注等の自動化には日単位での精度の高い来店客数予測が必要となる. そこで, このチェーン店で数日前から把握可能な来店予約数を来店客数の予測に用いることにした. すなわち, 各店舗の予約なしに来店した客数のみを機械学習で予測し, (3)式にて予測値を算出した(表 3).

Z:総来客数(予測値)
X:予約なし来客数(予測値)
Y:予約客数(実績値)

$$Z = X + Y \quad (3)$$

改善後の予測結果を表 3 に示す. 表 3 より, 表 2 の予測率よりも高くなっていると共に, いずれの店の予測率も 80%以上になった. 銀座一丁目店はいずれの手法も予測率が 90%前後である. 最も予測率の高い Bayesian を用いた銀座一丁目店の来店客数の予測結果を図 3 に示す. 予測値と実績値にほとんど差がないことが分かる.

表 3 改善後の予測結果

	Bayesian	Boosted	Decision
銀座一丁目店	91.0%	89.0%	90.5%
銀座四丁目店	89.4%	88.8%	88.8%
上野店	84.3%	83.4%	84.2%
川崎店	81.0%	83.0%	83.9%
立川店	86.7%	88.3%	87.8%



図 3 銀座一丁目の改善後の来店客数予測

4. 材料発注量の予測

本研究の目標の 1 つは材料発注を自動化することにより，店長の作業的業務を削減し，顧客対応などの高度な業務にシフトすることである．そこで本手法を用いた材料発注量の予測を試みた．レストランチェーン A の銀座一丁目店，銀座四丁目店，上野店，川崎店，立川店の’15/5/1 ~’16/4/30 の発注量データの中で，発注回数が多かった飲物類，鮮魚，冷凍品に限定し，発注回数が最も多い品目を抽出し，’16/5/1~’17/4/30 の発注量を予測した．結果を表 4 に示す． 予測率は 30%～90%の範囲であると共に，60%前後の結果となった予測率が多く，実用レベルではない．

最も予測率の高い Decision を用いた銀座四丁目店の冷凍品のグラフを図 4 に示す．予測値平均と実績値平均の差は，0.8 であり，年間総数の差は総来客実績の約 30%である．予測と実績には乖離があり，特に発注 0 の日が予測できていない．

表 4 発注量の予測結果

		Bayesian	Boosted	Decision
飲物類	銀座一丁目店	61.4%	73.3%	68.3%
	銀座四丁目店	52.0%	61.5%	59.4%
	上野店	33.1%	39.0%	39.5%
	川崎店	58.0%	56.9%	57.6%
	立川店	58.2%	65.5%	66.3%
冷凍品	銀座一丁目店	54.8%	83.0%	55.9%
	銀座四丁目店	92.7%	86.2%	92.7%
	上野店	59.3%	58.6%	58.9%
	川崎店	61.4%	61.6%	63.9%
	立川店	60.7%	61.7%	63.5%
鮮魚	銀座一丁目店	55.4%	63.1%	59.1%
	銀座四丁目店	63.4%	63.7%	68.5%
	上野店	58.4%	58.9%	56.8%
	川崎店	56.2%	58.8%	59.3%
	立川店	68.0%	69.8%	70.9%

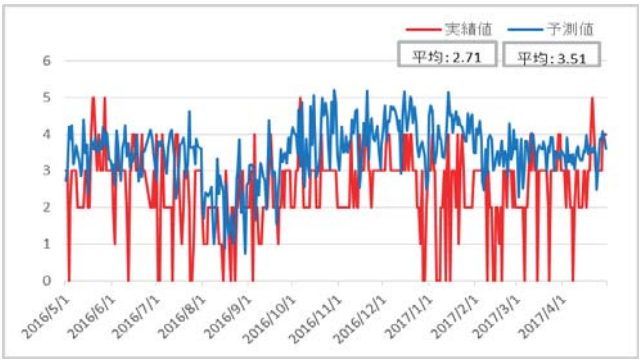


図 4 銀座四丁目店の冷凍品の発注量予測

本予測結果を基に，立川店の店長へヒアリングを行った．その結果，(4)式に示す在庫量が発注点 θ より少なくなれば，発注を行うことが分かった．在庫量変動が少ない本研究の発注方式より，材料入庫の回数が少なく，作業の手間が少ない現行方式が良いとの結論となった．今後，現行発注方式に基づく発注量予測モデルと発注の自動化について研究を進める．

A:在庫量
B:閉店後の在庫量
C:先数日分の予約に基づく材料使用量
D:既発注量

$$A = B - C + D \quad (4)$$

5. まとめ

対面重視型サービス産業を対象とした店舗運営高度化支援の研究の一環として，機械学習を主体とした，「来店客数」と「発注量」の予測を行った．レストランチェーン A から入手した 5 店舗分の来客データを用いた予測結果は，以下の通りである．

(1)来店客数の予測

5 店舗ともに予測率が 80%以上と良好な結果であった．来店予約データを予測に用いることが有用であることが判明した．

(2)発注量の予測

5 店舗ともに，予測率は 60%程度であった．レストランチェーン A で実施している発注点方式を取り入れていない事が，予測率が低い要因である．レストランチェーン A との打ち合わせの結果，在庫量変動が少ない本研究の予測方式より，材料入庫時の作業回数が少ない現行方式が良いとの結論となった．

今後はレストランチェーン A が行っている発注方法を踏まえ，機械学習等を用いて発注量予測の精度を向上する．本研究結果を実用化し，各店舗の店長が行う在庫発注業務の負荷を軽減し，店長がより高度な業務遂行が可能になることを目指す．

謝辞

本研究の一部は JSPS 科研費 16H02909 の助成を受けたものです

参考文献

[本村 02] 本村陽一：ベイジアンネットソフトウェア，人工知能学会論文誌，Vol.13，No.5，p.1，2002.

[本村 03] 本村陽一：ベイジアンネットワーク，信学技報，pp. 1-4. Vol.103,No.285,pp.25-30，2003.

[鈴木 18] 鈴木恒一：大量データから知識を抽出するベイジアンネットワークとその応用，https://www.cac.co.jp/softechs/pdf/st3201_04.pdf(参照 2018.2.6).

[渡辺 02] 渡辺斉：ランダムフォレスト，情報処理学会研究報告，Vol.2012-CVIM-182 No.31，pp.2-3，2002.

[ヨアブ・フロイント 99] ヨアブ・フロイント ロバート・シャピリ 訳：安倍直樹：ブースティング入門，人工知能学会誌，p.771，1999.

Oral presentation | Organized Session | [Organized Session] OS-24

[2O3-OS-24b] [Organized Session] OS-24

Wed. Jun 6, 2018 3:20 PM - 4:40 PM Room O (2F Kaimon)

[2O3-OS-24b-01] Estimation of Unwritten Cooking Utensils in Recipe Texts

Kento Nagai[1], ○Takafumi Ohsugi[1], Mitsunori Matsushita[1] （1. Kansai University）

3:20 PM - 3:40 PM

[2O3-OS-24b-02] How can we apply AI to a study of gastronomy?

○Hiroshi Izawa[1], Ryosuke Yamanishi[1] （1. Ritsumeikan University）

3:40 PM - 4:00 PM

[2O3-OS-24b-03] Improving Labor Productivity for Japanese Cuisine Restaurant Using Simulator

○Takeshi Shimmura[1], Nobutada Fujii[2], Tomomi Nonaka[3] （1. Ritsumeikan University, 2. Graduate School of System Informatics, Kobe University, 3. Department of Industrial and System Engineering, Aoyama Gakuin University）

4:00 PM - 4:20 PM

[2O3-OS-24b-04] Community Value Co-creation Using Home Cooking-plate Network Systems

○Nobutada Fujii[1], Tomomi Nonaka[2], Akio Shimomura[3], Jun Yamadera[4] （1. Kobe University, 2. Ritsumeikan University, 3. Shimomura Shikkiten Co., Ltd., 4. Eyes, JAPAN Co., Ltd.）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

料理レシピ中に記載されていない調理器具の推定に関する基礎検討

Estimation of Unwritten Cooking Utensils in Recipe Texts

永井 廉人 *1　大杉 隆文 *2　松下 光範 *1
Kento Nagai　Takafumi Ohsugi　Mitsunori Matsushita

*1関西大学 総合情報学部
Faculty of Informatics, Kansai University

*2関西大学大学院 総合情報学研究科
Graduate School of Informatics, Kansai University

The purpose of this research is to present a recipe suitable for the user ' s cooking environment from various recipe candidates. Since recipes on user-contributed recipe sites are increasing yearly, it is not easy to find recipes suitable for his/her cooking environment such as cooking utensils and condiments that the user has. In particular, cooking utensils do not always described in the recipe texts. To solve the problem, a search system is desired that can retrieve cooking recipes by estimating omitted cooking utensils and condiments into account. As the first step of the system, this paper proposes a method to estimate omitted cooking utensils. This method uses foodstuffs and cooking behaviors appeared in the recipe texts as clues. As the result of applying this method, we confirmed that it is possible to infer the candidates for cooking utensil omitted in the recipe. This result shows the effectiveness of the proposed method.

1. はじめに

インターネットやスマートフォンの普及により cookpad（https://cookpad.com/）や楽天レシピ（https://recipe.rakuten.co.jp/）のようなユーザ投稿型料理レシピサイトが手軽に利用可能になっている．ユーザ投稿型料理レシピサイトとは，料理レシピを閲覧するだけでなく，自身が作成した料理レシピを投稿することができるものである．料理レシピの豊富さや利用の手軽さから，多くのユーザに利用されている．クックパッド株式会社による料理に関する意識・実態とレシピサイトの利用動向調査（https://info.cookpad.com/pr/news/press_2013_0723）では，料理をする際の「最も参考にする情報源」は料理レシピサイトであることが報告されており，ユーザ投稿型料理レシピサイトに注目が集まっていることがわかる．ユーザ投稿型料理レシピサイトには数多くの料理レシピが投稿されており，掲載される料理レシピ数は日々増加している（https://info.cookpad.com/ir/library_indetx/presentation/）．ユーザはその中から任意のキーワードで料理レシピを検索し，閲覧する．

しかし，料理レシピは投稿者によって書き方が異なり，すべてのユーザにわかりやすく書かれているわけではない [3]．調理手順が詳細に書かれている料理レシピや，調理器具や細かい調理動作を省略して書かれている料理レシピなど様々なものが投稿されている．料理レシピの投稿者とユーザの調理環境は同じであるとは限らないため，そのような書き方が異なる膨大な数の料理レシピの中からユーザが自身の調理環境で調理可能な料理レシピを見つけることは容易ではない．調理環境とは，ユーザが所持する食材や簡単に入手可能な食材，調理器具や調理技術を意味する．

ユーザ個人の詳細な調理環境を考慮した料理レシピの提示を実現するためには，所持する食材の状況や調理動作の難易度，所持調理器具を考慮する必要がある．特に調理器具は，料理レシピを検索する際に考慮すべき重要な要素である．例えば，ユーザは作りたい料理の名称で料理レシピの検索・閲覧を行う

連絡先: 永井 廉人，関西大学総合情報学部
大阪府高槻市霊仙寺町 2-1-1，k743469@kansai-u.ac.jp



図 1: 調理器具が省略されていない料理レシピの例（クックパッド株式会社：cookpad より図引用，https://cookpad.com/recipe/4826011）

が，ユーザの所持していない調理器具が料理レシピで用いられている場合，調理を行うことができない．更に，ユーザ投稿型の料理レシピでは調理器具を省略して書かれることが多い．料理レシピの調理手順には食材と食材に対する調理動作が記述されているが，本来は調理動作の前に調理器具が入り，「何（食材）を何（調理器具）で何（調理動作）をする」という形で表記されることが理想である．図 1 のような調理器具を省略していない料理レシピであれば，どの調理器具を用いるのか明確であり，調理可能であるかの判断は容易である．しかし，実際には図 2 のように調理器具を省略している料理レシピが多く，用いる調理器具が明確ではない．そのため，閲覧した料理レシピが調理可能であるかをユーザが判断することは難しい．

本研究では，料理レシピ調理時におけるユーザの負担軽減のため，使用調理器具の情報を料理レシピに付与することで，調理環境を考慮した料理レシピの提示を目指す．本稿ではその端緒として，料理レシピで省略されている調理器具の候補を推測する手法について検討する．

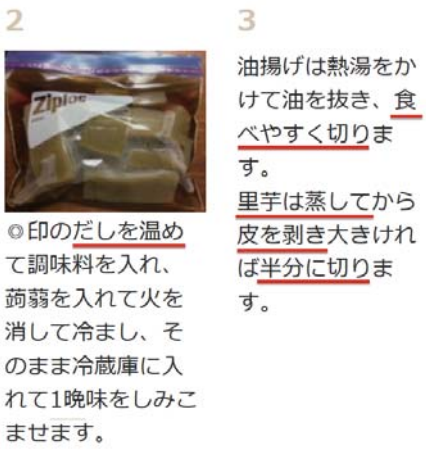


図 2: 調理器具が省略されている料理レシピの例（クックパッド株式会社：cookpad より図引用，https://cookpad.com/recipe/4725878）

2. 関連研究

赤澤らは，冷蔵庫内の食材を考慮した料理レシピ検索を提案した [1]．従来の料理レシピ検索にはユーザの冷蔵庫にある食材の状態が関連付けられておらず，冷蔵庫内の食材で調理可能であるのか，どの食材が不足しているのかといった情報を取得することが困難であるという問題を指摘している．この問題に対し，「冷蔵庫内の食材で調理可能な料理は何か」，「不足している食材は何か」，「何を買い足せば調理可能になるか」といった冷蔵庫内の食材を考慮した料理レシピ検索を提案している．「検索キーワード」，「料理のジャンル」，「使用したい冷蔵庫食材」，「使用したくない冷蔵庫食材」，「作りたい料理の分量」，「冷蔵庫には無い食材の数」など，多方向から条件を指定することで，ユーザの冷蔵庫内の食材状況に適した料理レシピ検索の提案を行った．

岩本らは調理の難易度を考慮した料理レシピ検索システムを提案している [4]．従来の料理レシピ検索サイトでは料理レシピの難易度が考慮されていないことを指摘している．ユーザの調理技能で最後まで十分に調理可能な料理レシピが見つからなければ，何らかのアレンジや料理レシピの再検索を行わなければならない．この問題を解決するため，あらかじめ調理動作の難易度を家庭調理技能検定に基づいて定義し，難易度を考慮した検索の検討を行っている．

牧野らは，料理レシピ検索の際に調理の難易度が考慮されておらず，うまく調理できないといった問題を指摘している [5]．この問題を解決するため，家庭科の教科書を用いて調理動作の難易度を客観的に定義し，調理動作に基づく料理レシピの難易度算出法の提案を行っている．

橘らは料理レシピの「簡単」や「楽ちん」といったネーミングコンセプトを抽出するため，食材と調理器具に着目している [7]．料理の典型的な食材と調理器具を取得し，同一料理における差異の特徴パターンを発見することでネーミングコンセプトを抽出する手法を提案している．考察では，典型的な調理器具の抽出の際に，調理器具が料理レシピ文中で省略されているため，正確に抽出できなかったとしている．この結果を受け，「茹でる」なら「鍋」，「炒める」なら「フライパン」のように，調理動作によって使用される調理器具を推測して補うことが必要であると述べている．

これらの先行研究では，調理動作や食材，調理器具に着目して料理レシピの検索や難易度の決定を行っている．しかし，これらは料理レシピに記述されている調理動作や食材，調理器具のみ扱っており，記述されていないものに関しては扱えていない．本研究では，料理レシピに記述されていない調理器具の候補を推測し，個人の調理環境を考慮した料理レシピの提示を目指す．

3. 提案手法

本章では調理器具の推測のために必要である処理について述べる．料理レシピ文中で省略されている調理器具の候補を推測するための手法として，村田らによる文章へ語句を補う研究を参考にした [6]．村田らは，対話システムや機械翻訳システムの実現には自然言語において省略されている動詞を復元することが不可欠なことであるとし，省略されている動詞を表層表現と用例から補完する手法を提案している．表層表現とは，文の表層に現れている表現であり，この表層表現を手掛かりに，省略されている動詞を補っている．例えば助詞の “も” で文が終わっている省略文の場合，助詞の “も” という手掛かりがあることで前文の繰り返しであることが推測でき，前文の動詞を補うといった手法である．用例とは，ある語句を使用している例である．この手法に倣い，本稿では料理レシピ文中で表記されている食材と調理動作に着目し，それらの語句を手掛かりとすることで省略されている調理器具の候補を推測する．表記されている調理動作によって使用する調理器具の候補を決め，「切る」であれば「包丁」や「ハサミ」，「スライサー」を候補として挙げ，「焼く」であれば「フライパン」や「オーブン」を候補として挙げることができる．更に，食材の用例を考慮することで使用する調理器具の候補を決めることが可能である．例えば「ゴボウを切る」の場合，「ゴボウを包丁で切る」という用例と，「ゴボウをハサミで切る」という用例が存在すれば，「ゴボウを切る」では「包丁」もしくは「ハサミ」を使用する調理器具の候補として挙げることができる．「マグロを切る」の場合，「マグロを包丁で切る」という用例のみであれば，「マグロを切る」では「包丁」を使用する調理器具の候補として挙げることが可能である．このように料理レシピ文中で表記されている食材と調理動作を手掛かりとし，省略されている調理器具の候補を推測する．そのため，あらかじめ食材，調理器具，調理動作の三つ組をデータとして表 1 のように登録した辞書を構築する．この辞書を参照し，料理レシピ文中に出現する食材と調理動作の組が辞書内の組と一致した場合にその組の調理器具を表示することで，省略されている調理器具の候補を推測する．

4. 辞書の構築

本稿では料理レシピで使用されている調理道具と共起している食材と調理動作に着目し，辞書の構築を行う．構築した辞書を参照し，調理器具の候補の推測を行う，辞書の構築には料理オントロジー構築プロジェクト [*1] を参考にした．調理で頻繁に使用される器具として，本稿では「包丁」，「スライサー」，「ハサミ」，「フードプロセッサー」，「ピーラー」，「フライパン」，「オーブン」，「グリル」，「鍋」，「ボウル」の 10 種類を対象とした．調理動作は，対象とした各調理器具を使用する「切る」，「刻む」，「削ぐ」，「剥く」，「焼く」，「炒める」，「煮る」，「茹でる」，「混ぜる」とした．食材は，対象とした各調理器具名が表

*1 料理オントロジー構築プロジェクト http://www.ls.info.hiroshima-cu.ac.jp/cgi-bin/cooking/wiki.cgi/（2017/11/28 確認）

表 1: 三つ組の例

調理動作	調理器具	食材
切る	包丁	さざえ
切る	包丁	サザエ
切る	包丁	ささみ
切る	包丁	ササミ
切る	包丁	ささみ肉
切る	包丁	ささ身
切る	包丁	サツマイモ
切る	包丁	さつまいも
切る	包丁	さつま芋
切る	包丁	さつま揚げ
切る	包丁	さといも
切る	包丁	サニーレタス

表 2: 抽出した食材数

調理器具	抽出総数	異なり数	食材数
包丁	44,134	1,443	1,076
スライサー	14,612	345	278
ハサミ	5,079	573	425
フードプロセッサー	11,653	892	734
ピーラー	11,497	319	245
フライパン	12,721	919	691
オーブン	32,709	770	498
グリル	6,997	687	507
鍋	17,089	1,163	889
ボウル	23,190	1,102	848

記されている料理レシピ文を cookpad データセットから取得し，各文中で調理器具名より先に表記されている食材を抽出した．それぞれの調理器具ごとに抽出した食材数を表 2 に示す．調理動作，抽出した食材を各調理器具に対応付け，三つ組の作成を行った（e.g.,＜じゃがいも，包丁，切る＞，＜人参，フライパン，焼く＞）．また,「包丁」,「フライパン」,「鍋」は様々な調理動作で用いられるため，複数の調理動作を対応付けた．その結果，10,614 組のデータが辞書に登録された．

5. 評価実験

5.1 データ処理

Python（ver.3.6.3），MeCab（ver.0.996），Pandas を用いてデータの処理を行った．また，MeCab の辞書には ipadic-neologd[*2] を用いた．

まず料理レシピ文を「。」,「！」,「全角スペース」,「半角スペース」で区切り，一文ずつの処理を行えるようにする．次に，分割した一文を形態素に分け，語句を正規化する．その際,「みじん切り」や「細切り」等は「切る」という動作ではあるが「切る」として認識されないため，語句を置き換える処理を行う．置き換えを行った語句を表 3 に記す．その後，構築した辞書を参照することで，語句の正規化を行った文中に出現している食材と調理動作の組が辞書内の組と一致するかを検証し，一致した場合にその組の調理器具を表示することで推測を行う．

*2 mecab-ipadic-neologd https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md（2018/3/7 確認）

表 3: 置き換えを行った語句

置き換え後	置き換え前
切る	みじん切り 薄切り 細切り
	半月切り 小口切り 乱切り
	千切り ささがき ざく切り
	くし形切り ぶつ切り 斜め切り
	面取り 角切り スライス
	カット 切り落とす 切り取る
	切り刻む 切り離す 輪切り 等 36 語
煮る	煮込む 煮立てる 煮立つ 煮詰める
茹でる	ゆでる
剥く	むく
炒める	炒る いためる
混ぜる	まぜる

5.2 評価

cookpad データセットから料理レシピ 600 件をランダムに取得し，前節のデータ処理により調理器具の候補を推測した．料理レシピ 600 件におけるそれぞれの調理動作の出現回数と，それに対して調理器具を推測した数を表 4 の出現回数欄と推測数欄に示す．また，推測した結果に対し，各調理動作において適切な調理器具が一種類以上含まれている場合を「1」，そうでない場合を「0」とする評価を評定者 2 名により独立に行い，両者が「1」と評価した結果を正解とした．ここで，調理動作に対して適切と考えられる調理器具は複数想定される場合があるが，そのうち一つでも推測結果に含まれている場合は「1」と判断するように指示した．表 4 の正解数欄がこの結果に該当する．評定者 2 名によるこの評価の一致率を Cohen のカッパ係数 [2] を用いて求めた結果，$k = .57$ であった．また，推測数に対する正解数の割合を表 4 の正解率欄に示す．これらの結果から，提案手法の有効性が確認された．

5.3 考察

現状では，食材と調理動作の一組に対して複数の調理器具が辞書内にある場合，それら複数の調理器具を候補として推測している．しかし,「薄切り」や「乱切り」のような調理動作ごとや料理レシピごとに，どの調理器具を使用することが適切であるかを自動で判断する必要があると考える．例えば,「きゅうりを斜め薄切りする」では,「ハサミ」を使用するよりも「包丁」や「スライサー」を使用することで調理の難易度が低くなる．このように，使用する調理器具や食材，またそれらの組み合わせによって調理動作の難易度は変化する．調理器具への重み付けを行うことでこれらを考慮し，省略されている適切な調理器具を一意に推測することが期待できる．

6. おわりに

本研究では調理者の調理時の負担を軽減するために，調理環境を考慮した料理レシピの提示を目的としている．本稿では，調理器具に着目し，料理レシピで省略されている調理器具の候補を推測する手法について検討した．料理レシピ文中に出現する食材と調理動作の組み合わせから省略されている調理器具の候補を推測する手法を提案した．この提案手法を実際の料理レシピに対して試み，評価を行った．その結果，省略されている調理器具の候補を推測することが可能であると示唆された．

今後の展望として，料理レシピの提示方法の検討を行う．著者らは料理レシピをフローで表示する試みを行っている（図 3

表 4: 調理動作ごとの評価

調理動作	出現回数	推測数	推測率 (%)	正解数	正解率 (%)
切る	769	453	58.9	379	83.7
刻む	33	18	54.5	13	72.2
削ぐ	9	5	55.6	5	100.0
剥く	53	33	62.3	31	93.9
焼く	257	69	26.8	43	62.3
炒める	231	138	59.7	130	94.2
煮る	164	62	37.8	47	75.8
茹でる	145	67	46.2	65	97.0
混ぜる	554	183	33.0	169	92.3
平均値	246.1	114.2	48.3	98.0	85.7

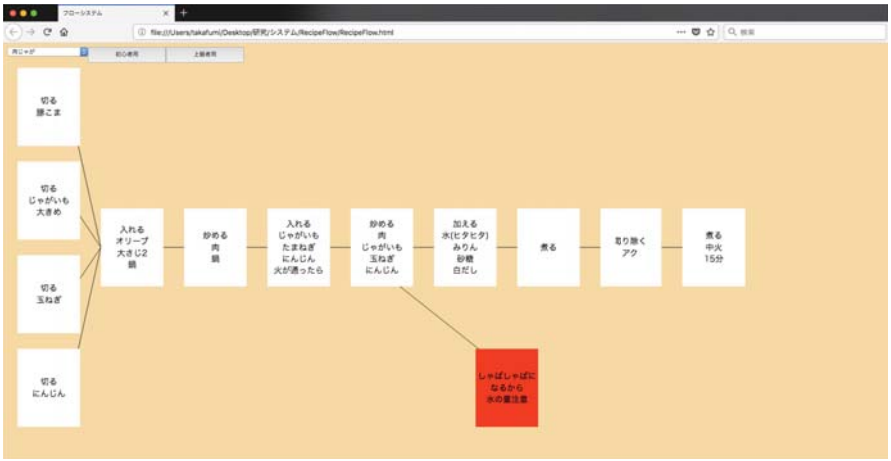


図 3: システム図（文献 [8] より参考）

参照）[8]．このシステムでは調理の動作に着目し，その動作に必要な食材や調理器具，量などを一単位としてブロックとして表示している．しかし，このシステムでは cookpad に投稿されている料理レシピを参考にしているため，調理器具が省略されて提示されている場合がある．そのため料理レシピで使用されている調理器具が推測できた場合，フローの中に調理時に必要となる調理器具の情報を組み込むことが可能になる．また，ユーザが必要な調理器具を所持していない場合に代替可能な調理器具を提示可能なことが示唆される．これらにより理解しやすい形式で料理レシピの提示を行う．

謝辞

本研究では，クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用しました．ここに感謝の意を記します．本研究の遂行にあたり，文部科学省科学研究費（課題番号:15H02780）の助成を受けた，記して謝意を表す．

参考文献

[1] 赤澤康幸, 宮森恒: 冷蔵庫食材を考慮した料理レシピ検索システムの提案, DEIM2011 論文集 E3-2 (2011).

[2] Cohen, J.: A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46 (1960).

[3] Ide, I., Shidochi, Y., Nakamura, Y., Deguchi, D., Takahashi, T. and Murase, H.: Multimedia Supplementation to a Cooking Recipe Text for Facilitating Its Understanding to Inexperienced Users, *2010 IEEE International Symposium on Multimedia*, pp. 242–247 (2010).

[4] 岩本純也, 宮森恒: 調理の難易度を考慮したレシピ検索システムの提案, DEIM2012 論文集 E1-3 (2012).

[5] 牧野望, 塩井隆円, 楠和馬, 波多野堅治: 調理動作に基づく料理レシピ検索のための難易度算出法の提案, DEIM2017 論文集 C4-5 (2017).

[6] 村田真樹, 長尾真: 日本語文章における表層表現と用例を用いた動詞の省略の補完, 自然言語処理, Vol. 5, No. 1, pp. 119–133 (1998).

[7] 橘明穂, 若宮翔子, 難波英嗣, 角谷和俊: 料理名の修飾表現の関係性に基づくレシピのネーミングコンセプト抽出, 電子情報通信学会技術研究報告, Vol. 113, No. 214, pp. 19–24 (2013).

[8] 大杉隆文, 永井廉人, 松下光範: 料理レシピ中の失敗しやすい工程の提示方法に関する基礎検討, 電子情報通信学会技術研究報告, Vol. 117, No. 374, pp. 109–114 (2017).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

食文化研究への AI 活用

How can we apply AI to a study of gastronomy?

井澤 裕司[*1]
Hiroshi Izawa

山西 良典[*2]
Ryosuke Yamanishi

[*1] 立命館大学食マネジメント学部
College of Gastronomy Management, Ritsumeikan University

[*2] 立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University

The research related to gastronomy has been treated as a non-scientific one. The situation may change, through the development of artificial intelligence (AI). Dietary assessment is formed by combinations and interactions of multiple factors. This is not familiar with the so-called modern scientific method, trying to extract the effect of a specific factor. In order to draw out the potential possibilities of such AI, it is essential to develop a database on food culture.

1. はじめに

長く「食事」に関わる研究は「雑学(非科学的な体系性のない知識の集合体)」扱いをされてきた. その状況を人工知能(AI)の発展が一変させる可能性がある.

食事に関わる研究が「雑学」のように扱われてきた要因は複数あるように思われるが, ここでもっとも強調したいのは, 食の評価は複数の要因の組合せと相互作用によって形成されることである. それは「他の要因を一定にして(*ceteris paribus*)」特定の要因の効果を抽出しようとする, いわゆる近代的な科学的方法(還元主義)には馴染まない. これは「デュエム-クワイン・テーゼ(Duhem-Quine thesis)」として知られている, 実証主義的な方法論の限界を示す一例でもある.

Ceteris paribus に基礎づけられた「科学的手法」が支配的であった最大の理由は, 人間の認知能力には限界にあり, 多数の要因を同時的に考慮した「法則性」を理解することが難しかったからである. AI の発達は, 雑学のように見えていた事柄に法則性と体系性を見出す可能性をもたらしうることを示した. これは, 顕微鏡や望遠鏡の登場によって, 人間の認知する世界が革命的に拡大した事例に例えることができるだろう.

AI を食に関わる研究に活用し,「食の体系化」を実現することは大きな付加価値をもたらす. 特に人に固有な能力と考えられてきた hospitality を AI が身に付けたとき, その応用領域は大きく広がる.

2. 食選択行動の特徴

2.1 合理性の限界

食選択が複数の要因の組み合わせによって定まり, そのため, 通常の「合理性」の公理を見たさないことを簡単な事例で示しておこう.

選好関係を“＞”で表わす. 仮に コメ＞パンであるとすれば, 同一の要素を加えた, コメ＋牛乳＞パン＋牛乳が成立しなければならない. さらに醤油＞ケチャップであるなら,

コメ＋牛乳＋醤油＞パン＋牛乳＋ケチャップ

が成立する. これらの推論は, すべて我々の「直感」に反している.

連絡先:井澤裕司, 立命館大学食マネジメント学部, 〒525-8577(固有郵便番号), TEL077-561-4664, izawa@ec.ritsumei.ac.jp.

このような事態が生じたのは, 食選択を構成する個々の要素(食材だけではなく, 環境や, 時間, 場所などを含む)は独立して評価することが難しく, 相互依存関係が極めて強いからである. 何と一緒に食べるかを捨象して, コメとパンのどちらが好きかを考えることに意味はない. これは現代の実証主義を支えている還元主義の大前提に反する.

2.2 要素還元主義の限界

「還元主義(reductionism)」とは,「有意味な命題はすべて 直接的な経験 を表す命題に還元することができる」とする考え方である. ここで「命題」とは個々に独立に存在しうる基本単位であり,「経験的な検証や反証の手続きにおいては, その対象 となるのはあくまでテスト命題と呼ばれる単独の命題であった. したがって, 検証されたり反証されたりする知識の基本単位は『命題』であると考えられてきた.」[野家 2015].

これを言い換えれば, 還元主義の立場に立てば, コメは単体として味覚を構成する要素として分析の対象となるのであり, 他の条件から独立して実証分析の対象となるというのが, 論理実証主義的な科学的手続きである.

これに意義を唱えたのが,「デュエム-クワイン・テーゼ 」として知られる立場である.「クワインの考えによれば, われわれの知識は個々の命題として独立に存在するのではなく, それらが相互に結びついて一つの体系を 構成しているのである. したがって, 知識に対する検証や反証 という手続きも, この体系全体を対象としてなされねばならない.」[野家 2015].

デュエムが指摘したとされるように, あらゆる実験は実験設備に依存する.(時間や場所まで含めて)物理的に同一の条件下で実験を行うことは不可能である. クワインはさらに, 知識や信念の体系は周辺の学問体系や経験, 感覚に依存する以上, あらゆる「命題」は, 単独ではなく, それをさせるシステム全体として検証されなければならないとしたのである.

2.3 AI による新たな分析への可能性

2.1 で見たように, デュエム-クワイン・テーゼは, (少なくとも)食選択行動の分析に対しては極めて妥当で直感に合致する指摘である.

他方, 還元主義が検証方法として長く支持されてきたのは, 主に人間の認知能力の限界に基づくものであったように思われる. 人は複数の要因を総合的に理解できるほどには, 計算能力や認知能力に優れているわけではなく, 相互依存的な複数の

要因に現れる「法則性」を理解することが難しい．このような統合的な知は「雑学」とされてきたのである．

AI の発達は，雑学のように見えていた事柄に法則性と体系性を見出す可能性をもたらしうることを示した．これは，顕微鏡や望遠鏡の登場によって，人間の認知する世界が革命的に拡大した事例に例えることができるだろう．

3. 共感型 AI へ向けて

3.1 食文化研究への応用

食のビッグデータ分析の先駆けとして， 文献 [Sajadmanesh 2017]を挙げることが出来る．そこでは，国際的なレシピ・サイト Yammly にアップされたデータを元に，多要素の風味（flavor)や食材などから情報量を定義し，各国料理の近似性を分析している．その結果，例えば日本料理は，風味ではアジア圏の中に含まれるものの，食材で世界の中でも異質な存在であることが示唆されている．

食のビッグデータ分析はさらに，人の属性データを組合せて嗜好性の「法則性」を探り当てる可能性を示唆しているだけではなく，レシピ・ビジネスの新たな展開へ道を拓くことになる．

ロボティクスや3D プリンターを活用した自動調理技術は目覚ましい発展を見せているが，このような「再生技術」が普及するためには，再生させる元のコンテンツ（レシピ・データ）が必要になる．従来，レシピにはコピーライトが無いものとされてきた．その一つの要因は課金方法にあった．このような「再生装置」を IoT 化するのは自然な発想であり，装置による「再生」の都度，ダウンロードされるレシピ・データに課金するのは容易である．

現在世界最大規模にまでなっているクックパッドなどの主な収入源は広告や会員収入であるが，再生装置の普及は，このようなレシピ・ビジネスのビジネスモデルの規模と内容を一新させる可能性が大きい．

3.2 共感

我々が目指すのは，人の選好や状況を読み取り，いくつかのシナリオを作成し，それぞれのシナリオのもとでの最適なメニューを提示し，さらに，それを選んだ理由を文化的背景を考慮しながら説明する AI である．本研究の革新性は，AI がなぜ複数の目的関数の中からをある関数を選びだしたのかを，文化的背景や行動科学的法則性を考慮しながら人に説明することができ，その選択過程を人が「納得」できるように説明することを目的としていることにある．これは，技術的には，通常 AI は単一の目的関数の最適化問題を解くように設計されているのに対して，多目的関数あるいは複数の目的関数の選択問題を，対象としている人の文化的背景を情報として取り入れながら，解かせることを意味する．説明を受ける人の納得感を規定する要因は，認知科学の方法によって実験データを収集して分析される．さらに AI が提示する選択肢は歴史的，文化的データの学習と解析の結果を参照して作成される．

それが可能になったとき，我々は「共感」が得られたと定義することができるだろう．ここで「共感」とは，経済学の父とされる Adam Smith が『道徳情操論』で論じた社会形成の基盤としての共感(sympathy)のことを意味する．共感は観察による是認により生じる．それらは社会生活を通じて身につけられていく．Adam Smith によれば，社会が成立するためには，他人の心が分かるという意味での「共感」が存在していなければならない．この意味で AI が社会で受容され大きな役割を果たしていくためには，社会をつくっている人々に共感できなければならない．そのためには，AI は社会の歴史・文化の史資料を学習し，それを考慮した意思決定を行うとともに，AI 自らが下した意思決定や判断を人に分かる論理と言葉で説明することが求められるのである．

3.3 共感型 AI の開発

このような「共感能力」を備えることは，AI をひとつの主体として社会に布置するための必須条件である．AI は多くの職能を人から代替すると言われているが，これらの職能の代替が行われるためには，その相手となる人の側に「納得感」が必要である．

特に，AI の判断や選択が人や社会のあり方に具体的な影響を及ぼすようになると，AI には自らの判断を人に理解できる枠組みで説明することが求められる．そのためには，人の認知や理解の根拠となっている歴史や文化を AI 自身が学び，理解していることが必要である．これからの社会は，好むと好まざるとに関わらず，AI の下した判断に多くを依存する社会にならざるを得ない．その時人は，AI に対して，何故そのような判断を下したのかを尋ねたくなるのは当然である．我々が AI の判断を受け入れるためにはその説明が説得的で納得感のあるものでなければならない．例えて言えば，これからの社会は単に勝負に勝つ将棋ソフトではなく，対局後に互いの手順を人と検討し対話することができる「感想戦ソフト」が求められるであろうことを意味している．

食に対する情報科学的アプローチについては国内では電子情報通信学会 HCG 食メディア研究会が中心となって，データベースや情報推薦などの研究が発表されている．また，株式会社クックパッドによるレシピ・データの研究者向けへの公開などを受け，料理レシピを用いた研究は活発化している．国外においても，Workshop on Multimedia for Cooking and Eating Activities などが開催されており，情報科学研究においてレシピや調理活動を対象とした研究は盛り上がりを見せている．しかしながら，食のホスピタリティを高めるための「共感」を目的関数とした食メディア研究は見当たらず，従来の情報推薦研究で延長線上に留まっており機械的な評価指標に基づいた研究がほとんどである．

食文化に対する関心の高まりは世界的な傾向であり，レシピ・ビジネスの将来性は極めて大きいと予想されている．レシピ・ビジネスが成立させるためのプラットフォーム形成は国民経済の成長戦略上も重要である．

参考文献

[野家 2015] 野家啓一. 科学哲学への招待 (ちくま学芸文庫) Kindle 版，2015.

[Sajadmanesh 2017] Sina Sajadmanesh, Sina Jafarzadeh, Seyed Ali Osia, Hamid R. Rabiee, Hamed Haddadi, Yelena Mejova, Mirco Musolesi, Emiliano De Cristofaro, Gianluca Stringhini: Kissing Cuisines: Exploring Worldwide Culinary Habits on the Web. 2017 International World Wide Web Conference Committee, doi:10.1145/3041021.3055137.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

シミュレータを用いた日本料理レストランの労働生産性改善

Improving Labor Productivity for Japanese Cuisine Restaurant Using Simulator

新村猛[*1]
Tskeshi Shimmura

藤井信忠[*2]
Nobutada Fujii

野中朋美[*3]
Tomomi Nonaka

立命館大学[*1]
Ritsummeikan University

神戸大学[*2]
Kobe University

青山学院大学[*3]
Aoyama Gakuin University

This document describes a case study for enhancing labor productivity of Japanese cuisine restaurant using kitchen simulator. Kitchen simulator is developed, and applies it for changing cooking system and kitchen layout. Results show that it is possible to reduce work hour, and correlation of coefficient of sales and labor input, and also possible to enhance sales per man-hour.

1. はじめに

先進諸国を中心に経済のサービス化が急激に進展しており，例えばアメリカでは GDP の 80%，日本では 75%をサービス産業が占めている．経済成長が一定水準に達するとサービス化するとされているため，サービス経済化という現象自体は国民にとって歓迎すべき事象である．加えて，モノ作りのみで収益を得ることが困難になった製造業はサービスによる収益モデルを組み込んだため，製造業にとってもサービス産業は重要な要素となっている（Product Service System）．

しかし，サービス産業の労働生産性は製造業よりも低いため，製造業とサービス産業の労働生産性を加重平均すると全体の労働生産性が低下する．ゆえに，サービス経済の比率が向上した先進諸国や，PSS にシフトした製造業の生産性が低下してしまうというジレンマに陥っている．加えて，サービス産業の低生産性は従業員の低賃金や長時間労働の要因となっており，結果としていわゆる“ブラック企業”発生の温床ともなっている．国全体，企業，労働者いずれの観点においても，サービス産業の労働生産性向上は喫緊の課題である．

サービス産業の低労働生産性要因はサービス財の特性に起因する．サービスは無形であるため在庫できず（無形性）（消滅性），顧客の需要発生と同時にサービスを生産する必要がある（同時性）．加えて，サービスの品質や生産量は人のスキルに依拠することが多く（変動性），生産効率や品質が安定しない．サービス財の特性は労働集約型サービス産業においてより顕著に表れるため，同産業の労働生産性はサービス産業の中でもさらに低位である．

このような社会的要請にこたえるため，21 世紀初頭に工学分野の知見をサービス産業に適用して生産性向上を図るサービス工学が提唱され，十数年にわたって技術開発や現場適用が実践されてきた．サービスは無形であるため，まず可視化してサービスの生産構造を明らかにするという手法が有効であり，サービスシミュレータによる可視化技術やサービス CAD といった設計技術が研究されてきた．また，従来サービス設計は人間の勘と経験に依拠していたため，AI や OR 手法を用いたサービス設計技術も提案されている．本稿では，労働集約型サービス産業の 1 典型である外食産業を対象とした，シミュレータを用いた労働生産性向上を図る試みについて概説する．

連絡先：新村猛，立命館大学，住所，+81-6-6308-2288，E-mail:t-shinmura@gankofood.co.jp

2. 調理場シミュレータを用いたレイアウト設計による生産性向上の取り組み

2.1 調理場レイアウト，生産システムの再設計

外食産業の調理場は，人間の手作業と機械加工の工程を組み合わせたマニュファクチャー的生産システムである．加えて調理場は焼物，煮物などの料理ジャンルごとに作業分担して生産するため，ライン加工的な性格を帯びている．さらに，料理の品質を維持するため，顧客からの注文を受けてから生産する料理が多く，受注生産の比率が高い．調理場の労働生産性を向上させるためには，このような生産システムの持つ特徴を考慮しつつ，生産性低下要因を除去して再構築することが望ましい．

人間の生産能力はスキルに依拠するため習熟を待つしかないが，調理設備の生産能力と料理の注文量とを最適化することで，適切な生産能力を維持しつつ過剰な，あるいは過少な設備投資を回避することが可能である．また，料理ジャンルごとに調理する場合，繁忙時は分業による効率が確保できるが，閑散時期には料理ジャンルと同数の従業員配置が必要であるため，セル生産のように少量多品種の生産システムと併用することが望ましい．さらに，一般客の料理を受注生産する部門と，予約客の料理を計画生産するモデルを組み合わせることで，計画生産ラインの効率化を図ることは可能である．

しかし，このような仮説の有用性を検証するためには実際に設備投資を実行してレイアウト変更をおこない，生産システムを変更する必要があるが，仮説が正しくなかった場合，当該投資は無駄となり，経営効率を下げるとともに従業員への負担増となる．ゆえに，調理場レイアウトと人員配置状況の相違によって生産状況がどのように変化するのかを事前にシミュレートし，定量的な検証を実施してから設備投資を実行することが求められる．

そこで，POSデータ（顧客の注文情報），勤怠管理データ（従業員の出退勤および作業場所），調理データベース（調理作業場所，作業時間），設備データベース（設備ごとの生産能力およびリードタイム）などで従業員ごとの，あるいは設備機器ごとの生産量およびリードタイムを推定することができる調理場シミュレータを開発し，現状の調理場レイアウトにおけるリードタイムや稼働率を求めてその課題を検討した．検討結果に基づき，調理機器量や配置，ラインとセルの組み合わせ，受注生産区画と計画生産区画を分割した調理場レイアウト改善案を作成し，再度シミュレーションして労働時間削減への有用性を検証した後，実際

に設備を変更した. その後, 同店の労働時間 1 時間当たりの売上高(人時売上高)を KPI としてその有用性を確認した.

2.2 結果

レイアウト変更前の 1 日当たり平均労働時間は 70.3 時間(SD＝8.7 時間), 人時売上高は 12,274 円(SD＝3,279 円), 売上と労働時間の相関係数は 0.68((*p*<0.01)であった. レイアウト変更後の 1 日当たり平均労働時間は 56.7 時間(SD＝9.6 時間), 人時売上高は 15,562 円(SD＝3,790 円), 売上と労働時間の相関係数は 0.60((*p*<0.01)であった(図 1-1.1-2).

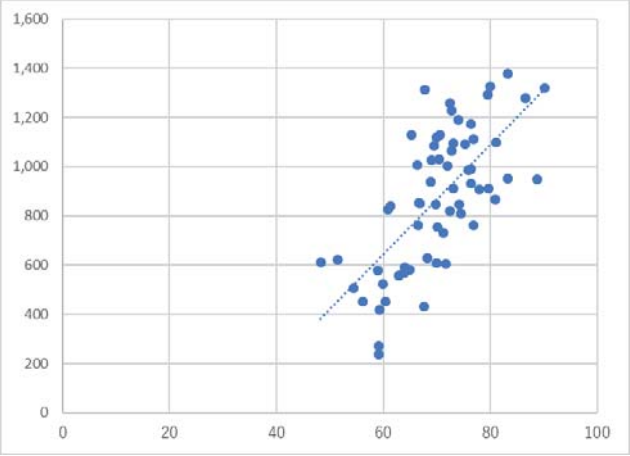


図 1-1:レイアウト改善前の 1 日当たり売上, 労働時間の散布図
(縦軸＝売上高(千円), 横軸＝労働時間(H))

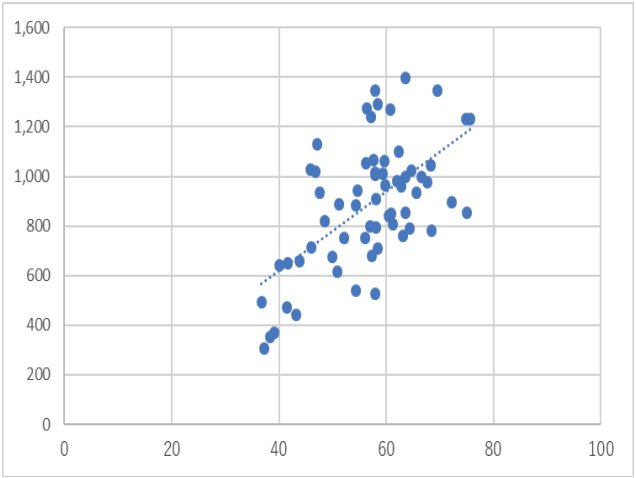


図 1-2:レイアウト改善後の 1 日当たり売上, 労働時間の散布図
(縦軸＝売上高(千円), 横軸＝労働時間(H))

2.3 考察

調理場レイアウト改善前の人時売上高は 12,274 円, レイアウト変更後の人時売上高は 15,562 円であり, 126.8%効率が改善している. 労働時間自体も 70.3 時間から 56.7 時間時間に低下しており, シミュレーション結果に基づいたレイアウト変更が労働生産性向上に寄与していることが確認できる. 一方, 売り上げと労働時間の相関係数は 0, 68 から 0.60 へと低下しており, 本研究によるレイアウト改善が顧客の需要と投入労働量の同時性向上(およびその結果としての相関係数向上)には寄与していない.

従来, サービス財の特性上顧客の需要発生に応じて投入労働量を同期し, 相関を向上させることがサービス産業における労働生産性向上にとって重要であるとされてきた. 当然, その有用性が現在において有効であることに間違いはないが, 一方でサービスの需要と供給の同時性を低下させることで生産性向上を実現することも可能であることを本研究の結果は示している.

従来, サービスの生産と消費の同時性を下げる試みは主としてコンテンツ型サービス産業でおこなわれてきた. 音楽サービスでは演奏を録音または録画し, 音楽サービス生産時点と顧客の視聴時点との同時性を下げることで生産性向上を図ったり, 教育サービスにおいては熟練教員の授業サービスを録画・保存し, 生徒が希望する時間帯に再生することでサービスの同時性を下げるとともに顧客満足度向上を図ったりするサービス再設計などがその典型である. 労働集約型サービス産業においても, IoT や人工知能, サービスロボットを用いてサービス生産システムを再構成し, サービスの同時性を低下させることで労働生産性向上を図るシステムを構築することが望まれる.

3. おわりに

本研究では, 労働集約型サービス産業の 1 典型である外食産業を対象として, 労働生産性向上を図ることを目的として調理場シミュレータを活用した料理生産システム, および調理場レイアウトの変更を試みた. 労働生産性を向上させるため, 調理機器の数量や配置, ラインとセルの組み合わせ, 受注生産区画と計画生産区画とを分割した調理場レイアウト改善案を作成し, 再度シミュレーションして労働時間削減への有用性を検証した後, 実際に設備を変更した. その後, 同店の労働時間 1 時間当たりの売上高(人時売上高)を KPI としてその有用性を確認した. その結果, 調理場シミュレータを活用した調理場レイアウト変更, 調理システム変更によって人時売上高を向上させるとともに, 労働時間を削減することが可能であること, また, 顧客の需要は発生と投入労働量の同時性を下げることが可能であることがわかった. 今後は, 労働生産性向上とって重要であるもう 1 つの要素, 付加価値や品質の向上を実現するための手法を開発して併用することで, 付加価値向上と効率化との両立を可能にすることが望まれる.

謝辞

この研究は, 神戸大学大学院システム情報工学研究科藤井研究室, 青山大学工学部水山研究室との共同研究, および日本科学技術振興機構の科研費(16H02909)の助成を受けて実施しました. 関係各位に感謝申し上げます.

参考文献

[Wölfl 2003] A. Wölfl, Productivity Growth in Service Industries an Assessment of Recent Patterns and the Role of Measurement, OECD Science, Technology and Industry Working Papers, OECD Publishing, 2003.

[内藤 2009] 内藤耕 (編) : サービス工学入門, 東京大学出版会, 東京, 2009.

[今野 2015] 今野晴貴, ブラック企業の沿革と展望, 大原社会問題研究所雑誌, No. 681, pp.6-21, 2015

[鄭 2007] 鄭森豪, サービス産業におけるサービス財の特性, 九州産業大学商經論叢, Vol.48, No.1, pp.119-131 , 2007.

[貝原 2016] 貝原俊也, 藤井信忠, 新村猛, 外食産業を対象とした共創的デザインによる生産性革新統合モデルの構築, 2016 年度サービス学会 第 4 回 国内大会, 1-8, 2016.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

個別クッキングプレート・ネットワーク・システムによる地域価値共創

Community Value Co-creation Using Home Cooking-plate Network Systems

藤井　信忠[*1], 野中　朋美[*2], 下村　昭夫[*3], 山寺 純[*4]
Nobutada Fujii, Tomomi Nonaka, Akio Shimomura, Jun Yamadera

[*1] 神戸大学大学院システム情報学研究科, [*2] 立命館大学食マネジメント学部,
[*3] 株式会社下村漆器店, [*4] 株式会社 Eyes, Japan
Graduate School of System Informatics, Kobe University, College of Gastronomy Management, Ritsumeikan University,
Shimomura Shikkiten Co., Ltd., Eyes, JAPAN Co., Ltd.

This paper proposes home cooking-plate network systems to collect and analyze food-related data toward value co-creation in a community. The system consists of the following three subsystems: (1) individual cooking plate network, (2) information management / planning system, and (3) data analysis / machine learning system. The proposed system will be verified by demonstration experiment in Kawada area in Sabae City.

1. はじめに

少子高齢化や人口の都市への集中などに起因し，特に車などの移動手段を持たない地方の一人暮らしの高齢者の健康状態をいかに管理・サポートするかが重要な課題であることは論を俟たない．さらに近年，高齢者の栄養評価において，「サルコペニア＝筋肉の衰え」「フレイル＝運動能力だけでなく心も含めた心身の衰え」が散見されているが，これは移動手段に問題がある地方だけの傾向ではなく，都市部においても同様であり，高齢化とともに食がおろそかになり，健康的な生活が阻害される傾向にある．これらの状況を打破し，地域として高齢者をサポートしながら高齢者の健康を維持し，健康寿命を延伸することで地域活性化に繋げていかなければならない[厚生労働省 17].

本研究では，これらの背景をもとに高齢者の食に着目する．食を改善することで健常高齢者の健康を増進するだけでなく，入院していた病院を退院し在宅介護の必要な高齢者が再び入院するリスクを低下させるなど，高齢者の健康寿命を延伸することが可能となると考えられる．そこで，一人暮らしの高齢者を対象とし，日常の生活に欠かせない食を起点としながら食の履歴・健康データを収集・分析するための個別クッキングプレート・ネットワーク・システムを提案する．食に関する客観的データだけでなく，配膳する人による対面コミュニケーションによる情報的・物理的両面からの支援を実現することで利用者の健康を維持し，地域の活性化を実現する価値共創[竹中 08]のしくみづくりを提案する．

2. 個（戸）別クッキングプレート・ネットワーク・システム

個（戸）別クッキングプレート・ネットワーク・システムは，下村ら（下村漆器店）がこれまで開発・導入を図ってきている病院・施設向けのインカート・クックシステムを拡張し開発する．インカート・クックシステムをもとに個（戸）別システム化に向けて開発中の個別クッキングプレート（図 1）を用いて，それらをネットワークで結合した個別クッキングプレート・ネットワークを構成し，食事に関するデータを取得し栄養管理するだけでなく，かかりつけ医師・薬局・管理栄養士（在宅 NST（Nutrition Support Team））や民生委員らの協力を得て物理的サポート（対面）と情報的サポート（データ取得・分析）を組み合わせ，総合的に健康サポートを行うシステムである．



図 1 個別クッキングプレート

2.1 インカート・クックシステム

インカート・クックシステムとは，IH 調理が可能な食器を用いてカート内で食材を生から調理することで，できたての料理を高い衛生基準で安全タイムリーに提供できるシステムである（図 1）．生の食材を食器に盛り付けて加熱するため，ひとりひとり個別に計算された栄養量を，そのまま提供することができる．また，素材の持つ水分を利用して加熱調理するため，食材の旨味と栄養成分をあますことなく食せる．

本システムでは，食材の準備，下処理，盛り付け，加熱，提供と従来のクックサーブとは加熱のタイミングが異なる「ニュークックサーブ方式」を採用する．「ニュークックサーブ方式」は，食材の下処理でしか人手を介さず，喫食直前に加熱し，蓋を閉じたままの状態でユーザに提供できるため，加熱調理過程において菌・ウィルスが死滅し衛生面のメリットがある．HACCAP における重要管理点は，この加熱のタイミングだけでよく，コストや手間を軽減することができる．また，必要な材料を必要な各人の分だけ盛り付けることができるため，食材を適量使用でき，食品残渣が軽減できるなどのメリットがあることが確認されてきている．減塩食・低タンパク食，嚥下食など各個人にあわせた食事を一度に調理できるというメリットもある．

従来，病院や施設などの給食施設では，食材を準備し，下処理，加熱，盛り付け，提供を順次行う「クックサーブ方式」が採用されてきた．しかし，加熱後に盛り付けてからの食中毒のリスクや，食事提供時間の一定時間前に調理を行うことが必要なため，大量調理を一度に実施するための人件費やエネルギーコストなど課題を多く抱えていた．インカート・クックシステムでは，総労

連絡先：藤井信忠，神戸大学大学院システム情報学研究科，神戸市灘区六甲台町 1-1，078-803-6133，nfujii@phoenix.kobe-u.ac.jp

働時間の短縮や朝食を準備するための早朝勤務を必要としない人件費削減，廃棄ロスや正確な計量による食材コストの低減，光熱費や空調費などのランニングコストの低減といった効果も確認してきている．

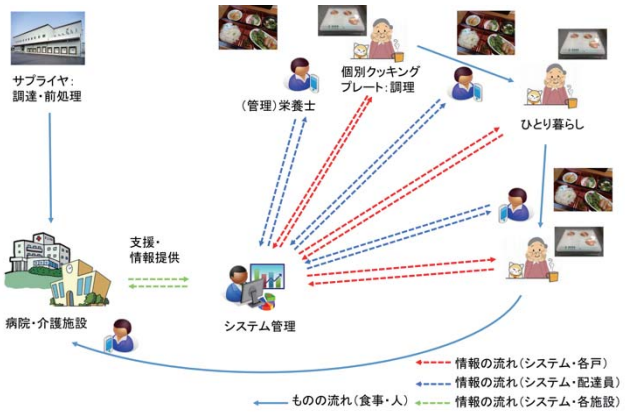


図 2 個別クッキングプレート・ネットワーク・システム

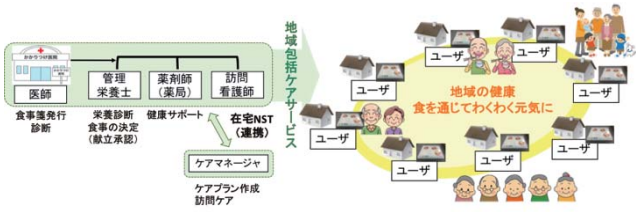


図 3 在宅 NST 連携

2.2 システム全体像

個別クッキングプレート・ネットワーク・システムの全体像を図 2 に示す．個別クッキングプレートが各戸に配置され，ネットワークで接続された状況を想定している．

まず図中のものの流れから説明する．サプライヤから調達した原材料を病院・介護施設などの施設にて下処理，盛り付けを行う．各お膳は協力管理栄養士が各戸に配膳し，前日の膳を回収するとともに食事残渣をチェックし，ユーザとコミュニケーションを行い健康チェックや栄養指導を行う．ユーザ過程全てを回り配膳を終了し，回収した膳の洗浄を行う．

次に情報の流れを説明する．各ユーザの食事調理のコントロールを行うとともに，食事履歴等のデータをシステム管理へと送信する．また，各戸に配膳される配達員が食事残渣等のチェックを行い，それらの情報を手持ちのハンディ端末でシステムへと送付するとともに，投薬管理等の指示に関する情報も提示する．システム管理へと集められた情報は，協力病院，介護施設，管理栄養士，薬剤師，ケアマネージャなどの在宅 NST と情報連携（図 3）し，健康チェックが行われる．

2.3 システム構成

本システムは，次のサブシステムから構成される．

(1) 個別クッキングプレート・ネットワーク

下村らが開発を進めてきている個別クッキングプレートにネットワーク・インターフェイスを導入し，ネットワーク接続可能になるように拡張する．ネットワーク接続が可能になった個別クッキングプレートを相互に接続するためのネットワークシステムの構築の際には，情報の秘匿性・頑健性に配慮してブロックチェーン技術を用いて構築し，すべてのトランザクションを分散保持するようにする．ブロックチェーン技術は，FinTech 分野などでその情報の秘匿性・頑健性が評価されてきており，本システムにおいても情報流通システムの技術基盤として採用する．

(2) 情報管理・計画システム

個別クッキングプレート・ネットワークを用いて，各ユーザにカスタマイズした料理を提供するための管理システムを構築する．これまで構築されている調理レシピ・データベースを参照し，かかりつけ医が発行する食事箋のデータをもとに調理条件を求めて実行する機能を実装する．食材が盛り付けされたプレートの配達を効率よく行うための人員シフトスケジューリング，配送計画スケジューリングについても管理システムに統合する．

(3) データ解析・機械学習システム

本システムの特徴は，システム管理者がデータのハブとなることにある．各ユーザの食事データや食事残渣（現状では目視で確認）など，データを収集し各ユーザの状態の分析を行う．収集したデータを対象に機械学習を用いて，データからユーザの好みを抽出するとともに，食事箋とユーザの思考の両方を考慮したレシピ提案システムを構築する．

3. パイロット地域における実証実験

福井県鯖江市の河和田地区を対象に実証実験を行う．地域のかかりつけ医の協力を得てモニタ家庭を 10 軒程度選定し，個別クッキングプレートを配置する．食材を盛り付けた膳の配送，遠隔調理加熱，喫食，回収といった一連の流れを行い，実際に個別クッキングプレート・ネットワークシステムの運用を行い，システムの有効性と問題点を確認する・

実証実験結果は，各モニタ家庭毎にデータをとりまとめ，統計的手法により分析を行う．かかりつけ医，管理栄養士ら在宅 NST 協力者とともに，データの解釈と，生理データを含むモニタユーザの健康状態チェックを行い，提案システムの効果を検証する．さらに，モニタユーザにはアンケート，ヒアリング調査を行い，ユーザの主観的なシステム評価もあわせて行う．

4. おわりに

本研究では，日常の生活に欠かせない食を起点としながら食の履歴・健康データを収集・分析するための個別クッキングプレート・ネットワーク・システムを提案した．構築したシステムを実証実験により検証し，地域活性化のための社会実装および汎用化につなげていく予定である．

謝辞

本研究の一部は JSPS 科研費 基盤研究（A）17H00841 の助成を受けたものであり，謝意を表する．

参考文献

[厚生労働省 17] 厚生労働省: 地域高齢者等の健康支援を推進する配食事業の栄養管理の在り方検討会報告書価値共創に向けたサービス研究戦略，http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000158814.html (2017).

[竹中 08] 竹中毅，内藤耕，上田完次: 価値共創に向けたサービス研究戦略，情報処理学会論文誌，Vol. 49, No. 4, pp.1539-1548 (2008).

[2P3] [General Session] 2. Machine Learning

座長:木村 圭吾（NEC）

Wed. Jun 6, 2018 3:20 PM - 5:00 PM Room P (4F Emerald Lobby)

[2P3-01] Feasibility Study on PUC for Measurement Noise Reduction

○Takeshi Yoshida[1], Takashi Washio[1], Takahito Ooshiro[1], Masateru Taniguchi[1] （1. The Institute of Scientific and Industrial Research Osaka University）

3:20 PM - 3:40 PM

[2P3-02] Bayesian network classifier learned by conditional marginal likelihood

○Shouta Sugahara[1], Masaki Uto[1], Maomi Ueno[1] （1. The University of Electro-Communications）

3:40 PM - 4:00 PM

[2P3-03] Influence of Deterministic Decision Rules in Fair Logistic Regression

○Toshihiro Kamishima[1], Shotaro Akaho[1], Hideki Asoh[1], Jun Sakuma[2,3] （1. National Institute of Advanced Industrial Science and Technology (AIST), 2. University of Tsukuba, 3. RIKEN Center for Advanced Intelligence Project）

4:00 PM - 4:20 PM

[2P3-04] Data Fusion Method with Kernel Matching for Estimated Canonical Variables

○Masaki Mitsuhiro[1,2], Takahiro Hoshino[3,4] （1. Nikkei Research Inc., 2. Graduate School of Economics, Keio University, 3. Department of Economics, Keio University, 4. RIKEN Center for Advanced Intelligence Projec）

4:20 PM - 4:40 PM

[2P3-05] Multi-label Logistic Regression Using Relative Density Ratio

○Masaaki Okabe[1], Jun Tsuchida[2], Hiroshi Yadohisa[3] （1. Graduate School of Cluture and Infomation Science, Doshisha University, 2. Faculty of Engineering, Tokyo University of Science, 3. Faculty of Culture and Information Science, Doshisha University）

4:40 PM - 5:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

計測ノイズ除去に向けたPU分類学習原理の検討

Feasibility Study on PUC for Measurement Noise Reduction

吉田 剛 Takeshi Yoshida　鷲尾 隆 Takashi Washio　大城 敬人 Takahito Ooshiro　谷口 正輝 Masateru Taniguchi

大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research Osaka University

The needs to employ machine learning is increasing for accurate estimation and noise reduction in recent advanced measurement where its output data is enormous, complex and noisy. Particularly, the recently emerging Positive and Unlabeled Classification (PUC) can be used to classify target objects and contaminants in the measurement. However, the existing standard machine learning is based on Bayesian estimation which assumes invariance of the target population distributions, whereas they are very different depending on the objects in the measurement. In this study, we investigated the PUC to overcome this issue. We applied the method to an actual measurement problem and confirmed its significant noise reduction.

1. はじめに

これまで半教師あり学習に関する研究が多数行われている [Chapelle 10]．この中でも特に，多数のラベルなし事例が簡単に取得できる一方で非常に限られた個数の正例しか得られないという我々がよく直面する問題について，正例とラベルなし事例から学習し正負例を分類する問題 (PU Classification, PUC) [Elkan 08, Ward 09] が，近年とりわけ盛んに研究されてきている．

一方，IoT センシングや自然科学における観測など多くの計測技術において，対象の状態 Y を観測結果 X から正確に推定する技術が非常に重要になってきている．このような計測の場面においては，各観測対象の事前確率密度関数 $p(Y)$ は不明であることがほとんどである．しかし，例えば，温度計は空気や熱水などの計測対象の分布がどうであろうとも，その対象の正しい温度を計測する必要がある．最先端の一分子計測装置では，計測のたび様々に混合比率が変化する目的分子と夾雑物とを分類し目的分子の計測を行う必要がある [Tsutsui 10]．これを達成するために，ほとんどの場合においてセンサーのような計測装置は，$p(Y)$ の変化に対して Y のロバストな推定ができるように，不変な $p(X|Y)$ を安定的に実現するよう設計されている．

これに対して，標準的な機械学習におけるベイズ推定 $p(Y|X) \propto p(X|Y)p(Y) = p(X,Y)$ は，学習用事例集合とテスト用事例集合の同時確率分布 $p(X,Y)$ が同一であることを仮定し，通常はこれら学習用／テスト用事例集合のクラス事前確率 $p(Y)$ も変わらないと仮定している．

これまで提案された PUC 手法はすべてこの設定で開発されている [Elkan 08, Ward 09]．$p(X,Y)$ の全体または $p(Y)$ と $p(X|Y)$ のうち一方が変化する状況をコンセプトドリフト（concept drift）と呼ぶが，通常の教師あり学習において徐々にコンセプトドリフトが起こる場合に対して多くの適応的学習法が提案されている [Gama 14]．より大きなコンセプトドリフトについては，テスト用事例集合のための新しい推定器を学習用事例集合から得る多くの手法が，トランスダクティブ転移学習 [Pan 10] および共変量シフト適応 [Pfeffermann 98] において研究されている．しかし，これらの手法はすべて，新しく得られるテスト用事例集合の分布に関する情報を必要とし，状態 Y が不明な単一の観測値 X が新たに得られるというような計測タスクには適用できない．

このような特徴をもつ計測タスクでは，上記のようなベイズ推定における困難を回避するため，正負ラベル付き学習用事例集合から学習した $p(X|Y)$ の不変性に基づく X による Y の最尤推定が広く採用されている [Washio 17]．しかし我々が知る限り，正例集合とラベルなし事例集合から $p(X|Y)$ を正確に推定する原理は知られていない．このような原理に基づく最尤 PUC が利用可能になれば，後述するノイズ除去技術を含んだ，計測を指向した様々な新しい情報処理技術の開発が実現可能となる．以上の背景から，本研究では以下の成果をもたらす検討を行った．

- 我々は，計測タスクを指向した最尤推定 PUC の新しい原理を検討した．これは，2 標本問題においてクラス事前確率を必要としない初めての PUC である．
- 検討 PUC 手法は，ラベルなし事例が多数得られる一方でラベル付き正例が少数しか得られないようなインバランスな学習用事例集合について正確に正負例分類ができ，現実の多くの問題に適用できる．
- 検討手法は，最先端一分子計測技術におけるノイズ除去という，現実の応用問題において非常に有効に機能することが確認された．

2. 問題設定および関連先行研究

$X \in \mathcal{X} \subseteq \mathbb{R}^d$ $(d \in \mathbb{N})$，$Y \in \{P, N\}$ をそれぞれ特徴ベクトルおよび正負例ラベルとする．学習用に与えられた正例集合とラベルなし事例集合をそれぞれ D_{LP}，D_{LU} とし，計測のたびに取得される事例集合 D_{TU} をテスト用ラベルなし事例集合とする．D_{LP} の事例は正例周辺分布 $p_{LP}(X|Y=P)$ から i.i.d. 標本抽出され，D_{LU}，D_{TU} の事例はそれぞれ周辺分布 $p_{LU}(X)$，$p_{TU}(X)$ から i.i.d. 標本抽出される．ラベル付き負例は存在しない．

ここで，$p(X|Y)$ に関して以下のように仮定する．

Assumption 1. $p_{LP}(X|Y=P)$，$p_{LU}(X)$，$p_{TU}(X)$ は同一の不変な分布 $p(X|Y)$ から構成される．□

連絡先: tyoshida@ar.sanken.osaka-u.ac.jp

これは特殊な仮定ではなく，過去のすべての PUC 手法においても全事例集合にわたって共通の $p(X|Y)$ を仮定している [Elkan 08, Ward 09]．この仮定により，$p_{LP}(X|Y=P) = p(X|Y=P)$ が成り立つので，$Y=P, N$ についての共通の $p(X|Y)$ と正負例のクラス事前確率 $\pi_L = p_{LU}(Y=P)$，$\pi_T = p_{TU}(Y=P)$ を用いて，$p_{LU}(X)$ と $p_{TU}(X)$ を以下のように表せる．

$$p_{LU}(X) = \pi_L p(X|Y=P) + (1-\pi_L)p(X|Y=N), \quad (1)$$

$$p_{TU}(X) = \pi_T p(X|Y=P) + (1-\pi_T)p(X|Y=N). \quad (2)$$

さらに $p(X|Y)$ を正で有界かつ $\mathcal{X}$ において 2 回微分可能であるとし，$\pi_L, \pi_T \in (0,1)$ は不明だが独立に与えられるとする．

我々の問題は，D_{LP} と D_{LU} のみから，D_{TU} から抽出された未知の事例についての正確なノンパラメトリック分類器を学習することである．D_{LP} と D_{LU} は独立に与えられ，かつ相互に異なる未知のクラス事前確率に従うので，これは PUC における 2 標本問題である [Ward 09]．

この問題に取り組む上で，計測タスクに特有の問題が 2 つある．1 つは，D_{TU} に関する情報はその事例を分類する以前には何一つ得られないということである．前述のようにこれまでに提案された PUC では $p(X,Y)$ の不変性を仮定するベイズ推定を用いているために，このような条件下で適用できない．伝統的な正負例の 2 値分類問題において，コンセプトドリフト [Gama 14] や学習／テスト事例集合間の分布が大きく違う場合 [Pan 10, Pfeffermann 98] に，分類器を適応させる手法がこれまでに多数研究されてきている．PUC においては，Li らがデータストリームに対するマイクロクラスターのラベルが徐々に変化する場合における PUC を提案した [Li 09]．ラベル付き負例も学習に用いる通常の半教師あり学習において，Saerens らはクラス事前確率が異なるラベル付き学習用事例集合とラベルなしテスト用事例集合に対して共変量シフト適応を提案した [Saerens 02]．より最近の研究としては，トランスダクティブ転移学習を用いてクラス事前確率を推定するものもある [Plessis 14b]．しかしこれらはすべて，未知のテスト用事例を分類する以前に D_{TU} を必要とし，我々の問題設定において適用することができない．

第 2 の問題点は，π_L と π_T は未知であるということである．したがって，考案すべき PUC は，そもそもクラス事前確率を用いないものであるか，またはクラス事前確率の推定も同時に行うようなものでなければならない．Elkan らの PUC [Elkan 08] はベイズ推定に基づいており我々の問題設定には使えないが，彼らの手法はクラス事前確率を用いないため前者の PUC である．この方法では，D_{LU} の分布に D_{LP} の分布を部分的に一致させることによりクラス事前確率を暗に推定している．そのために 1 標本問題を仮定しており，$p_{LU}(X)$ から抽出される正例はある一定確率でラベル付けされ D_{LP} をなし，残りの正負例を D_{LU} としている [Plessis 15]．しかし D_{LP} と D_{LU} は通常，個別に生成されるので，この 1 標本問題は計測タスクにおいて常に適用できるとは限らない．

我々の問題には不適切であるが，後者のクラス事前確率の推定を同時に行う PUC が，D_{LP} と D_{LU} が独立に提供される 2 標本問題において提案されている [Ward 09, Plessis 14a, Niu 16]．さらに，D_{LU} のクラス事前確率の推定手法も多数，提案されている [Blanchard 10, Plessis 14a, Scott 15, Ramaswamy 16]．これらの手法はすべて $p(X|Y)$ の非還元性，すなわち $p(X|Y=N)$ と他の X の確率密度関数のいかなる線形結合によっても $p(X|Y=P)$ は実現できない，という仮定に基づいている．しかし計測タスクにおいて，この $p(X|Y)$ の非還元性が保証されるとは限らない．さらにこの手法では，未知の事例を分類する以前に π_T の推定のため D_{TU} が必要であり，我々の問題設定には適用できない．

このような問題を解決すべく次節において，D_{TU} を事前に必要ともせず，π_L，π_T の推定も必要としない，最尤推定原理に基づく新しい PUC を検討する．

3. 検討手法

3.1 原理

π_L と π_T の違いにロバストな分類器を構成するため，クラス事前確率の影響を受けない最尤推定原理を用いる分類基準を採用する．あるラベルなしテスト事例 $x \in D_{TU}$ の最尤な Y は，仮定 1 より以下で与えられる．

$$y = \begin{cases} P & \text{if } p(x|Y=P) \geq p(x|Y=N), \\ N & \text{if } p(x|Y=P) < p(x|Y=N). \end{cases} \quad (3)$$

この基準に関連して，以下の補題が成立する．

Lemma 1. 任意の $\pi \in (0,1)$ を有する

$$p_\pi(X) = \pi p(X|Y=P) + (1-\pi)p(X|Y=N)$$

について，以下の 2 つの不等式は等価である．

$$p(X|Y=P) \geq p_\pi(X) \Leftrightarrow p(X|Y=P) \geq p(X|Y=N). \square$$

Proof. $p(X|Y=P) \geq p_\pi(X)$ に上記の $p_\pi(X)$ の定義を代入し，以下を導く．

$$\begin{aligned} & p(X|Y=P) \geq \pi p(X|Y=P) + (1-\pi)p(X|Y=N) \\ & \Leftrightarrow (1-\pi)p(X|Y=P) \geq (1-\pi)p(X|Y=N) \\ & \Leftrightarrow p(X|Y=P) \geq p(X|Y=N) \textit{ s.t. } \pi \in (0,1). \end{aligned}$$ ■

この補題と仮定 1 から，直ちに我々の検討手法の中核原理である以下の定理を得る．

Theorem 1. 任意の $\pi_L \in (0,1)$ の下で与えられる以下の式は，π_L と独立に与えられる任意の $\pi_T \in (0,1)$ をもつ $p_{TU}(X)$ に従う事例 $x \in D_{TU}$ の最尤分類基準である．

$$y = \begin{cases} P & \textit{if } p(x|Y=P) \geq p_{LU}(x), \\ N & \textit{if } p(x|Y=P) < p_{LU}(x). \end{cases} \quad (4)$$

□

Proof. *Eq.(1), (2)* と補題 *1* より，以下が成立する．

$$\begin{aligned} & p(X|Y=P) \geq p_{LU}(X) \Leftrightarrow p(X|Y=P) \geq p_{TU}(X) \\ & \Leftrightarrow p(X|Y=P) \geq p(X|Y=N). \end{aligned}$$

仮定 *1* より最後の不等式は π_L,π_T に依存しないので，任意の $\pi_L, \pi_T \in (0,1)$ について，これら 3 つの不等式は等価でありかつ不変な条件を表す．したがって，*Eq.(3)* より，*Eq.(4)* は $p_{TU}(X)$ に従う $x \in D_{TU}$ の最尤分類基準である． ■

D_{LP}，D_{LU} からそれぞれ $\hat{p}(x|Y=P) := \hat{p}_{LP}(x|Y=P)$ と $\hat{p}_{LU}(x)$ をノンパラメトリック推定し，Eq.(4) を用いて $x \in D_{TU}$ のラベル y を最尤推定する PUC を構成できる．

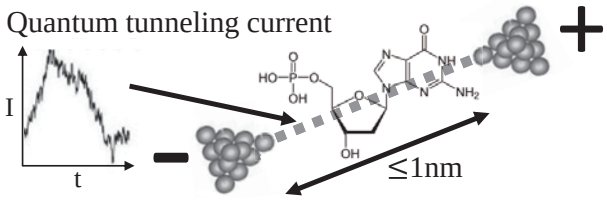


図 1: ナノギャップ電極の模式図

手法	$\pi_L \simeq \pi_T$	$\pi_L < \pi_T$	$\pi_L \ll \pi_T$
NB-E&N	0.572	0.582	0.592
KD-E&N	0.318	0.322	0.305
検討 PUC	0.715	0.671	0.654

表 1: 各手法におけるノイズ除去効果の比較

3.2 パラメータ最適化と性能指標

$\hat{p}(x|Y=P)$, $\hat{p}_{LU}(x)$ のノンパラメトリック推定にはカーネル幅 h のガウスカーネル密度を用い，この h の最適化は以下のように行う．D_{LP} と D_{LU} からランダム抽出したうち 90% をそれぞれ学習用 D^L_{LP}，D^L_{LU} とし，残りの 10% をテスト用 D^T_{LP}，D^T_{LU} として，10CV 法を用いる．このように最適化した h を用いて，すべての D_{LP}，D_{LU} により $\hat{p}(x|Y=P)$，$\hat{p}_{LU}(x)$ の推定をする．

ところで我々の問題設定ではラベル付き負例が得られないため，上述の 10CV 法においてラベル付きの負例を用いない分類性能指標が必要となる．このような性能指標としては，分類器のパラメータ選択ではなく分類器のクラスの評価法として，AUC と同様な性能指標が多数提案されているが [Hajizadeh 14, Menon 15, Jain 17]，パラメータ選択のための指標はほんの数例しか提案されていない．Lee らは 1 標本問題において，precision と recall の幾何平均と同等な指標を提案した [Lee 03]．Calvo らは，2 標本問題において，真の π_L がわかっている場合には通常の F-measure に一致するような，疑似 F-measure を提案した [Calvo 10]．2 標本問題に適用できるので，我々は後者の指標を採用した．疑似 F-measure は以下で定義される．

$$\tilde{F} = \frac{2\tilde{\pi}|\hat{D}^T_{LP}|/|D^T_{LP}|}{|\hat{D}^{TP}_{LU}|/|D^T_{LU}| + \tilde{\pi}} \tag{5}$$

ここで，D^{TP}_{LU} は D^T_{LU} 中の真の正例集合であり，$\hat{D}^T_{LP}$ と $\hat{D}^{TP}_{LU}$ はそれぞれ D^T_{LP}，D^T_{LU} 中で正例と分類された事例集合である．この疑似 F-measure は，$\hat{D}^{TP}_{LU} = D^{TP}_{LU}$ のとき，つまりラベルなし事例中の正例を正しく分類できた場合に最大となることが示せる．この指標は，h の最適化だけではなく，後に述べるノイズ除去効果の性能指標としても用いる．

4. 実データによるノイズ除去

我々は，現実の一分子計測装置 [Tsutsui 10] におけるノイズ除去のため，検討 PUC 手法を適用し性能評価を行った．図 1 はこの計測装置の模式図である．2 つのナノ電極ギャップの間を物体が通り抜けると，量子トンネル効果により通過物体に応じた電流パルスが計測される．我々の目的は，溶液中の目的分子と夾雑物などのノイズとを，それらのパルス概形に基づいて分類することである．ここでパルス概形は，パルス開始時点から終了時点までを 10 分割した各計測電流値の平均値を成分とする 10 次元特徴ベクトルにより表現する．計測初期の短時間におけるバックグラウンド値の計測により，ノイズのみによるパルス概形の集合を，ラベル付き正例集合として得ることができる．引き続き一定時間による計測により，ノイズと目的分子が混在したパルス概形集合を，ラベルなし事例集合として得られる．この PU 分類問題を解くことによって未知のパルス概形を，ノイズまたは目的分子に分類することができる．

性能比較対象として，Elkan らの PUC [Elkan 08] を用い，$\hat{p}_{LP}(X|Y=P)$ と $\hat{p}_{LU}(X)$ の推定にはガウシアンナイーブベイズ (NB-E&N) と，ガウスカーネル密度を用いたベイズ推定 (KD-E&N) の 2 通りを採用した．

真の正負例ラベルは得られないので，3.2 節で導入した疑似 F-measure をその性能指標として用いる．すなわち，テスト用の正例集合 D_{TP}，ラベルなし事例集合 D_{TU} のうちで正例に分類された事例集合をそれぞれ $\hat{D}_{TP}$，$\hat{D}^P_{TU}$ とするとき，$\tilde{F}=2\tilde{\pi}|\hat{D}_{TP}|/|D_{TP}|/(|\hat{D}^P_{TU}|/|D_{TU}|+\tilde{\pi})$ とする．

各事例集合について，$|D_{LP}|=20$，$|D_{LU}|=800$ を学習用に，$|D_{TP}|=20, |D_{TU}|=100$ をテスト用に取得した．表 1 に，各 PUC 手法の疑似 F-measure の値を示す．D_{TU} として，計測の初期段階（$\pi_L \simeq \pi_T$），しばらく時間が経って夾雑物が多くなった頃（$\pi_L < \pi_T$），非常に夾雑物が増えた頃（$\pi_L \ll \pi_T$）の 3 通りについて調べた．疑似 F-measure は通常のものとは異なり $[0,1]$ に規格化されないが，この値が大きいほど分類性能が高いことを示す．結果，検討手法は，学習用正例数が少ないにも関わらず，π_T の値によらず既存手法よりはるかに高性能を示すことがわかった．

5. 結論

本研究において，対象の母集団分布に依存しない，計測系を指向した新しい PU 分類手法を検討した．検討手法は既存手法に比べ，ラベル付き正例とラベルなし事例の数がインバランス，かつクラス事前確率が大きく異なる場合において，特に高精度・高ロバスト性を示すことがわかった．このような，ラベル付き正例が少数しか得られない一方で，様々なクラス事前確率をもつ多数のラベル付き事例集合がたやすく得られるという，現実に起こりやすいケースに対して，検討手法は非常に有効であることがわかった．

6. 謝辞

本研究は，JST CREST JPMJCR1666 「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用（研究課題名：機械学習と最先端計測技術の融合深化による新たな計測・解析手法の展開）」の支援を受けたものである．

参考文献

[Blanchard 10] Blanchard, G., Lee, G., and Scott, C.: Semi-Supervised Novelty Detection, *J. Machine Learning Research*, Vol. 11, pp. 2973–3009 (2010)

[Calvo 10] Calvo, B., Inza, I., Larrañaga, P., and Lozano, J. A.: Wrapper Positive Bayesian Network Classifiers, *Knowledge and Information Systems*, Vol. 33, No. 3, pp. 631–654 (2010)

[Chapelle 10] Chapelle, O., Schlkopf, B., and Zien, A.: *Semi-supervised learning*, The MIT Press (2010)

[Elkan 08] Elkan, C. and Noto, K.: Learning classifiers from only positive and unlabeled data, in *Proc. KDD08: the 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 213–220 (2008)

[Gama 14] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A.: A survey on concept drift adaptation, *ACM Computing Surveys (CSUR)*, Vol. 46, No. 4, pp. 44:1–44:37 (2014)

[Hajizadeh 14] Hajizadeh, S., Li, Z., Dollevoet, R. P., , M.J, D., and Tax, : Evaluating Classification Performance with only Positive and Unlabeled Samples, in *Proc. S+SSPR14: Structural, Syntactic, and Statistical Pattern Recognition*, Vol. LNCS 8621, pp. 233–242 (2014)

[Jain 17] Jain, S., White, M., and Radivojac, P.: Recovering True Classifier Performance in Positive-Unlabeled Learning, in *Proc. AAAI17: the 31st AAAI Conf. on Artificial Intelligence*, p. 3060 (2017)

[Lee 03] Lee, W. S. and Liu, B.: Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression, in *Proc. ICML03: the 20th Int. Conf. on Machine Learning* (2003)

[Li 09] Li, X.-L., Yu, P. S., Liu, B., and Ng, S.-K.: Positive Unlabeled Learning for Data Stream Classification, in *Proc. SDM09: the 2009 SIAM Int. Conf. on Data Mining*, pp. 259–270 (2009)

[Menon 15] Menon, A., Rooyen, B. V., Ong, C. S., and Williamson, B.: Learning from Corrupted Binary Labels via Class-Probability Estimation, in *Proc. ICML15: the 32nd Int. Conf. on Machine Learning*, Vol. 37, pp. 125–134 (2015)

[Niu 16] Niu, G., Plessis, du M. C., Sakai, T., Ma, Y., and Sugiyama, M.: Theoretical Comparisons of Positive-Unlabeled Learning against Positive-Negative Learning, in *Proc. NIPS16: Advances in Neural Information Processing Systems*, Vol. 29, pp. 1199–1207 (2016)

[Pan 10] Pan, S. J. and Yang, Q.: A survey on transfer learning, *IEEE Trans. on Knowledge and Data Engineering*, Vol. 22, pp. 1345–1359 (2010)

[Pfeffermann 98] Pfeffermann, D., Skinner, C., Holmes, D. J., Goldstein, H., and Rasbash, J.: Weighting for unequal selection probabilities in multilevel models, *J. the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 60, No. 1, pp. 23–40 (1998)

[Plessis 14a] Plessis, du M. C., Niu, G., and Sugiyama, M.: Analysis of Learning from Positive and Unlabeled Data, in *Proc. NIPS14: Advances in Neural Information Processing Systems*, Vol. 27, pp. 703–711 (2014)

[Plessis 14b] Plessis, du M. C. and Sugiyama, M.: Semi-Supervised Learning of Class Balance under Class-Prior Change by Distribution Matching, *Neural Networks*, Vol. 50, pp. 110–119 (2014)

[Plessis 15] Plessis, du M. C., Niu, G., and Sugiyama, M.: Class-prior Estimation for Learning from Positive and Unlabeled Data, in *Proc. ACML15: the 7th Asian Conf. on Machine Learning*, Vol. 45, pp. 221–236 (2015)

[Ramaswamy 16] Ramaswamy, H. G., Scott, C., and Tewari, A.: Mixture Proportion Estimation via Kernel Embedding of Distributions, in *Proc. ICML16: the 33rd Int. Conf. on Machine Learning*, Vol. 5, pp. 2996–3004 (2016)

[Saerens 02] Saerens, M., Latinne, P., and Decaestecker, C.: Adjusting the Outputs of a Classifier to New a Priori Probabilities: A Simple Procedure, *Neural Computation*, Vol. 14, No. 1, pp. 21–41 (2002)

[Scott 15] Scott, C.: A Rate of Convergence for Mixture Proportion Estimation, with Application to Learning from Noisy Labels, in *Proc. AISTATS15: the 18th Int. Conf. on Artificial Intelligence and Statistics*, Vol. 38, pp. 838–846 (2015)

[Tsutsui 10] Tsutsui, M., Taniguchi, M., Yokota, K., and Kawai, T.: Identifying Single Nucleotides by Tunneling Current, *Nature Nanotechnology*, Vol. 5, pp. 286–290 (2010)

[Ward 09] Ward, G., Hastie, T., Barry, S., Elith, J., and Leathwick, J. R.: Presence-only data and the em algorithm, *Biometrics*, Vol. 65, No. 2, pp. 554–563 (2009)

[Washio 17] Washio, T., Imamura, G., and Yoshikawa, G.: Machine learning independent of population distributions for measurement, in *Proc. DSAA17: the 4th IEEE Int. Conf. on Data Science and Advanced Analytics*, pp. 212–221 (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

条件付き周辺尤度を用いたベイジアンネットワーク分類器学習

Bayesian network classifier learned by conditional marginal likelihood

菅原聖太 Shouta Sugahara　宇都雅輝 Masaki Uto　植野真臣 Maomi Ueno

電気通信大学大学院情報理工学研究科
Graduate school of Informatics and Engineering, The University of Electro-Communications

This study proposes a new score for discriminative classifier of Bayesian network. Our proposed score approximates the conditional marginal likelihood. Our experiments showed our proposed score outperformed previous discriminative scores, and it is more suitable for classification than previous generative scores.

1. まえがき

ベイジアンネットワークは，離散確率変数をノードとし，ノード間の条件付き従属関係を非循環有向グラフ（Directed Acyclic Graph: DAG）で表し，同時確率分布を各ノードの親ノード集合を所与とした条件付き確率パラメータの積に分解する，確率的グラフィカルモデルである．ベイジアンネットワークにおける一つのノードを目的変数とし，その他のノードを説明変数とした分類器モデルは，ベイジアンネットワーク分類器（Bayesian Network Classifier: BNC）と呼ばれ，高い分類精度を持つことが知られている [Friedman 97]．

ベイジアンネットワークの DAG 構造は一般にデータから推定する必要があり，この問題をベイジアンネットワークの構造学習と呼ぶ．構造学習では，候補構造から最適な学習スコアを持つ構造を探索するスコアベースアプローチが従来から行われてきた．一般に，学習スコアとして構造を所与とした時の学習データの周辺尤度を用いる．周辺尤度はパラメータの事前分布にディリクレ分布を仮定すると閉形式で表すことができ，この時の周辺尤度は Bayesian Dirichlet（BD）と呼ばれる．また，周辺尤度を近似する情報理論の最小記述長（Minimum Description Length: MDL）[Rissanen 89] もしばしば学習スコアとして用いられる．MDL は，構造のデータへのフィッティングを反映する対数尤度の項と，構造の複雑さに対するペナルティ項の和で表される．BD や MDL で学習した BNC は，全変数の同時確率をモデル化した生成モデルであり，目的変数を所与として説明変数を確率推論するなど，分類問題だけでなく多くの用途に柔軟に対応できる．

一方で，説明変数を所与とした目的変数の条件付き確率をモデル化する識別モデルの方が，漸近的な分類精度が生成モデルより高いことが知られている [Ng 01]．BD や MDL で学習した BNC 構造が，識別モデルの BNC 構造として最適である保証がないことが指摘されており [Friedman 97]，識別モデルの学習スコアとして，MDL のフィッティング項の尤度を，説明変数を所与とした目的変数の条件付き尤度（Conditional Likelihood: CL）に置き換えた Conditional MDL（CMDL）が提案された [Grossman 04]．MDL が周辺尤度の近似であるように，CMDL は条件付き周辺尤度（Conditional Marginal Likelihood: CML）の近似と解釈できる．しかし，実際に CMDL は CML から導出されておらず，CMDL が CML の高精度な近似である保証はなく，数学的なスコアの意味も明確ではない．

さらに，CMDL のフィッティング項の CL を最適にするパラメータ推定式は閉形式で表せないため，パラメータを勾配法で近似的に推定しなければならない．このため，構造学習の全候補構造に対して勾配法によるパラメータ推定を行う必要があり，学習時間が膨大になってしまう．

本論では，CML を定義し，それを実時間で計算可能なように直接近似した approximate CML（aCML）を識別モデルの学習スコアとして提案する．aCML は CML の直接の近似であるため，数学的なスコアの意味が明確であり，分類精度の向上が期待できる．また，CMDL の計算時に行われる勾配法によるパラメータ推定が必要ないため，学習時間の大幅な減少が期待できる．さらに，リポジトリデータベースを用いた評価実験により，aCML の有効性を示す．

連絡先: 電気通信大学大学院情報理工学研究科，調布市調布ヶ丘 1-5-1，042-484-8585，
{sugahara,uto,ueno}@ai.lab.uec.ac.jp

2. ベイジアンネットワークと分類

2.1 ベイジアンネットワーク

ベイジアンネットワークは，確率変数をノードとし，ノード間の条件付き従属関係を非循環有向グラフで表し，各ノードの親ノード集合を所与とした条件付き確率で表現される確率的グラフィカルモデルである．今，$n+1$ 個の離散確率変数集合 $\{X_0, X_1, \cdots, X_i, \cdots, X_n\}$ において，各変数 X_i は r_i 個の状態集合 $\{1, \cdots, r_i\}$ から一つの値をとるとし，各変数 X_i が値 k をとるとき，$X_i = k$ と書く．また，ベイジアンネットワークの構造を G とし，G における変数 X_i の親変数集合を Π_i とする．さらに，θ_{ijk} を Π_i が j 番目のパターンをとったとき（$\Pi_i = j$ と書く）に $X_i = k$ となる条件付き確率 $P(X_i = k \mid \Pi_i = j, G)$ を示すパラメータとし，$\Theta = \bigcup_{i=0}^{n} \bigcup_{j=1}^{q_i} \bigcup_{k=1}^{r_i} \{\theta_{ijk}\}$ とする．ベイジアンネットワークでは，次式のように同時確率分布 $P(X_0, X_1, \cdots, X_n \mid G, \Theta)$ を各変数の条件付き確率パラメータの積に分解して表せる．

$$P(X_0, X_1, \cdots, X_n \mid G, \Theta) = \prod_{i=0}^{n} P(X_i \mid \Pi_i, G, \Theta)$$

2.1.1 ベイジアンネットワークの構造学習

ベイジアンネットワークの構造はデータから推定する必要があり，この問題を構造学習と呼ぶ．構造学習では，候補構造から最適な学習スコアを持つ構造を探索するスコアベースアプローチが従来から行われてきた．一般に学習スコアとして周

辺尤度を用いる．今，全変数に値が割り当てられたサンプルが N 個あり，t 番目のサンプルを $\mathbf{d}^t = \langle x_0^t, x_1^t, \cdots, x_n^t \rangle$ と表し，学習データを $D = \langle \mathbf{d}^1, \cdots, \mathbf{d}^t, \cdots, \mathbf{d}^N \rangle$ と表す．パラメータの事前分布がディリクレ分布と仮定すると，周辺尤度は次のように閉形式で表される．

$$P(D \mid G) = \prod_{i=0}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N^D_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N^D_{ijk})}{\Gamma(N'_{ijk})} \quad (1)$$

ここで，N^D_{ijk} は D において $X_i = k$ かつ $\Pi_i = j$ となる頻度を表し，N'_{ijk} はディリクレ事前分布のハイパーパラメータを表す．また，$N^D_{ij} = \sum_{k=1}^{r_i} N^D_{ijk}$，$N'_{ij} = \sum_{k=1}^{r_i} N'_{ijk}$ である．式（1）の周辺尤度は Bayesian Dirichlet（BD）と呼ばれる．Buntine[Buntine 91] はハイパーパラメータを $N'_{ijk} = N'/(r_i \cdot q_i)$ とした尤度等価を満たすスコアを提案しており，このスコアは Bayesian Dirichlet equivalent uniform（BDeu）と呼ばれ，最もよく用いられるスコアである．

一方，次式に示される，周辺尤度の近似である最小記述長（Minimum Description Length: MDL）[Rissanen 89] は，ベイジアンネットワークと学習データ D の同時記述長を表す．

$$MDL(D \mid G, \Theta) = \frac{\log N}{2}|\Theta| - \log P(D \mid G, \Theta) \quad (2)$$

MDL を用いた学習では，MDL を最小にする構造を最適解とする．式（2）の第一項は構造の複雑さに対するペナルティ項である．式（2）の第二項は構造のデータへの当てはまりを反映するフィッティング項を表す対数尤度である．

BD と MDL は各変数 X_i とその親変数集合 Π_i からなる局所構造のスコアについての総積（MDL の場合は総和）が構造全体のスコアに一致する分解可能なスコアであり，効率的な学習ができる．

2.2 ベイジアンネットワーク分類器

ベイジアンネットワークにおける一つのノードを目的変数とし，それ以外のノードを説明変数とした分類器モデルはベイジアンネットワーク分類器（BNC）と呼ばれ，高い分類精度を持つことが知られている [Friedman 97].

2.2.1 ベイジアンネットワーク分類器のモデル

一般に，BNC の構造学習で探索する候補構造はとりうる全ての構造であり，そのような候補構造に対して BD や MDL などを最適化して学習される BNC は General Bayesian Network（GBN）と呼ばれる．しかし，BD や MDL で学習された GBN は目的変数の子変数が少ない構造をとることがあり，分類のためのベイズ学習のためのデータが少なく，分類精度が低くなる傾向がある [Jensen 07]．この問題を回避できる BNC として，各説明変数が目的変数のみを親に持つと仮定する Naive Bayes や，各説明変数が目的変数を親に持ち，説明変数間で木構造をとると仮定した Tree-Augmented Naive Bayes（TAN）[Friedman 97] などが知られている．また，Naive Bayes や TAN を一般化した，より表現力の高いモデルとして，各説明変数が目的変数を親に持つことのみを仮定する Augmented Naive Bayes（ANB）[Friedman 97]．が知られている．

2.2.2 ベイジアンネットワーク分類器の学習

BD や MDL で学習した BNC は同時確率をモデル化した生成モデルであり，目的変数を所与として説明変数を確率推論するなど，分類問題だけでなく多くの用途に柔軟に対応できる．

一方で，説明変数を所与とした目的変数の条件付き確率をモデル化する識別モデルの方が，漸近的な分類精度が生成モデルより高いことが知られている [Ng 01]．そこで，識別モデルの学習スコアとして MDL のフィッティング項の尤度を，説明変数を所与とした目的変数の条件付き尤度（Conditional Likelihood: CL）に置き換えた Conditional MDL（CMDL）が提案された [Grossman 04]．MDL が周辺尤度の近似であるように，CMDL は条件付き周辺尤度（Conditional Marginal Likelihood: CML）の近似と解釈できる．しかし，実際に CMDL は CML から導出されておらず，CMDL が CML の高精度な近似である保証はない．

さらに，CMDL のフィッティング項の CL を最適にするパラメータ推定式は閉形式で表せないため，パラメータを勾配法で近似的に推定しなければならない．このため，構造学習の全候補構造に対して勾配法によるパラメータ推定を行う必要があり，学習時間が膨大になってしまう．

そこで，Carvalho ら [Carvalho 13] は，分解可能で効率的に計算できるよう対数条件付き尤度を近似した aCLL（approximate Conditional Log Likelihood）を学習スコアとして提案した．aCLL は対数条件付き尤度の最小分散不偏推定量であるため，MDL のフィッティング項を aCLL に置き換えた approximate CMDL（aCMDL）は実時間で計算可能であり，CMDL を精度良く近似できると考えられる．

しかし，CDML そのものが CML の近似として導かれておらず，数学的意味が不明確である．さらに，勾配法によるパラメータ推定が必要であり，その計算時間は膨大である．そこで本論では，CML を定義し，それを実時間で計算可能なように直接近似したスコアを提案する．

3. 条件付き周辺尤度

本節では，条件付き周辺尤度（CML）を定義し，その近似スコアの推定法を提案する．今，$\mathbf{x}_i = \langle x_i^1, \cdots, x_i^N \rangle$ とし，$\mathbf{c} = \langle c^1, \cdots, c^t, \cdots, c^N \rangle, (c^t \in \{1, \cdots, r_0\})$ とすると，CML は以下で定義できる．

定義 3..1 学習データ D に対する構造 G の *CML* は以下で定義できる．

$$P(\mathbf{x}_0 \mid \mathbf{x}_1, \cdots, \mathbf{x}_n, G) = \frac{P(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_n \mid G)}{\sum_{\mathbf{c}} P(\mathbf{c}, \mathbf{x}_1, \cdots, \mathbf{x}_n \mid G)} \quad (3)$$

ここで，$\sum_{\mathbf{c}}$ はデータ列 $\mathbf{c}$ のとりうる値の全パターンに対する総和であり，サンプルサイズに対してパターン数が指数的に増えてしまう．例えば，目的変数が 2 値をとる場合，2^N パターンのデータ列 $\mathbf{c}$ が存在し，サンプルサイズが大きくなると，CML は現実的な時間で計算できない．そこで，本論では現実的な時間で計算できるように CML を近似した学習スコアを提案する．

今，$\mathbf{x}_i^{t;N} = \langle x_i^t, \cdots, x_i^N \rangle$，$D^{1;t} = \langle \mathbf{d}^1, \cdots, \mathbf{d}^t \rangle$ とし，CML の計算量が膨大になる要因をより詳しく解析するため次のように CML を展開する．

$$\begin{aligned} & P(\mathbf{x}_0 \mid \mathbf{x}_1, \cdots, \mathbf{x}_n, G) \quad (4) \\ &= \prod_{t=1}^{N} P(x_0^t \mid x_0^1, \cdots, x_0^{t-1}, \mathbf{x}_1, \cdots, \mathbf{x}_n, G) \\ &= \prod_{t=1}^{N} P(x_0^t | x_1^t, \cdots, x_n^t, D^{1;t-1}, \mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}, G) \end{aligned}$$

式（4）の最右辺のように $\mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}$ を用いて確率推論するには，次式のように $\mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}$ に対応した目

的変数のデータ列 $\mathbf{c}^{t+1:N} = \langle c^{t+1}, \cdots, c^N \rangle$ のとりうる値の全パターンに対して確率推論をしなければならならず，計算量がサンプルサイズに対して指数的に増加してしまう．

$$P(\mathbf{x}_0 \mid \mathbf{x}_1, \cdots, \mathbf{x}_n, G) = \prod_{t=1}^{N} \sum_{\mathbf{c}^{t+1}} P(x_0^t | x_1^t, \cdots, x_n^t, D^{1;t-1}, D_{\mathbf{c}^{t+1}}^{t+1;N}, G) \times P(\mathbf{c}^{t+1} | D^{1;t-1}, \mathbf{x}_1^{t;N}, \cdots, \mathbf{x}_n^{t;N}, G)$$

ここで，$D_{\mathbf{c}^{t+1}}^{t+1;N}$ は $D^{t+1;N}$ の目的変数のデータ列を $\mathbf{c}^{t+1}$ に置き換えたものである．したがって，CML の計算量を膨大にしているのは式（4）の最右辺の条件付き確率の条件部 $\mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}$ である．一方で，$\mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}$ は対応した目的変数のデータがない欠損データと捉えることができ，欠損値を含まない完全データである $D^{1;t-1}$ と比べて確率推論に与える影響は小さい．そこで，現実的な時間で計算できるように，式（4）の最右辺の確率の条件部から $\mathbf{x}_1^{t+1;N}, \cdots, \mathbf{x}_n^{t+1;N}$ を除くことで CML を近似する approximate CML（aCML）を次式のように提案する．

$$P(\mathbf{x}_0 \mid \mathbf{x}_1, \cdots, \mathbf{x}_n, G) \approx \prod_{t=1}^{N} P(x_0^t \mid x_1^t, \cdots, x_n^t, D^{1;t-1}, G) = \prod_{t=1}^{N} \frac{P(x_0^t, x_1^t, \cdots, x_n^t \mid D^{1;t-1}, G)}{\sum_{c=1}^{r_0} P(c, x_1^t, \cdots, x_n^t \mid D^{1;t-1}, G)}$$

また，$c \in \{1, \cdots, r_0\}$ に対して $\mathbf{d}_c^t = \langle c, x_1^t, \cdots, x_n^t \rangle$ とし，aCML の具体的な計算式を次の定理で示す．

定理 3..1 *aCML* は，パラメータの事前分布にディリクレ分布を仮定すると，次のように表せる．

$$\prod_{t=1}^{N} P(x_0^t \mid x_1^t, \cdots, x_n^t, D^{1;t-1}, G) \tag{5}$$

$$= \prod_{t=1}^{N} \frac{\prod_{i=0}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\frac{N'_{ijk} + N_{ijk}^{D^{1;t-1}}}{N'_{ij} + N_{ij}^{D^{1;t-1}}} \right)^{1_{\mathbf{d}^t, ijk}}}{\sum_{c=1}^{r_0} \prod_{i=0}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\frac{N'_{ijk} + N_{ijk}^{D^{1;t-1}}}{N'_{ij} + N_{ij}^{D^{1;t-1}}} \right)^{1_{\mathbf{d}_c^t, ijk}}}$$

ここで，$1_{\mathbf{d},ijk}$ は変数列 $\mathbf{d}$ において $X_i = k$ かつ $\Pi_i = j$ の時に 1 をとり，それ以外の時は 0 をとる変数である．また，尤度等価となるように $N'_{ijk} = N'/(r_i q_i)$ とする．

aCML は CML の直接の近似であるため，スコアとして数学的な意味が明確であり，分類精度の向上が期待できる．さらに，CMDL の計算時に行われる勾配法によるパラメータ推定が必要ないため，学習時間の大幅な減少が期待できる．次節で，aCML で学習したベイジアンネットワーク分類器（BNC）と従来のスコアで学習された BNC の分類精度の比較と，aCML を用いた学習と CMDL を用いた学習の実行時間の比較を行う．

4. 評価実験

4.1 実験手順

UCI リポジトリデータベースから，24 個の学習データを用いて，従来のスコアで学習した BNC と提案した aCML で学習した BNC の比較実験を行った．比較する手法は，Naive Bayes と， TAN に対し対数尤度を最適化した BNC（TAN），ANB に対し BDeu を最適化した BNC（ANB-BDeu），GBN に対し BDeu を最適化した BNC（GBN-BDeu），ANB に対し aCMDL を最適化した BNC（ANB-aCMDL），ANB に対し CMDL を近似的に最適化した BNC（ANB-CMDL），GBN に対し CMDL を近似的に最適化した BNC（GBN-CMDL），ANB-aCML: ANB に対し aCML を近似的に最適化した BNC（ANB-aCML），GBN に対し aCML を近似的に最適化した BNC（GBN-aCML）である．本論では，GBN-aCML を提案手法とする．

解析の妥当性を示すため，構造学習と分類に対して 10 分割交差検証を行った．表.1 の「分類精度」に，各学習データに対する各手法の分類時の正答率を載せた．また，提案手法の分類精度の有意性を示すため，提案手法と各比較手法に対してウィルコクソンの符号順位検定を行った．帰無仮説を提案手法と比較手法の分類精度に差がないとし，対立仮説を提案手法の方が比較手法より分類精度が高いとした．検定結果として z 値と p 値を表.1 の「検定」に載せた．また，提案手法の学習時間の有意性を示すため，ANB-CMDL，GBN-CMDL，ANB-aCML，GBN-aCML の学習時間を測定し，表.2 に載せた．さらに，GBN-BDeu と提案手法の学習構造における目的変数の子変数数（Number of Children of the Class variable: NCC）の 10 分割交差検証における平均値を表.3 に載せた．

4.2 結果と考察

表.1 より，提案手法である GBN-aCML は従来の識別モデルである ANB-aCMDL，ANB-CMDL，GBN-CMDL よりも有意水準 0.05 のもとで有意に分類精度が高かった．この結果から，提案スコアである aCML は CML を直接近似するため CMDL，aCMDL を用いるより適切な BNC 構造を学習できている．

また，表.2 より提案スコアは従来の識別モデルのスコアである CMDL よりも学習時間が少ない．この理由は，CMDL スコアでは探索する候補構造ごとに ELR アルゴリズムによるパラメータ推定を行わなければならず，推定値の局所的最適解を得るのに大きな時間を要するためである．

提案手法と従来の生成モデルである GBN-BDeu の分類精度の有意差は認められなかったが，GBN-BDeu は学習データ 4 番と 16 番では他の比較手法より分類精度が大きく劣っている．表.3 を見ると学習データ 4 番と 16 番における GBN-BDeu は，他の学習データと比較して，目的変数の子変数数（NCC）が少ない構造を学習していることがわかる．このため，2 節で述べたように NCC が少ない構造では分類のためのベイズ学習のためのデータが少なく，GBN-BDeu は学習データ 4 番，16 番の分類精度が著しく悪くなった．一方，提案手法は GBN-BDeu より NCC が多いため分類のためのベイズ学習のためのデータが多く，分類精度が他の比較手法より著しく悪くなることがない．したがって，従来の生成モデルである GBN-BDeu は学習データによって分類精度が不安定であるが，提案手法の分類精度は安定している．

以上の結果から，提案スコアは CML の直接の近似であるため従来の識別モデルのスコアよりも分類精度が高く，提案手法は従来の識別モデルのスコアである CMDL よりも学習時間が少ない．また，提案手法は従来の生成モデルである GBN-BDeu よりも分類のためのベイズ学習のためのデータが多く，分類精度が安定している．

表 1: 各ベイジアンネットワーク分類器の分類精度と，提案手法と他手法の比較検定結果（太字は最大の分類精度を示す）

			Naive Bayes	TAN	ANB-BDeu	GBN-BDeu	ANB-aCMDL	ANB-CMDL	GBN-CMDL	ANB-aCML	GBN-aCML
分類精度	1	lymphography	**0.8446**	0.7297	0.7770	0.7500	**0.8446**	**0.8446**	0.7432	0.7703	0.8108
	2	Breast Cancer Wisconsin	**0.9751**	0.9649	**0.9751**	**0.9751**	**0.9751**	**0.9751**	0.8660	0.9722	0.9707
	3	crx	0.8377	0.8484	0.8515	**0.8591**	**0.8591**	0.8545	0.8346	0.8545	**0.8591**
	4	zoo	**0.9802**	0.9604	0.9505	0.9307	**0.9802**	**0.9802**	0.9604	0.9703	0.9703
	5	Breast Cancer	0.7401	0.7184	0.7040	0.7184	0.7365	**0.7437**	0.6047	0.6787	0.6968
	6	ClimateModel	0.9222	0.9315	0.8426	0.9000	0.9222	0.9222	**0.9407**	0.8741	0.8685
	7	Image Segmentation	0.7290	0.7515	0.8225	0.8156	0.7290	0.7788	0.8190	**0.8255**	0.8203
	8	Congressional Voting Records	0.9095	0.9483	0.9483	0.9655	0.9095	0.9095	**0.9698**	0.9224	0.9224
	9	Solar Flare	0.7811	0.7941	0.8229	**0.8431**	0.7963	0.7840	0.8294	0.7862	**0.8431**
	10	Tic-Tac-Toe	0.6921	0.7620	0.8497	0.8340	0.6858	0.8434	0.8831	0.9499	**0.9530**
	11	Contraceptive Method Choice	0.4671	0.4698	0.4650	0.4542	0.4711	0.4508	0.4372	0.4664	**0.4820**
	12	Nursery	0.9033	0.9250	0.9181	0.9340	0.9197	0.9377	**0.9612**	0.9351	0.9378
	13	Car Evaluation	0.8571	0.9375	**0.9427**	0.9416	0.8571	0.9421	0.9421	**0.9427**	0.9416
	14	Balance Scale	**0.9152**	0.8576	**0.9152**	**0.9152**	**0.9152**	**0.9152**	**0.9152**	**0.9152**	**0.9152**
	15	EEG	0.5778	0.6305	0.6864	0.6814	0.6699	0.6528	0.6782	0.6856	**0.6872**
	16	mux6	0.5469	0.5781	0.5469	0.4531	0.5469	0.5313	0.3594	**1.0000**	0.9844
	17	threeOf9	0.8164	0.8496	0.8730	0.8887	0.8242	0.9219	0.9121	**0.9902**	0.9805
	18	MONK ' s Problems	0.7500	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
	19	parity5+5	0.3633	0.2998	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
	20	mofn-3-7-10	0.8505	0.9252	0.8784	**1.0000**	0.9381	0.8505	0.8474	**1.0000**	**1.0000**
	21	LED Display Domain	0.7294	0.7306	0.7294	0.7294	0.7294	**0.7363**	**0.7363**	0.7294	0.7294
	22	MAGIC Gamma Telescope	0.7482	0.7768	**0.7874**	0.7873	0.7734	0.7807	0.7859	0.7862	0.7864
	23	HTRU2	0.8966	**0.9141**	**0.9141**	0.9112	0.9115	0.9112	0.9117	**0.9141**	0.9059
	24	banknote authentication	0.8433	**0.8819**	0.8812	0.8812	0.8812	0.8433	0.8433	0.8783	0.8724
		平均	0.7782	0.7994	0.8367	0.8404	0.8282	0.8379	0.8242	0.8686	**0.8724**
検定		z 値	-2.9706	-2.7069	-1.9973	-1.1376	-2.0524	-1.651	-2.2766	-0.4024	-
		p 値	0.00149	0.00336	0.02275	0.12714	0.02018	0.04947	0.0113	0.34458	-

表 2: CMDL，aCML を用いた学習時間（秒）

		ANB-CMDL	GBN-CMDL	ANB-aCML	GBN-aCML
1	lymphography	225.4	65.6	8.7	4.4
2	Breast Cancer Wisconsin	133.8	313.2	0.9	0.6
3	crx	270.0	254.4	10.3	6.1
4	zoo	7.3	12.4	7.1	3.4
5	Breast Cancer	22.3	12.0	0.4	0.4
6	ClimateModel	387.3	512.7	20.1	12.7
7	Image Segmentation	10000 以上	10000 以上	220.8	78.3
8	Congressional Voting Records	181.4	59.7	5.5	2.8
9	Solar Flare	2153.5	1085.3	8.3	6.4
10	Tic-Tac-Toe	241.3	166.7	1.7	0.9
11	Contraceptive Method Choice	118.1	136.0	2.1	1.6
12	Nursery	10000 以上	10000 以上	23.7	13.8
13	Car Evaluation	178.8	207.0	0.8	0.9
14	Balance Scale	50.4	26.9	0.1	0.1
15	EEG	758.6	4432.3	211.8	148.1
16	mux6	0.3	0.2	0.1	0.1
17	threeOf9	30.6	47.1	1.3	0.9
18	MONK ' s Problems	3.6	4.3	0.2	0.2
19	parity5+5	1.7	6.3	1.1	1.5
20	mofn-3-7-10	53.8	279.9	9.3	5.5
21	LED Display Domain	121.7	94.4	3.0	2.5
22	MAGIC Gamma Telescope	694.3	1263.6	34.3	32.4
23	HTRU2	454.6	340.5	11.0	7.0
24	banknote authentication	5.9	6.1	0.2	0.2

表 3: ベイジアンネットワーク（GBN）の目的変数の子変数数（NCC）

		GBN-BDeu	GBN-aCML
1	lymphography	5.9	13.8
2	Breast Cancer Wisconsin	8.0	7.2
3	crx	2.8	8.0
4	zoo	1.6	15.7
5	Breast Cancer	0.3	2.5
6	ClimateModel	4.4	8.7
7	Image Segmentation	10.4	13.6
8	Congressional Voting Records	2.6	13.8
9	Solar Flare	0.2	0.5
10	Tic-Tac-Toe	2.2	7.8
11	Contraceptive Method Choice	0.8	2.1
12	Nursery	3.0	7.7
13	Car Evaluation	3.0	3.4
14	Balance Scale	3.6	3.8
15	EEG	8.2	12.2
16	mux6	0.0	5.5
17	threeOf9	2.1	8.6
18	MONK ' s Problems	0.0	3.0
19	parity5+5	0.9	7.2
20	mofn-3-7-10	0.0	9.6
21	LED Display Domain	6.1	6.1
22	MAGIC Gamma Telescope	6.1	9.6
23	HTRU2	4.2	5.1
24	banknote authentication	2.0	3.0

参考文献

[Buntine 91] Buntine, W.: Theory Refinement on Bayesian Networks, in *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 52–60, San Francisco, CA, USA (1991), Morgan Kaufmann Publishers Inc.

[Carvalho 13] Carvalho, A. M., AdÃčo, P., and Mateus, P.: Efficient Approximation of the Conditional Relative Entropy with Applications to Discriminative Learning of Bayesian Network Classifiers, *Entropy*, Vol. 15, No. 7, pp. 2716–2735 (2013)

[Friedman 97] Friedman, N., Geiger, D., and Goldszmidt, M.: Bayesian Network Classifiers, *Machine Learning*, Vol. 29, No. 2, pp. 131–163 (1997)

[Grossman 04] Grossman, D. and Domingos, P.: Learning Bayesian Network classifiers by maximizing conditional likelihood, in *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, pp. 361–368 (2004)

[Jensen 07] Jensen, F. V. and Nielsen, T. D.: *Bayesian Networks and Decision Graphs*, Springer Publishing Company, Incorporated, 2nd edition (2007)

[Ng 01] Ng, A. Y. and Jordan, M. I.: On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes, in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, pp. 841–848, MIT Press (2001)

[Rissanen 89] Rissanen, J.: *Stochastic Complexity in Statistical Inquiry Theory*, World Scientific Publishing Co., Inc. (1989)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

公平ロジスティック回帰での確定的決定則の影響

Influence of Deterministic Decision Rules in Fair Logistic Regression

神嶌 敏弘 *1 Toshihiro Kamishima　赤穂 昭太郎 *1 Shotaro Akaho　麻生 英樹 *1 Hideki Asoh　佐久間 淳 *2 Jun Sakuma

*1産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

*2筑波大学／理化学研究所 革新知能統合研究センター
University of Tsukuba; and RIKEN Center for Advanced Intelligence Project

The goal of fairness-aware classification is to categorize data while taking into account potential issues of fairness. For example, when applying data mining technologies to university admissions, admission criteria must be fair with regard to sensitive features, such as gender or race. We developed logistic regression satisfying such a fairness constraint. In this paper, we show the trade-off between prediction accuracy and fairness can be drastically improved by explicitly considering the influence of a deterministic decision rule.

1. はじめに

公平性配慮型分類とは，採用の可否などの判定の過程から，性別などの公平性の観点から影響してはならない情報を除外するような制約下で行うクラス分類問題である．我々は，今までにロジスティック回帰について正則化項を加える方法を提案していた [Kamishima 12]．ここでは，確定的な決定則の影響を明示的に考慮する実独立性と呼ぶ公平性の規準を満たすことで予測精度と公平性のより良いトレードオフを実現できることを示す．さらに，この実独立性を満たすロジスティック回帰法の高速化についても予備実験を行った．

2. 章では公平配慮型分類問題の形式的定義と実独立性の概念を示したのち，この問題に対処できるように修正したロジスティック回帰について述べる．3. 章では実独理性を達成することでより小さな予測精度の低下でより公平な分類が可能であることを実験的に検証する．4. 章は高速化についての予備的検討で，5. 章はまとめである．

2. 公平ロジスティック回帰

公平配慮型分類問題を定義したのち，モデルベース独立性と実独立性の概念を示す．最後に，これらの独立性を達成する公平ロジスティック回帰モデルについて述べる．

2.1 表記と問題設定

潜在的な公平性の問題に配慮しつつデータを分類するのが公平配慮型分類の目的である．確率変数 S と $\mathbf{X}$ は，それぞれセンシティブ特徴と非センシティブ特徴を表す．公平性を保証すべき情報はこのセンシティブ特徴で表し，それ以外が非センシティブ特徴である．例えば，採用の可否を決めるときには，法的に配慮すべき求職者の性別や人種の情報をセンシティブ特徴とする．ここではセンシティブ特徴は二値のスカラー変数で，非センシティブ特徴は m 次元の実数値ベクトルとする．クラス変数 Y は，採用の可否といった分類クラスを表し，ここでは二値分類を扱う．さらに，真の分布でのクラスの事後分布を近似したものを予測モデルとする．この予測モデルから確率的に生成されたクラスラベルを $\hat{Y}$ で表し，真の分布から生成されたラベル Y とは区別しておく．ここで，実際の予測ラベルは，確率的には生成されず，誤分類リスクを最小化するように次の決定則によって確定的に生成される．

$$\tilde{y} = \arg\max_{\hat{y}} \Pr[\hat{Y}{=}\hat{y}|\mathbf{X}{=}x, S{=}s] \tag{1}$$

この実際の予測ラベルを変数 $\tilde{Y}$ で表す．

次に，分類での 3 種類の公平性を紹介する．一つ目は，$\hat{Y} \perp\!\!\!\perp S \mid \mathbf{X}$ の条件付き独立性で，センシティブ特徴を単純に予測モデルから削除した場合に相当する．このようにセンシティブ特徴を予測モデルから削除しても，センシティブ特徴と相関のある他の変数からの間接的な影響のために不公平な決定がなされる場合がある．これを red-lining 効果という [Calders 10]．二つ目の条件である，予測クラスとセンシティブ特徴の条件なし独立性 $\hat{Y} \perp\!\!\!\perp S$ は，この red-lining 効果をを回避するのに有効である．この公平性では，訓練データ中のラベル情報は潜在的に不公平に判断に基いていると仮定している．このデータ中のラベル情報は公平であると仮定しているのが三つ目の公平性である [Hardt 16, Zafar 17]．この公平性は予測誤差がセンシティブ特徴にはよらないというもので，観測されたラベルが与えられたときの予測クラスとセンシティブ特徴の独立性 $\hat{Y} \perp\!\!\!\perp S \mid Y$ として形式的には定義される．これらの規準のうち，本稿では $\hat{Y} \perp\!\!\!\perp S$ を扱う．

公平性に配慮した分類問題の前に，標準的な分類問題について述べる．真の分布から得られた実現値の対 $(\mathbf{x}, s)$ で各対象は表される．この対象のクラスの実現値 y は真の分布 $\Pr[Y|\mathbf{X}{=}\mathbf{x}, S{=}s]$ から生成されるものとする．なお，この真の分布は，センシティブ特徴に依存した潜在的に不公平なラベルを生成することがあることに注意されたい．この真の分布自体を知ることはできないが，この真の分布から得られたデータは観測できる．これらのデータを集めたものが（訓練）データ集合 $\mathcal{D} = \{(y_i, \mathbf{x}_i, s_i)\}, i = 1, \ldots, n$ である．さらに，センシティブ特徴の値が s であるデータを集めた $\mathcal{D}$ の部分集合を $\mathcal{D}_s$ と記す．モデル分布の族 $\Pr[\hat{Y}, \mathbf{X}, S]$ も与えられたとき，この中から真の分布を最もよく近似するものを特定することが，標準的な分類タスクの目標となる．

では，公平性配慮型分類問題に移る．本論文では，予測クラスとセンシティブ特徴の条件なし独立 $\hat{Y} \perp\!\!\!\perp S$ が公平性の規準

連絡先: ホームページ http://www.kamishima.net

である場合を扱う．この場合では，訓練データ中のラベルは潜在的に不公平で，公平性に配慮した真のラベルの分布は観測できないだけでなく，そこからデータをサンプリングすることですらできない．それゆえ，公平なラベルは公平性規準を満たしているとの仮定を導入する．潜在的に不公平な訓練データ集合，モデル分布の族，および公平性規準が与えられたとき，モデル分布の族中で公平性規準を満たす分布の中から，真の分布を最も良く近似する公平モデル分布を見つけることが公平性配慮型分類問題の目的である．公平性制約の影響で予測に利用可能な情報は一般的に減少するため，予測精度と公平性はトレードオフ関係にある．

2.2 モデルベース独立性と実独立性

ここでは，モデルベース独立性と実独立性の概念を導入する [Kamishima 18]．モデルベース独立性では，モデル分布族の中の分布から直接的にクラスラベルは生成される．一方で，実独立性では，モデルバイアスと決定則を考慮した分布からクラスラベルは生成される．モデルベース独立性ではなく実独立性を満たすことで，予測精度と公平性のよりよいトレードオフを実現できることを 3. 章の実験で示す．

識別モデル [Bishop 08, 1.5.4 節] であるロジスティック回帰をここでは対象としているので，識別モデルの場合での 2 種類の独立性を紹介する．まず，予測モデルからクラスラベルが確率的に生成される場合であるモデルベース独立性から始める．形式的には，この独立性を次式で定義する．

$$\hat{Y} \perp\!\!\!\perp S, \text{ where } (\hat{Y}, S) \sim \Pr[\hat{Y}, S] \tag{2}$$

条件付き分布 $\Pr[\hat{Y}|\mathbf{X}, S]$ を直接的にモデル化するのが識別モデルである．この識別モデルに対しては，$\mathbf{X}$ 上の期待値を標本平均によって近似することで，分布 $\Pr[\hat{Y}, S]$ を得る．

$$\Pr[\hat{y}, s] \approx \frac{|\mathcal{D}_s|}{n} \frac{1}{|\mathcal{D}_s|} \sum_{\mathbf{x} \in \mathcal{D}_s} \Pr[\hat{y}|\mathbf{x}, s] = \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{D}_s} \Pr[\hat{y}|\mathbf{x}, s] \tag{3}$$

なお，ここで標本平均を用いて真の分布を近似しているので，モデルバイアスは除去されており，決定則の影響のみが残っている．

もう一つの実独立性は，予測クラスとセンシティブ特徴の間の独立性である点についてはモデルベースの独立性と同じである．しかし，実独立性では，クラスラベルはモデル分布から生成されるのではなく，決定則の影響をも考慮した分布から生成される．実独立性の形式的定義は次式である．

$$\tilde{Y} \perp\!\!\!\perp S, \text{ where } (\tilde{Y}, S) \sim \Pr[\tilde{Y}, S] \tag{4}$$

予測クラス $\tilde{Y}$ は，確率的に生成されるモデルベース独立性の場合とは異なり，式 (1) の決定則で確定的に生成される．モデルベースの場合の式 (3) と同様に，$\Pr[\tilde{Y}|\mathbf{X}, S]$ の標本平均をとることで分布 $\Pr[\tilde{Y}, S]$ を得る．

$$\Pr[\tilde{y}, s] = \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{D}_s} \Pr[\tilde{y}|\mathbf{x}, s] \tag{5}$$

確定的にラベルを生成する分布 $\Pr[\tilde{Y}|\mathbf{X}, S]$ では，各実現値を生成する確率質量が 0 または 1 のいずれかの値になる．

$$\begin{cases} \Pr[\tilde{Y}{=}1|\mathbf{x}, s] = \begin{cases} 1, & \text{if } \Pr[\hat{Y}{=}1|\mathbf{x}, s] \geq \Pr[\hat{Y}{=}0|\mathbf{x}, s] \\ 0, & \text{otherwise} \end{cases} \\ \Pr[\tilde{Y}{=}0|\mathbf{x}, s] = 1 - \Pr[\tilde{Y}{=}1|\mathbf{x}, s] \end{cases} \tag{6}$$

ただし，$\Pr[\hat{Y}|\mathbf{X}, S]$ は元の識別モデルである．なお，この式の $\Pr[\tilde{Y}|\mathbf{x}, s]$ は $\Pr[\hat{Y}{=}1|\mathbf{x}, s] - \Pr[\hat{Y}{=}0|\mathbf{x}, s]$ にステップ関数を適用したものに相当する．

以上のように，モデルベース独立性と実独立性の二つの公平性制約は，クラスラベルを生成する分布でモデルバイアスや決定則の影響を考慮しているかどうかが異なる．

2.3 実独立な公平ロジスティック回帰

本論文では，偏見除去正則化項 (prejudice remover regularizer) 付きロジスティック回帰 [Kamishima 12]（PR モデルと略す）と呼ぶ公平分類モデルについて扱う．モデルベース独立性と実独立性それぞれの制約を満たす二つの PR モデルを述べる．このモデルの目的関数は，ロジスティック回帰の目的関数に公平性を強化するための制約項を加えたものである．通常のロジスティック回帰の予測モデルは次式である．

$$\Pr[\hat{y}|\mathbf{x}; \mathbf{w}] = y \ \mathrm{sig}(\mathbf{x}^\top \mathbf{w}) + (1 - y)(1 - \mathrm{sig}(\mathbf{x}^\top \mathbf{w})) \tag{7}$$

ただし，$\mathrm{sig}(\cdot)$ はシグモイド関数であり，$\mathbf{w}$ は重みベクトルである．一般性を失うことなく，バイアス項を扱うため入力 $\mathbf{x}$ の最初の要素 $x^{(1)}$ は定数 1 であると仮定しておく．

このモデルを公平性を扱えるように修正する．センシティブ特徴に予測モデルが依存するようにするために，センシティブ特徴のそれぞれの値ごとにロジスティック回帰モデルを作る．

$$\Pr[\hat{y}|\mathbf{x}, s] = \Pr[\hat{y}|\mathbf{x}; \mathbf{w}^{(s)}]$$

重みパラメータ $\mathbf{w}^{(s)}$, $s \in \{0, 1\}$ はセンシティブ特徴の各値ごとに必要となる．PR モデルでは，過学習を避けるための L_2 正則化項 $\|\mathbf{\Theta}\|_2^2$ と，公平性を強化する偏見除去正則化項 $\mathrm{R}_{\mathsf{PR}}(Y, S)$ の 2 種類の正則化項を採用する．負の対数尤度関数にこれら二つの正則化項を加えたものが PR モデルの目的関数である．

$$\mathrm{loss}(\{\mathbf{w}^{(s)}\}; \mathcal{D}) = - \sum_s \mathcal{L}(\mathcal{D}_s) + \eta \, \mathrm{R}_{\mathsf{PR}}(Y, S) + \frac{\lambda}{2} \sum_s \|\mathbf{w}^{(s)}\|_2^2 \tag{8}$$

ただし，λ と η は正の正則化パラメータで，$\mathcal{L}(\cdot)$ は対数尤度関数である．

文献 [Kamishima 12] のモデルベース独立性を満たす PR 法の場合，$\hat{Y}$ と S の非独立性を測るためこれらの変数の相互情報量を用いた．

$$\mathrm{R}_{\mathsf{PR\text{-}MI}}(Y, S) = n \sum_{\hat{Y}, S} \Pr[\hat{Y}, S] \ln \frac{\Pr[\hat{Y}, S]}{\Pr[\hat{Y}] \Pr[S]} \tag{9}$$

なお，n 倍してあるのは，尤度項とオーダーを揃えるためである．式中の $\Pr[\hat{Y}, S]$ は式 (3) から導出でき，この分布 $\Pr[\hat{Y}, S]$ から他の分布 $Pr[\hat{Y}]$ と $\Pr[S]$ も導くことができる．この正則化項は解析的に微分可能なので，目的関数 (8) は効率的な勾配降下型の手法で最適化できる．このモデルを PR-MI と略記する．

この偏見削除正則化項を実独立性を満たすように修正する．そこで，式 (9) の $\Pr[\hat{Y}, S]$ を $\Pr[\tilde{Y}, S]$ と置き換えて，次式を得る．

$$\mathrm{R}_{\mathsf{PR\text{-}AI}}(Y, S) = n \sum_{\tilde{Y}, S} \Pr[\tilde{Y}, S] \ln \frac{\Pr[\tilde{Y}, S]}{\Pr[\tilde{Y}] \Pr[S]} \tag{10}$$

同時分布 $\Pr[\tilde{Y}, S]$ は，式 (5) と (6) から導出できる．このモデルを PR-AI と略記する．わずかな修正ではあるが，これにより

表 1: 通常のロジスティック回帰（LR）と公平ロジスティック回帰（PR-MI 法と PR-AI 法）の比較

Methods	Adult dataset			Dutch dataset		
	Acc	CVS	NMI	Acc	CVS	NMI
LR	0.862	0.170	4.36×10^{-02}	0.819	0.171	2.20×10^{-02}
PR-MI	0.822	0.055	1.81×10^{-02}	0.792	0.162	2.30×10^{-02}
PR-AI	0.825	0.008	6.03×10^{-05}	0.715	0.001	1.77×10^{-06}

3. 章のように公平性を大きく改善できる．しかし残念ながらこの偏見削除正則化項 $\mathrm{R}_{\text{PR-AI}}(\tilde{Y}, S)$ は，式 (6) に不連続な変換があるため微分できない．そのため，この目的関数を最適化するには勾配がなくても適用できる最適化手法を用いる必要がある．しかし，こうした手法では，パラメータ数を $|\boldsymbol{\Theta}|$ として，目的関数を $O(|\boldsymbol{\Theta}|^2)$ 回評価する（[Bishop 08] の 5.2.3 節などを参照）しなくてはならない．これは，勾配を用いる最適化手法の評価回数 $O(|\boldsymbol{\Theta}|)$ よりも多いため，このモデルの最適化は一般に非効率的である．

3. 公平ロジスティック回帰の性能評価

モデルベース独立性ではなく，実独立性を満たすようにすることで公平ロジスティック回帰の予測精度と公平性の間のトレードオフが改善されるかを検証する．

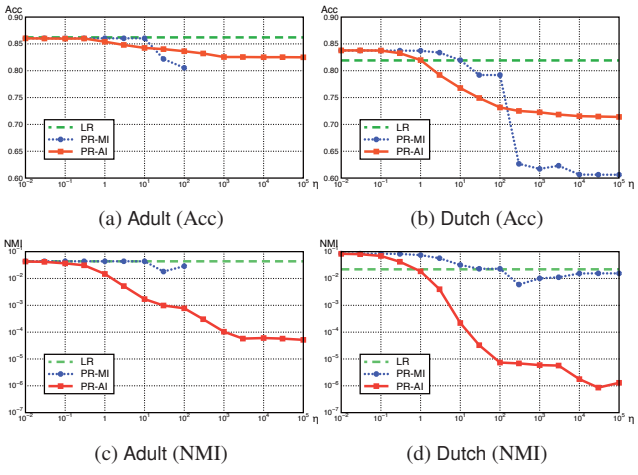
3.1 実験条件

実験に用いたベンチマークデータ[1]は文献 [Žliobaitė 11] で用いられたものである．一つ目は adult データ（別名 census income データ）であり，元データは URI レポジトリ [Frank 10] で配布されている．このデータ集合を Adult で参照する．クラス変数は個人の収入が高いかどうかの二値であり，センシティブ特徴は個人の性別である．データ数は 15,696 個，非センシティブな特徴数は 12 個で，どの特徴も離散である．二つ目は Dutch census で，これを Dutch で参照する．クラス変数は個人の職業が高収入のものか，そうでないかを表し，センシティブ特徴は個人の性別である．データ数は 60,420 個，非センシティブ特徴数は 10 個で，どの特徴も離散である．

5 分割の交差確認を行い，文献 [Kamishima 12] で用いた評価指標を求めた．公平ロジスティック回帰の性能評価のため，どれだけ正しくクラスラベルを予測できたかだけでなく，どれだけ厳密に公平性制約を満たすことができたかも評価する必要がある．なぜなら，予測精度と公平性はトレードオフの関係にあるからである．予測精度の評価には，正しくラベル付けできた標本の割合である正解率 (Acc) を用いた．正解率が高いほど，より正確にクラスが予測できている．公平性の評価には 2 種類の指標を用いた．一つ目は，S=1 で正ラベルになる割合から S=0 での正ラベルの割合を引いた CV スコア (CVS) で，0 に近づくほどクラス変数はセンシティブ特徴と独立になる．二つ目は，正規化相互情報量 (NMI) で，$\tilde{Y}$ と S の相互情報量を $[0, 1]$ の範囲になるように正規化したものである．NMI が小さくなると，より公平な決定がなされたことにになる．

3.2 実験結果

実験結果を表 1 に示す．LR は，センシティブ情報を取り除いた通常のロジスティック回帰である．なお，予測精度と公平性のトレードオフを調整するパラメータ η は，PR-MI では 3×10^1 に，PR-AI では 1×10^4 に設定した．



(a) Adult (Acc) (b) Dutch (Acc)

(c) Adult (NMI) (d) Dutch (NMI)

図 1: η に伴う予測精度 Acc と公平性 NMI の変化

NOTE: 横軸はパラメータ η，縦軸はキャプションに示した統計量である．緑の破線，青の丸付き点線，および赤の四角付き実線はそれぞれ LR，PR-MI，および PR-AI の結果である．Acc は大きいほどより正確な，MNI は小さいほどより公平な決定ができていることを示す．

まず，通常のロジスティック回帰（LR）と公平ロジスティック回帰（PR-MI と PR-AI）とを比較する．LR では，公平性指標 CVS と NMI に注目すると十分な公平性は達成できていない．このように単にセンシティブ情報をモデルから取り除くだけでは red-lining 効果のため公平な決定はできないことが分かる．それに対し，Dutch での PR-MI の場合を除き，公平ロジスティック回帰は通常のロジスティック回帰より公平な決定をしている．一方で，センシティブ特徴に含まれる情報を予測に使わないようにしているため，予測精度は低下している．

次に，モデルベース独立性の代わりに，実独立性を達成することの利点を検証する．公平性に関しては PR-AI が PR-MI よりどちらの指標でも非常に改善されている．表では PR-AI の予測精度は，Dutch では PR-MI より悪いが，Adult では良い．しかし，Dutch の場合でも，公平性が表の PR-MI と同等である η=3 の状況では PR-AI の予測精度は 0.792 であり，PR-MI と同等である．以上のことから，実独立性を達成することで，同等の公平性ではより予測精度の高い分類器が得られているといえる．

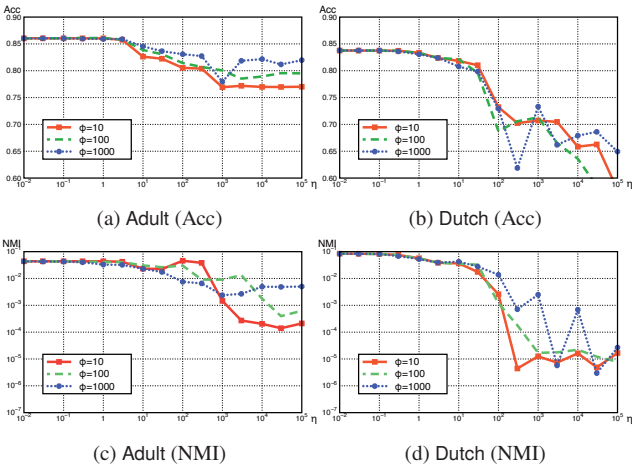
予測精度と公平性のトレードオフについてさらに検証する．釣り合いを調整するパラメータ η を変えたときの予測精度 Acc と公平性 NMI の変化を図 1 に示す．PR-MI では η が一定以上になると極端に予測精度が低下してしまい，それ以上は公平性を強化できなくなる問題生じるが，PR-AI ではそのような現象は見られず，η に応じて公平性は向上させることができる．

以上の実験結果をまとめておく．公平ロジスティック回帰は通常のロジスティック回帰より公平な決定ができるが，それに伴って予測精度はやや低下する．モデルベース独立性の代わりに実独立性を達成することで，同水準の公平性でより正確な予測が実現できる．

4. PR-AI モデルの最適化手法の改良

実験の結果，モデルベース独立性の代わりに実独立性を達成することで，より良い予測精度と公平性のトレードオフが実現できることが分かった．しかし，実独立性を達成する PR-AI

*1 https://sites.google.com/site/conditionaldiscrimination/



(a) Adult (Acc) (b) Dutch (Acc)

(c) Adult (NMI) (d) Dutch (NMI)

図 2: 平滑化を用いた緩和手法での予測精度 Acc と公平性 NMI の変化

NOTE: 横軸はパラメータ η，縦軸はキャプションに示した統計量である．赤の四角付き実線，緑の破線，および青の丸付き点線は，それぞれ ϕ が 10，100，および 1000 である場合の結果である．

モデルは 2.3 節の最後で述べたように，目的関数が微分できないため，効率的に最適化できない問題がある．ここでは，この目的関数を平滑な関数で近似して微分可能にすることで効率的に最適化する手法について予備的検討を行う．

4.1 平滑化した近似目的関数

前述のように，PR-AI モデルの目的関数（式 (8)）には勾配を用いた最適化手法を適用できない．これは，式 (10) には，不連続なステップ関数を含む式 (6) があるため，この目的関数は $\Pr[\hat{Y}{=}1|\mathbf{x}, s] = \Pr[\hat{Y}{=}0|\mathbf{x}, s]$ なる点で不連続になり，微分できないことが理由である．この問題を避けるため，この式 (6) のステップ関数をシグモイド関数で置き換えて平滑化する．しかし，この置き換えは $\Pr[\tilde{Y}|\mathbf{X}{=}x, S{=}s]$ を $\Pr[\hat{Y}|\mathbf{X}{=}x, S{=}s]$ に置き換えることと等価であり，$\mathrm{R}_{\text{PR-MI}}(Y, S)$ と等しい偏見除去正則化項になってしまい，明らかに無意味である．

そこで，$\Pr[\tilde{Y}|\mathbf{X}{=}x, S{=}s]$ をモデル化するときに，よりよくステップ関数を近似できるように，より急激に変化する関数を用いる．ここで，ϕ を大きな正定数とすれば，シグモイド関数 $\mathrm{sig}(\phi x)$ はより急激に変化するようになる．この修正したシグモイド関数を用いて，式 (6) を近似する．

$$\Pr[\tilde{Y}|\mathbf{X}{=}x, S{=}s] \approx$$
$$y\ \mathrm{sig}(\phi \mathbf{x}^\top \mathbf{w}^{(s)}) + (1-y)(1-\mathrm{sig}(\phi \mathbf{x}^\top \mathbf{w}^{(s)})) \quad (11)$$

ϕ が正の無限大であれば，式 (11) は (6) に等しくなるため，ϕ が大きな値である方が望ましい．一方で，大きすぎると計算中にオーバーフローを生じて計算できない．よって，ϕ はオーバーフローとならない程度に大きな値にする必要があり，この ϕ この調整が微妙である点が，この平滑化を用いた近似手法の短所である．一方で，目的関数は微分可能であるため，効率的な最適化手法を適用できる．

4.2 緩和手法の効果の検証実験

図 2 は，平滑化による緩和手法の予測精度と公平性の変化である．この図から得られる結果をまとめる．Adult では PR-MI と比べてあまり改善はされていないが，Dutch ではより公平な予測が実現できている．一方で，PR-AI と比べると，計算は速かったが，どちらのデータ集合でも同等の公平性では予測精度は悪く，これらの間のトレードオフは悪い．さらに，パラメータ ϕ の値に対して結果は大きく変動するためこれをうまく調整する必要があることや，特に ϕ が大きいときに指標の η に対する変化も不安定という問題点もある．効率的に実独立性を達成できる分類器を学習するには，ϕ を適応的に調整する手段が必要になるだろう．

5. まとめ

本稿では，公平配慮型分類問題でのモデルベース独立性と実独立性の概念を示し，モデルベース独立性の代わりに実独立性を達成することで予測精度と公平性よりよいトレードオフを実現できることを実験的に確認した．今後は，予備的検討を行った高速化について研究を進める予定である．

謝辞：本研究は JSPS 科研費 JP24500194，JP15K00327，および JP16H02864 の助成を受けた．

参考文献

[Bishop 08] Bishop, C. M.: パターン認識と機械学習 — ベイズ理論による統計的予測, 上下, シュプリンガー・ジャパン (2007–2008), [監訳：元田 浩 他；翻訳：神嶌 敏弘 他]

[Calders 10] Calders, T. and Verwer, S.: Three naive Bayes Approaches for Discrimination-free Classification, *Data Mining and Knowledge Discovery*, Vol. 21, pp. 277–292 (2010)

[Frank 10] Frank, A. and Asuncion, A.: UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences (2010), ⟨`http://archive.ics.uci.edu/ml`⟩

[Hardt 16] Hardt, M., Price, E., and Srebro, N.: Equality of Opportunity in Supervised Learning, in *Advances in Neural Information Processing Systems 29* (2016)

[Kamishima 12] Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J.: Fairness-aware Classifier with Prejudice Remover Regularizer, in *Proc. of the ECML PKDD 2012, Part II*, pp. 35–50 (2012), [LNCS 7524]

[Kamishima 18] Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J.: Model-based and Actual Independence for Fairness-aware Classification, *Data Mining and Knowledge Discovery*, Vol. 32, pp. 258–286 (2018)

[Zafar 17] Zafar, M. B., Valera, I., Rogriguez, M. G., and Gummadi, K. P.: Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment, in *Proc. of the 26th Int'l Conf. on World Wide Web*, pp. 1171–1180 (2017)

[Žliobaitė 11] Žliobaitė, I., Kamiran, F., and Calders, T.: Handling Conditional Discrimination, in *Proc. of the 11th IEEE Int'l Conf. on Data Mining* (2011)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

正準変量のカーネルマッチングによるデータ融合法

Data Fusion Method with Kernel Matching for Estimated Canonical Variables

光廣 正基 *1*2　　星野 崇宏 *3*4
Masaki MITSUHIRO　　Takahiro HOSHINO

*1（株）日経リサーチ
Nikkei Research Inc.

*2慶應義塾大学大学院経済学研究科
Graduate School of Economics, Keio University

*3慶應義塾大学経済学部
Department of Economics, Keio University

*4理化学研究所革新知能統合研究センター
RIKEN Center for Advanced Intelligence Project

To clarify the relationship between variables obtained from different sources, we integrate two source data by statistical matching. In this method, missing values of certain data are predicted from covariates and observed values of another data. However, with a large number of covariates, it is difficult to find objects with the same or close covariates. For this reason, statistical matching after reducing covariates in the low-dimensional subspace is useful. In this manuscript, we propose a data fusion method combining kernel canonical correlation analysis and kernel matching. This proposed method can fusion two data sets by kernel matching for estimated canonical variables in the common space reflecting the relationship between covariates and outcome variables.

1. はじめに

回答者や調査モードなどが異なる複数の多変量データを統合する方法として，統計的マッチングやモデルベースの手法がある [Kamakura 97]．中でも，統計的マッチングは，あるデータの欠測しているアウトカム変数に，そのアウトカム変数が観測されているもう一方のデータから，共変量が同じもしくは一番近い個人の値を割り当てる方法であり，共変量が複数ある場合はマハラノビスマッチング法がよく使用される．

マハラノビスマッチング法は，共変量の数が多い場合は次元の呪いの影響を受けることが知られており [Rubin 78] [Stuart 10]，この欠点を解決するため，マッチングする前に共変量を次元縮約することが多い．共変量の縮約方法には，正準相関分析法 [Hotelling 36] が挙げられるが，2 つのデータセットの個人が同一である必要があったり [Yu 13]，個人が異なっていたとしても教師や事前情報により個人間の関係がわかっている必要がある [Yan 07] [Shimodaira 14]．さらに，カーネル正準相関分析法 [Akaho 06] を使用すれば，共変量とアウトカム変数の非線形な関係も捉えることが可能である．

また，欠測値にはもう一方のデータで観測された値しか代入できないため，2 つのデータの共変量に偏りがある場合に，マハラノビスマッチング法の予測精度が落ちてしまう．そこで，カーネルマッチング法により，各個人との重みを推定し，アウトカム変数の重み付き平均を計算することで，観測していない値も代入可能となる [Fan 93] [Heckman 98]．

本研究では，個人の異なる 2 つの多変量データの統合するため，カーネル正準相関分析法とカーネルマッチング法を組み合わせたデータ融合法を提案する．この手法は，各データの共変量とアウトカム変数の非線形な関係を反映した共通の低次元空間に共変量を縮約し，推定された正準変量に対して，カーネルマッチング法を適用することで，マハラノビスマッチング法の欠点を改善する．

連絡先: 光廣正基，（株）日経リサーチ，東京都千代田区内神田 2-2-1 鎌倉河岸ビル，masaki_mitsuhiro@nikkei-r.co.jp

2. 提案手法

ソースの異なる 2 つのデータ行列の一方を $\boldsymbol{X}_1 = (x_{ij1})$ $(i = 1, 2, \cdots, n_1; j = 1, 2, \cdots, p^{\mathrm{cov}} + p^{\mathrm{out}})$ とする．さらに $\boldsymbol{X}_1$ は共変量 $\boldsymbol{Z}_1 = (z_{ij1})$ $(i = 1, 2, \cdots, n_1; j = 1, 2, \cdots, p^{\mathrm{cov}})$ と，観測されているアウトカム変数 $\boldsymbol{Y}_{11} = (y_{ij11})$ $(i = 1, 2, \cdots, n_1; j = 1, 2, \cdots, p_1^{\mathrm{out}})$，欠測しているアウトカム変数 $\boldsymbol{Y}_{12}^* = (y_{ij12}^*)$ $(i = 1, 2, \cdots, n_1; j = 1, 2, \cdots, p_2^{\mathrm{out}})$ に分割できる．もう一方のデータ行列も同様に定義でき，2 つのデータ行列を統合すると $\boldsymbol{X} = (x_{ij})$ $(i = 1, 2, \cdots, n; j = 1, 2, \cdots, p^{\mathrm{cov}} + p^{\mathrm{out}})$ となる．この統合データを作成するために，欠測している $\boldsymbol{Y}_{12}^*, \boldsymbol{Y}_{21}^*$ に推定値となる $\hat{\boldsymbol{Y}}_{12}, \hat{\boldsymbol{Y}}_{21}$ をそれぞれ代入し，$\boldsymbol{X}$ は欠測のない完全データとなる．

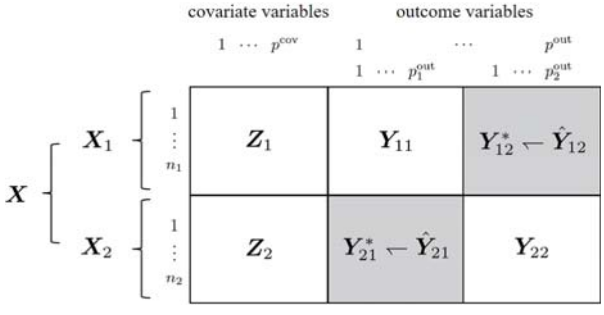


Figure 1: ソースの異なる 2 つの多変量データ $\boldsymbol{X}_1, \boldsymbol{X}_2$

データ行列 $\boldsymbol{X}_1$ における $\boldsymbol{Z}_1$ と $\boldsymbol{Y}_{11}$ のカーネル正準相関分析と，データ行列 $\boldsymbol{X}_2$ における $\boldsymbol{Z}_2$ と $\boldsymbol{Y}_{22}$ のカーネル正準相関分析を同時に行い，共通の低次元空間で推定される共変量側の正準変量に対して，カーネルマッチング法を適用するデータ融合法を提案する．この手法により，個体が異なるデータ同士の統合が可能となる．

2.1 STEP1: 次元縮約

非線形関数 $\phi_{\boldsymbol{z}}(\cdot), \phi_{\boldsymbol{y}_1}(\cdot), \phi_{\boldsymbol{y}_2}(\cdot)$ を用いて，データ行列 $\boldsymbol{X}_1, \boldsymbol{X}_2$ の共変量とアウトカム変数を高次元特徴空間に写像し，特徴ベクトルを抽出する．

$$\begin{aligned}\boldsymbol{z}_i &\mapsto \phi_{\boldsymbol{z}}(\boldsymbol{z}_i),\\ \boldsymbol{y}_{i11} &\mapsto \phi_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11}),\\ \boldsymbol{y}_{i22} &\mapsto \phi_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22}).\end{aligned}$$

推定したい共変量側の正準変量を $u(\boldsymbol{z}_i)$，アウトカム変数側の正準変量をそれぞれ $v(\boldsymbol{y}_{i11}), v(\boldsymbol{y}_{i22})$ とする．このとき，中心化した特徴ベクトル $\tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_i), \tilde{\phi}_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11}), \tilde{\phi}_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22})$，および中心化したカーネル関数 $\tilde{k}_{\boldsymbol{z}}(\boldsymbol{z}_i,\cdot), \tilde{k}_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11},\cdot), \tilde{k}_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22},\cdot)$ を用いて，リプレゼンター定理より，各正準変量は下記の通りである．

$$\begin{aligned}u(\boldsymbol{z}_i) &= \tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_i)^{'}\boldsymbol{A}_i\\ &= \tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_i)^{'}\sum_{j=1}^{n_k}\tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_j)\boldsymbol{c}_{\boldsymbol{A}j}^{'}\\ &= \sum_{j=1}^{n_k}\tilde{k}_{\boldsymbol{z}}(\boldsymbol{z}_i,\boldsymbol{z}_j)\boldsymbol{c}_{\boldsymbol{A}j}^{'},\\ v(\boldsymbol{y}_{i11}) &= \tilde{\phi}_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11})^{'}\boldsymbol{B}_{i1}\\ &= \tilde{\phi}_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11})^{'}\sum_{j=1}^{n_1}\tilde{\phi}_{\boldsymbol{y}_1}(\boldsymbol{y}_{j11})\boldsymbol{c}_{\boldsymbol{B}_1 j}^{'}\\ &= \sum_{j=1}^{n_1}\tilde{k}_{\boldsymbol{y}_1}(\boldsymbol{y}_{i11},\boldsymbol{y}_{j11})\boldsymbol{c}_{\boldsymbol{B}_1 j}^{'},\\ v(\boldsymbol{y}_{i22}) &= \tilde{\phi}_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22})^{'}\boldsymbol{B}_{i2}\\ &= \tilde{\phi}_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22})^{'}\sum_{j=1}^{n_2}\tilde{\phi}_{\boldsymbol{y}_2}(\boldsymbol{y}_{j22})\boldsymbol{c}_{\boldsymbol{B}_2 j}^{'}\\ &= \sum_{j=1}^{n_2}\tilde{k}_{\boldsymbol{y}_2}(\boldsymbol{y}_{i22},\boldsymbol{y}_{j22})\boldsymbol{c}_{\boldsymbol{B}_2 j}^{'}.\end{aligned}$$

ここで，$\boldsymbol{A} = (\boldsymbol{A}_1^{'}, \boldsymbol{A}_2^{'}, \cdots, \boldsymbol{A}_n^{'})^{'}$ は共変量から得られる高次元特徴ベクトルの負荷量を表し，$\boldsymbol{B}_1 = (\boldsymbol{B}_{11}^{'}, \boldsymbol{B}_{2k}^{'}, \cdots, \boldsymbol{B}_{n_1 1}^{'})^{'}$ や $\boldsymbol{B}_2 = (\boldsymbol{B}_{12}^{'}, \boldsymbol{B}_{22}^{'}, \cdots, \boldsymbol{B}_{n_2 2}^{'})^{'}$ はアウトカム変数から得られる高次元特徴ベクトルの負荷量である．さらに，$\boldsymbol{C_A} = (\boldsymbol{c}_{\boldsymbol{A}1}^{'}, \boldsymbol{c}_{\boldsymbol{A}2}^{'}, \cdots, \boldsymbol{c}_{\boldsymbol{A}n}^{'})^{'}$ は共変量側のカーネル関数に対する負荷量であり，$\boldsymbol{C}_{\boldsymbol{B}_1} = (\boldsymbol{c}_{\boldsymbol{B}_1 1}^{'}, \boldsymbol{c}_{\boldsymbol{B}_1 2}^{'}, \cdots, \boldsymbol{c}_{\boldsymbol{B}_1 n_1}^{'})^{'}$ と $\boldsymbol{C}_{\boldsymbol{B}_2} = (\boldsymbol{c}_{\boldsymbol{B}_2 1}^{'}, \boldsymbol{c}_{\boldsymbol{B}_2 2}^{'}, \cdots, \boldsymbol{c}_{\boldsymbol{B}_2 n_2}^{'})^{'}$ はアウトカム変数側のカーネル関数に対する負荷量である．

共変量側の中心化した特徴ベクトルをデザイン行列 $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{Z}} = (\tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_1)^{'}, \tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_2)^{'}, \cdots, \tilde{\phi}_{\boldsymbol{z}}(\boldsymbol{z}_n)^{'})^{'}$ で表し，アウトカム変数側も同様とする．このとき，高次元特徴ベクトルの負荷量は，データ次元の負荷量 $\boldsymbol{C_A}, \boldsymbol{C}_{\boldsymbol{B}_1}, \boldsymbol{C}_{\boldsymbol{B}_2}$ を用いて，下記のように表す．

$$\begin{aligned}\boldsymbol{A} &= \tilde{\boldsymbol{\Phi}}_{\boldsymbol{Z}}\boldsymbol{C_A},\\ \boldsymbol{B}_1 &= \tilde{\boldsymbol{\Phi}}_{\boldsymbol{Y}_1}\boldsymbol{C}_{\boldsymbol{B}_1}\\ \boldsymbol{B}_2 &= \tilde{\boldsymbol{\Phi}}_{\boldsymbol{Y}_2}\boldsymbol{C}_{\boldsymbol{B}_2}.\end{aligned}$$

高次元な負荷量を特徴ベクトルの内積となる中心化グラム行列を $\tilde{\boldsymbol{K}}_{\boldsymbol{Z}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{Z}}^{'}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{Z}}, \tilde{\boldsymbol{K}}_{\boldsymbol{Y}_k} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{Y}_k}^{'}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{Y}_k}$ とする．共変量側の中心化グラム行列 $\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}$ とアウトカム変数側の中心化グラム行列 $\tilde{\boldsymbol{K}}_{\boldsymbol{Y}} = \mathrm{diag}(\tilde{\boldsymbol{K}}_{\boldsymbol{Y}_1}, \tilde{\boldsymbol{K}}_{\boldsymbol{Y}_2})$ が与えられたとき，データ行列 $\boldsymbol{X}_1$ でのカーネル正準相関分析とデータ行列 $\boldsymbol{X}_2$ でのカーネル正準相関分析を同時に行い，共変量側の正準変量を共通の低次元空間に縮約する．このとき，アウトカム変数側の負荷量を $\boldsymbol{C_B} = (\boldsymbol{C}_{\boldsymbol{B}_1}^{'}, \boldsymbol{C}_{\boldsymbol{B}_2}^{'})^{'}$ とすると，各パラメータを推定する目的関数は下記の通りである．

$$\begin{aligned}&f(\boldsymbol{C_A}, \boldsymbol{C_B}|\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}, \tilde{\boldsymbol{K}}_{\boldsymbol{Y}}, \lambda_{\boldsymbol{A}}, \boldsymbol{\Lambda_B})\\ &= \left\|\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}\boldsymbol{C_A} - \begin{pmatrix}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}_{11}}\boldsymbol{C}_{\boldsymbol{B}_1}\\ \tilde{\boldsymbol{K}}_{\boldsymbol{Y}_{22}}\boldsymbol{C}_{\boldsymbol{B}_2}\end{pmatrix}\right\|^2\\ &+ \lambda_{\boldsymbol{A}}\boldsymbol{C}_{\boldsymbol{A}}^{'}\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}\boldsymbol{C_A} + \boldsymbol{\Lambda_B}\begin{pmatrix}\boldsymbol{C}_{\boldsymbol{B}_1}^{'}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}_1}\boldsymbol{C}_{\boldsymbol{B}_1}\\ \boldsymbol{C}_{\boldsymbol{B}_2}^{'}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}_2}\boldsymbol{C}_{\boldsymbol{B}_2}\end{pmatrix},\end{aligned} \tag{1}$$

subject to

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{A}}^{'}(\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}^2 + \lambda_{\boldsymbol{A}}\tilde{\boldsymbol{K}}_{\boldsymbol{Z}})\boldsymbol{C_A} &= \boldsymbol{I}_d,\\ \boldsymbol{C}_{\boldsymbol{B}}^{'}(\tilde{\boldsymbol{K}}_{\boldsymbol{Y}}^2 + \boldsymbol{\Lambda_B}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}})\boldsymbol{C_B} &= \boldsymbol{I}_d.\end{aligned}$$

ここで，$\lambda_{\boldsymbol{A}}, \lambda_{\boldsymbol{B}_1}, \lambda_{\boldsymbol{B}_2}$ は正則化パラメータであり，アウトカム変数側の正則化パラメータは $\boldsymbol{\Lambda_B} = \mathrm{diag}(\lambda_{\boldsymbol{B}_1}\boldsymbol{I}_{\mathrm{n}_1}, \lambda_{\boldsymbol{B}_2}\boldsymbol{I}_{\mathrm{n}_2})$ とする．また，$\boldsymbol{I}_d$ は d 次元の単位行列であり，正準変量に直交制約が課せられる．

上記の目的関数を制約条件の下で最小化するため，特異値分解を行う．

$$(\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}^2 + \lambda_{\boldsymbol{A}}\boldsymbol{K}_{\boldsymbol{Z}})^{-\frac{1}{2}}\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}}(\tilde{\boldsymbol{K}}_{\boldsymbol{Y}}^2 + \boldsymbol{\Lambda_B}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}})^{-\frac{1}{2}} = \boldsymbol{S}\boldsymbol{\Xi}\boldsymbol{R}^{'},$$

ここで，$\boldsymbol{S}^{'}\boldsymbol{S} = \boldsymbol{R}^{'}\boldsymbol{R} = \boldsymbol{I}_n$ は特異ベクトル，$\boldsymbol{\Xi} = \mathrm{diag}(\xi_1, \xi_2, \cdots, \xi_{\mathrm{n}})$ は特異値であり，ξ_i $(i = 1, 2, \cdots, n)$ は降順に並べられている．特異値分解の結果を用いて，パラメータの推定式は下記の通りである．

$$\begin{aligned}\hat{\boldsymbol{C}}_{\boldsymbol{A}} &= (\tilde{\boldsymbol{K}}_{\boldsymbol{Z}}^2 + \lambda_{\boldsymbol{A}}\tilde{\boldsymbol{K}}_{\boldsymbol{Z}})^{-\frac{1}{2}}\boldsymbol{S}_d,\\ \hat{\boldsymbol{C}}_{\boldsymbol{B}} &= (\tilde{\boldsymbol{K}}_{\boldsymbol{Y}}^2 + \boldsymbol{\Lambda_B}\tilde{\boldsymbol{K}}_{\boldsymbol{Y}})^{-\frac{1}{2}}\boldsymbol{R}_d.\end{aligned} \tag{2}$$

ここで，$\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ と $\hat{\boldsymbol{C}}_{\boldsymbol{B}}$ は各グラム行列の d 次元の負荷量の推定値を表す．

2.2 STEP2: マッチング

推定した $\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ から，データ行列 $\boldsymbol{X}_1$ の共変量側の正準変量 $\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_1}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ とデータ行列 $\boldsymbol{X}_2$ の共変量側の正準変量 $\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_2}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ を求め，これらに統計的マッチングを適用し，データ行列 $\boldsymbol{X}_1, \boldsymbol{X}_2$ の欠測を補完する．

実際には，$\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_1}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ をレシピエントデータの共変量，$\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_2}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ をドナーデータの共変量として，レシピエントデータのアウトカム変数 $\hat{\boldsymbol{Y}}_{12}$ を推定する．一方，レシピエントとドナーを入れ替え，$\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_2}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ をレシピエントデータの共変量，$\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_1}\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ をドナーデータの共変量として，レシピエントデータのアウトカム変数 $\hat{\boldsymbol{Y}}_{21}$ を推定する．

$\hat{\boldsymbol{C}}_{\boldsymbol{A}}$ と $\hat{\boldsymbol{C}}_{\boldsymbol{B}}$ において，次元数が $d > 1$ であれば，マハラノビスマッチング法などが適用でき，$d = 1$ であれば，最近傍マッチング法やカーネルマッチング法などが適用できる．本研究では，カーネルマッチング法を使用する．

3. シミュレーション

提案手法と従来手法の欠測しているアウトカム変数の予測精度の比較を行う．提案手法は，推定する正準変量を 1 次元に設定し，カーネル正準相関分析法によって得られた共変量側の正準変量に対して，カーネルマッチング法を適用する．従来手法には多変量データのマッチング手法で代表的なマハラノビスマッチング法を用いる．

3.1 事前設定

STEP1 のカーネル正準相関分析法において，カーネル関数はガウスカーネルを用い，バンド幅は入力値の 2 乗誤差の中央値とする．正則化パラメータは 5-fold クロスバリデーションによって決定する．正則化パラメータの候補は $(0.1, 0.5, 1.0)$ とし，この中から最適な値の組を選ぶとする．

データ行列を g 個に分割したとき，その一つをテストデータとし,，残りの $(g-1)$ 個をトレーニングデータとする．$\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_1}^{(g)}$ や $\tilde{\boldsymbol{K}}_{\boldsymbol{Y}_1}^{(g)}$ はテストデータの中心化グラム行列を表し，$\hat{\boldsymbol{C}}_{\boldsymbol{A}}^{(-g)}$ や $\hat{\boldsymbol{C}}_{\boldsymbol{A}}^{(-g)}$ はトレーニングデータの推定値を表す．クロスバリデーションの評価関数は正準変量間の誤差を用い [Hardoon 09]，データ行列 $\boldsymbol{X}_1$ の正準変量間の誤差とデータ行列 $\boldsymbol{X}_2$ の正準変量間の誤差の平方和の 5 回平均とする．

$$\varepsilon(\lambda_{\boldsymbol{A}}, \boldsymbol{\Lambda_B}) = \frac{1}{5}\sum_{g=1}^{5}\Big(||\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_1}^{(g)}\hat{\boldsymbol{C}}_{\boldsymbol{A}}^{(-g)} - \tilde{\boldsymbol{K}}_{\boldsymbol{Y}_1}^{(g)}\hat{\boldsymbol{C}}_{\boldsymbol{A}}^{(-g)}||^2 + ||\tilde{\boldsymbol{K}}_{\boldsymbol{Z}_2}^{(g)}\hat{\boldsymbol{C}}_{\boldsymbol{B}}^{(-g)} - \tilde{\boldsymbol{K}}_{\boldsymbol{Y}_2}^{(g)}\hat{\boldsymbol{C}}_{\boldsymbol{B}}^{(-g)}||^2\Big).$$

また，STEP2 のマッチング法の設定として，カーネルマッチング法ではガウスカーネルを用いてアウトカム変数を予測する．共変量が同じもしくは近い個人を決定するマハラノビスマッチング法では，1 度マッチングされたドナーデータの個人は何回も利用できるとする．

3.2 シミュレーションデータ

50 個のシミュレーションデータ $\boldsymbol{X}$ を作成し，2 つのデータ行列に分割した後，各データ行列で欠測変数を作成する．共変量はアウトカム変数と非線形な関係を持つような 2 変数と 10 変数の $N(0, 1^2)$ に従う正規ノイズから作られる．2 つの共変量の生成は下記の通りである．

$$\begin{aligned}
z_{i1} &= \theta + \varepsilon_{i1}^{(z)},\\
z_{i2} &= 5 + 2\theta + 0.8\theta^2 - 3\theta^3 + \varepsilon_{i2}^{(z)},\\
\theta &\overset{\text{i.i.d.}}{\sim} U(-1, 1),\\
\varepsilon_{ij}^{(z)} &\overset{\text{i.i.d.}}{\sim} N(0, 0.1^2)\\
&(i = 1, 2, \cdots, n; j = 1, 2).
\end{aligned}$$

続いて，共変量の θ を用いて，アウトカム変数を 4 変数作成する．アウトカム変数の生成は下記の通りである．

$$\begin{aligned}
y_{i1} &= 1 + \theta + 5\theta^2 + \varepsilon_{i1}^{(y)},\\
y_{i2} &= 1 - 3\theta + 3\theta^2 + \varepsilon_{i2}^{(y)},\\
y_{i3} &= 2 + \theta + \theta^4 + \varepsilon_{i3}^{(y)},\\
y_{i4} &= \theta^2 - 0.5\theta^3 + \varepsilon_{i4}^{(y)},\\
\varepsilon_{ij}^{(y)} &\overset{\text{i.i.d.}}{\sim} N(0, 0.1^2)\\
&(i = 1, 2, \cdots, n; j = 1, 2, 3, 4).
\end{aligned}$$

生成した 1000×16 のデータ行列 $\boldsymbol{X}$ を 2 分割する．分割方法は，ロジスティック関数で得られる確率 $p_i = 1/(1+e^{-\theta})$ に従って，$n_1 = 500$ の $\boldsymbol{X}_1$，残りを $\boldsymbol{X}_2$ とし，$n_2 = n - n_1 = 500$ となる．$\boldsymbol{X}_1$ については，y_{i3}, y_{i4} を欠測させ，$\boldsymbol{X}_2$ については，y_{i1}, y_{i2} を欠測させる．この 4 つのアウトカム変数の欠測の値を予測する．

3.3 比較結果

50 個のシミュレーションデータに対して，提案手法と従来手法を適用し比較する．まず，提案手法とマハラノビスマッチング法の相関係数を比較する．提案手法の真値と推定値との相関係数とマハラノビスマッチング法の真値と推定値の相関係数を計算し，差分をとる．

$$\rho_j = \mathrm{corr}(\boldsymbol{y}_j, \hat{\boldsymbol{y}}_j) - \mathrm{corr}_{\mathrm{MM}}(\boldsymbol{y}_j, \hat{\boldsymbol{y}}_j).$$

各変数の差分が正の場合 $(\rho_j > 0)$ は提案手法の予測精度が良く，負の場合 $(\rho_j < 0)$ は提案手法の予測精度が従来手法よりも悪いと評価する．Figure 2 は相関係数の比較結果である．Figure 2 から，提案手法は 4 つのアウトカム変数すべてにおいて，相関係数の差分が正に大きく，従来手法のマハラノビスマッチング法よりも予測精度が改善されている．

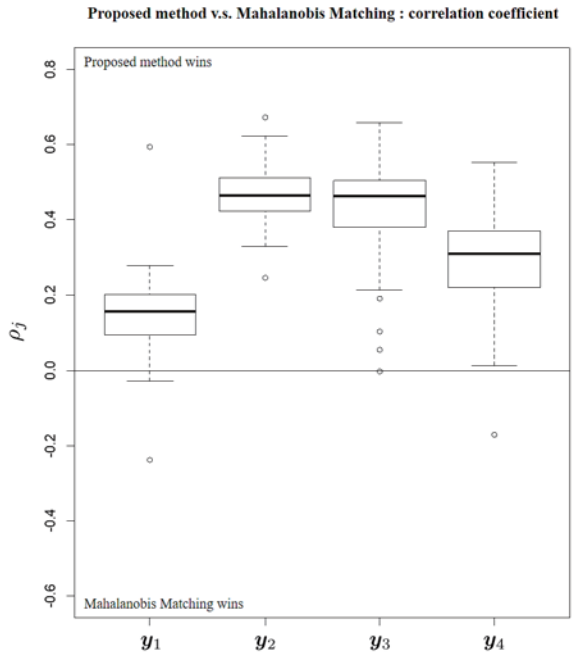


Figure 2: 提案手法とマハラノビスマッチング法の相関係数の差分 ρ_j

続いて，提案手法とマハラノビスマッチング法の平均二乗誤差 (MSE) を比較する．各アウトカム変数に対して，MSE を計算し，マハラノビスマッチング法との相対比を計算する．MSE の比が 1 より小さい場合は提案手法の予測精度が良く，1 より大きい場合は提案手法の予測精度が従来手法よりも悪いと評価する．Figure 3 は MSE の比較結果である．Figure 3 から，提案手法は 4 つのアウトカム変数すべてが 1 より小さく，従来手法のマハラノビスマッチング法よりも予測精度が改善されている．

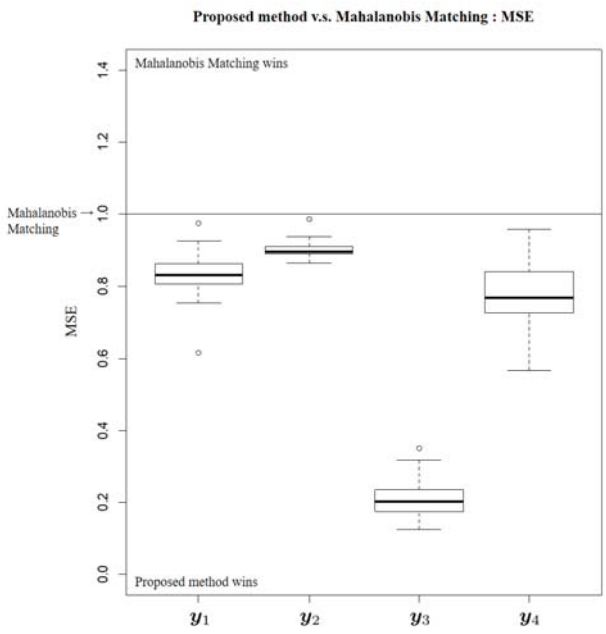


Figure 3: 提案手法とマハラノビスマッチング法の MSE の相対比

4. おわりに

カーネル正準相関分析法とカーネルマッチング法を組み合わせたデータ融合法を提案した．2 つのデータ行列のカーネル正準相関分析を同時に行い，共変量側の正準変量については，共通の低次元空間に縮約した．推定された正準変量に対して，カーネルマッチング法を行い，ドナーデータにない値もレシピエントデータのアウトカム変数に代入できるようにした．シミュレーションデータを用いて，従来手法のマハラノビスマッチング法と比較し，提案手法のほうが相関係数が大きく，MSE が小さいという結果となり，予測精度の改善がみられた．

参考文献

[Akaho 06] Akaho, S. (2006). A kernel method for canonical correlation analysis. arXiv preprint cs/0609071.

[Fan 93] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 196–216.

[Hardoon 09] Hardoon, D. R. and Shawe-Taylor, J. (2009). Convergence analysis of kernel canonical correlation analysis: theory and practice. *Machine learning*, **74**, 23–38.

[Heckman 98] Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *The review of economic studies*, **65**, 261–294.

[Hotelling 36] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321–377.

[Kamakura 97] Kamakura, W. A. and Wedel, M. (1997). Statistical data fusion for cross-tabulation. *Journal of Marketing Research*, 485–498.

[Rubin 78] Rubin, D. B. (1978). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *ETS Research Report Series*, **1978**.

[Shimodaira 14] Shimodaira, H. (2014). A simple coding for cross-domain matching with dimension reduction via spectral graph embedding. arXiv:1412.8380.

[Stuart 10] Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, **25**, 1.

[Yan 07] Yan, S., Xu, D., Zhang, B., Zhang, H. J., Yang, Q., and Lin, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE transactions on pattern analysis and machine intelligence*, **29**, 40–51.

[Yu 13] Yu, S., Tranchevent, L. C., De Moor, B., and Moreau, Y. (2013). *Kernel-based data fusion for machine learning*. Springer-Verlag Berlin An.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

相対密度比を用いたマルチラベルロジスティック回帰について

Multi-label Logistic Regression Using Relative Density Ratio

岡部 格明 *1　土田 潤 *2　宿久 洋 *3
Masaaki Okabe　Jun Tsuchida　Hiroshi Yadohisa

*1同志社大学大学院文化情報学研究科
Graduate School of Cluture and Infomation Science, Doshisha University

*2東京理科大学工学部
Faculty of Engineering, Tokyo University of Science

*3同志社大学文化情報学部
Faculty of Culture and Information Science, Doshisha University

Multi-label classification is a supervised learning problem where multiple labels may be assigned to each instance. The main baseline for multi-label classification is binary relevance method, which is estimate the binary classification model for each label. In binary classification, there are cases where poor results are data when the class is imbalance.

In this paper, we propose a multi-label classification model used relative density ratio. In this model, we used relative F-measure by relative density ratio for weight of error function to solve the class imbalance problem.

1. はじめに

機械学習における分類問題のひとつにマルチラベル分類問題がある．マルチラベルとは，各対象に複数のラベルが付与されているような状況であり，マルチラベル分類問題は，対象ごとに付与されるであろうラベルを推定するような問題である．この問題は，文書や画像の分類などに応用されている．

マルチラベル分類問題を解決するための方法には，各ラベルを一つのクラスとしてみなし，マルチクラスの問題として解く，Label Power Set 法 [Tsoumakas 07] と各ラベルをそれぞれのクラスに属するかそうでないかを 2 値分類の問題として解く，Binary relevance 法 [Boutell 04] がある．Label Power Set 法は，データ内に存在しているラベルの組み合わせに対して，一つのラベルを割り当てるため，ラベル同士の関係性を反映した判別を実現することができる．一方で，データに存在しないラベルの組み合わせに対しては，対応できないという欠点がある．また，ラベルの種類が 2 の累乗の形になってしまうため，対象の数が少ないデータセットに適用したときに，過適合を起こす場合がある．Binary relevance 法は，それぞれのラベルのみを考慮するため，ラベル間の関係性に着目することができないが，計算コストが少ないという利点がある．

一般に，マルチラベル分類問題では，データの数に比べ，あるラベルが付与されている割合が極端に小さいという，ラベルのクラス不均衡の問題が起こりやすいと言われている [Sun 13]. このような場合に，評価指標として単純な正解率を用いると，全てのデータにラベルが付いていないという予測結果が得られたとしても，高い正解率を得てしまう場合がある．正解率のほかの評価指標として F-measure があり，この指標は，先ほどのような状況に対して,「悪い」判別であることを表すという性質がある．F-measure の算術平均が macro F-measure であり，ある特定のクラスの判別が「悪い」場合にも対応する指標である．

そこで，本発表では，Binary relebance 法による，近似された macro F-measure を重みとして用いた，ロジスティック回帰を提案する．この手法では，各ラベルの混同行列における関係から，相対 F-measure を相対密度比によって近似する．それを誤差関数の重みとして用いることで，正解率だけでなく，F-measure を考慮した最大化問題を解くことになり，ラベルのクラス不均衡の問題を解決する．

連絡先: 岡部格明，同志社大学大学院文化情報学研究科，京都府京田辺市多々良京都谷 1-3，0774-65-7657，mokab.0328@gmail.com

2. 問題設定と評価指標

データ行列 $\boldsymbol{X} = (\boldsymbol{x}_1,..,\boldsymbol{x}_i,..,\boldsymbol{x}_N)' \in \mathbb{R}^{N\times D}$，各対象のラベル行列 $\boldsymbol{Y} = (\boldsymbol{y}_1,..,\boldsymbol{y}_i,..,\boldsymbol{y}_N)' \in \{0,1\}^{N\times C}$ が与えられたとする．Binary relebance 法では，対象 i のデータ $\boldsymbol{x}_i$ が与えられたとき，その対象 i がラベル c を持つか否かを表す関数 $f_c : \mathbb{R}^D \to \{0,1\}$ $(c = 1,...,C)$ を推定することを目的とする．

また，推定結果の評価指標として，各ラベル c に属するか否かを表した，混同行列 (表 1) を得る．混同行列の各要素 $\mathrm{TP}(c)$, $\mathrm{FN}(c)$, $\mathrm{FP}(c)$, $\mathrm{TN}(c)$ はそれぞれ，$y_{ic} = 1$ であり，$f_c(\boldsymbol{x}_i) = 1$ であった個数，$y_{ic} = 1$ であり，$f_c(\boldsymbol{x}_i) = 0$ であった個数，$y_{ic} = 0$ であり，$f_c(\boldsymbol{x}_i) = 1$ であった個数，$y_{ic} = 0$ であり，$f_c(\boldsymbol{x}_i) = 0$ であった個数を表している． このとき，ラベル c に対する

表 1: ラベル c に対する混同行列

		予測ラベル		合計
		c	c でない	
正解ラベル	c	$\mathrm{TP}(c)$	$\mathrm{FN}(c)$	n_c
	c でない	$\mathrm{FP}(c)$	$\mathrm{TN}(c)$	$N - n_c$
合計		m_c	$N - m_c$	N

Precision(再現率) と Recall(適合率) は次のように定義される．

$$\mathrm{Precision}(c) = \frac{\mathrm{TP}(c)}{\mathrm{TP}(c)+\mathrm{FP}(c)}, \quad \mathrm{Recall}(c) = \frac{\mathrm{TP}(c)}{\mathrm{TP}(c)+\mathrm{FN}(c)}.$$

一般に，Precision と Recall はトレードオフの関係にあるため，総合指標として，Precision と Recall に対して重みをつけた調和平均である，相対 F-measure という指標がつぎのように定義される．

$$\mathrm{F\text{-}measure}(\alpha, c) = \left(\alpha \frac{1}{\mathrm{Recall}(c)} + (1-\alpha)\frac{1}{\mathrm{Precision}(c)}\right)^{-1} \tag{1}$$

ただし，$0 \leq \alpha \leq 1$ である．

マルチラベル問題としての総合的な判別の評価の指標はそれぞれの指標の平均が用いられる．中でも，(1) 式の算術平均をとったものが macro F-measure であり，つぎのように定義される．

$$\text{F-measure}_{\text{macro}}(\boldsymbol{\alpha}) = \frac{1}{C}\sum_{c=1}^{C} \text{F-measure}(\alpha_c, c)$$

macro F-measure は，出現頻度の少ないラベルがあるときに，小さな値を取りやすいため，判別全体の「よさ」をうまく測ることができる．

3. 提案手法

データ行列 $\boldsymbol{X} \in \mathbb{R}^{N\times D}$，ラベル行列 $\boldsymbol{Y} = (\boldsymbol{y}_i) \in \{0,1\}^{N\times C}$ が与えられたとする．

このとき，各ラベルに対して，ロジスティック回帰モデルをつくるため，対象 i にラベル c が付く確率を次のように表す．

$$\text{P}(y_{ic} = 1|\boldsymbol{x}_i) = p_{ic} = \frac{1}{1+\exp(-\boldsymbol{\beta}_c'\boldsymbol{x}_i)}. \quad (2)$$

ここで，$\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_C) \in \mathbb{R}^{D\times C}$，とすると，$\boldsymbol{B}$ について最小にするべき目的関数は以下のように表される．

$$L_{\boldsymbol{\beta}}(\boldsymbol{B}|\boldsymbol{\Theta}) = -\sum_{c=1}^{C}\sum_{i=1}^{N} w_\alpha^c(p_{ic}|\hat{\boldsymbol{\theta}}_c) \{y_{ic}\log p_{ic} + (1-y_{ic})\log(1-p_{ic})\} + \lambda_{\boldsymbol{\beta}}\sum_{k=1}^{C}||\boldsymbol{\beta}_c||^2. \quad (3)$$

ここで，$w_\alpha^c(p_{ic}|\hat{\boldsymbol{\theta}}_c)$ は相対密度比 [Yamada 13]，$\lambda_{\boldsymbol{\beta}} > 0$ は正則化パラメータ，$||\cdot||$ はフロベニウスノルムである．各クラスへの所属を表す y_{ic}，予測された各クラスへの所属確率 p_{ic} がそれぞれ連続型確率分布 P_{y_c}，P_{p_c} に独立同一に従うと仮定すると，確率分布 P_{y_c}，P_{p_c} の相対密度比は，それぞれの確率密度関数を f_{y_c}, f_{p_c} としたとき，以下のように定義される．

$$w_\alpha^c(p) = \frac{f_{p_c}(p)}{\alpha f_{y_c}(p) + (1-\alpha) f_{y_c}(p)}.$$

クラス c に対して，「理想的な判別」が行われた状況を考えると，混同行列における FN(c) $= 0$ となることから，このときの相対 F-measure を以下のように表すことができる．

$$\text{F-measure}(\alpha, c) = \frac{n_c}{\alpha n_c + (1-\alpha)m_c} \quad (4)$$

n_c，m_c をそれぞれ f_{y_c}，f_{p_c} と置く．すると，それぞれ各対象 1 つ 1 つに対して「分類に寄与する度合い」のようなものと解釈することができ，これは相対 F-measure の形になる．目的関数 (3) 式の $w_\alpha^c(p_{ic}|\hat{\boldsymbol{\theta}}_c)$ の c に関する和の部分が全体での macro F-measure に対応している．

次に，相対密度比の推定については，[Kanamori 09] で提案されている Unconstrained Least-Squares Importance Fitting (uLSIF) に基づく方法で推定する．いま，相対密度比を $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_N)' \in \mathbb{R}_+^N$ として以下の通りにモデル化する．

$$w(p|\boldsymbol{\theta}) = \sum_{i=1}^{N} \theta_i K(p, p_i).$$

ここで，$K(p, p_i)$ は，カーネル関数であり，通常は RBF カーネルが使われる [Kanamori 09]．$\boldsymbol{\theta}_c$ の推定値は，

$$L_{\boldsymbol{\theta}}(\boldsymbol{\theta}_c|\boldsymbol{\beta}_c) = \frac{1}{2}\boldsymbol{\theta}_c'\hat{\boldsymbol{H}}_c\boldsymbol{\theta}_c - \hat{\boldsymbol{h}}_c'\boldsymbol{\theta}_c + \frac{\lambda_{\boldsymbol{\theta}}}{2}||\boldsymbol{\theta}_c||^2,$$

最小にすることで得られ，その解は $\hat{\boldsymbol{\theta}}_c = \max(\boldsymbol{0}, (\hat{\boldsymbol{H}}_c + \lambda_{\boldsymbol{\theta}}\boldsymbol{I}_N)^{-1}\hat{\boldsymbol{h}}_c)$ によって相対密度比の条件を満たす形で得られる．ここで，$\lambda_{\boldsymbol{\theta}}$ (≥ 0) は正則化パラメータであり，$\boldsymbol{I}_N$ は N 次単位行列である．さらに，$\hat{\boldsymbol{H}}_c \in \mathbb{R}^{N\times N}$ は第 (ℓ, m) 要素 $\hat{H}_{c\ell m}$ が $\hat{H}_{c\ell m} = \frac{\alpha}{N}\sum_{i=1}^{N} K(p_{ic}, p_{\ell c})K(p_{ic}, p_{mc}) + \frac{(1-\alpha)}{N}\sum_{i=1}^{N} K(y_{ic}, p_{\ell c})K(y_{ic}, p_{mc})$ である行列で，$\hat{\boldsymbol{h}}_c \in \mathbb{R}^N$ は，第 ℓ 要素に $\hat{h}_{c\ell} = \frac{1}{N}\sum_{i=1}^{N} K(p_{ic}, p_{\ell c})$ を持つベクトルである．正則化パラメータ $\lambda_{\boldsymbol{\theta}}$，RBF カーネルのバンド幅はクロスバリデーションによって定められる．

推定を行うためのアルゴリズムはつぎの通りである．

入力 データ行列 $\boldsymbol{X} \in \mathbb{R}^{N\times D}$，ラベル行列 $\boldsymbol{Y} = (\boldsymbol{y}_i) \in \{0,1\}^{N\times C}$，相対密度比に対するパラメータ $\boldsymbol{\alpha}$

出力 パラメータ行列 $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_C) \in \mathbb{R}^{P\times C}$

STEP0 各クラス $y_c \in \{0,1\}^D$ を目的関数としたロジスティック回帰のパラメータベクトル $\boldsymbol{\beta}_c$ $(c = 1, ..., C)$ を求め，$\boldsymbol{B}_0 = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_C)$ とする．

STEP1 得られたパラメータ行列 $\boldsymbol{B}_0$ を用いたモデル ((2) 式) による予測結果 $\hat{\boldsymbol{P}} = (p_{ic}) = (1/1+\exp(-\boldsymbol{\beta}_c'\boldsymbol{x}_i))$ を求める．

STEP2 各クラス $(c = 1, ..., C)$ に対するラベルベクトル $\boldsymbol{y}_c$ と各対象のクラス c への所属確率のベクトル $\boldsymbol{p}_c$ の $\boldsymbol{\alpha}$-相対密度比に関するパラメータ $\boldsymbol{\theta}_c$ を uLSIF によって推定し，$\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_C)$ とする．

STEP3 前の STEP で推定したパラメータ $\boldsymbol{\Theta}$ を用いて，(3) 式の $L(\boldsymbol{B}|\boldsymbol{\Theta})$ を最小にするパラメータ行列 $\boldsymbol{B}$ を求める．

STEP4 $||\boldsymbol{B}_0 - \boldsymbol{B}||^2 < \varepsilon$ ならば終了する．そうでない場合は $\boldsymbol{B}_0 = \boldsymbol{B}$ として，STEP1 に戻る．

参考文献

[Boutell 04] Boutell, M. R., Luo, J., Shen, X., and Brown, C. M. (2004). Learning multi-label scene classification. Pattern recognition, **37**, 1757-1771.

[Kanamori 09] Kanamori, T., Hido, S., and Sugiyama, M.(2009). A Least-squares Approach to Direct Importance Estimation, *Journal of Machine Learning Research*, **10**, 1391–1445.

[Sun 13] Sun, L., Ji, S., and Ye, J. (2013). Multi-label dimensionality reduction. CRC Press.

[Tsoumakas 07] Tsoumakas, G., Vlahavas, I.(2007). Random k-labelsets: An ensemble method for multilabel classification., Proceedings of the 18th European Conference on Machine Learning (ECML 2007), Warsaw, Poland (2007) 406–417.

[Yamada 13] Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., and Sugiyama, M.. (2013). Relative density-ratio estimation for robust distribution comparison. *Neural computation*, **25**, 1324–1370.

[2P1] [General Session] 7. Agent

座長:田中 友紀子（NEC）

Wed. Jun 6, 2018 9:00 AM - 10:40 AM Room P (4F Emerald Lobby)

[2P1-01] Building Cooperative Relationship by Continuous Observation of Interaction between Two Agents

○Junya Karasaki[1], Yoshimasa Ohmoto[1], Toyoaki Nishida[1] （1. Graduate School of Informatics, Kyoto University）

9:00 AM - 9:20 AM

[2P1-02] Detecting Suspects by Observing Behavioral Changes of Surrounding Pedestrians

○Reina Saitou[1], Ei-Ichi Osawa[1] （1. Future University Hakodate）

9:20 AM - 9:40 AM

[2P1-03] Parallelization of evolution of reinforcement learning agents using GPGPU

○Yoshiki Senga[1], Kouichi Moriyama[1], Atsuko Mutoh[1], Tohgoroh Matsui[2], Inuzuka Nobuhiro[1] （1. Nagoya Institute of Technology, 2. Chubu University）

9:40 AM - 10:00 AM

[2P1-04] Hedonic games considering agents sociality

○Kazunori Ota[1], Nathanael Barrot[2], Yuko Sakurai[3], Makoto Yokoo[1,2] （1. Graduate School of Information Science and Electrical Engineering, Kyushu University, 2. RIKEN, Center for Advanced Intelligence Project AIP, 3. AIST Artificial Intelligence Research Center International Symposium）

10:00 AM - 10:20 AM

[2P1-05] Intelligent agents supporting to estimation of cyber threat intelligence

○Hidetoshi Kawaguchi[1], Yuichi Ishihara[1] （1. NIPPON TELEGRAPH AND TELEPHONE CORPORATION）

10:20 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

エージェント間インタラクションの継続的な観察による協調関係の構築

Building Cooperative Relationship
by Continuous Observation of Interaction between Two Agents

唐崎準也 Junya Karasaki　大本 義正 Yoshimasa Ohmoto　西田 豊明 Toyoaki Nishida

京都大学大学院 情報学研究科
The School of Information and Mathematical Science, Faculty of Engineering, Kyoto University

We aim to develop a cooperative relationship between people and agents through interaction. To develop this, we have to make people recognize agents as intentional being, and make people interact with agents voluntarily. So we designed two agent models. The feature is to let people observe the situation, where two agents, the superior role and the subordinate role, interact with each other in the form of questions and answers. A psychological hurdle for interaction falls down and interacts voluntarily, by a man observes the state of making question and answering. I prepared experiment tasks to evaluate the performance of proposed agents. By the experiment, I proved that these proposed methods causes psychological attitude to feel the intention of the agent and voluntary interaction, and they are useful to develop cooperative relationship with agent.

1. はじめに

1.1 背景

人とインタラクションを行うエージェントに求められている役割とは，人とエージェントがインタラクションを通して社会の様々な問題を解決することだと言える．特定の個人と密接に関わる問題を解決するために，社会的パートナーとしてのエージェントが期待されている．

そのためエージェントとの協調関係を構築する研究が求められている．協調とは，異なる環境や立場に存する複数の者が互いに助け合ったり譲り合ったりしながら同じ目標に向かって任務を遂行することと定義されている．インタラクションを通してお互いの意図を確認し，目標を達成するために協力し合えるエージェントを作成することが本研究の最終的に目指すところである．

人と協調関係を構築するエージェントを実現するためには，いくつかの問題点を解決しなければならない．その一つは人のエージェントに対する心理的姿勢についての問題である．具体的な例として，エージェントを意思のある存在だと認識せず，エージェントの意思を考慮した行動をしないというものがある．このような他者の振る舞いに対する心理的姿勢を表す表現として，設計姿勢と志向姿勢という概念が存在する [Dennett 89]．設計姿勢をとっている状態とは設計者の定めた行動原理を考えて他者を捉える姿勢であり，志向姿勢は相手が何を考えているかに注目して他者を捉える姿勢である．人がエージェントに対して設計姿勢を取っているままでは，エージェントの意思に配慮する協調関係が生まれることはない．協調関係の構築と維持を続けるためには，エージェントとの継続的なインタラクションを通して志向姿勢を誘発し，志向姿勢を維持する必要がある．

協調関係の実現には，人がエージェントに対して自発的にインタラクションを行わないという問題点も解決する必要がある．将来の関係が親密になるか否かは出会ってから 1, 2 週間程度の初期のインタラクションが重要であることが示されている [山中 94] ことを踏まえると，人とエージェントとの協調関係を実現するためには，まずインタラクションを円滑に行うことができる関係を構築する必要がある．しかし初対面の人同士が会話を続けることは難しいように，人は関係を構築していない相手に対して自発的にインタラクションを行うことに抵抗を感じてしまう．特にエージェントに対してはより強い抵抗感を感じてしまうため，その心理的なハードルを取り除く手法を検討する必要がある．

エージェントは意思のある存在であると認識させ志向姿勢を取らせることと，人がエージェントに対して自発的にインタラクションを行うことは，エージェントとの協調関係を構築するために必要である．そこで本研究ではこの 2 点を達成するためのエージェントモデルを設計した．このモデルは，二人のエージェントがインタラクションを行う様子を継続的に観察させることが心理的なハードルを下げることにつながり，自発的なインタラクションを促すという仮説に基づいて設計が行われた．

1.2 関連研究

説得エージェントが被説得エージェントを説得する様子を被験者が横から漏れ聞いた場合，説得エージェントに直接説得された場合と比較して，被験者は説得方向により態度を変容する [Walster 62] [鈴木 04]．態度が変化する原因として，鈴木らは人は他者の存在を基準にして自らの態度を決めることを理由としてあげている．このことから，エージェント同士が協調的なインタラクションを取る様子を横から観察させることで人も協調的な態度になると考えられる．しかしこれらの研究では人とエージェントとのインタラクションについての言及はないため，エージェントに対する自発的なインタラクションを行わせるために，どのようなインタラクションを横から観察させればよいかを考える必要がある．

プレゼンテーションの場において聴衆からの質問を促すための手法をとして，聴衆からの質問を求める前に「別の講演ではこのような質問があった」などと聴衆の興味がありそうな簡単な質問を話者が代弁し，それに答えるという方法が存在する [開米 16]．話者が質問の例を提示することで，聴衆は質問を行った第三者と興味が一致したことによる安心感が得られる．その結果，質問行為に対する心理的ハードルが下がり，積極的な質問を行うようになる．開米はエージェントに関わる記述は行なってないが，この手法はエージェントが行なった場合

連絡先: 唐崎準也，京都大学大学院情報学研究科，karasaki@ii.ist.i.kyoto-u.ac.jp

でも人からの質問を促すことができると考えられる．

これらの研究を踏まえると，エージェントが人に向かって直接質問の例を提示する方法よりも，「質問エージェントが他のエージェントに質問する」という協調的なインタラクションの様子を横から観察させて，間接的に質問の例を提示する方が優れていると言える．なぜなら質問の例を提示して心理的ハードルを下げる効果と，エージェント同士が協調的なインタラクションを取る様子を観察させて人も協調的な態度を取らせる効果の二つを同時に行えるからである．

そこで本研究では鈴木らと開米の二つの手法を組み合わせることで，人がエージェントに対して自発的にインタラクションを働きかけるよう促し，協調的な態度を取らせることが可能かを検証する．

2. 提案モデル：協調関係構築のための 2 人のエージェントモデル

このモデルの要点は，二人のエージェントが質問と返答を伴うインタラクションを行うこと，その様子を観察させることで自発的なインタラクションを促すことである．

2.1 上司／部下エージェント

エージェント同士が質問とその返答を伴うインタラクションを人に観察させることで，心理的ハードルを下げ，エージェントに対して自発的なインタラクションを引き起こすためのエージェントを提案する．異なる立場であることを端的に示し，共同でタスクを行う際に違和感のない関係として，上司と部下の役割を二人のエージェントに割り当てることが適している．部下エージェントは人と同じ立場に立っており．上司エージェントは人と部下エージェントの二人よりも上の立場に立っている．

タスクにおける上司エージェントの役割は，タスク全体の流れを管理し，部下エージェントの質問に答えることでタスク遂行に必要な情報を提示することである．部下エージェントの役割は，上司エージェントに質問する様子を人に観察させることで，エージェントとのインタラクションに対する心理的ハードルを下げ，自発的なインタラクションを引き起こすことである．

2.2 インタラクションの観察

上司と部下の二人のエージェント同士は人に観察させるためのインタラクションを行う．エージェント同士が質問と返答を伴うインタラクションを行うことは前述した通りだが，この節では質問の内容について触れる．

エージェントが提示するべき質問は，人の興味を引き，簡単なものである必要がある．人がエージェントと共同でタスクを行う際，タスクに対する動機付けが適切に行われている場合，人はタスクに関わる内容に興味を示すと考えられる．そこでエージェント同士が行うインタラクションでは，部下エージェントがタスク遂行に関する疑問点を上司エージェントに尋ね，上司エージェントがそれに回答するという内容の会話を行う．

3.1 節と 3.2 節の内容をまとめたものを図 1 に示す．

3. 評価実験

人とエージェントとの間に協調関係を構築するにあたって，エージェント同士が質問と返答を通してインタラクションする様子を継続的に観察することで自発的なインタラクションを促す，という手法が有効であるかを検証する実験を行なった．

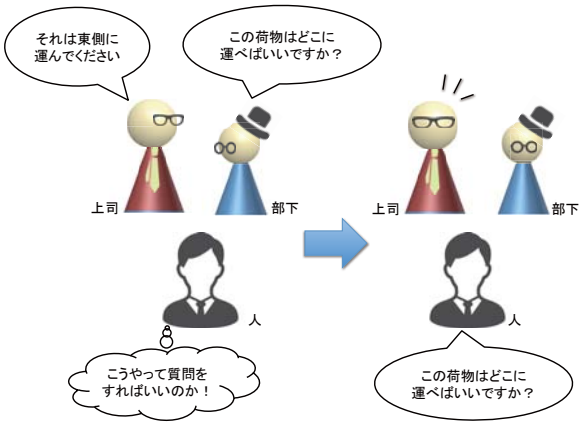


図 1: エージェントを観察して自発的インタラクションを促す様子

3.1 実験タスク

タスクには建設ゲームを用いた．前半タスクでは，エージェントが質問を行う様子を見せ，インタラクションに対する心理的ハードルを下げることで自発的インタラクションを促し，協調関係の構築を目指す協調関係を構築する．後半タスクでは，エージェントから参加者に対する働きかけを最小限に抑えた上で，エージェントとの協調関係が維持されていたかを検証した．全体のタスクの流れを図 2，各セッションで建築する家の完成予想図を図 3 に示す．

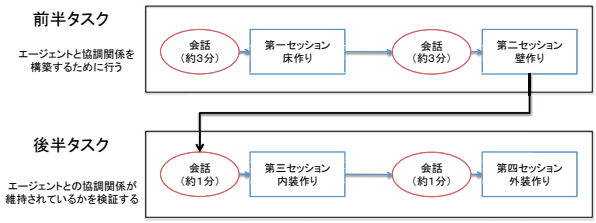


図 2: タスクの流れ

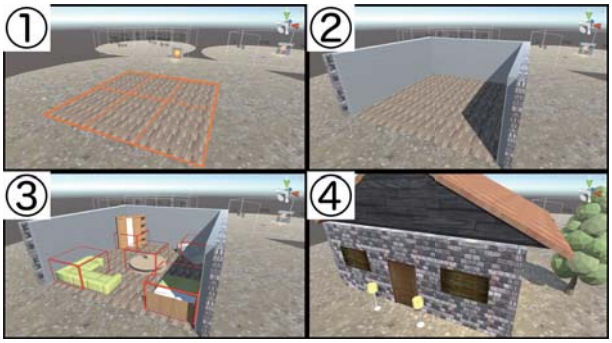


図 3: 建設タスクの各セクションにおける完成図

3.1.1 ルール

プレイヤーは二人のエージェントと協力して複数の資材を正しい目的地に運び，家を建設する．資材には一人で運ぶ資材と

二人で運ぶ資材があり，二人で運ぶものはエージェントたちと協力して運ぶ必要がある．資材を運んだ種類・個数・場所に応じて，家は自動的に建築される．人とエージェントは音声による会話を行うことができる．

3.1.2 前半タスク

資材には複数の種類があり，タスクの進行状況に応じて運ぶ目的地が決められている．人はエージェントの行動や発言からそれぞれの資材の正しい目的地を推測し，運ぶ必要がある．エージェントは資材の目的地などを尋ねられたら，それに返答する．

提案エージェントの性能を評価するため，以下のエージェントを比較に用いた．

提案エージェント

部下エージェントが目的地について質問し，上司エージェントがそれに答える様子を人に観察させることで，インタラクションに対する心理的ハードルを下げる．また部下エージェントは上司エージェントの他に，人に対しても質問する．

比較対象エージェント

上司エージェントが資材の目的地などの情報を，人に直接伝える．

3.1.3 後半タスク

資材には複数の種類があり，人は好きな場所に資材を設置できる．自由度の高いタスクの中で，部下エージェントが設置場所などに関する妥当な提案と妥当でない提案を行う．それらの提案に関する反応が，実験群と対照群で変化するかを検証する．

3.2 参加者

本実験は，情報工学に携わっていないという条件で本学学生から参加者を募集した．参加者は 19 歳から 35 歳までの 25 人で，年齢の平均は 21.60 歳（分散 3.274），内男性が 15 人，女性が 10 人であった．実験群には男性 7 人と女性 5 人，対照群には男性 8 人と女性 5 人を割り当てた．

3.3 実験結果

3.3.1 参加者の発話の割合

参加者が自分の運びたい家具を伝えるための発話を行った人数を，後半タスクの各セッションごとに表 1 にまとめた．後半タスクにおいて実験群と対照群の人数を比較するためにカイ二乗検定を行なったところ，有意水準 $p < 0.05$ で実験群のほうが高いという有意差が見られた（$p = 0.0469, p = 0.006691$）．

第三セッション	実験群	対照群
発話あり	5	1
発話なし	7	12
第四セッション	実験群	対照群
発話あり	7	1
発話なし	5	12

表 1: 参加者が自分の運びたい家具を伝える発話を行なった人数

次に，エージェントがどの家具を運びたいかを尋ねる発話を行なった人数を，後半タスクの各セッションごとに表 2 にまとめた．後半タスクにおいて実験群と対照群の人数を比較するためにカイ二乗検定を行なったところ，有意水準 $p < 0.05$ で実験群のほうが高いという有意差が見られた（$p = 0.04691$）．第四セッションにおいては有意差は見られなかった（$p = 0.2383$）．

第三セッション	実験群	対照群
発話あり	5	1
発話なし	7	12
第四セッション	実験群	対照群
発話あり	3	1
発話なし	9	12

表 2: エージェントがどの家具を運びたいかを尋ねる発話を行なった人数

3.3.2 参加者による主観評価の結果

参加者の主観評価を調査するために 7 段階のリッカート尺度を用いてアンケートを行った．本研究において特筆すべき以下の項目について，本実験における意味と結果を述べる．

1. 外装を作る際，全体のレイアウトを意識して設置したか．
   （1:全く意識しなかった 7:常に意識した）

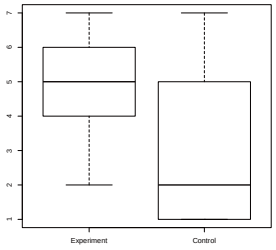


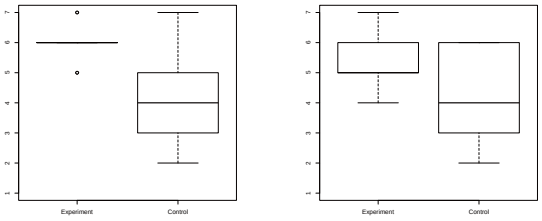
図 4: 全体のレイアウトを意識して設置したかという項目のアンケート結果

   第四セッションにおける参加者のタスクに対する意欲を評価する質問である．エージェントとの協調関係が築けていればこの評価は高くなると考えられる．参加者がエージェント同士のインタラクションを模倣を行なった実験群と行わなかった対照群のアンケート結果においてマン･ホイットニーの U 検定を行なったところ，有意水準 $p < 0.05$ で実験群のほうが高いという有意差が見られた（$p = 0.04338$）（図 4）．

2. 上司役のエージェントに対してどのくらい好感を感じたか．

3. 部下役のエージェントに対してどのくらい好感を感じたか．
   （1:全く好感を得なかった 7:非常に好感を得た）

   エージェントに対する好感度を評価する質問である．エージェントと協調関係を築けている場合，エージェントに対する仲間意識は高まるはずであり，エージェントに対する好感度は高くなると予想される．実験群と対照群のアンケート結果においてマン・ホイットニーの U 検定をそれぞれの項目に行なった．上司役エージェントの好感



(a) 上司役のエージェントの好感度のアンケート結果　(b) 部下役のエージェントの好感度のアンケート結果

度では，有意水準 $p < 0.01$ で実験群のほうが高いという有意差が見られた（$p = 9.103 * 10^{-4}$）（図 5a）．部下役エージェントの好感度では，有意水準 $p < 0.10$ で実験群のほうが高いという有意傾向が見られた（$p = 0.07063$）（図 5b）．

また実験群の上司エージェントと部下エージェントの好感度で違いがあるかを調べたところ，有意水準 $p < 0.10$ で実験群のほうが高いという有意傾向が見られた（$p = 0.0887$）．

4. 完成した家にはどのくらい満足したか．
（1:全く満足しなかった 7:非常に満足した）

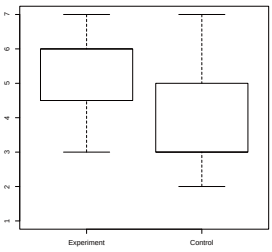


図 6: 完成した家に対する満足度のアンケート結果

完成した家の満足度を評価する質問である．タスクに対する意欲が高ければ建設作業を真剣に行い，満足する家を完成させるはずである．実験群と対照群のアンケート結果においてマン・ホイットニーの U 検定をそれぞれの項目に行なったところ， 有意水準 $p < 0.05$ で実験群のほうが高いという有意差が見られた（$p = 0.04239$）（図 6）．

4. 実験考察

以上の結果を踏まえると，上司エージェントと部下エージェントの好感度は実験群の方が対照群よりも有意に高く，特に上司エージェントについてその結果が顕著に得られた．エージェントに対する好感度の高さは，エージェントと協調関係を構築する上で重要な点であるため，好感度の面から提案手法の有効性は示すことができた．ただし部下エージェントに対する好感度は変化が小さかった．その原因として，後半タスクに妥当でない提案をしたからという理由が考えられる．

参加者の意欲の評価について検証すると，実験群の方が対照群よりも有意に高かった．第四セッションの最中とタスク終了時の両方で意欲が高いことが示されているため，前半タスクで構築した協調関係が最後まで維持されていたことが示された．

参加者の発話内容について検証を行う．参加者が意見を伝えるための発話を行なった人数は有意な差があった．一方でエージェントの意見を尋ねる発話を行なった人数は，第三セッションでは有意な差があったものの，第四セッションでは実験群と対照群でともに人数が減り，有意差が見られなくなった．この原因は，後半タスクではエージェントは人に追従する動作をする場面が多く，参加者がエージェントの行動意図をあまり感じられなくなったからだと考えられる．この問題を解消するためには，エージェントに目的志向性を持たせてエージェント自身の意思を示す動作を増やす，などの手法が考えられる．

5. 結論

本研究ではエージェントと協調関係を構築するために，エージェントを意思のある存在だと思わせることと，人がエージェントに対して自発的にインタラクションを試みることを目指した．そのために，二人のエージェント同士で行うインタラクションを観察させることでインタラクションに対する心理的ハードルを下げ，エージェントに対する自発的なインタラクションを引き起こす手法を提案した．

評価実験の結果，実験群では自発的なインタラクションを行う人数が増加し，エージェントに対する好感度の向上や相手の意思に配慮した発言の増加も見られた．これらのことから，提案手法によって協調関係の構築と維持が可能であることが示唆される．

今後の課題として，エージェントに対する配慮を長期にわたるタスクの終盤まで維持するための手法を研究する必要がある．そのためエージェントが何らかの目的意識を持って行動していることをより強力に提示するエージェントモデルを構成し，そのモデルがエージェントに対する配慮を長期にわたり維持するために有効であるかを検証する必要がある．

6. 謝辞

本研究は，科研費 若手研究 (B) (KAKENHI No.16K21113) の助成を受けています．

参考文献

[Dennett 89] Dennett, D. C.: *The intentional stance*, MIT Press (1989)

[山中 94] 山中一英 :対人関係の親密化過程における関係性の初期分化現象に関わる検討. *The Japanese Journal of Experimental Social Phychology*, Vol. 34, No. 2, pp. 105-115 (1994)

[Walster 62] Walster, E., Festinger, L. :the effectiveness of “overheard” persuasive communications, *Journal if Abcormal and Social Psychology*, Vol. 65 No. 6, pp. 395-402 (1962)

[鈴木 04] 鈴木聡, 山田誠二 :Persuasion through overheard communication by life-like agents, *Proceedings. IEEE/WIC/ACM International Conference*, pp. 225-231 (2004)

[開米 16] 開米瑞浩 :エンジニアを説明上手にする本 相手に応じた技術情報や知識の伝え方, 翔泳社 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

周辺歩行者の挙動変化に基づく不審者検出

Detecting Suspects by Observing Behavioral Changes of Surrounding Pedestrians

齊藤　伶奈　　大沢　英一
Reina Saitou　　Ei-Ichi Osawa

公立はこだて未来大学システム情報科学部複雑系知能学科
Department of Complex and Intelligent Systems, School of Systems Information Science, Future University Hakodate

The recent attacks in towns have increased demands for technology to find suspects in a crowd. A typical way to find suspects is to use facial recognition systems, however, those require a huge amount of personal data in advance. In this paper we propose a method to detect suspects in a crowd by observing pedestrians' behavior without using any specific personal data. The fundamental idea is that if a pedestrian finds a suspect, they might stop walking, or change the direction. We have devised a method to find such changes of behavior of pedestrians based on a Kalman filter and a hidden Markov model. The filter is used to detect a change, and the HMM is for assuming the intention of each observed pedestrian. Agent simulation results shows that the method works sufficiently well, especially where people are walking not in a single direction but in various different directions.

1. はじめに

昨今，テロに代表されるような往来での無差別殺人への対処は安全保証の観点において重要な案件となっている．もし，多数の人間が歩いている雑踏の中から事前に不審者を検出することが出来れば，将来の犯罪抑止の一助となるだろう．このような目的のための不審者検出に関するこれまでの研究では，不審者本人の行動検出や顔検出といった方法が様々用いられている [1, 2]. しかし，これらの手法では犯罪者データベースなどが必要となる．加えて，不審者自身の行動を認識できない場合には検出することができないという問題がある．

そこで本研究では，それらの手法の補助を目的として，不審者を回避した歩行者の回避行動に着目して，不審者を検出する手法を提案する．本研究では，提案手法の有効性を群衆の中の不審者とそれに気づいた人のモデルを構築し，このモデルを利用したエージェントシミュレーションをすることで実験考察を行なった．

2. 提案手法

本章では，提案する不審者検出で用いる手法について詳しく書く．まず，不審者を見つけた歩行者の行動変化の検出に使用するカルマンフィルタを紹介する．次に，検出した行動変化が不審者を避けるための意図によって引き起こされているかどうかを調べるために利用する隠れマルコフモデルを紹介する．

2.1 カルマンフィルタ

本研究では，まず回避行動と思われる行動を検出するために，歩行者の行動変化を検出する．そのために，カルマンフィルタを用いて次状態の予測を行い，実際の次状態との予測誤差から変化を検出する．ここで，歩行者の時刻 t での位置を x_t，速度を v_t とする．この歩行者の位置 x_t を状態とし，次状態への間隔を Δ とすると，状態は $x_t = x_{t-1} + v\Delta$ と更新することができる．実際の観測がノイズを含むと想定すると，この式にガウスノイズ σ を加えることで，観測 z_t は式 (1) で得ることができる．

$$z_t = x_{t-1} + v\Delta + \sigma_t \tag{1}$$

カルマンフィルタはこの更新式から，事前に状態を推定する予測ステップと，観測によって事後推定値を求めるフィルタリングステップを行う．ベイズの定理を用いると，観測と状態は以下のような式で表すことができる．

$$P(x_t|z_t) = \frac{P(z_t|x_t)P(x_t)}{P(z_t)} \tag{2}$$

時刻 t の時，状態 x_t の条件付き確率 $P(x_t|Z^{t-1})$，さらに観測 z_t の条件付き確率は $P(z_t|Z^{t-1})$ となり次の関係式を得ることができる．Z^{t-1} は時刻 $t-1$ までの観測の集合を表す．

$$P(x_t|Z^t) = \frac{P(z_t|x_t)P(x_t|Z^{t-1})}{P(z_t|Z^{t-1})} \tag{3}$$

$$P(x_{t+1}|Z^t) = \int P(x_{t+1}|x_t)P(x_t|Z^t)dx_t \tag{4}$$

ここで，式 (3) からは，時刻 t の状態 x_t の状態推測値を得ることができ，式 (4) からは，時刻 $t-1$ での状態 x_t の状態推測値を得ることができる．この 2 つの式を，交互に繰り返すことで時刻 t での状態予測を行うことができる．実際はさらに，時刻 t での状態 x_t の推定値 $\hat{x}_{t|t}$ はカルマンゲイン K を用いて，

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(z_t - \hat{x}_{t|t-1}) \tag{5}$$

と更新を行う．このカルマンゲインは，観測と状態の関わりの強さを表し，

K_t = 時刻 t における x の推定値の誤差分散 / 時刻 t における x の観測値の誤差分散

となっている．

2.2 HMM

つぎに，行動変化があった歩行者が，本当に不審者に対して回避行動を取っているかどうかを判断するために，隠れマルコフモデル (以下 HMM:Hidden Markov Model) を用いた．HMM

連絡先: 齊藤伶奈，公立はこだて未来大学，〒 041-8655 北海道函館市亀田中野町 116-2，b1014094@fun.ac.jp

は，直接観測できない隠れた状態とそこから出力される観測できる状態からなるモデルで，$< Q, A, O, B, \Pi >$のタプルで表現される．Q は隠れ状態の有限集合，q_i は i 番目の隠れ状態，A は遷移確率の有限集合，O は観測可能な状態の有限集合，B は観測確率の有限集合，Π は初期状態確率の有限集合を表す．また，a_{ij} は q_i から q_j への遷移確率 $a_{ij} = Pr(s_{t+1} = q_j | s_t = q_i)$，$o_i$ は部分的に観測可能な q_i の状態，$b_{ik} = Pr(o_k | q_i)$ は状態 q_i で o_k を観測する確率，$\Pi_i = Pr(s_0 = q_i$ は初期状態が q_i にある確率を表す．

HMM の隠れ状態 Q を歩行者の意図とし，観測可能な状態 O をカルマンフィルタを用いて得た予測誤差と，実際に観測した歩行者の座標の傾きの変化量とした．図 1 はその HMM の構造である．

ただし，Yue らのマルチエージェントの目標認識についての論文 [5] によると，マルチエージェントシミュレーションは状態空間がとても大きく，HMM のパラメータの教師なし学習は不可能とある．そこで，状態空間が大きすぎることに対して，それぞれ連続値である O の予測誤差と，座標の傾きの変化量をそのまま扱うのではなく，離散値にして表現することで大きさを抑えた上で，Baum-Welch アルゴリズムを用いて事前に行なったエージェントシミュレーションの結果から HMM の教師ありパラメータ学習を行った．

出力値の離散化は，あらかじめ行なった実験から得られたデータからそれぞれの最大値と最小値をもとめ，その値を等間隔で 5 等分して行なった．出力はそれぞれの区間に従って $[0, 1, 2, 3, 4]$ のいずれかに離散化され，離散化された値の組みが改めて観測可能な状態 O として得られる．

また，図 1 では簡略化のため隠れ状態 Q を二つにしているが，実際は Q が 5 つのモデルを用いた．これは歩行者の意図を，後述するエージェントの回避行動として想定した 4 種類の行動の意図とその他に分けた場合で表現している．

学習後の HMM を用いて，カルマンフィルタで行動変化が観測された歩行者が回避行動を取っていると推測された場合，その周辺の人物を不審者の疑いがあるとしてタグづけを行い，タグが一定以上の歩行者を不審者とした．

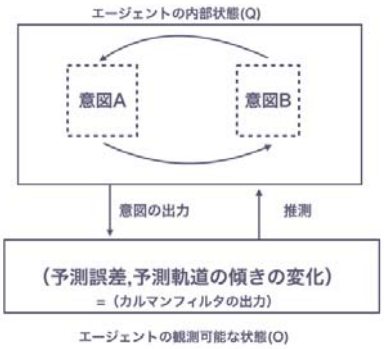


図 1: 研究で使用した HMM の構造

3. エージェントシミュレーション

本章では，手法の検証のために行なったエージェントシミュレーションについて詳述する．なお，エージェントシミュレーション上で不審者検出を行なっている研究は，Patrix らの研究や [5]，Kaluza らの研究 [4] がある．これらは，歩行者を複数人まとめて扱う点や．それぞれ不審者を直接認識する点で本手法とは異なる．

3.1 環境

実験は，歩行者のみで交通量の多い道路を想定しエージェントシミュレーションを行った．環境はグリッドワールドで表現し，実際に使用した環境は図 2 に示した．図 2 では，エージェントの目的地を赤色，エージェントを速度ごとにその他の色で視覚化している．グリッドは (60×30) マスになっており，環境中のエージェントは同じマスに重複しないようになっている．

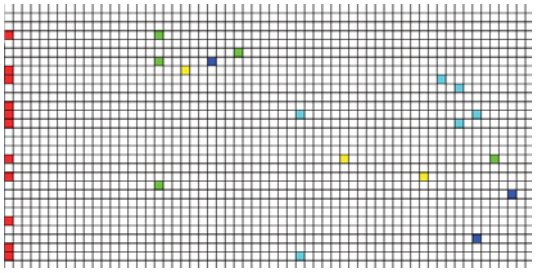


図 2: 実際に使用した環境

3.2 エージェント

エージェントは不審者の役割を持つものと，それ以外の普通の歩行者を持つものが存在する．エージェントは出発地から反対側の辺上の目的地に向かって環境内を歩き，目的地にたどり着いた時点で環境からいなくなる．シミュレーションは，不審者が環境から居なくなった時終了する．

エージェントはそれぞれ次の 7 つの要素を持っている．

- id:エージェントの名前
- $intention$:エージェントの意図
- $position$:エージェントの現在地
- $goal$:エージェントの目的地
- $speed$:エージェントごとの歩行速度
- map:エージェントが観測した情報を保存するための地図
- $fear$:エージェントの恐怖

また，それぞれの要素は以下のように表現されている．

- id = 重複しない文字列型の変数
- $intention =< A, H, S, T, N >$ のどれかをとる文字列型の変数 (詳細は後述)
- $position = (x, y)$ で保存されるシーケンス型の変数
- $goal = (x, y)$ で保存されるシーケンス型の変数
- $speed = 1 \sim 3$ の値をとる整数型の変数
- $map = [x][y]$ で周囲の様子を保存する整数型の 2 次元配列
- $fear = 0 \sim 100$ の値をとる整数型の変数

エージェントは，図 3 に示した範囲の視野をもち，視覚情報を用いて一歩進み，その後視野を更新することを繰り返しながら目的地を目指す．エージェントの周囲の情報は，map に記録されていく．もし，視界内に不審者をとらえた時はその位置を記録する．

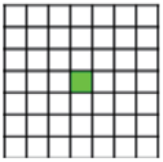


図 3: エージェントの視野

エージェントが不審者を発見できるかどうかは，fea の値で決まる．$fear$ は，不審者の恐怖への敏感さ，確信度合いを表したものでランダムな値をとる．$fear$ の値が決められた閾値を超えた時に，エージェントは不審なエージェントに気づくことができるようになっている．

3.3 回避行動

確率的に不審者に気づいた歩行者は回避行動を取るか取らないかを選択する．ここでの回避行動は，避ける・急ぐ・立ち止まる・引き返して逃げるの 4 つを想定しており，それぞれ不審者に遭遇した歩行者が取るであろう行動を仮定している．この回避行動は，行動が行われやすいと考えられるものほど選ばれやすくなるように設定し，それぞれの行動は以下のように定義した．

- 避ける:$fear$ の値に応じて不審なエージェントの周りを忌避する行動
- 急ぐ:$fear$ の値に応じて目的地に向かって急ぐ行動
- 立ち止まる:その場で立ち止まる行動
- 逃げる:初期位置に引き返すことで環境から脱出する行動

回避行動の選択は，エージェントの内部の意図 $intention$ の遷移であるとして，4 つの回避行動の意図と特に回避行動を取らない意図を合わせた計 5 つの意図を状況に用いた重み付き確率で遷移させることで表現している．$intention$ との対応は，避けるが A，急ぐが H，立ち止まるが S，逃げるが T，その他が N となっている．ただしエージェントの意図の初期状態は，回避行動ではない意図 N から始まる．図 4 に意図の遷移関係を示した．矢印の周囲の数字は遷移確率を表しており，状況に応じてもっともらしくなるように 2 つの遷移関係を用意した．

4. 実験・考察

本章では実際に行なった実験と考察について述べる．

4.1 実験

まず，通路のような環境を想定して，エージェントが環境の左端から歩き出し一方向に歩く場合で検出実験を行った．実験は，環境における歩行者エージェントの人数を変更するごとに 15 回ずつ行い，不審者の検出率と誤検出率を算出した結果を表 1 に示した．次に，左右二方向から生成され通行する状況で不審者の検出実験を行なった．結果は表 2 に示した．また，同

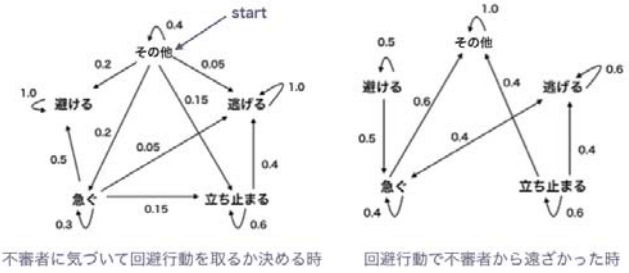


図 4: 意図の遷移関係

様の状況で環境中に不審者が複数いる場合の検出実験を行い，表 3 に結果をまとめた．

歩行者	不審者	検出率 (不審者：歩行者)
100 人	1 人	20.00%：0.13%
150 人	1 人	13.33%：0.64%
200 人	1 人	13.33%：0.79%

表 1: 一方向に歩く場合

歩行者	不審者	検出率 (不審者：歩行者)
150 人	1 人	33.33% ： 4.49%
200 人	1 人	86.00% ： 12.4%
250 人	1 人	80.00% ： 19.2%

表 2: 二方向通行の場合

加えてエージェントが上下左右四方向から生成され通行する場合の検出実験と，加えてこれまで実験を行なっていた環境が長方形であることから，上下方向に移動するエージェントの移動距離が短くなってしまうことから，大きさを変え正方形の環境を用いた場合での実験を行なった．それぞれの実験結果は表 4 と表 5 に示した．

表 1 と表 2 を比較すると，表 2 の方が明らかに検出率が高くなっている．これは表 1 の場合，環境に歩行者が増えれば増えるほど歩行者がすし詰めになってしまい，回避行動の取れない歩行者が増えてしまうからだと考えられる．加えて，環境を大きくした表 5 の実験では表 4 の実験と比較して，歩行者が少ない場合の検出率が低く歩行者が増えた場合には検出率が高い．これは，元々の環境では上底または下底に不審者が生成された場合に，環境から不審者がいなくなるまでの時間が短くなるため検出が難しかった．そのため，環境を大きくすることで不審者がどの場所に生成された場合でも同様に検出できるようになったからである．一方で，環境中に歩行者が少ない時には不審者の周囲で回避行動をとる歩行者が少なくなってしまったため検出率が低くなった．

4.2 考察

結果から，提案手法は環境中に人が少なすぎる場合や多すぎる場合は検出できない．さらに，ある程度の時間不審者が環境にとどまっていなければ検出できないことがわかった．環境

歩行者	不審者	検出率 (不審者：歩行者)
150 人	2 人	36.67% ： 5.54%
200 人	2 人	56.00% ： 11.58%
250 人	2 人	90.00% ： 20.5%

表 3: 不審者複数の場合

歩行者	不審者	検出率 (不審者：歩行者)
200 人	1 人	46.00% ： 1.85%
250 人	1 人	40.00% ： 4.72%
300 人	1 人	53.00% ： 7.62%

表 4: 進行方向が 4 方向の場合

の 1 マスを $1m^2$ としたとき，検出率が 8 割を超えた場合の人密度は，実験全体を通して最小で 0.08 人 $/m^2$, 最大で 0.13 人 $/m^2$ 程度だった．環境の大きさを拡大縮小した場合でも，環境中の歩行者密度がこれに近い中程度の密度であるとき，システムは不審者を検出可能であると考えられる．

しかし，環境全体の歩行者の密度が中程度であったとしても，不審者の周辺が極端に密度が高い場合やまた低い場合には検出がうまくいかない可能性が高い．例えば，不審者がグループで複数人でまとまって行動している場合では提案手法での検出は難しいだろう．ただし，表 3 の結果からもわかるように環境中の複数に不審者がいる場合でも検出可能であるため，それぞれ単独である場合には提案手法は十分有効だと考えられる．

また，進行方向を増やしても検出可能だったことから歩行者がある程度複雑に行き交う環境でも検出可能であることがわかった．さらに，歩行者が複雑に移動する場合の検出可能性にも期待できるだろう．ただし，環境中に障害物があるような複雑さを持つ環境の場合，現状の提案手法では障害物を避ける行動と回避行動とが区別できず誤検出が増えることが予測されるため，事前に障害物の位置がわかる場合にはその位置を記録し，その周辺では回避行動と判別しづらくするなどの操作が必要になると考えられる．

5. まとめと展望

5.1 まとめ

本研究は，群衆中での犯罪を事前に防ぐため，歩行者行動の変化を用いて群衆中から不審者を検出する手法を提案し，エージェントシミュレーションで手法の有効性の検出を行った．提案手法は，特殊なデータベースが必要なく，直接不審者が観測しづらい環境でも対応できることを目指した．人のみが通行する歩行者天国のような環境を使用してエージェントシミュレーションを行った．シミュレーションでは，まず大量の歩行者エージェントとそれに紛れた不審者エージェントを用意し，不審者エージェントに気づいた歩行者エージェントが回避行動をするようにした．次に，回避行動の原因となった不審者エージェントの特定を行い不審者の検出実験を行った．

検出は，カルマンフィルタを用いた行動変化の検出と HMM を用いた意図の予測を用いて行なった．実験の結果から，提案手法は不審者の周辺が中程度の歩行者密度の時によく不審者を検出することができ，歩行者の進行方向が増えた場合や，環境が大きくなった場合でも同様だった．

歩行者	不審者	検出率 (不審者：歩行者)
200 人	1 人	26.67% ： 1.89%
250 人	1 人	53.34% ： 1.35%
300 人	1 人	80.00% ： 4.29%

表 5: (60×60) の環境 4 方向の場合

5.2 課題と今後の展望

今回作成した不審者検出システムでは，局所的に人密度が高い時や低いときに弱く，不審者が集団の中に紛れてしまった場合や人混みを避けて歩いている場合には検出することができなかった．そこで，環境中に極端に密度が違う場所があるときに，その密度に従って行動変化の検出を緩やかにしたり厳しくしたりするような工夫をする必要がある．

また，実験を行なった環境は全て障害物がない環境だったため，障害物を避ける動作も加わってきた場合には回避行動との区別も重要な課題になるだろう．加えて，本手法は現在不審者検出で，実用的に使われている他手法の補助的な検出方法として考えられたものであるが，実際に他手法と併用した場合にどの程度有効なのかについてはまだ分かっていない．

したがって今後は，より普遍的な環境に対する対処と検出実験を行うことに加えて，他手法と併用した場合に対してのシミュレーションや，実際の歩行データを用いた実験を行っていきたい．

参考文献

[1] 佐藤 敦: 安全安心な社会を支える画像認識技術, 人工知能学会論文誌 29 巻 5 号, pp. 408-455(2014).

[2] テクノスコープ Technology File(012) 監視カメラシステム 不審者の”オーラ”を検知 ELSYS ジャパン キヤノンパナソニック，日経ビジネス (1835), 72-74(2016).

[3] Jeremy Patrix, Abdel-Illah Mouaddib, Simon Le Gloannec, Dafni Stampouli, and Marc Contat: Discrete relative states to learn and recognize goals-based behaviors of groups, AAMAS ’13 Proceedings of the 2013 inter-national conference on Autonomous agents and multi-agent systems, pp. 933-940(2013).

[4] Kaluza, Bostjan and Kaminka, Gal and Tambe, Milind: Towards Detection of Suspicious Behavior from Multiple Observations, AAAI Workshop - Technical Report, (2011).

[5] S. Yue, K. Yordanova, F. Krger, T. Kirste, and Y. Zha: A decentralized partially observable decision model for recognizing the multiagent goal in simulation systems. Discrete Dynamics in Nature and Society, vol. 2016, Article ID 5323121, pp. 15(2016).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

GPGPUを用いた強化学習エージェントの並列進化シミュレーション

Parallelization of evolution of reinforcement learning agents using GPGPU

千賀 喜貴 *1 Yoshiki Senga　森山 甲一 *1 Koichi Moriyama　武藤 敦子 *1 Atsuko Mutoh　松井 藤五郎 *2 Tohgoroh Matsui　犬塚 信博 *1 Nobuhiro Inuzuka

*1名古屋工業大学 大学院工学研究科 情報工学専攻
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

*2中部大学 生命健康科学部 臨床工学科
Department of Clinical Engineering, College of Life and Health Sciences, Chubu University

GPGPU is a parallel computation technology using GPU that has huge number of processor cores for parallelly calculating colors of pixels on a monitor. Owing to its parallel performance, GPGPU is being used for multiagent simulation that contains multiple independent but interdependent agents. In a previous work, we used GPGPU to parallelize many runs of reinforcement learning agents for calculating their fitness in a simulation of evolution. It speeded up the simulation surprisingly. However, the evolution part was sequentially run in CPU and the communication between CPU and GPU happened in every generation. Hence, this work uses GPGPU to parallelize the evolution part in addition to the fitness calculation. It makes the simulation even faster due to parallelism and the reduction of latency between CPU and GPU.

1. はじめに

マルチエージェントシミュレーションの分野では，エージェント数の増大に伴って計算時間が増加するという問題がある．この問題の解決方法として，画像処理等に用いられる GPU の計算性能を利用して並列に計算を行う GPGPU という分野が注目されるようになった．GPU は CPU に比べ一つのコアの計算性能は低いが数が多く，多数のコアで並列処理をすることで CPU 単体で計算するよりも高速な計算を行うことができる．GPGPU では計算に必要なデータと命令を CPU から GPU に転送する．CPUGPU 間のデータ転送は低速でボトルネックとなりやすい [1]．GPU 内のデータ転送は CPUGPU 間のデータ転送よりも数倍大きなメモリバンド幅で行われる．本研究では，2 人ゲームにおける強化学習の適応度計算を GPGPU によって並列実行するシミュレーション [2] を対象として，逐次計算している進化計算の部分を GPGPU によって並列化ならびに，シミュレーション全体を GPU 上で実行することによる CPUGPU 間通信を削減の 2 つにより速度向上を図れるか検証した．

2. General-Purpose computing on GPU (GPGPU)

2.1 GPU

GPU とは画像処理用のプロセッサであり，主にディスプレイに表示するピクセルの色を並列に計算することに使用される．CPU と比較すると，CPU は一般的に 4〜64 のコアを有していて，コア数は少ないがコア一つあたりの計算性能は高く連続的な処理に適しているのに対し，GPU はコアあたりの計算性能は低いがコア数が 1000 以上と非常に多く，並列的な処理に適している．大量のデータを VRAM と呼ばれる GPU のメモリに転送し多数のコアで並列処理することで，CPU 単体で処理するよりも高速に演算を行うことが可能である．一方で，GPU は 1 つの命令で複数のデータを処理する SIMD(Single Instruction Multiple Data) 形式で並列実行を行うため，分岐処理が存在する場合分岐する複数の命令を同時に実行することができない．また，CPU と GPU の間のデータ転送は GPU 内のデータ転送に比べ低速である．そのため実際の GPU を使った計算時間はデータの転送速度やその回数にも依存する．

連絡先: 千賀　喜貴，名古屋工業大学工学研究科情報工学専攻，名古屋市昭和区御器所町，y.senga.381@stn.nitech.ac.jp

2.2 GPGPU

GPGPU とは，高い演算能力を持つ GPU の演算領域を画像処理以外の数値計算に応用する技術である．GPGPU 用の開発環境として 2008 年に NVIDIA から CUDA が発表された．これにより，計算時間の膨大なプログラムの一部を GPU に委任することで，全体の処理効率を上げることが可能となった．その直後に GPGPU 言語として Khronos Group が提供する OpenCL が発表された．CUDA が NVIDIA 製の GPU に特化しているのに比べ，OpenCL は NVIDIA や AMD の GPU，Intel CPU のオンボードグラフィックスなど多くの環境下で動作し様々なデバイスに対応が可能であることが特徴である．

3. 準備

本章では本研究で使用する，囚人のジレンマゲーム，効用利用 Q 学習，囚人のジレンマゲームを用いた効用利用 Q 学習の進化計算の説明を行う．

3.1 囚人のジレンマゲーム

囚人のジレンマゲームは 2 人 2 行動ゲームの一種であり，2 人のプレイヤー A と B がそれぞれ C(協調) または D(裏切り) の行動を選択し，その組み合わせから表 1 の利得 $T, R, P, S \in \mathbb{R}$ を得る．囚人のジレンマゲームでは，利得 T, R, P, S 間に以下の関係を持つ．

$$T > R > P > S$$

B が C を選んだ場合，A は C を選ぶと得られる利得は R，D を選ぶと得られる利得は T なので A は D を選択したほうが

表 1: 囚人のジレンマゲームの利得表

$A\backslash B$	C	D
C	R, R	S, T
D	T, S	P, P

良い．B が D を選んだ場合，A は C を選ぶと得られる利得は S，D を選ぶと得られる利得は P，よって A は D を選択したほうが良い．以上の 2 つより A は B がどの行動を取るかに関わらず，行動 D を選択することが最適な選択になる．B にとっても同じのため行動の組み合わせが (D, D) となり，結果として両者は P の利得を得ることになる．しかし，仮に両者が C の行動を取った場合，それぞれは R の利得を得る．$R > P$ より個人が得る利得も増えるため，両者にとって良い結果となる．お互いが協調しあって大きい利得を得ることが望ましいのだが，自分の利得を追求するとお互いに裏切り合うという望ましくない結果になることからジレンマと呼ばれている．また，今回扱うのは囚人のジレンマゲームを繰り返し行う「繰り返し囚人のジレンマ」というゲームがである．繰り返し囚人のジレンマゲームでは以下の条件が加わる．

$$2R > T + S$$

この条件を加えることでどちらかが D を選択する状態が繰り返されるよりも互いに C を繰り返したほうが利得が大きくなる．

3.2 効用利用 Q 学習

エージェントはある時刻 t において状態 $s_t \in S$ を知覚し，方策 π に基づいて行動 $a_t \in A(s_t)$ を選択する．ここで S は環境の取りうるすべての状態を，$A(s_t)$ は状態 s_t の時に取りうる行動の集合を表す．行動後にエージェントは報酬 $r_{t+1} \in \mathbb{R}$ を受け取り，新たな状態 s_{t+1} を知覚する．強化学習の一種である Q 学習 [3] は，最適方策 π^* における行動価値を Q^* として行動価値関数 Q を Q^* に近づけるように以下の更新式に基づいて学習を行う．

$$Q_{t+1}(s,a) = \begin{cases} Q_t(s_t,a_t) + \alpha\delta_t & \text{if}(s,a) = (s_t,a_t) \\ Q_t(s,a) & \text{otherwise} \end{cases}$$

$$\delta_t \equiv r_{t+1} \max_{a \in A(s_{t+1})} Q_t(s_{t+1},a) - Q_t(s_t,a_t)$$

上記の α は学習率と呼ばれるパラメータで $0 < \alpha \leq 1$ の値を取り，γ は割引率と呼ばれるパラメータで $0 \leq \gamma < 1$ の値を取る．δ_t は TD (Temporal-Difference) 誤差と呼ばれ，$Q_t(s,a)$ が π^* における Q^* に近づくに連れて 0 に近づくものである．Q 学習の初期段階では $Q_t(s,a)$ が $Q^*(s,a)$ とは異なった値であるが，エージェントがすべての状態への訪問とすべての行動の選択を無限回繰り返し，α_t が以下の条件を満たす時，この誤差は 0 に収束する．$n^i(s,a)$ を i 回目に $(s,a) = (s_t,a_t)$ となった時の t とする．

$$\sum_{i=1}^{\infty} \alpha_{n^i}(s,a) = \infty, \sum_{i=1}^{\infty} [\alpha_{n^i}(s,a)]^2 < \infty, \forall s,a$$

最適方策における行動価値関数 Q^* が既知の場合，状態 s における最適な行動 a^* を以下のように表せる．

$$a^* = \underset{a' \in A(s)}{\text{argmax}}\, Q^*(s,a')$$

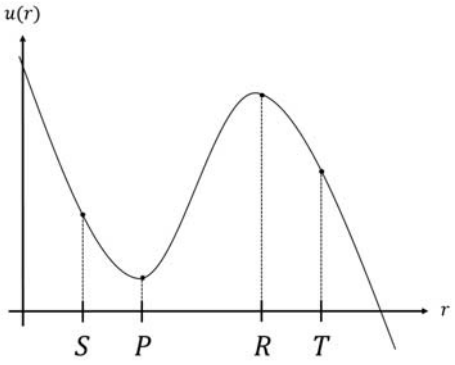


図 1: 効用導出関数 $u(r)$ のグラフのイメージ

エージェントが学習途中にこのような行動選択を行う場合，エージェントが行わない行動，未探索の状態が発生し，Q_t が局所解となる可能性がある．これを避けるため，ϵ-greedy 法などの手法を用いて行動選択を行う．ϵ-greedy 法は確率 $1-\epsilon$ で最大の Q_t を持つ行動を，確率 ϵ でランダムな行動を選択する．
効用利用 Q 学習 [4] では，エージェント内部に何らかの情動機構の存在を仮定し，その出力である主観的効用 u を Q 学習の報酬として利用する．以下では，情動機構として環境から得られる客観的報酬 r による関数を想定し，それを効用導出関数と呼ぶ．

3.3 囚人のジレンマゲームを用いた効用利用 Q 学習の進化計算

本稿では 2 人ゲームにおける強化学習シミュレーションとして囚人のジレンマゲームを用いた効用利用 Q 学習の進化計算を扱う [4]．このシミュレーションは囚人のジレンマゲームを繰り返し行うエージェントの情動機構の進化を疑似的に表現することを目的としている．報酬 r に対する効用を $u(r)$ とする．囚人のジレンマゲームの報酬 T, R, P, S に対して図 1 のように

$$u(R) > u(T) \quad \text{and} \quad u(S) > u(P) \quad \text{and} \quad u(R) > u(P)$$

となれば，協調行動 C が優位になる．この条件を表現する効用導出関数を

$$u \equiv u(r) \equiv ar^3 + br^2 + cr + d$$

と仮定し，この関数の係数 a, b, c, d をまとめて染色体として遺伝的アルゴリズム (GA) で求める．シミュレーションの流れは以下のようになる．N はエージェント数，G は世代数を表す．

1. N 個の染色体をランダムな値で生成し，それを元に初期エージェントを生成する．
2. N 体の内 2 体のエージェントに囚人のジレンマゲームを行わせ，報酬から効用を計算し，エージェントが Q 学習を行うことを有限回数繰り返す．
3. 2 を一回の試合として，全てのエージェントの組み合わせで行う．
4. 各エージェントが得た報酬の合計をそれぞれの適応度として遺伝的操作を実行し，次世代のエージェントを作成する．
5. 2〜4 を指定した世代数繰り返す．

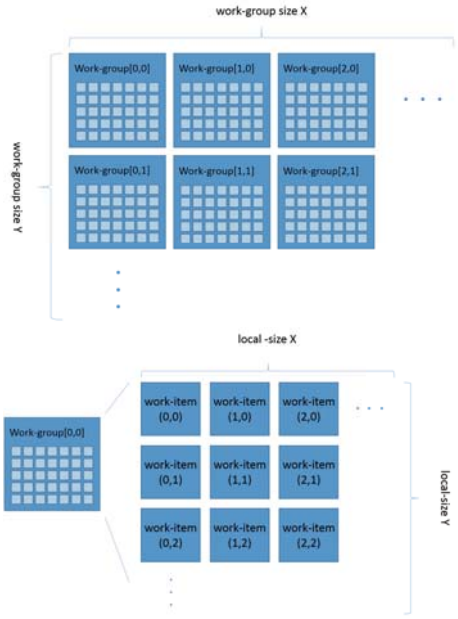


図 2: work item および work group 内の構造概念図

4. GPGPU を用いたシミュレーションの並列化

黒木ら [2] は 3 章で述べたシミュレーションの繰り返し囚人のジレンマゲームを行う部分を GPGPU により並列化した．しかし，GA は CPU 上で逐次計算されるためエージェント数が増えることで実行時間も大幅に増加する．そこで，GA についても GPU 上で並列に処理して実行速度の向上を図る．また，先行研究では遺伝子操作を実行し次世代のエージェントを作成する度に GPU に染色体データを転送し，囚人のジレンマゲームを実行した後その結果を CPU に返して適応度を計算する必要がある．これにより，先行研究では GA を実行する世代数の 2 倍の数 CPU と GPU で通信する．GA を並列化して一連のシミュレーション全体を GPU 上で実行することで CPUGPU 間の通信回数を削減することができる．GA を並列に処理することと、それによりシミュレーション全体を GPU 上で実行するという 2 つの観点から実行速度の向上を図る．本研究では先行研究同様 GPGPU のためのフレームワークとして OpenCL[5] を採用した．OpenCL では，以下のような work item，work group といった仕組みが扱われる．work item，work group の構造概念図は図 2 のようになる．

- work item: OpenCL における処理対象データの最小単位．この work item に対してカーネルと呼ばれるプログラムが実行される．work item 内でのみ共有できるプライベートメモリと呼ばれる記憶領域を持つ．
- work group: work item をいくつかのグループにしたもの．work group 内で共有できるローカルメモリと呼ばれる記憶領域を持つ．

GA の並列化手法として島モデルが知られている [6]．本研究では GPGPU による GA の並列化に島モデルを使用した．島モデルとは，染色体集合を島と呼ばれるいくつかの部分集合に分配し，それぞれの島で並列に GA を実行するとともに，

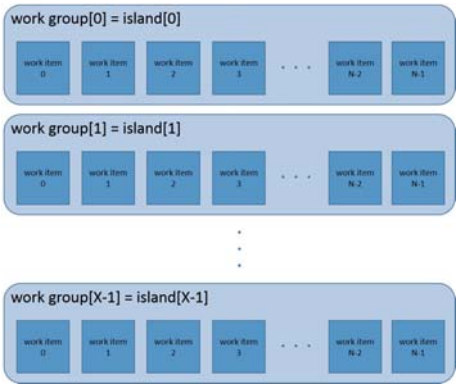


図 3: 島構造の概念図

一定世代数実行するなどの条件を満たした後，各島から一部の染色体を選んで移動させる手法である．本研究では，島の数を work group 数，島の染色体の数を work group 内の work item 数として並列化していく．work group 内は 1 次元の要素数 N 個とし，work group の概念図は図 3 のようになる．シミュレーションの流れは以下のようになる．

1. シミュレーションに必要な遺伝子情報，パラメータを GPU に転送する．
2. GPU 上で繰り返し囚人のジレンマゲームを実行し，その後染色体を島に順番に分配していく．
3. 適応度を元にルーレット法により島内の染色体を選択する．
4. 選択された染色体を交叉率 P_c で交差し，次世代の遺伝子とする．
5. 次世代の遺伝子それぞれに対し，突然変異率 P_m で突然変異を実行する．
6. 繰り返し囚人のジレンマゲームを実行し，次世代のエージェントの適応度を計算する．
7. 前回の移動から K 世代が経過していたら，適応度を元にルーレット法により選択された島内の染色体数の 5%の染色体をを隣の島の同じインデックスの染色体に上書きする．
8. 2〜6 を一定世代数繰り返す．
9. シミュレーション終了後に結果を GPU から CPU に転送する．

黒木らは全ての囚人のジレンマゲームが終了した時点で結果を CPU へ返し，CPU 上で遺伝的操作をした後に次世代のエージェントを GPU に転送し，囚人のジレンマゲームを行わせたが，本研究では上記のようにシミュレーション全体が GPU 上で実行される．

5. 実験

実験に使用したサーバーのデバイス情報は表 2 のとおりである．また，以下の実験において囚人のジレンマゲームの報酬を $T = 5$, $R = 3$, $P = 1$, $S = 0$ と設定し，Q 学習のパラメータを $\alpha = 0.25$,$\gamma = 0.5$,$\epsilon = 0.05$ 遺伝的アルゴリズムの交叉確率，突然変異確率を $P_c = 0.9$, $P_m = 0.01$ と設定する．

表 2: サーバーの CPU と GPU

ハードウェア	名称	個数
CPU	Intel Xeon E5-2650 v4	2
GPU	GeForce GTX 1080	1

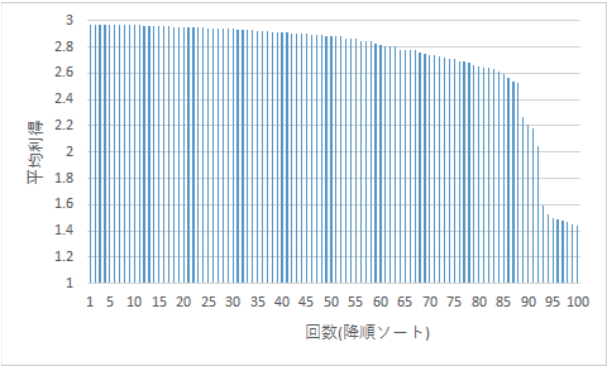


図 4: 平均利得

5.1 妥当性検証実験

まず，提案した手法の実行結果が正しいものであるかについての検証を行う．先行研究と同様に文献 [4] に従い，シミュレーションのエージェント数を 100，島内の染色体数を 100，島数を 1，繰り返し囚人のジレンマゲームを行う回数を 1000 回，世代数を 10000 と定めてシミュレーションを 100 回行った．相互協調が起こるとされる平均利得 2.7 以上の回数を見る．

結果は図 4 の通りである．グラフは 100 回行った実験を平均利得の降順にソートしたものである．100 回中 75 回が相互協調が起こるとされる平均利得 2.7 以上となった．文献 [4] では平均利得 2.7 以上の達成率が 8 割なので少し低い結果になったがほぼ同等の結果が得られた．GPGPU による GA の並列化は先行研究と同じように動作していると考える．

5.2 速度比較実験

従来手法と提案手法で実行速度を比較した．シミュレーションのエージェント数を 100，繰り返し囚人のジレンマゲームを行う回数を 1000 回，世代数を 1000，K を 5 と設定し，エージェント数を 100 ずつ増やして実行速度を計測して先行研究との比較を行った．今回の実験では最初に島を一つに設定し，島ごとのエージェント数を 100 に設定して島の数を 1 つずつ増加させて実験を行った．

速度比較の結果は図 5 の通りである．全てのエージェント数において本研究の実行速度が先行研究より向上した．先行研究では，GA を実行する際に適応度計算で囚人のジレンマゲームのシミュレーションを行う．つまり GA の世代数の 2 倍の数 CPU と GPU の間でデータ転送を行う．エージェント数 N とすると一回あたりの GA では，4 つの遺伝子を含むエージェントが N 体，エージェントの総当りの対戦の報酬や対戦したエージェントの id などをを一時保存するデータが $N(N-1)$ 個，その他パラメータのデータ転送が行われる．本研究では適応度計算を含む GA が全て GPU 上で行われているので上記のデータ転送は一回である. また，本研究では先行研究で逐次計算されていた GA が並列に処理されている．本研究では GA はエージェント数と同じ数の work item 上で並列に行われ

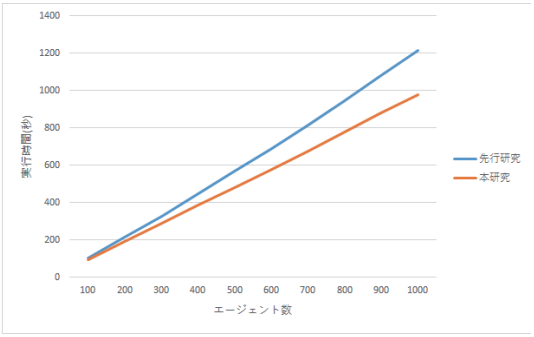


図 5: 速度比較

ており，エージェント数分並列化できていることになる．実際に CPU と GPU 間の通信速度が計測できたわけではないが，CPUGPU 間通信の回数を減らし，GA を並列化することでシミュレーションの実行速度の向上ができたと考えられる．

6. おわりに

GPGPU を用いて遺伝的アルゴリズムの並列化およびアルゴリズム全体を GPU 上で動かすことによる CPUGPU 間通信の削減により，シミュレーションの実行速度の向上が確認された．しかし，今回行ったアルゴリズム全体の並列化において，データ転送回数の削減と GA の並列化という 2 つの面において実行速度の向上が見られたが，それぞれにおいてどの程度速度が向上されたのか検証はできていない．GA を並列化することによる実行速度の向上とデータ転送回数の削減による実行速度の向上のどちらの恩恵が大きいのかの検証が今後の課題となる．

謝辞

本研究の一部は，JSPS 科研費 JP16K00302，堀科学芸術振興財団，および栢森情報科学振興財団の助成を受けて行われた．

参考文献

[1] ゼロからはじめる GPU コンピューティング,http://www.gdep.jp/page/view/248, (2018 年 2 月 1 日閲覧)

[2] 黒木是治:GPGPU を用いた 2 人ゲームにおける強化学習の高速化 情報処理学会第 79 回全国大会，2017

[3] Christopher J.C.H. Watkins and Peter Dayan: Technical Note: Q-learning. *Machine Learning,* Vol. 8, pp. 279–292, 1992

[4] Koichi Moriyama, Satoshi Kurihara, and Masayuki Numao: Evolving Subjective Utilities: Prisoner's Dilemma Game Examples. *Proc. 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS),* pp. 233–240, 2011

[5] The OpenCL Specification Version 2.2 Revision 06, https://www.khronos.org/registry/OpenCL/specs/opencl-2.2.pdf, 2016

[6] 棟朝雅晴: "遺伝的アルゴリズム —その理論と先端的手法—" 森北出版 2008

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

エージェントの社会性を考慮したヘドニックゲーム

Hedonic games considering agents sociality.

大田　一徳 *1　Nathanaël Barrot*2　櫻井　祐子 *3　横尾　真 *4
Kazunori Ota　Nathanaël Barrot　Yuko Sakurai　Makoto Yokoo

*1*4九州大学 Kyushu University　*2*4理化学研究所 RIKEN　*3産業技術総合研究所 AIST

We study hedonic games when each agent considers other agents as friends, enemies, or neutrals. In hedonic games, each agent has her preference over all coalitions that she can join. In this paper, we propose hedonic games under friends appreciation by allowing an agent to take into account the number of neutrals in her coalition. We show that the existence of neutrals does affect the stability of coalition structures, even though the impact of neutrals on preference is so small as to be meaningless compared to the impact of friends and enemies. More specifically, when each agent prefers coalitions with more neutrals, we show that neither a core stable coalition structure nor an individual table coalition structure can exist.

1. 序論

自律エージェントの集団に効率的に作業を実施させるために，社会的に望ましい分割方法を決定する問題は提携構造形成問題と呼ばれ，マルチエージェントシステムの研究者らによって盛んに研究が行われている．ヘドニックゲームは提携構造形成問題の一種であり，各エージェントは自身が所属する可能性がある提携に対してのみ選好を持つ場合を考えるゲームである [Dreze 80]．応用事例として，ソーシャルネットワークの分類，共同研究チーム形成等がある．

ヘドニックゲームでは，エージェント数の増加に伴って，各エージェントが所属可能な提携数が指数関数的に増加するため，エージェントの選好の記述量が課題となっている．そこで，ゲームの簡潔記述法として，他のエージェントを友達／中立者 (無関心な人) ／敵対者に分類し，提携内の友達が多い方が嬉しく，また，友達が同数の場合は敵対者が少ない方が嬉しいという比較方法に基づいて提携間の順序を決定する方法 (友達優先方式) が提案されている [Ohta 17]．

本論文では，従来の友達優先方式では考慮されていなかった，各エージェントの社会性を導入することで新たなヘドニックゲームを提案する．より具体的には，提携内に友達が同数であっても，中立者が多いか／少ないかによって選好が異なるというエージェントの社会性を考慮する．

また，提携構造問題では決定された提携構造の安定性について議論が行われており，ヘドニックゲームにおいても様々な安定性の概念が存在し，研究が行われている．本論文では，中立者の影響は友達や敵対者の影響よりも十分に小さいにも関わらず，新たなヘドニックゲームの下では，既存の友達優先方式では保証されていた安定性が保証できなくなることを示す．

2. モデル

本章では，本論文で扱うモデルを説明する．まず，記法の定義を与える．

エージェント集合を N，$C \subseteq N, C \neq \emptyset$ を満たすエージェント集合 C を提携，N の分割 π を提携構造とする．エージェント i を含む提携の集合を $\mathcal{N}_i = \{C \subseteq N \mid i \in C\}$，提携構造 π の要素のうち，エージェント i を含む提携を $\pi(i)$ とする．任意のエージェント i は $\mathcal{N}_i$ 上の選好関係 $\succeq_i$ を持つ．$\succeq_i$ は，i が属する全ての提携に対する好ましさを表現し，反射性，完備性，推移性を満たすものとする．$P = (\succeq_1, \succeq_2, \ldots, \succeq_n)$ は全ての i の選好の組を示し，$\mathcal{P}$ を全選好の組の集合とする．ヘドニックゲームは，エージェントの全体集合 N と全てのエージェント $i \in N$ の選好の組 P の組合せ (N, P) で表現される．

次に，本論文で議論する安定性を紹介する．

定義 1 (個人安定) ヘドニックゲーム (N, P) と提携構造 π を与える．$j \in C$ に対して，$C \cup \{i\} \succ_i \pi(i)$ かつ $C \cup \{i\} \succeq_j C$ を満たすような，$i \in N$ かつ $C \in \pi \cup \{\emptyset\}$ が存在しないならば，提携構造 π は個人安定性を満たすという．

定義 2 (ブロック) ヘドニックゲーム (N, P) と提携構造 π を与える．ある $C \subseteq N$, 全ての $i \in C$ に対して，$C \succ_i \pi(i)$ ならば π は C に強くブロックされるという．また，ある $C \subseteq N$, 全ての $i \in C$ に対して $C \succeq_i \pi(i)$ かつ，少なくとも 1 人以上の $j \in C$ に対して $C \succ_j \pi(j)$ ならば π は C にブロックされるという．

定義 3 (コア安定性／強コア安定性) ヘドニックゲーム (N, P) と提携構造 π を与える．π が任意の C に強くブロックされないならば提携構造はコア安定性を満たすという．また，π が任意の C にブロックされないならば提携構造は強コア安定性を満たすという．

以下, エージェントの選好を定義する．エージェント i に対して, $F_i = F(\succeq_i) = \{j \in N \backslash \{i\} \mid \{i, j\} \succ_i \{i\}\}$ とする．エージェント i 自身は F_i の要素とする．$j \in F_i$ である場合 j は i の友達とみなす．次に, $E_i = E(\succeq_i) = \{j \in N \backslash \{i\} \mid \{i\} \succ_i \{i, j\}\}$ とする．$j \in E_i$ である場合 j は i の敵対者とみなす．最後に, $\perp_i = N \setminus (F_i \cup E_i)$ とする．$j \in \perp_i$ である場合 j は i の中立者である．

定義 4 (社会性を考慮した友達優先方式) $P \in \mathcal{P}$ について，i かつ全ての $C, C' \in \mathcal{N}_i$ に対して $C \succeq_i C'$ が成り立つとき，以下の 3 つの条件のうちのいずれかが成り立つ P の決定方式を社交

連絡先: 大田一徳，九州大学大学院システム情報科学府，812-0395 福岡県福岡市西区元岡 744 番地，(092)802-3576，ota@agent.inf.kyushu-u.ac.jp

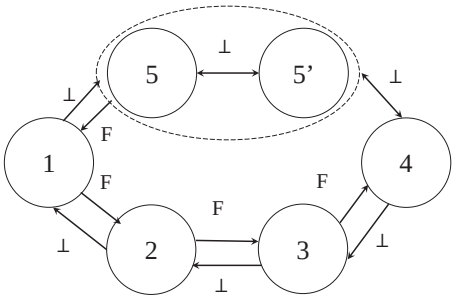


図 1: コア安定／個人安定な提携構造が存在しない社交的なエージェント間の選好関係

的な (内向的な) エージェント間での友達優先方式と呼ぶ．(1). $|C \cap F_i| > |C' \cap F_i|$，(2). $|C \cap F_i| = |C' \cap F_i| \wedge |C \cap E_i| \leq |C' \cap E_i|$, または (3). $|C \cap F_i| = |C' \cap F_i| \wedge |C \cap E_i| = |C' \cap E_i| \wedge |C \cap \perp_i| > |C' \cap \perp_i|$ ($|C \cap \perp_i| < |C' \cap \perp_i|$)

本論文では，ゲームを有向グラフ $G_{F\perp} = (V, A_F \cup A_\perp)$ で表現する．頂点の集合を $V = N$ とし，有向辺の集合を $A_F = \{(i,j) \in N \times N \mid i \neq j, j \in F_i\}$, $A_\perp = \{(i,j) \in N \times N \mid i \neq j, j \in \perp_i\}$, とする．友達への有向辺を F, 中立者への有向辺を $\perp$ とする．

3. 社交的なエージェントの場合

本章では，エージェントの社交性を考慮することで既存モデルでは保証されていた安定性が保証できなくなることを示す．

定理 1 エージェントが社交的な場合の友達優先方式では，コア安定な提携構造が常に得られることを保証しない．

証明 $N = \{1,2,3,4,5,5'\}$ とする．$i \in \{1,2,3\}$ について，i は $i+1$ を友達とみなし，$i+1$ は i を中立者とみなす．$4, 5, 5'$ はお互いを中立者とみなす．$5, 5'$ は 1 を友達とみなし，1 は $5, 5'$ を中立者とみなす．他の選好関係は敵対者とする．

提携構造 π がコア安定であると仮定して矛盾を導く．(1).5 と 5' が別の提携に属する場合 (2).5 と 5' が同じ提携に属する場合に場合分けを行う．まず，5 と 5' が分かれている場合に，5 が属す可能性のある提携は {5}, {1, 5}, {4, 5'} の 3 通りである．選好関係は対称的なので，5' が属す可能性のある提携も上記の 3 パターンである．しかし，この全ての組合せにおいて {4, 5, 5'} がブロッキングコアリションとなる．よってコア安定な提携構造は存在しない．次に (2) のケースでコアを満たす π は存在しないことを示す．5 と 5' は同じ提携に属し，他のエージェントへの選好関係も同一であるため，単独提携とみなす．3 エージェント以上が同じ提携に属す場合，そのうち少なくとも 1 エージェントは友達が 0 人で敵対者が 1 人いる状態となるため，単独提携へ逸脱を起こす．よってコア安定な π は存在しない．次に，2 エージェント，2 エージェント，単独提携と提携を組む場合を考える．この場合，1 (2, 3, 4) が単独提携だった場合，$\{1,5,5'\}$ ($\{1,2\}, \{2,3\}, \{3,4\}$) がブロッキングコアリションとなり，$\{5,5'\}$ が単独提携だった場合は，$\{4,5,5'\}$ がブロッキングコアリションとなる．従って，各エージェントの選好が図 1 によって表現される場合に，コアを満たす π は存在しない．

定理 2 エージェントが社交的な場合の友達優先方式では，個人安定な提携構造が常に得られることを保証しない．

証明 各エージェントの選好関係は定理 1 で示した選好関係を用いる．提携構造 π が個人安定であると仮定して矛盾を導く．

まず，5 と $5'$ が異なる提携に属す場合を考える．この場合，$\pi(4) \cap \{1,2\} = \emptyset$ でない場合は，4 が単独提携へと逸脱を起こす．また，3 も同様に，$\pi(1)$ と同じ提携に属す場合，3 が単独提携へと逸脱を起こす．もし，1 が $\pi(2)$ が同じ提携に属さないならば，$\{1,5\}, \{1,5'\}$, または $\{1,5,5'\}$ へと逸脱を起こす．よって，1 は $\pi(2)$ と同じ提携に属さなければならないため，$\pi(1) \cap \{5,5'\} = \emptyset$ が成り立つ．また，5 または 5' が単独提携へと逸脱を起こすため，3 は $\pi(5)$ と同じ提携に属さない．よって，5 と 5' はどちらも単独提携となるか，どちらか一方が $\pi(4)$ と同じ提携に属す．もし，5 と 5' が単独提携だったならば，{5, 5'} へと逸脱を起こす．また，$\pi(4) \cap \{5,5'\} \neq \emptyset$ の場合，{4, 5, 5'} へと逸脱を起こす．したがって，5 と 5' は同じ提携に属す必要がある．この場合は，定理 1 と同様の議論によって，個人安定の非存在性を示すことが可能である．

4. 内向的なエージェント

本章では，内向的なエージェントのみが存在する場合において，強コア安定な提携構造が常に存在することを示す．

定理 3 (1) エージェントが内向的である場合の友達優先方式では，強コア安定性を満たす提携構造の存在を常に保証し，(2) グラフ $G_F = (N, A_F)$ から強連結成分を取り出す操作によって，多項式時間で計算可能である．

紙幅の都合上，本証明は省くが，文献 [Dimitrov 06] と同じアイデアを用いることで証明が可能である．

5. 結論

本論文では，エージェントの社会性を考慮した新たなヘドニックゲームを提案し，提携構造の安定性に関する議論を行った．今後の課題として，中立者が多い状況が望ましいとき，エージェント間の選好に対称性が存在する場合のコア安定／個人安定な提携構造の存在性の解析が挙げられる．

謝辞

本研究は JSPS 科研費 (JP17H00761) の助成を受けたものです．深く感謝致します．

参考文献

[Dimitrov 06] Dimitrov, D., Borm, P., Hendrickx, R., and Sung, S. C.: Simple priorities and core stability in hedonic games, *Social Choice and Welfare*, Vol. 26, No. 2, pp. 421–433 (2006)

[Dreze 80] Dreze, J. H. and Greenberg, J.: Hedonic coalitions: Optimality and stability, *Econometrica*, pp. 987–1003 (1980)

[Ohta 17] Ohta, K., Barrot, N., Ismaili, A., Sakurai, Y., and Yokoo, M.: Core stability in hedonic games among friends and enemies: impact of neutrals, in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)* (2017)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

脅威情報の評価を支援する知的エージェント

Intelligent agents supporting to estimation of cyber threat intelligence

川口英俊　　石原裕一
Hidetoshi Kawaguchi　Yuichi Ishihara

NTT ネットワークサービスシステム研究所
NTT Network Service Systems Laboratories

In this paper, we will propose method of estimating cyber threat intelligence with intelligent agents. Population of security operators are insufficient though cyber threat intelligence is increasing. Therefore, there is a possibility that the cyber threat intelligence can not be estimated properly. Our proposed intelligent agent is composed of three modules, Estimate module, Split module and Database. We will define a problem of Estimate module and explain how to apply methods of machine-learning. And we will experiment with real data to verify feasibility of our proposed intelligent agent.

1. はじめに

本稿では，脅威情報の評価を支援する知的エージェント技術の提案を行い，実データを用いた実験とその考察を行う．

脅威情報とは，サイバー攻撃やマルウェアなどの悪意ある通信やソフトウェアについての構造化された情報のことを指す．具体的には，IPS(Intrusion Prevention System) に用いるシグネチャ，STIX[STIX 18] や MISP[MISP 18] の形式で記述された情報が該当する．

脅威情報の評価とは，脅威情報に対して組織のセキュリティ運用者がその運用対象への意味を見出すことである．例えば，IPS のシグネチャにおける評価の場合では，ロギングや遮断などの通信とのマッチ時の挙動の設定方法の判断のことを指す．

セキュリティ市場の盛り上がりから脅威情報自体の数が増えることが予想され，加えてセキュリティ人材が不足するという調査結果も出ている [METI 16]．そのため，将来的には脅威情報に対して十二分に評価を行えなくなり，セキュリティインシデントが増加する問題の発生が考えられる．

本研究では，より多くの脅威情報への評価を行うことを目的に，その評価を自動化するソフトウェアを実現していく．その自動化ソフトウェアを知的エージェントと呼称する．

2 章と 3 章では本研究の全体像として，知的エージェントの動作イメージを説明しその構成について提案する．4〜6 章では，知的エージェントを構成する一部のモジュールについて問題設定を行い脅威情報の変換方式および評価の手法について述べる．7 章では，実現可能性を検証するための実データを用いたシミュレーション実験とその考察について述べる．

2. 知的エージェントの動作イメージ

知的エージェントは脅威情報とセキュリティ運用者の行った評価のデータセットから自動評価のためのパターンを自動で生成し，セキュリティ運用者が評価を行わなかった脅威情報に対して代理で評価を実行する．このパターンを生成することを本稿では学習と表現する．動作イメージを図 1 に示す．

脅威情報の評価は『A』と『B』の 2 種類があるものとする．

連絡先: 川口英俊，NTT 武蔵野研究開発センタ，〒 180-8585 東京都武蔵野市緑町 3-9-11，E-mail: hidetoshi.kawaguchi.my@hco.ntt.co.jp

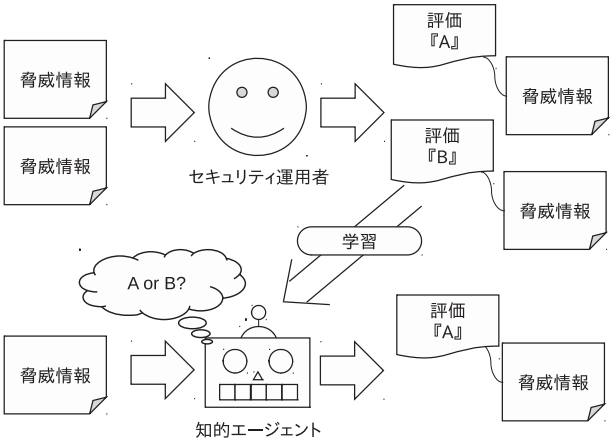


図 1: 知的エージェントの動作イメージ

セキュリティ運用者は，脅威情報を受け取ったあと，その情報を読み取り,『A』か『B』のどちらかを選択する．

知的エージェントは脅威情報とセキュリティ運用者による評価を基に学習する.『A』と評価された脅威情報を取得した場合はセキュリティ運用者と同様に『A』という評価を行うようにパターンを作る.『B』の場合も同様である．

学習が進み，許容できる精度を担保できた場合は，脅威情報をエージェントへ転送し，評価作業の一部を委譲する．セキュリティ運用者の評価数を減らし，結果としてエージェントなしであれば手が回らなかった脅威情報への評価も可能となる．

学習後も生成したパターンの陳腐化を防ぐために，継続的な学習が必要となる．そのため，セキュリティ運用者もある程度は脅威情報を評価する必要がある．

3. 知的エージェントの構成

知的エージェントの構成とデータフローを図 2 に示す．エージェントを「評価モジュール」「割り振りモジュール」「データベース」の 3 つのモジュールで構成する．

3.1 評価モジュール

データベースに保存された脅威情報とセキュリティ運用者自身による評価の対に基づきパターンを生成し，そのパターンを

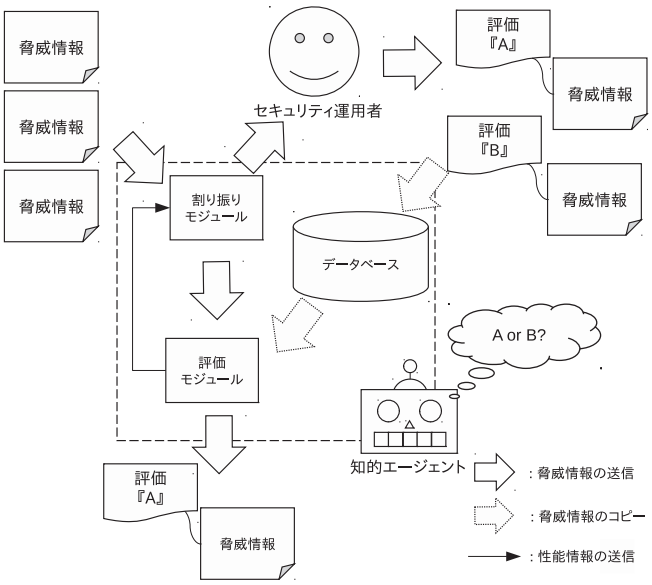


図 2: 知的エージェントの構成とデータフロー

用いて脅威情報に対して評価を行うモジュールである．このパターンのモデリングとその生成方法が研究課題となる．

また，割り振りモジュールに対して評価を行うパターンの性能を算出するために用いる性能情報の送信も行う．パターンの生成に機械学習を用いた場合は，そのパターン生成の過程で算出された精度や混同行列などが考えられる．

3.2 データベース

脅威情報とセキュリティ運用者自身による評価の対を記憶する媒体である．データベースの構造は実装時の課題となる．

3.3 割り振りモジュール

評価モジュールからのパターンの性能を算出するためのデータを受け取り，性能を算出し，その算出した性能の結果を基にセキュリティ運用者か評価モジュールへの脅威情報の振り分けを行うモジュールである．この際，セキュリティ運用者へ評価を多く回すとよりエージェントの性能は良くなるが，その分自動化の効果が薄れるジレンマを抱えていると考えられる．

性能は定量的な指標であり，脅威情報の評価の内容により算出方法を設計する必要がある．機械学習の評価に用いられる正解率，適合率，再現率などが単純な性能の例として考えられる．

4. 評価モジュールの問題設定

知的エージェントを実現するための方法を考える上で，まずは以下のように評価モジュールに着目して問題を整理する．脅威情報は逐次生成されるため，時系列を意識した問題設定を行う必要があるが，問題をシンプルにするために時系列を意識しないものとする．

$$Y' = \mathcal{F}(X'; Y, X) \tag{1}$$

$\mathcal{F}$ は評価の付けられていない脅威情報を受け取り，それに対応する評価を導出する評価モジュールを関数として表した．本稿では，この $\mathcal{F}$ を明確にしていく．X' は評価対象となる脅威情報のリスト，Y' は X' の脅威情報に対応して推定された評価のリストを表す．Y と X はそれぞれ，脅威情報とセキュリティ運用者による評価のデータセットである．

評価モジュールを実現するための課題は，以下の 2 点である．

- 【課題 1】脅威情報の特徴ベクトルへの変換
- 【課題 2】評価モジュールの内部アルゴリズム

【課題 1】セキュリティ運用者の評価を推定する上では，脅威情報を計算できるデータの形式へ変換する必要がある．

【課題 2】$\mathcal{F}$ の内部アルゴリズムを検討する必要がある．セキュリティ運用者の評価 Y と脅威情報 X は与えられているものとして，X' を得た時に X' の各要素に対応する評価を Y' として推定する問題と合致するように手法を構成する．

5. 脅威情報の特徴ベクトルへの変換

STIX や MISP の記述形式，実際に運用されている IPS のシグネチャの形式を確認し，以下の 2 つの特徴を抽出した．

- 【特徴 1】多くの名義特徴量で構成されている．
- 【特徴 2】自然言語で説明されている属性がある．

【特徴 1】にある名義特徴量は，IP アドレス，ポート番号，マルウェアのタイプや名前などのことを指している．IP アドレスやポート番号などは数値ではあるが，その数値の大きさに意味はなく記号のひとつでしかない．脅威度や優先度などの順序特徴量もあったが，それも細かな数字ではなく low,middle,high などの 3 段階で表していることが多い．いずれにしても数値ではなく記号的表現であるため，なんらかの方法で数値へ変換する必要がある．

【特徴 2】については，その脅威情報の概要や補足などを説明している文章のことである．シグネチャの場合は，ルールメッセージや概要などである．STIX の場合では，攻撃者の情報や，攻撃の手口などの情報が項目ごとにテキストで説明されている．脅威情報の中でも様々な種類があるが，人が見ることが前提となっているからであると考えられる．

これらの特徴を変換する手法はいくつか存在するが，本稿では名義特徴量に対しては One-hot エンコーディングを，自然言語の属性には TF-IDF を用いて変換することを提案する．

5.1 One-hot エンコーディング

各シンボルの特徴量の一意な値毎にダミー特徴量を新たに作成する手法である．例えば，ある特徴量は A,B,C の 3 つのいずれかの記号を取るものとする．その場合は，本来は 1 つの特徴量であるものが 3 つ分の特徴量になる．

5.2 TF-IDF

TF-IDF は自然言語で記述された文章を特徴ベクトルへ変換する手法のひとつである．複数の文章において，単語の出現頻度が高いものは重みを減らすなどの工夫がなされている．

6. 評価モジュールの内部アルゴリズム

X' に対応する Y' を導出するために使える情報は X と Y である．実現する方法はいくつも考えられるが，本稿では二つの案を紹介する．

一つ目の案としては，X と Y を対応付ける関数 $\mathcal{G}$ を推定し，X' の各要素を $\mathcal{G}$ の入力として，その出力のリストを Y' とすることである．

二つ目の案としては，X と X' を連結したリスト上でグループ分けを行い，X' の各要素について，同一グループ内に存在する X の要素に対応する Y の要素を Y' の要素の評価とみなして，Y' を構成する方法である．

一つ目の案については「教師あり学習＋パラメータ調整」を，二つ目の案については「制約付きクラスタリング」を適用できる．以下にそれぞれの手法の説明を述べる．

6.1 教師あり学習＋パラメータ調整

教師あり学習は，与えられた入力データリストとそれに対応する出力データリストの関係をつくる関数を導出する方法論であり，【課題 2】で示した問題の特徴に合致した手法である．ただし，導出するためにはハイパーパラメータを設定する必要がある．本来このハイパーパラメータは機械学習の知見を持つエンジニアが調整するものである．しかし，本研究のエージェント技術の目的を考えるとエンジニアが調整する労力は可能な限り削減したいため，ヒューリスティックにハイパーパラメータの調整を行う機構が必要となる．

6.2 制約付きクラスタリング

「制約付きクラスタリング」とは，与えられたデータをグループ分けするクラスタリング技術へ must-link と cannot-link という制約条件を加えたものである [Kamishima 06]．

セキュリティ運用者から取得した評価と脅威情報の対から，must-link と cannot-link を構成することができる．例えば，同一の評価が付けられた脅威情報同士は must-link を構成し，違う評価が付けられた脅威情報同士は cannot-link が構成される．これらの 2 つの制約を加えた上で，X と X' をまとめて制約付きクラスタリングを適用することにより，未評価の脅威情報へ評価を導出することができる．

7. シミュレーション実験

IPS のシグネチャとその評価を対象に，知的エージェントの実現可能性を検証するシミュレーション実験を行う．評価精度と学習に用いたデータ数の相関を確認することによる効果の確認や，適用するアルゴリズム毎の特性に基づいた考察を行うことにより，今後の足がかりとなる知見を得る．

特徴ベクトルへの変換方法は，名義特徴量のみ (One-hot エンコーディングを適用)，自然言語の属性のみ (TF-IDF を適用)，全属性 (One-hot エンコーディングと TF-IDF をそれぞれ適用) の 3 種類を試行する．これにより，それぞれの変換方法の影響を確認する．

「教師あり学習＋パラメータ調整」と「制約付きクラスタリング」はそれぞれ無数に手法が存在する．本実験では代表的な手法と考えられる「教師あり学習＋パラメータ調整」へは多クラス分類線形 SVM(Support Vector Machine) と 10-交差検証を，「制約付きクラスタリング」へは COP-kmeans を用いる．

特徴ベクトルへの変換は 3 種類，適用する内部アルゴリズムの手法は 2 種類であるため，その全組み合わせの合計 6 種類の方法を試行することになる．

7.1 対象となる実験データとその評価について

対象となる実験データは，実際に運用されている IPS に用いられるシグネチャとその評価の実データセット 2579 件である．そのシグネチャは IPS のセキュリティベンダより送信されてくるものである．ただし，IPS のセキュリティベンダは送信先，つまり受け取り手の環境については一切考慮せず送信を行う．そのため，必ずしもセキュリティ運用者にとって必要であるとは限らない．

セキュリティ運用者はその IPS のシグネチャの情報を確認しながらひとつずつシグネチャの評価を行っている．本実験データでの評価は，シグネチャを 12 種類に分類する作業を指すものとする．全データセットの評価の内訳は 1123,12,570,364,105,133,39,55,66,75,16,21 である．比較的偏りのある問題と考えられる．加えて，それぞれの分類には「ロギング」「通知」「遮断」のそれぞれ ON,OFF が設定されており，この分類を行うことで IPS への設定方法が決定されるのである．以降，本実験で指す脅威情報の評価とは，IPS のシグネチャを分類することと同義とする．

本実験データのシグネチャは，1 つの自然言語と 13 個の名義特徴量，そして low, medium, high の 3 つの記号で表される Priority という順序特徴量の属性 1 つの合計 15 個で構成される．ID などの明らかに評価と相関がないと考えられるものは省いている．

7.2 シミュレーション設定

本実験は，すでに評価された実データのセット 2579 件を用いて実際にエージェントに評価を代行させた場合に，どの程度の精度を得ることができるかを測定するシミュレーションに相当する．全データセットを学習データとテストデータの 2 つにランダムに分け，学習データを用いてエージェントを学習させ，テストデータを用いて精度を測定する．学習データ数，つまり人手で評価を行ったと想定する数は 1,2,3,5,7,10,20,30,50,70,100,200,300,500,700,1000,2000 のパターンで行う．テストは全データセット数 2579 のうち，学習データとして用いられなかったデータを用いる．そのため，テストデータ数は 2579 件から学習データ数を減じた値となる．ここでのテストデータ数は，即ちエージェントが自動で評価した脅威情報の数となり，技術による支援効果の定量的数値となる．翻すと，学習データ数の少なさは，支援効果の大きさを表す．当然ながら，学習データ数が 10 以下の場合は 12 種類の分類は確保できない．十分な学習データ数であっても完全にランダムに学習データとするので，いずれの学習データ数でもすべての分類を確保できるとは限らない．ただし，学習データ数によっては適用できない手法もある．

7.3 適用するアルゴリズム

7.3.1 多クラス線形分類 SVM ＋ 10-交差検証

線形 SVM は教師あり学習の中でも広く持ちいられている基本的かつ強力な予測モデルである．通常の SVM 単体では，その仕組み上 2 クラスを分類することしかできない．そこで本実験では，線形識別器を，学習データに含まれる分類数 n の全組み合わせ $\frac{n(n-1)}{2}$ 個用意し，それを基に分類を推定する One-Versus-One という方法をとる．本実験での SVM のハイパーパラメータは逆正則化パラメータ C のみである．

それぞれの線形識別器を学習する線形 SVM に対して，ハイパーパラメータの最適化手法の一つであるグリッドサーチを用いる．グリッドサーチは教師あり学習を行う上でのハイパーパラメータの最適化手法のひとつである．指定したパラメータの組み合わせを総当りで試行し，その中から最も精度の良いハイパーパラメータを導出する．

C は 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0 の 8 種類で性能評価し，パラメータを探索する．また性能評価には 10-交差検証を用いるものとする．

7.3.2 COP-kmeans

COP-kmeans はクラスタリングアルゴリズムである K-means を背景知識を考慮できるように拡張した手法である [Wagstaff 01]．通常の K-means はハイパーパラメータとしてクラスタ数を設定するが，COP-kmeans はそれに加えて must-link と cannot-link の 2 つを制約として加えてクラスタリングを行う．must-link は必ず同じクラスタに属すデータのペアを規定し，cannot-link は同じクラスタに属さないデータのペアを

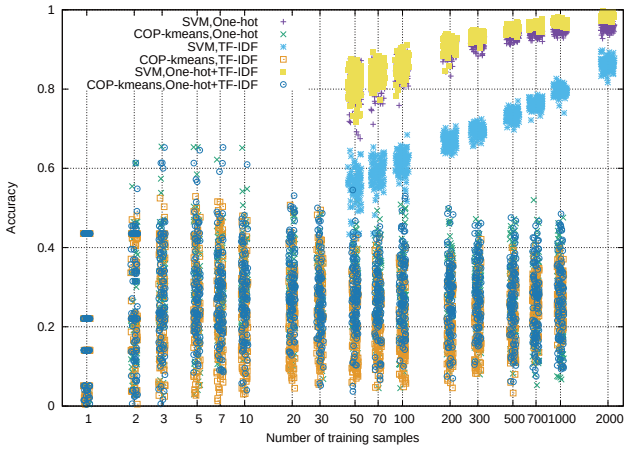


図 3: 実験結果 (精度と学習データ数の関係)

規定する．COP-kmeans の基本的なアルゴリズムは K-means と同じである．しかし，must-link を基に初期クラスタを形成し，その後は cannot-link を満たす範囲内でセントロイドを更新しながらクラスタを形成していく．

7.4 実験結果

実験結果を図 3 に示す．縦軸が精度であり，横軸が人手で評価したと想定する脅威情報の数，つまり学習データ数である．また，精度の分布の視認性を高めるという理由で，横軸については少しだけずらしているが，横軸のある最も近傍の学習データ数で実験を行っている．ひとつの点が 1 シミュレーションに相当し，学習データは 1 シミュレーション毎に全データから無作為にサンプリングしている．

多クラス線形分類 SVM+10-交差検証については，学習データ数が少なすぎたことが原因で 30 以下の学習データ数では実験を行えなかった．COP-kmeans は学習データ数 2000 の場合は計算が収束しなかったため実験を行えていない．

7.5 考察

本実験の代表的な結果は，内部アルゴリズムへは多クラス線形分類 SVM+10-交差検証を，特徴ベクトルへの変換方法は全属性を対象としたもの (One-hot エンコーディングと TF-IDF をそれぞれ適用) である．精度 90% で十分であれば，2579 件中 2379 件の脅威情報を評価でき，精度 98% を求められるのであれば，579 件 の脅威情報を評価できたという結果となった．実運用へ適用する場合は，現場のセキュリティ運用者が求める精度をヒアリングする必要があるが，本実験対象においては充分な実現可能性があると考えられる．

以下に，特徴ベクトルへの変換と適用した内部アルゴリズムについての 2 つの側面から詳細な考察を述べる．

7.5.1 特徴ベクトルへの変換について

3 種類のデータ変換方法を試行した結果，全属性を対象とした One-hot エンコーディングと TF-IDF の組み合わせが最も良い結果となった．COP-kmeans については違いが分かりづらいが，多クラス線形分類 SVM+10-交差検証ではその違いが顕著である．本問題の場合は，全属性へ対して多クラス線形 SVM が充分な表現力を持っていたことが考えられる．

One-hot エンコーディングと TF-IDF を比較したとき，前者のほうが精度が良いことがわかる．このセキュリティ運用者の場合は，より名義特徴量を重視しながら評価を行っていると考えられる．しかし，学習データ数の増加にともない TF-IDF の精度も上昇していることから，少なからず自然言語の属性も加味していることがわかる．

7.5.2 内部アルゴリズムについて

基本的には多クラス線形 SVM+10-交差検証を使えば良いと考えられる．学習データ数が 30 以下の場合は自動評価を行えない場合があるが，それも全データ 2579 件に対して極端に少ない場合であるため無視しても良いと見なせる．加えて，学習データ数が 2000 の場合でも 579 件は自動評価できるため，充分な支援効果があると考えられるためである．

COP-kmeans については，いずれの結果も学習データ数 10 までわずかに全体的に精度が向上し，20 以上ではほぼかわらない結果となった．加えて，シミュレーション毎に精度にバラツキがある．これは基となった K-means が各クラスタの分布が同程度かつ球状になるという前提に則った手法であるからと考えられる．本実験データの分類は，上述したように評価毎にかなりの偏りがある．そのため，クラスタ毎に過不足が発生し精度が伸びなかったと考えられる．バラツキがあるのは COP-kmeans の評価のサンプリングという側面もあるが，COP-kmeans の初期値も依存していると考えられる．特徴ベクトルがスパースな空間であることも，精度が伸びなかった原因のひとつの可能性がある．

8. おわりに

本稿では，研究の全体像として脅威情報を評価するための知的エージェントの概要を述べた．知的エージェントを構成するモジュールについて述べ，まずは時系列を意識せずに問題設定を行った．その問題設定に，機械学習分野の代表的な手法を適用することを示した．実現可能性を検証をするための実験を行い，知的エージェント実現へ向けての足がかりとなる知見を得ることができた．

今後は技術的側面と実問題への適用的側面の 2 つの側面から課題へ取り組む必要がある．技術的側面としては，より少ない学習データ数での精度改善と本稿で扱った実験データとは別の実験データでの検証が必要である．実問題への適用的側面としては，時系列を意識した問題設定と実際のセキュリティ運用者との協調方法の検討が残されている．

参考文献

[METI 16] IT 人材の最新動向と将来推計に関する調査結果, 経済産業省
http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_report_summary.pdf

[STIX 18] STIX-Structured Threat Information Expression
https://stixproject.github.io/

[MISP 18] MISP-Malware Information Sharing Platform and Threat Sharing
http://www.misp-project.org/

[Kamishima 06] 神嶌 敏弘. 絶対クラスタリングと相対クラスタリング. 人工知能学会全国大会 (第 20 回) 論文集,2A1-01 (2006)

[Wagstaff 01] Wagstaff, K., Cardie, C., Rogers, S.,and Schroedl, S.: Constrained K-means Clustering with-background Knowledge, in Proc. of The 18th Int'l Conf. on Machine Learning, pp.577-584(2001)

[2P2] [General Session] 7. Agent

座長:落合 友四郎（大妻女子大学）

Wed. Jun 6, 2018 1:20 PM - 3:00 PM Room P (4F Emerald Lobby)

[2P2-01] Accelerate Deep Q Network by weighting experiences

○Kazuhiro Murakami[1], Koichi Moriyama[1], Atsuko Mutoh[1], Tohgoroh Matsui[2], Nobuhiro Inuzuka[1]
（1. Nagoya Institute of Technology, 2. Chubu University）

1:20 PM - 1:40 PM

[2P2-02] Effect of Asset Price Fluctuation and Inter-bank Lending and Borrowing Network on Chain Bankruptcy of Bank

○Ryo Hamawaki[1], Kiyoshi Izumi[1], Hiroki Sakaji[1], Hiroto Yonenoh[1] （1. The University of Tokyo）

1:40 PM - 2:00 PM

[2P2-03] Trade-off between search time and responsiveness of MCTS in a ghting video game

○Wataru Kamegaki[1], Koichi Moriyama[1], Atsuko Mutoh[1], Nobuhiro Inuzuka[1] （1. Nagoya Institute of Technology）

2:00 PM - 2:20 PM

[2P2-04] 3D L-Systems Path-finding

○Djuned Fernando Djusdek[1], Yoshiteru Ishida[1] （1. Toyohashi University of Technology）

2:20 PM - 2:40 PM

[2P2-05] Effects Analysis of CAR Regulations on Financial Markets using Artificial Market Simulations

○Masanori HIRANO[1], Hiroto YONENOH[2], Kiyoshi IZUMI[2] （1. Faculty of Engineering, The University of Tokyo, 2. School of Engineering, The University of Tokyo）

2:40 PM - 3:00 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

経験データ重み付けによるDeep Q Networkの高速化

Accelerate Deep Q Network by weighting experiences

村上 知優 *1 Kazuhiro Murakami　森山 甲一 *1 Koichi Moriyama　武藤 敦子 *1 Atsuko Mutoh　松井 藤五郎 *2 Tohgoroh Matsui　犬塚 信博 *1 Nobuhiro Inuzuka

*1名古屋工業大学　大学院工学研究科　情報工学専攻
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

*2中部大学　生命健康科学部　臨床工学科
Department of Clinical Engineering, College of Life and Health Sciences, Chubu University

Deep Q Network (DQN) is a reinforcement learning method using deep neural network to approximate Q-function. Literature tells us that DQN can choose better actions than humans. However, it needs much time to learn such actions. DQN learns its actions by using tuples of (state, action, reward, next state), called experiences, sampled from its memory. DQN samples them uniformly randomly, but the experiences are skewed. It results in slow learning because frequent experiences are redundantly sampled while infrequent ones hardly.

This work mitigates the problem by weighting experiences based on their frequency and manipulating their sampling probability. In a video game environment, the proposed method learned good actions faster than DQN.

1. はじめに

近年，工場における部品の取り付け作業や事務作業など，様々な場面で自動化が進められている．従来は人間が動作ロジックを全て定義することで行われてきたが，現実世界のより複雑なものを自動化するにあたって，動作ロジックを全て定義することは非常に難しく，現実的でない．こうした背景のもと，動作ロジック自体を機械が獲得する機械学習の分野が注目されており，中でも与えられた環境下で最適な行動を獲得する強化学習が大きな成果を挙げている [1]．また，近年画像認識や自然言語処理，音声認識などの分野で深層学習が大きな成果を挙げている [2]．深層学習は入力データから特徴を抽出することに長けており，人間には認識できないような特徴を抽出することができる可能性を秘めている．そして深層学習と強化学習を組み合わせた深層強化学習の手法として，Deep Q Network(DQN)[3][4] が提案された．DQN は Atari2600[5] のゲーム画面を入力とし，得られたスコアを報酬とすることで，上級者を上回るスコアを出せる行動を獲得するに至っている．しかし，深層強化学習は学習速度が遅く，その原因の一つに観測したデータ群（経験データ）の出現率が異なることが挙げられる．DQN は観測したデータ群（経験データ）を有限なメモリ内に保存し，そこから等確率でランダムにサンプリングした経験データを用いて学習を進めていくが，経験データはそれぞれ出現率が異なるため，出現率の低い経験データを重要な経験データとして扱い，サンプリングされる確率を操作することで学習効率を高めることができると考えられる．そこで本研究では，メモリ内の経験データそれぞれに重み付けを行うことで，DQN の学習速度の向上を図る．

連絡先: 村上知優，名古屋工業大学　大学院工学研究科　情報工学専攻，愛知県名古屋市昭和区御器所町，k.murakami.638@nitech.jp

2. Deep Q Network

2.1 Deep Q Network のアルゴリズム

Deep Q Network(DQN) は入力を状態 s，出力をその状態において選択することの出来る行動の行動価値とするディープニューラルネットワークである．本研究ではこのディープニューラルネットワークを「Q ネットワーク」と呼ぶことにする．経験データを蓄積するメモリを「ReplayMemory」といい，ReplayMemory からランダムに M 個サンプリングした経験データを用いて Q ネットワークの重みを更新する過程を「Experience Replay」という．NIPS 版 DQN[3] のアルゴリズムを図 1 に示す．図 1 にある f はエピソードの終端かどうかを表すフラグである．また，t は教師データを，r は報酬を表し，E は教師データとネットワークの出力の二乗誤差を表す．

Nature 版 DQN[4] ではこれに加えて学習を安定させるために次の 2 つの工夫を行う．本研究では Nature 版 DQN を扱い，本論文では単に DQN と表記されていた場合は Nature 版 DQN を指すものとする．

- ターゲットネットワークの導入
  教師データを生成する際に Q ネットワークの出力を用いている．しかし，Q ネットワークの重みは毎回更新されていくので，ある状態 s を入力した際の教師データが毎回変わってしまう．この問題を解決するために，ターゲットネットワークと呼ばれる Q ネットワークと全く同じ構造の別のネットワークを用意する．ターゲットネットワークの重みは毎回更新されず，Q ネットワークの重みが定期的にコピーされる．教師データ生成の際にターゲットネットワークの出力を用いることで，ある状態 s に対する教師データを固定し，学習を安定させる．したがって，教師データ生成の式を以下のように変更する．

$$t_i = \begin{cases} r_i & (f_i = True) \\ r_i + \gamma \max_{a'_i} Q_{target}(s'_i, a'_i) & (f_i = False). \end{cases}$$

  ここで，Q_{target} はターゲットネットワークの出力値を意味する．

- エラーのクリッピング
  図 1 の誤差の計算で，教師データと Q ネットワークの出力の差の部分（誤差信号）が二乗されている．これにより，誤差信号の絶対値が 1 より大きいか 1 以下であるかで誤差が大きく変化する．誤差が大きく変化するということは勾配が大きく変化するということであるから，学習が不安定になってしまう．これを防ぎ，学習を安定させる方法がエラーのクリッピングと呼ばれる手法である．以下の場合分けに従って新たな誤差信号 e_i' を計算し，誤差の計算は e_i' を用いて行う．

$$e_i = t_i - Q(s_i, a_i)$$

$$e_i' = \begin{cases} 1 & (e_i \geq 1) \\ e_i & (-1 < e_i < 1) \\ -1 & (e_i \leq -1). \end{cases}$$

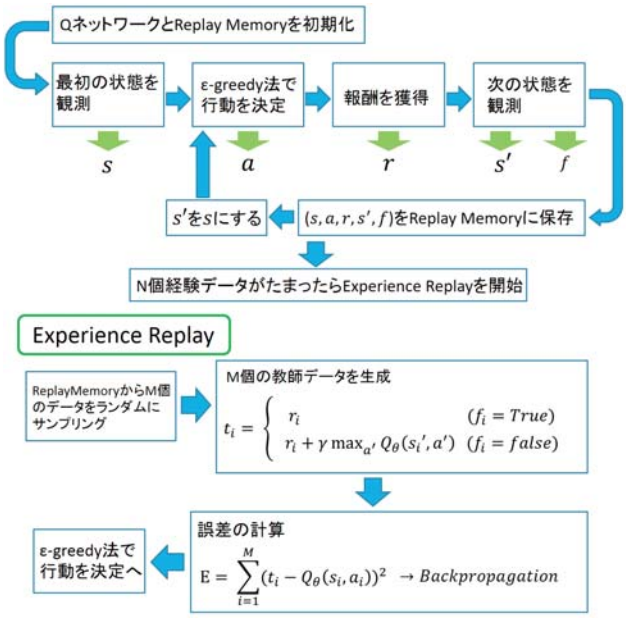


図 1: NIPS 版 DQN のアルゴリズム

2.2 Deep Q Network の問題点

Q ネットワークが最適な行動価値関数へ収束するためには出現する可能性のあるすべての経験データに対して十分な学習が必要であるが，出現率の低い経験データはサンプリングされる確率が低く，学習がなかなか進まないため，最適な行動価値関数への収束が遅くなってしまう．また，一度選ばれただけではその経験データに対する学習が十分に行われないため，出現率に応じて適度に選ばれることが好ましい．よって本研究では経験データそれぞれに重み付けを行い，出現率の低い経験データであるほどサンプリングされやすくすることでこの問題の解決を図る．

3. 提案手法

学習データに対して重み付けを行うことは機械学習の分野ではよく行われており，強化学習では Prioritized Sweeping[6] という手法が有名である．この手法は TD 誤差を指標として経験データに優先度を与え，キューに優先度順に保存し，優先度の高いものから順に経験データを取り出して学習に使用するというものである．本手法では経験データの重みを TD 誤差とは別の指標を用いて与え，かつ重みが大きいものが必ず学習に使用されるようにするのではなく，重みをもとに確率的に選択されたものが使用されるようにする．

DQN の構成に加えて新たに次の状態 $\boldsymbol{s}'$ を予測する A ネットワークを導入する．本手法では環境から与えられる状態がベクトルで表されるとする．図 2 に提案手法のネットワーク構成を示す．

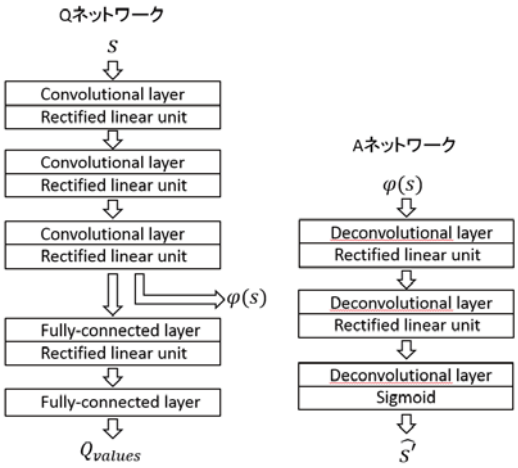


図 2: 提案手法のネットワーク構成

Q ネットワークの中間層の出力 $\boldsymbol{\varphi(s)}$ を A ネットワークへの入力とする．$\boldsymbol{\varphi(s)}$ は Q ネットワークの中間層の出力であるため，Q ネットワークの学習度合いと入力 $\boldsymbol{s}$ に応じて変化する．Q ネットワークに今までほとんど入力されていない状態 $\boldsymbol{s}$ が入力された場合，深層学習の特性 [8] から，$\boldsymbol{\varphi(s)}$ も今までほとんどとることのなかった値になり，A ネットワークの予測が大きく外れることになる．従って，A ネットワークの予測誤差を経験データの重みとすることで，出現率の低い経験データに大きな重みを与えることができる．

ReplayMemory からサンプリングした経験データの内 i 番目の重み $weight_i$ を次の式で計算する．

$$weight_i = \sum_j (s'_{i,j} - \hat{s'_{i,j}})^2.$$

ここで，$s'_{i,j}$ は i 番目の経験データに含まれる次の状態 $\boldsymbol{s}'$ の j 番目の要素である．そして $\hat{s'_{i,j}}$ は i 番目の経験データに対する A ネットワークの出力の j 番目の要素である．また，出現率の低い経験データによる学習が行われていくにつれて，その経験データに対する予測誤差は低下していくので，重みは徐々に小さくなっていくと考えられる．最終的に ReplayMemory 中の k 番目の経験データがサンプリングされる確率 P_k は，各経験データの重み W_k を用いて以下のように計算する．

$$P_k = \frac{W_k}{\sum_n W_n}.$$

A ネットワークの層数は予備実験の結果，3 層とした時が最も良い結果が得られた．また，Q ネットワークから $\boldsymbol{\varphi(s)}$ を取り出す場所は，勾配消失の関係上，出力層に近いほどその変動が

大きく望ましいが，全結合層の出力では入力の位置的な情報が失われてしまうため，A ネットワークでの状態 s' の予測が非常に困難になる．そこで本研究では，Q ネットワークの最も出力側の畳み込み層の出力をその場所とした．

4. 実験

4.1 環境設定

卓球を模したゲームである Atari2600 の Pong[5] を用いて実験を行った．エージェントが得点すると報酬 +1 が与えられ，逆に得点されると報酬 −1 が与えられる．これら以外は全て 0 の報酬が与えられる．エージェントの対戦相手は人間の記述した動作ロジック通りに動き，エージェントが適切な方向に球を打ち返さない限り対戦相手は失点しない作りになっている [5]．先に 21 点先取した方の勝利となり，この時 1 エピソード終了となる．エージェントのとることのできる行動は，上に移動，下に移動，その場で静止の 3 つである．また，どちらかに得点が入ると球は画面中央の初期位置に設置され，ランダムな方向に打ち出される．

この環境を選択した理由は，経験データの出現率に大きな違いがあるからである．特に学習序盤ではエージェントはほぼ無得点で負け続けることが予想されるので，+1 の報酬が得られた経験データの出現率が極めて低い．−1 の報酬が得られた経験データも 0 の報酬の経験データと比較すると出現率が低い．さらに，報酬 0 の経験データであっても，球が初期位置にある状態は得点あるいは失点後に必ず現れるため，出現率が他の状態に比べて高くなる．

本実験では，この環境に対して提案手法を適用することで，それぞれの経験データに対して出現率にもとづいた適切な重み付けができているか，学習速度が改善されるかを検証する．

予備実験の結果得られた最適なハイパーパラメータを表 1 に示す．ミニバッチサイズは ReplayMemory からサンプリングする経験データの個数 M で，ReplayMemory サイズは ReplayMemory に蓄えることのできる最大の経験データの個数である．Experience Replay 開始サイズはその個数の経験データを獲得した時点で Experience Replay を開始するということである．ターゲットネットワークの更新頻度はその回数だけ状態観測を行ったらターゲットネットワークを更新するということである．また，Q ネットワークの最後の全結合層の重みを全て 0 で初期化することで，Q ネットワークの学習が始まる前の出力値が全て 0 になるようにした．それ以外の層の重みは乱数による 0 から 1 の値で初期化した．使用した計算機は Intel Xeon E5-2650 v4 を 2 基，Nvidia GeForce GTX 1080Ti を 2 基持ち，メインメモリは 64GB である．

割引率 (γ)	0.99
ϵ の初期値	1.0
ϵ の最終値	0.1
ミニバッチサイズ (M)	32
ReplayMemory サイズ	100000
Experience Replay 開始サイズ (N)	10000
ターゲットネットワークの更新頻度	10000
Q ネットワークの学習率	0.00025
A ネットワークの学習率	0.00025

表 1: ハイパーパラメータ

4.2 実験結果

4.2.1 経験データの重みの可視化

図 3，図 4 は 100000 個の経験データ，図 5 は 16000000 個の経験データを獲得した時点での ReplayMemory 内の一部の重みの様子を可視化したものである．横軸が ReplayMemory 内の経験データのインデックス，縦軸が経験データの重みである．青色は報酬 0，橙色は報酬 −1，赤色は報酬 +1 の経験データを表す．また，経験データはインデックスが小さい方から時系列順に保存されており，図 5 の横軸は下 5 桁のみ表示されている．

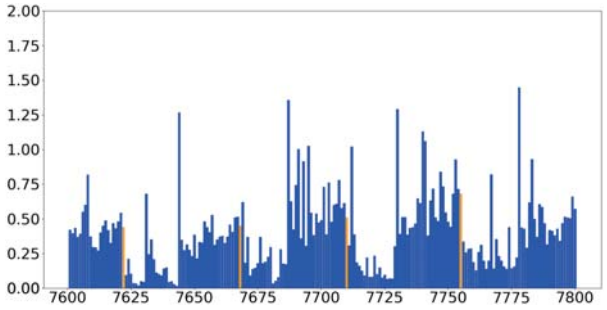


図 3: 開始から 7600〜7800 個の経験データ

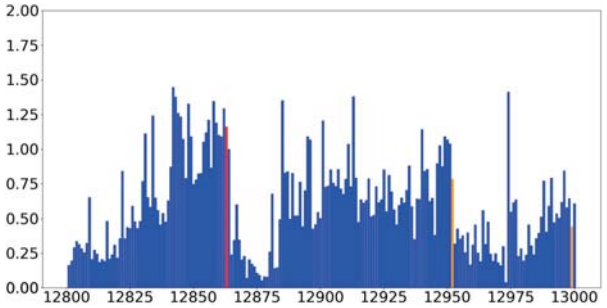


図 4: 開始から 12800〜13000 個の経験データ

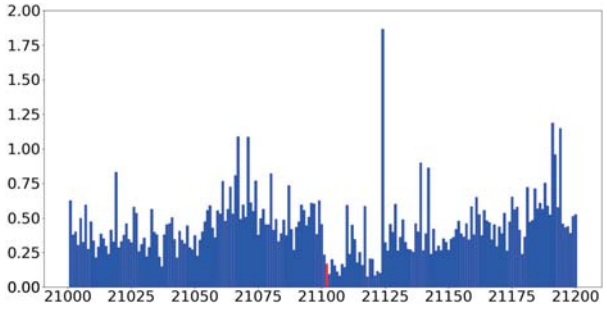


図 5: 開始から 15921000 個〜15921200 個の経験データ

図 3 には赤色の経験データがなく，橙色の経験データが 4 個存在する．これは，エージェントが 1 点も獲得することなく相手に 4 連続得点されているということである．また，一番左側以外の橙色の経験データがほぼ等間隔で並んでいることから，球は初期位置から右側のバー側へ打ち出され，そのまま打ち返すことができず，失点するということが 3 回連続で起こっ

ていると考えられる．さらに，−1 の報酬が得られた経験データに近づくにつれて経験データの重みが大きくなり，−1 の報酬が得られた経験データの後は重みが小さくなっている傾向があることがわかる．これは報酬 0 の経験データの中でも出現率が違うことが影響していると考えられる．また，図 4 より学習序盤において特に貴重な経験データとなる +1 の報酬が得られた経験データと，それが得られる過程の経験データに対して特に大きな重みが与えられていることがわかる．さらに図 5 より，出現率の低い経験データはいつまでも大きな重みではなく，重みが小さくなっていることがわかる．以上の結果より，それぞれの経験データに対して出現率にもとづいた適切な重み付けができていると考えられる．

4.2.2 合計報酬の推移

図 6 の横軸はこれまでにエージェントが学習中に観測したゲーム画面の合計フレーム数，縦軸はエージェントが獲得した報酬の合計値である．青色が DQN，緑色が提案手法のグラフである．

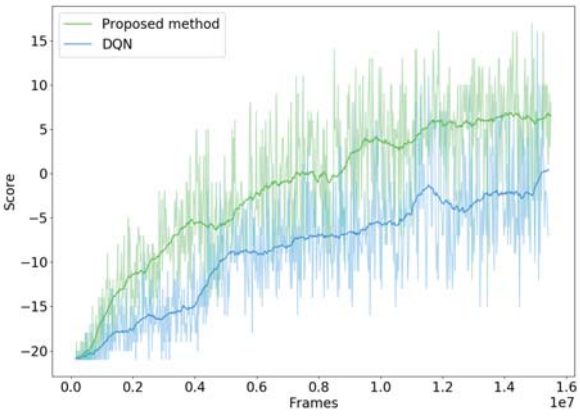


図 6: 合計報酬の推移

図 6 から，学習序盤で大きく差が開いていることが確認できる．DQN に比べて提案手法の方が少ないフレーム数で高いスコアを獲得できており，学習効率が改善されていることがわかる．また，移動平均が初めて 0 以上の値となるまでの実時間を比較すると，提案手法の方が約 32%高速化することができた．

4.3 Q 値の最大値の推移

図 7 の横軸はこれまでにエージェントが学習中に観測したゲーム画面の合計フレーム数，縦軸は Q ネットワークの出力したそれぞれの行動価値の最大値の 1 エピソード間の平均値である．行動価値の初期値が 0 で，学習率が小さいので，行動価値が徐々にしか上昇しない．よって，このデータが学習の進行を測る指標になる．青色が DQN，緑色が提案手法のグラフである．

図 7 から，DQN に比べて提案手法の方が早い段階で高い平均値となっている．よって，提案手法の方が報酬 +1 の経験データに対する行動価値関数の学習および，+1 の報酬を獲得する過程の状態行動対の組み合わせに対する行動価値関数の学習が優先的に行われていると考えられる．

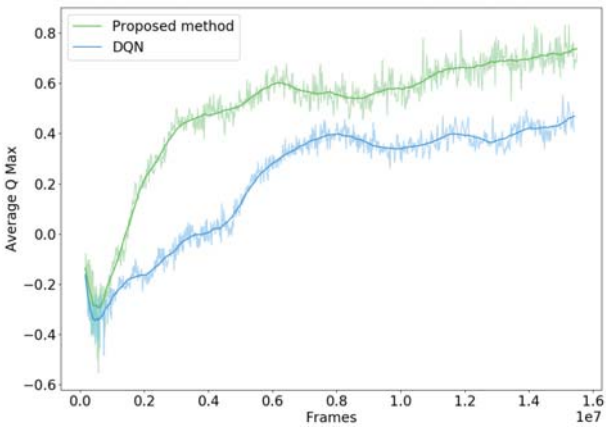


図 7: Q 値の最大値の推移

5. まとめ

本研究では DQN の経験データに対して出現率にもとづく重み付けを行うことで，学習速度を改善する手法を提案した．Atari2600 の Pong という経験データの出現率に大きな違いがある環境で実験した結果，経験データの出現率に基づいた動的な重み付けを行うことで，DQN に比べて少ない学習回数で高い性能を獲得することができ，学習速度を改善することができた．

今後の課題としては，経験データの出現率の違いの度合いが異なる環境での比較検証が挙げられる．

参考文献

[1] Pieter Abbeel et al.: Autonomous Helicopter Aerobatics through Apprenticeship Learning, *International Journal of Robotics Research*, pp. 1–31 (2010).

[2] Kaiming He et al.: Deep Residual Learning for Image Recognition, arXiv:1512.03385v1 [cs.CV] (2015).

[3] Volodymyr Mnih et al.: Playing Atari With Deep Reinforcement Learning, *NIPS Deep Learning Workshop* (2013).

[4] Volodymyr Mnih et al.: Human-level control through deep reinforcement learning, *Nature*, Vol. 518, pp. 529–533 (2015).

[5] Greg Brockman et al.: Openai gym. arXiv:1606.01540 (2016).

[6] Andrew W. Moore and Christopher G. Atkeson: Prioritized sweeping: Reinforcement learning with less data and less time, *Machine Learning*, Vol. 13, pp. 103–130 (1993).

[7] David Silver et al.: Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 484–489 (2016).

[8] David E. Rumelhart et al.: Learning representations by back-propagating errors, *Nature*, Vol. 323, pp. 533–536 (1986).

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

「資産市場の価格変動と銀行間貸借ネットワークの形状が銀行の連鎖倒産に与える影響」

Effect of Asset Price Fluctuation and Inter-bank Lending and Borrowing Network on Chain Bankruptcy of Bank

濱脇　諒[*1]
Ryo Hamawaki

和泉　潔[*1]
Kiyoshi Izumi

坂地　泰紀[*1]
Hiroki Sakaji

米納　弘渡[*1]
Hiroto Yonenoh

[*1] 東京大学大学院 工学系研究科
School of Engineering, The University of Tokyo

When a company goes bankruptcy, the effect can make another company go bankruptcy. This is called chain bankruptcy. In this research, we analyzed effect of asset price and inter-bank lending and borrowing network on chain bankruptcy using agent simulation. We have two results: (1) as the rate of change in asset price is higher, the final number of bankruptcy is higher; (2) as the density of links is higher, the final number of bankruptcy is lower. These results suggest that factor causing bankruptcy is both asset price fluctuation and inter-bank lending and borrowing network.

1. 序論

ある企業の倒産が原因となり，その他の企業まで倒産し，ドミノ倒しのように倒産が続くことを連鎖倒産と呼ぶ．連鎖倒産のように金融機関の破綻が金融システム全体に波及するリスクのことを「システミック・リスク」と言い，近年注目が集まっている．

システミック・リスクに関する研究として今久保ら[今久保 2008]や前野[前野 2013]の研究がある．しかし，今久保らは短期間の資金繰りに焦点を当てていて限定的であり，連鎖倒産は扱われていない．また，前野は市場性資産を介して金融ショックを与えているが，その資産市場の価格変動については触れておらず，ネットワークについても，1 種類のネットワークのみで形状は変化させていない．そこで，本研究では，エージェントシミュレーションを用いて，「銀行の保有する市場性資産の価格変動」と「銀行間の貸借ネットワーク形状」を変化させた時の銀行の連鎖倒産への影響を調査した．

2. 手法

本モデルでは，各試行開始時に銀行を 1 行強制的に倒産させ試行の終了時に銀行が何行倒産したかを数える．これにより，連鎖倒産数が多くなるのはどのような状況であるかを調査するのが本研究の目的である．先行研究として菊地ら[菊地 2015]を参考にバランスシートの作成などを行う．

モデルの構成要素としては: 1.銀行エージェントが 100 行; 2. 銀行間ネットワーク; 3. 市場性資産の市場の 3 つがある．

2.1 モデルの概要

各試行開始時に強制的に倒産させる銀行をスタート銀行と呼ぶ．スタート銀行は大銀行の 1 行とした．まず，初期設定として以下の 3 つを作成する:1.銀行間ネットワーク; 2. 各銀行のバランスシート; 3.市場性資産市場．各ステップで銀行は以下の 3 つを順番に行う: 1.外部資産の売買; 2.市場性資産の価格更新; 3.バランスシートの更新; 4.倒産判定．各試行は 100 ステップで終了する．これらをフローチャートで表したものが図 1 である．

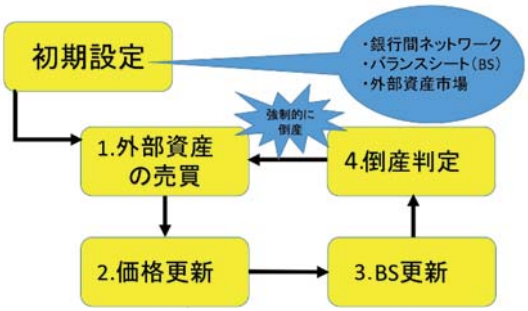


図 1　1 ステップの進行

2.2 銀行間ネットワーク構造

各銀行はほかの銀行と貸借関係を持っている．各銀行をノードとし，貸借関係をリンクとしたネットワークを銀行間ネットワークと呼ぶ．銀行間ネットワークは貸出借入の向きがあるので有向グラフとなる．貸出借入の向きは確率 2 分の 1 で決定する．

銀行間ネットワークの構造は国や時期によって異なるが，多くは次の 3 つのネットワーク構造と同じ特徴を持つ．

(1) コアペリフェラルネットワーク

以下の手順で，ネットワークを作成する: 1. 大銀行 10 個同士を完全グラフで結合する; 2. 小銀行 90 行を 10 個のグループに分ける; 3. 大銀行 1 個と1小銀行グループを一対一に対応させ，大銀行と各小銀行の間にリンクを作成する; 4. 小銀行グループの中でも 4 つの小銀行を完全グラフで結合し，クラスタを形成する; 5. 最後に，各銀行は現在リンクを結んでない銀行とランダムにリンクを結ぶ．

(2) スケールフリーネットワーク

スケールフリーネットワークにおいて，各ノードのリンク数には偏りがあり，リンクを多く持つ少数のハブとリンクが少ない多数の

連絡先:
濱脇　諒，東京大学大学院工学系研究科，東京都文京区本郷 7-3-1，03-5841-1867，b2017hamawaki@socsim.org

ノードに大別される. この性質をスケールフリー性と言い, ウェブサイトや論文引用数など現実のネットワークで観測される[増田 2005]. このスケールフリーネットワークの作成の手順を以下に述べていく.

1. 大銀行 10 行を完全グラフで結合する; 2. このネットワークに 1 つずつ小銀行となるノードを追加していくことでネットワークを形成する. ノードを追加していく過程について説明する.

追加する銀行は一定数のリンクを持っている. これを既存のネットワークにつないでいくが, 銀行 i につながる確率p_iは以下の式で決定する.

$$p_i = \frac{Link_i}{2 * LinkSum} \qquad (1)$$

$Link_i$は銀行 i が持っている総リンク数, LinkSum はネットワーク中の総リンク数を表している.

(3) ランダムネットワーク

本ランダムネットワークのリンクの結び方はコアペリフェラルネットワークやスケールフリーネットワークのように恣意的に決めるのではなく, ランダムに決める. この際, モンテカルロ法を用いる.

2.3 市場性資産

各銀行は株式などの市場性資産を保有し, 時刻 t の資産価格p_tは(2)式で決まる.

$$p_t = p_{t-1} + \alpha \cdot p_{t-1}\frac{(n_{b,t-1} - n_{s,t-1})}{N} \qquad (2)$$

p_tは時刻 t の価格, $n_{b,t}$は時刻 t に購入された市場性資産の数, $n_{s,t}$は時刻 t に売却された市場性資産の数, N は総株数, α は価格変動係数である. ここで, $\alpha = 0$ とした時には $p_t = p_{t-1}$が成り立つので, いかなる t においても価格は固定になる. 市場性資産は 1 種類でシミュレーションを行う.

2.4 倒産の判定と処理

倒産の判定には, (3)式で算出した自己資本比率を用いる.

$$CAR_t = \frac{NetWorth_t}{VaR * StockNum_t} \qquad (3)$$

CAR_tは時刻 t での自己資本比率, $NetWorth_t$は時刻 t での自己資本, $StockNum_t$は時刻 t での持ち株数を表している. VaR はバリューアットリスクのことで, 該当株式がもつリスクを定量的に表したものである. つまり, 分母が意味しているのは該当株式のリスクと株数の積であり, 保有する市場性資産の全リスクということになる. よって, 自己資本比率増加が起こる事象としては以下の 2 つが挙げられる: 1.自己資本の増加; 2.持ち株数の減少. この自己資本比率の値が 0.04 を下回った時, 各銀行は倒産となる. 倒産の処理については次節で述べる.

(1) 倒産の処理

銀行が倒産した時, 倒産した銀行に貸出を行っていた銀行はその額を回収することができなくなり, 一方倒産した銀行から借入を行っていた銀行は返済する義務がなくなる. 市場性資産は市場に売り出さず, 倒産しても株式は保有し続けるが, 株式の売買には参加できないものとする.

3. 結果

本研究では, 各パラメータセットについて 100 試行シミュレーションを行い, その平均をとった.

パラメータについて: 1. 市場性資産による影響; 2. ネットワークの形状による影響の 2 つに分けて, 次節以降記していく.

3.1 市場性資産による影響

市場性資産の価格を決定する式(2)の価格変動係数 α をパラメータとする. α を: 1. 0; 2. 0.25; 3. 0.5; 4. 1.0; 5. 1.5 の 5 通りで変化させ, 実験を行った. 結果を図 2 に示す.

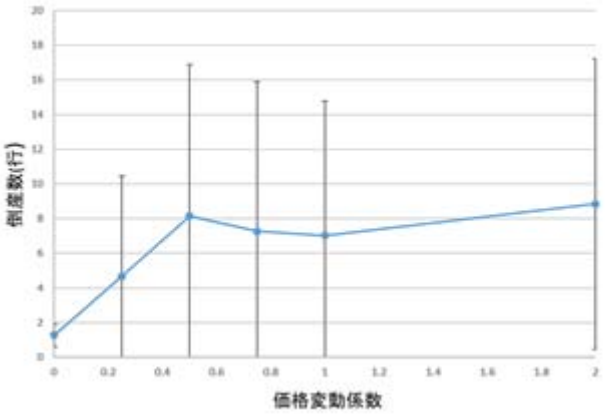


図 2 最終的な倒産数と価格変動係数の関係

価格変動係数の増加に伴い, 最終的な倒産数も増加した.

3.2 ネットワークの形状による影響

各ネットワーク構造についてパラメータと結果を示す.

(1) コアペリフェラルネットワーク

コアペリフェラルネットワークでは恣意的な構造を作成した後, 最終的にランダムにつなぐリンク数をパラメータとした. 恣意的な構造のリンク数も合わせたリンク数を大銀行で以下の 3 通り: 1. 20 本; 2. 30 本; 3. 50 本, 小銀行で以下の 4 通り: 1. 5 本; 2. 8 本; 3. 10 本; 4. 20 本となるようにパラメータを設定し, それぞれの組み合わせ 3×4＝12 通りについて実験を行った. 結果を図 3 に示す.

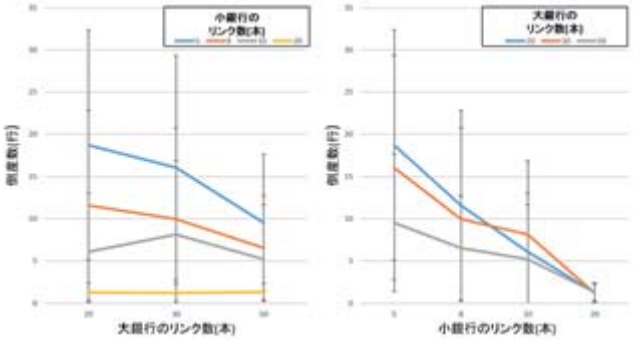


図 3 コアペリフェラルネットワークにおけるリンク数と最終的な倒産数の関係

一部を除き, リンク数が増加するにつれて, 平均倒産数は減少した.

(2) スケールフリーネットワーク

スケールフリーネットワークでは, 追加する銀行が持つリンク数をパラメータとして設定する. リンク数は以下の 5 種類である: 1. 2 本; 2. 3 本; 3. 5 本; 4. 8 本; 5. 10 本. 結果を図 4 に示す.

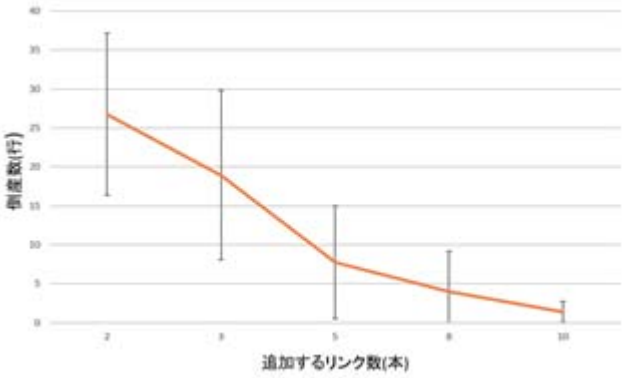


図 4 スケールフリーネットワークにおけるリンク数と
最終的な倒産数の関係

コアペリフェラルネットワークと同様に, リンク数が増えるにつれて, 倒産数は減少した.

(3) ランダムネットワーク

ランダムネットワークでは大銀行と小銀行のリンク数の組み合わせを以下の 4 通りで実験した: 1. 6 本と 3 本; 2. 10 本と 5 本; 3. 20 本と 10 本; 4. 50 本と 20 本. 結果を図 5 に示す.

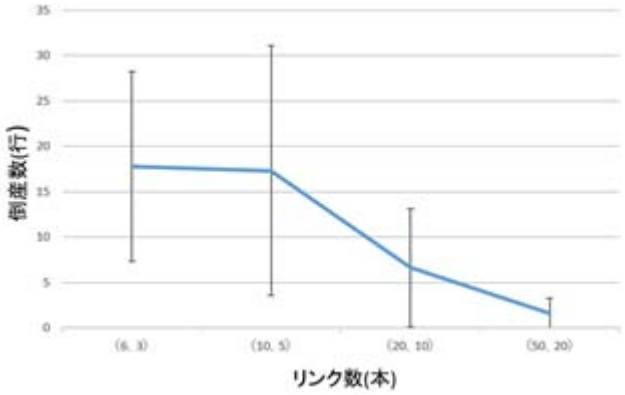


図 5 ランダムネットワークにおけるリンク数と
最終的な倒産数の関係

こちらもリンク数が増えるに連れて, 最終的な倒産数は減少した.

4. 考察

結果と同様に: 1. 市場性資産による影響; 2. ネットワークの形状による影響の 2 つに分けて, 次節以降に記し, 最後にまとめと将来の展望を述べる.

4.1 市場性資産による影響

図2において, $\alpha=0$ の場合と $\alpha=1$ や 1.5 の場合を比較すると, 価格変動が銀行のリスクを増大し倒産数が増大した. しかし, $\alpha=1$ と 1.5 のときを比べると, 倒産数に大きな違いは見られない. α が一定以上になると外部資産を一定以上持っている銀行はほぼ潰れ, それ以上 α を増やしても倒産数はそれ以上増加しないためだと考えられる.

4.2 ネットワークの形状による影響

(1) コアペリフェラルネットワークの結果

図 3 において, 小銀行のリンク数が 10 本や 20 本と多くなった時は大銀行のリンク数を増やしてもほとんど倒産数に変化がなかった. これはある程度小銀行と小銀行のリンク数が増えた時, 大銀行と小銀行のリンク数の本数は倒産数に影響を与えないことを示唆している.

(2) 倒産の処理

この節では, 3 つのネットワーク構造に共通して見られたリンク数が増えるほど最終的な倒産数が増加する, という結果についてその理由を考察する.

本モデルでは, 各銀行が保有する資産の総額は乱数を用いて一定の範囲のある値とした. よって, リンク数の大小によらず, 各銀行の貸出額はほぼ同じくらいの額になる. ここで, 具体的に銀行 X の持つリンク数が 3 本と少ない場合(Ⅰ), リンク数が 20 本と多い場合(Ⅱ)を考える. 今, 銀行 X の貸出額は図 6 の緑のように分割されていて, 各ブロックが 1 つの銀行に対して貸出している額に相当する.

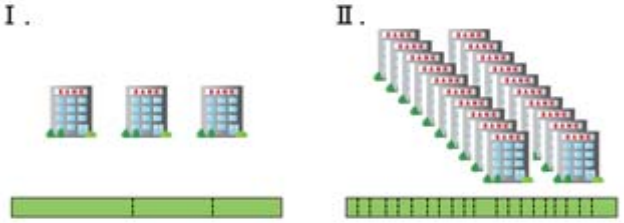


図 6 銀行 X の貸出額のイメージ

次に, この銀行 X とリンクでつながった銀行のうち 1 行が倒産した場合を考えるとき, ⅠとⅡでは回収不能になる額が異なり, Ⅰの方がより多くなるということが分かる. よって, 他の銀行とのリンク数が少ないと, その数少ない取引先が倒産した時に大きなダメージを受けてしまうということが分かる. この結果から, 銀行は同程度の額の貸出を行う際, なるべく多くの銀行に少額の貸出を行うべきだということが示唆される.

4.3 まとめと将来の展望

以上より, 得られた結論は: 1. 市場性資産の価格の変動率が高いほど最終的な倒産数も増加する; 2. リンク密度が高いほど最終的な倒産数は減少する, の 2 つであった.

将来の展望としては: 1. 銀行間ネットワークの大きさを今回 100 行としたが, 規模を変えるとどうなるか; 2. 倒産数の平均で議論したが, 倒産した銀行の総資産額から被害総額を求める; 3. 銀行間ネットワークと企業ネットワークを結合する, の 3 つを考えている.

参考文献(論文誌と同じスタイルを推奨)

[今久保 2008] 今久保 圭, 副島 豊: コール市場の資金取引ネットワーク, 日本銀行ワーキングペーパーシリーズ, 2008.

[前野 2013] 前野義晴: 金融システムを安定化する最適な投融資ポートフォリオと銀行間ネットワーク, SIG-FIN Vol.10,No.8, 2013.

[菊地 2015] 菊地 剛正, 高橋 大志, 寺野 隆雄:エージェントシミュレーションを用いた金融規制と金融機関の投資行動が破綻伝播に与える影響の分析, 経営情報学会 2015 年秋季全国研究発表大会要旨集,pp. 251-254 , 2015.

[増田 2005] 増田 直紀, 今野紀雄: 複雑ネットワークの科学, 産業図書, 2005.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

格闘ゲームにおけるモンテカルロ木探索の探索時間と即応性のトレードオフ

Trade-off between search time and responsiveness of MCTS in a fighting video game

亀垣 航 Wataru Kamegaki　森山 甲一 Koichi Moriyama　武藤 敦子 Atsuko Mutoh　犬塚 信博 Nobuhiro Inuzuka

名古屋工業大学
Nagoya Institute of Technology

In real-time strategy games, it is difficult for a computer to defeat a human player under the condition that the computer has same recognition ability to the player. It is because many factors in the environment are changing in a very short time during which the computer has to choose an action. Although the Monte Carlo Tree Search (MCTS) algorithm obtains decent results in such games by searching a good action from simulation, it will be better if it has more time. On the other hand, it may be worse due to delay of response to the environment. In this work, we investigated the trade-off property in a fighting video game.

1. はじめに

近年、人工知能 (AI) にコンピュータゲームをプレイさせる試みが盛んに行われている。その中でも囲碁や将棋などいったプレイヤーの行動順が明確に決まっているターン制戦略ゲームの分野では研究が大きく発展している。例を挙げると、Google DeepMind の「AlphaGo」[1] という囲碁の対局プログラムが人間のプロプレイヤーに勝利するといった報告がされている。

一方、格闘ゲームやシューティングゲームといった行動順に規定の無いリアルタイム戦略ゲームの分野では、人間の能力に合わせてコンピュータの認識能力に制限をかけた条件下では、人間のプレイヤーに勝利することは困難である [2]。人間とコンピュータが平等な条件下においてコンピュータが人間に勝利することが困難なのは、常に多くのパラメータがごくわずかな時間で変動し続けるため、限られた時間内で膨大な量の情報を処理して行動を選択することが求められるからである。

本研究では、リアルタイム戦略ゲームの中でも格闘ゲームを対象とし、人間を超えない程度の認識能力になるよう制限をかけた条件下において、コンピュータが対戦を行う際の行動選択能力の向上を目的とする。

格闘ゲーム AI の競技会では、モンテカルロ木探索を用いた AI が優秀な成績を収めている。しかし、モンテカルロ木探索でも計算時間が不足し、行動選択のための探索が十分に行われていない [8]。そのため、本研究では応答時間を延長させることによって、モンテカルロ木探索の探索時間増加による行動選択能力の向上と、ゲーム環境の変化への対応能力である即応性の兼ね合いを観察する。題材として FightingICE [3] と呼ばれる研究用の格闘ゲームを使用した。

本論文は全 6 章で構成されており、第 2 章では格闘ゲームと今回使用する FightingICE、第 3 章ではモンテカルロ木探索と既存研究について紹介する。第 4 章では提案手法、第 5 章で実験の内容と結果について記述し、最後の第 6 章でまとめと今後の課題について述べる。

連絡先: 亀垣 航，名古屋工業大学工学部情報工学科
名古屋市昭和区御器所町，
w.kamegaki.653@nitech.jp

2. 格闘ゲーム

2.1 格闘ゲーム

格闘ゲームとは、プレイヤーがキャラクターを操作して殴る蹴るといった格闘技や武器などを駆使して、対戦相手を倒すことを目的としたリアルタイム戦略ゲームの一種である。基本的には人間同士又は人間対コンピュータで対戦を行う。

格闘ゲームでは先に対戦相手の操作するキャラクターの体力を 0 にすることで対戦相手を倒したプレイヤーの勝利となり、制限時間内に終わらなかった場合は制限時間が 0 になった時点でのキャラクターが受けたダメージ量で判定を行い、受けたダメージの少ないプレイヤーの勝利となる。

コンピュータが格闘ゲームにおいて対戦を行う場合、多くの場合、予め作成された行動規則に従って動くので行動パターンが単調になりやすく、熟練した人間相手では何戦か対戦することで行動を簡単に予測されてしまうため、勝利するのは極めて困難である。

2.2 FightingICE

2.2.1 概要

FightingICE [3] とは、研究目的で開発された 2 次元格闘ゲームである。ゲーム内のキャラクターを操作するプログラムを Java 言語で作成可能である。なお本稿では以後このプログラムのことをエージェントと称する。

1 試合は 3 ラウンド、1 ラウンドの制限時間は 60 秒で構成されている。1 秒を 60 分割したものをフレームと呼び、フレームごとにその時の各キャラクターの情報が記録され、エージェントはその情報を取得することができる。

キャラクターの情報としては以下の表 1 の「HP」,「Energy」,「X,Y」,「Speed」,「State」,「Motion」がある。

そして、対戦結果の評価指標として対戦スコア Score がある。対戦スコアはラウンドごとに相手から受けたダメージ量である HP によって (1) 式によって算出される。

$$Score = \frac{oppHP}{oppHP + myHP} \times 1000 \quad (1)$$

Score はラウンド内でどちらが多くのダメージを受けたか判断するものであり、*Score* が 500 の場合引き分けとなる。

表 1: FightingICE におけるキャラクターデータ

情報	データ型	詳細
HP	int	対戦相手からどれだけダメージを与えられたかを表す。
Energy	int	キャラクターが所持するエネルギーの合計量を表す。エネルギーはキャラクターの一部の技を使用する際に必要なポイントであり、技ごとに必要なエネルギーは異なり、使用するとその分だけエネルギーは減少する。
X,Y	int	キャラクターの位置座標を表す。ゲーム画面の左上を $(X, Y) = (0, 0)$ として水平方向を X 方向、垂直方向を Y 方向とする。
Speed	int	キャラクターの移動量を表す。水平方向を X 方向、垂直方向を Y 方向とする。この値を現在の座標に加算することで次のフレームでの座標が算出される。
State	String	キャラクターの現在の状態を表す。状態は 4 種類あり、STAND(直立)，CROUCH(しゃが み)，AIR(空 中)，DOWN(転倒) となっている。
Motion	String	キャラクターが行っている行動を表す。 行動は大まかに Base(基本状態)，Move(移動行動)，Guard(防御行動)，Recovery(復帰行動)，Action(攻撃行動) の 5 つに分類でき、全 55 種類の行動がある。

2.2.2 認識能力の制限

FightingICE では、エージェントがキャラクターの位置や行動などのゲーム情報を取得する際は、15 フレーム (約 0.25 秒) 以前の情報しか取得することができない。これは人間が視覚から情報を取得して反応するまでにかかる時間が約 0.2 秒 [5] であるので、コンピュータの認識能力を人間の認識能力の限界にあわせるためである。このルールが無い場合、エージェントは 1 フレーム (約 0.017 秒) 前の状況を見て判断することができてしまうため、人間側が不利になってしまう。

2.2.3 行動入力

FightingICE では、ゲーム環境が 1 フレームごとに更新されていくため、エージェントは行動を 1 フレーム以内に入力しなければならない。行動が入力されなかった場合、エージェントが最後に入力した行動が再度入力されたものとして扱う。

3. モンテカルロ木探索

3.1 概要

モンテカルロ木探索 [6] とは、評価関数の代わりに確率的シミュレーションを使用する探索手法の 1 つであり、評価関数の作成が困難な問題に対して有効な探索アルゴリズムである。

探索は図 1 のような木構造を作成していくことで行っていく。ゲームの状況を根ノードに渡し、実行可能な行動を 1 つずつ子ノードに分けて木を作成する。そして有望なノードに子ノードを展開して木を成長させることで深く探索していく。ゲーム終了までランダムに行動を行うシミュレーションの勝率または評価値を求めた結果をプレイアウトと称する。

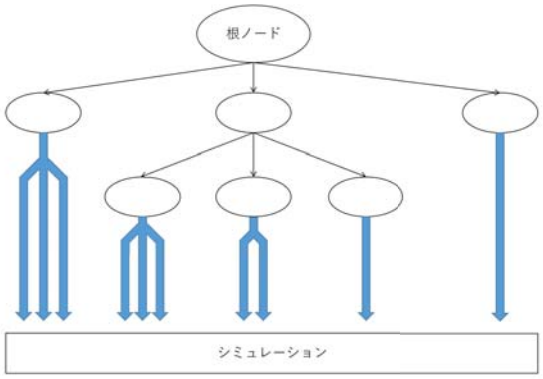


図 1: モンテカルロ木探索の例

有望なノードの指標として $UCB1$ があり、これは以下の (2) 式によってノードごとに算出される。$\overline{X}$ はそのノードのプレイアウトの平均、C は定数、N^p はそのノードの親ノードが選択された回数、N はそのノードが選択された回数である。

$$UCB1 = \overline{X} + C\sqrt{\frac{2\log N^p}{N}} \tag{2}$$

3.2 手順

モンテカルロ木探索では、以下の 1〜4 の手順を一定回数又は一定時間の間繰り返した後に根ノードの子ノードの中で最もプレイアウトの数が多いノードが表す行動を返す。

1. 選択 $UCB1$ が最大の子ノードを選択する。これを展開済みの末端ノードになるまで繰り返す。

2. 拡張 選択された回数が閾値を超えたノードで子ノードを展開する。展開した子ノードの中から $UCB1$ が最大のノードを選択する。

3. シミュレーション 1、2 で選択したノードの行動、ランダムに選択した行動の順でシミュレーションを行って、プレイアウトを求める。

4. 逆伝播 3 の結果から 1、2 で選択したノードの $UCB1$ を更新する。

3.3 FightingICE での実装

吉田ら [7] は、エージェントが操作するキャラクターの行動選択をモンテカルロ木探索によって行った。格闘ゲームにおいては、計算量が膨大になるためプレイアウトをラウンド終了まで求めるのは困難であるので、$\overline{X}$ を (3) 式によって算出する。

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N}(difHP_i^{my} - difHP_i^{opp}) \tag{3}$$

N はノードの総探索回数、$difHP_i^{my}$ で i 番目のシミュレーション前後での自身の HP の変化量、$difHP_i^{opp}$ で対戦相手の HP の変化量を表す。

モンテカルロ木探索において設定するパラメータは表 1 があり、吉田らは $C = 3$, $N_{max} = 10$, $D_{max} = 2$, $T_{sim} = 60$, $T_{max} = 16.5$, $A_{num} = 5$ と設定した。

表 2: モンテカルロ木探索におけるパラメータ

パラメータ	詳細
C	$UCB1$ を算出する際に使用する定数。数値が高いほど選択回数が少ないノードが選ばれやすくなる。
N_{max}	子ノードを展開する際の閾値。選択された回数がこの閾値を超えると子ノードを展開する。
D_{max}	木の深さの最大値。この値まで木を展開することができる。
T_{sim}	シミュレーション時間。シミュレーションをこの値フレーム先まで行う。単位はフレーム
T_{max}	探索時間。この時間の間モンテカルロ木探索で探索を続ける。単位は秒
A_{num}	行動数。シミュレーションする際にエージェントが行う行動の数

FightingICE ではゲームの状況が 1 フレーム (1/60 秒) 毎に変化し続けるため、1 フレーム以内に行動を選択しなければならない。吉田らのモンテカルロ木探索による行動選択では、その時間の制約から図 2 のように、ゲーム環境から自身と対戦相手のキャラクターデータなどのフレームデータを 1 フレーム毎に取得し、対戦相手の行動予測やエージェントが行動を選択するための計算を行っている。計算時間が 1 フレームしかないためプレイアウトの回数が不足し、行動予測や行動選択を行うための探索が足りていない [8] と思われる。

4. 提案手法

本研究では、プレイアウトの回数を増加させるために、応答時間を延長することでモンテカルロ木探索の探索時間を増加させる手法について述べる。加えて、延長したフレーム数が多いほど、フレームデータを取得した時と行動を入力した時とのゲーム環境の差が大きくなり、推定した状況と実際の状況との差が大きくなってしまうことが予測されるため、応答時間を延長するフレーム数と即応性のトレードオフを観察する。

提案手法では、図 3 のように応答時間を 1 フレームから N フレーム (N は自然数) に延長する。モンテカルロ木探索の探索時間に使用できる時間を増加させることによって、行動選択能力の向上を図った。ただし、応答時間を延長するほどフレームデータを取得した時と行動を入力した時とのゲーム環境の差が大きくなってしまうため、選択した行動が実行する状況に適していない場合が考えられる。そのため、応答時間を延長するフレーム数とゲーム環境の即応性のトレードオフを観察する。

なお、FightingICE の仕様上、エージェントは 1 フレーム毎に行動を入力しなければならないので、図 2 の斜めの矢印のように応答時間を延長して計算時間が 1 フレームを超える毎に、一番最後に入力した行動と同じ行動をゲーム環境に入力される。

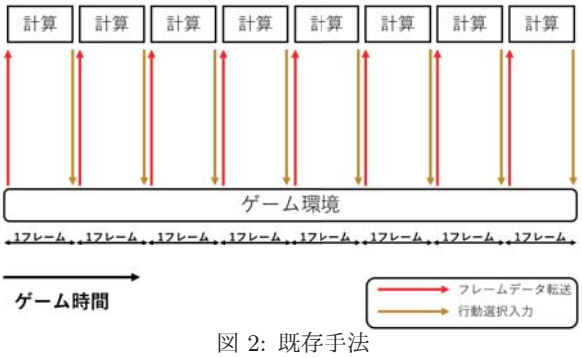


図 2: 既存手法

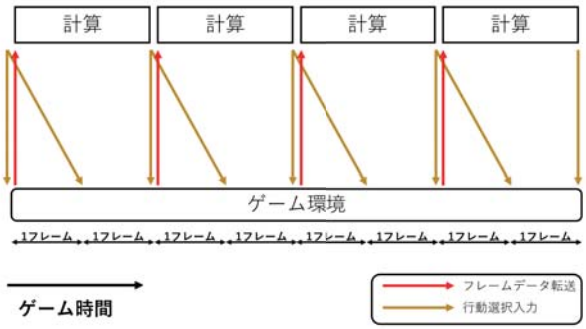


図 3: 提案手法 ($N = 2$ の場合)

5. 実験・考察

応答時間を延長したモンテカルロ木探索を用いて行動選択を行うエージェントの性能評価を行い、実験結果から考察を行った。今回、提案手法はモンテカルロ木探索によって行動選択を行う MctsAI [3] の応答時間を延長することで実装した。実験では吉田らや MctsAi と同様のパラメータである、$C = 3$, $N_{max} = 10$, $D_{max} = 2$, $T_{sim} = 60$, $T_{max} = 16.5$, $A_{num} = 5$ とした。

5.1 対戦スコア比較

対戦スコアを計測する際の対戦相手として MctsAI [3] を用いた。MctsAi パラメータは提案手法と同様にしたため、応答時間が 1 フレームの場合は MctsAi と同様となる。各応答時間ごとに 10 試合を行い、その対戦スコアを算出した。

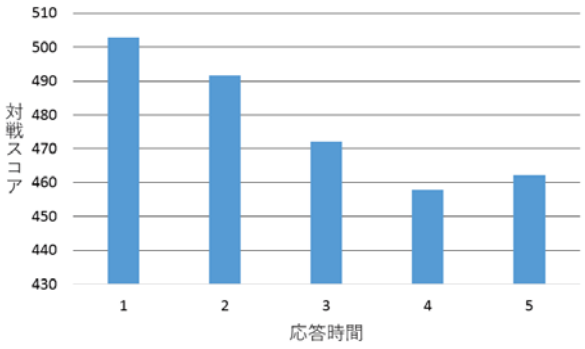


図 4: 対戦スコア

図 3 の対戦結果から応答時間が大きくなるほど対戦スコアが低下していく傾向が読み取れる。応答時間を延長することでモンテカルロ木探索の探索時間に使用できる時間を増え、プレ

イアウトの回数が増加し行動選択能力が上昇すると思われた。しかし、応答時間を延長することによってゲームの情報を取得してから実際に行動するまでの時間が大きくなり、推定した状況と実際の状況との差が大きくなってしまうことで状況に適した行動ではなくなってしまうことの影響の方が大きかったと推測される。

5.2 探索スコア比較

応答時間を延長することが行動選択にどのような影響を与えているのかを検証するために、モンテカルロ木探索によって行動を決定した時の探索スコア（(3) 式の $\overline{X}$) を比較する。キャラクターの位置や状況などが同様な状況下である方が好ましいため、対戦相手として単一ルールによって行動を選択する Jay Bot[3] を用いた。各応答時間毎に 5 試合行い、応答時間内で行った探索回数の平均と選択した行動の探索スコアの平均を求めた。

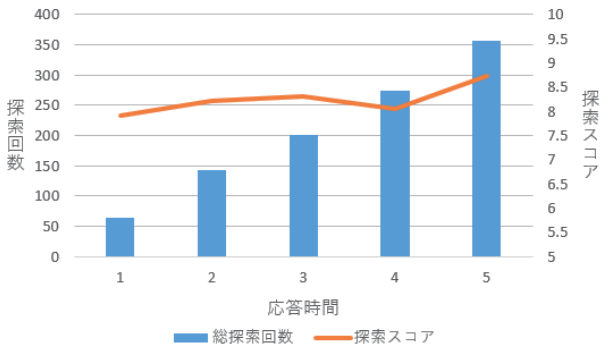


図 5: 探索スコア比較

図 4 の実験結果から、応答時間が増加するにつれて探索回数は増加したが探索スコアはそれほど上昇しない傾向が見られた。原因としては、探索回数を増加しても同じノードを探索して他のノードをあまり探索していない可能性が考えられる。そのため、この実験での総探索回数のうち最も探索回数が多いノードが占める割合を求めた。

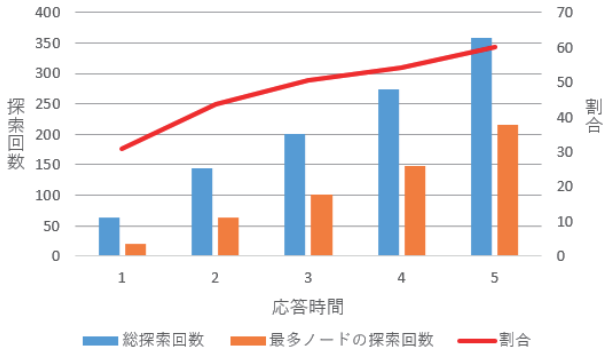


図 6: 探索回数割合比較

図 5 より応答時間を延長するのに伴って、総探索回数のうちで最も探索回数が多いノードが占める割合が上昇しているのが見て取れる。このことから探索回数の増加分が最多ノードの探索に費やされており、他のノードをあまり探索していない可能性があることが推定される。

5.3 考察

実験より、探索回数を増やしただけでは探索スコアの上昇よりも、応答時間の延長によるデメリットの方が大きく、対戦に悪影響を及ぼしてしまうことが判明した。探索スコアが上昇しにくかった原因として、今回利用したモンテカルロ木探索のパラメータが応答時間を延長した場合に適していなかったことが推測される。よって延長するフレーム数ごとに適したパラメータを設定する必要があると推察する。

6. おわりに

6.1 まとめ

本研究では、格闘ゲーム AI の性能向上を目的として応答時間を延長することで計算時間を確保し、モンテカルロ木探索の探索回数を増加させた場合における試合への影響を確認した。結果は、延長するフレーム数を増やすほど対戦スコアは減少してしまった。延長するフレーム数と即応性のトレードオフが見られなかった原因としては、モンテカルロ木探索で設定するパラメータが応答時間を延長した場合に適していなかったことが挙げられる。

6.2 今後の課題

今後の課題は、応答時間に対応した適切なパラメータの設定により探索効率を向上することでトレードオフを観察し、最適な計算時間を求めることが挙げられる。

参考文献

[1] David Silver et al. Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 484–489, 2016.

[2] 遠藤雅伸, 「ディジタルゲームにおける”AI” の役割」, 情報処理学会, Vol. 53, No.2, pp.146-152 2012.

[3] http://www.ice.ci.ritsumei.ac.jp/˜ftgaic/(2018 年 1 月 23 日参照)

[4] Man-Je Kim, Kyung-Joong Kim, ”Opponent Modeling based on Action Table for MCTS-based Fighting Game AI”, *IEEE Conference on Computational Intelligence and Games (CIG)*, 2017.

[5] 井上由紀子, 山根一人, 大河俊博, 中原啓晶, 高橋由起子, 落保子「単純反応時間の学習効果：特に健常人に対する検索」, 理学療法学, 17, p. 51, 1990.

[6] 美添一樹 「モンテカルロ木探索」, 「人工知能学大辞典」, pp. 1193–1195, 共立出版, 2017.

[7] Shubu Yoshida, Makoto Ishihara, Taichi Miyazaki, Yuto Nakagawa, Tomohiro Harada and Ruck Thawonmas, “Application of Monte-Carlo Tree Search in a fighting game AI,” *IEEE Global Conference on Consumer Electronics*, 2016.

[8] Man-Je Kim, Kyung-Joong Kim, ”Opponent Modeling based on Action Table for MCTS-based Fighting Game AI”, *IEEE Conference on Computational Intelligence and Games (CIG)*, 2017.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

3D L-Systems Path-finding
Case study: Walking Agent's Explorations in Simulations

Djuned Fernando Djusdek[*1] Yoshiteru Ishida[*1]

[*1] Toyohashi University of Technology

This paper presents a technique for exploring map or surface using extended L-Systems based on L-Systems' grammar, other than L-Systems rules, we need this derivative rule for exploring a walkable area. The rule that used is flipping approach, which is a technique for changing the angle of generating grammar, then will impact the direction. The results are validated and done in simulation systems.

1. Introduction

Lindenmayer Systems or L-Systems [Lindenmayer 1968] is a great invention, not only for representing a simple living object, yet can for representing plants. Inspired by biological systems, L-Systems was developed more, i.e., Stochastic L-System and Context-sensitive L-Systems, which is used for represented plant more natural. Not only that, by combining L-System with Genetic Algorithm (GA), the plant representation is unique [Prusinkiewicz 1990] [Ochoa 1998] [Kurth 2007].

Even more than that, recent L-Systems is also used for the path-finding solver. Some research was published, e.g. [Areyan 2012]. In this research, they use Ants Colony as a base algorithm, combined with Genetic Algorithm. In case of uses share memory "*stigmergic*" that ants left as pheromones, also uses individual memory that contained in its genetic information. This genetic information stored as genes and represented by simple L-Systems' grammar.

However, not available yet for exploring the map, which has no route or path, e.g., Mars's surface. In this research, we implement a 3D exploration's simulations for walking agents path-finding solver in an unexplored land. This research objection is used L-Systems for path-finding. Hopefully, this research can support human for exploring the unexplored land.

The rest of this paper is structured as follows. Section 2 describes the previous research which relates to L-Systems path-finding. Section 3 presents the method we propose which the simulation and experimental result is provided in Section 4 and 5. Finally, the conclusion is drawn in Section 6.

2. Related Work

This research uses Cellular Automaton (CA) and combines it with L-Systems. Related to [Alfonseca 2003], in this research was explained the formal definitions of deterministic and probabilistic CA, with special attention to the problem of updating the probabilistic information each automaton in the grid. In this article, also introduced a formal notation for probabilistic L-Systems and the language generated by them.

Contact: Djuned Fernando Djusdek, Toyohashi University of Technology, 1-1 Hibarigaoka Tempaku Toyohashi Aichi 441-8580 JP, +81 80 9986 4087, djuned@sys.cs.tut.ac.jp

For support this research, there is another research, e.g. [Koopman 2016], that try to solve path-finding in 3D space for the indoor environment. Many path-finding algorithm and approach discussed in this research. Concluding to use voxel location, and use "*voxelized*" model for representing a 3D indoor environment. The agent also divided into several types, i.e., walking and flying agents.

Rather than focus on an existing problem like exploration by robots, e.g. [Shen 2012], we focus on algorithms to explore unexplored land in simulations. In this related work, the autonomous flying vehicle was used as agents for indoor exploration.

Moreover, about L-Systems exploration, this problem can be classified as space-filling curve problems. Described in [Nair 2017], they also done exploration task by implementing a systemic strategy for Hilbert's space-filling curve. The exploration task is to find obstacles (or holes) and implemented online, and their future goal is to use it for path planning and searching.

3. Method

As described in related work, we use L-Systems with CA to solve this problem and related to our previous work [Ishida 2017]. The environment that used is simulation systems that have been developing in OpenGL using Qt wrapper. The simulation environments are used to simulate a small world, i.e., land or island(s).

```
1.  algorithm 1: loadHeightMap( grayScale image )
2.  w ← width(image)
3.  h ← height(image)
4.  c ← heightDivider ← 10.0f
5.  for i ← 0…w
6.    for j ← 0…h
7.      g ← greyColor( imageAtPoint( i, j ) ) / c
8.      storeToMatrix( g )
9.  return
```

Figure 1. The Pseudocode of Height-map Loader.

In this case, we use a height-map as input for generating the land surface. This surface will see as a plain land without anything, except water area, i.e., a sea. The pseudocode for load the height-map is described in Fig. 1. For enriching the surface, we use Game of Life [Gardner 1970] for distributing the tree(s), and by randomly generated seed, and combining with the height of areas that can be placed, i.e., walkable area. The constraint to determining the walkable area is defined in Eq. 1, while the *h* is the height of the walkable area.

$$waterArea < h < mountainArea \quad (1)$$

L-Systems' grammar that used shown in Fig. 2. This grammar produces four directions in the beginning and will create a branch on each iteration. In the generating steps on each direction, will be generate a branch like described in Fig. 3, that *A* is equal to 1/4 from *B* and *B* is equal to 3/4 from start point in each iteration, and *A* plus *B* is equal to *F*. Either *A*, or *B* will produce new *F* and do similar way in each iteration. This grammar will ensure that four directions will be explored.

However, this grammar cannot guarantee that all areas will be explored if there are obstacles. To solve this problem, we introduce a flip approach to change the angle for branching implementation. The flipping technique causes the branch to grow left and then right in continues iteration, with some rules. This approach can guarantee almost all walkable area will be explored.

In this research, L-Systems is mainly used to predict the next point and the edge that is connecting from the current point, i.e., vertex, to the next point. This step is done by using CA. At this step, the CA try to connect a vertex by uses heuristic function to determine the path, which is possible to passing or not. If not appropriate, we need to move to the next point as near as possible from the previous position of next point. The pseudocode for this algorithm is described in Fig. 4, and the steps for creating edge is described in Fig. 5.

Different from [Nair 2017], in this previous research, the algorithm is based on Hilbert's space-filling curve, where just have one start and one endpoint. They modified Hilbert's space-filling curve so can do in space, which has obstacles (or holes). Comparing with this research, our approach cannot reach all area in space, i.e., walkable area. However, our approach can be run concurrently and independently, but still depend on the neighbor in each cell, i.e., point, when generating the path.

4. Simulation

The simulation systems built in MacBook Pro 2017 and Acer M5 481TG, this will ensure that our simulator can run in multi-platform. The specifications for this simulation systems are described:

4.1 Hardware Specifications

In this research, the simulation built to focus on:

- MacBook Pro (13-inch, 2017), with specifications:
  - Processor: 3.1 GHz Intel Core i5-7267U
  - Memory: 8 GB 2133 MHz LPDDR3
  - Graphics: Intel Iris Plus 650 1536 MB
- Acer M5 481TG, Windows 10, with specifications:
  - Processor: 1.7 GHz Intel Core i5-3317U
  - Memory: 6 GB DDR3 SDRAM
  - Graphics: Nvidia GeForce GT 640M LE 1 GB

$$R \rightarrow [F][+F][++F][+++F] \qquad \delta = 90°$$
$$F \rightarrow (3/4)[+F](1/4)[F] \qquad \theta = 45°$$

Figure 2. The L-Systems' Grammar.

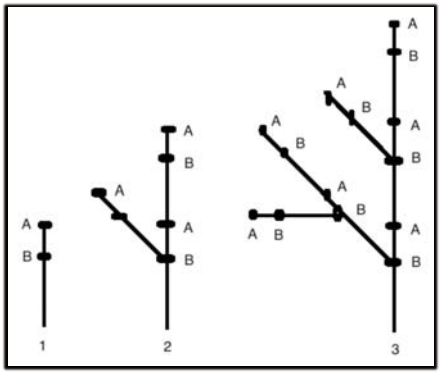


Figure 3. The L-Systems' Grammar in action.

```
1.  algorithm 2: pathMaker( current node )
2.  f ← locationOfCurrent( node )
3.  s ← locationOfCurrentParent( node )
4.  cur ← s
5.  while cur != f
6.      i, d ← findNearest( cur )
7.      if i = false
8.          return
9.      setVisited( d )
10.     storeToPath( d )
11.     cur ← d
12. return
13.
14. algorithm 3: findNearest( current point )
15. for n ← neighbor
16.     if nearestThan( n, point )
17.         t ← n
18. if isVisited( t ) || isDestination( t )
19.     return ( false, t )
20. return ( true, t )
```

Figure 4. The Pseudocode of Path-maker.

4.2 Software Specifications

This simulation system builds on:

- Mac OS 10.13 High Sierra & Windows 10 Home;

- Qt Creator 4.4.1, based on Qt 5.9.2 (Clang 7.0 (Apple), 64 bit) for Mac OS;
- Qt Creator 4.5.0, based on Qt 5.10.0 (MSVC 2015, 32 bit) for Windows;
- Modern OpenGL Shading Language (GLSL), which is implemented in Qt wrapper and native function; and
- Modern C/C++ programming language.

5. Experiments

The experiments were running on MacBook Pro and Acer M5 481TG too, with similar specifications.

5.1 Parameters

The parameters in detail:

- Game of Life generations (iterations) for determining tree(s) locations;
- The seed is for initiating Game of Life will be generated randomly. For experiment will be statically defined;
- The fixed location for start point *S*;
- The fixed locations for destination object *D*;
- Height-map that will be used;
- The height divider *div* for height-map to 3D;
- The walkable area constraint;
- The length *F*;
- The angle θ; and
- The maximum-paths that will be shown.

5.2 Environments

The environments were built refer to the approach that described in methodology, with some characteristics:

- Generated in the first run and uses the defined parameters;
- The trees, the destination and the initial point will generate by using the results of Game of Life map, which reduces to 1:5. So, in 5x5 points area will draw only one tree or destination's object or start point *R* for L-Systems;
- The destination is blindly known by *R*. The area for the destinations' object is 15x15 points; and
- The solutions paths will be drawn in 3 colors: red, green and blue.

5.3 Testing Scenario

In this research, we try to run our experiments as general as possible. First, we compare the result between Mac OS 10.13 and Windows 10. Second, we will focus on the results of L-Systems' grammar and this derivative.

To be fair, we consider using the smaller height-map and static seed for generating the environments. Then for testing the L-Systems' grammar and this derivative, we will consider the parameter θ for creating the branch and adding some rules for growing the paths.

5.4 Results

In this sub-chapter, we try to explain briefly about the experiment result. This experiment used, as described:

- Generation = 324;
- Seed = 0;

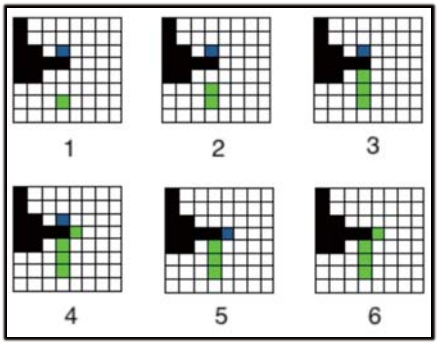


Figure 5. The Path that create by Cellular Automaton.

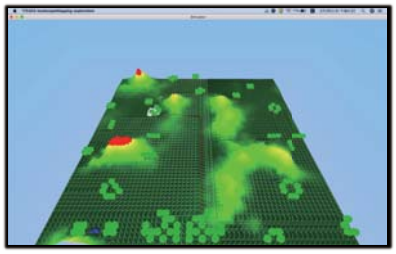


Figure 6. The Results when running in Mac OS 10.13.

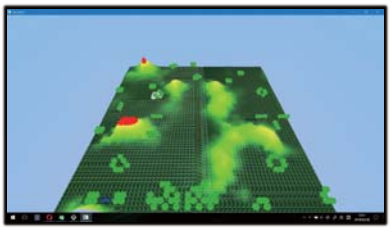


Figure 7. The Results when running in Windows 10.

- *S* = 75 (the tree at position 75th will be replaced);
- *D* = 30 (the tree at position 30th will be replaced);
- The height-map size is 257x257 pixels[1]. So, the surface will have size 257x257 points area;
- *div* = 10.0f (floating points);
- The walkable area constraints:
- Water Area < 1; and
- Mountain Area >= 6;
- *F* = 4 points (steps);
- θ = 45°; and
- The maximum-paths = 3.

First, we consider the environments that generated in Mac OS 10.13 and Windows 10. As described in Fig. 6 and Fig. 7, the environments are similar even in the different platform. This experiment gives a positive result that the paths successfully reached the destinations. For efficient time usage, we consider using Mac OS 10.13 for testing the L-Systems' grammar and this derivative rather than use both.

Consider the Fig 8, the paths also successfully reached the destinations. However, in Fig. 6, Fig. 7 and Fig. 8 cannot guarantee that all walkable areas are explored. To solve this problem, we consider changing the rule for generating branch and focus on

[1] http://www.rastertek.com/pic4005.gif

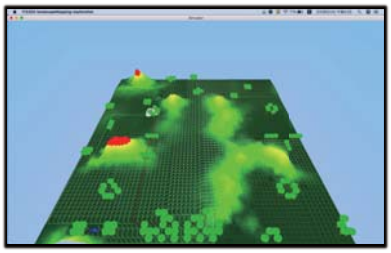


Figure 8. The Results when running in Mac OS 10.13 (θ = - 45°).

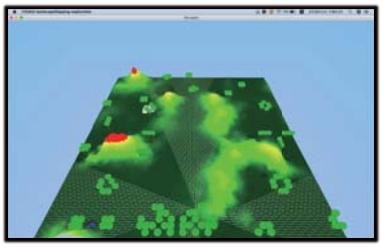


Figure 9. The Results when running in Mac OS 10.13 (θ = - 45°, flip).

angles that used. Previously, we use similar angle for generating on each iteration. Then we try to use some mechanism to make the angle flipping in next iteration and flip again. The flipping is only from -45° become 45° and vice-versa.

The results of flipping approach are described in Fig. 9, Fig. 10 and Fig. 11. The first rule that implemented in Fig. 9 is flip, that means always flipping in each iteration. Than in Fig. 10, the rule is flip-keep-flip-keep, that means in the first iteration will be flipping and in the second iteration will similar with the first iteration, then flip again and so on. Also, in Fig. 11, the rule is similar with Fig. 10, but the time for flip-keep is different, where the rule is keep-flip-keep-flip.

From Fig. 9, Fig. 10 and Fig. 11, the positive results are reached when we use the second and the third rule, which is described in Fig. 10 and Fig. 11. However, in the Fig. 9, the first rule cannot explore all area, even to create the path also impossible when surrounded by obstacles.

6. Conclusions

Path-finding in an unexplored land is a new area for developing an L-Systems based path-finding. Many possible solutions and challenges are still wide open to solve this problem.

The L-Systems' grammar that used was shown, that L-Systems can be used for exploring the surface. Then, can generate the path from source to destination. By combining with CA, growing the paths is more flexible and aware of the environments, which is if there is an impassable area, e.g., barriers such as a tree(s), or a visited location, we can reallocate the next point that will be created into a possible area. However, this grammar cannot guarantee that all walkable area will be explored. But this problem was solved by implementing the flipping mechanism that allows changing the direction of growing branch in each iteration, that can guarantee almost all walkable area explored.

Future work entails the implementation of GA as L-Systems' grammar modifier. Then, improve the algorithm so that it can be applied to the flying agent.

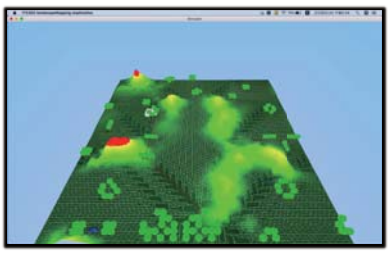


Figure 10. The Results when running in Mac OS 10.13 (θ = - 45°, flip-keep-flip-keep).

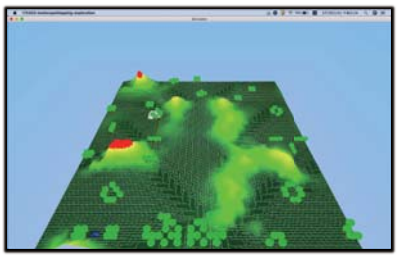


Figure 11. The Results when running in Mac OS 10.13 (θ = - 45°, keep-flip-keep-flip).

References

[Alfonseca 2003] M. Alfonseca, A. De la Puente and A. Suárez, “Cellular Automata and Probabilistic L Systems: An Example in Ecology,” 2003, 111-120.

[Areyan 2012] E. Areyan, “EvolvingAnts,” School of Informatics and Computing, Indiana University, 2012, in press.

[Gardner 1970] M. Gardner, “Mathematical Games - The fantastic combinations of John Conway’s new solitaire game “life””, Scientific American, 1970, pp.120-123.

[Ishida 2017] Y. Ishida, “Path-finding and path-finding system based on the extended L-system”, Innovation Japan, 2017.

[Koopman 2016] M. Koopman, “3D Path-Finding in A Voxelized Model of an Indoor Environment,” TU Delft, 2016.

[Kurth 2007] W. Kurth, “Specification of morphological models with L-systems and relational growth grammars,” Image – Journal of Interdisciplinary Image Science, 2007, vol.5.

[Lindenmayer 1968] A. Lindenmayer, “Mathematical models for cellular interaction in development, Parts I and II”, Journal of Theoretical Biology 18, 1968, pp.280–315.

[Nair 2017] S. H. Nair, A. Sinha and L. Vachhani, “Hilbert’s Space-filling Curve for Regions with Holes”, IEEE 56th Annual Conference on Decision and Control (CDC), 2017.

[Ochoa 1998] G. Ochoa, “On genetic algorithms and lindenmayer systems,” in International Conference on Parallel Problem Solving from Nature V, Amsterdam: Springer, 1998, pp. 335-344.

[Prusinkiewicz 1990] P. Prusinkiewicz and A. Lindenmayer, “The Algorithmic Beauty of Plants", New York: Springer-Verlag GmbH, 1990, pp.1-50.

[Shen 2012] S. Shen, N. Michael and V. Kumar, “Autonomous indoor 3D exploration with a micro-aerial vehicle”, International Conference on Robotics and Automation (ICRA), 2012, pp.9-15.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

人工市場シミュレーションを用いた 自己資本規制の金融市場に与える影響分析

Effects Analysis of CAR Regulations on Financial Markets using Artificial Market Simulations

平野 正徳 *1　米納 弘渡 *2　和泉 潔 *2
Masanori HIRANO　Hiroto YONENOH　Kiyoshi IZUMI

*1東京大学工学部
Faculty of Engineering, The University of Tokyo

*2東京大学大学院工学系研究科
School of Engineering, The University of Tokyo

Basel regulatory framework, one of CAR (capital adequacy ratio) regulations, is said to make markets destabilized in a previous study. But the previous study included some inappropriate assumptions. So, this study assessed this destabilizing effects with a new model. In my model, FCN agents and 2 kinds of portfolio agents, CAR regulated ones and not regulated ones, were included. Using this model, some simulations were run. As results, the simulations revealed some facts: 1. Asset management using portfolio stabilizes markets and the stabilizing effect are significant if there are a lot of markets included in the portfolio; 2. CAR regulation destabilizes markets and vanish the stabilizing effects of portfolio. In addition, the results of my simulations suggest that CAR regulation does not only raise the chance of price crashes but also depress whole price.

1. はじめに

国際金融システムにおいて，金融機関の連鎖破綻が引き起こす市場不安定化 (システミックリスク) を防ぐため，金融機関のリスク管理に関する規制が存在する．[BCBS 96]

バーゼル規制のようなリスク管理を行う銀行等が，株式市場等の資産市場で取引を行う場合，価格変動の大きい時期においては資産リスクが向上し，リスク管理の結果，資産売却を行い，市場を不安定化する可能性がある．[Hermsen 10]

このリスク管理が市場に与える影響について調査している [Yonenoh 12] の研究においては，現実の金融市場を反映した人工市場モデルにより，日中の株価変動を考慮したシミュレーションを行なっているが，日中の株価変動に関していくつか考慮されていない点があった．さらに，規制が適用された場合の金融機関エージェントの行動において，簡略化のため，大域的最適解にたどり着くことが難しい行動を行うことがあった．

そこで，本研究においては，ポートフォリオによる資産運用の最適化を図るエージェントが存在する市場における，自己資本 (CAR) 規制が市場に及ぼす影響について，既存の研究の問題点を解消する新たな人工市場モデルを構成することで，規制が市場の安定性に与える影響を検証することとした．

2. モデル

まず，市場のモデルは，ザラ場 (連続ダブルオークション (Coutinuous Double Auction)) とし，銘柄ごとに市場が存在するとし，銘柄 D_i の理論価格 (Fundamental Price) は，その対数リターンが平均 $\mu_{D_i} = 400$，分散 $(\sigma_{D_i})^2 = (0.001)^2$ に従うような幾何ブラウン運動 (ランダムウォーク) になるように設定し，各銘柄間の相関はないものと仮定した．

市場に存在するエージェントは，①標準的な FCN エージェント，②規制なしポートフォリオエージェント，③規制ありポートフォリオエージェントの三種類とした．基本的な初期設定を以下のようにした．

連絡先: 〒 113-8656 文京区本郷 7-3-1 東京大学工学部システム創成学科　システムデザイン&マネージメントコース
和泉・坂地研究室　平野正徳　HP: https://mhirano.jp

- 各エージェントが最初に持っている株：全銘柄 50 株ずつ
- 各ポートフォリオエージェントの資産：8000
- 各ポートフォリオエージェントのレバレッジ：10

2.1 FCN エージェント

マーケットの価格形成を促す標準的なエージェントとして，FCN エージェントを採用した．

従来の [Yonenoh 12] とは異なり，1 財を扱うポートフォリオエージェントではなく，予算制約のない FCN エージェントを利用することで，シミュレーションにおいて，綺麗な価格形成が期待できると考えた．

FCN エージェントとは，銘柄 D_i の期待リターン (価格変動率) をファンダメンタル項 (Fundamental)，テクニカル項 (Chartist)，ノイズ項 (Noise) の 3 つの加重平均として計算をし，売買をするエージェントである．[Chiarella 02][Torii 15]

$$r = \frac{1}{w_F + w_C + w_N}(w_F \cdot F + w_C \cdot C + w_N \cdot N) \quad (1)$$

$$F = \frac{1}{\tau_r} \ln (p_t / p_t^*) \quad (2)$$

$$C = \frac{1}{\tau} \sum_{i=1}^{\tau} \ln \left(\frac{p_{t-i}}{p_{t-i-1}} \right) \quad (3)$$

$$N \sim N(0, \sigma_N) \quad (4)$$

p_t:市場価格，p_t^*:ファンダメンタル価格，τ_r:平均回帰速度 ($[50, 150]$ の区間の一様分布)，τ:過去のどれくらいのチャートを参照するか ($[100, 200]$ の区間の一様分布)，σ_N:ノイズの正規分布の標準偏差 (0.001)

ここで，$w_F, w_C, w_N \geq 0$ は各項への重みであり，今回は，$w_F = 10.0, w_C = 1.0, w_N = 10.0$ と設定した．これは，市場の fat tail 性を評価できる尖度が現実に合うようにパラメーターサーチで決めた．

これらに基づき計算されたリターン r に基づいて，期待価格を $p_{exp} = p_t \cdot \exp(r \cdot \tau_h)$ と計算する．τ_h は，どれくらい現在のポジションを維持するかというパラメータであり，$[100, 200]$ の区間の一様分布になるように設定した．

この期待価格が現在価格よりも高い時には p_{exp} で 1 株の買い注文を出し，現在価格よりも低い時には p_{exp} で 1 株の売り注文を出すという行動を毎ステップで行う．ただし，最良気配値の方が条件が良い場合はその価格で約定する．

また，このエージェントはあくまで価格形成のための標準的なエージェントであるため，価格の流動性を高めるために，他のエージェントとは異なり，空売り，信用買を無制限に行える．

2.2 規制なしポートフォリオエージェント

規制なしポートフォリオエージェントとは，資産管理をポートフォリオ最適化という形で行うエージェントのことである．[Yonenoh 12] で使用されている複数財取引のリスク管理なしエージェントをベースにして作成した．

このポートフォリオエージェントは全ての銘柄をポートフォリオにしてリスク管理を行いつつ取り扱う．

また，ポートフォリオ保有期間 τ_P がそれぞれのエージェントで決まっており，平均，分散が 150 になるような指数分布に従っている．その τ_P ごとにポートフォリオを組みなおす．基本的に τ_P ごとにしか行動せず，その際の行動は次の様にした．①まだ約定していない注文をキャンセルする．②FCN エージェントと同様に期待価格を計算する．③ポートフォリオを組み直す (再計算)．④再構成したポートフォリオに合わせて現在のポジションとの差分の売買をする．

ここで，ポートフォリオの最適化においては，制約条件を予算制約と空売り禁止制約とし，目的関数は，[Markowitz 52] の平均分散アプローチに基づき，以下のように設定した．

$$\mathrm{EU}(\mathbb{x}) = {}^tP_{exp}\mathbb{x} - \frac{1}{2}{}^t\mathbb{x}\Omega\mathbb{x} \tag{5}$$

ただし，ここで，$\mathbb{x}$ は全銘柄のポジションの行列で，各要素は正の整数のみを取り，P_{exp} は全銘柄の期待価格行列，Ω は τ の期間の全銘柄間の分散・共分散行列である．この時，目的関数 $\mathrm{EU}(\mathbb{x})$ を最大化するような，制約条件を満たす $\mathbb{x}$ が，最適化後のポートフォリオである．

2.3 規制ありポートフォリオエージェント

規制ありポートフォリオエージェントは，2.2 節で述べたエージェントに加え，市場リスク規制導入後のバーゼル I[BCBS 96] および修正後バーゼル II[BCBS 06] の自己資本比率規制を参考にして作成した．

これらにおける自己資本比率規制は，過去の価格変動率の分布に基づく Value at Risk(VaR) を利用して，リスクを計算することで，今後の資産の損失可能性を自己資本と比較して，規制するものである．

$$(\text{自己資本比率 (CAR)}) = \frac{(\text{自己資本})}{12.5 \times (\text{市場リスク})} \tag{6}$$

$$(\text{市場リスク}) = (\text{時価総額}) \times (1 - \exp(\mathrm{VaR})) \tag{7}$$

$$(\mathrm{VaR}) = -1.0 \cdot (\text{99\%片側信頼区間}) \cdot \sqrt{T} \cdot \sqrt{\mathbb{x}\Omega\mathbb{x}} \tag{8}$$

ここでいう時価総額はあくまでも，ポートフォリオとして組んでいる株式の分のみである．VaR の計算においては，99% 片側信頼区間を上側に取った時の下側限界を採用し，ルート T 倍法を採用している．[FFR+ 08] $\mathbb{x}$ は全銘柄のポジション行列であり，Ω は全銘柄間の分散・共分散行列である．さらに，今回は，リスク資産の保有期間 T を 10 に設定し，Ω を過去，250 ステップの集計とした．これらは，市場リスク規制導入後のバーゼル I[BCBS 96] で初めて導入された基準に従っている．

もともと，約 250 営業日のデータを分布に使用し，保有期間は 10 日間とするとされていることに基づいているが，今回は，これを，単純に 250 ステップと 10 ステップとした．これは，資本市場における自己相似性 (フラクタル性)[Masahiro 00][Shozo 97] や，そもそも，ブラウン運動は Fractual Brownian Motion(fBm) のハースト指数 (H) を 1/2 にしたものであること，VaR は全体で時間スケールに依存していないことに基づいている．

これらの計算に基づき，CAR が 0.08 以下になるようにポートフォリオを規制する．

もし，規制に引っかかった場合は，以下のアルゴリズムにより，ポートフォリオを再構成し，CAR を小さくすることとした．基本的な方法としては，ポートフォリオの総額を下げていく (=レバレッジを下げていく)．

1. 現在のポートフォリオ構成 $\mathbb{x}$ に基づいた CAR を計算する．以下，CAR($\mathbb{x}$) とする．
2. 目標の規制基準の CAR=0.08 との比 R を以下のように定義し，計算する．
$$R = \frac{\mathrm{CAR}(\mathbb{x})}{0.08} \tag{9}$$
3. 現在のポートフォリオ構成 $\mathbb{x}$ の期待価格での時価総額を計算する．以下，Val($\mathbb{x}$) とする．
4. 次式に基づき，再度ポートフォリオを構成する時の予算制約 B を以下のように定義し，計算する．
$$B = \frac{1}{\frac{R-1}{2}+1} \cdot \mathrm{Val}(\mathbb{x}) \tag{10}$$
5. 予算制約 B のもとで，規制なしポートフォリオエージェント同様にポートフォリを最適化し，$\mathbb{x}$ を更新する．
6. この新しい $\mathbb{x}$ が $\mathrm{CAR}(\mathbb{x}) \leq 0.08$ を満たしているかチェックする
7. 上記の式を満たしていれば，これを最終的な規制に違反しない最適なポートフォリオのポジションとし，満たしていなければ，1 に戻る．

規制が適用される状況でのエージェントの行動アルゴリズムは [Yonenoh 12] の手法と比べ，より最適化が進むと考えている．そもそも，[Yonenoh 12] ではチェビシェフ距離が 1 になるポジションのなかで最適なものを取っていき，規制に引っかからないところにたどり着くまで繰り返すという手法を使っていたが，これでは，あくまで局所最適に落ちるだけであり，大域的最適にはたどり着けない．さらに，実際に規制に引っかかった時の人の行う対応を考えると，最小単位ずつ売って行くということをしているとは考えにくい．

3. 実験方法

本研究の中心である，CAR 規制が市場に与える影響についての実験を説明する．

以下で説明する全ての実験において，評価する結果としては，[Yonenoh 12] と同様に理論価格 (ファンダメンタル価格) との下への乖離度 $\mathrm{DV_t} = \ln(p_t/p_t^*)$ を使った．これに関して，$\mathrm{DV_t} \leq -0.05$ となっているステップ数を N_{DV} と定義した．これは，[Yonenoh 12] と閾値が異なっている．

さらに，全てのパラメータセットにおける実験は 10 試行ずつ行った．

また，試行ステップは 60000 ステップとした．

以下で，実際に行った 2 つの実験を説明する．

3.1 ポートフォリオエージェントの与える影響について

最初に，規制なしポートフォリオエージェントと FCN エージェントのみを使って，ポートフォリオエージェントの市場に与える影響について実験した．

- FCN エージェント数：10000(固定)
- 規制なしポートフォリオエージェント数
  ：0-100(10 刻みで変更)
- 銘柄 (マーケット) 数：1-10(1 刻みで変更)

3.2 規制が与える影響について

次に，ポートフォリオエージェントの合計，つまり，規制なしポートフォリオエージェントと規制ありポートフォリオエージェントの合計を 100 に固定し，比率を変えて実験した．

- FCN エージェント数：10000(固定)
- 規制なしポートフォリオエージェント数
  ：0-100(10 刻みで変更)
- 規制ありポートフォリオエージェント数
  ：100 – (規制なしポートフォリオエージェント数)
- 銘柄 (マーケット) 数：1-10(1 刻みで変更)

4. 結果・議論

4.1 ポートフォリオエージェントの与える影響について

結果のグラフは図 1 の通りである．

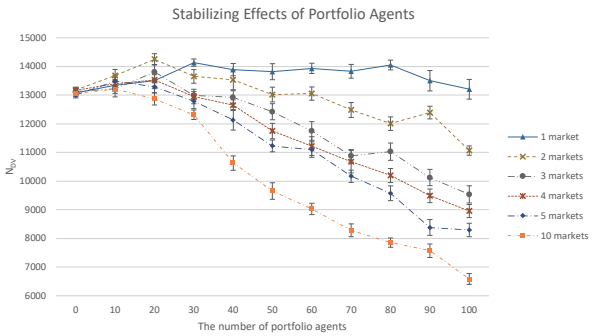


図 1: ポートフォリオエージェントが市場に与える影響のグラフ．縦軸が理論価格から閾値以上の下への乖離が起きているステップ数．横軸はポートフォリオエージェントの数．銘柄 (マーケット) 数ごとにプロットしている．エラーバーは 1σ である．

ポートフォリオエージェントが増えるに従って，理論価格との下への乖離が少なくなっている．これは，ポートフォリオ運用が，市場を安定化させ，効率化させることを示唆している．

また，銘柄 (マーケット) 数が増えるほど，ポートフォリオエージェントが有効になっていることが確認できる．[Yonenoh 12] の研究においては，銘柄 (マーケット) 数が 1 である時と 2 である時の差は大きかったが，2 以上に増えても，大きな違いが確認できなかった．しかし，今回の結果では，3 銘柄 (マーケット) や 4 銘柄 (マーケット) の場合でも同様に違いが発生しており，さらには 10 銘柄 (マーケット) でも十分に乖離の低下が確認できた．

このような違いが発生したかという理由だが，[Yonenoh 12] では，価格形成をする標準的なエージェントを導入しておらず，1 銘柄で [Markowitz 52] の手法でポートフォリオ最適化をするエージェントを導入していた．そのため，どのエージェントも非常に似た動きをしてしまい，ポートフォリオの効果が急激にサチレーションしてしまっていたのだと考えている．

また，FCN エージェントが 10000 であるのに対して，ポートフォリオエージェントが高々100 しか存在しないにも関わらず，充分な影響が確認できていることからも，FCN エージェントとポートフォリオエージェント比率の妥当性が確認できる．

4.2 規制が与える影響について

結果のグラフは図 2 の通りである．

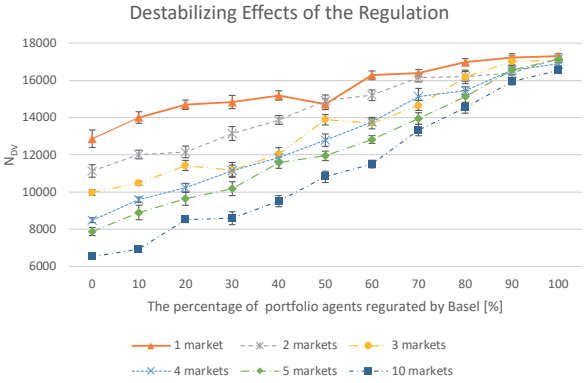


図 2: 規制が市場に与える影響のグラフ．縦軸が理論価格から閾値以上の下への乖離が起きているステップ数．横軸がポートフォリオエージェントの中での規制ありエージェントの割合．銘柄 (マーケット) 数ごとにプロットしている．エラーバーは 1σ である．

規制ありポートフォリオエージェントがポートフォリオエージェントの中で占める割合を上げると，理論価格からの下への乖離が大きくなることが確認できる．

規制ありポートフォリオエージェントがポートフォリオエージェントの中で占める割合を上げて行くと，ポートフォリオエージェントが一切いない状況よりも，市場が不安定化する．

さらに，規制ありポートフォリオエージェントがポートフォリオエージェントの中で占める割合が増えるほど，銘柄 (マーケット) 数の増加に伴う市場安定性が小さくなっていることが確認できる．

最終的に図 2 の “The percentage of portfolio agents regulated by Basel” が 100%の時においては，ほとんど銘柄 (マーケット) 数による変化がないことから，規制ありポートフォリオエージェントは，ポートフォリオエージェントが通常持つ，4.1 節で明らかにした，ポートフォリオエージェントの市場安定化効果を完全に持っておらず，逆に，市場を不安定化することがわかる．

これがどのようなメカニズムであるのかということを考えると，市場価格が落ち，リスクが上がったような場合に，規制に引っかかり，手元のポジションを下げなければならない，というような現象が発生し，売りが売りを呼ぶような現象が起きていると考えられる．

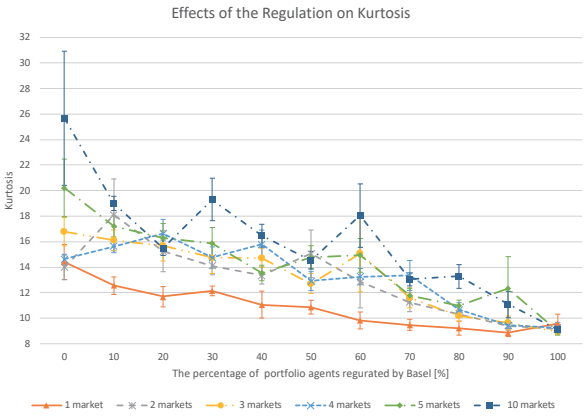


図 3: 規制ありエージェントの割合と尖度の関係．縦軸が尖度，横軸がポートフォリオエージェントの中での規制ありエージェントの割合．銘柄 (マーケット) 数ごとにプロットしている．エラーバーは 1σ である．

この図 3 は興味深い結果を示唆している．

規制ありポートフォリオエージェントの比率が大きくなるにつれて，尖度が小さくなるのである．これは，言い換えれば，規制ありポートフォリオエージェントの比率が大きくなると，fat tail 性が小さくなるというものである．

理論価格から下への乖離が大きくなるとともに，fat tail 性も小さくなるということは，逆に上への乖離が小さくなっているということを示唆していると考える．つまり，規制の導入により，連鎖的に売りが発生し，価格崩壊が起こるという考えは従来あったものの，実際には全体として，価格を押し下げているということが示唆されているのではないだろうか．

これが，どのような現象を示しているのかを考えてみる．VaR を使用している以上，上への価格の変動が増加した場合には，リターンの標準偏差が増加するため，リスク分布の裾野が広がることとなる．これは，VaR を正規分布で計算していることの支障であるが，リスク分布の裾野が広がれば，それは，下落のリスクも増加したことになってしまう．このとき，VaR を利用すると，99%信頼区間における，下落可能性幅も増加する．つまり，リスクが増加し，CAR の分母が小さくなってしまい，CAR は大きくなってしまう．その結果，市場価格が上昇していても，買い控え，あるいは，売りが発生してしまうのではないだろうか．

尖度と規制ありポートフォリオエージェントの比率の相関係数も，マーケット数毎の平均が，-0.91 ± 0.033 となっており，かなり強い傾向性が確認できており，影響は否定できない．

これらの結果から，CAR 規制は，市場価格の下落を増加させるだけでなく，市場価格の上昇を抑えてしまい，市場に悪い影響を及ぼしていると考えられる．

5. まとめ

本研究においては，バーゼル規制のような CAR 規制が市場に不安定性を与えると想定し，ポートフォリオ最適化および規制がザラ場 (連続ダブルオークション) 方式の市場に与える影響について調査した．その結果，ポートフォリオ最適化に基づいた資産運用は市場を安定化させ，銘柄 (マーケット) 数が増えるほどその市場安定化効果は大きくなる．さらに，CAR 規制の存在は，市場に不安定性をもたらし，ポートフォリオの市場安定化効果をなくしていくということがわかった．これらの結果は，どれも，[Yonenoh 12] の研究よりも明確な結果が出ており，今回のモデルの良さが表れていた．

さらに，CAR 規制の存在は，市場価格を押し下げ，価格上昇も抑えてしまうということが示唆された．これはこれまでになかった視点であり，かなりの強い傾向性を確認できている．これについては，さらに，この論を裏付ける検証をすべきだと考えている．

これらの結果から，複数の側面で CAR 規制は市場に悪影響を及ぼしている，と考えられる．

また，今回は，従来のバーゼル II までの規制に従ってモデルを組んだが，バーゼル III に対応したモデルを作成して，最新の規制がより良いのかどうかの検証をすべきだと考えている．

参考文献

[BCBS 06] Basel Committee on Banking Supervision : International Convergence of Capital Measurement and Capital Standards, http://www.bis.org/publ/bcbs128.pdf (2006)

[BCBS 96] Basle Committee on Banking Supervision : Amendment to the Capital Accord to incorporate market risks, https://www.bis.org/publ/bcbs24.pdf (1996)

[Chiarella 02] Chiarella, C. and Iori, G.: A simulation analysis of the microstructure of double auction markets, *Quantitative Finance*, Vol. 2, No. 5, pp. 346-353 (2002)

[FFR+ 08] FFR+ : ＶａＲの計測と検証, http://www.ffr-plus.jp/material/pdf/0006/risk_meter_quantification_03.pdf (2008)

[Hermsen 10] Hermsen, O.: Does Basel II destabilize financial markets? An agent-based financial market perspective, *The European Physical Journal B*, Vol. 73, No. 1, pp. 29-40 (2010)

[Markowitz 52] Markowitz, H.: PORTFOLIO SELECTION, *The Journal of Finance*, Vol. 7, No. 1, pp. 77-91 (1952)

[Masahiro 00] Masahiro, K.: 価格変動のフラクタル性への統計物理的アプローチ, Bachelor Thesis, The University of Tokyo (2000)

[Shozo 97] Shozo, T., Yoshikazu, I., and Wataru, T.: 自己相似性を用いたフラクタル時系列の予測とその応用, 情報処理学会研究報告数理モデル化と問題解決（MPS）, Vol. 1997, No. 95(1997-MPS-015), pp. 33-38 (1997)

[Torii 15] Torii, T., Izumi, K., and Yamada, K.: Shock transfer by arbitrage trading: analysis using multi-asset artificial market, *Evolutionary and Institutional Economics Review*, Vol. 12, No. 2, pp. 395-412 (2015)

[Yonenoh 12] Yonenoh, H. and Izumi, K.: 人工市場を用いた自己資本比率規制に基づく市場リスク管理が複数資産市場に与える影響の分析, *JPX Working Paper*, Vol. 18 (2012)

[2Z2] [General Session] 5. Web Intelligence

座長:西村 悟史（産業技術総合研究所）

Wed. Jun 6, 2018 1:20 PM - 2:40 PM Room Z (3F Matsu Take)

[2Z2-01] Synthetic Accessibility Assessment Using Auxiliary Responses

○Shun Ito[1], Yukino Baba[1], Tetsu Isomura[2], Hisashi Kashima[1,3] （1. Graduate school of Informatics, Kyoto University, 2. Mitsubishi Chemical Holdings Corporation, 3. RIKEN Center for Advanced Intelligence Project）

1:20 PM - 1:40 PM

[2Z2-02] Crowd Neural Network

○Yusuke Sakata[1], Yukino Baba[1], Kashima Hisashi[1,2] （1. Kyoto University Department of Intelligence Science and Technology, 2. RIKEN Center for AIP）

1:40 PM - 2:00 PM

[2Z2-03] LOD Surfer API: A web API for LOD exploration based on class-class relationships

○Atsuko Yamaguchi[1], Norio Kobayashi[2], Hiroshi Masuya[2], Yasunori Yamamoto[1], Kouji Kozaki[3] （1. Research Organization of Information and Systems, 2. RIKEN, 3. Osaka University）

2:00 PM - 2:20 PM

[2Z2-04] A Preliminary Approach on a Cooperative Query Editing System using Dynamic Ontology Mappings

○Takuya Adachi[1], Naoki Fukuta[2] （1. Graduate School of Integrated Science and Technology, Shizuoka University, 2. College of Informatics, Academic Institute, Shizuoka University）

2:20 PM - 2:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

補助回答を利用した意見集約法による化合物の合成可能性判定

Synthetic Accessibility Assessment Using Auxiliary Responses

伊藤 駿 *1 Shun Ito　馬場 雪乃 *2 Yukino Baba　磯村 哲 *3 Tetsu Isomura　鹿島 久嗣 *1*4 Hisashi Kashima

*1京都大学 Kyoto University　*2筑波大学 University of Tsukuba　*3三菱ケミカルホールディングス Mitsubishi Chemical Holdings Corporation

*4理化学研究所 革新知能統合研究センター RIKEN Center for Advanced Intelligence Project

Despite of the recent advances in computational approaches to discovering new compounds, synthetic accessibility assessment of chemical compounds is still an AI-hard task. Therefore human computation approaches are potentially promising. We introduce a new aggregation method of semi-expert judgments that consist of five-grade synthetic accessibility scores along with annotations of obstructive substructures to their synthesis. Our proposed two-stage method treats the obstructive substructures as auxiliary responses. The first stage divides both the semi-experts and the atoms in the substructures into clusters using stochastic block models. The second stage generates five-grade responses based on the clusters. Our experiments show that the use of the auxiliary responses improves the prediction performance.

1. はじめに

創薬における分子設計プロセスは計算機による自動化が進んでおり，新しい候補化合物を短時間で大量に生成することが可能となっている．計算機により生成された化合物は合成可能性が保証されていないため，それを検証する工程が必要であるが，過去に提案された自動化手法 [1, 2] は精度や計算時間の面で実用レベルに達しておらず，専門家による判定作業に依存しており，プロセス全体の高速化のための課題となっている．少数の作業者に依存することによる非効率化に対処するためのアプローチとして，クラウドソーシングの仕組みを活用し多数のワーカで作業を分担する方法が存在する．合成可能性判定においても，専門家が行なっている判定作業を専門家ではない多数のワーカが分担して行うことで，より効率的な判定が可能となる．しかし，合成可能性判定は高度な専門知識を必要とする作業であるため，化学に関する知識を全く持たない人による回答から正しい判定が得られることは期待できない．この問題に対処するため，回答者として化学にある程度知識を持つ準専門家を導入することが提案されており [3]，本研究でも準専門家の回答を集約することによって合成可能性の予測を行う．

本研究では，着眼点の違いから準専門家の能力や化合物の性質などの情報を得ることを目的とし，判定回答とは別の補助回答として，適切でないと考えた部分構造を選択するように依頼する．つまり，準専門家は図 1 に示されるような 2 種類の回答を行う．(1) 5 段階での判定では，判定対象の化合物が合成可能であると判断するほど高い値を回答する．(2) 悪い部分構造の選択では，判定対象の化合物のうち合成するのに不適切な部分構造を構成している原子を 0 個以上選択する．

部分構造の選択で、似た能力の準専門家による原子の選び方・似た性質の原子の選ばれ方は，それぞれ似たような傾向になると仮定し，補助回答に対して確率的ブロックモデルを適用することで，準専門家と原子をいくつかのクラスタに分類する．これを第 1 段階とし，第 2 段階で，各準専門家が独立して持つ潜在パラメータと，準専門家・原子のクラスタを考慮した潜在パラメータの合計 3 種類のパラメータを用いて判定回答の生成をモデル化する，という 2 段階での予測手法を提案する．潜在クラスタを考慮したモデルには，ワーカの能力をクラスタに依存して決定したもの [4] や，ワーカとタスクの両方のクラスタを考慮したもの [5] 等が存在するが，これらの既存手法とは異なり，本研究はクラスタを補助回答から推定し，判定回答の生成モデルを構築する．

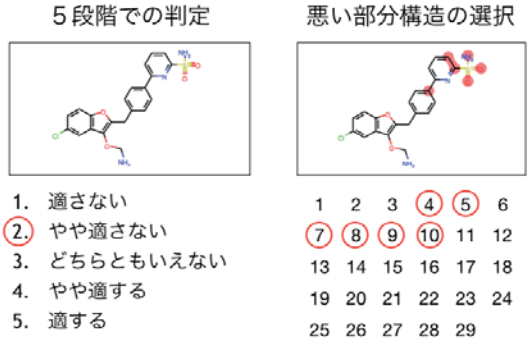


図 1: 準専門家の回答の例.

2. 提案手法

補助回答を観測データとする第 1 段階と，判定回答を観測データとする第 2 段階とで構成される，2 段階での合成可能性の予測手法を提案する．第 1 段階では確率的ブロックモデル (Stochastic Block Model; SBM) [6] または混合メンバシップ確率的ブロックモデル (Mixed Membership Stochastic Blockmodels; MMSB) [7] によるクラスタリングを行い，第 2 段階では 3 種類の潜在パラメータによる判定回答の生成モデルを構築する．観測データを元にこれらのモデルの推定を行い，化合物の合成可能性の予測値を得る．以下では，準専

連絡先: 伊藤 駿，京都大学大学院情報学研究科知能情報学専攻，s.itoh248@ml.ist.i.kyoto-u.ac.jp

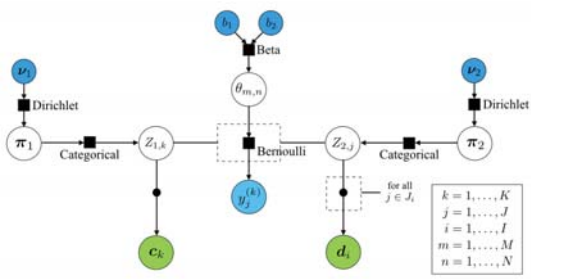


図 2: SBM によるクラスタリング.

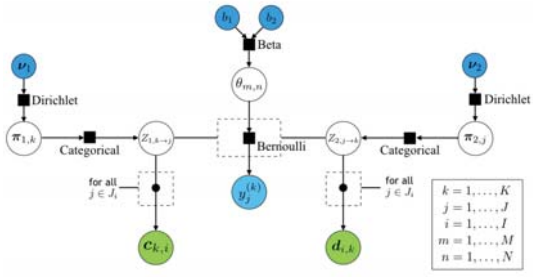


図 3: MMSB によるクラスタリング.

門家を $k = 1, \ldots, K$，化合物を $i = 1, \ldots, I$, 化合物 i の真の合成可能性を $t_i \in \{0, 1\}$，化合物 i の構成原子の集合を $\mathcal{J}_i$，全化合物に対する原子を $j = 1, \ldots, J$ $(J = \Sigma_i |\mathcal{J}_i|)$，とする．ただし，$t_i = 1$ が化合物 i が合成可能であることに対応する．観測データは，準専門家 k の化合物 i に対する判定回答 $x_i^{(k)} \in \{1, \ldots, 5\}$ と原子 j に対する回答 $y_j^{(k)} \in \{0, 1\}$ で，$y_j^{(k)} = 1$ の時に準専門家 k が原子 j を選択したことを表す．

2.1 クラスタリング

第 1 段階では，補助回答データに SBM または MMSB を適用することで，準専門家と原子をいくつかのクラスタに分類する．SBM は，関係ネットワークにおいて，あるノードとノードとの間にエッジが存在する確率は，それぞれのノードが所属するクラスタに依存するという仮定をおくモデルである．SBM のもとで補助回答 $y_j^{(k)}$ は次のベルヌーイ分布に従って生成される．

$$y_j^{(k)} | \theta_{m,n}, Z_{1,k}, Z_{2,j} \sim \text{Bernoulli}(y_j^{(k)} | \theta_{Z_{1,k}, Z_{2,j}})$$

ここで $Z_{1,k} = 1, \ldots, M$, $Z_{2,j} = 1, \ldots, N$ はそれぞれ準専門家・原子の所属クラスタで，$\theta_{m,n}$ はクラスタ m に属する準専門家がクラスタ n に属する原子を選択する確率を表す．これらのパラメータは次のような分布によって生成される．

$$Z_{1,k} | \boldsymbol{\pi}_1 \sim \text{Categorical}(Z_{1,k} | \boldsymbol{\pi}_1)$$
$$Z_{2,j} | \boldsymbol{\pi}_2 \sim \text{Categorical}(Z_{2,j} | \boldsymbol{\pi}_2)$$
$$\theta_{m,n} \sim \text{Beta}(\theta_{m,n} | b_1, b_2)$$

$\boldsymbol{\pi}_1, \boldsymbol{\pi}_2$ はディリクレ分布により生成されるパラメータで，それぞれ各次元が各クラスタ m, n に所属する確率を表す．以上の過程により準専門家と原子のクラスタ $\{Z_{1,k}\}_k, \{Z_{2,j}\}_j$ が得られるが，第 2 段階のためにこれらを準専門家と化合物のクラスタ $\{\boldsymbol{c}_k\}_k, \{\boldsymbol{d}_i\}_i$ に変換する．化合物 i に対するクラスタは，原子 $j \in \mathcal{J}_i$ が所属しているクラスタを複合したものとして決定されると仮定し，次のように変換を行う．

$$\boldsymbol{c}_k = (c_{k,1}, \ldots, c_{k,M})^\top$$
$$c_{k,m} = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} \mathbb{I}(Z_{1,k} = m)$$
$$\boldsymbol{d}_i = (d_{i,1}, \ldots, d_{i,N})^\top$$
$$d_{i,n} = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} \mathbb{I}(Z_{2,j} = n)$$

ここで $\mathbb{I}(\cdot)$ は指示関数を表す．以上の SBM により準専門家と原子，化合物のクラスタが得られる過程を図 2 に示す．

一方，MMSB は SBM の拡張として提案されたモデルで，SBM がインデックス毎にクラスタリングを行うのに対し，MMSB はインデックスのペア毎によってクラスタリングを行うため，一般に一つのインデックスは複数のブロックに割り当てられる．MMSB のもとで補助回答 $y_j^{(k)}$ は次のベルヌーイ分布に従って生成される．

$$y_j^{(k)} | \theta_{m,n}, Z_{1,k \to j}, Z_{2,j \to k} \sim \text{Bernoulli}(y_j^{(k)} | \theta_{Z_{1,k \to j}, Z_{2,j \to k}})$$

ここで $Z_{1,k \to j} = 1, \ldots, M$, $Z_{2,j \to k} = 1, \ldots, N$ はそれぞれ準専門家・原子の所属クラスタであるが，SBM の場合とは異なりどちらも k と j のペア毎にクラスタが決定しており，次のカテゴリカル分布から生成される．

$$Z_{1,k \to j} | \boldsymbol{\pi}_{1,k} \sim \text{Categorical}(Z_{1,k \to j} | \boldsymbol{\pi}_{1,k})$$
$$Z_{2,j \to k} | \boldsymbol{\pi}_{2,j} \sim \text{Categorical}(Z_{2,j \to k} | \boldsymbol{\pi}_{2,j})$$

$\boldsymbol{\pi}_{1,k}, \boldsymbol{\pi}_{2,j}$ は SBM と同様にディリクレ分布により生成されるパラメータであるが，MMSB では各準専門家 k，原子 j 毎に異なる値をとる．また，$\theta_{m,n}$ は SBM と同様にベータ分布から生成される．以上の過程により準専門家と原子のクラスタ $\{Z_{1,k \to j}\}_{k,j}, \{Z_{2,j \to k}\}_{j,k}$ が得られるが，これらも第 2 段階のために準専門家と化合物のクラスタ $\{\boldsymbol{c}_{k,i}\}_{k,i}, \{\boldsymbol{d}_{i,k}\}_{i,k}$ に変換する．化合物 i に対するクラスタは，原子 $j \in \mathcal{J}_i$ が所属しているクラスタの複合として決定される．

$$\boldsymbol{c}_{k,i} = (c_{k,i,1}, \ldots, c_{k,i,M})^\top$$
$$c_{k,i,m} = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} \mathbb{I}(Z_{1,k \to j} = m)$$
$$\boldsymbol{d}_{i,k} = (d_{i,k,1}, \ldots, d_{i,k,N})^\top$$
$$d_{i,k,n} = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} \mathbb{I}(Z_{2,j \to k} = n)$$

以上の MMSB により準専門家と原子，化合物のクラスタが得られる過程を図 3 に示す．

2.2 判定回答生成モデル

第 2 段階では，第 1 段階で得られた準専門家と化合物に対するクラスタリング結果を考慮し，5 段階の判定回答の生成を確率モデルで表現する．本手法では，判定回答を生成する潜在パラメータとして (1) 各準専門家 k が独立に持つパラメータ $\boldsymbol{\alpha}_t^{(k)}$，(2) 準専門家のクラスタ m 毎に共通するパラメータ $\boldsymbol{\beta}_{t,m}$，(3) 化合物のクラスタ n 毎に共通するパラメータ $\boldsymbol{\gamma}_n$ の 3 種類を仮定する．ここで $t \in \{0, 1\}$ は真の合成可能性を表し，$t = 1$ が合成可能な場合に対応する．これら 3 種類のパラメー

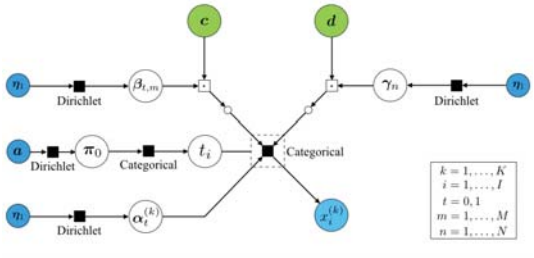


図 4: 判定回答生成モデル.

タは全て 5 次元ディリクレ分布から生成され，各次元が 5 段階での各段階の生成確率を表す．つまり，$\boldsymbol{\alpha}_t^{(k)}$ は真の合成可能性が t の時の準専門家 k の回答傾向，$\boldsymbol{\beta}_{t,m}$ は真の合成可能性が t の時のクラスタ m に属する準専門家の回答傾向，$\boldsymbol{\gamma}_n$ はクラスタ n に属する化合物の被回答傾向となる．$\boldsymbol{\gamma}_n$ は化合物自体の性質を表すため，真の合成可能性には依存しないと仮定する．これらのパラメータの生成は次のようになる．

$$\boldsymbol{\alpha}_t^{(k)} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha}_t^{(k)}|\boldsymbol{\eta}_1)$$
$$\boldsymbol{\beta}_{t,m} \sim \mathrm{Dirichlet}(\boldsymbol{\beta}_{t,m}|\boldsymbol{\eta}_2)$$
$$\boldsymbol{\gamma}_n \sim \mathrm{Dirichlet}(\boldsymbol{\gamma}_n|\boldsymbol{\eta}_3)$$

一方，化合物 i の真の合成可能性 t_i は 2 次元パラメータ $\boldsymbol{\pi}_0$ を事前分布として，次のカテゴリカル分布から生成される．

$$t_i|\boldsymbol{\pi}_0 \sim \mathrm{Categorical}(t_i|\boldsymbol{\pi}_0)$$

以上のパラメータを考慮することで，判定回答 $x_i^{(k)}$ が生成される．

$$x_i^{(k)}|t_i, \boldsymbol{\alpha}_{t,k}, \boldsymbol{B}_t, \boldsymbol{\Gamma}, \boldsymbol{c}_k, \boldsymbol{d}_i \sim \mathrm{Categorical}\left(x_i^{(k)} \middle| \frac{\boldsymbol{\alpha}_{t_i}^{(k)} + \boldsymbol{c}_k^\top \boldsymbol{B}_{t_i} + \boldsymbol{d}_i^\top \boldsymbol{\Gamma}}{\omega}\right)$$

ここで ω は正規化定数，$\boldsymbol{B}_t \in [0,1]^{M\times 5}, \boldsymbol{\Gamma} \in [0,1]^{N\times 5}$ はそれぞれ $\{\boldsymbol{\beta}_{t,m}\}_m, \{\boldsymbol{\gamma}_n\}_n$ を並べた行列であるとする．

$$\boldsymbol{B}_t = \begin{pmatrix} \boldsymbol{\beta}_{t,1} \\ \vdots \\ \boldsymbol{\beta}_{t,M} \end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_N \end{pmatrix}$$

判定回答を観測データとして以上のモデルを推定することにより，合成可能性の予測値 $\{\hat{t}_i\}_i$ を得る．第 2 段階の過程を図 4 に示す．

3. 実験

提案手法を実際に準専門家から得られた回答データに対して適用し，性能を検証する．今回用いるデータセットは，分子設計研究に携わる 9 名を準専門家とし，全員が 5 段階の判定回答と，部分構造の選択を行なったものである．対象となる新規化合物は，公開データベース上の化合物に対してランダムな部分構造置換を行うことによって生成されたものである．新規化合物の真の合成可能性は，創薬の専門家 5 名が 95 個の新規化合物に対して 2 度の回答を行い，化合物 i に対して 5 名全員が 3 以下の判定をした場合には $t_i = 0$，全員が 3 以上の判定をした場合には $t_i = 1$ となるように決定した．今回はこの方法で正解が得られた 59 個の新規化合物を対象に実験を行う．

表 1: 提案手法と比較手法の実データに対する予測性能.

手法	AUC
MV	0.886
Raykar	0.939
Only-α	0.925
SBM-$(\alpha+\beta)$	0.924
SBM-$(\alpha+\gamma)$	0.944
SBM-$(\beta+\gamma)$	0.887
SBM-$(\alpha+\beta+\gamma)$	**0.955**
MMSB-$(\alpha+\beta)$	0.927
MMSB-$(\alpha+\gamma)$	0.947
MMSB-$(\beta+\gamma)$	0.875
MMSB-$(\alpha+\beta+\gamma)$	**0.959**

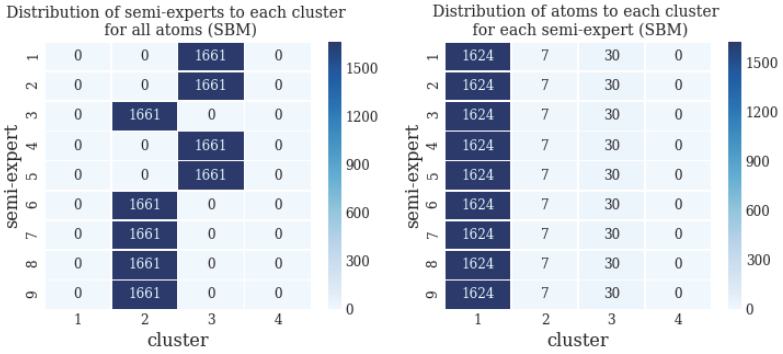
比較手法として，補助回答を用いない手法を 3 つ，補助回答を用いる手法を 6 つ用意した．補助回答を用いない（5 段階の判定回答のみを用いる）手法は，9 人の判定回答の平均値を予測値として用いる手法 (MV)，回答が段階ラベルで得られる場合に感度と特異度によってワーカの能力をモデル化した手法 [8] (Raykar)，本手法での $x_i^{(k)}$ の生成を準専門家依存のパラメータ $\boldsymbol{\alpha}_t^{(k)}$ のみを用いて行う手法 (Only-α) の 3 つを用意する．補助回答を用いる手法は Only-α と同様に提案手法の第 2 段階で用いるパラメータを変化させた手法であり，$\boldsymbol{\alpha}_t^{(k)}$ と準専門家クラスタ依存のパラメータ $\boldsymbol{\beta}_{t,m}$ を用いる手法 (SBM-$(\alpha+\beta)$, MMSB-$(\alpha+\beta)$)，$\boldsymbol{\alpha}_t^{(k)}$ と化合物クラスタ依存のパラメータ $\boldsymbol{\gamma}_n$ を用いる手法 (SBM-$(\alpha+\gamma)$, MMSB-$(\alpha+\gamma)$)，準専門家・原子クラスタ依存のパラメータ $\boldsymbol{\beta}_{t,m}, \boldsymbol{\gamma}_n$ を用いる手法 (SBM-$(\beta+\gamma)$, MMSB-$(\beta+\gamma)$) の計 6 つを用意する．

推論は第 1 段階のクラスタリング部分と第 2 段階の判定回答生成モデル部分のそれぞれで，最大事後確率推定を独立して行う．準専門家と原子のクラスタ数は $M = 4, N = 4$ に設定した．

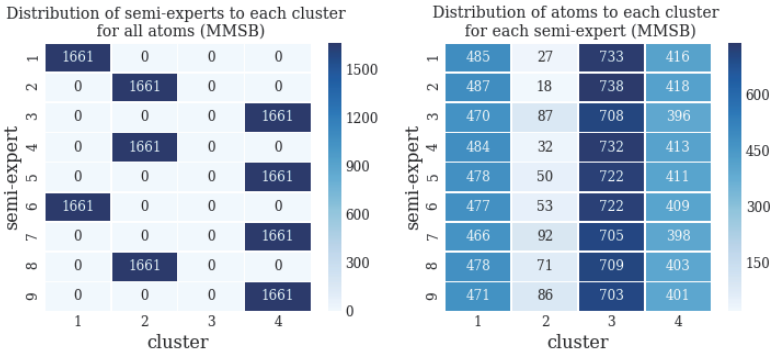
4. 実験結果

提案手法と比較手法を実際のデータセットに適用した結果を AUC により評価したものを表 1 に示す．補助回答に対するクラスタリングに SBM と MMSB のどちらを用いた場合にも，提案手法の SBM-$(\alpha+\beta+\gamma)$, MMSB-$(\alpha+\beta+\gamma)$ が最も高い AUC の値を得られた．また，判定回答の生成パラメータとして α, β を使っている手法と α, γ を使っている手法を比較すると，γ を考慮した手法の方が高い AUC が得られている．これは，α と β が準専門家に関する潜在パラメータであるのに対して γ が化合物に関する潜在パラメータであるため，準専門家と化合物の両方の性質を同時に考慮することでより正確なモデルとなったことが理由として考えられる．さらに，3 つのパラメータ全てを考慮した提案手法が最も高い AUC の値を得られていることから，β が α からは得られない準専門家同士に共通する能力を抽出していると考えられる．

続いて SBM と MMSB によって得られたクラスタリングの結果について図 5 を用いて検証する．まず準専門家のクラスタリング結果を見ると，SBM による結果の方が MMSB によ



(a) SBM



(b) MMSB

図 5: SBM と MMSB によるクラスタリング結果．左側が準専門家 9 人の全原子 1661 個に対するクラスタ，右側が原子 1661 個の準専門家 9 人に対するクラスタを表す．準専門家は個人の判定回答の AUC が低い順に上から並べられている．

る結果よりも明確であるが，いずれの場合でも AUC の比較的高い準専門家のクラスタと低い準専門家のクラスタが生成されており，能力による分類がうまく行われていることがわかる．続いて原子のクラスタリング結果を比較する．SBM による結果ではほとんど全ての原子がクラスタ 1 に割り当てられているのに対し，MMSB による結果ではより詳細な分類結果が得られていることがわかる．また，MMSB でのクラスタ 2 に注目すると，ペアとなる準専門家の AUC が高くなるにつれてクラスタ 2 に属する原子の個数が増加する傾向にあるため，悪い部分構造を構成している可能性が高い原子が MMSB によるクラスタリングによりうまく分類されていると考えられる．

5. むすび

補助回答から得たクラスタリングの結果を考慮し，準専門家の判定回答を統合することによって化合物の合成可能性判定を行うモデルを提案し，提案手法の有効性を実験により示した．今回は，回答者が適切でない部分構造を発見することにより化合物が合成不可能であると判定するような一種の減点方式であったが，適切である，もしくは特に良いと思った部分構造があれば判定対象に良い判定を与えるというような加点方式の設定も考慮していくことが，今後の展望として挙げられる．

参考文献

[1] P. Ertl and A. Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, Vol. 1, pp. 1–11, 2009.

[2] K. Boda, T. Seidel, and J. Gasteiger. Structure and reaction based evaluation of synthetic accessibility. *Journal of Computer-Aided Molecular Design*, Vol. 21, pp. 311–325, 2007.

[3] Y. Baba, T. Isomura, and H. Kashima. Wisdom of crowds for synthetic accessibility evaluation. *Journal of Molecular Graphics and Modelling*, 2018. in press.

[4] M. Venanzi, J. Guiver, Kohli Kazai, G., P., and M. Shokouhi. Community-based bayesian aggregation models for crowdsourcing. In *Proceedings of the the 23rd World Wide Web Conference*, pp. 155–164, 2014.

[5] H. Lakkaraju, J. Leskovec, J. Kleinberg, and S. Mullainathan. A bayesian framework for modeling human evaluations. In *Proceedings of the SIAM International Conference on Data Mining*, pp. 181–189, 2015.

[6] K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, Vol. 96, pp. 1077–1087, 2001.

[7] E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, Vol. 9, pp. 1981–2014, 2008.

[8] V. C. Raykar and S. Yu. Ranking annotators for crowdsourced labeling tasks. In *Advances in Neural Information Processing Systems 24*, pp. 1809–1817. 2011.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

人間ニューラルネットワーク

Crowd Neural Network

坂田 雄亮 *1　馬場 雪乃 *2　鹿島 久嗣 *1*3
Yusuke Sakata　Yukino Baba　Hisashi Kashima

*1京都大学 Kyoto University
*2筑波大学 University of Tsukuba
*3理化学研究所 革新知能統合研究センター RIKEN Center for AIP

Feature values are indispensable for almost all machine learning methods to make predictions; however, the definitions of the features themselves are sometimes too abstract so that they cannot be extracted automatically. Human-in-the-loop machine learning is a promising solution to such cases where humans extract feature values for machine learning models. In this research, we use crowdsourcing for feature extraction and consider a problem to aggregate the feature values given by crowdsourcing workers to improve the accuracy of machine learning classifiers. We propose a novel neural network model called *Crowd Neural Network* that learns both the capabilities of human feature extractors and the weights of a neural network classifier by applying the idea of the convolution neural network to feature aggregation. Our experiments using four datasets show the proposed method outperforms the baseline method using an existing unsupervised aggregation method in some datasets.

1. はじめに

機械学習技術の近年の発展は著しく、様々な分野において機械学習の応用が行われているが、いずれにおいてもまずは学習対象の情報を計算機可読な特徴量によって表現する必要がある。しかし、その特徴量の抽象性が高い場合には、その機械的抽出もまた困難であり、従って機械学習の適用も困難となる。一方で、機械的処理では抽出困難な特徴であっても、人間にとっては比較的容易に抽出可能な特徴量も存在する。このような場合、まずは人間が特徴量を抽出し、これを用いて機械学習を行うといった、いわゆる人間参加型 (human-in-the-loop) のパイプラインが有効であろう。従来はこのようなパイプラインの実現には、人的リソースのコストが大きかったが、近年のクラウドソーシングプラットフォームの台頭により人的リソースの大量かつ機械的な獲得が可能となり、人間を特徴抽出機構と見なし機械学習アルゴリズムに組み込む事が容易となった。

機械と人間の協調が必要となる問題の例としては、絵画の作者を推定する問題や、動画中の人物が作り笑いしているか否かを見分ける問題などが考えられる。これらは機械にとっても専門知識のない人間にとっても比較的困難な問題であろう。一方、絵画が明るいイメージであるかとか、動画内の人物が肩を震わせて笑っているかなどといった、最終的な判別に寄与すると期待されるやや抽象的な特徴は、依然として機械にはその判断が難しい反面、人間ならば比較的簡単に判断できる。このような抽象的な特徴量抽出をクラウドソーシングによって行うことで前述のパイプラインを実現できると期待できる。

しかし、クラウドソーシングにおいては、ワーカの能力ややる気のばらつきがあるため、結果として得られる成果物の品質にばらつきが生じる。そこで頑健性の確保の為に複数のワーカから回答を得て統合を行うことで信頼度の高い結果を得るという手法が広く用いられている。1 つのタスクについて複数のワーカからラベリングが行われている時、最も簡単な意見統合手法としては多数決や平均を取ることが挙げられる。

しかし単純に多数決を取ることにはいくつかの問題がある。あるタスクについて正解と不正解のラベルがある状況を仮定すると正解を見極める能力を持つワーカが少数の場合、単純に多数決を行なってしまうと不正解のラベルが選択されてしまう。そのためワーカの能力を考慮して意見統合を行う手法が広く研究されている。Dawid と Skene によって提案されたワーカの感度と特異度を用いて意見統合を行う手法はこの分野の先駆け的手法である [Dawid 79]。

また別の問題としてスパムワーカの存在がある。スパムワーカとはクラウドソーシングなどにおいて報酬のみを目的にタスクを意図されない形でこなすワーカなどを指し、スパムワーカの回答は大抵の場合実験の意図に則さないデータ、即ちノイズデータとなる。もちろんプラットフォーム側でも対策は行なっているものの完全な排除は難しくデータにスパムワーカによるものが混入することは考慮されなければならない。1 つのタスクに参加するワーカの数が多ければ多数決でもノイズは無視できるが、ワーカの数が少ない場合は結果に大きく影響が出てしまう。この場合もワーカの能力を考慮に入れることでスパムワーカの意見の影響を抑える事が可能であり結果的にノイズの除去が可能となる。

一般に広く知られる意見統合手法は統合したラベルを得る事が目的となるため、意見統合の際に正解を参考にできない一種の教師無し学習と捉える事ができる。一方、統合したラベルを学習に用いる本研究の設定の場合ラベル統合の明示的な正解ラベルは存在しないものの、ラベルを特徴量として使用した分類器の出力に正解が存在するので間接的に教師データを持つと言える。従って一度にワーカの能力と分類器を学習する手法は既存の意見統合手法と分類器の学習を組み合わせたものより優れた精度を実現しうると考えられる。

本研究ではクラウドソーシングを用いた機械学習の一手段として複数のクラウドワーカによってデータの特徴抽出を行った際、ワーカの能力まで学習対象として分類器の学習を行うことでより精度の高い分類器を作成する手法を提案する。ここで意見統合自体はワーカの能力を係数とする重み付き足し合わせで表現される。複数のクラウドワーカより得られる特徴ラベルの重み付き足し合わせを 1 次元の畳み込みを用いて実現し、畳み込みニューラルネットワークを用いて各ワーカの能力と各ノードとの重みを同時に学習する。

連絡先: 坂田　雄亮，京都大学大学院情報学研究科知能情報学専攻，yskskt75@ml.ist.i.kyoto-u.ac.jp

提案手法の分類精度及び適応範囲を確認するため、絵画の作者の識別、ニュース記事の公共性の識別、ホテルのレビュー文の真偽の識別、動画上の人物の作り笑いの識別の 4 つのデータセットに対しての実験を行ない、提案手法が既存手法に優る例を示した。

2. 関連研究

2.1 機械学習におけるクラウドソーシングの活用

クラウドソーシングの台頭によって不特定多数の人間に安価に作業を依頼することが出来る様になり機械学習にもクラウドソーシングを活用する研究が行われてきた。機械学習においてはまず学習に用いる教師データの収集の目的でクラウドソーシングが用いられた [Ipeirotis 10]。しかしクラウドソーシングを用いて得られるデータは品質のばらつきが大きい為、何らかの処理によって品質管理を行う必要がある。品質管理手法の一例として、一つのタスクに関して複数のワーカから結果を収集し統合する事で頑健な結果を得る手法が挙げられる。最も単純な意見の統合は多数決だが、より信頼度の高い結果を得る為に様々な手法が提案されている。

Dawid と Skene が提案したワーカの能力をワーカが正しく回答する確率によってモデル化する手法もその一つである [Dawid 79]。ワーカの能力、正解ラベルの二つが最終的に推定したいものであり、片方が判明していればもう片方が推定できる事を利用して、EM アルゴリズムによってこれらを交互に推定する。推定値の初期値としては仮の正解ラベルとして多数決によるものが使用される事が多い。

2.2 クラウドソーシングによる特徴ラベル生成

あるタスクの答えを得るためにクラウドソーシングを用いる場合、タスクの答えそのものを募集するのが最もシンプルな手法となる。しかしタスクが抽象的でかつある程度の専門性を必要とする課題である時、上記の手法には専門的知識を持つワーカが必要となる。そこで専門的知識を持たないワーカでも答えられる簡単な、しかし機械的に抽出する事が難しいような質問を用意しその回答を特徴ラベルとして教師付き機械学習に用いる手法が提案されている。これは本実験の問題設定に最も近い関連研究である。上述したタスクの例として画像を見て野鳥の種類を特定するタスクが考えられる。この場合、「お腹の色」や「くちばしの形」などの質問によって専門的知識を持たないワーカからも特徴ラベルを得る事ができる [Branson 10]。

2.3 クラウドソーシングによる特徴ラベル定義

機械学習の際にどの様な特徴量を定義するかは予測精度に大きく影響を及ぼすので、技術者は適切な特徴量の設定を行いたいが、しばしば適切な特徴量の設計には専門的知識や多大な労力が必要とされる事がある [Domingos 12]。そこで特徴ラベル生成に用いる質問そのものもクラウドソーシングを用いて作成する手法も提案されている。Flock と呼ばれるこのシステムはクラウドワーカに正例と負例を比べさせてその差異を記述させることで特徴の定義を行う [Cheng 15]。特徴ラベルの収集もまたクラウドソーシングによって行われる。より精度上昇に貢献しうる特徴定義を適当的に獲得する方法として AdaFlock が提案されている [Takahama 18]。AdaFlock ではブースティングの考え方を用いて、現在の分類器で誤分類されているデータをうまく分類出来るような特徴定義を逐次的に得る。

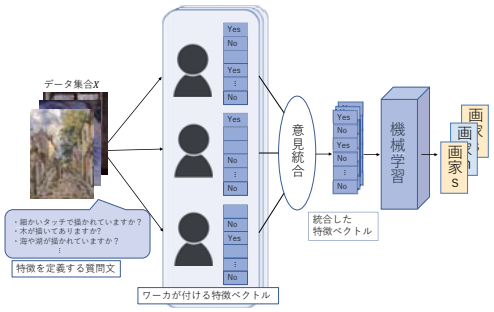


図 1: ラベリングタスクのイメージ図。例として Paintings データセットを用いた。

3. 問題設定

はじめに予測モデルの定式化を行う。本問題では、二値分類問題を基盤とする。X をサンプルの集合、$Y = \{-1, +1\}$ をラベルの集合とする。N 個のラベル付きサンプルからなる訓練データセット $D = \left\{(x_i, y_i) \in X \times Y\right\}_{i=1}^{N}$ を入力として、分類器 $H : X \to Y$ を獲得する事が目的となる。ここで、入力 X に関して特徴ベクトルは与えられておらず、X に対してクラウドワーカから特徴ラベルを収集する必要がある。結果として特徴ベクトルとしては i 番目のデータの k 番目の特徴に j 番目のワーカがつけた特徴ラベル X_{kj}^{i} が得られる。ただし各ワーカは全ての特徴をラベリングする必要はなく、欠損が存在していても構わない。また前提としてテストデータセットの特徴ラベリングに参加したワーカは全て訓練データセットの特徴ラベリングに参加している。即ち訓練データの学習を行うことで全ワーカの能力が学習できる状況を想定している。

なお、特徴は 'Yes'、'No' で答えられる質問にワーカが答える形でラベリングされるものとする。質問文は事前に準備したものを用いる。質問文の作成の詳細は Takahama ら [Takahama 18] に譲る。絵画のデータセットを例にあげると、X としては絵画の画像データの集合しか与えられていない。よって絵画の各画像データに対してクラウドワーカに「空が大きき書かれていますか？」や「輪郭がぼやけていますか？」といった質問により特徴ラベルを生成する。図 1 に特徴ラベリングタスクのイメージ図を示す。

4. Crowd Neural Network

特徴ラベリングタスクを経て各データについて複数のワーカから特徴ラベルが得られた。この複数のワーカからなるラベルを良く統合するためのワーカの能力の推定と、統合されたラベルを用いる分類器の学習を同時に行うニューラルネットワークである Crowd Neural Network を提案する。以下 Crowd Neural Network を Crowd NN と表記する。

4.1 ワーカ畳み込み層

提案手法では分類器の学習と同時にワーカの特徴抽出能力の推定を行う。ワーカの能力の定義にはいくつかの手法が存在するがここでは重み $\alpha = \alpha_1, \ldots, \alpha_J$ によって定義し、データの特徴ベクトルとして各ワーカの付けたラベルを重み付きで足し合わせたものを用いる事を考える。これは確率的モデルではなく、能力の高いワーカの意見を重視する単純な意見統合モデルとなっている。

以上の構造を実装するためにクラウドワーカの回答統合を行うワーカ畳み込み層を提案する。これは一般に用いられる畳

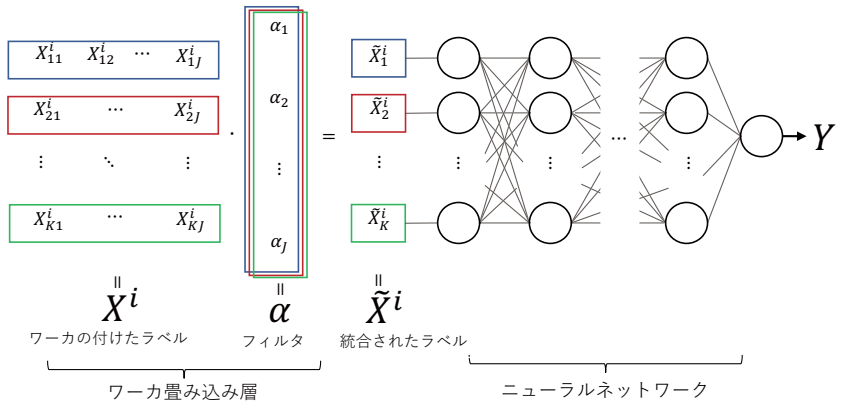


図 2: Crowd NN のイメージ図。同色の行とフィルタ α を掛け合わせて $\tilde{X}^i$ を計算しニューラルネットワークの入力とする。

み込みニューラルネットワークの畳み込み層を応用したものである。以下にワーカ畳み込み層内での処理を述べる。

まず特徴ラベルの重み付き足し合わせを畳み込みによって実現する構造について述べる。畳み込みの 1 ステップはフィルタという重みを共有する概念によってフィルタの範囲内の入力データをフィルタの重みに応じて足し合わせる事である。よって i 番目のデータの k 番目の特徴についてクラウドワーカがつけたラベル $X^i_{k1}, \ldots, X^i_{kJ}$ に対して、ワーカの能力に対応する 1 次元のサイズ J のフィルタとして α を用いて畳み込みを行う事でワーカの能力に応じた重み付き足し合わせを表現でき、統合した特徴ラベル $\tilde{X}^i_k$ を得られる。欠損している特徴ラベル、つまり該当ワーカが回答していないラベルについては 0 で埋められているため、ラベルの統合に影響しない。同様に全ての特徴に畳み込みを適用する事で統合した特徴ベクトルである $\tilde{X}^i$ が得られる。ここでフィルタの重みは全データ、全特徴で共有される事に注意されたい。

4.2 Crowd NN の学習

前述 のワーカ畳み込み層によって統合された特徴ベクトル $\tilde{X}^i$ は後段のニューラルネットワークの入力となる。特徴ベクトル $\tilde{X}^i$ に基づいて分類を行い発生した誤差の逆伝搬によってフィルタ、及び各ノードの重みを学習する。ニューラルネットワークの次元数、層数、及び活性化関数は後述するチューニングを行って決定する。

I をデータの総数、J をワーカの総数、K を特徴の総数とし i, j, k をそれぞれのインデックスとするとワーカ畳み込み層について順伝播の式は以下の様に表せる。

$$\tilde{X}^i_k = \sum_{j=1}^{J} X^i_{kj}\alpha_j$$

ここで得られた $\tilde{X}^i_k$ を上述したニューラルネットワークの順伝搬に適用する。一方逆伝播の式は誤差関数を E で表すと

$$\begin{aligned}\frac{\partial E}{\partial \alpha_j} &= \sum_{k=1}^{K} \frac{\partial E}{\partial \tilde{X}^i_k}\frac{\partial \tilde{X}^i_k}{\partial \alpha_j} \\ &= \sum_{k=1}^{K} \frac{\partial E}{\partial \tilde{X}^i_k} X^i_{kj}\end{aligned}$$

以上の様に表す事が出来る。ここで $X^i_{kj} = 0$ の時、すなわちデータ i の特徴ラベリングタスク k についてワーカが応答をしていない時はデータ i の特徴 k の誤差にワーカの能力が影響されない事がわかる。またバックプロパゲーションの誤差

$$\delta^i_k = \frac{\partial E}{\partial \tilde{X}^i_k}$$

はニューラルネットワークの一層目のレイヤーより逆伝播されている為、上式によってパラメータを更新する事が出来る。最後に Crowd NN のイメージ図を図 2 に示す。

4.3 特徴ラベルの取り扱い

特徴ラベル X^i_{kj} は欠損を含んでいても良く、欠損値は 0 で穴埋めされる。ここで質問文の返答として本質的意味合いを持つのはポジティブな反応である 'Yes' でありネガティブな反応である 'No' は大きな影響を持たないと考える事も出来る。その為ワーカの返答の入力ラベルの符号化を $\{1, 0, -1\}$ と $\{1, 0\}$ の二種類考える事とする。$\{1, 0\}$ の符号化を用いる場合、欠損データも 0 で埋めているのでネガティブな反応である 'No' は応答しなかった事と同義に扱われる事となる。

5. 評価実験

5.1 特徴ラベルデータセット

4 種類の特徴抽出タスクを対象にして、それぞれ 1 つの特徴量に対して特徴ラベリングを行うワーカ数を 10 人とし特徴ラベルを収集した。これらの特徴抽出タスク（使用する画像等のデータ、質問文）は、Takahama ら [Takahama 18] が設計したものと同じである。4 種類のタスクは、絵画の著者識別の Paintings、動画の人物の自然な笑顔と作り笑いの識別の Smiles 、ホテルに関する偽造レビューの識別の Reviews、記事がトップページに乗るほど重要か否かを識別する Articles である 。全てのデータセットは共通して正例、負例共に 200 件ずつ合計 400 件。各タスクごとの特徴量は 100 種とした。

5.2 ベースライン

提案手法と比較するベースラインとして以下の 2 つの意見統合手法を用いる：

1. 多数決: データ毎の特徴量をワーカの付けたラベルの単純な多数決によって得たもの。

2. Dawid&Skene (D&S): ワーカの感度と特異度ならびに特徴ラベルを EM アルゴリズムにより推定し特徴量とした

表 1: テストデータに対する精度の比較表。10 回の交差検証による精度の平均値と標準偏差を示す。また各データセットで最も精度が良かったものを太字で示す。

	Accuracy			
Dataset	Crowd NN	Pos. Crowd NN	多数決	D&S
Paintings	0.765 ±0.051	**0.773** ±0.039	0.732 ±0.065	0.762 ±0.066
Smiles	0.785 ±0.081	0.765 ±0.060	0.617 ±0.101	**0.787** ±0.073
Reviews	0.645 ±0.072	0.650 ±0.074	0.520 ±0, 064	**0.700** ±0.064
Articles	0.645 ±0.045	**0.710** ±0.087	0.550 ±0.031	0.660 ±0.055

もの。条件を揃えるために訓練データのみを用いてワーカの能力を推定する。

特徴ラベルの正解は存在しないため、統合したラベルを入力としたニューラルネットワークの精度で比較する。ニューラルネットワークのハイパーパラメータは提案手法と同様のチューニングを行う。

5.3 評価方法

各モデルの評価はハイパーパラメータのチューニングも含めた 10 分割交差検証を用いて行った。softmax の値に基づく精度の平均と標準偏差を評価値とした。ハイパーパラメータ及びその探索範囲として、各層のサイズを 100 から 300 まで 50 刻み、層数を 2 から 7 まで 1 刻み、活性化関数を *relu* または *leaky_relu* から選択、エポック数を 20 から 100 まで 10 刻み、バッチサイズを 20 から 200 まで 20 刻みとした。

5.4 実験結果

結果を表 1 に示す。なお、以降 $\{1, 0, -1\}$ の符号化を用いた提案法を Crowd NN、$\{1, 0\}$ の符号化を用いた提案法を Positive Crowd NN と呼ぶ。全てのデータセットを通じて Crowd NN は多数決より優れた精度となった。一方、D&S と比べて Painting データセットでは僅かに上回ったが Smiles データセットでは僅かに下回り、Articles データセット、Reviews データセットでは大きく下回る結果となった。この結果から Crowd NN は単純な多数決より良く意見統合が出来ている事が確認出来る。一方、D&S とは比較して必ずしも常に高い精度ではない結果となった。Positive Crowd NN は全てのデータセットを通じて多数決より優れた結果となった。Paintings、Articles データセットにおいては Crowd NN、D&S を上回る結果となった。Reviews データセットについて Crowd NN を僅かに上回ったが Dawid&Skene を下回り、Smiles データセットにおいて Crowd NN、D&S の両方に劣る結果となった。

Crowd NN が必ずしも Dawid&Skene よりも高い精度を得られなかった理由について考察する。まず挙げられるのは過学習の可能性であろう。今回の実験で用いた 4 つのデータセットはそれぞれ 400 件程度と比較的小規模である。既存手法に比べて、Crowd NN が学習するべきパラメータの数は多く、従って過学習が起きている可能性が指摘できる。

$\{1, 0\}$ の符号化を用いると $\{1, 0, -1\}$ の符号化を用いた場合に比べて精度が上がったデータセットでは、不要な特徴を省く事で学習が容易となり結果として過学習の軽減に繋がったと考えられる。ただしネガティブなラベルである 'No' が比較的価値をもつデータセットにおいてはこの情報が失われるため精度が下がったものと思われる。

6. 結論

本研究ではクラウドソーシングを利用して生成された特徴ラベルを用いた機械学習の精度向上を目指して、特に特徴ラベルの意見統合に焦点を当てた手法としてクラウドワーカの能力をフィルターに用いた畳み込みニューラルネットワークを提案した。この提案手法の性能を測るため 4 つのデータセットを用いて実験を行い提案手法の長所と短所を確認した。

今後の研究の展望としては予測精度への貢献性の低い特徴量の削減が考えられる。貢献性の低い特徴量は分類器の学習自体にはそもそも大きく影響することはなく、ワーカの能力推定に対してはノイズになっていると考えられる。通常の意見統合手法の場合は統合ラベルが目的であるが、統合ラベルを特徴量に機械学習を行う今回の設定の場合は統合ラベルは目的でないため精度予測への貢献性が低いラベルを無視する事が出来る。

データ数を増やす事で過学習を回避する事も考えられる。特徴生成にクラウドソーシングを用いているためコストとのトレードオフとなってしまうが、データ数の増加によって提案手法のより正確な性能を確認出来る事が期待される。

今回の実験では訓練データから特徴生成に参加した全ワーカの能力を推定する事が出来るという前提としているが、実際にはテストデータの特徴生成に新規ワーカが参加する可能性は無視できない。その様な事前に能力推定が出来ないワーカの能力の取り扱いについても検討の余地がある。

参考文献

[Branson 10] Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., and Belongie, S.: Visual Recognition with Humans in the Loop, in *Proceedings of the 11th European Conference on Computer Vision*, pp. 438–451 (2010)

[Cheng 15] Cheng, J. and Bernstein, M. S.: Flock: Hybrid Crowd-Machine Learning Classifiers, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 600–611 (2015)

[Dawid 79] Dawid, A. P. and Skene, A. M.: Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, No. 1, pp. 20–28 (1979)

[Domingos 12] Domingos, P.: A Few Useful Things to Know About Machine Learning, *Communications of the ACM*, Vol. 55, No. 10, pp. 78–87 (2012)

[Ipeirotis 10] Ipeirotis, P. G.: Analyzing the Amazon Mechanical Turk Marketplace, *ACM XRDS*, Vol. 17, No. 2, pp. 16–21 (2010)

[Takahama 18] Takahama, R., Baba, Y., Shimizu, N., Fujita, S., and Kashima, H.: AdaFlock: Adaptive Feature Discovery for Human-in-the-loop Predictive Modeling, in *Proceedings of the 32nd Conference on Artificial Intelligence* (2018)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

LOD Surfer API: クラス間関係に基づく LOD 探索のためのウェブ API

LOD Surfer API: A web API for LOD exploration based on class-class relationships

山口敦子[*1] 小林紀郎[*2] 桝屋啓志[*3] 山本泰智[*1] 古崎晃司[*4]
Atsuko Yamaguchi Norio Kobayashi Hiroshi Masuya Yasunori Yamamoto Kouji Kozaki

[*1] 情報・システム研究機構 ライフサイエンス統合データベースセンター
Database Center for Life Science, Research Organization of Information and Systems

[*2] 理化学研究所 情報基盤センター
Advanced Center for Computing and Communication, RIKEN

[*3] 理化学研究所 バイオリソースセンター
BioResource Center, RIKEN

[*4] 大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

In life-sciences domain, Linked Open Data (LOD) is being increasingly used when publishing databases. For flexible use of such databases, we employ a method that uses federated query search along a path of class-class relationships. To demonstrate our strategy, we implemented a prototype system accessible via a web API as a preliminary trial. We have been collecting SPARQL Builder Metadata (SBM) that describe a data schema for each SPARQL endpoint. Using the SBM, we constructed a graph called a merged class graph based on the class-class relationships in LOD. Using this graph, our system can provide the information required to construct a federated search query, such as the SPARQL endpoints that include a class–class relationship, and paths of the class-class relationships between two classes.

1. はじめに

日々生産される生命科学のデータは巨大かつ多岐にわたる．実験によって得たデータを解釈し，生物学的に新たな知識を得るためには，これらのデータ全体を統合し，自在に組合わせることが必須となってきている．このような状況のもと，数多くの研究施設が生産した生命科学データの RDF 化をすすめ，Linked Open Data (LOD)として公開してきている．LOD 上でデータベースを公開する際には，RDF の標準的な検索言語である SPARQL を用いた検索が可能なウェブ API である，SPARQL エンドポイントが提供されることが多い．その結果，LOD 上の SPARQL エンドポイントは次々と新たに提供されており，ユーザがその全体像を理解するのは困難である．

一方，著者らはセマンティックウェブ技術にあまり詳しくないユーザを対象に，SPARQL クエリ生成支援システム SPARQL Builder を開発し，2013 年 10 月からサービス運用を行ってきた[1,2]．その際，高速にクラス間関係提示を行うため，予め LOD 上の SPARQL エンドポイントをクロールし，SPARQL エンドポイントで提供されるデータのデータスキーマを記述するメタデータを取得し蓄積してきた．

そこで，これらのメタデータを利用し，LOD 全体の構造，および，どのエンドポイントから必要な情報が得られるかをユーザが把握する必要がなく，柔軟に LOD からデータを切り出すことを可能とする情報基盤 LOD Surfer を提案する．本発表では特に，LOD Surfer を実現するにあたり，必要な機能と課題を検討するために，プロトタイプ実装を行ったウェブ API，LOD Surfer API について紹介する．

連絡先：山口敦子，情報・システム研究機構 ライフサイエンス統合データベースセンター，〒277-0871 千葉県柏市若柴 178-4-4-6F，atsuko@dbcls.rois.ac.jp

2. SPARQL Builder Metadata

2.1 SPARQL Builder

SPARQL Builder とは，SPARQL 言語の知識がなくとも，また対象データセットの構造を知らなくても，対話的な GUI を介して SPARQL クエリを生成することができることを目指して開発されたウェブ上のサービスである[4].

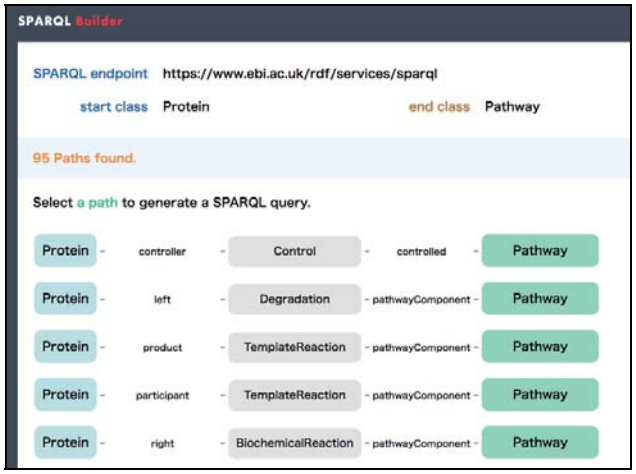


図 1 SPARQL Builder

SPARQL Builder はインスタンス間がどのようなプロパティでつながっているかというクラス間の関係をベースとして SPARQL クエリの記述補助を行う．ユーザはまず，入力クラスと出力クラスをそれぞれクラスのリストから選ぶ．入出力の二つのクラスが確定すると，それらのクラス間の関係でデータ内に含まれるものが

ユーザに提示される. ここで, ユーザが望むクラス間の関係は, データ内ではどのプロパティを用いて記述されているか分からないことに加え, 複数のトリプルを用いた多段の関係により記述されている可能性もある. そこで, ユーザの欲しい関係を提示できるよう, 入出力クラス間の多段の関係, つまりクラス間関係のパスを提示する. 現在の SPARQL Builder では, 最大 4 段のクラス間関係パスをすべて提示する. ユーザがパスを一つ選ぶと, そのパスのクラス間関係に対応する SPARQL クエリが自動生成される. 生成された SPARQL クエリはそのまま検索へ利用することができ, また, 生成された SPARQL クエリを編集して利用することも可能なシステムとなっている.

2.2 SPARQL Builder Metadata の概要

ユーザにデータセットやクラスの情報を提示するために必要なデータ, および, パスの計算を高速に行うために必要なデータは, 事前に SPARQL エンドポイントから取得し蓄積しておくことが望ましい. この目的で設計されたメタデータが SPARQL Builder Metadata (SBM)である. SBM はそれ自身が RDF で記述される.

SBM のスキーマ仕様は, データセットのメタデータ記述語彙 VoID[1] および, SPARQL エンドポイントのウェブ上のサービスとしてのメタデータ記述語彙 SPARQL 1.1 Service Description[2] の拡張として定義されている. 大まかには, SPARQL エンドポイント→エンドポイントに含まれるデータセット→データセットのメタデータの階層構造になっており, 各メタデータ部分にはクラスリスト, プロパティリスト, クラス-プロパティ-クラス関係, さらにそれらに関連するトリプル数等の統計情報が含まれる. 詳しくは http://www.sparqlbuilder.org/doc/sbm_2015sep/ を参照されたい.

SPARQL Builder を運用するにあたり, 現在, 生命科学分野の 43 の SPARQL エンドポイントから 76 のデータセットのメタデータを取得し蓄積している. メタデータ取得は主に, SPARQL エンドポイントに対し, 一連の SPARQL クエリを投げ, その結果から SBM を構成する. SBM 取得プログラムは GitHub にて公開されており, https://github.com/sparqlbuilder/metadata から入手可能となっている.

3. LOD Surfer API

3.1 LOD Surfer 概要

L OD Surfer は, LOD から自分のデータに紐付いたデータの中から, 欲しいデータがどの SPARQL エンドポイントに含まれるか気にすることなく, 興味がある部分を柔軟に切り出すことを可能とすることを目指して設計されたシステム基盤である. 対象データはウェブ上に分散しているため, LOD Surfer を実現するためには, データの所在を自動的に見つけ, さらにそのデータを適切かつ現実的な時間で取得する必要がある.

まず, データが所在する SPARQL エンドポイントを見つけるためには, 先述の SBM が利用可能である. SPARQL エンドポイントから必要なデータを適切かつ現実的な時間で取得することを担保するためには, Umaka-YummyData[3]を利用する. Umaka-YummyData は, 生命科学データを中心とした SPARQL エンドポイントのリストに対し, 毎日アクセスを行い, サービスの死活, 更新頻度, VoID 等のメタデータ提供状況, クエリ返答速度などを監視し, その結果をスコア付けするシステムである. このシステムのウェブ API を利用し, 稼働率が低い SPARQL エンドポイントやクエリ返答速度が遅い SPARQL エンドポイントをフィルタすることで, 検索先のサービスが返答せず, データが取得できないことを回避する.

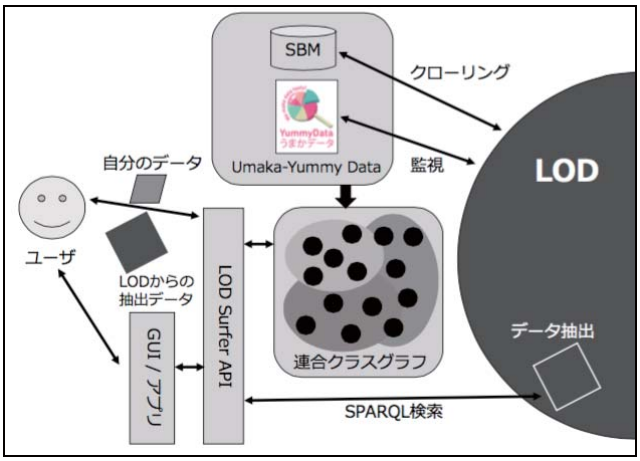


図 2 LOD Surfer システム構成

図 2 は LOD Surfer のシステム構成設計の概要である. Umaka-Yummy Data によりフィルタされクエリ応答が担保された SPARQL エンドポイントのリストに対し, 対応する SBM 形式のメタデータを利用し, LOD 全体のクラス間関係を示すグラフ, 連合クラスグラフを構築する. ユーザが LOD Surfer API を通じて自分のデータを入力すると, そのデータのクラスを始点としたクラス間関係パスを, 連合クラスグラフを用いて計算し, 到達可能なクラスのリストを示す. それらのクラスのリストから, ユーザが興味をもったクラスを指定すると, クラス間パスからフェデレート検索クエリを生成し, クエリを LOD 上で実行し, その結果をユーザへ返すことができる.

計算機に詳しくなく, LOD Surfer API を直接利用することが難しいユーザに対しては GUI やアプリケーションソフトウェアを通じて API を利用するように, 今後開発を進める予定である.

3.2 連合クラスグラフと連結成分

LOD をクロールした結果, SBM には LOD 内のクラスおよびクラス間関係情報が蓄積される. そこで, 蓄積した全ての SBM に対し, 一度でも現れるクラスを頂点にし, クラス間関係を辺としたグラフを考える. このグラフを連合グラフという.

LOD Surfer におけるフェデレート検索クエリ構築もクラス間関係をベースとするため, 効果的なフェデレート検索の設計のためには連合クラスグラフの構造の解析が必要である. そこで, 現在 SPARQL Builder で利用中の 43 の SBM を利用し, 連合クラスグラフを構築し, その解析を試みた. 特に, クラス間の到達可能を考慮するために連結成分を中心に解析を行った. ちなみに, グラフの連結成分とは, グラフを互いにパスを持つ極大頂点集合に分割したものである. 言い換えるならば, 同じ連結成分内のクラス間には何らかのパスがあるといえる.

SBM から構築した連合グラフは, 頂点数 14,147, 辺 18,738 のグラフとなった. このグラフの連結成分を計算し, その頂点数(=クラス数)の分布を調べた. 連結成分の数は 6,857 個であり, そのうち, 6,814 個が孤立点であった. 一方, 最大の連結成分は 5,327 個のクラスを含み, 孤立しない 7,333 個のクラスのかなりの部分を最大の連結成分が占めることがわかる.

[1] https://www.w3.org/TR/void/
[2] https://www.w3.org/TR/sparql11-service-description/

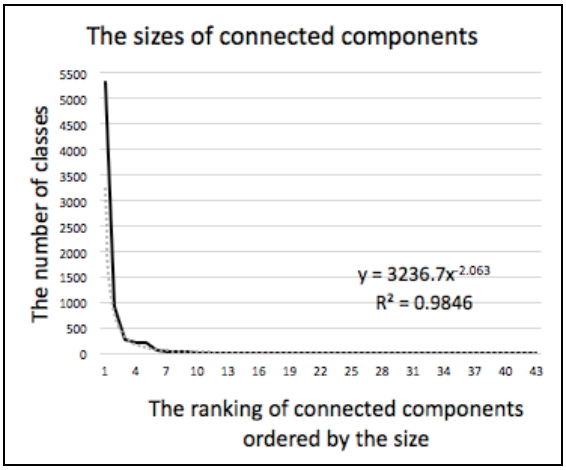


図 3 連合グラフの連結成分数の分布

図 3 は，孤立点でない連結成分について，連結成分の頂点数をサイズが大きいものから順に示したものである．上下二つのグラフは Y 軸に関し，上は線形軸，下は対数軸となっている．分布は冪乗則に従うように見え，事実，最小二乗法でフィッティングすると $y=3236.7x^{-2.063}$ に対し，その決定係数 R2 は 0.9846 と十分 1 に近い値を取る．

このことから，(1)連合クラスグラフは連結成分を予め計算し，連結成分ごとに別に管理することで，探索空間を大幅に減らすことができる，(2)孤立点となるクラスが多いので，上位クラスでまとめる工夫が必要である，ということができる．

3.3 API 実装

LOD Surfer API はユーザがクラス間関係を利用して LOD から効率良くかつ柔軟にデータを切り出すためのウェブ API である．この API を利用して，対象となる SPARQL エンドポイント一覧，クラス一覧，あるクラスから到達可能なクラスの一覧，二つのクラスに対し，連合クラスグラフ上のクラス間パスのリスト，クラス間パスに対し，SERVICE 句付きのフェデレート検索クエリを出力することができる．

これらの機能を実現するウェブ API を下記のように設計した．

- /eplist?class1={class1 URI}
  class1 をデータ内に含む SPARQL エンドポイントのリストが JSON 形式で出力される．class1 が指定されない場合，全ての対象 SPARQL エンドポイントのリストが出力される．
- /clist?class1={class1 URI}
  class1 をデータ内に含む SPARQL エンドポイントのリストが JSON 形式で出力される．class1 が指定されない場合，全ての対象 SPARQL エンドポイントのリストが出力される．
- /path?class1={class1}&class2={class2}
  class1 から class2 へのパスのリストが JSON 形式で出力される．
- /sparql?path={path}
  path に対応する SERVICE 句付き SPARQL クエリがテキスト形式で出力される．

これらのウェブ API は，先述の連合クラスグラフおよびその連結成分を用いて高速に実行できる．例えば，到達可能性は連結成分の頂点のリストと一致するため，これを用いて瞬時に計算できる．パスも連結成分内のみに探索空間が限定されるため，現実的な時間で計算できる．

LOD Surfer API の有効性を確認するため，現在 SPARQL Builder で利用中の 43 の SBM に対し，プロトタイプを作成した．実装にあたり，このウェブ API の諸機能のための計算は直接 RDF 形式の SBM を利用するわけではなく，パス探索や連結成分計算などのグラフアルゴリズムが適用しやすいように，一旦，連合クラスグラフを表す隣接リスト構造へ変換したものを利用した．さらに，パス探索の効率性を上げるために，予め連結成分を計算したうえで，成分ごとに頂点のハッシュ集合で保持した．これらの構造上で，上記の機能を持つウェブ API を作成した．

LOD Surfer API をプロトタイプ実装したサービスは http://lodsurfer.dbcls.jp/api/で試すことができる．例えば，全クラスのリストは http://lodsurfer.dbcls.jp/api/clist で出力することができる．また，LOD Surfer API において使われているシステムのコードは https://github.com/LODSurfer/lodsurfer-api から取得可能である．

4. まとめと今後の課題

クラス間関係をベースとした一つのグラフを構築することで，LOD 上のデータを柔軟に切り取る手法を提案した．また，この手法のプロトタイプ実装として，SPARQL Builder 運用において収集した SBM を用い，連合クラスグラフを構築して，連合クラスグラフ上でパス探索などの諸機能を実現し，ウェブ API を開発した．

今後の課題は，現在の枠組みで検索対象にならない，あるいはなりにくいデータへの対応である．連合クラスグラフの解析結果より，インスタンスが少ないクラスや孤立したクラスが数多くあることがわかった．これらのクラスに対しては，より上位クラスで下位クラスをまとめたうえでクラス間関係を検索することが望ましいと思われる．しかしながら，現時点での SBM は rdfs:subClassOf への対応が十分ではなく，SBM の情報だけでは，下位クラスを適切に上位にまとめることが難しい．そのため，今後，SBM を LOD Surfer 向けに設計を直す必要があり，また，そのためのクローラシステムの開発も必須である．

また，今回の LOD Surfer API のプロトタイプ実装では，最終的にフェデレート検索クエリの出力で留まり，クエリの処理は他の SPARQL エンドポイントへ任せる設計になっている．今後，SBM 内の統計情報などを利用することにより，データの結合を効率的に行える順序を計算することができると思われる．その手法を使ったクラス間関係ベースのフェデレート検索システムの提案も考えたい．

謝辞

本研究のは独立行政法人科学技術振興機構(JST)、バイオサイエンスデータベースセンター (NBDC) の助成，および科学研究費補助金基盤(C) 17K00434，17K00424，基盤(B) 17H01789 の助成による．

参考文献

[Yamaguchi 2014] Yamaguchi, A., Kozaki, K., Lenz, K., Wu, H., Kobayashi N.: An Intelligent SPARQL Query Builder for Exploration of Various Life-science Databases, The 3rd International Workshop on Intelligent Exploration of Semantic Data (IESD 2014), CEUR Workshop Proceedings 1279, 2014

[Yamaguchi 2016] Yamaguchi, A., Kozaki, K., Lenz, K., Yamamoto, Y., Masuya, H., Kobayashi N.: Semantic Data Acquisition by Traversing Class-Class Relationships Over Linked Open Data, The 6th Joint International Semantic

Technology Conference (JIST 2016), LNCS 10055, 136-151, 2016

[Yamamoto 2018] Yamamoto, Y., Yamaguchi, A., Splendiani, A.: YummyData: providing high-quality open life science data. Database, doi: 10.1093/database/bay022, 2018

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

SPARQL クエリ編集者と編集補助者との編集支援機構の動的オントロジーマッピング手法の検討

A Preliminary Approach on a Cooperative Query Editing System using Dynamic Ontology Mappings

足立拓也 *1　福田直樹 *2
Takuya Adachi　Naoki Fukuta

*1静岡大学大学院
Graduate School of Integrated Science and Technology, Shizuoka University

*2静岡大学学術院情報学領域
College of Informatics, Academic Institute, Shizuoka University

There are several approaches to support the coding process of a SPARQL query for users who are unfamiliar with the used ontologies and the stored data. We are implementing a cooperative query editing system using dynamic ontology mappings. In this paper, we present how dynamic ontology mapping mechanism will be able to target anonymous helpers for support, that uses endpoint different from that used by SPARQL query editors.

1. はじめに

Linked Open Data(LOD) の公開や利用が拡大しており，2018 年 3 月 16 日現在少なくとも 216 個の SPARQL エンドポイントが利用可能である *1．LOD は Resource Description Framework(RDF)*2 や Web Ontology Language(OWL)*3 で記述することができ，W3C 勧告されているクエリ言語 SPARQL*4 で SPARQL エンドポイントから RDF 形式で記述されたデータを検索することができる．

SQL や XQuery などの構造化されたクエリ言語は表現力がある一方，クエリ言語や構文，ドメインスキーマに関する技術的なスキルや知識が必要である [Soylu 16]．SPARQL のクエリ言語や構文，ドメインスキーマに関する技術的なスキルや知識が十分ではない場合を考慮して，様々なアプローチが提案されている．具体的には，クエリベースの検索 [Ferré 14] やセマンティックファセット検索 [Arenas 14]，ビジュアルクエリシステム [Soylu 16]，ユーザーインターフェース [Ermilov 17] といった GUI を用いた SPARQL エンドポイント検索や，重み付きオントロジーマッピング [Atencia 12] を用いた拡張 SPARQL クエリ文法 [Fujino 13] といったオントロジーマッチング手法 [Euzenat 13] によって得られたオントロジーマッピングを用いた SPARQL クエリ変換手法が提案されている．

我々はクラウドベースの SPARQL クエリ実行および編集支援システムとして，MCHA SPAIDA の実装を進めている [足立 18]．本システム上で SPARQL クエリ編集者が編集したクエリを実行し，結果を確認する過程を収集していくことで，将来における SPARQL クエリ作成のノウハウを蓄積するための環境を提供することを目的としている．SPARQL クエリ編集者が SPARQL に関する技術的なスキルや知識が十分でなかった場合，SPARQL クエリ編集者と編集補助者をマッチングし，SPARQL クエリ編集者が記述したクエリを編集補助者が編集支援する仕組みの用意を進めている．

本研究では，SPARQL クエリの編集に他者からの支援を受けるシナリオにおいて，SPARQL クエリ編集者が用いるエンドポイントとは異なるエンドポイントの編集補助者を支援の対象にできるようなオントロジーマッピング手法を検討する．

連絡先: 足立拓也，静岡大学，静岡県浜松市中区城北 3-5-1, adachi.takuya.17@shizuoka.ac.jp

∗1 http://sparqles.ai.wu.ac.at/availability
∗2 https://www.w3.org/RDF/
∗3 https://www.w3.org/OWL/
∗4 https://www.w3.org/TR/rdf-sparql-query/

2. MCHA SPAIDA

我々はクラウドベースの SPARQL クエリ実行および編集支援システムとして，MCHA SPAIDA の実装を進めている [足立 18]．本システム上で SPARQL クエリ編集者が編集したクエリを実行し，結果を確認する過程を収集していくことで，将来における SPARQL クエリ作成のノウハウを蓄積するための環境を提供することを目的としている．

SPARQL クエリ編集者が SPARQL に関する技術的なスキルや知識が十分でなかった場合，SPARQL クエリ編集者と編集補助者をマッチングし，SPARQL クエリ編集者が記述したクエリを編集補助者が編集支援する仕組みの用意を進めている．

SPARQL クエリ編集者が SPARQL クエリの編集に他者からの支援を受けたい場合，SPARQL クエリ編集者が用いるエンドポイントとは異なるエンドポイントの編集補助者を支援の対象になるようなケースが考えられる．

例えば，SPARQL クエリ編集者が対象エンドポイントを公開したくない場合である．SPARQL クエリの編集に他者からの支援を受けるシナリオにおいて，編集補助者は SPARQL クエリ編集者が公開している情報や編集支援を依頼したクエリを確認することで，SPARQL クエリ編集者が取得したクエリの対象や SPARQL クエリ編集者自身，その所属組織における機密情報を推測することができてしまう恐れがある [足立 18]．また，SPARQL に関する技術的なスキルに対しての編集支援を対象にしている場合である．具体的には，集約（COUNT や SUM など）や横断的な検索（Federated Query）など SPARQL1.1*5 で追加された構文を用いる SPARQL クエリの編集を支援してもらう場合である．編集補助者が SPARQL クエリ編集者が用いるエンドポイントに十分な知識を有しているが，複雑なクエリを記述したことがなかった場合，SPARQL クエリに関する技術的なスキルに対しての支援が受けられないことが考えられる．

∗5 https://www.w3.org/TR/sparql11-query/

3. 動的オントロジーマッチング手法の検討

SPARQL クエリ自身を編集補助者からの編集補助を受ける目的に損なわない範囲で匿名化する際，クエリ対象が匿名化されていること，匿名化する SPARQL クエリが編集者の対象エンドポイント上で実行可能であれば匿名化したクエリも編集補助者の対象エンドポイント上で実行可能であることの 2 つの制約が満たされる必要がある．これらの制約を満たすようなクエリに変換することができる動的オントロジーマッピング手法を検討することが求められる．

また，動的オントロジーマッピング手法は SPARQL クエリを実行することによって，クエリを記述したエンドポイントから選択したマッチング手法に必要な情報を取得し，クエリ対象のエンドポイントからそれらの情報を用いてマッチング候補を取得する仕組みである．動的オントロジーマッピング手法を多くの回数実行してしまうと，SPARQL エンドポイントに負荷がかかってしまう．実行に負荷のかかる SPARQL クエリを一度に大量処理することになった場合，SPARQL エンドポイントがサーバーダウンを起こすことがある [Yamagata 15]．そのため，SPARQL エンドポイントに負荷がかからないように，動的オントロジーマッピング手法の実行回数を限られた回数に抑えることが求められる．

さらに，SPARQL クエリ編集者が用いるのとは異なるエンドポイントの編集補助者を支援の対象にできるようにするためには，異なる 2 つのオントロジーの対応関係を導く一般的なオントロジーマッピング [Noy 09] が必要になる．このようなオントロジーマッピングを生成するオントロジーマッチング手法は様々あり [David 11][Jiménez-Ruiz 11][Bella 17]，どのオントロジーマッチャーが SPARQL クエリの匿名化に適しているのか判断することは難しい．

このような複数の選択肢がある中で，限られた回数で目的に合った最適な組み合わせを見つけることが必要である．本研究では，組み合わせ最適化手法の一つである遺伝的アルゴリズム [Eiben 15] を用いて SPARQL クエリ編集者が用いるエンドポイントとは異なるエンドポイントの編集補助者を支援の対象にできるようなオントロジーマッピング手法を検討する．

4. おわりに

本研究では，SPARQL クエリ編集者が用いるエンドポイントとは異なるエンドポイントの編集補助者を支援の対象にできるようなオントロジーマッピング手法を検討した．

今後の課題として，Federated Query を利用した複数エンドポイントを介した SPARQL 検索における SPARQL 編集者と編集補助者とのマッチング手法の検討，SPARQLoid[Fujino 13] や Wikidata などで用いられている拡張 SPARQL クエリを本機構に適用することが挙げられる．

参考文献

[Arenas 14] Arenas, M., Grau, B. C., Kharlamov, E., Marciuška, Š., Zheleznyakov, D., and Jiménez-Ruiz, E.: SemFacet: Semantic Faceted Search over Yago, in *Proc. of the 23rd International Conference on World Wide Web (WWW2014)*, pp. 123–126 (2014)

[Atencia 12] Atencia, M., Borgida, A., Euzenat, J., Ghidini, C., and Serafini, L.: A Formal Semantics for Weighted Ontology Mappings, in *Proc. of the 11th International Semantic Web Conference (ISWC2012)*, pp. 17–33 (2012)

[Bella 17] Bella, G., Giunchiglia, F., and McNeill, F.: Language and Domain Aware Lightweight Ontology Matching, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 43, No. 1, pp. 1–17 (2017)

[David 11] David, J., Euzenat, J., Scharffe, F., and Santos, dos C. T.: The Alignment API 4.0, in *Semantic Web Journal 2(1)*, pp. 3–10, IOS Press (2011)

[Eiben 15] Eiben, A. E. and Smith, J. E.: *Introduction to Evolutional Computing*, Springer-Verlag Berlin Heidelberg, 2 edition (2015)

[Ermilov 17] Ermilov, T., Moussallem, D., Usbeck, R., and Ngomo, A.-C. N.: GENESIS A Generic RDF Data Access Interface, pp. 125–131 (2017)

[Euzenat 13] Euzenat, J. and Shvaiko, P.: *Ontology Matching*, Springer-Verlag Berlin Heidelberg, 2 edition (2013)

[Ferré 14] Ferré, S.: Sparklis: a SPARQL Endpoint Explorer for Expressive Question Answering, in *Proc. of the 13th International Semantic Web Conference (Posters & Demos) (ISWC2014)* (2014)

[Fujino 13] Fujino, T. and Fukuta, N.: Utilizing Weighted Ontology Mappings on Federated SPARQL Querying, in *Proc. of the 3rd Joint International Semantic Technology Conference (JIST2013)* (2013)

[Jiménez-Ruiz 11] Jiménez-Ruiz, E. and Grau, B. C.: LogMap: Logic-based and Scalable Ontology Matching, in *Proc. of the 10th International Semantic Web Confernece (ISWC2011)*, pp. 273–288 (2011)

[Noy 09] Noy, N. F.: Ontology Mapping, in Staab, S. and Studer, R. eds., *Handbook on Ontologies*, pp. 573–590, Springer-Verlag Berlin Heidelberg (2009)

[Soylu 16] Soylu, A., Giese, M., Jimenez-Ruiz, E., Vega-Gorgojo, G., and Horrocks, I.: Experiencing OptiqueVQS: a multi-paradigm and ontology-based visual query system for end users, *Universal Access in the Information Society*, Vol. 15, No. 1, pp. 129–152 (2016)

[Yamagata 15] Yamagata, Y. and Fukuta, N.: An Approach to Dynamic Query Classification and Approximation on an Inference-enabled SPARQL Endpoint, *Journal of Information Processing*, Vol. 23, No. 6, pp. 759–766 (2015)

[足立 18] 足立 拓也, 福田 直樹: SPARQL クエリ編集者と編集補助者との匿名化マッチングおよび編集支援機構のオントロジーマッピング手法を用いた試作, 第 44 回セマンティックウェブとオントロジー研究会 (SIG-SWO-044) (2018)

[2Z3] [General Session] 5. Web Intelligence

座長:大向 一輝（国立情報学研究所）

Wed. Jun 6, 2018 3:20 PM - 4:40 PM Room Z (3F Matsu Take)

[2Z3-01] Quantification of Diverse Personal Attributes in Tweets

○take yo[1], Ayahito Saji[1], Kazutoshi Sasahara[1] （1. nagoya university）

3:20 PM - 3:40 PM

[2Z3-02] Analysis of Minor Women's Online Risk

○Yuichi Hirano[1], Fujio Toriumi[1], Masanori Takano[2], Kazuya Wada[2], Ichiro Fukuda[2] （1. The University of Tokyo, 2. CyberAgent, Inc.）

3:40 PM - 4:00 PM

[2Z3-03] Preliminary Investigation for Analysis of Browsing News Pages on Smartphone App and Wikipedia

○Mitsuo Yoshida[1], Yoshifumi Seki[2] （1. Toyohashi University of Technology, 2. Gunosy Inc.）

4:00 PM - 4:20 PM

[2Z3-04] Comparison of Classification Methods for Ascii Art

○Kazuyuki Matsumoto[1], Akira Fujisawa[2], Minoru Yoshida[1], Kenji Kita[1] （1. Tokushima University, 2. Aomori University）

4:20 PM - 4:40 PM

©The Japanese Society for Artificial Intelligence

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ツイートが運ぶ多様な個人属性の定量化

Quantification of Diverse Personal Attributes in Tweets

余 岳 *1 Take Yo　佐治 礼仁 *2 Ayahito Saji　笹原 和俊 *1*3 Kazutoshi Sasahara

*1名古屋大学大学院情報学研究科 Graduate School of Informatics, Nagoya University
*2名古屋大学情報学部 School of Informatics, Nagoya University
*2名古屋大学大学院情報学研究科, JST さきがけ Graduate School of Informatics, Nagoya University, JST PRESTO

We studied personal attributes represented in tweets, such as gender, occupation, and age groups. First, we examined how much these basic attributes can be predicted from the texts of tweets, each of which was vectorized by a word2vec-based method for machine learning. The results showed that machine learning algorithms can predict all three attributes with 60-70% accuracy. We also confirmed that differences in word usage between males and females (related to semantic differences) affect the predictive accuracy of gender. Furthermore, we quantified other personal attributes, such as Big 5 and values, using IBM Personality Insights.

1. はじめに

近年, 情報のハブかつ意見交換のプラットフォームであるソーシャルメディアから，人間行動に関するビッグデータを収集できるようになった．このような時代背景から，計算社会科学 (Computational Social Science) と呼ばれる新しい学際科学が誕生し，現在盛んに研究が行われている [Lazer 09]．例えば，Twitter では「ゆるいつながり」による多様なコミュケーションを観測することができ，ソーシャルデータの入手が容易なため，人間行動や社会現象の理解に利用されている [Sasahara 13, Takeichi 15].

ソーシャルデータを用いた研究の中でも，個人属性の推定に関する研究は活発に行われている．ソーシャルデータに含まれる大量の言語表現には，パーソナリティーを反映した情報が潜在していると考えられる [Pennebaker 15]．Schwartz らは，Facebook の投稿に含まれる単語とトピックの使用頻度に基づいて辞書を作成し，性格，性別，年齢を判別できることを示した [Schwartz 13]．Kosinski らは，Facebook の「いいね！」の頻度に基づいて，線形回帰などの比較的簡単な方法でも個人属性を推定できることを示した [Kosinski 13]．Liu たちは，中国のソーシャルメディア Weibo の投稿に基づいて，autoencoder を学習させ，投稿者のビッグ・ファイブと呼ばれる人間の性格因子（神経症傾向，外向性，経験への開放性，協調性，誠実性）を推定できることを示した [Liu 16]．これらのように，ソーシャルデータから個人情報を数値化・推定する手法は，多くの研究者が取り組んでいるものの，まだ確立した一般性の高い方法は存在しない．

本研究の目的は，ソーシャルデータのテキストのみから個人属性をどの程度推定できるかに関するベースラインの知見を獲得し, その理由を調査することである．そこで，Twitter のデータを用いて，機械学習のアルゴリズムによって性別，職業，年齢層の 3 つの個人情報を予測する実験を行った．さらに，性別を例とし，男女の共通単語の予測精度への影響を調べた．

連絡先: 余 岳, 464-8601 名古屋市千種区不老町
名古屋大学大学院情報学研究科, yotake1987@nagoya-u.jp

2. 方法

2.1 データ収集

本研究では予測モデルを構築するために，Web 検索で性別と職業が確認でき，投稿されたツイート数が 3000 以上のアクティブアカウントを 120 人特定した．また，これらのアカウントの年齢を Web 検索で可能な限り特定した（年齢は非公開な場合が少なくない）．これらのアカウントのうち，100 人をトレーニング用，残り 20 人をテスト用にした．

120 人のアカウントの特徴は，次のようにまとめられる．男性と女性のアカウント数は等しく，偏りがない．職業は 10 種類を選択し，それぞれの職業に属するアカウント数は等しく，偏りがない．年齢層は，1980 年以後の生まれ（デジタルネイティブ）か，それ以前の生まれ（デジタル移民）の 2 種類にした．

次に，これらのアカウントから Twitter の公式 API を用いて，リツイートや返信を含むユーザータイムラインを可能な限り収集した．その結果，トレーニング用のアカウントからは 314382 個，テスト用のアカウントからは 64027 個のツイートが収集された．

2.2 データ処理

次の手順でデータ処理を行った．収集したツイートを，日本語形態素解析ツール MeCab と日本語辞書 NEologd を用いて分かち書きの処理をした．分かち書きしたツイートから，長さが４未満の情報量が少ないツイートを削除した．クリーニング後に残ったツイートは，トレーニング用が 312169 個，テスト用が 63454 個である．トレーニング用の全ツイートをコーパスとして（全単語数は 11308535 個，異なり語は 395491 個），word2vec [Mikolov 13] を用いて (window size は５，イテレーション回数は 20)，全ての異なり語を単語ベクトルの辞書に変換した．

各ツイートのベクトルは，そこに含まれる単語のベクトルの平均値を使用した．単体のツイートに含まれる情報量は少ないため，複数のツイートを束にして学習させる方が精度が向上する可能性があると考えられる．そこで，複数のツイートを束にしたものをツイートブロックと呼び，そこに含まれる各ツイー

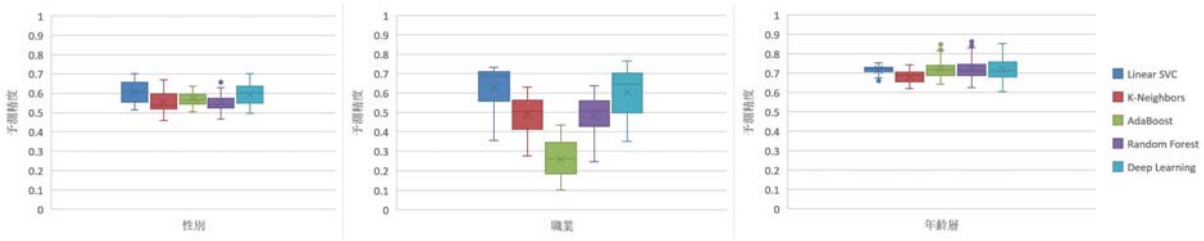


図 1: 各課題における予測精度の分布

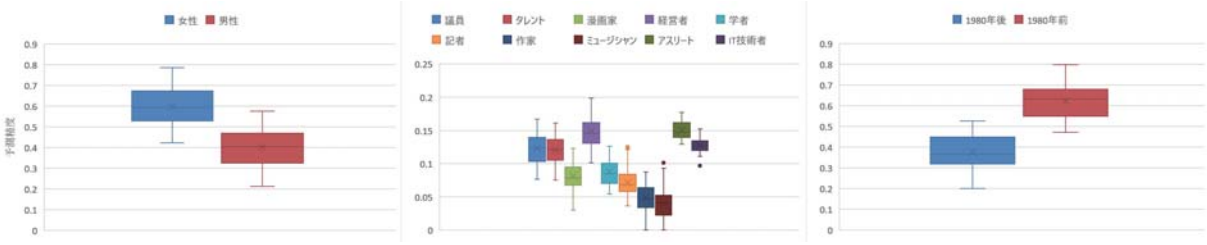


図 2: 個人属性における各項目の予測のしやすさ

トをベクトル的に平均化したものツイートブロックのベクトルとして扱った．

2.3 基本的な個人属性の予測

以上の処理から得られたツイートブロックのベクトル値を入力として，機械学習を用いてモデルを構築し，個人属性の予測精度を調べる．使用した学習アルゴリズムは，Linear Support Vector Classification（Linear SVC），K-Neighbors，AdaBoost，Random Forest の 4 つと深層学習である．

2.4 その他の個人属性の定量化

本研究の 120 人の Twitter のアカウントに対し, IBM Watson の Personality Insights を使って得られたデータを定量化した. Personality Insights の公式 API を用い, 各アカウントにおけるビッグ・ファイブ（5 項目）, ニーズ（12 項目）, バリュー（5 項目）の合計 22 項目の個人属性のデータを取得した. 本研究では, 特に職業に注目して, 22 項目の個人属性を 22 次元のベクトルとして各アカウントを特徴付ける. 各職業ごとに 12 名のアカウントの平均値をとり, その職業の特徴ベクトルにする. そして, 主成分分析の手法で各職業の 2 次元の空間配置を観察する.

2.5 個人属性の予測に影響する要因の分析

個人属性の予測が成功する要因を調べるために， 次のようなデータ分析を行う．例えば, ツイート・コーパスに存在している男女が共通して使用する単語は，性別の予測するときに重要な役割を果たすと推測される. ここでは，性別を例としてデータを分析し，この考えを検証する．トレーニング用のツイート・コーパスを性別ごとに分け，前節と同様の処理で単語の分散表現を作り，男女ごとにそれぞれ単語ベクトルの辞書を作成した．男性の単語ベクトル辞書（単語数 237141）と女性の単語ベクトル辞書（単語数 241645）に基づき，共通単語の集合（単語数 83295）を抽出した．そして，男女の共通単語において，コサイン距離を測った．

予測精度への影響を調べるため，トレーニング用のコーパスから（単語数が 376577）から，重み W (コサイン距離と出現頻度の積) の大きい順に従って単語を取り除き，新しい単語ベクトル辞書を作成した．この辞書を用い，トレーニング用の 100 人とテスト用の 20 人のすべてのツイートをベクトル化した．また，非共通単語をランダムに取り除き，同様の処理をして，比較対象とした．

3. 結果

3.1 基本的な個人属性の予測

図 1 は，word2vec の埋め込み次元 N とツイートブロックのサイズ L の全ての組み合わせにおいて，3 つの個人属性の推定における精度の分布を示したものである．年齢層は，どの学習アルゴリズムにおいても平均して 70%程度の予測精度を示していることから，ツイートのテキストのみから年齢層を推定することは，他の 2 つと比べて容易であることがわかる．また今回の実験では，Linear SVC と深層学習は他のアルゴリズムと比べて，安定して高い精度が得られることがわかった．

3.2 個人属性の項目ごとの予測

性別，職業，年齢層の 3 つの個人属性には，予測しやすい（しにくい）項目があると考えられる．それを調べたのが図 2 で，深層学習（全ての N と L の組み合わせ）によって予測が成功した項目の内訳を示している．性別の場合は，男性より女性の方が予測しやすい．職業の場合は，経営者とアスリートは予測しやすく，作家とミュージシャンは予測しにくい．年齢層の場合は，デジタルネイティブよりデジタル移民の方が予測しやすいことがわかった．

3.3 その他の個人属性の定量化

Personality Insights の結果に基づき, 22 次元で表現した各職業の個人属性 (平均値) を, 主成分分析で二次元にマップしたのが 図 3 である．これを見ると, 学者と記者, タレントとミュージシャンは個人属性が似ていることがわかる. 一方, IT 技術者とアスリート, 漫画家と経営者は, 他の職業とは個人属性が似ていないことが読み取れる.

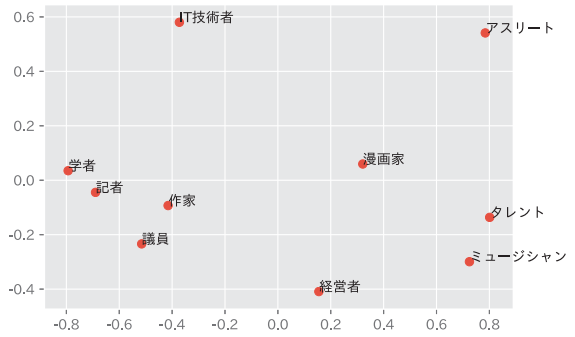


図 3: Personality Insights の個人属性のデータに基づく職業の類似度

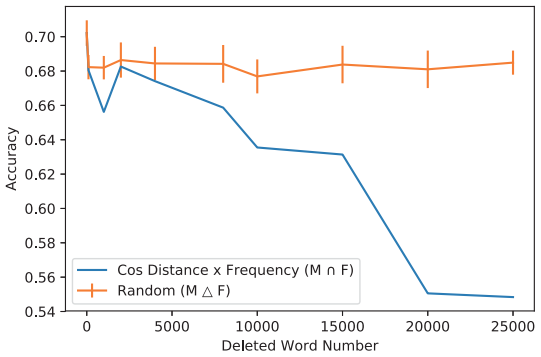


図 4: 男女の共通単語と非共通単語を取り除いた際の予測精度への影響

3.4 性別予測に影響する要因の分析

トレーニング用のツイート・コーパスから，重み W の大きい順に共通単語を 1 つずつ取り除いて，性別の予測精度を調べた．同様に，非共通単語をランダムに 1 つずつ取り除いて予測精度（10 回の施行の平均値）を比較した．ここでは、深層学習のアルゴリズムを用い，前の実験結果による N=500, L=50 という性別予測に最適なパラメーターで行った．その結果が図 4 である．取り除かれた共通単語の個数を増やすにつれて，予測精度は最初の 70%から 55%に大幅に下がった．一方，比較対象の非共通単語の場合は，取り除かれた単語数を増やしても予測精度がずっと 68%程度に安定している．

4. まとめ

本研究では，まず，ツイートのテキストのみから性別，職業，年齢層という 3 つの個人属性を予測する実験を機械学習を用いて行い，ベースラインの結果を得た．

予測精度を比べると，年齢層（デジタルネイティブか否か）は，性別や職業よりもアルゴリズム的に推定しやすいことが示された．これは，言葉の選び方や使い方が年齢によって異なるからではないかと考えられる．テキストデータのみから，60〜70%の精度で個人属性を推定できるということは注目に値する．ツイートから個人属性を予測する場合，単体のツイートよりも複数のツイートをまとめてブロックとして使用した方が精度が顕著に上がった．この結果は，ツイートの束をまとめてベクトル化したものが，個人属性の推定に必要な情報を比較的低次元で表現していることを示唆している．換言すると，テキストのみから個人属性を推定するためには，ある程度の量のデータが必要である．Twitter の場合は，60%程度の予測精度を得るためには，50 個程度がその下限だということになる．その他の個人属性の定量化から，個人属性が類似する職業や，逆に類似しない職業があることがわかった．

個人属性を予測できる理由を調べたところ，男女が共通して頻繁に使用する単語の中で意味が異なる単語は，性別の予測精度に大きい影響を与えることが示された．共通単語の場合は，取り除かれた単語数が増えていくにつれ，予測精度が著しく下がった．さらに，取り除かれた単語数は 15000 から 20000 の間にある時，最も予測精度が下がった．これは，一部の共通単語が他より予測精度に強い影響を与えていることを示唆している．非共通単語の場合は，取り除かれた単語数が増えても予測精度が大きく変動することがなかった．これらの結果によると，性別を予測する場合は，非共通単語と比べて共通単語が予測精度に大きく影響することが分かった．今後，性別以外の個人属性でも同様の調査をする予定である．

本研究の技術が確立すれば，ソーシャルデータから個人属性を精度良く推定することが可能になり，計算社会科学の分析やマーケティングへの応用など幅広い利用が期待できる．

謝辞

本研究は JSPS 科研費 (JP16K16112, JP15H03446, JP17H06383JST), JST さきがけ (JPMJPR16D6), JST CREST(JPMJCR17A4) の助成を受けたものです．

参考文献

[Kosinski 13] Kosinski, M., Stillwell, D., and Graepel, T.: Private traits and attributes are predictable from digital records of human behavior, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 15, pp. 5802–5805 (2013)

[Lazer 09] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Alstyne, M.: Computational Social Science, *Science*, Vol. 323, No. 5915, pp. 721–723 (2009)

[Liu 16] Liu, X. and Zhu, T.: Deep learning for constructing microblog behavior representation to identify social media user ' s personality, *PeerJ Computer Science*, Vol. 2, p. e81 (2016)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013)

[Pennebaker 15] Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., and Booth, R. J.: The Development and Psychometric Properties of LIWC2007 (2015)

[Sasahara 13] Sasahara, K., Hirata, Y., Toyoda, M., Kitsuregawa, M., and Aihara, K.: Quantifying collective attention from tweet stream, *PLoS ONE*, Vol. 8, No. 4, p. e61823 (2013)

[Schwartz 13] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., et al.: Personality, gender,

and age in the language of social media: The open-vocabulary approach, *PLoS ONE*, Vol. 8, No. 9, p. e73791 (2013)

[Takeichi 15] Takeichi, Y., Sasahara, K., Suzuki, R., and Arita, T.: Concurrent Bursty Behavior of Social Sensors in Sporting Events, *PLoS ONE*, Vol. 10, No. 12, pp. e0144646–13 (2015)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

未成年女性のネットリスク分析

Analysis of Minor Women's Online Risk

平野　雄一 *1　鳥海　不二夫 *1　高野　雅典 *2　和田　計也 *2　福田　一郎 *2
Yuichi Hirano　Fujio Toriumi　Masanori Takano　Kazuya Wada　Ichiro Fukuda

*1東京大学　*2株式会社サイバーエージェント
The University of Tokyo　CyberAgent Inc.

With the spread of smartphones and SNS sites, online communication has become indispensable for social life. However,the risk of online communication for minors such as sexual crime and cyberbullying is increasing. In particular, many minor women suffer sexual crime through the Internet. In this research, we analyzed the communication among sex and age group using online chat service data. As a result, we showed that what kind of communication has high risk for minor women.

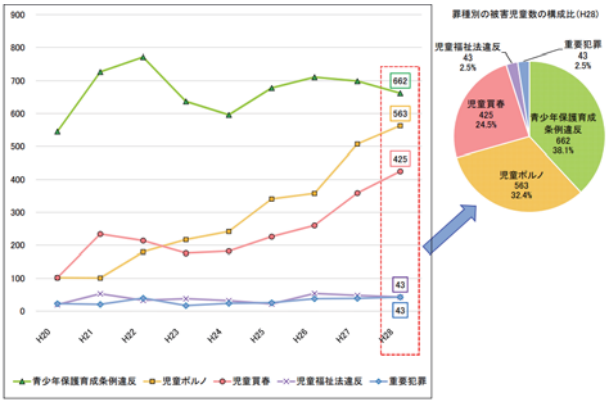


図 1　被害児童数の推移 2(警察庁資料 [3] より抜粋)

1.　イントロダクション

現在、世界中で普及しているインターネットは、情報を簡単に取得できるだけでなく、世界中の誰とでも簡単に会話できるといった利便性があることから、コミュニケーション手段としての重要度が高くなっている。メールや SNS、チャットアプリ等を始めとしたインターネットを通じたコミュニケーションは社会生活上、必要不可欠なものとなってきている。

この傾向は中学生、高校生を初めとした未成年者にとっても同様のことがいえる。総務省「平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書」[2] によると日本国で代表的なコミュニケーションツールの一つである LINE の 10 代における普及率は 79.3% であり、多くの未成年者がインターネットをコミュニケーションの場に利用しているといえる。

しかしながら、オンラインでのコミュニケーションには、不用意な発言から炎上やネットいじめ [1] に発展したり未成年者が悪意を持った成人に誘い出され性犯罪の事件に発展したりする等、未成年者にとってのリスクも含んでいる。警察庁「平成 28 年におけるコミュニティサイト等に起因する事犯の現状と対策について」[3] によると、図 1 にあるように、オンラインでのコミュニケーションによって被害に会う未成年者の数は性犯罪を中心として年々増えている。

このような問題に対して、Toriumi らの研究 [5] では、コミュニケーションペアに着目し送受信した文章の数や文字量等の情報を元に、ユーザーの年齢を含めずに潜在的な被害者と加害者の候補の抽出を行った。しかしながら、Wolak らの研究 [4] によると、オンラインで誘い出しが行われ性犯罪に至ったケースの多くが、未成年女性と 26 歳以上の成人男性であったので、潜在的な被害者や加害者特定のためには年齢や性別を考慮に入れることは重要である。一方で、ユーザーが偽りの年齢でサービスに登録している場合や、年齢を登録する必要のないサービスである場合等、運営会社側にとって正しい年齢が分からない場合も多い。

未成年女性にとって成人男性の交流全てが危険というわけではなく、オンラインでの自由な交流により多様な価値観に触れることができる等、彼女らの成長に繋がる部分も多いと考えられる。従って、彼らのマッチングのうち、リスクの高い交流と安全な交流を分離することが重要である。

本研究では、まずユーザーの年齢を推定した上で未成年女性と成人男性の交流を検出し、次に、彼らの交流形態がどのようになっているのかを求め、さらに、どういったユーザー同士のマッチングがリスクが高いのかについて発見することを目指す。

2.　データ

2.1　サービス

本研究では株式会社 7gogo によって提供されている「755」*1 を用いた。「755」は、ユーザーが設立した「トーク」に、参加したユーザーが投稿したり、参加していないユーザーでも「やじうまコメント」と呼ばれるコメントを付与することにより、各ユーザーが交流することのできるサービスである。「755」の特徴は芸能人のユーザーが多いことと、「トーク」がオープンであり、どのユーザーからも投稿内容を見ることができることである。「755」においては一般的に、はじめに「トーク」の設立者と他のユーザーが「やうじうまコメント」を通じて仲良くなり、その後そのユーザーが「トーク」に招待され、参加し、他の参加者たちと交流する。即ち、「トーク」での投稿は仲の良いユーザー同士の交流であることが多いと推察できる。

また、LINE や Twitter を始めとした他のサービスの ID やメールアドレスなど個人情報や、公序良俗に反する内容等を含む投稿を「やじうまコメント」として投稿した場合は運営会社

*1　https://7gogo.jp/

表 1　利用データ

名前	データ量	備考
ユーザー	191,448	投稿内容を含む
トーク	153,903	トークに対するやじうまコメントを含む

表 2　年齢分布

属性	未成年男性	未成年女性	成人男性	成人女性	合計
人数	126	193	241	319	879

表 3　年齢性別推定結果

label	適合率	再現率	F 値	データ数
未成年男性	0.65	0.22	0.33	126
未成年女性	0.67	0.82	0.74	319
成人男性	0.69	0.81	0.74	241
成人女性	0.64	0.54	0.59	193
平均	0.67	0.67	0.65	879

表 4　年齢性別推定分布

属性	未成年男性	未成年女性	成人男性	成人女性
人数	3,203	14,260	18,863	6,340

から削除される。本研究では、運営会社から削除されたコメントを「NG コメント」と定義し、NG コメントを投稿したユーザーを「NG ユーザー」と定義する。

2.2　使用データ

本研究では 2015 年 2 月 1 日から 2015 年 2 月 7 日までの 1 周間におけるユーザーによるチャット部屋 (トーク) と、「トーク」へのコメント (やじうまコメント) への投稿文章データを用いた。但し、次節の求める年齢性別推定で用いたデータは学習データ量の確保のため、2014 年 12 月 1 日から 2015 年 3 月 31 日までのデータを用いた。

「755」ではトークにサービス上参加していなくてもコメントを通じて交流しているユーザーは多くいること、またサービス上参加していても殆ど活動していないユーザーもいるが、本研究でトークに参加したユーザーの定義は、トークに対して投稿した、もしくはコメントを行ったユーザーであるとする。

2.3　年齢性別推定

本研究ではトークに対して 20 件以上投稿したユーザの年齢性別ラベルを、Hirano ら [6] と同様の手法で次のように取得した。

1. 年齢性別正解ラベルを投稿文章から取得
2. 正解データから TF-IDF とロジスティック回帰を用い [7]、年齢性別推定分類器を学習
3. 分類機から全ユーザーの年齢性別ラベルを推定

本研究で用いた正解データ数は表 2 に、10 分割交差検証による精度は表 3 に、推定したユーザーの年齢性別ラベルは表 4 に示した。

なお、本研究では未成年者の定義は 18 歳以下、成人は 19 歳以上とした。

3.　未成年女性のネットリスク

本章は、多数対多数のコミュニケーションの中で、どのような特徴を持つ交流で未成年者が高いリスクに晒されているかを解明することを目的とする。まず、本研究におけるネットリスクを定義し、そのネットリスクについての考察を行った。続いて、現状のサービスではどういったコミュニケーションにおいてリスクが高くなるか分析した。最後に、潜在的にリスクの高いコミュニケーションに着目し、その言語的特徴について分析した。

表 5　投稿文章からの NG ユーザー推定

	適合率	再現率	F 値	データ数
一般ユーザー	0.75	0.76	0.76	3151
NG ユーザー	0.76	0.75	0.75	3151
平均	0.76	0.76	0.76	6302

表 6　NG ユーザーの特徴語 (上位)

順位	単語
1	LINE
2	カカオ
3	寝
4	ライン
5	寝る

3.1　ネットリスクの定義

本研究では、期間内に「1 件以上 NG コメントを投稿したことのあるユーザー」との交流をネットリスクであると定義する。

3.2　NG ユーザーの性質

まず初めに、NG ユーザーがどのような発言が多いか、単語ベースでの分析を行った。NG ユーザーと一般ユーザーそれぞれ 3,151 名ずつについて、年齢性別推定と同様に、投稿文章を TF-IDF からベクトル化し、ロジスティック回帰を用いて、10 分割交差検証で学習し、各評価指標を求めた。結果として表 5 を得た。

特徴語の分析から表 6 を得た。ここで、「LINE」や「ライン」は LINE 株式会社の LINE[*2]、「カカオ」はカカオ社のカカオトーク [*3] というサービスを指していると考えられる。いずれのサービスも、利用者の非常に多いチャットアプリケーションであり、特に LINE は 10 代の 79.3% が利用している [2]。これらのアプリケーションではチャットが主要な機能であり、会話内容を第三者に閲覧されること無く利用できる。実際にオフラインで出会う場合、第三者に見られない場所で住所や待ち合わせの場所等の情報を交換すると考えられるため、これらのアプリケーションへの招待は誘い出しの第一歩であると考えられる。本結果から、それらの NG ユーザーには誘い出しを目的としたユーザーが多いと考えられる。さらに、図 1 にあるように、児童買春や児童ポルノといった誘い出しと関連性の深い要因が近年増大していることから、彼らと未成年者のマッチングの危険性が非常に高いことが分かる。

次に、NG ユーザーの年齢性別と、各年齢性別層に占める NG ユーザーの割合を求めた。分類結果を表 7、表 8 に示した。NG ユーザーの数については、成人男性が、1 ユーザーあたりの NG ユーザーの割合については未成年男性、成人男性が多いという結果になった。これらの結果は、本サービスにおいて、成人男性を中心として注意を払う必要があることを意味している。

*2　https://line.me/ja/
*3　http://www.kakaotalk.jp/

表 7　NG ユーザー年齢性別推定分布

属性	未成年男性	未成年女性	成人男性	成人女性
人数	352	583	2,013	203

表 8　1 ユーザーあたり NG ユーザー年齢性別割合

属性	未成年男性	未成年女性	成人男性	成人女性
人数	11.0%	4.09%	10.7%	3.20%

表 9　部屋の分類

グループ名	未成年女性の割合	成人男性の割合	トーク数
グループ 0	>0.5	>0	736
グループ 1	> 0	>0.5	841
グループ 2	> 0	>0	7,933
グループ 3	>0.5	0	4,295
グループ 4	0	>0.5	2,631
グループ 5	その他	その他	9,101

表 10　各グループでの NG コメントが書かれたトークの割合

	NG トーク数	NG トーク割合
グループ 0	9	1.22%
グループ 1	36	4.28%
グループ 2	24	0.30%
グループ 3	26	0.61%
グループ 4	16	0.61%
グループ 5	14	0.15%

3.3　トークの分割

トークにおいて、成人男性と未成年女性がどのような交流が行われているかを理解することを目的に、未成年女性と成人男性が参加しているかどうか、未成年女性か成人男性どちらかが参加人数において過半数を占めているかどうかという 2 つの基準を用いて 2 人以上が参加しているトークを 6 つに分割し、結果を表 9 に示した。例えば、グループ 1 は未成年女性、成人男性が参加していて、かつ成人男性が参加人数の過半数を占めているトークを表していて、グループ 4 は成人が参加しているが、未成年女性が参加していないことを示している。

表 9 の結果を元に、どのグループのリスクが高いかについて直接的リスク、潜在的リスクという 2 つの観点から調べた。

3.4　直接的リスクの分析

まず 1 つ目は、実際にリスクに遭遇した未成年女性がどこのグループに多いかを調査することを目的とした。そのために、各グループにどれくらい直接 NG コメントが書かれたかを調べ、表 10 を得た。表 8 において、例えば、グループ 0 では NG コメントが投稿されたトークが 9 件であり、これはグループ 0 の全 736 個トークのうち、1.22% を占めていることを表している。NG コメントは個人情報や社会的問題のあるものを初めとした、実際に運営会社に削除されたコメントであることから、NG コメントがされたトークに参加していた未成年女性は実際に危険に巻き込まれたことを意味している。表 10 から、グループ 1 が量、割合両方の観点からリスクが高いことを示している。このことは、成人が多いトークに参加した未成年者は危険な交流に巻き込まれる可能性が高いことを示唆している。

表 11　各グループでの NG ユーザーの参加割合

グループ	未成年男性	未成年女性	成人男性	成人女性
0	0.00%	2.30%	1.63%	0.40%
1	3.45%	4.99%	10.34%	1.78%
2	0.19%	0.41%	0.50%	0.08%
3	0.02%	0.65%	0.00%	0.02%
4	0.19%	0.00%	1.36%	0.04%
5	0.07%	0.03%	0.06%	0.08%

3.5　間接的リスクの分析

次に、2 つ目は、潜在的に危険に巻き込まれる未成年者がどこに多いかということを調査することを目的とした。そのために、期間内に一度でも NG コメントを投稿したことのある NG ユーザーがどのグループに多いかを調査し、表 11 を得た。表 11 は各グループのトークの中で、どの年齢性別層の NG ユーザーが含まれていたかを表しており、例えば、グループ 1 のトークのうち、未成年女性の NG ユーザーが参加していたトークが 4.99% 存在し、成人男性の NG ユーザーが参加していたトークは 10.34% 存在することを表している。NG ユーザーは実際に NG コメントを投稿しことのあるユーザーであることから、ユーザーの性質的に今後も機会があればそのような投稿をする可能性の高いユーザーであると考えられるため、未成年女性にとって NG ユーザーとの交流は現時点で被害がなくても潜在的なリスクの高い交流であると言える。

表 11 から、どの NG 年齢層においてもグループ 1 が最も高い値となる事がわかった。特に、未成年女性が含まれているかという点で条件が異なるグループ 4 との比較から、成人男性のうち、危険なユーザーは参加する相手として、成人男性同士で交流しているトークよりも、未成年女性が参加しているトークの方を意図して選択して参加している可能性が高いということが分かる。従って、成人男性が多く参加しているトークに参加している未成年者は危険に巻き込まれる確率が潜在的にも高いというこが分かる。

また、グループ 0 との比較により、成人男性は参加するものの、未成年女性が中心のトークよりも成人男性が中心のトークでの方が、未成年女性自身から不適切な投稿をするユーザーが多いと考えられることが分かる。グループ 1 において、多くの NG 成人男性が存在していること未成年男性や成人女性においても高い値となることと合わせて考察すると、この原因には不適切な投稿が多い空間では不適切な投稿をする心理的障壁が小さくなるから、不適切な投稿を促すユーザーに誘導されてしまったから等といった可能性が考えられる。

3.6　考察

以上 2 つの分析から、成人男性が多く参加しているグループ 1 が実際に被害に合っている未成年女性が多い、危険性が高いトークであることがわかり、さらに潜在的にも高いリスクを抱えているトークであるということが分かる。実際に被害があったかどうかだけでなく、潜在的リスクの高い交流に注意を払うことも重要であると考えられる。

3.7　リスクの高いトークルームの特徴

前節まででグループ 1 がリスクの高いトークであるということが示された。本節ではグループ 1 の中でもどのようなトークにリスクが高いかを調べるために、NG ユーザーが 1 人以上参加したグループ 1 のトーク (NG トーク) と、1 人も参加していないトーク (一般トーク) の特徴語を調べた。それぞれ 90 トークずつについて、TF-IDF とロジスティック回帰を用いて分類

表 12　NG トーク推定

	適合率	再現率	F 値	データ数
一般トーク	0.71	0.77	0.74	90
NG トーク	0.75	0.69	0.72	90
平均	0.73	0.73	0.73	180

表 13　NG トークの特徴語 (上位)

順位	単語
1	?
2	し
3	!
4	です
5	こんばんは
6	～
7	(=ω)
8	ライン
9	たい
10	LINE

した。分類結果を示した表 12 から、直接 NG コメントが投稿されたとは限らない NG トークについても単語的な特徴によって判別することが可能であることが分かる。

また、回帰係数上位の単語を示した表 13 から、NG ユーザーと同様に「LINE」や「ライン」といった単語が上位に来ることが分かる。加えて、「?」が最上位であることから、NG トークは一般のトークに対して疑問文が多いということが推察できる。疑問文は一般的に、「好きな食べ物は何ですか?」、「どのアイドルが好きですか?」といった形で、相手から情報を引き出すために用いられるため、疑問文が多いチャット環境の方が各参加者の年齢や趣味といった個人の性質情報、LINEID 等の個人情報が流出しやすいと考えられる。即ち、「LINE」等や疑問文の係数が高いことから、NG トークは言語的特徴という観点から考えても未成年保護という観点からは好ましくない環境と言える。

これらの分析結果から、多数の大人男性と少数の未成年女性が参加しているトークの中でも NG ユーザーが参加しているかどうかによって、トークの単語的特徴に違いが見受けられることが分かり、NG トークは未成年女性にとって言語的に好ましくない環境であり、前節で考察したように直接的リスクだけでなく間接的リスクにも注意が必要ということが示された。

4.　結論

本研究は多数対多数のコミュニケーションについて直目し、どういったチャット空間が未成年女性にとってリスクが高いのかを分析した。本研究によって、年齢性別層の組み合わせの違いによって、ネットリスクの差異があることが示された。特に、多数の成人男性と少数の未成年女性が交流するチャット空間のリスクが高いことが示された。

さらに、多数の成人男性と少数の未成年女性が交流するチャット空間の中でも言語的分析によって危険な発言をしたことのあるユーザーが参加しているチャット空間について分析し、そのチャット空間が未成年女性にとて危険であることを示した。

これらの結果から、本研究は特に多対多のコミュニケーションにおける未成年女性のネットリスク理解に意義が大きく、ネットリスクの軽減に繋がる結果を得ることが出来たと言える。

今後の研究の方向性としては、次のようなものが考えられる。まず、時系列的特徴に着目し、どういった環境が危険なユーザーのトークへの参加を促しているのかについての分析が考えられる。続いて、他サービスとの比較である。他の多数体多数のコミュニケーションを行うことのできるサービスを分析し、本研究の結果と比較することにより、

参考文献

[1] B,Belsey. *Cyberbullying, An Emerging Threat to the 'Always On' Generation.*Recuperado el Vol 5, 2005

[2] 総務省平成 *28* 年情報通信メディアの利用時間と情報行動に関する調査報告書.`http://www.soumu.go.jp/main_content/000357570.pdf`.22.11.2017 取得

[3] 警察庁平成 *28* 年におけるコミュニティサイト等に起因する事犯の現状と対策について `https://www.npa.go.jp/cyber/statics/h28/h28_community_shiryou.pdf`22.11.2017 取得

[4] J,Wolak,;D, Finkelhor; and K.,Mitchell. *Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national*" Journal of Adolescent Health, vol35, no 5, pp424–e11. 2004.

[5] Fujio,TORIUMI;Takafumi,NAKANISHI; Mitsuteru,TASHIRO;Kiyotake,EFUCHI. *Encounters between Predators and Their Targets in Private Chat.*Journal of Transformation of Human Behavior under the Influence of Infosocionomics Society Vol1.2016.

[6] Yuichi,Hirano;Fujio,Toriumi;Mitsuteru,TASHIRO; Kiyotaka,EGUCHI. *Finding Minors Faced wtih Online Risk.*Journal of Transformation of Human Behavior under the Influence of Infosocionomics Society Vol1.2018.

[7] Genkin,A;Lewis,D,D;Madigan,D. *Sparse logistic regression for text categorization.*DIAMCS Working Group on Monitoring Message Streams,Project Report.2005.

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

スマートフォンアプリとWikipediaにおけるニュースに関する閲覧行動の予備調査

Preliminary Investigation for Analysis of Browsing News Pages on Smartphone App and Wikipedia

吉田 光男 *1　関 喜史 *2
Mitsuo Yoshida　Yoshifumi Seki

*1豊橋技術科学大学　*2株式会社 Gunosy
Toyohashi University of Technology　Gunosy Inc.

We conducted a preliminary investigation for the relationship between the number of browsing pages by the news application (impressions and clicks) and page views of Wikipedia (PC accesses and mobile accesses). As a result, we confirmed the tendency that the number of mobile accesses is larger than the number of PC accesses for the Wikipedia entry title included in the news article titles distributed in the news application. The correlation coefficients of these numerical values were not high, on the click numbers, the coefficients of mobile accesses were larger. When utilizing Wikipedia page view data for news applications, it is necessary to separate PC access frequency and mobile access frequency. Furthermore, it suggests that mobile access frequency is more useful.

1. はじめに

スマートフォンの普及に伴い，ニュースの閲覧行動も大きく変化している。2016 年 3 月に発表された意識調査によれば，84%がニュース閲覧にスマートフォンを利用しており，テレビの 61%を大きく上回っている [LINE 18]。また，ユーザがコンテンツを投稿するソーシャルメディアも普及し，自身の生活に関わる情報のほかに，ニュースなどの情報を「シェア」する行動も一般的になった。ユーザがあるニュースを気に入ったとき，そのニュースをシェアするだろうか。あるいは，あるニュースをより深く知るために，そのニュースに関するキーワードでウェブ検索エンジンを利用するだろうか。我々は，ニュースの閲覧と，ニュースを閲覧する以外の行動との関係について興味を持っている。

ニュースの閲覧，ソーシャルメディアの投稿，ウェブ検索エンジンの利用，それぞれの関係の分析は，ウェブリソースを横断した分析と言える。吉田と荒瀬 [吉田 16] は，各種のウェブリソースにおいて，トレンドキーワードがどのように出現したのかを調査しており，ニュースの記事数とソーシャルメディアでの投稿数，あるいは検索頻度とには，緩やかな相関があることを示している。また，Wikipedia の閲覧数（アクセス回数）とソーシャルメディアでの投稿数および検索頻度との間には，高い相関があることも示している。一方，ニュースに関する報告は，記事数のみが取り上げられており，読者の行動となる「閲覧数」については触れられていない。

ニュースの閲覧とそれ以外の行動との関係を明らかにすることを目指し，本稿では，グノシーにおけるニュースの閲覧数と Wikipedia の閲覧数との関係に関する予備調査を報告する。

2. データ・調査方法

本稿では，スマートフォンアプリにおけるニュースの閲覧数と Wikipedia の閲覧数との関係について調査する。スマートフォンアプリにおける記事閲覧数のデータとして，情報キュレーションサービスであるグノシー *1 の閲覧データを利用し，Wikipedia によるけ記事閲覧数のデータとして Wikistats pageview files*2 の閲覧データを利用する。利用するデータの期間は 2017 年 6 月 1 日から 2017 年 11 月 30 日の 6ヶ月間である。

グノシーのデータからは，記事タイトル，表示回数，クリック数，配信タイプを用いる。表示回数はスマホアプリ上で記事のタイトル（およびサマリー画像）が表示された回数であり，クリック数はそのタイトルをクリックするなどして記事本文を表示した回数である。表示回数は，アプリ上でスクロールするなど，ユーザに対して実際に記事タイトルが表示した回数が記録されている。配信タイプはスマホアプリ上で記事を配信したか，プッシュ通知によって配信したかの二値の情報である。地震などの速報ニュースについては，1 日に数件，プッシュ通知によって配信されており，この場合，スマホアプリ上で記事タイトルを表示しないため，表示回数にはカウントされない。なお，プッシュ通知によって配信し，ユーザが記事を閲覧した場合には，クリック数にカウントされる。これらの仕様により，正確なクリック率の算出が困難であることから，本稿ではクリック率を指標として用いていない。

Wikipedia（Wikistats pageview files）のデータからは，見出し語，PC アクセス回数，モバイルアクセス回数を用いる。見出し語とアクセス回数との対応の取り方については，[吉田 15] を参照されたい。本稿では，対応をとる際，言語情報（元データの第一カラム）が “ja” である場合は PC アクセス回数，“ja.m” である場合はモバイルアクセス回数とした。

グノシーおよび Wikipedia のデータは，それぞれ，その性質が異なるため，何らかの方法で照合が必要である。今回，グノシーの記事タイトルに含まれる Wikipedia の見出し語に着目し，その見出し語のアクセス回数の合計をニュースに関係するアクセス回数とした。例えば，記事タイトルが「人工知能が地球を救う」であり，Wikipedia の見出し語として「人工知能」「地球」が存在し，それぞれのアクセス回数が 10 と 5 であるならば，この記事に関する Wikipedia のアクセス回数は 15 であると見なす。

グノシーおよび Wikipedia のデータは，いずれも 1 時間単位で存在するが，本稿では，記事ごとに日別に再集計して利用する。この際，記事ごとに，グノシーにおけるクリック数が 10

連絡先: 吉田光男，豊橋技術科学大学，愛知県豊橋市天伯町雲雀ヶ丘 1-1，yoshida@cs.tut.ac.jp

*1 https://gunosy.com/

*2 https://dumps.wikimedia.org/other/pagecounts-ez/

回以上の時間のみを集計対象とし，その時間帯の Wikipedia のアクセス回数のみを利用する。例えば，6 月 1 日 13 時に 5 クリック，14 時に 50 クリック，15 時に 20 クリックされた記事の場合，10 クリックを下回る 13 時のクリック数を算入せず，6 月 1 日のクリック数は 70 であると見なす。また，Wikipedia におけるアクセス回数の再集計においては，14 時と 15 時のデータのみを用いる。Wikipedia におけるアクセス回数の再集計で，PC アクセス回数またはモバイルアクセス回数が 0 の記事は除外した。

3. 調査結果

まず，ニュースに関する Wikipedia の閲覧数について，PC アクセスが多いのかモバイルアクセスが多いのかを調査した。その結果を，表 1 に示す。総記事数は，別の日に同じ記事が出現した場合は二重にカウントされるものの，本稿における調査対象となる記事数と同等である。大半に記事について，モバイルアクセスの方が多く，スマートフォンによるニュースの閲覧と，スマートフォンによる検索や Wikipedia の閲覧が強く関係していることがうかがえる。

次に，ニュースに関する Wikipedia の閲覧数について，表示回数やクリック数との相関があるかを，ピアソン積率相関係数を用いて調査した。その結果を，表 2 に示す。全体的に値は低く，相関しているとは言いがたいことが分かった。アプリ配信について見ると，表示回数では PC アクセスの方が値が高くなるものの，クリック数ではモバイルアクセスの方が値が高くなる傾向がある。また，クリック数について，アプリ配信とプッシュ配信を比べると，プッシュ配信の方が値が高くなる傾向がある。これらは，記事一覧で表示されるという受動的な行動と，記事をクリックするという能動的な行動という風に見ることもでき，また，Wikipedia の閲覧も能動的な行動であることから，スマートフォン上での能動的な関係同士は，そうでない関係よりも，なんらかの相関が存在することが示唆される。

前節で述べたように，グノシーのデータと Wikipedia のデータとは，記事タイトルに見出し語が含まれるかどうかで，関連付けている。この照合処理において，見出し語が極端に短い場合，適切な関連付けができていない可能性があることから，使用する見出し語の長さを 2 文字以上，3 文字以上と制限をかけて調査した。この調査結果を，表 3，表 4 に示す。文字数の制限をかけるほど，値が下がる結果となった。これは，誤った照合よりも，照合されないことの方が影響が大きいことが考えられる。

4. おわりに

本稿では，スマートフォンアプリによる閲覧数（表示回数およびクリック数）と Wikipedia における閲覧数（PC アクセス回数およびモバイルアクセス回数）との関係について，予備調査を行った。その結果，スマートフォンアプリで配信されたニュースの記事タイトルに含まれる Wikipedia 見出し語について，PC アクセス回数よりもモバイルアクセス回数の方が多い傾向が確認できた。また，それらの数値の相関を調べたところ，相関は高くないものの，クリック数については，モバイル回数の方が多いことが確認できた。スマートフォンアプリに対して Wikipedia の閲覧数データを活用する際には，PC アクセス回数とモバイルアクセス回数とを分ける必要があり，モバイルアクセス回数の方が有用であることを示唆している。

今回の調査では，ニュースのカテゴリを考慮していない。例えば，政治に関するニュースとエンターテイメントに関するニュースとでは，閲覧の性質が異なる [関 17]。このような性質の異なりは，Wikipedia へのアクセス（あるいはウェブ検索エンジンの利用）に影響を与えている可能性があり，今後，カテゴリを考慮した調査を進める。また，日別に相関係数を算出すると，0.428（11 月 30 日・表示・モバイル）を記録するなど，比較的，高い値が出ることもあった。今回の調査では，すべての日を一括して相関係数を算出しており，日別の全体傾向（グノシーの利用者の増加など）が考慮されていない。この点についても，データを細分化するなどして，より詳細な調査を進める。

表 1: Wikipedia における PC アクセス回数よりもモバイルアクセス回数の多かった記事数

	総記事数	モバイルが多い記事数
アプリ配信	917,740	813,542 (88.6%)
プッシュ配信	4,726	4,469 (94.6%)

表 2: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語の制限なし）

		PC	モバイル	PC/モバイル
アプリ	表示	0.166	0.162	0.169
	クリック	0.132	0.149	0.149
プッシュ	クリック	0.163	0.196	0.190

表 3: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語 2 文字以上）

		PC	モバイル	PC/モバイル
アプリ	表示	0.129	0.124	0.130
	クリック	0.108	0.121	0.121
プッシュ	クリック	0.157	0.189	0.184

表 4: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語 3 文字以上）

		PC	モバイル	PC/モバイル
アプリ	表示	0.108	0.109	0.112
	クリック	0.098	0.113	0.112
プッシュ	クリック	0.154	0.187	0.181

参考文献

[LINE 18] LINE 株式会社：世代間のニュースサービス利用に関する意識調査を実施 (2018) https://linecorp.com/ja/pr/news/ja/2016/1267

[関 17] 関 喜史, 潮 旭, 米田 武, 松尾 豊：クリックベイトなニュース記事の特定に向けたユーザ行動分析, 信学技報, Vol. 117, No. 207, pp. 65–70 (2017)

[吉田 15] 吉田 光男, 荒瀬 由紀, 角田 孝昭, 山本 幹雄：検索頻度推定のための Wikipedia ページビューデータの分析, 第 29 回人工知能学会全国大会 (2015)

[吉田 16] 吉田 光男, 荒瀬 由紀：トレンドキーワードに関するウェブリソースの横断的分析, 情報処理学会論文誌. データベース, Vol. 9, No. 1, pp. 20–30 (2016)

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

アスキーアート分類手法の比較検討

Comparison of Classification Methods for Ascii Art

松本 和幸 *1 Kazuyuki Matsumoto　藤澤 日明 *2 Akira Fujisawa　吉田 稔 *1 Minoru Yoshida　北 研二 *1 Kenji Kita

*1徳島大学 Tokushima University　*2青森大学 Aomori University

In recent years, a lot of non-verbal expressions have been used on social media. Ascii art (AA) is an expression by visual technique using characters. In this paper, we set up an experiment to classify AA pictures by using character features and image features. We try to clear which feature is more effective for the method to classify AA pictures. We proposed four methods; 1) character frequency based method, 2) character importance value based method and 3) image feature based method, 4) character's image feature based method. We trained the neural networks by using these four features. As the experimental result, the best classification accuracy was obtained with the feed forward neural networks using character's image feature.

1. はじめに

非言語表現の一つとして，電子掲示板などでアスキーアートという表現が用いられる．アスキーアートは，1 行内に収まる顔文字や絵文字といった非言語表現とは異なり，複数行にわたって表現されるため，文字ベースでの解析が困難である．しかしながら，1 枚のアスキーアートが表現できる意味内容は非常に多様であり，Web 上のデータ解析において無視することができない表現といえる．アスキーアートは文字を点や線の代わりに用いて絵を視覚的に表すため，使用される一つ一つの文字は，アスキーアート中の人物の台詞やキャプションを除けば，ほとんど意味を持たない．したがって，アスキーアートは文字集合として扱うよりも，画像として扱うほうが適していると考えられる．本研究では，アスキーアートのカテゴリ分類に画像特徴量が有効かどうかを検証するため，アスキーアートの文字特徴量（文字出現頻度，文字重要度）と，アスキーアートを画像化して得た画像特徴量をそれぞれ機械学習により学習させてアスキーアートのカテゴリ分類器を作成する．これらの分類器を実験により評価することで，有効な特徴量について考察する．

2. 関連研究

谷岡ら [Tanioka2005] は，バイトパターンと形態素解析による品詞解析結果を用いてサポートベクターマシンによるアスキーアートを含むテキストとそうでないテキストとに分類する手法を提案し，90%の分類精度を得ている．林ら [Hayashi2009] は，アスキーアートを言語に依存せず抽出する手法として，同じ文字が連続して 2 回現れる回数に着目し，さらに，テキストから LZ77 圧縮による圧縮率，RLE 圧縮による圧縮率を素性として用いることにより，アスキーアートの抽出実験の結果，90%を超える高い F 値を得ている．鈴木 [Suzuki2011] は，林らと同様に圧縮率に着目しており，機械学習アルゴリズム C4.5 を用いて特定の言語に依存しない抽出を実現している．いずれの研究においても，アスキーアートかそうでないかの 2 値分類タスクを対象としたものであり，アスキーアートの意味理解については考慮されていない点で本研究とは異なる．

連絡先: 徳島大学理工学部，南常三島町 2-1, matumoto@is.tokushima-u.ac.jp

藤澤ら [Fujisawa2016] は，アスキーアートの持つ形状情報を活用するための手法として，アスキーアートを画像に変換することで，画像特徴量を用いたアスキーアートの類似度評価手法を提案した．画像特徴量を用いることで，顔文字をもとに作成されたアスキーアートに対して，使用されている文字やサイズの違いに影響されず画像特徴量の類似性を確認することに成功している．藤澤ら [Fujisawa2017] は，顔文字と，その顔文字をもとに作られた大型アスキーアートに対し，画像特徴量を抽出して類似度比較を行っている．この研究において，藤澤らは，形状特徴量である HOG(Histograms Oriented Gradients) が画像特徴量として有効であると述べている．本研究では，アスキーアートのカテゴリを文字特徴量と画像特徴量の 2 通りの特徴量を用いた分類器を機械学習により構築することで，従来はほとんど対象とされなかったアスキーアートのカテゴリ分類手法について比較検討する．

3. アスキーアートからの特徴抽出

3.1 画像特徴量を用いたアスキーアートの分類

1 枚のアスキーアート全体から画像特徴量を抽出し，それを機械学習させることでアスキーアート分類器を作成する．アスキーアートの画像への変換には，AAtoImage[AAtoImage] を用いる．画像に変換されたアスキーアートは，N×N ピクセルのサイズの画像に正規化した後，二値化処理をおこないエッジ画像に変換する．本論文では，N= 128 と設定した．

前述の手順でカテゴリごとに準備したアスキーアートのエッジ画像を，畳み込みニューラルネットワーク (CNN) により学習し，アスキーアート分類器を作成する．画像の枚数による学習の偏りの影響を避けるため，あらかじめカテゴリ間の偏りが出ないように訓練データ数を揃えておく．

また，文字単位で画像特徴量 (HOG) を抽出し，その特徴量を k-means 法によりクラスタリングすることで得た文字画像のクラスの出現頻度ベクトル（文字画像ベクトル）をフィードフォワードニューラルネットワーク (FFNN) に学習させてアスキーアート分類器を作成する手法も試す．クラスタリングのクラスタ数を 1000 とした．

3.2 文字特徴量を用いたアスキーアートの分類

アスキーアートから，使用されている文字の頻度を用いる文字頻度ベクトル特徴を FFNN に学習させる方法と，単語重要度計算に用いられる TF∗IDF の考え方に基づく文字重要度を特徴とした行列を CNN により学習させる方法を検討する．文字重要度 (CF∗IAF) は，式 1,2,3 により計算できる．文字重要度を文字単位で計算することで，画像におけるピクセルの画素値と同じ効果を期待する．

$$CF * IAF = CF \times IAF \tag{1}$$

$$CF = freq_c^i \tag{2}$$

$$IAF = log(\frac{N}{AF} + 1.0) \tag{3}$$

CF は，あるアスキーアート AA_i における文字 c の出現頻度を示す．N は，重要度の計算をおこなうアスキーアートの総数を示し，AF は，文字 c が幾つのアスキーアートに出現したかを示す．図 1 に，アスキーアートから文字重要度行列を抽出し，CNN により学習させる過程を簡単に示す．アスキーアートによって 1 行あたりの文字数や行数が異なることから，1 行における最大文字数を 100 文字，最大行数を 100 行と設定し，このサイズに満たない場合は，空白文字によりパディング処理を施した．また，このサイズよりも大きいものは，各行の先頭から 100 文字，1 行目〜100 行目までを使用して特徴抽出をおこなった．

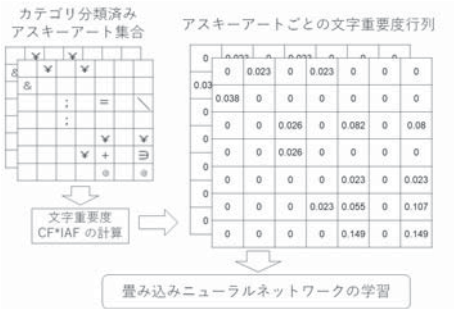


図 1: アスキーアートから文字重要度行列の抽出

同じ作品に登場するキャラクターをモチーフとした AA は，同じ作家が作成している場合が多く，同一キャラクターを描く際には差分のみを変更して作成することが多い．このため，文字頻度ベクトル特徴を用いるだけでも，同一作品カテゴリの判定が可能であると予想できる．しかし，同じ作品においても，デフォルメされたキャラクターであったり，キャラクターの人気や複雑度などから作られるアスキーアート数が少ないキャラクターも存在する．このことから，アスキーアートの種類によっては，文字特徴のみでのカテゴリ分類が困難なことが予想される．

4. 評価実験

4.1 実験設定

画像特徴による AA 分類器（2 種類）と文字特徴による AA 分類器（2 種類）がどの程度アスキーアートのカテゴリ分類に適しているかを確認するため，同じアスキーアートを基に作成したデータに対する評価実験をおこなう．AA 分類器の学習に用いたニューラルネットワーク (NN) の各層のパラメータは，表 1, 2, 3 に示すとおりである．学習試行回数はいずれも 50 回とし，ミニバッチサイズを 256 と設定して学習をおこなった．畳み込み層におけるフィルタは，3×3 のサイズでフィルタ数を 20, 32 とする．中間層の活性化関数に ReLU，最適化アルゴリズムに Adam を使用し，学習率を 0.0001 とする．出力層（全結合層）の直前のドロップアウト層でのドロップアウト率を 0.5 とし，出力層の活性化関数に Softmax を用いる．

表 1: 文字頻度/文字画像ベクトルを学習させる FFNN の構造

Layer	出力	パラメータ数
全結合層-1	(500)	3374000
全結合層-2	(128)	64128
全結合層-3	(44)	5676

表 2: 文字重要度を学習させる CNN の構造

Layer	出力	パラメータ数
畳み込み層-1	(98, 98, 20)	200
畳み込み層-2	(96, 96, 20)	3620
プーリング層-1(Max pooling)	(48, 48, 20)	0
畳み込み層-3	(46, 46, 20)	18496
プーリング層-2(Max pooling)	(23, 23, 20)	0
Flatten 層	(10580)	0
全結合層-1	(1024)	58983424
全結合層-2	(44)	45100

表 3: 画像特徴量を学習させる CNN の構造

Layer	出力	パラメータ数
畳み込み層-1	(126, 126, 32)	320
畳み込み層-2	(124, 124, 32)	9248
プーリング層-1(Max pooling)	(62, 62, 32)	0
畳み込み層-3	(60, 60, 64)	18496
プーリング層-2(Max pooling)	(30, 30, 64)	0
Flatten 層	(57600)	0
全結合層-1	(1024)	58983424
全結合層-2	(44)	45100

4.2 実験データ

本研究では，Web サイト「AAMZ Viewer[*1]」からダウンロード可能なまとめ zip ファイルに含まれる漫画やアニメ，ゲームなどの作品のキャラクターをモチーフに作成されたアスキーアートを対象とし，作品のシリーズ名をカテゴリとみなす.「匿名掲示板 2 ちゃんねる」のオリジナルアスキーアートや「その他作品」に分類されるアスキーアートは実験対象から除いた．この理由として，これらをカテゴリの分類に含めると，様々な作品やキャラクターをモチーフにしたものが同一カテゴリに含まれることになってしまい，それらが訓練の際のノイズとなる可能性があると考えたためである．一方で,「原哲夫作品」などといった，シリーズを通して作風・絵柄が類似するキャラクターが登場する作品群に関しては，一つのカテゴリとして扱う．各アスキーアートの数が 900 以上のカテゴリ（計 44 カテゴリ）を実験対象とする．訓練データと検証データを，カテゴリごとに均等になるように 8:1 の比率で分け，残りのアスキーアートを評価データとする．データの内訳を表 4 に示す．

以上のように準備したデータをもとに，文字特徴量に基づく分類器（2 種類）と，画像特徴量に基づく分類器（2 種類）をそれぞれ作成する．

4.3 評価方法

ニューラルネットワークにより出力された分類カテゴリごとの確率値の上位 k=3,5,10 件以内に，正解カテゴリが出力されれば分類成功とする場合の分類成功率を求める．また，上位 1 件に出力された場合のみを成功とした場合の完全一致率（精

∗1 http://aa.yaruyomi.com/

表 4: データの内訳

訓練データ	検証データ	評価データ
35200	4400	20309

度）も求める．さらに，k=10 のときに分類成功した事例の平均順位について求めることで，分類器の性能を比較する．

4.4 実験結果

表 5 に，各分類器における分類成功率，完全一致率および平均順位を示す.「文字頻度」は，文字頻度ベクトルを FFNN により学習した分類器,「文字重要度」は，文字重要度行列を CNN により学習した分類器,「画像特徴」は，エッジ画像に変換して CNN により学習した分類器を示している．また,「文字画像」は，文字単位の HOG 特徴量をクラスタリングして得た文字画像ベクトルを FFNN により学習した分類器を示している．

表 5: 実験結果

	文字頻度	文字重要度	画像特徴	文字画像
分類成功率 k=3(%)	41.19	21.61	32.96	45.22
分類成功率 k=5(%)	49.57	28.92	39.18	52.65
分類成功率 k=10(%)	64.30	44.16	51.72	65.96
完全一致率 (%)	27.59	11.12	25.64	31.97
平均順位	3.307	4.307	3.299	3.046
作品別精度平均 (%)	31.41	13.13	45.50	35.33

実験の結果，文字画像ベクトルによる手法が，他の手法よりも高い分類成功率，完全一致率および平均順位が得られた．また，画像特徴に基づく手法は，作品別精度平均で最も高い結果が得られた．

5. 考察

精度上位および下位作品の例を示しながら，各手法ごとの違いについて考察する．表 6，7，8, 9 は，各手法における精度が上位 10 件，下位 10 件の作品とその精度を示したものである．上位 10 件だけをみると，画像特徴を用いた手法はすべての作品において 70%以上の分類精度を示しており，一方で，文字重要度に基づく手法では，60%以上の成功率の作品は存在しないことがわかる．

図 2 に，4 手法を通して最も高い精度を得た作品「BLEACH」のアスキーアートを示す．アスキーアートからはある程度複雑な絵柄であることがわかる．また，このカテゴリに属する他のアスキーアートをみると，ほぼ同じアスキーアートで，少しの差分があるだけのものも多くみられた．この作品のアスキーアートの種類は全部で 913 種類であり，すべてのデータを学習および評価に用いた．このため，絵柄や作風が独特であるということよりも，同じようなアスキーアートが同一カテゴリに含まれることが結果に大きく影響したといえる．しかし，画像特徴を用いた手法は，文字頻度や文字重要度を用いた手法とは異なる作品が上位に多く出現していることからも，文字の種類に大きく影響を受けなかったと考えられ，汎用性の高い手法であるといえる．

4 手法に共通して高い精度が得られた作品として,「原哲夫作品」「銀河英雄伝説」などがあげられる．これについて，各作品カテゴリに属する小分類の数を比較すると,「原哲夫作品」が 19 種類,「銀河英雄伝説」が 24 種類であり，大きな差は無かった．さらに，キャラクターの表情だけが異なるものや，全身・頭部のみの違い，一つのアスキーアートにキャラクターが複数，単独で描かれているといった違い，台詞の有無の違いなど，類似するアスキーアートが同一カテゴリに多量に含まれている点で共通している．いずれの場合においても，アスキーアートのサイズが異なっていても，用いられる文字の種類は同じであるため，文字を特徴とした場合にも高い精度が得られたと考えられる．

一方で，精度が低かった作品について，画像特徴量，文字画像ベクトルを用いた手法は他の 2 手法と比較して平均して 2 倍以上の精度が得られている．画像特徴量から得られる形状特徴により作品の絵柄の特徴を上手くとらえることが出来た結果であると考える．これらのことから，全体的な精度が低くなった原因は，評価データに多く含まれているカテゴリの作品が，類似するアスキーアートが多いタイプであったため，文字頻度特徴が有利になった結果であると考えられる．

また，どの手法においても下位 10 件に含まれた「機動戦士ガンダム」のように，登場人物だけでなく，登場するメカ類が含まれる場合に，同じカテゴリとして分類することが困難であったと考えられる．これは，アスキーアートのように基本的に線のみで構成された絵の場合，まったく異なる対象（例えばメカ類と人間）を描いた作品に対し，作風や絵柄で共通点を見

表 6: 完全一致率（文字頻度）

順位	作品名	精度	順位	作品名	精度
1	機動戦艦ナデシコ	83.59	35	Rozen Maiden	8.29
2	銀河英雄伝説	73.36	36	コードギアス	6.51
3	魔人探偵脳噛ネウロ	66.67	37	銀魂	6.45
4	恋姫無双	65.91	38	ストライクウィッチーズ	5.57
5	プリキュア	64.25	39	BLEACH	5.26
6	原哲夫作品	62.99	40	VOCALOID・UTAU	3.44
7	さよなら絶望先生	56.59	41	機動戦士ガンダム	3.43
8	ゾイド	55.36	42	女神転生シリーズ	3.42
9	ジョジョの奇妙な冒険	51.3	43	TYPE-MOON 作品	2.96
10	福本伸行作品	47.11	44	魔法少女まどか☆マギカ	2.27

表 7: 完全一致率（文字重要度）

順位	作品名	精度	順位	作品名	精度
1	恋姫無双	52.27	35	コードギアス	6.13
2	逆転裁判	42.86	36	魔法少女リリカルなのは	6.09
3	プリキュア	30.33	37	Rozen Maiden	5.53
4	原哲夫作品	29.92	38	涼宮ハルヒの憂鬱	5.29
5	ドラえもん	29.5	39	咲-Saki-	4.86
6	銀河英雄伝説	27.1	40	東方 project	4.68
7	ポケットモンスター	22.23	41	機動戦士ガンダム	3.84
8	魔人探偵脳噛ネウロ	21.97	42	特撮	3.75
9	機動戦艦ナデシコ	17.56	43	魔法少女まどか☆マギカ	3.54
10	スクライド	17.39	44	遊戯王	3.07

表 8: 完全一致率（画像特徴）

順位	作品名	精度	順位	作品名	精度
1	BLEACH	92.63	35	魔法少女リリカルなのは	17.34
2	逆転裁判	91.84	36	機動戦士ガンダム	17.15
3	銀魂	87.1	37	魔法少女まどか☆マギカ	17.14
4	化物語	80.81	38	遊戯王	16.67
5	キン肉マン	80.73	39	特撮	15.27
6	原哲夫作品	74.8	40	プリキュア	14.37
7	恋姫無双	73.3	41	TYPE-MOON 作品	13.46
8	さよなら絶望先生	71.32	42	涼宮ハルヒの憂鬱	13.25
9	銀河英雄伝説	70.56	43	東方 project	12.86
10	スクライド	70.29	44	Rozen Maiden	11.91

表 9: 完全一致率（文字画像）

順位	作品名	精度	順位	作品名	精度
1	機動戦艦ナデシコ	83.97	35	TYPE-MOON 作品	19.31
2	銀河英雄伝説	71.03	36	とある魔術の禁書目録	18
3	プリキュア	70.87	37	ストライクウィッチーズ	16.04
4	魔人探偵脳噛ネウロ	68.94	38	BLEACH	15.79
5	恋姫無双	67.05	39	化物語	12.12
6	原哲夫作品	66.14	40	Rozen Maiden	11.78
7	さよなら絶望先生	64.34	41	VOCALOID・UTAU	10.64
8	ゾイド	63.1	42	コードギアス	9.58
9	ドラえもん	50.36	43	女神転生シリーズ	4.79
10	ジョジョの奇妙な冒険	48.1	44	魔法少女まどか☆マギカ	4.25



図 2: 分類成功したアスキーアート（画像特徴）

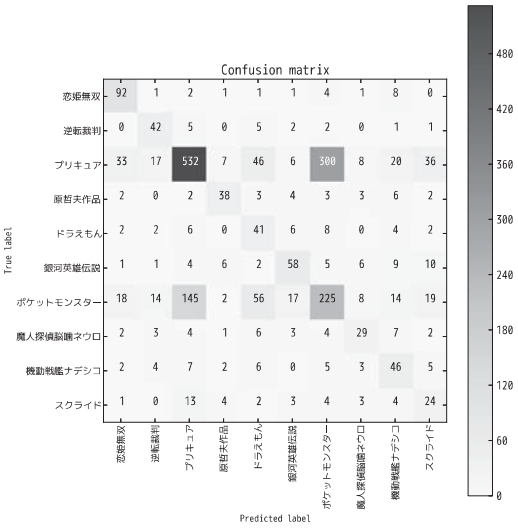


図 3: 完全一致率上位 10 件間の混同行列（文字重要度）

出すことは人間であっても難しいことからも容易に推測できる．アスキーアートのように，収集が比較的困難なものを訓練データとする場合には，人工的なノイズ付加によるデータの拡張や，線画からソースとなる画像を推測し，カテゴリ分類するなどといった工夫が必要と考えられる．

図 3 に，最も精度が低かった文字重要度による手法における完全一致率上位 10 件の作品間での混同行列，図 4 に，最も作品別精度平均が高かった画像特徴量による手法における完全一致率下位 10 件の作品間での混同行列を示す．図からわかる通り，文字重要度を用いた手法は，誤分類が多くみられる．とくに,「ポケットモンスター」,「プリキュア」,「ドラえもん」が混同されやすい傾向がみてとれる．これらの作品はすべて子供向けアニメ作品であることが共通しているため，絵柄に類似点があったと考えられる．一方で，画像特徴量を用いた手法の場合，偏った誤分類傾向はあまり見られず，どの作品にも同程度に誤分類しやすい結果となった．

6. おわりに

本論文では，アスキーアートの分類手法として，文字特徴による手法と，画像特徴による手法を比較検討した．アスキーアートのカテゴリとして，そのアスキーアートのモチーフとなった漫画やアニメ，ゲーム作品の作品名またはシリーズ名を設定し，ニューラルネットワークを用いて分類器の学習をおこなった．実験の結果，文字単位の HOG 特徴量による文字画像ベクトルを FFNN により学習した分類器を用いた場合に全体的な成功率が最も高くなり，画像特徴量を CNN により学習した分類器を用いた場合に，作品ごとでみたときに高い精度（最

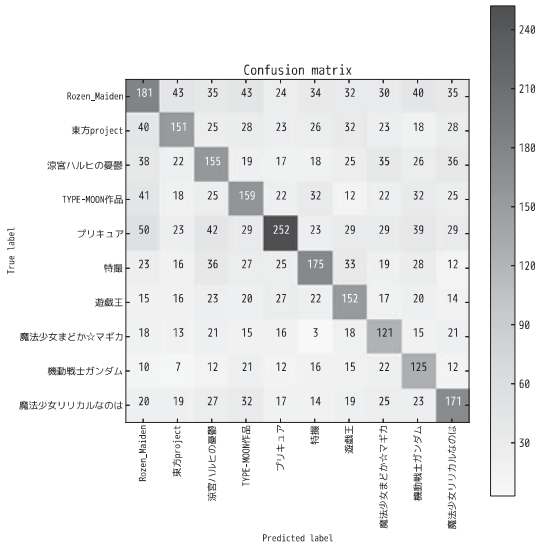


図 4: 完全一致率下位 10 件間の混同行列（画像特徴）

大 92.63%）を得ることが出来た．

今後の課題として，文字頻度ベクトルを用いた手法では不利と考えられる異なる文字で同じキャラクターを表現しているような場合における画像特徴量の優位性を，評価実験をおこなうことで実証したい．また，作品による分類だけでなく,「デフォルメ」か「非デフォルメ」の分類や，モチーフとなっている物体のカテゴリ認識についても検討したい．さらに，写真やイラスト画像をエッジ画像とすることにより，非アスキーアートを訓練データとして用いてアスキーアートのカテゴリ分類器を作成し，その評価をおこないたい．

謝辞

本研究の一部は，科学研究費補助金 15K16077，15K00425，15K00309 によりおこなわれた．

参考文献

[Tanioka2005] 谷岡 広樹, 丸山 稔: 形態素解析に基づく SVM を用いたアスキーアートの識別, 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, 2005-02-25, Vol.104, No. 668, pp. 25–3, 2005.

[Hayashi2009] 林 和幸, 鈴木 徹也: テキスト圧縮を用いた言語に依存しないテキストアート抽出法, 研究報告デジタルドキュメント（DD）, 情報処理学会, 2009-09-18, Vol.2009, No.3, pp.1–6, 2009.

[Suzuki2011] 鈴木 徹也: データ圧縮率を用いるテキストアート抽出法におけるテキストの正規化手法，言語処理学会第 17 回年次大会発表論文集，pp.842–845, 2011.

[Fujisawa2016] 藤澤 日明，松本 和幸，吉田 稔，北 研二: 画像特徴量を用いたアスキーアートからの顔文字検出, 人工知能学会全国大会，3H4-OS-17b-1, 2016.

[Fujisawa2017] 藤澤 日明，松本 和幸，吉田 稔，北 研二: 画像特徴量を用いた大型アスキーアートの分類手法の提案, 人工知能学会全国大会，2O3-OS-22b-3, 2017.

[AAtoImage] http://www.nicovideo.jp/watch/sm20296302，最終アクセス日 2016-3-28.