

Thu. Jun 6, 2019

Room R

[3Rin2] Interactive Session 1

10:30 AM - 12:10 PM Room R (Center area of 1F Exhibition hall)

[3Rin2-01] Traffic Risk Estimation from On-vehicle Video by Region-based Spatio-temporal DNN trained using Comparative Loss

OKwong Cheong Ng¹, Yuki Murata¹, Masayasu Atsumi¹ (1. Dept. of Information Systems Sci., Graduate School of Eng., Soka University)

10:30 AM - 12:10 PM

[3Rin2-02] Bi-directional multimodal generation via estimating conditional distribution of latent variables obtained from pre-trained generative models

OShigeaki Imakiire¹, Masanao Ochi¹, Junichiro Mori¹, Ichiro Sakata¹ (1. University of Tokyo School of Engineering Department of Technology Management for Innovation)

10:30 AM - 12:10 PM

[3Rin2-03] Binarized Variational Information Bottleneck

OMakoto Kawano¹, Yu Oya², Satoshi Yagi², Jin Nakazawa¹ (1. Keio University, 2. NTT Corporation)

10:30 AM - 12:10 PM

[3Rin2-04] Semi-supervised Domain Adaptation using Prediction Models in Associated Domains

OYasuhiro Sogawa¹, Tomoya Sakai¹ (1. NEC Corporation)

10:30 AM - 12:10 PM

[3Rin2-05] A study on measures in multi-armed bandit problem with hidden state.

OKouhei Kudo¹, Takashi Takekawa¹ (1. Kogakuin University)

10:30 AM - 12:10 PM

[3Rin2-06] An Approach to Unseen Classs Classification with In-Service Predictors

OTomoya Sakai¹, Yasuhiro Sogawa¹ (1. NEC Corporation)

10:30 AM - 12:10 PM

[3Rin2-07] Multi-armed bandit algorithm applicable to stationary and non-stationary environment using self-organizing maps

ONobuhito Manome^{1,2}, Shuji Shinohara², Kouta Suzuki^{1,2}, Kosuke Tomonaga^{1,2}, Shunji Mitsuyoshi²

(1. SoftBank Robotics Corp., 2. Graduate School of Engineering, The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-08] Learning Interpretable Control Policies with Decision Trees via the Cross-Entropy Method

OYukiko Tanaka^{1,2}, Takuya Hiraoka^{1,2}, Yoshimasa Tsuruoka^{2,3} (1. NEC, 2. National Institute of Advanced Industrial Science and Technology, 3. The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-09] Development of Embedded System for Recognizing *Kuzushiji* by Deep Learning

Masahiro Takeuchi¹, OTaichi Hayasaka¹, Wataru Ohno¹, Yumie Kato², Kazuaki Yamamoto³, Mamoru Ishima⁴, Tetsuya Ishikawa⁴ (1. National Institute of Technology, Toyota College, 2. Tsurumi University, 3. National Institute of Japanese Literature, 4. TRC-ADEAC, Inc.)

10:30 AM - 12:10 PM

[3Rin2-10] Stochastic Regularization for Residual Networks: Shake-ResDrop and Shake-SENet

OJunya Shirahama¹, Kazuhiko Kawamoto¹ (1. Chiba University)

10:30 AM - 12:10 PM

[3Rin2-11] Fairness-aware Edit of Thresholds in a Learned Decision Tree Using a Mixed Integer Programming Formulation

OKentaro Kanamori¹, Hiroki Arimura¹ (1. Hokkaido University)

10:30 AM - 12:10 PM

[3Rin2-12] Predicting Laughters in Comedy Drama with Subtitles and Facial Expression

OYuta Kayatani¹, Mayu Otani², Chenhui Chu³, Yuta Nakashima³, Haruo Takemura¹ (1. Graduate School of Information Science and Technology, Osaka University, 2. CyberAgent, Inc., 3. Institute for Datability Science, Osaka University)

10:30 AM - 12:10 PM

[3Rin2-13] Application of Aspect-based Sentiment Analysis using Self-Attention Mechanism to Japanese Sentences

ORyuichi Akai¹, Masayasu Atsumi¹ (1. Graduate School of Engineering, Soka University)

10:30 AM - 12:10 PM

[3Rin2-14] Estimating Emotion Intensities in Japanese Tweets Using Emotion Intensity Lexicon

OTatsuki Akahori¹, Kohji Dohsaka¹, Hidekatsu Ito¹,
Masaki Ishii¹ (1. Akita Prefectural University)

10:30 AM - 12:10 PM

[3Rin2-15] An Investigation of Effective Features for
Toponym Resolution of Words in Newspaper
Articles

ORyo Seki¹, Takashi Inui¹ (1. University of
Tsukuba)

10:30 AM - 12:10 PM

[3Rin2-16] Cluster analysis of Twitter Data, using
Interactive Data visualization Tool

OShinichiro Wada¹ (1. Graduate School of
Sociology, Rikkyo University.)

10:30 AM - 12:10 PM

[3Rin2-17] Comparative Analysis of the Effect of
Additional Training between Multiple Domains
by Clustering of the Embeddings of Parsing
Errors

OTakuya Hara¹, Takuya Matsuzaki¹, Hikaru Yokono²,
Satoshi Sato¹ (1. Graduate School of Engineering,
Nagoya University, 2. Fujitsu Laboratories Ltd.)

10:30 AM - 12:10 PM

[3Rin2-18] Vending Machine Drink Recognition with Deep
Learning for IoT Device.

OKatsuhiro Araya¹, Takayuki Osa², Shigemitsu
Yamaoka², Kazuhiko Nishi², Masayuki Nakao² (1.
Arkth Inc., 2. University of Tokyo, School of
Engineering)

10:30 AM - 12:10 PM

[3Rin2-19] Robust Eye Contact Detection for Multi-Party
Conversational Systems

OKenjiro Nogawa¹, Shinya Fujie², Tetsunori
Kobayashi¹ (1. Waseda University, 2. Chiba
Institute of Technology)

10:30 AM - 12:10 PM

[3Rin2-20] Successive estimation of the asteroid shape and
probe motion using sequential images

OToma Suzuki¹, Takehisa Yairi¹, Naoya Takeishi¹,
Yuichi Tsuda², Naoko Ogawa² (1. The University of
Tokyo, 2. Japan Aerospace Exploration Agency)

10:30 AM - 12:10 PM

[3Rin2-21] Image-to-image Translation from Apparel Item
Image Placed Flat to Image Put on Using Deep
Neural Networks

Saki Tsumugiwa¹, OYoshiaki Kurosawa¹, Kazuya
Mera¹, Toshiyuki Takezawa¹ (1. Hiroshima City

University)

10:30 AM - 12:10 PM

[3Rin2-22] Noise reduction of live image in scanning
electron microscope

OFuminori Uematsu¹, Masahiko Takei¹, Mitsuyoshi
Yoshida¹ (1. JEOL Ltd.)

10:30 AM - 12:10 PM

[3Rin2-23] Truthful Dynamic Pricing Mechanisms for On-
demand Mobility Services

OKeiichiro Hayakawa¹, Eiji Hato² (1. Toyota
Central R&D Labs., Inc., 2. The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-24] Multi-agent simulation tool incorporation
group evacuation behavior model

ORyusei Ishida¹, Masanori Akiyoshi¹ (1. Kanagawa
University)

10:30 AM - 12:10 PM

[3Rin2-25] A study of observation fluctuation reduction
method for ear acoustic authentication

OMasaki Yasuhara¹, Takayuki Arakawa², Takafumi
Koshinaka², Shohei Yano¹ (1. National Institute of
Technology, Nagaoka College, 2. NEC Corporation)

10:30 AM - 12:10 PM

[3Rin2-26] New similarity scale to recognize bird calls and
abnormal sounds of concrete/machine

OMichihiro Jinnai¹, Edward James Pedersen² (1.
Nagoya Women's University, 2. Central Queensland
University, Australia)

10:30 AM - 12:10 PM

[3Rin2-27] An Investigation of Controllable Neural
Conversation Model with Dialogue Acts

OSeiya Kawano¹, Koichiro Yoshino^{1,2}, Satoshi
Nakamura¹ (1. Nara Institute of Science and
Technology, 2. Japan Science and Technology
Agency)

10:30 AM - 12:10 PM

[3Rin2-28] Improvement of Knowledge Graph Completion
Using Label Characters for Questions to
Acquire Knowledge in Dialog Systems

OYuma Fujioka¹, Katsuhiko Hayashi¹, Mikio
Nakano², Kazunori Komatani¹ (1. The Institute of
Scientific and Industrial Research, Osaka University,
2. Honda Research Institute Japan Co.,Ltd.)

10:30 AM - 12:10 PM

[3Rin2-29] Acquisition and Utilization of Trivia for
Conversational News Contents Delivery

OHiroaki Takatsu¹, Yoichi Matsuyama¹, Hiroshi Honda², Shinya Fujie^{1,3}, Tetsunori Kobayashi¹ (1. Waseda University, 2. Honda R&D Co.,Ltd., 3. Chiba Institute of Technology)

10:30 AM - 12:10 PM

- [3Rin2-30] Generative Adversarial Networks toward Representation Learning for Image Captions
 OYuki Abe¹, Takuma Seno¹, Shoya Matsumori¹, Michita Imai¹ (1. Keio University)

10:30 AM - 12:10 PM

- [3Rin2-31] Speech extraction from conversation based on image-to-image translation using deep neural networks

OKosuke Takaichi¹, Yoshio Katagami², Yoshiaki Kurosawa¹, Kazuya Mera¹, Toshiyuki Takezawa¹ (1. Graduate School of Information Sciences Hiroshima City University, 2. School of Information Sciences Hiroshima City University)

10:30 AM - 12:10 PM

- [3Rin2-32] An anti-noise performance comparison between acoustic features in detecting voice pathology using machine learning
 OKouta Suzuki^{1,2}, Shuji Shinohara², Nobuhito Manome^{1,2}, Kosuke Tomonaga^{1,2}, Shunji Mitsuyoshi² (1. SoftBank Robotics Corp., 2. Graduate School of Engineering, The University of Tokyo)

10:30 AM - 12:10 PM

- [3Rin2-33] Toward proofreading support using Word2vec
 OMasato Maruyama¹, Takashi Takekawa¹ (1. Kogakuin University)

10:30 AM - 12:10 PM

- [3Rin2-34] Construction of Corpus for Text Simplification by Sentential Alignment based on Decomposable Attention Model
 OKoichi Nagatsuka¹, Masayasu Atsumi¹ (1. Soka University)

10:30 AM - 12:10 PM

- [3Rin2-35] What's Here Like? Analysis of Web Search Log Based on User's Location
 OTatsuru Higurashi¹, Kouta Tsubouchi¹ (1. Yahoo Japan Corporation)

10:30 AM - 12:10 PM

- [3Rin2-36] Relevance Analysis among Domestic Enjoyments based on News Site App Users' Interest
 OKota Kawaguchi^{2,1}, Tatsukuni Inoue¹, Seiya

Osada¹, Tatsuo Yamashita¹ (1. Yahoo! JAPAN, 2. University of Tsukuba)

10:30 AM - 12:10 PM

- [3Rin2-37] Implementation and Evaluation of an Interpretable Fake News Detector

OKazuya Yamamoto¹, Satoshi Oyama¹, Masahito Kurihara¹ (1. Hokkaido University)

10:30 AM - 12:10 PM

- [3Rin2-38] Automatic Evaluation for Cyberbullying Detection Method based on Statistical Scale
 OMasaki Arata¹, Fumito Masui¹, Michal Edmond Ptaszynski¹ (1. Kitami Institute of Technology)

10:30 AM - 12:10 PM

- [3Rin2-39] Automatic Impression Indexing based on Appraisal Dictionary from Tweet

ORuna Yamada¹, Sho Hashimoto¹, Atsuhiko Yamada¹, Noriko Nagata¹ (1. Kwansei Gakuin University)

10:30 AM - 12:10 PM

- [3Rin2-40] Catch Me if Yahoo Can: Hotel Recommendation for Potential Travelers using Transit App Log
 OMikiya Maruyama¹, Kotaro Takahama¹, Kota Tsubouchi¹, Teruhiko Teraoka¹ (1. Yahoo Japan Corporation)

10:30 AM - 12:10 PM

- [3Rin2-41] Official document simplification using neural machine translation approach
 OTakumi Maruyama¹, Kazuhide Yamamoto¹ (1. Nagaoka University of Technology)

10:30 AM - 12:10 PM

- [3Rin2-42] Neural Error Detection for Weather Forecast Manuscript by Pseudo Error Corpus
 ONaruhisa Shirai¹, Masatsugu Hangyo², Mamoru Komachi¹ (1. Tokyo Metropolitan University, 2. WEATHERNEWS INC.)

10:30 AM - 12:10 PM

- [3Rin2-43] Neural Sequence-Labelling Models for ASR Error Correction
 OTaishi Ikeda¹, Hiroshi Fujimoto¹, Takeshi Yoshimura¹, Yoshinori Isoda¹ (1. NTT DOCOMO, INC.)

10:30 AM - 12:10 PM

- [3Rin2-44] Construction of a Diagnosis Representation Model of Person with Dementia Based on ConceptNet for Deeper Understanding of

Physical and Mental Conditions

○Naoki Kamiya¹, Takumi Yoshizawa¹, Shogo Ishikawa¹, Hideki Ueno^{2,4}, Mia Kobayashi^{2,4}, Minoru Maeda³, Chiaki Nishiyama³, Yujun Murakami³, Shinya Kiriya^{1,4}, Yoichi Takebayashi^{1,4} (1. Shizuoka University, 2. Chiba University Hospital, 3. Orange Cross Foundation, 4. The Society of Citizen Informatics for Human Cognitive Disorder)

10:30 AM - 12:10 PM

- [3Rin2-45] Evaluation and Analysis of design for life environment with a high regard for self-reliance based on the representation of a self of people with dementia

○Mika Teramen¹, Shogo Ishikawa¹, Shinya Kiriya¹, Tadasuke Kato², Takeshi Ide², Yoichi Takebayashi^{1,3} (1. Shizuoka University, 2. Aoicare Co., Ltd, 3. Citizen Informatics for Human Cognitive Disorder)

10:30 AM - 12:10 PM

- [3Rin2-46] Development of Electroencephalogram brain-machine interface using convolutional neural network

○Masaki Kato¹, Sotaro Shimada² (1. Meiji University Graduate School, 2. Meiji University)

10:30 AM - 12:10 PM

- [3Rin2-47] Evaluation of Automatic Monitoring of Instillation Adherence Using Eye Dropper Bottle Sensor and Deep Learning in Patients with Glaucoma

○Hitoshi Tabuchi^{1,2}, Kazuaki Nishimura¹, Shunsuke Nakakura¹, Hiroki Masumoto¹, Hirotaka Tanabe¹, Asuka Noguchi¹, Ryota Aoki¹, Yoshiaki Kiuchi² (1. Tsukazaki Hospital, 2. Hiroshima University)

10:30 AM - 12:10 PM

- [3Rin2-48] Sleep/wake classification using remote PPG signals

Yawen Zhang¹, ○Masanori Tsujikawa², Yoshifumi Onishi² (1. HKUST, 2. NEC)

10:30 AM - 12:10 PM

- [3Rin2-49] Biomarker discovery from gene expression data of mixed tumor samples

○Katsuhiko Murakami¹ (1. Fujitsu Laboratories Ltd.)

10:30 AM - 12:10 PM

[3Rin2] Interactive Session 1

Thu. Jun 6, 2019 10:30 AM - 12:10 PM Room R (Center area of 1F Exhibition hall)

[3Rin2-01] Traffic Risk Estimation from On-vehicle Video by Region-based Spatio-temporal DNN trained using Comparative LossOKwong Cheong Ng¹, Yuki Murata¹, Masayasu Atsumi¹ (1. Dept. of Information Systems Sci., Graduate School of Eng., Soka University)

10:30 AM - 12:10 PM

[3Rin2-02] Bi-directional multimodal generation via estimating conditional distribution of latent variables obtained from pre-trained generative modelsOShigeaki Imakiire¹, Masanao Ochi¹, Junichiro Mori¹, Ichiro Sakata¹ (1. University of Tokyo School of Engineering Department of Technology Management for Innovation)

10:30 AM - 12:10 PM

[3Rin2-03] Binarized Variational Information BottleneckOMakoto Kawano¹, Yu Oya², Satoshi Yagi², Jin Nakazawa¹ (1. Keio University, 2. NTT Corporation)

10:30 AM - 12:10 PM

[3Rin2-04] Semi-supervised Domain Adaptation using Prediction Models in Associated DomainsOYasuhiro Sogawa¹, Tomoya Sakai¹ (1. NEC Corporation)

10:30 AM - 12:10 PM

[3Rin2-05] A study on measures in multi-armed bandit problem with hidden state.OKouhei Kudo¹, Takashi Takekawa¹ (1. Kogakuin University)

10:30 AM - 12:10 PM

[3Rin2-06] An Approach to Unseen Classs Classification with In-Service PredictorsOTomoya Sakai¹, Yasuhiro Sogawa¹ (1. NEC Corporation)

10:30 AM - 12:10 PM

[3Rin2-07] Multi-armed bandit algorithm applicable to stationary and non-stationary environment using self-organizing mapsONobuhito Manome^{1,2}, Shuji Shinohara², Kouta Suzuki^{1,2}, Kosuke Tomonaga^{1,2}, Shunji Mitsuyoshi² (1. SoftBank Robotics Corp., 2. Graduate School of Engineering, The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-08] Learning Interpretable Control Policies with Decision Trees via the Cross-Entropy MethodOYukiko Tanaka^{1,2}, Takuya Hiraoka^{1,2}, Yoshimasa Tsuruoka^{2,3} (1. NEC, 2. National Institute of Advanced Industrial Science and Technology, 3. The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-09] Development of Embedded System for Recognizing *Kuzushiji* by Deep LearningMasahiro Takeuchi¹, OTaichi Hayasaka¹, Wataru Ohno¹, Yumie Kato², Kazuaki Yamamoto³, Mamoru Ishima⁴, Tetsuya Ishikawa⁴ (1. National Institute of Technology, Toyota College, 2. Tsurumi University, 3. National Institute of Japanese Literature, 4. TRC-ADEAC, Inc.)

10:30 AM - 12:10 PM

[3Rin2-10] Stochastic Regularization for Residual Networks: Shake-ResDrop and Shake-SENet○Junya Shirahama¹, Kazuhiko Kawamoto¹ (1. Chiba University)

10:30 AM - 12:10 PM

[3Rin2-11] Fairness-aware Edit of Thresholds in a Learned Decision Tree Using a Mixed Integer Programming Formulation○Kentaro Kanamori¹, Hiroki Arimura¹ (1. Hokkaido University)

10:30 AM - 12:10 PM

[3Rin2-12] Predicting Laughters in Comedy Drama with Subtitles and Facial Expression○Yuta Kayatani¹, Mayu Otani², Chenhui Chu³, Yuta Nakashima³, Haruo Takemura¹ (1. Graduate School of Information Science and Technology, Osaka University, 2. CyberAgent, Inc., 3. Institute for Datability Science, Osaka University)

10:30 AM - 12:10 PM

[3Rin2-13] Application of Aspect-based Sentiment Analysis using Self-Attention Mechanism to Japanese Sentences○Ryuichi Akai¹, Masayasu Atsumi¹ (1. Graduate School of Engineering, Soka University)

10:30 AM - 12:10 PM

[3Rin2-14] Estimating Emotion Intensities in Japanese Tweets Using Emotion Intensity Lexicon○Tatsuki Akahori¹, Kohji Dohsaka¹, Hidekatsu Ito¹, Masaki Ishii¹ (1. Akita Prefectural University)

10:30 AM - 12:10 PM

[3Rin2-15] An Investigation of Effective Features for Toponym Resolution of Words in Newspaper Articles○Ryo Seki¹, Takashi Inui¹ (1. University of Tsukuba)

10:30 AM - 12:10 PM

[3Rin2-16] Cluster analysis of Twitter Data, using Interactive Data visualization Tool○Shinichiro Wada¹ (1. Graduate School of Sociology, Rikkyo University.)

10:30 AM - 12:10 PM

[3Rin2-17] Comparative Analysis of the Effect of Additional Training between Multiple Domains by Clustering of the Embeddings of Parsing Errors○Takuya Hara¹, Takuya Matsuzaki¹, Hikaru Yokono², Satoshi Sato¹ (1. Graduate School of Engineering, Nagoya University, 2. Fujitsu Laboratories Ltd.)

10:30 AM - 12:10 PM

[3Rin2-18] Vending Machine Drink Recognition with Deep Learning for IoT Device.○Katsuhiro Araya¹, Takayuki Osa², Shigemitsu Yamaoka², Kazuhiko Nishi², Masayuki Nakao² (1. Arkth Inc., 2. University of Tokyo, School of Engineering)

10:30 AM - 12:10 PM

[3Rin2-19] Robust Eye Contact Detection for Multi-Party Conversational Systems○Kenjiro Nogawa¹, Shinya Fujie², Tetsunori Kobayashi¹ (1. Waseda University, 2. Chiba Institute of Technology)

10:30 AM - 12:10 PM

- [3Rin2-20] Successive estimation of the asteroid shape and probe motion using sequential images
 ○Toma Suzuki¹, Takehisa Yairi¹, Naoya Takeishi¹, Yuichi Tsuda², Naoko Ogawa² (1. The University of Tokyo, 2. Japan Aerospace Exploration Agency)
 10:30 AM - 12:10 PM
- [3Rin2-21] Image-to-image Translation from Apparel Item Image Placed Flat to Image Put on Using Deep Neural Networks
 Saki Tsumugiwa¹, ○Yoshiaki Kurosawa¹, Kazuya Mera¹, Toshiyuki Takezawa¹ (1. Hiroshima City University)
 10:30 AM - 12:10 PM
- [3Rin2-22] Noise reduction of live image in scanning electron microscope
 ○Fuminori Uematsu¹, Masahiko Takei¹, Mitsuyoshi Yoshida¹ (1. JEOL Ltd.)
 10:30 AM - 12:10 PM
- [3Rin2-23] Truthful Dynamic Pricing Mechanisms for On-demand Mobility Services
 ○Keiichiro Hayakawa¹, Eiji Hato² (1. Toyota Central R&D Labs., Inc., 2. The University of Tokyo)
 10:30 AM - 12:10 PM
- [3Rin2-24] Multi-agent simulation tool incorporation group evacuation behavior model
 ○Ryusei Ishida¹, Masanori Akiyoshi¹ (1. Kanagawa University)
 10:30 AM - 12:10 PM
- [3Rin2-25] A study of observation fluctuation reduction method for ear acoustic authentication
 ○Masaki Yasuhara¹, Takayuki Arakawa², Takafumi Koshinaka², Shohei Yano¹ (1. National Institute of Technology, Nagaoka College, 2. NEC Corporation)
 10:30 AM - 12:10 PM
- [3Rin2-26] New similarity scale to recognize bird calls and abnormal sounds of concrete/machine
 ○Michihiro Jinnai¹, Edward James Pedersen² (1. Nagoya Women's University, 2. Central Queensland University, Australia)
 10:30 AM - 12:10 PM
- [3Rin2-27] An Investigation of Controllable Neural Conversation Model with Dialogue Acts
 ○Seiya Kawano¹, Koichiro Yoshino^{1,2}, Satoshi Nakamura¹ (1. Nara Institute of Science and Technology, 2. Japan Science and Technology Agency)
 10:30 AM - 12:10 PM
- [3Rin2-28] Improvement of Knowledge Graph Completion Using Label Characters for Questions to Acquire Knowledge in Dialog Systems
 ○Yuma Fujioka¹, Katsuhiko Hayashi¹, Mikio Nakano², Kazunori Komatani¹ (1. The Institute of Scientific and Industrial Research, Osaka University, 2. Honda Research Institute Japan Co.,Ltd.)
 10:30 AM - 12:10 PM
- [3Rin2-29] Acquisition and Utilization of Trivia for Conversational News Contents Delivery
 ○Hiroaki Takatsu¹, Yoichi Matsuyama¹, Hiroshi Honda², Shinya Fujie^{1,3}, Tetsunori Kobayashi¹

(1. Waseda University, 2. Honda R&D Co.,Ltd., 3. Chiba Institute of Technology)

10:30 AM - 12:10 PM

[3Rin2-30] Generative Adversarial Networks toward Representation Learning for Image Captions

○Yuki Abe¹, Takuma Seno¹, Shoya Matsumori¹, Michita Imai¹ (1. Keio University)

10:30 AM - 12:10 PM

[3Rin2-31] Speech extraction from conversation based on image-to-image translation using deep neural networks

○Kosuke Takaichi¹, Yoshio Katagami², Yoshiaki Kurosawa¹, Kazuya Mera¹, Toshiyuki Takezawa¹ (1. Graduate School of Information Sciences Hiroshima City University, 2. School of Information Sciences Hiroshima City University)

10:30 AM - 12:10 PM

[3Rin2-32] An anti-noise performance comparison between acoustic features in detecting voice pathology using machine learning

○Kouta Suzuki^{1,2}, Shuji Shinohara², Nobuhito Manome^{1,2}, Kosuke Tomonaga^{1,2}, Shunji Mitsuyoshi² (1. SoftBank Robotics Corp., 2. Graduate School of Engineering, The University of Tokyo)

10:30 AM - 12:10 PM

[3Rin2-33] Toward proofreading support using Word2vec

○Masato Maruyama¹, Takashi Takekawa¹ (1. Kogakuin University)

10:30 AM - 12:10 PM

[3Rin2-34] Construction of Corpus for Text Simplification by Sentential Alignment based on Decomposable Attention Model

○Koichi Nagatsuka¹, Masayasu Atsumi¹ (1. Soka University)

10:30 AM - 12:10 PM

[3Rin2-35] What's Here Like ? Analysis of Web Search Log Based on User's Location

○Tatsuru Higurashi¹, Kouta Tsubouchi¹ (1. Yahoo Japan Corporation)

10:30 AM - 12:10 PM

[3Rin2-36] Relevance Analysis among Domestic Enjoyments based on News Site App Users' Interest

○Kota Kawaguchi^{2,1}, Tatsukuni Inoue¹, Seiya Osada¹, Tatsuo Yamashita¹ (1. Yahoo! JAPAN, 2. University of Tsukuba)

10:30 AM - 12:10 PM

[3Rin2-37] Implementation and Evaluation of an Interpretable Fake News Detector

○Kazuya Yamamoto¹, Satoshi Oyama¹, Masahito Kurihara¹ (1. Hokkaido University)

10:30 AM - 12:10 PM

[3Rin2-38] Automatic Evaluation for Cyberbullying Detection Method based on Statistical Scale

○Masaki Arata¹, Fumito Masui¹, Michal Edmond Ptaszynski¹ (1. Kitami Institute of Technology)

10:30 AM - 12:10 PM

[3Rin2-39] Automatic Impression Indexing based on Appraisal Dictionary from Tweet

○Runa Yamada¹, Sho Hashimoto¹, Atsuhiro Yamada¹, Noriko Nagata¹ (1. Kwansei Gakuin)

University)

10:30 AM - 12:10 PM

[3Rin2-40] Catch Me if Yahoo Can: Hotel Recommendation for Potential Travelers using Transit App Log

○Mikiya Maruyama¹, Kotaro Takahama¹, Kota Tsubouchi¹, Teruhiko Teraoka¹ (1. Yahoo Japan Corporation)

10:30 AM - 12:10 PM

[3Rin2-41] Official document simplification using neural machine translation approach

○Takumi Maruyama¹, Kazuhide Yamamoto¹ (1. Nagaoka University of Technology)

10:30 AM - 12:10 PM

[3Rin2-42] Neural Error Detection for Weather Forecast Manuscript by Pseudo Error Corpus

○Naruhisa Shirai¹, Masatsugu Hangyo², Mamoru Komachi¹ (1. Tokyo Metropolitan University, 2. WEATHERNEWS INC.)

10:30 AM - 12:10 PM

[3Rin2-43] Neural Sequence-Labelling Models for ASR Error Correction

○Taishi Ikeda¹, Hiroshi Fujimoto¹, Takeshi Yoshimura¹, Yoshinori Isoda¹ (1. NTT DOCOMO, INC.)

10:30 AM - 12:10 PM

[3Rin2-44] Construction of a Diagnosis Representation Model of Person with Dementia Based on ConceptNet for Deeper Understanding of Physical and Mental Conditions

○Naoki Kamiya¹, Takumi Yoshizawa¹, Shogo Ishikawa¹, Hideki Ueno^{2,4}, Mia Kobayashi^{2,4}, Minoru Maeda³, Chiaki Nishiyama³, Yujun Murakami³, Shinya Kiriya^{1,4}, Yoichi Takebayashi^{1,4} (1. Shizuoka University, 2. Chiba University Hospital, 3. Orange Cross Foundation, 4. The Society of Citizen Informatics for Human Cognitive Disorder)

10:30 AM - 12:10 PM

[3Rin2-45] Evaluation and Analysis of design for life environment with a high regard for self- reliance based on the representation of a self of people with dementia

○Mika Teramen¹, Shogo Ishikawa¹, Shinya Kiriya¹, Tadasuke Kato², Takeshi Ide², Yoichi Takebayashi^{1,3} (1. Shizuoka University, 2. Aoicare Co., Ltd, 3. Citizen Informatics for Human Cognitive Disorder)

10:30 AM - 12:10 PM

[3Rin2-46] Development of Electroencephalogram brain-machine interface using convolutional neural network

○Masaki Kato¹, Sotaro Shimada² (1. Meiji University Graduate School, 2. Meiji University)

10:30 AM - 12:10 PM

[3Rin2-47] Evaluation of Automatic Monitoring of Instillation Adherence Using Eye Dropper Bottle Sensor and Deep Learning in Patients with Glaucoma

○Hitoshi Tabuchi^{1,2}, Kazuaki Nishimura¹, Shunsuke Nakakura¹, Hiroki Masumoto¹, Hirotaka Tanabe¹, Asuka Noguchi¹, Ryota Aoki¹, Yoshiaki Kiuchi² (1. Tsukazaki Hospital, 2. Hiroshima University)

10:30 AM - 12:10 PM

[3Rin2-48] Sleep/wake classification using remote PPG signals

Yawen Zhang¹, Masanori Tsujikawa², Yoshifumi Onishi² (1. HKUST, 2. NEC)

10:30 AM - 12:10 PM

[3Rin2-49] Biomarker discovery from gene expression data of mixed tumor samples

Katsuhiko Murakami¹ (1. Fujitsu Laboratories Ltd.)

10:30 AM - 12:10 PM

Traffic Risk Estimation from On-vehicle Video by Region-based Spatio-temporal DNN trained using Comparative Loss

Kwong Cheong Ng Yuki Murata Masayasu Atsumi

Dept. of Information Systems Sci., Graduate School of Eng., Soka University

Abstract: We propose a method to estimate the traffic risk during road navigation based on the region-based spatio-temporal deep neural network (DNN) trained by the comparative loss function. In this method, moving object regions are extracted using the object detector YOLO and their features are clipped out from the middle layer of the detector. Then, these feature sequence is used to estimate the traffic risk by the spatio-temporal DNN followed by the risk estimation network. Experiments were conducted using the KITTI and Dashcam Accident dataset images and we have shown that it is possible to estimate a dangerous traffic situation using the proposed risk estimation network.

1. Introduction

In the research of advanced driver-assistance systems (ADAS) and autonomous driving, numerous studies have been conducted to estimate the risk of traffic situation from images of driver's viewpoints [1, 2, 7]. In this paper, we propose a traffic risk estimation method based on the region-based spatio-temporal deep neural network (DNN) trained using the comparative loss function. In this method, first, moving object regions are detected using the object detector YOLO [4, 5] from each frame of a video and their features are extracted from a middle layer of the detector. Then, these object feature sequence is input into the spatio-temporal pattern encoding network which consists of a convolutional neural network (CNN) and a long short-term memory (LSTM) and the risk is estimated by the risk estimation network trained by the comparative loss function. The accuracy of the risk estimation is evaluated using a dataset of on-vehicle cameras.

2. Related Work

2.1 Traffic Risk Estimation

Several studies have been proposed in the research of estimating upcoming traffic incidents. Chan *et al.* [1] proposed a dynamic-spatial-attention Recurrent Neural Network (RNN) that could anticipate accidents before they occur. Suzuki *et al.* [2] extended the work of Chan by introducing an Adaptive Loss for Early Anticipation (AdaLEA) method which allows a model to learn earlier incident anticipation as training progresses. In addition, a quasi-recurrent neural network (QRNN) is adopted in the base model to enable stable output from temporal convolution on sequential data such as videos. However, these methods are specialized in anticipating traffic accidents, while our proposed method takes into account learning risk estimation of any traffic situation including accidents, risky incidents, congestion and so on through the comparative loss function. On the other hand, Mark *et al.* [6] suggested a Deep Predictive Model that can learn its own, task-specific filters which improves prediction performance. The model uses a Bayesian

convolutional long short-term memory (ConvLSTM) method to process spatio-temporal visual data, proprioceptive data and steering commands to identify potential impending collisions. In addition, Ernest Cheung *et al.* [8] proposed another approach called Trajectory to Driver Behavior Mapping (TDBM) that accounts the driving behavior of neighboring vehicles to perform risk assessment.

2.2 Traffic Datasets for Risk Estimation

Several datasets are available for studies of the traffic risk estimation. In [1], the Dashcam video dataset is proposed to evaluate the proposed method in which 678 dashcam videos captured across six cities in Taiwan are used. This dataset is also used in this paper to estimate risk for traffic accidents. To focus on near-miss traffic incidents, Suzuki *et al.* [7] introduced a Large Scale Near-Miss Traffic Incident database (NIDB) that comprises over 6.2K videos and 1.3M frames, many of which are incident scenes and are classified into seven classes, including low/high risk for bicycles, pedestrians, and vehicles, as well as a background class. However, as the NIDB dataset is not open-source, it is not used in this paper. Mark *et al.* [6] uses a robotic simulation platform proposed in [9] to simulate experiment data, whereby the training dataset is generated using a series of dynamic street scenes involving two vehicles in a sparsely simulated environment. Due to the dataset is not generated from a real environment, the approach is not used in this paper.

3. Risk Estimation Method

3.1 Overview

As shown in Figure 1, the proposed network consists of the object detector YOLOv2, a moving objects' spatial pattern encoding CNN, a moving objects' spatio-temporal pattern encoding LSTM, and a risk estimation network. First, moving object regions are detected using the YOLO object detector from on-vehicle camera images and their features are extracted from its middle layer. Then, by inputting the feature sequence of these moving object regions into the spatio-temporal pattern encoding network which consists of a CNN and a LSTM, spatio-temporal feature of moving objects are extracted, and the traffic risk is estimated based on the risk estimation network. The risk estimation network is trained by the comparative loss function

Contact: Kwong Cheong Ng, e18m5252@soka-u.jp

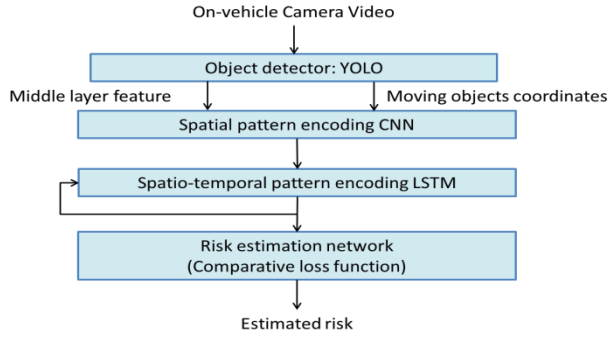


Figure 1: Overview of proposed network

using a set of pairs of images with different relative risk levels.

3.2 Object Feature Extraction

To estimate the risk level on an on-vehicle camera images, it is necessary to detect objects which are involved in traffic risk. Although it is necessary to attend various objects during driving, the important attention targets are moving objects. Therefore, in this research, a pre-trained YOLO using the COCO dataset is further trained using the KITTI dataset [3] and the Dashcam Accident dataset [1], which are on-vehicle camera datasets, to detect moving objects. The target objects to be detected are classified into the following seven categories – car, truck, person, tram, bicycle, motorbike and bus. These categories were unified between the two datasets.

After detecting the moving objects, their features are extracted using the coordinates of them from middle convolutional layers of the YOLO. The 21st layer is used as the middle layer for YOLOv2.

3.3 Moving Objects' Spatio-temporal Feature

The moving objects' spatio-temporal feature is extracted by applying the spatio-temporal pattern encoding network to a composite feature map in which only the convolutional features of moving object regions are clipped from a feature map of the middle layer of YOLO (Figure 1). The network consists of the spatial pattern encoding CNN followed by a spatial pyramid pooling layer and the spatio-temporal pattern encoding LSTM. Figure 2 shows the detailed configuration of the network. In the figure, the (2, 2) of the max pooling layer represents the kernel size and the stride. The (512, 3, 1, 1) of the convolutional layer represents the number of output channels, the kernel size, and the padding width. The (3) in the spatial pyramid pooling layer is the number of pyramid levels.

3.4 Risk Estimation and Comparative Loss Function

In general, it is difficult to evaluate traffic risk by objective numerical values. On the other hand, it is easier to evaluate which is dangerous between a situation where an accident is likely to occur and a situation where no accident occurs or which is dangerous between a congested situation and a non-congested situation. Therefore, we introduce the comparative loss function that learns a correct risk estimation function through relative

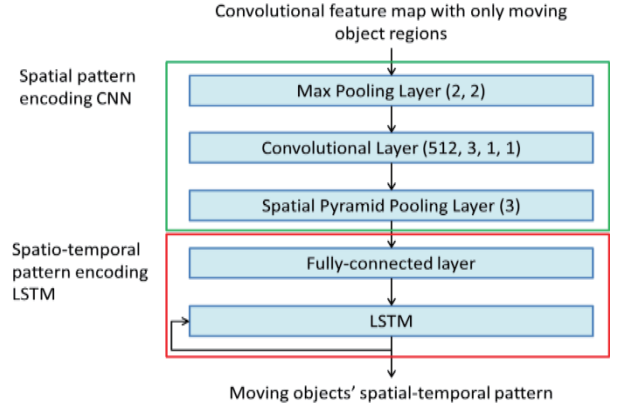


Figure 2: Spatio-temporal pattern encoding network of moving objects

comparison of the risk level between pairs of two spatio-temporal patterns of moving objects. In training, each estimated risk is tuned according to a supervised signal that gives which is more dangerous. Let two camera images be labelled as v_1, v_2 , their relative risk be labelled as $R(v_1, v_2)$, and r_1, r_2 be estimated risks by a sigmoid activation function of the risk estimation network (a perception) respectively. The $R(v_1, v_2)$ is a supervised signal which takes a value of 1 when v_1 is more dangerous than v_2 , -1 when v_1 is safer than v_2 , and 0 when both are comparatively equal. Then, the comparative loss function is defined by the following equation:

$$L(r_1, r_2, R(v_1, v_2)) =$$

$$\begin{cases} \max(r_2 - r_1 + m, 0) & \dots R(v_1, v_2) = 1 \\ |r_2 - r_1| & \dots R(v_1, v_2) = 0 \\ \max(r_1 - r_2 + m, 0) & \dots R(v_1, v_2) = -1 \end{cases} \quad (1)$$

Here, m is a parameter that provides a margin of relative comparison.

In risk estimation, the degree of risk is computed by the risk estimation network from the moving objects' spatio-temporal pattern in each frame interval of a video.

4. Dataset

In experiments, both the KITTI dataset [3] and Dashcam Accident dataset [1] were used for training the object detection network YOLO, and the Dashcam Accident dataset was used for risk estimation experiments. The Dashcam Accident dataset contains videos with accidents and videos without accidents, and only videos with accidents were used in experiments. Each video consists of 100 frames and accidents occur in the last 10 frames. Therefore, in risk prediction experiments, each video was divided into a non-accident part of the former 50 frames and an accident part of the latter 50 frames. The dataset used for the risk estimation is configured as shown in Table 1.

Table 1. Configuration of a dataset. Videos with accidents (positive) samples and videos without accidents (negative) samples. Numerical values in each cell represent the number of videos/ the number of frames

	Training	Test
positive	455/22750	164/8200
negative	455/22750	164/8200

5. Experiments

5.1 Outline of Experiments

The evaluation of risk estimation was performed for two cases: one was performed between videos with accidents and videos without accidents, and the other was performed between two videos without accidents. In the former, the comparative loss function specifies the relative comparison in which videos with accidents are more dangerous. Here, the relative comparison margin was set to 0.5. In the latter, the comparative loss function specifies relative comparison in which videos with more moving objects are more dangerous. Here in this case, the relative comparison margin was set to 0.1.

5.2 Results

The detection performance of moving objects using YOLOv2 was evaluated by the mean average precision (mAP) and the intersection over union (IOU). As a result, the mAP was 0.2902 and IOU was 0.5996. Since the result of detection performance was not so accurate, we evaluated the risk prediction performance by extracting moving object features from YOLOv2 using ground truth object region boxes.

Table 2 shows the result of risk estimation between videos with accidents and videos without accidents.

In table 2, BG_ZERO and BG_GN represent that the region other than moving objects is filled with zero (for BG_ZERO) and filled with Gaussian noise (for BG_GN) respectively. In addition, ‘DO’ represents that a dropout is applied to the output of the fully connected layer in Figure 2.

Table 3 shows the result of risk estimation between two videos without accidents but with the different numbers of moving objects.

Table 2. Accuracy of risk estimation between videos with accidents and videos without accidents

	Training (%)		Test (%)	
	w/ DO	w/o DO	w/ DO	w/o DO
BG_ZERO	99.771	97.706	69.565	62.733
BG_GN	97.706	94.839	70.807	72.050

Table 3. Accuracy of risk estimation between videos without accidents but with the different number of moving objects

	Training w/o DO (%)	Test w/o DO (%)
BG_ZERO	93.349	86.957

Based on the results of these experiments, we found that the proposed method achieved a better result when BG_GN was used. This means that the Gaussian noise is useful to achieve robust training for risk estimation. In addition, we found that the proposed method was able to estimate dangerous situation not only caused by accidents but also triggered by congestion.

6. Conclusion

In this paper, we have proposed a traffic risk estimation DNN which is trained by the comparative loss function. This network encodes a spatio-temporal pattern of moving objects based on YOLO and the spatio temporal network and estimates traffic risk using the risk estimation network. Then, in the experiments, we have shown that it is possible to estimate traffic risk by the proposed network. As a future work, we are going to improve the accuracy of preliminary risk prediction by extending the risk estimation network. To improve the efficiency of the feature extraction, we are going to study the latest object detector YOLOv3 in hope of replacing YOLOv2 for better performance. In addition, we hope to study the effect of detecting traffic signs in traffic risk estimation and add them as a new object detection category.

References

- [1] Chan, F-H., et al., Anticipating Accidents in Dashcam Videos, ACCV, 2016
- [2] Suzuki, T., et al., Anticipating Traffic Accidents with Adaptive Loss and Large-scale Incident DB, CVPR, 2018
- [3] Geiger, A., et al, Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite, CVPR, 2012
- [4] Redmon, Joseph., Divvala, Santosh., et al., You Only Look Once: Unified, Real-Time Object Detection, 2016
- [5] Redmon, Joseph., Farhadi, Ali., YOLO9000: Better, Faster, Stronger, 2018
- [6] Strickland, Mark., Fainekos, Georgios., et al., Deep Predictive Models for Collision Risk Assessment in Autonomous Driving, 2017
- [7] Suzuki, T., et al, Drive Video Analysis for the Detection of Traffic Near-Miss Incidents, 2018
- [8] Cheung, Ernest., Bera, Aniket., et al., Efficient and Safe Vehicle Navigation Based on Driver Behavior Classification, CVPR, 2018
- [9] Rohmer, Eric., Singh, Surya P. N., et al., V-REP: a Versatile and Scalable Robot Simulation Framework, 2013

訓練済み深層生成モデルの潜在変数間相互条件付きサンプリングによるマルチモーダル双方向生成モデル

Bi-directional multimodal generation via estimating conditional distribution of latent variables obtained from pre-trained generative models

今給黎 成彬 大知 正直 森 純一郎 坂田 一郎
Shigeaki Imakiire Masanao Ochi Junichiro Mori Ichiro Sakata

東京大学大学院工学系研究科技術経営戦略学専攻
University of Tokyo School of Engineering Department of Technology Management for Innovation

In recent years, research on multimodal generation that mutually converts between different data such as images and sentences has attracted attention from the viewpoint of applicability to real service such as automatic annotation of images and subtitles of audio.

Meanwhile, in the field of machine learning research, reusable trained models trained using large-scale data sets are being opened to the public, and the number is expected to increase in the future.

Therefore, in this research, we aim to realize multimodal generation with small data by utilizing this trained model. In this paper, we propose a multimodal generation method using a trained generation model in which latent variables of individual modality can be inferred and a small amount of data set. We realized multimodal generation by estimating the conditional distribution of latent variables obtained from trained models using small number of train data.

1. はじめに

近年、画像や文章などの異なるデータ間を相互に変換するマルチモーダル生成の研究が実サービスへの応用可能性の観点から注目されている。マルチモーダル生成は、画像の自動注釈、音声の字幕生成、など最近でも大きな成果を得ており産業応用上重要な技術である一方で、大規模なデータセットが必要であるという課題がある。

一方で、機械学習研究分野は、大規模なデータセットを用いて訓練した再利用可能な訓練済みモデルを一般に公開するようになってきており、その数は今後増加すると見込まれる。

そこで本研究では、この訓練済みモデルを活用することで、小規模なデータでマルチモーダル生成を実現することを目的とする。本稿では、個別モダリティの潜在変数が推論可能な訓練済み生成モデルと少量のデータセットによってマルチモーダル生成を行う手法を提案する。

本稿では画像とラベルからなるマルチモーダルデータに提案手法を適用する実験を行い、与えられた訓練済み生成モデルが十分訓練されていれば提案手法によって少数の訓練データでそれらのモデルと同等程度の精度でマルチモーダル双方向生成が可能なことを示した。

2. 関連研究

2.1 前提となる代表的な深層生成モデル

Kingma らによる Variational Autoencoders[1](以下, VAE) は代表的な深層生成モデルである。VAE は入力データから潜在変数を推論する Encoder と潜在変数から入力データを復元する Decoder という 2 つのニューラルネットワークを組み合わせた深層生成モデルである。VAE の対象とするデータ構造は画像だけでなく、テキスト [2] や音声 [3] などに対するモデルも派生的に提案されている。

Generative Adversarial Networks[4](以下, GAN) は、潜在変数からデータを生成する Generator とその生成データの真偽を判定する Discriminator を互いに学習させることで、現実

のデータに近いデータ生成を可能にする深層生成モデルである。Goodfellow らが提案した元の GAN には入力データから潜在変数の分布を推論する過程は備わっていないが、Dumoulin らによる ALI[5] などそれを可能にするモデルが提案されている。

2.2 単方向のマルチモーダル生成モデル

単方向のマルチモーダル生成モデルが扱う問題は画像からテキストの生成 [6]、テキストから画像の生成 [7]、テキストから音声の生成 [8] など多岐に渡っている。単方向のマルチモーダル生成の主要なアプローチは、一方のモダリティのデータの特徴を抽出する Encoder と抽出された特徴を入力として対応モダリティを出力する Decoder を構築するという方法である。特徴抽出とそのデコードをいかに行うかが重要であるため、モデルが複雑化し、一般的には大規模なデータセットが必要となる。

2.3 双方向のマルチモーダル生成モデル

双方向のマルチモーダル生成モデルは単方向のマルチモーダル生成モデルよりも学習のコストが低いという点で優れている。双方向マルチモーダル生成モデルの主なアプローチは異なるモダリティの共有表現を獲得する方法である。共有表現の獲得の方法としては Deep Boltzman Machine を用いるモデル [9] や AutoEncoders を用いるモデル [10]、VAE を用いるモデル [11] などが存在する。

2.4 本研究の位置づけ

本研究ではマルチモーダルデータについて各モダリティについて潜在変数の推論が可能な生成モデルの訓練済みモデルが与えられた時に、それらの潜在変数の相互の条件付きサンプリングを通じてマルチモーダル双方向生成を行うモデルを提案する。

3. 提案手法

本節では、各モダリティ毎に潜在変数の推論が可能な訓練済み生成モデルが所与であるものとし、一方のモダリティの潜在変数で条件付けた異なるモダリティの潜在変数の分布を推定し、その推定分布からサンプリングされた潜在変数と所与のモ

デルの生成過程によって異なるモダリティデータの生成を行う手法を提案する。

(x, y) という複数のモダリティを持つデータセット $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$ が存在し、それぞれのモダリティについて潜在変数 z_x, z_y が推論可能な訓練済みの生成モデルが所与であるとする。本研究では、 x のみが観測された時に対応する y を獲得する問題と y のみが観測された時に対応する x を獲得する問題の 2 つの問題を考える。本稿では、この問題に潜在変数 z_x, z_y の相互の条件付き分布のモデリングするというアプローチを提案する。

3.1 訓練済み生成モデルの潜在変数間の条件付き分布推定によるマルチモーダル生成

x と y は対称性があるため、まずは x を観測した下で y を獲得する問題を考える。

y の生成過程を考慮すれば、 x が与えられた下での z_y の分布 $p(z_y|x)$ を考えればよいが、 x は低次元の z_x により生成されたものであるため z_x を z_y に対応付ける方が複雑さが小さい。したがって z_x が与えられた下での z_y の分布 $p(z_y|z_x)$ を考える。すると、以下の手順でモダリティ x のあるデータ x_i が観測された時に以下の手順で対応する y_i を生成できる。

1. 観測した x_i で条件付けた z_x の分布 $p(z_x|x_i)$ から z_{x_i} をサンプリングする
2. サンプリングした z_{x_i} で条件付けた $p(z_y|z_{x_i})$ から z_{y_i} をサンプリングする
3. サンプリングした z_{y_i} で条件付けた $p(y|z_{y_i})$ から y_i をサンプリングする

$p(z_x|x)$ は潜在変数 z_x の推論過程、 $p(y|z_y)$ は y の生成過程であり、それぞれ所与の訓練済みモデルで十分よく近似されているとする。よって本稿で取り扱うのは $p(z_y|z_x)$ の推定である。

より具体的に $p(z_y|z_x)$ をモデル化する方法として、 $q_{\psi_x}(z_y|z_x)$ を真の $p(z_y|z_x)$ の近似分布として、 z_x を入力として q の分布パラメータを出力する関数 h_{ψ_x} を考える。ただし、 q は所与のモデルにて z_y の事後分布に仮定されている分布と同一形状の分布とする。すなわち、例えば $p(z_y|y)$ に多変量正規分布が仮定されているならば、 $q_{\psi_x}(z_y|z_x)$ も多変量正規分布を仮定する。ここで、その分布パラメータを γ_y すると $q_{\psi_x}(z_y|z_x)$ は以下のように書き表すことができる。

$$q_{\psi}(z_y|z_x) = \text{Dist}_y(z_y|\gamma_y = h_{\psi_x}(z_x; \psi_x)) \quad (1)$$

h のパラメータ ψ_x を最適化すれば所与の訓練済みモデルと組み合わせ x を観測した下で対応する y の生成が可能になる。同様に y を観測した下で対応する x を獲得したい場合は、 z_y を入力として z_x が従う分布のパラメータを出力する関数 h_{ψ_y} を考える。

3.2 最適化問題

まずは h_{ψ_x} の最適化を考える。 (x_i, y_i) というペアのデータのうち x_i のみが観測された場合に、関数 h_{ψ_x} がどのような性質を満たせば y_i を生成できるかを考える。3.1 節で記した通り、 y_i を生成するならば、分布 $p(z_y|y_i)$ からサンプリングされた潜在変数 z_{y_i} が必要である。よって、 x_i によって z_{x_i} が推論された下で z_{y_i} を獲得するためには、 $q_{\psi_x}(z_y|z_{x_i})$ と $p(z_y|y_i)$ を一致させればよい。すなわち、 h_{ψ_x} は z_{x_i} を入力として $p(z_y|y_i)$ に近い分布のパラメータを出力する関数であればよい。

$D[p||q]$ を確率分布 p と q の間に定義され、分布間の距離を評価できるダイバージェンス関数とすると、パラメータ ψ_x はそのダイバージェンス関数を最小化するような値であればよい。これは以下の式を満たすことと等しい。

$$\min_{\psi_x} D[q_{\psi_x}(z_y|z_{x_i})||p(z_y|y_i)] \quad (2)$$

ただし、 $z_{x_i} \sim p(z_x|x_i)$

$D[q_{\psi_x}(z_y|z_{x_i})||p(z_y|y_i)]$ は x_i, y_i, ψ_x の関数として表せるためこれを $L_i(\psi_x, x_i, y_i)$ とする。ここで、目的は特定の x_i だけでなく、データセット $\{(x_1, y_1), \dots, (x_N, y_N)\}$ の各点において ψ_x が式 2 を満たすことである。よって、 $L_i(\psi_x, x_i, y_i)$ の期待値の最小化を考える。

$$\min_{\psi_x} \mathbb{E}_{(x_i, y_i) \sim p_{data}} \left[L_i(\psi_x, x_i, y_i) \right] \quad (3)$$

式 3 の期待値は $X = \{x_1, x_2, \dots, x_N\}, Y = \{y_1, y_2, \dots, y_N\}, \psi_x$ の関数として表されるためこれを $L(\psi_x; X, Y)$ とする。 N が十分に大きいと L の計算はメモリ効率上困難となるため、バッチ学習を提案する。すなわち、データセット $\{(x_1, y_1), \dots, (x_N, y_N)\}$ からランダムな M 個のサンプル X^M, Y^M を抽出し、以下の関数 L^M で L を近似する。

$$\begin{aligned} L(\psi_x, X, Y) &\simeq L^M(\psi_x, X^M, Y^M) \\ &= \frac{N}{M} \sum_{i=1}^M L_i(\psi_x, x_i, y_i) \end{aligned} \quad (4)$$

ダイバージェンス関数やモデルパラメータの最適化には様々な選択が考えられるが、本稿においてはダイバージェンス関数には KL ダイバージェンスを選択し、モデルパラメータの最適化には勾配降下法を用いる。

4. 実験の概要

本実験では、第 3. 節のモデルを画像とラベルからなるデータセットに適用してそれらの相互生成を行った。本実験の目的は提案手法の有効性を検証するために以下の 3 点を確認することである。

1. 提案手法によって異なるモダリティの潜在変数間の条件付き分布が正しく推定され、所与の訓練済み生成モデルと組み合わせマルチモーダル双方向生成が可能となること
2. 提案手法においてモデルパラメータの最適化が少数のデータセットでも可能であること
3. 双方向の生成が与えられた訓練済み生成モデルと同程度の精度で行われること

4.1 データセット

使用したデータは MNIST データセットである。MNIST は 70,000 件の手書き数字画像データとその画像に書かれた数字を表すラベルデータからなるデータセットである本実験では全データセットのうち 60,000 件を各モダリティの生成モデルの訓練に利用し、 N 件を提案モデルの訓練に利用し、残りの $(10,000 - N)$ 件を評価で利用した。

4.2 訓練済み生成モデル

本実験では事前に各モダリティの VAE の訓練を行い、その後各モダリティの潜在変数間の条件付き分布の推定を行う提案モデルを訓練した。

どちらのモダリティについても潜在変数の分布は正規分布とし、潜在変数の次元は画像は 64 次元、ラベルは 2 次元とした。どちらのモデルも十分に学習させた。

4.3 提案モデルの訓練詳細

提案モデルに用いた構造は、2 層の BatchNormalization 及び LeakyReLU 活性化関数付きの線形全結合層と潜在変数の事後分布パラメータを出力する全結合層からなる多層パーセプトロンである。隠れ層の出力ユニット数は順に 64, 128 とし、最終層の出力ユニット数はターゲットのモダリティの潜在変数の次元数とした。また、LeakyReLU 関数の negative slope は 0.2 とした。

モデルのパラメータ最適化に必要なデータ数を確認するために訓練データ数 $N = \{256, 1024, 8192\}$ で訓練を行った。

5. 結果と考察

5.1 学習の進展

図 1 は提案モデルを訓練中の目的関数の値の推移である。

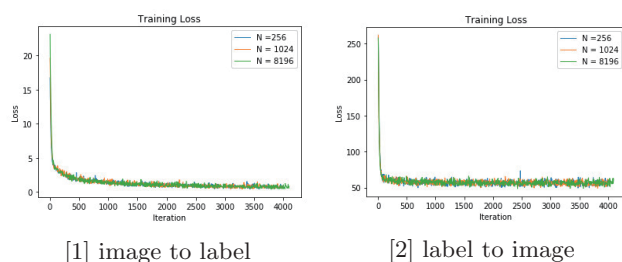


図 1: 目的関数の値の推移

目的関数の値は学習が進むにつれて減少していることから異なるモダリティの潜在変数間の条件付き分布が推定されていることが分かる。また、学習の速度と最終的な収束値は訓練データ数によって変わっていないことから訓練データ数の増減が提案モデルの学習に及ぼす影響は小さいと考えられる。

5.2 ラベルデータから画像データの生成

0 ~ 9 のラベルデータから画像データを生成した結果と訓練済み画像 VAE による生成をまとめたものが図 2 である。

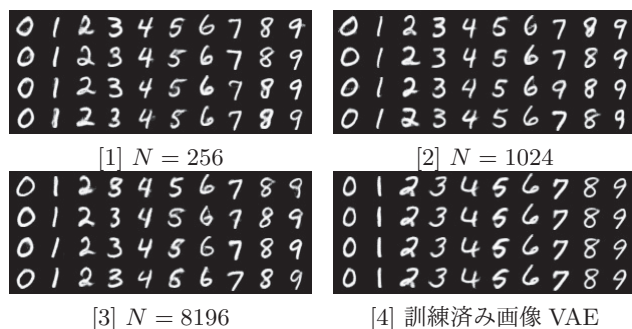


図 2: ラベルから生成した画像と訓練済み画像 VAE によって生成した画像

どの訓練データ数においても入力ラベルに合致した画像が生成され、生成された画像の質も訓練済み VAE によるものと同等精度であることが分かる。

次に、画像データ x の対数尤度 $\log p(x)$ の推定値による生成精度の定量評価を行う。本実験の目的は提案手法によるデータの生成が所与の訓練済みモデルと同程度であることの確認であるため、対数尤度の比較は所与の訓練済みモデルによる生成と提案手法の生成の間で行う。表 1 はその結果である。

表 1: 提案手法と訓練済み画像 VAE による対数尤度の比較

モデル	$\log p(x)$
$N = 256$	-351.52
$N = 1024$	-354.59
$N = 8192$	-355.56
訓練済み画像 VAE	-89.99

どの訓練データ数においても提案手法は元の訓練済みモデルよりも低い対数尤度となった一方で、訓練データ数による大きな差は見られない。

5.3 画像データからラベルデータの生成

潜在変数の分布が正しく推定されているかを検証するために、画像データの潜在変数からラベルデータの潜在変数を二次元平面上に可視化した結果を確認する。図 3 はその結果である。

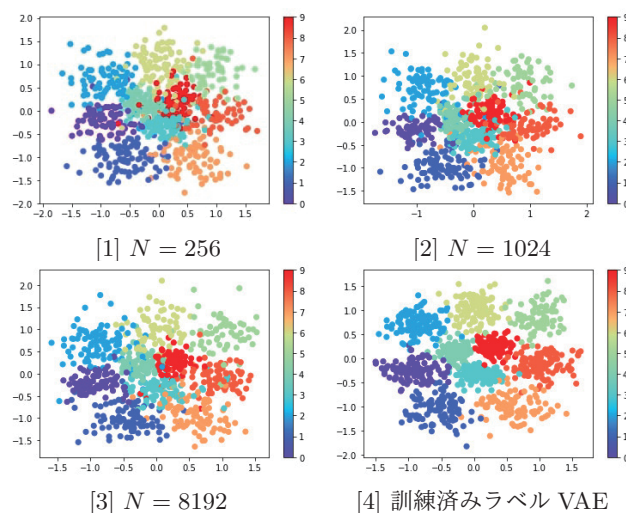


図 3: 画像から推定したラベルの潜在変数と訓練済みラベル VAE で推定した潜在変数のプロット

どのデータ数においても訓練済み VAE と同様のプロットとなっていることが分かる。

次に、ラベルデータ y の対数尤度 $\log p(y)$ の推定値による生成精度の定量評価を行う。結果は表 2 である。対数尤度には推定値を用いているため、値の大小関係が変動しうることに注意されたい。

表 2: 提案手法と訓練済みラベル VAE による対数尤度の比較

モデル	$\log p(y)$
$N = 256$	-0.0456
$N = 1024$	-0.0456
$N = 8192$	-0.0457
訓練済みラベル VAE	-0.0456

訓練データ数間での差はなく、かつ訓練済みモデルと同等精度であることが分かる。

5.4 所与の訓練済みモデルの学習が不十分な場合

双方向生成が所与の訓練済みモデルと同等精度で行われることを確認するために、画像データの VAE に 3 エポックのみしか訓練していない学習途中の重みを使って提案手法を学習した。目的は双方向生成が所与の訓練済みモデルと同等精度で行われるかの確認なので、訓練データ数の比較は行わず $N = 1024$ のみで実験を行った。

図 4 はラベルデータから画像データを生成した結果である。ラベルと対応した画像が生成されているが、生成の質は画像 VAE よりもわずかに劣っている。

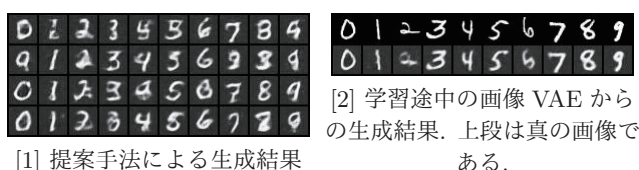


図 4: 学習途中の画像 VAE を用いた場合の生成結果

図 5 は画像データからラベルデータの潜在変数を推定した結果である。十分に訓練した場合である図 3 と比較すると異なるラベルの潜在変数が重なりあっていることが分かる。

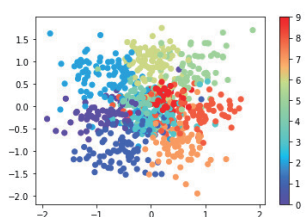


図 5: 提案手法によって推定した潜在変数のプロット

5.5 考察

以上の実験結果に対する考察を行う。5.2 節においてラベルデータから対応する画像データが生成されたこと、5.3 節において多くの画像から対応するラベルデータが正しく生成されたこと、5.3 節において提案手法によって画像データから推定したラベルデータの潜在変数が訓練済みラベル VAE によって推定した潜在変数と同様の分布をしたこと、これら三点によって提案手法によってモダリティの潜在変数間の条件付き分布の推定を介してマルチモーダル生成が可能になったことが分かった。

5.1 節における学習の進展速度と目的関数の収束値にデータ数による違いがないこと、5.2 節と 5.3 節における定性及び定量評価にデータ数による差が見られなかったこと、これらによって提案手法は少数の訓練データで学習が可能になったことが分かった。

5.2 節において提案手法によってラベルから生成した画像が質の面で訓練済み画像 VAE による生成画像と同等程度であったこと、5.3 節において提案手法の尤度評価が訓練済み VAE と同等程度なこと、5.4 節において与えられたモデルが未学習ならば生成の質が落ちること、これら三点からは所与の訓練済み生成モデルの潜在変数の推論過程が十分パラメータ化されていれば提案手法による生成は与えられたモデルによる生成と同等程度の精度で行われることが分かった。

6. 結論

様々なモダリティについて個別に高精度な生成モデルが提案され、かつ訓練済みモデルの公開が一般的となりつつある現状に着目して、訓練済み生成モデルと小規模なデータセットを組み合わせることで十分に高精度なマルチモーダル双方向生成を達成することが本研究の目的であった。

まず複数モダリティについてそれぞれ訓練済みの生成モデルを所与として、その潜在変数間の相互の条件付き分布を推定しその分布からのサンプリングによって得られた潜在変数と所与のモデルの生成過程を用いることによってマルチモーダル生成を行うモデルを提案した。そして、特に所与の訓練済み生成モデルが Variational Autoencoders である場合に、潜在変数間の相互の条件付き分布を多層パーセプトロンによって推定することでたしかにマルチモーダル生成が可能であることを実験によって示した。また、従来手法とは異なり潜在変数間の条件付き分布を推定するのみなので、モデルを簡略化することによってパラメータ数を減らし小規模なデータセットによってモデルパラメータの最適化が可能であることを確認した。さらに、所与の訓練済み生成モデルが十分に訓練されているならばそれと同等程度の生成が可能であることを確認した。

以上より、本研究はマルチモーダルデータについての訓練済み生成モデルが得られた一方で対応する訓練データセットは少数しか得られない場合に十分よく双方向生成を達成する手法を提案し、その有効性を示した。

参考文献

- [1] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [2] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. Toward controlled generation of text. *arXiv preprint arXiv:1703.00955*, 2017.
- [3] Wei-Ning Hsu, et al. Learning latent representations for speech generation and transformation. In *Inter-speech*, pp. 1273–1277, 2017.
- [4] Ian J. Goodfellow, et al. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 2672–2680, Cambridge, MA, USA, 2014. MIT Press.
- [5] Vincent Dumoulin, et al. Adversarially learned inference. *CoRR*, Vol. abs/1606.00704, , 2016.
- [6] Quanzeng You, et al. Image captioning with semantic attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4651–4659, 2016.
- [7] Elman Mansimov, et al. Generating images from captions with attention. In *ICLR*, 2016.
- [8] Aaron van den Oord, et al. Wavenet: A generative model for raw audio. In *SSW*, 2016.
- [9] Nitish Srivastava, et al. Multimodal learning with deep boltzmann machines. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pp. 2222–2230. Curran Associates, Inc., 2012.
- [10] Jiquan Ngiam, et al. Multimodal deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pp. 689–696, USA, 2011. Omnipress.
- [11] Masahiro Suzuki, et al. Improving bi-directional generation between different modalities with variational autoencoders. *arXiv preprint arXiv:1801.08702*, 2018.

二値化ニューラルネットワークへの変分情報ボトルネックによる正則化

Binarized Variational Information Bottleneck

河野 慎^{*1}
Makoto Kawano

大屋 優^{*2}
Yu Oya

八木 哲志^{*2}
Satoshi Yagi

中澤 仁^{*1}
Jin Nakazawa

^{*1}慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University

^{*2}NTT ソフトウェアイノベーションセンタ
NTT Software Innovation Center

Deep neural networks are utilized in various applications in real worlds, thanks to their capabilities. One of the fashions of it is their deployment on edge devices. With edge devices, deep neural networks can be used in the context of IoT. However, the specification of those edge devices is often poor so that deep neural networks cannot be deployed. Binary neural networks, whose weights and activations are binarized, is one of the solutions. There is a well-known issue, the drastic drop in accuracy compared to its full precision networks. We consider that this is because the binary neural networks can only represent a subset of discrete functions so that they become sensitive to the input perturbation: the lack of robustness for inputs. In this paper, we propose a regularization approach that helps to alleviate the over-fitting problem by introducing variational information bottleneck. We show ablation studies on CIFAR-10 that reduce loss value though accuracy is maintained on AlexNet-like networks with different binary activation functions.

1. はじめに

深層学習は、その性能の高さから様々な分野で応用されている。特に画像認識の分野の発展は著しく、大量の GPU/TPU と数千～数億枚の画像を学習に用いることで、近年は画像分類タスクだけではなく、高品質・高画質な画像生成 [Brock 18] が可能となっている。一方で、道路の損傷点検 [Kawano 17] のように、高性能計算機ではなく RaspberryPi3 や NVIDIA JetsonTX2 などエッジデバイスでの実行を想定した研究が注目されている。エッジデバイスで深層学習を実行する場合、その性能の低さゆえに深層学習モデルを小さくする必要がある。

深層学習モデルを圧縮することで、エッジデバイスのような低性能計算機でも実行可能にする技術が研究されている。本研究では、圧縮技術の一つである量子化に注目する。量子化は、32 ビットの単精度浮動小数点数で表現されている深層学習モデルのパラメータを 16 ビットや 8 ビットなど低演算精度で表現するものである。特に、重みパラメータと層間の信号^{*1}を 1 ビットで表現された binary neural networks (BNNs) [Courbariaux 16] や XNOR-Net [Rastegari 16] は、メモリを最大で 1/32 の節約と速度の向上が期待される。さらに、BNNs は FPGA 上での実行において、その優位性が発揮される。

一方で、BNNs には、単精度浮動小数点数で表現された深層学習モデルと比べて、一段と精度が下がってしまう問題がある。これは、学習の逆伝播時にパラメータの表現力が不足していることが主な原因とされている [Lin 17]。つまり、重みパラメータの量子化（二値化）は情報量の欠落による表現力の低下を引き起こし、重みパラメータ値の急激な変動（+1 → -1 や -1 → +1）によって、入力の変動に対する頑健性が失われており、過学習しやすくなってしまっている。

過学習を抑制する手法として、重みパラメータへの制約として L1 正則化や L2 正則化が挙げられるが、BNNs において重みは +1 もしくは -1 になっているため、その効果を受けるこ

とができない。そこで本研究では、確率的にニューラルネットワークを扱う変分情報ボトルネックを BNNs に導入し、BNNs の過学習抑制を試みる。

2. 二値化変分情報ボトルネック

2.1 二値化ニューラルネットワーク

BNNs は、重みおよび信号を +1 もしくは -1 のいずれかの値に制約することで、モデルのサイズも小さくなり、また推論時に論理演算で行うことが可能となる。重みもしくは信号 x^b の二値化は

$$x^b = \text{Sign}(x) = \begin{cases} +1 & x \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

である。ただし、 x は、実数値を表しており、Sign は符号関数を表している。BNNs の順伝播時、式 (1) によって二値化された重みパラメータ W^b を用いて計算される。層間の活性化関数にも符号関数が用いられ、二値の信号が出力される。しかしながら、符号関数の勾配はほとんど至るところで 0 であるため、逆伝播時に勾配が消失してしまい、学習ができない。この問題を解決するため、BNNs では、Hinton の講義 (2012) で紹介された straight through estimator (STE) :

$$\frac{\partial \text{Sign}}{\partial x} = 1_{|x| \leq 1}$$

が用いられることが多い。なお、STE は hard tanh とみなすことも可能であることが報告されている [Courbariaux 16]。また、符号関数を滑らかな関数として近似した SignSwish :

$$\text{SS}_\beta(x) = 2\sigma(\beta x)[1 + \beta x\{1 - \sigma(\beta x)\}] - 1$$

関数を用いることで、より学習が安定して行われることが報告されている [Darabi 18]。なお、 σ はシグモイド関数を表す。

STE により BNNs の学習が可能になる一方で、依然として過学習が起きやすい。この原因として、BNNs

連絡先: 河野 慎, 慶應義塾大学, 神奈川県藤沢市遠藤 5322, makora9143@gmail.com

^{*1} 活性化信号。層から出力され、次の層の入力になるもの。

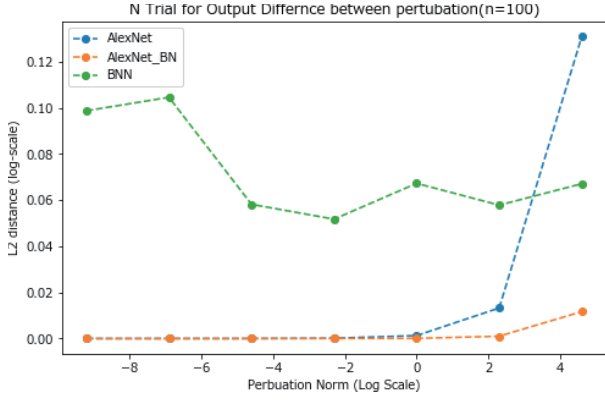


図 1: 入力にノルム γ の摂動を与えた時の出力の誤差。

は摂動に対して弱い問題が挙げられる。摂動に弱い様子を図 1 に示す。学習前の 3 種類のモデル (AlexNet, AlexNet+BatchNormalization[Ioffe 15], BNN) に MNIST の同一データにノルム γ の摂動を追加した時の追加前との誤差を示している。32 ビット浮動小数点数で表現されている深層学習モデルに比べ、BNN は摂動の大きさに関係なく誤差が生じてしまっている。そこで本研究では、この摂動に対してロバスト性を向上するための正則化項を提案する。正則化項の技術として、変分情報ボトルネック (variational information bottleneck, VIB) を導入する。

2.2 変分情報ボトルネック

変分情報ボトルネック (VIB) [Alemi 16] は、情報ボトルネック [Tishby 15] の変分下界を学習する手法である。情報ボトルネックでは、教師あり学習を表現学習としてみなす。すなわち、入力データ X をラベル Y の予測に利用可能な潜在表現 Z への確率的な写像関数を求める。

二つの確率変数 Z, Y 間における相互情報量を $I(Z; Y)$ とすると、情報ボトルネックは、次のように表現される：

$$\max I(Z; Y) \text{ subject to } I(Z; X) \leq R. \quad (2)$$

ただし、 R は、ボトルネック (=定数) である、式 (2) を非制約ラグランジュ最適化問題として書き直すと、

$$\max I(Z; Y) - \eta I(Z; X)$$

となる。情報ボトルネックは一般的に計算困難で知られているため、Alemi ら [Alemi 16] は変分法を適用し、

$$\max_{\theta, \phi, \psi} \mathbb{E}_{p(x, y) e_{\theta}(z|x)} \left[\log q_{\psi}(y|z) - \eta \log \frac{e_{\theta}(z|x)}{m_{\phi}(z)} \right] \quad (3)$$

とすることで、変分下界を導出している。ただし、 q は分類時の尤度、 $\frac{e}{m}$ は任意の潜在表現空間 m に関連した潜在表現 e の長さを罰則するレートを表している。また、 $e_{\theta}(z|x)$ は、入力 X を潜在表現 Z に変換する確率的なエンコーダを表している。さらに、 $q_{\psi}(y|z)$ は、潜在表現 Z からラベル Y を予測する変分分類器 (もしくはデコーダ) であり、 $m_{\phi}(z)$ は、変分近似分布を示している。変分情報ボトルネックを導入することで、ニューラルネットワークが敵対的事例 [Goodfellow 15] を含めた摂動に対し、頑健になることが報告されている [Alemi 18]。

2.3 BNN への変分情報ボトルネックの導入

本研究では、BNN の最終層手前の層を、変分オートエンコーダ [Kingma 13] のようにガウス分布 $\mathcal{N}$ の平均 μ と分散 σ^2 を

Activation	$-\log q_{\phi}(y z) \downarrow$	Accuracy $\uparrow$
BNN(HT)	$1.199 \pm 3.93 \times 10^{-3}$	61.592 ± 0.267
BVIB(HT)	$1.186 \pm 9.15 \times 10^{-3}$	61.646 ± 0.178
BNN(SS $_{\beta=5}$)	$1.302 \pm 5.40 \times 10^{-3}$	61.270 ± 0.410
BVIB(SS $_{\beta=5}$)	$1.290 \pm 10.82 \times 10^{-3}$	61.134 ± 0.475

表 1: CIFAR10 の実験結果。矢印は、高いスコア/低いスコアのどちらが良いのかを示している。表中の HT は HardTanh, SS は SignSwish を表す。

出力するため、2 種類の出力をするように変更する：

$$q_{\phi}(y|z) = \mathbb{E}_{z \sim e_{\theta}(z|x)} [\text{softmax}(W_1 z + b_1)]$$

$$\log e_{\theta}(z|x) = \log \mathcal{N}(z; \mu, \sigma^2 I)$$

$$\text{where } \mu = W_2 h + b_2 \quad (4)$$

$$\sigma = \text{softplus}(W_3 h + b_3)$$

$$h = \text{NN}(x)$$

ただし、 $I \in \mathbb{R}^{zdim \times zdim}$ は単位行列、 W_1, b_1 は、ソフトマックス層の全結合層、 W_2, b_2 と W_3, b_3 はそれぞれ、平均と分散を出力する全結合層のパラメータを示し、 $\text{NN}(\cdot)$ は最終層から 2 層手前までのニューラルネットワークを表す。

3. 実験

提案手法の有効性を検証するため、本実験では、CIFAR10 を用いた損失関数の値および精度の比較を行う。まず、CIFAR10 のデータセットに対して本研究では、正規化のみを行った。これは、CIFAR10 において、左右反転やランダムクロップなどのデータ拡張処理は、精度に非常に大きな影響があることが報告されているため [Graham 14]、本研究では平均 0、分散 1 への正規化のみを行った。

本実験では、深層学習モデルの基本的な構造として AlexNet[Krizhevsky 12] を用いる。ただし、AlexNet の畳み込み層部を式 (4) の NN とし、全結合層は 3 層ではなく 2 層とし、 $\mu, \sigma \in \mathbb{R}^{256}$ とする。さらに BatchNormalization を畳み込み層の直後に導入する。また活性化関数 (Hard-Tanh/SignSwish) については、Rastegari ら [Rastegari 16] が指摘するように Max プーリング層の直前に入れてしまうと出力される値がほとんど +1 になってしまうことが懸念されるため、畳み込み層の直前に挿入した。SignSwish のハイパーパラメータは、[Darabi 18] で固定値の中で最も精度が高かった $\beta = 5$ とした。バッチサイズは 128 とし、オプティマイザには Adam[Kingma 14] を採用、学習率を 10^{-3} 、エポック数を 50 とし、10 エポックごとに 0.1 倍にした。本実験の結果を表 1 に示す。なお、異なるランダムシードを 5 種類用い、その平均と標準偏差を載せた。

4. 考察・まとめ

本研究では、BNNs が入力の摂動に対して敏感であるために過学習が起きてしまう問題に対し、変分情報ボトルネックを導入することで、正則化を試みた。

表 1 に示す通り、変分情報ボトルネックを入れることで、HardTanh と SignSwish いずれの活性化関数においても、目的関数であった負の対数尤度は下回っている。精度においては

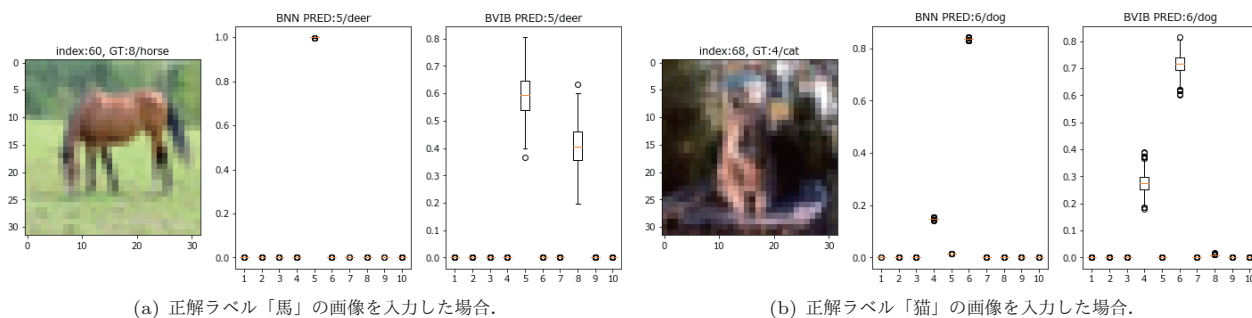


図 2: 任意の画像を入力した時の出力結果の違い. それぞれ, 入力画像, BNN, BVIB の各出力結果となっている. BVIB は変分情報ボトルネックの性質によりサンプリングが可能であるため, 出力は箱ひげ図 (サンプリング数 1000) となっている.

HardTanh を用いた場合は上回っている一方で, SignSwish を用いた場合は下回ってしまった. 目的関数と精度が一致していない理由として, 分類結果の出力の仕方があげられる. 一般的に, 分類結果の出力は, 最終層のソフトマックス層出力で最も大きい値を出力している次元のラベルとみなされる. したがって, モデルの学習がより上手く行っているとしても, つまり目的関数がより小さい値になったとしても, 最終的に最も大きい値を出力するラベルが変わらなければ精度は変わらない.

BNN と BVIB 共に不正解であった CIFAR10 データおよびモデルの出力の一例を図 2 に示す. BVIB の出力は $y \sim p(y|z)p(z|x)$ であるため, サンプリング数を 1000 とし, 結果には箱ひげ図を用いた. 図 2(a) において, 入力画像の正解ラベルが「馬」なのに対し, BNN の出力は「鹿」が最大となっており, またそのほかのラベルの値はほぼ 0 になっている. 一方で, BVIB では BNN と同様に「鹿」が最大となっているが, 「馬」のラベルが次に高い値を示しており, 依然として不正解ではあるが, BNN よりも正しく認識できていると言える. また, 図 2(b) において, 入力画像の正解ラベルが「猫」に対して, BNN と BVIB 共に一番大きい値となったラベルは「犬」であり, 二番目に大きい値となったラベルは「猫」となっている. しかし, BNN に比べると BVIB の方が「猫」ラベルの値は大きく, より正しく認識できている.

以上のことから, BNN に変分情報ボトルネックの枠組みを取り入れることで, 正則化され, 過学習が抑えられていることが確認できた. 一方で, 32 ビット浮動小数点数で表現されたニューラルネットワークに比べると依然として精度は低い. 過学習を抑えつつ, 精度を高めていくことが今後の課題である.

参考文献

- [Alemi 16] Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K.: Deep Variational Information Bottleneck (2016)
- [Alemi 18] Alemi, A. A., Fischer, I., and Dillon, J. V.: Uncertainty in the variational information bottleneck, *arXiv preprint arXiv:1807.00906* (2018)
- [Brock 18] Brock, A., Donahue, J., and Simonyan, K.: Large scale gan training for high fidelity natural image synthesis, *arXiv preprint arXiv:1809.11096* (2018)
- [Courbariaux 16] Courbariaux, M., Hubara, I., Soudry, D., El-Yaniv, R., and Bengio, Y.: Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1, *arXiv preprint arXiv:1602.02830* (2016)
- [Darabi 18] Darabi, S., Belbahri, M., Courbariaux, M., and Nia, V. P.: BNN+: Improved binary network training, *arXiv preprint arXiv:1812.11800* (2018)
- [Goodfellow 15] Goodfellow, I., Shlens, J., and Szegedy, C.: Explaining and Harnessing Adversarial Examples, in *International Conference on Learning Representations* (2015)
- [Graham 14] Graham, B.: Spatially-sparse convolutional neural networks, *arXiv preprint arXiv:1409.6070* (2014)
- [Ioffe 15] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015)
- [Kawano 17] Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T., and Nakazawa, J.: Road marking blur detection with drive recorder, in *Big Data (Big Data), 2017 IEEE International Conference on*, pp. 4092–4097 IEEE (2017)
- [Kingma 13] Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114* (2013)
- [Kingma 14] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014)
- [Krizhevsky 12] Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, pp. 1097–1105 (2012)
- [Lin 17] Lin, X., Zhao, C., and Pan, W.: Towards accurate binary convolutional neural network, in *Advances in Neural Information Processing Systems*, pp. 345–353 (2017)
- [Rastegari 16] Rastegari, M., Ordonez, V., Redmon, J., and Farhadi, A.: Xnor-net: Imagenet classification using binary convolutional neural networks, in *European Conference on Computer Vision*, pp. 525–542 Springer (2016)
- [Tishby 15] Tishby, N. and Zaslavsky, N.: Deep learning and the information bottleneck principle, in *Information Theory Workshop (ITW), 2015 IEEE*, pp. 1–5 IEEE (2015)

関連ドメインの予測モデルを活用した半教師付きドメイン適合

Semi-supervised Domain Adaptation using Prediction Models in Associated Domains

十河 泰弘 坂井 智哉
Yasuhiro Sogawa Tomoya Sakai

NEC データサイエンス研究所
NEC Corporation

Semi-supervised domain adaptation which trains a prediction model so that it adapts to novel domains from a few labeled and relatively large unlabeled observations. In this talk, we consider semi-supervised domain adaptation and propose a model embedding method. Unlike the conventional semi-supervised domain adaptation, our work utilizes prediction models in source domains. Moreover, our method can generate a pseudo label to unlabeled data without any special assumption on data distribution. Through experiments, we confirm the effectiveness of our proposed model embedding approach.

1. はじめに

人はわずかな事例を観測するだけで、自身が保有する知識や過去の経験から次に何が起こるかを予測することができる。このような特性を機械学習の分野において実現するために、近年ではドメイン適合と呼ばれる手法の研究が盛んに行われている [3, 5, 6]。ドメイン適合は、潤沢なラベル付きサンプルが存在するドメイン（元ドメイン）から得られた知識をラベル付きサンプルが限られたドメイン（目標ドメイン）に適用することで少数サンプルからでも精度が高く予測が可能なモデルを学習することを目的としており、さらにその発展として、目標ドメインでラベル付きサンプルに加えてラベル無しサンプルが存在する場合に、それを活用する半教師付きドメイン適合と呼ばれる手法が提案されている [4]。

本研究では、モデル埋め込みと呼ばれるアプローチに基づいた半教師付きドメイン適合による予測モデルの学習法を提案する。従来のドメイン適合が、元ドメインのデータや元ドメインのモデルパラメータが必要なのにに対し、提案法は元ドメインで学習済みのモデルを用い、そのモデルの特性を表すモデル属性を抽出することでドメイン適合を実現する。また、提案法でラベル無しサンプルへの近似ラベル付与をデータクラスタ性等の仮定なしに実現し、半教師付き学習による予測モデル推定を可能とする。

2. 問題設定

2.1 半教師付きドメイン適合

まず、入力となる d 次元の特徴ベクトルを $\mathbf{x} \in \mathbb{R}^d$ とし、それに対応する出力を $\mathbf{y} \in \mathcal{R}$ とする。元ドメインの数を T とし、予測モデルを構築したい対象ドメインの数を T' と表す。本研究では目標ドメインの $t' \in \{T+1, T+2, \dots, T+T'\}$ に対し、それぞれ $N_{t'}$ 個のラベル付きサンプル $\{(\mathbf{x}_i^{(t')}, y_i^{(t')}) | i = 1, \dots, N_{t'}\}$ が与えられるとする。さらに、これらのラベル付きサンプルに加え、目標ドメインそれぞれについて、 $N'_{t'}$ 個のラベル無しサンプル $\{\mathbf{x}_j | j = 1, \dots, N'\}$ が与えられているとする。一般的な半教師付きドメイン学習では、 T 個の元ドメインで得られるデータまたはモデルのパラメータを基に、 T' 個の目標ド

メインにおける予測モデル $\{f_{t'}\}_{t'=T+1}^{T+T'}$ を構築することである。

2.2 関連ドメインの予測モデル活用

本研究では、従来のドメイン適合と異なり、元ドメインにおいてデータから入出力関係を学習した T 個の予測モデル $\{h_t(\mathbf{x}) | t = 1, \dots, T\}$ が得られていることを仮定する。ここで T は元ドメインの数を、 t は各タスクのインデックスを表し、 $h_t: \mathbb{R}^d \rightarrow \mathbb{R}$ である。これらの予測モデルは任意の学習アルゴリズムによって得られたもので実運用に資するものであり、ある程度の予測精度が保証されているものとする。

本研究での目的は、元ドメインで構築された任意の予測モデル $\{h_t(\mathbf{x}) | t = 1, \dots, T\}$ と目標ドメインのデータ $\{(\mathbf{x}_i^{(t')}, y_i^{(t')}) | i = 1, \dots, N_{t'}\}$ および $\{\mathbf{x}_j | j = 1, \dots, N'\}$ を用いて、目標ドメインにおける予測モデル $\{f_{t'}\}_{t'=T+1}^{T+T'}$ を構築することである。

3. 提案手法

本節では、学習済みモデルを活用した半教師付きドメイン学習を実現するための、モデル埋め込み法を提案する。

3.1 モデル属性とモデル間共通入力表現の獲得

本研究において鍵となるアイディアは、各ドメインでの学習済みモデルの性質を表すモデル属性とそれらに対応するサンプルの変換、モデル間共通入力表現を獲得した上で、目標ドメインにおける予測モデルを構築することである。すなわち、各学習済みモデルを

$$g(\mathbf{z}_i, \mathbf{a}_t) \approx h_t(\mathbf{x}_i)$$

という形で近似することを考える。ここで、 $g: \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ 、 $\mathbf{z}_i \in \mathbb{R}^p$ は、 $\mathbf{x}_i$ に対応する p 次元の共通入力表現であり、 $\mathbf{a}_t \in \mathbb{R}^p$ はモデル h_t の特性を表す p 次元のモデル属性ベクトルである。各ドメインに起因した特性は、モデル属性によって表現され、共通入力表現は、モデル属性と合わせて予測を実現するのに用いられる。つまり、ドメイン間で共通した情報は、この共通入力表現が保持する。

まずモデル属性と共通入力表現を推定するために、ラベル付きサンプルとラベル無しサンプルを用いて g と h の近似誤

差を

$$L(\mathbf{Z}, \mathbf{Z}', \mathbf{A}) := L_U(\mathbf{Z}, \mathbf{A}) + \sum_{t'=T+1}^{T+T'} L_L^{(t')}(\mathbf{Z}', \mathbf{A})$$

と定義する。ここで、

$$L_U(\mathbf{Z}, \mathbf{A}) = \sum_{j=1}^{N'} \sum_{t=1}^T \ell(h_t(\mathbf{x}_j), g(\mathbf{z}_j, \mathbf{a}_t)),$$

$$L_L^{(t')}(\mathbf{Z}', \mathbf{A}) = \sum_{i=1}^{N_{t'}} \sum_{t=1}^T \ell(h_t(\mathbf{x}_i^{(t')}), g(\mathbf{z}_i^{(t')}, \mathbf{a}_t))$$

であり、 $\ell: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ は $\ell_S(s, t) = (s-t)^2$ のような任意の損失関数を表し、 $\mathbf{Z} := (\mathbf{z}_1, \dots, \mathbf{z}_{N'})$, $\mathbf{Z}' := (\mathbf{z}_{T+1}, \dots, \mathbf{z}_{T+T'})$, $\mathbf{Z}_t := (\mathbf{z}_1, \dots, \mathbf{z}_{N_{t'}})$, $\mathbf{A} := (\mathbf{a}_1, \dots, \mathbf{a}_T)$. このモデル近似誤差を $\mathbf{z}$ と $\mathbf{a}$ について最小化することで、モデル属性と共通入力表現を推定することができる。また、目標ドメイン t' のモデル属性については、それぞれ上記で得られた共通入力表現を用いて

$$\hat{\mathbf{a}}_{t'} = \arg \min_{\mathbf{a}_{t'}} \sum_{i=1}^{N_{t'}} (y_i^{(t')} - g(\hat{\mathbf{z}}_i^{(t')}, \mathbf{a}_{t'}))^2$$

として求めることができる。

上記の具体例として、 g を下記のモデル属性と共通入力表現の内積で表すことを考える。すなわち、

$$g(\mathbf{z}, \mathbf{a}) = \mathbf{z}^\top \mathbf{a} \quad (1)$$

としたとき、前述のモデル属性及び共通入力表現は、辞書学習 [7] と同様にして推定することができる。

3.2 モデル属性を用いた半教師付きドメイン適合

本節で、推定したモデル属性と共通入力表現を用いた半教師付きドメイン適合による目標ドメインの予測モデル $f_{t'}$ の推定法について述べる。ラベル付きサンプルのみから学習する場合、目標ドメインの予測モデル $f_{t'}$ は、 $f_{t'}$ のモデルパラメータについて下記の経験損失を最小化することで得られる。

$$R_L^{(t')}(f_{t'}) := \sum_{i=1}^{N_{t'}} \ell(y_i^{(t')}, f_{t'}(\mathbf{x}_i^{(t')})).$$

しかしながら、ドメイン適合の問題設定においてはラベル付きサンプルの数が限定されており、正確な予測モデルの学習が難しい。そこで、ラベル無しサンプルの活用を考える。一般的な半教師付き学習 [2] ではラベルなしサンプルへのラベル付与にデータのクラスタ構造等の仮定を導入する必要があるが、本研究では前節で推定したモデル属性を用いてラベル無しサンプルに対し疑似的なラベルを付与することが可能である。結果として、下記の経験損失を最小化することで、目標ドメイン t' の予測モデルを推定することができる。

$$R^{(t')}(f_{t'}) := (1 - \gamma) R_L^{(t')}(f_{t'}) + \gamma R_U(f_{t'}).$$

ここで、

$$R_U(f_{t'}) := \sum_{j=1}^{N'} (g(\hat{\mathbf{z}}_j, \hat{\mathbf{a}}_{t'}) - f_{t'}(\mathbf{x}_j))^2$$

であり、 $g(\hat{\mathbf{z}}_j, \hat{\mathbf{a}}_{t'})$ がラベル無しサンプルに対する疑似ラベルである。また、 γ はラベルありサンプルとラベル無しサンプルのトレードオフパラメータである。

表 1: 100 回試行の平均予測誤差（二乗誤差）。ここで Linear はタスク数 T に依存しないため、各 T に対し共通の値である。

	提案法	Stack	Linear
T=10	18.57	19.54	
T=20	17.50	19.67	19.33
T=40	17.45	21.43	

4. 評価実験

マルチタスク学習 [1] の評価に倣って人工データを生成し、提案手法の評価を行った。本実験では各ドメインのデータ生成に用いる線形モデルの 10 次元重みベクトルを $N(\mathbf{0}_{10}, \Sigma)$, $\Sigma := \text{Diag}\{1, 0.25, 0.1, 0.5, 2.0, 1, 0.25, 0.1, 0.5, 2.0\}$ の 10 次元正規分布から、15 次元の入力特徴ベクトルを $U(-2.5, 2.5)$ の一様分布からランダムに生成した。15 次元の入力特徴ベクトルのうち、ドメイン間で共通の 10 次元が重みベクトルと結びついて出力を与える。比較手法には、リッジ回帰 (Linear) 及びスタッキング (Stack) を用意した。スタッキングのモデルは、 $\min_{\mathbf{w}} \sum_{i=1}^{N_{t'}} (y_i - \sum_{t=1}^T w_t h_t(\mathbf{x}_i))^2$ を解くことで得られる、 $\sum_{t=1}^T w_t h_t(\mathbf{x}_i)$ である。本実験において提案法のパラメータはそれぞれ $p = 5$, $\gamma = 0.1$ とし、モデルの近似に (1) を用いた。また、提案法のモデルおよび元ドメインのモデルにはリッジ回帰モデルを用いた。最終的に 6 個のラベル付きサンプル、200 個のラベル無しサンプルを用いて、学習を行った。

表 1 に 1000 個のテストサンプルに対する各手法の予測誤差を示す。Linear に比べ、提案法がより高精度な予測モデルを学習できていることや、タスク数増加に伴って提案法の精度が改善することがわかる。これらは複数の元ドメインからうまくモデル属性と共通入力表現を獲得して、目標ドメインの学習を行っているためと考えられる。

5. おわりに

本研究では、学習済みモデルを用いたドメイン適合法の提案を行った。提案法により、元ドメインのモデルからモデル属性を抽出することで、半教師付きドメイン適合が可能となり、この有用性を人工データ実験によって確認した。

参考文献

- [1] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Convex multi-task feature learning. *Machine Learning*, Vol. 73, No. 3, pp. 243–272, 2008.
- [2] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. The MIT Press, 2006.
- [3] Hal Daume III. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 256–263, 2007.
- [4] Hal Daumé III, Abhishek Kumar, and Avishek Saha. Frustratingly easy semi-supervised domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pp. 53–59. Association for Computational Linguistics, 2010.
- [5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pp. 1126–1135, 2017.
- [6] Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 999–1006. IEEE, 2011.
- [7] Julien Mairal, Francis Bach, and Jean Ponce. Task-driven dictionary learning. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 34, No. 4, pp. 791–804, 2012.

隠れ状態を持つ多腕バンディット問題における方策の検討

A study on measures in multi-armed bandit problem with hidden state

工藤 亘平^{*1}
Kouhei Kudo

竹川 高志^{*1}
Takasi Takekawa

^{*1} 工学院大学情報学部
Faculty of Informatics Kogakuin University

Abstract: The Bandit problem is a matter of maximizing the current reward by selecting one out of the options and acquiring the reward, while limiting it to one state. Reinforcement learning is a problem of maximizing rewards earned in the future by performing various actions from options, in the presence of multiple states. The difference between the two is that state information is known, and multiple states are taken into account. In this simulation, we consider a model in which the current state and state transition information is unknown, maintaining one state for a certain period of time and then transitioning to another state. Regarding this model, we compare the general Bandit problem policy and reinforcement learning policy by cumulative reward. As a result, the cumulative reward was higher for the reinforcement learning policy than for the Bandit problem policy.

1. はじめに

近年はインターネットの普及によってインターネットショッピングやニュースの利用や、インターネット広告を目にする機会が増えている。総務省の調べによるインターネットショッピングの個人利用率は全年代平均で約 7 割を超える結果[1]となっていることから、老若男女に関係なく多くの人が利用していることがわかる。このため企業側も利益を最大化する目的で個人ごとの購買意欲などを高めるためにパーソナライズを行っている。パーソナライズは、ユーザーが過去に選択したコンテンツの利用履歴や年齢や性別などの個人情報に合わせてシステムが推薦するコンテンツを最適化する手法である。

パーソナライズに用いる手法としてバンディット問題があり、バンディット問題によるニュースのパーソナライズ化 [2]による個人の趣味嗜好にあったニュースの推薦などバンディット問題を活用したパーソナライズ化の方法が多く考えられている。バンディット問題は探索と知識利用のトレードオフを解決する問題であり、1つの状態に限定した中で、選択肢の集合から1つを選びその選択肢から報酬を獲得してその報酬から次の選択を決定することで、現在獲得できる報酬を最大化する問題である。一般的にバンディット問題では選択肢をアーム、アームを選択する戦略を方策と呼ぶ[3]。

またバンディット問題に似た問題設定を持つ手法として強化学習がある。強化学習は複数の状態が存在し全ての状態を知っている中で、選択肢の集合から様々な試行を繰り返していくことで、未来で獲得する報酬を最大化する問題である。バンディット問題と強化学習の違いは、状態の情報が既知で複数の状態を考慮しているかである。

バンディット問題でパーソナライズを行う場合はユーザー情報や過去の評価などによってコンテンツを推薦する。しかし人は気分や状況など状態の変化によって異なる行動を起こすため、同様のコンテンツを推薦していても良い評価を得られない場合がある。本論文ではこのような人の行動を現在の状態や状態遷移の情報が未知の設定で、一定時間は1つの状態を維持して、その後他の状態に遷移する対人行動モデルとして考える。この

モデルに対してバンディット問題の方策と強化学習の方策を用いてシミュレーションを行い、一般的なバンディット問題の設定と比較をした。

2. シミュレーション設定

2.1 モデル作成

人の感情や状況で行動が変化する性質を各状態で高い確率のアームが異なる設定で、毎回各状態の推移確率によって状態遷移を繰り返すが一定の時間は1つの状態を維持しその後他の状態に遷移することでアームの確率が変化するモデルとして作成する。本モデルは現在いる状態や状態数などの情報については方策側に提示せずに獲得した報酬によって方策に状態を推定させる。そして遷移確率は「初期状態」「確率分配」「状態遷移」の3つの状況がある。

「初期状態」は複数の状態が存在する中でいずれか1つの状態の推移確率が100%となる状況である。「確率分配」が起こるまでの一定時間はこの状況が続く。

「確率分配」は一定時間経過後に毎回獲得する報酬の結果によって各状態の推移確率を増減させる状況である。この状況は「状態遷移」が起こるまで続く。

「状態遷移」は増減した推移確率で他の状態へ遷移した場合の行動である。遷移した状態の推移確率を100%、他の状態の推移確率を0%にして「初期状態」を行う。

2.2 方策

今回のシミュレーションで用いる方策のバンディット問題の ϵ -グリーディ法、UCB 方策、Thompson Sampling と強化学習の Q 学習と Q-Network について説明する。

ϵ -グリーディ法は、パラメータ ϵ によって定められた回数で最適なアームを探索し、残りの回数で探索された最適なアームを選択する。

UCB 方策は、毎回各アームで報酬の平均とアームの選択回数をを用いた補正項を合わせた UCB スコアを計算して、そのスコアが最も高いアームを選択する[4]。

Thompson Sampling は、各アームの報酬の当たりとはずれの回数によって事後分布を作成し、その分布より生成される乱数が最大となるアームを選択する[5]。

連絡先: 竹川高志, 工学院大学情報学部, 163-8677 東京都
新宿区西新宿 1-24-2, 033340-0103,
takekawa@cc.kogakuin.ac.jp

Q 学習は各アームに行動価値関数 Q 値を設定して獲得した報酬で Q 値を最適化していく方法である。各 Q 値とソフトマックス関数を用いて確率的にアームを選択する。

Q-Network は、Q 学習とニューラルネットワークを合わせた学習法である。ニューラルネットワークでは行動の履歴と報酬を入力とし、各アームの報酬の期待値を出力として Q 値を教師データとして学習を行う方策である。またアームの選択はソフトマックス関数を用いて確率的に選択する。

2.3 設定

本シミュレーション用いるバンディット問題の設定としてアームに用いる分布をベルヌーイ分布、報酬は当たりで 1、はずれで 0 とする。また 1 回の試行回数を 10000 回、1 つの方策で試行を繰り返す回数 500 回とする。アーム数を 3 本とし、各アームの当たりの確率は、(0.2, 0.2, 0.75) と設定する。このときの試行回数 10000 回での最大累積報酬は 7500 となる。モデルに用いる設定として状態数は 3 として各状態で高い確率のアームが異なる設定を持ち、1 つの状態を維持する時間を 100~1000 回の範囲で乱数を用いてランダムに設定する。方策のパラメータは ϵ -グリーディ法の ϵ を 0.2, Q 学習の割引率 γ を 0.1, 温度 β を 10.0, Q-Network の履歴数を 1, 学習率 α を 0.8, 温度 β を 100.0, 割引率 γ を 0.8 とする。

3. シミュレーション結果

3.1 各方策の累積報酬

図 1 は UCB 方策の各アームの UCB スコアと Q 学習の各アームの Q 値とモデルの状態推移である。これよりバンディット問題と強化学習の方策でスコアの減少に差があることがわかる。図 2 は今回の方策に一般的なバンディット問題の設定とモデルを用いた場合での累積報酬の結果である。結果、左の一般的な場合には Thompson Sampling と UCB 方策が約 7500 と最も報酬を獲得し、Q 学習と Q-Network が約 7400 と報酬を獲得した。右のモデルを適用した場合には Q 学習と Q-Network が約 7200 と最も報酬を獲得し、UCB 方策が約 7000 と報酬を獲得した。

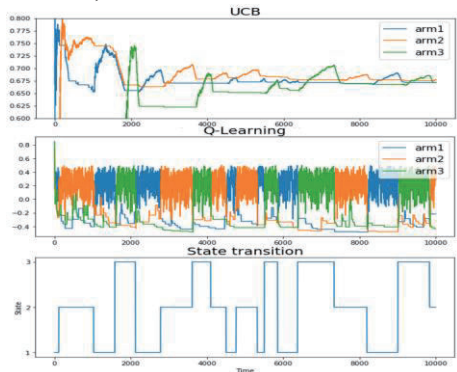


図 1 モデルの状態推移

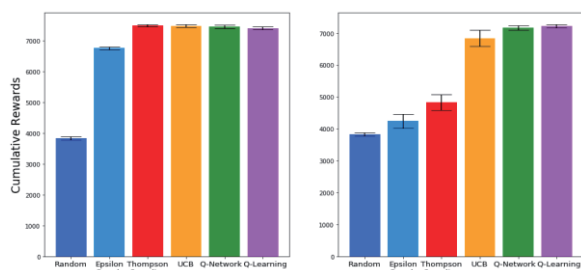


図 2 一般的な場合(左)とモデル(右)の方策の累積報酬

3.2 状態遷移回数の増加による累積報酬

今回のモデルは 1 つの状態を維持する時間が 100~1000 回と長く設定されている。この範囲を限定して状態遷移の回数を増加させた場合の各方策の累積報酬の変化をシミュレーションした。図 3 は待機する時間を変化させた場合の各方策の累積報酬である。横軸は維持する時間の最大値であり、左側ほど状態遷移回数が多い。結果、1 つの状態を維持する時間の最大が 300 回以下でどの方策も累積報酬が大きく減少した。

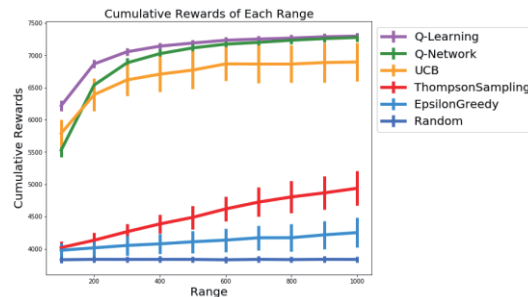


図 3 状態遷移回数の増加による各方策の累積報酬の変化

4. まとめ

一般的な設定とモデル適用時を比較すると Thompson Sampling に大きな差があった。これは Thompson Sampling が当たりの回数など報酬の履歴に影響してアームを選択しているため、変化が起これとその対応に大幅な時間がかかるためだと考えられる。その他にモデル適用時にはバンディット問題の方策よりも強化学習の方策の累積報酬が高い結果となった。これはアームの確率を変更し報酬を受け取りにくくなった場合に、Q 学習の更新式による Q 値の減少率が高く、またソフトマックス関数によって確率的にアームを選択することで他のアームを選択しやすくなるためだと考える。また待機時間を減少させていくとすべての方策で累積報酬が大きく減少した。状態遷移の回数が増加することで累積報酬の高かった方策であっても状態を推定しきれなくなっていると考えられる。

今後の課題として、状態が連続で変化していく待機時間が短い場合には高い累積報酬を獲得した方策がなかった。そのためこのような場合でも高い累積報酬が獲得できるような方策を考える必要がある。また今回のモデルは状態数や次に遷移する状態などの設定を限定した中で状態遷移を行っている。しかし、日常的問題には様々なシチュエーションや状態が存在する。よって様々な問題設定のモデルで各方策の累積報酬を明らかにしていく必要がある。以上の 2 つにおいてシミュレーションしていく。

5. 参考文献

- [1] 総務省. (2015). インターネットショッピングの利用状況. 参照先: 総務省 | 平成 27 年版 情報通信白書: http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc12_2400.html
- [2] Lihong Li, Wei Chu, John Langford, Robert E. Schapire. (2012). A Contextual-Bandit Approach to Personalized News Article Recommendation.
- [3] 本多淳也, 中村篤祥. (2016). バンディット問題の理論とアルゴリズム. 講談社.
- [4] Peter Auer, Nicolò Cesa-Bianchi, Paul Fischer. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. Machine Learning.
- [5] Olivier Chapelle, Lihong Li. (2011). An Empirical Evaluation of Thompson Sampling

運用中の予測器を使って未知のクラスを分類する方法

An Approach to Unseen Classs Classification with In-Service Predictors

坂井 智哉 十河 泰弘

Tomoya Sakai Yasuhiro Sogawa

NEC データサイエンス研究所

NEC Data Science Research Laboratories

The goal of zero-shot learning is to recognize a novel class that did not appear in training. In this talk, we introduce a novel approach to zero-shot learning. Our approach reuses in-service predictors which are often available in practice. Unlike most of the existing methods, our method does not require to replace in-service predictors with new predictors specifically designed for zero-shot learning.

1. はじめに

機械学習技術が一般的になるにつれ、様々なデータ分析ツールが利用可能になっている。これにより、従来の課題の多くが容易に解決されるようになりつつあるが、一方で、これまでの機械学習手法を単純に適用するだけでは解けない問題も認識されるようになった。その一つの例として、訓練データには存在しない未知のクラスがテスト時に現れるゼロショット学習 [1] が挙げられる。

ゼロショット学習では、既知と未知のクラスにまたがる属性と呼ばれる共通の意味情報が与えられると仮定することで、未知クラスを含めた分類を行う [1]。これまでに様々なゼロショット学習法が提案され、その有用性が示されてきた [2-6]。

一方で、従来のゼロショット学習法の多くが、ゼロショット学習を導入する以前に既知クラスに対する予測器が運用されている状況を考慮していなかった。そのため従来の方法を利用すると、既に運用中の予測器をゼロショット学習のための新しい予測器と置き換えることになり、多くの費用と時間がかかることになる。また、運用中の予測器に依存した他のシステムがあれば、その更新費用も必要となる。

本研究では、モデル再利用法に基づいたゼロショット学習法を提案する。提案法により、運用中の予測器を置き換えることなく、クラス属性を準備するだけで未知クラスを含む予測が可能となる。計算機実験により、提案法はゼロショット学習専用の予測器を一から構築しないにも関わらず、既存のゼロショット学習手法の性能に匹敵する性能が得られることを示した。

2. 予備知識

2.1 問題設定

入力を d 次元の特徴ベクトル $\mathbf{x} \in \mathbb{R}^d$, 対応するクラスラベルを $y \in \mathcal{Y}$ とする。クラス t に対する予測器を $f^{(t)}: \mathbb{R}^d \rightarrow \mathbb{R}$ とし、テストデータ点 $\mathbf{x}'$ に対する予測ラベルを $\hat{y} = \arg\max_{t' \in \mathcal{Y}} f^{(t')}(\mathbf{x}')$ で得る。

訓練およびテスト時におけるクラスラベルのドメインを、それぞれ $\mathcal{Y}_S = \{1, \dots, k_S\}$ および $\mathcal{Y}_U = \{k_S + 1, \dots, k_S + k_U\}$ とする。ここで、 k_S と k_U は、それぞれ既知および未知クラスのクラス数であり、 $\mathcal{Y}_S \cup \mathcal{Y}_U = \mathcal{Y}$ および $\mathcal{Y}_S \cap \mathcal{Y}_U = \emptyset$ である。ゼロショット学習の目標は、訓練時にあたえられる既知クラスの教師情報から、テストデータ点が既知クラスと未知クラス

を含むクラスのうち、どのクラスに属するかを予測することである。したがって、如何にして未知クラスに対する予測器の出力を得るかが鍵となる。

2.2 属性を用いた予測

ゼロショット学習を可能にする一つの方法が、属性と呼ばれる既知および未知クラスすべてに共通する意味情報を用いることである [1]。

クラス t における属性を m 次元の属性ベクトル $\mathbf{a}^{(t)} \in \mathbb{R}^m$ で表す。既知クラスおよび未知クラスに対する属性ベクトルの集合を、それぞれ $\mathcal{A}_S := \{\mathbf{a}^{(t)}\}_{t=1}^{k_S}$ と $\mathcal{A}_U := \{\mathbf{a}^{(t')}\}_{t'=k_S+1}^{k_S+k_U}$ で表す。

従来のゼロショット学習のほとんどは、特徴および属性ベクトルを入力とする予測器 $G(\mathbf{x}, \mathbf{a}): \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}$ を、属性ベクトルと既知クラスに関するラベル付きデータを用いて訓練する。予測器 G に入力する属性ベクトル $\mathbf{a}^{(t)}$ を固定すれば、 $g_{\mathbf{a}^{(t)}}(\mathbf{x}) := G(\mathbf{x}, \mathbf{a}^{(t)})$ はクラス t に対する予測器

$$f^{(t)}(\mathbf{x}) \approx g_{\mathbf{a}^{(t)}}(\mathbf{x})$$

とみなせる。ここで、 $g_{\mathbf{a}^{(t)}}$ を属性固定予測モデルと呼ぶ。

未知クラスの属性を用いることで、未知クラスに対する予測器の出力が得られるため、未知クラスを含めた分類が

$$\hat{y} = \arg\max_{t' \in \mathcal{Y}_S \cup \mathcal{Y}_U} g_{\mathbf{a}^{(t')}}(\mathbf{x}')$$

により可能となる。

2.3 運用中の予測器

本論文では、従来のゼロショット学習法と異なり、既知クラスに対する予測器

$$\{h^{(t)}: \mathbb{R}^d \rightarrow \mathbb{R}\}_{t=1}^{k_S}$$

が得られることを仮定する。

運用中の予測器が利用可能という仮定は現実においてしばしば満たされる。例えば、データ分析者が教師付き学習を用いて用意した予測器が運用されていることが十分考えられる。これらの運用中の予測器を新たなゼロショット学習用の予測器に置き換えるには、関連するシステムの更新費用や時間など多くのコストがかかるため、置き換えなしに未知クラスを含めた予測ができることが望ましい。本研究の目的は、この運用中の予測器 $\{h^{(t)}\}_{t=1}^{k_S}$ を再利用しながらゼロショット学習法を構築することである。

連絡先: {t-sakai,y-sogawa}@ah.jp.nec.com

3. 提案手法

3.1 特徴固定予測モデル

従来法の多くは、属性ベクトルを固定することで予測モデル $G(\mathbf{x}, \mathbf{a})$ を対応するクラスに対する予測器とみなした。それに対して我々は、特徴ベクトルを固定した特徴固定予測モデル $\eta_{\mathbf{x}'_j}(\mathbf{a})$ を考える。ここで、 $\mathbf{x}'_j$ は予測対象となるデータ点であり、予測ラベルは

$$\hat{y}_j = \operatorname{argmax}_{t' \in \mathcal{Y}_S \cup \mathcal{Y}_U} \eta_{\mathbf{x}'_j}(\mathbf{a}^{(t')})$$

により得られる。

3.2 モデル再利用予測

次に特徴固定予測モデルの推定法について説明する。テストデータ点には様々なパターンがあり得るため、あり得るすべてのテストデータ点に対して特徴固定予測モデルを準備することは現実的ではない。そこで、提案法は特定のテスト点に対する予測モデルをテスト時に学習する。

まず、テストデータ点 $\mathbf{x}'_j$ に対する運用中の予測器の出力 $s_j^{(t)} = h^{(t)}(\mathbf{x}'_j)$ を得る。鍵となるアイデアは、特徴固定予測モデル $\eta_{\mathbf{x}'_j}(\mathbf{a}^{(t)})$ を、その出力が運用中の予測器の出力に近くなるように訓練することである：

$$s_j^{(t)} \approx \eta_{\mathbf{x}'_j}(\mathbf{a}^{(t)}).$$

そのような $\eta_{\mathbf{x}'_j}(\mathbf{a}^{(t)})$ は、 $\{(\mathbf{a}^{(t)}, s_j^{(t)})\}_{t=1}^{k_S}$ を訓練データとして既製の回帰用ソフト（例えばリッジ回帰を行う関数）を利用することで容易に得られる。得られた予測器 $\hat{\eta}_{\mathbf{x}'_j}$ を用いることで、予測ラベル $\hat{y}_j = \operatorname{argmax}_{t' \in \mathcal{Y}_S \cup \mathcal{Y}_U} \hat{\eta}_{\mathbf{x}'_j}(\mathbf{a}^{(t')})$ が得られる。

4. 評価実験

4.1 人工データ

人工データを [3] に従い生成し、提案法の性能を評価した。我々の実験では、 $d = 10$, $m = 20$, $n^{(t)} = 50$, $k_U = 100$ とした。

運用中の予測器として、ロジスティック回帰（LR）とサポートベクトル分類器（SVC）を検討し、予測器の違いによる提案法の性能への影響も調査した。提案法における、特徴固定予測モデルの推定には ℓ_2 -正則化最小二乗法を利用した。

既知クラス数 k_S を変化したときの提案法の分類精度の変化を図 1 に示す。 $k_S < 50$ では、既知クラス数の増加に伴い分類精度が増加し、 $k_S \geq 50$ であれば、分類精度に大きな変化は見られなかった。運用中の予測器に用いられる予測モデルに関しては、LR を利用の方が平均的に少し高い分類精度が得られた。

4.2 ベンチマークデータ

次に、4 つのベンチマークデータを用いて提案手法の性能を評価した。ベンチマークデータとして、Attribute Pascal and Yahoo (aPY) [7], Animals with Attributes (AWA) [1], Caltech-UCSD-Birds 200-2011 (CUB) [8], SUN [9] を利用した。画像特徴量として、101 層の ResNet [10] から得られた 2048 次元ベクトル [6] を利用した。

表 1 に平均 Top-1 分類精度を示す。提案法はゼロショット学習に特化した分類器を一から構築しないにも関わらず、従来のゼロショット学習専用の分類器の分類精度に匹敵する精度が得られた。

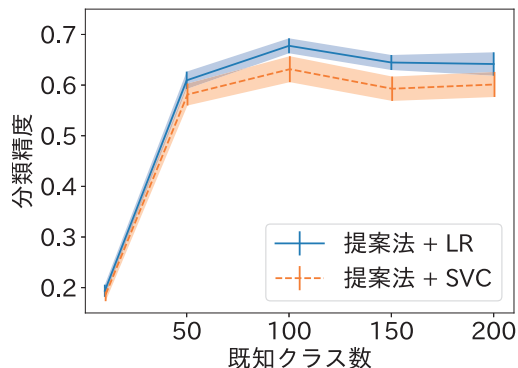


図 1: 既知クラス数 k_S を変化させたときの平均精度と標準誤差（試行回数 5 回）。

表 1: 平均 Top-1 分類精度。既存手法の結果は、既存研究 [6] より結果を引用した。

手法	aPY	AWA	SUN	CUB
ALE [2]	39.7	59.9	58.1	54.9
ESZSL [3]	38.3	58.2	54.5	53.9
SynC [4]	23.9	54.0	56.3	55.6
LatEm [5]	35.2	55.1	55.3	49.3
提案法	38.7	54.5	55.4	53.4

5. おわりに

本研究では、モデル再利用に基づいたゼロショット学習法を提案した。多くの従来法と異なり、提案法は運用中の予測器を活用して未知クラスを含めた予測ができる。実験により、提案法はゼロショット学習に特化した分類器を一から構築しないにも関わらず、従来法に匹敵する性能が得られることを確認した。

謝辞

原稿にコメントをくれた大坂直人さんに感謝します。

参考文献

- [1] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009.
- [2] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for attribute-based classification. In *CVPR*, 2013.
- [3] Bernardino Romera-Paredes and Philip Torr. An embarrassingly simple approach to zero-shot learning. In *ICML*, 2015.
- [4] Soravit Changpinyo, Wei-Lun Chao, Boqing Gong, and Fei Sha. Synthesized classifiers for zero-shot learning. In *CVPR*, 2016.
- [5] Yongqin Xian, Zeynep Akata, Gaurav Sharma, Quynh Nguyen, Matthias Hein, and Bernt Schiele. Latent embeddings for zero-shot classification. In *CVPR*, 2016.
- [6] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning - the good, the bad and the ugly. In *CVPR*, 2017.
- [7] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. Describing objects by their attributes. In *CVPR*, 2009.
- [8] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical report, California Institute of Technology, 2010.
- [9] Genevieve Patterson and James Hays. SUN attribute database: Discovering, annotating, and recognizing scene attributes. In *CVPR*, 2012.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

自己組織化マップを用いた 定常・非定常環境に適応可能な多腕バンディットアルゴリズム

Multi-armed bandit algorithm applicable to stationary and non-stationary environment using self-organizing maps

馬目 信人^{*1*2} 篠原 修二^{*2} 鈴木 康大^{*1*2} 朝長 康介^{*1*2} 光吉 俊二^{*2}
Nobuhito Manome Shuji Shinohara Kouta Suzuki Kosuke Tomonaga Shunji Mitsuyoshi

^{*1} ソフトバンクロボティクス株式会社
SoftBank Robotics Corp.

^{*2} 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

A communication robots aiming to satisfy the users facing them needs to take appropriate behavior more rapidly. However, user requests often change while these robots are determining the most appropriate behavior for these users. Therefore, it is difficult for robots to derive an appropriate behavior. Such problems are formulated as a multi-armed bandit problem. To solve this problem, we proposed a multi-armed bandit algorithm capable of adaptation to stationary and non-stationary environments using self-organizing map. In this study, numerous experiments were conducted considering a stochastic multi-armed bandit problem in both stationary and non-stationary environments. Consequently, the proposed algorithm demonstrated equivalent or improved effectiveness in stationary environments with numerous arms and consistently strong capability in non-stationary environments regardless of the number of arms in contrast with existing UCB1, UCB1-Tuned, and Thompson Sampling algorithms.

1. はじめに

対面するユーザを満足させることを目的とするコミュニケーションロボットは、ロボットの持つ多くの行動選択肢から、より早くそのユーザに適した行動を出力する必要がある。しかし現実には、ロボットがそのユーザに適した行動を模索している最中にユーザの求めるものが変化してしまうことが多々ある。そのため、ロボットの最適な行動出力は難しい。

このような問題は、多腕バンディット問題[Robbins 1952]として定式化される。多腕バンディット問題とは、ある確率で報酬が得られる腕が複数ある環境下において、逐次的に腕を選択し、得られる報酬を最大化させることを目的とする問題である。この問題は、報酬確率の高い腕を見つける探索と報酬確率が高いと確信する腕を引く知識利用のトレードオフを内在するモデルであり、強化学習の最も基本的な問題とされている[Sutton 1998]。

多腕バンディット問題における代表的なアルゴリズムとしては、 ϵ -greedy [Sutton 1998], Softmax policy [Sutton 1998], Upper-Confidence Bound (UCB) policies [Auer 2002]などが存在する。これらのアルゴリズムは、定常環境下においては良い性能を示す。しかし、非定常環境下においては良い性能を示さないことから、動的な環境に特化したアルゴリズムも提案されている[Kocsis 2006][Garivier 2011]。このように定常環境下において優れた性能を出すアルゴリズムが、非定常環境下においても優れた性能を出すとは限らず、どちらの環境下においても有用なアルゴリズムの開発は容易でない。

我々は、以前 Kohonen の自己組織化マップ (Self-Organizing Map: SOM) [Kohonen 1995]を用いた多腕バンディット問題のための新しいアルゴリズム (SOM-based Algorithm) を提案している[馬目 2019]。本稿では、定常環境下、非定常環境下両者における提案アルゴリズムの性能をみるために、多腕バンディット問題の中でも腕の報酬が腕ごとに関連付けられた確率分布に従

い与えられる確率的バンディット問題を対象として、数値実験を行い、アルゴリズムの性能を検証する。

2. 多腕バンディットアルゴリズム

本稿では、提案アルゴリズムと ϵ -greedy や Softmax policy より良い性能を示すことが知られている UCB1 [Auer 2002], UCB1-Tuned [Auer 2002], Thompson Sampling [Thompson 1933] との性能比較を行う。そこで、本章ではこの 3 つの比較アルゴリズムと提案アルゴリズムについて述べる。

2.1 UCB1

多腕バンディット問題において、得られる報酬を最大化するためには、報酬確率が高い腕を見つける探索と報酬確率が高いと確信する腕を引く知識利用のバランスが重要である。UCB1 は、この探索と知識利用のバランスが優れており、確率的バンディット問題における標準的なアルゴリズムとされている。

このアルゴリズムでは、最初に全ての腕を 1 回ずつ選択し、その後は式(1)で定義される評価値が最も高い腕 j を選択していく。

$$j = \operatorname{argmax}_i \left(\bar{X}_i + \sqrt{\frac{2 \log n}{T_i(n)}} \right) \quad (1)$$

このとき、 $\bar{X}_i$ は腕 i の期待値、 n は全ての腕の選択回数、 $T_i(n)$ は腕 i の選択回数である。

2.2 UCB1-Tuned

UCB1-Tuned は、UCB1 の改良モデルである。

このアルゴリズムでは、UCB1 同様最初に全ての腕を 1 回ずつ選択し、その後は式(2)で定義される評価値が最も高い腕 j を選択していく。

$$j = \operatorname{argmax}_i \left(\bar{X}_i + \sqrt{\frac{\log n}{T_i(n)} \min \left\{ \frac{1}{4}, V_i(T_i(n)) \right\}} \right) \quad (2)$$

連絡先: 馬目 信人, ソフトバンクロボティクス株式会社 / 東京大学大学院工学系研究科, manome@bioeng.t.u-tokyo.ac.jp

このとき、 $\bar{X}_i$ は腕 i の期待値、 n は全ての腕の選択回数、 $T_i(n)$ は腕 i の選択回数であり、 V_i は式(3)で定義される。

$$V_i(s) = \left(\frac{1}{s} \sum_{\gamma=1}^s X_{i,\gamma}^2 \right) - \bar{X}_{i,s}^2 + \sqrt{\frac{2 \log n}{s}} \quad (3)$$

このとき、 $X_{i,\gamma}$ は γ 時点での腕 i の報酬である。

2.3 Thompson Sampling

Thompson Sampling は、腕の報酬パラメータ μ_i が何らかの確率分布に従い生成されると仮定して考える確率一致法の一つである。また、大抵の場合 UCB policies に比べ有限の試行回数内で良い結果を出せることが知られている[Kaufmann 2012]。

このアルゴリズムは、各腕について、ある確率分布に従って生成される乱数 $\tilde{\mu}_i$ をサンプリングし、この値が最も高い腕を選択していく。

本稿では、各腕の報酬パラメータ μ_i がベータ分布に従うとして実装した。アルゴリズムを次に示す。

Algorithm 1: Thompson Sampling

For each arm $i = 1, \dots, N$ set $n_i = 0, m_i = 0$.

foreach $t = 1, 2, \dots$, **do**

For each arm $i = 1, \dots, N$, sample $\tilde{\mu}_i$ from the Beta($1 + n_i, 1 + m_i$) distribution.

Play arm $i = \operatorname{argmax}_{i \in \{1, \dots, N\}} \tilde{\mu}_i$ and observe reward r_t .

If $r_t = 1$, then $n_i = n_i + 1$, else $m_i = m_i + 1$.

End

2.4 提案アルゴリズム

本節では、はじめに Kohonen の最も基本的な SOM について説明する。次に SOM を用いた多腕バンディット問題のためのアルゴリズムについて説明する。

(1) SOM

SOM は、Kohonen により発明された教師なし学習を行う人工ニューラルネットワークであり、主にデータマイニングの分野でよく用いられる[Li 2018][Belkhir 2018]。SOM のアーキテクチャと学習アルゴリズムの概要を図 1 に示す。

SOM は入力層と競合層の 2 層で構成される。SOM の目的は、高次元の観測データセットを競合層の低次元空間へ写像することである。SOM の競合層は、正方格子状に分割され、それぞれにノードが配置される。また、各ノードは重みと呼ばれる観測データの次元数と同じ大きさのベクトルを持っている。SOM の学習アルゴリズムを次に示す。

Step 1. 各ノードの重みに初期値 $\mathbf{y}_k(0)$ を与える。そして、学習回数 $t = 1$ とする。

Step 2. 観測データセットから一つのデータ $\mathbf{x}$ を入力ベクトルとして選択する。

Step 3. 入力ベクトル $\mathbf{x}$ に対して、最も近い重みを持つ勝者ノード c を式(4)により決定する。

$$c = \operatorname{argmin}_k \|\mathbf{x} - \mathbf{y}_k(t-1)\|^2 \quad (4)$$

Step 4. 各ノードの重みを式(5)により更新する。

$$\mathbf{y}_k(t) = \mathbf{y}_k(t-1) + \alpha(t) h_{ck}(t) \{\mathbf{x} - \mathbf{y}_k(t-1)\} \quad (5)$$

Step 1. Initialize each node's weight.
Step 2. Choose an input vector from data set.
Step 3. Calculate the Best Matching Unit.
Step 4. Modify the weights of nodes.

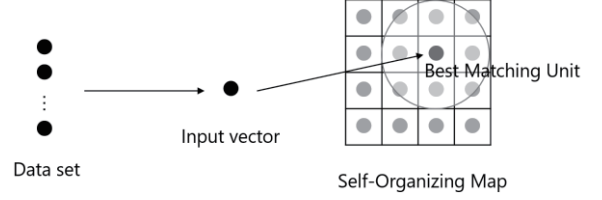


図 1. SOM のアーキテクチャと学習アルゴリズムの概要

このとき、 α は学習率、 h_{ck} は近傍関数であり、それぞれ式(6) – (8)により定義される。

$$\alpha(t) = \alpha_0 \left(1 - \frac{t}{T} \right) \quad (6)$$

$$h_{ck}(t) = \exp \left(-\frac{d_{ck}^2}{2\sigma(t)^2} \right) \quad (7)$$

$$\sigma(t) = \max \left\{ \sigma_0 \exp \left(-\frac{t}{\tau} \right), \sigma_{\min} \right\} \quad (8)$$

このとき、 α_0 は学習率の初期値、 T は学習率の縮小スピードを決める時定数、 d_{ck} は勝者ノード c と近傍ノード k のユークリッド距離である。また、 σ は近傍半径であり、 σ_0 は近傍半径の初期値、 $\sigma_{\min}$ は近傍半径の最小値、 τ は近傍半径の縮小スピードを決める時定数である。そして、学習回数 $t = t + 1$ として Step 2 に戻り処理を繰り返す。

(2) SOM-based Algorithm

提案アルゴリズムは、腕 i を選択したときの報酬を腕 i に対応する SOM^i の入力とすることで腕の報酬確率を推定し、選択する腕を逐次的に決定していくものである。アルゴリズムを次に示す。

Step 1. 腕 i に対応する SOM^i を腕の数分用意する。この際、 SOM^i の各ノードの重みの初期値 $\mathbf{y}_k^i(0)$ は、全て 1 次元ベクトルで $\mathbf{y}_k^i(0) = (1)$ とする。そして、時刻 $t = 1$ とする。

Step 2. 式(9)で定義される評価値が最も高い腕 j を選択する。

$$j = \operatorname{argmax}_i \frac{1}{N^i} \sum_{k=1}^{N^i} \|\mathbf{y}_k^i(t-1)\| \quad (9)$$

このとき、 N^i は SOM^i のノード数である。また、評価値が最大となる腕が複数あった場合、その中で腕番号 i が最も小さい腕を選択する。

Step 3. 腕 j を選択したときの報酬を確認する。得られた報酬が 1 の場合 $\mathbf{x} = (1)$ 、報酬が 0 の場合 $\mathbf{x} = (0)$ とする。

Step 4. 選択した腕 j に対応する SOM^j について、 $\mathbf{x}$ に対する勝者ノード c を式(10)により決定する。

$$c = \operatorname{argmin}_k \|\mathbf{x} - \mathbf{y}_k^j(t-1)\|^2 \quad (10)$$

このとき、勝者ノードとなる候補が複数あった場合、その中でノード番号 k が最も小さいノードを選択する。

Step 5. SOM^j について、各ノードの重みを式(11)により更新する。

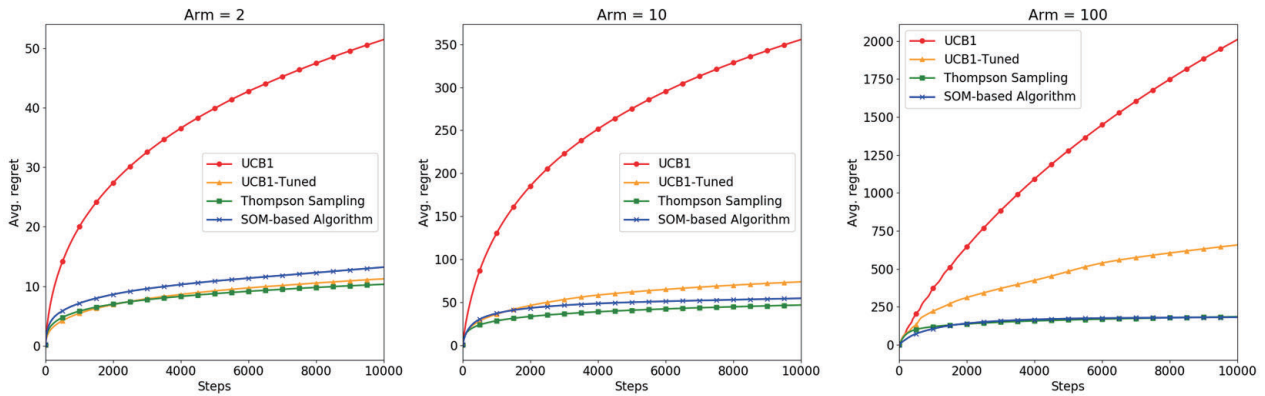


図 2. 定常環境下における UCB1, UCB1-Tuned, Thompson Sampling, 提案アルゴリズムの regret の比較結果. 左から順に腕の本数 2, 10, 100 の場合における regret の 10,000 回の平均値を示す.

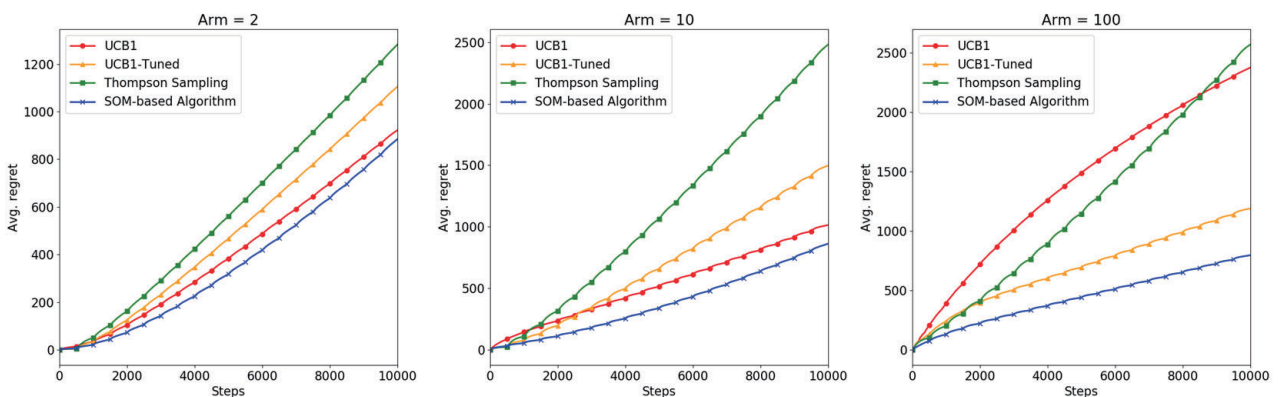


図 3. 非定常環境下における UCB1, UCB1-Tuned, Thompson Sampling, 提案アルゴリズムの regret の比較結果. 左から順に腕の本数 2, 10, 100 の場合における regret の 10,000 回の平均値を示す.

$$\mathbf{y}_k^j(t) = \mathbf{y}_k^j(t-1) + \alpha(t^j) h_{ck}(t^j) \{\mathbf{x} - \mathbf{y}_k^j(t-1)\} \quad (11)$$

このとき、 t^j は腕 j の選択回数であり、 α および h_{ck} は式 (6) – (8) より計算する。選択されなかった腕 $i (i \neq j)$ については $\mathbf{y}_k^i(t) = \mathbf{y}_k^i(t-1)$ とする。そして、時刻 $t = t+1$ として Step 2 に戻り処理を繰り返す。

3. 数値実験

本稿では、確率的バンディット問題を対象として、定常・非定常環境下それぞれについて数値実験を行った。

腕の報酬は、腕 i ごとに設定された報酬確率 P_i に基づいて決定される。プレイヤーは腕 i を選択すると確率 P_i で報酬 1.0、確率 $1 - P_i$ で報酬 0.0 を得ることができる。このとき、腕を選択する回数を step と呼び、プレイヤーが 1 step で選択できる腕は 1 本のみである。定常環境下における各腕の報酬確率 P_i は、試行ごとに $[0,1]$ 区間の一様乱数によって決定した。また、非定常環境下における各腕の報酬確率 P_i は、試行ごとに $[0,1]$ 区間の一様乱数によって決定する他、500 step ごとに全ての腕の報酬確率を $[0,1]$ 区間の一様乱数によって再編した。

アルゴリズムの性能評価には、多腕バンディット問題における代表的な評価指標である regret を用いた。regret とは、全ての step において報酬確率が最も高い腕を選択した場合と実際に選択した腕の報酬期待値の差であり、この値が小さいほど良いアルゴリズムと考える。

本実験では、アルゴリズムの性能評価のために、腕の本数 2, 10, 100 の場合について、それぞれ 10,000 step のシミュレーションを 10,000 回試行し、各 step における regret の平均値を算出した。また、提案アルゴリズムにおける SOM の競合層は、全て大きさ 10×10 、非トーラス型の正方格子とし、 $\alpha_0 = 0.1$, $T = 10,000$, $\sigma_0 = 10$, $\sigma_{min} = 5$, $\tau = 50$ とした。

4. 結果

腕の本数 2, 10, 100 の場合について、定常環境下における regret の結果を図 2 に、非定常環境下における regret の結果を図 3 に示す。

図 2 より定常環境下において、腕の本数 2 の場合については UCB1-Tuned と Thompson Sampling の regret が同じ程度で最も小さい値となった。腕の本数 10 の場合については Thompson Sampling の regret が最も小さい値となった。腕の本数 100 の場合については、Thompson Sampling と提案アルゴリズムの regret が同じ程度で最も小さい値となった。

図 3 より非定常環境下においては、腕の本数 2, 10, 100 全ての場合について提案アルゴリズムの regret が、環境が変化する 500 step 目以降最も小さい値となった。

5. 考察

結果から、提案アルゴリズムは定常環境下において、UCB1, UCB1-Tuned, Thompson Sampling と比べ、腕の本数が多い場合について同等以上の性能をあげることが読み取れる。また、

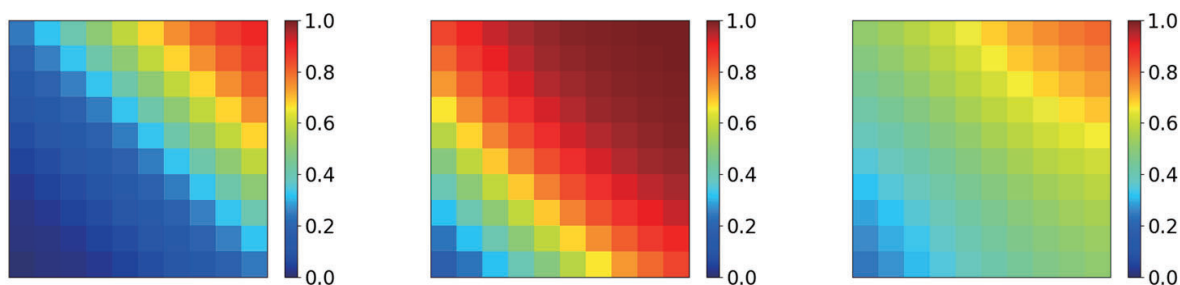


図 4. 提案アルゴリズムにおける SOM の可視化結果. 左の図は報酬 0.0, 1.0 が得られる確率がそれぞれ 0.8, 0.2 の腕を, 中央の図は報酬 0.0, 1.0 が得られる確率がそれぞれ 0.2, 0.8 の腕を, 右の図は報酬 0.0, 0.5, 1.0 が得られる確率がそれぞれ 0.2, 0.6, 0.2 の腕を 1,000 回選択した後の SOM の可視化結果を示す. 格子上的色はノードの重みを表す.

非定常環境下においては, UCB1, UCB1-Tuned, Thompson Sampling と比べ, 腕の本数に関わらず常に報酬確率の高い腕を選択できることが読み取れる.

本稿では, 提案アルゴリズムについて Kohonen の SOM をそのまま用いた. このメリットとデメリットについて述べる.

提案アルゴリズムにおけるメリットは, SOM の学習率, 近傍半径を縮小させることで有限の時間内に腕の報酬確率を効率よく推定できる点にある. 本来 SOM を用いる際は, 観測対象のデータが変動しないことを前提としている. そのため, 観測対象のデータが時々刻々と変化する場合には, 学習率, 近傍半径を縮小させない方が良い. 一方で, 提案アルゴリズムの式(11)の 2 項目から分かるように, 現時点の SOM のノードの重み(腕の評価値に該当する)と実際に得られた報酬に差があればあるほど, 大きく値を更新するため, 定常環境下のみならず非定常環境下でも優れた性能を出せたと考えられる.

また, データマイニングの観点におけるメリットも存在する. SOM はデータの特徴を可視化することに優れている. そのため, SOM を可視化することで, どの報酬がどの程度の割合で出現するかという腕の情報を分かりやすく捉えることができる. SOM の可視化結果の一例を図 4 に示す. 図 4 は, 報酬が 0.0, 1.0 の 2 値を出力する腕と報酬が 0.0, 0.5, 1.0 の 3 値を出力する腕を 1,000 回選択した後の SOM の可視化結果である. 図 4 に示すいずれの SOM においても, 直感的に, どの報酬がどの程度の割合で出現するかという腕の特徴が見て分かる.

提案アルゴリズムにおけるデメリットとしては, 腕の数と同じ数分の SOM を用意しなければならない点がある. そのため, 問題における腕の本数が 100 の場合, SOM の数も 100 用意する必要がある, メモリの消費量が多くなる. したがって, 多腕バンディット問題をスマートに解くことだけを考えると SOM の計算式のみを用いた単純なモデルを考えた方が良い. ただし, 計算速度については, 1 step につき 1 つの SOM のノードの重みを更新するのみなので, さほど遅くはならない. また, 決めなければならないパラメータが多い点も課題である.

6. まとめ

本稿では, 確率的バンディット問題を対象として, 定常・非定常環境下それぞれについて数値実験を行った. その結果, 提案アルゴリズムは既存アルゴリズム UCB1, UCB1-Tuned, Thompson Sampling と比べ, 定常環境下においては腕の本数が多い場合について同等以上の性能を, 非定常環境下においては腕の本数に関わらず常に良い性能をあげることを示した.

今後は, SOM のパラメータを変えた際の分析と SOM の計算式をベースとしたより洗練されたアルゴリズムの開発に取り組む.

参考文献

- [Auer 2002] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine Learning*, vol. 47, no. 2-3, pp. 235–256, 2002.
- [Belkhir 2018] L. Belkhir, L. Mouni, A. Tiri, T. S. Narany, and R. Nouibet, “Spatial analysis of groundwater quality using self-organizing maps,” *Groundwater for Sustainable Development*, vol. 7, pp. 121–132, 2018.
- [Garivier 2011] A. Garivier, and E. Moulines, “On UpperConfidence Bound Policies for Non-stationary Bandit Problems,” *Algorithmic Learning Theory*, pp. 174–188, 2011.
- [Kaufmann 2012] E. Kaufmann, N. Korda, and R. Munos, “Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis,” *Algorithmic Learning Theory*, pp. 199–213, 2012.
- [Kocsis 2006] L. Kocsis, and C. Szepesvári, “Discounted UCB,” In: *2nd PASCAL Challenges Workshop*, 2006.
- [Kohonen 1995] T. Kohonen, “Self-organizing Maps,” *Springer*, 1995.
- [Li 2018] T. Li, G. Sun, C. Yang, K. Liang, S. Ma, and L. Huang, “Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes,” *Science of The Total Environment*, vol. 628–629, no. 1, pp. 1446–1459, 2018.
- [Robbins 1952] H. Robbins, “Some aspects of the sequential design of experiments,” *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.
- [Sutton 1998] R. S. Sutton, and A. G. Barto, “Introduction to Reinforcement Learning,” *MIT Press*, 1998.
- [Thompson 1933] W. R. Thompson, “On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples,” *Biometrika*, vol. 25, pp. 285–294, 1933.
- [馬目 2019] 馬目 信人, 篠原 修二, 鈴木 康大, 朝長 康介, 光吉 俊二: SOM に基づく多腕バンディットアルゴリズム, in *Proceedings of the 81st National Convention of IPSJ*, 2019.

決定木と Cross-Entropy 法を用いた解釈可能な制御方策の学習

Learning Interpretable Control Policies with Decision Trees via the Cross-Entropy Method

田中 友紀子 *1*2

Yukiko TANAKA

平岡 拓也 *1*2

Takuya HIRAOKA

鶴岡 慶雅 *2*3

Yoshimasa TSURUOKA

*1 NEC セキュリティ研究所

NEC Security Research Laboratories

*2 産業技術総合研究所 人工知能研究センター

National Institute of Advanced Industrial Science and Technology, Artificial Intelligence Research Center

*3 東京大学

The University of Tokyo

Learning interpretable policies for control problems is important for industrial requirements for safety and maintenance. A common approach to acquiring interpretable policies is to learn a decision tree that imitates a black-box (e.g., neural network-based) policy trained to maximize the expected reward in a given environment. However, such approximated decision tree policies are suboptimal in the sense that they do not necessarily maximize the expected reward. In this paper, we propose a method for learning a decision tree policy that directly maximizes the reward using the cross-entropy method. Our experimental results show that our method can acquire interpretable decision tree policies that perform better than baseline policies learned by the imitation approach.

1. はじめに

解釈性のある制御方策 (以降、方策と呼ぶ) を学習することは、産業応用における安全性やメンテナンス容易性の観点で重要である [Lipton 18]。人間が方策を解釈することができれば、実際に方策に従って操作を実行する前に、操作の安全性を確認することができる。また、操作の過程で問題が起きた場合でも、解釈性のある方策は人の手によって修正することができる。

深層強化学習が多くの制御タスクで人間を介さずに高い性能の方策を獲得することに成功した [Silver 17, Andrychowicz 18] 一方で、これらの方策はニューラルネットワークに基づいているため、一般に解釈性が低い。そのような問題に対して、方策の学習に決定木のような解釈性の高いモデルを使う方法が存在する。Sammur らは、熟練操作員の操作ログから、教師あり学習を用いて決定木に基づく方策を構築している [Sammur 92]。Verma らは、ニューラルネットワークでモデル化した方策を模倣するように解釈性のある方策を学習している [Verma 18]。Liu らは、Q-Learning で得られたブラックボックスの Q 関数を近似することで、決定木でモデル化された Q 関数を構築している [Liu 18]。これらの模倣に基づく手法における主要な問題点のひとつに、学習済みのブラックボックス方策や Q 関数を近似するように学習した決定木方策は、モデルの複雑度の差によって、元のブラックボックス方策や Q 関数よりも性能が低下することが挙げられる。他方、Hein らは解釈性のある木構造の方策を、ブラックボックス方策の模倣を介さずに、遺伝プログラミングを用いて期待報酬を直接最大化して学習している [Hein 18] が、このアプローチでは方策パラメータを離散化して扱う必要があり、そのことが学習で得た方策の性能低下を引き起こす可能性がある。

本研究では、Cross-Entropy 法 (CEM) [Szita 06] を用い、

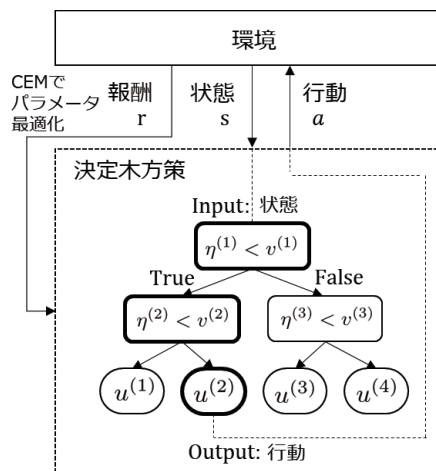


図 1: 決定木方策の学習

環境から与えられる期待報酬を直接最大化することで、解釈性のある決定木方策を獲得する手法を提案する (図 1)。本論文では、まず CEM により決定木方策が直接最適化できるよう、決定木方策のパラメータ表現を定義する。そして、我々の手法を評価するため、複数の制御タスクで実験をする。実験の結果より、ニューラルネットワークに基づく方策のふるまいを模倣することで決定木方策を構築するベースライン手法よりも我々の手法がよい性能となることを示す。さらに、我々の手法で獲得した方策について考察し、それらがどのような解釈性をもつかを示す。

2. 検証準備

2.1 強化学習

強化学習の目標は、期待累積報酬を最大化するエージェントの方策を学習することである [Sutton 98]。一般に、エー

連絡先: 田中 友紀子, NEC セキュリティ研究所, y-tanaka@jz.jp.nec.com

エージェントが臨む問題は、マルコフ決定過程 (Markov Decision Processes; MDPs) で表される。MDPs では、各時間ステップにおいて、エージェントが環境から状態 $s_t \in S$ を観測し、環境に行動 $a_t \in A$ を返す。ここで、 S は状態空間で、 A は行動空間である。次に環境が次の s_{t+1} に遷移し、エージェントが報酬 $r_{t+1} \in \mathbb{R}$ を受け取る。エージェントは方策 π に従い、 s_{t+1} における a_{t+1} を決定する。本問題の目的は、エピソード周期 T における期待累積報酬 $R(\pi^*) = E_{\pi^*} \left[\sum_{t=1}^T r_t \right]$ を最大化する最適な方策 π^* を獲得することである。なお、本論文では、feature 関数 $\phi(s)$ を用い、状態 s は F 次元ベクトル表現 $[\phi^{(1)}(s), \dots, \phi^{(F)}(s)]$ に変換されると仮定する。

2.2 Cross-Entropy 法

本研究では、最適方策の獲得に CEM を用いる。具体的には、パラメータ $\theta \in \mathbb{R}^d$ で方策 π をパラメータ化し (パラメタライズされた方策は π_θ と表記)、 $\theta^* = \arg \max_{\theta} R(\pi_\theta)$ とする最適パラメータ値 θ^* をみつけるために CEM を使う。

CEM では、1) パラメータサンプル生成、2) パラメータサンプル評価、3) パラメータ生成確率分布更新、を繰り返すことで θ^* を探索する。ステップ 1 では、 b 個のパラメータサンプル $\theta_1, \dots, \theta_b$ が確率分布から生成される。本研究では、確率分布として、平均ベクトル μ と対角共分散行列 Σ をもつ多変量ガウス分布 $N(\mu, \Sigma)$ を用いる。ここで、 $\sigma^{(i,i)}$ を、 Σ の i 番目の対角成分と定義する。ステップ 2 では、 $\theta_1, \dots, \theta_b$ がそれぞれに従う方策を実行することで得られる累積報酬 $R(\pi_{\theta_1}), \dots, R(\pi_{\theta_b})$ に基づいて評価される。そして、高い累積報酬を得た上位 m 個のサンプルが $\theta_1, \dots, \theta_b$ から選択される。これらの選ばれたサンプル集合を M と定義する。ステップ 3 では、サンプル生成用の確率分布が M によって更新される。本研究では、前記多変量ガウス分布の平均ベクトルは次のように更新される。

$$\mu = \frac{1}{m} \sum_{\theta' \in M} \theta'. \quad (1)$$

また、 Σ の i 番目の対角成分は次のように更新される^{*1}。

$$\sigma^{(i,i)} = \frac{1}{m} \sum_{\theta' \in M} (\theta^{(i)'} - \mu^{(i)})^2 + \epsilon, \quad (2)$$

ここで、 $\theta^{(i)'}$ は θ' の i 番目の要素であり、 $\mu^{(i)}$ は μ の i 番目の要素である。また、 ϵ は学習早期における収束を防ぐための探索ノイズの定数である。このような手順を通し、CEM は θ^* を探索する。

3. 決定木方策の学習

この章では、決定木方策の学習手法について述べる。始めに決定木方策のパラメータ表現について説明し、次に CEM を用いたパラメータ更新手順について示す。

3.1 パラメータ表現

この節では、決定木方策のパラメータ表現について定義する。本研究では、図 2 の右側に示すような二分木構造をもつ決定木を方策の表現として用いることとする。この方策において、非終端ノードは条件を表し、葉ノードは選択される行動を表す。また、ノード間のリンクは条件の評価結果を表す。この

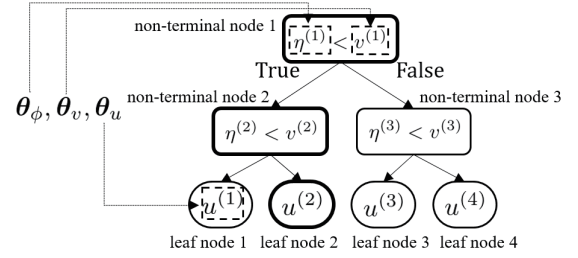


図 2: 決定木方策のパラメータ化

方策は、ある状態が与えられたとき、木の根ノードから葉ノードまでを辿ってひとつの行動を選択する。ここで、 n を非終端ノードのインデックス番号として表記し、非終端ノードの数を l 、決定木の最大の深さを D と表記する^{*2}。

我々は、図 2 の左側に示すように、 $\theta_\phi, \theta_v, \theta_u$ で決定木方策をパラメータ化する。 θ_ϕ は各非終端ノードで評価される $\phi(s)$ の要素を決定するために用いられる。 θ_ϕ は $(2^D - 1) \times F$ 行列として表現される:

$$\theta_\phi = \begin{bmatrix} \theta_\phi^{(1,1)} & \dots & \theta_\phi^{(1,f)} & \dots & \theta_\phi^{(1,F)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_\phi^{(n,1)} & \dots & \theta_\phi^{(n,f)} & \dots & \theta_\phi^{(n,F)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_\phi^{(2^D-1,1)} & \dots & \theta_\phi^{(2^D-1,f)} & \dots & \theta_\phi^{(2^D-1,F)} \end{bmatrix}, \quad (3)$$

ここで、 $\theta_\phi^{(n,f)}$ は n 番目の非終端ノードでの評価における状態ベクトル $\phi(s)$ の f 番目の要素に対応する値である。 n 番目の非終端ノードにおいて、評価に用いられる状態ベクトルの要素 $\eta^{(n)}$ は、 θ_ϕ を用いて次のように決定される: 始めに、 θ_ϕ の n 番目の行にある全ての要素を取り出す (これらの要素を $\{\theta_\phi^{(n,1)}, \dots, \theta_\phi^{(n,F)}\}$ とする)。次に、 $\{\theta_\phi^{(n,1)}, \dots, \theta_\phi^{(n,F)}\}$ の中で最も値が高い要素を選択し、対応する feature の要素を $\eta^{(n)}$ として用いる。例えば、 $\{\theta_\phi^{(n,1)}, \dots, \theta_\phi^{(n,F)}\}$ から $\theta_\phi^{(n,f')}$ が選択された場合、 $\eta^{(n)}$ は $\eta^{(n)} := \phi^{(f')}$ と決定される。

θ_v は各非終端ノードにおける閾値を決定するために用いられる。 θ_v は $(2^D - 1)$ 次元ベクトルで表現される:

$$\theta_v = [\theta_v^{(1)}, \dots, \theta_v^{(n)}, \dots, \theta_v^{(2^D-1)}], \quad (4)$$

ここで $\theta_v^{(n)}$ は n 番目の非終端ノードにおける閾値 $v^{(n)}$ を表す。 $v^{(n)}$ は $v^{(n)} = (v_{max}^{(n)} - v_{min}^{(n)})g(\theta_v^{(n)}) + v_{min}^{(n)}$ で算出される。ここで、 $v_{max}^{(n)}$ と $v_{min}^{(n)}$ は、 $v^{(n)}$ の上限値と下限値である。また、 g はクリッピングを行うための関数であり、本研究では sigmoid 関数を用いる。

θ_u は葉ノードにおける行動を決定するために用いられる。我々は、離散行動問題と連続行動問題それぞれのための θ_u の表現を提案する。

離散行動問題を想定した場合、 θ_u は $2^D \times |A|$ 行列として

*1 ただし、ステップ 3 において、 Σ のすべての対角成分は式 (2) で更新される

*2 つまり、非終端ノードの総数は $(2^D - 1)$ 、葉ノードの総数は 2^D である

表される:

$$\theta_u = \begin{bmatrix} \theta_u^{(1,1)} & \dots & \theta_u^{(1,k)} & \dots & \theta_u^{(1,|A|)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_u^{(l,1)} & \dots & \theta_u^{(l,k)} & \dots & \theta_u^{(l,|A|)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_u^{(2^D,1)} & \dots & \theta_u^{(2^D,k)} & \dots & \theta_u^{(2^D,|A|)} \end{bmatrix}, \quad (5)$$

ここで $\theta_u^{(l,k)}$ は l 番目の葉ノードにおける k 番目の行動候補 $a^{(k)} \in A$ に対応する値を表す。方策実行時において、 $\{\theta_u^{(1,1)}, \dots, \theta_u^{(l,|A|)}\}$ のうち最も値の大きい要素に対応する行動候補が l 番目の葉ノードにおける行動 $u^{(l)}$ として選ばれる。

連続行動問題を想定した場合、 θ_u は 2^D 次元ベクトルで表される:

$$\theta_u = [\theta_u^{(1)}, \dots, \theta_u^{(l)}, \dots, \theta_u^{(2^D)}], \quad (6)$$

ここで $\theta_u^{(l)}$ は l 番目の葉ノードにおける行動 $u^{(l)} \in \mathbb{R}$ に対応する値を表す。 $u^{(l)}$ は $u^{(l)} = \alpha h(\theta_u^{(l)})$ で算出される。ここで関数 h はクリッピングを行うための関数であり、本研究では $\tanh$ 関数を用いる。また、 α は $u^{(l)}$ を適切な範囲に収めるための係数である。

3.2 パラメータ更新

前節で説明した決定木方策のパラメータ (すなわち、 $\theta_\phi, \theta_v, \theta_u$) を報酬を最大化するように更新する。パラメータの更新手順を、Algorithm 1 に示す。ここでは、 $\theta_\phi, \theta_v, \theta_u$ のすべての要素を結合して θ として扱い、CEM を適用する。

Algorithm 1 Cross-Entropy 法による決定木方策の学習

Require: initial mean vector μ_0 , initial diagonal covariance matrix Σ_0 , number of elite m , constant exploration noise ϵ

- 1: $\mu \leftarrow \mu_0$
- 2: $\Sigma \leftarrow \Sigma_0$
- 3: **for** each update **do**
- 4: Generate b parameter samples $\theta_1, \theta_2, \dots, \theta_b$ from $N(\mu, \Sigma)$
- 5: Evaluate each parameter samples using $R(\pi_{\theta_1}), \dots, R(\pi_{\theta_b})$
- 6: Select m parameter samples with the highest evaluations.
- 7: Update μ by Equation (1)
- 8: Update all diagonal elements in Σ by Equation (2)
- 9: **end for**

4. 実験

4.1 タスク

今回、2つの制御タスクにて手法を評価する。タスク設定のサマリを表 1 に示す。

CartPole: ボールが倒れないよう、台車をコントロールするタスクである。今回、OpenAI Gym [Brockman 16] の「CartPole-v0」環境を用いる。状態空間は連続で 4 次元、行動空間は 2 つの離散行動 (台車に加える力の方向) である。

Pendulum: トルクをコントロールし、振り子を倒立状態に維持するタスクである。今回、OpenAI Gym の「Pendulum-v0」環境を用いる。状態空間は連続で 3 次元、行動空間は 1 次元の連続値 (トルク) である。

表 1: タスク設定

タスク	F	行動
CartPole	4	離散、 $ A = 2$
Pendulum	3	連続、 $a \in \mathbb{R}$

表 2: ベースライン手法と提案手法の平均累積報酬

タスク	D	ベースライン	提案
CartPole	2	13.2	200.0
(NN: 200.0)	3	24.4	200.0
	4	10.55	200.0
Pendulum	2	-1091.4	-848.96
(NN: -342.5)	3	-1331.1	-290.98
	4	-1219.94	-542.57

CartPole、Pendulum の両タスクにおいて、各エピソードの初期状態はランダムに初期化される。

4.2 実験設定

ベースライン手法としてブラックボックス方策のふるまいを模倣する決定木を用いる [Sammur 92, Verma 18, Liu 18]。まず始めに、ニューラルネットワークに基づく方策 (ブラックボックス方策) を PPO [Schulman 17] を用いて学習する。次に、ニューラルネットワークに基づく方策を用いて 20 エピソード分の試行をし、訓練用のサンプル (状態と行動のペア) を集める。最後に、訓練用のサンプルを用いて、教師あり学習で決定木方策を構築する。なお、ニューラルネットワークは 64 ノードの隠れ層 2 層を用いる。また、PPO は 環境との相互作用を 10^6 回通して学習される。

また、提案手法として Algorithm 1 を用いて決定木を構築する。ハイパーパラメータには、 $b = 20$, $m = 4$, $\epsilon = 0.0001$ を用いる。

5. 評価

表 2 に定量的な評価結果を示す。今回、20 エピソードの平均累積報酬に基づいて前記手法を評価した。ベースライン手法は、5 つの乱数シードを用いた試行のうち、学習終了時の方策の成績 (平均累積報酬) がもっともよい方策を用い、決定木を構築した。提案手法は、5 つの乱数シードを用いた試行のうち、学習時における方策の成績 (平均累積報酬) がもっともよい更新回の方策を評価に用いた。なお、表 2 に示す評価値は、学習時とは異なる新たな 20 エピソード分の試行を用いて算出したものである。表 2 内の D は決定木の最大の深さである。また、表 2 の各タスク名の下に「NN」として記載している手法は、PPO で学習したニューラルネットワークに基づく方策を用いたものである。表 2 の結果より、我々の手法はベースライン手法の成績を上回っていることがわかる。

決定木方策の解釈性についての定性的な評価をする。図 3 に、Pendulum タスクで学習した決定木方策の可視化例を示す。方策がある状態を環境から受け取った場合、図 3 に示されるような決定木方策の各非終端ノードでの条件分岐結果に従い、次の時間ステップで取るべき行動を選択する。非終端ノードでの条件文は、振り子の位置や速度のような物理量を用いた表現で構成されており、タスクを達成する方法 (どのように振り子を倒立した状態に維持するか) が人間からみて理解しやすい形式となっている。また、図 4 に CartPole タスクで学習した決定木方策の例を可視化した図を示す。図 4 では、2 番目の

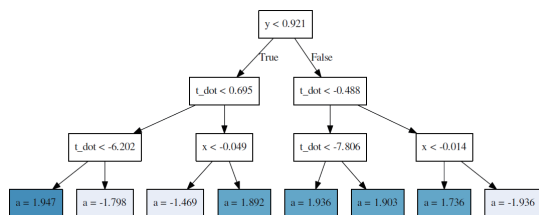


図 3: Pendulum タスクにおける Cross-Entropy 法を用いた決定木方策 $D = 3$.

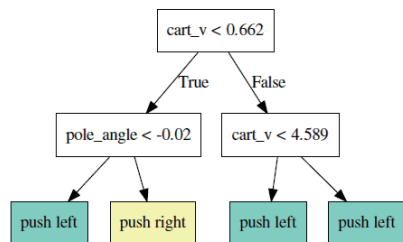


図 4: CartPole タスクにおける Cross-Entropy 法を用いた決定木方策 $D = 2$.

深さの右に位置する非終端ノードの条件「 $\text{cart_v} < 4.589$ 」の分岐結果はいずれも「push left」となっている。このような結果は、図 5 のようなより簡潔な表現にすることができる。図 5 の例も、図 3 と同様に車の速度やボールの傾き角のような物理量を用いて方策が表現されており、人間からみて理解しやすい形で可視化されている。本研究では、Pendulum、CartPole の両タスクで図 3、図 5 のような比較的浅い木構造で学習結果を確認することができた。このような決定木方策は、数百のノードを含むニューラルネットワークに基づく方策と比較し、学習で得られた方策の確認に要する時間や労力が軽減される。

6. まとめ

今回、CEM を用いて報酬を直接最大化する、解釈性のある決定木方策の獲得手法を提案した。2 つの単純な制御タスクにおける実験結果から、ニューラルネットワークに基づく方策を模倣するベースライン手法より我々の手法がよい性能となることが分かった。将来的には、よい性能を出すためにより深い木構造が必要となるような、より複雑なタスクでの検証をすることが望ましい。深い決定木において、最適な方策パラメータを求めることは難しいことが予想される。今後は、解釈性を保ちつつ複雑なタスクに適用できる手法に拡張する必要がある。

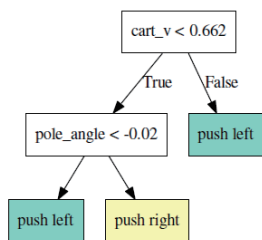


図 5: 簡潔な表現にした決定木方策

参考文献

- [Andrychowicz 18] Andrychowicz, O. M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J. W., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L., and Zaremba, W.: Learning Dexterous In-Hand Manipulation, Vol. arXiv:1808.00177, (2018)
- [Brockman 16] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W.: OpenAI Gym, Vol. arXiv:1606.01540, (2016)
- [Hein 18] Hein, D., Udluft, S., and Runkler, T. A.: Interpretable Policies for Reinforcement Learning by Genetic Programming, *Engineering Applications of Artificial Intelligence*, Vol. 76, pp. 158–169 (2018)
- [Lipton 18] Lipton, Z. C.: The Mythos of Model Interpretability, *Queue*, Vol. 16, No. 3, pp. 30:31–30:57 (2018)
- [Liu 18] Liu, G., Schulte, O., Zhu, W., and Li, Q.: Toward Interpretable Deep Reinforcement Learning with Linear Model U-Trees, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 414–429, Springer (2018)
- [Sammur 92] Sammut, C., Hurst, S., Kedzier, D., and Michie, D.: Learning to fly, in *Machine Learning Proceedings 1992*, pp. 385–393, Elsevier (1992)
- [Schulman 17] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O.: Proximal Policy Optimization Algorithms, Vol. arXiv:1707.06347, (2017)
- [Silver 17] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al.: Mastering the game of Go without human knowledge, *Nature*, Vol. 550, No. 7676, p. 354 (2017)
- [Sutton 98] Sutton, R. S. and Barto, A. G.: *Introduction to Reinforcement Learning*, MIT Press, Cambridge, MA, USA, 1st edition (1998)
- [Szita 06] Szita, I. and Lörincz, A.: Learning Tetris using the noisy cross-entropy method, *Neural computation*, Vol. 18, No. 12, pp. 2936–2941 (2006)
- [Verma 18] Verma, A., Murali, V., Singh, R., Kohli, P., and Chaudhuri, S.: Programmatically Interpretable Reinforcement Learning, in *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80 of PMLR, pp. 5045–5054, Stockholmssan, Stockholm Sweden (2018)

ディープラーニングによるくずし字認識組み込みシステムの開発

Development of Embedded System for Recognizing *Kuzushiji* by Deep Learning

竹内正広^{*1} 早坂太一^{*1} 大野亙^{*1} 加藤弓枝^{*2} 山本和明^{*3} 石間衛^{*4} 石川徹也^{*4}
Masahiro Takeuchi, Taichi Hayasaka, Wataru Ohone, Yumie Kato, Kazuaki Yamamoto, Mamoru Ishima, and Tetsuya Ishikawa

^{*1} 豊田工業高等専門学校 ^{*2} 鶴見大学 ^{*3} 国文学研究資料館 ^{*4} TRC-ADEAC 株式会社
National Institute of Technology, Toyota College, Tsurumi University, National Institute of Japanese Literature, and TRC-ADEAC, Inc.

Abstract: Effective utilization of “Pre-modern Japanese book database” constructed by the project supervised by Center for Collaborative Research on Pre-Modern Texts, National Institute of Japanese Literature, will push forward the development of the inter-field study. It may become an obstacle for the researchers with a little knowledge of classical literature, however, because historical Japanese texts have been written by *Kuzushiji* (*Hentaigana* and cursive kanji). In this article, we report an attempt of recognizing *Kuzushiji* by deep learning. Using Raspberry Pi which is the low-cost small single board computer, we developed the embedded system with enough recognizing rate and speed. In addition, manifold images of *Kuzushiji* written in a text image of the spread can be extracted. Our system will be effective for school children or elderly people not possessing mobile devices who want to come in touch with *Kuzushiji*.

1. はじめに

国文学研究資料館により平成 26 年度より開始された「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」[1]では、研究基盤整備として約 30 万点の歴史的典籍を画像データ化し、既存の書誌情報データと統合させたデータベースの構築を行っている。これを有効活用できれば、人文科学のみならず、自然科学系分野を融合させた研究の展開も期待される。しかしながら多くの研究者にとっては、それらに書かれている文字が「くずし字」であることが障壁となる。

本研究では、ディープラーニングを用いたくずし字の自動翻刻の試みを行っている。本稿では、既に WWW アプリケーションとして実装されている学習モデル[2]を応用し、小型で比較的安価なシングルボードコンピュータ Raspberry Pi を用いて開発されたくずし字認識組み込みシステムについて報告する。

2. システム概要

図 1 に本研究で開発したシステムの概要を示す。Raspberry Pi 用カメラモジュール Picamera から対象の古典籍画像を撮影する。本研究で用いるディープラーニングにより学習されたモデルは、くずし字 1 文字に対して認識を行うため、Python3 および

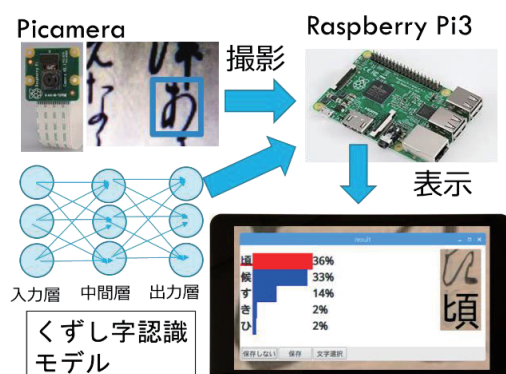


図 1 システムの概要図

連絡先: 早坂太一, 豊田工業高等専門学校 情報工学科,
〒471-8525 愛知県豊田市栄生町 2-1, TEL(0565)36-
5861, FAX(0565)36-5926, hayasaka@toyota-ct.ac.jp

OpenCV3 を用いて認識する文字の選択, 画像処理などを行う。認識は機械学習用ライブラリ TensorFlow Lite を用いて行われ、結果は 7 インチのタッチディスプレイに表示される。

2.1 くずし字領域の検出

古典籍画像に対する前処理を OpenCV3 によって行う。1 文字の選択は、タッチスクリーンに指を触れてから離すまでの x-y 座標を取得し、文字領域を 62×62 ピクセルにリサイズし、2 値化を行った後に、学習モデルに入力する。しかしながら、1 文字ずつ選択して認識する方法では、時間と手間がかかるため、OpenCV3 のモルフォロジ処理によって、古典籍画像から自動で 1 文字ずつの領域を検出することを試みた。

まず入力画像に対して、図 2(a)に示すように縦方向に対して膨張処理を行い、上下の字をつなげることで、1 行ごとの集まりを作成し、領域検出を行う。次に、検出した領域範囲を用いて、図 2(b)に示すように元の入力画像を横方向に膨張処理を行い、1 行の中から 1 文字の領域を検出する。図 2(c)に検出例を示す。

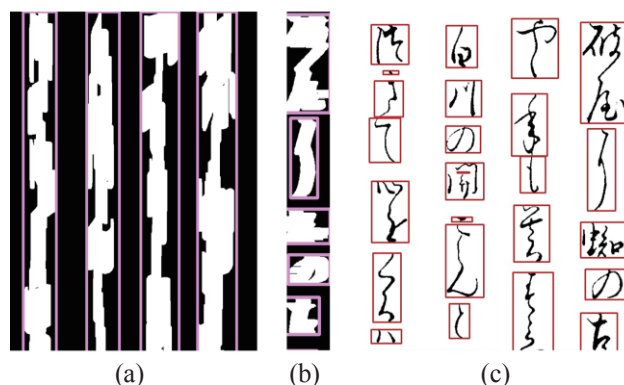


図 2 文字領域の検出過程と結果例

2.2 くずし字認識モデル[2]

本研究では、ディープラーニングで学習したくずし字認識モデルを用いる。このモデルは、4 層の畳み込み層と 3 層の全結合層から構成される畳み込みニューラルネットワークである。このモデルに対して、ディープラーニング用ライブラリ Caffe を用

いて学習を行った。学習には 3,985 種類の 418,831 文字(平均 105 文字/種類)のくずし字画像データ[3]を用いた。そのうち、変体仮名(48 文字+濁点・半濁点)は 75 種類の 262,547 文字(平均 3,501 文字/種類)、その他の漢字・カタカナ・記号などは 3,910 種類の 156,284 文字(平均 40 文字/種類)である。このモデルは、テストデータの源氏物語絵巻「桐壺」[4]11,286 文字(変体仮名 10,026 文字、その他 1,260 文字)に対して、正解率 92.4%(変体仮名 95.2%, その他 70.1%)を得ている。

2.3 モデルの量子化

本研究では当初、Raspberry Pi で学習モデルを実行させるために、代表的な機械学習ライブラリである TensorFlow を使用した[5]。しかしながら、学習モデルのサイズは 298.2MB、認識時間は約 3.0 秒と、モデルの読み込みおよび認識にある程度の時間がかかった。そこで、モバイル端末や組み込み機器向けの TensorFlow Lite を使用した。TensorFlow Lite では、学習済みモデルの重みなどを浮動小数点形式から固定小数点形式に量子化し、モデルサイズを小さくすることで、能力がやや劣る機器での演算の高速化、および読み込みにかかる時間を短縮することができる。本研究では、入力のみ 32bit の浮動小数点形式とし、他のパラメータは 8bit の固定小数点形式に量子化した。その結果、先行研究と同精度のまま、モデルサイズは 78.4MB に軽量化され、1 文字あたりの認識にかかる時間は約 0.4 秒と、TensorFlow による認識時間から約 1/8 に短縮することができた。

3. 実行例

開発したくずし字認識システムの表示画面を図 3 および図 4 に示す。開発したシステムでは、GUI アプリケーションから、カメラ撮影、ファイル読み込み、文字検出、切り取り、自動認識のそれぞれのボタンを操作することで処理を行うことができる。

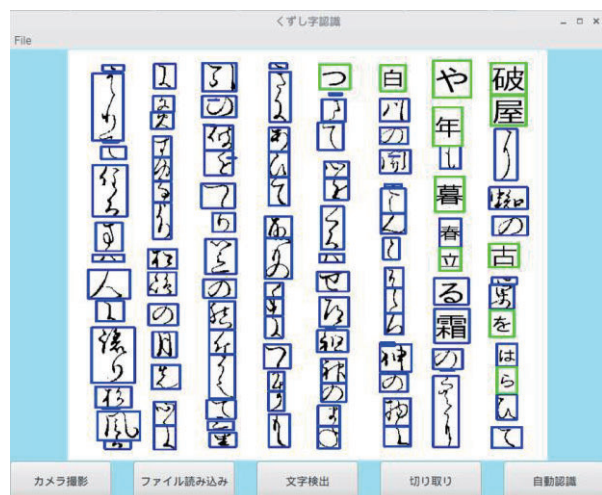


図 3 開発したシステムの実行例

ディスプレイの背面に設置した Raspberry Pi 用カメラモジュールは、焦点距離を約 30cm とし、3 秒間のプレビュー後に古典籍画像を撮影するよう設定した。既に保存済みの古典籍画像データ、もしくは保存を行った過去の画像データのどちらかを選択して、ファイル読み込みを行うこともできる。認識後は図 4 に示すように結果を表示し、認識結果の保存と元画像のくずし字を認識結果に置き換えることができる。その後、認識した文字を再びタッチすることで結果を何度でも参照することが可能であり、利用者が候

補となる 5 文字の中から認識結果を選択することが可能である。認識結果はテキストデータとして保存され、画像と同時に読み込むことで、いつでも認識結果を参照することができる。

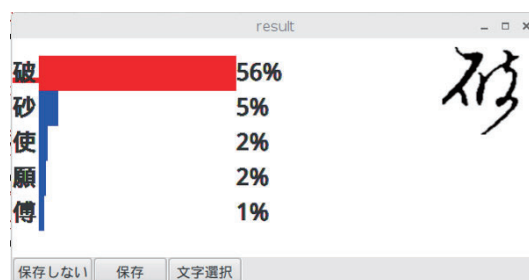


図 4 認識結果の表示例

また、ディスプレイを 2 本指でタッチすることで、図 3 に示すように、認識を行った文字は緑色枠で、文字検出を行った結果の領域は青色枠で囲まれて表示される。検出した領域をタッチすることで認識が行われ、結果が表示される。任意の領域の切り取りを行いたい場合は、ディスプレイからタッチおよびドラッグにより選択し、その後に表示される認識ボタンを押すと認識を行うことができる。ボタンにより再度選択を行うことも可能である。

4. おわりに

本研究では、Raspberry Pi を用い、カメラで撮影した画像などからくずし字の認識を可能とするシステムを開発した。量子化ニューラルネットワークを用いたところ、1 文字を認識するのにかかる時間や精度は先行研究とほぼ変わらず、高性能なコンピュータを用いなくても、Raspberry Pi のようなシングルボードコンピュータで問題なく動作することを示すことができた。

スマートフォンを持ち込むことができない小・中学校や、普段モバイル機器を持ち歩かない高齢者の方々でも、くずし字に触れたいという場面に遭遇することは少なくないと考えられる。ネットワーク環境のない場所も含めて、そうした現場で支援ツールとして活躍することができるよう、本システムの操作性や認識精度をより高めることが今後の課題である。

謝辞

本研究は JSPS 科研費 JP16K02433 および TRC-ADEAC 株式会社からの助成を受けたものです。

参考文献

- [1] 国文学研究資料館：歴史的典籍に関する大型プロジェクト, <https://www.nijl.ac.jp/pages/cijproject/>, 2015 年 10 月 14 日参照。
- [2] 早坂太一 他：2017 年度人工知能学会全国大会(第 31 回)論文集, 3Q1-2in1, 2017.
- [3] 人文学オープンデータ共同利用センター：日本古典籍字形データセット(国文研所蔵・CODH 加工), <http://codh.rois.ac.jp/char-shape/>, 2016 年 12 月 8 日参照。
- [4] 人文学オープンデータ共同利用センター：日本古典籍データセット(国文研所蔵), 源氏物語, <http://codh.rois.ac.jp/pmjt/book/200003803/>, 2017 年 2 月 15 日参照。
- [5] 竹内正広 他：第 17 回情報科学技術フォーラム講演論文集, vol.4, pp.293-294, 2018.

Residual Networks に対する確率的正則化の提案： Shake-ResDrop と Shake-SENet

Stochastic Regularization for Residual Networks: Shake-ResDrop and Shake-SENet

白濱 淳也 川本 一彦
Junya Shirahama Kazuhiko Kamoto

千葉大学
Chiba University

Recently, residual networks (ResNets) and their improvements, such as stochastic regularization, have proven to be able to reduce overfitting during training processes. In this paper, we propose two stochastic models which combines stochastic regularization and attention mechanism. The two models are based on ShakeDrop, combining either SENet or Stochastic Depth with ShakeDrop itself. Both of our methods were able to improve existing ShakeDrop results on CIFAR-100.

1. はじめに

深層学習を用いた画像認識タスクでは、畳み込みニューラルネットワーク (CNN) が主流となっている。Residual networks (ResNet) [He et al. 2016] は、ショートカット接続をもつ CNN で、2015 年に 150 層以上という大幅な深層化を実現した。ResNet は深層化に欠かせないため、現在でも広く利用されている。最近では、ResNet の汎化能力を向上させるために確率的な正則化が提案されている。Dropout [Srivastava et al. 2014] のようにニューロンではなく、ResNet の層を確率的に落とす Stochastic Depth (ResDrop) [Haung et al. 2016] や、外乱を加えデータ拡張効果をもたらす Shake-Shake [Gastaldi et al. 2017] などの確率的手法が提案されている。さらに、ResDrop のように層を落とすのではなく Shake-Shake のように外乱を加えるか加えないか確率的に変化させ、ShakeDrop [Yamada et al. 2018] も提案されている。ShakeDrop は CIFAR-100 を用いた画像分類で当時の最高精度を記録した。さらに、ResNet にアテンション機構を導入した SENet [Hu et al. 2018] も提案されている。

本研究では、このようなアテンション機構や確率的正則化のモデルを組み合わせることでより高性能な確率的正則化を提案する。

2. 提案手法

既存の確率的正則化手法である ShakeDrop を軸に ResDrop と SENet を組み合わせたアーキテクチャを提案する。

2.1 ShakeDrop と ResDrop

ShakeDrop と ResDrop を組み合わせる。ShakeDrop に層を落とす機構を加えるため、図 1 のように、residual モジュール、ショートカット接続 (出力に 0 をかけるため)、出力に乱数 $\alpha \in [-1, 1]$ をかける ShakeDrop モジュール、の 3 つの構造が学習中に現れることになる。その結果、ShakeDrop よりも強いアンサンブル学習の効果を得ることが期待できる。さらに、ショートカット接続が加わることで、パラメータを減らすこともでき、学習時間を短縮することもできる。また、ResDrop や ShakeDrop と同様に層が深くなるにつれて、residual モジュールが現れる確率を低くし、残りの 2 つの確率

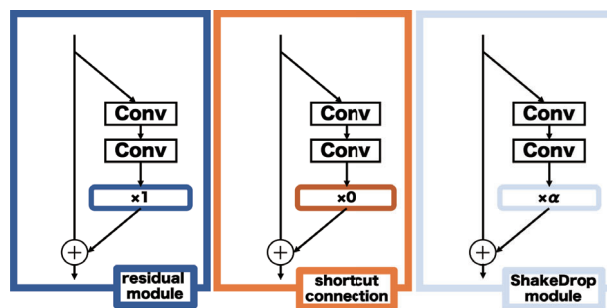


図 1: ShakeDrop と ResDrop を組み合わせて現れる 3 つの構造

が高くなるようにする。この確率の変化は、ネットワークが 5 層のとき図 2 に示すようになる。

2.2 ShakeDrop と SENet

この 2 つを組み合わせる方法としては図 3 に示すようになり、ShakeDrop させた後に SE Block を入れるか (図 3 左)、前に入れるかとなる (図 3 右)。

ShakeDrop の後に加えた場合は ShakeDrop の特徴空間でのデータ拡張と SENet のアテンション機構を組み合わせることができる。この組み合わせでは、図 3 左に示すように、畳み込み層の出力に $\alpha \in [-1, 1]$ をかけるため符号反転も起こる。学習を乱しすぎることになると考えられる。乱数をかけない場合 (residual モジュールになる) は SENet の本来のアテンション機構を生み出すことができるため、後に加えた場合も精度向上が見込まれる。

ShakeDrop の前に加えた場合は ShakeDrop のでのデータ拡張に SENet のアテンション機構を組み合わせることができる。ShakeDrop の乱数と同時に、畳み込み層からの出力を SE Block の重みでスケールするため、チャンネルのアテンションを含むデータ拡張を行うことができる。乱数が負を含むことによってチャンネルのアテンションの符号反転を入れることもできるため、負の値での強い摂動を更に与えられると考えられる [Yamada et al. 2018]。また前に加える場合後に加えた場合と同様に、乱数をかけないときは SENet を本来の形で与えられる。

連絡先: 川本一彦, 千葉大学大学院工学研究院, 千葉市稲毛区
弥生町 1-33, kawa@faculty.chiba-u.jp

表 1: 実験設定

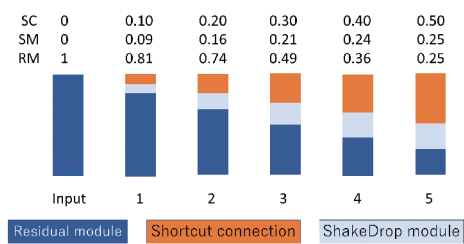


図 2: 5 層である場合の 3 種類の現れ方

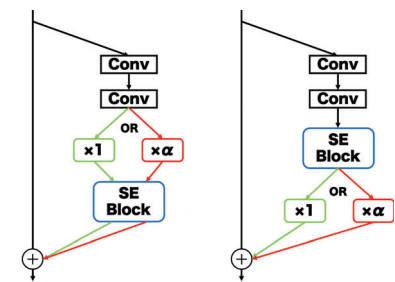


図 3: 左 : ShakeDrop の後 右 : ShakeDrop の前

3. 実験結果と考察

一般画像認識データセットの CIFAR-100 を用いて誤識別率で比較する。CIFAR-100 は 100 クラス各 600 枚、合計 60000 枚の画像を抽出したデータセットで、各クラス 500 枚を学習用データ、100 枚を評価用データとする。実験の設定を表 1 に記す。

表 2 に実験結果を示す。表 2 より、ShakeDrop 単独の手法よりも、3 つの提案手法で誤識別率は低くなっている。ResDrop と組み合わせでは、アンサンブル学習の効果を強めることができ性能向上できたと考えている。さらに、パラメータ数を減らすことができるため学習時間の減少も確認できる。ShakeDrop の前に SENet を組み込む手法が最も低い誤識別率を達成している。このことは、チャンネルのアテンションを含むデータ拡張をすることが性能向上に貢献することを示している。ShakeDrop の後に SENet を組み込む手法も、ShakeDrop 単独に比べて誤識別率は低くなっており、アテンション機構を単純に確率的正則化に組み合わせても有効であることを示している。

4. おわりに

本研究では、確率的正則化手法の ShakeDrop に、確率的正則化手法の ResDrop やアテンション機構の SENet を組み合わせることの有効性を従来手法と比較することにより実証した。今後は検証した確率的正則化方法とアテンション機構を同時に組み合わせることを検討している。

謝辞

本研究は JSPS 科研費 JP16K00231 の助成を受けたものです。

表 2: CIFAR-100 に対する実験結果

手法	誤識別率 (%)	学習時間 (h)
ShakeDrop	13.16	66.8
ShakeDrop and ResDrop	12.76	63.0
ShakeDrop after SENet	12.42	77.9
ShakeDrop before SENet	12.83	76.8

参考文献

[He et al. 2016] K.He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. CVPR, pp.770-778, 2016.

[Srivastava et al 2014] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” J. of Machine Learning, vol.15, pp.1929-1958, 2014.

[Haung et al. 2016] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Weinberger, “Deep networks with stochastic depth,” arXiv preprint arXiv:1603.09382v3, 2016.

[Gastaldi et al. 2017] X. Gastaldi, “Shake-shake regularization,” arXivpreprint arXiv:1705.07485v2, 2017.

[Yamada et al. 2018] Y. Yamada, M. Iwamura, and K. Kise. “Shakedrop regularization,” arXiv preprint arXiv:1802.02375, 2018.

[Hu et al. 2018] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu. “Squeeze-and-excitation networks,” Proc. CVPR, pp.7132-7141, 2018.

[Zoung et al 2016] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” arXiv preprint arXiv:1708.04896, 2017.

[Loshchilov et al. 2017] I. Loshchilov , F. Hutter. “stochastic gradient descent with restarts.” arXiv preprint arXiv:1608.03983, 2016.

Fairness-aware Edit of Thresholds in a Learned Decision Tree Using a Mixed Integer Programming Formulation

Kentaro Kanamori Hiroki Arimura

Graduate School of Information Science and Technology, Hokkaido University

Fairness in machine learning is an emerging topic in recent years. In this paper, we propose a post-processing method for editing a given decision tree to be fair according to a specified discrimination criterion by modifying its branching thresholds in internal nodes. We propose a mixed integer linear programming (MIP) formulation for the problem, which can deal with several other constraints flexibly and can be solved efficiently by any existing solver. By experiments, we confirm the effectiveness of our approach by comparing existing post-processing methods.

1. Introduction

Background and Motivation In the application of machine learning models to the actual decision making, problems other than their prediction accuracy, such as *interpretability* [10] and *fairness* [7], attract increasing attention. If their prediction results are unexplainable or discriminative, they are no longer usable in the actual decision making, even if they achieve high accuracy.

In this paper, we focus on *decision tree models* [4], and study a post-processing method [7] for decision trees. More specifically, we consider a problem of editing a decision tree by modifying its branching thresholds in internal nodes so as to satisfy fairness constraints w.r.t. a sensitive attribute such as gender and race, named *re-thresholding problem*. We formulate it as a *mixed integer linear programming (MIP)* problem, which we can obtain an optimal solution by powerful off-the-shelf solvers such as CPLEX*¹ and Gurobi*². Our approach has the following advantages:

- **Interpretability:** Decision tree models are known as one of the interpretable machine learning models since their predictions are based on a set of rules that human can understand easily [10, 11].
- **Adaptivity:** Post-processing methods can deal with the situation that a sensitive attribute or fairness criterion is given after learning [6, 7, 9]. In the actual decision making, they are not always given in advance.
- **Flexibility:** Our MIP formulation can deal with additional constraints defined by users without implementing designated algorithms, if these constraints can be expressed as linear equations or inequalities [2, 3, 11].

Contribution Our contributions are as follows:

1. We formulate a post-processing problem of editing a given decision tree so as to satisfy fairness constraints by modifying its branching thresholds, named *re-thresholding problem*, as an MIP problem.

2. We formulate an edit distance [12] of a decision tree to measure dissimilarity between given and modified decision trees as edit limitation constraints.
3. By experiments on real datasets, we confirm the effectiveness of our proposed method by comparing other existing post-processing methods.

2. Preliminary

2.1 Notation

For $n \in \mathbb{N}$, we denote by $[n] = \{1, \dots, n\}$. For a proposition ψ , $\mathbb{I}[\psi]$ denotes the indicator of ψ , i.e., $\mathbb{I}[\psi] = 1$ if ψ is true, and $\mathbb{I}[\psi] = 0$ if ψ is false.

In this paper, we consider a binary classification problem, and assume its input space is normalized to $[0, 1]^D$ without loss of generality. Let a pair of an input and an output $(x, y) \in [0, 1]^D \times \{0, 1\}$ be an *example*, and $S = \{(x^{(j)}, y^{(j)})\}_{j=1}^N$ be a *dataset* with N examples. For a *prediction model* $h : [0, 1]^D \rightarrow \{0, 1\}$, the *empirical loss* on S is defined by $L(h | S) := \frac{1}{N} \sum_{j=1}^N \mathbb{I}[h(x^{(j)}) \neq y^{(j)}]$.

In addition, we consider a *sensitive attribute* $z \in \{0, 1\}$, such as gender and race. Let $z^{(j)}$ be the sensitive attribute value w.r.t. j -th example $(x^{(j)}, y^{(j)}) \in S$, and $Z = \{z^{(j)}\}_{j=1}^N$ be the set of its values w.r.t. S .

2.2 Decision trees

The *decision tree* [4] is a prediction model that consists of a set of prediction rules expressed by the full binary ordered tree structure. It makes the prediction according to the label of the leaf node that the input x reaches, and the corresponding leaf node is determined by traversing the tree from the root. Each internal node has a pair of parameters $(d, b) \in [D] \times [0, 1]$, where d is a *branching feature* and b is a *branching threshold*, and the input $x = (x_1, \dots, x_D)$ is directed to one of two child nodes depending on whether the statement $x_d \leq b$ is true or not.

Then, the decision tree can be expressed as follows:

$$h(x) = \sum_{k=1}^K l_k \prod_{m \in a_k^{(L)}} \mathbb{I}[x_{d_m} \leq b_m] \prod_{m \in a_k^{(R)}} \mathbb{I}[x_{d_m} > b_m],$$

Contact: kanamori@ist.hokudai.ac.jp

*1 <https://www.ibm.com/analytics/cplex-optimizer>

*2 <http://www.gurobi.com/>

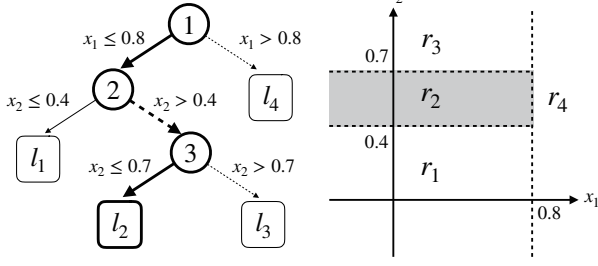


Figure 1: An illustration of a decision tree. For 2-nd leaf node, $a_2^{(L)} = \{1, 3\}$, $a_2^{(R)} = \{2\}$, and its corresponding region $r_2 = (-\infty, 0.8] \times (0.4, 0.7]$.

where $K \in \mathbb{N}$ is the total number of leaf nodes, $l_k \in \{0, 1\}$ is a predictive label of k -th leaf node, $(d_m, b_m) \in [D] \times [0, 1]$ is a branching rule in m -th internal node, and $a_k^{(L)}$ ($a_k^{(R)}$) $\subseteq [K-1]$ is a set of left(right)-branching internal node indexes on the pass from the root node to k -th leaf node. Note that a decision tree with K leaf nodes has $K-1$ internal nodes since it is expressed as the full binary ordered tree structure. We denote a region corresponding to k -th leaf node by $r_k \subseteq [0, 1]^D$. Then, a set of the regions $\{r_k\}_{k=1}^K$ expresses a partition of the input space. Figure 1 illustrates an example of a decision tree.

We denote a set of all possible decision trees by $\mathcal{H}$. In this paper, since we consider the problem of modifying branching thresholds $B := (b_1, \dots, b_{K-1}) \in [0, 1]^{K-1}$ of a given decision tree, we denote a decision tree with B by $h_B \in \mathcal{H}$.

2.3 Discrimination scores

We use two major criteria named *demographic parity*(DP) [5] and *equal opportunity*(EO) [8] to evaluate the discrimination of the model. We denote an empirical probability on a dataset S and sensitive attribute Z by $\hat{P}$.

Definition 1 (DP score) DP score of a model h on a dataset S w.r.t. a sensitive attribute Z is defined by

$$\delta_{DP}(h | S, Z) := |\hat{P}(h(x) = 1 | z = 1) - \hat{P}(h(x) = 1 | z = 0)|.$$

Definition 2 (EO score) EO score of a model h on a dataset S w.r.t. a sensitive attribute Z is defined by

$$\delta_{EO}(h | S, Z) := |\hat{P}(h(x) = 1 | y = 1, z = 1) - \hat{P}(h(x) = 1 | y = 1, z = 0)|.$$

In this paper, we call DP and EO scores *discrimination score* together. These values approach 1 as the model h tends to make the predictions unfairly for z , while they approach 0 if the model makes the predictions fairly.

2.4 Problem formulation

Here, we define our problem named *re-thresholding problem*. We assume that a decision tree h_B with branching thresholds B is already given, and the goal is to reduce the discrimination score on a given dataset S by modifying branching thresholds in B without changing the given decision tree significantly.

Problem 1 (Re-thresholding) Given a dataset S , sensitive attribute Z , decision tree h_B with branching thresholds $B = (b_1, \dots, b_{K-1})$, discrimination score $\text{disc} \in \{\text{DP}, \text{EO}\}$, discrimination threshold $t \in [0, 1]$, dissimilarity measure of decision trees $\Delta : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$, and parameters $\lambda \geq 0$ and $N_{\text{minsup}} \in [|S|]$, re-thresholding problem is defined as follows:

$$\underset{\theta \in [0, 1]^{K-1}}{\text{minimize}} \quad L(h_\theta | S) + \lambda \cdot \Delta(h_B, h_\theta)$$

$$\text{subject to} \quad \delta_{\text{disc}}(h_\theta | S, Z) \leq t$$

$$\forall k \in [K] : |\{(x, y) \in S \mid x \in r_k\}| \geq N_{\text{minsup}}$$

Note that we modify only branching thresholds, and do not modify branching features and predictive labels in leaf nodes. As a dissimilarity measure of decision trees Δ in Problem 1, we propose an *edit distance of a decision tree* Δ_{ED} , which will be defined in the next section.

3. Proposed Method

Our formulation is based on *OCT*, the MIP framework for learning optimal classification trees proposed by Bertsimas and Dunn [3]. In order to adapt it to our editing problem for decision trees, we formulate an edit distance of decision trees and fairness constraints.

3.1 Program variables

For $m \in [K-1]$, $j \in [N]$, $k \in [K]$, we introduce some variables for formulating Problem 1 as follows:

- $\theta_m \in [0, 1]$ is a modified branching threshold in m -th internal node.
- $\phi_{j,k} \in \{0, 1\}$ indicates whether j -th input $x^{(j)}$ reaches k -th leaf node, i.e., $\phi_{j,k} = \mathbb{I}[x^{(j)} \in r_k]$.
- $\psi_k \in \{0, 1\}$ indicates whether some example reaches k -th leaf node, i.e., $\psi_k = \mathbb{I}[\exists j \in [N] : x^{(j)} \in r_k]$.
- $\epsilon_m^{(L)}$ ($\epsilon_m^{(R)}$) $\in \{0, 1\}$ indicates whether no example reaches leaf nodes in the left (right) subtree of m -th internal node $c_m^{(L)}$ ($c_m^{(R)}$) $\subseteq [K]$, i.e., $\epsilon_m^{(L)} = \mathbb{I}[\forall k \in c_m^{(L)} : \psi_k = 0]$ ($\epsilon_m^{(R)} = \mathbb{I}[\forall k \in c_m^{(R)} : \psi_k = 0]$).
- $\xi_m \in [0, 2]$ expresses the cost corresponding to edit operations for m -th internal node.

$\phi_{j,k}$, ψ_k , $\epsilon_m^{(L)}$, $\epsilon_m^{(R)}$ and ξ_m are auxiliary variables. The total number of the variables is $\mathcal{O}(NK)$.

3.2 Edit distance of decision trees

We define an *edit distance of a decision tree* based on the standard tree edit distance [12]. For two ordered trees T and T' , the tree edit distance between them is defined as the minimal length of the sequence of editing operations to transform T into T' . Available edit operations are insertion, deletion, and relabeling.

In our problem, an explicit edit operation is relabeling threshold values in internal nodes. We define $|b_m - \theta_m| \in$

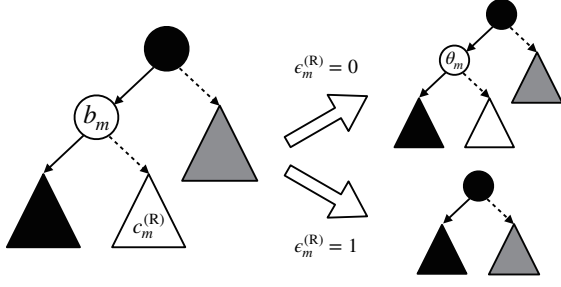


Figure 2: An example of deleting and relabeling operations. If no example reaches leaf nodes in the right subtree of m -th internal node ($\epsilon_m^{(R)} = 1$), m -th internal node and its right subtree are deleted from the tree, and the total number of these edit operations are measured as an edit distance. Otherwise ($\epsilon_m^{(R)} = 0$), $|b_m - \theta_m|$ occurs as a cost for the relabeling operation.

$[0, 1]$ as the cost of the relabeling operation for m -th internal node. However, when a modified branching threshold exceeds a specific value, no examples in S reaches its child leaf node, which is denoted by $\psi_k = 0$. By deletion of all leaf nodes with $\psi_k = 0$ and their parent internal nodes in the decision tree h_θ , we can obtain a unique tree structure whose prediction results on S are equivalent to h_θ .

Then, we define our edit distance Δ_{ED} between h_θ and h_B by the total number of these delete operations and relabeling costs. For m -th internal node, if no example reaches leaf nodes in its left or right subtree, i.e., $\epsilon_m^{(L)} = 1$ or $\epsilon_m^{(R)} = 1$, the editing cost ξ_m is 2 because m -th internal node and its subtree are deleted. Note that if $\epsilon_m^{(L)} = 1$ ($\epsilon_m^{(R)} = 1$), $\xi_{m'} = 2$ holds for any m' -th internal nodes included in left (right) subtree of m -th internal node because $\psi_k = 0$ for any $k \in c_m^{(L)}$ ($c_m^{(R)}$) and $\epsilon_{m'}^{(L)} = 1$ or $\epsilon_{m'}^{(R)} = 1$ hold, and sum of these $\xi_{m'}$ is equivalent to the cost of delete operations for the subtree. It can be expressed as follows:

$$\Delta_{ED}(h_B, h_\theta) = \sum_{m=1}^{K-1} \max\{|b_m - \theta_m|, 2\epsilon_m^{(L)}, 2\epsilon_m^{(R)}\}.$$

Figure 2 shows an example of edit operations in our problem. By using ξ_m , ψ_m , $\epsilon_m^{(L)}$, and $\epsilon_m^{(R)}$, it can be expressed as a linear function and constraints.

3.3 Fairness constraints

We can express both DP and EO scores by using variables $\phi_{j,k}$ as follows:

$$\delta_{disc}(h_\theta | S, Z) = \sum_{k=1}^K l_k \sum_{j=1}^N d_{j,k}^{(disc)} \phi_{j,k}.$$

$d_{j,k}^{(disc)}$ is a constant value determined automatically when S and Z are given such that

$$d_{j,k}^{(disc)} = \begin{cases} \frac{z^{(j)}}{|S_1^{(DP)}|} - \frac{1-z^{(j)}}{|S_0^{(DP)}|} & (disc = DP), \\ \frac{y^{(j)}z^{(j)}}{|S_1^{(EO)}|} - \frac{y^{(j)}(1-z^{(j)})}{|S_0^{(EO)}|} & (disc = EO), \end{cases}$$

where $S_z^{(DP)} := \{(x^{(j)}, y^{(j)}) \in S \mid z^{(j)} = z\}$ and $S_z^{(EO)} := \{(x^{(j)}, y^{(j)}) \in S \mid y^{(j)} = 1 \wedge z^{(j)} = z\}$ for $z \in \{0, 1\}$.

3.4 Overall formulation

Now, we can formulate Problem 1 as the following MIP problem:

$$\text{minimize } \frac{1}{N} \sum_{k=1}^K \sum_{j=1}^N c_{j,k} \phi_{j,k} + \lambda \sum_{m=1}^{K-1} \xi_m \quad (1)$$

$$\text{subject to } \sum_{k=1}^K \phi_{j,k} = 1, \forall j \in [N] \quad (2)$$

$$\phi_{j,k} \leq \psi_k, \forall k \in [K], j \in [N] \quad (3)$$

$$\sum_{j=1}^N \phi_{j,k} \geq N_{\text{minsup}} \cdot \psi_k, \forall k \in [K] \quad (4)$$

$$x_{d_m}^{(j)} \leq \theta_m + (1 - \phi_{j,k}), \quad \forall k \in [K], j \in [N], m \in a_k^{(L)} \quad (5)$$

$$x_{d_m}^{(j)} - e_m \geq \theta_m - (1 + e_{\max})(1 - \phi_{j,k}), \quad \forall k \in [K], j \in [N], m \in a_k^{(R)} \quad (6)$$

$$- \xi_m \leq b_m - \theta_m, \forall m \in [K-1] \quad (7)$$

$$b_m - \theta_m \leq \xi_m, \forall m \in [K-1] \quad (8)$$

$$\xi_m \geq 2\epsilon_m^{(L)}, \forall m \in [K-1] \quad (9)$$

$$\xi_m \geq 2\epsilon_m^{(R)}, \forall m \in [K-1] \quad (10)$$

$$1 - \epsilon_m^{(L)} \leq \sum_{k \in c_m^{(L)}} \psi_k, \forall m \in [K-1] \quad (11)$$

$$\sum_{k \in c_m^{(L)}} \psi_k \leq (1 - \epsilon_m^{(L)})|c_m^{(L)}|, \forall m \in [K-1] \quad (12)$$

$$1 - \epsilon_m^{(R)} \leq \sum_{k \in c_m^{(R)}} \psi_k, \forall m \in [K-1] \quad (13)$$

$$\sum_{k \in c_m^{(R)}} \psi_k \leq (1 - \epsilon_m^{(R)})|c_m^{(R)}|, \forall m \in [K-1] \quad (14)$$

$$\sum_{k=1}^K l_k \sum_{j=1}^N d_{j,k}^{(disc)} \phi_{j,k} \leq t \quad (15)$$

$$- \sum_{k=1}^K l_k \sum_{j=1}^N d_{j,k}^{(disc)} \phi_{j,k} \leq t \quad (16)$$

where $e_m = \min\{|x_{d_m}^{(i)} - x_{d_m}^{(j)}| \mid i, j \in [N], x_{d_m}^{(i)} \neq x_{d_m}^{(j)}\}$, $e_{\max} = \max\{e_m\}_{m=1}^{K-1}$, and $c_{j,k} := l_k(1 - y^{(j)}) + (1 - l_k)y^{(j)}$. We can express $L(h_\theta | S)$ and $\Delta_{ED}(h_B, h_\theta)$ by $\sum_{k=1}^K \sum_{j=1}^N c_{j,k} \phi_{j,k}$ and $\sum_{m=1}^{K-1} \xi_m$, respectively. Equation (2) and inequalities (3-6) are the same constraints with OCT [3]. Inequalities (7-15) and (16,17) are constraints for expressing our edit distance $\Delta_{ED}(h_B, h_\theta)$ and fairness constraint $\delta_{disc}(h_\theta | S, Z) \leq t$.

4. Experiments

Experimental setup We used the COMPAS dataset [1]. It has $N = 6172$ examples and the total number of features is $D = 9$. Its output label $y^{(j)} \in \{0, 1\}$ indicates whether j -th person recidivates within two years. We use the attribute "African.American" as its sensitive attribute $z^{(j)} \in \{0, 1\}$.

Table 1: Experimental results averaged over 5 trials. "Loss" and "DP" denote the average empirical loss with its standard deviation and DP score for training and test datasets. "time" denotes the average running times.

method	Training		Test		time[s]
	Loss	DP	Loss	DP	
NB-Modified	0.387 ± 0.022	0.103	0.384 ± 0.005	0.118	2.028
Relabeling	0.412 ± 0.018	0.044	0.412 ± 0.028	0.063	0.704×10^{-4}
MIP (proposed)	0.396 ± 0.011	0.095	0.389 ± 0.012	0.112	316.72

We compared our method (**MIP**) with the existing two post-processing methods: (1) modifying naive Bayes (**NB-Modified**) [6], and (2) relabeling for decision trees (**Relabeling**) [9]. In our experiments, we randomly split the dataset into training (50%) and test (50%) datasets, and report the average statistics over 5 trials. We obtained initial prediction models and applied each methods by using training datasets. Decision trees were learned by CART [4] with the constraint on these height less than 3. We used the threshold value of the fairness constraint $t = 0.1$ for all methods, and $\lambda = 0.1$ and $N_{\text{minsup}} = 100$ for Problem 1. All codes were implemented in Python 3.6 with scikit-learn^{*3} and IBM ILOG CPLEX Optimization Studio v12.8. All experiments were conducted on 64-bit macOS Sierra 10.12.6 with Intel Core i5 2.90GHz CPU and 8GB Memory.

Experimental results Table 1 shows the experimental results for DP scores. The DP scores of naive Bayes and decision trees before modification were 0.462 and 0.226, respectively. Our method maintained slightly lower loss than the relabeling method for decision trees on both training and test datasets, and comparable accuracy with the modifying naive Bayes method. We note that the average DP score attained by the modifying naive Bayes method exceeds $t = 0.1$ since it sometimes failed to obtain a model satisfying the fairness constraint. On the other hands, the running time of our method was longer than other two methods. This result implies that we need to improve our formulation, e.g., reducing program variables and constraints.

5. Conclusion and Discussion

We studied a fairness-aware post-processing method for decision trees, and proposed an MIP formulation of the re-thresholding problem, which makes a given decision tree fair by modifying their branching thresholds. Also, we formulate an edit distance of a decision tree so as to avoid that a learned model is changed significantly. By experiments on real datasets, we confirmed the effectiveness of our methods by comparing with the existing post-processing methods.

As future work, we will try to extend our framework so as to modify branching features, and deal with user-defined constraints more flexibly. It is important that maintaining user's prior knowledge contained in the model, e.g., order of branching features on a pass from its root node to a leaf node, while improving fairness by editing operations.

Acknowledgements This work was partially supported by JSPS KAKENHI(S) 15H05711 and JSPS KAKENHI(A)

^{*3} <https://scikit-learn.org/>

16H01743.

References

- [1] J. Adebayo. FairML: Auditing black-box predictive models. <https://github.com/adebayoj/fairml>, 2018.
- [2] S. Aghaei, M. J. Azizi, and P. Vayanos. Learning optimal and fair decision trees for non-discriminative decision-making. In *Proc. AAAI 2019, Hawaii*, 2019 (to appear).
- [3] D. Bertsimas and J. Dunn. Optimal classification trees. *Mach. Learn.*, pages 1039–1082, 2017.
- [4] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.
- [5] T. Calders, F. Kamiran, and M. Pechenizkiy. Building classifiers with independency constraints. In *IEEE ICDM 2009 Workshops*, pages 13–18, Dec 2009.
- [6] T. Calders and S. Verwer. Three naive bayes approaches for discrimination-free classification. *Data Min. Knowl. Discov.*, pages 277–292, 2010.
- [7] S. Hajian, F. Bonchi, and C. Castillo. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proc. ACM KDD 2016, San Francisco*, pages 2125–2126, 2016.
- [8] M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In *NIPS 2016, Barcelona*, pages 3315–3323.
- [9] F. Kamiran, T. Calders, and M. Pechenizkiy. Discrimination aware decision tree learning. In *IEEE ICDM 2010*, pages 869–874, 2010.
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin. "Why should I trust you?": Explaining the predictions of any classifier. In *Proc. ACM KDD 2016, San Francisco*, pages 1135–1144, 2016.
- [11] C. Rudin and S. Ertekin. Learning customized and optimized lists of rules with mathematical programming. *Mathematical Programming Computation*, 10(4):659–702, Dec 2018.
- [12] K. Zhang and D. Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, 18(6):1245–1262, 1989.

コメディドラマにおける字幕と表情を用いた笑い予測

Predicting Laughters in Comedy Drama with Subtitles and Facial Expression

萱谷 勇太^{*1}
Yuta KAYATANI

大谷 まゆ^{*2}
Mayu OTANI

Chenhui Chu^{*3}

中島 悠太^{*3}
Yuta NAKASHIMA

竹村 治雄^{*1}
Haruo TAKEMURA

^{*1}大阪大学 大学院 情報科学研究科
Graduate School of Information Science and Technology, Osaka University

^{*2}株式会社サイバーエージェント
CyberAgent, Inc.

^{*3}大阪大学 データビリティフロンティア機構
Institute for Datability Science, Osaka University

In this paper, we propose a model to predict whether an utterance leads to laughter in a comedy TV show. And we counts for facial expression that actors/presenters make. The model with subtitles and facial expression constructed as input was able to obtain accuracy, precision, recall, f-score more than model which input subtitle only or model which input only facial expression.

1. はじめに

笑いは、コミュニケーションにおいて非常に重要な役割を持つ。その笑いを引き起こすジョークは、プレゼンテーションにおいても聴講者の関心を示し、引き込むことができる。それだけでなく、ジョークは普段の会話においても会話を盛り上げ、その場を和ませることができる。実際に、これらのジョークによる笑いが生み出す効果は、暗黙的に周知されており、人々は会話の中で笑いを誘発しようとしている。

現在、機械による笑いの予測に向けた研究も進められており、例えば Bertero ら [1] の研究が挙げられる。Bertero らは、笑いが起こる発話はその発話以前の発話が起因となっていると仮定し、時系列情報を扱うことを得意とする LSTM を使用したモデルを構築した。この Bertero らの手法は字幕のみを入力としており、登場人物の表情などの視覚情報について考慮されていない。しかし、我々は笑いを引き起こすことにおいて視覚情報は重要な手がかりであると考えた。例えば、図 1 は登場人物の発言が終了した時点の表情の変化によって笑いを誘発している。そこで、我々は Bertero らと同様に、笑いを予測することができるモデルを考案し、Bertero らが複数の発話を入力していることに対して、1 つの発話のみを用い、その上で、字幕だけでなく、視覚情報である発話者の表情の特徴量である Action Unit[2] を使用し、それが笑い予測において重要であることを示す。結果として、字幕だけでなく、視覚情報である表情も組み合わせることで、最も高い精度を得ることができた。



図 1: 表情の例 (The Big Bang Theory, Season 3, Episode 12 (CBS) より.)

表 1: Action Unit

AU No.	内容	AU No.	内容
1	眉の内側を上げる	14	えくぼを作る
2	眉の外側を上げる	15	唇両端を下げる
4	眉を下げる	17	オトガイを上げる
5	上まぶたを上げる	20	唇両端を横に引く
6	頬を上げる	23	唇を固く閉じる
7	瞼を緊張させる	25	顎を下げて唇を開く
9	鼻にシワを寄せる	26	顎を下げて唇を開く
10	上唇を上げる	45	瞬く
12	唇両端を上げる		

2. 提案手法

提案手法として、モデルへの入力を表情・字幕とし、その二つの入力に対するモデルの出力によって、笑いが起こった確率を得る。

2.1 表情の取得

本研究では、表情を取得する手法として、OpenFace[3]を用いる。OpenFace では、目・鼻・口などの顔のパーツの形状や位置、頭の向き、視線の向き、顔の表情 (Action Unit) などを検出することができる。Action Unit とは、顔面筋肉の解剖学

的知見を基礎とした、44 の動作単位の事を指し、例えば、笑顔を表す時、Action Unit は 6 + 12 となる。OpenFace では 17 種類の Action Unit の強度 (0 - 5) が取得できる。OpenFace で得られる Action Unit は表 1 の通りである。

2.2 ネットワーク構成

本研究におけるネットワークモデルを図 2 に示す。二つの入力における詳細は以下の通りである。

● 字幕

字幕に関しては、単語分散表現である GloVe[4] を使用する。この時、発話ごとに単語のベクトル値の平均値を計算したものを使用し、また入力の発話は 1 つのみとする。

連絡先: 萱谷 勇太, 大阪大学 大学院 情報科学研究科,
yuta.kayatani@lab.ime.cmc.osaka-u.ac.jp

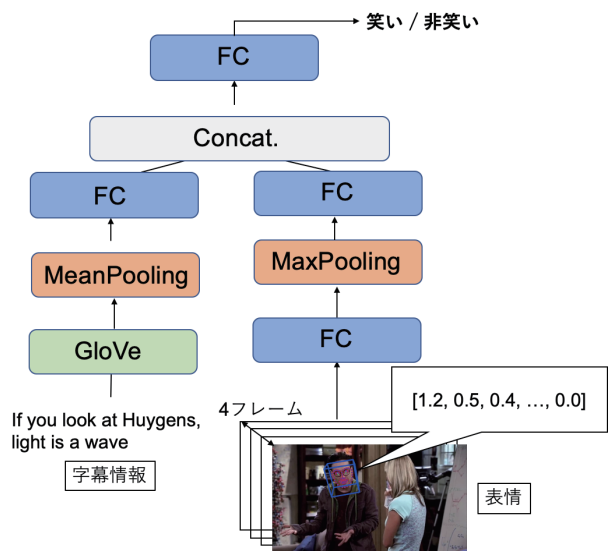


図 2: ネットワークモデル図

表 2: データの内訳

エピソード数	159
発話数	84191
笑いの割合	43.2%
非笑いの割合	56.8%

● 表情

表情に関しては OpenFace を用いて出力された Action Unit の強度情報を使用する。この時、発話終了時点の時間のフレームだけでなく、「発話終了時点の時間 + 3 秒内のフレーム」、「発話終了時点の時間 - 3 秒内のフレーム」を入力としたモデルをそれぞれ構築した。それぞれは、時系列方向へのマックスプーリングを行うことで 1 次元に圧縮しこれを使用する。この時、発話者の Action Unit ではなく、その発話の時間において OpenFace により出力される、信頼度の最も高い登場人物の Action Unit において、マックスプーリングを行う。

3. 実験

3.1 実験設定

本研究では、笑いを予測するタスクのデータセットとして海外コメディドラマである The Big Bang Theory を使用した。データにおける笑い・非笑いのラベリング方法としては発話が終わった時間から 1 秒以内の効果音から笑いが検出された時、笑いが起こった（ラベルを 1）とした。また、データにおける 80% をトレーニングセット、10% をバリデーションセット、10% をテストセットとした。データの内訳を表 2 に示す。また、モデルのパラメータ設定として、エポック数を 80、バッチサイズを 32、最適化手法を AdaDelta とした。

3.2 実験結果

字幕のみ・表情のみ・字幕と表情を使用した場合の実験結果を表 3 に示す。字幕のみ・表情のみを用いた結果に比べて、字幕と表情の両方を用いた場合、-3 秒、+3 秒両方において性

表 3: 実験結果

手法	Acc	Pre	Rec	F
字幕のみ	63.5	61.6	63.5	61.5
表情のみ	62.2	55.9	62.2	51.3
字幕 + 表情 (-3 秒)	65.7	63.9	65.7	62.0
字幕 + 表情 (+3 秒)	66.1	64.5	66.2	63.1

能が向上し、+3 秒の時が最も高い性能を達成することができた。この結果から、笑い予測のタスクにおいて字幕だけでなく表情も組み合わせることは精度の向上において有効であることが確認できた。

3.3 考察

字幕と表情を用いた場合において、+3 秒・-3 秒の両方について評価を行ったが、+3 秒時点の方がモデルの精度が高かったことから、笑い予測のタスクにおいては発話が終了した後の表情が重要であることが認識できる。しかし、予測という観点において、笑いが起こった後の情報を学習に使用することは適切ではないと考えられる。だが、-3 秒でのモデルでも性能が向上していることから、表情の組み合わせは有効であり、適切なモデルを選定することでさらなる精度の向上が見込まれる。

4. 結論と今後の課題

本論文では、笑い予測のために字幕だけでなく表情を用いたモデルの提案を行った。結果としては字幕だけ、表情だけを用いたモデルよりも両方を用いたモデルのほうが全ての評価指標において高い数値を得ることができることを示した。今後の課題としては、本論文では字幕に付与する情報として表情のみを扱っているが、他にも発話者の骨格情報や姿勢の情報も使用することなどが考えられる。

本研究の一部は科研費 No. 18H03264 による。

参考文献

[1] Dario Bertero and Pascale Fung. A long short-term memory framework for predicting humor in dialogues. In *NAACL-HLT*, pp. 130–135, 2016.

[2] Paul Ekman. Pictures of facial affect. *Consulting Psychologists Press*, 1976.

[3] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *WACV*, pp. 1–10. IEEE, 2016.

[4] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *EMNLP*, pp. 1532–1543, 2014.

自己注意機構を利用したアスペクトベースの感情分析の 日本語文への適用

Application of Aspect-based Sentiment Analysis using Self-Attention Mechanism to Japanese Sentences

赤井 龍一^{*1}
Ryuichi Akai

渥美 雅保^{*1}
Masayasu Atsumi

^{*1} 創価大学大学院工学研究科情報システム工学専攻
Information Systems Science, Graduate School of Engineering, Soka University

Sentiment analysis is a task to estimate emotions from information such as sentences. As SNSs such as LINE and Twitter which communicate mainly using sentences have developed, technology to estimate emotion from sentence information is increasingly important. In recent years, a method using aspect information has attracted attention as a sentiment analysis method using context information. The aspect-based sentiment analysis is realized in three stages. It categorizes sentences at the first stage, estimates word position which is a concrete description of aspect at the second stage, and outputs the polarity at the last stage. In this research, a neural network-based method using the self-attention mechanism for performing the aspect-based sentiment analysis is applied to Japanese sentences to evaluate its performance in comparison with English sentences.

1. はじめに

感情分析は文章などの情報から感情を読み取ろうとするタスクである。LINE や Twitter に代表される、主に文を使ってコミュニケーションを行う SNS が発達してきているため、文情報から感情を読み取る技術が重要性を増している。近年、文脈情報を利用した感情分析手法として、アスペクト情報を利用した手法が注目されている[Pontiki 2015]。これら研究では、アスペクトベースの感情分析は 3 つのステージに分けて実現されている。最初のステージで文をアスペクトカテゴリーに分類し、二番目のステージで文中のその文のアスペクトを規定する単語の位置の推定を行い、最後のステージで極性を出力する。本研究では、自己注意機構を利用して文のアスペクト情報に注意を向けて感情分析を行うニューラルネットワークベースの手法[Lin 2017]を日本語文に適用する。そして、日本語文への適用の精度を英語文に対する適用の精度と比較する。

2. 関連研究

2.1 アスペクトベースの感情分析

アスペクトベースの感情分析とはアスペクト情報を利用した感情分析のことである。アスペクト情報とはその文のカテゴリに当たるものであり、例えば、製品とその属性の組み合わせ等、エンティティタイプと属性タイプの組み合わせによって定義される。[Pontiki 2015]では 3 つのステージに分けてアスペクトベースの感情分析を行っている。初めに、あらかじめ設定しておいたアスペクトのカテゴリに文を分類する(図 1)。次に、文とカテゴリ情報を入力データにしてアスペクトの具体的な記述である単語の位置を出力する(図 2)。最後に、文、カテゴリ、単語の位置を入力値として極性の値を得る(図 3)。この手法を日本語[Kubo 2018]に適用すると 2 番目のステージの精度が他のステージと比べて低い点が問題である(表 1)。そこで本研究では、日本語文の感

情分析に 2 番目のステージの精度をあげることが期待される[Lin 2017]の手法を適用することによって、3 番目の極性の出力の精度を向上させることを目指す。

表 1 アスペクトベースの感情分析のステージごとの F 値

	英語	日本語
ステージ 1	0.73	0.46
ステージ 2	0.72	0.19
ステージ 3	0.88	0.74

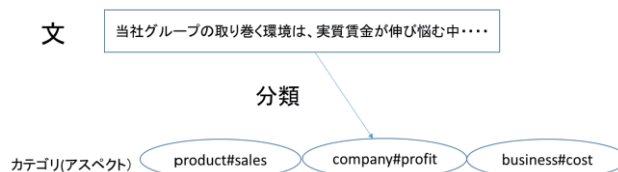


図 1 文のカテゴリ分類

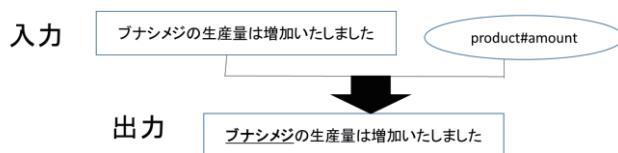


図 2 単語位置の出力

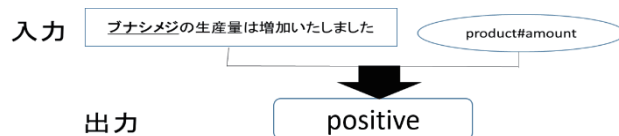


図 3 極性の出力

連絡先: 赤井 龍一, 創価大学大学院工学研究科,
〒192-8577 東京都八王子市丹木町 1-236, Tel: 042-691-2211

2.2 自己注意機構

注意機構はエンコーダデコーダモデルでよく使われる手法である。一般的なエンコーダデコーダの注意は、エンコーダの隠れ層を Source, デコーダの隠れ層を Target として次式によって表される。

$$\text{attention}(\text{Target}, \text{Source}) \\ = \text{Softmax}(\text{Target} \cdot \text{Source}^T) \cdot \text{Source}$$

これは、Target を query とみなし、Source を Key と Value に分離すると、次のように表される。

$$\text{attention}(\text{query}, \text{Key}, \text{Value}) \\ = \text{Softmax}(\text{query} \cdot \text{Key}^T) \cdot \text{Value}$$

このとき、注意は、query に対して各 key と各 value が 1 対 1 対応するペアの配列、すなわち辞書オブジェクトから value を取り出す操作とみなせる。自己注意は、query, key, value がすべて同じ層から来る場合の注意を指す。

3. 日本語文の感情分析実験

3.1 実験概要

自己注意機構を使ったニューラルネットワークによる英語文の感情分析の先行研究[Lin 2017]では、文中の単語間の依存関係を得るために、bidirectional LSTM で文を変換し、各単語に対応する隠れ層 H を入力として、その単語に注目すべき(自己注意)確率を次式により計算する。

$$a = \text{softmax}(w_{s2} \tanh(W_{s1} H^T))$$

得られる重みベクトル a は文の特定の単語やフレーズに焦点を当てるものである。そのため、アスペクト情報を与える文脈情報の抽出が期待できる。文中に複数の構成要素がある場合、特に長い文の場合は文の異なる部分に焦点を合わせた複数の注意重みベクトルが必要となる。このとき出力される注意が異なるように、ペナルティ項を導入して注意の多様性を促進する。最後に、注意の重み付きで各単語に対応する隠れ層の出力を足し合わせたものを入力として、全結合層と Softmax 層で文の極性ラベルの予測を行う。この自己注意機構を用いた感情分析モデルを日本語文へ適用し、日本語文の感情分析を行った。

3.2 データセット

今回の実験で使用する日本語のデータセットは KNB コーパス[橋本 2011]のうち評判情報の感情タグが付いた 492 文である。学習用データセットとして 392 文、テスト用データセットとして 100 文を使用する。まず文を形態素解析して単語に分解し、単語の出現頻度順に上位の単語からインデックスを付ける。文頭を表すインデックスを 1 とした。また、助詞などの付属語のインデックスを 2 とした。そして、その他の単語につけるインデックスは 3 からとした(表 2)。このインデックスを用いて単語を one-hot ベクトルにより表現する。ポジティブとネガティブをそれぞれ 1, 0 として二値分類を行った。

比較のための英語のデータセットは IMDB という映画レビューサイトのデータセット¹を使用する。このデータセットは単語の出現頻度順にインデックスが 3 から付けられている。文の初めにインデックス 1 が付けられ、出現頻度が上位 1000 位未満の単語のインデックスには 2 が付けられている(表 3)。そして、日本語のデータセットと同じようにポジティブとネガティブが 1, 0 でラベル付けされている。

表 2 日本語データセットのインデックス

インデックス	
0	Padding
1	文の開始位置
2	助詞などの付属語
3 以降	インデックスを 2 にした単語以外の単語

表 3 IMDB データセットのインデックス

インデックス	
0	Padding
1	文の開始位置
2	出現頻度順が 1000 位未満
3 以降	出現頻度順が 1000 位以上

3.3 実験結果と考察

日本語と英語の実験結果として、感情分析の正確度を示す(表 4)。日本語も英語と同程度の精度を達成している。英語のデータセットは学習用とテスト用それぞれ 25000 文ずつあるので、英語と日本語の差は学習用データセットの量の差であると考えられる。

表 4 感情分析の正確度

英語	91%
日本語	85%

英語文の場合、文にポジティブとネガティブな意味が含まれる単語が入っているとその単語に注目して正解している。皮肉やあまり使われない用法で使われた単語が含まれている文は間違ってしまうことが多かった。日本語文の場合、データセットの量が少ないため出現頻度の少ない単語に注目がいく傾向があった。

4. むすび

本論文では、アスペクトベースの感情分析手法の中のステージ 2 の精度の向上をはかることを目的に、自己注意機構を利用したニューラルネットワークモデルの日本語文のセンチメント分析への適用について述べた。今後は注意がアスペクト情報にいかに向けられているか確かめて、ネットワークモデルを改良していく。また一文に対する感情分析だけでなく、1 つのレビューなどの文章に対する感情分析も行いたい。

参考文献

- [Pontiki 2015] Maria Pontiki et al: Semeval-2015 Task12: Aspect Based Sentiment Analysis, Semval, 2015.
- [Kubo 2018] Kubo T. and Nakayama H.: chABSA: Aspect Based Sentiment Analysis dataset in Japanese, <https://github.com/chakki-works/chABSA-dataset>, TIS 2018.
- [Lin 2017] Zhouhan Lin, et al: A Structured Self-attentive Sentence Embedding, ICLR, 2017.
- [橋本 2011] 橋本 力, et al: 構文・照応・評判情報つきブログコーパスの構築, 自然言語処理 Vol.18, pp175-201, 2011

¹: <https://www.imdb.com/interfaces/>

感情強度辞書を利用した日本語ツイートの感情強度の推定

Estimating Emotion Intensities in Japanese Tweets Using Emotion Intensity Lexicon

赤堀 立樹^{*1}
Tatsuki Akahori

堂坂 浩二^{*2}
Kohji Dohsaka

伊東 嗣功^{*2}
Hidekatsu Ito

石井 雅樹^{*2}
Masaki Ishii

^{*1} 秋田県立大学 大学院
Graduate School, Akita Prefectural University

^{*2} 秋田県立大学
Akita Prefectural University

This paper presents a method for estimating emotion intensities in Japanese tweets using emotion intensity lexicon. We intend to utilize this method to create dialogue systems that can regulate user's emotion. Such dialogue systems can be expected to be useful for maintaining the good mental condition and interpersonal relationships of the users. We first created a crowdsourced emotion intensity lexicon for eight categories of emotion (joy, sadness, anger, fear, trust, disgust, anticipation, surprise), which captures word-emotion intensities using best-worst scaling. To the best of our knowledge, there is no previous works to construct emotion intensity lexicon in Japanese. Then we carried out an experiment to evaluate the effect of the emotion intensity lexicon in the estimation of emotion intensities in Japanese tweets. The experimental results show that the emotion intensity lexicon can improve the accuracy of estimating emotion intensities in tweets.

1. はじめに

対話システムには、ユーザとの対話を通して道案内や予約など特定のタスクを遂行するタスク指向型対話システムと、雑談などコミュニケーション自身を目的とする非タスク指向型対話システム(雑談対話システムとも呼ぶ)があり [中野 2015]. 近年は雑談対話システムの研究が盛んに行われている. 人間同士の雑談において、対話参加者は相手の感情を考慮し、相手の感情を調整(促進・維持・抑制)しながら対話に従事する. 例えば、悲しんでいる人に対して慰める発話をするにより、相手の「悲しみ」を和らげたり、「喜び」の感情をもっている人に対して賞賛や共感するような発話をするにより、相手の「喜び」を促進する. 雑談対話システムにおいても、ユーザの感情を考慮し、ユーザの感情を調節することができれば、ユーザの良好な心身状態や対人関係の維持、また感情調節困難な方の支援に貢献できることが期待される.

ユーザの感情を調節する対話システムを構築するためには、ユーザの感情強度やその変化を測定することが必要である. 感情強度の推定に関する研究は盛んになりつつあり、WASSA-2017 [Mohammad 2017] や SemEval 2018:Task 1 Affect in Tweets [Mohammad 2018a] では、4感情 (anger, sadness, fear, joy) を対象として、Twitter における英語ツイートの感情強度を推定するタスクが設定されており、様々な手法が提案されている. これらの研究は英語の感情強度推定が目的であるが、その一方、佐藤らは日本語ツイートの感情強度を推定する手法を提案している [佐藤 2018].

ツイートの感情強度推定のためには、感情辞書が重要な役割を果たす. 英語の感情強度推定の場合、英語の言語表現に対して感情強度を付与した感情強度辞書 [Mohammad 2018b] が利用されている. 日本語の感情辞書としては、日本語表現を10種類の感情に分類した感情表現辞典 [中村 1993] や単語に感情極性 (positive・negative) の値を実数値 [-1.0, 1.0]で付与した単語感情極性対応表 [高村 2006] などがあるが、管見の限りでは日本語の感情強度辞書は構築されてこなかった.

そこで、本研究では、日本語の感情強度辞書を構築し、感情

表 1. 感情分類

Plutchik の 8 感情分類	中村の 10 感情分類
喜び	喜
受容	好・安
嫌悪	厭
悲しみ	哀
怒り	怒
恐れ	怖
予期	NRC-Emotion-Lexicon
驚き	驚

強度辞書を用いた日本語ツイートの感情強度の推定法を評価することを目的とする. 以下において、第 2 節で日本語感情辞書の構築について説明する. 第 3 節で感情強度辞書を使った日本語ツイートの感情強度推定の評価実験について述べ、第 4 節で実験結果について考察する.

2. 日本語感情強度辞書の構築

本研究では、まず、感情強度辞書に登録する感情表現を収集する. その感情表現の強度を数値化し、感情強度辞書を構築する.

2.1 感情表現の収集

本研究では、中村の「感情表現辞典」 [中村 1993] から感情カテゴリ(喜,好,安,厭,哀,怒,怖,驚,恥,昂)ごとに感情表現を収集し、中村の10分類をPlutchik [Plutchik 2001] の8分類 (喜び, 受容, 嫌悪, 悲しみ, 怒り, 恐れ, 予期, 驚き) に再分類した. 再分類は他研究との比較や応用しやすくするためである. 表 1 に再分類の振り分けを示す. ここで、中村の感情分類の「好」と「昂」は、他の複数の感情に分類できるため、適宜分類した. 例えば、「昂」の「飛び上がる」や「踊る」などは「喜び」に分類し、「熱り立つ」や「怒り狂う」は「怒り」に分類した. また、Plutchik の感情分類の「予期」は中村の感情分類には該当する感情がないため、Plutchik の感情分類に対応する言語表現を人手により収集した

連絡先: 堂坂浩二, 秋田県立大学 システム科学技術学部 情報工学科, dohsaka@akita-pu.ac.jp

表 2. 感情表現 BWS 結果(喜び, 受容, 嫌悪, 悲しみ)

喜び		受容		嫌悪		悲しみ	
嬉し泣き	0.979	信頼する	0.944	反吐が出る	0.924	悲痛な叫び	1.000
歓喜	0.974	敬愛する	0.875	目障り	0.905	咽び泣き	0.984
天にも上る気持ち	0.974	相思相愛	0.854	生理的に無理	0.896	泣き崩れる	0.923
大喜び	0.970	信仰する	0.854	殺意を抱く	0.894	泣き叫ぶ	0.875
狂喜	0.917	尊ぶ	0.852	吐き気がする	0.875	見るのもつらい	0.875
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
おかしい	0.119	ゆるゆる	0.143	羨ましい	0.167	しける	0.109
笑	0.119	のどか	0.119	困惑	0.146	くすんくすん	0.078
気軽	0.095	無事	0.111	残念に思う	0.125	デメリット	0.047
吉	0.083	ふかふか	0.104	無念	0.095	苦手	0.042
クスクス	0.083	ゆったり	0.097	可哀想	0.000	ガーン	0.000

NRC-Emotion-Lexicon [Mohammad 2013] を日本語に翻訳した表現集を使った。

次に、多様な表現を収集するために、収集した感情表現の類語を収集した。すべての感情表現の類語を収集するのは、コストがかかるため、Twitter API を使って 2017 年 3 月から 2018 年 9 月までの間に収集した約 2000 万ツイートの上で出現回数を調べ、感情カテゴリごとに出現回数上位 30 語程度の感情表現の類語を収集することとした。類語の収集には Weblio 類語辞典¹を使用した。類語の収集の際には、明らかに感情表現と意味合いが異なる類語は除外した(例:「怒り」の感情表現「雷」から稲妻・電光などの類語は収集しない)。類語を持たない語(例:爆発する)も除いた。類似した類語(例:激怒、憤怒、憤慨)は 1 つを残して除いた。最後に、類語を含めた感情表現の出現回数を再度調べ、出現回数 1 回以上の語を抽出した。

2.2 節で説明するように、本研究では、Best-Worst Scaling (BWS) [Louviere 1991] を用いて、感情表現に感情強度を付与する。BWS は相対評価法であるため、感情強度に極端な偏りがあると適切に算出できない。そのため、抽出した感情表現の強さを手動で 3 段階に分類し、収集した感情表現の強度に極端な偏りが無いことを確認した。感情強度の 3 段階の分類は、筆者の一人が各表現に対して 0~100 の実数値を付与することによって行った。3 段階は大(68~100)、中(34~67)、小(0~33)とした。以上の結果、1051 語の感情表現を収集した。

2.2 感情表現の強度の数値化

2.1 で収集した各感情表現に対して BWS によって感情強度を数値化した。BWS は Mohammad らの手法 [Mohammad

表 3. 感情表現 BWS 結果(怒り, 恐れ, 予期, 驚き)

怒り		恐れ		予期		驚き	
ぶっ殺す	1.000	戦慄	0.938	予兆	0.938	言葉が出ない	0.929
殺す	0.984	叫喚	0.938	予測する	0.917	言葉を失った	0.881
殺意	0.911	阿鼻叫喚	0.913	兆候	0.917	想像を絶する	0.871
怒鳴り散らす	0.909	生きた心地がしなかった	0.859	兆し	0.905	夢にも思わない	0.854
怒り狂う	0.900	畏怖	0.859	予想する	0.833	驚愕する	0.833
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
頬を膨らます	0.141	渋る	0.125	宿命	0.208	ビクビク	0.185
くしゃくしゃ	0.141	決めかねる	0.113	釈然としない	0.208	認められない	0.167
つまらん	0.125	期待できない	0.094	理想	0.200	たまたま	0.042
ふくれっ面	0.089	気遣い	0.094	大志	0.167	うつとり	0.042
つまらない	0.075	堂々巡り	0.094	ドリーム	0.167	はわわ	0.024

2018b] を参考に行った。感情カテゴリごとに 4 人の作業者に $2 \times N$ (N は感情表現数)の質問に回答してもらった。各質問では、特定の感情カテゴリに属する 4 つの感情表現が与え、最も感情を強く感じる感情表現 (Best) と最も感情を弱く感じる感情表現 (Worst) の 2 つを回答することを求めた。このとき、Best と Worst は異なる感情表現を選択してもらった。各感情表現は $2 \times N$ の質問中に異なる質問で 8 回出現し、同一の選択肢の質問は現れないようにした。感情表現 e の感情強度の値は次の式によって計算する。

$$\frac{P_{Best}(e) - P_{Worst}(e) + 1}{2}$$

このとき、 $P_{Best}(e)$ 、 $P_{Worst}(e)$ は感情表現 e がそれぞれ Best、Worst に選択された確率である。 $P_{Best}(e)$ 、 $P_{Worst}(e)$ の値域は $[0, 1]$ であり、 $P_{Best}(e) - P_{Worst}(e)$ の値域は $[-1, 1]$ となる。本研究では値域を $[0, 1]$ とするために線形変換を行う。感情強度は値が 1 に近いほど強い感情を示し、0 に近いほど弱い感情を示す。

BWS はクラウドソーシング Lancers を使って実施した。作業の質を確保するために全質問の 5% の質問を回答が明らかに一意に定まる質問 (gold question) として設定し、gold question の正解率が 70% 未満の作業者の回答は使用しなかった。gold question は、BWS の作業者とは別の 2 名の作業者に感情強度「大」「小」からそれぞれ 1 つと「中」から 2 つから成る質問に対して BWS と同様に回答してもらい、2 人の回答が一致した回答の中から作成した。

BWS は単一の質問で複数の項目間の順序関係を得られるため、簡便かつ有効な評価方法であると考えられる。表 2, 3 に各感情の感情強度の値が上位と下位それぞれ 5 つずつの結果を示す。

¹ <https://thesaurus.weblio.jp>

2.3 感情強度辞書に関する考察

2.2 節の方法により、1051 語からなる感情強度辞書の作成を行った。感情強度の値は概ね直感と一致する結果となったが、予期の感情では一部直感とは異なる結果が見られた。

予期の感情では、感情強度が高いと考えられる「期待」、「宿命」、「理想」といったの表現が強度 0.67 以下の中・下位に入った。一方、感情強度が低いと考えられる「知らせ」、「見込み」などの表現が、感情強度 0.67 以上の高位に入った。このようになった要因として、作業者が「予期」の感情の意味を限定的に捉えた可能性がある。予期の感情には「予想」、「知らせ」、「見込み」といった未来の予見や未知の情報を表す表現の他に、「宿命」といった決まった未来を表す表現や、「期待」、「理想」といった希望・意図を表す表現がある。作業者は予期の感情について、未来の予想や未知の情報を表す意味に限定して考えたため、その他の意味を表す表現の強度が相対的に低くなった可能性がある。

3. ツイートの感情強度の推定と評価

3.1 感情強度辞書を使った感情強度推定方法

本研究では、感情強度辞書を使ったツイートの感情強度推定法の開発の第一歩として、簡易な方法により感情強度の推定を行い、感情表現辞書の効果について考察する。

前提として、ツイートの感情分類が成されているものとした。また、感情強度辞書に登録された感情表現を含むツイートを対象とした。ツイート中に含まれる否定や仮定、強調表現 (!! , オオオオオなど) は考慮しなかった。

ツイートの感情強度の推定は、感情強度辞書を用いて、ツイートに含まれる感情表現を手がかりとして行う。ツイートには複数の感情表現が含まれる場合もあるため、以下に示す、First, Last, Average の 3 つの方法と、ランダムに感情強度値を生成する方法の計 4 つの手法により推定を行う。各推定方法について以下に示す。

- First : ツイート中で最初に現れた感情表現の感情強度値をツイートの感情強度値とする
- Last : ツイート中で最後に現れた感情表現の感情強度値をツイートの感情強度値とする
- Average: ツイート中の感情表現の感情強度値の平均をツイートの感情強度値とする
- Random: ランダムに感情強度値を生成し、ツイートの感情強度値とする

3.2 評価実験

感情強度辞書を使ったツイートの感情強度推定の評価実験を行った。評価の対象となるツイートは「喜び」と「悲しみ」の感情に事前に分類された各 50 ツイートとした。ツイートの感情強度値の正解データを作成する作業者の負担を考慮して、ツイートは 20~50 文字程度のツイートのみを使用した。

ツイートの感情強度値の正解データ (測定値) を作成するため、事前に 3.1 節と同様の手順によりツイートの感情強度値を BWS により測定した。gold question の正解率が 70% の作業者の BWS の結果を使用し、正解データを作成した。

評価は、各方法により推定された感情強度値を大(0.67~1), 中(0.33~0.67), 小(0~0.33)の 3 段階への分類した際の正解率を使って行った。

表 4, 5 にツイートの感情強度推定の評価結果を示す。「喜び」の感情では、最初の感情表現の感情強度をツイートの感情

表 4. ツイートの感情強度推定評価結果(「喜び」)

	小	中	大	全体
a) First	60%	59%	69%	62%
b) Last	56%	50%	56%	54%
c) Average	60%	55%	60%	58%
d) Random	27%	48%	26%	34%

表 5. ツイートの感情強度推定評価結果(「悲しみ」)

	小	中	大	全体
a) First	37%	56%	54%	48%
b) Last	44%	56%	50%	50%
c) Average	44%	57%	54%	52%
d) Random	30%	47%	29%	35%

強度にする手法が正解率 62% で最も高い結果となった。また「悲しみ」の感情では、ツイート中の感情表現の感情強度の平均をツイートの感情強度とする手法が正解率 52% で最も高い結果となった。感情で評価結果を比較すると、「喜び」の感情のほうが全体的に高い正解率となった。

4. 考察

「喜び」の感情強度推定は 60% 程度、「悲しみ」の感情強度推定は 50% 程度の正解率となった。Random との比較ではあるが、感情強度辞書は一定の効果があると言える。「喜び」の感情の方が全体的に高い正解率となったが、作業者の感想から、「喜び」の感情に比べて「悲しみ」の感情は、感情表現の感情強度よりも、ツイート内容や顔文字の種類に影響されやすいことが考えられる。そのためツイートに含まれる感情表現のみを手がかりとする推定方法では「喜び」の感情の正解率が高くなったことが考えられる

表 6, 7 にそれぞれ「喜び」と「悲しみ」の感情強度推定が不正解となった場合の特徴的な例を示す。測定値とは、3.2 節で述べたように、事前に作業者の BWS により測定した感情強度の正解値である。推定値とは、感情強度辞書を使ってツイートの感情強度を推定した値である。「喜び」の場合は方法 First, 「悲しみ」の場合は方法 Average による推定値である。

まず、表 6 の「喜び」の場合、特徴的な例として、文中に否定を含む場合、辞書に含まれない感情を表す表現を含む場合、強調表現を多用している場合があった。まず、文中に否定を含む例として、表 6 のツイート番号 13 では、釣りに行ったが釣れなかったという否定内容が含まれていた。否定的な表現の影響

表 6. 感情強度推定の不正解例 (喜び)

ツイ ート 番号	ツイート内容	測定 値	推定 値
13	初めてうなぎつりやった てか、 昨晚は鯉つり w ミナミの釣り堀 〜 釣れなかったけど w 久々に 糸垂らして恍惚としてたよ	0.271	0.771
18	デザフェス当選した！！ヤッター ー！！！！！！	0.875	0.667
26	念願の！！！！チュロス(｡◕◕｡)この ベリーソース神ってる!! 店員 のお姉さんと、少しお話もして、 充実した朝時間になりました㇏ (●´▽`●)ﾉ	0.833	0.563

により、感情強度の測定値の方が、否定を考慮せずに推定された推定値よりも低くなったと考えられる。

ツイート番号 26 は、感情強度辞書に含まれていない感情表現を含む例である。「念願」や「神ってる」は辞書に含まれていないため、測定値の方が推定値よりも高くなったと考えられる。

ツイート番号 18 は強調表現「!!」を多用している例である。また、ツイート自体が短いため、強調表現による影響を受けやすい可能性がある。そのため、測定値の方が推定値よりも高い結果となったと考えられる。

表 7 に示す「悲しみ」の場合、不正解となった要因としては、顔文字の使用、ツイート内容、辞書に含まれない感情表現による影響が考えられる。まず、顔文字とツイート内容について考察する。表 7 のツイート番号 1 では感情強度の強い感情表現「咽び泣く」を使用しているが、ツイート内容または顔文字の影響により、悲しみの感情が弱く感じられたため、測定値が推定値より小さくなったと考えられる。一方、ツイート番号 12, 28, 32 では、測定値が推定値よりも高くなっており、顔文字が感情強度を高めている可能性がある。また、ツイート番号 36, 42, 45 でも、測定値が推定値よりも 0.2 程度高くなっているが、これらのツイートは、ツイートの内容の影響を受け、含まれる感情表現の感情強度値より高い測定値になった可能性がある。

次に、感情強度辞書に含まれない感情表現による影響について考察する。その例として表 17 のツイート番号 15, 47 がある。特にツイート番号 15 では、「阪神淡路大震災」がツイートに含まれている。現在作成した感情強度辞書は主に形容詞と動詞によって構成されている。今後は、感情強度値を付与した名詞を増やすことで推定性能が向上する可能性がある。

5. おわりに

本研究では、中村の「感情表現辞典」を基に類語の収集を行い、BWS により感情強度を測定し、1051 語からなる感情強度辞書を作成した。概ね直感と一致する辞書の作成を行うことができたが、予期の感情では一部直観と異なる結果となった。例として、予期の感情では、未来の予見や未知の情報を表す表現が上位の感情強度となり、決まった未来や希望・意図を表す表現が中・下位の感情強度となった。要因としては、作業者が予期の感情を未来の予想や未知の情報に限定して捉えたため、その他の意味をもつ表現の強度が相対的に低くなったと考えられる。感情強度辞書を使ったツイートの感情強度推定では喜び、悲しみともに 50～60% の正解率となった。感情強度辞書は一定の効果があることが示唆された。

今後の課題としては、感情強度値付きの大量のツイートからの学習による感情強度推定の性能向上、感情表現を含まないツイートの感情強度の推定が考えられる。また、名詞への感情強度値の付与や、否定や強調表現などを考慮した感情強度推定を行うことを検討する。

謝辞

本研究は科研費(16K00355)の助成を受けたものである。

参考文献

- [Louviere 1991] Jordan J. Louviere: Best-worst scaling: A model for the largest difference judgments, Working Paper, 1991.
- [Mohammad 2013] Saif M. Mohammad and Peter Turney: Crowdsourcing a Word-Emotion Association Lexicon, Computational Intelligence, Vol.29, No.3, pp.436-465, 2013.
- [Mohammad 2017] Saif M. Mohammad and Felipe Bravo-Marquez: WASSA-2017 shared task on emotion intensity. In:

表 7. 感情強度推定の不正解例(悲しみ)

ツイート番号	ツイート内容	測定値	推定値
1	なによりあのぱっちりお目目と色白の肌とすっぴんとの髪が手に入るかと思うと色黒一重天然パーマな私は咽び泣きます(;▽;)ノ	0.396	0.984
12	まじっすか……(´ω`lll)寂しいです……やっぱ心痛みます…(´;ω`)	0.708	0.589
15	阪神淡路大震災から 23 年。被害に遭われた方々へ 哀悼の意を表します。この日を決して忘れません	0.771	0.672
28	災害よりもモリ・カケ、ですか。情けない。(ノ 口)。(ノ 口)	0.646	0.516
32	ミク、コヒメが卒業発表するシーンでコヒメから貰い泣きしてしまった(;;)無理いい(;;)	0.625	0.406
36	家で珈琲を入れるときにフレンチプレスってのていれてるんですけど、うっかりフィルターの紙を入れるのを忘れてた時があって、朝からがつくりしたことあります…	0.500	0.328
42	昔絡んでた方々は見えないところに行ってしまったよ(しんみり)	0.542	0.313
45	職場の人で「自分がやられて嫌なことを、自分は他の人にする派」がいてしょげる	0.500	0.268
47	自分の無知さを恥じる	0.583	0.234

Proc. the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2017.

[Mohammad 2018a] Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh and Svetlana Kiritchenko: SemEval-2018 Task 1: Affect in Tweets, In: Proc. the 12th International Workshop on Semantic Evaluation, pp.1-17, 2018.

[Mohammad 2018b] Saif M. Mohammad: Word Affect Intensities, In Proc. LREC-2018, 2018.

[中村 1993] 中村明: 感情表現辞典, 東京堂出版, 1993.

[中野 2015] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子, 奥村学 (監修): 対話システム, コロナ社, 2015.

[Plutchik 2001] Robert Plutchik: The nature of emotions, American Scientist, Vol.89, No.4, pp.344-350, 2001.

[佐藤 2018] 佐藤一輝, 尾崎知伸: 複数の表現学習手法を用いた日本語ツイートの感情強度推定, 人工知能学会研究会資料, SIG-KBS-B802-02, 2018.

[高村 2006] 高村大也, 乾孝司, 奥村学: スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌ジャーナル, Vol.47, No.2, pp. 627-637, 2006.

新聞記事中の地名に対する地理的位置推定における有効な素性の調査

An Investigation of Effective Features for Toponym Resolution of Words in Newspaper Articles

関 龍^{*1} 乾 孝司^{*1}
Ryo Seki Takashi Inui

^{*1}筑波大学大学院システム情報工学研究科
Graduate School of System and Information Engineering, University of Tsukuba

In this paper, we investigate which features are efficient on toponym resolution of words in newspaper articles. Speriosu's TRIPDL algorithm is used in the investigation, and an extension version of TRIPDL is also used to diminish data sparseness problem. We focus on nouns and four types of named entity classes (ORGANIZATION, PERSON, LOCATION, and ARTIFACT), which are used in standard named entity recognition task. Through the experiments, it turns out that using only LOCATION words achieves better performance in terms of both accuracy and computational complexity.

1. はじめに

近年, SNS などの様々なサービスが普及し, 人々が気軽に情報を発信できる時代になってきている. これら発信された文書内にある, 場所を示す表現と実世界上での地理的位置を結びつけたいというニーズが存在する. SNS の代表とも言える Twitter では投稿時の位置情報をつぶやきに付与することが可能であるが, Middleton ら [1] によればこの機能を利用しているものは全体の 1% にも及ばない. 文書内の表現に対して実世界上でのその地理的位置が同定できれば, 特定区画を対象を限定したテキストベースの社会分析などがこれまで以上に高精度・高被覆で実現できるようになる.

文書内の場所を示す単語と実世界上での地理的位置を結びつけるには様々な問題があるが, その中でも場所を示す単語の曖昧性の問題がある [2]. 曖昧性とは例えば, 「中央区」という単語があった場合に, 日本に数ある中央区のうちのどの位置を実際に参照しているのかわからない状況のことである. 実際には数多くの中央区が存在するが, 簡単のため「大阪市中央区」と「東京都中央区」に絞った例を図 1 に示す. この地理的位置推定の研究は以前から行われており, どのように場所を示す単語の曖昧性を解消するかの方法には様々な手法が提案されてきた. 文書レベルでの地理的位置推定や単語レベルでの地理的位置推定, 教師データの有無などによって手法は異なるが, 本研究では単語レベルかつ教師データを用いる地理的位置推定について検討する.

一般にこの地理的位置推定のタスクではデータに SNS データが用いられることが多い. Lianhua ら [3] はデータに Twitter を用い, 地理的位置推定には地名やハッシュタグ, メンションの情報が有効であることを示した. 本研究ではデータに新聞記事を利用するときに, どのような素性, 特にどの固有表現が地理的位置推定に有効なのか調査する.

本論文では文書中に表れる実世界上の地理的位置を指す単語を「地名」と呼び, その地名に対応する可能性のある実世界上の住所を「候補地」あるいは単に「候補」と呼ぶ.

2. 関連研究

文書中の地名と実世界上の位置を関連付ける研究は数多く行われてきたが, 問題設定の違いにより場合分けを行うことができる. まず, 文書中に出現する「地名」つまり単語を位置と結



図 1: 中央区に対する曖昧性

びつけるタスクを toponym resolution と呼ぶ. 次に「文書」の記述内容を位置と結びつけるタスクを document geolocation と呼ぶ. また, ある「ユーザ」と関連付けられた文書集合と位置を結びつけるタスクを user geolocation と呼ぶ. この分類に従うと, 本研究は toponym resolution に該当する.

Speriosu は TRIPDL [4] を提案した. この手法は実世界をセルで区分けし, それらと文書との対応付けを行っている. 入力文書の単語確率分布と学習済みのセルの単語確率分布間の KL ダイバージェンスを測り, より近いセルを出力とする方法である. 本研究ではベース手法として TRIPDL を採用する. 手法の詳細については 5. 章で述べる.

オリジナルの TRIPDL の問題設定は document geolocation であり, 実世界上の位置にはセル (地理的位置を格子状に区切った領域) を採用している. セルを採用している理由としては入手が簡単, 扱いやすいという点が挙げられる. まず, 場所に対する経度緯度情報さえあれば場所とそれに対応するセルがすぐに関連付けできるから入手が簡単であるといえる. また, セルとは地図を正規化したものであるとも言え, 複雑な地形などもすべて格子状に落とし込むことができるため扱いやすいといえる. しかし, 本研究ではセルは使用せず, 行政地区を使用する. 行政地区を使用するメリットとしては非常に出力が直感的で他の用途に使いやすい点が挙げられる. 例えばラベル付けを考えた場合, ある曖昧性のある地名に対して出力がセルだとなんの情報かわかりにくい, 出力が行政地区の場合はラベル付けもしやすく, 見る側にとってもわかりやすい.

表 1: 住所 DB の形式

地名	候補数	候補 1	候補 2	...
松屋町	8	岐阜県_岐阜市_松屋町	京都府_京都市_下京区_松屋町	京都府_京都市_上京区_松屋町
		京都府_京都市_中京区_松屋町	京都府_京都市_伏見区_松屋町	大阪府_寝屋川市_松屋町
		大阪府_大阪市_中央区_松屋町	香川県_丸亀市_松屋町	
本田	7	山形県_鶴岡市_本田	埼玉県_深谷市_本田	埼玉県_南埼玉郡_宮代町_本田
		新潟県_新潟市_本田	富山県_射水市_本田	岐阜県_瑞穂市_本田
		大阪府_大阪市_西区_本田		
亀田	3	福島県_郡山市_亀田	千葉県_富津市_亀田	新潟県_新潟市_江南区_亀田
旭区	2	神奈川県_横浜市_旭区	大阪府_大阪市_旭区	
パルプ町	1	北海道_旭川市_パルプ町		
芝浦	1	東京都_港区_芝浦		

3. コーパス

本研究で使用するコーパスは拡張固有表現タグ付きコーパス [5] の毎日新聞ジャンルである。このコーパスには固有表現として様々なタグが付与されているが、本論文では地名を扱うためタグは City タグのみを利用する。このコーパスに存在する事例数、つまり曖昧性を解決すべき地名の数はのべ 1723、異なりで 373 である。また異なりで曖昧性を解決すべき地名の平均候補数は 10 である。曖昧性の定義は 4. 章で説明する。

4. 住所データベース

本研究では候補の情報を得るために住所データベース（以降、住所 DB）を参照する。住所 DB とは日本中全ての住所を含んだデータベースであり、実世界上の住所と一致していなくてはならない。しかし、住所は市区町村の合併などにより流動性をもったものであるから必ずしも住所 DB と実際の住所が合致しているとは限らない。本論文では住所 DB のおおよそのデータとして国土交通省の発行している「街区レベル位置参照情報」*1 を利用した。今回使用するこのデータのサンプリングは 2016 年である。

本論文では一つの地名 t に対してどのような候補が存在しているのかを知りたいため、上記データに前処理を施し、表 1 の形式のデータを得た。例えば、地名「パルプ町」の場合、この名前をもつ地理的位置は北海道に 1 地点存在するだけであるので曖昧性がない。一方、「旭区」は 2 地点存在するため曖昧性の解消が必要である。なお、候補は都道府県名からの住所を含んでいるため候補間に曖昧性はない。住所 DB のエントリ数は 152,937 である。この内、候補数が 2 以上になる（曖昧性がある）エントリの割合は約 5% であり、これらエントリにおける平均候補数は 4.49 である。

5. TRIPDL

本研究のベース手法である TRIPDL について述べる。まず文書 d_k 中にあらわれる単語 w_j の生起確率 $\tilde{\theta}_{d_{kj}}$ を計算する。式 (1) が求める式である。

$$\tilde{\theta}_{d_{kj}} = \frac{\#(w_j, d_k)}{\sum_{w_l \in V} \#(w_l, d_k)} \quad (1)$$

$\tilde{\theta}_{d_{kj}}$ は実際に訓練データから計算する単語の生起確率だが訓練データが少ない場合に、この値だけでは多くの単語の生起確率が 0 となってしまうためスムージングを施した $\theta_{d_{kj}}$ を考える。このスムージングは good-turing 推定の考え方に基

づく。まず、係数となる α_{d_k} を求めるために式 (2) を計算する。これは文書 d_k 中に一度だけ出現した単語 w_j の生起確率である。

$$\alpha_{d_k} = \frac{|w_j \in V.s.t. \#(w_j, d_k) = 1|}{\sum_{w_j \in V} \#(w_j, d_k)} \quad (2)$$

次に単語の生起確率 $\tilde{\theta}_{d_{kj}}$ が 0 のときに使用する値 $\theta_{D_j}^{(-d_k)}$ である。その式が式 (3) である。なお、式 (3) で使用している θ_{D_j} は使用する全ての文書集合 D における単語 w_j の生起確率である。

$$\theta_{D_j}^{(-d_k)} = \frac{\theta_{D_j}}{1 - \sum_{w_l \in d_k} \theta_{D_l}} \quad (3)$$

最後に式 (2) と式 (3) を用いて求めた $\theta_{d_{kj}}$ である。

$$\theta_{d_{kj}} = \begin{cases} \alpha_{d_k} \theta_{D_j}^{(-d_k)}, & \text{if } \tilde{\theta}_{d_{kj}} = 0 \\ (1 - \alpha_{d_k}) \tilde{\theta}_{d_{kj}}, & \text{o.w.} \end{cases} \quad (4)$$

セル c_i 中にあらわれる単語 w_j の生起確率 $\theta_{c_{ij}}$ について求める。まず以下の式 (5) を計算し、訓練データにおけるセル中にあらわれる単語の生起確率 $\tilde{\theta}_{c_{ij}}$ を求める。

$$\tilde{\theta}_{c_{ij}} = \frac{\sum_{d_k \in c_i} \#(w_j, d_k)}{\sum_{d_k \in c_i} \sum_{w_l \in V} \#(w_l, d_k)} \quad (5)$$

この後の処理は式 (2) から (4) と同様に行う。それが式 (6) から式 (8) である。

$$\alpha_{c_i} = \frac{|w_j \in V.s.t. \#(w_j, c_i) = 1|}{\sum_{w_j \in V} \#(w_j, c_i)} \quad (6)$$

$$\theta_{C_j}^{(-c_i)} = \frac{\theta_{C_j}}{1 - \sum_{w_l \in c_i} \theta_{C_l}} \quad (7)$$

$$\theta_{c_{ij}} = \begin{cases} \alpha_{c_i} \theta_{C_j}^{(-c_i)}, & \text{if } \tilde{\theta}_{c_{ij}} = 0 \\ (1 - \alpha_{c_i}) \tilde{\theta}_{c_{ij}}, & \text{o.w.} \end{cases} \quad (8)$$

式 (4) と式 (8) から文書中とセル中における単語の生起確率が求められたので、最後に入力文書 d_k がどの候補セル c_i と最も KL 距離が近いかを求めそのセルが出力 $\hat{c}$ となる。つまり式 (9) となる。

*1 <http://nlftp.mlit.go.jp/isj/index.html>

$$\hat{c} = \arg \min_{c_i \in C_{d_k}} KL(\theta_{d_k} || \theta_{c_i}) \quad (9)$$

$$= \arg \min_{c_i \in C_{d_k}} \sum_{w_j \in V_{d_k}} \theta_{d_k j} \log \frac{\theta_{d_k j}}{\theta_{c_i j}} \quad (10)$$

候補 c_i は住所 DB のエントリ数だけあり、その数は非常に多い。そのため、疎データの状態に陥りやすい。

6. 拡張手法

5. 章に示した TRIPDL を 2. 章で述べたようにセルでなく行政地区を用いて実装する場合、ある 1 文書に対応する可能性のある候補は 4. 章より、いくつかの候補住所に一意に決まる。TRIPDL ではこれらの候補住所（セル）の単語確率分布は教師がある限り必ず 1 つはあるが、今回の設定では「いくつかの」候補住所に絞られているので、どの候補住所も単語確率分布を持たない可能性がある。その場合は候補間の確率分布がない、つまり等距離の状態となり出力なしになってしまう。これを解決するために、1 文書に対して複数の候補を学習させる方法を考える。

TRIPDL では式 (5) の学習の際、1 文書 d_k に対して 1 セル c_i を学習させる。その部分を 1 セルではなく複数のセルに広げたのが拡張手法である。つまり式 (5) を書き換えると式 (11) になる。ここで、 $\text{Widen_city}()$ とは 1 セルを与えたときにそのセルを広げた、複数のセルを返す関数である。

先に述べたように今回の設定ではセルではなく行政地区なので $\text{Widen_city}()$ には大字レベルまたは市区町村レベルの住所が入力される。その場合はその住所の属する市区町村に属する全ての住所候補が返される。例えば「茨城県つくば市天久保」という大字レベルの住所が入力された場合、「茨城県つくば市」に属する住所候補（「茨城県つくば市」、「茨城県つくば市天久保」、「茨城県つくば市桜」など）が全て返される。

また、式 (12) については関数部分のみが違うだけで $\text{Widen_city}()$ が $\text{Widen_pref}()$ に替わっている。 $\text{Widen_pref}()$ とは市区町村でなく入力候補の都道府県に属する住所候補全てを返す関数である。拡張手法では式 (5) の代わりに式 (11) あるいは式 (12) を用いる。つまり、市区町村レベルや都道府県レベルまで正解の解釈を拡げて学習するのが拡張手法である。

$$\hat{\theta}_{c_{ij}} = \frac{\sum_{c_m \in \text{Widen_city}(c_i)} \sum_{d_k \in c_m} \#(w_j, d_k)}{\sum_{c_m \in \text{Widen_city}(c_i)} \sum_{d_k \in c_m} \sum_{w_l \in V} \#(w_l, d_k)} \quad (11)$$

$$\hat{\theta}_{c_{ij}} = \frac{\sum_{c_m \in \text{Widen_pref}(c_i)} \sum_{d_k \in c_m} \#(w_j, d_k)}{\sum_{c_m \in \text{Widen_pref}(c_i)} \sum_{d_k \in c_m} \sum_{w_l \in V} \#(w_l, d_k)} \quad (12)$$

7. 評価実験

7.1 実験の設定

7.1.1 実験条件

教師なし学習の先行手法である POPULATION[4] と文脈参照法 [6] と MENTION.COUNT[6]、教師あり学習の先行手法である TRIPDL、拡張手法の市区町村レベル、拡張手法の都道府県レベルについて正解率・拡大正解率を測定する実験を行っ

た。また TRIPDL と拡張手法では、以下のように素性を使う単語を変えてもそれぞれ実験を行った。数詞を除く名詞全てを使った場合、固有表現 ORGANIZATION (O) のみを使った場合、固有表現 PERSON (P) のみを使った場合、固有表現 LOCATION (L) のみを使った場合、固有表現 ARTIFACT (A) のみを使った場合、固有表現 O,P,L,A(OPLA) のみを使った場合について実験を行った。文脈参照法において参照する文字数 k を指定する必要がある。今回はコーパスの 1 記事あたりの平均文書長 $L = 1,657$ のとき、前方文脈を k_1 、後方文脈を k_2 とすると $k_1 = k_2 = \frac{L}{2} = 829$ とした。今回の設定ではコーパス内にあるターゲットは入力として与えられる。また TRIPDL は本来 document geolocation の手法であるが、本研究では toponym resolution を扱うため、その間の違いを吸収しなくてはならない。そのため、入力の toponym に対してその前後 829 文字（文脈参照法の k と同様）を結合した文字列を擬似 document とし、TRIPDL で解いた。

7.1.2 データセット

評価を行なうコーパスは 3. 章に示したものを使用する。候補データ生成に用いる住所データベースは 4. 章に示したものを利用する。POPULATION に用いる人口 DB は政府が発行している人口統計データ「e-stat 統計で見る日本」*2 の調査年 2015 年のデータを用いる。MENTION.COUNT で用いる言及回数 DB を作成する際に使用する生コーパスには毎日新聞社のコーパス [7]1 年分を用いる。

正解データは人手で作成した。拡張固有表現タグ付きコーパスに付与された City タグが付く地名のうち、住所 DB を参照して複数の候補が得られるものについて、その近隣文脈を読むことで正解候補を選択した。今回は毎日新聞ジャンルのみの実験であるため、毎日新聞ジャンル中の候補が複数ある City タグ全 1732 個について正解を付与した。また、作業は基本的に一人で行ったが、アグリーメントのため全 City タグのうちランダムに抽出した 200 個をもう一人も作業を行い、どの程度二人のアノテーション結果が一致するかを検証した。その結果一致しなかったタグは全 200 個中 3 個であり、このアノテーションは一般性を確保できていると考えられる。

7.1.3 評価尺度

評価尺度には正解率と拡大正解率を用いる。正解率とは手法により選択した候補と正解ファイル中の候補とを比較し、一致していた場合に正解、そうでない場合に不正解としたときの候補が複数ある全 City のうちの正解の割合である。拡大正解率とは手法により選択した候補が属する都道府県と正解ファイル中の候補が属する都道府県とを比較し、一致していた場合に正解、そうでない場合に不正解としたときの候補が複数ある全 City のうちの正解の割合である。

7.2 実験結果

実験の結果を表 2 に示す。ここで、不正解を誤選択、正解なし、出力なしの 3 つに分けた。誤選択とは正解と出力がどちらも存在するが一致しない場合であり、使用した手法によって正しい候補を選ぶことができなかったことをあらわす。正解なしとは候補中に正解が存在しない場合である。これは合併などによる原因が考えられる。出力なしとは候補があるにも関わらず、使用した手法で出力を出せなかった場合である。また正解なしの場合は出力の有無はドント・ケアであり、出力がなくても出力なしにはカウントされない。

表 2 より、まず POPULATION と文脈参照法と TRIPDL(名詞) を比較すると、教師なし学習である

*2 <https://www.e-stat.go.jp/regional-statistics/ssdsview>

表 2: 実験結果

手法	拡大正解率	正解率	正解	誤選択	出力なし	正解なし
POPULATION	57.0	56.9	985	207	388	152
文脈参照法	63.5	56.2	974	279	327	152
TRIPDL(名詞)	67.4	67.0	1,161	124	295	152
TRIPDL(O)	62.1	61.8	1,071	190	319	152
TRIPDL(P)	60.1	59.8	1,035	233	312	152
TRIPDL(L)	68.9	68.6	1,188	97	295	152
TRIPDL(A)	55.4	55.2	956	205	419	152
TRIPDL(OPLA)	68.0	67.6	1,171	114	295	152
拡張手法 市区町村 (名詞)	59.4	57.2	991	505	84	152
拡張手法 市区町村 (O)	44.8	43.0	745	735	100	152
拡張手法 市区町村 (P)	32.5	30.5	528	956	96	152
拡張手法 市区町村 (L)	59.9	58.0	1,004	492	84	152
拡張手法 市区町村 (A)	32.6	31.3	542	847	191	152
拡張手法 市区町村 (OPLA)	59.8	57.5	996	500	84	152
拡張手法 都道府県 (名詞)	61.8	53.7	930	650	0	152
拡張手法 都道府県 (O)	43.7	38.3	664	916	0	152
拡張手法 都道府県 (P)	30.8	25.6	443	1,137	0	152
拡張手法 都道府県 (L)	64.0	56.3	975	605	0	152
拡張手法 都道府県 (A)	22.2	17.8	309	1,261	10	152
拡張手法 都道府県 (OPLA)	64.2	56.8	983	597	0	152
MENTION_COUNT	72.1	66.7	1,156	364	60	152

POPULATION, 文脈参照法に比べ教師あり学習である TRIPDL (名詞) は正解率が高いことがわかる。

次に, 同じ TRIPDL でも使用する素性が違う場合をそれぞれ比べる。L が一番正解率が高いことがわかる。これは固有表現の中で一番 LOCATION が地理的位置推定に有効だということを示しており, 直感通りの結果となった。ここで, それぞれの素性の数は, 名詞:271011, O:5736, P:7933, L:16302, A:1593 である。名詞の数は他の素性に比べてかなり多くあるにもかかわらず L に正解率で及んでいないのは全ての名詞を利用するよりも固有表現 LOCATION のみを使ったほうがより小さい計算量で正解率が向上することを示していると考えられる。

次に TRIPDL と拡張手法を比較すると, TRIPDL が一番正解率がよく, 市区町村, 都道府県と範囲を拡げる毎に正解率が悪くなっていることがわかる。不正解の内訳から, 出力なしの数は範囲を拡げる毎に確実に減っており, 候補間が等距離になる問題は解決されていることがわかる。しかしそれと同時に正解は減り誤選択は増えている。

最後に一番正解率が高かった TRIPDL(L) と MENTION_COUNT を比較する。MENTION_COUNT は教師なし学習にもかかわらず, 教師あり学習の TRIPDL と同じ程度の正解率をだしており, 非常に興味深い結果となった。これは MENTION_COUNT では大量の教師なしデータを考慮できる点が性能へ反映されたからと考えられる。

8. おわりに

本論文では新聞記事の地理的位置推定のタスクにおける地名の曖昧性問題について, 曖昧性解消に有効な素性の調査を行った。実験より, 固有表現の中では LOCATION が一番有効な素性であり, 全名詞も同じくらい有効ではあるが, 計算量を考えると全名詞よりも LOCATION のほうが良いという結果が得られた。今後の課題と検討として, 次のものが挙げられる。

- 拡張手法によって正解率が下がってしまったがその原因の追求

- MENTION_COUNT と TRIPDL(L) とのアルゴリズムの突き合わせを行い, 結合方法を検討する

謝辞

本研究の一部は科研費 (18K11982) の助成を受けて実施されました。

参考文献

- [1] Stuart Middleton, Lee Middleton, and Stefano Modafferi. Real-time crisis mapping of natural disasters using social media. 2014.
- [2] 北本 朝展. G 空間 オープンな地名情報システム GeoNLP. THE JOURNAL OF SURVEY 測量, pp6-10, 2014.
- [3] Lianhua Chi, Kwan Hui Lim, Nebula Alam and Christopher J. Butler. Geolocation Prediction in Twitter Using Location Indicative Words and Textual Features. The 2nd Workshop on Noisy User-generated Text. 2016.
- [4] Michael Adrian Speriosu. Methods and Applications of Text-Driven Toponym Resolution with Indirect Supervision. THE UNIVERSITY OF TEXAS AT AUSTIN. 2013.
- [5] 橋本 泰一, 乾 孝司, 村上 浩司. 拡張固有表現タグ付きコーパスの構築. 情報処理学会研究報告自然言語処理, pp113-120, 2008.
- [6] 関 龍, 乾 孝司. 局所文脈と関連文書を用いた地名に対する地理的位置の同定. 人工知能学会, 2018.
- [7] 毎日新聞社. CD-毎日新聞 2004 データ集.

インタラクティブなデータ・ヴィジュアルライゼーション・ツールを用いた Twitter データのクラスタ分析

Cluster analysis of Twitter Data, using Interactive Data visualization Tool.

和田 伸一郎^{*1}
WADA, Shinichiro

立教大学大学院社会学研究科
Graduate School of Sociology, Rikkyo University.

Abstract: This study attempts cluster analysis of Twitter data posted on Tokyo Governor's Election held in 2016, using Python (July 13 - August 1, 2016, 4.8 million tweets, 170 million words). For cluster analysis, words were vectorized using gensim version word2vec algorithm which is a library of Python, and attempt to visualize clusters in three dimensions using t-SNE (t-distributed Stochastic Neighbor Embedding) which is dimensionality reduction algorithm. In particular, in this research, we used the data visualization tool Embedding Projector for clustering. By using this tool, we attempted to visually identify clusters by moving the three-dimensional space interactively while visualizing the dynamic learning process in the three-dimensional space. As a result, we could identify multiple clusters with high accuracy. This made it possible to clarify what in this election Twitter users were interested in.

1. はじめに

本研究では、2016 年 7 月に行われた東京都知事選挙に関する Twitter データのうち、選挙期間中(とその前後一日ずつを含む)に投稿されたものを全数収集し、クラスタ分析を行うことによって、選挙に関する、どのような投稿がなされたのかを分析した。

2. データと前処理について

2.1 収集した Twitter データについて

Twitter データは、ユーザーローカル社(東京都港区)の特別な協力を得て、2016 年 7 月 13 日～8 月 1 日の間の以下の検索ワードを含む全数データを収集することができた。検索ワードは次のとおりである。「小池 OR 増田 OR 鳥越 OR 百合子 OR 寛也 OR 俊太郎 OR 都知事選 OR 都知事選挙 OR 知事選挙 OR 知事選」。その結果、表1のようなデータを csv ファイルにて収集することができた。RT は公式リツイートのみ数、OT はオリジナルツイートを指す。

表1

	総ツイート数	語彙数
ALL	4,825,560	199,287
RT	3,588,302	123,192
OT	1,237,258	187,394

なお ALL データ(RT データと OT データを足したすべてのデータ)の単語数は、177,439,525 語となった。

この表からも分かるように、このデータのうち、約 74%がリツイートからなっていた。語彙数(単語の種類)でみても、RT の語彙数が OT よりも少ないのは、同じツイートが数多くリツイートされているためである。したがって、Twitter データの場合、しばしばテキストマイニングで行われる、単語の出現回数を出しても、あまり意味がない。多くリツイートされた文章に含まれる単語の出現回数が多くなってしまったためである。

そこで、単語の類似度(意味の近さ)を計算する、Python のライブラリ gensim に実装されている word2vec を使って、単語のベ

クトル化を行い、どのような類似する単語群、つまりクラスタがあるかを調べることにした。なお、単語ベクトル化をする際に設定した word2vec のオプションの値は以下の通りである。sg=1, size=300, min_count=5, window=10, hs=0, negative=5, iter=10, sample=0.001 (オプションについての詳しい説明については、GitHub で公開されている Python ライブラリ gensim のソースコード(word2vec.py)に記載されている)。

2.2 データの前処理(形態素解析)

その前に、これらのデータをそれぞれ形態素解析する必要がある。日本語の形態素分析エンジンとして、最も有名なのは Mecab である(Kudo (2013))。またそれとセットにしばしば使われる辞書が、Mecab-ipadic 辞書である。ここで問題になるのは、とりわけ SNS テキストデータには、それぞれの SNS プラットフォームに固有のスラングや、多岐にわたるトピックごとに多種多様な語彙群が存在することである。なぜ問題なのかといえ、ipadic は、標準的な辞書レベルの語彙を十分網羅的に含んでいるが、特殊な語彙を欠いているからである。これを解決するために本研究では、Sato (2015)によって、いまなお定期的に更新され続けている、こうしたネット上のスラングなどを多く含む Mecab-ipadic-NEologd 辞書を用いた。

とりわけ、ipadic 辞書が SNS データ分析にとって致命的なのは、氏名を一つの単語として認識しないことである。都知事選の場合でいえば、「小池」とカウントされた単語が「小池百合子」なのか共産党議員の「小池晃」なのかが分からない。つまり、「小池」と「百合子」、「晃」が分解されて学習されてしまう。さらには、「桜井」とカウントされた単語が、候補者の1人であった元在特会会長の「桜井誠」なのか、一時期候補すると噂された「桜井俊」(アイドルグループ嵐のメンバーの桜井翔の父親)なのか、保守派論客である「桜井よしこ」なのか、が分からないということが起きる(これらの氏名はすべて分解されてしまう)。こういったことがクラスタリングで大きな欠陥となりうる理由は、それらの氏名が、相当異なる文脈に出現する可能性が高い以上、氏名が分解されてしまうと、それぞれの文脈の差異が学習不能になってしまうからである。これだと、いくら word2vec, t-SNE などのアルゴリズムの精度が高くても、学習結果が混乱したものになる可能性が高くなってしまふ。

連絡先: 和田伸一郎 立教大学大学院社会学研究科, 東京都豊島区西池袋3-34-1

3. word2vec と Embedding Projector による可視化

約 20 万もある語彙の分布のどこにクラスタがあるのかを調べることは、単語ごとに類似する単語をリスト化してくれる word2vec の Python スクリプトを実行するだけでは難しい。例えば、約 20 万もの単語すべてに類似単語リストをつくることは、現実的に難しい。

そこで、コーパス全体を可視化した上で、その全体の中に、どのようなクラスタが現れているかを視覚的に把握するツールとして、Google 社がオープンソースで提供しているディープラーニングフレームワークである TensorFlow のパワフルな可視化ツール、TensorBoard のスタンドアローン版 WebUI である Embedding Projector を用いることにした。

word2vec を使って 300 次元で単語ベクトル化したコーパスを、Embedding Projector 上で、次元圧縮アルゴリズムである PCA、t-SNE を用いて、選定した特徴語とそれに類似する 1000 語からなる、三次元のローカル空間を可視化し、クラスタを出すことを試みた。特徴語として、選挙期間中に大きな争点となっていた「待機児童」、「介護」、また、全国的に注目されていた「オリンピック／パラリンピック」などを選定した。なお、「待機児童」、「介護」、「オリンピック／パラリンピック」を含んだツイートは、それぞれ、43,794, 26,955, 78,006 存在した。

図1は、「待機児童」をターゲット単語として設定し、2000 回学習した結果、収束した、その単語と意味の近い 1000 語からなる三次元のローカル空間である。

4. 分析結果と課題

いくつか特徴語を選定し、学習を行った結果、日本語でも高い精度でクラスタを確認することができた。この精度の高さは、まずは word2vec, PCA, t-SNE のアルゴリズムの精度の高さによりもたらされたものであるが、加えて、NEologd 辞書が、例えば、候補者や有識者などの名前や政党名、政治団体名、イベント名、地名、市民会館などの建物の名前といった固有名詞を一語で拾ってくれたことからもたらされたものでもあることが分かった。

他にも、Instagram のテキストデータを学習させ、高い精度でクラスタが出ることも確認できた。

課題としては、今回、word2vec に用意されているオプションの値を変えて、それぞれのデータで、最も精度が高い学習結果を比較する余裕がなかったことを挙げるができる。これについては、今後検討したい。

参考文献

- [Smilkov 16] Google Developers (Smilkov, Daniel, Viégas, Fernanda, Wattenberg, Martin.) : A.I. Experiments: Visualizing High-Dimensional Space. <https://www.youtube.com/watch?v=wvsE8jmlGzE&feature=youtu.be> (2016)
- [木田 17] 木田勇輔 : ソーシャルメディアとポピュリストの動員 —2016 年東京都知事選挙における Twitter データの分析から—, 文化情報学部紀要, 相山女学園大学, 第 17 巻, pp.83—92 (2017)
- [Kudo 13] Kudo Taku. : MeCab:Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/> (2013)
- [Maaten 08] Maaten, Laurens van der, and Hinton, Geoffrey. : Visualizing data using t-SNE. Journal of Machine Learning Research, Vol 9(Nov), pp. 2579—2605. (2008)
- [Mikolov 13a] Mikolov, Tomas, Chen, Kai, Corrado, Greg, Dean, Jeffrey. : Efficient estimation of word representations in vector space”. CoRR, abs/1301.3781. (2013)
- [Mikolov 13b] Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Gregory S., Dean, Jeffrey. : Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States, pages 3111—3119 (2013)
- [Sato 15] Sato Toshinori. : Neologism dictionary based on the language resources on the Web for mecab-ipadic”, <https://github.com/neologd/mecab-ipadic-neologd/> (2015)
- [Smilkov 16] Smilkov, Daniel, Thora, Nikhilt, Nicholson, Charles, Reif, Emily, Viégas, Fernanda, Wattenberg, Martin. : Embedding Projector: Interactive Visualization and Interpretation of Embeddings. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain (2016)
- [山縣 18] 山縣史哉, 梅原英一 : 平成 28 年度東京都知事選挙の Twitter 分析, 信学技報, 電子情報通信学会 (2018)

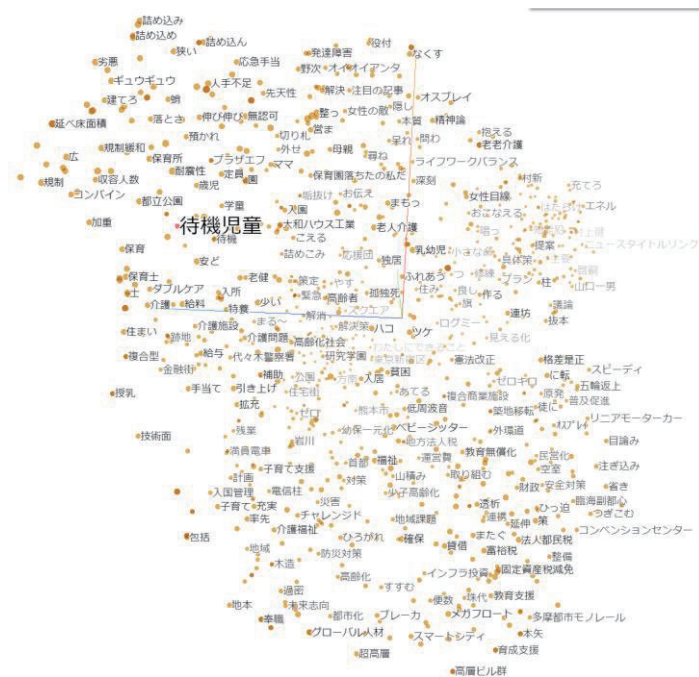


図1. 「待機児童」ローカル空間

Embedding Projector では、学習中であってもインタラクティブに三次元空間を左右上下自由に回転させることができ、また、ズームイン、ズームアウトすることによって、どの辺りにクラスタが出現しつつあるかを、直感的に目視で判断することができる。このツールの大きなメリットとして挙げることができるのは、機械学習や深層学習の欠点としてしばしば指摘される、学習の「ブラックボックス」性を、このツールが一定程度、取り除いてくれることである。

係り受け誤り埋め込み表現のクラスタリングによる 複数のドメイン間での追加訓練の効果の比較分析

Comparative Analysis of the Effect of Additional Training between Multiple Domains by
Clustering of the Embeddings of Parsing Errors

原 拓也^{*1}

Takuya Hara

松崎 拓也^{*1}

Takuya Matsuzaki

横野 光^{*2}

Hikaru Yokono

佐藤 理史^{*1}

Satoshi Sato

^{*1}名古屋大学工学研究科

Graduate School of Engineering, Nagoya University

^{*2} (株) 富士通研究所

Fujitsu Laboratories Ltd.

We propose a method for studying the effect of additional training of Japanese dependency parsers across multiple domains from a bird's-eye view. We collected the parsing errors of a parser before and after additional training using target domain data. We then conducted cluster analysis of the parsing errors represented as dense real vectors, which are obtained from the internal states of the parser. Through quantitative and qualitative analysis of the clusters, we could grasp the types and the numbers of the parsing errors across multiple target domains.

1. はじめに

機械学習における課題の一つにドメイン適応問題がある。多くの場合、訓練に用いたデータのドメイン（以下、訓練ドメイン）と、適用の対象となるデータのドメイン（以下、評価ドメイン）が異なるとタスクの精度が低下する。精度の低下の原因は訓練ドメインと評価ドメインに含まれている文の構造や単語、あるいは正解の分布が異なっているからである。この精度の低下を防ぐという課題をドメイン適応問題という。

本研究では、ドメイン適応問題の対象となるタスクとして係り受け解析を取り上げる。係り受け解析とは文を構成している文節同士の係り受け関係（修飾・被修飾）を解析するタスクである。係り受け解析の例を図1に示す。図の中の四角が文節を表しており、矢印が係り受け関係を表している。それぞれの矢印は係り文節（修飾文節）から受け文節（被修飾文節）へと伸びている。

ドメイン適応問題は、評価ドメインデータを用いて追加訓練を行うことである程度解決できることが多い。評価ドメインデータの量は多くの場合限られており、それらを効率的に用いるための方法に関する研究は、多数存在する。例えば Daumé III と Marcu [Daumé III 06] は、追加訓練を行う際の評価ドメインデータへの適切な重み付け方法を提案した。また Daumé III [Daumé III 07] はドメイン固有の情報とドメインに依存しない情報を分けて訓練を行えるように素性を設計した。他にも、効率的な学習を行うために、次にアノテーションすべきデータを逐次的に見つける方法（アクティブラーニング）に関する研究も多数ある。

上で述べた研究では、適用する評価ドメインそれぞれに対し、追加訓練用データを用意する必要がある。しかし、評価ドメインのデータを多数のドメインについて用意することはコストの点から現実的でない。一方、追加訓練がなぜ有効なのかは、追加訓練によって解消された誤りや、解消されなかった誤りを一つ一つ見ていてもよくわからないことが多い。これに対し、追加訓練の効果を俯瞰的に分析することで、より少ないアノテーションで、効果的に適応を行うための知見が得られると期待できる。そこで、本研究では係り受け解析器のドメイン適応を対象に、追加訓練の効果を俯瞰的に分析するための手法として、係り受け誤りのクラスタリングを提案する。



図 1: 係り受け解析の例

クラスタリングを行うためには係り受け誤りの間の類似性を数量的に捉える必要がある。そのため、本研究では係り受け解析器の内部状態を用いて係り受け誤りを埋め込み表現（密ベクトル表現）に変換する。

本論文では特に、複数の評価ドメインの間での追加訓練の効果の違いについて分析する。この分析は、評価ドメインデータを用いた追加訓練の前後における係り受け誤りを埋め込み表現に変換し、それらに対しクラスタリングを行い、得られた結果を複数ドメイン間で比較することで行われる。本論文では訓練ドメインとして新聞記事、評価ドメインとして中学校理科教科書及び特許文書を用いてクラスタリングを行った。クラスタに含まれる誤りの数の分布や、クラスタごとの誤りの性質を調べることで、ドメインごとの特徴を比較・分析することができた。

以下、2 節で提案手法について説明する。3 節では実際のデータを用いた分析を行う。

2. 係り受け誤りのクラスタリング

2.1 係り受け解析器

本研究では、LSTM に基づく係り受け解析器の誤りを分析する。また、誤りの埋め込み表現を得るために、この解析器の内部状態を用いる。本研究で用いる解析器は松野ら [松野 18] のものをベースにした。松野らのモデルは大きく分けて、(i) 単語層 LSTM による単語表現の獲得、(ii) 文節層 LSTM による文節表現の獲得、(iii) 多層パーセプトロンによる次元圧縮、(iv) Biaffine 変換による係り受け関係スコアの計算の 4 段階に分けられる。本研究では (i) 及び (ii) はそのまま、(iii) 及び (iv) の代わりに多層パーセプトロンによる係り受け関係スコアの計算を行う。

具体的には以下のモデルを用いる。最初に文の各単語の品詞（4 レベル）・活用形・活用タイプ・基本形をそれぞれ埋め込み層に通し、それらを結合させることで単語分散表現とする。文の各単語の単語分散表現をまとめて系列にしたものを x^{word}

として単語層 LSTM に入力する。この時 (i) の操作は以下の式のように行われる。

$$(\overleftarrow{\mathbf{y}}_i^{word}, \overrightarrow{\mathbf{y}}_i^{word}) = \text{BiLSTM}^{word}(\mathbf{x}_i^{word}) \quad (1)$$

t 番目の文節が i 番目から j 番目の単語で構成されている時、文節層 LSTM への入力 $\mathbf{x}_t^{chunk}$ は以下のように計算される。

$$\overleftarrow{\mathbf{x}}_t^{chunk} = \overleftarrow{\mathbf{y}}_i^{word} - \overleftarrow{\mathbf{y}}_{j+1}^{word} \quad (2)$$

$$\overrightarrow{\mathbf{x}}_t^{chunk} = \overrightarrow{\mathbf{y}}_{i-1}^{word} - \overrightarrow{\mathbf{y}}_j^{word} \quad (3)$$

$$\mathbf{x}_t^{chunk} = \overleftarrow{\mathbf{x}}_t^{chunk} \oplus \overrightarrow{\mathbf{x}}_t^{chunk} \quad (4)$$

t 番目の文節に対応した文節表現を $\mathbf{y}_t^{chunk}$ とする。このとき (ii) の操作は以下の式のように計算される。

$$(\overleftarrow{\mathbf{y}}_t^{chunk}, \overrightarrow{\mathbf{y}}_t^{chunk}) = \text{BiLSTM}^{chunk}(\mathbf{x}_t^{chunk}) \quad (5)$$

$$\mathbf{y}_t^{chunk} = \overleftarrow{\mathbf{y}}_t^{chunk} \oplus \overrightarrow{\mathbf{y}}_t^{chunk} \quad (6)$$

t 番目の文節が u 番目の文節に係る際 ($t < u$) の係り受け関係スコア $s_{t,u}$ は、以下のように定義される。

$$s_{t,u} = \text{MLP}(\mathbf{y}_t^{chunk} \oplus \mathbf{y}_u^{chunk} \oplus \mathbf{d}(t, u)) \quad (7)$$

$$\mathbf{d}(t, u) = [u - t = 1] \oplus [u - t \leq 5] \quad (8)$$

ただし $[\cdot]$ は内部の条件が真ならば要素が 1 のベクトルを、偽ならば要素が 0 のベクトルを表す。

この係り受け関係スコアモデルを用いて内元ら [内元 99] のアルゴリズムで訓練・解析を行う。内元らのアルゴリズムでは、後方の文節から順に係り先を決定する。その際、ある係り文節に対し、候補となる受け文節の集合の中から 1 つの文節を選択する必要がある。訓練時は、係り文節と正しい受け文節のペアの係り受け関係スコアが、係り文節とその他の受け文節の候補のペアの係り受け関係スコアよりも高くなるように訓練する。具体的には、ある係り文節 t に対し、正しい受け文節を u 、候補文節集合を C_t とするとき、損失関数を以下のように定義する。

$$loss = - \sum_{u' \in C_t} \log \frac{\exp(s_{t,u})}{\exp(s_{t,u}) + \exp(s_{t,u'})} \quad (9)$$

解析時は、それぞれの係り文節に対する受け文節候補集合のうち、係り受け関係スコアが最も高くなるものを選択する。

2.2 係り受け誤りの収集法

本研究では評価ドメインデータによる追加訓練を行う前後で、解析結果の誤り方がどのように変化するかを観察する。この目的のため、以下のように訓練データ及び解析対象のデータが異なる 3 タイプの誤り例を収集した。

(1) **訓練ドメイン誤りの収集** 最初に、訓練ドメインのデータの半分を用いてベースモデルを作る。訓練ドメインのデータの残り半分をベースモデルで解析した時の誤りを収集する。

(2) **追加訓練前の評価ドメイン誤りの収集** 評価ドメインデータの全てをベースモデルで解析した時の誤りを収集する。

(3) **追加訓練後の評価ドメイン誤りの収集** 評価ドメインデータに対し 4 分割交差検証を行う。ただし、ベースモデルの多層パーセプトロン (式 (7)) の部分のみを、ベースモデルのパラメータ値を初期値として、評価ドメインデータを用いて訓練する (その結果を以降では追加訓練済みモデルと呼ぶ)。その結果得られた誤りを収集する。

追加訓練において多層パーセプトロンのみを訓練することで、すべてのモデルについて文節の埋め込み表現は共通となる。これにより係り受け誤りの埋め込み結果 (2.3.1 項) を比較することができる。なお、本研究では追加訓練済みモデルで訓練ドメインデータを解析した際の誤りは用いない。理由は本研究の目的が追加訓練による評価ドメイン誤りの変化の観察であり、訓練ドメイン誤りの変化には興味がないためである。

2.3 係り受け誤りの分析

係り受け誤りを俯瞰的に観察するための手法としてクラスタリングを行いたい。クラスタリングを行うためには、誤りの間に距離を定義する必要がある。そこで最初に係り受け誤りを埋め込み表現 (ベクトル) に変換し、ユークリッド距離を誤り間の距離とする。

2.3.1 係り受け誤り埋め込み表現

係り受け誤り埋め込み表現 (以降、係り受け誤り表現) は係り受け誤りの 3 要素である係り文節・正しい受け文節・誤った受け文節とそれらの間の距離素性を埋め込んだものである。

2.1 項で述べた係り受け解析器には、中間層として文節表現 ((6) 式の $\mathbf{y}_t^{chunk}$) がある。文節表現は文節及び周辺文脈から、係り受け解析に必要な情報を抽出したものだ考えることができる。そこで係り文節・正しい受け文節・誤った受け文節に対応する文節表現を結合し、さらに距離素性を結合したものは係り受け誤り全体の特徴を表現するものと考えられる。係り文節を t 、正しい受け文節を u 、誤った受け文節を $\tilde{u}$ として、係り受け誤り表現 $\mathbf{e}_{t,u,\tilde{u}}$ は以下のように作成する。

$$\mathbf{e}_{t,u,\tilde{u}} =$$

$$\begin{cases} \mathbf{y}_t^{chunk} \oplus \mathbf{y}_u^{chunk} \oplus \mathbf{y}_{\tilde{u}}^{chunk} \oplus \mathbf{d}(t, u) \oplus \mathbf{d}(t, \tilde{u}) & (u < \tilde{u}) \\ \mathbf{y}_t^{chunk} \oplus \mathbf{y}_{\tilde{u}}^{chunk} \oplus \mathbf{y}_u^{chunk} \oplus \mathbf{d}(t, \tilde{u}) \oplus \mathbf{d}(t, u) & (u > \tilde{u}) \end{cases}$$

上式は、二つの受け文節候補 u と $\tilde{u}$ のうち、正解がどちらであるかとは無関係に、文中での出現順に従って埋め込み表現を構成することを表す。これにより、係り受けアノテーションされていないデータについても、例えば係り受け関係スコア 1 位と 2 位のペアを用いて、同様に埋め込み表現を作成できる。

2.3.2 クラスタリング

本研究では、係り受け誤り表現を以下の手順で分析する。

1. 訓練ドメイン誤りの埋め込み表現を k-means 法でクラスタリングする。
2. 評価ドメイン誤りそれぞれの埋め込み表現 e について、1. で得られたクラスタのうち、 e からクラスタ重心までの距離が最も近いものに e を割り当てる。

この手順は数の多い訓練ドメイン誤りを類型ごとにグループ化した後、評価ドメイン誤りのそれぞれを、近い類型をもつクラスタへと分類することを意図している。

3. 実験

この節では最初に実験に用いたデータの統計について述べ、次に係り受け誤りの収集に関する詳細を述べる。最後にクラスタリングの結果とその分析について述べる。

3.1 利用データ

訓練ドメインデータとして新聞記事 (京都大学テキストコーパス [黒橋 97]) を用いた。また、評価ドメインデータとして理科教科書及び特許文書を用いた。理科教科書データは東京書籍「新編新しい科学 (中学 1 年生、中学 2 年生)」のテキス

表 1: データの統計

データ	文数	文節数	文節数/文
新聞記事	38,400	372,130	9.7
理科教科書	2,460	22,923	9.3
特許文書	953	11,157	11.7

表 2: 係り受け誤りの収集結果

対象	モデル	誤り文節数	精度
新聞記事	ベースモデル	13,878	91.39
理科教科書	ベースモデル	2,429	88.13
	教科書訓練済みモデル	1,671	91.83
特許文書	ベースモデル	1,465	85.64
	特許訓練済みモデル	1,143	88.80

ト、特許文書は高分子化合物に関する特許の実施例の箇所にそれぞれ係り受け情報を付与したものである。それぞれのデータの統計値を表 1 に示す。表に示すように、本論文における実験では、理科教科書及び特許文書の量は新聞記事の 10 分の 1 未満と少量である。

3.2 係り受け誤りの収集

2.1 項で述べた係り受け解析器の訓練は以下のように行った。新聞記事の全体の 50 % をベースモデルの訓練データ、残りの 50 % を開発データ兼評価データとして用いた。学習データの中で出現数が 1 回だった単語は UNK トークンに置き換えた。単語の品詞 (4 レベル)、活用形、活用タイプはそれぞれ 20 次元に、単語の基本形は 200 次元に埋め込んだ。単語層・文節層 LSTM は隠れ層 1 層 (300 次元) のものを用いた。多層パーセプトロンは隠れ層 1 層 (300 次元) のものを用いた。訓練アルゴリズムは学習係数 0.01 の AdaGrad で、ミニバッチサイズは 32 文、重み減衰係数は 0.00001、勾配の大きさは 5 に制限して訓練を行った。またエポック数は最大 32 エポックだが、開発データに対する損失が最も少ないものを最終結果とした。実装には Chainer [Tokui 15] を用いた。追加訓練は理科教科書・特許文書それぞれについて 4 分割交差検証の要領で行った。^{*1}

次に 2.2 項で述べた方法で係り受け誤りを収集した。その結果を表 2 に示す。ベースモデルによる新聞記事に対する解析精度は、松野ら [松野 18] と同程度であった。適応先ドメインデータによる追加訓練を行うことで精度が向上していることがわかる。ベースモデルによる解析精度が理科教科書と特許文書で 2.5 ポイントほどの差があるので、特許文書は理科教科書より解析が難しい (新聞記事との違いが大きい) と考えられる。

3.3 クラスタリング結果と分析

2.3 項で述べた方法で新聞記事、理科教科書、特許文書それぞれの係り受け誤りを埋め込み表現に変換した。またクラスタ数を 30 としてクラスタリングを行った。さらに追加訓練前後における誤りの傾向の変化を調べるため、各クラスタに含まれていた追加訓練前の誤りの数・追加訓練後の誤りの数・追加訓練によって解消できた誤りの数を調べた。その結果を図 2 (理科教科書)、図 3 (特許文書) に示す。グラフの縦軸は各クラスタに含まれる新聞記事誤りの数に対する評価ドメインデータ誤りの比 (百分率) を表している。つまり、この比が相対的

^{*1} 新聞記事・理科教科書は文単位で分割し、訓練に用いたが、特許文書は文書単位で分割した。その理由は特許文書には同じ文書内では似た文が連続するという特徴があり、それらが訓練データと評価データに分かれると、汎化性能を正しく評価できないからである。

表 3: 評価データにおける「…/係り文節/…/名詞+と/用言」という文型を含む文の数 (出現数) と正しい係り先

データ	出現数	正しい係り文節	
		名詞+と	用言
新聞記事	594	378	216
理科教科書	274	36	238
特許文書	12	5	7

に高い値になるクラスタは、評価ドメインに特有の誤りの類型に対応すると考えられる。グラフから、ほぼ全てのクラスタにおいて、追加訓練によって誤りの数が減少していることがわかる。またクラスタ間で誤りの分布は様ではなく偏っており、ドメインごとに固有の分布になっていることもわかる。本論文では、クラスタ内の誤りの傾向が分かりやすかったクラスタ 13 とクラスタ 24 についてその特徴を述べる。以降では係り受け誤りの例を表す時、係り文節を下線、正しい受け文節を ○ のついた波線、誤った受け文節を × のついた波線で表す。また誤り例の通し番号の最初の文字が「N」ならば新聞記事を、「T」ならば理科教科書を、「P」ならば特許文書を表す。

クラスタ 13 クラスタ 13 に属する典型的な新聞記事誤りと理科教科書誤りを以下に示す。

N-13 :
立証負担が/軽く、/証拠も/そろえやすい/ことから/東京
地検は/「製造の/企て」に/○絞_り込_んだ_と/×聞_く。
T-13 :
ドルトンは、/19世紀の/初めごろ、/物質は/それ以上/分割する/ことの/できない/小さな/粒子で/×できていると/○考_えた。

クラスタ 13 に属する誤りは、正しい受け文節と誤った受け文節のいずれか一方が引用を表す助詞「と」を含み、他方が直後の動詞であるパターンが多かった。理科教科書ではその中でも誤った受け文節が助詞「と」を含んでおり、正しい受け文節がその直後であるパターンの方が多かった。評価データにおける、この文型の出現数を表 3 に示す。表から、新聞記事では「名詞+と」の文節にかかる割合がその直後の文節に係る割合の 2 倍程度だが、理科教科書では用言文節にはるかに大きく偏っていることがわかる。これは理科教科書では説明文 (「…を/…と/いう」) が非常に多いからである (238 個中 213 個)。ここから、図 2 に見られるようにクラスタ 13 に属する多数の理科教科書誤りのうち、多くが追加訓練で解消されたのは、頻出する説明文の構文を学習したからだと考えられる。またクラスタ 13 に属する特許文書誤りの数が少なかったのは、この文型の出現回数の少なさが原因であることがわかる。

クラスタ 24 クラスタ 24 に属する典型的な新聞記事誤り、理科教科書誤り及び特許文書誤りを以下に示す。

N-24 :
ところが/阪神大震災では/一番/重要な/情報収集が/遅れ、/自衛隊の/初動問題に/×代_表さ_れる_ように/首相官邸だけでなく/政府全体の/対応が/後手に/○回_つた。
T-24 :
図 1 のように、/光源から/出た/光は/四方八方に/広がり、/私たちの/目に/直接/×届_くか、/何かに/当たって/はね返って/私たちの/目に/○届_く。
P-24 :
次いで、/圧力を/常圧から/13.3 kPa に/○下_げて、/加熱槽温度を/190℃まで/1時間/○上_げせ_せながら、/発生する/フェノールを/反応容器外へ/×抜_き出_した。

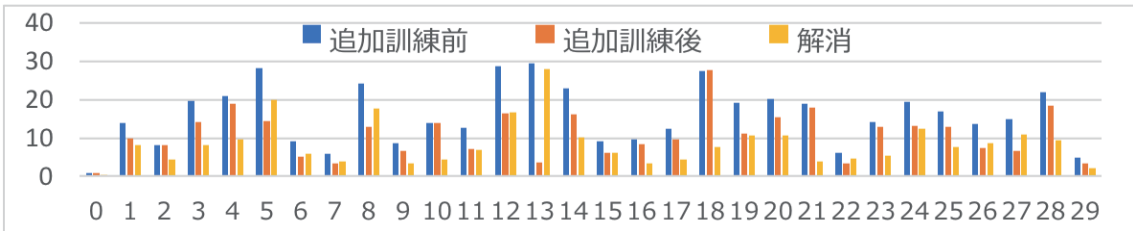


図 2: 追加訓練による、各クラスタに含まれる理科教科書誤りの割合の変化

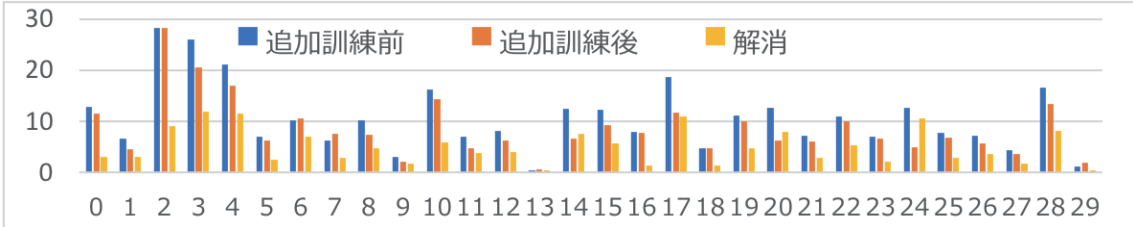


図 3: 追加訓練による、各クラスタに含まれる特許文書誤りの割合の変化

表 4: クラスタ 24 に属する係り受け誤りにおける誤った係り先の位置の分布

データ	ベースモデル		追加訓練済み	
	前	後	前	後
新聞記事	280	398	—	—
理科教科書	83	2	5	37
特許文書	13	59	12	7

クラスタ 24 に属する誤りは係り文節と正しい受け文節、誤った受け文節がすべて用言文節（節の末尾）であるパターンが多かった。3つの節 A, B, C がこの順番で並んでいる時、意味的な構造は ((A B) C) と (A (B C)) の 2 パターン存在する*2。この 2つの構造のうち、どちらが正しいか判定するためには、本来は節間の関係性を認識する必要があり、難しい判断を必要とする場合も多いと考えられる。しかし、図 2 と図 3 から、追加訓練によって、いずれのドメインでも比較的多くの誤りが解消されていることがわかる。その理由は、ドメインごとに出現しやすい構造に偏りがあるからだと考えられる。実際にそれぞれのドメインでクラスタ 24 に属する誤りについて、正しい受け文節と誤った受け文節のどちらが前方にあるかを調べた。結果を表 4 に示す。理科教科書データの解析では、追加訓練前のモデルは正しい受け文節よりも前方に存在する文節を選択する傾向があったが、追加訓練によりその傾向が改善されたことがわかる。また特許文書は理科教科書と対照的な結果である。ここから、理科教科書は T-24 のような (A (B C)) という形の階層的な並列関係を取る場合が多く、節 A はより速くの文節 C に係ることが多いと考えられる。一方、特許文書は P-24 のような手続きを表す文を多数含み、3 つ以上の節が単純に並列された構文が多いと考えられる。

4. おわりに

本研究では係り受け解析器の内部状態を用いて係り受け誤りを埋め込み、クラスタリングすることで、追加訓練の効果を、

*2 T-24 では「図 1 のように、…広がり、」、「私たちの目に直接届くか」、「何かに…目に届く。」という 3 つの節に分けられるが、最初の節は後ろの 2 つの両方に意味的に係るので、(A (B C)) のパターン。P-24 は節が順番に接続しているので、((A B) C) のパターン。

複数のドメインに渡って俯瞰的に分析した。結果として追加訓練によって解消される誤りのタイプをいくつか特定でき、さらに、解析が難しい文の特徴のドメインごとの差を比較することができた。今後の課題としては、解消されない誤りのタイプとその原因の特定、重心から誤りへの距離と誤りの性質の関係の調査、ドメイン適応技術への応用などがあげられる。

謝辞

本研究では、JSPS 科研費 16H01819 で開発された教科書アノテーションデータの提供を受けた。ここに記して謝意を表す。また本研究は名古屋大学と富士通研究所の共同研究の成果である。

参考文献

[Daumé III 06] Daumé III, H. and Marcu, D.: Domain adaptation for statistical classifiers, *Journal of Artificial Intelligence Research*, Vol. 26, pp. 101–126 (2006)

[Daumé III 07] Daumé III, H.: Frustratingly Easy Domain Adaptation, in *Proc. of ACL*, pp. 256–263, Association for Computational Linguistics (2007)

[Tokui 15] Tokui, S., Oono, K., Hido, S., and Clayton, J.: Chainer: a next-generation open source framework for deep learning, in *Proceedings of workshop on machine learning systems (LearningSys) in NIPS 29*, Vol. 5, pp. 1–6 (2015)

[黒橋 97] 黒橋禎夫, 長尾眞: 京都大学テキストコーパス・プロジェクト, 人工知能学会第 11 回全国大会論文集, pp. 58–61 (1997)

[松野 18] 松野智紀, 能地宏, 松本裕治: ニューラルグラフ型係り受け解析器のための文節ベクトル表現, 言語処理学会第 24 回年次大会発表論文集 (2018)

[内元 99] 内元清貴, 関根聡, 井佐原均 他: 最大エントロピー法に基づくモデルを用いた日本語係り受け解析, *情報処理学会論文誌*, Vol. 40, No. 9, pp. 3397–3407 (1999)

IoT デバイスのための深層学習を用いた自動販売機の購入商品の識別 Vending Machine Drink Recognition with Deep Learning for IoT Device.

新屋勝啓^{*1}
Katsuhiro Araya

長隆之^{*2}
Takayuki Osa

山岡成光^{*2}
Shigemitsu Yamaoka

西和彦^{*2}
Kazuhiko Nishi

中尾政之^{*2}
Masayuki Nakao

^{*1} 株式会社 Arkth
Arkth, Inc.

^{*2} 東京大学大学院工学系研究科
University of Tokyo, School of Engineering

In recent years, Internet of Things (IoT) with deep learning is used for various purposes. The balance of edge computing, data traffic, cloud computing and data volume is important for IoT. In this study, we present a framework that automatically recognizes bottles purchased from a vending machine in camera images with a convolutional neural network (CNN).

1. はじめに

近年 IoT と深層学習の連携が様々な用途や分野で活用され始めている。しかし現状ではデバイス、ネットワーク、深層学習による認識率など様々な面で課題があり、多くの研究者によって多岐にわたり研究されている。例えば一般的に IoT デバイスは大量に利用されるため、なるべく安価に構築する必要がある。しかし、安価なデバイスではスペックが低くなるため、計算量の大きい分析については制限がかかってしまう。そのため、深層学習による計算量を削減するための前処理やアルゴリズムの調整、ハードウェアの構成の改善などを行う必要がある。さらに、IoT デバイスが増えるとネットワークへの負荷も高くなるため、なるべく少ない通信コストで動作させる必要がある。サーバーに蓄積するデータも IoT デバイスの量に比例して増加するためデバイス側でデータ圧縮や分析を行いたい。このように IoT と深層学習の連携では、エッジ AI コンピューティング、ネットワークの転送量、クラウドでのデータ蓄積と分析の強調を、認識率やデバイスの価格などとの兼ね合いを鑑みてどのように設定するかが重要である。

IoT と深層学習を連携したシステムの中でも、本研究では近年増加している自動販売機での購入者や購入商品の分析に着目してシステムを構築する。

現状の自動販売機の購入動向の分析システムでは店舗や機器に改修を加えたり入れ替えたりする必要があるため、導入コストが高い課題がある。本研究で構築するシステムは既存の機器の内部に改修を加えることなく、安価で外付けできることを特徴として開発を進めている。

本稿ではそのシステムの中でも特にエッジ AI コンピューティングによる飲料の分類に着目し、安価で低スペックなデバイスと深層学習を用いて飲料自動販売機に設置したカメラから撮影された動画像中の購入商品を認識することを目的としている。

2. 本システムの全容

本システムでは購入者の年齢・性別と購入商品を認識すると共に、気象センサによる気象情報を取得してクラウドに保存する。ネットワークには LoRaWAN を利用することを前提としているため、通信するデータは飲料 ID や年齢、性別フラグ、気温など最低限となるようにエッジで分類を行う。

また、本システムでは購入者にリアルタイムに広告表示などのインタラクションを行うため、ディスプレイや表示用のデバイス

を設置し、クラウドから最新のデータを取得できるようにする。
図 1 に本システムのシステム構成を示す。

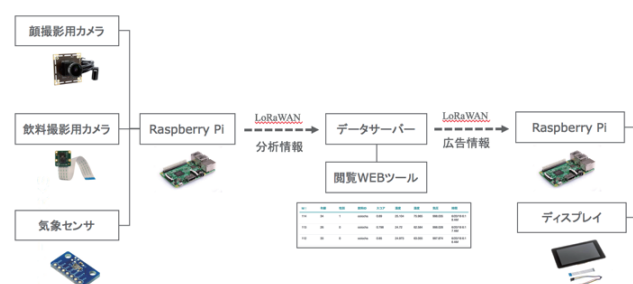


図 1 システム全容の構成図

3. 飲料の認識

3.1 デバイスについて

図 1 左側のセンサー情報の取得や、飲料の分類、サーバーへのデータ転送には Raspberry Pi 3 Model B+ を利用した。気象センサに BME280 を利用し、温度・湿度・気温の 3 つの環境情報を取得できる。カメラは購入者の撮影と飲料の撮影用にそれぞれ設置し、購入者撮影用には広角に撮影できるカメラを利用した。

3.2 飲料の認識

本研究では、深層学習を用いた物体検出技術を用いる。ニューラルネットワークは Single Shot Multibox Detector (SSD) を採用し [Liu 16], 深層学習のフレームワークとして TensorFlow を利用した。SSD はニューラルネットワークの構造に VGG16 [Simonyan 14] を用いた手法であり、入力画像が小さくても高い精度の物体検出が可能である。

3.3 認識の前処理

自動販売機の取り出し口に向けて設置したカメラから SSD によって飲料を認識する際に、背景を飲料に誤認識するフレームが多く発生した (図 2)。

そこで本手法では、認識範囲のトリミングを行うと共に、背景差分によって取り出し動作の検出を行った。

具体的には 320x240 の画像を用いた場合、取り出し口付近の背景差分を用いて取得した動きのあったピクセル数が一定のしきい値を超えた場合に飲料の認識を行うようにした。これにより認識を軽減すると共に、認識を行うフレーム数を削減できる。

背景差分を取得する範囲も図 2 に矩形で示す。

連絡先: 新屋勝啓, 株式会社 Arkth, info@arkth.co.jp

3.4 学習データのデータオーギュメンテーション

自動販売機に設置したカメラから取得される画像は、設置場所や時間、気象条件によって明るさが様々に変化する(図 3)。そこで、学習データに対してデータ拡張を行った。データの拡張はノイズ追加、コントラスト調整、ガンマ変換による明るさ調整、拡大縮小、反転を行った。データオーギュメンテーションを行うことで、限られた数の教師データからでも明るさなどの変化にロバストな認識器を構築することができる。



図 2 背景の誤認識と背景差分取得範囲

4. 実験

4.1 自動販売機とカメラの設置

本実験に利用する自動販売機は、大学構内の屋根のある屋外に設置した。飲料認識用のカメラは図 3 のように、自動販売機前面に取り出し口付近を撮影するように設置した。カメラは Raspberry Pi 用のカメラモジュールを用いた。



図 3 飲料認識用カメラの設置

動画撮影時のフレームレートは設定やプログラム上では 20FPS と設定しているが、デバイスのスペックによる制限によって実際は 7~8FPS となった。

4.2 学習データの取得

19種類の飲料を対象に自動販売機に取り出し口に向けたカメラから、手に持った飲料の画像を収集した。

背景による影響を考慮し、過学習を避けるために屋内の白い背景で取得した学習データも追加し学習を行った。



図 4 学習データ例

学習データは各飲料で 100 枚ずつ程度取得した。その後データオーギュメンテーションを用いて各画像を 13 枚変換し合計 26600 枚を学習データとして利用した。

4.3 実験結果

飲料が写っている画像を 19 種類の飲料ごとに 100 枚ずつ検証に利用した。その結果を下の表にまとめる。

合計画像数	正認識数	誤認識数	未検出数
1900	1651(87%)	137(7%)	112(6%)

飲料ごとの認識率では、緑茶や烏龍茶などのお茶系の認識率が低く、一方で缶コーヒーなどの缶飲料の認識率が高い傾向があった。これはお茶系の飲料は外観の色味が似ている物が多いのに対し、缶コーヒーなどはパッケージの色彩などの差異が大きいためと考えられる。

また、1枚の画像から飲料を認識するのに、Raspberry Pi 3B+では 10 秒程度を要した。

5. 結果の考察

上記の通り本実験では 85%を超える認識率を達成した。しかし、本システムの狙いは様々な環境や種類の自動販売機に設置して、エッジ AI コンピューティングにより分析を行うことである。

本実験では1箇所に設置した自動販売機で同じ季節に撮影したデータが多いため、様々な環境に設置した自動販売機で認識を行った場合よりも高い認識率が出ていると予想される。

また、自動販売機の飲料は無数に存在すると共に、商品ラインナップの入れ替えや商品ラベルの変更も数多く発生するため環境の変化だけでなく商品の多様性にも対応する必要がある。本実験は深層学習やデバイスの実装に加え、学習データの取得に多くの手間がかかった点も今後の課題である。

市販されている飲料はメーカーにラベルやボトルのデータがあるため、今後は Zero shot Learning や Few Shot Learning[Sachin 17]の模索も重要となる。

6. まとめ

本稿では IoT デバイスによる飲料認識の手法を提案し、実験を行った。その結果、限定的な環境であれば安価なエッジデバイスで高い認識率を実現できた。今後、更に多くの環境や飲料で検証を進めると共に学習データを集め、ハイパーパラメータのチューニングや本手法で用いた以外のデータオーギュメンテーション、Few Shot Learning などの学習手法の研究を進める。

また、深層学習による画像認識は顔や動物などの認識についての先行研究が多く、飲料に特化した文献はあまり見られなかった。しかし、今後の IoT の拡大につれてあらゆるものやことを認識する需要が増えてくると予想できるため、様々な物体検出に応用できる手法を目指していく。

謝辞

本研究に利用した自動販売機は株式会社伊藤園の協力により設置いただきました。心より感謝申し上げます。

参考文献

[Simonyan 14] Karen Simonyan and Andrew Zisserman: Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[Liu 16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C.: Ssd: Single shot multibox detector, in European conference on computer vision, pp. 21-37Springer (2016)

多人数会話システムのための 話者の状態変化に頑健な注視状態検出

Robust Eye Contact Detection for Multi-Party Conversational Systems

野川賢二郎^{*1}

Kenjiro Nogawa

藤江真也^{*2}

Shinya Fujie

小林哲則^{*1}

Tetsunori Kobayashi

^{*1}早稲田大学

Waseda University

^{*2}千葉工業大学

Chiba Institute of Technology

Eye contact detection method for multi-party conversational systems is proposed. Detecting eye contact is an important function for multi-party conversational systems because eye contact is an essential cue for deciding whether system should respond to a user utterance or not. Accuracy of a conventional eye contact detector given a single frame decreases because of various noises (e.g. blink, facial expression change, etc.) occurred in real-time environment. The eye contact detection method that utilizes multiple frames is proposed for solving this problem. The system extract the feature for detecting eye contact by CNN from the image contains both eyes of a target user. Then, the feature is given to LSTM with attention mechanism as an input. The result of the experiment conducted using the gaze data in multi-party conversation shows effectiveness of the proposed method.

1. はじめに

多人数会話システムにおいてユーザのシステムに対する注視状態を検出する手法について検討する。

近年、スマートスピーカや会話ロボットなどの対話システムが普及している。これらのシステムは主に一対一の対話を扱うものであるのに対し、複数人が会話に参加するような環境で動作する多人数会話システムの研究が行われている。多人数会話システムにおいて、視線はシステムがユーザの振る舞いを理解する上で重要な情報を与える。多人数会話システムでは、ユーザの発話がシステムに向けたものか、そうでないかを識別する必要がある、会話システムの分野で応答義務推定（受話者推定）という問題として研究が行われている。応答義務推定において、ユーザがシステムのことを注視しているか（注視状態かどうか）という情報が精度の向上に寄与することが報告されている。Turnhout らは応答義務推定において、ユーザがシステムのことを注視しているか（注視状態かどうか）を人手でアノテーションして特徴量として用いることで精度が向上すると報告した [Van Turnhout 05]。実際の会話システムを動かすためには注視状態を自動検出が必要となる。視線の代用として比較的推定が簡単なユーザの顔向きが用いられてきた [Shriberg 12, Tsai 15, 馬場 13, 杉山 16]。しかし、多人数会話において、ユーザの顔向きを視線の近似として用いることは不十分であると報告されている [Vrzakova 16]。以上のことから、多人数会話システムにおいて注視状態を自動で検出する技術は重要である。

画像処理の分野において、アイコンタクト検出に関する研究が行われてきた。アイコンタクト検出では、ノイズのない単一の画像から虹彩が対象物体に向けられているかどうかを識別する。Fujie らは会話ロボットでの利用を目的に対象者の目画像を入力とするアイコンタクト検出器を PCA や LDA などの次元圧縮法と Gaussian Mixture Model の組み合わせにより構築した [Fujie 06]。また、Smith らはアイコンタクト検出タス

クのためのデータセット（Columbia Gaze Data Set）を公開し PCA と MDA による次元圧縮と線形 SVM による手法を提案した [Smith 13]。また、Parekh らは Convolutional Neural Network（CNN）を用いた手法を提案した [Parekh 17]。アイコンタクト検出では、ノイズのない単一の画像から虹彩が対象物体に向けられているかどうかを識別する。一方、会話システムの利用時には、前処理の段階で生じるノイズ（顔検出の失敗、顔特徴点推定の誤差）や話者の状態変化によるノイズ（まばたき、表情の変化、動きによるブレ）が生じるため、単一の画像を入力とする手法では注視状態を検出するのは難しい。

そこで本研究では、多人数会話システムにおいて注視状態を高精度に検出することを目的とし、過去の入力画像を利用することでノイズに対して頑健に注視状態を検出する手法を検討する。

本稿では、2. で提案手法である画像系列を入力とする注視状態検出手法を説明する。3. では会話中の視線データの収集方法とデータの詳細を述べる。4. では収集したデータを用いた注視状態検出の実験結果を報告する。最後に、5. で本研究のまとめを述べる。

2. 画像系列を入力とする注視状態検出器

時々刻々と入力される画像を逐次処理し、その都度過去の履歴を考慮してユーザの注視状態を検出するシステムを構築した。まず、入力された全方位画像から前処理を経て人の目元の画像を抽出する。次に、得られた目元の画像を Convolutional Neural Network (CNN) に入力し、その中間層にある画像特徴量を Long-Short Term Memory (LSTM) に入力する。LSTM の現在の中間状態から、過去の状態に対する注意重みを計算し、中間状態と注意重みの重み付き和で得られたベクトルから注視状態を検出する。システムの全体像を図 1 に示す。以下、前処理 (図 1.a) と注視状態検出 (図 1.c) について説明する。

2.1 前処理

まず全方位カメラによって撮影された全方位画像を極座標変換によりパノラマ展開する。得られたパノラマ画像から人物領域を切り出す。次に、人物領域から顔領域の検出を行い、領域を元に 68 点の顔特徴点の検出を行う。顔領域と顔特徴点の検

連絡先: 野川賢二郎, 早稲田大学 理工学術院 知覚情報システム研究室, 〒162-0042 東京都新宿区早稲田町 27, nogawa@pcl.cs.waseda.ac.jp
(所属先は執筆・投稿時 2019 年 2 月のもの)

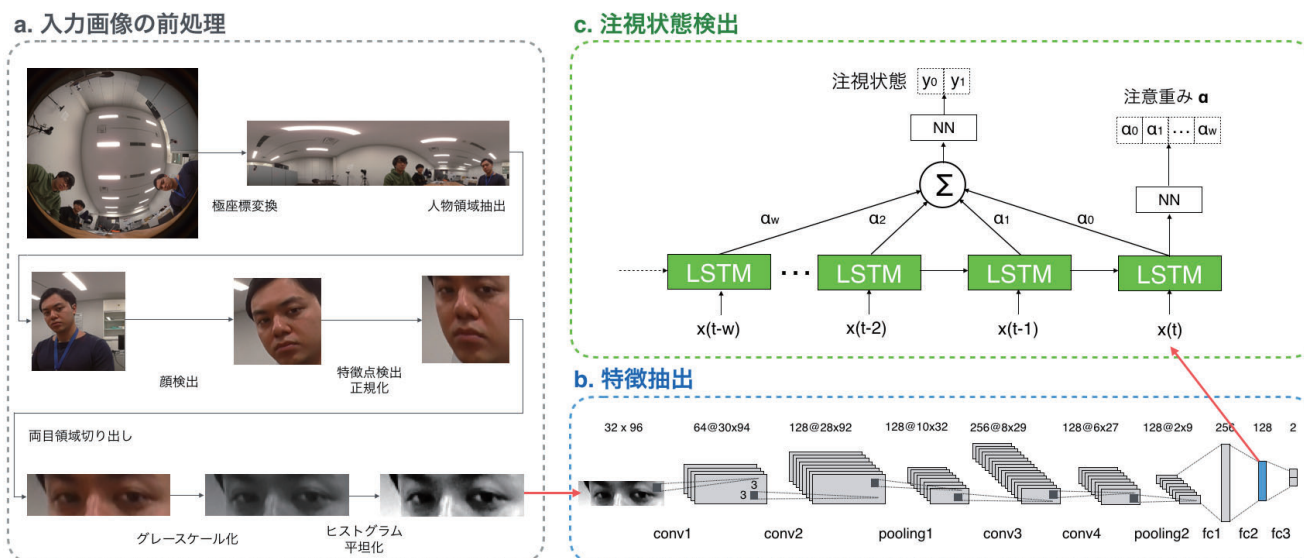


図 1: 画像系列を入力とする注視状態検出器のシステム概要図。入力画像に対して前処理を行い (a), CNN の入力とする。CNN の全結合層 (fc2) で出力された値を LSTM に入力する (b)。LSTM の中間層から履歴に対する注意重みを計算した後、履歴と注意重みによる重み付き和を全結合層の入力とし、注視状態検出結果を出力する (c)。

出には機械学習ライブラリである dlib^{*1} を利用した。推定された顔特徴点の両目尻と鼻の先端の 3 点を元に顔領域画像に対してアフィン変換を適用する。アフィン変換された顔画像から両目部分の画像領域を切り出す。切り出した両目領域画像に対してグレースケール化とヒストグラム平坦化を行う。以上の方法により得られた両目領域画像を CNN の入力とする。

2.2 識別器

識別器として注意機構を持った LSTM を用いる。会話中に生じるノイズに対して頑健に注視状態を検出するためには、ノイズらしきものが入力された場合は、入力による状態の更新をせずに過去の状態を引き継ぎ、ノイズでないものが入力された場合は、状態を更新するといった処理が必要だと考えられる。LSTM は忘却ゲートと入力ゲートを持つため、入力と状態に応じた状態の更新を学習することに適している。そのため、LSTM を用いることでノイズに対して頑健に注視状態を検出することができると考えられる。また、会話中に生じるノイズは、種類ごとに継続する長さが異なる。例として、まばたきは平均で 100-150 ミリ秒なのに対し、表情の変化は数秒継続する。そこで、LSTM に注意機構を持たせ、現在の状態に応じた過去の状態利用方法を明示的に学習させることにより、多様なノイズに頑健に注視状態を検出することを期待する。

3. 視線データ収集

本研究では、複数人のユーザがダイニングテーブルに設置されたスピーカ型会話システムを含めて会話を行い、意思決定を行う状況を想定してデータ収集を行った。

今回は、会話システムとして Amazon 社製の Amazon Echo^{*2} を用いた。Amazon Echo の上部に全方位カメラを設置して会話中のユーザの様子を撮影した。会話参加者に対して、他の会話参加者と Amazon Echo と会話しながら指定されたタスクを遂行するようにあらかじめ指示した。タスクとし

て、会話参加者同士で観に行く映画を決定するタスクと外食に行く先の飲食店を決定するタスクの 2 種類を用意した。1 タスクにつき、少なくとも 2 回はスピーカに対する質問を行うように指示を行った。会話が滞ることを防ぐために、各タスクごとに Amazon Echo への質問例を会話参加者に対してあらかじめ提示した。また、ファシリテータが会話に参加した。ファシリテータは会話が滞った場合に会話参加者に対して質問を行うなど、円滑にタスクが進むように行動を行った。それ以外の場合は他の会話参加者と同様にタスクに取り組んだ。ファシリテータの振る舞いは他の対話参加者と異なるため注視状態の検出対象から除外した。研究室の学生 13 人を 5 グループに分割し、それぞれのグループにつき 2 つのタスクに取り組んでもらい合計 42 分のデータを収集した。1 人のアノテータが、撮影した動画データに対し、各会話参加者が会話システムを注視しているかどうかのアノテーションを行った。アノテーションの際に、以下の点を注意してラベルを付与した。会話参加者がシステムを注視している状態でまばたきをして、まばたきのあとも注視している場合はまばたきの区間を含めて注視状態とした。これは、まばたきはあくまでも一時的に目を閉じているのであり、会話参加者の注意はシステムに向けられているためである。一方、会話参加者がシステムを注視している状態でまばたきをして、まばたきのあとはシステムを注視していない場合は瞬きの前までを注視状態とした。

会話視線データセットに加え、単一画像を入力とする注視検出器を事前学習するための視線データセットを収集した。視線データセットは、多様な顔形状、顔向き、注視点を含むデータセットである。会話視線データセットと同一の環境に、複数の注視点を設置し、収集対象者が注視点を注視しながら顔を動かしている様子を、Amazon Echo 上部に設置された全方位カメラで動画撮影をし、収集を行った。カメラのレンズを注視している動画とスピーカに設置された注視点を注視している動画を正例、その他の注視点を注視している動画を負例とした。研究室の学生 17 人を対象にデータを収集し、正例画像が 4955 枚、負例画像が 9479 枚となった。

両データセット共に Kodak 社製の全方位カメラである PIX-

*1 <http://dlib.net/>

*2 <https://www.amazon.co.jp/dp/B071ZF5KCM/>

PRO SP360 4K *³ を用い、フレームレート 30fps、解像度は 2880 × 2880 で撮影を行った。

4. 会話視線データセットにおける注視状態検出実験

提案手法の有効性を調査するため、以下の 4 種類の手法で比較を行った。評価指標として F 値と AUC (Area Under the Curve) を用いた。

- CNN: 視線データセットを用いて学習した CNN。過去のフレームの情報は利用しない。
- CNN Average: 視線データセットを用いて学習した CNN を利用し、過去の x フレーム分の結果の平均を出力とする。
- CNN-LSTM: 視線データセットを用いて学習した CNN の全結合層のベクトルを入力とする LSTM を会話視線データセットを用いて学習する。
- CNN-Attention LSTM: 視線データセットを用いて学習した CNN の全結合層のベクトルを入力とする注意機構付き LSTM を会話視線データセットを用いて学習する。現在のフレームを入力したときの LSTM の中間層のベクトルから過去 w フレームに対する注意重みを決定する。

4.1 実験条件

4 つの全ての手法において、視線データセットを用いて CNN の学習を行った。全 17 人分の視線データセットの 14 人分を学習、1 人分を検証、2 人分をテストに用いた。CNN-LSTM と CNN-Attention LSTM において、全 5 グループ分の会話視線データセットの 3 グループ分を学習、1 グループ分を検証、1 グループ分をテストに用いた。会話視線データセットのテストデータ中に含まれる人は、視線データセットの学習データと検証データには含まれていない。CNN-LSTM と CNN-Attention LSTM の学習時は CNN の重みは固定した。また、CNN Average の x は会話視線データセットの学習データと検証データでもっとも AUC が高い 3 を採用した。CNN-Attention LSTM の w は会話視線データセットの検証データでもっとも AUC が高い 32 を採用した。

4.2 実験結果

各手法のテストデータでのスコアを表 1 に示す。詳細な分析を行うために、会話中に生じるノイズである、瞬き、表情の変化、顔検出の失敗が生じている時間付近に対する各手法の推定結果と、平常時と、各ノイズ発生時の注意重みを可視化したものをそれぞれ図 2、図 3、図 4 に示す。

画像系列を入力とする全ての手法 (CNN Average, CNN-LSTM, CNN-Attention LSTM) が、単一画像を入力とする手法 (CNN) の精度を上回っていることから、画像系列を入力とすることの有効性が示された。

表 1 より、CNN-LSTM が CNN Average のスコアを上回っている。図 2 より、CNN-LSTM が CNN Average に比べてノイズである瞬きの画像が入力された場合でもうまく推定できていることがわかる。また図 3 より、表情が変化している入力と比較的長期に渡って入力された場合でも CNN-LSTM がうまく推定できていることがわかる。以上より、LSTM によ

表 1: 会話視線データセットにおける各手法の正しい例の F1 スコアと AUC

手法	F1 値	AUC
CNN	0.831	0.923
CNN Average	0.847	0.938
CNN-LSTM	0.880	0.958
CNN-Attention LSTM	0.880	0.960

て状態を保持し、過去の入力の情報を利用することによりノイズに頑健な注視推定が行えることが確認できた。

表 1 より、AUC において CNN-Attention LSTM の精度が CNN-LSTM の精度を上回っている。図 3 より、ユーザの表情が変化している区間において、CNN-Attention LSTM が CNN-LSTM に比べてうまく推定できていることがわかる。また、図 4 より、顔検出が失敗している区間において、CNN-Attention LSTM が他の手法に比べてうまく推定できていることがわかる。どの時点での履歴を利用するかを明示的に扱う注意機構の有効性が確認できた。

CNN-Attention LSTM と CNN-LSTM の精度差が小さい原因として、状態の推定に必要な履歴情報が LSTM のゲート機構によってうまく保持できていることが考えられる。図 2, 3 の注意重みの可視化結果を見ると平常時とノイズ発生時の両者の現在の状態に対する重みが高い。対して、図 4 の注意重みの可視化結果を見るとノイズ発生時の現在の状態に対する重みが低くなり、過去の状態に対する重みが上昇している。以上のことから多くの場合において LSTM のゲート機構によって状態の推定に必要な履歴情報が保持できるが、顔検出の失敗などの特殊な場合において注意機構が有効に作用すると考えられる。

5. まとめ

本研究では、多人数会話システムにおいて注視状態を高精度に検出することを目的とし、過去の入力画像の情報を利用することでノイズに対して頑健に注視状態を検出する手法を検討した。提案手法の有効性を検証するために、会話中のユーザの視線データを収集し、収集したデータを用いて実験を行った。評価実験の結果、画像系列を入力とする手法が、単一画像を入力とする場合に比べて精度が向上した。LSTM を利用した手法と Attention LSTM を利用した手法は、過去の画像の注視状態検出結果の平均を出力とするベースライン手法よりも良い結果となった。また、LSTM に注意機構を持たせ、入力に応じた履歴の利用方法を明示的に学習させることにより、精度が向上することが確認できた。詳細な分析の結果、会話中に生じるノイズである、瞬き、表情の変化、顔検出の失敗が生じた場合において、提案手法がベースライン手法に比べて頑健に注視状態を検出できることがわかった。

*3 <https://kodakpixpro.com/AsiaOceania/jp/cameras/vrcamera/sp3604k/>

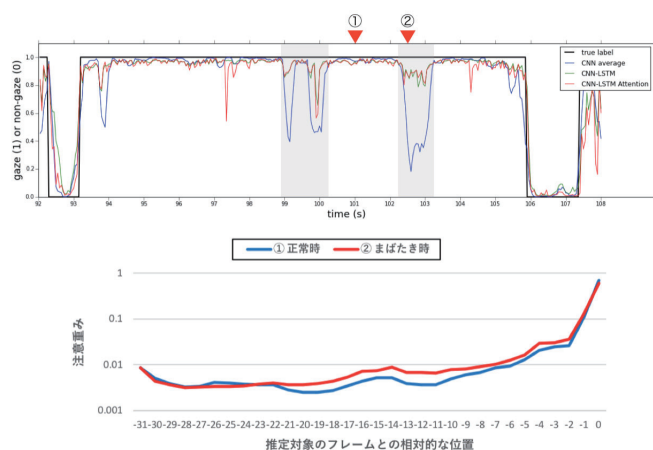


図 2: ユーザが瞬きをしている時間付近における各手法の注視状態検出結果。灰色部分がユーザが瞬きをしている時間。下部は①, ② 時点での過去 32 状態に対する注意重みを可視化したもの。

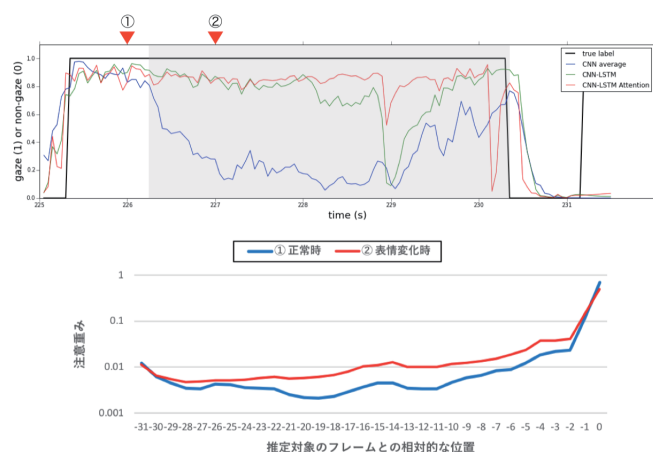


図 3: ユーザの表情が変化している時間付近における各手法の注視状態検出結果。灰色部分がユーザの表情が変化している時間。下部は①, ② 時点での過去 32 状態に対する注意重みを可視化したもの。

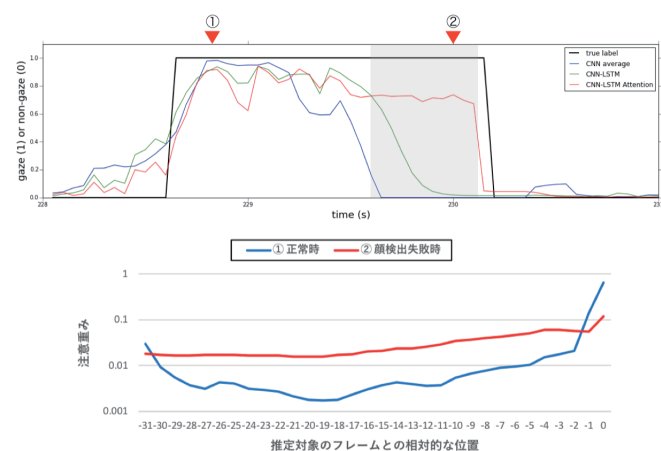


図 4: 顔検出に失敗している時間付近における各手法の注視状態検出結果。灰色部分が顔検出に失敗している箇所。下部は①, ② 時点での過去 32 状態に対する注意重みを可視化したもの。

参考文献

- [Fujie 06] Fujie, S., Yamahata, T., and Kobayashi, T.: Conversation Robot with the Function of Gaze Recognition, in *2006 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy, December 4-6, 2006*, pp. 364–369 (2006)
- [Parekh 17] Parekh, V., Subramanian, R., and Jawahar, C. V.: Eye Contact Detection via Deep Neural Networks, in *HCI International 2017 - Posters' Extended Abstracts - 19th International Conference, HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I*, pp. 366–374 (2017)
- [Shriberg 12] Shriberg, E., Stolcke, A., Hakkani-Tür, D. Z., and Heck, L. P.: Learning When to Listen: Detecting System-Addressed Speech in Human-Human-Computer Dialog, in *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association, Portland, Oregon, USA, September 9-13, 2012*, pp. 334–337 (2012)
- [Smith 13] Smith, B. A., Yin, Q., Feiner, S. K., and Nayar, S. K.: Gaze locking: passive eye contact detection for human-object interaction, in *The 26th Annual ACM Symposium on User Interface Software and Technology, UIST'13, St. Andrews, United Kingdom, October 8-11, 2013*, pp. 271–280 (2013)
- [Tsai 15] Tsai, T. J., Stolcke, A., and Slaney, M.: A Study of Multimodal Addressee Detection in Human-Human-Computer Interaction, *IEEE Trans. Multimedia*, Vol. 17, No. 9, pp. 1550–1561 (2015)
- [Van Turnhout 05] Van Turnhout, K., Terken, J., Bakx, I., and Eggen, B.: Identifying the intended addressee in mixed human-human and human-computer interaction from non-verbal features, in *Proceedings of the 7th international conference on Multimodal interfaces*, pp. 175–182ACM (2005)
- [Vrzakova 16] Vrzakova, H., Bednarik, R., Nakano, Y. I., and Nihei, F.: Speakers' head and gaze dynamics weakly correlate in group conversation, in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pp. 77–84ACM (2016)
- [杉山 16] 杉山貴昭, 船越孝太郎, 中野幹生, 駒谷和範: 多人数対話におけるユーザの状態に着目したロボットの応答義務の推定, *人工知能学会論文誌*, Vol. 31, No. 3, pp. C-FB2.1–9 (2016)
- [馬場 13] 馬場直哉, 黄宏軒, 中野有紀子: 人対会話エージェントとの多人数会話における頭部方向と音声情報を用いた受話者推定機構, *人工知能学会論文誌*, Vol. 28, No. 2, pp. 149–159 (2013)

画像列を用いた小惑星形状・探査機相対運動の逐次推定

Successive estimation of the asteroid shape and probe motion using sequential images

鈴木 透馬^{*1}

Toma Suzuki

矢入 健久^{*1}

Takehisa Yairi

武石 直也^{*1}

Naoya Takeishi

津田 雄一^{*2}

Yuichi Tsuda

尾川 順子^{*2}

Naoko Ogawa

^{*1}東京大学

The University of Tokyo

^{*2}宇宙航空研究開発機構

Japan Aerospace Exploration Agency(JAXA)

In order for a probe to safely land on an asteroid, it is necessary to select the landing point and navigate the probe based on information of the asteroid shape and the probe motion. The procedures for extracting such information from images taken by the probe is called global mapping. In the Hayabusa project of JAXA, the global mapping was conducted with much manual manipulation. As the manual global mapping is too costly, automating the procedures is strongly desired in the future projects. In this paper, we propose a new approach for the automation of global mapping by using the technique of image recognition and robotics. Throughout the experiment using the dataset provided by JAXA, we checked the validity of the method which we proposed.

1. はじめに

太陽系には小惑星と呼ばれる小さな天体が無数に存在している。小惑星は太陽系初期の情報を持っている「始原天体」とされ、太陽系の誕生や進化、そして我々生命がどのようにして生まれてきたのかを知る大事な手がかりの一つとされている。

小惑星探査ミッションでは探査機を精密に誘導したり、安全に小惑星に着陸させるために小惑星の詳細な形状と運動を把握するグローバルマッピングと呼ばれる作業が不可欠である。しかし、それらの情報はミッション開始段階では未知なことがほとんどであり、宇宙航空研究開発機構による「はやぶさ」、「はやぶさ2」プロジェクトでは探査機が小惑星付近に到着してから人間が手動で小惑星画像を解析し Ground Control Point と呼ばれる地点を多数設定することで小惑星の形状や探査機の姿勢の推定を行った [Shirakawa 06]。この作業は多大な人的資源及び時間を消費するものであるため、この作業の自動化は今後の小惑星探査ミッションを円滑に行う一助となり得ると考える。

本研究ではグローバルマッピングの自動化に関する新たなアプローチを提案することを目的としている。提案手法では探査機の位置姿勢及び小惑星の相対運動パラメータについては Extend Kalman Filter(EKF) を用いて確率的に推定し、小惑星形状を Point Cloud を用いて確定的に表現し、小惑星形状による制約を用いて Adaptive Moment Estimation(Adam) による逐次最適化を行うことで過去に提案された手法と比較して精度を多少落としながらも高速に大規模な点群を推定する。提案手法の有効性を示すために小惑星探査機「はやぶさ2」が撮影した小惑星「リュウグウ」の複数の画像列を用いて実験を行った。実験の結果、本手法のアプローチが有効であることを確認した。

2. 問題設定

2.1 グローバルマッピングの流れ

グローバルマッピング全体の大まかな流れとしては以下のようになる。

1. 小惑星全体を俯瞰出来るような位置に探査機を誘導し、画像を多数枚撮影することで小惑星全体が含まれるような画像列を得る。
2. 画像に対して何らかの特徴量を考えることで、画像間の点と点の対応付を考える。
3. カメラの観測モデルや小惑星及び探査機のダイナミクスについて検討することで、小惑星の形状及び探査機の相対運動を推定する。

本論文では1の工程は満足されているとし、2, 3の工程について検討する。2の工程については高速かつ高精度な Scale-Invariant Feature Transform (SIFT) 特徴量 [Lowe 04] を考えることで行う。グローバルマッピング作業全体を効率化するためには3の工程の効率化が特に重要である。小惑星形状及び探査機相対位置が未知である場合、両者はカメラを介した非線形な観測を通して相互に依存するため、全体を最適化することは大規模な非線形最適化問題となっている。

2.2 本研究における問題設定

上記を踏まえ、本研究では以下の入出力を行うアルゴリズムの作成を問題とする。

表 1: 提案アルゴリズムの入出力関係

Input:	一定間隔で撮影された小惑星の画像列
Output:	各時刻の探査機の相対位置姿勢, 小惑星形状, 小惑星のスピン軸, 小惑星の相対角速度

なお、小惑星形状についてはポイントクラウドを用いて表現する。

さらに、本研究では推定対象に関する制約条件として以下の3点を設けるものとする。

1. 探査機は小惑星赤道上空一定高度を小惑星に対して一定の姿勢を保ち周回するように制御されている。
2. 小惑星は球形に近い形状である。

連絡先: 鈴木透馬, 東京大学 知能工学研究室, toma-aerospace@g.ecc.u-tokyo.ac.jp

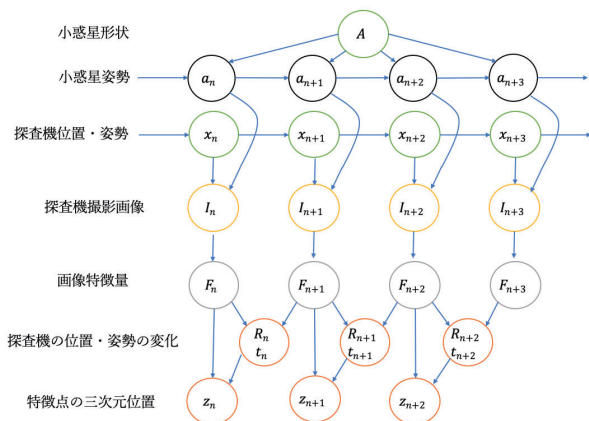


図 1: 本問題のモデル

3. 小惑星は非タンプリングである.

仮定 1 については探査機が小惑星全体を初期観測する際に全域を最も効率良く観測することが出来る条件であり, 妥当な仮定であると考えられる. 仮定 2 については実際に存在する小惑星の中でもそのような条件を満たすものは数多く存在し, ここではそのような制約を置いて大きな問題にはならないと考える. 仮定 3 については本手法は画像列のみを入力とする推定を考えているが, 仮定 1 の制約を満たして撮影された画像列のみからでは小惑星がタンプリングしているかを判断することは難しく, 本研究では小惑星がシングルスピンであるという仮定を置くこととする.

3. 提案手法

本研究では従来手法に比べ高速かつ一定の精度を保ち推定を行うアルゴリズムとして「Adaptive Moment Estimation(Adam)[Kingma 14] による小惑星形状の局所最適化と Extended Kalman Filter(EKF) による推定を用いた小惑星形状及び探査機の相対運動の逐次推定」を提案する.

3.1 提案手法の流れ

提案手法全体の流れは以下の通りである.

1. Adam による観測の最適化.
2. EKF による小惑星形状パラメタを除く状態変数の推定.
3. 三次元形状の再復元.
4. 剛体制約による形状の最適化.
5. 1 から 4 を時系列入力に対して繰り返し適用する.

以下では, 手順 1 から 4 についての概要を述べる. なお, 世界座標系は初期時刻におけるカメラ座標系と一致させた. すなわち, 初期時刻の観測の画像座標系の x 軸, y 軸を世界座標系の x 軸, y 軸とし, その時のカメラの光軸方向に z 軸を取るような系を考えるものとした. 今回の問題においては小惑星は世界座標系の一点に固定されており, 探査機が世界座標系において小惑星周囲を円運動するようなモデルを考えることになる.

3.2 各ステップの概要

3.2.1 Adam による観測の最適化

まず, 時系列に沿って画像が入力されるごとに SIFT 特徴量を用いた画像特徴点のマッチングを行う. 観測モデルにピンホールカメラを採用することでエピポーラ幾何を利用してカメラの位置関係及び各カメラ座標系における画像特徴点の三次元位置が復元される. しかし, この観測には SIFT 特徴量の誤マッチング等に起因する誤差が含まれていると考えられるので以下の方法によって観測値を補正する.

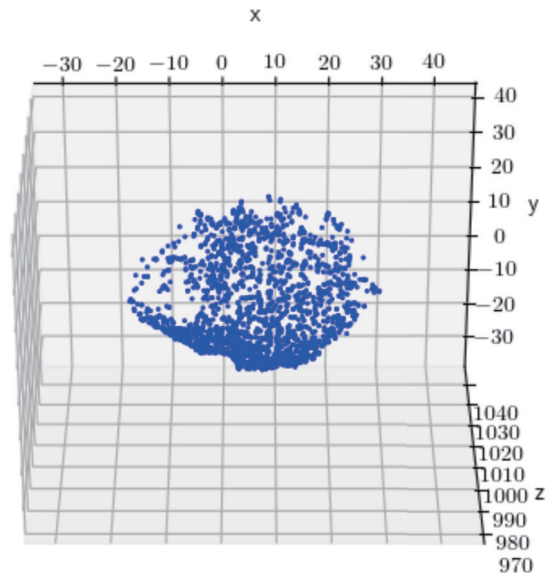


図 2: 連続する観測から復元された小惑星の部分形状

観測について i 番目の観測と $i+1$ 番目の観測による復元 (復元 i) と, $i+1$ 番目の観測と $i+2$ 番目の観測による復元 (復元 $i+1$) を考える. それぞれの三次元復元についてスケールは任意であるから, 復元 i と復元 $i+1$ スケール比を表す変数を s_i とする. また, 世界座標系におけるある点 z の i 番目のカメラ座標系における表示を z_i と書くことにし, 観測より得られる 2 つのカメラ座標系の変換を R_i, t_i とすると理想的には次の関係式が成立する.

$$s_i z_{i+1} = R_i^T (z_i - t_i) \quad (1)$$

そこで, 連続する観測の両方で検出された点全体について以下の誤差関数を最小化する観測値 R_i, t_i 及び s_i を考えることで観測値を補正する. このように最適化する対象を各座標系の変換のみに絞り, 各復元内の点の相対関係は固定することで計算量を削減している.

$$I_1 = \frac{1}{2} \sum \|s_i z_{i+1} - R_i^T (z_i - t_i)\|_2 \quad (2)$$

最適化には Adam を用いる.

上記のようにして補正された値を探査機の変位及び小惑星形状の観測値とする. また, 仮定 2 より補正された点群がある球面付近に位置していると考え以下に誤差関数を最小とする座標系を考えることで小惑星の重心 c_i を観測するものとする.

$$I_{center} = \frac{1}{N} \sum (r_{ave} - \|z_i - c_i\|)^2 \quad (3)$$

$$r_{ave} = \frac{1}{N} \sum \|z_i - c_i\| \quad (4)$$

小惑星の相対角速度 ω_i については探査機が各時刻において局所的に小惑星重心に対して円運動していると考えること以下
の式より観測するものとする。

$$\mathbf{v}_i = \mathbf{t}_i - \mathbf{c}_i \quad (5)$$

$$\omega_i = \frac{\mathbf{v}_i \cdot \mathbf{t}_i}{|\mathbf{v}_i| |\mathbf{t}_i|} \quad (6)$$

なお (2), (3), (4) における総和は連続する観測の両方で検出された点に関する総和を意味し、 N はそのような点の個数である。

3.2.2 EKF による小惑星形状パラメタを除く状態変数の推定

ここでは拡張カルマンフィルタを用いて状態量の更新を行う。観測方程式としては上述の式を用いることとする。姿勢表現に関しては簡単のため R をオイラー各表現 α , β , γ (順に y , x , z 軸周りの回転) に分解して考える。本問題では問題設定で述べた制約条件より探査機の取りうる姿勢は大まかに限定されているのでオイラー角による表現を用いても姿勢表現に問題はない。

状態遷移方程式としては、同様に制約条件より探査機は各時刻で小惑星重心に対して等速円運動しようとしているとモデル化出来るため、次のものを用いる。状態の表現は全て世界座標系における表現である。 r は各時刻における円軌道の半径の推定値、 Δy は探査機が軌道面から遠ざからないための位置に関する制御を仮想的にモデル化したものである。

$$x_{i+1} = x_i - r_i \sin \omega_i + w_{x,i} \quad (7)$$

$$y_{i+1} = y_i + \Delta y_i + w_{y,i} \quad (8)$$

$$z_{i+1} = z_i - r_i (1 - \cos \omega_i) + w_{z,i} \quad (9)$$

$$\mathbf{c}_{i+1} = \mathbf{c}_i \quad (10)$$

$$\alpha_{i+1} = \alpha_i + \omega_i + w_{\alpha,i} \quad (11)$$

$$\beta_{i+1} = \beta_i + w_{\beta,i} \quad (12)$$

$$\gamma_{i+1} = \gamma_i + w_{\gamma,i} \quad (13)$$

$$w_{i+1} = w_i \quad (14)$$

w は状態遷移ノイズを表す。

上記の状態遷移モデルは本来の物理モデルを正確に反映しているものとは言えないが、今回は等時間間隔で撮影される画像の観測と探査機は小惑星赤道面を小惑星重心を中心とする等速円運動をするように制御されるという事前情報だけが得られ、太陽系における探査機の位置やその他センサの情報は観測として得られないと仮定しているため、実際に起こるダイナミクスを単純かつ矛盾無く反映するような上記のモデルを採用した。

上記のモデルによる拡張カルマンフィルタにより観測値をフィルタリングすることで、その時刻の状態の推定値を得る。

3.3 三次元形状の再復元

カルマンフィルタによって推定された探査機の変位を用いて i 番目のカメラ座標系から見たカメラ行列 P_{i+1} を復元し、それを用いて再び時刻 i のカメラ座標系における三次元形状を再復元する。復元されたカメラ行列は始めに観測されたカメラ行列よりも尤もらしいものとなっていると考えられる。

世界座標系と i 番目のカメラ座標系への変換は推定されているため、その変換を用いて再復元された形状を世界座標系での表示に変換することが出来る。各時刻における形状の世界座標

系表現を順次足し合わせていくことで、小惑星の形状がポイントクラウドとして復元されていく。

しかし、本手法では時刻 i と時刻 $i+1$ の関係性のみを議論して推定を行っているため、このままでは復元して重ね合わせた小惑星形状等に誤差が蓄積していくと考えられる。よって、手順 4 では復元した形状について順次スムージングを行うことでその問題を解決する。

3.4 剛体制約による形状の最適化

ここでは小惑星のある点とある点の相対的位置関係は時不変である、すなわち小惑星は剛体であるという制約を用いて誤差を減少させる。観測の最適化と EKF による推定による推定が誤差無く行われていた場合、世界座標系から観測 i におけるカメラ座標系への変換を $R_{w,i}$, $\mathbf{t}_{w,i}$ とすると以下の等式が成立する。

$$R_{w,i+1} \mathbf{z}_{i+1} + \mathbf{t}_{w,i+1} = R_{w,i} \mathbf{z}_i + \mathbf{t}_{w,i} \quad (15)$$

しかし、実際には残留誤差のためにこの等式は成立しない。そこで、EKF により推定された R_i , $\mathbf{t}_i$ 及び s_i に対して微小な変化を許し以下の誤差関数を考える。

$$I_2 = \frac{1}{2} \sum w \|\mathbf{s}_i \mathbf{z}_{i+1} - R_i^T (\mathbf{z}_i - \mathbf{t}_i)\|_2 \quad (16)$$

w は復元された三次元点の位置に関する尤度であり、本手法ではカメラによる写像の歪みを考慮して画像上に投影された小惑星上で中心に近い点は大きい尤度を与え、外側の点には低い尤度を与えるようにした。最適化は Adam を用いて行い、求めたパラメタを EKF の推定に伝播させて補正することで誤差の蓄積を防ぐ。

4. 実験及び評価

実験に用いる画像列としては「はやぶさ 2」が実際に小惑星「リュウグウ」の赤道上空を周回しながら撮影した 2 種類の画像列 (2018 年 6 月 30 日撮影分, 同年 7 月 10 日撮影分) を用いた。なお、推定の際の全体のスケールに関しては、初期時刻の観測における小惑星表面から測った探査機の高度が 1000 になるようにスケールを正規化した。

推定の際のハイパーパラメタ等は以下の通りである。

表 2: ハイパーパラメタ等

名称	意味	値
$\mathbf{f}$	焦点距離	9348.1[px]
$\mathbf{c}$	光軸中心	512[px]
r_{th}	SIFT マッチングのための距離比	0.4
$v_{x,i}, v_{y,i}, v_{z,i}$	位置の観測に関するノイズの分散	5.0
$v_{\alpha,i}, v_{\beta,i}, v_{\gamma,i}$	姿勢の観測に関するノイズの分散	0.5[deg]
$v_{c_{x,i}}, v_{c_{y,i}}, v_{c_{z,i}}$	小惑星重心の観測に関するノイズの分散	10.0
$v_{\omega,i}$	相対角速度の観測に関するノイズの分散	0.3[deg]
$w_{x,i}, w_{y,i}, w_{z,i}$	位置の状態遷移に関するノイズの分散	2.0
$w_{\alpha,i}, w_{\beta,i}, w_{\gamma,i}$	姿勢の状態遷移に関するノイズの分散	0.1[deg]
$ \Delta y_i $	y 軸方向の微小な制御	0.5

どちらの画像列についてもほぼ同じ結果が得られたため本稿には 2018 年 6 月 30 日撮影分の画像列で撮影した結果を付する。

4.1 実験結果

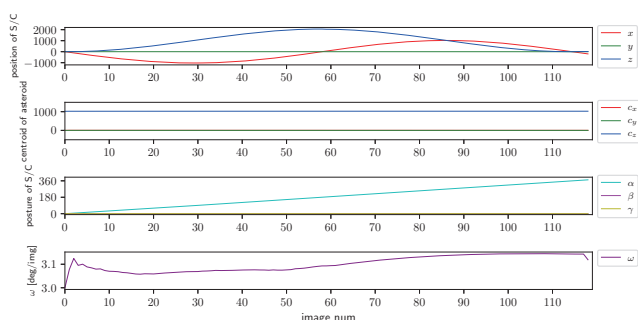


図 3: 各パラメタの推定結果

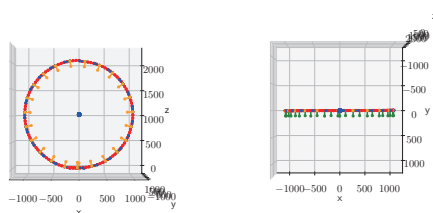


図 4: 全体の俯瞰図 1

図 5: 全体の俯瞰図 2

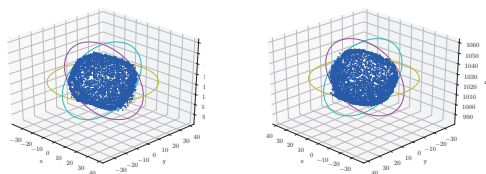


図 6: 小惑星形状 1

図 7: 小惑星形状 2

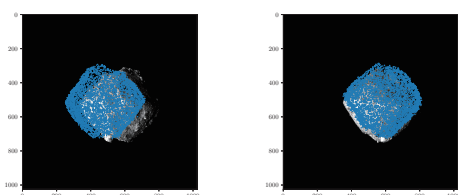


図 8: 再投影結果 (30 ステップ目)

図 9: 再投影結果 (60 ステップ目)

得られたポイントクラウドを構成する点の数は計 118778 個であり、ここでは図が煩雑になるのを避けるため適当に 5000 個を選び表示させている。

4.2 評価

各種の値について今回のデータセットより期待されている値を概ね満足な精度で推定出来ているように思われる。小惑星の

形状については正解データが無いため定量的な比較は出来ないものの、使用したデータセットから想像されるような形状を推定出来ており、本手法を用いることで小惑星形状及び探査機の相対運動が推定出来ていることが確認された。また、推定結果を用いて形状を画像上に再投影することで視覚的に推定誤差を大まかに計り知ることが可能である。再投影した画像を確認してみると、図 9 のように再投影が大体一致しているものもあり、図 8 のように再投影が大きくずれているものもあり、精度の改善の余地があることが分かった。

5. 結論と今後の展望

5.1 本研究のまとめ

本研究では、ある条件下で運動する探査機が撮影した画像列に基づき、小惑星の形状及び探査機の相対運動を推定する新たな手法の構築を目的としていたが、それに関してはある程度達成された。

5.2 今後の展望

今回の研究でさらに検討すべき点は以下の 2 点が主に考えられる。

1. 最適化アルゴリズムの精査。
2. 誤差関数の精査。

1 について、本手法では最適化のアルゴリズムとして十分に高速かつ安定的に良好な解を得られることが多かった Adam を選択したが最適化のアルゴリズムは他にも多く存在するので、それらについても検討してみるべきである。最適化の速度や大域的最適解への収束性を向上させることはオンライン推定に実用的な段階への大きな一歩になると思われる。

2 について、本手法では最適化の際の指標として空間内での点のユークリッド距離を考えたが精推定法である Bundle Adjustment[Triggs 00] のように再投影誤差についての評価を誤差関数として用いることで精度が向上するのではないかと考えられる。

参考文献

- [Shirakawa 06] K. Shirakawa et al. Accurate landmark tracking for navigating Hayabusa prior to final descent. In *Advances in the Astronautical Sciences*, volume 124, pages 1817–1826, 2006.
- [Lowe 04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, Nov 2004.
- [Kingma 14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [Triggs 00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment — a modern synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Vision Algorithms: Theory and Practice*, Springer, Berlin, 2000, pp. 298–375.

深層学習を用いたアパレルアイテム平置き画像から着装状態への変換

Image-to-image Translation from Apparel Item Image Placed Flat to Image Put on Using Deep Neural Networks

積際 早紀^{*1}
Saki Tsumugiwa

黒澤 義明^{*2}
Yoshiaki Kurosawa

目良 和也^{*2}
Kazuya Mera

竹澤 寿幸^{*2}
Toshiyuki Takezawa

^{*1} 広島市立大学 情報科学部
School of Information Sciences
Hiroshima City University

^{*2} 広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences
Hiroshima City University

This paper deals with image-to-image translation of apparel items. The images are difficult to be translated because the items are variously set, when they are took photos: being placed flat, being put on the mannequin and so on. We try to investigate and improve the previous work also known as ‘pix2pix’ based on deep neural networks, especially deep convolutional generative adversarial network (DCGAN). We propose a new two-stage procedure. Some experimentation revealed that our proposed method was superior to the previous work, evaluated using structural similarity index. Moreover, we confirmed it generated item details (zipper, button) and patterns (dot) as the result of visual confirmation. This knowledge is very important because the fault image of the item without buttons should be completely different from the original item image.

1. はじめに

インターネットの発展に伴い、各種通販サイトが開設されてきた。アパレル分野も例外ではなく、各メーカー/ブランドのサイトに加え、ZOZOTOWN, MAGASEEK 等の大規模複合サイトが開設され、今に至っている。

これら複合サイトの利点は、様々なブランドの商品を一度に閲覧し、また購入できることである。その一方で、ブランド毎に用意される画像が統一しておらず、ユーザのイメージ想起を困難にすることがある。

例えば、サーキュラースカート(図 1(a))のように広がり大きいスカートは、着装した時に「フレア感」と呼ばれる印象を形成する。この点、図 1(a)左よりも図 1(a)右の方が印象を伝えやすい。しかし、短いサイクルでデザインから縫製、販売まで手掛ける現代のファストファッションでは、写真撮影に時間を割いてばかりもいられない。このため、『基本的な写真(例えば、平置き画像)からの、自動的なバリエーション画像(例えば、着用画像)変換』が望ましいと考えられる。

そこで本研究は、このような異なるバリエーション画像を、深層学習を用いて自動で変換し、ユーザの印象形成を容易にすることを目的とする。特に、図 1(a)のように変形の大きい対象についても変換できる手法の確立を目指す。



図 1 アパレルサイト内バリエーションの例

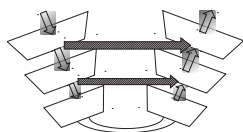


図 2 U-Net の模式図

2. 関連研究

本研究はアパレルアイテム画像から、様々なバリエーション画像の変換生成を目的とする。深層学習を用いた画像変換の研究として、いわゆる pix2pix が挙げられる[Isola 17].

本研究はこの pix2pix を利用して、大きな変形を伴う画像変換を実現する。そこで、まず本研究にとって重要な pix2pix の特徴と検討項目を挙げた上で、関連研究を 2 つ挙げる。

2.1 pix2pix

pix2pix は conditional Generative Adversarial Networks (GAN) の一種に基づく画像変換器である[Isola 17]. 線画への着色やセグメンテーション等、様々な対象に利用されている。

次に pix2pix の特徴を 2 つ挙げる。

(1) Generative Adversarial Networks (GAN)

第一の特徴は、2 つの畳み込みニューラルネットワーク(生成器 G(Generator)と判別器 D(Discriminator))を使用する点である。

まず生成器 G は、訓練画像に似た画像(fake)を生成するように訓練される。次に判別器 D は、D に入力された画像が訓練画像か、G が生成した fake かを判定する。そして、G は D を騙すように訓練され、D は G に騙されないように訓練される。このように、G と D が互いに争い合うことで画像の変換精度を高める。

(2) U-Net

次の特徴は、畳み込みネットワークに U-Net を使用している点である。このネットワークは、医療分野での画像セグメンテーション(細胞の着色 etc)を目的として開発された[Ronneberger 15]. 重要な点は、U-Net という名前が示す構造にある(図 2)。

U 字の左右をつなぐ結合(skip connection)がある(右向き車線矢印)。この結合により、U 字の右側では下層から出力された特徴マップに、左側の対応する層からの特徴マップが加わる。これがお手本のようになり、結合のない単なる Encoder-Decoder モデルよりも、速く正確に学習する。一般的に、GAN による画像生成は細部が破綻することも多く、U-Net を使う利点は大きい。

連絡先: 黒澤 義明, 広島市立大学大学院 情報科学研究科,
広島市安佐南区大塚東 3-4-1, 082-830-1500(代),
kurosawa@hiroshima-cu.ac.jp

2.2 pix2pix における要検討項目

[Isola 17]は車や人のセグメンテーション、線画に対する着色等、様々な変換事例を紹介している。しかし、pixel 同士の対応を前提としており、変形を伴う対象を検討していない。よく似たCycleGAN [Zhu 17]では、変形の問題が指摘されているため、吟味する必要があると言える。

本研究が対象とするアパレルアイテムは、図 1(a)に挙げたような変形を含む。こうしたアイテムの変形が可能であれば、背景からのアイテム抽出や、アイテムの変形加工等の複数の処理を、モデルを変えるだけで実行できるという利点がある。

以上の観点から、pix2pix が変形に対応可能かどうか議論することを本研究の課題の 1 つとする。問題があるとすればどのような点か、そしてどのように解決するかに検討する。

次に、アパレルアイテムに関する関連研究について、以下に補足しておく。

2.3 Pixel Level Domain Transfer

pix2pix 以外にも変換を試みる研究がある。[Yoo 16]は、本研究同様、アパレルアイテム画像の様態変換を行っている。この研究も GAN を利用している。

異なる点は、U-Net ではなく単純な Encoder-Decoder モデルを使用していること、G と D に加えて、Domain 判別用の別のネットワークが用意されている点、の 2 点である。先に述べたように、Encoder-Decoder モデルより U-Net が優れる。

2.4 Detailed Garment Recovery

単なる着用画像の変換だけではなく、動画化を試みた研究もある[Yang 16]。フレアスカートのフレアを揺らすことができる等、ユーザのイメージ想起に有効に働く手法である。

ただ、彼らの手法の問題は、型紙から衣服を作るかのような詳細なアイテム記述が必要なことであり、コストがかかる点である。短いサイクルで様々なアイテムが発売されるファストファッションの場合には、このようなコストをかけることは困難である。したがって、シンプルな画像変換による方法の方が、コストが少なく、より有効であると考えられる。

3. 提案手法

本研究の提案は、pix2pix を用いた変形を伴う画像変換に対する解決法である。以下に、その手法について述べる。

先述の通り、pix2pix の手法は、画像間のピクセルについて対応が取れているとき、または対応のずれが小さいときには有効である。一方で、変形が大きく、対応のずれが大きくなると、学習が困難になると考えられる。

この理由としては、pix2pix の利点となっている skip connection に原因がある。すなわち、お手本として振る舞うべきであるのに、対応のずれが災いする。つまり、オリジナルの画像情報が反映されない学習結果となるわけである。

そこで本研究は、お手本となるべき入力を新たに導入するため、2 段階の処理を採用し、つなぎ目に処理を加える方法を提案する(図 3)。

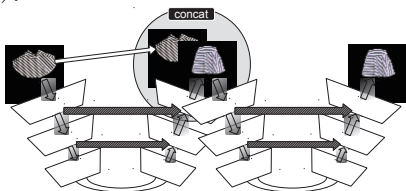


図 3 本研究の提案手法

本研究の手法はシンプルであり、2 つの pix2pix をつなぎ、そのつなぎ目で、1 つめの pix2pix の出力と、最初の入力(白抜き矢印)を重ね合わせて(concat), 2 つめの pix2pix への入力とする。これにより、オリジナル画像の情報を有した 6 次元の画像による学習が可能となる。これにより、変形への対応を目指す。

4. 実験と考察

本節では、pix2pix の変形にの可否に関する調査を行い、その問題点を解消する実験について報告する。

4.1 本研究で使用するデータ

本稿では対象をスカート、また、平置き画像(図 1(a, b)左)から着装状態(図 1(a, b)右)への変換に限定する。に限定した理由は、変形量が大きく、検証に有効と考えたからである。図 1(a)が変形の大きい例、(b)が小さいタイトスカートの例である。

以下に、本研究で利用したデータの詳細について述べる。

(1) 撮影

1 枚のスカートに付き、平置き画像(床に広げ、上方から撮影)と着装状態画像(マネキンに着用)の 2 枚ずつを撮影した。その後、それぞれ背景を削除した。アイテム数は 283 であった。

(2) Data Augmentation

上記アイテム数は、深層学習に十分な量とは言えないため、data augmentation を行った。以下の表 1 の命令からランダムに 4 種類を選び、色等を変更する。なお、オプションランダムに選ぶ。

上の手続きを 1 枚のアイテムにつき 21 回行った上で、さらに左右反転を行った。これにより、オリジナルの画像も含め、22 (21+1) x 2 (反転) 倍の画像となった (283x44=12,452 画像)。

(3) 画像サイズ調整

平置きと着装状態は別々に撮影しているため、サイズの調整を行う。それぞれ、輪郭抽出を行った後、外接四角形を決定する。そして、着装画像の縦サイズに合わせ、縦横比を維持したまま、平置き画像を縮小/拡大し、縦横 256 pixel の画像を得た。

4.2 実験 1 ~基礎的な変形に関する検討~

本研究は pix2pix の実装として、原著者らが公開している実装を用いた。なお、特に断らない限り、デフォルトのパラメータを使用する。GPU は Geforce GTX 1080Ti を使用した。

まず基本的な変形に関する情報収集のため、判別器 D の層数を 5 に変更(デフォルト:3)し、バッチサイズを 10 で 1,000 epoch (同:200) 繰り返す実験を行った。学習時の loss の推移を図 5 に示す。学習開始直後の L1 の値が大きく見づらいため、補正した図になっている。

表 1 data augmentation に使用した命令とオプション

命令	オプション		
画像の伸縮	[0.9], [1.05], [1.1]		
回転	[3], [5], [7] (度)		
透視変換	[0.015] (画像幅 x0.015) , [0.03], [0.05], [0.07],		
色交換	[0, 2, 1, 3] (BGR→BRG) , [1, 0, 2, 3], [1, 2, 0, 3], [2, 0, 1, 3], [2, 1, 0, 3]		
色軽減	[0.95] (Bx0.95, Gx0.95, Rx0.95) , [0.9]		
特定色軽減	[0, 0.95] (Bx0.95) , [1, 0.95], [2, 0.95]		
色除去/半減	[0, 0] (Bx0) , [0, 0.5], [1, 0], [1, 0.5], [2, 0], [2, 0.5]		
HSV 円環上の移動	[-40], [-30], [-20], [-10], [10], [20], [30], [40]		
平滑化/ 先鋭化	[0], [1], [2]		
	[1]: Gaussian Filter Kernel Size(11x11)		
	[1]: [-1, -1, -1] [-1, 9, -1] [-1, -1, -1]		
	[2]: [0, -1, 0] [-1, 5, -1] [0, -1, 0]		

ここで、「G_GAN」は『生成器 G が判別器 D を騙しているか』についての指標、「G_L1」は『生成器 G が生成した fake(predict) と正解(ground truth)との L1 距離』に関する指標である。今回の学習では収束条件が L1 となっている。また、「D_real」は『正解に対する、判別器 D の判定』、「D_fake」は『fake に対する、判別器 D の判定』について表している。

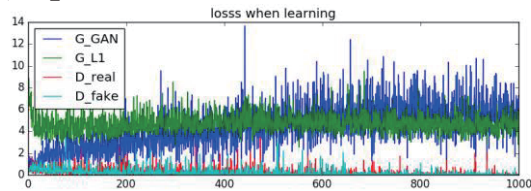


図 5 loss の推移 (補正後)

G_L1, G_GAN 共に変動が大きいことがわかる。また、L1 が横這いである点も学習が不十分である可能性を示唆している。このため今回の検討は、以下の実験を含め、全てクロズドテストで行うこととする。学習時のアイテムをモデルに入力し、生成した画像を挙げる(図 6～図 8)。

変形の必要なサーキュラスカート(図 6)やフレアスカート(図 7)では、アイテム領域の内外に破綻が見られる(ex. 600 epoch)。ただ、変形はおおむね成功しているように見える。また、シンプルなタイトスカート(図 8)でも変形はできていても、テクスチャの変換には成功していない。仮説としては、ネットワークの重み学習が変形を中心に行われて、余裕がなかったのではないかと、つまり学習資源に関するパラメータ設定が必要と考えられる。

4.3 実験 2 ～パラメータ変更～

前項の仮説を確認するため、生成器 G が持つカーネルの数を増加させる。デフォルトが 512 の層 1024 にする等、各層 2 倍にする¹。また、性能向上のため、判別器 D の性能向上を目指しカーネルの大きさを 1 にした実験を行う²。アイテムを学習モデルに入力した例を、次の図 9～図 11 に示す。

実験 1 では成功していなかったストライプ(図 6)の復元に成功している(図 9)。また図 10 では、フレアスカートのフレア感も再現している。一方で、ドット柄の再現については不十分であり、さらに図 11 からわかるように、ディテール(ボタンやジッパー)がどの epoch でも再現されていない。

つまり、重み学習のためにパラメータの変更～特に、カーネル数等の増加～を行うことは重要である。しかし、それでも細かい模様やディテールを再現できないという問題が残る。

4.4 実験 3 提案手法 ～2 段階処理への拡張～

3 節に述べた提案手法に基づく実験を行う。第一段階には前節で学習した中で、200 epoch のモデルを選んだ。200 はデフォルトの値である。

このモデルにより得られた変換画像を 6 次元に合成した上で、2 段階目の学習を行った。結果を図に示す(図 12, 図 13)。

図 10 に比べた図 12 では、細かいドット柄を表現できていることがわかる。また、図 11 で不十分だったディテールが、図 13 では表現できていることもわかる。

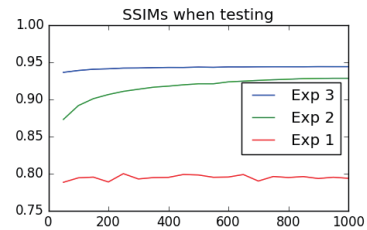


図 4 3 実験比較を目的とした SSIM 値の推移

4.5 考察

変換がどの程度向上したか、以下の Structural Similarity (SSIM) を用いて評価する。画素中の小領域の平均・分散・共分散を用いて、局所的な破綻に敏感であり、最良値は 1 である。

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

撮影画像 283 枚それぞれについて、50 epoch ごとに平置き・着装間の平均 SSIM を求めた(図 4)。

実験 1 と比べ、実験 2 と実験 3(提案手法)との値が高くなっていることがわかる。また実験 3 が優れていることも明らかである。

ただ、実験 2 との差はあまりなく、pix2pix のパラメータを変更しただけであるので、「提案手法のようなコストをかけなくてもよいのでは」という指摘もあろう。しかし、前節で述べたように、見た目の変化は大きい(図 10 vs. 図 12, 図 11 vs. 図 13)。特に、ボタン等のディテールが再現できなければ、全く別のデザインになることは指摘しておきたい。その意味で、数値以上に提案手法が優れていると言える。

5. おわりに

アパレルアイテム画像の様態変換を目的に、pix2pix を利用した 2 段階の学習方法を提案した。

実験の結果、SSIM による数値による比較で、本提案が優れていることが明らかになった。pix2pix は細かい模様やディテールを再現できない一方、提案手法では再現できたからである。

今後の課題としては、アイテム数を増やした上で、さらなる検討を試みる事が挙げられる。また SSIM 値は、ディテールの有無等を反映していないこともわかった。さらなる改善のため、適切な指標を検討することも必要であろう。

謝辞

この研究の一部は、国立研究開発法人科学技術振興機構(JST)の研究成果展開事業「センター・オブ・イノベーション(COI)プログラム」・広島市立大学特定研究費(先端学術研究費 H27～29, 30 年度科研費獲得支援費)の補助を得ている。

参考文献

- [Isola 17] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros : Image-to-Image Translation with Conditional Adversarial Networks, In IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [Ronneberger 15] O. Ronneberger, P. Fischer, T. Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597, 2015.

¹ 関数 UnetGenerator の ngf を 64 から 128 に変更した。

² 関数 PixelDiscriminator. なお、実験 1 と 2 の間にはカーネル

のサイズを変更した複数の実験がある。ページの都合により今回は割愛する。

[Yang 16] S. Yang, T. Ambert, Z. Pan, K. Wang, L. Yu, T. Berg, M. C. Lin: Detailed Garment Recovery from a Single-View Image, arXiv:1608.01250v4, 2016.

[Yoo 16] D. Yoo, N. Kim, S. Park, A. S. Paek, I. S. Kweon: Pixel-Level Domain Transfer, arXiv:1603.07442, 2016.

[Zhu 17] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, In IEEE International Conference on Computer Vision, 2017.

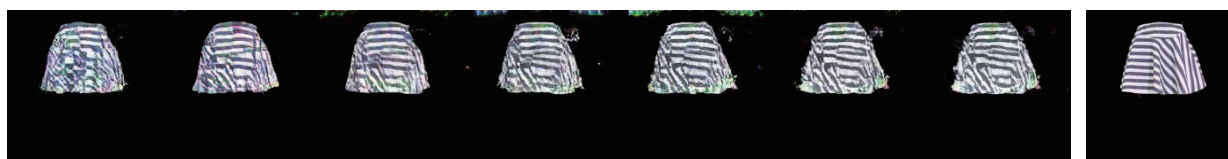


図 6 実験 1 item002: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)



図 7 実験 1 item215: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)



図 8 実験 1 item200: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)



図 9 実験 2 item002: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)



図 10 実験 2 item226: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)

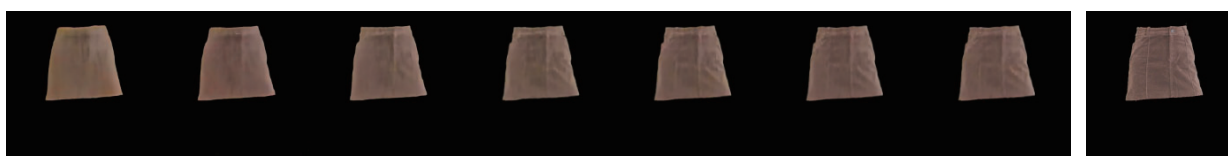


図 11 実験 2 item200: epoch 毎生成画像 (左から 50, 100, 200, 400, 600, 800, 1000 epoch, ground truth)



図 12 実験 3 item226: epoch 毎生成画像 (左から 50, 100, 200, 500, 550, 600, 800 epoch, ground truth)



図 13 実験 3 item200: epoch 毎生成画像 (左から 50, 100, 200, 500, 550, 600, 800 epoch, ground truth)

走査電子顕微鏡におけるライブ像のノイズ低減

Noise reduction of live image in scanning electron microscope

植松 文徳^{*1}
Fuminori Uematsu

武井 雅彦^{*1}
Masahiko Takei

吉田 光良^{*1}
Mitsuyoshi Yoshida

^{*1} 日本電子株式会社
JEOL Ltd.

A real time display called a live image is used to search for a region to be observed, when observing a sample with a scanning electron microscope (SEM). This live image is usually an image with high noise and poor visibility. This makes it difficult to find an appropriate observation area. Therefore, in this research, a noise reduction model by deep neural network was created with the aim of improving the visibility of live images. We incorporated this model into the SEM and succeeded in obtaining a live image with less noise. This makes it possible to efficiently search for the observation region.

1. 背景と動機

走査電子顕微鏡(SEM)で試料を観察する際には、試料の観察したい領域が電子顕微鏡の表示領域(視野)に含まれるようユーザーが装置を操作する必要がある。これを視野探しと呼び、具体的には試料ステージの移動や観察倍率の変更などを行う。視野の画像がリアルタイムで表示されるので、視野探しで表示される像をライブ像と呼ぶ。

ライブ像では視野探しの操作に対して十分な応答性が得られるよう、高いフレームレートを使用するのが一般的である。すなわち、電子ビームの走査が速く、検出器で観測される時間当たりの電子が少ない。そのため検出信号にノイズが発生しやすい。また、応答性が落ちるという理由から、ライブ像では複数フレームの積算によるノイズ除去を無効にすることも多い。その結果、ライブ像はノイズの多い画像となるのが常態である。

SEMの熟練ユーザーが扱い慣れた試料を観察する場合であれば、ノイズの多いライブ像であっても、ノイズの無い状態にある程度推測しながら視野探しを行える。しかし、そのような推測は経験の浅いユーザーにとって困難であり、視野探しに長い時間を要したり、観察したい対象物を見落としたりする可能性がある。

ライブ像のノイズを低減できれば、視野探しをより効率的に見落としなく行うことが可能となり、高い有用性が期待できる。本研究ではライブ像のノイズ低減にディープラーニングを用いることを提案する。

2. 提案手法

2.1 モデル設計

ノイズの多い SEM の観察像をノイズの少ない観察像に変換するモデル(ノイズ低減モデル)を作成し、それを SEM に組み込んでライブ像のノイズを低減させる。SEM でノイズの少ない観察像を得るには、電子ビームの走査を遅くし、さらに複数フレームの積算によるノイズ除去を用いれば良い。これは静止画を保存するときの観察条件であり、このようにして得られる像をフォト像と呼ぶ。従って本研究ではライブ像からフォト像への変換を学習させることで、ノイズ低減モデルを作成する。

ライブ像はフォト像の一部がノイズによって隠された状態と見

なせるため、ライブ像をノイズ除去した結果は一意に定まらない。そのため Encoder-Decoder モデルを二乗誤差で学習させてノイズ低減モデルを作成しても、ライブ像から推定される多数のフォト像を平均したような、不明瞭な像が得られると予想される。そこで本研究ではノイズ低減モデルを学習させる手法として、Conditional GAN を採用した。この手法ではライブ像をフォト像に変換する生成器と、フォト像が生成された物か本物かを識別する識別器の二つのモデルを相互に競わせて学習する。すなわち、生成器は自身が変換した画像が本物であると識別器に誤認させようとする。一方、識別器は与えられたフォト像を正確に識別しようとする。その際、生成したフォト像が不明瞭であると本物ではないことが容易に判別できてしまうので、明瞭な画像を生成するように生成器の学習が進むと期待できる。

モデル設計には pix2pixHD[Wang 17]で提案されている手法を取り入れた。これは高品質で高解像度の画像が生成可能な Conditional GAN の手法であり、次のような特徴を有している。

- 階層化された生成器
- マルチスケールに対応した識別器
- 複数の誤差関数の組み合わせ

pix2pixHD ではセマンティックラベルマップから写真のようにリアルな画像を生成しているが、本研究は同様のモデルを SEM 観察像のノイズ除去に適用した。

(1) 生成器のアーキテクチャ

本研究では pix2pixHD と同様に粗い解像度の生成器(Global Generator)と密な解像度の生成器(Local Enhancer)を階層的に結合した生成器を用いた。ここでは Johnson ら[Johnson 16]や CycleGAN[Zhu 17]や pix2pixHD で用いられる表記の慣習に従って次のように表現する。

- フィルター数 k でストライド 1 の 7×7 畳み込み層-InstanceNorm[Ulyanov 16]層-ReLU 層を $c7s1-k$ と表現する
- フィルター数 k でストライド 2 の 3×3 畳み込み層-InstanceNorm 層-ReLU 層を dk と表現する
- フィルター数 k の 3×3 畳み込み層を二つ含む Residual ブロックを Rk と表現する
- フィルター数 k でストライド 1/2 の 3×3 転置畳み込み層-InstanceNorm 層-ReLU 層を uk と表現する

生成器のアーキテクチャは、Global Generator を c7s1-32, d64, d128, d256, d512, R512, R512, R512, R512, R512, R512, R512, R512, u256, u128, u64, u32, c7s1-1 とし、Local Enhancer を c7s1-16, d32, R32, R32, R32, u16, c7s1-1 とした。

(2) 識別器のアーキテクチャ

識別器においても pix2pixHD と同様に画像サイズの異なる複数の識別器を結合して用いた。識別器の個数は、Global Generator の学習では 2 個、Local Enhancer を結合した学習では 3 個とした。ここで次のように表現する。

- フィルター数 k でストライド 2 の 4×4 畳み込み層-InstanceNorm 層-LeakyReLU 層を C_k と表現する

いずれの識別器も $C32$ - $C64$ - $C128$ - $C256$ とし、その後にフィルター数 1 でストライド 1 の 4×4 畳み込み層を繋げたアーキテクチャとした。ただし、 $C32$ の InstanceNorm 層は省略した。

2.2 教師データ

教師データには同一視野を観察したライブ像とフォト像のペアが必要となるが、これを大量に取得するには長時間の SEM 操作が必要となり、容易ではない。そこで本研究ではフォト像のみを取得し、それに人工的なノイズを付加することによって、疑似ライブ像を作成して用いた。本手法であれば既に所有しているフォト像を教師データとして活用できる利点がある。SEM の二次電子像において支配的なノイズは、検出器で観測される電子がポアソン分布に従っていることに由来するショットノイズである。そこでフォト像にこのようなショットノイズを付加して疑似ライブ像を作成した。また教師データの数を増やすため、フォト像に拡張・回転・並進・反転といった画像変換を行ってデータ拡張した後に疑似ライブ像を作成した。

3. 実験

3.1 学習

対象とする画像の解像度は、日本電子製の SEM である JSM-7900F のフォト像と同じ 1280×960 ピクセルとした。教師データとして疑似ライブ像とフォト像のペアを約 2 万データ作成し、GeForce GTX 1080(NVIDIA 社製)を 1 基搭載した Ubuntu サーバーで学習を行った。Global Generator を 40 エポック学習した後、Local Enhancer を結合して 20 エポック学習した。フレームワークには Microsoft Cognitive Toolkitを使用した。

3.2 評価

学習を行って作成したノイズ低減モデルを JSM-7900F に組み込み、ノイズ低減されたライブ像を観察して効果を評価した。視野探しに必要な応答性を確保するため、SEM の制御用 PC に GeForce GTX 1080 を 1 基搭載し、画像辺りの変換時間が 100ms 以下となるようにした。

同一視野のライブ像(変換前)、ライブ像(変換後)、フォト像を図 1 に示す。ライブ像のノイズが低減され、フォト像と比較しても自然な画像となっており、視認性が向上していることがわかる。一方で、本手法ではライブ像をフレーム単位で変換しているため、各フレームのノイズの偏りが変換後の画像に僅かに影響を与える。これによって、リアルタイム表示したときに試料表面が波打って見えるなどの違和感を生じる場合があった。

4. 結論と今後の課題

本研究ではディープラーニングを用いることで走査電子顕微鏡におけるライブ像のノイズ低減が可能であることを確認し、効率的な視野探しへの寄与が期待できる結果が得られた。

今回は効果の測定を目視による主観に頼ったが、今後は非参照メトリックなどを用いた定量的な知覚品質の測定に取り組む必要がある。また、vid2vid[Wang 18]で行われているように、変換結果に時系列での一貫性を持たせるなどして、試料表面が波打って見える現象を解消することが今後の課題となる。

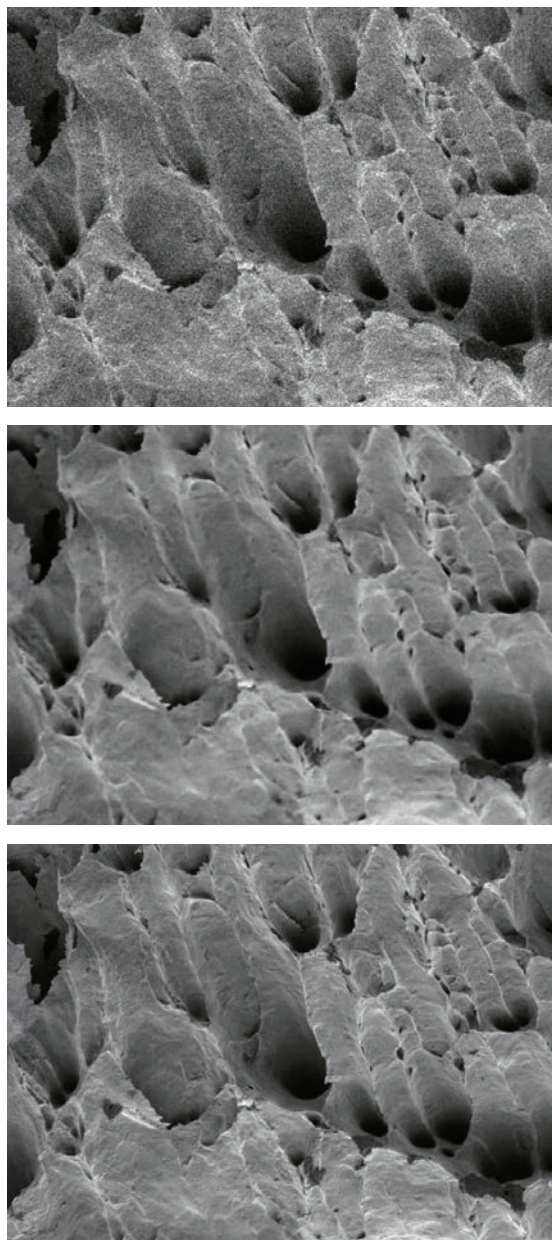


図 1 SEM ライブ像の変換例。上段:ライブ像(変換前)、中段:ライブ像(変換後)、下段:フォト像

参考文献

- [Wang 17] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz and Bryan Catanzaro: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, arXiv:1711.11585, 2017.

- [Johnson 16] Justin Johnson, Alexandre Alahi and Li Fei-Fei: Perceptual losses for real-time style transfer and super-resolution, arXiv:1603.08155, 2016.
- [Zhu 17] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, arXiv:1703.10593, 2017.
- [Ulyanov 16] Dmitry Ulyanov, Andrea Vedaldi and Victor Lempitsky: Instance Normalization: The Missing Ingredient for Fast Stylization, arXiv:1607.08022, 2016.
- [Wang 18] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz and Bryan Catanzaro: Video-to-Video Synthesis, arXiv:1808.06601, 2018.

交通サービスのための耐戦略性を有する動的課金メカニズム

Truthful Dynamic Pricing Mechanisms for On-demand Mobility Services

早川 敬一郎^{*1}

Keiichiro Hayakawa

羽藤 英二^{*2}

Eiji Hato

^{*1}豊田中央研究所

Toyota Central R&D Labs., Inc.

^{*2}東京大学大学院

The University of Tokyo

Transportation services under which multiple traffic mode services are provided by a service operator through mobile apps are often called Mobility as a Service (MaaS). Commonly, the mobile app has the function of activity loggers, which represents the heterogeneity of each user. Thus, the traffic allocations and pricing algorithms that properly handle this heterogeneity of users is required. In this paper, we present a conceptual framework for the pricing mechanisms for such dynamic on-demand traffic services.

Specifically, we aim to establish Bayesian-Nash incentive compatible mechanisms, by which the dynamic system optimal state is achieved by the best response strategy of selfish agents. We introduce two common mechanisms to our MaaS settings, one which guarantees non-negative *ex-post* revenue and the other non-negative *ex-post* utility of agents. In the numerical study, we show that these two mechanisms have a trade-off between the revenue of the service operator and the benefit of customers.

1. はじめに

近年、複数の交通モードを統合した移動をサービスとして提供する MaaS (Mobility as a Service) の取り組みに注目が集まっている。MaaS の多くは顧客に対してケータイアプリを提供し、顧客はアプリを通して交通資源（例えば、列車の指定席やライドシェアへの相乗りなど）を予約する。一方でケータイアプリは顧客の日々の行動を記録するロガーとしても活用されており、顧客の多様性を考慮した有効な交通資源配分が実現可能である。人々の多様な嗜好と制約の下で、有限の交通資源を上手く配分するアルゴリズムの確立が求められる。

このような問題に対して、[Hayakawa 18b] は、Activity-based model [Kitamura 96] を用いて利用者のトリップチェーンを考え、利用者の持つ時間と場所に関する制約 [Hägerstrand 70] を考慮した交通サービス設計の考え方を示している。具体的には、受け入れた全ての利用者の時空間制約と交通容量制約を将来にわたって守ることを保証する交通資源割り当てメカニズムを、*RC(Resource-Customer)-feasible* アルゴリズムと定義し、その中でも特に、各時刻の割引総社会効用 [Rust 94] を最大化するアルゴリズムを *RC-optimal* アルゴリズムと定義している。

本研究では、顧客の利己的な行動選択を前提として、利用者および交通事業者双方の多くの制約の下での最適状態が導かれるような課金制度を検討する。具体的には、オークション理論の分野で知られる *Dynamic pivot* メカニズム [Bergemann 10] と、*Online VCG* メカニズム [Parkes 04] の二つのメカニズムを導入し、その特性を評価する。

2. オンデマンド型交通サービスモデル

本研究で想定する交通サービスを示す。まず、複数の交通資源（列車、バス、タクシーなど）を有するサービスオペレータは、利用者に対してケータイアプリを提供する。利用者は、交通サービスを利用する際に、ケータイアプリ上の予約入力画面

から、自身の希望（「嗜好」と「制約」）を入力する。ケータイアプリ上に実装された User Agent は、入力された情報と、日々の行動観測から得られた利用者の嗜好の情報などを踏まえて、トリッププランの候補を複数生成し、オペレータに送信する。オペレータは、まずその利用者を受け入れるか否かを判断し、受け入れた全ての利用者に対してトリッププランを提示し、保有する交通資源を動的に割り当てる。さらに、割り当てられた交通資源に応じて、利用者にサービス利用料金を課金する。

本研究では、既往研究 [Hayakawa 18b, 早川 18] と同様に、離散時間 $t \in T = \{0, 1, \dots, \bar{T}\}$ を考える。交通ネットワークは、ノードの集合 $\mathcal{N}$ とエッジの集合 $\mathcal{E}$ からなる片方向グラフ $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ で表されるものとする。このグラフ上では、同一地点間を結ぶ異なる交通モードは異なるエッジで表される [Sheffi 85]。エージェントの集合を I とし、エージェントのモデルとしては、[Hayakawa 18b] と同様の状態遷移モデルを用いる。すなわち、エージェントのタイプは、報酬関数は、 $R_i : \mathcal{S}_i \times \mathcal{A}_i \times T \rightarrow \mathbb{R}$ と、実行可能なトリッププランの集合 L_i を用いて、 $\theta_i = \{R_i, L_i\}$ で表される。ここで、エージェントのトリッププラン $l_i \in L_i$ は、活動時間帯 $T_i = [t_i^B, t_i^E]$ における一連の状態とアクションの遷移として表されるものとする。ただし、状態 $s_{i,t}$ は位置や交通モードなどを含む多次元変数であり、アクション $a_{i,t}$ は移動もしくは滞在を表すものとする。エージェントは、出発時刻 t_i^B 以降の任意の時刻に、サービスオペレータに対して自らのタイプを申告し、オペレータは、当該時刻までに得られた申告に基づいて、各時刻ごとに交通リソースの割り当てと課金額を決定するものとする。

3. メカニズム

本研究では、前章に示した交通サービスに対して、Dynamic pivot [Bergemann 10] と Online VCG [Parkes 04] の二つのメカニズムを適用する。これらのメカニズムは、共に、各時刻の割引総社会効用 [Rust 94] を最大化する資源割り当てメカニズムを用いるが、課金方法が異なる。Dynamic pivot メカニズムは、各時刻に各エージェントが社会に与える外部性の期待

連絡先: 早川 敬一郎, 豊田中央研究所, 愛知県長久手市横道 41-1, kei-hayakawa@mosk.tytlabs.co.jp

値に相当する額を課金する。一方、Online VCG は申し込み時点で計算された外部性の期待値に相当する額を基本としてその後の割り当てに応じて課金額が調整される仕組みとなっている。

3.1 効率的な交通資源の割り当て

まず、交通資源の割り当て方法について考える。この方法は、二つのメカニズムに共通であり、割引総社会効用を最大化することが求められる。その割り当て方法は、以下の式で表される。

$$\pi_t = \operatorname{argmax}_{\pi'_t \in A(s_t)} \mathbf{E} \left[\sum_{t \in \bar{T}} \beta^{t'-t} R(s_t, \pi'_t) \right]. \quad (1)$$

ベルマンの最適性原理により、この式は以下のように書き換えられる。

$$\pi_t = \operatorname{argmax}_{\pi'_t \in A(s_t)} [R(s_t, \pi'_t) + \beta \cdot V(\mathcal{T}(s_t, \pi'_t))], \quad (2)$$

ただし、 $V(s_t)$ は、以下の式で与えられる価値関数である。

$$V(s_t) = \mathbf{E} \left[\sum_{t'=t}^{\bar{T}} \beta^{t'-t} R(s_{t'}, \pi_{t'}) \right], \quad (3)$$

既往研究 [Hayakawa 18b, 早川 18] で示したように、この計算は ZDD [Minato 93] を用いて効率的に実行できる。

3.2 Dynamic pivot mechanism の課金方法

Dynamic pivot mechanism [Bergemann 10] の課金方法は、以下の式で与えられる。

$$\begin{aligned} x_{i,t} = & -R_{-i}(a_t) - \beta \cdot \mathbf{E}[V_{-i}(s', \pi_{\theta_{-i}}^*) | s' = \mathcal{T}(s_t, a_t)] \\ & + V_{-i}(s_t, \pi_{\theta_{-i}}^*), \end{aligned} \quad (4)$$

ここで π^* は前節で示した効率的な交通資源割り当てを、 R_{-i} はエージェント i を除く全エージェントの報酬の合計を、 V_{-i} はエージェント i を除く全エージェントの価値関数を示す。この式はすなわち、「各エージェントは各時刻に社会に与えた迷惑料を支払う」、ということを表している。

このメカニズムは、Bayesian-Nash incentive compatible かつ *ex-post* budget balanced である。各エージェントのが利己的にベストレスポンス戦略を実行するという仮定の下で、割引総社会効用が最大化される。また、メカニズムに参加する（交通プランをリクエストする）ことによって負の効用が生じないことは、確率的に保証される。ただし、条件によっては、エージェントはメカニズムに参加することで負の効用を得る（損をする）可能性がある。

3.3 Online VCG mechanism の課金方法

Online VCG mechanism [Parkes 04] の課金方法は、以下の式で与えられる。

$$x_{i,t} = \begin{cases} R_i(a_{i,t}) - V(s_t, \pi_{\theta}^*) + V_{-i}(s_t, \pi_{\theta_{-i}}^*) & (t = t_i^B) \\ R_i(a_{i,t}) & (\text{otherwise}). \end{cases} \quad (5)$$

この式はすなわち、「各エージェントは、最初の申告の時点で社会に与える迷惑料を支払った上で、実際に発生した事象に応じて支払額を調整する」ということを表している。

このメカニズムは、Bayesian-Nash incentive compatible かつ *ex-post* individual rational である。各エージェントのが利

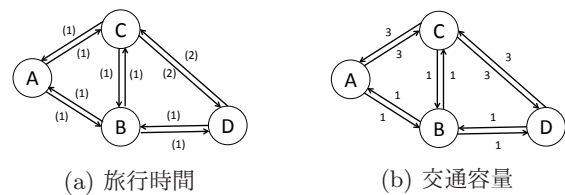


図 1: Sample network

己的にベストレスポンス戦略を実行するという仮定の下で、割引総社会効用が最大化される点は、Dynamic pivot mechanism と同じであるが、メカニズムに参加する（交通プランをリクエストする）ことによって負の効用が生じないことが確定的に保証されている。

4. 数値実験

4.1 実験条件

今回の実験では、制御対象時刻は $T = \{0, 1, \dots, 8\}$ の 8 タイムステップとし、時間割引率は $\beta = 1$ とした。以下に、交通ネットワークおよびエージェントに関する実験条件を示す。

4.1.1 交通ネットワーク

実験では、図 1 に示すネットワークを考えた。このネットワークは、 $N = \{A, B, C, D\}$ の 4 ノードと、10 本の片方向リンクを有している。図 1(a) にリンク旅行時間を、図 1(b) にリンク容量を示す。実験においては、ノード A を住居エリア、ノード D をオフィスエリアと考え、ノード B およびノード C は娯楽エリアと考えた。さらにノード B に位置する娯楽施設の営業時間は $b_B = \{3, 4, 5\}$ とした。なお、ノード B 以外の施設については営業時間を限定せず、 $\forall n \in N \setminus B; b_n = T$ とした。

4.1.2 エージェントの条件設定

エージェント $i \in I$ の報酬関数については、以下のように設定した。

$$\forall a_{i,t} \in A_i^M : R_i(s_t, a_{i,t}) = 0 \quad (6)$$

$$\forall a_{i,t} \in A_i^S : R_i(s_t, a_{i,t}) = \begin{cases} v_i^n & (t \in b_n) \\ 0, & (\text{otherwise}) \end{cases} \quad (7)$$

つまり、各エージェント i は「移動」行動においては報酬を得ず、各ノード $n \in N$ 上の施設の営業時間内に当該ノードに滞在したときに報酬 v_i^n を得るものとした。

今回の実験ではエージェントとして、滞在型エージェント I_C と通過型エージェント I_P を考えた。ただし、 $I_C \cup I_P = I$ である。観光トリップを想定した滞在型エージェントの目的は、エリア内を周遊し各地点における滞在を楽しむことである。一方で、通勤トリップを想定する通過型エージェントは、出発地と目的地の間を短い時間で移動することを目的としている。

以下、それぞれについて、今回の数値実験で設定したパラメータを示す。

滞在型エージェント：滞在型エージェントはノード A を出発してノード A に戻る周遊トリップを行うものとする。これは、ノード A に宿泊する観光客の観光行動を想定した設定である。滞在型エージェントの主目的は、ノード B に位置する施設を訪問することとする。すなわち、全ての滞在型エージェントは、トリップがキャンセルされた場合を除き、ノード B の施設の営業時間 $b_B = \{3, 4, 5\}$ の間に 1 時間ステップ以上ノード B に滞在する必要がある。

滞在型エージェントについては、各ノードで得る報酬を以下の正規分布に従って設定した。

$$\begin{aligned}\hat{v}_i^A &\sim \mathcal{N}(50, 50), & \hat{v}_i^B &\sim \mathcal{N}(200, 200), \\ \hat{v}_i^C &\sim \mathcal{N}(100, 100), & \hat{v}_i^D &\sim \mathcal{N}(50, 50)\end{aligned}\quad (8)$$

ただし、 $\mathcal{N}(\mu, \sigma^2)$ は、平均 μ 、分散 σ^2 の正規分布を表している。この報酬関数によると、滞在型エージェントは、ノード B およびノード C に滞在することで、ノード A やノード D に滞在するよりも大きな報酬を得ることができる。また、滞在型エージェント $i \in I_C$ の活動可能時間帯 $T_i = \{t_i^B, \dots, t_i^E\}$ については、活動開始時刻 t_i^B が一様分布 $\{0, 1, 2\}$ に従い、活動終了時刻 t_i^E については、 $\forall i \in I_C; t_i^E = t_i^B + 6$ と設定した。

通過型エージェント：通過型エージェントについては、ノード A を出発してノード D に向かうエージェントと、逆にノード D を出発してノード A に向かうエージェントの二通りを考えた。それぞれのエージェントの集合を I_P^{AD} および I_P^{DA} で表す。ただし、 $I_P^{AD} \cup I_P^{DA} = I_P$ である。ノード A からノード D へ向かうエージェント $i \in I_P^{AD}$ については、各ノードで得る報酬を、以下の正規分布に従って設定した。

$$\begin{aligned}\hat{v}_i^A &\sim \mathcal{N}(200, 200), & \hat{v}_i^B &= 0, \\ \hat{v}_i^C &= 0, & \hat{v}_i^D &\sim \mathcal{N}(500, 500)\end{aligned}\quad (9)$$

また、逆にノード D からノード A へ向かうエージェント $i \in I_P^{DA}$ については、各ノードで得る報酬を、以下の正規分布に従って設定した。

$$\begin{aligned}\hat{v}_i^A &\sim \mathcal{N}(500, 500), & \hat{v}_i^B &= 0, \\ \hat{v}_i^C &= 0, & \hat{v}_i^D &\sim \mathcal{N}(200, 0)\end{aligned}\quad (10)$$

すなわち、通過型エージェントは早い時刻に目的地に到着することで大きな報酬を得る。通過型エージェント $i \in I_P$ の活動可能時間帯 $T_i = \{t_i^B, \dots, t_i^E\}$ については、ノード D へ向かうエージェント $i \in I_P^{AD}$ の活動開始時刻 t_i^B は一様分布 $\{0, 1, 2\}$ に、また、ノード A へ向かうエージェント $i \in I_P^{DA}$ の活動開始時刻 t_i^B は一様分布 $\{2, 3, 4\}$ に従うものとし、活動終了時刻 t_i^E については、 $\forall i \in I_P; t_i^E = t_i^B + 4$ と設定した。

なお、今回の実験においては、全てのエージェントについて、同一リンクの同方向への移動は活動時間帯の間に一度しか実行しないものとした。

エージェント間の干渉と交通管制：上述した設定では、滞在型エージェントは、ノード A から出発するときとノード A へ戻ってくるときに、それぞれノード A からノード D へ向かう通過型エージェントおよびノード D からノード A へ向かう通過型エージェントとの間で交通リソースの競合が発生するため、適切な交通管制が必要となる。本実験では、交通管理者は、本節に示したエージェントの需要に関する統計モデルを事前情報として有しているものとした。ただし、各エージェントの出発時刻および報酬に関する確定値については、通知を受けるまで未知とする。

4.2 実験結果

本研究で示した二つのメカニズムは、共に、逐次的に総社会効用を最大化する割り当てを実現する。その効率については、既往研究 [Hayakawa 18b, 早川 18] の通りである。本研究では、二つの課金メカニズムによって実現する収益の違いに着目して、実験を行った。実験では、通過型エージェントの数を

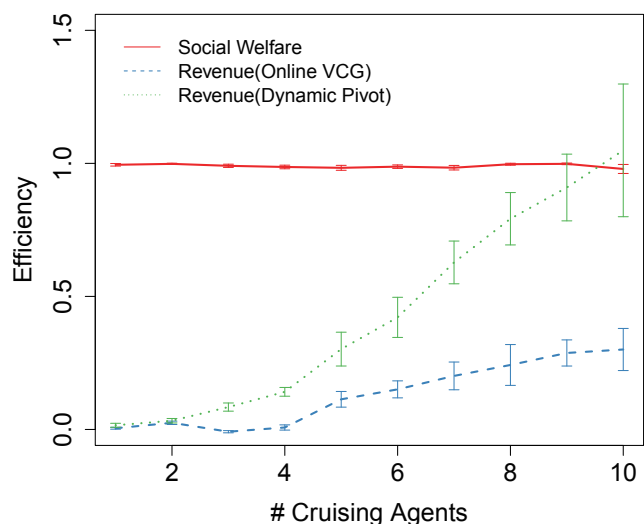


図 2: Revenue achieved by two pricing algorithms

2 とし、滞在型エージェントの数を 1 から 10 まで変化させた。乱数を変化させながら、各条件における計算時間に応じて 10 回から 100 回試行した。

まず、二つのメカニズムの収益を図 2 に示す。図の縦軸は、交通資源配分の効率性を表し、将来の全エージェントの情報が既知であった場合の最適交通資源配分で得られる総社会効用を 1.0 として正規化された値である。図の赤実線は、両メカニズムで得られる総社会効用を示している。一方で、緑点線および青破線は、Dynamic pivot mechanism および Online VCG mechanism で得られるサービスオペレータの収益を表している。全利用者の効用の合計は、総社会効用とサービスオペレータの収益の差としてみることができる。この図によると、Dynamic pivot メカニズムでは、利用者の干渉が大きい時に利用者の効用の合計値が負の値となっている。一方で、Online VCG mechanism は、利用者の干渉が比較的小さい一定条件下において、サービスオペレータの収益が負となっていることが分かる。

5. まとめ

オンデマンド型の交通サービスを想定し、利用者の希望の異質性を活用してより良い交通資源配分を実現する割り当ておよび課金のメカニズムを提案した。今後は、事業者利益とサービスクオリティを両立させるメカニズム [Hayakawa 18a] の確立を目指す。

参考文献

- [Bergemann 10] Bergemann, D. and Välimäki, J.: The dynamic pivot mechanism, *Econometrica*, Vol. 78, No. 2, pp. 771–789 (2010)
- [Hägerstrand 70] Hägerstrand, T.: What about people in regional science?, *Papers in regional science*, Vol. 24, No. 1, pp. 7–24 (1970)
- [Hayakawa 18a] Hayakawa, K., Gerding, E. H., Stein, S., and Shiga, T.: Price-based Online Mechanisms for Settings with Uncertain Future Procurement Costs and

Multi-Unit Demand, in *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 309–317 (2018)

[Hayakawa 18b] Hayakawa, K. and Hato, E.: Dynamic traffic resources allocation under elastic demand of users with space-time prism constraints, *arXiv preprint arXiv:1806.10719* (2018)

[Kitamura 96] Kitamura, R., Pas, E. I., Lula, C. V., Lawton, T. K., and Benson, P. E.: The sequenced activity mobility simulator (SAMS): an integrated approach to modeling transportation, land use and air quality, *Transportation*, Vol. 23, No. 3, pp. 267–291 (1996)

[Minato 93] Minato, S.-i.: Zero-suppressed BDDs for set manipulation in combinatorial problems, in *Proceedings of the 30th international Design Automation Conference*, pp. 272–277 ACM (1993)

[Parkes 04] Parkes, D. C. and Singh, S. P.: An MDP-Based Approach to Online Mechanism Design, in *Advances in Neural Information Processing Systems 16 (NIPS)*, pp. 791–798 (2004)

[Rust 94] Rust, J.: Structural estimation of Markov decision processes, *Handbook of econometrics*, Vol. 4, pp. 3081–3143 (1994)

[Sheffi 85] Sheffi, Y.: Urban transportation networks: equilibrium analysis with mathematical programming methods, *Traffic engineering control. Prentice-Hall, ISBN 0-13-93-972* (1985)

[早川 18] 早川敬一郎, 羽藤英二: グラフ列挙手法により容量制約付き総社会効用最大配分を実現する動的アルゴリズム, 土木計画学研究・講演集, Vol. 57, (2018)

集団避難行動モデルを組み込んだ マルチエージェントシミュレーションツール

Multi-agent simulation tool incorporating group evacuation behavior model

石田龍星 秋吉政徳
Ryusei Ishida Masanori Akiyoshi

神奈川大学工学部情報システム創成学科
Department of Information Systems Creation, Faculty of Engineering, Kanagawa University

Evacuation simulation by multi agent model is used to make plans for reducing damages at disaster. In this research, we specially focused on emergent phenomena that aims to provide notices for unexpected evacuation behaviors by expressing embedded individual behavior rules concerning interactions in groups. To construct such simulation tool, we introduced three models as follows; 2-dimensional continuous planes as spacial model, evaluation agents with eyesight and individual behavior rules, and potential models as deciding moving target positions based on mutual behaviors. In the experimental studies, we set two types of agents, that is, agents with notice of evacuation exits and agents without it. Simulation results shows significance of agents with notice of evacuation exits and effects of group sizes as to evacuation behaviors.

1. はじめに

災害時には迅速かつ安全に避難することが重要であり、避難安全計画や建物設計の支援のために、避難シミュレーションが用いられる [山下 2014]。避難シミュレーションでは、エージェント・ベース・シミュレーションが主流となっており、マルチエージェントは簡単なルールからミクロ・マクロのループにより明示されていなかった新たな行動や現象が起こる創発現象を発生させることができる。

本稿では、避難者間の相互作用によって創発される避難群衆行動の発生を目的とし、シミュレータ使用者にとって予期しなかった避難行動への気づきを与える集団避難行動シミュレーションモデルを提案する。

2. 提案モデル

2.1 空間モデル

避難シミュレーション空間を構築する空間モデルにはネットワークモデルやセル空間モデルなどが考えられるが、避難エージェントの行動空間は、行動をより精緻に表現するために2次元連続平面での空間としている。ただし、障害物となる壁や建物は、複雑な形状であっても避難エージェントとの距離を計算するために2次元のグリッド状の形状で構成している。

2.2 避難者モデル

避難者モデルには、エージェント・ベース・モデリングとして、マルチエージェントモデルを基本とする。エージェントの行動には、外部の環境を認識し、意思決定行動ルールにより行動を行うとした。環境認識には視野を実装し、障害物や出口、他の避難者を認識する。次の動作を決定する意思決定行動ルールは、if-then形式により表現する。そして、避難行動として他の避難者を避けたり、移動経路上の障害物を迂回するような行動を表現するためにポテンシャルモデルを用いる。

2.2.1 視野の設計

避難エージェントが、他の避難者や出口を認識するために視野を実装する。本シミュレーションで実装した避難エージェン

連絡先: 石田龍星, 神奈川大学, 神奈川県横浜市神奈川区六角橋3丁目27-1, 045-481-5661, r201504099sb@jindai.jp

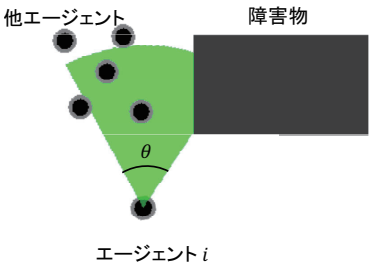


図 1: 避難エージェントの視野

表 1: 避難エージェントの意思決定行動ルール

避難者の種類	避難行動ルール
出口を知っている避難者	1. 出口を発見したら向かう 2. 知っている出口に向かう
出口を知らない避難者	1. 出口を発見したら向かう 2. 集団を追従 3. 集団を一時的に追従 4. ランダムに移動

トの視野機能を図1に示す。視野角と視野の最大距離により視野機能を構成し、避難エージェントは視野内に入ったものを認識することができる。

2.2.2 意思決定行動ルールの設計

避難エージェントとしては、出口を知っている避難エージェントと出口を知らない避難エージェントより構成する。避難エージェントの行動ルールを表1に示す。

出口を知っている避難エージェントは、避難空間における脱出口を知っているため、その出口方向に移動する。しかし、移動中に他の出口を発見した場合には、その出口に移動するといった行動をとるルールとした。出口を知らない避難エージェントは、基本としていかなるルールにより行動していたとしても出口を発見することがあれば、出口を知っている避難エージェントと同様にその出口に向かう行動をとることとした。現

実の避難者の行動は予め定まっていなかったことから、その他の行動ルールは時間ステップごとにランダムに発火することとした。集団を追従するルールでは集団を視野内におけるエージェント数が k 人以上であることと定義し、追従ルールを発火させることとした。集団を発見した場合、視野内における避難エージェントの一人を追従し、集団を発見することができなかった場合には、ランダム行動をとる。なお、集団に追従するルールが発火後は、追従先の避難エージェントが視野内から外れることがない限り、時間ステップごとにランダムに発火せず、ルールを固定する。

2.3 ポテンシャルモデル

避難シミュレーションの行動モデルや、移動ロボットの障害物回避モデルとしてポテンシャルモデルがある。他の行動モデルとして、目標点までの引力と障害物への斥力とした運動方程式で局所的な働きを表現する Helbing らの Social Force Model などが考えられる [Helbing 1995]。しかし、目標地点までの経路上に障害物がある場合に、障害物の形によっては移動中に立ち往生する問題がある。空間モデルをネットワークモデルとし、エージェント行動に制約を与えることや、柳沢らの提案する移動経路にサブゴールの経由点を決めることによりこの問題を解決できるが、エージェントの行動ルールが複雑になってしまう可能性がある [柳沢 2006]。

そこで、ポテンシャルモデルはサブゴールを設定することなく、目標地点まで経路設計を行うことができるため、ポテンシャルモデルを用いる。目黒らはポテンシャル場から避難エージェントの進行方向を決定する避難行動シミュレーション手法を提案しているが、シミュレーション空間はグリッド上のセル空間によって構成されており、進行方向はムーア近傍の 8 方向にしか移動することができない [目黒 1995]。本シミュレーションモデルでは、避難行動をより精緻に扱うため空間モデルを二次元連続平面での空間で構成しているため、ポテンシャル関数の最急降下法によって求められるポテンシャルベクトルと、避難エージェントの移動ベクトルを合成することにより進行方向を決める。

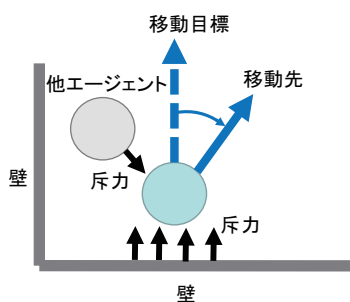


図 2: ポテンシャルモデルによる進行方向の変更

ポテンシャル関数には島倉らを参考に D.H.Kim らによって提案されたポテンシャル関数を用いる [島倉 2009][Kim 2006]。斥力ポテンシャル U_o は次式により定義される。

$$U_o(x, y) = \sum_{i=1}^n c_o \exp\left\{-\frac{(x+x_{oi})^2 + (y+y_{oi})^2}{l_o^2}\right\} \quad (1)$$

ここで、 x_{oi} , y_{oi} は、斥力となる障害物の座標、 c_o はポテンシャル関数の影響度、 l_o はポテンシャルの山の大きさを調整するパラメータとして用いる。ポテンシャルベクトルには、最急

降下法をも用いることによって、避難エージェントの移動方向ベクトルと合成する斥力ベクトルを決める。最急降下法 d は、次式によって定義される。

$$d = -\text{grad}U_o. \quad (2)$$

3. 群衆移動実験

3.1 群衆実験の条件

閉鎖空間内で 1 つの出口に向かう群衆実験 A と、2 つの出口に同時に向かう群衆実験 B を行った。シミュレーションを行う空間は以下の図 3 に示すように縦幅 15m、横幅 10m の空間に幅 1m の出口を二つとした。実験は群衆実験 A、群衆実験 B ともに避難エージェント数を 30 人、40 人、50 人の条件で行う。

群衆実験 A はすべての避難エージェントが右の出口を知っている状態で行動をする。群衆実験 B は避難エージェントの半分が左の出口を知っており、残る半分が右の出口を知っている状態で行動を開始する。群衆実験は両実験ともにエージェントの初期位置が上部 3m 以内に収まる範囲でランダムに配置している。また、避難エージェントの視野角は 60 度、視野の最大距離を 4m とし、歩行スピードは 1.5m/s としている。

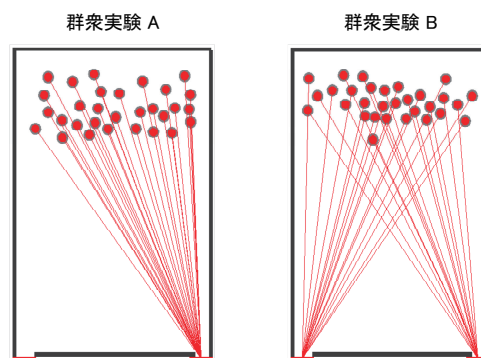


図 3: シミュレーション空間

3.2 群衆実験 A の結果

群衆実験 A における避難エージェントが脱出した出口について以下の表 3 に示す。エージェント数 30 人ではすべてのエージェントが自分の知っている出口へ脱出しているが、エージェント数 40 人、エージェント数 50 人では、複数のエージェントが自らの知ってる出口ではない出口で脱出している。この行動は出口を知っているエージェントの意思決定行動ルールにおける出口が視野に入った場合はその出口に向かうルールが発火しているためだと考えられる。

避難エージェントが移動した軌跡を図 4 に示す。エージェント数 30 人ではポテンシャルモデルにおける斥力ポテンシャルの影響が少なく、左の出口が視野に入ることは無かったが、初期位置が比較的左の位置で移動を開始した避難エージェントは左の出口を発見している。このように、エージェント数によりポテンシャルの影響が異なり、他の脱出口を発見する創発現象を軌跡で確認することができる。

3.3 群衆実験 B の結果

群衆実験 B の実験結果を表 3 に示す。3 つのすべてのシミュレーションで、他のゴールで脱出する避難エージェントが多くいることがわかる。エージェント数が 30 人、50 人のシミュレーションでは、3 割程度が他のゴールで脱出している。

表 2: 実験結果 A

エージェント数	知っている出口で脱出	他の出口で脱出
30	30	0
40	37	3
50	45	5

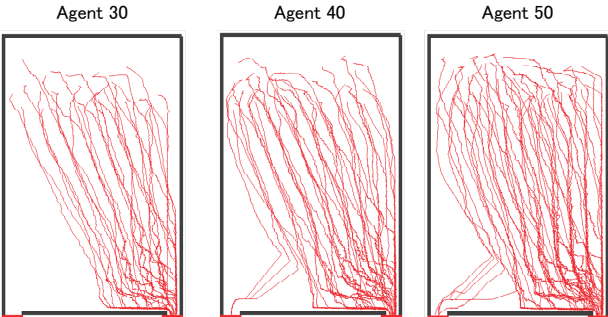


図 4: エージェントの行動軌跡

群衆実験 B における避難エージェントの移動軌跡を図 5 に示す。群衆実験 A の移動軌跡と比べると、移動途中で出口を発見し、その出口へ向かったような軌跡ではないことがわかる。異なる出口を目指すエージェント集団に囲まれていた場合は、そのまま集団の流れにより、他の出口で脱出している。

3.4 群衆実験の考察

群衆実験 A・B より、初期位置によって群衆の流れに逆らうような動きをすることができず、結果として自らの知らない出口を発見したり、他の避難群衆に導かれるようにして他の出口を発見するような創発現象を確認することができた。この結果について、現実でも人の過度な密集状態での移動では、自分の思う方向に進むことが難しいことが表現できている。通勤ラッシュ時の駅のホームのような狭い通路などで人が密集している状態では、群衆として一方へ移動の流れができている場合、逆方向への移動はできずその流れに巻き込まれてしまうことがよくある現象を本実験においても確認することができた。

4. 集団行動ルールを組み込んだ避難行動実験

4.1 シミュレーション条件

図 6 に示すような単純な長方形の空間を用意し、シミュレーションを行う。シミュレーション条件として、出口を知らない避難者が 30 人の条件と、出口を知っている避難者が 10 人、出口を知らない避難エージェントが 20 人の合計 30 人の避難エージェントを空間に配置する。初期位置としてランダム分布 (R)、または中央分布 (C) に分ける。空間は出口が空間の角に 1 つ (A)、または 2 つ (B) としてシミュレーションを行う。なお、60step で 1 秒としており、全てのエージェントが脱出できなくても 60 秒 (3600step) でシミュレーションを終了することと

表 3: 実験結果 B

エージェント数	知っている出口で脱出	他の出口で脱出
30	20	10
40	32	8
50	35	15

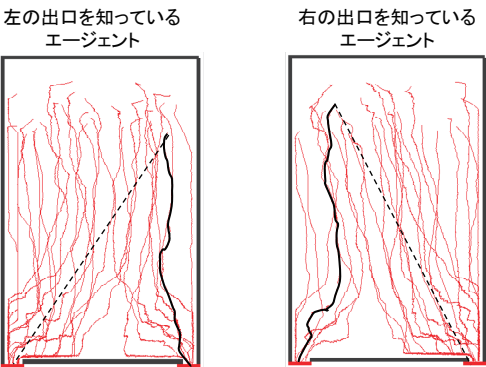


図 5: エージェント数 50 人の移動軌跡

した。また、避難エージェントの視野角は 60 度、視野の最大距離を 4m とし、歩行スピードは 1.5m/s としている。

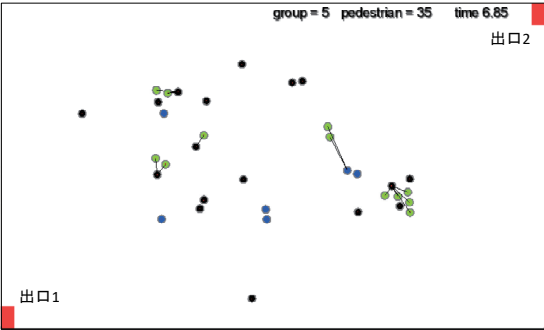


図 6: シミュレーション空間

4.2 シミュレーション結果

条件ごとに避難シミュレーションを 10 回行った。同じ条件で 10 回のシミュレーションを行ったが、ランダム行動というルールが避難エージェントに実装されているため、揺らぎの影響により、避難エージェントの脱出人数がシミュレーションごとに大きく変わることもあった。それぞれのシミュレーション条件で脱出人数が最大だったシミュレーションのステップ数における脱出人数のグラフを図 7 に示す。

出口を知っている避難エージェントが 10 人のシミュレーション条件では 500step から 1500step までの間で多くの避難エージェントが出口から脱出していることがわかる。このステップ間で 20 人以上の避難エージェントが脱出していることから、出口を知っている避難者は全員脱出し、出口を知らない避難エージェントも 10 人以上が脱出していることがわかる。全ての避難エージェントが出口を知らない条件でのシミュレーション結果は、出口数が 2 つの BC0 と BR0 の合計脱出人数が 15 人、14 人となっていることから、避難エージェントの初期位置で脱出人数の変化はあまりなかったと言える。また、出口を知っている避難者のいる条件に比べ、段階的に脱出していることがわかる。また、AC0 と AR0 における脱出人数は 3 人と 5 人の避難エージェントしか脱出することができていない。10 回のシミュレーションの中でも全てのエージェントが脱出できていない場合も何度かあった。

シミュレーション中で追従ルールが発火している避難エージェント数と全体の集団数を表すグラフを図 8、図 9 に示す。出口を知っている避難エージェントのいる条件では、1step 目

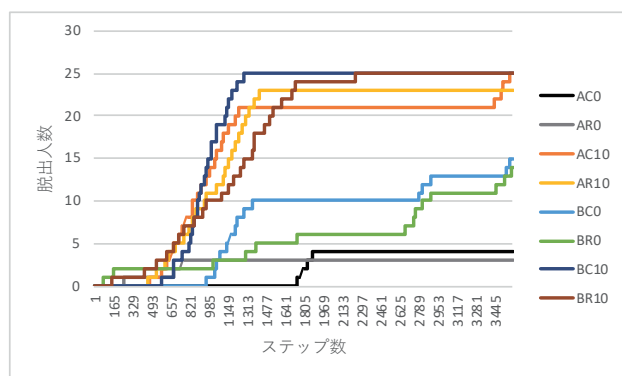


図 7: 脱出人数

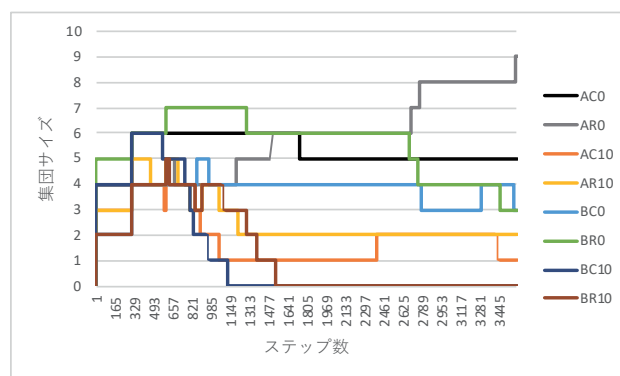


図 9: 集団サイズ

から 10 人近くの出口を知らない避難エージェントの追従ルールが発火し、600step あたりから減少しているため、出口を知っている避難エージェントを追従することで出口から脱出していることがわかる。AC0 のシミュレーションでは 20 人近くの多数の追従が発生しているが、追従先の避難エージェントがランダム行動で出口を見つけることができなかったため、ほとんどが脱出することができなかった。集団サイズは、AR0 のシミュレーションでは集団サイズが 9 つになっている。追従ルールでは、5 人以上の避難エージェントを視野内で見たときに追従するルール設計をしているため、視野内における避難エージェントが集団でないことがあるため、AR0 のような多く追従集団が形成された。

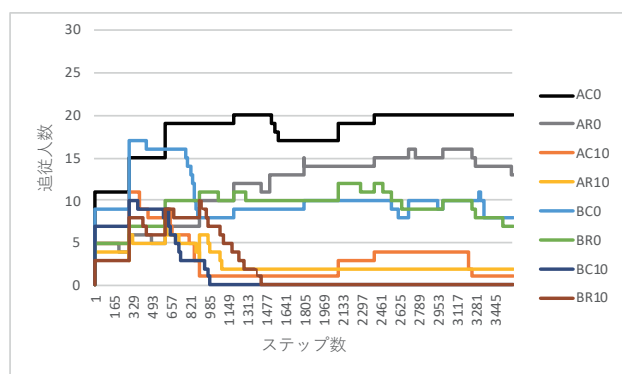


図 8: 追従人数

4.3 考察

出口を知らないエージェントが出口を知るエージェントに追従し、群れが形成されることによって、追従者が増加し、脱出人数も増加するが、出口を知らない避難者を追従してしまった場合には、多くの避難エージェントが脱出することができない場合があることがわかる。現在の追従ルールでは、視野から追従先の避難エージェントが外れない限り、追従をやめることのないルール設計とし、集団サイズの制限はしていないため、追従先が出口の知らない避難者であったとしても追従を続けているためほとんどの避難エージェントが出口から脱出できないような結果になった。このことは、避難行動における出口に向かう避難行動者の重要性や、適切な集団サイズを促す必要性を示唆している。

5. まとめ

本稿では、マルチエージェントモデルを用いて、避難者行動の創発現象を表現することで、シミュレータ使用者に気づきを与える集団避難行動シミュレーションモデルを提案した。ポテンシャルモデルによるシミュレーション実験により、創発現象を確認することができた。また、集団行動ルールを組み込んだ避難エージェントがいるシミュレーション実験では、避難エージェントの追従ルールによる群れが形成されるが、追従先の避難エージェントが出口を知っているか否かにより、脱出人数が変化した。今後は、障害物のある空間でのシミュレーションやエージェント内部のダイナミクスの変化を起こす機能を持ったエージェントモデルを構築し、創発現象の発生条件について研究を進める予定である。

参考文献

- [山下 2014] 山下倫央, 野田五十樹, “避難シミュレーションの実社会への応用”, 情報処理, pp 572-578, 2014
- [Helbing 1995] Helbing, D. and Molnar, “Social force model for pedestrian dynamics”, Physical Review E, Vol. 51, pp. 4282-4286, 1995
- [柳沢 2006] 柳沢豊, 山田辰美, 平田圭二, 佐藤哲司, “視線に基づくサブゴール決定過程を取り入れた歩行者モデル”, 情報処理学会論文誌 47(7), pp. 2160-2167, 2006
- [目黒 1995] 横山秀史, 目黒公郎, 片山恒雄, “避難行動解析へのポテンシャルモデルの応用”, 土木学会論文集 1995(513), pp. 225-232, 1995
- [島倉 2009] 島倉諭, 項警宇, 稲垣伸吉, 鈴木達也, “ポテンシャルを組み込んだパーティクルコントロールによる自律移動ロボットの障害物回避制御”, ロボティクス・メカトロニクス講演会講演概要集 2009(0), pp. 2A1-G05, 2009
- [Kim 2006] D. H. Kim, S. Shin, “Local path planning using a new artificial potential function configuration and its analytical design guidelines”, Advanced Robotics 20, pp. 111-135, 2006

耳音響認証における観測ゆらぎ軽減手法の一検討

A study of observation fluctuation reduction method for ear acoustic authentication

安原 雅貴^{*1}

Masaki Yasuhara

荒川 隆行^{*2}

Takayuki Arakawa

越仲 孝文^{*2}

Takafumi Koshinaka

矢野 昌平^{*1}

Shohei Yano

^{*1}長岡工業高等専門学校

NIT, Nagaoka College

^{*2}日本電気株式会社

NEC Corporation

Ear acoustic authentication is a type of biometric authentication that uses the acoustic characteristics of the ear canal. A special earphone which has a driver and a microphone is used in this system. The measurement data has error due to the attaching and detaching of earphone each time. In our previous study, we proposed a special earphone which has a driver and two microphones. And multiple features are concatenated. However, the accuracy didn't improve. It's conceivable that the concatenation method is not good. In this study, we conceive that multiple features obtained by multiple microphones in one measurement complement observation fluctuation. As a result, the accuracy was improved. By analysis of variance, we can say that the features were interpolated, and the learning data increased by using multiple microphones. We conclude that the method using the earphone which has multiple microphones is effective.

1. はじめに

指紋や顔による生体認証は普及が進んでいる。多くの場合、認証はサービス開始時に行われる。そのため途中でユーザが入替わる「なりすまし検知」が困難である。我々は、新たな生体情報として耳穴（外耳道）に着目し、耳音響認証を提案した。耳音響認証はマイクロホン内蔵イヤホンにより外耳道の音響伝達特性を取得し、特徴量として用いる。イヤホンの装着具合等により特徴量に差異を生じる。これを観測ゆらぎと呼ぶ。我々は過去に、2つのマイクロホンを90degに配置したイヤホンを作成し、角度の異なるマイクロホンからの特徴量を利用することで、特徴量の角度データを補間し、観測ゆらぎ等に対して強靱な認証になると考え検討を行った。しかし、画期的な精度向上は得られなかった。特徴量の扱い方と角度の補間に問題があると考えられる。本研究では、さらに角度データを補間するため、90degずつに4つのマイクロホンを内蔵したイヤホンを作成し、各マイクロホンからの特徴量の取り扱い方法と補間方法とを検討することで、観測ゆらぎ軽減により精度向上の可能性を実験的に検討する。また、観測ゆらぎの要因について、特徴量の分散の観点から考察を行う。

2. 実験方法

耳音響特性は図1(a)に示す測定システムを用いて、MLSインパルス応答測定法により測定する。実験は10~40歳の男女38人に対し、左耳測定、4ch、10回/人、毎回着脱をして測定を行う。90degずつにマイクロホンを配置できるように図1(b)のように設計し、3Dプリンタで筐体出力する。測定した耳音響特性に対し、最小位相系を用いてインパルス応答が先頭になるように処理する。このとき、信号長は16384次元(0.3715秒)を先頭から256次元で切り出す次元削減処理をしたものを特徴量として用いる。特徴量に対して、SVM(linear kernel, $C = 100$, $\gamma = 0.001$)を用いて識別精度を算出する。識別性能指標として等誤り率(equal error rate: EER)を用いる。EERは、誤って他人を本人と識別した確率(fail acceptance rate

: FAR)と誤って本人を他人と識別した確率(fail recognition rate: FRR)とが等しくなる値であり低いほど良い性能であることを示す。

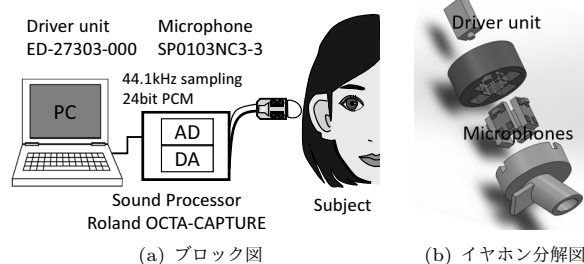


図 1: 耳音響特性の測定システム

本研究では、4ch分の特徴量の取り扱いとして次の4手法を用いる。手法①として各chを個別に用いる手法。手法②として4ch分の特徴量を時間軸で連結し用いる手法。さらに、本報告において提案する手法③および手法④でEERを算出し比較検討する。手法①においては、図2のようにデータセットを分けてEERを算出する。図2では、Sub0を本人、Sub1を学習しない他人、その他を学習する他人として分ける場合を示している。各ユーザ10個の耳音響特性のうち、本人データとして5個を学習、残り5個をテストとして用いる。手法②では4ch分の特徴量を時間軸で連結させたものを特徴量を用いる。手法③では図3のようにデータセットを分けて、4ch分の特徴量を同一のラベルとして取扱う。図3において、テストのデータ数は4ch分を同一ラベルとして取扱うためSub0で20個、Sub1で40個の合計60個となる。手法①とデータ数を合わせるため、Sub0より5個、Sub1より5個を無作為に抽出しEERを算出する。Sub0~Sub37を変更した全組合せでEERを算出しその平均EERを導出する。手法④では、ch間のデータを足し合わせ2で割る補間処理により補間データを作成する。補間処理は45deg毎に設置された8個のマイクロホンからの耳音響特性を得ることを模擬している。補間処理で得られたデータは学習に用い、テストデータには用いない。他手

連絡先: 長岡高専電気電子システム工学科音響工学研究室

E-mail: syano@nagaoka-ct.ac.jp

法と同様に EER を算出し平均 EER を導出する。手法①～④において学習データとテストデータとは重複はない。それ以外の特徴量は提案手法①と同様に学習させ、テストデータには本人 5 個、他人 5 個で無作為に抽出を行ったものを用いる。

さらに手法①, ③, ④において、特徴量の周波数振幅特性から、周波数に対するクラス内分散 V_W とクラス間分散 V_B を導出し、耳道音響特性に含まれる観測ゆらぎの要因について検討を行った。

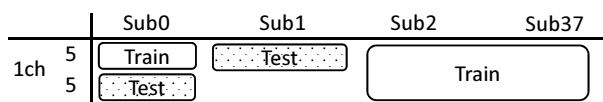


図 2: 手法①の学習データとテストデータの分割

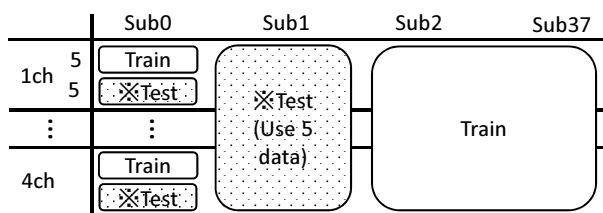


図 3: 提案手法③の学習データとテストデータの分割

3. 実験結果

図 4 に、手法①～④の平均 EER を示す。本報告での提案手法である手法③, ④が、他手法の EER より 2~3%ほど低下し精度が向上している。また、両耳の特徴量を用いた際に識別精度向上に有効であった手法②よりも、精度が向上している。これより提案した手法③, ④は有効であると考えられる。図 5, 6 に手法①, ③, ④の各 ch 特徴量のクラス内分散, クラス間分散と周波数との関係を示す。図 5, 6 において、手法④は手法③の特徴量の線形的な拡張であり同じ分散を示す。クラス間分散は複数のチャンネルを同一ラベルとして扱っても各チャンネルと同じ分散を示している。クラス内分散は提案手法のときに大きくなっており、各チャンネルで得られる特徴量に差異があることがわかった。この結果より、マイクロホンの角度を変えたことで得られる特徴量に変化することが明らかになった。クラス内分散が手法③, ④では悪くなっていることから、この手法では SVM への学習の質が悪くなっているといえる。しかし、識別精度は向上している。これは、角度データが多くなったことで観測ゆらぎを補間したことによるものと、各チャンネルで特徴量を扱うよりも手法③, ④の方が学習データが多いことによるものと 2 つが考えられる。提案した手法③と手法④を比較すると、手法④の方が精度が向上しているが、0.2%ほどである。このことから、角度データが多くなったことで観測ゆらぎを補間できることが示唆される。また、今回の実験で手法③, ④の精度が向上したのは、学習データ数が多くなったことに大きく寄与していると考えられる。また、図 5 より、マイクロホンの違いによる周波数特性の違いは低周波に現れることがわかった。

4. まとめ

耳音響認証における測定時の観測ゆらぎの軽減手法を検討した。90 deg ずつに配置したマイクロホンを 4 つ内蔵したイ

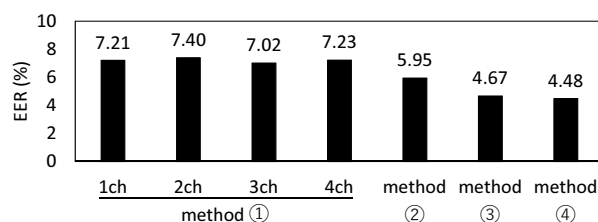


図 4: 特徴量の扱い方と精度の関係

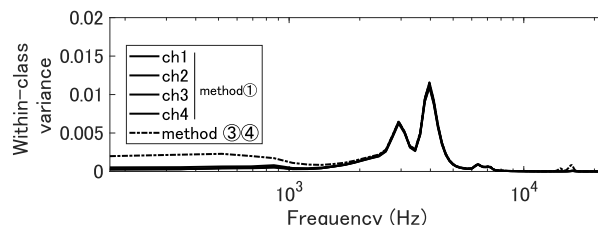


図 5: 周波数軸上の特徴量のクラス内分散

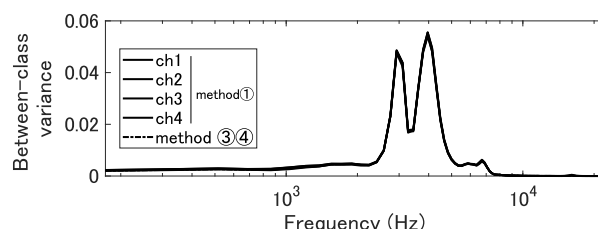


図 6: 周波数軸上の特徴量のクラス間分散

ヤホンを用いることで、異なる特徴量を得られることがわかった。異なる特徴量を同一に学習させると、精度が悪くなりそうだが、今回の実験では精度は良くなった。角度データが多くなったことで観測ゆらぎを補間したことによるものと、各チャンネルで特徴量を扱う手法①よりも提案した手法③, ④の方が学習データが多いことによるものとが考えられる。複数のマイクロホンを内蔵したイヤホンを用いる手法は有効であることがわかった。加えて、既成品のイヤホンにマイクを 1 つ取り付けた場合の EER は 2%なので、イヤホンの設計が重要であることがわかった。4 ch イヤホンを作成した小林毅範氏に感謝する。本研究は JSPS 科研費 JP16K00182 の助成を受けたものです。

参考文献

- [Yano 17] 矢野昌平, 荒川隆行, 越仲孝文, 今岡仁, 入澤英毅, “誤差の周波数拡散と加算平均処理による耳音紋認証の精度向上,” 信学論 (A), Vol.J-100A, no.4, pp.161-168, Apr.2017.
- [Kobayashi 17] 小林毅範, 矢野昌平, 荒川隆行, 越仲孝文, “複数マイクロホンを用いた耳音響認証の認証精度向上の検討”, 信学技報, P.164, 2017.
- [Yasuhara 18] 安原雅貴, 矢野昌平, 荒川隆行, 越仲孝文, “両耳での耳音響認証の特徴量設計”, 信学技報, P.77, 2018.

New similarity scale to recognize bird calls and abnormal sounds of concrete/machine
Development of pattern matching software using multi-CPU

Michihiro Jinnai ^{*1}

Edward James Pedersen ^{*2}

^{*1} Nagoya Women’s University

^{*2} Central Queensland University, Australia

A new similarity scale called the Geometric Distance, that numerically evaluates the degree of likeness between the standard pattern and the input pattern is proposed. Traditionally, the similarity scales known as the Euclidean distance and cosine similarity have been widely used to measure likeness. Traditional methods do not perform well in the presence of noise or pattern distortions. In this paper, a mathematical model for similarity is proposed to overcome these limitations of the earlier models, and a new algorithm based on a one-to-many point mapping is proposed to realize the mathematical model. Using the new similarity scale, experiments in bird call recognition were carried out in noisy environments. Furthermore, experiments in abnormal sound recognition of concrete structure were carried out. In all cases a significant improvement in recognition accuracy is demonstrated.

1. Introduction

Human beings, dogs, cats, and other such mammals exhibit a “sense of similarity” in their perception of sounds and sighted objects. To emulate this sense of similarity algorithmically in a “similarity scale” is an important objective for developing computer intelligence.

In an acoustic similarity scale, the degree of likeness between an acoustic standard pattern (control) and an undetermined input pattern is evaluated as a “distance” between the two patterns. This process arbitrarily emulates a human perceiving a sound and comparing that autonomously with ‘templated’ or remembered sounds. Ostensibly, a software-based similarity scale would return a short ‘distance’ for two patterns that humans would consider as similar to, or the same as each other, and a long ‘distance’ for two patterns that humans would consider as dissimilar.

Euclidean distance and cosine similarity are widely used to measure likeness. Conventional similarity scales compare patterns using one-to-one mapping. The result of one-to-one mapping is that the distance metric is highly sensitive to noise, and the distance metric changes in a staircase pattern when a difference occurs between peaks of the standard and input patterns. As an improvement, we have developed a new similarity scale called “Geometric Distance (GD)” [1]. GD is more accurate than the conventional similarity scales in noisy environments.

The GD similarity scale and allied detection software have proven applicable in a broad range of dynamics: from automated call detection of endangered species [2] from within ‘big’ environmental data sets; to real-time fault detection in concrete structures and operating machinery [3]. In this paper, we describe the underlying mathematical model for similarity; the GD algorithm; and we introduce automatic recognition software for bird vocalisations that uses the GD algorithm.

2. The LPC spectrogram of bird sounds

The lower diagram of Fig. 1 shows the waveform of a Noisy Miner Chur call (Paul G. McDonald; University of New England, Australia 2012, *Cooperative bird differentiates between the calls*

of different individuals, even when vocalisations were from completely unfamiliar individuals. *Biology Letters* 8: 365-368).

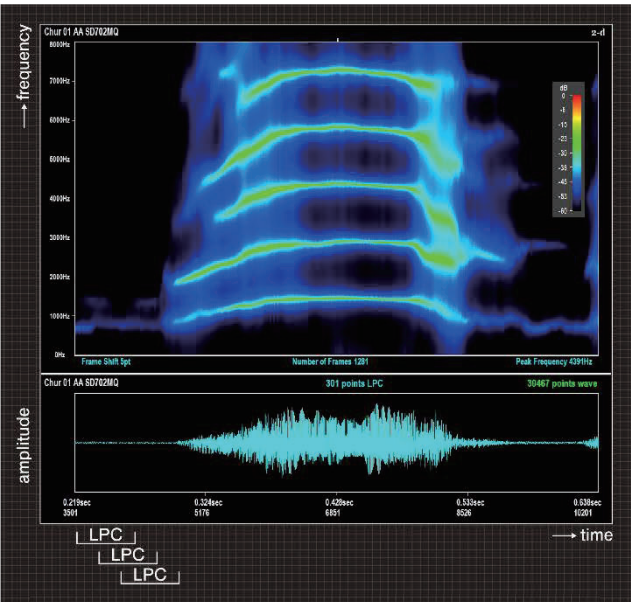


Fig. 1 Spectrogram of bird vocalisation

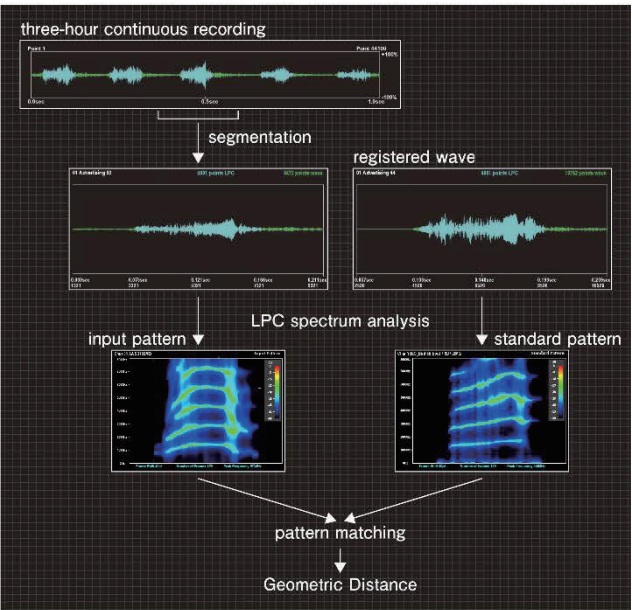


Fig. 2 Processing procedure in recognition system

Contact: Michihiro Jinnai, mjinna@nagoya-wu.ac.jp

The upper diagram of Fig. 1 shows a spectrogram (time-frequency-power) of the exact same signal. In Fig. 1, the waveform has been segmented with 18.8msec frame width and 0.31msec frame period, and the LPC spectrum has been calculated for each frame. Next, the spectrogram has been coloured according to logarithmic power of the LPC spectrum. We have set the software analysis parameters for this bird vocalisation recorded at 16kHz sampling frequency, 16bit quantization, to an LPC order of 12; restricted the spectral frequency range to 0Hz to 8000Hz, with an 11.5Hz frequency resolution; and set a dB threshold filter of 0dB to -60dB logarithmic power spectrum. To analyze transient signals such as bird vocalisations, we set a fractional frame width (with respect to total signal period) as shown at the bottom of Fig. 1. LPC spectrum analysis is suitable for spectral modeling of transient signals.

3. Automatic recognition procedure

Fig. 2 shows the software procedural stages for automated signal detection and recognition. First, the software differentiates potential signals from background noise (segmentation). Second, the software extracts the segmented signals and establishes the segmented signal's spectral characteristics (time-frequency-power) using LPC spectral analysis. Third, the software compares the spectral characteristics of the extracted signal (the input pattern) with a previously registered standard pattern of the focal signal (the signal to be automatically detected). Comparison is effected using the GD similarity scale. To expedite the process, the software executes parallel processing using multiple CPUs [4][5].

4. Mathematical model for similarity

For a functional similarity scale, we need first to develop a mathematical model for similarity, that can perform numerical processing by computation. In the GD process, a mathematical model incorporating the following two characteristics is used:

- < 1 > A distance metric which shows good immunity to noise.
- < 2 > A distance metric which increases monotonically when a difference increases between peaks of the standard and input patterns.

Figs. 3 and 4 graphically demonstrate the underlying computational and algorithmic processes. The upper diagram of Fig. 3 shows an example of the “difference” where the standard pattern has two peaks in the spectrogram, and input patterns 1, 2, and 3 have a different position on the first peak. Note that both the standard and input patterns have the same volume. Fig. 4 shows an example of a “wobble” where the standard pattern has a flat spectrogram. Input patterns 4 and 5 have a “wobble” on the flat spectrogram, and input pattern 6 has a single peak. Each pattern however, is assumed to have variable α in the relationship shown in Fig. 4. Therefore, the standard and input patterns always have the same volume.

Bar graphs at the bottom right of Figs. 3 and 4 express the characteristics < 1 > and < 2 > of the mathematical model diagrammatically.

5. New algorithm for similarity scale

A new algorithm based on one-to-many point mapping is proposed to realize the mathematical model. In the GD algorithm, when a “difference” occurs between peaks of the standard and input patterns with a “wobble” due to noise, the “wobble” is

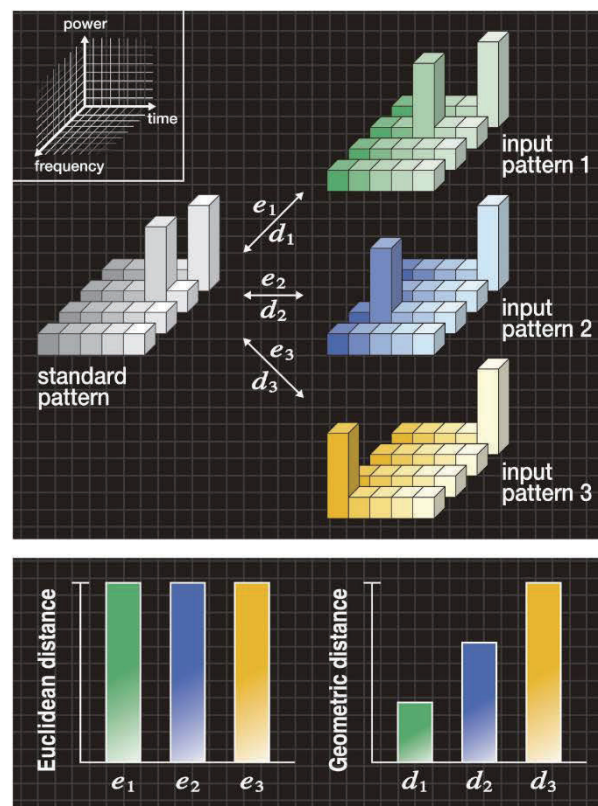


Fig. 3 Typical example of “difference”

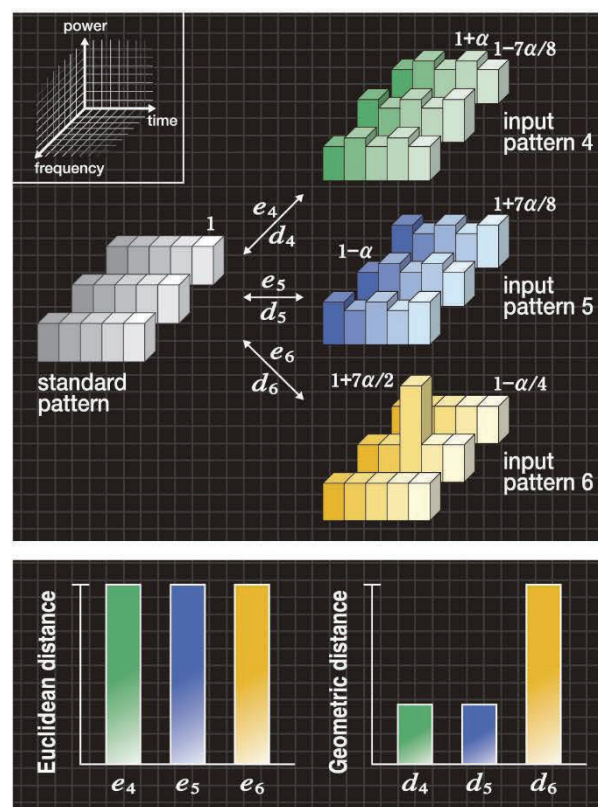


Fig. 4 Typical example of “wobble”

absorbed and the distance metric increases monotonically according to the increase of the “difference”.

In statistical analysis, a normal distribution is usually used as a model for a phenomenon. Then, a “kurtosis” and a “skewness” are used to verify whether the phenomenon obeys the normal distribution or not. Here, the kurtosis ‘ a ’ and the skewness ‘ b ’ are statistics, and we explain them using Figs. 5 and 6. If a probability distribution of the phenomenon follows the normal distribution, then $a=3$ (Fig. 5(b)). If it has flatness relative to the normal distribution, then $a<3$ (Fig. 5(a)). Conversely, if it has peakedness relative to the normal distribution, then $a>3$ (Fig. 5(c)). Also, if a probability distribution of the phenomenon is symmetrical about the mean μ , then $b=0$ (Fig. 6(b)). If the tail on the left side of the probability distribution is longer than the right side, then $b<0$ (Fig. 6(a)). Conversely, if the tail on the right side of the probability distribution is longer than the left side, then $b>0$ (Fig. 6(c)).

In this section, we explain the GD algorithm using Figs. 7 and 8. Fig. 7 shows the spectra (frequency-power) extracted from a Macleay's Fig Parrot (*Cyclopsitta diophthalma macleayana*) vocalisation. Fig. 7 shows standard and input patterns that have been created using the momentary power spectrum (frequency-power) of standard and input sounds. Figs. 8(a)-(e) respectively show typical examples of the standard and input patterns. Note that the power spectrum is generated from the output of a filter bank with m frequency bands. The i -th power spectrum values (where, $i = 1, 2, \dots, m$) of the standard and input sounds are divided by their total energy, so that normalized power spectra s_i and x_i have been calculated, respectively. At this moment, the standard and input patterns have the same area size. Moreover, Figs. 8(a)-(e) respectively show reference patterns that have the initial shape r_i of a normal distribution.

With the GD algorithm, a difference in shapes between standard and input patterns is replaced by the shape change of the reference pattern using the following equation.

$$r_i \leftarrow r_i + (x_i - s_i) \quad (i = 1, 2, 3, \dots, m) \quad (1)$$

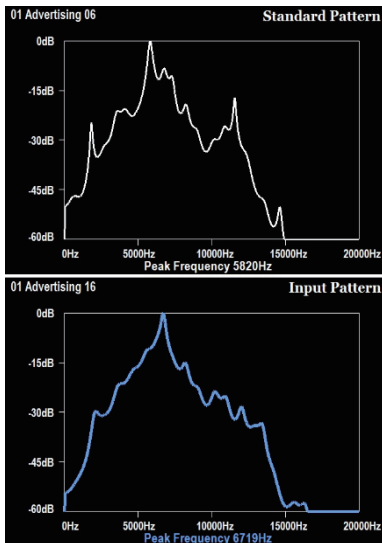


Fig. 7 Standard and input patterns

Next, we explain Eq. (1) using Figs. 8(a)-(e).

- Fig. 8(a) gives an example of the case where the standard and input patterns have the same shape. Because values r_i of Eq. (1) do not change during this time, the reference pattern shown in Fig. 8(a) does not change in the shape from the normal distribution.
- Figs. 8(b)-(d) respectively show examples exhibiting a small, medium, and large “difference” of peaks between the standard and input patterns. If Eq. (1) is represented by the shapes, as shown in Figs. 8(b)-(d), value r_i decreases at peak position i of each standard pattern. At the same time, value r_i increases at peak position i of each input pattern.
- Fig. 8(e) typically shows the standard pattern having a flat shape and the input pattern where a “wobble” occurs in the flat shape. Because values r_i increase and decrease alternatively in Eq. (1)

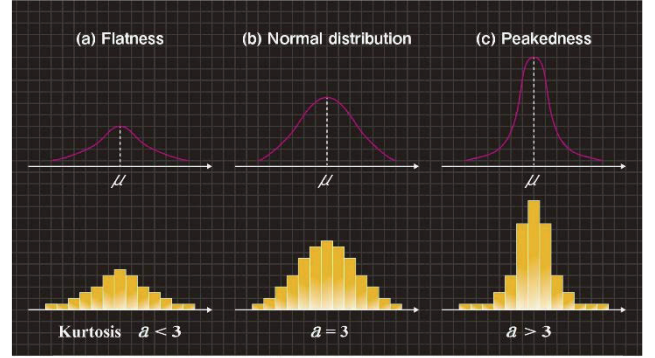


Fig. 5 Shape change and kurtosis value ‘ a ’

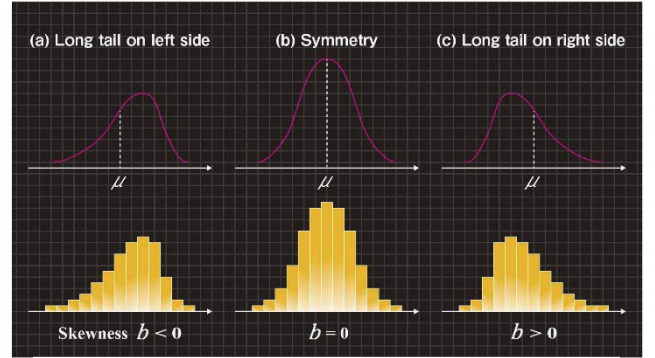


Fig. 6 Shape change and skewness value ‘ b ’

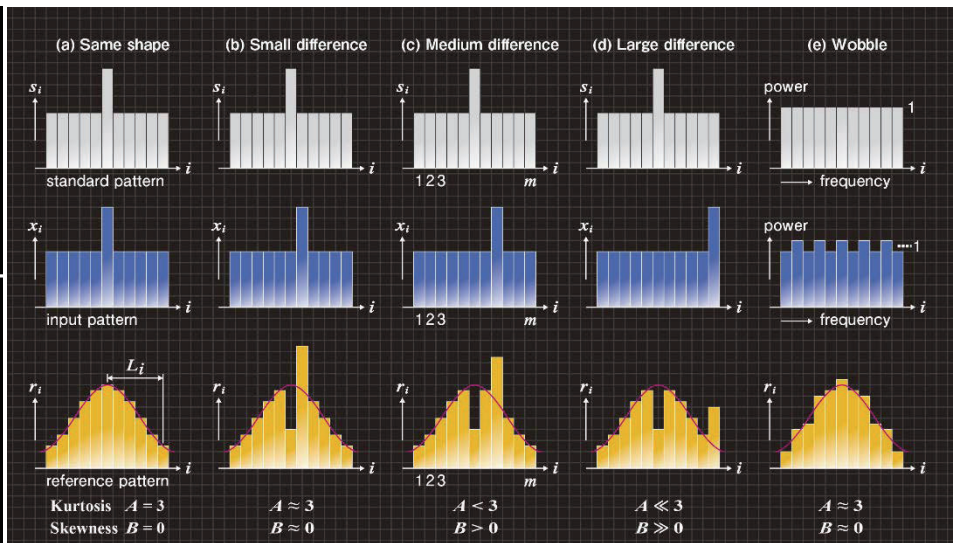


Fig. 8 Shape change of reference patterns

during this time, the reference pattern shown in Fig. 8(e) has a small shape change from the normal distribution.

With the GD algorithm, we replace the mean μ shown in Figs. 5 and 6 with the centre axis of the normal distribution (reference pattern) shown in Fig. 8(a). Then, we replace the kurtosis ' a ' and the skewness ' b ' with a kurtosis ' A ' and a skewness ' B ' shown in the following equations.

$$A = \frac{\left\{ \sum_{i=1}^m r_i \right\} \cdot \left\{ \sum_{i=1}^m (L_i)^4 \cdot r_i \right\}}{\left\{ \sum_{i=1}^m (L_i)^2 \cdot r_i \right\}^2} \quad B = \frac{\left\{ \sqrt{\sum_{i=1}^m r_i} \right\} \cdot \left\{ \sum_{i=1}^m (L_i)^3 \cdot r_i \right\}}{\left\{ \sum_{i=1}^m (L_i)^2 \cdot r_i \right\}^{\frac{3}{2}}} \quad (2)$$

Where, L_i ($i = 1, 2, \dots, m$) is a deviation from the centre axis of the normal distribution as shown in the reference pattern of Fig. 8(a). Then, numerical experiments were carried out to study the relationships between the kurtosis ' a ' and the kurtosis ' A ' or between the skewness ' b ' and the skewness ' B '. As a result of the experiments, we have confirmed that they have the same characteristics [1][3].

For the reference pattern whose shape has changed by Eq. (1), the magnitude of shape change is numerically evaluated as the variable of kurtosis A and skewness B . The kurtosis and the skewness of the reference pattern can be calculated using Eq. (2). Figs. 8(a)-(e) show how A and B vary with r_i .

- In Fig. 8(a), the values r_i do not change. The kurtosis becomes $A = 3$ and the skewness becomes $B = 0$.
- In Fig. 8(b), the position i of the decreased r_i and that of the increased r_i are close. Because the effect of an increase and a decrease is cancelled out, the kurtosis becomes $A \approx 3$ and the skewness becomes $B \approx 0$.
- In Fig. 8(d), because the shape of the reference pattern is flattened relative to the normal distribution and the shape of the reference pattern has a long tail to the right side, the kurtosis becomes $A \ll 3$ and the skewness becomes $B \gg 0$.
- In Fig. 8(c), because the shape of the reference pattern is an intermediate state between (b) and (d), the kurtosis becomes $A < 3$ and the skewness becomes $B > 0$.
- In Fig. 8(e), the reference pattern has a small shape change from the normal distribution, and the kurtosis becomes $A \approx 3$ and the skewness becomes $B \approx 0$.

From Figs. 8(a)-(d), we can understand that the values $|A|$ and

$|B|$ respectively increase monotonically according to the increase of the "difference" between peaks of the standard and input patterns. Also, from Fig. 8(e), it is clear that $A \approx 3$ and $B \approx 0$ for the "wobble". In this method, when a "difference" occurs between peaks of the standard and input patterns with a "wobble" due to noise, the "wobble" is absorbed and the distance metric increases monotonically purely in accord with the increase of the "difference". On this basis, we verify that the GD algorithm matches the characteristics $\langle 1 \rangle$ and $\langle 2 \rangle$ of the mathematical model. GD is defined using both the kurtosis A and the skewness B [3]. We have both one-dimensional GD and two-dimensional GD. In addition, we have a fast calculation GD algorithm [1][3].

6. Evaluation experiments

To authenticate the effectiveness of the GD algorithm described in Section 5, we performed evaluation experiments for the vocalisations of Macleay's Fig-Parrot. Fig. 9 shows that, using the GD algorithm, pattern matching even in a noisy environment is accurate. The same GD algorithm has been successfully used to locate cavities in concrete structures by comparing the acoustic response to controlled surface tapping above integral concrete and concrete compromised by erosion cavities. Recognition accuracy comparing taps arising from integral and cavity-compromised concrete is 17/20 [5]. These applied experiments verify the effectiveness of the GD algorithm.

7. Conclusions and future work

We have described the GD algorithm and introduced associated automatic recognition software for bird vocalisations. The software executes parallel processing using multiple CPUs. In our future work, we will continue to improve the recognition software.

References

- [1] M. Jinnai, S. Tsuge, S. Kuroiwa and M. Fukumi, "A New Geometric Distance Method to Remove Pseudo Difference in Shapes", *International Journal of Advanced Intelligence*, Volume 2, Number 1, pp. 119-144 (2010) <http://iaia-i.com/ijai/index.html>
- [2] M. Jinnai, N. Boucher, M. Fukumi and H. Taylor, "A new optimization method of the geometric distance in an automatic recognition system for bird vocalisations", *International Congress on Acoustics*, 105, (April 2012, Nantes, France) <http://www.soundid.net/>
- [3] M. Jinnai, Y. Akashi, K. Hashimoto and S. Hayashi, "Method for detecting abnormal sound and method for judging abnormality in structure by use of detected value thereof", *Patent No.* US9552831(2017), AU2016200487(2017), CA2918533(2018), JP5956624(2016)
- [4] M. Jinnai, "A new segmentation method of bird vocalisations recorded in the distance", *2018 Autumn Meeting, Acoustical Society of Japan*, 3-1-5
- [5] M. Jinnai, K. Hashimoto, S. Hayashi and E. J. Pedersen, "Recognition of abnormal vibrational responses of concrete structures", *2019 Spring Meeting, Acoustical Society of Japan*, 1-6-15

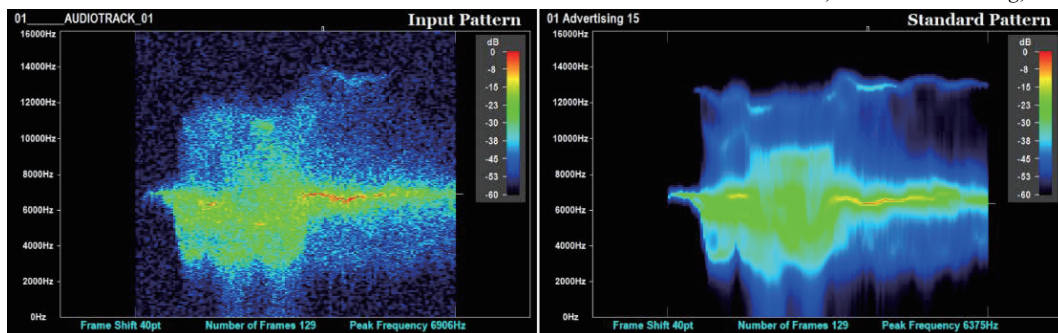


Fig. 9 Result of pattern matching for bird call recognition in a noisy environment

対話行為を用いた制御可能なニューラル対話モデルの検討

An Investigation of Controllable Neural Conversation Model with Dialogue Acts

河野 誠也 ^{*1}

Seiya Kawano

吉野 幸一郎 ^{*1*2}

Koichiro Yoshino

中村 哲 ^{*1}

Satoshi Nakamura

^{*1} 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

^{*2} 科学技術振興機構

Japan Science and Technology Agency

Dialogue act is known as an essential component of the dialogue system, which captures the user's intention and produces the appropriate response. In this paper, we propose a controllable response generation model given dialogue acts. Recent neural conversation models are based on the end-to-end approach that learns a mapping a mapping between dialogue histories and response utterances. However, it was difficult to control the contents of the response generated by the model. Several models tackled the problem of generating responses under the specified dialogue acts as a condition; however, these models still have problems on conditioned generations. In this paper, we introduced an extended framework of the generative adversarial network that optimizes both conditioned generator and discriminator which explicitly classifies dialogue act classes. Experimental results showed that our conditional response generation model improved both the response quality and controllability of neural conversation generation.

1. はじめに

対話行為とは、話者が発話において持つ何かしらの「意図」あるいは発話における「機能」であり、その意図や機能の種類として対話行為タグが定義される [Boyer 10, 河野 18]. 対話行為は、対話モデルにおける基本単位の一つとして利用されており、特に近年、対話における発話間の相互作用をモデル化する上で有用であることが知られている [Yoshino 15]. しかし、近年広く用いられているニューラル対話モデル (Neural Conversation Model; NCM) では、こうした対話行為によってシステム発話を明示的に操作することが困難である。そこで本研究では、システム応答が持つ対話行為の情報を条件として用いた敵対的生成学習の枠組みを NCM に導入する。具体的には、与えられた対話行為に基づいて応答を生成する Generator と、Generator が生成した応答が指定した対話行為に基づいた適切なものであるかを判別する Discriminator を構築し、これらの 2 つのモデルを交互に敵対的に訓練する。このような学習の枠組みの導入により、NCM が任意の対話行為に基づいた適切な応答を生成することが可能かどうかの評価を行った。

2. 関連研究

対話行為の情報をを用いて発話生成を行った研究として、Wen や Zhao らの研究がある [Wen 15, Zhao 17]. Wen らは、レストラン案内対話のドメインにおいて、任意の対話行為 (e.g. Inform) とそのスロット情報 (e.g. food=Chinese, name=Seven.Days) を用いて発話生成 (e.g. “Seven Days serves Chinese.”) を行うニューラル言語モデル (Semantically Conditioned LSTM; SC-LSTM) を提案している。SC-LSTM では、LSTM 言語モデルに「読み込みゲート」と呼ばれるゲーティング機構を新たに導入することで、単語の生成時に、指定した対話行為とスロット情報のうちの情報を使うかの制御を行う。しかしながら、SC-LSTM は限られたドメインのタスク達成を対象としているため、取り扱う発話の範囲が膨大なオー

ンドメインの非タスク志向対話システムにおいて適用しようとした場合、対話行為とそのスロット情報の両方をフレームとして保持することは現実的ではない。

また、Zhao らは NCM が過度に一般化された応答を返すという問題 [Li 16a] に対して、対話行為の情報をを用いることでより多様かつ高品質な応答生成を行う KgCVAE (Knowledge-guided Conditional Variational Autoencoder) を提案している。KgCVAE では、Encoder の出力結果から応答の対話行為を予測し、これを Decoder への条件として使用している。また、同様のアプローチとして、NCM における Decoder に感情ラベルや話者ラベルを条件として利用し、NCM における応答内容を制御しようとするモデルも提案されている [Zhou 18, Li 16b]. これらの NCM は、SCE (Softmax Cross Entropy Loss) に基づいて Decoder の各単語生成ステップにおける単語予測の最適化を行う。しかしこれらのモデルは必ずしも生成された応答が指定したラベルに基づいた適切な応答であることを保証しない。この傾向は特に、低頻度のラベルで顕著である。これは、訓練データにおけるラベルの分布の偏りに起因する [Zhou 18].

一方で、NCM に敵対的生成ネットワーク (Generative Adversarial Net; GAN) を導入することで、より多様かつ高品質な応答生成モデルの構築が検討されている [Li 17a]. 具体的には、応答発話を生成する Generator と、与えられた応答が学習データか Generator のどちらに由来するものかを識別する Discriminator を構築する。これらの 2 つのモデルを交互に敵対的に訓練することで、より自然な発話を生成しようとする試みである。このような、学習の枠組みが従来の NCM と比較して高品質かつ多様な応答を生成したことが報告されている。本研究はこうした GAN に基づく学習の枠組みを、指定した対話行為らしい応答を生成することを目的として導入する。

3. 提案手法

本研究では、前節で述べた NCM における GAN に基づく発話生成の枠組みを拡張し、応答の対話行為を条件として用いる NCM (条件付き NCM) に応用する。具体的には、与えられた対話行為に基づいて応答を生成する Generator (G) と、G が

連絡先: 河野誠也, 奈良先端科学技術大学院大学, 〒 630-0101, 奈良県生駒市高山町 8916 番地の 5, kawano.seiya.kjo@is.naist.jp

生成した応答が指定した対話行為に基づいた適切なものであるかを識別する Discriminator (D) を構築し、これらの 2 つのモデルを交互に敵対的に訓練する。本手法は、G による応答生成と D による応答発話の評価の枠組みに、応答の対話行為の情報を明示的に導入している点で従来手法 [Li 17a] と異なる。本研究では、このような応答の対話行為の情報を利用した GAN の枠組みの導入により、NCM が任意の対話行為に基づいた適切な応答を生成することが可能かどうかの評価を行う。

3.1 対話行為による条件付き NCM

対話行為による条件付き応答生成では、対話履歴 $M = \{M_{i-1}, M_{i-2}, \dots, M_{i-n}\}$ と応答の対話行為 d_i を用いて、応答発話 $R_i = \{w_1, w_2, \dots, w_I\}$ を生成する。ここで、 n は対話長、 I は発話長である。本研究では、最も素朴な NCM の拡張として図 1 のように、階層的な Encoder-decoder モデルに基づく NCM [Tian 17] に対して、発話のデコード時に明示的に対話行為の情報を与えるモデルを構築した。

通常の Encoder-Decoder モデルでは、Decoder における各単語出力ステップ t において、直前の隠れ状態 h_{t-1} と現在の単語 w_t のベクトル表現 x_t が入力として使用される。一方で、本モデルでは、直前の隠れ状態 h_{t-1} に加えて、対話行為 d_i のベクトル表現 e と x_t の連結ベクトル $v_t = e \oplus x_t$ を入力として使用することで、指定した対話行為に基づいた応答を NCM が生成することを期待する *1。

RNN には、LSTM や GRU と比較して高速で訓練が可能であることが知られている SRU (Simple Recurrent Unit) [Lei 17] を用いた。評価実験では、Encoder 及び Decoder で使用する SRU の層数は 2、中間層の次元は 1024、単語埋め込みベクトルの次元は 256、対話行為埋め込みベクトルの次元は 100 とした。また、Encoder における最大対話長 n は 5 に設定した。モデルの訓練は SCE により行い、開発セットにおける Perplexity が最も下がったモデルを評価に使用する。

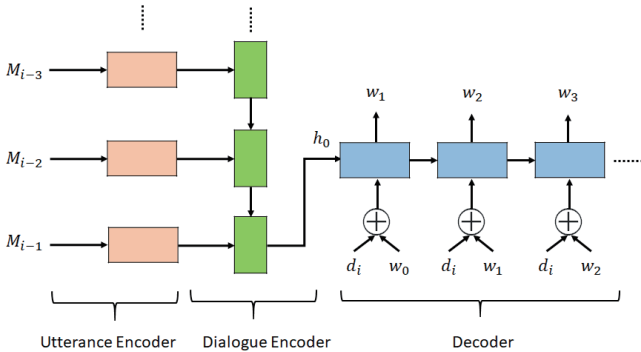


図 1: 対話行為による条件付き NCM の概要

3.2 対話行為による条件付き敵対的生成ネットワーク

本研究では、敵対的生成学習 (GAN) の手法として SeqGAN (Sequence Generative Adversarial Net) [Yu 17, Li 17a] の枠組みを採用し、対話行為による条件付き応答生成へ応用する。発話のような離散系列データの生成に GAN を適用する場合、Generator の出力は単語の離散値の系列を前提とするため、Discriminator から Generator への勾配の更新を行うことが難しい。この問題に対して、SeqGAN では、強化学習によるアプローチを Generator の更新に用いることで、離散的

な系列データに GAN を適用できることを示している。以下に GAN が持つ Generator, Discriminator の構成を示す。

3.2.1 Generator (G)

Generator では、応答生成における単語選択のプロセスをマルコフ決定過程とみなし、各ステップにおける単語生成がある政策に基づいて選択された行動によって決定されると考える。SeqGAN では、Discriminator によって推定された応答文の妥当性についてのスコア $Q(M, R, d)$ を Generator の報酬として利用し、この期待報酬を最大化するように Generator を方策勾配法により更新する [Williams 92]。本研究における対話行為による条件付き応答生成においては、その目的関数 J と勾配を次式で定義する *2。ここで、 π は応答 R の生成確率である。本研究では、前節で述べた対話行為による条件付き NCM を Generator として用いる。

$$J(\theta) = \mathbb{E}_{R \sim p(R|M, d)}(Q(M, R, d)|\theta) \quad (1)$$

$$\nabla J(\theta) \approx Q(M, R, d) \nabla \log \pi(R|M, d) \quad (2)$$

$$= Q(M, R, d) \nabla \sum_t \log \pi(w_t|M, d, w_{1:t-1}) \quad (3)$$

3.2.2 Discriminator

本研究では、以下の 2 種類の Discriminator について検討を行う。

- Implicit-Discriminator; 対話履歴 M 、応答発話 R 、対話行為 d を入力として、与えられた R が本物 (学習データ由来のもの) か偽物 (Generator 由来のもの) であるかの 2 クラス分類を行う。Generator によって生成された R が本物であると推定される確率を $Q(M, R, d)$ として利用する。これは、対話行為タグを素朴に Discriminator の入力として用い、生成結果が制御されることを期待するものである。
- Explicit-Discriminator; 対話履歴 M 、対話行為 d に対応する応答発話 R を入力として、 R が Generator 由来のものであるとき偽物、学習データ由来のものであるときその対話行為を分類するように Discriminator を訓練する。つまり、対話行為の種類数を N としたとき、 $N+1$ クラスの多クラス分類を行う。ここで、 $Q(M, R, d)$ は、Generator によって生成された R が学習データ由来の対話行為 d であると推定される確率である。生成文が各対話行為の特徴を反映しているかを Discriminator が明示的に予測することにより、各対話行為の特徴が生成文に反映されることを期待する。

Discriminator には、図 2 に示すような Utterance Encoder と Dialogue Encoder で構成される階層型 RNN を Discriminator として用いる。また、Generator の場合と同様に、RNN には SRU を Discriminator の訓練の高速化のために採用する。評価実験においては、SRU の層数は 1、中間層の次元は 1024、単語埋め込みベクトルの次元は 256、対話行為埋め込みベクトルの次元は 100、全結合層における中間層の次元は 512 に設定した。また、Discriminator の更新における損失関数には SCE を用いる。

4. 評価実験

4.1 データセット

対話行為による条件付き NCM の訓練と評価には DailyDialog Corpus [Li 17b] を用いた。本コーパスには、13,118 対話における発話に、Inform (4,653 発話)、Questions (29,428 発話)、Directives (17,29 発話)、Commisive (9,724 発話) の

*1 入力単語 w_t 及び対話行為 d は埋め込み層を用いて密ベクトル表現 x_t, e へと変換される。

*2 本研究では簡単化のために文全体で単一の報酬を使用している点に留意されたい。

表 1: 対話行為による条件付き NCM の評価

モデル	Perplexity	Average	Greedy	Extreme	Distinct-1	Distinct-2	Accuracy
条件付き NCM (greedy)	36.6614	0.7912	5.5075	0.5298	0.0336	0.0160	0.8644
Adversarial-Implicit (greedy)	39.2864	0.7882	5.6239	0.5394	0.0297	0.0069	0.8585
Adversarial-Explicit (greedy)	39.6993	0.7867	5.6661	0.5395	0.0311	0.0010	0.8843
条件付き NCM (sampling)	36.6614	0.7917	5.1785	0.4905	0.0767	0.2650	0.8176
Adversarial-Implicit (sampling)	39.2864	0.7870	5.2566	0.5035	0.0594	0.2093	0.8062
Adversarial-Explicit (sampling)	39.6993	0.7865	5.2782	0.5041	0.0583	0.1186	0.8573
条件付き NCM (beam)	36.6614	0.7847	5.5680	0.5381	0.0325	0.0006	0.8707
Adversarial-Implicit (beam)	39.2864	0.7793	5.6668	0.5411	0.0267	0.0031	0.8615
Adversarial-Explicit (beam)	39.6993	0.7796	5.7167	0.5431	0.0274	0.0017	0.8865

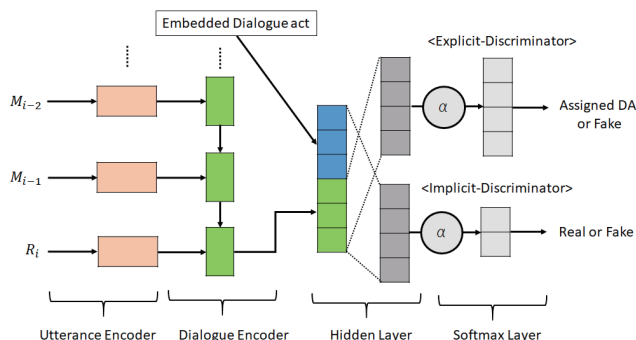


図 2: Implicit-Discriminator, Explicit-Discriminator の概要

4 つの対話行為タグがアノテーションされている。本研究では、このコーパスを訓練 (11,118 対話)、開発 (1,000 対話)、評価用データ (1,000 対話) に分割し、対話行為による条件付き NCM の訓練と評価に用いた。語彙サイズは 2,5000 に設定し、未知語は特殊記号 “UNK” に置き換えうえで訓練を行う。対話行為による条件付き敵対的生成ネットワークの訓練方法については、Li らの手法 [Li 17a] を参考に以下の手順で行う。D-step, G-step におけるステップ数はそれぞれ、4, 20 に設定し、batch size は 32 とした。また、Discriminator の更新には SGD (学習率 $1e-3$)、Generator の更新には Adam (学習率 $1e-5$) を用いた *3。

Algorithm 1 条件付き敵対的生成ネットワークの訓練

```

1: for number of iteration do
2:   for number of D-step do
3:     sample  $(M, R, d)$  from training data
4:     generate response  $\hat{R}$  using  $G$  on  $(M, d)$ 
5:     update  $D$  using  $(M, \hat{R}, d)$  and  $(M, R, d)$ 
6:   for number of G-step do
7:     sample  $(M, R, d)$  from training data
8:     generate response  $\hat{R}$  using  $G$  on  $(M, d)$ 
9:     compute reward  $r$  for  $(M, \hat{R}, d)$  using  $D$ 
10:    update  $G$  on  $(M, \hat{R}, d)$  using  $r$ 

```

4.2 対話行為による条件付き応答生成の自動評価

対話行為による条件付き応答生成の評価については、関連性 (Relevance)、多様性 (Diversity)、制御性 (Controllability) の 3 つの観点から多面的に評価を行う。

- 関連性 (Relevance); 生成された応答発話とレファレンスの応答発話における単語分散表現の類似度により評価を行う。類似度の算出方法には Embedding Average, Greedy Matching, Vector Extreme の三つの評価指標 [Liu 16] を採用した *4。

*3 800 iteration 実行した後のモデルを評価に使用した

*4 これらの詳細については原論文 [Liu 16] を参照されたい。本研究

- 多様性 (Diversity); 対話システムの分野で広く採用されている Distinct-1 及び Distinct-2 により評価を行う。Distinct-1 (Distinct-2) は、テストデータに対して生成された応答発話集合におけるユニークな unigram (bigram) の数を、生成されたすべての unigram (bigram) の数でスケーリングした指標である [Li 16a]。

- 制御性 (Controllability); 入力として使用した対話行為と、実際に生成された応答文における対話行為の推定結果の一致率 (Accuracy, Precision, Recall, F1) に基づいて評価する。応答文の対話行為については、事前訓練した対話行為分類器 *5 による推定結果 (評価用データにおける推定精度は、0.8303) を正解として用いる。

5. 実験結果

表 1 に対話行為による条件付き NCM の自動評価結果について示す。ここで、条件付き NCM は SCE により訓練されたモデルであり、Adversarial-Implicit モデル、Adversarial-Explicit モデルはそれぞれ、Implicit-Discriminator, Explicit-Discriminator を使用した GAN により訓練されたモデルである。応答文の生成においては、貪欲的に出力単語を選択する場合 (greedy) と多項分布に基づいてサンプリングを行う場合 (sampling), beam search (ビーム幅 5) を行う場合 (beam) の 3 通りを試した。

5.1 応答生成結果の評価

GAN を導入した Adversarial-Implicit モデルでは、いずれの decoding 手法を用いた場合においても、GAN を用いない条件付き NCM と比較して、関連性尺度の Vector Extreme が向上していることが確認できる。一方で、多様性・制御性については改善が確認できなかった。また、sampling を使用する場合は、greedy に応答を生成する場合と比較して、多様性は改善する一方で、制御性は 0.0523 ポイントもの低下が確認できた。しかしながら、Adversarial-Explicit モデルでは、いずれの decoding 手法を用いた場合でも、GAN を用いない条件付き NCM と比較して関連性 (Greedy Matching & Vector Extreme) と制御性が一貫して改善している。特に、制御性においては、greedy に応答を生成する場合は 0.0199 ポイント、sampling を使用する場合は 0.0396 ポイント、beam search を使用する場合は 0.0157 ポイントもの改善が確認でき、本モデルの有効性が示唆される結果となった。

Adversarial-Explicit モデルでは、応答文が Generator か学習データのどちらに由来するものかを判別すると同時に、明示的に対話行為の推定を行っている。これに対して、Adversarial-Implicit モデルでは、生成された応答の対話行為を明示的に

では、wikipedia データで訓練された fastText による単語分散表現 (300 次元) を用いた。

*5 Explicit-Discriminator と同様のアーキテクチャを採用した。

推定しておらず制御性のスコア改善につながらなかったと考えられる。また、GAN を導入した 2 つのモデルでは、GAN を用いない条件付き NCM と比較して Perplexity の改善が確認できなかった。GAN によるテキスト生成モデルが必ずしも Perplexity を改善しないことは先行研究 [Tevet 18] においても指摘されている。これは、敵対的生成学習は単語予測の最適化を目的としないためである。

5.2 各対話行為タグ毎における制御性の評価

表 2, 表 3 はそれぞれ Adversarial-Explicit モデルにおいて、beam search と sampling を使用して応答生成を行った場合における、各対話行為タグ毎の制御性の評価結果について示している。括弧内は、GAN を用いない条件付き NCM におけるスコアとの差分である。ここで、テストセットにおける各対話行為タグの出現数は、Inform は 2639, Questions は 1679, Directives は 938, Commisive は 484 である。

表 2: 各対話行為タグ毎における制御性 (beam search)

対話行為タグ	Precision	Recall	F1
Inform	0.8855	0.9231	0.9039 (+0.0125)
Questions	0.9699	0.9976	0.9836 (+0.0060)
Directives	0.9263	0.6695	0.7772 (+0.0622)
Commisive	0.5976	0.7211	0.6536 (+0.0187)
Macro Avg	0.8448	0.8278	0.8296 (+0.0249)
Weighted Avg	0.8926	0.8864	0.8854 (+0.0111)

表 3: 各対話行為タグ毎における制御性 (sampling)

対話行為タグ	Precision	Recall	F1
Inform	0.8766	0.9075	0.8918 (+0.0256)
Questions	0.9086	0.9774	0.9418 (+0.0214)
Directives	0.7963	0.6375	0.7081 (+0.1126)
Commisive	0.6364	0.5930	0.6139 (+0.0587)
Macro Avg	0.8045	0.7789	0.7889 (+0.0546)
Weighted Avg	0.8526	0.8573	0.8530 (+0.0293)

表 2, 表 3 より、GAN を用いない条件付き NCM と比較して、Adversarial-Explicit モデルでは、すべての対話行為タグで F 値が改善していることが確認できた。また、各対話行為タグにおける F 値のマクロ平均、加重平均も同様に改善していることが確認できた。特に、sampling を使用した場合においては、各指標のより顕著な改善が確認できる。しかしながら、Questions タグにおいては、各対話行為タグの中で最も小さな F 値の改善にとどまった。これは、Questions タグが疑問符や 5W1H のような単語を用いたある程度パターン化された文構造を持つため、問題が簡単 (既にベースラインが比較的高い制御性を実現している) であることが原因だと推察される。一方で、Directives タグについては、beam search を使用する場合は 0.0622 ポイント、sampling を使用する場合は 0.1126 ポイントの F 値の顕著な改善が確認できる。Commisive タグにおいても同様に、beam search を使用する場合は 0.0187 ポイント、sampling を使用する場合は 0.0587 ポイントの F 値の改善が確認できた。

6. おわりに

本研究では、任意の対話行為に基づいて応答発話の生成を行う条件付き敵対的生成ネットワークを提案し、その評価を行った。その結果、提案モデルはベースラインモデルと比較して、関連性、制御性における客観評価尺度が改善されることを示した。しかしながら、本研究では客観評価指標による提案手法の

評価のみにとどまり、生成文の品質についてより厳密に評価するためには人手による評価を行うことが必要であると考えられる。さらに、Generator の更新の際に使用する報酬値の計算に多様性や発話長などの情報を導入することなど、敵対的生成ネットワークの訓練方法そのものの改善を行うことも重要であると考えられる。また、今後は提案手法の他タスク (感情やペルソナを考慮した応答生成など) における有効性の検討も進める予定である。

謝辞

本研究は、JST さきがけ (JPMJPR165B) および JST CREST (JPMJCR1513) の支援を受けた。

参考文献

- [Boyer 10] Boyer, K. E., Ha, E. Y., Phillips, R., Wallis, M. D., Vouk, M. A., and Lester, J. C.: Dialogue act modeling in a complex task-oriented domain, in *Proc. of SIGDIAL*, pp. 297–305 Association for Computational Linguistics (2010)
- [Lei 17] Lei, T., Zhang, Y., and Artzi, Y.: Training rnns as fast as cnns, *arXiv preprint arXiv:1709.02755* (2017)
- [Li 16a] Li, J., Galley, M., Brockett, C., Gao, J., and Dolan, B.: A Diversity-Promoting Objective Function for Neural Conversation Models, in *Proc. of NAACL-HLT*, pp. 110–119 (2016)
- [Li 16b] Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., and Dolan, W. B.: A Persona-Based Neural Conversation Model, *CoRR*, Vol. abs/1603.06155, (2016)
- [Li 17a] Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A., and Jurafsky, D.: Adversarial Learning for Neural Dialogue Generation, in *Proc. of EMNLP*, pp. 2157–2169 (2017)
- [Li 17b] Li, Y., Su, H., Shen, X., Li, W., Cao, Z., and Niu, S.: Daily-Dialog: A Manually Labelled Multi-turn Dialogue Dataset, *arXiv preprint arXiv:1710.03957* (2017)
- [Liu 16] Liu, C.-W., Lowe, R., Serban, I., Noseworthy, M., Charlin, L., and Pineau, J.: How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation, in *Proc. of EMNLP*, pp. 2122–2132 (2016)
- [Tevet 18] Tevet, G., Habib, G., Shwartz, V., and Berant, J.: Evaluating Text GANs as Language Models, *arXiv preprint arXiv:1810.12686* (2018)
- [Tian 17] Tian, Z., Yan, R., Mou, L., Song, Y., Feng, Y., and Zhao, D.: How to make context more useful? an empirical study on context-aware neural conversational models, in *Proc. of ACL*, Vol. 2, pp. 231–236 (2017)
- [Wen 15] Wen, T.-H., Gasic, M., Mrksic, N., Su, P.-H., Vandyke, D., and Young, S.: Semantically conditioned lstm-based natural language generation for spoken dialogue systems, *arXiv preprint arXiv:1508.01745* (2015)
- [Williams 92] Williams, R. J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Machine learning*, Vol. 8, No. 3-4, pp. 229–256 (1992)
- [Yoshino 15] Yoshino, K. and Kawahara, T.: Conversational system for information navigation based on POMDP with user focus tracking, *Computer Speech & Language*, Vol. 34, No. 1, pp. 275–291 (2015)
- [Yu 17] Yu, L., Zhang, W., Wang, J., and Yu, Y.: SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient., in *AAAI*, pp. 2852–2858 (2017)
- [Zhao 17] Zhao, T., Zhao, R., and Eskenazi, M.: Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders, in *Proc. of ACL*, Vol. 1, pp. 654–664 (2017)
- [Zhou 18] Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory, in *Thirty-Second AAAI Conference on Artificial Intelligence* (2018)
- [河野 18] 河野誠也, 吉野幸一郎, 中村哲: 条件付き敵対的生成ネットワークを用いたデータ拡張による対話行為分類法の検討, 情報処理学会研究報告音声言語情報処理 (SLP), No. 9 (2018-SLP-125), pp. 1–6 (2018)

対話システムにおける知識獲得質問のための ラベル文字列を用いた知識グラフ補完性能の向上

Improvement of Knowledge Graph Completion Using Label Characters for Questions to Acquire Knowledge in Dialog Systems

藤岡 勇真^{*1}
Yuma Fujioka

林 克彦^{*1}
Katsuhiko Hayashi

中野 幹生^{*2}
Mikio Nakano

駒谷 和範^{*1}
Kazunori Komatani

^{*1}大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

^{*2}(株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co., Ltd.

Dialogue systems cannot respond about information that is not explicitly described in their knowledge bases. Constructing a perfect knowledge base is practically impossible; that is, filling all the values in databases is quite labor-intensive. We are trying to construct a system that can acquire information that is not explicitly described in the knowledge base by inferring latent information from knowledge graphs. In particular, we complement the links in a knowledge graph by using an embedding into latent space. We use partial character sequences of labels (i.e. entity names) to improve of knowledge graph completion. We also show examples of queries generated using the latent information in our target knowledge graph.

1. はじめに

データベースを参照して応答を行うタイプの対話システムは、データベースに明示的に記述されていない情報について上手く応答できない。想定する応答に必要な情報が網羅したデータベースが構築できれば良いが、人手で構築することは不可能な場合が多い。

そこで我々は、データベースにない情報を自ら獲得出来る対話システムを構築することでこの問題の解決を試みている。そのようなシステムはユーザに質問して情報を聞き出そうとするが、この時に明らかに間違っている質問をしてしまうとユーザの対話意欲を削いでしまうという課題がある。

我々は、グラフ構造を持つデータベースである知識グラフにある潜在的な情報を利用して、できるだけ間違った内容の質問を避けることができる対話システムの構築を目指している[藤岡 18]。知識グラフはデータ間の様々な関係を表現するのに長けた知識モデルであり[Angles 08]、近年様々なシステムやサービスで用いられている。知識グラフに対する代表的な解析手法として、潜在空間への埋め込みがある[Kadlec 17]。埋め込み表現を用いることで知識グラフのリンクの補完が行えるが、この補完されたリンクの確信度に基づきユーザへの質問を生成する。

本稿では、知識獲得で利用する知識グラフ補完の精度を向上させるため、ラベルの文字列を用いた知識グラフ埋め込み手法を提案する。ここでラベルとは、知識グラフ上のエンティティが持つ、自身を識別する名称である。このラベルの文字列を部分文字列に分解し利用することで、類似したラベルを持つエンティティ間に関連性を持たせ、補完精度の改善を図る。この手法の効果を、実際の対話システムでの利用を目的として作成された知識グラフを用いて検証し、どのような質問が行えるかを例示する。

2. 知識獲得を行う対話システムの枠組み

2.1 知識グラフ埋め込みを利用した知識獲得

本研究では、対話システムが知識グラフをバックエンドデータベースとして持っていることを仮定する。一般に知識グラフはラベル付き有向グラフとして表される。有向グラフ上のエッジにはリレーションを表すラベルが付与されており、ノードはエンティティに相当する。知識グラフ上のエンティティの集合を $\mathcal{E}$ 、リレーションの集合を $\mathcal{R}$ とする。 $e_i, e_k \in \mathcal{E}$, $w_j \in \mathcal{R}$ に対して三つ組 (i, j, k) をトリプルと呼ぶ。この時 i, j, k をそれぞれ主語、述語、目的語と呼ぶ。トリプルは2つのエンティティ間の関係を表現する知識グラフの基本的な要素である。知識グラフ $\mathcal{G}$ はトリプルを要素とする集合として表せる。

図1にユーザから知識獲得を行う対話システムの枠組みを示す。知識グラフ上にない知識、すなわちトリプルをユーザから獲得するために質問リストを生成する(図1①)。知識グラフ上にないトリプルの存在をユーザに質問することで知識獲得を試みる。図の例では(もみじ丼, 料理種, ?)と目的語を穴埋めするように質問リストを生成している。本稿で生成する質問リストはこの例のように、目的語を穴埋めするような形式で生成されるものとする。この時、対話の文脈や話題に関連する質問リストを生成するようにし、急な話題転換を避けるようにする。そして既知の知識グラフから埋め込み表現を学習し(図1②)、埋め込み表現から質問リスト内のトリプルに対し存在尤度を表すスコアを計算し確信度として付与する(図1③)。付与した確信度に基づき質問リストをソートし、最も高い確信度かつその確信度がしきい値以上のトリプルの有無をユーザに質問し知識獲得を試みる(図1④)。確信度による順位だけでなく、確信度の絶対値もしきい値によって考慮する。これは、確信度が低い、すなわち推定結果に自信がない場合にもユーザに問いかけてしまうことを防ぎ、明らかに間違っているような質問でユーザの対話意欲を削がないようにするのが狙いである。システムの設計指針は、質問する内容をなるべく正しいものとするところである。知識グラフ埋め込みによる補完精度の向上は質問内容の正しさに直結するため、ラベル文字列を用いた知識グラフ埋め込みによる補完精度向上を3章で提案する。

連絡先: 藤岡 勇真, 大阪大学 産業科学研究所,
大阪府茨木市美穂ヶ丘 8-1, Tel:06-6879-8416,
E-mail:fujioka@ei.sanken.osaka-u.ac.jp

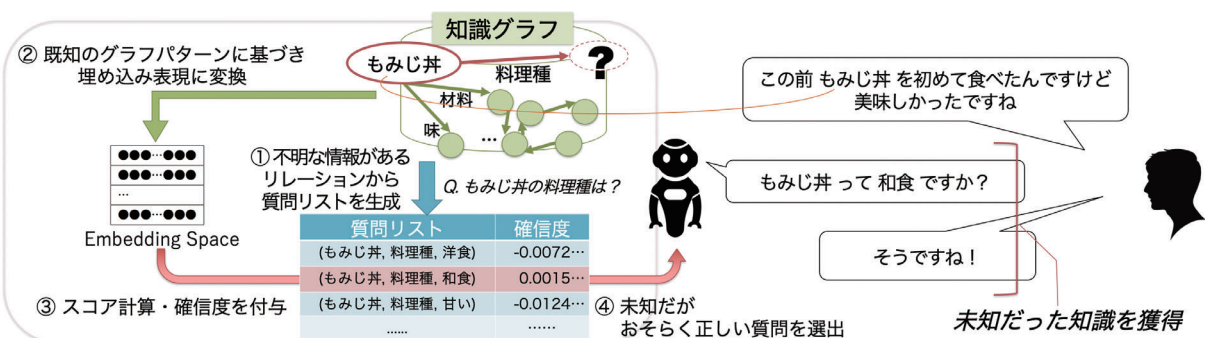


図 1: 知識獲得を行う対話システムの枠組み

2.2 知識グラフ埋め込みと ComplEx

2章で述べた枠組みでは、トリプルの存在尤度を計算する必要がある。その計算に知識グラフの埋め込み表現を利用する。知識グラフ埋め込みは、知識グラフに対する代表的な解析手法として知られている。低次元線形空間に知識グラフを埋め込み汎化させることで、グラフ上の欠損したリンクの有無を推論し補完する。知識グラフは、 $|\mathcal{E}| \times |\mathcal{R}| \times |\mathcal{E}|$ の3階テンソル $\mathcal{X}$ として表現することができ、 $\mathcal{X}$ の (i, j, k) 要素 $x_{i,j,k}$ は以下のように表される。

$$x_{i,j,k} = \begin{cases} 1, & (i, j, k) \in \mathcal{G}; \\ -1, & \text{otherwise.} \end{cases}$$

この表現を用いて、知識グラフ埋め込みではトリプル (i, j, k) が知識グラフ上に存在する確率 $P(x_{i,j,k} = 1)$ を、モデルに対応したスコア関数 ϕ を用いて以下のように表す。

$$P(x_{i,j,k} = 1) = \sigma(\phi(i, j, k; \Theta))$$

$\sigma(\cdot)$ はシグモイド関数、 Θ は各モデルにおけるパラメータを表す。

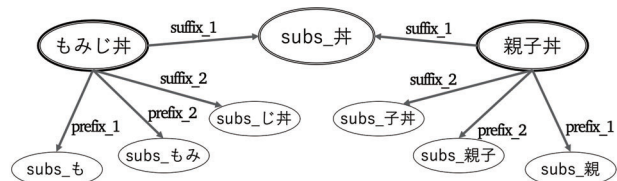
ComplEx[Trouillon 16] は知識グラフ埋め込みモデルの一種である。ComplEx は複素数表現とエルミート内積を利用したモデルとして知られる。ComplEx におけるスコア関数 ϕ_{Comp} は、 $\mathbb{C}^n$ を複素 n -次元空間、 $\mathbf{e}_i, \mathbf{e}_k \in \mathbb{C}^D$, $\mathbf{w}_j \in \mathbb{C}^D$ をそれぞれエンティティ・リレーションに関する D 次元の埋め込みベクトルとし、 $Re(x)$ を x の実部とすると、

$$\phi_{Comp}(i, j, k; \Theta) = Re(\langle \mathbf{w}_j, \mathbf{e}_i, \bar{\mathbf{e}}_k \rangle)$$

と書ける。ここでベクトル $\mathbf{v}$ の l 番目の要素を v_l とした時、 $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle := \sum_k a_k b_k c_k$ と定義する。 $\bar{\mathbf{v}}$ は $\mathbf{v}$ の複素共役ベクトルである。

3. ラベル文字列を用いた知識グラフ埋め込み

本章では、エンティティの持つラベル文字列を知識グラフの埋め込み学習に用いる手法を提案する。2.2 節で述べたモデルはエンティティ同士がどのようにリンクしているかを元に埋め込み表現を学習するが、すべてのエンティティが持つラベル文字列に関する情報は学習に組み込まれていない。そこで、知識グラフ上のエンティティのラベル文字列を部分文字列に分解し、先頭もしくは末尾の数文字を疑似エンティティとして作成することでラベル文字列を組み込む。似た部分文字列を持つエンティティ同士は似た性質を持つ可能性が高いという仮定のも

図 2: 「もみじ井」と「親子井」に対する $N = 2$ での展開例

とで、このようなエンティティ同士の関連性が強くなり補完精度の向上に結びつくことを期待する。

N 文字以下の先頭もしくは末尾の部分文字列で擬似的にエンティティを作成し新たにトリプルを構成することを、 N 文字で展開すると記述する。図 2 に $N = 2$ 、すなわち 2 文字で展開を行った例を示す。展開の対象となっている「もみじ井」と「親子井」の先頭・末尾 1, 2 文字が抽出され、部分文字列であることを示す $subs_$ を先頭に付けたラベルを持つエンティティが作成される。そして展開元のエンティティから作成した疑似エンティティに向けて $prefix_N$ もしくは $suffix_N$ というリレーションでトリプルを構成する。図 2 の例では、2 つの料理が「subs_井」という疑似エンティティを経由して繋がっていることがわかる。

4. ラベル文字列を用いた知識グラフ埋め込みによる補完精度

ラベル文字列を用いた知識グラフ埋め込みによる補完の精度を検証し、その効果を調査する。

4.1 使用データ

本実験で使用する知識グラフは、対話システムでの運用を目的として、料理に関する表形式のデータベースを元に作成されたものである。詳述すると、料理やその材料、料理種、味、食べられる場所などが格納されている。このデータベースは人手で作成されたものであり、部分的にしか情報がないため、現在も継続して情報が追加され続けている。このデータベースから作成される知識グラフに対し、2.1 節で述べた知識獲得の枠組みを適用して情報の拡充を行うという設定の下で議論する。

この知識グラフはエンティティ数 $|\mathcal{E}| = 7289$ 、リレーション数 $|\mathcal{R}| = 14$ であり、そのトリプル数は 22321 である。エンティティに付与されたラベルの長さの平均は 5.88 (文字)、標準偏差は 3.18 (文字) である。

表 1: 各設定における精度指標及び学習データの詳細

設定	Hits@					学習トリプル数	エンティティ数	リレーション数
	1	3	5	10	MRR			
Baseline	0.140	0.205	0.272	0.406	0.215	17857	6898	14
$N = 1$	0.187	0.306	0.394	0.521	0.286	31681	7865	16
$N = 2$	0.237	0.384	0.473	0.580	0.345	45361	12147	18
$N = 3$	0.257	0.436	0.524	0.621	0.377	57893	18534	20
$N = 4$	0.246	0.441	0.523	0.626	0.373	68477	25035	22
$N = 5$	0.241	0.448	0.532	0.627	0.373	76837	30970	24
$N = 6$	0.227	0.452	0.537	0.629	0.368	83271	35791	26
$N = 7$	0.221	0.472	0.550	0.641	0.373	87975	39381	28

表 2: リレーション毎の Hits@K. $N = 3$ とした場合を示している (括弧内の数値は Baseline での結果との差を表す).

リレーション名	Hits@							出現数	
	1		3		5		10		
is_a	0.157	(+0.140)	0.284	(+0.216)	0.331	(+0.240)	0.389	(+0.243)	795
材料	<u>0.056</u>	(+0.047)	0.108	(+0.075)	0.158	(+0.107)	0.238	(+0.115)	702
popularity	0.119	(+0.108)	0.377	(+0.238)	0.565	(+0.308)	0.736	(+0.169)	639
別名	<u>0.828</u>	(▼ −0.108)	0.980	(+0.010)	0.982	(+0.008)	0.986	(+0.012)	501
meal_type	0.578	(+0.516)	0.878	(+0.609)	0.916	(+0.477)	0.950	(+0.235)	417
料理種	0.130	(+0.115)	0.367	(+0.315)	0.536	(+0.382)	0.682	(+0.279)	330
温度	0.297	(+0.246)	0.572	(+0.381)	0.674	(+0.347)	0.801	(+0.220)	236
restaurant	0.385	(+0.261)	0.578	(+0.280)	0.646	(+0.211)	0.714	(+0.180)	161
味	0.239	(+0.159)	0.487	(+0.336)	0.522	(+0.301)	0.611	(+0.310)	113
場所	0.000	(±0.000)	0.030	(±0.000)	0.061	(+0.030)	0.091	(+0.030)	33
typical_side	0.056	(+0.056)	0.111	(+0.111)	0.222	(+0.167)	0.278	(+0.222)	18
代表料理	0.000	(±0.000)	0.000	(±0.000)	0.063	(+0.063)	0.063	(+0.063)	16
機会	0.000	(±0.000)	0.067	(+0.067)	0.133	(+0.067)	0.133	(+0.067)	15
時間	0.182	(+0.091)	0.364	(+0.273)	0.455	(+0.273)	0.455	(+0.182)	11

4.2 実験方法

精度検証のため 4.1 節で述べた知識グラフを無作為に 5 分割し, その内 4 つをトレーニングデータ $\mathcal{G}'$, 1 つをテストデータ $\mathcal{H}'$ とする 5 分割交差検証を行った. トレーニングデータを埋め込み, その埋め込み表現と $\mathcal{H}'$ から知識グラフ補完で一般に用いられる精度指標である Hits@K と MRR を算出した. Hits@K は $(i', j', k') \in \mathcal{H}'$ の内 k' を全エンティティと入れ替えてスコアを計算し, 得られるランキング中で $\phi_{Comp}(i', j', k'; \Theta)$ が上位 K 位に入る割合を指す. MRR は平均逆順位とも呼ばれ, 前述のランキングにおける $\phi_{Comp}(i', j', k'; \Theta)$ の順位の逆数の平均として表される.

ラベル文字列を用いた知識グラフ埋め込みの効果を検証するため, 展開文字数 $N = 1, 2, \dots, 7$ としてトレーニングデータを展開した場合に加え, ベースラインとして展開を行わない場合の Hits@K, MRR を算出した. 展開時のトレーニングデータに関する詳細を表 1 の右部に示す. テストデータ数は 4464 個である. 全ての場合で共通して ComplEx による埋め込みを適用し, 埋め込み次元は複素 200 次元としてロジスティック回帰による学習を行った. 学習率調整は [Trouillon 16] に倣い Adagrad を利用し, イテレーション数は 1000, 負例サンプリング数は 5 とした. 指標算出時のランキングについて, 展開して得られた部分文字列エンティティを含むトリプルはランキングの対象外とした. 同様に, 既にテストデータに含まれるトリプルに関してもランキングの対象外とした.

4.3 知識グラフ埋め込みによる補完精度

表 1 の左部に 5 分割交差検証の結果を示す. 展開文字数 N に関わらず, ベースラインに対して全ての指標が上昇していることがわかった. Hits@3, 5, 10 については, N に比例して上昇していた. 一方で Hits@1 は, $N = 3$ の時に最大値をとり, $N = 4, 5, \dots$ と増加するにつれ減少する傾向が見られた. MRR に関しても同様に $N = 3$ の時が最大であった.

この分割したデータで補完された実例を示す. 例えば (たらか茶漬け, is_a, お茶漬け) といった部分文字列がそのまま結びついた単純な補完例があり, 2987 位から 1 位に改善されて

いた. また (ホットケーキ, 味, 甘い) といったように, 部分文字列で結びついたエンティティの情報から改善されたと考えられる例もあり, 8 位から 1 位に改善されていた.

4.4 リレーション毎の補完精度

リレーション毎の精度を調査した. ある分割データに対し, 4.2 節で述べた方法に従って算出された Hits@K をリレーション毎に分類して算出した結果を表 2 に示す. 表 1 の結果 (Hits@1) を考慮し, $N = 3$ の場合とベースライン設定の 2 設定の比較として示す. Hits@3, 5, 10 の精度が悪化しているリレーションは無く, 全体としてはベースライン以上の精度を保持していることがわかった. 『is_a』や『料理種』といったような, 階層構造やタイプ等を表しているリレーションは, ベースラインに対し精度が大きく上昇していた. またそれ以外にも, 『味』や『温度』といったリレーションに関しても同等の精度上昇が確認できた.

一方で, ベースラインに対して大きな改善がされていないリレーションもあった. 出現数が少ないリレーション (100 以下) では精度が改善されていないものがほとんどである. また『別名』というリレーションでは, Hits@3, 5, 10 はベースラインと同程度の精度であるものの, Hits@1 の値が大幅に減少するという結果になった. 本質的にラベル文字列を取り扱うリレーションである『別名』にとって, 部分文字列のリンク情報がノイズとなってしまった可能性が考えられる. しかし精度の絶対値自体は他のリレーションに比べ比較的良好である. これはデータの特性に起因したものであり, 使用データ上で基本的に別名というリレーションがエンティティ間に双方向に張られる様に取り扱っており, 補完が容易になっているためと考えている. 出現数も多く重要なリレーションである『材料』の精度も大きな改善はされておらず, 精度の絶対値も低いことがわかる. 材料のように 1 対多の関係を表すリレーションは, 1 対 1 の関係を表すリレーションに比べて, ランキング上位に正しいトリプルが来るよう学習するのが容易ではないためだと考えられる.

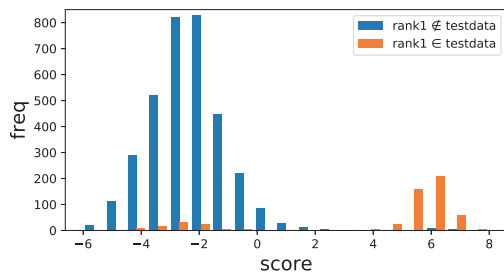


図 3: Baseline 設定でのスコア（確信度）の分布

5. 質問の精度と実例

5.1 スコア分布としきい値

2.1 節で述べた知識獲得を行う対話システムの枠組みにおける、ラベル文字列を用いた知識グラフ埋め込みの効果を検証する調査を行った。

推測がうまくいく場合、すなわちランキング上で 1 位に順位付けられるトリプルがテストデータ中に存在する場合のスコア（信頼度）の分布を調査する。つまり、テストデータ中に存在するトリプルは正例のみであるため、これが 1 位に順位付けられた場合、補完結果は正解であったとみなせる。4.4 節と同様に、知識グラフのある分割データを用いて、 $\mathcal{H}'$ 中のトリプルに関して順位を算出し、1 位のトリプルがテストデータにあるかどうかで分類して各々のスコア分布を確認した。

図 3, 4 に順位が 1 位だったトリプルのスコアで作成したヒストグラムを示す。このヒストグラムは、1 位と判定されたトリプルがテストデータ中に存在するかどうかで振り分けて度数を計測している。図 3 はベースライン設定、図 4 は $N = 3$ で展開した設定での結果である。テストデータに存在したトリプルが、部分文字列の展開によって増加していることが見てとれる。また、増加したトリプルは -2 付近を中心としたスコアを持っていることがわかる。スコアが 4~7 付近のトリプルはそのほとんどはリレーションが『別名』である。4.4 節でも述べたが、『別名』はエンティティ間に双方向に張られているリレーションであり、補完が容易なためスコアが比較的高くなったと考えられる。

表 3 に、適合率としきい値の関係を示す。ここでの適合率は、1 位に順位付けられたトリプルの中でテストデータ $\mathcal{H}'$ 中に存在し、しきい値以上のスコアを付与されたものの割合を表すもので、表の 3, 4 列目には割合ではなくその個数を示している。全データ数は 4464 個である。この適合率は、 $(i', j', k') \in \mathcal{H}'$ に対応する $(i', j', ?)$ を質問リストとして知識獲得を試みた時に、行う質問が正しい割合として捉えることができる。またしきい値以上かつテストデータ中に存在するトリプル数は、2.1 節で述べた知識獲得の枠組みを通して獲得できるトリプルの数と擬似的に捉えられる。適合率が高いほど正確に、表 3 中のトリプル数が多いほどより多くのトリプル（知識）の獲得が期待ができる。同じ適合率の下でそのトリプル数を見ると、適合率が 0.8 の場合は Baseline がやや上回っており、他の適合率設定では $N = 3$ が上回っている。高い適合率の下でこのトリプル数を増加させることに対し、ラベル文字列を用いる大きな効果は見られず、出来るだけ間違った質問をせず多くの知識を獲得するという目標に対しては課題が残る。

5.2 質問例

提案手法で生成できる質問例について述べる。学習方法は 4.2 節と同様であるが、知識グラフを分割せず、すべてを学習

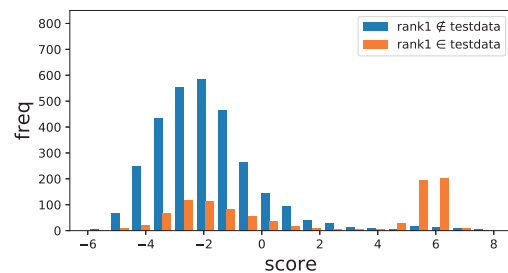
図 4: $N = 3$ とした時のスコア（確信度）の分布

表 3: 適合率としきい値の関係。BL は Baseline を表す。

適合率が x 以上	しきい値以上かつ $\mathcal{H}'$ 中に存在する トリプル数		適合率を 満たすしきい値	
	$N=3$	BL	$N=3$	BL
0.8	453	468	2.034	0.210
0.7	472	469	0.906	-0.333
0.6	497	472	0.160	-0.710
0.5	550	477	-0.592	-1.066
0.4	619	480	-1.280	-1.451

に用いた。ベースラインと $N = 3$ で展開した場合の 2 設定を取り上げる。

例えば（月見団子、味、甘い）という事実について、質問リストを（月見団子、味、?）として順位とそのスコアを算出した結果、ベースライン設定では 5 位、-4.00 で、 $N = 3$ とした場合は 1 位、-1.68 であった。「月見団子」と「甘い」が部分文字列の展開によって結びついたわけではないが、「団子」を末尾に持つ他のエンティティの味に関する情報を元に順位とスコアが改善されたものと考えられる。また（五目炊き込みご飯、温度、あたたかい）という事実に関しても同様に順位とスコアを算出すると、ベースライン設定では 7 位、-4.13 で、 $N = 3$ とした場合は 1 位、-1.66 であった。この例も、「ご飯」を末尾に持つ料理との関連性が上がったことが改善の理由だと推察出来る。

6. まとめ

本稿では、知識グラフの埋め込み表現を用いた対話システムにおける知識獲得のための質問生成の枠組みについて述べた。またラベル文字列を用いた知識グラフ埋め込みの補完精度の改善を提案し検証した。リレーション毎の補完精度、しきい値設定に関しても言及した。

今後の課題は、ランキング評価におけるルールの記述や対話内容からの質問リスト作成の枠組み設計、能動学習の導入などが挙げられる。

参考文献

- [Angles 08] Angles, R. and Gutierrez, C.: Survey of Graph Database Models, *ACM Comput. Surv.*, Vol. 40, No. 1, pp. 1:1–1:39 (2008)
- [Kadlec 17] Kadlec, R., Bajgar, O., and Kleindienst, J.: Knowledge Base Completion: Baselines Strike Back, in *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pp. 69–74 (2017)
- [Trouillon 16] Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., and Bouchard, G.: Complex Embeddings for Simple Link Prediction, in *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, pp. 2071–2080 (2016)
- [藤岡 18] 藤岡 勇真, 林 克彦, 中野 幹生, 駒谷 和範: 対話システムにおける知識グラフの埋め込み表現を用いた応答生成の試み, *SIG-SLUD*, Vol. B5, No. 02, pp. 88–89 (2018)

会話によるニュース記事伝達のためのトリビアの獲得と活用

Acquisition and Utilization of Trivia for Conversational News Contents Delivery

高津 弘明^{*1} 松山 洋一^{*1} 本田 裕^{*2} 藤江 真也^{*1*3} 小林 哲則^{*1}
Hiroaki Takatsu Yoichi Matsuyama Hiroshi Honda Shinya Fujie Tetsunori Kobayashi

^{*1}早稲田大学 Waseda University ^{*2}本田技術研究所 Honda R&D Co.,Ltd. ^{*3}千葉工業大学 Chiba Institute of Technology

Trivia has characteristics that draw people's interests and has been used as a strategy for attracting interests in contents in various conversational applications, such as entertainment and educational sites. In this paper, we propose a model that distinguishes trivia based on the surface representation of sentences. We incorporated the trivia presentation function into the spoken dialogue system we are developing and assessed the effect of presenting trivia in the introduction part of the topics in the news transmission task. As a result of the experiment, we confirmed that it is effective for enhancing users' interests in topics and promoting the feeling of rapport with system although we were not able to find significant improvements in the number of users' feedback and the Efficiency of Information Transfer (EoIT).

1. はじめに

文の表層表現に基づいてトリビアを識別するモデルを提案する。また、ニュースを伝える会話システムにおいて、話題の導入部分でトリビアを提示することの効果を検証する。

我々はニュース記事のようなまとまった量の情報を効率的に伝達する会話システムの開発を行っている [高津 18]。ここで「効率的」とは、伝達対象となる記事の中から、ユーザにとって不要な情報を除き、必要な情報だけを伝えることを意味する。我々のシステムの特徴は、あらかじめ主計画、副計画と呼ぶ複数のシナリオを用意しておき、このシナリオに沿って会話を進めることで、リズムの良い会話を実現するうえで必須となる迅速な応答を可能としたところにある。主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザからのフィードバックを理解し、必要に応じて副計画に遷移して補足情報を提示する。このようにユーザの興味や理解状態に応じて提示する情報を柔軟に切り替えながら会話を進めていく仕組みを持つ。一方で、高い情報伝達効率 (EoIT; Efficiency of Information Transfer) を実現するにはユーザからのフィードバックが必要不可欠である。本研究ではフィードバックに寄与する要素としてユーザのコンテンツに対する興味とシステムに対する親密感に着目する。

トリビアには人の興味を引く性質があり、従来よりコンテンツに対する興味を誘引する戦略として、会話主体のエンターテインメント (ゲーム、クイズショーなど) や教育現場などでトリビアが利用されてきた [Brown 84, Palincsar 84, ProMarketing Wizard 13, Voices Heard Media 13]。

人の興味を引くということに関して、落語の世界では聴衆を「話を聞かせる」状況に置くために、話の「マクラ」を導入部分に持つてくることが多い [広瀬 10]。ここでマクラは、次に来る話題への説明的な導入の機能を果たし、話し手と聴衆の間の親密感 (ラポール) の構築に寄与していると考えられる。

本研究では、話題の導入部分でのトリビアの提示がコンテンツに対する興味の誘引だけでなく、システムに対する親密感 (ラポール) の増進にも効果があると仮定する。興味や親密感が改善されることで、ユーザのフィードバックが増加し、情報伝達効率や継続利用可能性の向上が見込めるか、実験を通して検証する。

2. 関連研究

2.1 トリビアの自動獲得

トリビアには、人の興味を引く性質があり、ゲームなどのエンターテインメントや教育現場などで利用されてきた [Brown 84, Palincsar 84, ProMarketing Wizard 13, Voices Heard Media 13]。例えば、Brown と Palincsar らは、読解力の指導においてトリビアは学生の注意をコンテンツに向けさせる効果があると指摘している [Brown 84, Palincsar 84]。

トリビアを自動で獲得する手法に関する研究やトリビアを対話システムの発話生成に応用した研究がある。新名らは、ユーザを対話に引き込む効果を期待し、料理についての雑学を発話する対話システムを開発した [新名 17]。また、トリビアを含む文を「ある物事についての瑣末な知識であり、かつ人の興味を引くような内容を含む文」と定義し、主語の IDF や分散表現、主語と文中の名詞との類似度などを素性として文のトリビア度合いを回帰やランキング学習によって推定する手法を提案した [新名 18]。太田らは、対話を盛り上げるような発話の生成を目的として、Wikipedia の各文を TF-IDF、語の共起、文の長さによってスコア付けし、意外性のある文の抽出を行った [太田 09]。Prakash らは、固有表現の種類や Superlative Wordsなどを素性として Rank SVM を学習し、トリビアらしさのランキングを行った [Prakash 15]。Tsirel らは、トリビア度を人物とカテゴリの類似度の低さから計算される surprise とカテゴリに含まれる全ページ間の平均類似度の高さから計算される cohesiveness の積で定義し、人物の Wikipedia ページに付与されたカテゴリからトリビア度を計算する手法を提案した [Tsirel 17]。Fatma らは、(entity, predicate, object) の三つ組からなるデータを入力として、CNN でトリビアかどうかを識別する手法を提案した [Fatma 17]。また、人手で設計した特徴量と組み合わせることにより高い精度でトリビアを識別できることを示した。

本研究でも、従来トリビアと呼ばれてきたものと同様、トリビアを「ある物事についての瑣末な知識であり、興味をそそるもの」と定義する。また、Fatma らと同様にトリビアか否かを識別する 2 値分類問題として定式化するが、本研究では、文の単語系列そのものを入力として、双方向 LSTM と self-attention [Lin 17] からなるモデルによって識別を行う。そして、ニュース伝達タスクにおいて、トリビアの提示がユーザのコンテンツに対する興味を誘引する効果があると仮定し検証する。

連絡先: 早稲田大学 理工学術院 知覚情報システム研究室
〒162-0042 東京都新宿区早稲田町 27
E-mail: takatsu@pcl.cs.waseda.ac.jp

2.2 会話のマクラと話し手一聞き手間のラポール形成

人間同士のインタラクションを円滑にし、協働的なタスクの達成率を向上させるためには、対話相手同士の繋がりや親密さの感覚「ラポール (rapport)」が十分に形成されていることが重要であると、社会心理学や人工知能などの分野で指摘されてきた。Spencer-Oatey はラポールを対話の調和度合いと円滑さであると定義した [Spencer-Oatey 05]。特に、交渉 [Zhao 18], カウンセリング [Kang 12], 教育 [Zhao 14] など、相互の親密度が特定のタスクの達成に貢献するようなドメインでラポールの効果が報告されている。近年、ラポールの機能を部分的に実装した会話エージェントが提案されている。Pecune らは、情報推薦型のバーチャルアシスタントにラポール機能を導入し、ラポールの度合いと情報の受諾率などについて検討している [Pecune 18]。

ところで、「聴衆に話を聞かせるメディア」であるところの落語は、伝統的に「マクラ」、「本題」、「落ち」を基本構造として持つ。一般的に、落語家はいきなり落語の演目に入突することはほとんどなく、聴衆に語りかける雰囲気をつくるために、まずマクラとして、挨拶をしたり、世間話をしたり、軽い小咄を披露したりしてから本題に入っていく。マクラは本題への導入部であり、話し手と聴衆の心理的距離感を縮め、本題に関連する話題で聴衆の意識を物語の現場に引きつけるような効果があるとされている [広瀬 10]。このように、効果的に「話を聞かせる」状況をつくるために、マクラは話し手と聴衆の間のラポールの構築に寄与していると考えられる。

以上のことから、本研究では冒頭でのトリビアの提示がユーザのシステムに対する親密感を増進する効果があると仮定し検証する。

3. トリビアの識別

双方向 LSTM と self-attention [Lin 17] からなるモデルによってトリビアか否かを識別する手法を提案する。モデルの全体像を図 1 に示す。

3.1 トリビア識別モデル

識別対象の文が n 個の単語系列 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ から構成されているとする。 $\mathbf{x}_i$ は i 番目の単語の one-hot ベクトルである。まず、各単語を埋め込み層で d 次元に圧縮し、得られた単語の埋め込み表現を双方向 LSTM に入力する。ここで、LSTM の隠れ層の次元を u とする。

$$\mathbf{e}_i = W_{xe} \cdot \mathbf{x}_i \quad (1)$$

$$\vec{\mathbf{h}}_i = \text{LSTM}(\vec{W}_{eh} \cdot \mathbf{e}_i + \vec{W}_{hh} \cdot \vec{\mathbf{h}}_{i-1}) \quad (2)$$

$$\overleftarrow{\mathbf{h}}_i = \text{LSTM}(\overleftarrow{W}_{eh} \cdot \mathbf{e}_i + \overleftarrow{W}_{hh} \cdot \overleftarrow{\mathbf{h}}_{i+1}) \quad (3)$$

順方向 LSTM の出力と逆方向 LSTM の出力を連結し、それらを n 単語分並べて得られる行列を H とする。

$$\mathbf{h}_i = [\vec{\mathbf{h}}_i || \overleftarrow{\mathbf{h}}_i] \quad (4)$$

$$H = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n) \quad (5)$$

この行列 H に対して attention を計算するが、長い文において複数箇所 (r 箇所) に注目できるように $r \times d_a$ 次元の行列 W_{s2} を掛ける (d_a はハイパーパラメータ)。

$$A = \text{softmax}(W_{s2} \tanh(W_{s1} H^T)) \quad (6)$$

得られた attention 行列 A を入力文の隠れ層の行列 H と掛けあわせて文の埋め込み行列 S を得る。この $r \times 2u$ 次元の行列 S を $2ru$ 次元のベクトル $\mathbf{s}$ に変換し、文の埋め込みベクトルを得る。

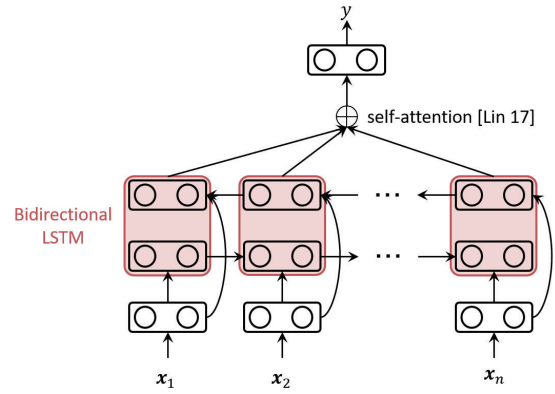


図 1: トリビア識別モデル

表 1: トリビアデータセットの統計

	訓練セット	開発セット	テストセット
正例	2500	100	126
負例	2500	100	126

表 2: トリビアと非トリビアの例

トリビア	非トリビア
童話「ウサギとカメ」で 2 匹が競争した距離は 2 8 0 m 3 3 c m 未満	最初は馬の姿で現れるが、命じられれば人間の姿になる
アルゼンチンには、昔の丸の内線が走っている	また、自動車関連メーカーにはより安全な自動車の開発を促す
ミツコビナマケモノは首を最大 2 7 0 度回転させられる	ネオジム塩を成分とする魴物の記事も含ます

$$S = AH \quad (7)$$

$$\mathbf{s} = [S[1] || S[2] || \dots || S[r]] \quad (8)$$

最後にこのベクトルをシグモイド関数を活性化関数とする出力層に与え、トリビアである確率を計算する。

$$y = \sigma(W_{sy} \mathbf{s} + b) \quad (9)$$

3.2 評価実験

Web から収集したトリビアを正例、Wikipedia からランダムに抽出してきた文を負例としてデータセットを作成し、トリビア識別モデルの性能を評価する。

3.2.1 実験設定

Web から収集したトリビアを正例、Wikipedia からランダムに抽出してきた文を負例 (非トリビア) とする。負例を抽出するにあたり、トリビアらしい文が含まれないように負例は次の 3 つの条件を満たす文の中からランダムに抽出した。(1) 最初の段落でない、(2) 段落の先頭の文でない、(3) 記事タイトルを含まない。作成したトリビアデータセットの統計を表 1 に示す。また、トリビアおよび非トリビアの例を表 2 に示す。

モデルのネットワーク構造として、埋め込み層の次元 d を 100、LSTM の隠れ層の次元 u を 100、attention の次元 d_a を 100、attention の数 r を 5 に設定した。訓練セットにおいて頻度が 2 以下の単語は特殊記号 (unk) に置き換え、入力語彙サイズを 4790 に制限した。

表 3: トリビアと非トリビアの識別精度 (%)

Accuracy	「トリビア」 識別精度	「非トリビア」 識別精度
97.2	97.6	96.8

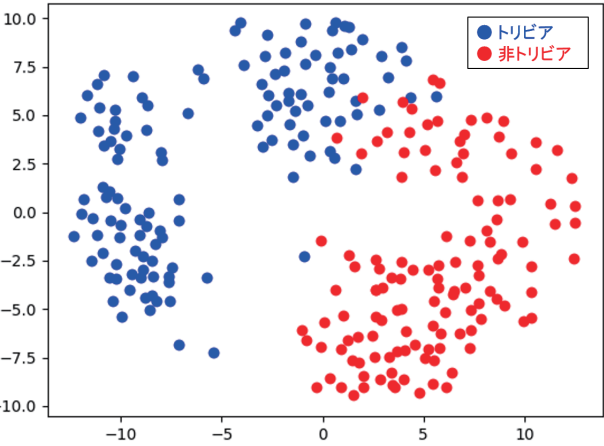


図 2: テストセットにおけるトリビアと非トリビアの埋め込み表現 s の分布を多次元尺度構成法で次元圧縮し可視化した結果

3.2.2 実験結果

テストセットにおけるトリビアと非トリビアの識別精度を 3 に示す。この結果から、提案モデルによって 97%と高い精度でトリビアか否かを識別できることがわかった。

テストセットにおけるトリビアと非トリビアの埋め込み表現 s の分布を多次元尺度構成法で次元圧縮し可視化した結果を図 2 に示す。この結果から、トリビアと非トリビアの境界がある程度明確に分かれるように学習できていることがわかった。

3.3 Wikipedia からのトリビアの獲得

学習した識別モデルを Wikipedia の記事に適用しトリビアを獲得する。ここでは、文脈依存性の高い文を回避するために以下の条件を満たす文に対して識別を行った。(1) 最初の段落である、(2) 段落の先頭の文である、(3) 記事タイトル+「は」を含む文である。Wikipedia の約 100 万記事に適用し、約 10 万文のトリビアを獲得した。

4. トリビアの会話システムへの導入効果の検証

ニュースを伝える会話システム [高津 18] においてトリビアを提示する効果を検証する。話題の導入部分でその話題に関連したトリビアを提示することで、ユーザのコンテンツに対する興味やシステムに対する親密感が向上すると仮定する。さらに、それによって、ユーザのフィードバックが増加し、情報伝達効率や継続利用可能性が向上するか、実験を通して検証する。

4.1 仮説

- 以下の仮説について検証を行う。
- 仮説 1: トリビアがトピックへの興味の増進に寄与する
 - 仮説 2: トリビアがラポール形成に寄与する
 - 仮説 3: トリビアがフィードバック数の増加に寄与する
 - 仮説 4: トリビアが情報伝達効率の向上に寄与する
 - 仮説 5: トリビアが継続利用可能性の向上に寄与する

表 4: 冒頭でのトリビア提示の例

発話タイプ	発話内容
トリビア	そういえば、天気予報の 177 って、昔は大隈重信の自宅の電話番号だったらいいよ
話題遷移	それで、その天気予報に関するニュースなんだけど
主計画	6 月から天気予報が大きく変わるみたいだよ 気象庁が計算能力を約 10 倍に高めたスーパーコンピュータを新たに導入するんだって
...	...

表 5: アンケート項目

ID	質問内容
I1	～の話について、対話しているうちに興味が強まった
C1	自分とシステムが同調していたと感じた
C2	システムと対話しているとき、あまり心地よくないことが言えなかった
C3	対話の間、フラストレーションがたまった
M1	対話システムが自分の言っていることに注意を向けてると感じた
M2	対話システムがユーザに対して敬意を払っていると感じた
P1	対話システムがフレンドリーと感じた
P2	対話システムに温かみを感じた
P3	対話システムは自分を気遣ってくれていると感じた
R1	このシステムをまた使用したい

4.2 比較条件

- 以下の 2 つの条件を比較する。
- 条件 1: ベースライン (主計画+副計画)
 - 条件 2: 条件 1 + トリビア提示機能
- 条件 2 における会話冒頭のトリビア提示の例を表 4 に示す。なお、条件 1 では冒頭の 2 行がなく主計画から会話が始まる。

4.3 実験設定

各条件ごとに、本システムを始めて利用する大学生 8 人ずつ、3 つのニューストピック (平均 33.7 文、主計画サイズ 9 文) について本システムと会話させ、会話終了後にアンケート調査を行った。

アンケート項目を表 5 に示す。I1 は各トピックについての興味に関する評価項目である。R1 は継続利用可能性に関する評価項目である。残りがラポールに関する評価項目である。

ラポールに関しては、Tickle-Degnen らが分類した同調性 (Coordination)、相互注意 (Mutual Attentiveness)、ポジティブ性 (Positivity) という 3 つの要素に従いアンケート項目を作成した [Tickle-Degnen 90]。ここで、C1,C2,C3 が同調性、M1,M2 が相互注意、P1,P2,P3 がポジティブ性に関する評価項目である。

各評価項目について 5 段階で評価させた (5 点: そう思う, 4 点: どちらかと言えばそう思う, 3 点: どちらでもない, 2 点: どちらかと言えばそう思わない, 1 点: そう思わない)。

また、被験者には、会話終了後にニュース記事を読ませ、各文について会話で伝えて欲しかったかどうかを 5 段階で評価させた (5 点: 伝えてほしい, 4 点: どちらかと言えば伝えてほしい, 3 点: どちらでもよい, 2 点: どちらかと言えばいらぬい, 1 点: いらない)。会話で提示できた 4 点以上の文の被覆率と会話で提示しなかった 2 点以下の文の除外率の調和平均から情報伝達効率を被験者ごとに計算した。

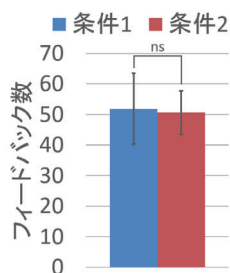


図 3: フィードバック数

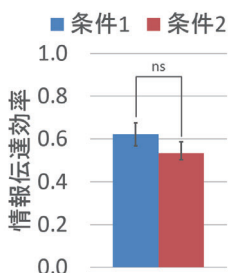


図 4: 情報伝達効率

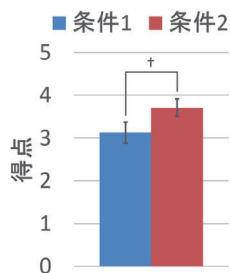
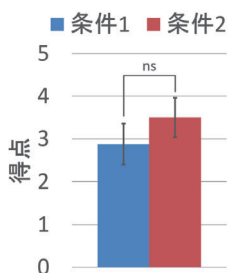
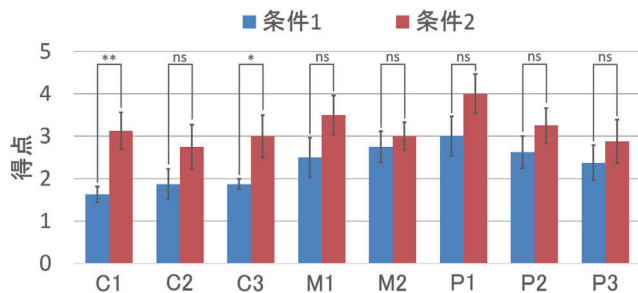
図 5: 興味増進効果 (†: $p < 0.1$)

図 6: 継続利用可能性

図 7: ラポール形成効果 (*: $p < 0.05$, **: $p < 0.01$)

4.4 実験結果

各条件について平均値を計算し、t検定による仮説検定を行った。被験者一人あたりの平均フィードバック数を図3に示す。被験者一人あたりの平均情報伝達効率を図4に示す。興味増進効果(I1)と継続利用可能性(R1)に関するアンケートの平均点を図5、図6に示す。ラポールに関する各評価項目の平均点を図7に示す。なお、C2とC3は同調的でないことを質問しているため、図は同調的であるほど値が高くなるように6から引いた値を示している。

フィードバック数と情報伝達効率、継続利用可能性については有意な差が得られず、仮説3と仮説4、仮説5の正しさは示されなかった。一方、興味増進効果とラポール形成効果(特に、同調性)については有意な差が得られ、仮説1と仮説2の正しさが示された。

5. おわりに

トリビアか否かを識別するモデルを提案し、97%と高い精度で識別できることを確認した。本研究では、文の表層表現のみを入力としてトリビアの識別を行ったが、先行研究で検討されているようなトリビアらしさを表現する特徴量も加えることでさらなる性能向上が見込めると考えられる。

検証実験の結果、情報伝達を目的とした会話システムにおいて、トリビアを話のマクラとして冒頭で伝えることにより、ユーザのトピックに対する興味とシステムに対する同調感が増進されることがわかった。すなわち「話を聞かせる会話メディア」

という新しいジャンルにおいて、トリビアの戦略的な使用が聞き手を効果的に「聞くモード」に置かせる方法として有効であることが示唆された。

しかしながら、本来の目的であった、高い情報伝達効率(EoIT; Efficiency of Information Transfer)を実現するためのフィードバックの促進効果については確認されなかった。

ユーザを「聞くモード」にしたうえで、他にどのような因子が直接的にフィードバックの促進に寄与するのかは、未だ解っていない。今回の検証実験の自由記述アンケートでは、そもそも期待した質問応答が得られなかった、音声合成の質が良くないと感じられる、間(ま)が悪いなどの、システム性能に対する信頼を疑う意見も散見された。今後の研究では、システム性能やラポール現象の網羅的な測定と分析を通してフィードバックを促進させるメカニズムをより系統的に明らかにしていく。

参考文献

- [Brown 84] Brown, A.L., Palincsar, A.S., and Armbruster, B.B.: Instructing comprehension-fostering activities in interactive learning situations, *Learning and Comprehension of Text*, pp. 255-286 (1984)
- [Fatma 17] Fatma, N., Chinnakotla, M.K., and Shrivastava, M.: The unusual suspects: Deep learning based mining of interesting entity trivia from knowledge graphs, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1107-1113 (2017)
- [広瀬 10] 広瀬和生: 現代落語の基礎知識, 集英社, ISBN 978-4-08-771365-7 (2010)
- [Kang 12] Kang, S.-H., Gratch, J., Sidner, C., Artstein, R., Huang, L., and Morency, L.-P.: Towards building a virtual counselor: Modeling non-verbal behavior during intimate self-disclosure, in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pp. 63-70 (2012)
- [Lin 17] Lin, Z., Feng, M., Santos, C.N., Yu, M., Xiang, B., Zhou, B., and Bengio, Y.: A structured self-attentive sentence embedding, in *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-15 (2017)
- [新名 17] 新名和也, 嶋田和孝: シズルワードから想起される料理の雑学を話す対話システム, 電子情報通信学会信学技報, Vol. 117, No. 367, NLC2017-42, pp. 77-82 (2017)
- [新名 18] 新名和也, 嶋田和孝: トリビア文抽出のためのトリビア度合いの推定, 電子情報通信学会信学技報, Vol. 118, No. 122, NLC2018-7, pp. 69-74 (2018)
- [大田 09] 太田知宏, 島海不二夫, 石井健一郎: 発話生成を目的とした Wikipedia から文抽出, 第 23 回人工知能学会全国大会論文集, 2G1-NFC5-11, pp. 1-4 (2009)
- [Palincsar 84] Palincsar, A.S. and Brown, A.L.: Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, *Cognition and Instruction* 1, No. 2, pp. 117-175 (1984)
- [Pecune 18] Pecune, F., Chen, J., Matsuyama, Y., and Cassell, J.: Field trial analysis of socially aware robot assistant, in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1241-1249 (2018)
- [Prakash 15] Prakash, A., Chinnakotla, M.K., Patel, D., and Garg, P.: Did you know? - Mining interesting trivia for entities from wikipedia, in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3164-3170 (2015)
- [ProMarketing Wizard 13] ProMarketing Wizard: Trivia game on facebook, <http://bit.ly/1HUhlRx> (2013)
- [Spencer-Oatey 05] Spencer-Oatey, H.: (Im)politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships, *Journal of Politeness Research. Language, Behaviour, Culture*, Vol. 1, No. 1, pp. 95-119 (2005)
- [高津 18] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会論文誌, Vol. 33, No. 1 (2018)
- [Tickle-Degnen 90] Tickle-Degnen, L. and Rosenthal, R.: The nature of rapport and its nonverbal correlates, *Psychological Inquiry*, Vol. 1, No. 4, pp. 285-293 (1990)
- [Tsourel 17] Tsourel, D., Pelleg, D., Guy, I., and Shahaf, D.: Fun facts: Automatic trivia fact extraction from wikipedia, in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 345-354 (2017)
- [Voices Heard Media 13] Voices Heard Media: Using trivia and quiz products to engage your customer, <http://bit.ly/1Htiv5V> (2013)
- [Zhao 14] Zhao, R., Papangelis, A., and Cassell, J.: Towards a dyadic computational model of rapport management for human-virtual agent interaction, in *Proceedings of the International Conference on Intelligent Virtual Agents*, pp. 514-527 (2014)
- [Zhao 18] Zhao, R., Romero, O.J., and Rudnicky, A.: SOGO: A social intelligent negotiation dialogue system, in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pp. 239-246 (2018)

画像キャプションに対する表現学習に向けた 敵対的生成ネットワーク

Generative Adversarial Networks
toward Representation Learning for Image Captions

阿部 佑樹
Yuki Abe

妹尾 卓磨
Takuma Seno

松森 匠哉
Shoya Matsumori

今井 倫太
Michita Imai

慶應義塾大学 理工学部
Department of Science and Engineering, Keio University

Captions generated from a single image are possibly different from each others as for representations (e.g. attention points or sentence expressions). However, a vast amount of image captioning datasets in the world have few or no annotations of latent variables. Learning latent variables of captions with no supervision is an important from perspectives of scalability and interpretability of conditional image captioning models. In this research, we propose a deep generative model to learn and leverage latent variables of image captions. In experiments, we used the task of image classification with several MNIST images and ground truth labels as down-scaled setting of image captioning, and we show that our proposed model acquired latent variables which represent sub-groups of labels.

1. はじめに

画像に対する適切な説明文（キャプション）を自動で生成することは画像キャプションングと呼ばれる。ひとつの画像に対するキャプションは、注目箇所の偏りや文章表現の違いにより複数通り存在することが考えられる。別の言い方をすると、キャプションは、画像および画像と独立したキャプションを特徴付ける要素（潜在変数）によって生成されたと考えることができる。潜在変数の利用や獲得は、コンテキストの考慮や分析といった実用上の有用性から、画像キャプションングにおける重要な課題のひとつとして位置付けられる。

これまでに、画像、キャプション、および潜在変数を用いた教師あり学習により、潜在変数の値に応じて生成するキャプションの特徴を変化させる手法が提案されてきている [1]。しかしながら、世の中に存在する画像キャプションのデータセットは、潜在変数がアノテーションされておらず未知であることが多い。

アノテーションのない訓練データから潜在変数を学習により獲得することは一般的に表現学習と呼ばれ、深層生成モデルを用いた手法が提案されてきている [2]。深層生成モデルを用いた画像キャプションに対する表現学習は、潜在変数による選択的な画像キャプションングの拡張性の向上に繋がる。また、画像のみからキャプションを生成する方法 [3] と比較して、意味的な潜在変数を用いて生成するキャプションを制御できることは、モデルの解釈性の観点から重要である。

本研究は、画像キャプションに対する表現学習に向けた、深層生成モデルのアーキテクチャの検討を行う。本提案モデルはテキスト生成モデル LaTextGAN [4] の拡張として表現される。画像キャプションングを画像で条件づけたテキスト生成として捉えることにより、conditional GAN [5] を参考に LaTextGAN を画像で条件付けする。また、InfoGAN [2] を参考に相互情報量に関する制約項を目的関数に加えることで、キャプションに対する潜在変数をアノテーションのない訓練データから獲得する。

キャプションの潜在変数の種類や区分は曖昧であるため、キャ

プションの特徴を捉えた潜在変数の定量的評価や定性的評価は困難である。本提案モデルの評価には画像キャプションングを直接扱わず、ある画像に対して複数の正解ラベルが存在する画像分類問題を、本提案モデルの有用性を示す検証課題として扱う。

2. 関連研究

2.1 Generative Adversarial Networks

Generative Adversarial Networks (GAN) は Goodfellow *et al.* [6] によって提案された深層生成モデルの学習フレームワークである。GAN は対立する 2 つのネットワークを利用する。生成ネットワーク G はノイズ変数 $z \sim p_z(z)$ をデータ $G(z)$ に変換する。識別ネットワーク D はデータが訓練データ $x \sim p_{data}(x)$ か G によって生成されたデータ $G(z)$ かを識別する。 G と D の学習は次式に示すミニマックスゲームによって行う。

$$\min_G \max_D \mathcal{L}_{adv}(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

2.2 LaTextGAN

LaTextGAN [4] は GAN をテキスト生成に応用した深層生成モデルである。LaTextGAN は Encoder-Decoder モジュールと GAN モジュールの 2 つのモジュールを利用する。Encoder-Decoder モジュールはテキストとテキストを表現する文章ベクトルの相互変換に利用される。Encoder-Decoder モジュールはオートエンコーダ [7] で構成され、学習はテキストの再構成誤差の最小化により行われる。GAN モジュールは文章ベクトルの生成に利用される。GAN モジュールは通常の GAN の枠組みで学習が行われる。訓練データにはテキストを直接用いるのではなく、事前学習済みの Encoder-Decoder モジュールから得られる文章ベクトルが用いられる。最終的な LaTextGAN の出力は、GAN モジュールの出力する文章ベクトルを Encoder-Decoder モジュールによりテキストに復号することにより得る。

2.3 conditional GAN

conditional GAN [5] は GAN を条件付き生成モデルに拡張したモデルである。ここで X を任意の種類の補足情報とする。

連絡先: 阿部 佑樹, 慶應義塾大学理工学部, 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1, E-mail: abe@ailab.ics.keio.ac.jp

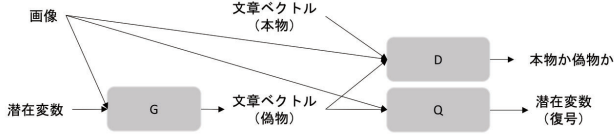


図 1: 本提案モデルの GAN モジュールの概要図。

補足情報には例えばクラスラベルやモダリティの異なる他のデータなどが挙げられる。条件付けは G および D の両方に追加の入力として X を与えることで行われる。

2.4 InfoGAN

InfoGAN[2] は GAN を表現学習に応用した深層生成モデルである。潜在変数を $c \sim p_c(c)$ とする。 $p_c(c)$ は任意の分布を用いる。 G はノイズ変数 z と潜在変数 c をデータ $G(z, c)$ に変換する。InfoGAN の目的関数は、式 1 で示される通常の GAN の目的関数に、潜在変数を再構成するネットワーク $Q(G(z, c))$ を用いた相互情報量制約項 $\mathcal{L}_I$ を加えた形で表現される。

$$\min_{G, Q} \max_D \mathcal{L}(G, D, Q) = \mathcal{L}_{adv}(G, D) + \lambda_I \mathcal{L}_I(G, Q) \quad (2)$$

$$\mathcal{L}_I = -\mathbb{E}_{z \sim p_z(z), c \sim p_c(c)} [\log Q(G(z, c))] \quad (3)$$

ここで λ_I はハイパーパラメータである。制約項 $\mathcal{L}_I$ の最小化は潜在変数 c とデータ $G(z, c)$ の間の相互情報量の変分下界の最大化に等しいため、訓練データの特徴付けの表現を潜在変数 c として獲得することが促される。

2.5 Variational AutoEncoder

Variational AutoEncoder (VAE)[8] はオートエンコーダの拡張モデルである。VAE のエンコーダは入力 x で条件づけられた事後分布 $q(z|x)$ で表される。すなわち VAE は入力 x の表現 z を分布で表現するモデルである。VAE では $q(z|x)$ が事前分布 $p(z)$ に近づくような制約が加えられる。VAE の目的関数は次式で表される。

$$\mathcal{L}_{VAE} = -\mathbb{E}_{q_\theta(z|x)} [\log p_\theta(x|z)] + D_{KL}(q_\theta(z|x) \| p(z)) \quad (4)$$

ここで θ は VAE のパラメータ、 D_{KL} はカルバックライブラ情報量である。第 1 項が再構成誤差に対応し、第 2 項が事後分布に対する制約に対応する。

3. 提案

本提案モデルは LaTextGAN を元に、Encoder-Decoder モジュールでの VAE の利用、および GAN モジュールの画像キャプショニング及び表現学習への拡張で表現される。

3.1 Encoder-Decoder モジュール

本提案モデルはテキストと文章ベクトルの相互変換に VAE を用いる。しかしながら、式 4 で表させる通常の VAE の目的関数は事後分布に対する制約が強すぎるため、テキストの再構成誤差が大きくなり相互変換に適さない。そこで制約係数 $0 < \lambda_{KL} < 1$ を用いて VAE の学習を行う。目的関数を次に示す。

$$\begin{aligned} \mathcal{L}_{VAE} = & -\mathbb{E}_{q_\theta(s|S)} [\log p_\theta(S|s)] \\ & + \lambda_{KL} D_{KL}(q_\theta(s|S) \| p(s)) \end{aligned} \quad (5)$$

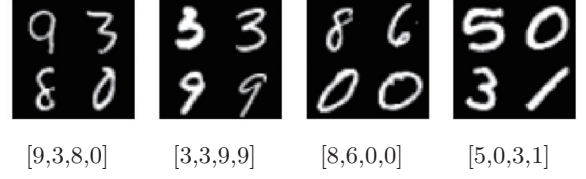


図 2: GridMNIST のサンプルの一例。

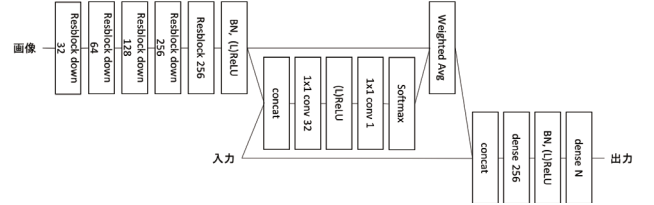


図 3: ネットワーク構成の概要図。BN はバッチ正規化、LReLU は LeakyReLU を示す。

ここで S はテキスト、 s は文章ベクトルである。事後分布 $q_\theta(s|S)$ には対角ガウス分布を用い、事前分布 $p(s)$ には標準正規分布を用いる。 $\lambda_{KL} = 0$ のときオートエンコーダの目的関数と一致し、 $\lambda_{KL} = 1$ のとき式 4 で示す VAE の目的関数と一致する。小さい λ_{KL} では再構成誤差の最小化が重要視される。

VAE の学習した文章ベクトル s を利用すると GAN モジュールの学習が安定する。文章ベクトル s は対角ガウス分布に従うノイズを含む。識別ネットワーク D の入力にノイズを加えることで GAN の学習を安定化する手法 [9] と同様に、訓練データの分布とモデル分布のサポートの重なりが増加する。また、生成ネットワーク G に対する近似誤差の緩和の効果が得られる。

3.2 GAN モジュール

図 1 に本提案モデルの GAN モジュールの概要図を示す。GAN モジュールは生成ネットワーク G 、識別ネットワーク D 、及び復号ネットワーク Q から構成される。全てのネットワークは画像 X を入力に受ける。 G は画像 X と潜在変数 c から文章ベクトル $\hat{s}$ を生成する。 D は画像 X に対する文章ベクトルが訓練データ s か生成データ $\hat{s}$ かを識別する。 Q は画像 X と文章ベクトル $\hat{s}$ を入力に受け G に入力された潜在変数 c を予測する。目的関数を次に示す。

$$\min_{G, Q} \max_D \mathcal{L}(G, D, Q) = \mathcal{L}_{adv}(G, D) + \lambda_I \mathcal{L}_I(G, Q) \quad (6)$$

$$\begin{aligned} \mathcal{L}_{adv} = & \mathbb{E}_{s \sim q_\theta(s)} [\log D(s, X)] \\ & + \mathbb{E}_{c \sim p_c(c)} [\log (1 - D(G(c, X), X))] \end{aligned} \quad (7)$$

$$\mathcal{L}_I = \mathbb{E}_{c \sim p_c(c)} [\|c - Q(G(c, X), X)\|^2] \quad (8)$$

ここで、式 8 の最小化は、 Q の出力がガウス分布に従うと仮定した場合の式 3 の最小化と同一である。画像キャプショニングを画像で条件づけられたテキスト生成として扱うことにより、 G は画像に対する適切な文章を生成するモデル、すなわち画像キャプションを生成するモデルとして学習が行われる。また、相互情報量制約 $\mathcal{L}_I$ により文章ベクトルに対する潜在変数 c の獲得が促される。

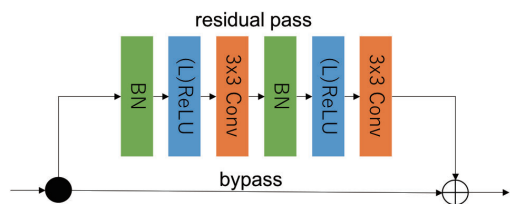


図 4: ResBlock の概要図。BN はバッチ正規化, LReLU は LeakyReLU を示す。ResBlock を用いてサイズを半分にする場合 (ResBlock down) は, bypass および residual pass の最後に平均プーリング層を挿入する。ResBlock を用いてチャンネル数を変化させる場合は, bypass の最初に 1×1 の畳み込み層を追加し, この畳み込み層と residual pass の最初の畳み込み層でチャンネル数を変化させる。画像に対して直接 ResBlock を適用する場合は, residual pass の最初の BN と (L)ReLU を取り除く。

4. 実験

潜在変数の評価の明確化と簡単化のため, 検証課題として複数の正解ラベルが存在する画像分類問題を用いる。画像キャプションにおいて, 潜在変数の値に応じて生成するキャプションの特徴を変化させることは, 画像分類問題において, 潜在変数の値に応じた特定のラベルを出力することと対応する。ひとつの潜在変数の値に対して複数の画像を用いたときの出力を評価することで, 潜在変数として獲得された表現を明らかにする。

4.1 データセット

複数の正解ラベルが存在する画像分類問題として, MNIST を元に GridMNIST を作成した。GridMNIST のサンプルの一例を図 2 に示す。GridMNIST は 64×64 の白黒画像と 4 つの正解ラベルを持った正解ラベル集合のペアから構成される。画像は 2×2 の格子状に配置された MNIST 画像であり, 正解ラベル集合の要素は画像中に存在する MNIST の正解ラベルである。MNIST の学習データとテストデータを元に, GridMNIST の学習データを 10000 件, テストデータを 1000 件作成した。

4.2 実装

Encoder-Decoder モジュールはクラスラベル L とクラスラベルを表現する潜在表現 l の相互変換に利用され, VAE を用いて構成される。クラスラベル L は 10 次元の Onehot ベクトルで表現し, 潜在表現 l の次元数は 2 次元とした。VAE のエンコーダおよびデコーダは 1 層の全結合層を用いた。制約係数 $\lambda_{KL} = 0.2$, バッチサイズ 64, エポック数 100 とし, 最適化手法に Adam[10] を用いた。

GAN モジュールの G , D , および Q のネットワーク構成を図 3 に示す。各ネットワークの計算処理は画像 X と追加の入力からある値を出力するという点で同一であり, Kazemi *et al.* [11] の提案するモデルを参考にした。学習の安定化のため, ネットワークの構成要素には, 図 4 に示す ResBlock[12], バッチ正規化 [13], G の活性化関数は ReLU[14], D および Q の活性化関数は負側の傾きが 0.2 の LeakyReLU[15] を用いた。また式 7 を負の対数尤度から最小二乗誤差に置き換えた [16]。すなわち G は $\mathbb{E}_{c \sim p_c(c)} [(D(G(c, X), X) - 1)^2]$ を, D は $\mathbb{E}_{s \sim q_\theta(s)} [(D(s, X) - 1)^2] + \mathbb{E}_{c \sim p_c(c)} [D(G(c, X), X)^2]$ を最小化するように学習する。潜在変数 c の分布 $p_c(c)$ は 3 次元の標準正規分布 $\mathcal{N}(0, I)$ とした。相互情報量制約の係数は $\lambda_I = 1.0$ とした。バッチサイズ 64, イテレーション数 400000

とし, 最適化手法に Adam を用いた。Adam のパラメータは, $lr = 0.0002$, $\beta_1 = 0.5$, $\beta_2 = 0.999$ とした。パラメータの更新は式 6 に示す最適化を 1 イテレーションごとに行った。

4.3 実験方法

ひとつの潜在変数の値に対して複数の画像を用いたときの出力を評価することで, 潜在変数として獲得された表現を明らかにする。ランダムに決定されたひとつの潜在変数の値に対して, 入力画像として GridMNIST のテストデータ 1000 件全てを用い, 本提案モデルの分類精度および出力ラベルの出現頻度比を評価する。ここで分類精度は, ある画像に対して出力したラベルが正解ラベル集合に含まれる場合を正解として計算される。分類精度は, いかなる潜在変数の値に対しても適切な出力を得ることができるかを評価する指標であり, 潜在変数とラベルが直接対応していないことを調べるために有用である。出力ラベルの出現頻度比は, ひとつの潜在変数の値に対する出力ラベルの傾向を調べるために有用である。出力ラベルの出現頻度比の比較手法には, 潜在変数を全てランダムに決定する方法を用いる。

4.4 実験結果

図 5 に実験結果を示す。異なるグラフでは異なる潜在変数の値を用いている。分類精度はいずれも 98% を超えており, 潜在変数とラベルが直接対応していないことがわかる。潜在変数を固定した場合にはラベルの出現頻度に偏りが見られる。例えば, 潜在変数の値 c_1 を用いた場合はラベル 0, 1, 2 を多く出力している。潜在変数の値 c_2 を用いた場合はラベル 3, 8 を多く出力しており, 潜在変数の値 c_3 を用いた場合はラベル 1, 7 を多く出力している。一方で, 潜在変数をランダムに決定した場合は全てのラベルを概ね均等に出力している。

分類精度を維持しながら特定のラベルを多く出力できるため, 潜在変数はラベルのサブグループに対応すると考えられる。例えば潜在変数の値 c_1 はラベル 0, 1, 2 を要素としたサブグループを表現し, 入力画像中の数字 0, 1, 2 を優先的に扱う。潜在変数の値 c_2 はラベル 3 と 8 のサブグループ, 潜在変数の値 c_3 はラベル 1 と 7 のサブグループと, 数字の形が似ているものはサブグループ化されやすいと考えられる。

5. おわりに

本研究では, 画像キャプションに対する表現学習に向けた, 深層生成モデルのアーキテクチャの検討を行った。画像キャプションを画像で条件づけたテキスト生成として扱うことで, 本提案モデルを既存の GAN のアーキテクチャの組み合わせとして構築した。本提案モデルの有用性を示す検証課題として画像分類問題を扱い, 本提案モデルが潜在変数としてクラスラベルのサブグループを獲得し, 生成するクラスラベルを選択的に変化させることが可能であることを示した。

将来研究として以下が挙げられる。画像キャプションは自然画像と自然言語を用いた問題であるのに対し, 今回扱った画像は自然画像ではない。画像キャプションに向け, 自然画像を用いた画像分類問題において本提案モデルの有用性を検証することは直近の課題のひとつである。また, 本提案モデルの出力の分析により潜在変数の表現を分析したが, 潜在変数の各次元や値が表現とどのように結びついているかは不明である。画像キャプションに対する表現学習の有効性の観点から, 潜在変数に対するより詳細な分析は必要である。モデルの解釈性の観点から, 人にとって解釈しやすいディスエンタングルな潜在変数を獲得することも重要である。

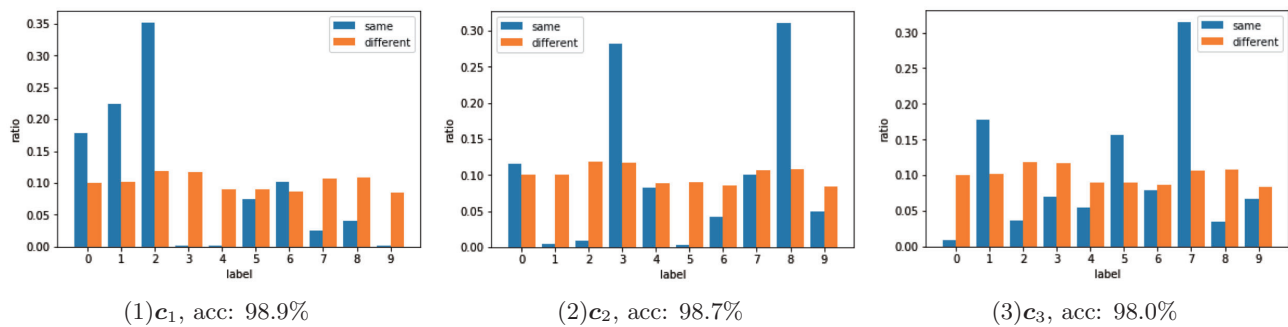


図 5: 入力画像 1000 件に対する出力ラベルの出現頻度比. 潜在変数の値を固定したものが same, 潜在変数の値がランダムなものが different である. 図 (1)(2)(3) では same は異なる潜在変数の値を用いている. acc は same に対する本提案モデルの分類精度を示す.

参考文献

- [1] Chuang Gan, Zhe Gan, Xiaodong He, Jianfeng Gao, and Li Deng. Stylenet: Generating attractive visual captions with styles. In *Proc IEEE Conf on Computer Vision and Pattern Recognition*, pp. 3137–3146, 2017.
- [2] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems*, pp. 2172–2180, 2016.
- [3] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3156–3164, 2015.
- [4] David Donahue and Anna Rumshisky. Adversarial text generation without reinforcement learning. *arXiv preprint arXiv:1810.06640*, 2018.
- [5] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 8, pp. 1798–1828, 2013.
- [8] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [9] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.
- [10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [11] Vahid Kazemi and Ali Elqursh. Show, ask, attend, and answer: A strong baseline for visual question answering. *arXiv preprint arXiv:1704.03162*, 2017.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [13] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [14] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.
- [15] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, Vol. 30, p. 3, 2013.
- [16] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 2813–2821. IEEE, 2017.

深層学習を用いた画像変換に基づく会話からの音声抽出

Speech extraction from conversation based on image-to-image translation using deep neural networks

高市 晃佑^{*1}
Kosuke Takaichi

片上 敬雄^{*2}
Yoshio Katagami

黒澤 義明^{*1}
Yoshiaki Kurosawa

目良 和也^{*1}
Kazuya Mera

竹澤 寿幸^{*1}
Toshiyuki Takezawa

^{*1} 広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences
Hiroshima City University

^{*2} 広島市立大学 情報科学部
School of Information Sciences
Hiroshima City University

We aim to separate sound sources by deep neural networks which has been active in recent years. We attempt to extract a certain human voice from usual conversation using the networks. We focus on image-to-image translation: pix2pix. The algorithm of pix2pix bases on purely procedure of the image processing. Therefore, we need an additional procedure, that is, we convert voice to spectrogram once. After that we perform to learn the networks to separate human voice, we especially pay attention to segmentation between the same sex and opposite sex. Form this point of view, we conducted two experiments using the sounds overlapped both sexes in this paper. Structure-Similarity (SSIM) index and color map representation were used as evaluation criteria. As a result, we confirmed the good extraction of the female voice from the one synthesized both sexes. However, we did not extract the female voice from same sex. Although we reached the conclusion that the separation did not work well, the generated voice seemed to be played naturally. This is not objective judgment. For this reason, it is our future work.

1. はじめに

これまでも音源分離の技術が開発されてきた。例えば、複数の音源を複数のマイクで録音したデータから、それぞれの音源を分離するときに使われるブラインド信号分離等である。しかし、音源が互いに独立であると仮定する等があり、これらが成り立たない場合、適切な結果が出ない。また、人間の声と音楽等の分解を行っていることが多く、人間の声同士の分解は例が少ない。

そこで、音源の重なりを近年盛んである深層学習を利用し分離することを目的とし、合成音から一人の音声の抽出を行う。また、音声はすべてスペクトログラムと呼ばれる画像に一度変換する。一度画像に変換する理由は、音声に関する研究よりも画像に関する研究の方が多く、画像に関する知見が使用できると考えたからである。そのため、本稿では、様々な画像に対応でき、画像から学習した通りに加工し画像を出力することのできる pix2pix を用いて、音源の抽出を行う。

2. 提案手法

先行研究に楽曲からボーカルと音楽を分離している研究がある[Jansson 17]。音楽と人間の声分離できるのであれば、人間同士の発話の重なりも分離できるのではないかと考えられる。

また、RNN (Recurrent Neural Networks) という手法で分離する研究も存在している[Huang 15]。音声を一度スペクトログラムに変換する必要はない。しかし、今回は画像に関する知見が使用できることが有用であると考え、画像を中心に学習を行う。

本稿では、音声をスペクトログラムへ変換し、pix2pix を用いて深層学習を行い、音声抽出を試みる。pix2pix は GAN (Generative Adversarial Network) を利用した画像変換アルゴリズムの一種であり、例えば、白黒写真からカラー写真を生成したり、地図から航空写真を生成したりできる。Generator と Discriminator の2つのモデルが敵対することで変換精度を高めていくモデルである。図1に pix2pix の変換例を示す。左が入力であり、右が出力(predict)である。

本稿は、まず女性と男性の声に現れる特徴の違いに着目する。人間の男性と女性の声には明確に高低差が存在する。具体的には、男性と女性の声では平均基本周波数が異なる。基本周波数とは、「周期的に生じる生体振動の時間間隔のうち、最短の間隔として与えられる基本周期」[森勢 18]であり、「人間が知覚する声の高さにおおむね対応する」[森勢 18]。本稿では、その差に注目する。すなわち、男性の声の平均基本周波数は約 120Hz であり、女性では 240Hz であると報告されている[寺澤 84]。同様に、男女の声の基本周波数と年齢の関係を報告している研究もある[粕谷 68]。その研究によると 14 歳以降では男女の声の平均基本周波数は大きく差がでている。

そこで本研究では、男女の声の平均基本周波数の違いから、男女の声には大きな特徴の差異があると仮定し、音声をスペクトログラムへ変換を行い、その違いを学習する。そして、同性間においてもある程度違いが表れると考え、同様に違いを学習する。基本周波数が異なる男女の声の分離は容易であると思われる。しかし、差が小さい女性同士の分離は困難であると思われる。

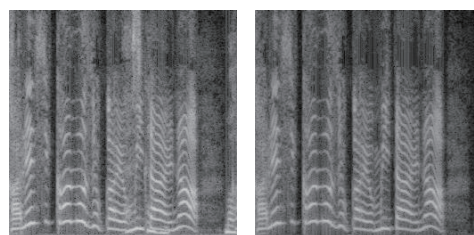


図1 pix2pix の入出力例

3. 実験と結果

本稿では、女性だけの音声と男女の合成音を変換したスペクトログラムをもとに、pix2pix による深層学習を行う実験(3.1 節)と女性だけの音声と女性同士の合成音を同様に変換し、深層学習を行う実験(3.2 節)を行う。

また、生成された画像は元の音声のスペクトログラムと比較して評価を行う。比較方法として SSIM (Structural Similarity) (式 1) とカラーマップ (図 2) を求め精度を測る。SSIM は人間が感じる違いに近い結果を返し、1 に近いほど同じと言える。

カラーマップは生成画像と元のスペクトログラム、2つの画像の差で表し、青(左)に近いほど未変換、赤(右)に近いほど過変換、中央値の緑に近いほど変換ができていていると言える。



図2 カラーマップの見方

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

3.1 男女の分離

実験の際に、Discriminator のネットワークパラメータを変更して3種の条件で実験を行い、各条件に対して validation, test を実行する。なお、本研究で使ったネットワークパラメータは、以下の通りである。

1. NLayerDiscriminator のカーネル数が 64
2. NLayerDiscriminator のカーネル数が 128
3. PixelDiscriminator のカーネル数が 64

1 と 2 の NLayerDiscriminator のネットワークと 3 の PixelDiscriminator のネットワークを図3と図4に記載する。なお、ここにある ngf はカーネル数のことである。

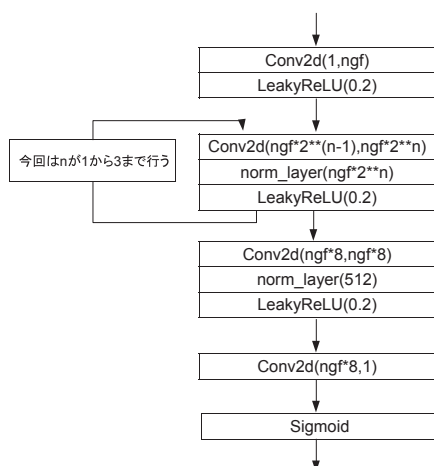


図3 NLayerDiscriminator のネットワーク

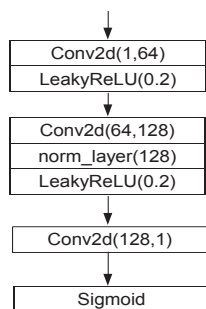


図4 PixelDiscriminator のネットワーク

NLayerDiscriminator と PixelDiscriminator の違いは畳み込みを行う際に、PixelDiscriminator は 1pixel 毎に読み込みと書き出しを行う点にある。

3 条件中で最も SSIM 値が高かった条件(0.806)について述べる。元音声のスペクトログラムと合成音のスペクトログラム(図5)、Discriminator が生成したスペクトログラムと SSIM を基にしたカラーマップ(図6)を示す。次に精度比較用に、最も精度が低かった条件(0.652)による、同一音声区間の生成画像とカラーマップ(図7)を示す。

図5、図6の比較から、元音声の特徴は正確に残しつつ、図5右側の左下付近、特に音が重なっている部分は変換ができた。このことは、図6のカラーマップからもわかる。しかし、図7の左

では左下部分は合成音が残っており、図7のカラーマップからも確認できる。つまり、ネットワークパラメータによって変換精度に大きく差があると言える。また、全体的に高周波域や無音区間の変換は過変換気味であることが、カラーマップから読み取れる。これにより SSIM 値が低下したと考えられる。

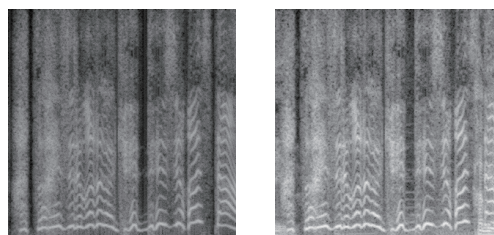


図5 元音声(左)と合成音(右)のスペクトログラム

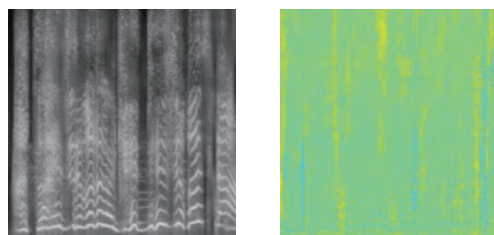


図6 生成されたスペクトログラム(左)とカラーマップ

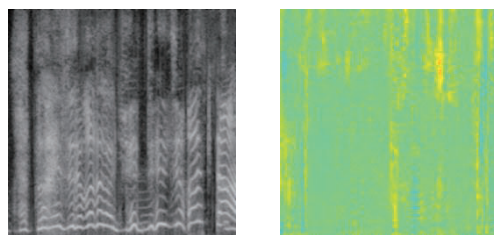


図7 別のネットワークの生成画像(左)とカラーマップ

3.2 女性同士の分離

女性同士の場合では、学習データを以下のように変え、学習を行った。

1. 単独の発話のみを学習
- 2.1. 対話からの発話を全体の 10% になるよう追加し学習
- 2.2. 対話からの発話を全体の 20% になるよう追加し学習
3. 無音区間を発話の前後に追加し、全体の 10% になるよう追加し学習

条件1では、学習データにラジオ CD から収集した単独の発話をスペクトログラムで 9,235 件収集し、これを学習した。他の条件ではこの 9,235 件に追加していく。各条件 1000epoch まで学習し、SSIM 値の推移とテスト結果にはそれぞれ最も良い世代の変換を記載する。図8、図9では、条件1(左上)、条件2.1(右上)、条件2.2(左下)、条件3(右下)とする。

SSIM 値については、図8より 100~300epoch においてピークを迎え、以降は緩やかに減少している。図9より、中央下の重なりが分離できていない等、変換できていない部分は同じであることがわかる。また、図9右下の条件3の結果は右上に過変換部分が見受けられる。これは右上が無音区間であると判断したことが原因だと考えられる。また、SSIM 値については、図8より、いずれも 0.6 ほどであり、右上以外の部分は変換できていると考えられる。しかし、男女間の実験と同様に高周波域の変換は過変換気味である。

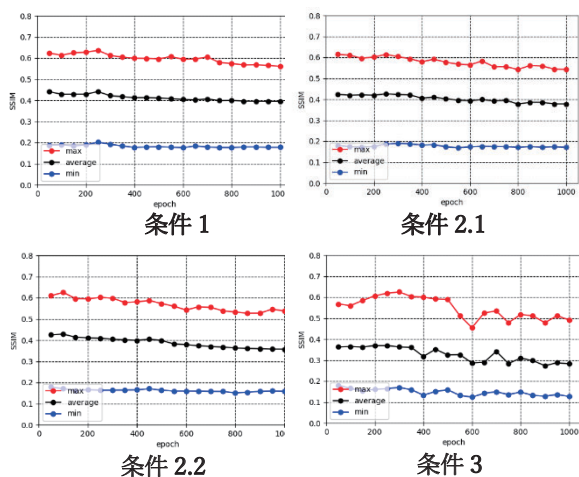


図8 各条件の SSIM 値の推移

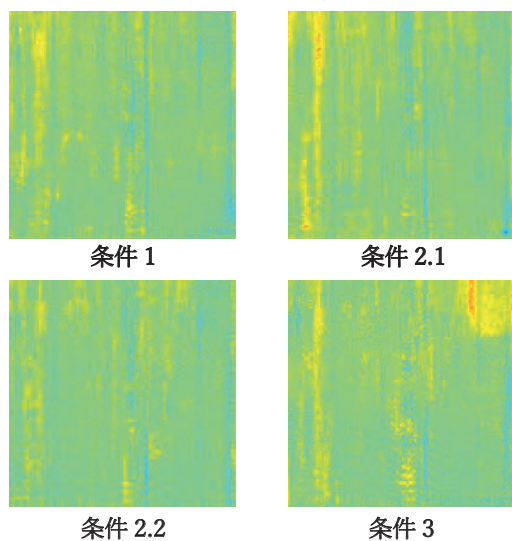


図9 各条件の SSIM 値が最も大きいテスト結果

Deep Recurrent Neural Networks for Monaural Source Separation, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 12, DECEMBER 2015

[Isola 17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, In CVPR 2017.

[Isola 17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, In ICCV 2017.

[Jansson 17] Andreas Jansson, Eric J. Humphrey, Nicola Montecchio, Rachel Bittner, Aparna Kumar, Tillman Weyde, Singing Voice Separation with Deep U-Net Convolutional Networks, Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR), 2017.

[粕谷 68] 粕谷 英樹, 鈴木 久喜, 城戸 健一, “年令, 性別による日本語 5 母音のピッチ周波数とホルマント周波数の変化,” 日本音響学会誌, 24(6), pp.355-364, 1968.

[森勢 18] 森勢 将雅, “音声分析合成, 一般社団法人 日本音響学会, pp47, コロナ社, 東京, 2018 年

[寺澤 84] 寺澤 り子 垣田 有紀 平野 実, “平均呼気流率, 声の基本周波数および声の強さの同時測定—正常成人男女各 30 名の成績—,” 音声言語医学, 25(3), pp189-207, 1984.

4. まとめ

本研究では, 重なった音源からの分離を目的とした. そのために本論文では pix2pix を用いてスペクトログラムから一人の発話を抽出する手法を提案した. 男女間では, SSIM 値とカラーマップからも抽出できていると判断できる. しかし, 同性間では, 黄色の箇所が残っており, 抽出できているとはいづらい. 学習データが少なかったことが原因と考えられる.

今後の課題として, 学習データに音声の前後ではなく途中に無音区間を考慮したデータを含めて再学習をし, 抽出の精度を高めていく必要がある. また, 学習に使用するネットワークに関しても, 適宜パラメータを調整し, 精度の向上を模索する必要がある. また, 音声を聞いた限り大きな差はないように感じられた点は補足しておく. 主観評価についても今後の課題である.

謝辞

この研究は, 国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」の補助を得ている.

参考文献

[Huang 15] Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, Senior and Paris Smaragdis, Joint Optimization of Masks and

機械学習を用いた構音障害分類における 音声特徴量間の耐ノイズ性能の比較

An anti-noise performance comparison between acoustic features in detecting voice pathology using machine learning

鈴木 康大^{*1*2} 篠原 修二^{*2} 馬目 信人^{*1*2} 朝長 康介^{*1*2} 光吉 俊二^{*2}
Kouta Suzuki Shuji Shinohara Nobuhito Manome Kosuke Tomonaga Shunji Mitsuyoshi

^{*1}ソフトバンクロボティクス株式会社
SoftBank Robotics Corp.

^{*2}東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

Developing communication robots requires to analyze human voice including various kinds of human biological information because the nonverbal information plays an important role in smooth communications between humans and robots. To analyze numerous voices available via the robot by using machine learning, we should take consideration of the existence of noises added to the voices. However, some acoustic features used for sensing human biological information is not designed for the noises. To validate the variation of the accuracy of classification when the voices includes the noises, we compare the classifications using voice indexes proposed for voice pathology estimation and using Mel-Frequency Cepstrum Coefficients(MFCC) in the classification problem of voice pathology as an early study. Experimental results show that classification using MFCC can detect voice pathology more precisely despite the noises while other voice indexes are adversely affected by the noises.

1. はじめに

人とのコミュニケーションを目的としたロボットが円滑に行動するためには、非言語情報を読み取ることが重要である。人から発せられた音声には多彩な生体情報が含まれており、健康状態や精神衛生状態を分類する試みが存在することから[光吉 2011] [Tokuno 2015] [Shinohara 2017], 声はコミュニケーションの際に重要な役割を果たすと考えられる。従って、コミュニケーションロボットの開発の際には、人から発せられた音声の解析が重要となる。

ロボットから取得される音声解析する手法として、機械学習が挙げられる。機械学習による音声分類の大まかな流れは以下の通りである。まず分類したい状態に対応する音声のある程度用意し、それぞれの状態における音声の特徴量を抽出する(前処理段階)。次に同じ状態の音声に現れる特徴量のパターンを見つける(学習段階)。このようにして状態ごとに得られたパターンを使って、新たに得られた音声に対してパターンマッチングを行い、最も適合するパターンに対応する状態を、その音声の状態と分類する(分類段階)。

ここで学習させる際の音声と分類したい音声の取得環境が異なる場合、人の状態が同じでも異なる特徴のパターンが見られる場合があり、分類性能に影響を及ぼす可能性がある。特に、ロボットなどを介して取得された音声の場合、音声を取得する周囲の環境を制御することが困難なため、周囲のノイズが音声に含まれることを考慮する必要がある。音声を用いて人の状態を分類することを目的とした機械学習において、いくつかの音声指標を導出して状態を分類することが提案されているが、これらは音声にノイズが含まれることを考慮していないものが多く、音声の一部の特徴しか用いないため、含まれるノイズ次第では指標に大きな影響を与える可能性がある[高野 2017]。

一方、近年計算機資源が増大することにより、深層学習をはじめとして膨大なデータを使った機械学習手法による分類が大きな成果を出している。これらは、得られた音声からヒューリスティックに指標を導出するのではなく、音声の様々な特徴を捉える分類を可能にする。ここで音声の特徴を表す普遍的な特徴量と

して、メル周波数ケプストラム係数(MFCC)が挙げられる。MFCCは音声認識でよく使われる特徴量であり、従来の提案されてきた指標よりも広い範囲で音声の特徴が含まれている。従って、音声にノイズが含まれている場合でも正確な分類が行えることが期待される。

本稿では、初期検討として、ノイズが含まれる環境下で音声から構音障害の有無を分類する問題を考え、音声からヒューリスティックに導出される指標と MFCC による分類を比較し、ノイズの強さ又は種類による分類性能の変動を検証した。

2. 関連研究

音声から構音障害の有無を分類するために、健常者の音声と比較して構音障害と診断された人のみに現れる音声指標がいくつか提案されている。Teixeira らは、音声長母音から導出される jitter, shimmer, HNR(Harmonic to Noise Ratio)が構音障害を持つ人と健常者と異なる分布を持つことを明らかにし、これを用いて構音障害を分類することを提案している[Teixeira 2014]。また、篠原らは、構音障害を持つ人の長母音からピッチを検出しづらいことを踏まえ、ピッチレートを用いて構音障害を分類することを提案している[Shinohara 2016]。

構音障害分類用に音声指標を設計するのではなく、音声特徴量を上手く使って機械学習を行い、構音障害を分類する試みもなされている。Londono らは、MFCCに加えて、音声波形から得られるリアプノフ指数や HNR を特徴量として、ガウス混合分布(GMM)及びサポートベクターマシン(SVM)で構音障害を分類する手法を提案し、マサチューセッツ工科大学が提供する構音障害データベース(MEEI)で 98.23%の分類性能を出した[Londono 2011]。Fang らは、MFCC のみを用いて深層学習で構音障害を分類することを提案し、Far Eastern Memorial Hospital が取得した構音障害データベースで、94.26%, 90.52%の分類性能を男性、女性の音声に対して出した[Fang 2018]。ALHUSSEIN らは音声をスペクトログラムに変換し、これを画像として扱い既存の画像分類用モデルである VGG16 や CaffeNet の一部分を使って学習し、先述の MEEI や saarbrücker が提供する構音障害データベース(SVD)で 93.9%の分類性能を出した[Alhussein 2018]。いずれの場合も音声にノイズが含まれていな

連絡先: 鈴木康大, ソフトバンクロボティクス株式会社/東京大学大学院工学系研究科, suzuki@coi.t.u-tokyo.ac.jp

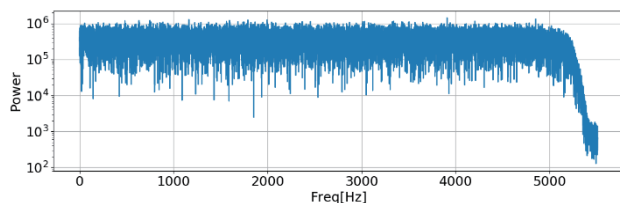


図 1. ホワイトノイズの周波数領域のパワー成分の分布

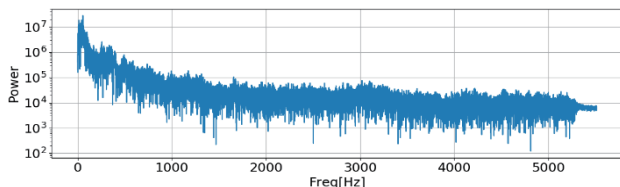


図 2. 車の走行音の周波数領域のパワー成分の分布

い場合に対する性能であり、これらが外部からノイズが混入した場合の性能については改めて検討する必要がある。

3. 実験概要

3.1 使用した音声データセット

今回、構音障害データベースとして、先述の MEEI を利用した。音声のサンプリングレートは 11kHz で、モノラルチャンネルで録音されている。このデータベースから、[Llorente 2006]に挙げられている、健常者 53 人、構音障害者 173 人の長母音 (ah の発音) を実験に利用するデータセットとして使用した。このデータセットは、健常者の発音は 3 秒間、構音障害者の発音が 1 秒間であり、年齢、性別などの偏りが健常者と構音障害者で同一であるため、健常者と構音障害者のデータ分布の偏りを最小限に抑えられている。このデータセットに含まれる音声を、ffmpeg-normalize を用いて音量の正規化を行った。ここでは、欧州放送連合が勧告している、EBU R128 に従った方法で行った。

3.2 付加するノイズ

先述のデータセットを 1:1 に分け、一方の音声にはノイズを付加させず、MFCC を用いた機械学習による訓練用のデータセットとして使用した。もう一方の音声は、ノイズを付加させて分類性能を評価するためのテスト用のデータセットとして使用した。付加するノイズの種類は 2 種類用意した。一つはホワイトノイズ (全周波数帯におけるレベルが均一なノイズ) で、もう一つは車の走行音 (低周波におけるレベルが高いノイズ) である。各ノイズの周波数領域でのレベルを図 1 図 2 に表す。ノイズの音源に対しても音声と同様の手法で正規化を行った。この正規化されたノイズ音源のレベルを変化させ、音声に付加させた。付加するノイズのレベルは 5 種類用意し、 $2 \times 5 = 10$ 種類の音声を用意した。

3.3 音声指標による構音障害分類

今回使用する音声指標として、jitter, shimmer, HNR を採用した。構音障害を持つ人の音声の特徴として、音声波形に乱れが生じることがある。この乱れ具合を定量的に評価できる指標として、jitter, shimmer, HNR が挙げられる。jitter は時間軸方向に発生する非常に短い変動の成分を表し、shimmer は音声の振幅方向に発生する変動の成分を表す。また、HNR は音声波形の基本周波数のパワーと、ノイズ成分のパワーの比を表す。今回この 3 つのパラメータを、音声特徴量抽出ソフトである OpenSMILE[Eyben 2010]を用いて音声ファイルから 3 つの特徴

量を抽出した。プリセットとして、avec2011.conf を使用し、音声ファイルごとに得られたパラメータのうち、中央値を表す値を採用した。この値を用いて健常者と構音障害者のラベル及び各指標の値を対応付け、ROC 曲線を描きその AUC の値を見ることで分類性能を評価した。

3.4 MFCC を用いた機械学習による構音障害分類

今回 MFCC を用いた分類を行うため、ニューラルネットワーク (DNN) および SVM の二つのモデルを構築した。二つのモデルの構築のために、ノイズが付加されていないセットで学習を行った。MFCC の導出にあたっては、Python のライブラリである librosa を利用した。音声から MFCC を抽出するにあたって、音声ファイルの先頭からウィンドウ幅 46 ms (512 サンプル) フレームシフト幅 12 ms (128 サンプル) と指定して切り出した。この切り出したウィンドウに対して、MFCC の 13 次元の係数と、その一次微分項の計 26 次元のベクトルを導出した。この 26 次元のベクトルを、DNN 及び SVM の入力として扱った。DNN の構築には chainer を利用した。DNN は、中間層として 300 ノードの全結合を 2 層用意し、出力層は 2 ノードの全結合を用意した。SVM は、python のライブラリである scikit-learn を用いた。パラメータとして、コストパラメータは 1.0、基底関数は RBF (動径基底関数) を指定した。この 2 つの学習済みモデルに対して、テスト用のデータセットから音声を入力すると、各ウィンドウに対して健常者又は構音障害者の分類が得られる。今回は全体のウィンドウ数に対して構音障害と分類されたウィンドウの数の割合を、その音声における構音障害のスコアとして利用し、このスコアを用いて健常者と構音障害者のラベル及び各指標の値を対応付け、ROC 曲線を描きその AUC の値を見ることで分類性能を評価した。

4. 実験結果と考察

図 3 図 4 にノイズレベルを変化させたときの AUC の値の推移、表 1 表 2 に AUC の数値を示す。音声指標を利用した分類において、jitter, shimmer, HNR などの音声指標を用いた場合の分類性能は、ノイズレベルを上昇させるといずれも低下した。ホワイトノイズを付加させた場合では、レベルが上昇するにつれて均一な下がり方をしており、車の走行音を付加した場合では、ノイズレベルが比較的高い時に性能が急激に悪化した。これは、この 3 つの音声指標はいずれも音の乱れが激しいほど高くなるため、ノイズの影響が無視できる際は健常者と構音障害者の声の乱れを捉えることが出来る指標となるが、ノイズレベルが大きい際は、ノイズから生じる音の乱れが指標に及ぼす影響が強くなるため、健常者と構音障害者の差が指標に含まれなくなる。そのためノイズレベルが高くなるにつれて、構音障害の分類性能が悪くなったと考えられる。

一方 MFCC を用いて DNN や SVM を用いた場合は、音声指標を用いた場合と比較して、ノイズレベルが上昇した場合でも分類性能が低下しなかった。DNN と SVM の間で性能及び性能の下がり具合に顕著な差は見られなかった。これらの結果は、MFCC から構音障害を持つ人に現れる音声特有の特徴を抽出し、テストデータにノイズが含まれていても読み取ることが出来る特徴を学習した結果が現れたものだと考えることが出来る。MFCC は音声特徴量として汎用的に用いることができるため、構音障害以外の状態分類に応用出来る可能性がある。

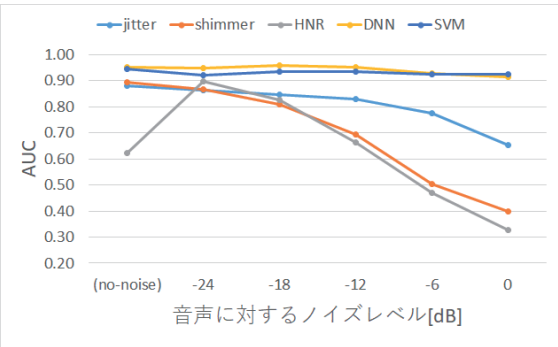


図 3. ホワイトノイズを付加させた際の AUC 推移

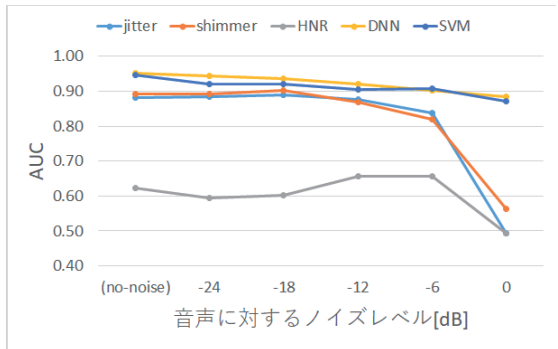


図 4. 車の走行音を付加させた際の AUC 推移

表 1. ホワイトノイズを付加させた際の AUC

音声に対する ノイズレベル[dB]	jitter	shimmer	HNR	DNN	SVM
(no-noise)	0.88	0.89	0.62	0.95	0.95
-24	0.86	0.87	0.90	0.95	0.92
-18	0.85	0.81	0.83	0.96	0.94
-12	0.83	0.69	0.66	0.95	0.93
-6	0.78	0.50	0.47	0.93	0.93
0	0.65	0.40	0.33	0.92	0.93

表 2. 車の走行音を付加させた際の AUC

音声に対する ノイズレベル[dB]	jitter	shimmer	HNR	DNN	SVM
(no-noise)	0.88	0.89	0.62	0.95	0.95
-24	0.88	0.89	0.59	0.94	0.92
-18	0.89	0.90	0.60	0.94	0.92
-12	0.88	0.87	0.66	0.92	0.91
-6	0.84	0.82	0.66	0.90	0.91
0	0.49	0.56	0.49	0.88	0.87

5. まとめ

ノイズが混在するような環境下で音声を取得する際に、MFCC を用いた分類を行うことで、既存の音声指標を用いた場合よりも分類性能を向上させる可能性が示唆された。今後の展望としては、今回は既存のデータセットに対してノイズを付加して評価したが、実際の録音時にノイズが含まれた場合の耐ノイズ性能の推移を評価することや、構音障害分類以外の問題での耐ノイズ性能の検証を行い、音声から人の内部状態をセンシングする際により強固な手法を確立していくことが挙げられる。

参考文献

[光吉 2011] 光吉俊二, 徳野慎一, 田中靖人: “音声感情技術 ST を使ったストレスへの応用”, 日本疲労学会誌, 第 6 巻, 第 2 号, pp.641-644, 2011.

[Tokuno 2015] Tokuno, Shinichi: Stress evaluation by voice: from prevention to treatment in mental health care, Econophysics, Sociophysics & other Multidisciplinary Sciences Journal, vol.5, pp30-35, 2015.

[Shinohara 2017] Shinohara, Shuji, et al.: Case studies of utilization of the mind monitoring system (MIMOSYS) using voice and its future prospects, Econophysics, Sociophysics, and Multidisciplinary Sciences Journal, vol.7, pp7-12, 2017

[高野 2017] 高野 毅 ほか, “走行中の自動車騒音が感情やストレス状態の分類に有用な音声構造解析に与える影響”, HCG シンポジウム 2017, 2017.

[Teixeira 2014] Teixeira, João Paulo, and Paula Odete Fernandes: "Jitter, Shimmer and HNR classification within gender, tones and vowels in healthy voices." Procedia Technology, vol.16 pp.1228-1237, 2014.

[Shinohara 2016] Shinohara, Shuji, et al. "Voice disability index using pitch rate." Biomedical Engineering and Sciences (IECBES), 2016 IEEE EMBS Conference on. IEEE, 2016.

[Londoño 2011] Arias-Londoño, Julián D., et al. "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients." IEEE Transactions on Biomedical Engineering, vol.58, pp.370-379, 2011

[Fang 2018] Fang, Shih-Hau, et al. "Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach." Journal of Voice, In press, 2018.

[Alhussein 2018] Alhussein, Musaed, and Ghulam Muhammad. "Voice pathology detection using deep learning on mobile healthcare framework." IEEE Access, vol.6, pp. 41034-41041, 2018.

[Llorente 2006] Godino-Llorente, Juan Ignacio, Pedro Gomez-Vilda, and Manuel Blanco-Velasco. "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters." IEEE transactions on biomedical engineering vol.53, pp.1943-1953, 2006

[Eyben 2010] Eyben, Florian, Martin Wöllmer, and Björn Schuller. "Opensmile: the munich versatile and fast open-source audio feature extractor." Proceedings of the 18th ACM international conference on Multimedia, 2010.

Word2vec を利用した文章校正支援に向けて

Toward proofreading support using Word2vec

丸山 正人^{*1}
Masato Maruyama

竹川 高志^{*1}
Takashi Takekawa

^{*1} 工学院大学情報学部
Faculty of Informatics, Kogakuin University

Abstract: Word2vec is a method of for learning distributed vector representations that semantic word relationships. In this paper, we aim for detection words misused from a sentence by distributed vector representations built using Word2vec. As a result, using linear discriminant analysis improved the detection rate of words misused.

1. はじめに

自然言語で扱われる単語の意味を計算機に理解させることは自然言語処理の分野において重要な課題である。この解決策として、分布仮説という同じ文脈で出現する単語は同じ意味を持つ傾向があるという事[1]に基づき、ある単語とその周辺に出現した単語を学習し単語間の関係をベクトルで表した分散表現という手法がある。これは現在では自動音声認識および機械翻訳への応用など広範囲の自然言語処理タスクへ応用されている[2]。近年 Tomas Mikolov らによって研究された Word2vec は分散表現を得る手法であり、従来の手法より高速な学習を可能にし[3]、単語同士の関係が $\text{vec}(\text{"Madrid"}) - \text{vec}(\text{"Spain"}) + \text{vec}(\text{"France"})$ の結果に最も近い単語ベクトルが $\text{vec}(\text{"Paris"})$ となるような線形構造を持つ[2]ことで近年話題になっている。

また、自然言語処理のタスクとして文章から誤りを検出、訂正する作業として校正があるが、検出、訂正される誤りは事前に予測ができる習慣的なものや、文から単語の分割を行う処理である形態素解析でわかるような表層的なものに限られており、文脈や一般常識を用いた解析により判断できるような、内容レベルの校正は困難とされている[4]。

そこで Word2vec を用い、実在している単語かつ文法として正しく用いられる固有有名詞の誤字を検出を本研究の目的とする。これは実在している単語であり文法が正しいために事前の予測のみでの解決や、形態素解析や文法からの検出が難しいと考えられるような内容レベルの校正になると考えられ、助詞や代名詞などの、文脈に制限されずに出現が考えられる品詞ではなく特定の文脈でのみ出現するような固有有名詞であるため、分布仮説に基づいた分散表現を出力する Word2vec を用いる事が適切と考えられるためである。

以上のことから、解析対象の文章に含まれていない固有有名詞を、文脈に出現しないような単語である誤字として、文章内の固有有名詞のベクトルの平均と各固有有名詞のベクトルについてユークリッド距離を用いた誤字の検出を行い、それを通して固有有名詞のベクトルの分布を調査した。

その結果、文章内の固有有名詞のベクトルの平均と各固有有名詞のベクトルについてのユークリッド距離が文章内の固有有名詞のベクトルと文章に含まれていない固有有名詞のベクトルで分布が重なっていることが分かり、文章内の固有有名詞について主成分分析を行ったものについても分布が重なっていた。これは軸が表す意味を無視して分離を行なっているのが原因と考えられ、文

章内の固有有名詞と文章に含まれない固有有名詞の分布を分離するような特徴を持つ軸を切り出す線形判別分析を行うと 2 つの分布を分離することができた。

2. 実験方法

本研究では、Word2vec を用いて学習を行った Web 上で公開されている分散表現のモデル[5]を用い、モデルの学習に用いられた Wikipedia の日本語記事を解析対象の文章に用いる。また、解析で用いる文章の単位は段落もしくは記事であり、単語の分割、品詞の特定についてはシステム辞書に mecab-ipadic-NEologd を用いた MeCab を用いた。

2.1 ユークリッド距離を用いた誤字の検出

Wikipedia の日本語記事から段落ごとに文章を取り出し、その中から固有有名詞 1 つを段落に含まれていない固有有名詞に交換された文章を解析対象とする。文章内の固有有名詞のベクトルの平均との差が最も大きかった固有有名詞が交換された固有有名詞であるか調べ、段落内の固有有名詞の数ごとにその正答率を計算した。

2.2 ユークリッド距離と主成分分析を用いた単語の分布

Wikipedia の日本語記事から記事ごとに文章を取り出し、記事内の固有有名詞と記事で用いられていない固有有名詞の分布をグラフに表し確認する。分布をグラフに表す方法は、記事内の固有有名詞のベクトルの平均との距離を記事内の固有有名詞と記事に用いられていない固有有名詞の 2 群つについてヒストグラムに表す方法、記事内の固有有名詞のベクトルについて 2 次元へ圧縮する主成分分析を行い、得られた変換行列を用いて記事内の固有有名詞と記事に用いられていない固有有名詞のベクトルに変換行列をかけ、散布図にプロットする方法である。

2.3 線形判別を用いた分離

Wikipedia の日本語記事から記事ごとに文章を取り出し、記事ごとに記事内の固有有名詞と記事外の固有有名詞を分離するような教師データとして線形判別分析を行い分離度を検証した。

3. 結果

3.1 ユークリッド距離を用いた誤字の検出

図 1 は段落内の固有有名詞の数ごとに文章に含まれていなかった単語の正解率を出したグラフであり、エラーバーは誤差分散を表している。文章内の交換されていない固有有名詞の数が 2 個のときの正解率がおおよそ 0.7 となり、単語数が多くなるにつれ

て正解率は低下し、交換されていない固有名詞の数が 30 個の時には正解率はおよそ 0.5 となった。

3.2 ユークリッド距離と主成分分析を用いた単語の分布

図 2 は Wikipedia の“日本語”という記事に対して記事内の固有名詞と記事に含まれていない固有名詞について分離を行ったグラフである。上から順に記事内の固有名詞のベクトルの平均との距離を用いた分離方法、主成分分析を用いた分離方法のグラフである。ユークリッド距離を用いて分離を表したグラフは記事内の単語のヒストグラムと記事に含まれていない単語のヒストグラムの面積が等しくなる様にグラフの設定をおこなっているため、実際には記事内の単語のヒストグラムは小さくなる。二つのグラフとも文章内の固有名詞の分布と文章に含まれていない固有名詞の分布が重なっていることが分かる。

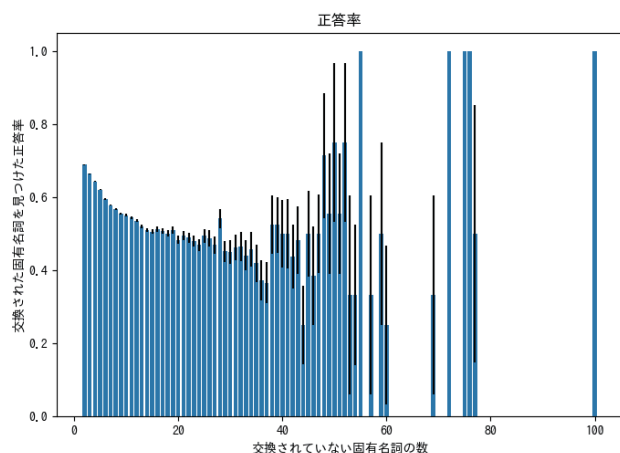


図 1. ユークリッド距離を用いて文章に含まれていなかった固有名詞を検出できた正答率

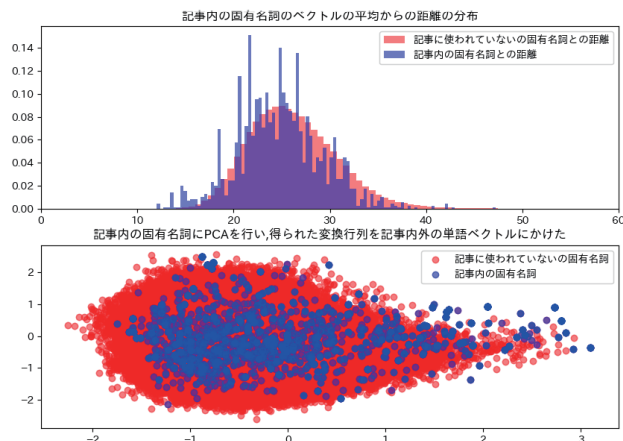


図 2. “日本語”記事に対しユークリッド距離を用いた分離、主成分分析を用いた分離を行なった分布

3.3 線形判別分析を用いた分離

図 3 は Wikipedia の“日本語”という記事に対して、図 4 は“石川県”という記事に対して記事に含まれる固有名詞と含まれていない固有名詞を教師データとして線形判別分析を行った結果である。この二つのグラフについても 2 つの分布の面積が等しくなる様にグラフの設定を行っているため、実際には記事に含まれる固有名詞のヒストグラムは小さくなる。ユークリッド距離を用いた分離や主成分分析を用いた分離より 2 つの分布の重なりが少ない事がわかる。

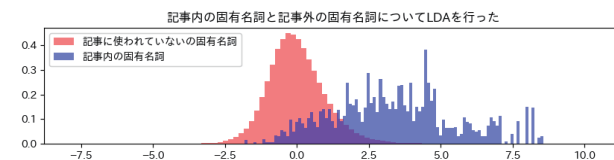


図 3. “日本語”の記事に対して線形判別を行った。

4. まとめ

ユークリッド距離を用いた誤字の検出では期待値よりは大きな正解率となったが、文章校正を行う上では不十分であった。そこで実際にベクトル同士の距離についての分布と主成分分析を行った分布について調査したところ、分離を行いたい 2 つの分布が重なっている事がわかった。これは Word2vec のモデルが 100 次元と高次元になっているのにも関わらず、距離を用いた分離では軸ごとの重みを考慮していないことから、主成分分析についても軸ごとの分布の意味を考慮していない事から起因するものだと考えられる。そこで 2 つの分布が分離する様な軸で次元を圧縮するために線形判別分析を行うと上の二つの手法よりよく分離している事がわかった。

また、図 5 は“日本語”の記事に対して線形判別分析を行い、得られた軸上に存在する単語を出力したグラフである。グラフに記載されている単語は横軸は線形判別分析により得られた軸であるが、縦軸に意味はない。このグラフにプロットされている単語は横軸で正の方向に進むほど文章の内容をよく表している単語と考えられられ、単語と文章の内容の関連度を数値化する事が期待できると考えている。

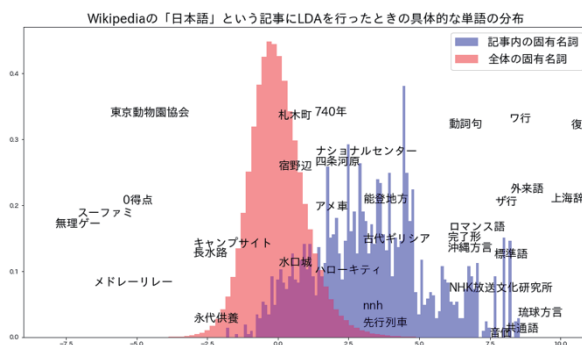


図 5. “日本語”の記事に対して線形判別分析を行った軸に対して単語をプロットした。

参考文献

- [1] Zellig S. Harris: Distributional structure, WORD, 10:2-3 .pp.146-162, 1954.
- [2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems 26. pp.3111-3119. 2013.
- [3] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at International Conference on Learning Representations, 2013.
- [4] 池原 悟, 小原 永, 高木 伸一郎. 文書校正支援システムにおける自然言語処理. 情報処理. 34 巻. 10 号. pp.1249-1258. 1993.
- [5] 鈴木 正敏, 松田 耕史, 関根 聡, 岡崎 直観, 乾 健太郎. Wikipedia 記事に対する拡張固有表現ラベルの多重付与, 言語処理学会第 22 回年次大会. pp.A5-2. 2016.

DAMを用いた文レベルのアライメントによる テキスト平易化コーパス構築手法の提案

Construction of Corpus for Text Simplification by Sentential Alignment based on Decomposable Attention Model

永塚 光一^{*1}
Koichi Nagatsuka

渥美 雅保^{*1}
Masayasu Atsumi

^{*1} 創価大学大学院 工学研究科
Graduate School of Engineering, Soka University

Text Simplification is a task to generate a sentence which is easier to understand than original. Text Simplification helps beginners such as children and foreigners learn languages. Recently, seq2seq models based on large scaled datasets have achieved state-of-the-art results in many areas including Machine Translation, Summarization, and Question Answering, to name a few. Although these model can be applied in text simplification as well, it requires a large number of parallel sentence pairs. Since available sentential corpora for text simplification are inadequate, building new corpus is so critical. In this paper, we suggest the application of neural textual entailment method to detection of simplified sentence pairs so that we are able to automatically construct text simplification dataset. In experiment, we evaluated the performance of identification of simplified sentences by using manually annotated dataset, and our proposed framework outperformed a baseline method.

1. はじめに

テキスト平易化とは、ある文が与えられた時に、読者にとってより分かりやすい文に表現を変換する処理のことを指す。テキスト平易化は、新たに言語を学び始めた子供や、外国人、難読症患者や、リテラシーの低い人々を支援するために重要な技術であると言える。ここで、ある文に対応する簡易な表現の文に変換するタスクは、単一言語における翻訳処理の一種であるとみなすことができる。この観点から、機械翻訳で成功を収めているニューラル seq2seq モデルをテキスト平易化へ応用した手法の研究が行われてきた[Nisioi 17]。

こうしたモデルの学習には入力文と出力文が 1 対 1 に対応している大規模なパラレルコーパスが必要とされる。しかしながら、テキスト平易化にはそのような文間でアライメントが取れたコーパスが構築されていない現状がある。テキスト平易化のためのパラレルコーパスが不足する理由として、2つの理由が挙げられる。まず、一つ目の理由として、ある同一のトピックに対して、複数のレベルで明確に平易化された文書の数が少ないということが挙げられる。これは、一般的に平易化という処理が翻訳に比べてニーズが低いために、平易化された文書のペアが手に入りにくいということに起因する。また、パラレルコーパスの構築が難しいもう一つの原因として、平易化された文が、元の文と 1 対 1 の対応関係を常に持つとは限らないということが挙げられる。テキスト平易化は、元の文章の重要な部分だけを残す処理であるという観点から要約の一種であり、1 文を短くするだけに留まらず、元の文章から重要でない文の削除が起きたり、2 文を 1 文に統合したりすることがある。したがって、平易化された文書ペアを取得することができた場合でも、文の対応関係を見つけることは容易ではない。これに対し、人間が手作業で対応関係にアノテーションを行うことは、質の高いコーパスのために理想的ではあるが、かなりのコスト及び時間を要するため、コーパス構築のボトルネックとなっている。

こうした中、先行研究[Xu 15]では、元文と平易化文のペア(これを平易化文ペアと呼ぶ)を自動的にアライメントする手法が提案されている。しかし、これまでの手法は、表層的な情報のみに基づいており、平易化の処理におけるパラフレーズや構文構造の変化を判定することが難しいという課題点が残されている。

そこで、本研究では、より深層的な意味特徴に着目した Decomposable Attention Model を平易化文ペアの検出に応用することを提案する。Decomposable Attention Model はテキスト含意認識の分野において使用されているモデルである。テキスト含意認識は、二つ文が与えられた時に、それらの間で含意関係が成り立つかどうかを判定するタスクである。ここで、平易化された文は元の文と等価か、少なくとも含意関係にあると言える。すなわち、平易化文ペアの検出は、含意関係認識の問題に置き換えることができる。含意関係の推論処理において、従来の表層的特徴ではなく、セマンティックな特徴を効果的に利用することができれば、パラフレーズが起きた平易化文ペアの検出が可能になると期待される。

2. 先行研究

2.1 テキスト平易化コーパス

これまで開発されてきたテキスト平易化のためのデータセットとして、Simple Wikipedia Corpus[Zhu 10]と Newsela Corpus[Xu 15]がある。

(1) Simple Wikipedia Corpus

Simple Wikipedia Corpus は最初に開発されたテキスト平易化のための大規模データセットである。データの構築には Wikipedia にある記事とその記事をより簡単な英語に置き換えた Simple English Wikipedia の記事が文書ペアとして使われている。Simple Wikipedia Corpus は、テキスト平易化の研究において貴重なベンチマークデータセットであるものの、平易化のための再編集をしているエディタがボランティアであることや、明確な平易化の統一基準が不足していることに起因する平易化処理の質の低さが問題点として指摘されている[Xu 15]。

(2) Newsela Corpus

こうした問題意識に基づき、その後 Newsela Corpus が作成された。Newsela Corpus は、児童のための学習教材からデータを収集しており、プロのエディタにより複数の学習者レベルに応じてテキストが編集されていることが特徴である。Newsela Corpus は Simple Wikipedia Corpus に比べて、テキスト平易化の質が高いコーパスであるものの、どちらのコーパスも依然として文レベルのアライメントがなされていないという課題点がある。

2.2 自動アライメント

文レベルのアライメントを自動的に検出する手法がいくつか提案されている。Newsela Corpus の作成者[Xu 15]は、最もシンプルな文間の類似度計算手法として Jaccard similarity を使用することを提案している。Jaccard similarity は以下の式で与えられる。

$$Jaccard(X, Y) = \frac{\text{count}(X \cap Y)}{\text{count}(X \cup Y)} \quad (1)$$

また、平易化文ペア間の 1 対 1 関係のみではなく、1 対 N や N 対 1 の対応関係の検出に特化した Vicinity-driven sentence alignment が提案されている。Vicinity-driven sentence alignment は TF-IDF に基づいた手法であり、実際に Newsela Corpus の自動アライメントに適用された研究が報告されている[Scarton 18]。しかしながら、パラフレーズや省略が頻繁に起こるテキスト平易化では、単語頻度により意味的な類似度を計算することには限界がある。

一方で、より深層の意味関係を推論するタスクにテキスト含意認識がある。テキスト含意認識では、ある文のペアを含意、矛盾、中立の 3 つの関係に分類する。

- Bob is in his room, but because of the thunder and lightning outside, he cannot sleep.
- Bob is awake.
- It is sunny outside.

例えば、最初の文章では、「he cannot sleep」から bob が起きている状態が推論されるので、二つ目の「bob is awake」という文意は最初の文意に含まれ、含意関係があると言える。一方で、最後の文の「it is sunny outside」という内容は最初の文の「the thunder and lightning」という描写に相反するので、矛盾関係にあると結論できる。本研究では、テキスト平易化文ペアの検出にテキスト含意認識モデルを適用する。

3. Decomposable Attention Model(DAM)

Decomposable Attention Model (DAM)[Parikh 17]は、テキスト含意認識を学習するニューラルネットワークモデルである。DAM は文間の意味推論をある単語レベルのアテンションに分解して、最終的な含意関係の予測を行う。このモデルではまず、アテンション機構を用いて 2 つの入力文 $S_a[w_{a1}, \dots, w_{ai}, \dots, w_{al_a}]$, $S_b[w_{b1}, \dots, w_{bj}, \dots, w_{bl_b}]$ 中の単語間の注意重み e_{ij} を計算する。その後、注意重みで単語分散表現 α_i 及び β_j を獲得する。

$$\alpha_i = \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ki})} \bar{a}_i \quad (2)$$

$$\beta_j = \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{ki})} \bar{b}_j \quad (3)$$

ここで、 l_a , l_b は文 S_a , S_b の単語長である。また、 $\bar{a}_i$ 及び $\bar{b}_j$ は、各入力文中の単語に対する事前学習された単語分散表現である。[Parikh17]は、GloVe[Pennington 14]を単語分散表現の事前学習手法として用いているが、今回のモデルでは、ELMo[Peters 18]を使用する。

次に、2 つ入力文の単語分散表現 $\bar{a}_i$, $\bar{b}_j$ と α_i , β_j との間で意味関係をフィードフォワードネット G により計算して、各単語と文間の関係を表すベクトル $v_{a,i}$, $v_{b,j}$ を求める。

$$v_{a,i} = G([\bar{a}_i, \beta_i]) \quad (4)$$

$$v_{b,j} = G([\bar{b}_j, \alpha_j]) \quad (5)$$

$$v_a = \sum_{i=1}^{l_a} v_{a,i} \quad (6)$$

$$v_b = \sum_{j=1}^{l_b} v_{b,j} \quad (7)$$

最終的に、フィードフォワードネット H により、2 つの文間の関係を表すラベル y を出力する。

$$y = H([v_a, v_b]) \quad (8)$$

この手法は、アテンションとフィードフォワードニューラルネットワークに、単語分散表現の事前学習を組み合わせた単純なモデルとなっており、従来の複雑なアーキテクチャに比べて少ないパラメータで学習することができる。

4. 実験

4.1 データセット

実験では、英語学習者向けのニュースサイト Breaking News English¹から収集したマルチレベルの平易化テキストを用いる。Breaking News English は、easy タイプの記事(level 0, 1, 2, 3)と hard タイプの記事(level 4, 5, 6)から構成されている。今回の実験では、収集サンプルから平易化文ペアであるものと平易化文ペアでないものを 50 ペアずつランダムに抽出し、計 100 事例を用意した。easy タイプの記事からはレベル 0 と 3 を、hard タイプの記事からはレベル 4 と 6 の文を使用する。これらのペアに人手でアノテーションを施すことで、モデルの平易化文ペア識別性能をテストする。ベースラインには、Jaccard Similarity を用いる。各モデルの閾値は、[Xu 15]に従い 0.5 とする。

4.2 実験結果

実験結果を表 1 に示す。表 1 から、F 値において、DAM がベースラインを大きく上回っていることが分かる。特に、Recall において、Jaccard Similarity に大きな差をつけていることから、平

¹ <https://breakingnewsenglish.com/>

表 1 平易化文ペア特定の実験結果

	Recall	Precision	F値
Jaccard	0.36	1.00	0.52
DAM	0.76	0.82	0.79

表 2 特定された平易化文ペアの例

元文
They said our sleeping brain is much more aware of the outside world than we thought.
平易化された文
The researchers said our sleeping brain is active.

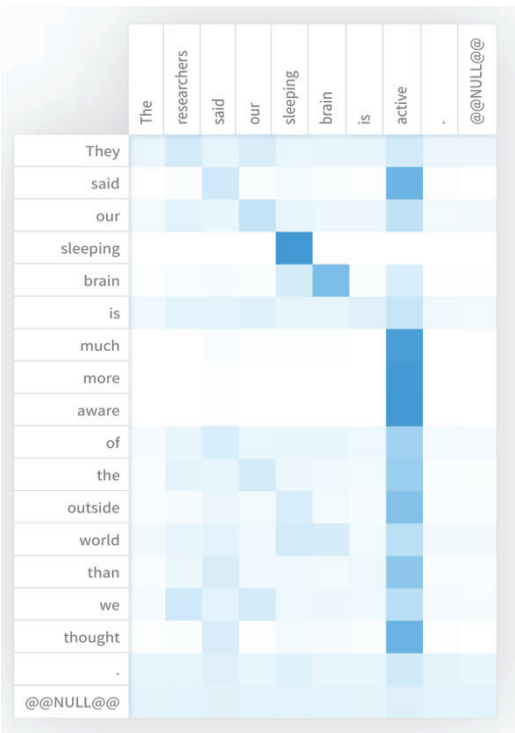


図 1 単語間のアテンション

易化文ペア間の共通単語頻度などの表層の特徴だけでは捉えきれない含意関係の認識に成功していることが分かる。

一方で、Precision においては、Jaccard Similarity が DAM よりも頑健な性能を示した。これは、ベースライン手法が文のペア間で共通する単語の重なりが一定の割合以上観測されない場合は全て負例に分類する方式であるためと考えられる。

実際に、Jaccard Similarity では検出に失敗したが、DAM によって検出に成功した平易化文ペアを表 2 に示す。この例では、平易化において、パラフレーズが起きているものの、含意関係が認識され検出に成功している。また図1に、単語間のアテンションの相関を示す。ここで、平易化された文中の”active”が元文中の”much more aware”というフレーズに強く対応していることがわかる。

5. まとめ

本研究では、テキスト平易化の平行コーパス構築のために、含意関係認識モデルを用いて、自動アライメントを行うことを提案した。実験結果より、従来の手法よりも含意関係認識モデルがより意味的な関係性を捉え、平易化文ペアの高精度なアライメントに貢献することがわかった。今後の課題の1つは、この枠組みを利用して、実際に作成中のコーパスに自動的にアノテーションを付与する機構を構築することが必要である。

参考文献

[Nisioi 17] : Sergiu Nisioi, Sanja Stajner, Simone Paolo Ponzetto, Liviu P. Dinu: Exploring neural text simplification models, In Proceedings of ACL. 2017.

[Xu 15] Wei Xu, Chris Callison-Burch, Courtney Napoles: Problems in Current Text Simplification Research: New Data Help, Transactions of the Association for Computational Linguistics, vol. 3, pp. 283–297, 2015.

[Zhu 10] : Zhu Z, Bernhard D, Gurevych I: A monolingual tree-based translation model for sentence simplification, In Proceedings of the 23rd International Conference on Computational Linguistics (COLING). 2010.

[Scarton 18] Carolina Scarton, Gustavo Henrique Paetzold, Lucia Specia: Text Simplification from Professionally Produced Corpora, In proceedings of Irec, 2018.

[Parikh 17] Ankur P. Parikh, Oscar Tackstrom, Dipanjan Das, Jakob Uszkoreit: A Decomposable Attention Model for Natural Language Inference, arXiv preprint:1606.01933, 2017.

[Pennington 17] Jeffrey Pennington, Richard Socher, Christopher D. Manning: GloVe: Global Vectors for Word Representation, Proceedings of EMNLP, 2014.

[Peters 18] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner: Deep contextualized word representations, In NAACL, 2018.

ココハドコ?: ユーザの位置が紐付いたウェブ検索履歴の解析

What's Here Like ? Analysis of Web Search Log Based on User's Location

日暮 立^{*1} 坪内 孝太^{*1}
Tatsuru Higurashi Kouta Tsubouchi

^{*1}ヤフー株式会社
Yahoo Japan Corporation

In recent years, with the spread of GPS-enabled mobile phones, a huge amount of users' historical location data is able to collect. Many studies on modeling with collected location data have been conducted. In this paper, we propose a new method to analyze the area characteristics based on user's web search query logs and location data. As a result of analyzing by the proposed system with various regions and periods, the feature of the area apper in users' search behavior.

1. はじめに

近年, GPS 機能を搭載したモバイル端末の普及に伴い, 膨大な量のモバイル端末位置履歴データが蓄積されてきている. モバイル端末に蓄積された位置履歴データを用いて, 都市の分析を行う研究や, ユーザの特徴を解析する研究が盛んである. 都市の分析を行うことで, 都市における需要把握などの都市計画への応用が期待できる. 本研究では位置履歴データを利用し, ユーザのウェブ検索行動から得られる都市や地域の特徴を解析する. Yahoo! JAPAN アプリ利用者の位置履歴データと Yahoo! 検索を利用したユーザの検索クエリログを用いて, 地域と期間を指定し機械学習により特徴的な検索ワードを推定するシステムを提案する. 学習の結果得られる, 地域毎の検索クエリの特徴を分析し, 指定期間の違いによる検索クエリの重みの変化から地域内で活動するユーザの興味の変化を分析する.

2. 関連研究

位置履歴データを活用して人々の活動パターンを抽出して解析する研究は数多くあり, パターン抽出結果を利用することで都市ごとの特性を把握することができる. [Fan 14] らは 1 日ごとの都市の活動人口のパターンが地域の性質を表現することに着目し, 位置履歴データからその代表的なパターンを抽出した. [Nishi 14] らは階層ベイズによるモデリング手法を用いて活動人口のパターン抽出を行なった. これらのパターン抽出の取り組みを応用して [Shimosaka 15] らは曜日や天気といった要因に対して変化する活動人口の予測に取り組み, [Okawa 17] らは路線毎の交通量予測という問題に取り組んだ. [Yabe 17] らは, 位置履歴データから都市ごとの人々の活動パターンを抽出し, 都市の災害に対する頑強性を示した.

大規模な位置履歴データから都市の状態を把握する研究が盛んであるが, 一方でユーザ個々の位置履歴データに着目しユーザを解析する研究も盛んである. [Wanaka 16] らは位置履歴データからユーザの興味関心を推定する方法を提案した. 位置履歴データからユーザ個人の状態を推定する研究は多いが, 推定の出力結果となるモデルがユーザの検索クエリで構成される推定モデルに関する研究は我々の知る限り存在しない.

3. 提案システム

本章では, 位置履歴データからユーザの特徴的な検索クエリを予測するためのシステムの説明を行う. 本提案システムでは, 任意の面積の地域と任意の期間を入力とし, 地域の特徴を現す検索クエリを出力とする. 本システムは, 地域と期間を入力し, 位置履歴データからユーザリストを抽出するユーザ抽出部と, 抽出した活動ユーザリストを入力して機械学習によって特徴的な検索クエリを予測するクエリ学習部で構成される. 以下に各システムの詳細を示す.

3.1 ユーザ抽出部

ユーザ抽出部は, 地域と期間を入力とし地域内で活動するユーザリストを取得する処理を行う. 以後ユーザ抽出部で出力するユーザを活動ユーザと呼ぶ. ユーザ抽出部では, 任意の面積の地域と任意の期間を入力とし, 蓄積された位置履歴データを元に入力した地域内の活動ユーザと活動ユーザの滞在スコアのリストを取得する. 滞在スコアの計算は位置情報の精度を加味するためにカーネル密度推定法による推定を行う. カーネル密度推定法とは, 密度を計算する地点を中心として任意に指定した検索半径内の点密度を, 計算地点からの距離減衰効果による重み付けを伴って計算する手法である.

あるユーザ蓄積されている位置履歴データのタイムスタンプを $T = \{t_1, t_2, \dots, t_n\}$ とし, 時刻 t_i の緯度を x_{t_i} , 経度を y_{t_i} , 位置情報の精度を a_{t_i} とし, ユーザの位置履歴データの集合を $D = \{(x_{t_i}, y_{t_i}, a_{t_i}) \mid t_i \in T\}$ とする. (x_{t_i}, y_{t_i}) が指定地域に含まれる場合, ユーザの滞在スコアは 1 となる. (x_{t_i}, y_{t_i}) が指定地域に含まれない場合, 指定地域の中心点と (x_{t_i}, y_{t_i}) の距離を d_{t_i} とすると, ユーザの滞在スコアはカーネル密度推定法により,

$$\begin{cases} \frac{1}{T} \sum_{i=1}^T K\left[\frac{d_{t_i}}{a_{t_i}}\right] & (a_{t_i} \geq h) \\ \frac{1}{T} \sum_{i=1}^T K\left[\frac{d_{t_i}}{h}\right] & (\text{otherwise}) \end{cases} \quad (1)$$

となる. d_{t_i} の計算はヒュベニの距離計算式を用い, カーネル密度推定法のカーネル関数 K はガウス関数を採用した. h はバンド幅である. 蓄積されているユーザの位置履歴データから点密度を計算し, 点密度が 0 より大きい上位の N 件のユーザを活動ユーザとして抽出する.

3.2 クエリ学習部

クエリ学習部では, ユーザ抽出部で抽出した活動ユーザリストとユーザ抽出部で指定した期間を入力として, 活動ユーザら

連絡先: 日暮 立, ヤフー株式会社, thiguras@yahoo-corp.jp

連絡先: 坪内 孝太, ヤフー株式会社, ktsubouc@yahoo-corp.jp

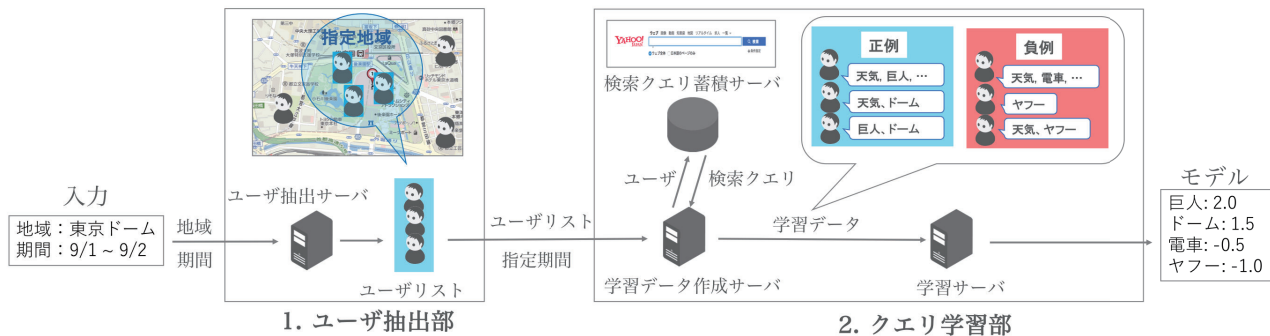


図 1: 提案システム概要

の特徴的な検索クエリを推定する。ユーザ抽出部にて出力した N 件の活動ユーザリストを $U_{pos} = \{p_1, p_2, \dots, p_N\}$ とする。また、 U_{pos} に含まれない、指定期間に Yahoo! 検索を利用したユーザをランダムに N 件サンプリングし、 $U_{neg} = \{n_1, n_2, \dots, n_N\}$ とする。

ユーザ p_i と n_i が指定期間に Yahoo! 検索にて検索を行なった検索クエリのリストを検索クエリログ蓄積サーバから取得し、取得した検索クエリの総数を Q とする。各ユーザ p_i と n_i が、指定期間に検索したクエリは 1、それ以外のクエリを 0 とする Q 次元の one-hot ベクトル $v \in \{0, 1\}^Q$ を特徴量ベクトルとする。ユーザ抽出部によって抽出した活動ユーザリスト U_{pos} を元に生成した特徴量ベクトルの集合 $D_{pos} = \{v_{p_i} \mid p_i \in U_{pos}\}$ を学習データの正例とし、ランダムにサンプリングしたユーザリスト U_{neg} を元に生成した特徴量ベクトル v の集合 $D_{neg} = \{v_{n_i} \mid n_i \in U_{neg}\}$ を学習データの負例とする。

生成した正例 D_{pos} と負例 D_{neg} を学習データとし、LIBLINEAR^{*1} を使用した二値分類学習を行う。学習の結果の学習モデルから得られる、 Q 個の素性として入力された検索クエリとその重みを出力する。学習の正規化パラメータ C は交差確認でパラメータ探索を行い、精度が最も良くなる C を採用するようにしている。

重みが高い素性は指定した地域内の活動ユーザ U_{pos} が特徴的に検索している検索クエリであることがわかり、重みが低い素性はユーザ U_{neg} が特徴的に検索している検索クエリであることがわかる。

4. データ

本章では、提案システムで使用するデータに関する詳細を示す。

ユーザ抽出部にて利用した位置履歴データは、Yahoo! JAPAN アプリ^{*2} を Andorid または iOS のスマートフォンにインストールし、位置情報を利用したサービス提供を受けるために「位置情報に基づいたコンテンツを表示する」という項目をオンにした Yahoo! JAPAN ID ログインユーザのデータを利用した。データはタイムスタンプと緯度経度と位置情報の精度の情報をもち、日本全国のユーザから取得されている。データは基本的に移動している状態の端末や基地局が切り替わった端末から取得されるため、人々の動きを表したデータと言える。スマートフォンで取得した位置履歴データは個人の位置の正確

な特定を避けるために 100m メッシュデータに変換している。利用時には k -匿名性を担保するため提案システムの抽出部において抽出人数には制限をかけている。

Android 端末のスマートフォンと iOS 端末のスマートフォンは位置情報の取得頻度が異なる。提案システムで利用する位置履歴データのタイムスタンプの間隔を共通化するために位置履歴データの内挿を行う。あるユーザの位置履歴データのタイムスタンプを t 、緯度を x 、経度を y 、精度を a とし、連続して蓄積された $t_0 < t_1$ となる履歴データをそれぞれ (t_0, x_0, y_0, a_0) 、 (t_1, x_1, y_1, a_1) とする。 S を 0 より大きい任意の秒数として $t_1 - t_0 > S$ の時、線形補間を行う。 S はシステム内で定義されるハイパーパラメータである。 $\{i \in \mathbb{N} \mid i < (t_1 - t_0)/S\}$ に対して、タイムスタンプが $t_i = t_0 + S * i$ 時の内挿データの x_i, y_i, a_i は以下のように算出される。

$$x_i = x_0 + (x_1 - x_0) * \frac{t_i - t_0}{t_1 - t_0} \quad (2)$$

$$y_i = y_0 + (y_1 - y_0) * \frac{t_i - t_0}{t_1 - t_0} \quad (3)$$

$$a_i = a_0 + (a_1 - a_0) * \frac{t_i - t_0}{t_1 - t_0} \quad (4)$$

クエリ学習部にて利用しているユーザの検索データは、ウェブブラウザもしくは携帯端末にインストールされているアプリで利用された Yahoo! JAPAN ID ログインユーザの Yahoo! 検索^{*3} のログを利用した。

5. 実験と結果

実験では $S = 300$ として内挿を行なった位置履歴データを利用した。提案システムのユーザ抽出部において、カーネル密度推定法のバンド幅 $h = 250$ 、抽出数 $N = 3000$ として活動ユーザを抽出し、クエリ学習部で使用した。3つのケースについて、提案システムによる解析例を以下に示す。

5.1 北海道地震 (北海道厚真町)

表 1 の例は、北海道勇払郡厚真町を含む 1600km² の地域を指定し、2018 年 9 月 6 日の 4 時から 2018 年 9 月 7 日の 17 時の期間を指定して提案システムによる学習を行った時の重みが上位の素性と重みが下位の素性の一部である。2018 年 9 月 6 日の 3 時 7 分には、北海道胆振地方中東部を震源として地震が発生しており、厚真町では最大震度 7 が観測されている。

*1 <https://www.csie.ntu.edu.tw/~cjlin/liblinear/>

*2 <https://promo-mobile.yahoo.co.jp/yjapp/>

*3 <https://search.yahoo.co.jp/>

表 1: 厚真町に滞在していたのユーザのクエリ

重みが高い検索クエリ		重みが低い検索クエリ	
検索クエリ	重み	検索クエリ	重み
新千歳空港	1.529	地下鉄	-2.57
道路交通情報	1.426	意味	-2.118
ラジコ	1.408	旭川	-1.779
運行状況	1.353	iphone	-1.397
ガソリンスタンド	1.238	ユーチューブ	-1.386
北海道電力	1.232	吉澤ひとみ	-1.343
苫小牧市停電情報	1.198	au	-1.249
鍋でご飯を炊く方法	0.97	おしゃれ	-1.242

学習結果から、指定した期間に北海道厚真町周辺に住む住民は「道路交通情報」、「ガソリンスタンド」や「北海道電力」などの検索クエリの重みが強く、災害時にはこれらの情報を特に必要としていることがわかり、一方で「iphone」「ユーチューブ」「おしゃれ」などの日常の娯楽を連想させる検索クエリは重みが低く、このような情報は災害時には特に必要としないということがわかる。実験から、提案システムによる検索クエリを予測することで、指定地域の災害状況を把握し住民が被災時にどのような状態かを素早く判断することができる。

5.2 安室奈美恵ライブ (沖縄県)

表 2 は沖縄県本島全域を含むを 8400km² を地域として指定し、指定期間を変化させた時の提案システムによる学習を行った時の重みが上位の素性の一部である。指定期間はそれぞれ 2018 年 9 月 12 日, 2018 年 9 月 15 日, 2018 年 9 月 18 日から 1 日間である。沖縄県では歌手の安室奈美恵が引退ライブを 2018 年 9 月 15 日に行っており、表 2 ではすべての日に「安室奈美恵」という素性が重みが強く、ライブに関する特徴がすべての日に現れている。各日付の「安室奈美恵」という検索クエリの重みに注目すると、ライブの前からライブ当日に向けて「安室奈美恵」という素性の重みは高く現れているが、ライブ終了後は重みが低くなっている。一方で「沖縄」のような地名を表す重みは指定期間の影響を受けることなく、常に高い重みで現れている。提案システムの出力結果からはイベントによる瞬間的に現れる検索クエリと、常時現れている検索クエリが存在することがわかる。瞬間的に現れる検索クエリを分析することで指定地域内のユーザの興味の変遷の把握が期待でき、常時現れる検索クエリを分析することで指定地域の特性の把握が期待できる。

5.3 東京ドームのイベント (東京都)

表 3, 表 4, 表 5 は東京ドームを含む地域 0.1km² を指定し、指定期間を変化させた時の提案システムによる学習を行った時の重みが上位の素性の一部である。

指定期間はそれぞれ、2018 年 11 月 17 日から 2019 年 1 月 20 日の間の土日の 2 日間である。東京ドームでは週末には様々なイベントが開催されており、アニメ「ラブライブ!サンシャイン!!」の音楽イベントやアイドルグループ「嵐」「Kis-My-Ft2」「SUPER JUNIOR」のドームライブ^{*4}、また音楽イベント以外にも全国の祭りやご当地の食を堪能できる「ふるさと祭り 2019」^{*5}、野球のファンイベント^{*6}、プロレスイベント^{*7} が行われていた。提案システムによって出力される検索クエリを

考察すると、東京ドーム周辺の地名以外に開催されているイベントに関連する検索クエリが特徴として現れている。提案システムによって出力される検索クエリから、指定した地域で行われているイベントを推定することが可能であることがわかる。また、表 4 のふるさと祭り東京の結果では「座席表」「時刻表」という検索クエリが特徴として強く現れている。ふるさと祭り東京は全国の祭りや食を堪能できるイベントとなっているため、全国から来場者が訪れる。来場者が全国各地から東京ドームに訪れる際に、新幹線や特急列車を利用している可能性が高いことが結果から考がえられる。このように、本提案システムからは、イベント抽出以外にも移動経路や移動方法の分析も期待できる。

6. まとめ

本論文では、Yahoo! JAPAN アプリを利用しているユーザの携帯端末の位置履歴データを利用して指定地域の活動ユーザを抽出し、ユーザの検索クエリを特徴量としてロジスティック回帰により学習することで、指定地域の特徴的なウェブ検索クエリを学習するシステムを提案した。様々な指定地域に対して提案システムによる分析を行なった結果、指定した地域にて行われたイベントに関する検索行動が検索クエリに現れることが示された。またイベント以外にも、指定地域で活動ユーザが必要としている情報を検索クエリから抽出することが可能であることがわかった。同じ指定地域において指定期間を変化させたところ、提案システムによって抽出された検索クエリの重みを比較することで、指定地域の活動ユーザの興味の変化を推定することが可能であることが示された。

地域ごとのウェブ検索クエリの解析の結果は、リスティング広告などのインターネット広告の配信への応用が期待される。今後は提案システムによって出力された地域毎のウェブ検索クエリモデルの実活用とその効果の検証を検討したい。

参考文献

- [Fan 14] Fan, Z., Song, X., and Shibasaki, R.: CitySpec-trum: A non-negative tensor factorization approach, UbiComp (2014)
- [Nishi 14] Nishi, K., Tsubouchi, K., and Shimosaka, M.: Extracting Land-use Patterns Using Location Data from Smartphones, URB-IOT (2014)
- [Okawa 17] Okawa, M., Kim, H., and Toda, H.: Online Traffic Flow Prediction Using Convolved Bilinear Poisson Regression, MDM (2017)
- [Shimosaka 15] Shimosaka, M., Maeda, K., Tsukiji, T., and Tsubouchi, K.: Forecasting Urban Dynamics with Mobility Logs by Bilinear Poisson Regression, UbiComp (2015)
- [Wanaka 16] Wanaka, S. and Tsubouchi, K.: Location History Knows What You Like: Estimation of User Preference from Daily Location Movement, Urb-IoT (2016)
- [Yabe 17] Yabe, T., Tsubouchi, K., and Sekimoto, Y.: CityFlowFragility: Measuring the Fragility of People Flow in Cities to Disasters Using GPS Data Collected from Smartphones, Ubicomp (2017)

*4 <https://www.tokyo-dome.co.jp/dome/event/artist/>

*5 <https://www.tokyo-dome.co.jp/furusato/>

*6 <http://meikyu-kai.org/bf2018/>

*7 <https://www.njpw.co.jp/tournament/163210>

表 2: 沖縄に訪れたユーザの検索クエリ

安室奈美恵のライブ3日前		安室奈美恵のライブ当日		安室奈美恵のライブ3日後	
検索クエリ	重み	検索クエリ	重み	検索クエリ	重み
沖縄	3.108	沖縄	3.644	沖縄	2.711
安室奈美恵	1.411	那覇	2.555	沖縄県	0.849
那覇	1.17	那覇市	1.772	古賀淳也	0.763
那覇市	0.763	安室奈美恵	1.769	那覇市	0.445
沖縄県	0.672	ほっともっと	1.757	那覇	0.417
直接謝罪	0.67	宜野湾	1.596	沖縄市	0.157
台風情報	0.404	地震速報	1.388	山本 kid 徳郁	0.116
沖縄宝島	0.351	読谷	1.38	沖縄タイムス	0.116
巨人杉内	0.335	今週の急上昇ワード	1.344	安室奈美恵	0.041
ひびきわたる	0.202	浦添	1.189	琉球新報	0.035

表 3: 東京ドームを訪れたユーザの検索クエリ (2018 年 11 月付近)

ラブライブ音楽イベント	野球のファンイベント	SUPER JUNIOR のライブ
ラブライブサンシャイン 東京ドーム ラブライブ aqours 新宿 アクセス 東京 文京区 秋葉原 紅白	名球会フェスティバル 2018 東京 東京ドーム 池袋 jra アクセス 銀座 上野 死去 名球会	新宿 super 東京ドーム 東京 水道橋 池袋 死去 今週の急上昇ワード 渋谷 秋葉原

表 4: 東京ドームを訪れたユーザの検索クエリ (2018 年 12 月付近)

嵐のライブ	Kis-My-Ft2 のライブ	嵐のライブ
嵐 東京ドーム 銀座 今週の急上昇ワード 池袋 水道橋 モード 東京 浅草 東京駅	キスマイ 東京ドーム アクセス 東京 秋葉原 東京駅 池袋 大恋愛 新宿 千葉	東京ドーム 嵐 銀座 神楽坂 水道橋 東京 みずほ銀行 浅草 今週の急上昇ワード 鶯谷

表 5: 東京ドームを訪れたユーザの検索クエリ (2019 年 1 月付近)

新日本プロレス興行	ふるさと祭り	ふるさと祭り
東京ドーム 池袋 news 祝日 東京 新日本プロレス 東京ドームシティ 高校サッカー 箱根駅伝 上野	ふるさと祭り東京 2019 東京ドーム 新宿 今週の急上昇ワード アクセス 池袋 天地総子 東京 北千住 東京駅	ふるさと祭り東京 2019 銀座 新宿 時刻表 悲痛コメント 渋谷 ランチ 座席表 東京ドーム 浅草

ニュースサイト閲覧者の興味関心を用いた 地域と観光要素の関連性抽出

Extraction of Relevance between Regions and Domestic Enjoyments
based on News Site App Users' Interest

川口輝太 ^{*1*2}

Kota Kawaguchi

井上裁都 ^{*2}

Tatsukuni Inoue

長田誠也 ^{*2}

Seiya Osada

山下達雄 ^{*2}

Tatsuo Yamashita

^{*1}筑波大学大学院 システム情報工学研究科

Graduate School of Systems and Information Engineering, University of Tsukuba

^{*2}ヤフー株式会社

Yahoo Japan Corporation

We proposed a method to discover domestic enjoyments information that are specific and original to individual domestic area. The proposed method is based on article follow information of news site users and themes given to articles. We extracted the relationship between regional theme and sightseeing theme using co-occurrence information of themes in each article. In addition, We pulled out the relationship between regional theme and sightseeing theme using themes actively selected by the user for personalization. By filtering with similarities between those two types of relevance, accuracy improvement was achieved. Moreover, by focusing on sightseeing theme with low similarity, error candidates for theme linking can be extracted.

1. はじめに

地域活性化のために、我が国の重要な成長分野である観光に注目が集まってきている。日本政府観光局の資料によると、2018 年は年間の訪日外国人数が 3119 万人と過去最高を記録した [1]。産業の観光化を推進するにあたり、観光地として訴求すべき観光コンテンツを正確に把握することは、地域への観光誘致のために重要であり、種々の研究がおこなわれている [2] [3]。

本研究では、実際のユーザの行動データを使って、各地域と観光要素の紐付けを行い、地域ごとの観光要素のクラスターを抽出することを目的とする。分析対象として、ニュースサイト「Yahoo!ニュース」[4]を用いた。Yahoo!ニュースの閲覧者は、興味のある地域や娯楽のテーマをフォローすることで、そのテーマが紐付けられた記事を優先的に閲覧できる。そこで、ニュース閲覧者の興味対象のテーマと実際の記事に付与されているテーマを併用した、観光要素別の地域ランキングと地域別観光要素ランキングの作成手法を提案する。

2. Yahoo!ニュースとテーマ

本研究でのターゲットである、Yahoo!ニュースについて説明する。Yahoo!ニュースは、新聞・通信社が配信する報道をはじめ、個人により執筆された記事や話題の動画、写真などを、ウェブサイトとアプリを通じて、提供するサービスである [4]。Yahoo!ニュースで用いられる「テーマ」とは、サービス側が事前に定義した興味関心の単位である。記事には複数のテーマが自動的に付与される。例えば、青森ねぶた祭りに関する記事は、「祭り」と伝統行事」、「青森市」、「青森ねぶた祭り」などのテーマが付与される。これにより、ユーザはテーマ別に記事閲覧できる。ユーザが上記のテーマを選択することを「フォロー」と呼ぶ。ユーザが各テーマをフォローをすることで、そのテーマが紐付けられた記事を優先的に閲覧できる。図 1 にニュース閲覧者とテーマと記事の関係を示す。

3. 地域テーマと観光テーマの関連性

本章では、「地域テーマ」と「観光テーマ」の関連性について述べる。地域テーマとは各都道府県と市区町村のテーマ、観

光テーマとは観光に関するテーマである。この二つの集合の各要素同士のつながりを明らかにすることで、地域ごとの特徴的な観光要素や観光要素ごとの特徴的な地域を抽出できると考える。これらのテーマ間の関連性を取り出すにあたり、ユーザがフォローするテーマ、記事に付与されるテーマの 2 種類のデータセットを用いた。

3.1 ユーザがフォローするテーマ

各ユーザがフォローする複数のテーマに着目した。ユーザがフォローするテーマはそのユーザの嗜好を反映している。そのため、同時にフォローされる複数のテーマはそれぞれに関連性を持つと言える。後述の手法により関連度を計算し、地域と観光要素のつながりを明らかにしていく。以下このテーマをユーザ由来テーマと呼ぶ。

3.2 記事に付与されるテーマ

次に配信される記事に同時に付与される複数のテーマに着目した。ニュース記事には、機械学習や言語処理技術 (エンティティリンキング等 [5]) を用いて複数のテーマを付与している。各記事に付与された地域テーマと観光テーマの共起を調べることで、それらの関連性が分かる。以下このテーマを記事由来テーマと呼ぶ。

3.3 関連性の意味

ここで関連性の意味について議論する必要がある。地域テーマをフォローするユーザは、その地域について興味を持つユーザではあるが、その興味を分類すると少なくとも次の 2 種類がある。

1. 地元民：その地域に住んでいる・係わりの深いユーザ
2. 観光民：その地域に観光等で行きたいユーザ

例えば、地域テーマ「渋谷区」と観光テーマ「スキー」を同時にフォローしているユーザがいるとする。渋谷区にはスキーを行う場所がないので、このユーザは 1 の地元民であると考えられる。そのため「渋谷区」と「スキー」の関連性は低くなるべきである。また、地域テーマ「長野県」と観光テーマ「スキー」を同時にフォローしているユーザの場合、「長野県」で「スキー」はメジャーであるため、このユーザは 2 の観光民の可能性が高い。そのため「長野県」と「スキー」の関連性は高くなるべきである。一方、地域テーマが付与された記事は、基

連絡先: ^{*1}s1820785@s.tsukuba.ac.jp, ^{*2}{tatinoue, sosada, tayamash}@yahoo-corp.jp

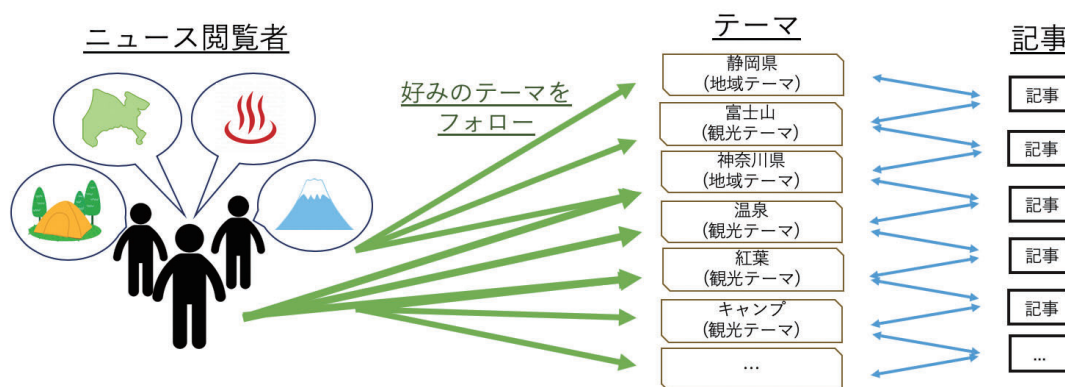


図 1: ニュース閲覧者とテーマと記事の関係

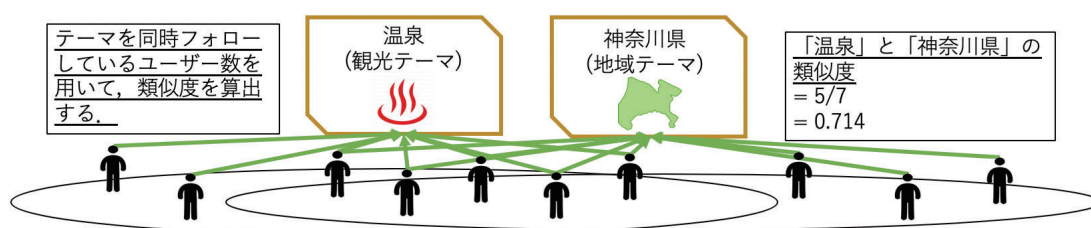


図 2: ユーザのフォローによるテーマ間の類似度の算出方法例

本的にその地域の話題であり、同時に付与された観光テーマがあればそれはその地域での観光要素であると言える。例えば、長野県でのスキー大会のニュース記事に、「長野県」「スキー」の両テーマが付与された場合、これらの関連性は高いと言える。ユーザのフォローするテーマのときは異なり、記事に付与された地域テーマと観光テーマの関連性は高くなる。

3.4 ユーザと記事におけるテーマ間関係の差

ユーザがフォローするユーザ由来テーマ、および、記事に付与される記事由来テーマから、それぞれ地域・観光間の関連性を得ることができる。ここで、テーマ間の関連度がユーザ由来と記事由来とでの間で離れていた場合、少なくとも下記の2つの可能性がある。

1. 誤った関連
2. ユーザの嗜好と供給されている記事とのミスマッチ

1 は、3.3 節の例で挙げた、ユーザ由来のデータから得られた「渋谷」と「スキー」のような、地域と観光要素の本来の組み合わせでないケースである。2 は、ある地域テーマとある観光テーマの関連性が、ユーザ由来では高く、記事由来では低い場合、ユーザの需要に対してニュース記事が供給不足の可能性があり、これらを実際の分析結果から確認する。

4. 関連性の抽出手法

本章では、地域テーマと観光テーマの関連性を明らかにするための関連度の計算方法について述べる。

4.1 抽出対象テーマの選択

関連性の抽出対象である地域テーマと観光テーマの選択手順について説明する。地域テーマは、47 都道府県 + 52 市区町村の計 99 テーマを人手で選定した。この 52 市区町村は、観光

に強く関連すると考えられる市区町村を主観に基づいて選定した。これらの地域テーマをフォローしているユーザが同時にフォローしているテーマの中から、旅・レジャー・観光に関する、フォロー数が 145 以上の 160 件を「観光テーマ」とした。表 1 に地域テーマと観光テーマの一例を示す。観光テーマの中には、移動手段などの間接的に観光と関係するテーマも含まれている。

4.2 関連度の計算

地域テーマ集合と観光テーマ集合のそれぞれのテーマ同士の関連度を計算する。関連度の計算には、Simpson 係数を採用した (図 2)。集合 X と集合 Y の間の Simpson 係数は、下記の式によって定義される [6]。

$$\text{Simpson}(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)}$$

例えば、ユーザ由来テーマを用いる場合、「 X = 長野県」「 Y = スキー」とすると、 $|X|$ は地域テーマ「長野県」をフォローしているユーザ数、 $|Y|$ は観光テーマ「スキー」をフォローしているユーザ数、 $|X \cap Y|$ は「長野県」と「スキー」を両方フォローしているユーザ数となる。また、記事由来テーマを用いる場合、同じく「 X = 長野県」「 Y = スキー」とすると、 $|X|$ は地域テーマ「長野県」が付与された記事数、 $|Y|$ は観光テーマ「スキー」が付与された記事数、 $|X \cap Y|$ は「長野県」と「スキー」の両方が付与された記事数となる。

5. 関連性の抽出

関連性の抽出にあたって、2018 年 11 月 8 日時点のユーザのフォローデータを使用した。対象ユーザ数は約 1000 万人であった。ニュース記事は、2018 年 12 月に配信された中から約

表 1: 地域テーマ, 観光テーマの一例

地域テーマ例	観光テーマ例
軽井沢市, 日光市, さいたま市, 相模原市, 新潟市, 倉敷市, 函館市, 川越市, 横須賀市, 西表島, 釧路市, 花巻市, 祇園, 難波, 松山市, 久留米市, 佐世保市, ...	航空業界, 東京ディズニーリゾート, 全日本空輸 (ANA), 日本航空 (JAL), ユニバーサル・スタジオ・ジャパン, 旅行・宿泊 (国内), 温泉, 釣り, 釣り情報, アウトドア, キャンプ場, オリエンタルランド, 登山家, 旅行・宿泊, ...
計 99 テーマ	計 160 テーマ

表 2: ユーザ由来のデータを用いた観光テーマ毎の関連度上位の地域 (括弧内の数字は関連度)

テーマ名/関連度スコア上位 3 位地域	1 位	2 位	3 位
温泉	東京都 (0.196)	箱根町 (0.175)	熱海市 (0.174)
民泊	京都市 (0.236)	東京都 (0.236)	京都府 (0.194)
パワースポット	京都市 (0.134)	京都府 (0.119)	大阪府 (0.104)
グランピング	北海道 (0.196)	兵庫県 (0.196)	京都府 (0.176)
スキューバダイビング	石垣島 (0.167)	東京都 (0.167)	北海道 (0.167)
日本庭園	京都市 (0.438)	大阪市 (0.250)	京都府 (0.250)
シュノーケリング	沖縄県 (0.462)	石垣島 (0.308)	名古屋市 (0.308)
平等院	京都府 (0.500)	京都市 (0.500)	祇園 (0.313)
グリーンツーリズム	北海道 (0.500)	山梨県 (0.500)	宮崎県 (0.500)
宿坊	京都府 (0.500)	山梨県 (0.250)	長野県 (0.250)

4000 件を使用した。

表 2 と表 3 にユーザ由来, 記事由来における観光テーマ毎の地域ランキングを示す。

表 2 では「温泉」に対して「箱根町」「熱海市」, 表 3 でも「大分県」「群馬県」といった温泉地として有名な観光地が抽出できた。また「宿坊」については, 表 2/表 3 とともに関連性が高い「京都市」が抽出できた。表 4 と表 5 に地域毎の観光要素ランキングを示す。表 4 では「宮崎県」に対して「海水浴」, 「北海道」に対して「トロッコ列車」, 表 5 では「京都市」に対して「紅葉」などの妥当な組み合わせが抽出できた。

6. ユーザ由来と記事由来テーマ間の類似度

記事由来テーマ間の関連性は, システムが自動的に記事に付与するテーマに基づくものであり, 実際に配信された記事の中身に基づく。一方, ユーザ由来テーマ間の関連性は, ユーザのニュースに対する興味関心に基づくものである。これら記事由来とユーザ由来の観光テーマを比較するための手法を述べる。5. 章で得られた, 観光テーマ毎の地域ランキング (表 2, 表 3) データに着目する。各観光テーマは, 地域テーマを要素としたベクトルを持つ。各地域テーマは各観光テーマとの関連度を持っている。これをここでは, 観光地域ベクトルと呼ぶこととする。 $\vec{v}_1$ は, ユーザ由来のデータによる観光地域ベクトル, $\vec{v}_2$ は, 記事由来のデータによる観光地域ベクトルである。観光地域ベクトル毎に, ユーザ由来テーマと記事由来テーマ間の類似度を算出した。類似度尺度には, $\cos$ 類似度を用いた。 $\cos$ 類似度は次のように定義される [6]。

$$\cos(\vec{v}_1, \vec{v}_2) = \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1||\vec{v}_2|}$$

表 3: 記事由来のデータを用いた観光テーマ毎の関連度上位の地域 (括弧内の数字は関連度)

テーマ名/関連度スコア上位 3 位地域	1 位	2 位	3 位
クルーズ客船	小樽市 (0.333)	長崎市 (0.222)	鹿児島県 (0.167)
トレッキング	さいたま市 (0.375)	屋久島 (0.250)	岩手県 (0.125)
パワースポット	宮崎市 (0.200)	川越市 (0.167)	宮崎県 (0.125)
宿坊	京都市 (0.667)	祇園 (0.500)	大阪市 (0.333)
キャンプ	宮崎市 (0.200)	宮崎県 (0.125)	金沢市 (0.067)
日本庭園	神戸市 (0.222)	長崎市 (0.111)	さいたま市 (0.071)
釣り	沼津市 (0.500)	相模原市 (0.125)	兵庫県 (0.111)
民泊	新潟県 (0.222)	新潟市 (0.200)	北海道 (0.187)
日本の観光列車	北海道 (0.333)	北九州市 (0.167)	札幌市 (0.067)
スキューバダイビング	大阪市 (0.219)	高知市 (0.167)	富山県 (0.063)
温泉	日光市 (0.857)	大分県 (0.650)	群馬県 (0.622)

v_1 と v_2 の次元数は, 地域テーマの数, つまり, 99 である。

7. 観光テーマ別地域ランキングの評価

観光テーマを中心に関連する地域テーマが正しく抽出できたか評価を行った。

7.1 評価用データ

評価対象は前述の観光テーマ 160 件のうち, 10 記事以上に付与された 71 テーマとした。記事由来とユーザ由来, それぞれにおいて, 観光テーマ別地域ランキングの TOP3 にその観光テーマで有名な地域が入っているか否かの 2 値のラベル付けの作業を行った。作業は 2 名で行い, ラベルが一致したテーマのみを評価に用いた。そのうち, ラベルが positive なものを正例とした。記事由来は正例 34 件, 負例 16 件, ユーザ由来では正例 37 件, 負例 11 件であった。

7.2 評価結果

観光テーマ別地域ランキングの記事由来での評価 (article theme) と, ユーザ由来での評価 (follow theme) を図 3 に示す。6. 章で得られた $\cos$ 類似度下限値を設けて, 下限値以上での判定結果における再現率, 適合率をプロットした。ユーザ由来と記事由来の類似度でフィルタリングしないときの精度はグラフ上, $Recall = 1$ に対応する $Precision$ であり, この時のユーザ由来と記事由来の $Precision$ は, それぞれ 0.771, 0.680 であった。ユーザの嗜好を直接反映していると考えられるユーザ由来の精度が, 両手法の類似度でフィルタリングすることで, 向上している。記事由来でもフィルタリングによる精度の向上が見られるが, ユーザ由来の精度には及ばないことがわかる。

8. ユーザ由来と記事由来テーマの差異分析

7.2 節で述べたように記事由来とユーザ由来の間で類似度が低いテーマをフィルタリングすると適合率が向上する。そこで, 類似度が低い観光テーマについて, 個別のテーマ毎に 3.4

表 4: ユーザ由来のデータを用いた地域テーマ毎の関連度上位の観光テーマ (括弧内の数字は関連度)

テーマ名/ 関連度 上位 3 件	1 位	2 位	3 位
北海道	トロッコ列車 (0.556)	グリーンツーリズム (0.500)	隠れ宿 (0.304)
静岡県	富士山 (0.272)	カヌー (0.200)	フライフィッシング (0.188)
京都府	宿坊 (0.500)	日本庭園 (0.250)	グリーンツーリズム (0.250)
宮城県	海水浴 (0.304)	カヌー (0.200)	シュノーケリング (0.154)
大阪市	観光競争力 (0.281)	海水浴 (0.217)	日本庭園 (0.125)
長野県	しなの鉄道 (0.667)	グリーンツーリズム (0.250)	宿坊 (0.250)
山梨県	グリーンツーリズム (0.500)	宿坊 (0.250)	トロッコ列車 (0.222)
金沢市	隠れ宿 (0.087)	リアル脱出ゲーム (0.083)	鮎釣り (0.071)
宮崎県	キャンプ (0.118)	鮎釣り (0.071)	フライフィッシング (0.063)
青森市	青森ねぶた祭り (0.353)	鮎釣り (0.071)	温泉 (0.051)
佐世保市	カヌー (0.200)	フライフィッシング (0.063)	海水浴 (0.043)
沼津市	フライフィッシング (0.125)	釣り大会 (0.105)	ワカサギ釣り (0.091)
熱海市	温泉 (0.174)	隠れ宿 (0.087)	秘湯 (0.087)
箱根町	温泉 (0.175)	紅葉 (0.100)	秘湯 (0.075)
日光市	温泉 (0.083)	日本の観光列車 (0.083)	リアル脱出ゲーム (0.083)

表 5: 記事由来のデータを用いた地域テーマ毎の関連度上位の観光テーマ (括弧内の数字は関連度)

テーマ名/ 類似度 上位 3 件	1 位	2 位	3 位
大分県	温泉 (0.650)	土産 (0.100)	温泉宿 (0.100)
金沢市	土産 (0.267)	紅葉 (0.267)	はとバス (0.167)
滋賀県	紅葉 (0.316)	鮎釣り (0.286)	土産 (0.105)
群馬県	温泉 (0.622)	秘湯 (0.333)	温泉宿 (0.077)
京都市	宿坊 (0.667)	紅葉 (0.250)	グリーンツーリズム (0.25)
北海道	ワカサギ釣り (1.000)	鉄道旅行 (0.600)	日本の観光列車 (0.333)
兵庫県	バーベキュー (0.333)	温泉 (0.270)	溪流釣り (0.222)
山梨県	富士急ハイランド (0.667)	富士山 (0.490)	溪流釣り (0.222)
大阪市	天王寺動物園 (0.9)	ユニバーサル・スタジオ・ジャパン (0.608)	宿坊 (0.333)

節であげた項目を検証した。

● 誤った関連

例 1 : 「東京都」「温泉」のような明らかな誤り
「東京都」と「温泉」テーマの間では、ユーザ由来テーマでは、関連度が高かったが、記事由来では、関連度は低かった (表 2, 表 3)。つまり、「東京都」をフォローしている人が同時に「温泉」テーマをフォローしていると考えられるので 3.3 節の 1 の地元民による「温泉」テーマのフォローであると考えられる。

例 2 : 「大阪府」「スキューバダイビング」のような記事テーマ付与の誤り候補

「大阪府」と「スキューバダイビング」のテーマの間では、ユーザー由来の関連度が低く、記事由来の関連度が高かった (表 2, 表 3)。実際に、「大阪府」と「スキューバダイビング」に同時に紐付く記事を見てみると、大阪府の水族館でコスプレをしたスキューバダイバーが水槽内に現れたという内容であった。ここでは、スキューバダイビングのテーマに紐付くのは、誤りであると考えられる。記事に対するテーマの紐付けの誤り候補を取り出すことができた。

● ユーザの嗜好と供給されている記事とのミスマッチ

ユーザ由来と記事由来間の類似度の低いテーマ「日本庭園」に着目する。「日本庭園」と「京都市」は、ユーザ由来のテーマでは関連が高いが、記事由来のテーマでは、関連が小さい (表 2, 表 3)。しかし、例 1 のような誤った地域は確認されず、配信側の記事テーマとユーザーの嗜好の間に差がある可能性がある。

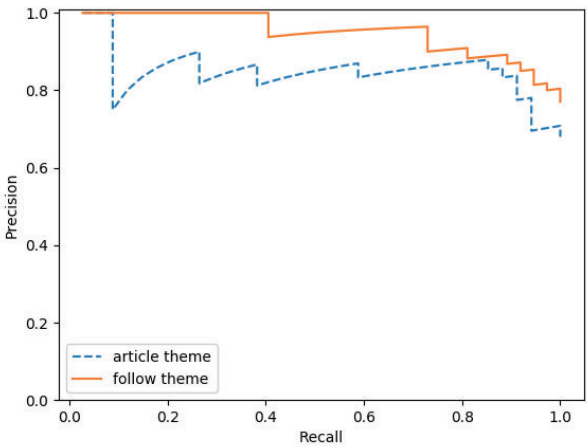


図 3: 観光テーマ別地域ランキングの人手評価結果

9. おわりに

本研究では、ニュースサイトにおいて、「ユーザによって自動的に選択されたテーマ」および「記事に内容に基づき自動的に付与されたテーマ」のそれぞれを用いて地域と観光要素の紐付けを行い、観光要素別地域ランキングと地域別観光要素ランキングを作成した。そして、両者の類似度を利用することにより精度の向上を確認できた。また、類似度が低いテーマを分析し、テーマ紐付けの誤りパターンの知見を得た。さらに、ユーザと記事配信の需要供給の間にミスマッチが存在すると思われるケースも発見した。今後は分析を深めて、より正確な地域・観光要素間の関連性の抽出を実現し、地域の意外な観光テーマの発見など、地域活性化に貢献していきたい。

参考文献

[1] 日本政府観光局 (JNTO). 国籍/月別 訪日外客数. https://www.jnto.go.jp/jpn/statistics/since2003_visitor_arrivals.pdf.

[2] 王怡青, 土井俊弥, 井上祐輔, 宇津呂武仁. ご当地グルメを題材とするクイズ・コンテンツの作成. 第 8 回 DEIM フォーラム論文集, 2016.

[3] 土井俊弥, 王怡青, 井上祐輔, 宇津呂武仁. 観光情報 PR のための旅ゲー風アプリの提案およびご当地グルメ版の作成. 言語処理学会第 22 回年次大会論文集, pp. 59–62, March 2016.

[4] Yahoo!ニュース. <https://news.yahoo.co.jp/>.

[5] 井上裁都, 末永圭吾, 長田誠也, 立石健二. Entity linking を用いたニュース記事に対する市区町村単位の地域情報の付与. 言語処理学会第 22 回年次大会論文集, pp. 45–48, March 2016.

[6] 相澤彰子. 大規模テキストコーパスを用いた語の類似度計算に関する考察. 情報処理学会論文誌, Vol. 49, No. 3, pp. 1426–1436, 2008.

解釈性のあるフェイクニュース検出器の実装と評価

Implementation and Evaluation of an Interpretable Fake News Detector

山本 和矢 ^{*1} 小山 聡 ^{*2} 栗原 正仁 ^{*2}
 Kazuya Yamamoto^{*1} Satoshi Oyama Masahito Kurihara

^{*1}北海道大学工学部
 School of Engineering, Hokkaido University

^{*2}北海道大学大学院情報科学研究科
 Graduate School of Information Science and Technology, Hokkaido University

Interpretability is an important element of fake news detection so that readers can assess the credibility of news by themselves. We implemented a naive Bayes fake news detection model proposed by Granik and Mesyur and evaluated it with the LIAR dataset in terms of recall, effect of stop words, and interpretability. The recall was affected by the imbalanced data and eliminating stop words did not improve the accuracy but slightly deteriorated it. Some high probability words were interpretable as reasons for fake news but longer phrases had better be considered as clues for fake news.

1. はじめに

SNS の普及により人々が手軽に情報を発信したり、入手することが可能になり我々の生活はより利便性を増した。一方でそうした情報の中には信憑性に欠けたものが含まれている。虚偽の情報でつくられたニュースをフェイクニュースと呼ぶ。フェイクニュースの目的は人を欺くことや衝撃的なタイトルで閲覧数を稼ぎ広告収入を得ること、また特定の人物や団体の根拠のない誹謗中傷をすることである。近年、このフェイクニュースを発端として様々な問題が発生している。その例として、2016 年 12 月 4 日にアメリカ大統領選中に SNS で拡散されたクリントン候補の陰謀説に関するフェイクニュースの真相を確かめようと銃を持った男がレストランに侵入するという事件が起こった ^{*1}。

あるニュースがフェイクニュースであるか人手で判別するには内容が虚偽であるか精査しなければならないため時間がかかる。また、メディアの多様化により大量のニュースが生み出されている。それゆえフェイクニュースをすべて人手で検出することは非常に困難である。そうしたことから、フェイクニュース自動検出の研究に対するニーズが高まっている。

一方で、フェイクニュースの自動検出器が構築できたとしても、それを無条件に信頼することにも問題がある。フェイクニュースの検出に誤りがあったり、検出器自体が「フェイク」である可能性も完全には否定できないからである。必要に応じて、読者が検出のプロセスを精査できることや、読者自身がニュースを批判的に読む能力を身に付けることも重要である。そのためには、なぜあるニュースがフェイクだと判定できたのか、その原因を調べることが可能な、解釈性の高い検出器が必要となる。

フェイクニュースの自動検出の試みの一つとして、[Granik 17] はフェイクニュースがスパムメッセージと類似点があるという仮定の元で、スパムフィルターに用いられるナイーブベイズ分類器を用いてフェイクニュースの自動検出を行い 74 % の精度でフェイクニュースを分類している。しかし、これには「データの偏ったデータでしか実験していない」、「ストップワードの除去を行っていない」、「使用しているモデルの解釈性が高いにもかかわらず検出の要因を分析していない」と

いった課題が存在している。

本研究の目的は [Granik 17] の用いた手法を実装し、ラベルの偏りによりどのような違いが生じるか調べることで、ストップワードの除去を行い分類結果の変化を検証すること、フェイクニュース検出の要因となった単語を分析することである。

2. 関連研究

[Granik 17] はフェイクニュースがスパムメッセージと、(1) 文法的な間違い、(2) 感情的に誇張された表現、(3) 読者の意見を操作、(4) 内容が虚偽、(5) 限定的な類似単語の使用、のような類似点があるという仮定の元で、スパムフィルターに用いられるナイーブベイズ分類器を用いてフェイクニュースの自動検出を行った。

この研究において使われたデータセットはラベルに偏りがありフェイクニュースは全体の 5 % ほどしか含まれていない。[Granik 17] はラベルの偏りが原因で検出の再現率が低くなったと主張している。また、ストップワードの除去を行えば分類の精度が向上すると主張している。しかし、いずれの主張も検証されておらず仮説にすぎない。また、単語的特徴から解釈性の高いモデルを用いてフェイクニュースの分類を行っているにもかかわらず、分類の要因となった単語を分析していない。

その他にもフェイクニュースの自動検出の研究が盛んに行われている。フェイクニュースの自動検出の手法として、ニュース記事の文章のみの情報を利用する言語的アプローチとソーシャルメディアや WEB 上の情報といった外部の情報を利用するネットワークアプローチがある。

欺瞞的な文章には代名詞や接続詞の頻度や出現パターン、否定的な感情を表現する言葉が多く用いられるといった言語的な特徴が存在する [Feng 13]。言語的アプローチはそうした特徴をニュースの文章から検出することを目的としている。

ソーシャルメディアで発信されたニュースに対して、ニュース記事の文章以外の特徴を用いてフェイクニュースの自動検出を行った例も存在する。[Tacchini 17] は Facebook の投稿がそれに「いいね!」を付けたユーザに基づいてデマであるかそうでないかに分類した。このように文書以外の情報を用いて分類する試みが存在するがフェイクニュースは必ずしも文書以外の情報が手に入るとは限らない。こういった手法は分類できる状況を限定してしまうという問題点がある。

連絡先: 山本和矢, yamamoto@complex.ist.hokudai.ac.jp

^{*1} <https://www.cnn.co.jp/usa/35093206.html>

3. 実験

3.1 使用手法

本実験では [Granik 17] が提案したナイーブベイズ分類器に基づくフェイクニュース検出器を実装して用いた。手法の詳細については [Granik 17] を参照されたい。

3.2 使用データ

先行研究で使われたデータは、入手不可能であったため代わりに形式の類似した別のデータを用いた。使用したデータは LIAR^{*2} である。LIAR は [Wang 17] によってつくられたフェイクニュース検出用のベンチマークデータセットである。アメリカの政治にまつわる声明のファクトチェックを行っているウェブサイト POLITIFACT.COM^{*3} の API で取得されたデータである。共和党及び民主党員の発言、ソーシャルメディアの投稿など約 1.2 万個の声明に対し、ID、6 段階の真偽ラベル (pants-fire, false, barely-true, half-true, mostly-true, true)、題目、発言者、発言者の職業、故郷の州、所属政党、発言者の声明に付けられたそれぞれのラベルの数、どこでの発言かといったタグが付与されている。実験では 6 段階のラベルのうち pants-fire, false, true のみを使用し、pants-fire と false をラベル Fake として扱った。またいくつかのデータの情報が欠損していたため除外した。

3.3 実験結果

ナイーブベイズ分類器を用いたフェイクニュース検出のデータの偏りによる再現率の変化を確認した。元のデータからラベルに偏りが生じるように抽出したデータとラベルに偏りがないように抽出したデータをデータ数の合計が同じになるように用意した。用意した 2 つのデータに対しナイーブベイズを用いたフェイクニュースの検出を行いそれぞれの再現率を比較した。偏りのあるデータの再現率は 0.04、偏りのないデータの再現率は 0.68 であった。これは [Granik 17] の仮説を支持する結果である。

Python の NLTK^{*4} コーパスのストップワードリストに含まれる単語を辞書から除去した場合とストップワードを辞書から除去しなかった場合の精度、再現率、適合率、F 値を比較したところ、ストップワードの除去によってこれらが改善されず、むしろ僅かながら低下する傾向があった。これは、フェイクニュース検出においてはストップワードは除去すべきものではなく、むしろ検出に有効な情報を含んでいる可能性があることを示している。

辞書に含まれた 9328 種類の単語のうち、単語がニュース記事中に存在するとき、フェイクニュースである確率が大きい上位 30 単語を表 1 に示す。このうち socialists, muslim, bad について、これらの単語が出現する訓練データの文書の例を表 2 に示す。表から分かるようにこれらの単語は文書中において特定の人物や組織を批判する目的で使われている。このことからフェイクニュース検出の要因となった単語には特定の人物や組織を批判する目的で使われる単語が含まれていることが分かる。フェイクニュースらしさの大きい他の単語は、単語自体ではなぜそれがフェイクニュースと判定する要因とするのが分かりにくいものが多い。しかし、たとえば responsible は is responsible for というフレーズで、他者を批判する文脈で用いられるなど、フレーズや文脈を考慮すれば、フェイクニュース判定の要因となった理由を解釈できるものもある。

face, governments, anything, responsible, began, reps, socialists, muslim, tell, scheme, cabinet, information, surplus, data, groups, census, benefit, reason, murphy, outside, loan, balanced, admits, parks, terry, provision, study, recently, bad, josh

表 1: フェイクニュース検出の要因となった単語

単語	単語が出現する訓練データの文書
socialists	Say Oregon Reps. Peter DeFazio and Earl Blumenauer are <i>socialists</i> who are openly serving in the U.S. Congress.
muslim	Says large majority of Republicans believe Obama is a <i>Muslim</i> and not U.S.-born.
bad	The Mexican government forces many <i>bad</i> people into our country.

表 2: 特定の単語が出現する訓練データの文書の例

4. まとめ

本稿では、先行研究で提案されたフェイクニュース検出の実装と評価を行った。これにより先行研究の仮説の検証と先行研究では着目されていなかったモデルの解釈性について分析を行いフェイクニュースに含まれる単語について知見を得ることができた。

本稿においては解釈性を活かすためナイーブベイズ分類器のみを用いたが、同様に解釈性のあるモデルに関して分類精度を比較する必要がある。検出の要因として単語だけでは解釈することが困難な場合が多いため、フレーズを抽出するなど、特徴のとり方を工夫するなどしてより人が解釈しやすくする必要がある。今回用いたフェイクニュース自動検出用のデータでは含まれる文書がアメリカの政治に関する限定的な題材であったが、様々なフェイクニュースを検出するためにはより多種多様なフェイクニュースに関するデータが必要であるため他の利用可能なデータの調査や SNS などを利用したデータの収集が必要である。将来的にはフェイクニュースの検出結果の根拠をユーザが解釈可能しやすい方法で可視化するシステムに関する研究を行いたいと考えている。

参考文献

- [Feng 13] Feng, V. W. and Hirst, G.: Detecting deceptive opinions with profile compatibility, in *IJNLP*, pp. 338–346 (2013)
- [Granik 17] Granik, M. and Mesyura, V.: Fake news detection using naive Bayes classifier, in *IEEE UKRCON*, pp. 900–903 (2017)
- [Tacchini 17] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., and Alfaro, de L.: Some like it hoax: Automated fake news detection in social networks, *arXiv preprint arXiv:1704.07506* (2017)
- [Wang 17] Wang, W. Y.: “liar, liar pants on fire”: A new benchmark dataset for fake news detection, *arXiv preprint arXiv:1705.00648* (2017)

*2 <https://www.cs.ucsb.edu/~william/data/liar.dataset.zip>

*3 <https://www.politifact.com>

*4 <https://www.nltk.org/>

統計的尺度に基づく有害表現抽出手法の自動評価

Automatic Evaluation for Cyberbullying Detection Method based on Statistical Scale

荒田 真輝^{*1} 梶井 文人^{*2} プタシンスキ ミハウ^{*3}
Masaki Arata Fumito Masui Michal Ptaszynski

^{*1}北見工業大学大学院 工学研究科 情報システム工学専攻
Kitami Institute of Technology Department of Computer Science

^{*2,3}北見工業大学 地域未来デザイン工学科
Kitami Institute of Technology School of Regional Innovation and Social Design Engineering

Various methods have been proposed to detect harmful information automatically as a support of cyberbullying countermeasures. However, there remain various problems such as detection performance deterioration over time and detection of information including private information. In this paper, we propose an automatic evaluation method for harmful expression extraction method using BLEU, an automatic evaluation method typically used in evaluation of machine translation methods.. In this method, for as a detected as a harmful by the harmful expression extraction method, the evaluation is performed based on word N-gram Precision using the harmful entry gold standard dataset. As a result of the evaluation, the proposed allowed assigning a high score to harmful entries, which confirmed the effectiveness of the method.

1. はじめに

インターネットの普及に伴い、Web 掲示板や SNS 等に特定個人への誹謗中傷や個人情報等を書込む「ネットいじめ」が社会問題となっている [1]。しかし「ネットいじめ」の対処には膨大なコストがかかることや、監視自体が困難になってきているなど多くの課題があり、これらに対処するための有害書き込み自動検出手法が数多く提案されている。

新田ら [2] は松葉ら [3] が提案した PMI-IR を用いた有害極性判定手法を拡張し、カテゴリ別関連度最大化手法 (CRM 手法) を提案している。新田らは松葉らが用いた種単語と呼ばれる有害極性単語を 3 カテゴリに分類し、種単語を各カテゴリに頻出の 3 単語ずつ含めることにより、90%を超える精度で有害書き込みを判定できたと報告している。

しかし、後に畠山ら [4] が CRM 手法を再現した際、精度が低下していたことが報告されている。畠山らは原因として種単語が時代に適さないものになっていたことを挙げ、有害極性単語の組合せや規模を変化させた結果、人間の判断による有害極性単語の更新が性能に良い影響を及ぼしたと報告している。

また、Zhang ら [5] は学習素性に音素を用いたニューラルネットワークによる有害書き込み検出手法を英語のために提案している。この手法は有害情報に見られる意図的なスペルミスや音素を用いて訂正しており、98.9%の精度で有害情報を判定が可能であった。

上記のように有害表現抽出手法は数多く提案されているが、課題も多く残されている。そこで、本稿では各手法における課題の整理に向けた、有害表現抽出手法の自動評価手法を提案する。具体的には、Papineni らにより提案された、機械翻訳における自動評価手法である BLEU [6] を有害表現抽出手法の評価へと応用、自動評価を行う。

2. 基本的な考え方

本章では、機械翻訳の自動評価手法 BLEU を活用した有害表現抽出手法の自動評価について説明する。

2.1 機械翻訳自動評価手法 BLEU

BLEU は Papineni らにより提案された機械翻訳における自動評価手法である。

BLEU では、機械翻訳文とプロの翻訳者による翻訳文の似ている度合いを数値で表すため、対象文を 2 つの文章間の N-gram 一致度 P_n (式 1) と、機械翻訳文の長さに応じたペナルティ係数 BP (式 2) を用いて表される式 3 により評価する。また、BLEU は 0~1 の範囲になり、0 は類似度なし、1 は完全な類似度を表している。

$$P_n = \frac{\text{ある参照訳との単語 } n\text{-gram 共有数の最大値}}{\text{MT 訳中の単語 } n\text{-gram 数}} \quad (1)$$

$$BP = \begin{cases} 1 & (c > r) \\ e^{(1-r/c)} & (c \leq r) \end{cases} \quad (2)$$

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N \frac{1}{N} \log P_n\right) \quad (3)$$

2.2 BLEU の有害表現抽出手法評価への応用

本研究でも、BLEU の評価の根拠であるプロの翻訳者による翻訳文が似ているほど機械翻訳文の質が良いという考え方と同様に、プロとしてネットいじめ書き込みを手作業で検索し通報するインターネットパトロールメンバーが集めた有害な文書と似た文書は有害であるとの考え方を根拠に、本提案手法による有害表現抽出手法の評価を行う。有害表現抽出手法が有害と判断した文書に対して、人間により有害であると判断されたネット上から抽出された文書を参照し、対象の文と有害な文の間の N-gram 一致度 P_n (式 4) を算出する。

$$P_n = \frac{\text{ある有害文書との単語 } n\text{-gram 共有数の最大値}}{\text{対象の評価文中の単語 } n\text{-gram 数}} \quad (4)$$

また、機械翻訳文の評価の場合とは異なり、有害な文章の評価では、とても短い文章の場合でも有害な文章である場合がある。そのため、有害な文章としての評価である本提案手法では、BLEU において評価対象の文の長さに応じてかけられるペナルティ係数 BP の適用は行わず、有害な文との近さを表す harmBLEU スコアは式 5 で求められる。

$$\text{harmBLEU} = \exp\left(\sum_{n=1}^N \frac{1}{N} \log P_n\right) \quad (5)$$

3. 提案手法

本章では、本提案手法の評価の妥当性について確認するため、松葉らが学校非公式サイト等のネット掲示板から抽出した書込み 2,998 件 (有害 1,509 件, 非有害 1,489 件) に対し、本提案手法による評価を行った。以下に評価の流れを示す。

1. データセットから評価対象の書込みを抽出する
2. 対象の書込み以外の有害書込み全てを参照文とし、対象の harmBLEU スコアを算出
3. データセット中の全書込みに対し、harmBLEU スコアを算出し、スコアの高い順に並べ替える
4. 順位毎にその順位以上の書込みを有害な書込み、順位未満の書込みを非有害な書込みとした際の適合率、再現率を算出

各順位における適合率、再現率を図 1 に示す。図 1 より、順位が 1000 位程度の閾値まで適合率が 0.8 以上と高い値が見られる。これより、有害な書込みは相対的にみて非有害な書込みより高い harmBLEU スコアを付与できており、本提案手法が有害書込みの評価に一定の有効性が有ることを示している。

4. 適用実験

次に、既存手法への評価の有効性について調査を行うため、同様のデータセットを用いた畠山らがアンケートを行い人手により有害と判断された単語を使用し改良を行った CRM 手法による有害書込みの判定実験の結果に対し、本提案手法を用いて評価を行った。

CRM 手法では、各書込みに対し有害度合いを表すスコア (SO 値) を付し、一定の閾値を用いて判定を行う手法であるため、本実験では各閾値毎の harmBLEU スコア平均を算出した。閾値による適合率、再現率と harmBLEU スコア平均の推移を図 2 示す。

図 2 より、harmBLEU スコア平均は閾値である SO 値の低下に伴い、緩やかに低下していることが読みとれる。これは改良 CRM 手法における SO 値が高いほど harmBLEU スコアも高い傾向にあることを示している。

5. 結果と考察

本適用実験の結果、本提案手法による改良 CRM 手法の評価は閾値が高い場合に高く、閾値を低くしていくにつれて評価も低くなっていることが分かった。これは、畠山らの改良 CRM 手法において、閾値が厳しい場合は適合率が約 8 割、閾値を緩くしていくと適合率が 5 割に近づき判定が機能なくなっていくという結果と一致しており、本提案手法による評価が有効である可能性を示している。

一方、閾値を低くした場合に着目してみると SO 値の上位約 1500 件を境に改良 CRM 手法の適合率の値が上昇を始めてい

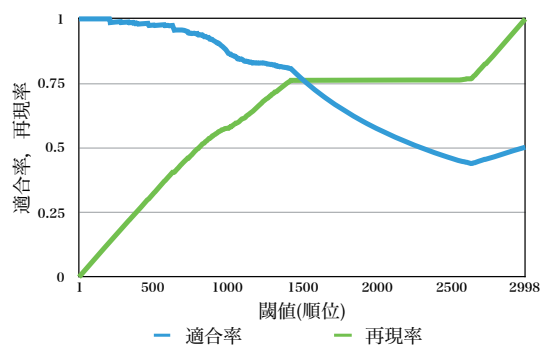


図 1: 各閾値における適合率と再現率

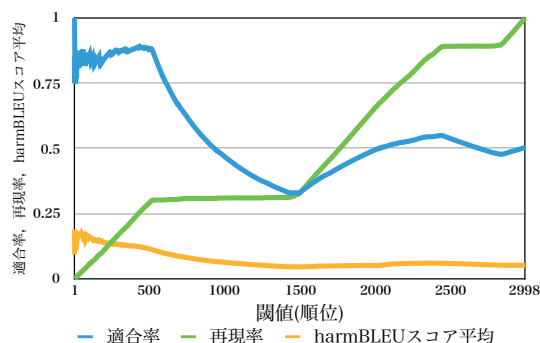


図 2: 各閾値における適合率、再現率と harmBLEU スコア平均

るが、harmBLEU スコア平均はほぼ一定の値を取っていることが読み取れ、本提案手法は有害である可能性が低い文書に対しては上手く評価が出来ていないことが分かった。

6. おわりに

今回、機械翻訳における自動評価手法である BLEU を有害表現抽出手法へと応用し、その手法を自動評価する手法を提案した。本提案手法による掲示板書込みの評価により、本提案手法が有害表現抽出手法の評価について一定の有効性を確認できた。しかし、今回の実験では参照文、対象文共に同じ掲示板の書き込みを利用しており、N-gram 一致度の算出の際において、個人名による影響が大きくなっていることが考えられるため、今後さらに異なるデータを用いた性能も確認する必要がある。

参考文献

- [1] 文部科学省:「「学校ネットパトロールに関する取組事例・資料集」(教育委員会等向け) 資料編第 2 章・巻末資料」, 文部科学省, (2012.9).
- [2] 新田大征, 榊井文人, ブタシンスキ・ミハウ, 木村泰知, ジェブカ・ラファウ, 荒木健治:「カテゴリ別関連度最大化手法に基づく学校非公式サイト上の有害書き込み検出」, 第 27 回人工知能学会全国大会発表論文集, (2013.6).
- [3] 松葉達明, 榊井文人, 河合敦夫, 井須尚紀:「学校非公式サイトにおける有害情報検出を目的とした極性判定モデルに関する研究」, 言語処理学会第 17 回年次大会発表論文集, P2-26(2011.3).
- [4] 畠山鈴生, 榊井文人, ブタシンスキ・ミハウ, 山本和英:「有害表現抽出に対する種単語の影響に関する一考察」, 第 30 回人工知能学会全国大会, (2016.6).
- [5] Xiang Zhang, Jonathan Tong, Nishant Vishwamitra, Elizabeth Whittaker, Joseph P.Mazer, Robin Kowalski, Hongxin Hu, Feng Luo:「Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network」, 15th IEEE International Conference on Machine Learning and Applications, (2016).
- [6] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu:「BLEU: a Method for Automatic Evaluation of Machine Translation」, 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, pp. 311-318, (2002.7).

Twitter データを用いたアプレイザル評価表現辞書に基づく 印象の自動指標化

Automatic Impression Indexing based on Appraisal Dictionary from Tweet

山田瑠奈
Runa Yamada

橋本 翔
Sho Hashimoto

山田篤拓
Atsuhiko Yamada

長田典子
Noriko Nagata

関西学院大学
Kwansei Gakuin University

With the development of social media, interest in text mining has been growing steadily during the past few years. Although these research studies have been conducted on product development and marketing, no research has been conducted to explore cultures that create new value not found in existing products. Therefore, we propose a method of extracting topics related to specific product domains and indexing impressions from tweets containing a wide range of descriptions of cultures and values. Our method is based on the appraisal dictionary as well as the structures of impressions for the target domain to become clear. A result of our method has been applied to the lawn culture. We extracted 13 ground topics and the impressions related to each of them. Moreover one of them was an unexpected concept. This result indicates the usefulness of the method in exploring a new concept of value related to a specific field.

1. はじめに

近年、Web 上に公開されている膨大なテキストデータから有益な情報を抽出するテキストマイニング技術に注目が集まっている。中でも、Twitter を代表とするソーシャルメディアの普及にともない、ソーシャルメディアから得られる商品やサービス、出来事などに対する記述文から評判分析を行い、商品開発やマーケティングに活用する研究が活発に行われている。特に Twitter では、評判分析の一つである感情分析の研究が広く行われており、ブランドごとに喜怒哀楽などの感情カテゴリ分類を行う研究 [Shukri 15] や、製品の構成要素に対する感情を推定する研究 [Hridoy 15] などが存在している。しかし商品開発、特にユーザのニーズを反映させた製品設計を行う上では、従来の Twitter データを用いた分析によって得られる「何が好き・欲しい」といったユーザの嗜好や感情に関する情報のみでは、その要因となる製品の特性や、機能に対する印象が明らかでなく、全体の設計への十分なフィードバックが行えない恐れがある。

一方で商品レビューを用いた研究 [山田 18, Hashimoto 19, 橋本 19] においては、製品に対する嗜好とそれに繋がる印象要因を明らかにする分析手法が提案されている。これらの研究では製品仕様（形態要素）と、そこから喚起される好き/嫌いといった「感情」との間に、製品の特徴や特性を表す「印象」を介した階層構造（図 1）が仮定されている。この階層構造に則って商品レビューから対象製品分野の印象構造を明らかにする印象の指標化 [山田 18, Hashimoto 19] と、製品の印象構造とユーザの感情との関係性を明らかにする感情の指標化を行っている [橋本 19]。上述のような印象を介した分析によって、製品に対するより詳細な情報を得ることができ、ユーザの感情を製品設計にフィードバックすることが可能となる。しかし商品レビューに記述されている内容は、既存の製品に関する印象や嗜好であり、周辺領域の記述やトレンドなど既存の製品にはない新たな価値を見出すことは困難である。そのため製品に対す

る印象だけでなく、その製品を使う理由や文化的背景、製品の使用を通して得られる印象など、対象製品ドメインに対する幅広い記述からの指標化が必要とされている。

そこで本研究では、文化や価値観など幅広い記述が含まれる Twitter データからの指標化の第一段階として、対象製品ドメインの印象構造を明らかにする印象の自動指標化を行う。



図 1: 感性の階層構造

2. 関連研究

1 章で述べたように商品レビューを用いた印象の指標化手法が山田ら [山田 18] や Hashimoto et.al. [Hashimoto 19] によって提案されている。これらの手法ではレビューから印象構造を構成する印象軸を抽出するにあたり、はじめに評価表現辞書に基づいた評価語（製品に対する評価を表す単語）の収集を行い、その後、収集した評価語をユーザの嗜好を表す感情語と製品の特徴や特性を表す印象語に分類している。

また印象軸の抽出においてはトピック抽出手法の一つである HDP-LDA [Teh 06] を用いている。LDA は、トピック分布にディレクレ事前分布を仮定しベイズ推定する手法であり、文書中の単語の「トピック」（話題やカテゴリ）を確率的に求める言語モデルである。さらに HDP-LDA は LDA をノンパラメトリックベイズに拡張した手法であり、LDA ではトピック数を指定する必要があるが、HDP-LDA では適切なトピック数が学習されるという特徴を有する。Hashimoto et.al. は LDA の特性を利用し、印象語のみを入力としたトピック抽出を適用することで得られた各トピックを対象製品分野の印象構造を表す印象軸として扱っている。

連絡先: 関西学院大学理工学部人間システム工学科

〒 669- 1337 兵庫県三田市学園 2 丁目 1 番地

E-mail: nagata@kwansei.ac.jp

3. 提案手法

3.1 Twitter データを用いた指標化の課題とアプローチ

本手法では Twitter データを用いた印象の指標化を行う。全体のアプローチとしては Hashimoto et.al.[Hashimoto 19] が提案した商品レビューに対する印象の指標化手法に則り、評価語の収集・分類を行った後に、印象語のみを用いて印象軸を推定する。しかし Twitter データへの印象の指標化手法の適用には 2 つの課題が存在する。1 つ目の課題は、評価対象となる文化や価値観があらかじめ特定できないことである。2 つ目の課題は、Twitter では 1 文書に含まれる印象語数が少ないため、先行研究において印象軸推定で用いている LDA は、適切に機能しないことである。

これらの課題を解決するため、本手法では特定製品ドメインにおける基本的な話題を、名詞を対象とした LDA より抽出し、各話題に対して印象語の分散表現を利用して推定した印象軸との関連付けを行う。

本手法の流れとして、対象製品ドメインにおける話題抽出、評価語の収集・分類、印象軸の推定、対象製品ドメインにおける話題と印象軸との関連度の算出、の 4 つの手順で印象の指標化を行う。また多様な文化における本研究の利用可能性の向上のために、分析は英語で記述された文書を対象に行う。そのため英語の評価表現辞書を作成する必要がある。そこでまず、指標化の基盤となるアプレイザル評価表現辞書の英語対応を行う。

3.2 英語アプレイザル評価表現辞書の作成

先行研究 [山田 18] では、評価表現のカテゴリ情報が付与されている日本語アプレイザル評価表現辞書に基づいて評価語を分類している。アプレイザル理論において評価表現には「外評価」と「内評価」という 2 種類のカテゴリがあるとされている [佐野 11]。内評価とは「嬉しい」や「楽しい」といった評価対象に対する評価者の感情や感情を表す行為を示す表現であり、外評価とは「やわらかい」や「きれいだ」といった評価対象の特徴を示す表現である。つまり図 1 において、内評価は感情層、外評価は印象層に位置づけられると考えられ、これらを分類した評価表現辞書もとに印象語と感情語の分類を行っている。

本研究では、クラウドソーシングサービスと和訳情報を用いて、Opinion Lexicon [Hu 04] 内に含まれる 6,789 語の評価語を「内評価」と「外評価」のカテゴリに分類することにより英語対応を行う。クラウドソーシングを用いた分類では 1 件の回答あたり 25 語の分類を行い、計 2,009 件の回答結果によって全体の約 74%にあたる 5,021 語を分類した。さらにクラウドソーシングを用いた分類できなかった評価語に関しては、各評価語の和訳情報と日本語アプレイザル評価表現辞書との対応により分類を行った。その結果、全体の約 93%にあたる 6,317 語を分類した。各分類結果の内訳は内評価が 1,276 語、外評価が 4,985 語、どちらにも当てはまるものが 53 語であった。本研究では、この分類結果を英語アプレイザル評価表現辞書として定義する。作成した辞書内の単語の例を表 1 に示す。

表 1: 英語アプレイザル評価表現辞書内の単語例

内評価	外評価
like love enjoy beautiful hate happy cool nice perfect well enjoy favorite	good great fresh best hard clean hot fast easy warm cold wrong

3.3 Twitter データを用いた印象の指標化手法

(1) 対象製品ドメインにおける話題抽出

はじめに名詞のみを対象としたトピック抽出により対象製品ドメインにおける話題を抽出する。トピック抽出には HDP-LDA を用いる。前節の課題で述べたように、一文書あたりにおける単語数が少ない（印象語だけで構成される文書など）場合、LDA は適切に機能しないが、名詞であれば十分な単語数が確保できると考えられるため、適用可能と判断した。

トピック抽出の入力文書群としては、ストップワード、ユーザ名、URL、ハッシュタグ、絵文字・記号の削除を行った前処理後のデータから、品詞が名詞と推定された単語を抽出したものを用いる。なお品詞の推定は形態素解析ツール NLTK を用いた。このとき、極端に単語数が少ない文書では適切なトピック抽出が行われないため、ツイート中の名詞数が 3 語以下の文書は分析対象としなかった。

また HDP-LDA によって得られるモデルには初期値や乱数による依存性があるため、同一のハイパーパラメータ設定で複数のモデルを比較し最適なモデルを選ぶ必要がある。そこでモデルの汎化性能を表す指標である Perplexity に基づいて、複数のモデルから最良（Perplexity が最も低い）のモデルを採用する。本研究では 10 回の学習のうち Perplexity の値が最も低いモデルを採用した。しかしながら、採択したモデルにおけるすべてのトピックが評価の基盤と言えるほどの頑健性があるとは考えづらい。そのため得られたトピックの中でも、10 モデル中の 5 モデルで共通して抽出されたトピックのみを頑健性の高いトピックとして採用した。

(2) 評価語の収集・分類

印象軸の抽出を行う上で必要な印象語を抽出するため、前処理後の Twitter データから評価語の収集・分類を行う。英語アプレイザル評価表現辞書によって内評価と分類された単語を感情語、外評価と分類された単語を印象語として評価語を収集・分類する。

(3) 印象軸の推定

印象語の中には類似した印象を表現する複数の印象語が存在しており、類似した方向性を持つ単語を束ねることで軸となる印象が推察されることが考えられる。そのため、本研究においては単語を分散表現に変換しクラスタリングすることで印象軸の推定を行う。

単語の分散表現の学習方法としては Word2Vec [Mikolov 11] を用いる。Word2Vec とは、Skip-gram モデルと呼ばれる二層ニューラルネットワークを用いて単語の分散表現を得る手法である。次に Word2Vec によって得られた印象語の分散表現によって印象語間のコサイン類似度を求め、それに基づいた距離を算出し、ウォード法による階層クラスタリングを行う。クラスタ数はクラスタの集中度と孤立度に関する指標 [斎藤 06] をもとに決定した。

(4) 対象製品ドメインにおける話題と印象軸の関連度の算出

名詞を対象としたトピック抽出によって得られた対象製品ドメインにおける話題と、印象構造を表す印象軸との関連度を算出する。これによって話題に対するユーザの印象を特定できたり、あるいはユーザが感じる印象を象徴する話題の推定が可能となる。

関連度は、話題として抽出した各トピックにおいて重要度の高いいくつかの名詞と、印象軸を構成する各印象語との分散表現のコサイン類似度に基づいて算出する。本手法ではトピッ

クにおける単語（名詞）の重要度として term-score を用いて、上位 10 語の名詞をそのトピックにおいて重要度の高い名詞とした。コサイン類似度は対象のすべての名詞と印象語間で計算し、その平均類似度を話題と印象軸との関連度として算出する。

4. 実験

本章では、実際の Twitter データに対して手法を適用し、手法の有用性を確認する。

4.1 対象製品ドメイン

本研究では芝刈機を対象製品とし、その背後にある芝生文化（芝文化）[北村 98] をドメインとする。芝文化とは、芝生に関わる価値観を反映した物心両面の行動様式を示しており、気候風土や歴史的背景の違いから国によって異なると考えられている。この芝文化の印象構造を明らかにすることで、芝刈機など芝に関連する製品のデザインの支援への活用が期待される。

4.2 データ概要

芝文化に関するキーワード検索（e.g. yard work, cut grass, lawn care）によって得られたツイートをデータとして扱った。本実験ではアメリカ、イギリス、欧州連合加盟を分析対象国とした。これらの国は日本に比べ芝の手入れに適した気候であることや、個人住宅での芝庭が多いことから、芝文化調査の対象国とした。取得したツイート数 122,877 件で、形態素解析の結果、総単語数は 1,178,039 語、総語彙数は 37,695 語となった。

4.3 結果・考察

芝文化に関する Twitter データに対する適用結果を印象の指標化の 4 つの手順ごとに示し、考察する。

(1) 対象製品ドメインにおける話題抽出

最も精度の高いモデルのハイパーパラメータは $\alpha = 2.0$, $\beta = 0.1$, $\gamma = 0.01$, Perplexity の値は 50.731 となり、19 個のトピック（話題）が抽出された。抽出されたトピックのうち、頑健性の高かった 13 個のトピックを表 3 に示す。トピック 1 や 3 のような文化的背景を連想させるトピックや、トピック 5 や 11 のように特定の話題が顕著に表れているトピックが抽出することができた。その他のトピックは芝庭・芝刈りに関連するトピックと考えられる。

表 2: 芝文化における話題抽出結果

	解釈結果	term-score の高い上位 5 単語
1	武器	people, gun, car, truck, weapon
2	人・家	man, woman, men, car, tire
3	低賃金労働	lawn, program, worker, people, food
4	近隣	lawn, people, home, neighbor, business
5	風景設計・造園	landscape, architect, designer, space, city
6	季節	spring, lawn, season, news, summer
7	庭の手入れ	yard, leaf, grass, neighbor, weed
8	子供	kid, school, car, paper, neighbor
9	天候	weather, yard, winter, week, snow
10	植物	plant, garden, lawn, flower, grass
11	スポーツ・ゴルフ	game, golf, shoe, shirt, yard
12	買い物	yard, dog, home, car, grocery
18	芝刈機	lawn, mower, gas, equipment, power

(2) 評価語の収集・分類

評価語の収集・分類結果の一部を表 2 に示す。収集された評価語は 2,083 語、うち内評価を表す評価語が 717 語、外評価を表す評価語が 1,366 語となった。抽出された評価語は、対象製品ドメインにおける話題への感情・印象を表している評価語であることがわかる。最も多くのツイートで用いられた

「good」（良い）という印象語は、商品レビューにおいても最も出現頻度が高かったことから、普遍的な印象語であるといえる。「fresh」という単語が「good」に次いで出現頻度が高いことから、芝生の新鮮さや芝刈りによってすっきりしたといった記述が多いことが推定可能である。感情語と印象語の出現頻度を比較すると、感情語の方が圧倒的に高い頻度で出現している。これは、Twitter 上には感情的な表現を投稿されることが多いという特性を示している。また、「like」や「love」などの個人の感情に関する評価語は適切に感情語として分類されており、手法の有効性が確認された。

表 3: 評価語の収集・分類結果（一部抜粋）

感情語		印象語	
like	68982	good	3169
love	22907	fresh	3019
well	11612	great	1910
nice	8481	right	1837
enjoy	7659	best	1076
fun	7190	bad	1068
hate	6896	free	1005
beautiful	6372	ready	974
happy	5956	hard	851
damn	5309	enough	777

(3) 印象軸の推定

0.05%以上のツイートに含まれる印象語 77 語を対象としたクラスタリングの結果、13 の印象軸が得られた。表 4 は各印象軸を構成する印象語と印象軸の解釈を示している。注目すべき点として、印象軸 1 や 11 のような印象語の中でも強い感情に結びつくような軸が表れた。これは山田ら [山田 2018] のようなレビュー文章を対象とした研究ではあまり見られない特徴であり、Twitter データの特性であると考えられる。また、印象軸 8 や 10, 12, 13 は物理的要因に対する印象を表す軸であり、印象語の集合として解釈可能な結果が得られていると考えられる。さらに印象軸 4 は文化特有の社会問題に対する印象と解釈できる。

(4) 対象製品ドメインにおける話題と印象軸の関連度の算出

対象製品ドメインにおける話題抽出で得られた各話題に対して 3.3.4 項で提案した手法を用いて関連度を算出した結果、特に関連度の高かったペアを印象軸と話題の 2 つの観点から取り上げたものを表 5 と表 6 に示す。

表 5 を見ると、庭の手入れには有用性が求められており、芝刈りの場面では植物に対して憤慨するような印象を持つことが推測できる。また、印象軸と話題のどちらの観点においても、印象軸 4 と話題 3 の関連性が見られたことから、芝刈りは低賃金労働であるという社会的背景があり、貧困性・犯罪性といった印象を持たれていることが示唆された。

このような知見はレビュー等を用いて製品同士を比較するなかでは得られがたい知見である。「植物に対してもいらつかない」「芝刈りは単純労働であるというイメージをくつつがえす」といった新しい価値はプロダクトデザインに有用であると考えられる。これらより、本研究の手法で特定製品ドメインにおける製品評価につながる可能性のある価値観や文化に対する新たな知見を獲得できることが示された。

5. おわりに

本研究では、Twitter 上に存在している製品分野に対する幅広い記述から、製品を使う理由、文化的背景、並びに製品の使用を通して得られる価値観に対する印象を指標化する手法を提案した。提案手法では、はじめにアプレイザル理論に基づいて

表 4: 印象語 77 語を対象とした印象軸推定結果

	解釈結果	印象軸を構成する印象語
1	憤慨	fuck, fake, dumb, shit, stupid
2	良否・正誤	good, right, wrong, weird, bad, clearly, positive, ridiculous, strong, terrible
3	高品質感	enough, pretty, worth, odd, decent, smart, afraid, expensive
4	貧困性・犯罪性	clean, rich, honest, fine, lazy, poor, cheap, illegal, criminal
5	適性	important, proper, properly, complimentary
6	有用性	available, leading
7	孤高さ	free, top, proud
8	安全性	safe, hard, tough, easy, difficult, useful, dangerous
9	完全性	dead, ideal
10	遅速・円滑性	enough, fast, weed, clear, slow
11	称賛	wonderful, lovely, great, best, excellent, awesome, amazing
12	寒暖・硬軟	ready, hot, fresh, warm, cold, bright, delicious, soft
13	生産性・静騒	super, lucky, quiet, loud, productive, wild, dark

表 5: 印象軸に対する話題の関連度算出結果

印象軸	関連度が高い話題
4【貧困性・犯罪性】	3【低賃金労働】
5【適正】	6【季節】
6【有用性】	5【風景, 建築】
7【孤高さ】	8【子供】
8【安全性】	1【武器】
10【遅速・円滑性】	7【庭の手入れ】
12【寒暖, 硬軟】	9【天候】

表 6: 話題に対する印象軸の関連度算出結果

話題	関連度が高い印象軸
3【低賃金労働】	4【貧困性・犯罪性】
4【近隣】	1【憤慨】
7【庭の手入れ】	6【有用性】
8【子供】	5【適正】
10【植物】	1【憤慨】
11【スポーツ・ゴルフ】	2【良否・正誤】
12【買い物】	4【貧困性・犯罪性】

英語アプレイザル評価表現辞書を作成した。次にドメインにおける基本的な話題をトピック抽出により推定し、その後、作成した辞書と単語の分散表現を利用した印象軸の抽出、および話題との関連度を算出した。

実験では芝刈機を対象製品とし、本手法を芝文化に関するツイートに適用した。トピック抽出によって得られた芝刈機分野における話題と印象軸の関連性を確認した結果、製品評価につながる可能性のある知見を獲得することができ、手法の有用性が示された。

今後の課題としてはトピック抽出の精度向上や、共起率を考慮した関連度の算出などを行うことにより、得られる情報の多様性・有用性を向上していくことが挙げられる。

謝辞

本研究の一部は国立研究開発法人科学技術振興機構（JST）の研究成果展開事業「センター・オブ・イノベーション（COI）プログラム」の支援によって行われた。また、共同研究を通してデータをご提供いただき、有益なご議論、ご提案をいただきました（株）本田技術研究所パワープロダクツ R & D センターの関係各位に深く御礼申し上げます。

参考文献

[Hashimoto 19] Hashimoto, S., Yamada, A., and Nagata, N: A Quantification Method of Composite Impression of

Products by Externalized Evaluation Words of the Appraisal Dictionary with Review Text Data, International Journal of Affective Engineering, Online ISSN 2187-5413, Advance publication Released January 25 (2019)

[橋本 19] 橋本翔, 山田篤拓, 長田典子: レビューデータを用いた評価表現辞書に基づく感性の自動指標化, 人工知能学会全国大会論文集 (2019)

[Hridoy 15] Hridoy, S. A. A., Ekram, M. T., Islam, M. S., Ahmed, F., and Rahman, R. M.: Localized twitter opinion mining using sentiment analysis, *Decision Analytics*, Vol.2, No.1, p.8 (2015)

[Hu 04] Hu, M. and Liu, B.: Mining opinion features in customer reviews, *AAAI 2004*, pp.755-760 (2004)

[北村 98] 北村文雄: 日本における芝生文化の展開, 芝草研究, vol.27, No.supplement1, pp.1-8 (1998)

[斎藤 06] 斎藤堯幸, 宿久洋: 関連性データの解析法, 共立出版 (2006)

[佐野 11] 佐野大樹: 『日本語アプレイザル評価表現辞書（態度表現編）』の構築-評価の多様性を捉えるための言語資源の開発-, 言語処理学会第 17 回年次大会発表論文集, pp.115-118 (2011)

[Shukri 15] Shukri, S. E., Yaghi, R. I., Aljarah, I., and Alsawalqah, H.: Twitter sentiment analysis: A case study in the automotive industry, *AEECT, 2015 IEEE Jordan Conference on*, pp.1-5 (2015)

[Teh 06] Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M.: Hierarchical Dirichlet processes, *Journal of the American Statistical Association*, Vol.101, issue 476, pp.1566-1581 (2006)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *ICLR Workshop 2013* (2013)

[山田 18] 山田篤拓, 橋本翔, 長田典子: レビューデータを用いた評価表現辞書に基づく印象の自動指標化, 日本感性工学会論文誌, Vol.17, No.5, pp.567-576 (2018)

潜在的な旅行者への宿泊施設の提案

Catch Me if Yahoo Can: Hotel Recommendation for Potential Travelers using Transit App Log

丸山 三喜也*¹
Mikiya Maruyama高濱 康太郎*¹
Kotaro Takahama坪内 孝太*¹
Kota Tsubouchi寺岡 照彦*¹
Teruhiko Teraoka*¹ ヤフー株式会社
Yahoo Japan Corporation

We suggest new hotel recommendation method that catches user context from transit log and recommends information before they need. People who are planning to travel or business trip tend to search a route to go to the destination beforehand. In this case, we use the transit search log to find out the potential travellers and recommend hotels near their destination. As a result, we confirmed that the method is more effective to reach the potential travellers than existing methods for potential user. This paper is consideration and summary of the method's verification.

1. はじめに

インターネットやスマホの普及により、情報がすぐ手に入るようになった反面、本当に必要とする情報を得るまでに時間と労力がかかることが少なくない。一般的に旅行は、宿泊施設、観光施設、飲食店、交通機関など、複合的な計画や調査が必要であり、ユーザにとって大変な手間のかかるものである。一連の調査や計画の早い段階で、ユーザの意図やコンテキストを読みとって、適切な情報を先回りして推薦できれば利便性が大きい。そこで我々は乗換検索に注目し、そこから潜在的な旅行者を予想して、宿泊施設の推薦を行うことを考えた。但し、移動距離や実際の移動日までの日数はユーザ毎にまちまちであり、どの程度の移動距離に対して、どのタイミングで、どのような施設を推薦すればよいかを決めることは難しい。

これらを明らかにするため、ヤフー乗換案内の検索ログを利用した実験を実施した。ルート検索の出発地と目的地から移動距離を計算し、一定以上の距離を移動する潜在的な旅行者にヤフートラベルを使って移動先周辺の宿泊施設の推薦をメールで配信した。

実験結果によれば、移動距離や検索日と実際の移動予定日の差が大きくなるほど、宿泊予約に至る傾向が見られた。

2. 関連研究

旅行や出張など長距離移動に伴う情報推薦では、観光場所や次に行くスポットの推薦[E. Palumbo 17]や、場所や時期、口コミからの宿泊施設の推薦[M. A-Ghossein 18]などの研究がある。本論文のように、ユーザが長距離移動の交通手段を検討中に、先読みして宿泊施設を推薦することは、あまり検討されていない。

3. データセット

今回考案した手法では以下のデータを用いる。

- 乗換案内ルート検索ログ
- ホテルマスタデータ

乗換案内ではユーザがルートを検索する毎に検索結果をログとして蓄積しており、いつでもどこに、どのようなルートで移動する予定かを知ることができる。このログを用いて長距離移動ユーザを抽出する。またレコメンドするホテルの選定のため、ホテルマスタデータを用いる。到着地に近いホテルをレコメンドするため、緯度経度で紐付けられるようになっている。

4. 考案手法

4.1 対象となるユーザーの抽出

乗換案内内のルート検索ログの出発地/到着地から距離を取得。そこから具体的に何 km 移動するかを算出する。長距離移動の定義は最低 100km 以上とし、その距離以上移動する予定、かつメールの配信日以降の日付で検索したユーザーのみレコメンド対象ユーザー群として扱う。

4.2 送信するメールコンテンツの生成

抽出した対象ユーザーがルート検索時に指定した到着地を基に、ホテルマスタデータからレコメンドするホテル情報を取得し、以下のようなメールコンテンツを生成する。



図 1. 生成したメールコンテンツ

またレコメンドするホテルは、到着地に指定された場所より半径 10km 以内に存在し、距離の近い順とした。生成したメールは、ユーザーがルートを検索した翌日、又は翌々日に配信されるよう設定した。長距離移動ユーザーの抽出からメール配信までの流れを下図に示す。

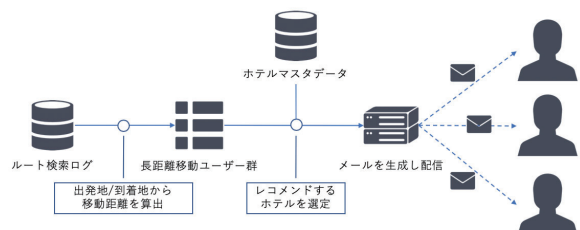


図 2. メール配信までの流れ

5. 検証結果

5.1 メール配信結果

3 日間メールの配信を行ったところ、それぞれ以下のような結果となった。

表 1. メール配信結果

配信日	集計期間	配信数	クリック数	予約数
12/6	10 日間	77,065	134(0.17%)	20(14.9%)
12/8	10 日間	53,235	106(0.20%)	15(14.2%)
1/19	10 日間	44,112	108(0.24%)	19(17.6%)

集計期間は、メールを配信してから何日以内に予約に至ったユーザーをメール配信効果のあったユーザーとみなすかを指す。今回は全て 10 日間としたので、それぞれ 12/6～12/15, 12/8～12/17, 1/19～1/28 の期間に、メールをクリックして予約に至ったユーザーの人数を予約数に記載している。予約数の括弧内の数値は、クリック数に対する CVR(コンバージョンレート)を指す。1/19 に関しては、対象者を 200km 以上移動する予定のユーザーに変更したので配信数は減っているが、予約数の CVR を比較すると向上していることが見て取れる。

5.2 予約に至ったユーザーの検索した条件

予約に至ったユーザーの移動距離と実際に移動するまでの期間をマトリクスにしたものを図に示す。

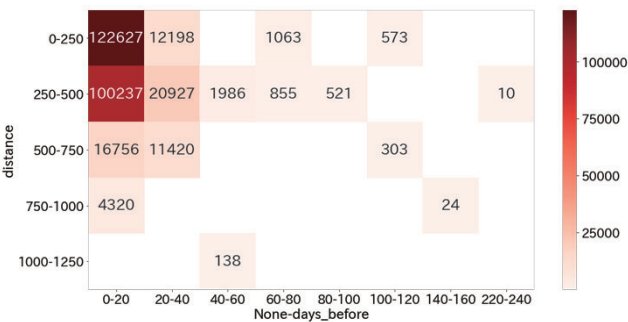


図 3. 配信ユーザー

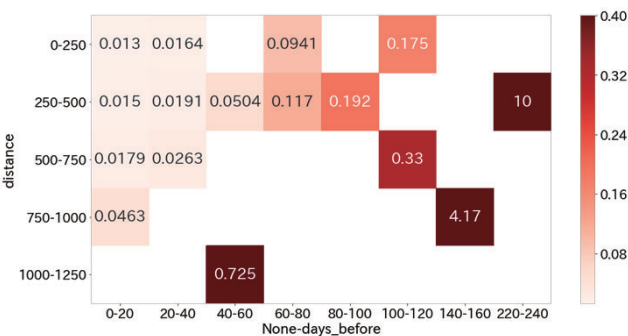


図 4. 配信したユーザーに対する予約ユーザーの比率

横軸が何日前に検索したか、縦軸が移動距離、枠内の数値が比率を指している。一瞥しただけでも、移動距離と実際に移動する日までの間が長いほど予約する可能性が高まるが見て取れる。

6. 考察とまとめ

6.1 考察

検証の結果、移動距離が伸びるほど予約に至る可能性が高まる傾向にあった。特に 40-60 日前にルート検索を行った予約ユーザーではその傾向が顕著に出ており、最小値と最大値で約 14 倍もの差がある(図 4 参照) また距離だけでなく、検索日と実際に移動する日までの差が大きいほど予約に繋がりがやすい傾向が見て取れた。各セグメント別に見ると、約 10 倍 - 1000 倍の差が出ていることが確認できる。今回は試行回数が少ないため一概には言えないが、回数を増やしても同様の結果が出てくるのではないと思われる。今回の結果をもとに、より効果的に宿泊情報を必要としているユーザーに対して、必要な情報を届けることができるようになった。

通常トラベルを来訪していないユーザへの訴求の場合、既存ユーザへの訴求よりも効果は乏しくなる傾向になるが、今回路線データを利用することにより、トラベル既存ユーザに迫るレコメンド施策の効果を確認できた。新規ユーザに対する施策の中では非常に高いスコアといえる。今回の結果を利用して、距離や検索日を最適化することで、さらに的確な潜在顧客にリーチできるようになると考えられる。

一方で、長距離移動ユーザーを抽出するために利用した乗換案内では当日検索するユーザーが最も多く、1 ヶ月以上先の検索をするユーザーの数はそれほど多くない。より多くの潜在的な旅行者を検知し情報を提供するには、乗換案内だけではなく他のデータを用いて検知する必要がある。例えば、旅行者が事前に検索するのはルートだけでなく、目的地周辺の観光施設や飲食店の情報といった可能性も高い。そういった検索クエリから旅行を計画していることが判別できれば、ルートを検索していなくとも潜在的な旅行者を抽出できると考えられる。

またユーザーへのアプローチも、今回メールという手段を用いたが、そもそもメールが使っていない、他のメールに埋もれ見えないユーザーも多数存在すると考えられる。予約に至ったユーザーの年齢層を調べたところ、40 代以上のユーザーが 90% 以上を占めていた。旅行に行きやすいのが金銭的に余裕の出るこの年代なのかもしれないが、乗換案内のユーザー層等を加味するとこの数値は非常に低いため、若年層がほとんどメールを利用しておらずリーチ出来ていない可能性を示唆している。乗換案内の場合、アプリや web でルートを検索した時点で宿泊施設をレコメンドするといったやり方の方が、全てのユーザーにリアルタイムで情報を届けられる。今後はメール以外のアプローチ方法や乗換案内以外のデータ活用も検討していきたい。

6.2 結論

移動距離、及び検索してから実際に移動するまでの期間に着目することで潜在的な旅行者を判別することができ、適切な宿泊情報を提供することで、高い確率で予約に繋がることが判明した。

この知見をもとに、サービス間の連携を強化し、より効果的なユーザーへのアプローチを試みる。

参考文献

[E.Palumbo 17] E.Palumbo, and G. Rizzo, "Predicting Your Next Stop-over from Location-based Social Network Data with Recurrent Neural Networks", pp.1-8, RecTour 2017 (2017).
[M. A-Ghossein 18] M. A-Ghossein, T. Abdesslem, and A. Barre, "Exploiting Contextual and External Data For Hotel Recommendation", pp.323-328, UMAP'18 (2018)

ニューラル機械翻訳による公的文書平易化

Official document simplification using neural machine translation approach

丸山 拓海^{*1} 山本 和英^{*1}
Takumi Maruyama Kazuhide Yamamoto

^{*1}長岡技術科学大学大学院 電気電子情報工学専攻
Electrical, Electronics and Information Engineering, Nagaoka University of Technology

Official documents are documents distributed at public facilities such as city halls, hospitals and schools. These documents contain a lot of important information for living. However, they are difficult for non-native speakers because they contain difficult vocabulary and expressions. Therefore, official documents must be simplified. We try to simplify official document using machine translation approach. We use a parallel corpus of the original and three kinds of simplified ones including literal translation, free translation and summary. They are rewritten by 40 Japanese teachers. We adapt several methods for low-resource machine translation such as pre-trained embeddings and sharing encoder, decoder and output embeddings (tied-embeddings). The result shows that Transformer can simplify official document using pre-trained embeddings and tied-embeddings in spite of low resource. Performance improvement using several methods of low resource machine translation shows that Transformer can improve performance more than other methods by extending training data.

1. はじめに

近年、たくさんの外国人が日本を訪れている。その数は年間で2,800万人^{*1}にのぼる。また、日本に住む外国人は232万人^{*2}であり、増加の傾向にある。このような状況の一方、日本社会においては、日本語で情報提供されることが多い。情報提供の最も理想的な方法は、すべての外国人に対してそれぞれの母語で情報を伝えることであるが、世界には多くの言語が存在するため、現実的には不可能である。このような言語の選択という問題を解決するための情報提供の手段として、近年、「やさしい日本語」という考え方が注目されている。ここで、「やさしい日本語」とは、簡単な語彙や文構造のみを用いて、日本語に不慣れな外国人にもわかりやすくした日本語のことを指す。また、国立国語研究所の全国調査では、日本に住む外国人のうち、英語を理解できる人の割合よりも、簡単な日本語を理解できる人の割合の方が多くと報告されており[岩田 10]、英語よりも簡単な日本語の方が外国人には伝わりやすいことが知られている。

本研究では、公的文書を対象とする。公的文書とは、市役所や病院、学校等の公共施設で配布される文書であり、これらの文書は生活する上で重要な情報を多く含んでいる。しかし、日本語初学者が学習する文に比べ、難解な語彙や公的文書に出現する固有の表現も含み、理解が困難であるため、平易化が必要な文書である。我々は、日本語教育者が公的文書の日本語を逐語訳、意識、要約の3段階の「やさしい日本語」に書き換えたものをコーパスとし、機械翻訳的なアプローチにより、文単位の平易化を試みる。

2. 関連研究

公的文書平易化の試みとして、空らの研究がある。彼らは次のような手順に従って、公的文書の平易化を行った[空 13]。

1. 重要部分の抽出
2. 短文化
3. 表現意図を用いた図示への変換
4. 「やさしい日本語」への変換

ステップ1からステップ3までの処理によって、文の構造を読みやすくしたのちに、彼らが構築した換言辞書を用いて、語彙的な平易化を行っている。平易化システムの出力を「日本語としての正しさ」と「やさしさ」という観点で人手により評価した結果、公的文書をやさしく書き換えることには成功したが、抽出した換言ルールを直接適応させているため、文法的な誤りや意味を保持できない場合が多いことが報告されている。また、直接的表現へ言い換えを行った研究[Moku 12]では、変換ルールの規模が小さく、平易化としての効果があり得られなかったことが報告されている。

一方で、松田らは、統計的機械翻訳を用いて公的文書を「やさしい日本語」へ変換する研究を行っている[松田 09]。BLEUによる精度評価では、20ポイントと低い値に留まった結果となった。また、原文と原文から語彙的なレベルで平易化を行った逐語訳、そして機械翻訳による翻訳結果の3つを人手で、1:比較的良質、2:解読不能・翻訳誤り、3:変化なし、の3段階に評価したところ、比較的良質な変換とされた翻訳結果は7.5%と少なく、そのうちのほとんどが見出しのような短い句であったことが報告されている。この理由として、対訳コーパスが少ないことや、公的文書特有の記号や括弧表記がノイズになっていることを指摘している。また、適切にドメインを設定することによって、性能改善ができる可能性についても述べられている。

機械翻訳の分野では、統計的機械翻訳の性能をはるかに上回るニューラル機械翻訳が提案されている[Bahdanau 14, Luong 15, Gehring 17, Vaswani 17]。近年、テキスト平易化においても、ニューラル機械翻訳を用いるアプローチが盛んに研究されており[Maruyama 17, Nisioi 17, Zhang 17, Surya 18, Zhao 18]、既存の統計的機械翻訳をベースとしたモデルより

連絡先: 丸山拓海, 長岡技術科学大学大学院電気電子情報工学専攻, 新潟県長岡市上富岡町 1603-1, maruyama@jnlp.org

^{*1} https://www.jnto.go.jp/jpn/statistics/visitor_trends

^{*2} <https://www.e-stat.go.jp>

も、高い平易化性能を記録している。ニューラル翻訳をベースとしたモデルでは、学習に大量のデータが必要となるが、テキスト平易化のための大規模な対訳コーパスを用意することは容易ではない。

本研究では、ニューラル機械翻訳のモデルが公的文書の平易化という言語資源が少ないタスクにおいて、どの程度効果を発揮できるのかを検証する。

3. データセット

本研究では、空らが利用した平易化コーパス [空 13] と同じもの (以下、「公的文書書き換えコーパス」と呼ぶ) を利用する。これは「やさしい日本語」のプロジェクトで作成されたものであり、約 40 名の日本語教師が、市役所や病院、学校等の公共施設で配布される公的文書を「やさしい日本語」に書き換えたものである。このコーパスは原文である公的文書 1,101 文書と共にその逐語訳、意識、要約という 3 段階の翻訳を含む対訳コーパスである。それぞれの翻訳の位置付けは下記の通りである。

- 逐語訳: 日本語文の難解な語彙や句をやさしい表現に、書き換えたもの。
- 意識: 文意等を損なわないように可能な限り、やさしい表現に書き換えたもの。
- 要約: 可能な限り文を平易化したもの。

これらは一定の文法基準 [庵 08] と旧日本語能力試験 2 級 (現試験における N2) レベルの語彙のみに制限されている。コーパスにおける「やさしい」の基準は日本語教師の主観である。各翻訳の例を以下に示す。

- 原文: ニュース等で報道されておりますように、世界的に新型 (豚) インフルエンザの流行が危惧されています。
- 逐語訳: ニュースなどにもあるように、世界中で新型インフルエンザの流行が心配されています。
- 意識: さて、ニュースでもありますが、世界中で新型インフルエンザが増えています。
- 要約: さて、世界中で新型インフルエンザが増えています。

上記の例では、原文に対して 3 種類の翻訳文が存在するが、場合によっては、翻訳するのではなく原文全体を削除する事例も存在する。そのため、原文と逐語訳・意識・要約のそれぞれの文数は必ずしも一致しない。本コーパスの文数、平均文長、語彙数を表 1 に示す。

表 1: 公的文書書き換えコーパスの統計量				
	原文	逐語訳	意識	要約
文数	35,861	35,809	32,841	27,588
平均文長	25.32	28.55	26.20	24.74
語彙数	14,848	11,337	10,243	9,259

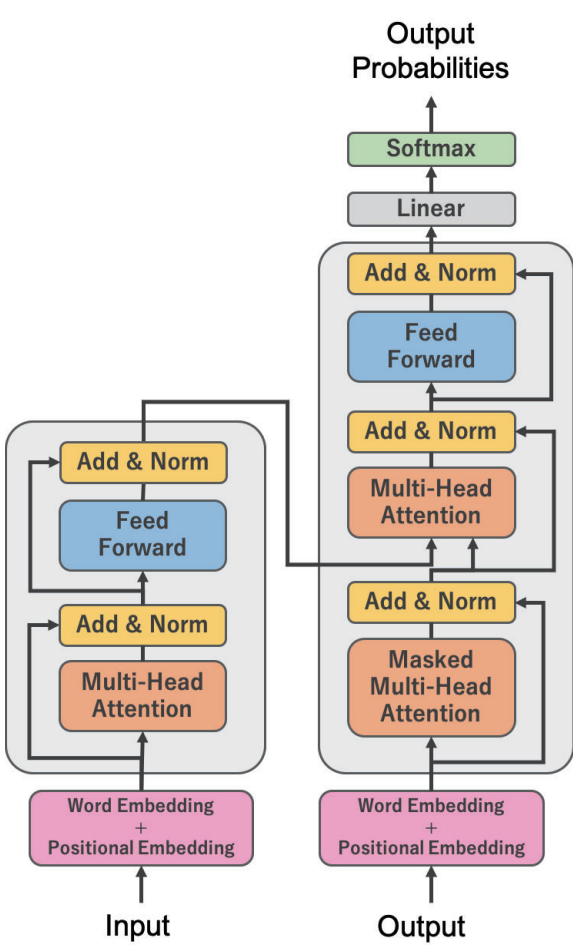


図 1: Transformer

4. 実験方法

今回は、機械翻訳モデルを利用し、原文から逐語訳・意識への文単位の変換を試みる。我々は公的文書書き換えコーパスを、学習データ 34,000 文、開発データ 1,000 文、評価データ 861 文のように分割した。これらのデータを用いて、翻訳モデル (4.1 節) をトレーニングし、モデルの出力結果を BLEU, SARI(4.2 節) により評価する。

4.1 Transformer

本研究では、翻訳モデルとして、Vaswani らが提案している Transformer を用いる [Vaswani 17]。モデルの概略図を図 1 に示す。Transformer の encoder (図 1 左側) は、multi-head self-attention と Feed Forward Neural Network で構成される層を L 段積み上げた構成となっている。multi-head self-attention は、次式に従って、前段の隠れ状態 $e_{(s',l)}$ から隠れ状態 $e_{(s,l)}$ を計算する。

$$e_{(s,l)} = \sum_{s'} \alpha_{(s',l)}^{enc} e_{(s',l-1)} \tag{1}$$

$$\alpha_{(s',l)}^{enc} = a(e_{(s',l)}, e_{(s',l-1)}, H) \tag{2}$$

ここで、 $\alpha_{(s',l)}^{enc}$ は、l 層目、s' ステップの attention distribution を意味する。a(・) を multi-head self-attention 表す関数、H はヘッド数を意味する。一方、decoder (図 1 右側) は、

encoder と同様な機構に加え、encoder の出力に対する attention 機構が存在する。1 段目の multi-head attention では、encoder と同様に次式に従って隠れ状態 $d_{(s,l)}$ を計算する。

$$d_{(s,l)} = \sum_{s'} \alpha_{(s',l)}^{dec} d_{(s',l-1)} \quad (3)$$

$$\alpha_{(s',l)}^{enc} = a(d_{(s',l)}, d_{(s',l-1)}, H) \quad (4)$$

2 段目の multi-head attention では、encoder の出力と前段の multi-head attention による隠れ状態から、次式により、文脈ベクトル $c_{(s,l)}$ を計算する。

$$c_{(s,l)} = \sum_{s'} \alpha_{(s',l)}^{dec2} e_{(s',l-1)} \quad (5)$$

$$\alpha_{(s',l)}^{enc} = a(d_{(s',l)}, e_{(s',l)}, H) \quad (6)$$

このモデルは、対数尤度 $P = \log P(O|I, \theta)$ を最大化するようにトレーニングされる。ここで、 O は平易文、 I は原文、 θ はモデルのパラメータを意味する。

エンコーダとデコーダの単語埋め込み層では、学習済みの単語ベクトル nwjc2vec[浅原 17](pre-emb.) を用いる。また、エンコーダの単語埋め込みとデコーダの入出力の単語埋め込み層を共有する (tied-emb.)。モデルのハイパーパラメータは、Zhao らの研究に倣い、エンコーダとデコーダを 4 層、それぞれのアテンション機構のヘッド数を 5、ドロップアウトを 0.3 に設定している。また、各単語埋め込み層の次元は、nwjc2vec に合わせ、200 次元としている。

4.2 評価方法

テキスト平易化では、一般的に流暢性、意味保持性、平易さの 3 つの観点からモデルの出力結果を評価する。ここで、流暢性とは、日本語として正しい文を出力できているかを測る評価尺度であり、意味保持性とは、原文 (モデルに対する入力文) とモデルが出力した文の意味が一致するかどうかを測る評価尺度である。また、平易さとは、出力文が入力文に比べ、どの程度簡単になっているかを示す評価尺度である。これらを自動で評価する方法として、BLEU と SARI[Xu 16] が用いられる。

BLEU とは、機械翻訳の評価で広く用いられる評価尺度であり、テキスト平易化においては、流暢性と意味保持性に関して正の相関があることが知られている [Vu 18]。

SARI は、最近提案された平易化の評価指標であり、平易化で行われる単語の追加、削除、保持の 3 つの操作における n-gram の precision 及び recall である。この尺度は、平易化すべき部分を適切に平易化した際に、高いスコアを与える仕組みとなっている。具体的には、モデルが参照文にのみ存在する単語を追加した場合や参照文に存在する単語をそのまま出力 (保持) した場合、参照文にない単語を削除した場合などに高いスコアを与える。一方、入力文をそのまま出力するようなモデルに対してはペナルティが与えられるように構成されている。実験的評価より、SARI が人間の平易さの判断とよく相関があることを示している [Xu 16, Vu 18]。

5. 結果と考察

原文から逐語訳への変換及び原文から意識への変換の結果を表 2 に示す。Origin は、原文をそのまま出力した際の結果である。NTS は、Nisioi らのニューラル機械翻訳によるテキス

ト平易化モデルである。ハイパーパラメータは、先行研究と同様に設定している。ただし、単語埋め込み層は、nwjc2vec の次元数に合わせて、200 次元に設定している。

表 2: 実験結果

	逐語訳		意識	
	BLEU	SARI	BLEU	SARI
Original	34.24	16.65	28.14	14.17
NTS	36.49	46.47	29.46	41.85
+ pre-emb.	34.60	45.17	28.87	41.31
+ tied-emb.	34.85	44.94	29.69	41.38
Transformer	33.88	45.94	18.61	35.56
+ pre-emb.	34.52	46.01	24.65	39.90
+ tied-emb.	41.89	48.43	29.41	41.63

実験結果より、NTS, Transformer どちらも、Original を上回る BLEU, SARI を記録しており、流暢性を保ちつつ、平易化できていることが分かる。NTS の結果に注目すると、逐語訳への変換、意識への変換どちらの場合においても、学習済みの単語ベクトルの使用 (NTS + pre-emb.) やエンコーダの単語埋め込みとデコーダの入出力の単語埋め込み層の共有 (NTS + tied-emb.) を行なっても、BLEU や SARI の向上にあまり効果がないことが分かる。一方で、Transformer の結果に着目すると、それらの工夫が大きく効果を示していることが分かる。特に、単語埋め込み層の共有 (Transformer + tied-emb.) の効果は大きく、逐語訳においては、そのままの Transformer に比べ、BLEU を 8.0 ポイント、SARI を 2.5 ポイント向上させている。意識においては、BLEU を 11 ポイント、SARI を 6.1 ポイントを改善させている。これらの結果は Transformer において、学習データを拡張させることでさらなる性能改善が可能であることを示している。

6. まとめ

我々は公的文書を対象に平易化を試みた。公的文書とは、市役所や病院、学校等の公共施設で配布される文書であり、生活する上で重要な情報を多く含んでいる。しかし、日本語初学者が学習する文に比べ、難解な語彙や公的文書に出現する固有の表現も含み、理解が困難であるため、平易化が必要な文書である。

本研究では、約 40 名の日本語教師が公的文書の日本語を逐語訳、意識、要約の 3 段階の「やさしい日本語」に書き換えたものをコーパスとし、機械翻訳的なアプローチにより、文単位の平易化を行った。また、学習済みの単語ベクトルの使用やエンコーダの単語埋め込みとデコーダの入出力の単語埋め込み層の共有を行い、モデル側で学習データの不足を補うことを試みた。

結果として、Transformer に学習済みの単語ベクトルやエンコーダの単語埋め込みとデコーダの入出力の単語埋め込み層の共有を利用することにより、小規模な学習データであっても適切に平易化できることを示した。学習済みベクトルや単語埋め込み層の共有による性能改善は、Transformer において、学習データの拡張によってさらなる改善が可能であることを示している。今後は、学習データの擬似的な拡張や転移学習といった方法を検討していきたい。

参考文献

- [Bahdanau 14] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio: Neural Machine Translation by Jointly Learning to Align and Translate, arXiv:1409.0473 (2014).
- [Gehring 17] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin: Convolutional Sequence to Sequence Learning, arXiv:1705.03122 (2017).
- [Luong 15] Minh-Thang Luong, Hieu Pham, Christopher D. Manning: Effective Approaches to Attention-based Neural Machine Translation, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.1412-1421 (2015).
- [Maruyama 17] Takumi Maruyama, Kazuhide Yamamoto: Sentence Simplification with Core Vocabulary, Proceedings of the International Conference on Asian Language Processing, pp.363-366 (2017).
- [Moku 12] Manami Moku, Kazuhide Yamamoto, Ai Makabi: Automatic Easy Japanese Translation for information accessibility of foreigners, Proceedings of the Workshop on Speech and Language Processing Tools in Education, pp.85-90 (2012).
- [Nisioi 17] Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, Liviu P. Dinu: Exploring Neural Text Simplification Models, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp.85-91 (2017).
- [Surya 18] Sai Surya, Abhijit Mishra, Anirban Laha, Parag Jain, Karthik Sankaranarayanan: Unsupervised Neural Text Simplification, arXiv:1810.07931 (2018).
- [Vaswani 17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), pp.5998-6008 (2017).
- [Vu 18] Tu Vu, Baotian Hu, Tsendsuren Munkhdalai, Hong Yu: Sentence Simplification with Memory-Augmented Neural Networks, Proceedings of NAACL-HLT 2018, pp.79-85 (2018).
- [Xu 16] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch: Optimizing Statistical Machine Translation for Text Simplification, Transactions of the Association for Computational Linguistics, Vol.4, pp.401-415 (2016).
- [Zhang 17] Xingxing Zhang, Mirella Lapata: Sentence simplification with deep reinforcement learning, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.584-594, (2017).
- [Zhao 18] Sanqiang Zhao, Rui Meng, Daqing He, Andi Saptono, Bambang Parmanto: Integrating Transformer and Paraphrase Rules for Sentence Simplification, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp.3164-3173 (2018).
- [浅原 17] 浅原 正幸, 岡 照晃:nwjc2vec:『国語研日本語ウェブコーパス』に基づく単語の分散表現データ, 言語処理学会第23回年次大会 発表論文集, (2017).
- [庵 08] 庵功雄: 「やさしい日本語」をめぐる. 多文化共生社会における日本語教育研究会, 第4回研究会, pp.1-12 (2008).
- [岩田 10] 岩田 一成: 言語サービスにおける英語志向: 「生活のための日本語:全国調査」結果と広島の実例から (特集:日本社会の変容と言語問題), 社会言語科学会, Vol.13, No.1, pp.81-94 (2010).
- [松田 09] 松田真希子: やさしい日本語への自動言い換えシステムの開発, 日本語教育学会大会 2009(平成 21) 年度春季大会予稿集, pp.91-93 (2010).
- [奎 11] 奎 真奈見, 山本 和英: 公的文書に対する「やさしい日本語」換言辞書作成のための調査, 言語処理学会 第17回年次大会 発表論文集, pp.376-379 (2011).
- [奎 13] 奎 真奈見, 山本 和英: 「やさしい日本語」変換システムの試作, 言語処理学会 第19回年次大会 発表論文集, pp.678-681 (2013).

擬似誤りコーパスを用いた天気予報原稿のニューラル誤り検出

Neural Error Detection for Weather Forecast Manuscript by Pseudo Error Corpus

白井稔久 *1

Naruhisa Shirai

萩行正嗣 *2

Masatsugu Hangyo

小町守 *1

Mamoru Komachi

*1 首都大学東京

Tokyo Metropolitan University

*2 株式会社ウェザーニューズ

WEATHERNEWS INC.

In this paper, we propose a neural method for detecting errors in Japanese weather forecast manuscripts. First, we analyze errors in weather forecasts to understand how native Japanese mistake. According to our analysis, we found native Japanese tend to cause errors in three types: particle error, conversion error and typo. In this paper, we focus on particle and conversion errors. However, any corpora written by native Japanese are not large enough for supervised learning. Therefore we use pseudo error corpus to augment training data. We generate pseudo particle errors by confusion matrix and we also generate pseudo conversion errors by back transliteration. As a result, we find pseudo corpus is effective for neural error detection for text written by native Japanese.

1. はじめに

天気予報原稿は一般的に人手で記述されているため、誤りを含んでいる場合がある。それらの誤りは公開する前に校正する必要がある。通常これらの誤りは人手での多重チェックなどで公開前に校正されているが、この校正には大きなコストがかかっている。

日本語母語話者が記述したテキストの自動誤り検出を行うには、テキストに誤りがアノテーションされているコーパスが少ないため、教師あり学習の手法をとることは困難である。また、日本語学習者が記述したテキストに比べて誤りが少なく、加えて誤り傾向が異なる可能性があり学習者のコーパスを学習にそのまま使用することは不適切であると考えられる。RNN 言語モデルを用いた誤りの自動検出 [9] では、誤警報率が非常に高く日本語母語話者の誤り検出が難しいことが報告されている。

そこで我々はまず誤り傾向を分析するためウェザーニューズ *1 の天気予報原稿コーパスを用いた。このコーパスは 2 年分の天気予報原稿があり、編集前の原稿と編集後の原稿がペアになっているため、擬似的に校正コーパスとみなすことができる。実際に含まれていた編集前後の文章対を図 1 に示す。このコーパスを用いた分析の結果、誤りの大部分は誤変換、助詞誤り、タイポの 3 つに分類できることが分かった。

そのため我々は教師あり学習が適用可能な誤変換と助詞誤りに着目しそれらの検出を行うことにした。しかし、それらを教師ありの手法を用いて検出するにはアノテーションされているデータ数が少ない。そこで我々は小規模データでの教師あり学習に有用であることが分かっている擬似コーパス [4] でコーパスを拡充するために、擬似誤りを生成した。

擬似誤りの生成は助詞の擬似誤りと誤変換の擬似誤りで分けて生成した。助詞の擬似誤りに関しては、助詞ごとに誤り傾向を分析し、その傾向に従って分布を作り擬似的な誤りを生成した。誤変換に関しては分析したコーパス全体で単語単位で誤変換をしている割合を基に、単語を別の同音の単語に変換して擬似誤りを生成した。

本研究の貢献は以下の 2 つである。

- 天気予報原稿の誤り検出のためのコーパスを分析した

連絡先: 白井 稔久, shirai-naruhisa@ed.tmu.ac.jp

*1 <https://weathernews.jp>

誤りが含まれる文章対

編集前

今日は雲優勢のスッキリしない空。
髪が乱れるほど風が強いのでご注意ください。
日差しが少なくてもムシ暑くなります。

編集後

今日は雲優勢のスッキリしない空。
髪が乱れるほど風が強いのでご注意ください。
日差しが少なくてもムシ暑くなります。

誤りが含まれない文章対

編集前

今日も雨が降り続きます。
激しく降る可能性があるので大きめの傘やレインコート・ブーツが良さそうです。
河川の増水・道路冠水・土砂災害にご注意下さい。

編集後

今夜～今日も雨が降り続きます。
激しく降る可能性があるので大きめの傘やレインコート・ブーツが良さそうです。
河川の増水・道路冠水・土砂災害にご注意下さい。

図 1: 編集前後の文章対の例。下線部が編集された部分である。

- 擬似誤り生成によるコーパスの拡充が日本語母語話者が記述したテキストのニューラル誤り検出に関して有用であることを示した

2. 関連研究

日本語母語話者が記述したテキストの誤り訂正の研究として、新納らは平仮名 n-gram を用いて誤りを検出し、訂正する手法を提案した [5]。また、南保らは文節内の特徴からルールを自動作成し、ルールベースで、日本語の助詞誤りを検出し、校正する手法を提案した [8]。これら 2 つの手法は我々と同じく、日本語母語話者が記述したテキストを対象にしている、特に南保らの研究とは助詞誤りに着目した点で我々の研究と共通する。一方我々の研究では教師あり学習を用いている。

また、日本語学習者が記述したテキストに対する自動訂正の研究も広く進められている。今村らは日本語学習者が助詞を間違えやすいことを指摘し、その助詞を、間違えやすい助詞の単語テーブルを用いることによって修正する手法を提案した [4]。また、今村らは小規模の誤りデータから擬似誤りを生成し、コーパスを拡充した。この研究は助詞に着目した点、擬似誤りを生成した点で我々と共通するが、我々の提案手法では助詞だけではなく、誤変換も対象としていて、誤りの検出のみで

訂正はしない。また、今村らは訂正に CRF を用いているが、本研究では Bi-LSTM を用いた RNN を使った。

水本らは、語学学習 SNS である Lang-8 から添削ログを抽出しコーパスを作成し、そのコーパスを用い、文字単位での修正と、文字-単語間での修正の二つの手法を提案した [6]。彼らが提案した手法は、統計的機械翻訳モデルを用いて誤りを修正するものである。我々の研究とコーパスを作成した点で共通するが、我々の提案手法では擬似誤りを生成してコーパスの拡充を図っている。また、本研究では RNN を用いている。

3. 擬似誤りコーパスの作成

我々は教師あり学習を行うには少ないコーパスを拡充するために、擬似誤りコーパスを作成した。擬似誤りコーパス内に含まれる擬似誤りは、助詞誤りと誤変換で異なる方法で作成した。擬似的な助詞誤りは実際に天気予報原稿コーパスで誤った助詞を基に誤りを作成した。擬似的な誤変換は天気予報原稿コーパスに含まれる誤変換の割合で元の単語を誤変換させて作成した。また、それらの誤変換は元の単語を平仮名にした後に再変換し、元の単語と異なるものに置換することで作成した。

3.1 天気予報原稿の誤り傾向分析

誤りの傾向を分析するために本研究ではウェザーニューズ社の天気予報原稿コーパスの分析を行った。このコーパスには 2014 年と 2015 年の 2 年分の天気予報原稿の、編集前の文章と編集後の文章対が入力されている。それらの文章対の総数は 100,931 対である。

これらの文章対における編集の内容は、文法誤りなどの校正だけでなく、原稿内容や表現を大きく書き換えるようなものも含まれる。それらを除外するために、編集前後の文章間の文字単位での編集距離が 1 以上 5 以下の文章対を抽出した。本研究で学習データとして用いた 2014 年のデータを対象に抽出された文章対は 2,575 対で、文対数は 7,765 文対だった。また、天気予報原稿内に 5 文字以下の文が含まれている可能性は非常に低いと考え、編集距離で抽出した文章対に編集前後で文数が異なるものは含まれていないと判断した。

その後、日本語形態素解析システム JUMAN7.0^{*2} を用いて文を形態素解析し、編集距離を用いた動的計画法で形態素単位でアライメントをとり、異なり形態素対およびその異なりが含まれる文を手で確認し、実際に誤りであると判断したものの編集前後の形態素対を記録し分析した。その結果誤りは大きく分けて誤変換、助詞誤り、タイポの 3 つに分類できることが分かった。

3.2 誤りタグ付きコーパスの作成

3.1 項と同様にして誤りであると判断した単語対にアノテーションを付与し、学習データとしてコーパスを作成した。誤りにアノテーション付与する際、以下のルールに基づきアノテーションを付与した。下記のルールに基づきアノテーションを付与する例を図 2 に示す。

置換 編集前の単語と編集後の単語を置換して文が成立する場合、その単語に誤りタグを付与

文の不成立 編集前の文が成立していない場合に原因と思われる単語に誤りタグを付与

余字 編集前の単語を削除すると文が成立する場合その単語に誤りタグを付与

置換

編集前: こまめに水分を **取って** 熱中症対策を万全にしてください。
編集後: こまめに水分を **摂って** 熱中症対策を万全にしてください。

文の不成立

編集前: 室内でも熱中症になることが **あり** 体調管理は万全に。
編集後: 室内でも熱中症になることがある体調管理は万全に。

余字

編集前: 今日の **今日** の朝は雨の可能性がありますが～
編集後: 今日の 朝 朝 朝は雨の可能性がありますが～

脱字

編集前: 今日は夏空 **広が** りますが急な雨もあります。
編集後: 今日は夏空 が ますが急な雨もあります。

図 2: ルールに基づいてタグを付与した例。下線部の単語に誤りのタグを付与する。編集後の文は編集前の文中における単語にアライメントを取っている単語の列であるため、実際の文ではない。また、編集前の文中における 1 単語に対して編集後の文中における複数の単語がアライメントを取っている場合、最も文頭に近い単語のみを出力している。

表 1: 学習・開発・評価データの詳細

	学習	開発	評価
誤変換	90	5	2
助詞誤り	76	5	6
タイポ	190	8	20
誤りの総数	356	18	28
総文数	7,765 文	3,842 文	2,971 文

脱字 編集前の文に明らかに単語が不足している場合、不足していると思われる位置の直後の単語に誤りタグを付与

上記に該当しない単語は誤っていないものとして誤りタグを付与しない

また、上記とほぼ同様の手順で誤りがアノテーションされているテストデータも作成した。相違点は、2014 年のデータではなく 2015 年のデータを対象にしたこと、編集距離が 1 以上 5 以下の文章対ではなく 0 以上 5 以下の文章対を用いたこと、そうして抽出された文章対の中から、季節や時期による文章の内容の偏りを防ぐために各月ごとに 200 対ずつランダムサンプリングしたことである。その結果 2,400 文章対からなる 6,813 文のアノテーション付きのデータを作成した。さらに作成したコーパスを各月から 100 対ずつ、合計 1,200 対ずつ開発データと評価データに分割した。学習データ、開発データ、評価データ誤りの種類毎の件数、総文数を表 1 に示す。

3.3 擬似誤りを含む文の生成

本研究では今回作成したコーパスが小規模であることから、小規模のコーパスでも誤りが検出できるよう擬似的に誤りを作成した。本研究では誤変換と助詞誤りを対象に擬似誤りを生成し、学習データの拡充を行う。

学習者が記述したテキストの擬似的な誤りの生成は一般的に実際に誤ったような誤りの生成割合で行われる [4]。しかし、予備実験によって本コーパスにおいてはこの手法はあまり有用でないことが分かった。そこで、本研究では 3.2 項で作成したコーパスの誤った単語対を基に、一様分布を用いて抽出前の全ての編集後の文章を誤らせることで擬似的に誤りを生成した。

*2 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

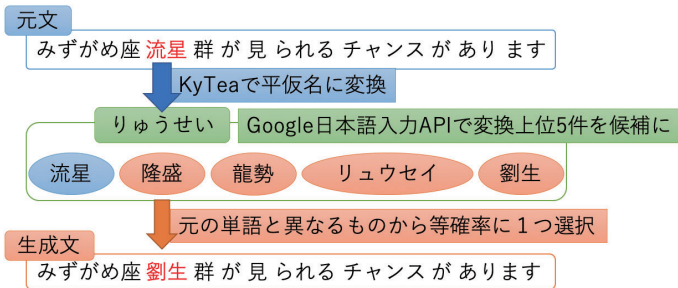


図 3: 擬似誤変換生成の例

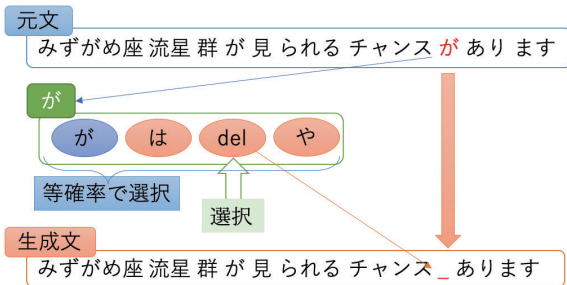


図 4: 擬似助詞誤り生成の例

我々は誤変換はどのような単語に対しても起こりうると思った。そこで、作成したコーパスの全単語を定めた割合に基づき再変換し、擬似的な誤変換を生成した。誤変換の作成は1度単語を京都テキスト解析ツールキット KyTea [2] を用いてかなに直し、その後 Google 日本語入力 API *3 を用いて、ランキング上位5件から無作為に元の単語と異なる単語を選び、擬似的な誤変換を作成した。擬似誤変換の生成の例を図3に示す。

助詞の誤りに関しては、2つの方法で擬似誤りを生成した。1つは3.2項で作成した学習コーパス中で出現した編集後の各助詞 w について編集前では誤っている各単語 $w_i = \{w_1, w_2, \dots, w_L\}$ と元の単語に対して一様に確率を割り当てる一様分布に従い擬似的な助詞の誤りを生成した。また、元の単語を選んだ場合誤りタグは付与しない。助詞の擬似誤り生成の例を図4に示す。もう1つは擬似誤変換の生成と同様に一定の割合で助詞誤りを発生させる方法である。全ての助詞に対して一定の割合に基づきその助詞を誤らせるか決め、誤らせる場合学習コーパス内で出現した誤り方から無作為に1つ選択肢誤らせる。また、学習コーパス内で誤らなかった助詞を誤らせることが選択された場合、その助詞を削除することで擬似助詞誤りを生成した。

4. 助詞誤りと誤変換の検出実験

4.1 誤り検出器

本研究では Bi-LSTM [1] を用いた誤り検出器で実験を行う。

入力文 $S = (w_1, w_2, \dots, w_n)$ の各単語 w_t は単語ベクトル $e_t \in \mathbb{R}^{de \times 1}$ に変換される。 n は文長であり、 de は単語ベクトルの次元である。単語ベクトルから LSTM により順方向の隠れ層 $\vec{h}_t \in \mathbb{R}^{dh \times 1}$ と逆方向の隠れ層 $\overleftarrow{h}_t \in \mathbb{R}^{dh \times 1}$ を作成する。 dh は隠れ層の次元とする。 $\vec{h}_t$ と $\overleftarrow{h}_t$ を連結することで最終的な隠れ層 $h_t^{(lstm)} \in \mathbb{R}^{2dh \times 1}$ を獲得する。隠れ層 $h_t^{(lstm)}$ を以下のように線形変換しソフトマックス関数を使い正誤タグの確率分布 $p_t \in \mathbb{R}^{t \times 1}$ を獲得する。 t はタグのサイズであり、正誤の

表 2: 実験に用いた各学習データの文数。「orig」は3.2項で作成したコーパス、「pp」は助詞の擬似誤りを生成したコーパス、「conv」は誤変換を擬似生成したコーパスを示す。

学習データ	文数
orig	7,765 文
orig+pp	115,744 文
orig+conv	115,744 文
orig+pp+conv	225,305 文

どちらかのタグを予測するためサイズは2である。

$$p_t = \text{softmax}(W_h h_t^{(lstm)} + b_h) \quad (1)$$

$W_h \in \mathbb{R}^{v \times dh}$ は重み行列であり、 $b_h \in \mathbb{R}^{v \times 1}$ はバイアスである。 v は語彙サイズの次元数である。

誤差関数である $loss$ は交差エントロピーによって以下の式を用いて計算される。

$$loss = - \sum y_t \log p_t \quad (2)$$

y_t は正解のタグであり学習データ内で正誤のどちらかが付与されている。

4.2 実験設定

入力にはあらかじめ学習済みの朝日新聞単語ベクトル [7] を用いてベクトル化した。このときの埋め込み層は300次元である。このモデルは PyTorch 1.0 で実装し、出力層の値を基に誤っている確率を出力する。この確率が0.5を超えているものを誤りとして検出する。隠れ層は開発データを用いて1層で1024次元に定めた。また、出力層は200次元、初期化は-0.1から0.1の間でランダムに初期化した。バッチサイズは64、最適化には ADADELTA [3] を用いて学習した。

学習には3.2項で作成したコーパスとそれに擬似誤りを生成したコーパスを用いる。また、今回の誤り検出では擬似誤りを加えたことによる有用性を確かめるために助詞誤りと誤変換のみを誤りとして検出する。よって学習データのタイプの誤りは誤りでないものとして学習した。開発データを用いて各エポックで再現率が最大のときに適合率が最も高いエポックのものを評価データの実験に用いた。また各データの文量を表2に示す。

評価には適合率と再現率を用いた。適合率はシステムが誤りだと判断した単語の内、実際に誤りである割合である。再現率はコーパス内の検出の対象である全ての誤りである単語の内、システムが検出した誤りの割合である。

4.3 実験結果

実験結果を表3に示す。表を見ると擬似誤りコーパスを学習データとして追加すると適合率、再現率ともに上昇していることがわかる。ただ、再現率は全体的に低く、誤りをほとんど検出できていないことがわかる。また、3.2項で作成したコーパスだけでは学習のための文数が非常に少ないため誤りを検出することができていないことが分かる。

また、コーパス内で擬似誤りを生成する割合を変えて実験した結果を表4に示す。この実験のデータ量は orig+conv, orig+pp と同一である。表を見ると擬似誤りを生成する割合を増やすと性能が向上していることが分かる。特に助詞誤りは生のが著しく向上している。また、予備実験で行った助詞の擬似誤りを実際に誤った分布に従って生成する実験で擬似誤りを生成する割合を増やしたが、特に性能の向上が見られなかった。

*3 <https://www.google.com/inputtools/try/>

表 3: 誤り検出実験の結果. pp*は元の誤り分布に従い助詞誤りを生成したコーパスで学習したものである. @以下の数字は擬似誤りの生成割合である.

学習データ	分割	適合率	再現率
orig	開発	0.00	0.00
	評価	0.00	0.00
orig+pp*	開発	0.00	00.0
	評価	0.00	00.0
orig+pp	開発	0.21	20.0
	評価	0.26	25.0
orig+conv	開発	1.75	20.0
	評価	0.00	0.00
orig+pp*+conv	開発	0.16	20.0
	評価	0.00	0.00
orig+pp+conv	開発	1.38	50.0
	評価	0.29	37.5
orig+pp@50.0%+conv@50%	開発	1.03	70.0
	評価	0.71	50.0

5. 分析

助詞誤りに関しては生成割合を一様分布に変えた結果、適合率は非常に低い再現実率は上がり、一定の生成割合で助詞誤りを生成した結果、適合率と再現率両方の上昇が見られた。このことから助詞の誤り検出に関しては母語話者の誤る割合よりも大きい割合で擬似誤りを生成した方が、性能が向上することが分かった。又、一定の割合での助詞誤りを生成したコーパスで学習したモデルは、“は”や“が”の助詞誤りは検出できる傾向にあった。これは“は”と“が”は元のデータ内でも誤り件数と誤り方の種類も多いため、ほとんどの誤り方を再現できたからではないかと考える。

誤変換に関しては生成割合を上昇させた結果、評価データ内の誤変換は2件とも検出できているが開発データ内の誤変換は2件検出できていない。これは両方とも“うだるような”が“うだる様な”に変換されてしまっているものだった。この変換を学習データ内には存在したが単語分割が開発データでは“よう_な”と分割されていたのに対し、学習データでは“ような”と分割されていたため、実質学習データ内に存在しない誤りになってしまったためだと考える。

助詞誤りの誤検出に関しては“は”、“の”、“を”の誤検出が多く見られた。この3つの助詞は全ての擬似助詞誤り生成法で誤りとして生成する割合が他の助詞よりも高いため、誤りとして誤検出してしまう傾向にあるのではないかと考える。

誤変換の誤検出に関しては“熱”という単語をよく誤検出してしまう傾向にあった。これは学習データ内の“熱”という単語ほとんどに誤りのアノテーションが付与されており、出現してしまうこと自体が誤りだと認識したためだと考える。

6. おわりに

本研究では天気予報原稿の誤り検出において教師あり学習を行うためのコーパスの作成と、そのコーパスを拡充するために実際に日本語母語話者が誤った確率を基に擬似誤りの生成を行なった。擬似誤りを生成する提案手法は結果として日本語母語話者が記述した日本語テキストのニューラル誤り検出に有用であることが分かった。加えて実際の誤る割合、および分布に従って擬似誤りを生成するよりも擬似誤りの生成割合を

表 4: 擬似誤りの生成割合を変えた実験結果

生成割合	分割	orig+conv		orig+pp	
		適合率	再現率	適合率	再現率
0.01%	開発	0.00	0.00	0.00	0.00
	評価	0.00	0.00	0.00	0.00
0.1%	開発	3.30	30.0	0.00	0.00
	評価	0.00	0.00	0.00	0.00
1.0%	開発	0.79	20.0	13.3	20.0
	評価	0.00	0.00	16.6	12.5
10.0%	開発	0.90	30.0	8.33	10.0
	評価	0.31	12.5	7.14	12.5
20.0%	開発	0.71	30.0	3.03	30.0
	評価	0.52	25.0	3.09	37.5
30.0%	開発	0.69	30.0	0.77	40.0
	評価	0.49	25.0	1.13	62.5
40.0%	開発	0.74	30.0	1.26	50.0
	評価	0.51	25.0	1.47	62.5
50.0%	開発	1.50	30.0	1.24	50.0
	評価	1.03	25.0	1.52	62.5

上げる方がモデルの性能が向上することが分かった。しかし、Bi-LSTMを用いたモデルでの検出は未知の誤りの検出が非常に困難であることが分かった。また、本研究の提案手法では適合率が低いいため、適合率を上げる手法を検討する必要がある。

参考文献

- [1] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*, Vol. 18, pp. 602–10, 07 2005.
- [2] Graham Neubig and Shinsuke Mori. Word-based partial annotation for efficient corpus construction. *LREC*, pp. 2723–2727, 2010.
- [3] Matthew D. Zeiler. ADADELTA: an adaptive learning rate method. *CoRR*, Vol. abs/1212.5701, , 2012.
- [4] 今村賢治, 齋藤邦子, 貞光九月, 西川仁. 小規模誤りデータからの日本語学習者作文の助詞誤り訂正. 言語処理学会論文誌, Vol. 19, No. 5, pp. 381–400, 2012.
- [5] 新納浩幸. 平仮名 n-gram による平仮名文字列の誤り検出とその修正. 情報処理学会論文誌, Vol. 40, No. 6, pp. 2690–2698, 1999.
- [6] 水本智也, 小町守, 永田昌明, 松本裕治. 日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得. 人工知能学会論文誌, Vol. 28, No. 5, pp. 420–432, 2013.
- [7] 田口雄哉, 田森秀明, 人見雄太, 西島羽二郎, 菊田洸. 同義語を考慮した日本語単語分散表現の学習. 情報処理学会第 233 回自然言語処理研究会, Vol. 2017-NL-233, pp. 1–5, 2017.
- [8] 南保亮太, 乙武北斗, 荒木健治. 文節内の特徴を用いた日本語助詞誤りの自動検出・校正. 情報処理学会第 181 回自然言語処理研究会, Vol. 2007, No. 94, pp. 107–112, 2007.
- [9] 白井稔久. RNNLM を用いた日本語テキストの誤字・脱字検出および再変換を用いた誤変換検出. 首都大学東京卒業論文, 2018.

ニューラル系列ラベリングを用いた音声認識誤り訂正

Neural Sequence-Labeling Models for ASR Error Correction

池田 大志

Taishi Ikeda

藤本 拓

Hiroshi Fujimoto

吉村 健

Takeshi Yoshimura

磯田 佳徳

Yoshinori Isoda

株式会社 NTT ドコモ

NTT DOCOMO, INC.

本稿では、音声認識処理の後処理として、音声認識誤りを訂正する手法を提案する。提案手法は、まず Bidirectional LSTM を用いて音声認識結果から音声認識誤り箇所を検出する。次に、誤り検出の結果と辞書を用いて単語ラティスを構築し、識別モデルにより最適な単語列を選択することで音声認識誤りを訂正する。評価実験の結果、提案手法により音声認識結果の文字誤り率と単語誤り率が改善することを示す。

1. はじめに

近年、音声認識システムと自然言語処理の技術を応用した音声エージェントや音声翻訳などのアプリケーションの普及を背景として、ディープニューラルネットワークによる音声認識手法が盛んに研究されている。一例として、電話会話を録音した Switchboard 音声コーパスを用いた実験では、人手による書き起こしに匹敵する認識結果が報告されている [1]。しかし、実世界で音声認識システムを利用する場合、集音マイクの性能や周囲の雑音など、使用環境による認識精度の低下が依然課題となっている。

このような音声認識誤りは、音声認識システムを利用したアプリケーションの自然言語理解（対話システムや機械翻訳）に悪影響を及ぼすことが報告されている。例えば、Sano らの研究 [2] では、音声エージェントにおいて、ユーザーとシステムによる対話のログデータを分析した結果、音声による入力には 31.7% の音声認識誤りが含まれており、自然言語理解における解析誤りの 57.2% は音声認識誤りが原因であると報告されている。また、Yoshikawa らの研究 [3] では、音声認識誤りが依存構造解析に悪影響を及ぼすことを報告されており、音声認識誤りは自然言語理解に必要な基礎解析の精度低下を招くため、その後段処理である応用アプリケーションの精度低下の原因となる。

一般的な音声認識システムは、音響モデル・単語辞書・言語モデルを組み合わせ構成されている。ある特定の誤りに対しては、各構成要素のいずれかの部分に学習データを追加することで、音声認識誤りの対策ができる。しかし、音声認識システムは各構成要素が相互作用しており、音響モデルや言語モデルにチューニングを施すことは、再学習の時間や精度検証など運用上のコストが掛かる。また、一般に API など提供されている音声認識システムは、ブラックボックスであることが多く、音声認識誤りを発見し、対策を施そうとしても、音声認識システム自体に手を加えることができない問題もある [4]。

そこで本研究では、音声認識処理の後処理として、音声認識誤りを訂正する手法を提案する。音声認識システムの 1-best の音声認識結果を入力として、音声認識誤りとなる単語を発話内容の正しい表記に変換するテキスト正規化の問題として、音声認識誤り訂正を考える。提案手法では、音声認識誤り訂正を二段階のパイプライン処理により行う。具体的には、まず

Bidirectional LSTM (BiLSTM) を用いて音声認識結果から音声認識誤り箇所を検出する。次に、辞書を用いて訂正候補を含む単語ラティスを構築し、構造化パーセプトロンを用いて、最適な単語列を選択することで音声認識結果を訂正する。

本研究では、提案手法の有効性を音声認識アプリケーションのログデータを用いた実験により検証した。その結果、提案手法を用いることで音声認識結果の文字誤り率と単語誤り率が、それぞれ 0.28 ポイントと 0.52 ポイント改善することを確認した。本稿では、本手法では解決することができなかった事例について考察し、今後の課題とその対策について述べる。

2. 関連研究

本研究では、テキスト正規化の問題として、系列ラベリングによる音声認識誤り検出と、構造化パーセプトロンによる識別モデルを用いた二段階のパイプライン処理により、音声認識誤り訂正を行う。そこで本章では、音声認識結果から音声認識誤りを検出する研究と日本語を解析対象としたテキスト正規化の研究について、それぞれ詳細を記述する。

2.1 音声認識誤り検出

Byambakhishig らの研究 [5] では、条件付き確率場を用いてコンフュージョンネットワーク上から音声認識誤りを検出し、リランキングにより音声認識誤り単語を訂正することで、音声認識システムの改善を行っている。それに対し、本研究では、ブラックボックスの音声認識システムを利用すると仮定し、コンフュージョンネットワークを利用せず、音声認識システムの 1-best の認識結果を入力とする音声認識誤り訂正タスクに取り組む。

Ghannay らの研究 [6] では、系列ラベリング手法を用いて、音声認識結果の各単語に対して、「正解」または「誤り」のラベルを付与することで、音声認識誤りを検出する手法を提案している。それに対し、本研究では、音声認識誤りを訂正することを目的とするため、音声認識結果の各単語に対して、「操作なし」「置換」「削除」の三種類のラベルに拡張したラベル付与を行う。これらのラベルと辞書を用いて訂正候補を含む単語ラティスの構築を行い、音声認識誤りの訂正を行う。

2.2 テキスト正規化

近年、Web テキストを対象としたテキスト正規化手法が盛んに研究されている [7, 8, 9]。テキスト正規化の研究では、入力文中の表記揺れや口語表現を辞書に存在する正規表記に変換

することを目的としている。本研究では、音声認識誤り訂正タスクを、音声認識結果の認識誤りを発話内容の正しい表記に変換するテキスト正規化の問題として考える。

テキスト正規化の手法としては、Encoder-Decoder モデルを用いて、音声認識誤りを正しい表記に翻訳する問題として取り込むことができる。しかし、音声認識結果と音声認識誤りを訂正したペアデータを大量に収集することは難しく、小規模なデータを用いて Encoder-Decoder モデルを学習する場合、データ拡張などの手法が必要となる [9, 10, 11]。そこで本研究では、Kaji ら [7] や Saito らの研究 [8] を参考として、辞書と識別モデルを用いた音声認識誤りの訂正を行う。

3. 提案手法

図 1 に提案手法の概要を示す。図 1 では、音声認識結果が「セキュリティ祖父とにログイン」、書き起こしが「セキュリティソフトにログイン」の例を示している。この例では、音声認識結果の「祖父と」の部分が音声認識誤りである。図 1 下部は、BiLSTM を用いて、音声認識結果から音声認識誤りを検出する過程を示している。図 1 上部は、辞書を用いて単語ラティスを構築し、識別モデルにより最適な単語列を選択する過程を示している。以下、それぞれ詳細を記述する。

3.1 系列ラベリングを用いた音声認識誤り検出

本研究では、音声認識結果から音声認識誤りを検出するため、形態素解析後の単語列に対して系列ラベリングを行う。また、音声認識誤りを検出するだけでなく、訂正することを目的とするため、音声認識結果の単語列に対して「操作なし」「置換」「削除」の三種類のラベル付与を行う。各単語に対するラベルは、動的計画法を用いて、音声認識結果と書き起こしのアライメントを求め、ラベル付与を行う。

図 2 にラベル付与の例を示す。具体的には、単語が一致する場合はスコア 0 を与え、削除と置換の操作が必要な場合にはスコア 1 を各単語に与える。各単語のスコアを足し合わせアライメントのスコアを計算し、ビタビアルゴリズムによりスコアが最小となるアライメントを求める。得られたアライメント結果を元に、対応する単語が一致する場合は「操作なし」ラベルを付与し、不一致となる場合は「削除」または「置換」ラベルを付与する。図 2 の例では、書き起こしの「セキュリティ」と音声認識結果の「セキュリティ」が一致するため「操作なし」ラベルを付与する。また、「ソフト」と「祖父」が不一致となるため「置換」ラベルを付与し、「と」に対してはアライメント対象が存在しないため「削除」ラベルを付与する。

上記の操作により、BiLSTM の入出力となる単語列 $W = w_1, w_2, \dots, w_n$ 、品詞列 $P = p_1, p_2, \dots, p_n$ 、ラベル列 $Y = y_1, y_2, \dots, y_n$ を得る。ただし、 n は単語数を示す。また、単語列 W と品詞列 P は、埋め込みベクトルに変換され、BiLSTM の入力となる。時刻 i の入力となる $\mathbf{x}_i \in \mathbb{R}^{d^{word}+d^{pos}}$ は、 d^{word} 次元の単語ベクトル $\mathbf{w}_i \in \mathbb{R}^{d^{word}}$ と d^{pos} 次元の品詞ベクトル $\mathbf{p}_i \in \mathbb{R}^{d^{pos}}$ を結合したものである。単語ベクトルは、Skip-gram などにより事前学習した単語埋め込みを初期値とし、品詞ベクトルは、一様分布によりランダムに初期化し、それぞれ学習中に更新する。BiLSTM は、各時刻の入力として $\mathbf{x}_i$ を受け取り、 d^{hidden} 次元のベクトル $\mathbf{h}_i \in \mathbb{R}^{d^{hidden}}$ を計算する。

$$\mathbf{h} = \text{BiLSTM}(\mathbf{x})$$

ただし、 $\mathbf{x} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ は埋め込みベクトル系列、 $\mathbf{h} = \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$ は BiLSTM の出力ベクトル系列である。各ラ

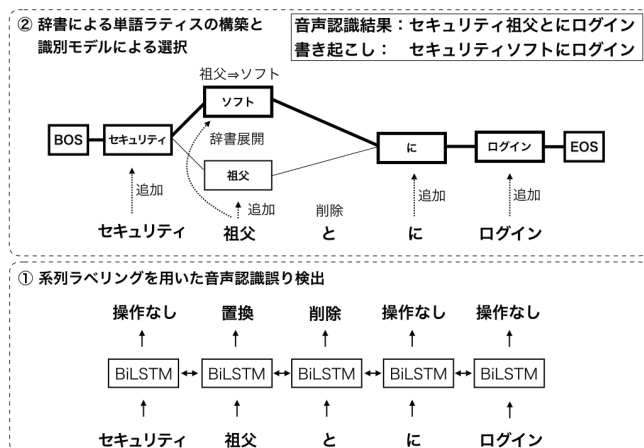


図 1: 提案手法の概要 (単語ラティスの太線は正解系列を表す)

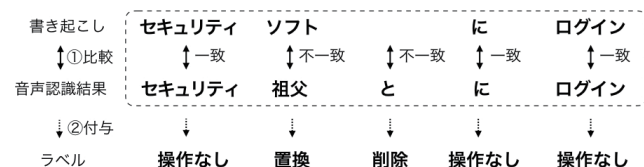


図 2: 動的計画法によるラベル付与

ベルの予測は、各時刻の $\mathbf{h}_i$ を入力として受け取り、アフィン変換とソフトマックス関数により各ラベルに対する確率値のベクトル $\mathbf{z}_i \in \mathbb{R}^{|T|}$ を得る。

$$\mathbf{z}_i = \text{Softmax}(\mathbf{W}_{out}\mathbf{h}_i + \mathbf{b}_{out})$$

ただし、 $|T|$ はラベル数、 $\mathbf{W}_{out} \in \mathbb{R}^{|T| \times d^{hidden}}$ は重み行列、 $\mathbf{b}_{out} \in \mathbb{R}^{|T|}$ はバイアスベクトルである。また、目的関数としてクロスエントロピー、パラメータの最適化には Adam ($\beta_1 = 0.9, \beta_2 = 0.999$) を用いる。

3.2 辞書による単語ラティスの構築と識別モデルによる選択

本研究では、辞書を用いて単語ラティスを構築し、構造化パーセプトロンにより最適な単語列を選択することで音声認識結果を訂正する。

まず、単語ラティスの構築方法について説明する。入力単語列 W と BiLSTM による予測ラベル列 $\hat{Y}$ が与えられたとき、BiLSTM により「操作なし」ラベルが付与された単語は、単語ラティスに追加し、「置換」ラベルが付与された単語は辞書により展開した訂正候補とともに単語ラティスに追加する。また、「削除」ラベルが付与された単語は単語ラティスには追加しない。図 1 上部の例では、「操作なし」ラベルが付与された「セキュリティ」は、単語ラティスに追加し、「置換」ラベルが付与された「祖父」は、辞書を参照し、訂正候補が存在するのであれば訂正候補の「ソフト」と「祖父」を単語ラティスに追加する。「削除」ラベルが付与された「と」は単語ラティスには追加しない。本研究では、人手による辞書構築を行い、辞書には音声認識誤りとその訂正後の表記のペアデータが 129 件含まれる。

次に、構造化パーセプトロンを用いて最適な入力単語列を

表 1: 本研究での実験で用いたコーパス概要

	学習	評価
文数	9,724	4,490
誤り文数	4,213	2,068
単語数 (音声認識結果)	113,993	55,265
単語数 (書き起こし)	112,699	54,896

選択する方法について説明する。本研究では、単語列 W と BiLSTM による予測ラベル列 $\hat{Y}$ に対し、音声認識誤りが訂正された単位列 $\hat{W}$ を求める問題を考える。この問題は次のように定式化できる。

$$\hat{W} = \arg \max_{\ell \in L(W, \hat{Y})} w \cdot f(\ell)$$

ここで、 $\ell \in L(W, \hat{Y})$ は、入力単語列 W と BiLSTM による予測ラベル列 $\hat{Y}$ により構築された単語ラティス (各ノードは単語) である。 $w \cdot f(\ell)$ は、重みベクトル w と素性ベクトル $f(\ell)$ の内積を表す。最適系列は $w \cdot f(\ell)$ の値にしたがって選択される。

素性は、学習データの書き起こし bi-gram 素性を用いる。ここでは、書き起こしの単語列を選択するよう学習を行い、構造化パーセプトロンにおける重みベクトル w の推定を行う。正解データの値 $w \cdot f(\ell)$ が、正解以外のどの値よりも大きくなるように $\hat{W}$ を求める。

重みベクトル w の推定は、平均化パーセプトロン学習に基づいて行う。平均化パーセプトロンでは、正解である単語系列 W が付与された N 個の文が与えられたとき、現在のパラメータ w^i に基づいて一文ずつ最適解 $\hat{W}$ を求め、もしこの系列が正解と異なる場合は重みパラメータ w^{i+1} を下記の式により更新する。

$$w^{i+1} = w^i + f(\hat{\ell}) - f(\ell)$$

ただし、 $f(\hat{\ell})$ は正解系列から得られる素性であり、 $f(\ell)$ はパラメータ w^i より求めた系列から得られる素性である。もし現在のパラメータに基づいて出力された最適解が正解と一致する場合にはパラメータの更新を行わない。最後に、文数と繰り返し回数の積で平均化した重みパラメータを計算する。以上の二段階のパイプライン処理により、音声認識誤り訂正を行う。

4. 実験

提案手法の有効性を検証するため、音声認識アプリケーションのログデータを用いて、音声認識誤り訂正の評価実験を行う。

4.1 実験設定

本研究では、重み付き有限状態トランスデューサを用いた音声認識システムを利用し、携帯電話に関する問い合わせに回答する音声認識アプリケーションからログデータを収集する。ここでは、一般に API などで提供されている音声認識システムを利用することを想定とする。そこで、ユーザーから収集した音声発話を用いて、1-best の音声認識結果を取得した。また、人手による音声発話の確認を行い、正解データとなる書き起こしの作成を行った。表 1 に学習と評価に用いたデータ数を示す。

評価データに対して、提案手法の誤り訂正処理を適応し、文字誤り率 (Character Error Rate: CER) と単語誤り率 (Word

表 2: 音声認識誤り検出結果

ラベル	適合率	再現率	F1 値	件数
操作なし	0.95	0.98	0.96	50,220
削除	0.51	0.19	0.28	1,691
置換	0.61	0.43	0.50	3,354

表 3: 音声認識誤り訂正結果

訂正方法	CER	WER	相対向上数
操作なし	10.10	11.59	0 (0-0)
削除	10.01	11.22	+10 (74-64)
置換	9.954	11.43	+34 (125-91)
削除+置換	9.819	11.07	+64 (188-124)
削除 (正解ラベル)	8.068	8.517	+298 (298-0)
置換 (正解ラベル)	9.302	10.91	+181 (181-0)
削除+置換 (正解ラベル)	7.248	7.820	+511 (511-0)

Error Rate: WER) がどの程度改善できるかを評価する。CER と WER は、音声認識結果と書き起こしの編集距離を計算し、文字単位の誤り数と単語単位の誤り数を元に求める。以下、CER と WER の定義を示す。

$$\text{CER} = \frac{\text{置換誤り文字数} + \text{削除誤り文字数} + \text{挿入誤り文字数}}{\text{全文字数}} \times 100$$

$$\text{WER} = \frac{\text{置換誤り単語数} + \text{削除誤り単語数} + \text{挿入誤り単語数}}{\text{全単語数}} \times 100$$

また、訂正によって音声認識結果が書き起こしと一致した件数と書き起こしと一致せず差分が発生した件数の差である相対向上数を表 3 に示す。

音声認識結果と書き起こしの形態素解析には、MeCab^{*1} とその辞書である Unidic^{*2} を利用した。BiLSTM のハイパーパラメータは、学習データの 5% を開発データとして利用し、 $d^{\text{word}} = 100$, $d^{\text{pos}} = 32$, $d^{\text{hidden}} = 100$ と設定した。開発データにおいて、WER が最小となるモデルを保存し、評価データで性能の評価を行う。また、BiLSTM のラベル予測には、三種類のラベルを利用するため、 $|T| = 3$ となる。

4.2 結果

音声認識結果から音声認識誤りをどの程度検出することが可能か検証するため、評価データに対して、音声認識誤り検出の精度評価を行った。表 2 には音声認識誤り検出の精度評価を示す。ここでは、音声認識誤りである単語の検出率を確認するため、「削除」ラベルと「置換」ラベルの再現率に注目する。「削除」ラベルの再現率が 0.19、「置換」ラベルの再現率が 0.43 と各ラベルとも低く、多くの音声認識誤りの検出に失敗していることがわかる。この結果より、音声認識誤り検出の再現率をさらに上げる方法の検討が必要になると考えられる。

次に、音声認識誤りの検出結果から音声認識誤りをどの程度訂正することが可能か検証するため、評価データに対して、音声認識誤り訂正の精度評価を行った。表 3 には音声認識誤り訂正の精度評価を示す。音声認識誤り訂正の効果を測るため、以下の方法を比較することで精度評価を行った。

操作なし 音声認識結果に訂正を一切行わない場合。

*1 <http://taku910.github.io/mecab/>

*2 <https://unidic.ninjal.ac.jp/>

削除 「削除」ラベルを用いて、音声認識誤りを削除した場合.

置換 「置換」ラベルを用いて、識別モデルによる訂正を行った場合.

削除+置換 削除と置換を同時に行った場合.

ここでは、操作なしと比較し、どの程度 CER と WER が改善したか確認することで、提案手法の有効性を検証する. 提案手法である削除+置換を評価データに適応した場合、操作なしの CER と WER と比較し、CER が 0.28 ポイントと WER が 0.52 ポイント改善したことから、提案手法により音声認識誤りの訂正が行われていることが確認できる. また、提案手法によって正しく訂正された例を示す. 一番目と二番目の例では、提案手法によって助詞の重複や助詞の間違いを正しく訂正している事例を確認できた.

正解 認識結果 訂正結果	暗証 番号 の 暗証 番号 を の 暗証 番号 の
正解 認識結果 訂正結果	サービス の 廃止 サービス が 廃止 サービス の 廃止

しかし、提案手法により悪化した事例も存在する. 例えば、一番目の例のように、不必要に単語を削除してしまう事例が多く存在した. また、提案手法では、二番目の例のような挿入操作に対応できないため、今後の課題として、挿入操作への対応を検討する必要がある.

正解 認識結果 訂正結果	五 月 十 五 日 五 月 十 五 日 月 五 日
正解 認識結果 訂正結果	アプリ が ダウンロード で き ま せ ん アプリ が た の で き ま せ ん アプリ が た の で き ま せ ん

また、表 3 では、BiLSTM の各単語の予測ラベルを全て本来の正解ラベルとした場合の CER と WER を示している. 削除 (正解ラベル) の場合は、操作なしと比較し、CER では 2.03 ポイント、WER では 3.07 ポイントと大幅に誤り率が減少している. しかし、置換 (正解ラベル) の場合は、操作なしと比較し CER では 0.79 と WER では 0.67 ポイントと誤り率が少ないことがわかる. この結果から、辞書の網羅性が低かったと考え、読み情報を利用した辞書構築など辞書構築方法の再検討が必要であると考えられる.

5. おわりに

本研究では、BiLSTM を用いて音声認識結果から音声認識誤り箇所を検出し、誤り検出の結果と辞書を用いて単語ラティスを構築後、識別モデルにより最適な単語列を選択することで音声認識誤りを訂正する手法を提案した. 音声認識アプリケーションのログデータを用いた実験では、提案手法を用いることで音声認識結果の文字誤り率と単語誤り率が改善することを確認した.

今後の課題として、音声認識誤り検出に対する再現率の向上と辞書の構築方法の検討が考えられる.

参考文献

- [1] Wayne Xiong, Lingfeng Wu, Fil Allea, Jasha Droppo, Xuedong Huang, and Andreas Stolcke. The microsoft 2017 conversational speech recognition system. In *Proc. ICASSP*. IEEE, 2018.
- [2] Shumpei Sano, Nobuhiro Kaji, and Manabu Sassano. Predicting causes of reformulation in intelligent assistants. In *Proc. SIGdial*, 2017.
- [3] Masashi Yoshikawa, Hiroyuki Shindo, and Yuji Matsumoto. Joint transition-based dependency parsing and disfluency detection for automatic speech recognition texts. In *Proc. EMNLP*, 2016.
- [4] Dong Yu, Jinyu Li, and Li Deng. Calibration of confidence measures in speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011.
- [5] E Byambakhishig, Katsuyuki Tanaka, Ryo Aihara, Toru Nakashika, Tetsuya Takiguchi, and Yasuo Ariki. Error correction of automatic speech recognition based on normalized web distance. In *Proc. ACL*, 2014.
- [6] Sahar Ghannay, Yannick Estève, and Nathalie Camelin. Task specific sentence embeddings for asr error detection. In *Proc. Interspeech*. ISCA, 2018.
- [7] Nobuhiro Kaji and Masaru Kitsuregawa. Accurate word segmentation and pos tagging for japanese microblogs: Corpus annotation and joint modeling with lexical normalization. In *Proc. EMNLP*, 2014.
- [8] Itsumi Saito, Kugatsu Sadamitsu, Hisako Asano, and Yoshihiro Matsuo. Morphological analysis for japanese noisy text based on character-level and word-level normalization. In *Proc. COLING*, 2014.
- [9] Taishi Ikeda, Hiroyuki Shindo, and Yuji Matsumoto. Japanese text normalization with encoder-decoder model. In *Proc. WNUT*, 2016.
- [10] Itsumi Saito, Jun Suzuki, Kyosuke Nishida, Kugatsu Sadamitsu, Satoshi Kobashikawa, Ryo Masumura, Yuji Matsumoto, and Junji Tomita. Improving neural text normalization with data augmentation at character-and morphological levels. In *Proc. IJCNLP*, 2017.
- [11] Jun Harashima and Yoshiaki Yamada. Two-step validation in character-based ingredient normalization. In *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*. ACM, 2018.

認知症の状態像理解を深めるための ConceptNet を用いた認知症見立て表現モデルの構築

Construction of a Diagnosis Representation Model of Person with Dementia Based on ConceptNet for Deeper Understanding of Physical and Mental Conditions

神谷直輝^{*1} 吉沢拓実^{*1} 石川翔吾^{*1} 上野秀樹^{*2,4} 小林美重^{*2,4} 前田実^{*3}
Naoki Kamiya Takumi Yoshizawa Shogo Ishikawa Hideki Ueno Mia Kobayashi Minoru Maeda

西山千秋^{*3} 村上佑順^{*3} 桐山伸也^{*1,4} 竹林洋一^{*1,4}
Chiaki Nishiyama Yujun Murakami Shinya Kiriya Yoichi Takebayashi

^{*1}静岡大学 Shizuoka University ^{*2}千葉大学医学部附属病院 Chiba University Hospital ^{*3}オレンジクロス Orange Cross Foundation

^{*4}みんなの認知症情報学会

The Society of Citizen Informatics for Human Cognitive Disorder

This paper describes construction of a diagnosis representation model of person with dementia based on ConceptNet to understand physical and mental conditions deeper. We constructed diagnosis knowledge of treatable dementia using ontology, as many of it remain tacit knowledge for familiar persons. The result of analyzing learners' diagnosis shows that our knowledge structure are useful to evaluate diagnosis by learner and represent diagnosis skills. Furthermore, the result suggest that practice of learning environment lead to understanding learning process on the basis of learner's diagnosis knowledge and improving accumulated knowledge structures.

1. はじめに

認知症は医学的な疾患が原因であり、その原因は 70 種類以上ある。その中には、原因を治療で取り除くことで脳機能を回復することができる治療可能な認知症があり、医療の関与は欠かせない。本稿では、認知機能の低下をもたらす多数の原因疾患の可能性について分析することを見立てと呼ぶ。これは、医師が原因疾患を鑑別診断するプロセスとは異なり、あくまでも医療職ではない身近な支援者が治療可能な認知症の可能性を検討することを意味する。しかし、関わりを主体とした介護現場では医学的に改善可能な側面が見逃されていることが多い。それに加えて見立てのプロセスを学ぶための方法が確立されていない。身近な支援者が医師の見立ての知識を学ぶことで、医師がどのような情報を必要としているかを理解することができ、適切な情報提供によって支援の質が改善されることが期待される。本研究では、介護従事者が見立てのプロセスを学ぶようにするため、見立てを表現する知識構造の設計を目指す。

2. 学びを活用した見立て知の構築

2.1 認知症の見立て

治療可能な認知症には、日常生活の中で気が付くためのポイントがある (表 1)。身近な人がそれらのポイント手掛かりに日々の関わりの中から治療可能な認知症を見つける必要がある。筆者らは、実践的な学びに繋げるためにケースメソッドアプローチの構成をベースに家族や介護、医療に従事する多職種が集い、ケースを通して認知症の見立てを実践的に学ぶ学習環境を設計した。学習会では、提示されたケースについて個人で見立てを行い、その結果をもとに 3 名から 5 名程度の少人数で見立てを行う。ケースメソッドアプローチではグループで出た意見を集約するが、本学習環境では、あらゆる改善可能な側面について検討し、知識を共有できるよう意見を集約させない所が大きな特徴である。

表 1: 治療可能な認知症

脳機能低下の原因	アセスメント方法
薬剤の副作用	薬の内服内容のチェック
うつ病, 精神的ストレス	うつ病の症状があるか
甲状腺機能低下症	むくみ, 食欲がないのに体重が増える
ビタミン B1, B12 欠乏	食事をきちんとしていない
正常圧水頭症	三主徴 (歩行障害, 尿失禁, 認知症)
意識障害, せん妄状態	状態の時間的変動, 特定の時間帯での増悪

2016 年より見立ての学習会を開催してきた結果、身近にいる人ならではの新たな解釈が得られることが分かり、見立て知の集積と洗練のための認知症の見立てコーパスの構築を進めている [神谷 2018]。学びを活用して多様な人々の知を収集・分析できる仕組みを構築することで、見立ての知を深め、皆がその知を学べる環境につながると期待される。

2.2 見立て表現モデルの設計

看護・介護に関する技術や行為に関する知識をオントロジーを用いて表現し、教育や知識共有で活用する取り組みが進んでいる [西村 2015][Hsieh2015]。筆者らは見立て知を収集するために、見立てを行う際に着目するケース中のポイントや判断プロセスを表現する見立て知を設計した。図 1 は、治療可能な認知症の一つ、せん妄の知識をオントロジーで表現したものである。身近な支援者は、日常生活の中でせん妄に気付く能力を習得する必要があるため、まずは、学習会のケース中から適切な情報を拾い上げることができたか評価する。そこで、せん妄の知識に対してケース中の語句をインスタンスとして結び付けておき、学習者が見立てた際の記述からせん妄に関する語句を抽出することでせん妄の知識構造との対応を分析した。

連絡先: 神谷直輝, 静岡大学 桐山研究室, 浜松市中区城北 3-5-1

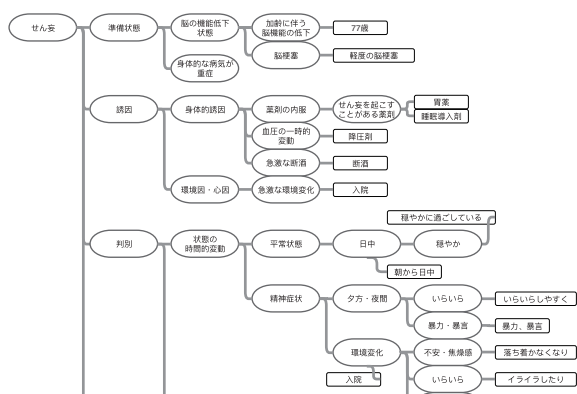


図 1: せん妄の知識の一部

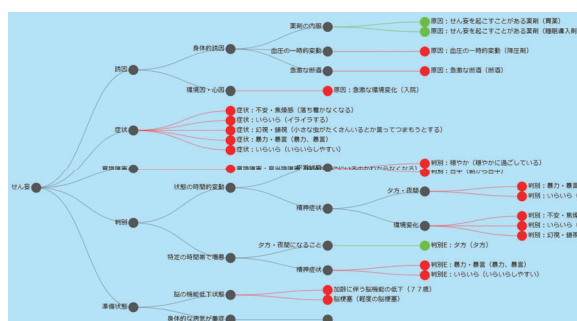


図 2: せん妄の構造を用いた見立ての評価

3. 見立て表現モデルを用いた学習プロセスの分析

3.1 学習者の見立ての評価

せん妄の見立て表現モデルを用いて学習者の回答を分析し、基本的なポイント全てが表示されるよう可視化すると図 2 のように表現できる。緑色のノードは回答に含まれていたポイントであり、赤色のノードは回答に含まれていなかったせん妄に関するポイントを意味する。例えば、「特定の時間帯で増悪」の下にある「判断 E: 夕方 (夕方)」は、学習者が「夕方」という語句を使って見立てた結果を記述しているため緑色のノードとなっており、その隣の「判断 E: 暴力・暴言 (暴力・暴言)」は「夕方」と「暴力」「暴言」「暴れる」という語句を使わずに見立てていたため赤色のノードとなっている。

3.2 学習者の見立て知の抽出

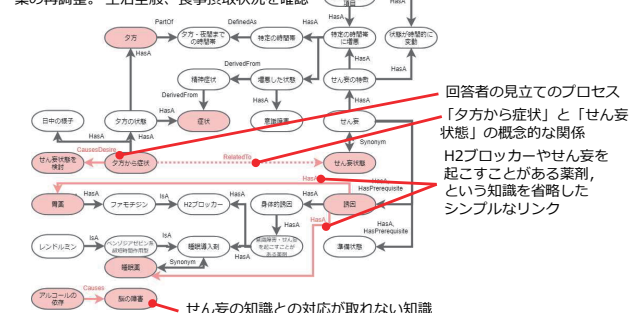
学びの評価を通して学習者がもつ見立て知を集積するために、学習者の見立てを ConceptNet [Liu2004] を用いて分析した。ConceptNet は、コモンセンス知識を扱うためのデータベースとして設計され、概念 (Concept) をノードとして表し、概念間の関係性 (Relation) を矢印で表す。IsA や PartOf, MotivatedByGoal といった関係性は 36 種類ある。

図 3 に図 2 と同じ学習者のひと月前後の回答結果とせん妄に関連する知識構造を示す。月に 1 度 2 時間の学習会を開催し、第 2 回目と第 3 回目の中で得られた結果である。回答の記述で下線を引いた部分が、ケースについて治療可能な側面があるか状態を分析した箇所であり、それ以外是对応方法に関する記述である。下線部を分析対象とし、回答に含まれていた語句を赤色のノードと関係性で示し、それらのノードに

せん妄の知識を黒色のノードと関係性で示した。得られたグラフを比較すると、ひと月後の方がせん妄に知識に関連する記述が増えており、せん妄の特徴の「夕方から症状」に気づいた事を起因として「せん妄状態を検討」、という見立てのプロセスも表現できた。また、誘因と薬剤の間にある知識を省略した関係性も見つかった。

1か月後の学習会での回答と構造

夕方から症状が出てくるという事なので、せん妄状態を検討する。誘因は、内服している胃薬、睡眠薬、が該当する可能性がある。服薬指導、薬剤を減らす。以前からのアルコールの依存による、脳の障害。



1か月前の学習会での回答と構造

服薬状況の確認。食事の状況確認。家庭内又は家族、近所との関係でストレスが生じていないか。血圧の値の確認。などを知り、良い状態へ改善する。規則正しい、生活リズムを持つ。

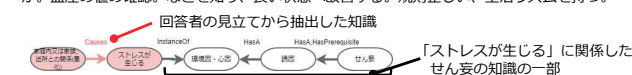


図 3: 見立て結果から抽出した知識構造の違い

4. まとめ

本稿では、見立て表現モデルを設計するために治療可能な認知症の知識を設計し、その表現モデルを活用することで学習者の見立ての評価と見立て知の表現に応用できることを示した。また、未解明な部分が多い見立てにおいて、学習者の常識的知識に基づいた見立ての分析や知の集積につながる事が示唆された。今後は見立ての協調学習会を継続的に開催しながら見立て知の集積と利用を進め、多様な人々による協調学習が創り出す知識を蓄積できるよう構造を洗練させる。

参考文献

- [神谷 2018] 神谷, 上野, 他: 認知症ケア高度化のための協調学習環境を活用した認知症の見立てコーパスの構築, 2018 年度人工知能学会全国大会 (第 32 回), 3Pin1-45(2018).
- [西村 2015] 西村, 笹嶋, 他: 目的指向の看護手順学習に向けた複数観点からの知識閲覧システム CHARM Pad と新人看護師研修への実践的活用, 人工知能学会論文誌, 30(1), pp.22-36(2015).
- [Hsieh2015] Hsieh, NC., Chiang, RD., et al., Ontology Based Integration of Residential Care of the Elderly System in Long-Term Care Institutions, Journal of Advances in Information Technology 6(3)(2015).
- [Liu2004] Liu, H. and Push S., "ConceptNet — a practical commonsense reasoning tool-kit, BT technology journal 22(4), pp.211-226(2004).

認知症のある人の個性表現に基づく 自立を重視した生活環境デザインの評価と分析

Evaluation and Analysis of design for life environment with a high regard for self-reliance
based on the representation of a self of people with dementia

寺面 美香^{*1} 石川 翔吾^{*1} 桐山 伸也^{*1} 加藤 忠相^{*2} 井出 猛^{*2} 竹林 洋一^{*1,3}
Mika Teramen Shogo Ishikawa Shinya Kiriyama Tadasuke Kato Takeshi Ide Yoichi Takebayashi

^{*1} 静岡大学
Shizuoka University

^{*2} 株式会社あおいけあ
Aoicare Co., Ltd.

^{*3} みんなの認知症情報処理学会
Citizen Informatics for Human Cognitive Disorder

This paper describes the evaluation of design for life environment with a high regard for self-reliance based on the representation of a self of people with dementia. We have structured care records in a care home where is pioneering efforts to support people with dementia, and constructed a model of personality expression. The personality expression tree makes it possible to see the connection of the life data and to compare and analyze the record of each facility. The result shows that the tree was possible to objectively evaluate the record of the facility, and learning with tree introduction is effective for reforming the consciousness of care practitioners toward self-reliance support.

1. はじめに

認知症は、何らかの脳機能低下の要因による認知機能障害とそれにより生じる生活障害で定義される。認知症とともに生きるには、生活に支障が出る状況において適切な生活環境を整えられるよう支援する必要がある。多くの当事者が声を上げている[佐藤 14, 丹野 17]ように、認知症の症状は千差万別で、ゆえに、型にはまった支援方法があるわけではない。生活は個人ごとに異なるため、当事者のパーソナリティ、身体的状況や精神的状況、家庭や地域における暮らし方(本稿ではこれらを個性を構成する要素として捉える)を理解しながら支援につなげる必要がある。介護においては作成したケアプランにそった支援が行われるが、目標の立て方、支援のための記録の作り方やその使い方には方法論がない[日総 12]のが現状である。

そこで本稿では、認知症のある人の個性表現に基づき、自立を重視した生活支援の評価と分析について述べる。

2. 個性表現モデルの基本構造

2.1 個性情報の構造化

本稿では、生活支援を評価するために介護記録に着目する。介護現場では、生活支援のためのケアプランや利用者の特徴を表すフェイスシート、日々の記録等、様々な個性に関する情報がある。しかし、これらは介護業界全体で統一された指標があるわけではなく、独自のフォーマットや書き方の作法がある。そこで、記録と支援の関係を整理するために、個性に基づく自立支援において先駆的な取り組みを実施している介護施設の記録を分析し、記録の特徴を表現するための構造化を行った。以下に手順を示す。

- (1)介護施設で収集している利用者に関する情報の調査
- (2)スタッフ 20 名が参加し、生活支援情報の特徴を抽出する検討会の実施
- (3)抽出した生活支援に関する情報を[大久保 13]のキーワードを活用して KJ 法でグルーピング

(4)項目間の関係を ConceptNet [H Liu 04]を活用し構造化
その結果、日々の変動の大きい日常生活動作(Activity of Daily Living: ADL)の把握以上にパーソナル情報の収集を重視することにより当事者を理解し寄り添う支援を実現させていることがわかった[寺面 18](図)。

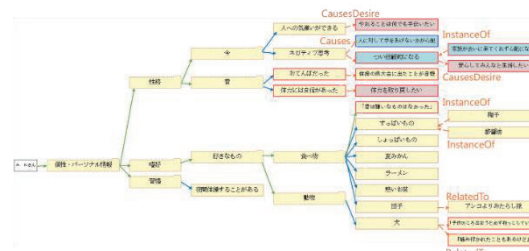
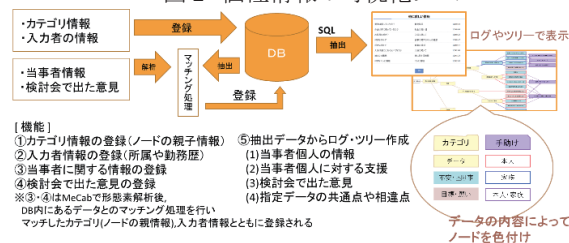


図 1 個性表現ツリー

2.2 個性情報可視化ツール

構造化した結果を元に情報の可視化ツールを作成した。ツールではデータベースにカテゴリ情報や入力者情報、当事者情報の登録を行うことで各情報のツリーやログを生成し状況を可視化することができる(図)。このツールを用いて施設のアセスメントの状況の評価・分析を行う。

図 2 個性情報の可視化ツール



3. 個性表現に基づく介護記録の評価

2.2節で作成したツールを活用し、施設ごとのデータを比較することで個性に基づいた生活支援の質を評価する。施設A、施設Bの二つの施設を対象に特定の利用者1名に関する施設の記録を収集し分析した。

図に比較結果を示す。①医療関係と②性格・嗜好に関する情報を可視化したものである。施設Bでは医療に関する項目は詳細に設定されており症状が丁寧に記載されている。それに対して、施設Aでは身体状況に関する本人の不安や家族の願いが表れるツリーとなった。また、施設Aでは性格・嗜好情報の収集を重視しており、味の好み等を細かく行っている他、それに関する本人の発言やエピソードも拾われている。施設Bでは性格や嗜好に関する記述はほとんど見られなかった(表)。このことから、施設Aは個性情報の中のパーソナリティに関することとその他のことを関係づけて支援をしており、当事者のこれまでの生活を踏まえながら自立支援に繋がっていると評価される。

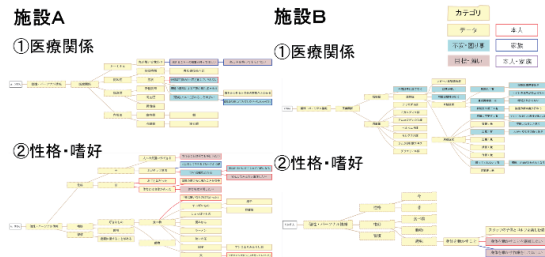


図3 施設A/施設Bの個性表現ツリー

表1 施設A/施設Bの個性情報に関するノード数

	施設A		施設B	
	①医療関係	②性格・嗜好	①医療関係	②性格・嗜好
ノードの深さ	7	7	8	6
全ノード数	18	26	30	4
本人の発言	0	6	0	0
本人の願い	0	3	0	1
本人の不安	1	2	0	0
家族の願い	2	0	0	1
家族の不安	1	1	0	0

4. 個性表現に基づくケース検討の分析

4.1 個性表現ツリーに基づく学習支援

個性表現ツリーが介護スタッフにどのような気付きを与えるかを検証するため、施設Bにて個性表現ツリーを用いた生活支援検討の場を設計した。12人の介護スタッフを3チームに分けグループワークを行い、個性表現ツリーを見ながら各項目に対して検討することとした。支援に関する検討として、以下の結果が得られた。

グループA:「できることはやってもらって、できていないことをサポートする形」と、ツリーから読み取れるケアに対して自立支援に繋がることをやっているという評価を行っていた。

グループB:「その場しのぎのケアが多い」と原因探索の必要性に対する議論が行われ、また、「個性が認知症に締められるのはおかしい」という評価も出ていた。

グループC:グループBと同様に「自立支援に向け成功事例や失敗事例の詳細を共有する必要がある」という議論がなされ、また、「本人の意向を無視して施設の都合の良ように動かしてしまっているかも」という意見も出ていた。

グループAと他のグループの意見が異なるが、これを全体で議論することによって考えが共有され、グループBやCの検討の方向で考えることが重要であると結論づけられた。以上のように、記録の意味や記録の作り方、活用の仕方について気づきが得られた結果が示された。

4.2 ConceptNetを用いた比較

生活支援は常識的思考が集約された場である。そこで、ConceptNetに着目し、実際の現場のケース会議における議論が、個性表現モデルと対応するかを分析することで支援の質の

評価につながるかを分析した。ConceptNet5のRelationに基づいて設計された個性表現ツリーに基づき施設Aと施設Bの議論をRelationの数で算出するとIsA,HasA以外の関係と数は表のようになった。グループBの議論で医療情報について触れられていたが、施設Bの医療情報はIsA,HasA,Causes関係のみで構成されていたのに対し、施設Aの医療情報では事象から引き出される望み(CausesDesire)や目標(MotivatedByGoal)が見て取れた(図4)。全体も思い(MotivatedByGoal, Desires), 原因探索(Desires, CausesDesire, ReceivesAction, ObstructedBy), 詳細化(InstanceOf, RelatedTo)に関わる記述が多くなっており、質の高い記録は単純なデータの数やノードの深さだけでなく情報の繋がりも重要となってくることが伺える。

以上のことから、個性を重視した介護記録を作るには利用者との関係を築いて人となりを理解することが重要となり、そこから見てきた本人や家族の意思の尊重・不安の解消を目指すことが質の高い自立支援に繋がることが示唆された。

①医療関係

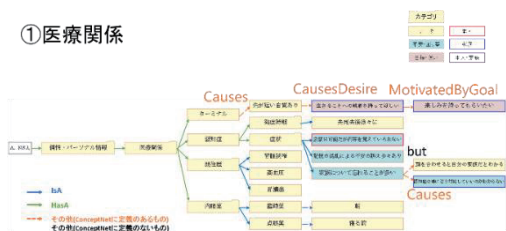


図4 ノードの関係記述(施設A①医療関係)

表2 施設A/施設Bのツリーで表出したノードの関係と数

ConceptNet 5 Relation URI	施設A	施設B
MotivatedByGoal	1	1
Desires	2	0
Causes	5	6
CausesDesire	5	3
ReceivesAction	1	0
ObstructedBy	3	1
InstanceOf	3	0
RelatedTo	5	1

5. おわりに

本稿では個性表現モデルの設計と、モデルを活用して生活環境デザインプロセスを分析した。モデルを個性表現ツリーとして可視化することにより客観的な比較・分析が可能となり、ツリーを導入した学習は質の高い支援に向けてのスタッフの意識変革を行うものとなった。生活環境デザインは、個性の理解と環境デザインに大別される。今後は、環境デザインの方法論も検討することで、自立重視の支援の実現に貢献したい。

参考文献

[H Liu 04] H Liu and P Singh: ConceptNet — a practical commonsense, BT Technology Journal, 2004.

[大久保 13] 大久保幸積, 他: 認知症ケアの視点が変わる「ひとときシート」活用ガイドブック, 認知症介護研究・研修東京センター, 2013.

[佐藤 14] 佐藤雅彦: 認知症になった私が伝えたいこと, 大月書店, 2014.

[丹野 17] 丹野智文: 丹野智文 笑顔で生きる -認知症とともに-, 文藝春秋, 2017.

[寺面 18] 寺面美香, 他: 認知症のある人の生活環境デザインのためのあおいけあナレッジの抽出と構造化, 2018 みんなの認知症情報学会, 2018.

[日総 12] 日本総合研究所: 介護支援専門員の資質向上と今後のあり方に関する調査研究ケアプラン詳細分析結果報告書, 2012.

畳み込みニューラルネットワークを用いた

脳波ブレイン・マシン・インターフェイスの開発

Development of Electroencephalogram brain-machine interface using convolutional neural network

加藤正起^{*1}
Masaki Kato

嶋田総太郎^{*2}
Sotaro Shimada

^{*1} 明治大学大学院
Meiji University Graduate School

^{*2} 明治大学
Meiji University

Brain-machine interface (BMI) is a system that manipulates machines directly from the brain activity data. Convolutional Neural Network (CNN) which enables end-to-end learning allows extracting information from brain activity [Antoniades 2016][Bashivan 2016]. In the previous study [Schirrmeister 2017], CNN was used to predict four kinds of motor imagery of right arm, left arm, foot and rest from electroencephalogram (EEG) data. In this study, we measured EEG when an individual performed reaching movements with his right arm. We propose a BMI system by using CNN that can classify four directions of reaching motion of the right arm from EEG data with about 78% accuracy.

1. はじめに

ブレイン・マシン・インターフェイス(BMI)は脳活動を解析することで、脳から直接、コンピュータや機械を操作する技術である。非侵襲的な脳波(EEG)や機能的核磁気共鳴計測(fMRI)による研究では、脳活動からコンピュータ画面のカーソルの操作[Wolpaw 1991]や車椅子の操作[Choi 2012]を可能にする研究が行われている。近年では脳波を含む脳活動の分析のために end-to-end 学習が可能である畳み込みニューラルネットワーク(Convolutional Neural Network : CNN)を用いた研究が行われている[Antoniades 2016][Bashivan 2016]。

先行研究[Schirrmeister 2017]ではCNNを用いて脳波から右腕、左腕、足の運動想起とレストの4種類を約90%の精度で予測した。しかし、それぞれの運動想起では握るなど具体的な運動を行っているわけではないため、四肢の詳細な運動を分類できるわけではない。

本研究では、右腕でリーチング運動を行った時の脳波を計測し、CNNを用いて、右腕の4種類の運動を分類することを可能にするBMIを開発した。

2. 実験方法

2.1 被験者

2名の健康な右利きの男性(23歳)が実験に参加した。被験者には、事前に実験方法などの説明を行い、書面にて実験参加の同意を得た。

2.2 実験環境

図1のようにプロジェクタ(EB-W420, EPSON)によって壁にゲーム開発ソフト(Unity, Unity Technologies)で作成したアバターアームを呈示した。このアバターアームの動作はジョイスティック(T.Flight Stick X, THRUSTMASTER)により制御され、被験者はアバターアームを右腕で操作することで、リーチング課題を行った。被験者は実験中、顎台に顎を置き、自分の手を見ないようにした。

実験装置を接続することによって生じる被験者の実際の手とアバターアームの運動の遅延は、ハイスピードカメラ(EX-F1, CASIO)で測定したところ約200msであった。

また、EEGの計測はg.tec社製のg.USBampを用いて、以下の条件で行った。

- 電極数: 60ch
(Fp1, Fpz, Fp2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2)
- 眼電: 4ch(左右の垂直水平成分)
- サンプリング周波数: 1200 Hz
- 帯域通過フィルタ: 0.5-100 Hz

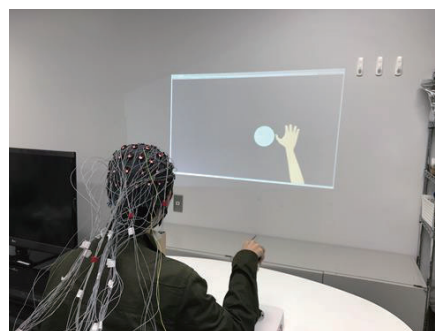


図1 実験風景

2.3 実験手順

1 トライアルの流れを図2に示す。被験者がアバターアームを一辺12cmのベースポジション上に置くことでトライアルの開始とした。ベースポジションの位置は毎回同じで、直径12cmの円のターゲットはベースポジションの位置から前後左右の20cm離れた場所に出現させた。ターゲットの4通りの位置(front, back, left, right)はランダムに出現させた。実験中は右腕のアバターアームだけが表示された。また、左腕は動かさないように被験者に指示をした。実験では2セッションを行い、セッション内には8つのブロックがあった。それぞれのブ

連絡先: 加藤正起, 明治大学大学院理工学研究科, 神奈川県
川崎市多摩区東三田 1-1-1, E-mail: ce181021@meiji.ac.jp

ロックでは 150 トライアル繰り返した。被験者には、できるだけ早く、正確にリーチング運動を行うよう教示した。なお事前に 1 ブロック分 (150 トライアル) の練習を行った。

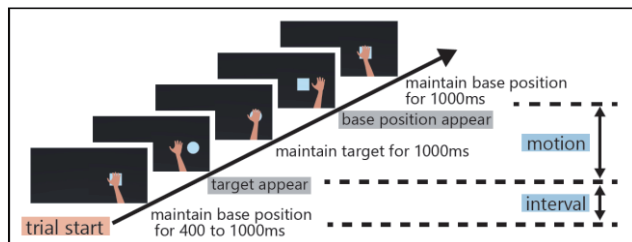


図 2 1 トライアルの流れ

3. 解析

3.1 前処理

測定した脳波データに対して、1Hz の高域通過フィルタをかけ、250Hz までダウンサンプリングを行った。1 セッション分のトライアルの中で $-100\mu\text{V} \sim +100\mu\text{V}$ の範囲を超える脳波データが 50% 以上含まれる電極 (Subject1: F6,T8,PO8,O1,Oz,O2, Subject2: T8,PO3) を解析から除外した。ターゲットが表示される 0.5 秒前から表示後 3.25 秒 (Subject2 は 3.2 秒) までを 1 トライアルとし脳波データを分割した。 $-100\mu\text{V} \sim +100\mu\text{V}$ または $-60\text{dB} \sim 40\text{dB}$ の範囲を超える脳波データを含むトライアルを除いた後、脳波データに含まれる眼電成分と筋電成分を取り除くために独立成分分析を行った。

Subject2 の脳波データは測定機器の不良により、セッション 1 と 2 の 5, 6, 7, 8 ブロックのデータにノイズが多く含まれていたため、解析から除外した。これらの前処理は脳波解析ソフトである EEGLAB (v14.1.2) で行った。

3.2 入力データの作成

トライアルごとに分割された脳波データに対し、それぞれに対応した種類のラベル (front, back, left, right) を付けた。1 トライアル分の脳波データを時間窓 2 秒、スライド幅 0.04 秒で分割することで 1 トライアルあたり 44 個 (Subject2 は 43 個) の脳波データを作成し、入力データとした。1 トライアルあたり 44 個または 43 個の入力データが作成されるが、それぞれのラベルについては分割する前のトライアルに付属されたラベルと同じものを使用した。入力データのサイズは時間ステップ数 $\times$ 電極数で、Subject1 は 500×53 、Subject2 は 500×58 であった。

また、脳波データを学習データ、検証データ、テストデータの 3 つに 8 : 1 : 1 で分割した。

3.3 CNN による運動方向分類モデル

先行研究 [Schirmmeister 2017] のモデルをもとに脳波から右腕の運動方向を分類するために図 3 のような CNN を深層学習のフレームワークである TensorFlow を用いて作成した。

CNN は主に、画像データに対して用いられているが、本研究では入力データとして時系列信号を用いた。これは、脳波をウェーブレット変換などの時間周波数解析によって、画像として学習することも可能であるが、学習させる特徴量が 1 つの電極につき周波数分、増えてしまう。そのため本研究では、脳波を時系列信号の状態での学習することで特徴次元数を減少させた。

脳波は時系列情報を持つ信号であり、脳内の複数の双極性電流源によって引き起こされる空間的に広がる電圧の線形重ね合わせに近似すると仮定されている [Nunez 2006]。時間的・空間的な情報をもつ脳波の特徴を捉えるため、第 1 層目で電極ごと

に脳波データを時間方向に畳み込んだ後、全電極分の特徴量を一度に畳み込むようなモデルを設計した。また、すべての畳み込み層のレイヤーにおいてドロップアウトを使用することで過学習を抑制した。最後の全結合層のソフトマックス関数によって入力データは 4 種類の運動のいずれかに分類された。

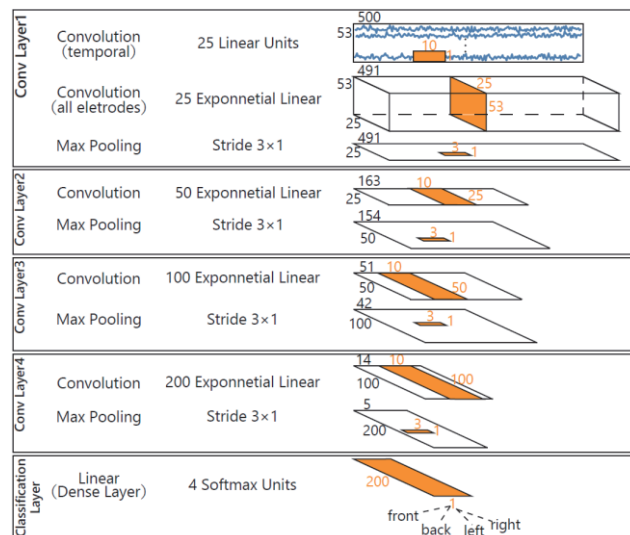


図 3 運動方向分類モデルの構造

4. 結果

テストデータを入力したときの混合行列を図 4 に示す。Subject1 と Subject2 の分類精度はそれぞれ 77.90% と 56.73% であった (チャンスレベルは 25%)。

本実験結果より、作成したモデルによって脳波から右腕の 4 種類の運動を最大で約 78% のという高い精度で分類できることが分かった。Subject2 の分類精度が Subject1 に比べ、20% 程低くなっている。これはノイズにより全データの半分しか学習データに使用できなかったため、汎化性能が低くなってしまったと考えられる。

predicted label	right	1749 90.3%	136 7.0%	82 4.2%	366 18.9%	74.97%
	front	75 3.9%	1497 77.3%	58 3.0%	384 19.8%	74.33%
	left	64 3.2%	245 12.4%	1778 89.8%	153 7.7%	79.38%
	back	48 2.4%	58 2.9%	62 3.1%	1077 54.4%	86.51%
	Sensitivity	90.34%	77.32%	89.80%	54.39%	Accuracy 77.90%
true label						Precision
right						
front						
left						
back						

(a) Subject1

predicted label	right	453 43.9%	8 0.8%	48 4.7%	59 5.7%	79.75%
	front	159 15.4%	873 84.6%	208 20.2%	312 30.2%	56.25%
	left	330 32.0%	100 9.7%	623 60.4%	268 26.0%	47.16%
	back	90 8.7%	51 4.9%	153 14.8%	393 38.1%	57.21%
	Sensitivity	43.90%	84.59%	60.37%	38.08%	Accuracy 56.73%
true label						Precision
right						
front						
left						
back						

(b) Subject2

図 4 テストデータ入力時の混合行列

5. まとめ

分類精度が約 78% と先行研究 [Schirmmeister 2017] の 90% と比べ低くなっているが、先行研究で行われた右腕、左腕、足の運動想起はそれぞれ、左脳の一次運動野、右脳の一次運動野、頭頂葉付近の一次運動野など脳の活動領域が異なる。それに比べ本研究で行った右腕のリーチング運動は左脳の一次運動野の中での活動であり、その狭い脳領域で 4 種類の運動の特徴を捉える必要があったため、精度が下がってしまったと考えられる。また、1 トライアルのなかでターゲットにアバターアームを

伸ばして戻すという運動を行うが、この一連の動作に対して1つのラベルを割り当てているためだと考えられる。

しかし、本研究では右腕の4種類の運動の分類を可能にした。今度は学習データとして右腕運動時の脳波だけでなく、左腕や両腕を動かしたときの脳波を用いて、BMIとしての分類数をさらに増やしていきたい。これによって、より現実的なBMIの開発を目指す。

参考文献

- [Wolpaw 1991] Wolpaw, J. R., McFarland, D. J., Neat, G. W., Forneris, C. A.: An EEG-based brain-computer interface for cursor control, *Electroencephalography and Clinical Neurophysiology*, Vol.78(3), pp.252-259, 1991.
- [Choi 2012] Choi, K.: Control of a vehicle with EEG signals in real-time and system evaluation, *European Journal of Applied Physiology*, Vol.112(2), pp.755-766, 2012.
- [Antoniades 2016] Antoniadis, A., Spyrou, L., Took, C. C., & Sanei, S.: Deep learning for epileptic intracranial EEG data, *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, 2016.
- [Bashivan 2016] Bashivan, P., Rish, I., Yeasin, M., & Codella, N.: Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks, *ICLR*, pp.1-15, 2015.
- [Schirmer 2017] Schirmer, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., ... Ball, T.: Deep learning with convolutional neural networks for EEG decoding and visualization, *Human Brain Mapping*, 38(11), 5391-5420, 2017.
- [Nunez 2006] Nunez, P. L., & Srinivasan, R.: *Electric Fields of the Brain, The Neurophysics of EEG Second Edition*, Oxford University Press, 2006.

点眼瓶センサーと Deep Learning による 緑内障患者点眼アドヒアランス自動把握能力の評価

Evaluation of Automatic Monitoring of Instillation Adherence Using Eye Dropper Bottle Sensor and Deep Learning in Patients with Glaucoma

田淵 仁志^{*1*2} 西村 和晃^{*1} 中倉 俊祐^{*1} 升本 浩紀^{*1} 田邊 裕貴^{*2}
Hitoshi Tabuchi Kazuaki Nishimura Shunsuke Nakakura Hiroki Masumoto Hirotaka Tanabe

野口 明日香^{*1} 青木 良太^{*1} 木内 良明^{*2}
Asuka Noguchi Ryota Aoki Yoshiaki Kiuchi

^{*1} 社会医療法人三栄会ツカザキ病院
Tsukazaki Hosipital

^{*2} 広島大学
Hiroshima University

Purpose: We developed and evaluated an eye dropper bottle sensor system comprising motion sensor with automatic motion waveform analysis using deep learning (DL) to accurately measure adherence of patients with antiglaucoma ophthalmic solution therapy. **Results:** The developed eye bottle sensor detected all 60 instillation events (100%). Mean difference between patient and eye bottle sensor recorded time was 1 ± 1.22 (range; 0–3) min. Additionally, mean instillation movement duration was 16.1 ± 14.4 (range; 4–43) s. Two-way ANOVA revealed a significant difference in instillation movement duration among patients ($P < 0.001$) and across days ($P < 0.001$). **Conclusion:** The eye dropper bottle sensor system developed by us can be used for automatic monitoring of instillation adherence in patients with glaucoma.

1. 諸言

緑内障患者は約 8 千万人いると推測され、年々その数は増え続けている。緑内障は世界的に失明の主要な原因の病気の 1 つである。緑内障点眼治療には患者アドヒアランスの問題が大きく存在する、例えば、新規にプロスタグランジン系の緑内障点眼薬を開始した患者の約 3 割が点眼治療を開始して 3 ヶ月の間で治療をドロップアウトしているという報告がある。こうしたアドヒアランス不良は治療の非効率化を招くと同時に、視野障害の進行するリスクが 6 倍以上になると報告されている。従って、緑内障点眼治療におけるアドヒアランス向上のための方法論は、臨床上重要な視点として、自己申告によるもの、点眼瓶の重量変化によるもの、点眼瓶ケースの挙動によるものなど、様々な報告されてきた。ただ煩雑さや信用性の低さなどの問題がそれぞれにあり、実際の臨床現場で一般的に定着している方法論が我々の知る限り、いまだ存在しない。

一方で、最近になって、モバイルヘルスを利用した薬物治療におけるアドヒアランス向上の取り組みが報告されている。モバイルヘルスは、スマートフォンやタブレット端末を用いて個別の情報をネットワーク接続し集積するという Internet of things の概念の範疇である。自動的に膨大なデータを緻密に収集できる上に、人手を介することがないためにコストパフォーマンスに圧倒的に優れている。すでに慢性の病気である高血圧における降圧剤の錠剤にセンサーを装着する事で患者の内服状況を監視するシステムなどが報告されている。IoT と同様に、医療への応用が広がり始めた技術が、Deep Learning である。DL は、それ以前の機械学習の方法論からは一線を画す高い精度の識別能力を持つ。眼科領域でも、画像識別の領域を中心に数多くの報告が出始めている。まだ定まった方法論がない緑内障治療における点眼アドヒアランス向上に、新しい技術である IoT や DL を応用した取り組みは我々の知る限り報告がない。

今回我々は点眼瓶の挙動そのものを自動的に検出する点眼瓶センサーを開発した。さらに、DL を用いたセンサー波形の点眼挙動自動識別モデルを作成し、システムとして組み合わせた。(以下、点眼瓶センサーシステム)。今回の検討では、緑内障患者に実際に点眼瓶センサーを使用してもらい、緑内障点眼に関する患者アドヒアランス自動把握の能力の評価を行った。

2. 方法

2.1 点眼瓶センサーの構造

点眼瓶センサーの構造は以下で構成される。点眼瓶、TWE-Lite2525A(以下センサー)、点眼補助フォルダ、処理端末部、電源コードからなる。点眼補助フォルダは点眼瓶を上に乗せるためのベースと点眼瓶を固定させるためのキャップと底蓋からなる。ベースは中が空洞になっており、センサーを装着する事が可能である。尚、センサーに CR1632 型リチウム電池をセットする必要がある。処理端末部の詳細は以下の通りである。(図 2)処理端末部のデータを記録するシールドにはデータロガーシールド(ADLSLD)を使用した。データロガーシールドには RTC(real time clock)機能、RTCを動かすための CR1220 型リチウム電池、SD ソケット、外部データインターフェイス、誤動作した際に押すとプログラムが元に戻るリセットボタンなど様々な機能が元々備わっている。さらに電源コードとつなげるための電源アダプター差し込み口や USB 差し込み口も存在する。これに TWE-Lite 無線マイコン、12C バス用双方向電圧レベル変換モジュール(PCA9306)、ケース内組込型アンテナ MW-A-P4208、データ受信確認用の LED やブザーも付け加えた。

2.2 点眼瓶センサーの仕組み

点眼瓶の挙動によりセンサーが検知し IEEE802.15.4 準拠の無線で処理端末部へデータを送信する。センサーは X,Y,Z の 3 軸の加速度センサーが搭載されており、約 0.08 秒刻み間隔で

重力加速度値を検出する。送信されたデータはケース内組込型アンテナ MW-A-P4208 で受信されて、TWE-Lite 無線マイコンにデータが送られる。無線マイコンでデータ进行处理し、ケーブルを通じて外部データインターフェイスに送られ、最終的に SD ソケット内の SD Memory Card にデータが記録される。TWE-Lite 無線マイコンの電源電圧が 5V で SD ソケット内の電源電圧が 3.3V のために 12C バス用双方向電圧レベル変換モジュール (PCA9306) を使用した。さらに、無線でデータが受信すると LED が光り、ブザーが鳴るなどデータの受信が確認可能な機能も併せて掲載した。データロガーシールドに RTC 機能が掲載されているため、縦軸が重力加速度、横軸が時刻のグラフ化された形で SD カードに記録される仕組みである。加速度センサーは重力加速度を検出するために、重力加速度がどの方向にどれだけかかっているのかを検出する事でセンサーがどれだけ傾いているか検出可能である。今回のシステムでは重力加速度が最小値となる時刻が点眼中である。重力加速度が+1G から小さくなっていく時刻を A、再び+1G に戻ってくる時刻を B として $(A+B) \div 2$ を点眼波形データの点眼時刻(秒単位まで)、 $(B-A)$ を点眼動作時間と定義した。

2.3 DL 点眼判定モデルの構築法

点眼瓶にセンサーを装着し、直立した状態を 1、そこから反転して下方を向いた状態を-0.1 として約 0.08 秒刻み間隔で取得した 1 次元の波形データを 216pixel×216pixel のフルカラー画像の線グラフとして、それを入力値として解析を行った。

学習モデルを生成するため健常者によるデータ 400 個を使用した。健常者に空のラタノプロスト点眼液 0.005%と点眼瓶センサーを渡し、両眼に点眼薬が入ったという仮定(シミュレーション点眼)を行ってもらい、点眼波形データを 200 個取得した。さらに点眼瓶を横に移動させるといったような点眼でない動作波形データも 200 個取得した。学習精度を上げるために波形全体のデータの他、波形の前部分を切り出したもの、波形の後方を切り出したもの、波形を平滑化したものを元のデータから新たに生成し、これも学習データとして使用した。

学習データを 64×64pixel の画像に変換し入力値として使用。各畳込み層(conv2d 1~conv2d 5)と、活性化関数(ReLU を使用)と pooling 層(max pooling2d 1~max pooling2d 4)を通り、平滑化(Flatten)の工程を経て、その後全結合層(dense 1、dense 2)と、その間に 10%の確率を設定した Dropout 処理を挟んで凡化性能を上げ、最終的に出力層で Softmax 関数を用いたクラス分類を行い、点眼動作をしている波形かを識別した。重みを最適化させるためのアルゴリズムは Adam を採用した。上記の学習サイクルを 30 回繰り返すことで内部パラメータを最適化した DL 点眼判定モデルを作成した。

2.4 対象患者と評価方法

2017.3/1~2017.9/30 の間でラタノプロスト点眼液 0.005%もしくはラタノプロスト・チモールマレイン酸塩配合点眼液を両眼に点眼中の開放隅角緑内障患者 20 人を対象にした。患者 20 名の平均年齢(標準偏差)は 64.1(12.06)歳で性別比は男 9 名女 11 名であった。点眼瓶センサーを自宅にて設置してもらった。(図 5)同一空間内で処理端末部からおよそ 10m 範囲内で 3 日間の点眼と点眼時刻の自己記載(分単位まで)をしてもらい、後日、SD カードと点眼時刻の書かれた紙を回収した。患者データから DL 点眼判定モデルを利用して自動取得された点眼と判定された波形データの点眼時刻と、自己記載点眼時刻を比較して 5 分以内であるものを検出成功として、検出率を算出した。

さらに各々の点眼動作時間についても算出した。点眼動作時間について個人間の差を分散分析にて検討した。Deep Learning モデルでの解析時間も算出した。使用コンピュータは MacBook Pro(Retina,13-inch,Early 2015)、外部装置と eGFX Breakway Box(eGPU Expansion System)[GPU-350W-TB3Z]と NVIDIA GeForce を使用した。

3. 結果

患者 20 名の各年齢及び性別については表 1 に示す。また、患者 20 名の MD 値の平均(標準偏差)が右眼は-8.5(7.48)デシベルで左眼は-6.9(6.55)デシベルであった。DL 点眼判定モデルから点眼と判定された波形データは 60 個であった。各患者における点眼波形データの点眼時刻と自己記載時刻の差の平均(標準偏差)1(1.22)分(最大 3 分~最小 0 分)であった。今回の検出成功基準(<5 分以内)にあてはめると検出率は 100%であり、実際の点眼回数と過不足なく自動取得できた。計 60 個の点眼動作時間の各患者の平均(標準偏差)が 16.1(14.41)秒(最大 43 秒~最小 4 秒)であった。点眼動作時間は一元配置の分散分析にて個人間で有意差があった。(ANOVA $P<0.0001$)

4. 考察

第一に、我々が開発した点眼瓶センサーシステムを緑内障患者 20 名に使用し、3 日間の点眼状況を 100%正確に自動抽出することに成功した。そのデータの解析時間は 1 分以内であった。まず、自己申告によらず、点眼状況が客観的に正確に把握できるという点に価値がある。なぜなら実際のアドヒアランスと自己申告にはほとんど相関がなく、点眼を本当はしていないのに、点眼をしていると報告する傾向があるからである。5) さらに短時間で情報を自動取得できる点は、実臨床への適応の障壁が低いことを示唆している。

第二に、点眼動作時間という我々の知る限りこれまでに報告のない新しい指標を計測することに成功し、個人間でその時間に統計学的に差があることを示した。今回のデータでは、最長の平均点眼時間は、個人 No14 の 43 秒であり、最小は個人 No3 の 4 秒であった。約 40 秒の差は 2 滴以上の点眼や正確に眼中に滴下することが困難など点眼動作に何らかの違いがあることが予想される。過去の報告では、点眼手技の問題がアドヒアランス不良の要因になっていると指摘されている。点眼動作時間という新しい指標が、手技の改善指導に役立つ情報となると私達は考えている。

今回の検討が 3 日間という短期間である点が、結果の解釈で最も注意が必要な点である。期間が長くなることで、機械的な問題が発生しやすくなり、患者さんのシステムの扱いもより雑になる可能性が否定できない。さらに、今回の検討は 1 種類の点眼での検討であり、多剤点眼使用者に対してどの程度システムが応用できるかについても今後の検討が必要である。点眼瓶センサーにかかる経費の詳細な検討も今回は行っていない。ただ、センサーの部品については携帯電話に用いられる大量生産されている低コストの部品を用いており、部品代については日本円で 1000 円程度である。

薬剤師が高齢者などの残薬を整理した所、医療費が約 80%削減されたと報告されている。先進国共通のテーマである社会保障費増大の問題に対して、残薬による医療費の浪費は小さくない問題である。点眼瓶センサーシステムの実用化は、アドヒアランス向上につながり、ひいては医療費の効率利用につながる可能性を持っている。

今回我々は、Deep learningと点眼瓶センサーを組み合わせて、点眼状況を客観的に正確に把握できるシステムを開発した。さらなる開発、検討を重ねて、実用化を目指していきたい。

参考文献

- [Quigley 2006] Quigley HA, Broman AT: The number of people with glaucoma worldwide in 2010 and 2020, British Journal of Ophthalmology, BMJ, 2006
- [Pascolini 2012] Pascolini D, Mariotti SP: Global estimates of visual impairment: 2010, British Journal of Ophthalmology, BMJ, 2012
- [Nordstrom 2005] Nordstrom BL, Friedman DS, Mozaffari E, Quigley HA, Walker AM: Persistence and adherence with topical glaucoma therapy, American Journal of Ophthalmology, American Academy of Ophthalmology, 2005
- [Okeke 2009] Okeke CO, Quigley HA, Jampel HD, Ying GS, Plyler RJ, Jiang Y, Friedman DS: Adherence with topical glaucoma medications monitored electronically the Travatan Dosing Aid study, Ophthalmology, American Academy of Ophthalmology, 2009
- [Sleath 2011] Sleath B, Blalock S, Covert D, Stone JL, Skinner AC, Muir K, Robin AL: The relationship between glaucoma medication adherence, eye drop technique, and visual field defect severity, Ophthalmology, American Academy of Ophthalmology, 2011
- [Lievens 2006] Lievens CW, Guvant P, Newman J, Gerstner M, Simpson C: Effect of Proview self-tonometry on pharmaceutical compliance, Clinical and Experimental Optometry, Wiley, 2006
- [Robin 2007] Robin AL, Novack GD, Covert DW, Crockett RS, Marcic TS: Adherence in glaucoma: objective measurements of once-daily and adjunctive medication use, American Journal of Ophthalmology, American Academy of Ophthalmology, 2007
- [Stegemann 2012] Stegemann S, Baeyens J-P, Cerreta F, Chanie E, Löfgren A, Maio M, Schreier G, Thesing-Bleck E: Adherence measurement systems and technology for medications in older patient populations, European Geriatric Medicine, Springer, 2012
- [Park 2016] Park N, Kang N: Mutual Authentication Scheme in Secure Internet of Things Technology for Comfortable Lifestyle, Sensors(Basel), MDPI, 2016
- [DiCarlo 2016] DiCarlo LA, Weinstein RL, Morimoto CB, Savage GM, Moon GL, Au-Yeung K, Kim YA: Patient-Centered Home Care Using Digital Medicine and Telemetric Data for Hypertension: Feasibility and Acceptability of Objective Ambulatory Assessment, The Journal of Clinical Hypertension(Greenwich), Wiley, 2016
- [Trebeschi 2017] Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, Peters NHGM, Beets-Tan RGH, Aerts HJWL: Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR, Scientific Reports, Springer, 2017
- [大杉 2017] 大杉秀治、田淵仁志、遠野宏季、石飛直史: Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment, Scientific Reports, Springer, 2017
- [Tamrat 2015] Tamrat L, Gessesse GW, Gelaw Y: Adherence to topical glaucoma medications in Ethiopian patients, Middle East African Journal of Ophthalmology, Wolters Kluwer, 2015
- [小柳 2013] 小柳香織、窪田敏夫、小林大介、木原太郎、吉田武夫、三井所尊正、斎藤友亮、打越英恵、高木淳一、瀬尾隆、島添隆雄: 節薬バッグ運動 外来患者の残薬の現状とその有効活用による医療費削減の取組み、薬学雑誌、日本薬学会、2013

リモート PPG 信号による睡眠／覚醒判別 Sleep/wake classification using remote PPG signals

張 雅文^{*1}
Yawen Zhang

辻川 剛範^{*2}
Masanori Tsujikawa

大西 祥史^{*2}
Yoshifumi Onishi

^{*1} Department of Electronic & Computer Engineering, Hong Kong University of Science and Technology

^{*2} Biometrics Research Laboratories, NEC Corporation

This paper proposes a remote sleep/wake classification method by combining vision-based heart rate (HR) estimation and convolutional neural network (CNN). Instead of directly inputting the estimated HR to CNN, we input remote PPG (Photoplethysmogram) signals filtered by a dynamic HR filter, which can overcome two main problems: low temporal resolution of estimated HR; much noise exists in the estimated remote PPG signals. Evaluation results show that the dynamic HR filter works more effectively compared to the static one, which helps improve AUC (area under the curve) index of the classification to 0.70, as good as the performance (0.71) of HR from a wearable sensor.

1. はじめに

眠気推定や睡眠／覚醒判別など覚醒度低下の定量化は、疲労のリスク管理、運転中、勉強中、仕事時の状態モニタリングや、ヘルスケア等の応用のために広く研究されている。眠気推定手法としては、装着型センサを利用する手法やカメラを利用する手法があるが、中でも顔映像から抽出した顔の動きを用いる手法は、非接触型で使い勝手が良く、また高精度に眠気を推定できる[Oliveira 2018, Tsujikawa 2018]。しかし、その手法は顔を閉じている状況(閉眼時)では常に眠いと判定されてしまうという問題があり、実際に寝ているのか、目を閉じているだけなのか判別できない。

閉眼時でも睡眠／覚醒を判別できる手法として、心電図(Electrocardiogram: ECG)信号や光電式容積脈波記録法(Photoplethysmogram: PPG)信号を用いる手法がある[Aktaruzzaman 2015, Ye 2016, Scherz 2017, Malik 2018]。ECG信号や PPG 信号から、連続した心拍の間隔(Interbeat interval: IBI)の時系列信号を抽出し、IBI の逆数である心拍数(Heart rate: HR)や、IBI の時間変動である心拍変動(Heart rate variability: HRV)を分析して、睡眠／覚醒を判別する。心臓の動きは自律神経により制御されており、覚醒中は交感神経が、睡眠中は副交感神経が支配的になるため、覚醒中は睡眠中より心臓が活発に動く[Somers 1993]。そのため、睡眠／覚醒判別において、心臓の動きを表す HR や HRV は重要な生理学的指標とされている。

近年、HR や HRV を算出するための脈波を含む信号を取得する手法として、離れて配置された(リモート)センサを利用する手法の研究が盛んである。文献[Rahman 2015]は、ドップラーレーダで抽出した HR と呼吸数、体動を用いて、睡眠をセンシングできるシステムを提案している。そのシステムにより、高精度に睡眠／覚醒を判別できると報告されている。しかし、前述の顔映像から眠気を推定するシステムとの統合のためには、リモートセンサとしてカメラを利用することが望ましい。

顔映像から HR を推定する手法は、ますます盛んに研究されている[Wang 2018, Utkarsh 2018]。それは、顔の輝度値に基づく推定手法と顔の動きに基づく推定手法の二種に大別されるが、

前者の研究が盛んである。カメラで捉えた顔の輝度値から PPG 信号(リモート PPG 信号)を抽出した後、高速フーリエ変換(Fast Fourier Transform: FFT)などの周波数分析を用いて、HR を推定する。心臓は周期的に動いているため、パワーが大きい周波数成分を HR 成分とする。欠点は、周波数分析の窓幅に相当する区間の HR しか計算できず、一拍一拍の瞬間的な HR (Instantaneous HR: IHR)の変化を高精度に捉えることが困難な点である。それは、短い周波数分析窓幅では HR 情報が、頭の動きや光の変化等で生じる大きなノイズに埋もれるためである。IHR の変化は HRV の算出において重要であり、その HRV は睡眠／覚醒判別の重要な指標の一つである。したがって、リモートセンサとしてカメラを利用して、睡眠／覚醒を判別する際の課題は、リモート PPG 信号から IHR の情報、すなわち高時間解像度の HR 情報を高精度に抽出することである。

上記の課題に対して、高時間解像度の HR 情報を含むリモート PPG 信号による睡眠／覚醒判別を提案する。頭の動きや表情変化、光の変化により生じる大きなノイズで埋もれた、高時間解像度の HR 情報を抽出するために、低時間解像度の HR の分布情報から動的に設計したフィルタを用いてノイズを除去する。ノイズ除去後のリモート PPG 信号は、手動での特徴抽出を避けるために、畳み込みニューラルネットワーク(Convolutional neural network: CNN)に直接入力し、睡眠／覚醒を判別する。2章で提案法を詳細に説明し、3章で、睡眠／覚醒判別の評価実験において提案法の有効性を示す。

2. 睡眠／覚醒判別法

2.1 従来法

図 1 に文献[Malik 2018]で提案されている従来の睡眠／覚醒判別手法の構成を示す。装着型センサである ECG センサから ECG 信号を得て、連続した心拍の間隔 IBI を算出し、その逆数である IHR を以下の式で算出する。

$$IHR(i) = 60/IBI(i) \quad (1)$$

ただし、 $IHR(i)$ は i 番目の IHR で単位は bpm (beat per minute)、 $IBI(i)$ は i 番目の IBI で単位は秒である。 $IHR(i)$ は不等間隔でサンプリングした時系列信号のため、補間処理により等間隔でサンプリングして HR の時系列信号を得た後、一定の時間窓で区切った HR の時系列信号を CNN に入力する。HR 信号から睡

眠／覚醒を判別できるように学習した CNN により、新たに入力された HR 信号に対して睡眠／覚醒の判別結果を出力する。

文献 [Malik 2018] では、二つの公開データベース ① DREAMS Subject Database¹, ② UCSDSADB²を用いて、学習したモデルの判別精度を評価している。①に対する AUC (Area under the curve) が 0.81, ②に対する AUC が 0.72 で判別可能と報告されている。

上記の従来法における装着型センサから算出した HR (Wearable HR) を、カメラから算出した HR (Camera HR) に置き換えると、次の問題がある。主問題は、Camera HR の時間解像度が低いことである。Camera HR は、一定の推定精度を得るために、通常、IBI より長い時間窓で PPG 信号を周波数分析して算出される。もう一つの問題は、ノイズに起因する推定誤差である。次節で説明する提案法では、高い時間解像度の HR 情報を含む PPG 信号に混在したノイズを動的なフィルタで除去し、それを CNN に入力することで上記の二つの問題を解決する。

2.2 提案法

図 2 は、提案する睡眠／覚醒判別手法の構成を示す。それは、文献 [Utkarsh 2018] の Camera HR 推定法と文献 [Malik 2018] で適用された CNN を組み合わせた手法である。その Camera HR 推定法は他手法と比較して高精度に HR を推定できると報告されている。まず、提案法ではカメラで取得した顔映像から PPG 信号 (Camera PPG) を抽出する。次に、Camera PPG から Camera HR を推定し、Camera HR の値に基づいて動的に設計したフィルタ (Dynamic Camera-HR filter) で Camera PPG に含まれるノイズを除去する。ノイズ除去後の Camera PPG を図 3 に示す CNN に入力し、睡眠／覚醒を判別する。以下、Camera PPG に含まれるノイズの除去手順と文献 [Malik 2018] の CNN とのモデル構造の違いについて説明する。その他に関しては、文献 [Utkarsh 2018] と [Malik 2018] を参照して欲しい。

Dynamic Camera-HR filter は、Camera-PPG に含まれるノイズを除去することで、HR 情報を強調するバンドパスフィルタである。それは Camera HR を用いて設計される。一定時間窓内の HR が正規分布に従うと仮定し、その平均を μ 、標準偏差を σ とすると、 $\mu \pm 3\sigma$ の範囲に 99.7% の値が含まれる。窓番号 I の Dynamic Camera-HR filter のカットオフ周波数の下限 $L_D(I)$ と上限 $H_D(I)$ を $\mu(I)$ と $\sigma(I)$ を用いて次のように算出する。

$$L_D(I) = \frac{\mu(I) - 3\sigma(I)}{60}, H_D(I) = \frac{\mu(I) + 3\sigma(I)}{60} \quad (2)$$

窓番号 I 毎に、推定した HR に応じてバンドパスフィルタの下限と上限を動的に調整する。その下限と上限を用いて、2 次のパワースバンドパスフィルタを設計する。

Dynamic Camera-HR filter の効果を検証するため、次章の評価で用いる学習データセット (DREAMS) の HR 情報から設計した静的なバンドパスフィルタ (Static HR filter) と比較する。Static HR filter においては、下限 $L_S = 0.53\text{Hz}$ 、上限 $H_S = 1.75\text{Hz}$ である。それらの値は次式で算出した。

$$L_S = \min[L_{\text{wake}}, L_{\text{sleep}}], H_S = \max[H_{\text{wake}}, H_{\text{sleep}}] \quad (3)$$

$$L_{\text{wake}} = \frac{\mu_{\text{wake}} - 3\sigma_{\text{wake}}}{60}, L_{\text{sleep}} = \frac{\mu_{\text{sleep}} - 3\sigma_{\text{sleep}}}{60} \quad (4)$$

$$H_{\text{wake}} = \frac{\mu_{\text{wake}} + 3\sigma_{\text{wake}}}{60}, H_{\text{sleep}} = \frac{\mu_{\text{sleep}} + 3\sigma_{\text{sleep}}}{60} \quad (5)$$

ただし、 μ_{wake} と μ_{sleep} は、それぞれ学習データにおいて覚醒状態または睡眠状態とラベル付けされた時刻の HR の平均、同様に σ_{wake} と σ_{sleep} は標準偏差である。それらを用いて、覚醒状態における HR の分布の下限 L_{wake} と上限 H_{wake} 、睡眠状態にお

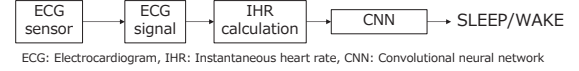


図1 従来の睡眠／覚醒判別手法の構成

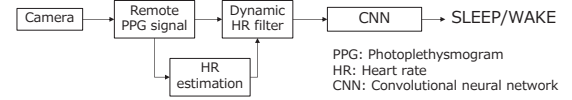


図2 提案する睡眠／覚醒判別手法の構成

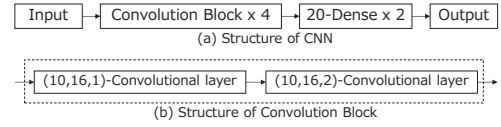


図3 (a) PPG 信号を入力する場合の 1 次元 CNN の構造。 m -Dense は m ノードの全結合層。 (b) 単一の畳込ブロックの構造。 (f, k, s) -convolution はフィルタ数 f 、カーネルサイズ k 、ストライド s の畳込層。

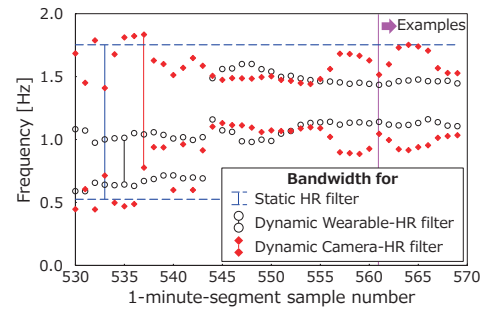


図4 Static HR filter と Dynamic HR filter の周波数帯域。

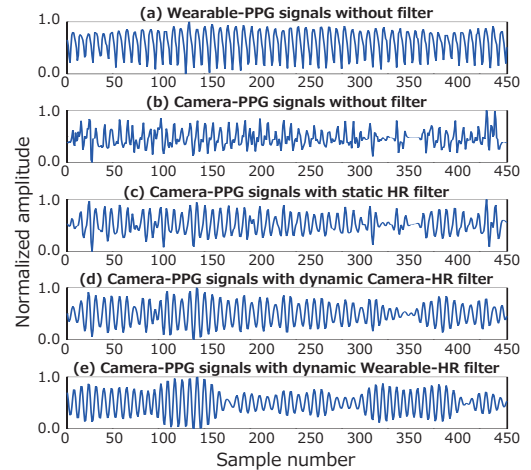


図5 Static HR filter と Dynamic HR filter の適用例。

ける下限 L_{sleep} と上限 H_{sleep} を算出し、最小値演算と最大値演算により、Static HR filter の下限 L_S と上限 H_S を算出した。

図 4 に Static HR filter と Dynamic HR filter の周波数帯域を例示する。比較のため、Dynamic HR filter は、Wearable HR から設計した Dynamic Wearable-HR filter と Dynamic Camera-HR filter の二種を示す。二種の Dynamic HR filter の周波数帯域は

¹ <http://www.tcts.fpms.ac.be/~devuyt/Databases/DatabaseSubjects>

² <https://physionet.org/pn3/ucddb>

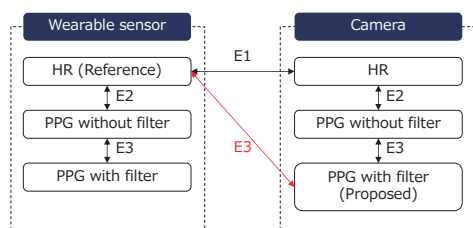


図6 三つの評価の関係。E1:評価1, E2:評価2, E3:評価3。

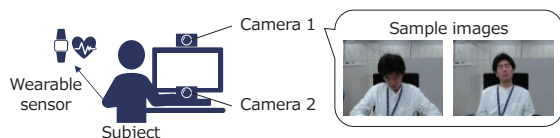


図7 検証用データの収集

表1 評価に用いたデータ

	Dataset	# of subjects	Recording length	Contents
Train	DREAMS Subjects Database+ DREAMS Patients Database	7 healthy subjects +12 patients	~8hours/person, 19182 1-min. segments	Wearable sensor: PPG signals, ECG signals.
Validate	Data collected by ourselves	25 healthy subjects	~30min/person, 1581 1-min. segments	Wearable sensor: PPG signals, IBI Camera: face videos

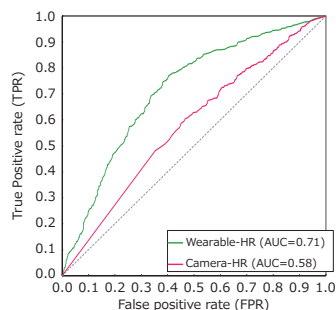


図8 ROCカーブ: Wearable HR vs Camera HR.

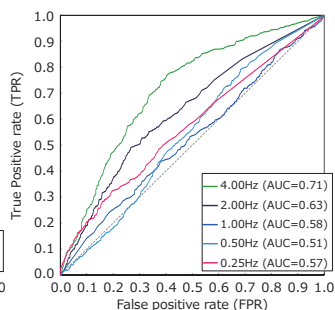


図9 ROCカーブ: Wearable HRにおける時間解像度 (0.25/0.5/1/2/4 Hz)の違い

Static HR filterと比較して狭い。人や状態(睡眠/覚醒)に応じてHRの分布が異なるため、その分布に応じて動的に算出することで周波数帯域が狭くなる。その周波数帯域の外側に存在するノイズを除去するため、帯域が狭いDynamic HR filterの方がノイズの除去量が多くなる。さらに、Dynamic Wearable-HR filterの方がDynamic Camera-HR filterより周波数帯域が狭い。Dynamic Camera-HR filterの周波数帯域はHRの推定誤差が含まれるためである。しかし、多くの場合にDynamic Camera-HR filterの周波数帯域はDynamic Wearable-HR filterの帯域を含んでおり、ノイズ除去量は劣るが、HR情報は殆ど欠落しない。

図5に、図4の窓番号 $I=561$ のCamera PPGに対して、(b)ノイズ除去フィルタ無しの場合、(c) Static HR filterを掛けた場合、

(d) Dynamic Camera-HR filterの場合、(e) Dynamic Wearable-HR filterの場合を例示する。なお、参考のため、(a)には装着型センサで得たPPG信号(Wearable PPG)を示す。(b)と(c)には多くのノイズが含まれるが、(d)と(e)はDynamic HR filterによりHR情報を残しつつノイズが除去されている。

ノイズ除去後のCamera PPGを図3に示すCNNに入力して睡眠/覚醒を判別する。図3に示すCNNの構造は文献[Malik 2018]のモデル構造に従う。そのモデル構造は睡眠/覚醒判別に必要なHRVを適応的に抽出できるものである。文献[Malik 2018]との違いは、入力サイズとカーネルサイズを2倍にした点のみである。それは、同等のHR情報を含むように、HRの時系列信号に対して、PPG信号のサンプリングレートを2倍に設定したためである。HRに関する情報量は同じであるため、PPG信号にノイズが含まなければ、文献[Malik 2018]と同等の精度で睡眠/覚醒を判別できると想定される。

3. 睡眠/覚醒判別評価

提案法の有効性を確認するため、CNNへの入力を変えて、睡眠/覚醒の判別結果に対するAUC指標を比較した。具体的には、以下三つの評価を行うことで、文献[Malik 2018]のWearable HRを入力した場合(従来法)に対する、Camera PPGを入力した場合(提案法)の有効性を確認した。各評価での確認事項は⇒印で示す通りである。

評価1: Wearable HR vs Camera HR

⇒HRの時間解像度の低さが問題であること

評価2: HR vs PPG信号

⇒PPG信号が高時間解像度のHR情報を含むこと

⇒PPG信号がノイズを含み、それが問題になること

評価3: PPG信号へのノイズ除去フィルタの適用

⇒Dynamic HR filterによりノイズ除去が可能なこと

⇒Camera PPGでも睡眠/覚醒判別が可能なこと

図6は評価1~3の関係図である。最終的に、図6の赤矢印の比較により、Camera PPG(提案法)でも、Wearable HR(従来法)と同等精度で睡眠/覚醒判別が可能なことを確認する。

3.1 評価に用いたデータ

表1に、評価に用いた学習データと検証データを示す。学習データは、公開データ(DREAMS Subject Database¹とDREAMS Patient Database²)から選択した計19名分である。それらにはWearable HRとWearable PPGの両方が含まれる。Wearable HR算出のためのECG信号とWearable PPGのオリジナルのサンプリングレートは200Hzであった。1分区切りの19182サンプルのうち覚醒(Positive)が4239サンプル、睡眠(Negative)が14942サンプルであった。

検証データは、我々が収集した。図7に示すように、装着型センサ(Empatica社E4)と上下の顔向き変動に対処するための二つのカメラ(640×480画素, 30fps)を用いた。収集は覚醒と睡眠の二つのステージを含む。覚醒ステージで、実験協力者は15分間スマートフォンを利用した。睡眠ステージの準備として単調な計算タスクを10分間行った。その後の40分間が睡眠ステージである。最後に、実験協力者は目覚めた時刻と、睡眠ステージ開始から何分ぐらいで眠りに落ちたかを質問紙に回答した。睡眠の平均時間は12.3分、標準偏差は6.6分であった。1分区切りの1581サンプルのうち覚醒(Positive)が898サンプル、睡眠(Negative)が683サンプルであった。

¹ <http://www.tcts.fpms.ac.be/~devuyt/Databases/DatabaseSubjects>

² <http://www.tcts.fpms.ac.be/~devuyt/Databases/DatabasePatients>

3.2 評価1: Wearable HR vs Camera HR

CNN への入力として, Wearable HR と Camera HR を比較した. 何れも 4Hz でサンプリングされた HR の時系列信号で, 窓幅は 1 分, つまり入力サイズは 240 で固定である. 睡眠/覚醒判別結果として, ROC(Receiver Operating Characteristic)カーブを図 8 に示す. Wearable HR の AUC は 0.71 で, Camera HR の 0.58 より高い. その理由として次の二つが考えられる. ① Camera HR は推定誤差が大きい. ② Camera HR の時間解像度が低い. まず理由①の確認のため, Wearable HR を正解として, Camera HR の推定誤差を算出した. RMSE は 11.1bpm, 相関は 0.49 であり, 比較的誤差が大きい.

理由②の確認のため, CNN の入力に含まれる HR 情報の時間解像度を低くして睡眠/覚醒判別の評価を行った. 具体的には, 4Hz の Wearable HR の算出において, 0.25/0.5/1/2 Hz でサンプリング後に, 4Hz でサンプリングした. 図 9 において, 最も高時間解像度である 4Hz の AUC が 0.71 と最も高く, 時間解像度が低いと AUC は低下する. 低時間解像度の HR 情報では, 睡眠/覚醒判別に有効な, HR と HRV の情報が欠落している.

3.3 評価2: HR vs PPG 信号

Camera HR では高時間解像度の HR 情報を得ることが困難なため, Camera PPG を CNN に入力し, Camera HR と比較した. Camera PPG のサンプリングレート 30Hz は顔映像のフレームレートと同じ 30fps だが, 比較する 4Hz の HR と情報量をそろえるため, 8Hz にダウンサンプリングした. CNN の入力サイズは 8Hz × 60 秒間で, 480 である. Wearable HR と Wearable PPG の比較も参考のために行った.

図 10(a)は装着型センサを, (b)はカメラを用いた場合の睡眠/覚醒判別結果である. (a)では Wearable HR の AUC(0.71)が Wearable PPG の AUC(0.62)より高いが, (b)では Camera PPG の AUC(0.66)が Camera HR の AUC(0.58)より高い. (b)の結果は高時間解像度の HR 情報が睡眠/覚醒判別において重要で, それが含まれる PPG 信号から抽出できることを示す. (a)の結果と, Camera PPG の AUC が Wearable AUC より低い原因としては, PPG 信号に含まれるノイズが問題と想定される. 次節で, ノイズ除去フィルタの効果を示す.

3.4 評価3: PPG 信号へのノイズ除去フィルタの適用

PPG 信号に対してノイズ除去フィルタを適用し, 睡眠/覚醒判別結果を比較した. 図 11 の(a)装着型センサと(b)カメラの両方で, ノイズ除去フィルタの適用による判別精度改善を確認できる. また, Dynamic HR filter は Static HR filter より判別精度の改善効果が高く, (a)と(b)の両方で, Wearable HR の AUC=0.71 とほぼ同じ精度を達成した.

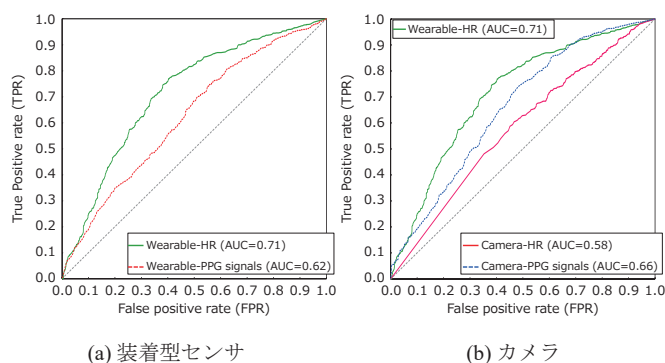


図10 ROC カーブ: HR vs PPG 信号

予想した通り, (b)において, Dynamic Wearble-HR filter の改善効果が Dynamic Camera-HR filter より少し高い. その理由は図 4 で示したように, Dynamic Wearble-HR filter の帯域幅が Dynamic Camera-HR filter より狭く, ノイズ除去効果が高いためである. Dynamic HR filter の改善効果が Static HR filter より高い理由も同じである.

4. まとめ

リモート PPG 信号(Camera PPG)による睡眠/覚醒判別方法を提案した. 提案法は, Camera HR と CNN を組み合わせた手法であり, まぶたの動きに基づく眠気推定方法と統合可能である. Camera HR を直接 CNN に入力する代わりに, Dynamic HR filter で処理した Camera PPG を CNN に入力する. それは, Camera HR の時間解像度の低さ, Camera PPG に含まれるノイズという二つの問題を解決する. 評価において, Dynamic HR filter が Static HR filter より効果的であり, 判別精度指標 AUC を 0.70 に改善できることを示した. それは, Wearable HR の AUC (0.71)と同等である.

参考文献

- [Oliveira 2018] L. Oliveira, et al.: Driver drowsiness detection : a comparison between intrusive and non-intrusive signal acquisition methods, 2018 7th European Workshop on Visual Information Processing (EUVIP), IEEE, 2018.
- [Tsujikawa 2018] M. Tsujikawa, et al.: Drowsiness Estimation from Low-Frame-Rate Facial Videos using Eyelid Variability Features, 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018.
- [Aktaruzzaman 2015] M. Aktaruzzaman et al.: The addition of entropy - based regularity parameters improves sleep stage classification based on heart rate variability, Medical & Biological Engineering & Computing, vol. 53, Issue 5, pp. 415-425, Springer, 2015.
- [Ye 2016] Yanqing Ye et al.: Automatic Sleep and Wake Classifier with Heart Rate and Pulse Oximetry : Derived Dynamic Time Warping Features and Logistic Model, 2016 Annu. IEEE Syst. Conf., IEEE, 2016.
- [Scherz 2017] W. D. Scherz et al.: Heart rate spectrum analysis for sleep quality detection, EURASIP Journal on Embedded Systems, EURASIP, 2017.
- [Malik 2018] J. Malik et al.: Sleep-wake classification via quantifying heart rate variability by convolutional neural network, Physiological Measurement, 29.8, 2018.
- [Somers 1993] V. K. Somers et al.: Sympathetic-Nerve Activity during Sleep in Normal Subjects, New England Journal of Medicine, 328.5, pp. 303-307, 1993.
- [Rahman 2015] T. Rahman et al.: DoppleSleep : A Contactless Unobtrusive Sleep Sensing System Using Short-Range Doppler Radar, Proceedings of the 2015 ACM international Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2015.
- [Wang 2018] C. Wang et al.: A Comparative Survey of Methods for Remote Heart Rate Detection From Frontal Face Videos, Frontiers in Bioengineering and Biotechnology, vol. 6, no. May, pp. 1-16, 2018.
- [Utkarsh 2018] S. Utkarsh et al.: Adaptive Heart rate estimation from face videos, The 8th Symposium on Biometrics, Recognition and Authentication, S2-9, IEICE, 2018.

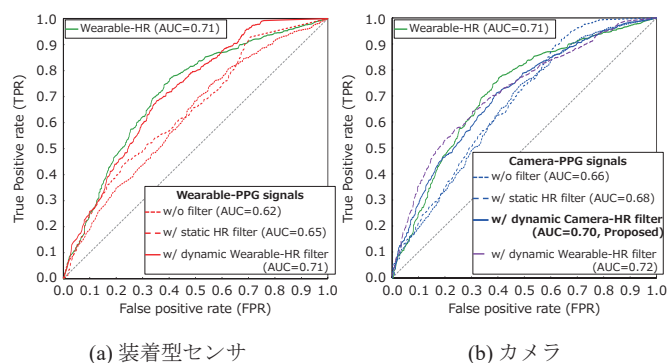


図11 ROC カーブ: PPG 信号に対するノイズ除去フィルタの効果

サンプル混合度を考慮した遺伝子発現量のがんバイオマーカー探索

Biomarker discovery from gene expression data of mixed tumor samples

村上 勝彦^{*1}
Katsuhiko Murakami

^{*1} 富士通研究所
FUJITSU LABORATORIES LTD.

To diagnose the state of tissues from cancer patients, utilization and discovery of biomarker with high discrimination accuracy is important. It is difficult to determine biomarkers that can cleanly separate cancer and normal tissues when normal tissues are mixed at the time of collecting a cancer sample. The purpose of the study is to provide method to discover biomarkers or characteristics of tissues with high accuracy for distinguishing between cancer tissues and non-cancer tissues. By utilizing a parameter λ , which indicates purity and status of each sample, highly accurate biomarker discovery became possible.

1. はじめに

細胞の状態を診断するため、判別精度が高いバイオマーカー探索方法は重要である。バイオマーカーとは、一般には病気の変化や治療に対する反応に相関し、指標となるものである。より具体的には、血液や尿などの体液や組織に含まれる、タンパク質や遺伝子などの生体内物質の存在、またはその量がバイオマーカーとなる。

本稿で扱う問題は、がん細胞と非がん細胞を判別するためのバイオマーカー（がん細胞で高発現する遺伝子セット）探索である。良い特徴量を選択できれば、新規の患者候補のサンプルが、癌であるかを精度良く判定できる。高精度の判定ができれば、治療の前後でサンプルを判定することで、治療がすすんでいるか、転移していないかを判定することにも用いられる。

バイオマーカー利用の際には、未知サンプルのバイオマーカー値を計測してサンプルの医学・生物学的な状態を推定する。探索時に構築した予測モデルを用いてバイオマーカー値を算定し、未知サンプルが、がんであるかの判定する。判定結果は、採取したサンプルが癌であるかの判定、癌種類や進行度の特定、臨床における治療効果の判定等に利用される。判別モデルは、回帰分析や LinearSVM [Abeel 2009] など通常の機械学習で学習、判別する。

実際のデータでは、がんサンプルを採取する際に正常細胞が混じったり、がん細胞内で特徴が大きくばらつく場合があり、がんと正常細胞をきれいに分離できるバイオマーカーを決定することが困難になることがある。そこでサンプル内の正常細胞とがん細胞の割合をパラメータとして統計モデルを作り ML(最尤法)で求める方法がある [Shen 2015]。しかしこれらは、「癌」と「正常」の2クラス分類で、学習や判定の際に、クラス内サンプルを等価に扱っている。「癌」も「正常」が均一であると仮定してこれら多様性が考慮されてない。また、癌には同種のがん(例えば「大腸がん」)でも様々なタイプや進行度がある。現在は同じ分類でもサブクラス分類が有効かもしれないが解析には考慮されていない。遺伝子発現量も安定な線形を仮定し、遺伝子の多様性を扱っていない。サンプルや遺伝子によって、外れ値のようにモデルにあわない部分がある。これらをノイズと扱うべき場合でも、等価に学習してしまうため、正しいパラメータを推定できなくなる。

本研究の目的は、提案した手法により混合度がどのくらい推定可能を評価することである。

2. データと方法

2.1 データセット

バイオマーカー探索の際には、癌サンプルと、(ネガティブ)コントロールとしてがんが発生している組織(例: 肺, 肝臓)の正常サンプルを採取する。組織サンプルごとに遺伝子 j の発現量を計測する。サンプル i に対して、測定値 j (遺伝子 j の発現量) x_{ij} とすると、データは行列形式に書ける。サンプルは大きく2つに分かれる。「がんサンプル」とラベルされたものと、「正常サンプル」である。しかしこれらはすべてが正しいものではない、という前提が本研究で重要なポイントである。解析の過程ではこれらのラベルは訂正される段階を経て、全体が解析される。

評価用のデータとして、NCBI GEO 遺伝子発現データベース (<http://www.ncbi.nlm.nih.gov/geo/>) から、アクセッション番号 GSE16515 として公開されている膵臓癌のデータを用いた。これには正常サンプル 16 個、がんサンプル 36 個、合計 52 サンプルのデータである。この実データをもとにして混合率の異なるデータセットを作成した。それらが以下に述べる手法で同定可能かどうかを検討した。

2.2 混合度の推定とバイオマーカー探索方法

「癌細胞として採取したサンプル内に混在する正常細胞の割合」を、隠れた変数として推定できるように、正常細胞の混在率(0 から 1) λ (正常度と呼ぶことにする)をモデルに組み込む。 λ は正常細胞とみなせる程度のパラメータとも解釈できる。

また、癌細胞はいくつかのステージで進行するものや分岐するなどの例が知られている。現在同じラベル(例「大腸がん」)といってもサブクラスで状態が異なる可能性がある。異なる状態には異なるバイオマーカーで状態を判定するのが適当である。 λ は細胞の正常度・非進行度とも解釈できる。

この状態をとらえる隠れ変数を導入して、データから細胞の状態を推定できるようにする。これは教師ラベル(癌, 正常)と高い相関があるが異なるものである。

正常度パラメータ λ が1に近いサンプル(第1グループ)は正常細胞とみなせる。一方、正常度パラメータ λ が0に近いサンプル(第2グループ)は癌細胞とみなせる。また λ が 0.5 に近いサンプル(第3グループ)は混合したサンプルか、ずれた正常細胞

連絡先: 村上勝彦, 富士通研究所, murakami.ktk@gmail.com

か、初期の癌細胞とみなせる。第3グループのサンプルは、判別の教師データとしてはのぞましくない。これを除いた(第1と第2グループからなる)学習データを作成し、再度判別モデルを構築すると、精度高い判別モデルが作成できる。同時に癌細胞の特徴をよく捉える特徴量のセット(バイオマーカー)が決定できる。

詳細には、以下の手順で行う(図1)。発現量など特徴量の変数を規格化する。SVDにより次元削減(約54,613次元から50次元程度へ)をする。5万という数はプローブの数であり、遺伝子数の数倍になっている。多くの場合はサンプル数が100個程度なので、それが次元の上限となる。

次にサンプルのクラスタリングを行う。例はガウス混合分布だが、k平均法など他の方法でもよい。

クラスター数 $n(>2)$ を決め、 n の数だけのガウス分布の重なりで表現する「混合ガウス分布(GMM)」のパラメータを求める。

ガウス分布のパラメータは、クラスター l に対して、平均 μ_l 、分散 σ_l 、中心位置ベクトル $\vec{c}_l$ ($l=1,2,\dots,n$) である。

各サンプル (x_i) の生成確率分布は、

$$P(\vec{x}_i) = \sum_{l=1}^n \pi_l C_l e^{-\frac{(\vec{x}_i - \vec{c}_l)^2}{\sigma_l}}$$

とする。ここで π_l はクラスター l の推定比率、 C_l は規格化定数。この全サンプルで和をとり、最大化するように求める。

常細胞サンプルだけから平均ベクトル $\vec{x}_v$ を求めておき、 $\vec{x}_v$ を含むクラスターと含まないクラスターに分ける。

$\vec{x}_v$ に近いクラスター中心を持つクラスターから順に、属するメンバーのクラスターを結合していき、教師データラベルが正常細胞のサンプルのうち、9割以上を含むクラスターが見つかったところで結合を停止する。その時点で結合されたクラスター群 n_n 個を「正常クラスター」、それ以外 $n_a (= n - n_n)$ 個を「異常クラスター」とみなす。よって各サンプル i 毎に

$$\lambda_i = \sum_{l \in N} \pi_{il}$$

が計算できる。正常度 λ_i が両極値を含まない領域(例えば0.2から0.8)のサンプルを、混在サンプルグループ M に属するとみなす。はじめのデータセットのサンプルのうちからサンプル群 M を除き、混合度の低い新規学習データ D_{new} を作成する。 D_{new} を用いてロジスティック回帰などのなんらかの判別器を作成する。結果、判別モデルと、バイオマーカーが得られる。

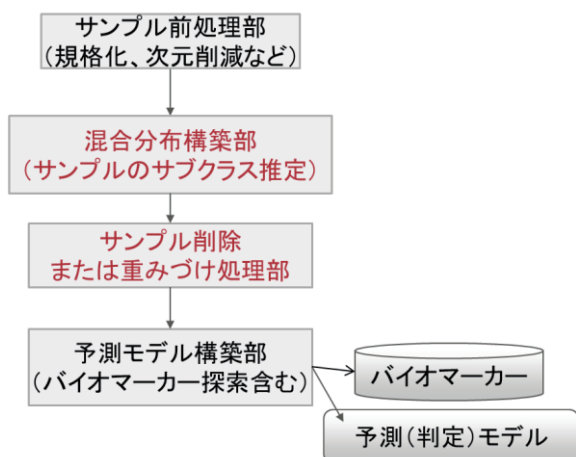


図1 解析方法のフローチャート。赤字が特徴的部分。

3. 結果

混合率の異なるデータセットごとに本手法を適用した結果、人工データにおいて混合サンプルを作成し、混合サンプルの70%を同定でき、それに基づくバイオマーカーを得た。

4. 議論

4.1 パラメータの意義

癌細胞と正常細胞の実サンプルには、それらの混在サンプルがあり得る。サンプルがマッピングされる特徴量空間で、2サンプル群の中間に位置する細胞は混在サンプルであることが想定される。これらのサンプルは、本来離れた位置にあるはずの2種のサンプルが混在した疑似サンプルであり、細胞としては実在しない。これらを、中間的な特徴量をもつサンプルとして学習するのは正しくない。これは生体サンプルでの特別な状況である。本方法ではパラメータ λ を導入して混在状況を把握できる。

混在サンプルが多い場合、通常のモデル構築では、精度の悪い結果を得るだけである。しかし本発明では、もし λ が中間値をとるようなサンプルが多い結果となれば、混合サンプルが多いという情報が得られる。

混合サンプルがあると、正しいモデル構築が非常に困難になる。ここでは混合サンプルをほぼ無視したきれいな学習データによるモデル構築ができる。その結果、高精度で安定的なバイオマーカー探索が行える。

5. おわりに

がん細胞を採取する際、目視で正常細胞と見分けることがしばしば困難であり、採取したサンプルに正常細胞が混在することがある。正常サンプルについても、その外れ値や異常性は考慮できなかった。混合サンプルがあるとそれらがノイズとなり特徴量を正しく学習できない。

本提案では、教師付学習の正解ラベル第1値(正常サンプル)の学習データ数の情報に基づいて、複数クラスを統合した統合クラスを生成することで、正常サンプルのラベル修正を可能とした。その結果、70%の混合サンプルを同定できた。結果、判別に有効なバイオマーカーを得た。今後、モデルを改良したり、さまざまなサンプルについて実験を重ねていく予定である。

参考文献

- [Abeel 2009] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," *Bioinformatics*, 2009.
- [Emmert 1996] M. R. Emmert-Buck, et al. "Laser Capture Microdissection", *Science*, 1996.
- [Zhao 2018] L. Zhao, V. H. F. Lee, M. K. Ng, H. Yan, and M. F. Bijlsma, "Molecular subtyping of cancer: current status and moving toward clinical applications," *Brief. Bioinform.*, 2018.
- [Ahn 2013] J. Ahn et al., "DeMix: Deconvolution for mixed cancer transcriptomes using raw measured data," *Bioinformatics*, 2013.
- [Shen 2016] Q. Shen, J. Hu, N. Jiang, X. Hu, Z. Luo, and H. Zhang, "contamDE: differential expression analysis of RNA-seq data for contaminated tumor samples," *Bioinformatics*, 2016.