Fri. Jun 8, 2018

Room P

Poster presentation | General Session | Interactive

[4Pin1] インタラクティブ(2)

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

[4Pin1-01] Generalization of CNN based deep learning method by featuring graph

○Kei Takahashi[1], Takumi Numajiri[2], Masaru Sogabe[2], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Shinji Yokogawa[1], Tohma Sogabe[1,2] （1. The University of Electro-Communications, 2. Grid Inc.）

9:00 AM - 10:40 AM

[4Pin1-02] Change detection across domains using GAN-based image translation

○kanji tanaka[1], yamaguchi kousuke[1], sugimoto takuma[1] （1. university of fukui）

9:00 AM - 10:40 AM

[4Pin1-03] Prototype and evaluation of active listening system using Recurrent Neural Network

○Kazuhiro Tamura[1] （1. CEC Customer Service Ltd.）

9:00 AM - 10:40 AM

[4Pin1-04] Segmentation in Stacks of Electron Microscopy Images Using Deep Learning

○Eichi Takaya[1], Yusuke Takeichi[2], Mamiko Ozaki[2], Satoshi Kurihara[1] （1. Graduate School of Science and Technology, Keio University, 2. Graduate School of Science, Kobe University）

9:00 AM - 10:40 AM

[4Pin1-05] Counting People via High Resolution Images using Deep Learning

○Yusuke Yamada[1], Makoto Kawano[1], Takuro Yonezawa[1], Jin Nakazawa[1] （1. Keio University）

9:00 AM - 10:40 AM

[4Pin1-06] Effect of Viewing Directions on Deep Reinforcement Learning in 3D Virtual Environment *Minecraft*

○Taiju Matsui[1], Satoshi Oyama[1], Masahito Kurihara[1] （1. Graduate School of Information Science and Technology, Hokkaido University）

9:00 AM - 10:40 AM

[4Pin1-07] A Decoding Method for Distorted Two-dimensional Barcodes Using Auxility Line Detection by Convolution Neural Network

○Kazuki Kuratsuwa[1], Makoto Kamizono[1], Hiroshi Kawasaki[2], Satoshi Ono[1] （1. Kagoshima University, 2. Kyushu University）

9:00 AM - 10:40 AM

[4Pin1-08] Domain Adaptation for Transferring Neural Question Answering from Simulation to Real

○Taiki Miyanishi[1], Motoaki Kawanabe[1,2] （1. Advanced Telecommunication Research Institute International, 2. RIEKN AIP）

9:00 AM - 10:40 AM

[4Pin1-09] Improving Neural Machine Translation by Incorporating Hierarchical Subword Features

○Makoto Morishita[1], Jun Suzuki[1], Masaaki Nagata[1] （1. NTT Communication Science Laboratories）

9:00 AM - 10:40 AM

[4Pin1-10] Conditional Cycle GAN based Transformation of Food Image Category

○Daichi Horita[1], Shu Naritomi[1], Ryosuke Tanno[1], Wataru Shimoda[1], Keiji Yanai[1] （1. The University of Electro-Communications, Tokyo）

9:00 AM - 10:40 AM

[4Pin1-11] Divide and Generate: Neural Generation of Complex Sentences

○Tomoya Ogata[1], Mamoru Komachi[1], Tomoya Takatani[2] （1. Tokyo Metropolitan University, 2. TOYOTA MOTOR CORPORATION）

9:00 AM - 10:40 AM

[4Pin1-12] Smart Grid Optimization by Deep Reinforcement Learning over Discrete and Continuous Action Space

○Tomah Sogabe[1], Dinesh Bahadu Malla[1], shota Takayama[1], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Thakur Praveen Singh[2], Masaru Sogabe[2] （1. The University of Electro-communications, 2. Grid Inc.）

9:00 AM - 10:40 AM

[4Pin1-13] Development of quantum autoencoder

○Yuta Kikawada[1], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Shinji Yokogawa[1], Tohma Sogabe[1] （1. The University of Electro-communications）

9:00 AM - 10:40 AM

[4Pin1-14] Deep CTR Prediction in Facebook Ads

○Yuki Iwazaki[1] （1. CyberAgent, Inc.）

9:00 AM - 10:40 AM

[4Pin1-15] Determination of Column Concepts in Tables using Multi-label Classification and Knowledge

©The Japanese Society for Artificial Intelligence

Base Embeddings

○Kunihiro Takeoka[1], Masafumi Oyamada[2], Shinji Nakadai[2], Takeshi Okadome[1] (1. Kwansei Gakuin University, 2. NEC Corporation)

9:00 AM - 10:40 AM

[4Pin1-16] Creators and Artificial Intelligence: Enabling Collaboration with Creative Processes and Meta-data for Four-scene Comic Story Dataset

○Miki Ueno[1] (1. Toyohashi University of Technology)

9:00 AM - 10:40 AM

[4Pin1-17] Movement Modeling for History Documents with Hidden Markov Models

○Yota Mizutani[1], Yoshimasa Tsuruoka[1] (1. Graduate School of Engineering, The University of Tokyo)

9:00 AM - 10:40 AM

[4Pin1-18] An entity linking method using multiple Doc2Vec

○Makoto Tsutsumi[1], Koji Murakami[2], Takashi Umeda[2] (1. Rakuten, Inc., 2. Rakuten Institute of Technology)

9:00 AM - 10:40 AM

[4Pin1-19] Emotional Style Transfer using CycleGAN

○Kazuhiro Ota[1] (1. CyberAgent, Inc.)

9:00 AM - 10:40 AM

[4Pin1-20] Persona Estimation using SNS based on Naive Bayes Method

○Hiroki Yamaoka[1], Mutsuo Sano[1] (1. Osaka Institute of Technology)

9:00 AM - 10:40 AM

[4Pin1-21] A Note on Similar Researcher Retrieval Considering Temporal Changes of Research Content and Affiliations

○Hiroyuki Nishizawa[1], Marie Katsurai[1], Ikki Ohmukai[2], Hideaki Takeda[2] (1. Doshisha University, 2. National Institute of Informatics)

9:00 AM - 10:40 AM

[4Pin1-22] Development of Calculus Solver software which provides solving hints and answer steps

○Otsuka Motohiro[1], Taira Hirotoshi[1], Shinkai Hisaaki[1] (1. Faculty of Information Science &Technology, Osaka Institute of Technology)

9:00 AM - 10:40 AM

[4Pin1-23] Multi-source neural grammatical error correction

○Cao Guolin[1], Hiroya Takamura[2,3], Manabu Okumura[2,4] (1. School of Engineering, Tokyo Institute of Technology, 2. Institute of Innovative Research, Tokyo Institute of Technology, 3. AIST, 4. RIKEN AIP)

9:00 AM - 10:40 AM

[4Pin1-24] Automatic Generation of Bilingual Corpus by Earth Mover's Distance and Senetence Vectors Using Word Embedding

○Ryo Tanoue[1], Hiroshi Echizen'ya[1], Kenji Araki[2] (1. Hokkai-Gakuen University, 2. Hokkaido University)

9:00 AM - 10:40 AM

[4Pin1-25] Neural Machine Translation with Kanji Decomposition

○Vishu Gupta[1], Akihiro Nakamura[1], Haruki Fukuda[1], Takashi Tsunakawa[1], Yoshinobu Kano[1], Masafumi Nishida[1], Masafumi Nishimura[1] (1. Shizuoka University)

9:00 AM - 10:40 AM

[4Pin1-26] Support for Searching for Articles to be Verified for Fact Checking

○Kaori Uchiyama[1], Kaito Suzuki[1], Tsubasa Tagami[1], Kazuaki Hanawa[1], Kentaro Inui[1,2], Atsushi Komiya[3], Atsuo Fujimura[3], Akinori Machino, Hitofumi Yanai[4], Ryo Yamashita[4] (1. Tohoku University, 2. RIKEN Center for Advanced Intelligence Project, 3. SmartNews, Inc., 4. Watchdog for Accuracy inNews-reporting, Japan)

9:00 AM - 10:40 AM

[4Pin1-27] Mining Significant Diamond Episodes from Event Sequences

○Yota Tani[1], Isamu Furuya[1], Kouichi Hirata[2], Hiroki Arimura[1] (1. Graduate School of Information Science and Technology, Hokkaido University, 2. Department of Artificial Intelligence, Kyushu Institute of Technology)

9:00 AM - 10:40 AM

[4Pin1-28] Public Release of a Web-based System for Visualizing Prefectural Assembly Minutes in Japan

○Hokuto Ototake[1], Keiichi Takamaru[2], Yuzu Uchida[3], Yasutomo Kimura[4] (1. Fukuoka University, 2. Utsunomiya Kyowa University, 3. Hokkai-gakuen University, 4. Otaru University of Commerce / AIP)

9:00 AM - 10:40 AM

[4Pin1-29] Utterance Intention Understanding for News

©The Japanese Society for Artificial Intelligence

Articles Transfer by Conversation

○Hiroaki Takatsu[1], Katsuya Yokoyama[1], Hiroshi Honda[2], Shinya Fujie[1,3], Yoshihiko Hayashi[1], Tetsunori Kobayashi[1] （1. Waseda University, 2. Honda R&D Co.,Ltd., 3. Chiba Institute of Technology）

9:00 AM - 10:40 AM

[4Pin1-30] The Estimation of Argument Structure using Deep Learning

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

9:00 AM - 10:40 AM

[4Pin1-31] Analogy comprehension between psychological experiments and word embedding models

○Shin Asakawa[1] （1. Tokyo Women's Cristian university）

9:00 AM - 10:40 AM

[4Pin1-32] Associating semantic structure in metaphor comprehension

○Miho Fuyama[1], Hayato Saigo[2] （1. Tamagawa university, 2. Nagahama Institute of Bio-Science and Technology）

9:00 AM - 10:40 AM

[4Pin1-33] Towards Interpretation as Natural Logic Abduction

○Naoya Inoue[1], Pontus Stenetorp[2], Sebastian Riedel[2], Kentaro Inui[1,3] （1. Tohoku University, 2. University College London, 3. RIKEN Center for Advanced Intelligence Project）

9:00 AM - 10:40 AM

[4Pin1-34] Innovative approach for prediction of local weather using machine learning

○Takao Yoshikane[1], Kei Yoshimura[1] （1. The University of Tokyo）

9:00 AM - 10:40 AM

[4Pin1-35] Construction of ground motion prediction equation using random forest

○Hisahiko Kubo[1], Takashi Kunugi[1], Shingo Suzuki[1], Wataru Suzuki[1], Shin Aoi[1] （1. National Research Institute for Earth Science and Disaster Resilience ）

9:00 AM - 10:40 AM

[4Pin1-36] Development of the motion direction classifiers for bimanual movement using electroencephalogram

○Masaki Kato[1], Sotaro Shimada[1] （1. Meiji University）

9:00 AM - 10:40 AM

[4Pin1-37] A computational system estimating perception evoked by arbitrary visual inputs on the basis of modeling the perceptual representation in the brain

○Satoshi Nishida[1,2], Shinji Nishimoto[1,2] （1. National Institute of Information and Communications Technology, 2. Osaka University）

9:00 AM - 10:40 AM

[4Pin1-38] Metal-Oxide-Semiconductor Gas Sensor Array Toward Artificial Olfaction

○Yuta Saito[1], Takamune Yokoyama[1], Takahisa Tanaka[2], Ken Uchida[2] （1. Faculty of Science and Technology, Keio University, 2. Department of Electronics and Electrical Engineering, Keio University）

9:00 AM - 10:40 AM

[4Pin1-39] Estimation model of bustle of shopping streets by integration of IoT sensor data

○Aoi Yoshino[1], Yuya Ieiri[1], Reiko Hishiyama[1] （1. Waseda University）

9:00 AM - 10:40 AM

[4Pin1-40] Examination of effectiveness of phased evacuation strategies in large-scale exhibition hall facility

○Noriaki Nishikawa[1], Yuichi Hirokawa[1], Takeshi Yamada[2], Junji Innami[2], Toshiyuki Asano[1] （1. Japan Agency for Marine-Earth Science and Technology, 2. Vector Research Institute, Inc.）

9:00 AM - 10:40 AM

[4Pin1-41] Why does the station get crowded irregularly in the near future?

○Tatsuo Yamashita[1], Kota Tsubouchi[1], Mikiya Maruyama[1], Yuuki Yamaura[1], Koichiro Okada[1] （1. Yahoo Japan Corporation）

9:00 AM - 10:40 AM

[4Pin1-42] Proposal of Lifestyle Analysis from Household Electricity Consumption Data

○Yu Shirai[1], Shunichi Hattori[2], Yasuhumi Takama[1] （1. Graduate School of System Design, Tokyo Metropolitan University, 2. Central Research Institute of Electric Power Industry）

9:00 AM - 10:40 AM

[4Pin1-43] A pilot study: Application of machine learning to 24 hour period data – three case studies

©The Japanese Society for Artificial Intelligence

○Isao Goto[1] (1. Sumitomo Electric Industries, LTD)

9:00 AM - 10:40 AM

[4Pin1-44] Introduction of idea generation support system "Hirameki engine"

○Masaki Ueyama[1], Keita Kobayashi[2], Taikai Ogawa[2], Yusuke Okutani[3], Ryosuke Sakuma[3] (1. Honda R&D Co.,Ltd. , 2. NOMURA Co.,Ltd., 3. KAYAC Inc.)

9:00 AM - 10:40 AM

[4Pin1-45] Toward Personal Memory Recall Support

○Harumi Murakami[1], Ryutaro Murakami[1], Yuta Ueminami[1] (1. Osaka City University)

9:00 AM - 10:40 AM

[4Pin1-46] Toward Utilization of Use Log of Employee Cafeteria

○Tatsuo Yamashita[1], Naomi Sasaya[1], Nobuhiro Kaji[1], Miho Ishige[1], Mizuki Numata[1], Takahiro Shima[1], Nobuyuki Shimizu[1] (1. Yahoo Japan Corporation)

9:00 AM - 10:40 AM

[4Pin1-47] Bit-pattern kernel filtering for duplication removing in cell annotation.

○Toru Nagasaka[1] (1. AI &Eye Co., Ltd.)

9:00 AM - 10:40 AM

[4Pin1-48] Construction of dataset for summarization based on the Twitter URL Paraphrase Corpus

○Koichi Nagatsuka[1] (1. Soka University)

9:00 AM - 10:40 AM

[4Pin1-49] Proposal of GUI usability evaluation method of smartphone application

○Takayuki Tanaka[1], Satoru Inoue[1] (1. Saitama Institude of Technology)

9:00 AM - 10:40 AM

[4Pin1-50] Generating Spatial Semantic Descriptions of Parking Instructions for Self-driving Cars

○Akari Inago[1], Hiroshi Tsukahara[2], Ichiro Kobayashi[1] (1. Ochanomizu University, 2. Denso IT Laboratory)

9:00 AM - 10:40 AM

[4Pin1-51] A Preliminary Study on Coevolutionary Simultaneous Design of Digital Watermark for Two-dimensional Codes and Coded Aperture

○Shingo Takeshita[1], Hiroki Hamasaki[1], Takeru Maehara[1], Kentaro Nakai[1], Toshiki Sonoda[2], Michihiro Mikamo[1], Hiroshi Kawasaki[2], Hajime Nagahara[3], Satoshi Ono[1] (1. Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, 2. Department of Advanced Information Technology, Graduate School of Information Science and Electrical Engineering, Kyushu University, 3. Institute for Datability Science, Osaka University)

9:00 AM - 10:40 AM

[4Pin1-52] Dataset Construction Method for Word Reading Disambiguation

○Koki Nishiyama[1], Kazuhide Yamamoto[1], Hideharu Nakajima[2] (1. Nagaoka University of Technology, 2. NTT Media Intelligence Laboratories)

9:00 AM - 10:40 AM

[4Pin1-53] An Activity Description Method by Multiple Events of Multiple Objects

○Takayuki Semitsu[1], Takashi Nammoto[1], Toshisada Mariyama[1] (1. Mitsubishi Electric Corporation)

9:00 AM - 10:40 AM

[4Pin1-54] Towards Annotating Problem Solutions and Evaluation Expressions in Information Science Papers

○Hono Shirai[1], Naoya Inoue[1], Kentaro Inui[1,2] (1. Tohoku University, 2. RIKEN Center for Advanced Intelligence Project (AIP))

9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

Poster presentation | General Session | Interactive

[4Pin1] インタラクティブ(2)

Fri. Jun 8, 2018 9:00 AM - 10:40 AM Room P (4F Emerald Lobby)

[4Pin1-01] Generalization of CNN based deep learning method by featuring graph

○Kei Takahashi[1], Takumi Numajiri[2], Masaru Sogabe[2], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Shinji Yokogawa[1], Tohma Sogabe[1,2] （1. The University of Electro-Communications, 2. Grid Inc.）

9:00 AM - 10:40 AM

[4Pin1-02] Change detection across domains using GAN-based image translation

○kanji tanaka[1], yamaguchi kousuke[1], sugimoto takuma[1] （1. university of fukui）

9:00 AM - 10:40 AM

[4Pin1-03] Prototype and evaluation of active listening system using Recurrent Neural Network

○Kazuhiro Tamura[1] （1. CEC Customer Service Ltd.）

9:00 AM - 10:40 AM

[4Pin1-04] Segmentation in Stacks of Electron Microscopy Images Using Deep Learning

○Eichi Takaya[1], Yusuke Takeichi[2], Mamiko Ozaki[2], Satoshi Kurihara[1] （1. Graduate School of Science and Technology, Keio University, 2. Graduate School of Science, Kobe University）

9:00 AM - 10:40 AM

[4Pin1-05] Counting People via High Resolution Images using Deep Learning

○Yusuke Yamada[1], Makoto Kawano[1], Takuro Yonezawa[1], Jin Nakazawa[1] （1. Keio University）

9:00 AM - 10:40 AM

[4Pin1-06] Effect of Viewing Directions on Deep Reinforcement Learning in 3D Virtual Environment *Minecraft*

○Taiju Matsui[1], Satoshi Oyama[1], Masahito Kurihara[1] （1. Graduate School of Information Science and Technology, Hokkaido University）

9:00 AM - 10:40 AM

[4Pin1-07] A Decoding Method for Distorted Two-dimensional Barcodes Using Auxility Line Detection by Convolution Neural Network

○Kazuki Kuratsuwa[1], Makoto Kamizono[1], Hiroshi Kawasaki[2], Satoshi Ono[1] （1. Kagoshima University, 2. Kyushu University）

9:00 AM - 10:40 AM

[4Pin1-08] Domain Adaptation for Transferring Neural Question Answering from Simulation to Real

○Taiki Miyanishi[1], Motoaki Kawanabe[1,2] （1. Advanced Telecommunication Research Institute International, 2. RIEKN AIP）

9:00 AM - 10:40 AM

[4Pin1-09] Improving Neural Machine Translation by Incorporating Hierarchical Subword Features

○Makoto Morishita[1], Jun Suzuki[1], Masaaki Nagata[1] （1. NTT Communication Science Laboratories）

9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

[4Pin1-10] Conditional Cycle GAN based Transformation of Food Image Category

○Daichi Horita[1], Shu Naritomi[1], Ryosuke Tanno[1], Wataru Shimoda[1], Keiji Yanai[1] （1. The University of Electro-Communications, Tokyo）

9:00 AM - 10:40 AM

[4Pin1-11] Divide and Generate: Neural Generation of Complex Sentences

○Tomoya Ogata[1], Mamoru Komachi[1], Tomoya Takatani[2] （1. Tokyo Metropolitan University, 2. TOYOTA MOTOR CORPORATION）

9:00 AM - 10:40 AM

[4Pin1-12] Smart Grid Optimization by Deep Reinforcement Learning over Discrete and Continuous Action Space

○Tomah Sogabe[1], Dinesh Bahadu Malla[1], shota Takayama[1], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Thakur Praveen Singh[2], Masaru Sogabe[2] （1. The University of Electro-communications, 2. Grid Inc.）

9:00 AM - 10:40 AM

[4Pin1-13] Development of quantum autoencoder

○Yuta Kikawada[1], Katsuyoshi Sakamoto[1], Kouichi Yamaguchi[1], Shinji Yokogawa[1], Tohma Sogabe[1] （1. The University of Electro-communications）

9:00 AM - 10:40 AM

[4Pin1-14] Deep CTR Prediction in Facebook Ads

○Yuki Iwazaki[1] （1. CyberAgent, Inc.）

9:00 AM - 10:40 AM

[4Pin1-15] Determination of Column Concepts in Tables using Multi-label Classification and Knowledge Base Embeddings

○Kunihiro Takeoka[1], Masafumi Oyamada[2], Shinji Nakadai[2], Takeshi Okadome[1] （1. Kwansei Gakuin University, 2. NEC Corporation）

9:00 AM - 10:40 AM

[4Pin1-16] Creators and Artificial Intelligence: Enabling Collaboration with Creative Processes and Meta-data for Four-scene Comic Story Dataset

○Miki Ueno[1] （1. Toyohashi University of Technology）

9:00 AM - 10:40 AM

[4Pin1-17] Movement Modeling for History Documents with Hidden Markov Models

○Yota Mizutani[1], Yoshimasa Tsuruoka[1] （1. Graduate School of Engineering, The University of Tokyo）

9:00 AM - 10:40 AM

[4Pin1-18] An entity linking method using multiple Doc2Vec

○Makoto Tsutsumi[1], Koji Murakami[2], Takashi Umeda[2] （1. Rakuten, Inc., 2. Rakuten Institute of Technology）

9:00 AM - 10:40 AM

[4Pin1-19] Emotional Style Transfer using CycleGAN

○Kazuhiro Ota[1] （1. CyberAgent, Inc.）

9:00 AM - 10:40 AM

[4Pin1-20] Persona Estimation using SNS based on Naive Bayes Method

○Hiroki Yamaoka[1], Mutsuo Sano[1] （1. Osaka Institute of Technology）

9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

[4Pin1-21] A Note on Similar Researcher Retrieval Considering Temporal Changes of Research Content and Affiliations

○Hiroyuki Nishizawa[1], Marie Katsurai[1], Ikki Ohmukai[2], Hideaki Takeda[2] （1. Doshisha University, 2. National Institute of Informatics）

9:00 AM - 10:40 AM

[4Pin1-22] Development of Calculus Solver software which provides solving hints and answer steps

○Otsuka Motohiro[1], Taira Hirotoshi[1], Shinkai Hisaaki[1] （1. Faculty of Information Science &Technology, Osaka Institute of Technology）

9:00 AM - 10:40 AM

[4Pin1-23] Multi-source neural grammatical error correction

○Cao Guolin[1], Hiroya Takamura[2,3], Manabu Okumura[2,4] （1. School of Engineering, Tokyo Institute of Technology, 2. Institute of Innovative Research, Tokyo Institute of Technology, 3. AIST, 4. RIKEN AIP）

9:00 AM - 10:40 AM

[4Pin1-24] Automatic Generation of Bilingual Corpus by Earth Mover's Distance and Senetence Vectors Using Word Embedding

○Ryo Tanoue[1], Hiroshi Echizen'ya[1], Kenji Araki[2] （1. Hokkai-Gakuen University, 2. Hokkaido University）

9:00 AM - 10:40 AM

[4Pin1-25] Neural Machine Translation with Kanji Decomposition

○Vishu Gupta[1], Akihiro Nakamura[1], Haruki Fukuda[1], Takashi Tsunakawa[1], Yoshinobu Kano[1], Masafumi Nishida[1], Masafumi Nishimura[1] （1. Shizuoka University）

9:00 AM - 10:40 AM

[4Pin1-26] Support for Searching for Articles to be Verified for Fact Checking

○Kaori Uchiyama[1], Kaito Suzuki[1], Tsubasa Tagami[1], Kazuaki Hanawa[1], Kentaro Inui[1,2], Atsushi Komiya[3], Atsuo Fujimura[3], Akinori Machino, Hitofumi Yanai[4], Ryo Yamashita[4] （1. Tohoku University, 2. RIKEN Center for Advanced Intelligence Project, 3. SmartNews, Inc., 4. Watchdog for Accuracy inNews-reporting, Japan）

9:00 AM - 10:40 AM

[4Pin1-27] Mining Significant Diamond Episodes from Event Sequences

○Yota Tani[1], Isamu Furuya[1], Kouichi Hirata[2], Hiroki Arimura[1] （1. Graduate School of Information Science and Technology, Hokkaido University, 2. Department of Artificial Intelligence, Kyushu Institute of Technology）

9:00 AM - 10:40 AM

[4Pin1-28] Public Release of a Web-based System for Visualizing Prefectural Assembly Minutes in Japan

○Hokuto Ototake[1], Keiichi Takamaru[2], Yuzu Uchida[3], Yasutomo Kimura[4] （1. Fukuoka University, 2. Utsunomiya Kyowa University, 3. Hokkai-gakuen University, 4. Otaru University of Commerce / AIP）

9:00 AM - 10:40 AM

[4Pin1-29] Utterance Intention Understanding for News Articles Transfer by Conversation

○Hiroaki Takatsu[1], Katsuya Yokoyama[1], Hiroshi Honda[2], Shinya Fujie[1,3], Yoshihiko Hayashi[1], Tetsunori Kobayashi[1] （1. Waseda University, 2. Honda R&D Co.,Ltd., 3. Chiba Institute of

©The Japanese Society for Artificial Intelligence

Technology）
9:00 AM - 10:40 AM

[4Pin1-30] The Estimation of Argument Structure using Deep Learning
○Ryo Kinoshita[1], Masaki Uto[1], Maomi Ueno[1] （1. The University of Electro-Communications）
9:00 AM - 10:40 AM

[4Pin1-31] Analogy comprehension between psychological experiments and word embedding models
○Shin Asakawa[1] （1. Tokyo Women's Cristian university）
9:00 AM - 10:40 AM

[4Pin1-32] Associating semantic structure in metaphor comprehension
○Miho Fuyama[1], Hayato Saigo[2] （1. Tamagawa university, 2. Nagahama Institute of Bio-Science and Technology）
9:00 AM - 10:40 AM

[4Pin1-33] Towards Interpretation as Natural Logic Abduction
○Naoya Inoue[1], Pontus Stenetorp[2], Sebastian Riedel[2], Kentaro Inui[1,3] （1. Tohoku University, 2. University College London, 3. RIKEN Center for Advanced Intelligence Project）
9:00 AM - 10:40 AM

[4Pin1-34] Innovative approach for prediction of local weather using machine learning
○Takao Yoshikane[1], Kei Yoshimura[1] （1. The University of Tokyo）
9:00 AM - 10:40 AM

[4Pin1-35] Construction of ground motion prediction equation using random forest
○Hisahiko Kubo[1], Takashi Kunugi[1], Shingo Suzuki[1], Wataru Suzuki[1], Shin Aoi[1] （1. National Research Institute for Earth Science and Disaster Resilience ）
9:00 AM - 10:40 AM

[4Pin1-36] Development of the motion direction classifiers for bimanual movement using electroencephalogram
○Masaki Kato[1], Sotaro Shimada[1] （1. Meiji University）
9:00 AM - 10:40 AM

[4Pin1-37] A computational system estimating perception evoked by arbitrary visual inputs on the basis of modeling the perceptual representation in the brain
○Satoshi Nishida[1,2], Shinji Nishimoto[1,2] （1. National Institute of Information and Communications Technology, 2. Osaka University）
9:00 AM - 10:40 AM

[4Pin1-38] Metal-Oxide-Semiconductor Gas Sensor Array Toward Artificial Olfaction
○Yuta Saito[1], Takamune Yokoyama[1], Takahisa Tanaka[2], Ken Uchida[2] （1. Faculty of Science and Technology, Keio University, 2. Department of Electronics and Electrical Engineering, Keio University）
9:00 AM - 10:40 AM

[4Pin1-39] Estimation model of bustle of shopping streets by integration of IoT sensor data
○Aoi Yoshino[1], Yuya Ieiri[1], Reiko Hishiyama[1] （1. Waseda University）
9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

[4Pin1-40] Examination of effectiveness of phased evacuation strategies in large-scale exhibition hall facility

○Noriaki Nishikawa[1], Yuichi Hirokawa[1], Takeshi Yamada[2], Junji Innami[2], Toshiyuki Asano[1] （1. Japan Agency for Marine-Earth Science and Technology, 2. Vector Research Institute, Inc.）

9:00 AM - 10:40 AM

[4Pin1-41] Why does the station get crowded irregularly in the near future?

○Tatsuo Yamashita[1], Kota Tsubouchi[1], Mikiya Maruyama[1], Yuuki Yamaura[1], Koichiro Okada[1] （1. Yahoo Japan Corporation）

9:00 AM - 10:40 AM

[4Pin1-42] Proposal of Lifestyle Analysis from Household Electricity Consumption Data

○Yu Shirai[1], Shunichi Hattori[2], Yasuhumi Takama[1] （1. Graduate School of System Design, Tokyo Metropolitan University, 2. Central Research Institute of Electric Power Industry）

9:00 AM - 10:40 AM

[4Pin1-43] A pilot study: Application of machine learning to 24 hour period data – three case studies

○Isao Goto[1] （1. Sumitomo Electric Industries, LTD）

9:00 AM - 10:40 AM

[4Pin1-44] Introduction of idea generation support system "Hirameki engine"

○Masaki Ueyama[1], Keita Kobayashi[2], Taikai Ogawa[2], Yusuke Okutani[3], Ryosuke Sakuma[3] （1. Honda R&D Co.,Ltd. , 2. NOMURA Co.,Ltd., 3. KAYAC Inc.）

9:00 AM - 10:40 AM

[4Pin1-45] Toward Personal Memory Recall Support

○Harumi Murakami[1], Ryutaro Murakami[1], Yuta Ueminami[1] （1. Osaka City University）

9:00 AM - 10:40 AM

[4Pin1-46] Toward Utilization of Use Log of Employee Cafeteria

○Tatsuo Yamashita[1], Naomi Sasaya[1], Nobuhiro Kaji[1], Miho Ishige[1], Mizuki Numata[1], Takahiro Shima[1], Nobuyuki Shimizu[1] （1. Yahoo Japan Corporation）

9:00 AM - 10:40 AM

[4Pin1-47] Bit-pattern kernel filtering for duplication removing in cell annotation.

○Toru Nagasaka[1] （1. AI &Eye Co., Ltd.）

9:00 AM - 10:40 AM

[4Pin1-48] Construction of dataset for summarization based on the Twitter URL Paraphrase Corpus

○Koichi Nagatsuka[1] （1. Soka University）

9:00 AM - 10:40 AM

[4Pin1-49] Proposal of GUI usability evaluation method of smartphone application

○Takayuki Tanaka[1], Satoru Inoue[1] （1. Saitama Institude of Technology）

9:00 AM - 10:40 AM

[4Pin1-50] Generating Spatial Semantic Descriptions of Parking Instructions for Self-driving Cars

○Akari Inago[1], Hiroshi Tsukahara[2], Ichiro Kobayashi[1] （1. Ochanomizu University, 2. Denso IT Laboratory）

9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

[4Pin1-51] A Preliminary Study on Coevolutionary Simultaneous Design of Digital Watermark for Two-dimensional Codes and Coded Aperture

○Shingo Takeshita[1], Hiroki Hamasaki[1], Takeru Maehara[1], Kentaro Nakai[1], Toshiki Sonoda[2], Michihiro Mikamo[1], Hiroshi Kawasaki[2], Hajime Nagahara[3], Satoshi Ono[1] （1. Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, 2. Department of Advanced Information Technology, Graduate School of Information Science and Electrical Engineering, Kyushu University, 3. Institute for Datability Science, Osaka University）

9:00 AM - 10:40 AM

[4Pin1-52] Dataset Construction Method for Word Reading Disambiguation

○Koki Nishiyama[1], Kazuhide Yamamoto[1], Hideharu Nakajima[2] （1. Nagaoka University of Technology, 2. NTT Media Intelligence Laboratories）

9:00 AM - 10:40 AM

[4Pin1-53] An Activity Description Method by Multiple Events of Multiple Objects

○Takayuki Semitsu[1], Takashi Nammoto[1], Toshisada Mariyama[1] （1. Mitsubishi Electric Corporation）

9:00 AM - 10:40 AM

[4Pin1-54] Towards Annotating Problem Solutions and Evaluation Expressions in Information Science Papers

○Hono Shirai[1], Naoya Inoue[1], Kentaro Inui[1,2] （1. Tohoku University, 2. RIKEN Center for Advanced Intelligence Project (AIP)）

9:00 AM - 10:40 AM

©The Japanese Society for Artificial Intelligence

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

特徴グラフを用いた汎用型 CNN 深層学習手法の開発

Development of general purpose CNN deep layer learning method using feature graph

高橋彗 [1], 沼尻 匠 [2], 曽我部 完 [2], 坂本克好 [1], 山口浩一 [1], 横川慎二 [3], 曽我部東馬*[1,2,3]

[1] 電気通信大学 基盤理工学専攻
[2] 株式会社 グリッド
[3] 電気通信大学 i-パワードエネルギー・システム研究センター

we propose a method of applying Convolutional Neural Networks (CNN) to non - image data. CNN has been successful in many fields such as image processing and speech recognition. On the other hand, it was difficult to adapt CNN to non-image data such as a csv file. The sequence of the data of the low dimensional grid structure such as the image has a meaning, and the CNN recognizes the order as the feature of the image and processes it. Therefore, CNN could not perform feature recognition on non-image data whose structure can be changed. We focused on a method to make CNN applicable by giving meaning to the sequence of non - image data, and demonstrated by adding improvements.

1. はじめに

様々な種類のデータの収集と保存は、あらゆる分野において基礎的かつ重要な活動である。収集され蓄積されるデータの量は日々、驚異的な割合で増加している。このデータを有用な知識にするために、大規模かつ複雑なデータセットを理解する方法を開発しなければならない。

その解析方法の新たな糸口として畳み込みニューラルネットワーク(Convolutional Neural Networks: CNNs)[1]が考えられている。従来のニューラルネットワークは過学習や勾配消失の問題から層を増やすことが容易ではないが、CNN は畳み込み層とプーリング層を交互に繰り返すことで層を増やし、ネットワークを深層化することが可能である。そのため近年、CNN はディープラーニングの主力とされ、画像処理、音声認識、コンピュータビジョン、言語処理など多くの分野で成功を収めてきた。一方で、CNN を変化するグラフ構造を持つデータに適応することは困難とされてきた。なぜなら CNN は画像などの低次元グリッド構造の幾何学的特性を利用する畳み込みをしているため、非画像であるグラフ構造やグリッド構造のデータに対しては畳み込みをしても、正しい特徴認識が行えないためである。

そうした理由から最近、非画像データに対して CNN を一般化する方法が研究されている。本稿では、データの相関を利用して畳み込みを可能とする Correlation GraphCNN という手法を元に、データ間の距離を利用した非画像データに対する CNN の応用方法を提案する。

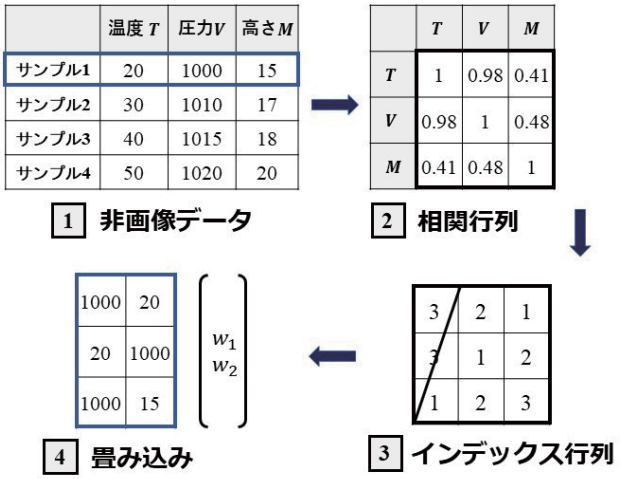
2. Correlation GraphCNN (C-GraphCNN)

CNN で最も重要な役割を果たしている畳み込み層は、フィルタと呼ばれる小さな行列を画像全体に渡って適用することで、画素ごとの数値の『順番』を特徴として学習している。CNN が得意とする画像処理の分野では、データの列や行の順番をランダムに入れ換えると、画像が本来の形から崩れて、データの持つ意味は変わってしまう。つまり、画像データの要素の順番には意味があるということであり、その意味がある要素の順番を CNN は特徴と捉えることで、学習を行っている。

図 1 GraphCNN の畳み込み手順例

連絡先：曽我部 東馬,電気通信大学 i-パワードエネルギー・システム研究センター, sogabe@uec.ac.jp

一方、csv ファイルのような非画像データは、列順を入れ換えたとしてもデータの持つ意味は変わらないため、データの順番



に意味はないということになる。そのため前述の理由から CNN は非画像データを学習することが出来ない。どうすれば、非画像データを CNN に適応できるかを考えたとき、非画像データの順番にも意味を持たせればいいということが考えられる。

そして最近、相関を用いて非画像データの順番に意味を持たせる Correlation GraphCNN[2] という手法が報告された。C-GraphCNN の畳み込みの手順例を図 1 に示す。C-GraphCNN では、最初に非画像データの特徴同士の相関係数を計算し、相関行列を作成する。相関行列は性質上、並べ替えることが出来ないため、データの順番に意味があるということになる。その相関行列から相関係数を強い順に並べたインデックス行列を作成する。ここで、データに対してあまり相関のない特徴変数を除外するため、インデックス行列を左側から任意の行だけ削除する。そしてサンプル 1 を学習する場合、インデックス行列の各行の要素の順番にサンプル 1 から特徴変数の値を取ることで、順番に意味がある行列を作成される。こうした行列を作成し、畳み込みを可能とするのが C-GraphCNN の手法である。

3. Distance GraphCNN (D-GraphCNN)

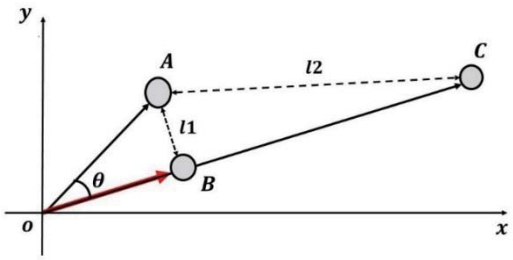
Correlation GraphCNN では、非画像データの順番に意味を与える手段として相関係数を用いていたが、相関係数にはある

欠点がある。ここで図 2 のような 3 つのデータ点がある場合を考える。相関係数 r は 2 つのベクトルのなす角度 θ の余弦で表されるため、ベクトル OA とベクトル OB の相関係数は $cos\theta$ となる。同様に、ベクトル OA とベクトル OC の相関係数もまた $cos\theta$ で表される。つまり、相関係数を基準に考えたとき、点 A,B,C は同じクラスタに分類される。

一方、データ同士の距離を基準に考えたとき、点 C は点 A,B から離れているため、一般的に点 A,B と同じクラスタと見なすことは出来ない。相関係数には、こうしたデータ間の実際の距離を考慮せずに関係性を決めてしまう特性があるため、データの関係性を表すのに最適とは言い切れない。そのため、本稿では非画像データの順番に意味を与える手段として相関係数ではなく、距離をコンセプトとした Distance GraphCNN という手法を 2 つに分けて提案する。

図 2 相関と距離によるクラスタリングの違い

(i) Kernel GraphCNN



Kernel GraphCNN では、全てのデータ間の距離を計算し、距離行列を作成することで、データの順番に意味を与える。計算にはガウシアンカーネルを用いる。ガウシアンカーネルは、

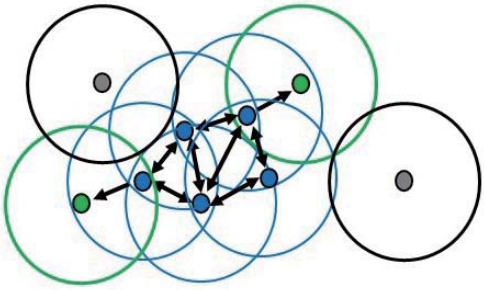
$$\mathrm{k}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

で定義される二変数関数であり、$\mathrm{k}(x_i, x_j)$はx_i, x_jの距離を表す。σはカーネル関数の拡がりを制御するパラメータであり、σ が小さいほど細かいクラスタリングになり、大きいほど単純なクラスタリングになる。

(ii) DBSCAN GraphCNN

DBSCAN GraphCNN では DBSCAN[3] という一般的なクラスタリングアルゴリズムを用いて、近い点同士のみの距離を計算し距離行列を作成する。特徴空間において、多くの点が近接しているような領域を高密度領域といい、DBSCAN はその領域に属する点を見つける。DBSCAN には minPts と eps という二つのパラメータがある。高密度領域内にあるデータをコアサンプルといい、距離 eps 以内に minPts 個以上のデータポイントがあるポイントをコアサンプルとして定義する。また、コアポイントから距離 eps 以内にあるデータを到達可能点、距離 eps 以内に minPts 個以上のデータポイントがない場合、そのデータを外れ点として扱う。図 3 に DBSCAN のクラスタリング例を示す。図 3 において minPts=3 であり、青点はコアポイントである。また緑点は距離 eps 内にコアポイントを含むため到達可能点となる。青点と緑点は単一のクラスタとして扱われ、黒点はコアポイントでも到達可能点でもないため外れ点となる。データに対してこのようなクラスタリングを行い、同クラスタのデータとの距離行列を作成する。

図 3 DBSCAN によるクラスタリング例



4. 実験および評価

Distance GraphCNN が非画像データに対して有効であるか、また C-GraphCNN との精度の違いを比較することを目的に 3 つのデータセットを用いて実験を行った。解析には入力層の後に、GraphCNN による畳み込み層と Dropout 層を 2 回繰り返され、Flatten 層、Dropout 層、全結合層、出力層と続く 9 層からなるニューラルネットワークを用いた。また比較のため 3 つの GraphCNN の他に、4 層のニューラルネットワークを加えた計 4 種の構成を用いた。

4.1 非画像データに対する回帰問題

(i) The Boston Housing Dataset

非画像データの例として、物件の人口統計に関する 12 つの情報から住宅価格を予想する回帰問題を扱う”The Boston Housing Dataset”[4] を用いた。データは 506 サンプルであり、分析の評価には決定係数を用いた。決定係数は

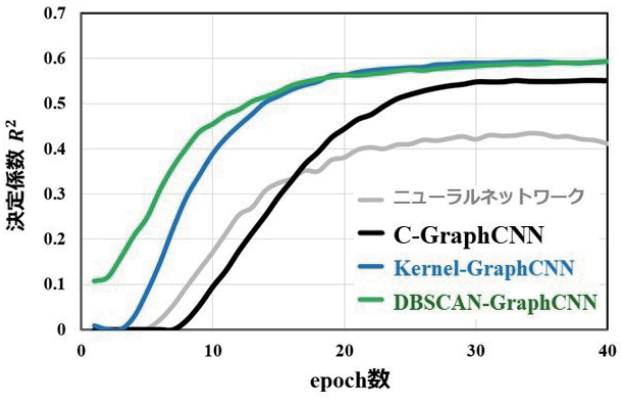
$$R^2 = \mathrm{Corr}(Y, \hat{Y})^2$$

で定義される関数であり、Y は実際の価格、$\hat{Y}$は予測された価格、R^2 は 1 に近いほど予想の精度が高いことを表す。3 つの GraphCNN の構成において、畳み込み 1 層目は 15 枚、2 層目は 20 枚のフィルタ、全結合層のユニット数は 150 個、Epoch 数は 40 回というパラメータ設定を共通に用いた。

実験結果を図 4 に示す。GraphCNN を用いた 3 つの構成の精度は、従来のニューラルネットワークの精度を大きく上回るということが分かる。また、C-GraphCNN はエポック数が 30 回付近で学習が収束しているのに対して、ガウシアンカーネル・高次元距離行列ではエポック数が 20 回付近で学習が収束している。これにより GraphCNN は相関行列を用いるよりも、距離行列を用いた方が予測の精度が良く、学習の速度も速いということが考えられる。

図 4 回帰問題における決定係数の推移

(ii) Merck Molecular Activity



次に、2153 個のも特徴と 6148 個のサンプルデータを持つ“Merck Molecular Activity Challenge”[5]に挑戦した。このデータセットはデータマイニングコンペティションサイト Kaggle で扱われていた題材であり、分子内の原子間の構造に基づいて、異なる分子の活性レベルを予測するという回帰問題である。

実験結果および様々な構成のR^2 結果を表 1 に示す。DNN と Random Forest は当時の Kaggle コンテストの勝者の 2 人のものであり、Spectral Networks[6] では最先端技術であるスペクトルアプローチを使用している。比較すると GraphCNN は Kaggle コンテストの優勝者よりも優れた性能を発揮していることが分かる。また DBSAN GraphCNN は最先端技術である Spectral Networks よりもわずかだが予測精度が良いことが分かる。

以上、2 つの実験結果より、GraphCNN は非画像データに対して有効的であり、我々の提案手法である D-GraphCNN は C-GraphCNN の精度を上回ることも実証された。

表 1　Merck の分子活性における各手法の予測精度

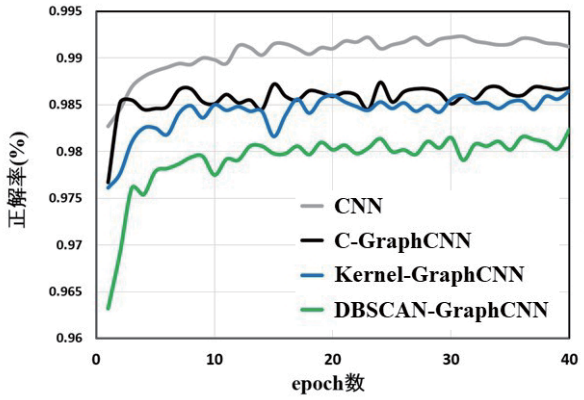
Method	R^2
Random Forest	0.232
Merck winner DNN	0.224
Spectral Networks	0.277
C-GraphCNN	0.264
Kernel GraphCNN	0.235
DBSCAN GraphCNN	0.280

4.2 画像の分類問題とフィルタの可視化

通常、CNN は学習後のフィルタの重みや畳み込み層から出力される値を可視化することで、データの何を特徴として学習しているかを知ることが出来る。しかし、非画像データは画像データのように、出力値を可視化しても何を表しているかを読み取ることは出来ない。そこで GraphCNN は非画像データだけではなく、画像データに対しても適用可能なため、画像データに GraphCNN を応用し、フィルタの可視化を試みた。

まず画像データに対して GraphCNN を用いて学習した場合と、9 層の標準的な CNN で学習した場合の違いを比較した。学習サンプルには MNIST データセット[7]を用いた。テストデータ 60000 サンプルを用いて画像の特徴を学習し、完成した識別機を用いてテストデータ 10000 サンプルに対して 10 クラスの分類問題を行い、正解率の比較およびフィルタの可視化を行った。

図 5　分類問題のおける正解率の推移



実験結果を図 5 に示す。標準的な CNN の正解率 99.12%には及ばなかったものの、GraphCNN はどの構成においても正解率 98%超えており、画像データに対しても GraphCNN は有効的であると判断できる。標準的な CNN はデータを 28×28 ピクセルの画像データとして学習しているのに対して、GraphCNN は 784×1 の数値データとして学習している。そのため、CNN が高い性能を発揮する画像データにおいては GraphCNN の精度は CNN よりもやや劣ると考えられる。

また、標準的 CNN と Kernel-GraphCNN の学習後の畳み込み 1 層目のフィルタの重み及び出力値の可視化したものを図 6 に示す。フィルタには 0 から 255 の数値が各ピクセルに割り当てられており、特徴として認識する形を緑色で表している。標準的 CNN が学習したフィルタからは、数字の横線や斜め線という特徴を抽出していることが分かり、それは出力される画像からも読み取ることが出来る。一方、GraphCNN で同じ画像データを学習した場合、フィルタはそれぞれ異なる特徴を抽出していることが分かる。出力された画像を見ると、数字の輪郭を認識する傾向があると考えられる。しかし、非画像データのおいては、出力値を画像として可視化することは出来ないため、フィルタが何を学習しているかはまだ分かっておらず、今後の課題として残っている。

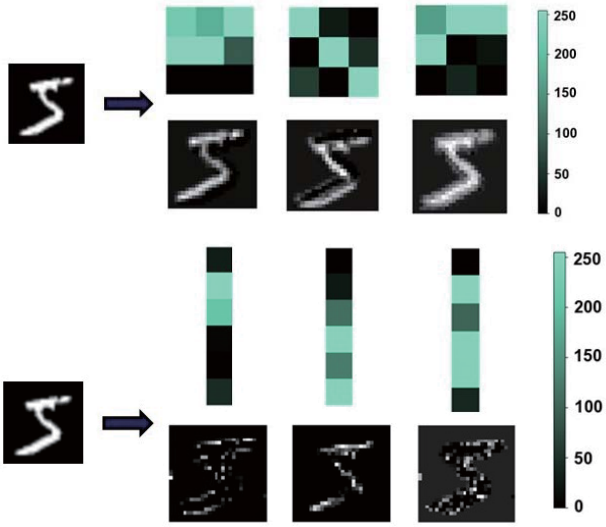


図 6 フィルタと出力値の可視化
(上図: 標準的 CNN 、下図: GraphCNN)

5. 強化学習への応用

強化学習とは試行錯誤を通じて、環境に適応する機械学習の一つである。強化学習が扱うデータは非画像データが多く、一般的にステータスの更新にはニューラルネットワークが用いられている。この強化学習のニューラルネットワークにおいても、勾配消失等の問題により層を増やすことが容易ではなく、4 層などの浅いネットワークでステータスの更新を行っていた。

しかし本実験で、非画像データに対しても GraphCNN は有効的であり、なおかつニューラルネットワークよりも優れた結果を出すことが実証された。そのため、強化学習におけるニューラルネットワークを GraphCNN に置き換えることでネットワークをより深層化し、学習精度の向上を試みる実験を行った。実験は Open AI gym のカートポールを利用したシミュレーター環境で行った。エピソードは 1000 回に設定し、4 層のニューラルネットワークと 8 層の C-GraphCNN の二つの構成を用いて、精度の違いを比較した。

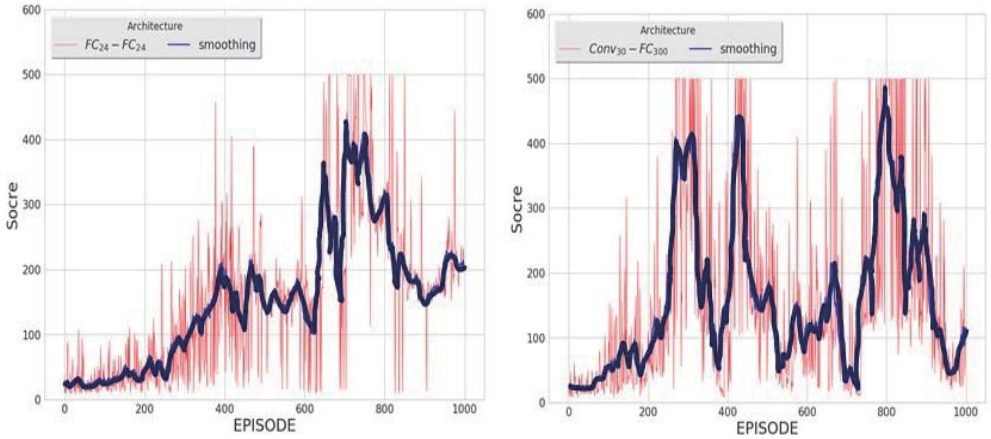


図 7　カートポールにおけるスコアの推移
(左図: ニューラルネットワーク、 右図:C-GraphCNN)

図 7 にエピソードごとのスコアの推移を示す。左の図がニューラルネットワーク、右図が C-GraphCNN で学習した結果である。赤線はエピソードごとのスコア、青線は 10 エピソードごとに平均してスムージングしたスコアを表す。GraphCNN で学習した場合は、ニューラルネットワークに比べて、スコアが最高値に到達するのが早く、また最高値を維持する回数も多いことが分かる。これより、強化学習においても GraphCNN は有効的であると考えられる。しかし、GraphCNN は行動ごとのスコアの差がやや激しいため、スコアを安定させることが改善点として残っている。

6. まとめ

本研究では、非画像データに対しての CNN の応用手法として Correlation GraphCNN に注目し、その手法の改善案として二つの距離行列を用いた Distance GraphCNN を提案した。提案手法の有効性を検証するため、非画像データの回帰問題に対する実験を行った。その結果、GraphCNN の予測精度は従来のニューラルネットワークよりも優れており、非画像データにも CNN が応用できることが明らかになった。また、提案手法である Distance GraphCNN は、Correlation GraphCNN の予測精度を上回り、距離行列の有用性が証明された。しかし、GraphCNN が非画像データに対して有効的であるのは証明されたが、データの何を特徴として捉え、学習しているのかは現段階では分かっておらず、今後の課題として残っている。また、GraphCNN は強化学習の一連の学習過程にも応用可能であり、学習精度の向上が期待される。

参考文献

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L.D. Jackel, “Backpropagation applied to handwritten zip code recognition”, Neural Computation, vol.1, pp.541-551, 1989.

[2] Yotam Hechtlinger, Purvasha Chakravarti, Jining Qin: A Generalization of Convolutional Neural Networks to Graph-Structured Data, arXiv preprint, arXiv:1704.08165 (2017)

[3] Martin Ester，Hans-Peter Kriegel，Jorg Sander，Xiaowei Xu..A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proccrdings of 2ndInternational Conference on Knowledge Discovery and Data Mining (KDD-96),1996.

[4] Harrison, D. and Rubinfeld, D.L. `*Hedonic prices and the demand for clean air*', J. Environ. Economics & Management, vol.5, 81-102, 1978.

[5] https://www.kaggle.com/c/MerckActivity

[6] Mikael Henaff, Joan Bruna, Yann LeCun, “Deep Convolutional Networks on Graph-Structured Data”, arXiv preprint, arXiv:1506.05163 (2015)

[7] Yann LeCun, Corinna Cortes, Christopher J.C. Burges, “THE MNIST DATABASE of handwritten digits”, http://yann.lecun.com/exdb/mnist/

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

Change detection across domains using GAN-based image translation

Tanaka Kanji[*1] Yamaguchi Yousuke[*1] Sugimoto Takuma[*1]

[*1] University of Fukui

The problem of visual change detection becomes a challenging one when query and reference images involve different domains (e.g., time of the day, weather, and season) due to variations in object appearance and a limited amount of training examples. In this study, we address the above issue by training a GAN-based image translator that maps a reference image to a virtual image that cannot be discriminated from query domain images, and experimentally verify efficacy of the approach.

1. Introduction

This paper addresses the problem of change detection from a novel perspective of domain adaptation. Change detection aims to identify interesting changes between a given query image and a reference image of the same scene taken at different times. This is a fundamental problem in computer vision and robotics with many important applications including visual navigation, novelty detection, surface inspection, and city model maintenance. This problem becomes a challenging one when query and reference images involve different domains (e.g., time of the day, weather, season). This is due to variations in object appearance in different domains, which makes it harder to discriminate changes of interest from nuisances. One of most basic schemes to handle this difficulty is to train a domain-specific change predictor for each possible domain pair. However, this requires a vehicle to collect and learn from a number of training images from both domains, which severely limits its applicability to new domains where large amounts of annotated training data are typically unavailable.

In this study, we address the above issue by leveraging a generative adversarial network (GAN) (Fig. 1). Our approach is motivated by the recent success of the image-to-image translation technique in the field of computer vision and graphics. The image-to-image translation technique aims to learn the mapping between an input image in one domain to a virtual image that cannot be discriminated from target domain images. The key concept we employ is to use a limited amount of training data to train a GAN-based image translator for reference-to-query domain translation. This enables us to treat the cross-domain change detection task as an in-domain image comparison and allows us to leverage the large body of literature on in-domain generic change detectors. In addition, we also consider the use of visual place recognition as a method for mining more appropriate reference images over the space of virtual images. Experiments validate efficacy of the proposed approach.

2. Related Work

Change detection has been widely studied in various task scenarios. These scenarios are categorized into 2D-to-2D matching [1], 3D-to-3D matching [2], 2D-to-3D matching [3]. [1] considers a 2D-to-2D matching on overhead imagery using an image registration and scale invariant feature transform (SIFT). [2] addressed a 3D-to-3D matching on point cloud data for novelty detection by patrol robots. [3] considered a 2D-to-3D matching between monocular image and cadastral 3D city models. Our approach belongs to 2D-2D matching, and our setting is far more complicated than the case of overhead imagery [1] as our vehicular applications require the capability to deal with more general 6 degrees-of-freedom freely-moving camera motion in a 3D space.

図 1: Change detection using image-to-image translation by generative adversarial network (GAN). Left: a reference image. Middle: virtual reference image. Right: query image.

Up until now, the most fundamental scheme reported for addressing this challenge is to directly compare each view image against the corresponding reference image. In [4], a scene alignment method for image differencing is proposed based on ground surface reconstruction, texture projection, image rendering, and registration refinement. In [5], a deep deconvolutional network for pixelwise change detection was trained and used for comparing query and reference image patches. However, a major concern of these direct methods is that they may fail to capture appearance variations caused by a domain shift.

Our approach is inspired by the domain adaptation and transfer learning approaches, ranging from parameter adaptation, feature transformation, and metric learning, to deep learning techniques, which have been applied to a wide variety of visual recognition tasks. In recent years, the application of GANs to domain adaptation tasks has attracted much attention. In our application domain,

連絡先: Tanaka Kanji, University of Fukui, 3-9-1 bunkyo fukui fukui japan, +81-776-27-8033, tnkknj@u-fukui.ac.jp

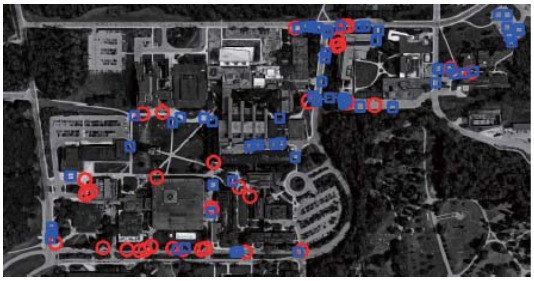


図 2: Experimental environments and viewpoints. (Red/blue circles: the viewpoints of the 100 pairs of summer/winter images in the test set.)

image-to-image translation has been one of mostly studied issues over the last two years [6]. Image-to-image translation aims to map a raw image in a source domain to a virtual image in such a way that the virtual image cannot be discriminated from raw images in the target domain. Typical training algorithms for the translation GAN take a set of paired source-target images as input. We address this basic setting and use a typical architecture in [6] for the current experimental system.

3. Problem

Although our approach is sufficiently general and applicable to various types of environments (indoor and outdoor for instance) and sensor modalities, in our experiments, we focus on the NCLT dataset [7]. The NCLT dataset is a large scale, long-term autonomy dataset for robotics research collected at the University of Michigan's North Campus by a Segway vehicle robotic platform. The Segway was outfitted with: a Ladybug3 omnidirectional camera, a Velodyne HDL-32E 3D lidar, two Hokuyo planar lidars, an IMU, a single axis fiber optic gyro (FOG), a consumer grade global positioning system (GPS), and a real-time kinematic (RTK) GPS. The data we used in the research includes image and navigation data from the NCLT dataset. The image data is from the front facing camera (camera#5) of the Ladybug3 omnidirectional camera. Fig. 2 shows a bird's eye view of the experimental environment and the vehicle's camera viewpoints.

During the vehicle's trips through the outdoor environments (Fig. 2), it encounters various types of changes, which originate from the movement of individuals, parking of cars, building construction, and opening or closing of doors. There are also nuisances that originate from illumination alterations, viewpoint-dependent changes of objects' appearances and occlusions, weather variations, falling leaves and snow. A critical and significant challenge in a substantial majority of change detection tasks is to discriminate changes of interest from nuisances. This renders our change detection task significantly more demanding.

Precision-recall is a standard performance index for image based change detection algorithms [8]. As a performance index, we use mean average precision (mAP), derived from the field of visual object detection [9]. A change detection algorithm outputs a prediction of the likelihood of change (LOC) for each cell on a $w \times h$ image grid of LOC values, as we will describe in 4.3. For the evaluation, the LOC value of each cell in the image grid is thresholded into a state of change or no-change, and then the binary change mask is compared against a ground-truth change mask. Different threshold values provide different tradeoff points between precision and recall. The mAP can be viewed as a summary of these different tradeoff points and approximates the value of receiver operating characteristic (ROC). Following the literature [9], we approximate the average precision (AP) based on the average of "interpolated" precision values for 11 different recall values, 0, 10, ..., 100 [%]. The mAP is obtained as the mean of the AP values over all the query images. Considering the fact that achieving 100% recall is not necessarily required in typical change detection applications, we also evaluate mAP@X%recall for different percentage points X =10, 20, 50, and 100.

4. Approach

Our domain-adaptive change detection (DA-CD) framework consists of three distinct steps: image adaptation, image description and image comparison. Image adaptation aims to map a given reference (or background) image to a virtual image, using the GAN-based image-to-image translation, in such a way that the virtual reference image cannot be discriminated from raw images in the target domain. Image description aims to extract a set of image feature descriptors from either image for image comparison. Image comparison aims to predict a change mask by comparing local features between the query and reference images. The three steps above are detailed in the following subsections.

4.1 Image Adaptation

In order to train a GAN that translates an image from the source domain to the corresponding virtual images in the target domain, we require a small number of paired source-target images as training data. We pair images based on their viewpoint locations. The task of pairing requires global viewpoint information in the source and target domains. Fortunately, in vehicular applications, it is often straightforward to obtain the small set of training images and the aforementioned global viewpoint information. First, a small number of training sets can be acquired by the vehicle's short range navigation in the target domain. Second, if the GPS measurement of the viewpoints are available, a target image with nearest neighbor viewpoint is selected as the corresponding image. Even in a GPS-denied environment, pseudo GPS information can be obtained with sufficiently high quality by using a modern visual place recognition (VPR) technique. For example, in [10], we also have developed a robust approach to cross-season VPR.

The GAN-based translation problem is often formulated as a per-pixel regression problem. However, this formulation treats the output space as "unstructured" in the sense that each output pixel is considered conditionally independent from all others given the input image. This issue is addressed by conditional GAN with a structured loss function, which penalizes a joint configuration of the output. In particular, the approach proposed in [6] aims for a general-purpose solution to image-to-image translation problems. The proposed approach consists of a generator based on U-Net which allows low-level information to take a shortcut across the network, and a discriminator based on PatchGAN that combines a typical L1 loss (that enforces correctness at low-frequencies) with an additional patch level loss (that models high-frequencies). We also refer to the results in [6], where the approach was found to be effective for a domain adaptation task of "day-to-night" image

translation, which is partly similar to the application of a cross-season image translation task that we present. In both the training and translation tasks, we resize each input image to 256×256.

4.2 Image Description

We follow the literature and use the standard local image feature descriptor, SIFT, as the basis for image comparison and change detection. In this study, we are particulary interested in how effective intensity-non-invariant feature descriptors like SIFT descriptors are when combined with our strategy of image adaptation. Intuitively, when the image adaptation is successful, the image comparison between the query image and the adapted background image can be viewed as an in-domain image comparison task rather than a cross-domain image comparison. Therefore, intensity-non-invariant feature descriptors are expected to be effective for such a pseudo in-domain image comparison. Each SIFT feature descriptor is L1 normalized.

We basically use a Harris-Laplace detector as a feature keypoint detector. This detector typically provides a set of well-localized keypoints. However, in low texture image regions, it fails to detect localized keypoints and the spatial density of keypoints tend to be very low. To compensate, we also employ an alternative dense sampling -based detector. The dense sampling -based detector outputs a large number of spatially uniform keypoints even in a low texture region. However, as a downside, such keypoints from dense sampling tend to be not well-localized and have a low invariance. This potentially yields a bias due to an increase in dissimilarity between local features at these dense keypoints, which then increases the false positive change detection. To deal with this issue, we decided to penalize the LOC values of these features by multiplying a pre-defined coefficient $\exp^{-d_a^2}$ ($d_a \gg 1$).

By default, the features from the virtual image are merged with those from the raw reference image, and then the merged set is used for the image comparison. In the experiments, we also consider an alternative setting where the virtual image set alone is used (without the merge) for image comparison. The above two strategies are respectively termed “w/ merge” and “w/o merge”.

4.3 Image Comparison

The image comparison step adopts the nearest neighbor (NN) search of local features. In offline mode, every feature in the reference image is indexed by an NN data structure. Once online, the NN search is performed for each query feature over the reference (raw/virtual) features. The search result provides a dissimilarity score of a query feature’s NN reference features with respect to the query feature, which can be interpreted to the LOC score of the query feature. We measure the dissimilarity by calculating the Euclidean distance $|f_i - f_g|$ between query feature f_i and its NN feature f_g. The search region on the image plane is defined as a radius 10-pixel circular region centered at the query keypoint of interest, for 256×256 query and reference images.

We also adopt a feature matching -based feature penalization. This strategy has been found to be effective in our previous research in [11]. The basic function of the feature matching approach is to conduct a similarity search for a given query feature over a collection of spatially relevant reference features. If a strong match is found, we can expect the actual possibility of the feature being changed to be very low. Therefore, we penalize the LOC values of these features by multiplying by pre-defined coefficient

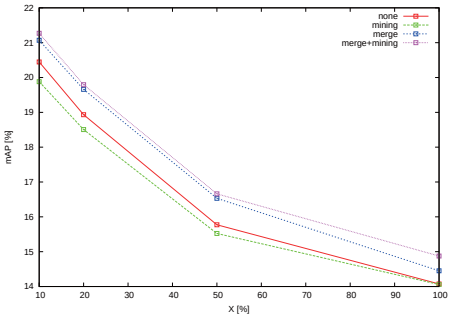


図 3: mAP@X[%]recall performance for different strategies.

$\exp^{-d_b^2}$ ($d_a \gg d_b \gg 1$). To detect strong matches, we use the SIFT ratio test [12] with a threshold value of 0.4. In preliminary experiments, we also tested different threshold values ranging from 0.4 to 0.8 and found that the difference among these threshold values yield a low significance in change detection performance. The feature matching is expected to effectively suppress false positive change detection even when ambiguous anomaly features exist.

In typical change detection applications, we are interested in pixel-level LOC, rather than feature-level LOC [8]. To obtain such pixel-level LOC values, we introduce a coarse $w \times h$ image grid of LOC values, where the width w and height h of the grid is set to $\lfloor W/10 \rfloor$ and $\lfloor H/10 \rfloor$ for a $W \times H$ image. The LOC value of each grid cell is computed from the LOC values of query features that belong to that grid cell by using max pooling.

In addition, we also explored the use of the VPR technique as a method for selecting appropriate background images, which we call a “background mining” strategy. In general, a virtual image generated by a GAN-based image-to-image translation is often inconsistent with the corresponding query image. This is natural because the query image is unseen and not available at the GAN training stage. To address this issue, here we use the VPR technique as a method for searching for a more relevant background image to a given query image, over the space of all the virtual images of the target domain. In this paper, the VPR algorithm adopts a local feature NN search with a Naive Bayes nearest neighbor (NBNN) distance metric, which was found to be effective in a cross-domain VPR task in our previous research [10]. In experiments, we consider and compare two distinct cases: one for the normal background image output by the GAN-based image-to-image translator, and one for the virtual background image selected by background mining.

5. Experiments

We evaluated the change detection strategies using an NCLT dataset [7]. As described in 3., the NCLT dataset is a large scale, long-term autonomy dataset for robotics research collected at the University of Michigan’s North Campus by a Segway vehicle robotic platform. During the vehicle’s trip through outdoor environments, it encounters various types of interesting changes and nuisances. A critical and significant challenge in a change detection task is to discriminate changes of interest from nuisances. We consider two different scenarios of domain-shift: winter-to-summer and summer-to-winter. Our winter image set is generated from the datasets “2012/1/22”, “2012/1/15”, and “2013/1/10” in

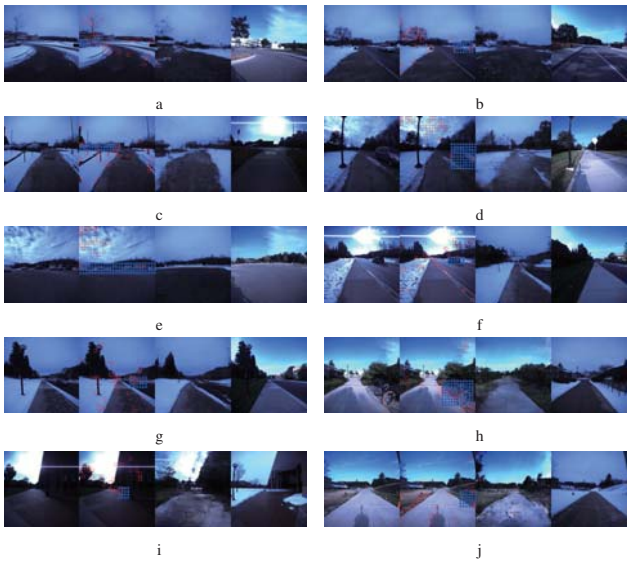


図 4: Examples of change detection.

NCLT, while our summer set is generated from "2012/8/4", by the following procedure. First, images from different seasons are paired using the GPS information available in the NCLT dataset. Then, a random subset of the paired images is used for training the GAN-based image-to-image translator. Two translators corresponding to winter-to-summer and summer-to-winter translation tasks are trained using 100 and 100 paired training images, which are independent from the test sets and randomly sampled from the both domains. Then, every paired image is manually examined and annotated with a bounding box of a change object. We obtain 100 annotated image pairs as the test set, which consist of 64 summer-to-winter pairs and 36 winter-to-summer pairs.

We categorized small changes (e.g., 10×10 [pixels]) that typically originate from distant objects, into no-change because it is challenging to detect such small changes via a visual change detection algorithm. As a result, such small objects are likely to cause pseudo false-positive detection by change detection algorithms.

The average time-overhead for image translation was around 1.0 sec and the time-overhead for change detection was 1.0 sec on a single PC (Intel Core i7-7700 3.2 GHz, GeForce GTX 1080).

Fig. 3 shows the mAP@X%recall performance for different change detection strategies. As mentioned in 4.2, we consider two settings, whether or not features from the virtual reference images are merged with features from the raw reference image, which are respectively called "w/ merge" and "w/o merge". In addition, we consider alternative settings and whether or not we use the background mining described in 4.3, which is termed "w/ mining" and "w/o mining". Figure 3 shows the mAP performance for different combinations of these three strategies. It can be observed that the combination "merge+mining" yields the highest mAP@100%recall performance.

Fig. 4 displays examples of change detection results. From left to right, each panel shows the input query image, prediction (red boxes) and ground-truth (blue points) change detection overlaid on the query image, a virtual reference image, and the input reference image. Red boxes indicate prediction results corresponding image grid cells with the top 5% LOC values. It can be seen that in these examples, the virtual reference image preserves the color and structure information and successfully generates a realistic virtual image, while the edge information is often blurred or lost. It is noteworthy that even with such erroneous edge information, the change detection was successful as reported in Fig. 3.

Acknowledgment

Our work has been supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) 17K00361.

参考文献

[1] W. Li, X. Li, Y. Wu, and Z. Hu, "A novel framework for urban change detection using vhr satellite images," in *Proc. Int. Conf. Pattern Recognition*, 2006, pp. 312–315.

[2] P. Drews, P. Núñez, R. Rocha, M. Campos, and J. Dias, "Novelty detection and 3d shape retrieval using superquadrics and multi-scale sampling for autonomous mobile robots," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3635–3640.

[3] A. Taneja, L. Ballan, and M. Pollefeys, "City-scale change detection in cadastral 3d models using images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 113–120.

[4] D. W. van de Wouw, G. Dubbelman, and P. H. de With, "Hierarchical 2.5-d scene alignment for change detection with large viewpoint differences," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 361–368, 2016.

[5] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, "Street-view change detection with deconvolutional networks." in *Robotics: Science and Systems*, 2016.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arxiv*, 2016.

[7] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *Int. J. Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[8] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. image processing*, vol. 14, no. 3, pp. 294–307, 2005.

[9] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. computer vision*, vol. 88, no. 2, 2010.

[10] K. Tanaka, "Cross-season place recognition using nbnn scene descriptor," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*. 2015, pp. 729–735.

[11] T. Murase, K. Tanaka, and A. Takayama, "Change detection with global viewpoint localization," in *IAPR Asian Conf. Pattern Recognition*. 2017, pp. 31–36.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. computer vision*, vol. 60, no. 2, 2004.

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

リカレントニューラルネットワークを用いた対話システムの試作と評価

Prototype and evaluation of active listening system using Recurrent Neural Network

田村和弘[*1]
Kazuhiro Tamura

[*1]株式会社シーイーシーカスタマサービス
CEC Customer Service Ltd.

In this paper, we tried to make a prototype active listening system using RNN (recurrent neural network) to be widely used by the elderly and the stressed young people and described on the elements necessary for the system. It was found that this system requires discussion on how to collect more effective learning data.

1. はじめに

超高齢社会である日本において、認知症や高齢者の孤独問題がますます顕著になることが予想される。また、若い世代において職場や学校でのいじめやストレスによって心理的に疲弊してしまう問題もある。

これら問題に対してカウンセリングにおける傾聴が有効であり病院や福祉施設などでカウンセラーがそれら問題を抱える人々のケアにあたるが、まだカウンセラーの人手不足でだれもが傾聴してもらえず、家庭でも気軽に傾聴してもらえることができない。

カウンセリングは「傾聴」「肯定的資質」「関係性」「意図性」の特徴をもっているとされる[福原 2007]。

ひとつは傾聴を土台としている点である。傾聴することは、クライエントのストーリーを引き出すのに非常に有効に作用する。その人が、どのような世界で生きてきたのか、そして今どのような問題を抱え、その問題をどの様に捉え、どのように感じ、どうしたいと思っているのかなどを理解するためにストーリーを傾聴によって引き出すことが重要となる。

肯定的資質とは、個人の可能性や潜在能力に目を向けることである。関係性は、個人の文化的背景、すなわち言語、ジェンダー、年齢、身体的問題、社会的経済的地位など様々であることから、それぞれ別々の文化的ヒストリーを持っている点に配慮したかかわりが行われる。意図性は個人の状況に応じてさまざまな理論や技法を柔軟に組み合わせて対応できることが求められていることである。

カウンセリングの技法としては、①かかわり行動、②基本的傾聴技法、③面接構造（ラポール、問題の定義化、目標の設定、選択肢を探求し不一致と対決、日常生活への般化）、④対決、⑤焦点のあてかた、⑥意味の反映、⑦積極技法（指示、論理的帰結、解釈、自己開示、フィードバック、カウンセラー発言の積極的要約、助言、情報提供、教示、統合、個人的スタイルと理論）があげられる。これらを一歩一歩積み重ねながら習熟していくことが重要で、それぞれの技法を試みながら応答の仕方を見直し、自分なりの援助スタイルを確立していく[福原 2007]。

基本的傾聴について下記が特に重要である

1）開かれた質問、閉ざされた質問

閉ざされた質問は、2～3 語の言葉や「はい」、「いいえ」だけで答えられる質問、開かれた質問は「どのように～？」や「何を～？」、「なぜ～？」と尋ねたり、「それについてもう少し詳しく話してくださいませんか？」といった表現を用いる。質問技法を習得するとカウンセラーはクライアントが自分の問題について話をするのを促し、自己探求することを助けることができる

2）はげまし、いいかえ、要約

はげまし、いいかえ、要約の技法は、共感の基礎となる。励まし技法は、「ええ」、「それで？」といったものや、クライアントが言葉を短く繰り返すものなど「言語」によるものとうなずきのような「非言語」によるものとがある。これにより、クライエントは、自分の気持ちや考えを探求し語ることを励まされる。いいかえ、要約は、クライアントが語った本質をフィードバックすることをいう。

3）感情の反映

クライエントが抱える問題は、クライエントがその根底にある自分の感情に気づき、向き合っていかなければ、なかなか解決に至らない。

感情の反映はクライエントの言語化されない感情を注意深く観察し、それを手がかりとしてクライエントの感情に注意を向けてフィードバックしていく。

表 1. カウンセリングの傾聴において重要とされる話し方

1. 語られたことをそのまま繰りかえす
2. 開かれた質問、閉ざされた質問
3. 語られた内容を言い換えて繰り返す
4. 語られた内容を要約する
5. もっと語るように問い返す
6. 話し手に共感し、気持ちを言葉にする
7. 相槌などにより聴いていることを示す

上記のカウンセリングの特徴・技法を踏まえ、高齢者やストレスを抱えた若い世代に広く利用してもらえるための RNN（リカレントニューラルネットワーク）を用いた対話システムの試作を通じて必要な要素を考えることにした。

対話システムとして、音声認識率の向上やスマートフォンの普及で Apple の Siri や Yahoo！の音声アシスト、NTTドコモのしゃべってコンシェルといった音声アプリケーションが登場し一般ユーザにも身近なものになっている。

連絡先：田村和弘，株式会社シーイーシーカスタマサービス，宮崎県都城市，kazuhiro.tamura@gmail.com

単語単位の音声入力や一問一答型対話により情報検索を行うタスク指向型対話システムに関してある一定の性能に達したと考えられている。しかし、これら対話システムはユーザが話す内容や発声の仕方(単語に区切るなど)を制限している。一方で雑談対話のような達成すべきタスクを設定しない対話システムも提案されている。

対話システムの先行研究として、大竹らの研究[大竹 2014]がある。対話システムの機能を傾聴に特化し、対話が破綻することなく継続でき、高齢者から若者まで満足感を得られるようなシステムを目指している先行研究[下岡 2017]、[三上 2017]もあり、これを支持しつつも、相談施設や病院、家庭や学校などで傾聴対話システムを利用してもらうことで、孤独化の防止やストレスによる心理的疲弊を和らげることを目的とし貢献したい。

2. リカレントニューラルネットワーク(RNN)を用いた対話システム

2.1 Sequence-to-Sequence

対話システムとして Sequence-to-Sequence を適用した。

Sequence-to-Sequence モデルは Sequence to Sequence Learning with Neural Networks の論文[Sutskever 2014] で提案され、「Seq2Seq モデル」「Encoder-Decoder モデル」「系列変換モデル」といった名前で呼ばれ、近年の自然言語処理においては、これをベースにしたモデルが多く提案されている。

モデルは Encoder と Decoder という二つの RNN が組み合わさってできており、Encoder が入力データ、Decoder が出力データの処理にそれぞれ対応しており、Encoder によって入力データを中間ノードに変換し, これをもとにして Decoder がデータを出力するといったものである。

2.2 対話コーパスの作成と学習

seq2seq 対話システムを学習させるための、発話とそれに対する返答のテキストデータを収集した。ここでは Twitter の会話を収集した。最終的に発話とそれに対する返答のテキストデータを約 20MB ずつ得た。その後、MeCab(形態素解析システム)で単語分割を行い, train data 90 % と test data 10 %でランダムに分割した。層のサイズは 256、層の数は 3、最大のボキャブラリ数は 100000 に設定した。学習時間は約 1 日程度だった。

3. 対話システムの実行

対話システムの実行は java アプリで実装してソケット通信を介した。図 1 に実行時の画面を表示している。

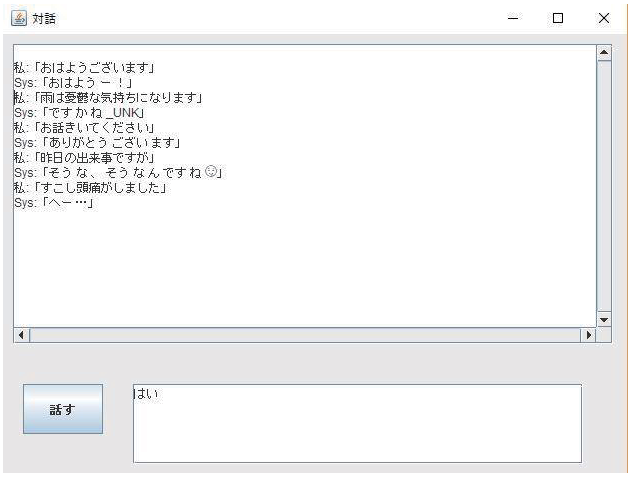


図 1:実行画面(デモ)

4. 考察とまとめ

本研究において試作した対話システムでは Twitter のデータを学習データとして用いたが、Twitter をはじめとするマイクロブログでは, ユーザにより日々多くの会話がなされており、Twitter における会話の応答文生成に関する研究はコミュニケーションが苦手な個人や日常的な会話を自動化したい法人等が抱える具体的な課題に対しても役に立つ可能性があるという意見[那須野 2016] があり我々も賛同できる。

だが、実際 Twitter 上の会話はくだけたものが多く、単語の表記が一定ではなかったり、顔(絵)文字が利用されたりするため Twitter における会話の応答文生成は傾聴対話システムの要件を満たすには難しさがあったことが理解できた。

傾聴対話システムの試作を通じてシステムの必要な要素について言及したが、今後改めてこれらの要素を満たすための実装方法の検討が必要であると考える。さらに、精緻された発話と応答データの有効な収集法においても議論が求められる。

参考文献

[福原 2007] 福原眞知子, マイクロカウンセリング技法- 事例場面から学ぶ-, 風間書房.

[大竹 2014] 大竹 裕也, 萩原 将文, 評価表現による印象推定と傾聴型対話システムへの応用,日本知能情報ファジィ学会誌, 26(2), pp.617-626.

[三上 2017] 三上 佳孝, 萩原 将文, 対話における文脈を考慮した分散表現生成ニューラルネットワーク, 知能と情報, 2017

[下岡 2017] 音声対話ロボットのための傾聴システムの開発, 自然言語処理, 言語処理学会 , 2017.

[Sutskever 2014] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to Sequence Learning with Neural Networks, https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf, arXiv, 2014.

[那須野 2016] 那須野薫, 松尾豊, 深層学習を用いた Twitter における会話の応答文生成の試み, 人口知能学会全国大会(第 30 回)

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

電子顕微鏡連続切片画像セグメンテーションにおける深層学習の利用に関する研究

Segmentation in Stacks of Electron Microscopy Images Using Deep Learning

高屋 英知 *1 Eichi Takaya　竹市 裕介 *2 Yusuke Takeichi　尾崎 まみこ *2 Mamiko Ozaki　栗原 聡 *1 Satoshi Kurihara

*1慶應義塾大学 大学院理工学研究科 Graduate School of Science and Technology, Keio University

*2神戸大学 大学院理学研究科 Graduate School of Science, Kobe University

In the research field of connectomics, serial section electron microscopy images are reconstructed in three dimensions for the purpose of observing neural structure. Automatic segmentation of neuronal cells is important to reconstruct the images, and in recent years a method using deep learning has got a lot of attention. However, in the application of deep learning, there are many optional methods to improve segmentation accuracy. In this paper, we focus on the tuning method of hyperparameters. Deep Contextual Network (DCN), which is expected to be processed with particularly high precision among existing methods, has some feature maps with different spatial frequencies. We have an eye to this structure and propose hyperparameter tuning method based on visualization of feature maps. Experiments using two kinds of datasets showed that DCN accuracy can be improved by changing the kernel size of the deconvolution layer by the proposed method.

1. はじめに

コネクトミクスと呼ばれる分野では，生物の脳や感覚器における神経系の構造や接続状態を詳細に調べ，その機能的な意味を理解することが目指されている [1]．神経構造を調べるための方法はスケールによって様々だが，特にミクロな構造を観察するためには，電子顕微鏡によって撮影された連続切片画像を取得し，それらを 3 次元に再構築する方法が用いられる．しかし，切片画像の再構築には，数百枚にわたる画像に対して図 1 のようなラベリングを行う必要があり，これを人手により行うことは多大なコストを要する．Kaynig ら [2] は，3 次元再構築を自動化するための画像処理パイプラインを採用しており，ここでは神経細胞のセグメンテーションと，3 次元構造における同一領域のグルーピングを段階的に行うことで，ラベリングの自動化を実現している．そして近年では，このパイプラインにおけるそれぞれのステージでの処理精度を向上させるために，機械学習に基づいた様々な手法が提案されている．本稿では，パイプライン方式において全体の精度に与える影響が大きいとされる，神経細胞セグメンテーションのタスクに注目する．

神経細胞セグメンテーションのタスクについては，関連したコンペティション [3] をきっかけに，様々な手法が提案されてきた．特に Ciresan ら [4] により Deep neural network を用いた手法が提案され，その精度の高さが浮き彫りになって以来，新たに提案される手法のほとんどは深層学習に基づくものとなっている．さらに，近年成果を挙げている例としては，Chen ら [5] による Deep contextual network（以下，DCN と呼ぶ）や Ronnenberger ら [6] による U-Net などが挙げられる．いずれもネットワークの構築方法に独自性がみられるが，Long らの提案した Fully convolutional network[7]（以下，FCN と呼ぶ）を基礎とし，中間層で得られた特徴マップを明示的に利用している点が共通している．しかし，あるデータセットにおいて高精度なセグメンテーションを行うためには，ただ上記のようなネットワークを適用するだけでは不十分である．例えば，

連絡先: 高屋英知，慶應義塾大学大学院理工学研究科，
〒 223-8522 神奈川県横浜市港北区日吉３－１４－１，
Email: takaya.wisdom0616@gmail.com

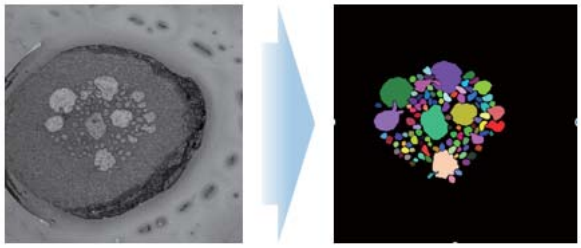


図 1: 連続切片画像のラベリング

ハイパーパラメータの適切な設定や，少ないデータを有効活用するための工夫，時系列性の考慮など，精度向上に寄与しうるメソッドが数多く存在する．本稿では，その中でもハイパーパラメータの設定に焦点を当て，既存のネットワークアーキテクチャのうち特に精度が高いとされる DCN の特性に基づいたハイパーパラメータ調整手法を検討する．具体的には，DCN の特徴マップの可視化に基づき，逆畳み込み層のフィルタサイズを変更するものである．以下では DCN のアーキテクチャについて概観し，続いて提案手法について述べる．

2. Deep Contextual Network

DCN は，すべての層が畳み込み層で構成される FCN を拡張した畳み込みニューラルネットワークである（図 2）．物体認識等で用いられる畳み込みネットワークのアーキテクチャは通常，複数の畳み込み層と全結合層によって構成されているが，FCN は「全畳み込みネットワーク」とも訳されるように，すべての層が畳み込み層で構成されている．そして，畳み込み層とプーリング層による特徴抽出を行うダウンサンプリング部と，抽出された特徴が逆畳み込み層によって入力画像と同サイズに復元されるアップサンプリング部を有することで，ピクセル毎の回帰および分類を行えるようになっている．一方，DCN では，ダウンサンプリング部で得られた段階的な特徴マップから複数のアップサンプリング部（以下ではスキップコネクションと呼ぶ）を構成し，最終的に足し合わせているの

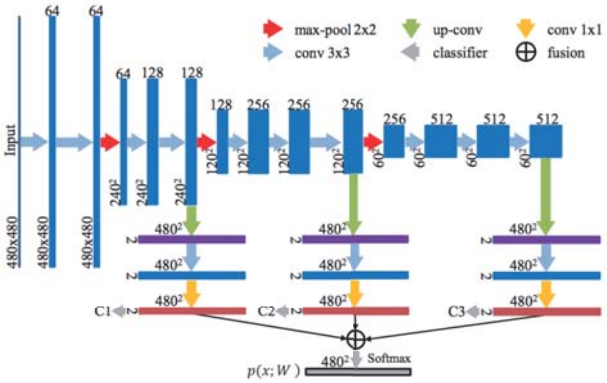


図 2: DCN のアーキテクチャ

が特徴的である．ここでは，深い層で得られた大域的な情報を分類に用い，浅い層で得られた局所的な情報を位置特定に用いることが意図されている [5]．DCN の有するスキップコネクションについては，特に神経細胞セグメンテーションのタスクでの重要性が指摘されており [8]，先述した U-Net などのアーキテクチャでも採用されている．

本研究では，予備実験として Chen ら [5] と同様のデータセットにて DCN の学習を行い，精度の検証および特徴マップの可視化を試みた．その結果，正則化手法として定評のあるバッチ正規化を導入することにより，特徴マップの可視化が明瞭に行えることが明らかになった．それに付随して，最終的に足し合わされる 3 種類の特徴マップは，各々が異なる粒度での特徴を保持しており，協調的に出力画像を生成していることが示唆された．さらに，スキップコネクションを欠落させることで，精度が低下することも明らかになった．

3. DCN 適用時におけるハイパーパラメータの決定

本研究では，DCN の能力を最大限に引き出し，高い精度での神経細胞セグメンテーションを行うためのパラメータチューニング方法を提案する．提案手法の模式図を図 1 に示す．まず，Chen ら [5] が提案したパラメータを用いて DCN による学習を行う．そして，チューニング用データを用いた推論の際に得られる 3 種類の特徴マップを目視によって観察し，十分に働いていない特徴マップを特定する．さらに，特定された特徴マップを出力する支流における，逆畳み込み層のフィルタサイズを変更し，再度学習を行う．

DCN においてはスキップコネクションが重要な要素であり，それらの働きの中心となるのは逆畳み込み層であることから，本研究では逆畳み込み層のハイパーパラメータであるフィルタサイズに注目している．なお，以下では 3 種類の特徴マップを浅い層から順に F1, F2, F3 と呼ぶこととする．

4. 実験

4.1 実験手順

まず，次項で述べるデータセット A を用いて，フィルタサイズの変更が，特徴マップの見た目やセグメンテーション精度にどのような影響を与えるかを検証した．次に，データセット B を用い，提案手法によるフィルタサイズの変更を行い，セグメンテーション精度の変化を検証した．

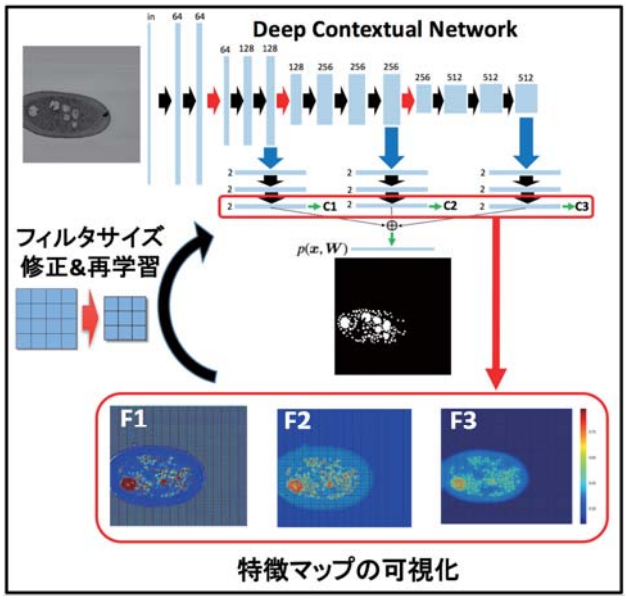


図 3: 提案手法の概要

4.2 データセット及び評価指標

実験には，「ショウジョウバエ 1 齢幼虫腹部連続切片画像」と，「クロオオアリ巣仲間識別感覚子連続切片画像」を用いた．以下ではこれらをデータセット A，データセット B と呼ぶこととする．

データセット A は，ISBI と呼ばれる国際会議のコンペで使用されたもので，DCN を含め，これまでに提案された深層学習アーキテクチャにおいて高い精度が保証されている．そのため，今回はフィルタサイズ変更の影響を検証するために利用するに留めた．訓練データ 30 枚のうち 3 枚を検証用データとし，汎化誤差の確認や特徴マップの可視化に用いた．画像サイズは 512 × 512 ピクセルである．

データセット B は，神戸大学理学研究科の尾崎研究室から提供された連続切片画像データで，377 枚の電子顕微鏡画像に対し，専門家によるラベル付けがなされている（図 1 右）．ここでは，ラベル画像で色の付いている部分のみが正解領域となるようにグレースケール変換したものを教師データとした．画像のサイズは 2048 × 2048 ピクセルであるが，実験では興味ある領域のみを 512 × 512 にクリッピングしたものを使用した．なお，今回は 377 枚の画像のうち，先頭から 100 枚を訓練データとし，続く 10 枚をテストデータとした．ただし，訓練データのうち最後の 10 枚は検証データとした．

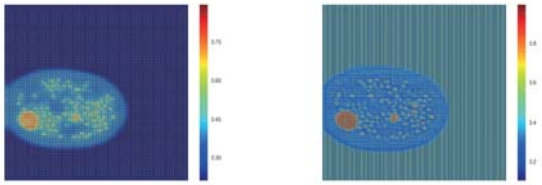
いずれのデータセットにおいても，その量を補うためにデータ拡張を適用している．具体的には，ミニバッチを作成する際に各画像に対し，(1)90 度回転，(2)180 度回転，(3)270 度回転，(4) 左右反転，(5) 上下反転，(6) 何もしない，のいずれかの処理をランダムに施した．

提案手法の評価には，画像セグメンテーションのタスクでしばしば利用される Intersection over Union (以下 IoU) を用いた．

4.3 実験結果

データセット A を用いた実験では，フィルタサイズを縮小した場合に，対応する特徴マップのコントラストが強くなり，拡大した場合には弱くなる傾向があることが明らかになった．

データセット B を用いた実験では，特徴マップの可視化に



(a) 縮小前　　(b) 縮小後

図 4: フィルタサイズの縮小による特徴マップの変化

(a) 縮小前　　(b) 縮小後

図 5: フィルタサイズの縮小による出力画像の変化

より，F3 にフィルタサイズ縮小による改善の余地があると判断された．そして，フィルタサイズを縮小して再度学習を行うことで，対応する特徴マップ F3 の変化がみられ（図 3），出力画像の改善もみられた（図 4）．また，その際の IoU の値も 0.6536 から 0.7041 へと改善しており，擬似的にランダムな変更を施した場合と比較しても同等以上の精度であった（表 1）．

5. まとめ及び今後の予定

本稿では，電子顕微鏡画像セグメンテーションのタスクにおける深層学習の適用方法の 1 つとして，特定のハイパーパラメータを決定する手法を検討した．DCN の特性をふまえ，特徴マップの可視化結果に基づいてフィルタサイズを変更し，再学習を行うことで，セグメンテーション精度の向上がみられた．今後の課題として，現状では目視で行うに留まっている，可視化結果からフィルタサイズを決定する部分の自動化や，各種ハイパーパラメータ探索手法との比較が挙げられる．

また，本稿では深層学習適用時のハイパーパラメータ決定に着目したが，今後は連続切片画像の時系列性を考慮したモデルや，少量のデータを有効に活用するための方法についても検討する．インタラクティブセッションでは，それらの実験結果や議論についても併せて報告する予定である．

謝辞

本研究では，NVIDIA 合同会社様のご厚意により，NVIDIA Tesla M40 GPU を使用させていただきました．心より感謝致します．

表 1: テストデータに対する精度評価（データセット B）．k の値は左から順に，F1, F2, F3 に対応する逆畳み込み層のフィルタサイズを意味する．

フィルタサイズ	IoU	変更前との差
$k = 4, 8, 16$（変更前）	0.6536	
$k = 4, 8, 10$（提案手法）	0.7041	+0.0505
$k = 4, 8, 12$（提案手法）	0.6859	+0.0323
$k = 4, 8, 14$（提案手法）	0.6625	+0.0089
$k = 2, 8, 16$	0.6871	+0.0335
$k = 6, 8, 16$	0.7097	+0.0561
$k = 4, 6, 16$	0.5198	-0.1338
$k = 4, 10, 16$	0.6934	+0.0398
$k = 4, 8, 18$	0.6163	-0.0373
$k = 4, 8, 20$	0.6801	+0.0265
$k = 2, 4, 10$	0.6693	+0.0157
$k = 2, 4, 20$	0.6573	+0.0037
$k = 2, 8, 20$	0.7063	+0.0527
$k = 4, 10, 18$	0.6611	+0.0075
$k = 2, 6, 16$	0.6419	-0.0117
$k = 2, 8, 18$	0.6736	+0.02

参考文献

[1] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS computational biology*, Vol. 1, No. 4, p. e42, 2005.

[2] Verena Kaynig, Amelio Vazquez-Reina, Seymour Knowles-Barley, Mike Roberts, Thouis R Jones, Narayanan Kasthuri, Eric Miller, Jeff Lichtman, and Hanspeter Pfister. Large-scale automatic reconstruction of neuronal processes from electron microscopy images. *Medical image analysis*, Vol. 22, No. 1, pp. 77–88, 2015.

[3] I Arganda-Carreras, HS Seung, A Cardona, and J Schindelin. Segmentation of neuronal structures in em stacks challenge–isbi 2012, 2012.

[4] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pp. 2843–2851, 2012.

[5] Hao Chen, Xiaojuan Qi, Jie-Zhi Cheng, Pheng-Ann Heng, et al. Deep contextual networks for neuronal structure segmentation. In *AAAI*, pp. 1167–1173, 2016.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[8] Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pp. 179–187. Springer, 2016.

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

深層学習を用いた高解像度画像からの人物カウント

Counting People via High Resolution Images using Deep Learning

山田 佑亮 *1 Yusuke Yamada　河野 慎 *2 Makoto Kawano　米澤 拓郎 *2 Takuro Yonezawa　中澤 仁 *1 Jin Nakazawa

*1慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University

*2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

We propose the method which is counting people in high resolution and wide areas images via deep neural network. There are many areas where there are no people in the wide area image. We improve the imbalance of this dataset using sampling algorithm. Our experiments demonstrate the out method effectiveness.We think that this method contribute to tourist spot. Because we can measure crowdedness in tourist spot and analyze the ability to attract customers.

1. はじめに

第一次産業やスポーツなどへの IT の導入が進む一方で，公共施設や観光地への IT の導入が進んでおらず，様々な問題が生じている. その問題の一つに，観光地の混雑状況を計測することができないことが挙げられる．混雑状況を自動で計測することができれば，その観光地の集客力などを定量的に分析することが可能となる．現在の人物カウントのシステムは，交通量把握のために，単一の交差点や狭いエリアにおいて，物体検出手法を用いたモデルを用いて行われることが多い．だが，広域の観光地の混雑状況を計測する場合，混雑時に人が重なってしまったり，広域であるがゆえに人が小さく写ってしまうなど，検出をする上での大きな弊害が存在する．そのため，従来の物体検出手法で検出することができない．実際に，HOG 特徴量を用いた人物検出 [Dalal 05] を今回のデータセットにおいて適用させた結果が，図 1 であり，検出ができていないことが分かる．

本研究では，観光地の広域捉えた高解像度の画像から密度推定を用いて人物カウントを自動で行うシステムを提案する．Zhangh らは，密度推定を用いた Crowd Counting CNN において，画像内の一部をパッチとしてランダムにサンプリングして学習させている．しかし，広域の観光地の画像は，広域であるがゆえに，人が一人もいないエリアが多くを占めるデータである．そのため，ランダムにパッチをサンプリングして学習をしまうと，人がいるエリアの学習が安定しない．本研究においては，データの不均衡性に対処するために，サンプリングアルゴリズムによって高解像度の画像からパッチを抽出して学習させた．提案手法の有効性を検証するため，人物カウントの精度比較の実験を行った．データセットには，江ノ島の展望台で実際に撮影された海岸沿いの高解像度画像 (図 2) を用いたものであり，解像度は 4000×6000 である．実験の結果，提案手法が既存研究よりも高い精度を得ることができた．

連絡先: 山田佑亮，慶應義塾大学
〒 252-0882 神奈川県藤沢市遠藤 5322 デルタ棟 S213,
yamad@ht.sfc.keio.ac.jp

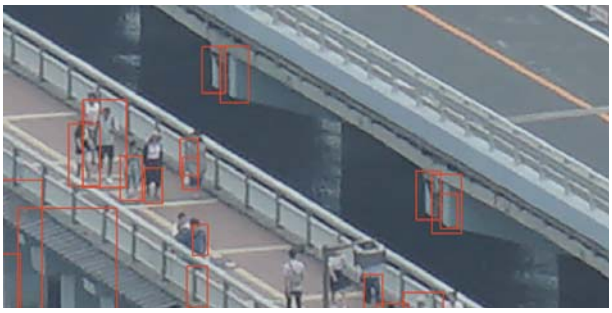


図 1: HOG 特徴量を用いた人物検出

図 2: データセットの例

2. 関連研究

人物カウント手法においては，Crowd Counting CNN[Zhang 15] という手法が挙げられる．画像を小さいパッチに分割し，そのパッチの中で，人物の密度マップを作成し，密度を積分すると画像内の人物の数になるように正規化する．それを CNN を用いた回帰モデルとして学習させる．予測時には，画像をパッチサイズに分割して，それぞれで得られた人物の数を合計する．このモデルの派生研究はいくつか存在する．Image Crowd Counting Using Convolutional Neural Network and Markov Random Field [Han 17] は，隣接するパッチは，密度に大きな差が存在しないという仮定を置き，パッチごとに密度マップを推定した後，その密度

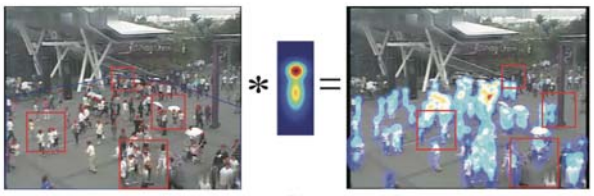


図 3: Crowd Counting CNN の概要図 [Zhang 15]

マップに対してマルコフ確率場により隣接するパッチの平滑化を行うことで，パッチ間の関係性を踏まえた人物カウントを行うことができるようになる．もう一つの派生研究として，Switching Convolutional Neural Network for Crowd Counting[Sam 17] がある．これは，分割したパッチを 3 種類に分類する Switching Network を学習させ，分類した後にそれぞれ違うアーキテクチャのネットワークで密度推定を行うというモデルである．これらの手法は密度推定による人物カウントにおいて，高い精度を出している．一方で，これらの研究で扱われている画像は，撮影しているエリアが狭域であるが，今回の研究におけるデータセットは高解像度で広域な画像であり，人がいないエリアと人がいるエリアのデータ量が不均衡である．そのため，既存手法をそのまま使用しても，高い精度を期待できない．

3. アプローチ

本研究で扱うデータは，既存研究のデータセットが捉えているエリアより広域であり，人がいない領域が大きい．そのため，今回扱う画像は不均衡データであると言え，学習時に，データサンプリング手法を用いる必要がある．不均衡データにおけるサンプリングの手法として，under-sampling[Kubat 97] が挙げられる．これは，クラス間でデータ数が不均衡である場合，データ数が多いクラスが，学習に使うデータ数を減らし，データ数が少ない方に数を合わせるという手法である．over-sampling[Chawla 02] は，データ数が少ない方のデータを増やし，データ数が多いクラスと同じ数にすることで，不均衡性に対処するという手法である．今回は，これらを組み合わせたサンプリング手法を用いる．

3.1 Crowd Counting CNN

本研究は，Crowd Counting CNN と同じく，密度推定を用いて人物カウントを行う．まず，学習用のデータを用意する．学習用のデータは，72×72 のパッチごとに作成する．学習用の画像全てにおいて，人物の頭の座標点のみアノテーションを行い，この座標点を基にガウス密度マップを生成する．任意のピクセル p の密度は，次の式によって算出される．

$$D(p) = \sum_{P \in P_i} \frac{1}{\|Z\|} (\mathcal{N}_h(p; P_h, \sigma_h) + \mathcal{N}_b(p; P_b, \Sigma)) \quad (1)$$

ここで，教師データにおける画像内の人の数と，密度マップの合計を等しい値にするために，正規化定数である $\frac{1}{\|Z\|}$ を使用する．P_h はアノテーションされた頭の座標点であり，P_b は体の座標点を示す．体の座標点は，人の身長を 1．75m，その画像内における 1 ピクセルあたりの長さ (m) を M とすると

$$P_b(x, y) = (P_h(x), P_h(y) + 1.75(M/2)) \quad (2)$$

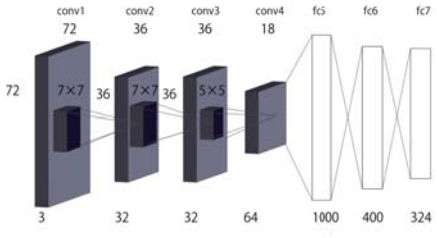


図 4: 提案手法のネットワーク構造

として計算される．この計算式は，頭の座標点から作成されるガウスカーネルの項と，体の座標点から作成されるガウスカーネルの項で構成され，結果として，図 3 で示されたヒートマップができる．また，σ の値は，σ_h を $0.2M$ として計算する．Σ の値は，σ_x を $0.2M$ として計算し，σ_y を $0.5M$ として計算する．

M の値は，画像内の特定の一人が y 方向に占めているピクセル数をアノテーションし，それを p として，

$$M = p/1.75 \quad (3)$$

として計算する．

この計算を，パッチの各ピクセルごとに行い，密度マップを生成する．これを教師データとして，CNN を学習させる．

3.2 ネットワーク

CNN は，72×72 の画像を受け取り，18×18 のサイズの密度マップを出力する構造である．具体的な構造は図 4 に示す．各層には，活性化関数として ReLU 関数を用いる．また，畳み込み層のあとには，ストライドが 2 の MaxPooling 層を用いている．ネットワークの最適化には，密度マップの各ピクセルの二乗誤差 L_D と，密度マップを合計して算出される人数の二乗誤差 L_C を使用し，

$$L = NL_D + L_C \quad (4)$$

で算出される．ここで，N は二つの誤差関数を調整するためのハイパーパラメータである．

3.3 サンプリング

先行研究においては，72×72 のパッチサイズに分割し，その中で人物検出を行なっていた．だが，今回取り扱う画像は，前述した通り，人がいない領域が広い．そのため，ランダムにパッチを作成した場合，ほとんどの場合に人がいないエリアがサプリングされるため，人がいるデータに対する学習がうまくいかない可能性がある．そこで今回は人がいるエリアといないエリアのデータ量を制御する確率 λ を導入し，λ をもとにデータをサンプリングする．画像はそれぞれ 72×72 の大きさでサンプリングされる．パッチが生成される際に 0 ～ 1 の間で乱数を発生させ，それが 0．7 以上であった場合，人がいる場所の中でランダムに一箇所選び，72×72 の大きさでサンプリングする．0．7 未満であった場合，画像内からランダムに 72×72 の大きさでサンプリングする．このとき，人がいる場所がクロップされる場合もある．このようにしてサンプリングされる箇所を制御して学習を行う．

4. 実験

今回の実験においては，サンプリングアルゴリズムを用いない Crowd Counting CNN と，サンプリングアルゴリズム

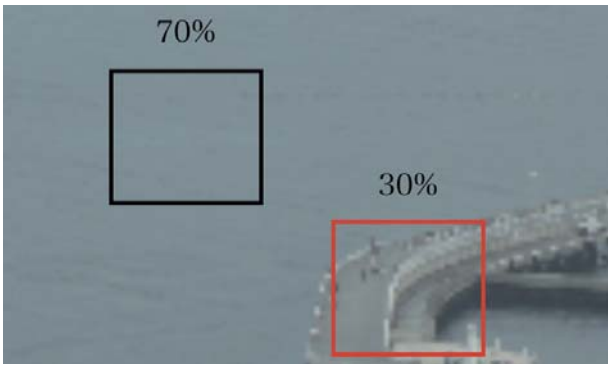


図 5: サンプリングの例

を導入して学習させた提案手法を比較し，精度の変化を見た．最適化手法は SGD を使用し，学習率は 0．01，エポック数は 7000 回とした．また今回は，バッチ処理で学習を行った．誤差関数において，$N = 10$ とし，サンプリングの確率において，$\lambda = 0.7$ を設定し，18 枚の 6000×4000 のデータセットに対して 6 枚のテストデータで K 分割交差検証法を行った．データセット一枚あたりの人物の数の平均は，132 人であった．

4.1 実験結果

表 1 に実験結果を示す．評価は，人物の数の正解値と，予測値の絶対二乗誤差の平均と絶対値誤差の平均の二種類で行う．結果，サンプリングアルゴリズムを導入したモデルの方が，既存のモデルの精度を大きく上回った．6000×4000 の画像を予測させるためには，CPU で 39.5 秒の時間がかかった．また，K 分割法における，人数の誤差の分散は 384.9 であり，標準偏差に直すと 19.6 であった．

4.2 考察

今回の実験では，サンプリングアルゴリズムを導入したモデルが，既存手法の精度を大きく上回った．実際，サンプリングアルゴリズムを導入していないモデルの予測結果の平均は，16．8 人，サンプリングアルゴリズムを導入したモデル (λ=0.7) の予測結果の平均は 163.1 人であった．これに対し，正解データの平均は 132 人であった．この結果より，サンプリングによってデータを操作することで，この高解像度のデータに対する予測値が大きく変化することがわかる．今回は，サンプリングアルゴリズムを導入したことによって不均衡なデータに対して学習をすることができたため，既存手法を大きく上回ることができたと考えられる．しかし，今回は平均で 31 人の誤検出をしてしまっている．今回のデータセットは，人であるかどうかわからないようなものが写り込むこともあるほど見分けがつかないものが多く，アノテーションの段階ではそのようなデータを全て抜いているので，正解データより多くカウントしてしまっている可能性も考えられる．

一方，予測にかかる時間は長い．これは，モデルの学習を 72×72 のパッチサイズで行ったため，6000×4000 の画像を分割してそれぞれを計算するという処理に時間がかかったものと考えられる。

5. 今後の課題

5.1 λ の探索

今回は，$\lambda = 0.7$ においてのみ学習をしたのだが，この λ はハイパーパラメータであり，この λ の値をグリッドサーチで探索することによって，より高い精度が得られる可能性もある．

表 1: 実験結果

手法	MAE	MSE
提案手法	30．5 ± 19.6	1315.1 ± 384.9
Crowd Counting CNN	115.8 ± 7.8	14403.8 ± 61.4

5.2 マルコフ確率場の導入

先行研究でも取り上げた，Crowd Counting CNN への Markov Random Field の適用 [Han 17] は，今回のデータセットにおいても有効な手法であると考えられる．今回のデータセットにおいてはビルや海のエリアなどが存在するため，これらのエリアの一つのパッチにおいて大きな誤検出が起きた場合に，隣接するピクセルが誤検出をしなければ，誤検出のパッチは平滑化される．つまり，人がいないエリアが非常に大きい今回の画像において，マルコフ確率場による平滑化は誤検出を抑える手法として非常に有効である．

6. まとめ

観光地で混雑状況を計測し，集客力などを定量的に分析するために，広域の人物カウントシステムは非常に価値のあるものである．本研究では，Crowd Counting CNN にサンプリングアルゴリズムを加えることによって，高解像度の観光地の画像から人物をカウントするための手法を提案した．この手法を用いることで，既存研究を大きく上回る精度を出すことができ，実際にこのシステムを観光地に導入することで，混雑状況を正確に知ることが期待できる．

参考文献

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

[Dalal 05] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 886–893IEEE (2005)

[Han 17] Han, K., Wan, W., Yao, H., and Hou, L.: Image Crowd Counting Using Convolutional Neural Network and Markov Random Field, *arXiv preprint arXiv:1706.03686* (2017)

[Kubat 97] Kubat, M., Matwin, S., et al.: Addressing the curse of imbalanced training sets: one-sided selection, in *ICML*, Vol. 97, pp. 179–186Nashville, USA (1997)

[Sam 17] Sam, D. B., Surya, S., and Babu, R. V.: Switching convolutional neural network for crowd counting, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, p. 6 (2017)

[Zhang 15] Zhang, C., Li, H., Wang, X., and Yang, X.: Cross-scene crowd counting via deep convolutional neural networks, in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 833–841IEEE (2015)

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

3D 仮想環境 Minecraft における深層強化学習への視線方向の影響

Effect of Viewing Directions on Deep Reinforcement Learning in 3D Virtual Environment *Minecraft*

松井　太樹　　小山　聡　　栗原　正仁
Taiju Matsui　　Satoshi Oyama　　Masahito Kurihara

北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

In recent years, deep reinforcement learning has attracted interests from AI researchers. Deep reinforcement learning is a method combining a deep neural network (DNN) and reinforcement learning (RL). By approximating a function in RL with a DNN, it enables an agent to learn in complex environment represented by low-level features, such as pixels given by a 3D video game. However, learning from low-level features is sometimes problematic. For example, a small difference in input pixels results in completely different behaviors of an agent. In this study, as an example of such problems, we focus on viewing directions of an agent in 3D virtual environment Minecraft and analyze the effect of them on the efficiency of deep reinforcement learning.

1. はじめに

近年, 人工知能 (AI) 技術は深層学習 (Deep Neural Network; DNN) により, 画像などの高次元データに対する情報処理能力を飛躍的に進化させた. これによりゲーム AI の分野においても, 画像データを用いてゲームの学習を行う Visual Learning の分野が開拓されてきた. 特に最近では google DeepMind によって提案された手法 Deep Q-Network (DQN) [Mnih 13] [Mnih 15] によって映像のみを入力とした学習で Atari2600 という簡単な 2D 環境のゲームにおいて人間と同等以上のスコアを獲得するに至った. また, 同チームは 2016 年に非同期な学習方式を提案し [Mnih 16], その内の一つである Asynchronous Advantage Actor-Critic (A3C) では必要スペックや計算量, 獲得した行動において DQN を圧倒する結果を記録している.

これらのアルゴリズムは深層強化学習と呼ばれている. その名の通り, 深層学習を用いた画像処理と試行錯誤から学習を行う強化学習を組み合わせた手法である. 深層学習を用いることで環境の状態を画像で定義することが可能となり, また深層学習の出力で強化学習の出力を近似することで, 少し工夫を加えた上で強化学習の手法を用いることができる. しかし, 画像データは状態としてはかなり複雑である. RGB の画像で考えると「画像の高さ」×「画像の幅」× 3(RGB) × 256(0 ∼ 255) もの表現が可能である. 実際に学習を行う環境においてはある程度のパターンに制限されることがほとんどであるが, 既存の表現に比べてこれが複雑であることに変わりはない. さらに 3D 仮想環境において一人称視点で学習を行う場合, 視界に入る情報が制限され自身の置かれた状態を完全には知ることができない. また画像を入力とした深層強化学習では, いくつかの画素の色や視点の少しの変化による入力画像の差が行動に大きな影響を与えてしまうことがよく起こる. この 3D 仮想環境における一人称視点からの行動獲得という問題を解決するためには, この複雑さを克服したアルゴリズムを考案することが必要とされる.

そこで本研究では, アルゴリズム考案の前段階として, 入力情報の差を視野の仰角の変化に限定し, 3D 仮想環境で自由度の高い環境設定と問題設定を行える Minecraft を用いて, 学習過程や学習結果にどのような変化がみられるか調査することを目的とする.

2. 関連研究

この章では, 本研究の関連研究として深層強化学習と利用環境について述べる.

2.1 深層強化学習

深層強化学習は深層学習の出力で強化学習の方策や価値関数を近似するアルゴリズムである. これにより, 深層学習と強化学習を組み合わせ, 画像のような複雑な環境を学習することが可能となった. しかし, 方策や価値関数をニューラルネットワークのような非線形近似器を用いて表現して行う強化学習は安定しないことが知られている. これは強化学習に用いる時系列に並んだデータと方策, 価値関数の間に相関があることが原因と知られている. 深層強化学習の手法である DQN と A3C はそれぞれ異なる方法でこの問題を解決している.

2.1.1 Deep Q-Network

DQN では学習の安定のために 2 つの主な工夫を採用している. まず 1 つは経験再生 (Experience Replay, ER) である. これは過去一定期間においてある時間の「状態, 行動, 報酬, 次の状態」のタプルを経験として ER メモリに保存し, 一定周期で ER メモリからランダムにサンプリングしたミニバッチデータを用いて学習を行う手法である. これによりデータとの相関を打ち消すことができるが, 過去の方策に基づいたデータも学習に用いられるため方策オン型のアルゴリズムが利用できない点, ER メモリに相当な容量が必要となる点に注意が必要である.

またもう 1 つの工夫は固定目標ネットワーク (fixed target network) である. DQN における学習は一定期間のデータ収集の度に行われるが, この際に少しの価値関数の変化により方策が大きく更新されてしまうことがあり学習が安定しない. これを回避するために一度の学習の間は最適化するパラメータ θ のコピー θ^- を用いて最適化を固定することで学習を安定化させる.

2.1.2 Asynchronous Advantage Actor-Critic

A3C ではデータの相関を打ち消すために複数のスレッド上で並列にシミュレーションを行い学習用データを収集する. これにより, DQN においてランダムにサンプリングしたデータで学習を行ったことと同様の結果を得ることができる. それぞれのスレッドの動作は次のようになる.

1. global network のパラメータ θ を自身の local network のパラメータ θ^- にコピーする.
2. 現在のパラメータに従い探索を行い, 各時間ステップについて TD 誤差により計算される誤差関数から勾配 $d\theta$ を計算し累加する.
3. 一定の探索期間の後, 勾配 $d\theta$ を global network に伝え, 勾配法に従い θ を更新する.
4. $d\theta$ をリセットし 1. に戻る.

以上の処理を非同期並列に行うことでデータの相関を打ち消して学習を行う. ここで 1 においてパラメータをコピーすることは DQN の固定目標ネットワークと同様の働きをする.

2.2 Minecraft (Project Malmo)

Minecraft は Microsoft より販売されているサンドボックスゲームである. サンドボックスゲームとはビデオゲームのジャンルの 1 つであり, ゲームにより定められた目標が無く, 自由度の高い環境で自由に目標を定められることが特徴である. 特に Minecraft は主に立方体のブロックで構成される点を除き, 現実の空間と非常に近い環境を提供することができることが魅力である. この環境は次のような特徴を有する.

- エージェントは重力の影響を受けるが, 一部を除いたブロックは重力に影響されない.
- 条件により複数種の敵性生物が自然発生する.
- 時間の概念があり, 空間の明るさが変化する.
- バイオーム (生物群系) の概念があり, バイオームごとに違った地形, ブロック, 生物が自動で配置される.

また環境内において, エージェントは次のことが可能である.

- ほぼ全てのブロックは攻撃し続けることで破壊し, 入手することができる (一部専用のアイテムが必要).
- 入手したブロックは任意のブロックに隣接したスペースに再配置することができる.
- 入手したブロックは組み合わせることで別のブロックまたはアイテムを作成することができる (クラフティング).
- 作成したアイテムを使用する.

また 2016 年, Microsoft Research は Minecraft を用いた人工知能開発を目的としたプラットフォーム Project Malmo [Johnson 16] を GitHub にて公開した. これにより Minecraft の環境と実装したエージェントとの相互作用が容易になり, 以前は困難であった Minecraft の環境を簡単に利用することが可能となった.

3. 実装手法

この章では, 本研究の実験において実装した手法について述べる.

3.1 Asynchronous Advantage Actor-Critic

Asynchronous Advantage Actor-Critic(A3C) は深層強化学習の手法であり, 前章で述べた非同期な学習方式として提案された手法の 1 つである. 当時の最先端手法であった DQN を圧倒する結果を記録した [Mnih 16]. 現在においても, 3D 環境の探索タスクで最高水準の手法である. 非同期な学習方式については前節で説明したため, ここでは Advantage Actor-Critic について述べる.

Actor-Critic 手法は方策を定める Actor と状態を評価する Critic からなる強化学習の手法である. 深層強化学習においては, これらは CNN の出力で $\pi(s,a) \approx \pi(s,a;\theta), V(s) \approx V(s;\theta_v)$ と近似的に表現される. ここで, θ, θ_v は CNN のパラメータである. これらのパラメータは REINFORCE アルゴリズム [Willia 92] の勾配法を用いて最適化されるが, この際 Advantage Actor-Critic では現在の状態推定に Advantage 関数を用いる. Advantage 関数は次の式で表される.

$$A_t = \sum_{i=0}^{k-1} \gamma^i r_{t+1} + \gamma^k V(s_{t+k};\theta_v) - V(s_t;\theta_v) \tag{1}$$

Advantage 関数を用いることで, Bellman 方程式より計算された TD 誤差よりも先のステップを考慮に入れることができるようになる. これを用いて勾配は次のように計算される.

$$d\theta \leftarrow d\theta + \alpha \nabla_\theta \log \pi(s_t,a_t;\theta) A_t + \beta \nabla_\theta H(\pi(s_t;\theta)) \tag{2}$$

$$d\theta_v \leftarrow d\theta_v + \alpha \nabla_{\theta_v} A_t^2 \tag{3}$$

ここで式 (2) の第 3 項は確率の正規化のために加算されたエントロピーである.

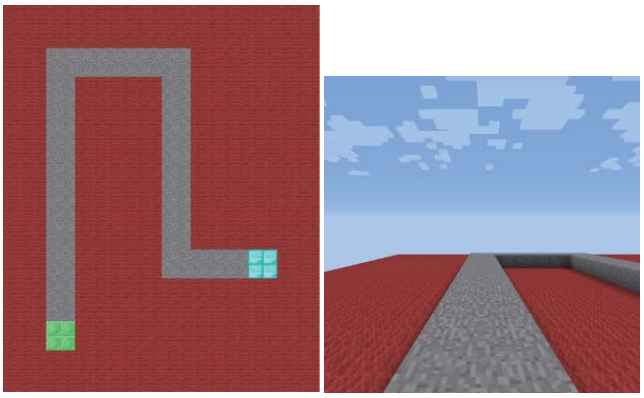
4. Minecraft による実験

4.1 実験目的

本実験では, 3D 仮想環境である Minecraft を用いた問題の学習において, 視野の仰角の変化によってどのような学習過程, 獲得行動の変化が見られるか調査を行うことを目的とする.

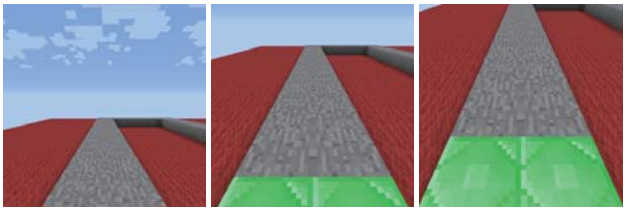
4.2 問題設定

この節では, 実際に Minecraft を用いた実験のエージェントと環境の設定について詳細を述べる. 本実験では視野の仰角の変化以外の変化を極力取り除くために簡単な問題設定として「一人称視点で一本道を落ちずにゴールまで進む」というタスクを用意した. 1 回のエピソードは, ゴールするか, 通路から落下するか, または制限時間に達することで終了し, Advantage 関数をスコアとして記録した. 環境は曲がり角を持つ一本道であり, エージェントは環境から一人称視点の RGB 画像 ($84 \times 84 \times 3$) を状態として受け取る (図 1). エージェントは状態から「前進する, 左を向く, 右を向く」を組み合わせた行動群から行動を確率的に選択し行動に対する結果として報酬を受け取る. 報酬は, ゴールした際に 1.0, 通路から落下した際に-1.0, またその場で回り続ける行動を獲得しないように, Minecraft の 1 ブロック分通路を進むたびに 0.1 の報酬を与えた. 実験ではこの問題をエージェントの視野の仰角を「$0°, -30°, -45°$」と変化させ, それぞれ 200 万ステップ (1 ステップはゲーム内時間で約 0.1 秒) 学習させた (図 2). 視点によって問題の根幹である通路について知ることのできる情報の量が変化するため, 調査内容にも変化が現れることが予想される.



(a) 環境の俯瞰図　　(b) エージェントの視点

図 1: 学習に用いた環境



(a) 仰角 0°　　(b) 仰角 −30°　　(c) 仰角 −45°

図 2: 視野の仰角を変化させた際の視点の変化

4.3 実験結果

4.3.1 学習過程への影響

まず 200 万ステップの学習において, 各 1 万ステップ毎にエピソードのスコアの平均の推移を調べた結果, 図 3 のようになった. より多くの通路に関する情報を視界に納めたエージェントほど, 早い段階で問題に適した行動を獲得していることが分かる.

4.3.2 獲得行動への影響

次に学習の結果獲得した行動について各視点のエージェントに 1000 回ずつエピソードを割り振り, 平均スコアとゴール到達を 100 とする平均進行度, ゴール到達率を調査した. 結果は図 4 のようになり, これもまた視界内における通路の情報の量に大きく影響を受けた結果となった. 実験 4.3.1 と比べてもおよそ対応した結果になっていることが見て取れる.

4.3.3 落下位置の比較

それぞれのエージェントについて落下位置についてプロットした結果, 図 5 のようになった. まず仰角 0° のエージェントは図 5 からもわかるように学習が不十分であり, 初めの直進という簡単な行動も学習する事ができていない. 仰角 −30°, −45° のエージェントに関してはゴールに到達可能なだけの行動を獲得している. しかし, これらのエージェントでも一定の割合で落下してしまっている. これは離散的かつ「する」か「しない」かの二値化された行動により, 曲がり角付近では向いている方向によって曲がり角を十分に認識できていないためであると考えられる.

4.4 考察

実験 4.3.1 より, 学習に関して視点の変化により学習の速度に大きな影響を与えることが伺える. 実験 4.3.2 に関しても獲得した行動に視点の変化による大きな影響があることがわかった. これらの結果より, 一人称視点の 3D 仮想環境の探索タスクを適切に学習するためには, 学習したい問題に関してより多くの情報が視界に含まれていることが望ましいとわかる. また実験 4.3.3 より, 視点を与えるだけでなく, 視点や行動を連続的または細かく多値化して制御することの必要性が考えられる. これらの実験結果より 3D 仮想空間における一人称視点での AI 開発のためには学習する行動に「視点の制御」を加えるべきであると考察する.

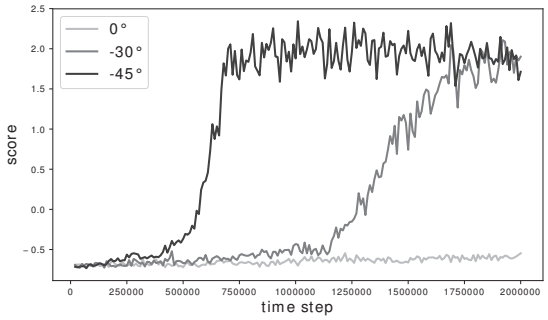


図 3: 学習ステップと平均スコアの遷移

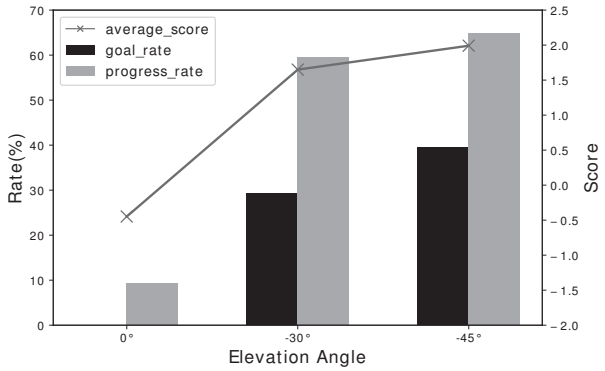


図 4: 1000 回の試行によるゴール率, 進行率, 平均スコア

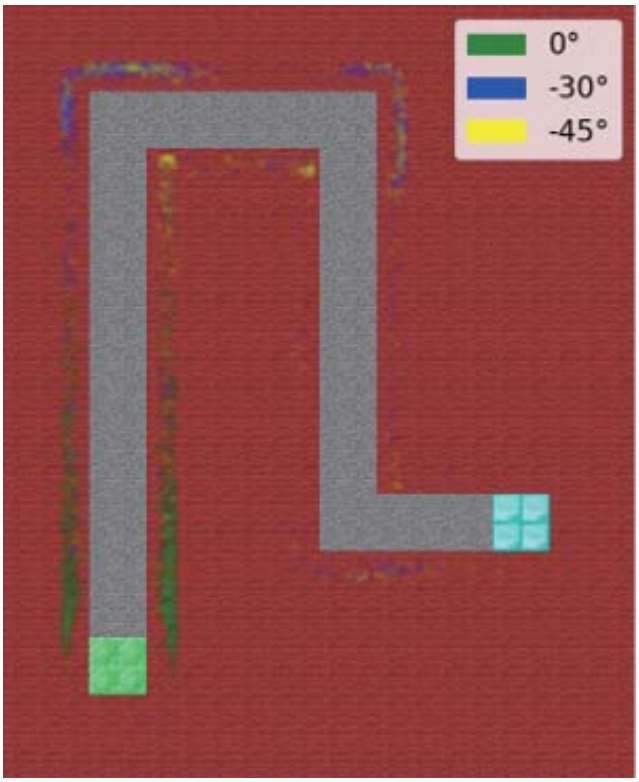


図 5: 視点の変化による落下位置の変化

5. まとめ · 今後の展望

本論分では, 強化学習を映像データ入力に対応させた深層強化学習の手法を紹介し, その手法の 1 つである A3C を用いて, 3D 仮想環境における一人称視点の行動獲得タスクの学習に視点の仰角の変化が与える影響を調査することが目的であった. その点において, 複数の視点に対して学習過程と学習結果について比較し調査することができたため, この目的は達成できた.

今後の展望として,「視点の制御」を加えた深層強化学習アルゴリズムの開発を行いたいと考えている. A3C アルゴリズムが 2016 年に発表されて以降, 様々な工夫を付け加えた手法が提案されている. その 1 つである自己教師付き予測による「好奇心」に導かれた探索を用いた研究 [Pathak 17] では, 行動の結果として環境から与えられる外部報酬とは別に, 入力である画像データから内部報酬を計算している. この内部報酬という考え方は 3D 仮想環境における深層強化学習及びその視点の制御において, 視点を評価するなど有効に働き得ると考えている. 今後はこの方法を第一に視点を制御するアルゴリズムを提案することを目標とする. 以上を今後の展望として本論分を締めくくる.

参考文献

[Johnson 16] Johnson, M., Hofmann, K., Hutton, T., and Bignell, D.: The malmo platform for artificial intelligence experimentation, in *IJCAI International Joint Conference on Artificial Intelligence*, Vol. 2016-Janua, pp. 4246–4247 (2016)

[Mnih 13] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M.: Playing Atari with Deep Reinforcement Learning, in *arXiv preprint arXiv:1312.5602*, pp. 1–9 (2013)

[Mnih 15] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D.: Human-level control through deep reinforcement learning., in *Nature*, Vol. 518, pp. 529–33, Nature Publishing Group (2015)

[Mnih 16] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., and Kavukcuoglu, K.: Asynchronous Methods for Deep Reinforcement Learning, in *International Conference on Machine Learning*, Vol. 48, pp. 1928–1937 (2016)

[Pathak 17] Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T.: Curiosity-Driven Exploration by Self-Supervised Prediction, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Vol. 2017-July, pp. 488–489 (2017)

[Willia 92] Willia, R. J.: Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning, in *Machine Learning*, Vol. 8, pp. 229–256 (1992)

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

畳み込みニューラルネットワークによる補助線検出を用いた歪みを含む2次元コードの復号方式

A Decoding Method for Distorted Two-dimensional Barcodes Using Auxility Line Detection by Convolution Neural Network

鞍津輪 一希 [*1] Kazuki Kuratsuwa　神薗 誠 [*1] Makoto Kamizono　川崎 洋 [*2] Hiroshi Kawasaki　小野 智司 [*1] Satoshi Ono

[*1]鹿児島大学大学院 理工学研究科 情報生体システム工学専攻
Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University

[*2]九州大学大学院システム情報科学研究院 情報知能工学専攻
Department of Advanced Information Technology, Graduate School of Information Science and Electrical Engineering, Kyushu University

Two-dimensional (2D) codes are widely used for various fields such as production, logistics, and marketing thanks to their larger capacity than one-dimensional barcodes. However, they are subject to distortion when printed on non-rigid materials, such as papers and clothes. Although general 2D code decorders correct uniform distortion such as perspective distortion, it is difficult to correct non-uniform and irregular distortion of the 2D code itself. This paper proposes a 2D code involving monochrome auxiliary lines that is robust against non-uniform, local distortion and its decode method that uses Convolution Neural Network for detection of auxiliary lines.

1. はじめに

2 次元コードは，従来の 1 次元コードよりも多くの情報を格納できることから生産，物流，販売等の幅広い分野において活用されている．また，2 次元コードの一種である QR コードは，発見が容易となるよう位置検出パターンを備えており，高速読み取りが可能で，汚損や欠損が生じた際に訂正や補完を行う機能を備える．このため 2 次元コードは，斜め方向から撮影を行った場合や円柱形状の物体に印刷された場合でも一般的なデコーダで復号を行うことができる．しかし，2 次元コードを印刷した紙や布などに非均一な凹凸が発生し，2 次元コード自体に複雑な歪みが生じた場合，復号することは困難である．

複雑な歪みの復号方式として Kamizuru らの先行研究では，カラー補助線を用いた 2 次元コードの復号方式が提案されている [Kamizuru 2015]．この方式では，あらかじめ QR コードに格子状の着色された補助線を付加し，補助線をもとに復号を行う．補助線の交点の位置を信頼度伝搬法で認識し，各セルの明暗を取得することで QR コードを復元し，デコーダによって復号する．この補助線は，デブルーイン系列の色パターンによって構成されており，連続する補助線 3 本を参照することで交点の局所的な位置推定を行い，その結果，局所的なセルの位置推定を行うことができる．そのため，局所的な歪みや遮蔽領域に頑健な復号方式となっているが，カラー画像を用いているため照明環境の影響を受けやすいことや印刷のコストがかかるといった問題がある．

また，白黒補助線を用いた復号方式も提案されている [Kamizuru 2017]．この方式では，あらかじめ QR コードに格子状の白黒二値の補助線を付加しており，補助線をもとに復号を行う．カラー補助線を用いた復号方式と同様の手順で QR コードを復元し，デコーダによって復号する．白黒二値の画像を用いているため，カラー補助線と比べると照明環境の変化に影響を受けにくく，印刷のコストを抑えることに成功している．しかし，補助線が白黒二値の単一パターンで構成されるため，補助線を用いた局所的なセルの位置推定が行えない．そのため，激しい歪みにより自己遮蔽が生じると復号が困難になる．

本研究では，上記の先行研究の問題の解決策として，複数種類の白黒二値の補助線を用いた復号方式を提案する．これら複数の補助線を検出するために画像認識に強い深層学習の畳み込みニューラルネットワーク (CNN) を用いる．これにより，[Kamizuru 2015] と同様に連続する補助線 3 本を用いた局所的なセルの位置推定を可能にし，遮蔽領域への対応を行う．また，白黒二値の画像を用いることによって照明環境の変化に影響を受けにくい復号方式を提案する．

連絡先: 鞍津輪一希，鹿児島大学大学院 理工学研究科 情報生体システム工学専攻，〒 890-0065 鹿児島市郡元 1-21-40，sc113017@ibe.kagoshima-u.ac.jp

2. 研究分野の概要

2.1 一般的なデコーダによる QR コードの復号

一般的な QR コードのデコーダは，軽度の幾何学的な歪みであれば射影変換などの単純な画像処理によって歪みを補正し，復号することが可能である．QR コードを復号する場合，デコーダは，撮影画像内から QR コードの 3 つの頂点に存在する位置検出パターンを検出し，位置検出パターンをもとにして図 1 に示すようなグリッドラインを作成する．作成されたグリッドラインの交点をサンプリング点として，各サンプリング点の座標の明暗を取得することで QR コードの復号を行う．しかし，QR コード上に歪みが発生している場合，図 2 に示すように実際のセルの位置とサンプリング点の座標に差異が生じてしまい，セルの明暗を正しく取得することが困難となる．このため，一般的なデコーダでは，複雑な歪みを含む 2 次元コードの復号は困難である．

2.2 関連研究

軽度の幾何学的な歪みを射影変換などの画像処理によって補正する方式や緩やかに湾曲した QR コードを復号する方式が提案されている [Iida 2000, Mizumoto 2004]．しかし，これらは，幾何学的な画像変換で補正不可能な歪みや遮蔽領域には対応していない．

幾何学的な画像変換で補正不可能な歪みや遮蔽領域に対応した復号方式として，先行研究ではカラー補助線を用いた復号方式が提案されている [Kamizuru 2015]．この方式では，図

図 1: グリッドライン

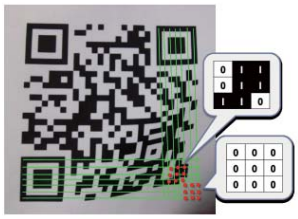


図 2: 歪みにより失敗した例

図 3: 従来方式で使用するカラー補助線入り QR コード

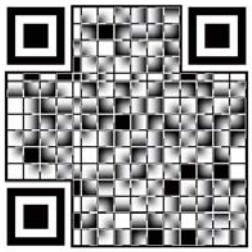


図 4: 従来方式で使用する白黒補助線入り QR コード

3 のような着色された補助線を付加した QR コードを使用しており，補助線は縦線と横線それぞれ連続する補助線 8 本を 1 周期とするデブルーイン系列 [Fredricksen 1977] に基づく色パターンで構成されている．縦線は赤 (R) と黄色 (Y) から構成され，Y-R-R-R-Y-R-Y-Y の周期的なパターンを持つ．横線は青 (B) とシアン (C) から構成され，C-B-B-B-C-B-C-C の周期的なパターンを持つ．デブルーイン系列により連続する補助線 3 本を参照することで補助線の位置推定を可能とする．付加される補助線は各セルの中心点への影響を避けており，既存のデコーダによる可読性を保つ．補助線の交点の ID 割り当てを信頼度伝搬法 [Frey 1998] を用いて行い，各セルの明暗を取得することで復号を行う．複雑な歪みや遮蔽領域にも対応した復号方式であるが，着色された補助線を使用しているため，いくつかの問題が挙げられる．まず，照明環境の問題である．補助線の色情報を取得する必要があるため，撮影時の照明環境によっては補助線の検出が困難となる．また，補助線の色を正しく取得するために事前に撮影するカメラと印刷するプリンタの色校正を行う必要もある．さらに，カラー画像を用いているため，印刷のコストが高くなってしまう．

また，他にも白黒補助線を用いた復号方式が先行研究として提案されている [Kamizuru 2017]．この方式では，図 4 に示すような単一パターンの白黒二値の補助線を付加した QR コードを使用しており，信頼度伝搬法を用いて補助線の交点に ID を割り当て，各セルの明暗を取得することで復号を行う．局所的な歪みに対応しており，白黒画像を用いているため，照明環境の変化に影響を受けにくい復号方式となっているが，単一パターンな補助線のため，カラー補助線のような局所的な補助線の位置推定が行えず，遮蔽領域に弱いといった問題がある．

3. 提案方式

3.1 研究の基本方針

本研究では，以下のアイデアに基づき歪みを含む 2 次元コードの復号方式を提案する．提案する方式では，既存のデコーダの可読性を保つ白黒二値の補助線を付加した QR コードを用いる (図 5)．本研究で使用する補助線および交点を図 7 に示す．QR コードにあらかじめ付加された補助線を検出し，補助線をもとに各セルの位置を推定することで復号を行う．この補助線は縦横それぞれ 2 種類で構成されており，縦線および横線は連続する補助線 8 本を 1 周期とするデブルーイン系列の周期的なパターンを持つ．これにより，連続する補助線 3 本を参照することで局所的に補助線の位置の推定ができ，その結果，局所的なセルの位置推定を行うことができる．

複数種類の補助線を検出および識別することは，単純なフィルタでは難しいため，本研究では，補助線の検出に画像認識で高い性能を発揮している CNN を用いる．

図 5: 提案方式で使用する白黒補助線入り QR コード

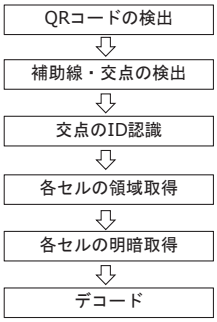


図 6: 処理手順

(i) 縦線 1 (ii) 縦線 2 (iii) 横線 1 (iiii) 横線 2 (iiiii) 交点

図 7: 使用する補助線と交点

本研究では，歪んだ QR コード画像を生成するためのシミュレータを構築する．これにより，大量な学習データを必要とする CNN の学習や方式の評価実験を行えるようにする．

3.2 処理手順

提案する方式の処理手順を図 6 に示す．以下に各処理について示す．

1. QR コードの検出
   撮映画像内から QR コードを検出する．QR コードの検出には Star 検出器を用いる．Star 検出器は Center Surround Extermas(CenSurE)[Agrawal 2008] から派生した特徴点検出器である．

2. 補助線および交点の検出
   入力画像から CNN で構築した検出器により補助線および交点の検出を行う．このとき，CNN は，入力画像をスライド窓で切り取ることによって生成された画像を入力データとし，その中心画素が補助線であるか，交点であるか，その他の領域であるかを認識し，クラス分類を行う．クラス分類の結果より補助線と交点の領域がそれぞれ異なる色で着色された画像を生成し，入力画像との合成を行う．補助線の縦線は黄と赤，横線は青と水色，交点はマゼンタに着色する．
   次に，生成された画像から白および黒のピクセルを除い

た領域を，補助線領域として抽出する．補助線抽出には HSV チャンネルのみ使用する．白領域の抽出は S チャンネルのみの二値化で行い，黒領域の抽出は V チャンネルのみの二値化で行う．
そして，抽出した補助線領域を RGB 成分を用いて縦線と横線に分ける．赤成分の強い領域を縦線領域，青成分の強い領域を横線領域とする．赤成分，青成分ともに強い領域は交点領域とする．縦線領域と交点領域を合わせて縦線，横線領域と交点領域を合わせて横線とし，それぞれオープニングとクロージングを行いノイズを除去する．このとき，縦線に関しては，縦方向に 3 ピクセルかつ横方向に 1 ピクセル，横線に関しては，縦方向に 1 ピクセルかつ横方向に 3 ピクセルの膨張・収縮を行うことで，効率よくノイズを除去する．
最後に，検出した縦線および横線の色を識別する．まず，縦線および横線からマゼンタ領域を除きラベリングを行う．ラベリングされた各領域の RGB 成分より色の識別を行う．検出した縦線および横線の色を識別した後，最終的に交点をノードとして抽出する．

3. 交点の ID 認識
検出された交点に ID を割り当てていく．交点の ID 認識には先行研究 [Kamizuru 2015] の処理を用いる．交点の ID 認識問題は MRF，即ち同時確率分布のグラフィカルモデルのラベル割り当て問題として定式化することができる．MRF において検出された交点をノードとし，二つのノードを結ぶ補助線をエッジとする．ノードのラベルは縦線 ID と横線 ID のペアとし，MRF のエネルギー関数 $E(\boldsymbol{x})$ は以下のように定義する．

$$E(\boldsymbol{x}) = \sum_{p \in \mathcal{N}} d_p(l_p) + \sum_{\{p,q\} \in \mathcal{E}} V_{pq}(l_p, l_q) \qquad (1)$$

$\mathcal{N}$ はノードの集合，$\mathcal{E}$ はエッジの集合である．p，q をノード番号，ノード p に割り当てられたラベルを l_p とする．また，l_p は縦線 ID と横線 ID のペア（$l_{p,h}$, $l_{p,v}$）で構成される．ラベル L の総数はノードの総数に等しい．各ノード p は観測座標（x_p, y_p）を持つ．
データコスト $d_p(l_p)$ は観測データと割り当てラベルの矛盾により定義する．観測された交点の座標とその交点に割り当てられたラベルから推定された座標の距離が大きい場合は，データコストが大きくなる．また，交点の縦線および横線の補助線の色が交点に割り当てられたラベルから推定される色と異なる場合にもデータコストが大きくなる．
不連続コスト $V_{pq}(l_p, l_q)$ は隣接している 2 つのノードのラベルの矛盾により定義する．同一補助線上にない場合や連続していない場合は不連続コストが大きくなる．また，ラベルの差とラベルの間隔が適切でない場合も不連続コストが大きくなる．

4. 各セルの領域と明暗の取得
まず，交点の ID 認識結果から各交点の座標を取得する．各交点を 1 つずつ注目し，注目する交点とその近傍 3 点を確認する．このとき，4 点のうち 1 点のみ欠損している場合は，取得した 3 点の座標から平行四辺形を作成するように 4 点目を補完する．全ての交点を確認し終えた後，補完前と補完後の取得交点数に変化がなければ補完処理を終了し，変化がある場合はさらに補完処理を行う．

図 8: シミュレータで生成した歪んだ QR コード画像

図 9: 教師データの作成に使用する画像

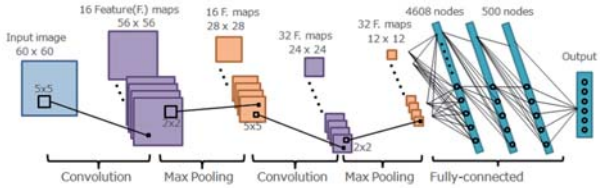


図 10: CNN のネットワーク構造

隣接する 4 つの交点で矩形を作成し，矩形内を適切なセル数に分割することで各セルの領域を取得する．
次に，4 つの交点で作成された矩形領域が黒および白のどちらのセルで構成されているかを判定する．各セルの領域の平均輝度値から明暗を取得することで黒および白のどちらのセルかを判定する．

5. デコード
取得した各セルの明暗情報から QR コードを復元し，復号処理を行う．

3.3 シミュレータの構築

本研究では，歪んだ QR コード画像を生成するシミュレータを構築する．シミュレータは，折り紙シミュレータと 3DCG ソフトである blender の二つのツールを組み合わせて構築する．まず，FreeFormOrigami[Tachi 2010] という折り紙シミュレータを用いて紙特有のしわや歪みを再現し，歪んだ紙の 3D データを生成する．次に，生成した歪んだ紙の 3D データを blender で読み込んで，読み込んだ 3D データに回転，スケール変更，光源の位置変更などの処理を行い，QR コード画像をテクスチャすることで図 8 に示すような歪んだ QR コード画像を大量に生成することができる．

CNN の学習データの作成は構築したシミュレータにより行う．シミュレータで作成した歪んだ QR コード画像からスライド窓で切り取った画像をネットワークの入力データとする．シミュレータで学習データを作成する際に，教師データの作成も行う．まず，補助線と交点の領域が着色された画像を用意し，学習データの作成と同様に図 9 のような歪んだ画像を生成する．生成された画像をスライド窓で切り取った画像の中心画素の色情報を用いて教師データを設定する．

3.4 CNN による検出器の構築

補助線および交点の検出器は，CNN を用いて構築する．本研究で用いる CNN のネットワーク構造を図 10 に示す．中間

層は畳み込み層とプーリング層がそれぞれ 2 層ずつ，全結合層が 3 層となっている．ネットワークの入力層では，シミュレータにより生成した歪んだ QR コード画像からスライド窓で切り取った画像を入力データとする．出力層は，6 クラス分類の形式となっており，入力データの中心画素が補助線，交点，その他のどの領域に属するかを認識し，クラス分類を行う．

4. 評価実験

提案方式を用いて複雑に歪んだ QR コードの復号を試みた．画像の撮影は産業用カメラ (Lynx acA1300-30gc，Fujinon DF6HA-1B，解像度 1,278 × 958) で行った．撮影した QR コードは 16cm 四方のものを紙に印刷し，40cm 程度の距離から解像度 1278 × 958 で撮影を行った．撮影時は 2 方向から照明を照射するものとした．

また，提案方式で用いる CNN の入力データは以下の設定で行う．シミュレータによって生成した歪んだ QR コード画像から 60 × 60 ピクセルのスライド窓で切り取った画像をネットワークの入力データとする．また，学習データの訓練データは 277,440 枚，テストデータは 36,992 枚を用いてネットワークの学習を行った．

提案方式を用いて複雑に歪んだ QR コードの復号を試みた結果を図 11 に示す．(a) は撮影画像，(b)(c) は補助線の横線と縦線の検出結果，(d) は復元結果，(e) の赤いセルは復元に失敗した誤りセルを表す．実験結果より，歪みの存在する領域は誤りを多く含むことがわかる．

提案方式の実験結果を図 12 に示す．グラフの縦軸は遮蔽量，横軸は歪み量を表す．遮蔽量は歪みなどにより視認できなくなったセル数，歪み量は 3 次元計測で求めた近似平面の平均 2 乗誤差とする．また，グラフの丸バツは復号の可否を表している．評価実験の結果から，提案方式では，遮蔽領域の存在する画像においても復号が可能であることがわかる．これは，提案方式が局所的なセルの位置の推定を行えるためである．

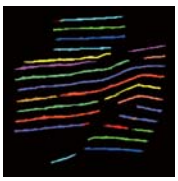
5. まとめ

本研究では，照明環境の影響を受けにくく，局所的な歪みや遮蔽領域にも対応した 2 次元コードの復号方式を提案した．提案方式は，補助線の検出に CNN を用い，補助線をもとにセルの位置を推定し復号を行う．実験結果から，遮蔽領域の存在する画像においても復号が可能であることを確認した．今後は，補助線の検出精度の向上や処理時間の短縮について検討する．

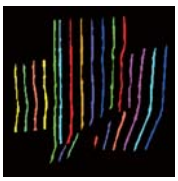
6. 謝辞

本研究の一部は科学研究費補助金 (課題番号：16k12490) による支援を受けて実施した．ここに感謝の意を表する．

(a) 撮影画像



(b) 横線検出結果



(c) 縦線検出結果

(d) 復元結果

(e) 復元結果 (誤りセル)

図 11: 実験結果例

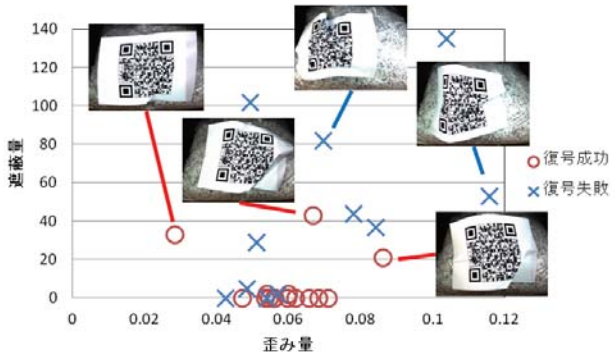


図 12: 復号可否の結果

参考文献

[Kamizuru 2015] Kamizuru, Kohei, et al. ”Belief-propagation-based robust decoding for two-dimensional barcodes to overcome distortion and occlusion and its extension to multi-view decoding.” ICIP, 2015.

[Kamizuru 2017] Kamizuru, Kohei, et al. ”A belief-propagation-based decoding method for two-dimensional barcodes with monochrome auxiliary lines robust against non-uniform geometric distortion.” QCAV, 2017.

[Iida 2000] 飯田正広，重草久志. ”２次元コード読取方法、２次元コード読取装置及び記録媒体”，特許出願公開番号　特開 2000-222517，2000.

[Mizumoto 2004] 水本裕之，瀬口陽太郎. ”二次元コード読取装置”，特許出願公開番号　特開 2004-f362053，2004.

[Fredricksen 1977] Fredricksen, Harold, and Irving Kessler. ”Lexicographic compositions and debruijn sequences.” Journal of Combinatorial Theory, Series A 22.1 (1977): 17-30.

[Frey 1998] Frey, Brendan J., and David JC MacKay. ”A revolution: Belief propagation in graphs with cycles.” Advances in neural information processing systems. 1998.

[Agrawal 2008] Agrawal, Motilal, Kurt Konolige, and Morten Rufus Blas. ”Censure: Center surround extremas for real-time feature detection and matching.” European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2008.

[Tachi 2010] Tomohiro Tachi, ”Freeform Rigid-Foldable Structure using Bidirectionally Flat-Foldable Planar Quadrilateral Mesh”, Advances in Architectural Geometry 2010, pp. 87–102, September 2010.

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

ニューラル質問応答モデルの仮想世界から現実世界のデータへのドメイン適応

宮西大樹 *1　　川鍋一晃 *2

*1*2国際電気通信基礎技術研究所
*2理化学研究所 革新知能統合研究センター

1. はじめに

近年の行動認識術の発展により，日常生活行動を表すセンサデータを時系列に並ぶ言語データに変換できるようになった [Krishna 17, Miyanishi 18]．このような日常生活を表す言語データを読み込んで，実世界の内容を問う様々な質問に解答できる実世界質問応答システムが実現すれば，現実世界の内容を理解するロボットや，室内での物探し，患者の医療モニタリングや個人の記憶補助に役立てることができる．

実世界質問応答の実現には，ユーザーによって与えられる質問と現実世界を表す時系列に並んだ複数の文（ストーリー）の関係から推論を行い，解答文となる単語または単語列を生成する必要がある．例えば,「サイトーさんがキッチンでコーヒーを作り，コーヒーを手に持った後、リビングに移動して，コーヒーを机の上に置いた．そしてサイトーさんは TV を付けた後にコーヒーを飲んだ.」というストーリーと「いまサイトーさんはどこにいる？」という質問文が与えられたとき，質問応答システムは「リビングルーム」という単語を出力する．本研究では，外部メモリの機構を持つニューラルネットワーク（以下，ニューラル質問応答モデル） [Kumar 16, Xiong 16] を用いてこの質問応答の課題に取り組む．

ニューラル質問応答モデルは，高精度な質問応答が可能であるものの，モデルの性能を発揮するために，質問・ストーリー・解答の三つ組の学習データが大量に必要である．しかし，実世界のデータはプライバシーの問題やラベリングに多大な労力を必要とするため，大量の学習データを用意することは困難である．

そこで，本研究ではライフシュミレーションで取得した仮想世界のデータを用いて質問応答の学習データを作成し，このデータで学習したニューラル質問応答モデルを敵対的損失関数を用いて現実世界のデータにドメイン適応させることで，この問題に対処する．我々はライフシミュレーターとして The Sims 4（図 1 参照）を使い仮想世界の質問応答データを取得し，住宅環境内で収集した日常生活行動に関する現実世界の質問応答データで評価を行う．

2. 実世界質問応答

本節では，実世界の内容を質問して解答を生成するニューラル質問応答モデルについて紹介する．図 2 に本手法の枠組みを図示する．提案手法では，質問と文書のエンコーダを仮想世界のデータ（ソース）と現実世界のデータ（ターゲット）の 2 種類に分け，仮想世界のデータで学習したエンコーダと解答を出力するデコーダ用いてソース側のデータで学習を行い，次にソース側のデータとエンコーダ，ターゲット側のデータとエンコーダを用いて，敵対的損失を通してドメイン適応を行う．最後に，ドメイン適応したターゲット側のエンコーダをソース側のデコーダと組み合わせて質問応答システムを構成して，評価を行う．本研究では，はエンコーダとして動的記憶ネットワーク（Dynamice Memory Networks; DMN），デコーダとしてゲート付き回帰ユニット（gated reccurent unit; GRU）を用いる．質問応答システムは，質問 Q とストーリー $\mathbf{D} = [D_1, D_2, \cdots, D_N]$ が与えられたとき，解答 A を返す．ここで，Q=$(w_1^Q, w_2^Q, \cdots, w_n^Q)$ を質問文の単語，D=$(w_1^D, w_2^D, \cdots, w_n^D)$ をストーリー中の文の単語，A=$(w_0^A, w_2^A, \cdots, w_n^A)$ を解答文中の単語に対応する単語ベクトルとする．

図 1: ライフシュミレーションゲーム The Sims 4 のプレイ画面．ゲーム内のキャラクターは自身の性格や欲求に従って，室内を半自動的に行動する．

2.1 デコーダ：ゲート付き回帰ユニット

エンコーダの出力から解答となる単語列を生成するため，下式で定義されるゲート付き回帰ユニット [Chung 14, Cho 14]（以下，GRU と略記）を用いる．

$$
\begin{aligned}
z_i &= \sigma(W_z x_i + U_z h_{i-1} + b_z) \\
r_i &= \sigma(W_r x_i + U_r h_{i-1} + b_r) \\
\tilde{h}_i &= \tanh(W x_i + r_i \circ U h_{i-1} + b_h) \\
h_i &= (1 - z_i) \circ h_{i-1} + z_i \circ \tilde{h}_i
\end{aligned}
$$

ここで，x_t は時刻 t の入力ベクトル，h_t は状態ベクトル，σ はシグモイド関数，$\circ$ はアダマール積を表す．また，$W_z, W_r, W \in \mathbb{R}^{n_H \times n_I}$ と $U_z, U_r, U \in \mathbb{R}^{n_H \times n_H}$，次元 n_I は入力ベクトルの大きさ，次元 n_H は状態ベクトルの大きさ，b_z, b_r, b_h はバイアス項を表す．以下，上記 GRU を表す関数を $h_t = \mathrm{GRU}(x_t, h_{t-1})$ と略記する．解答文を生成する際は，単語列の生成確率 $P(A) = \prod_{t=1}^{T+1} P(w_t^A | A_{[0,t-1]})$ に基づいて単語のサンプリングを行う．ここで，$P(A_t | A_{[0,t-1]}) = \mathrm{softmax}(W_o h_t^A + b_o) \cdot w_t^A$，$h_t^A = \mathrm{GRU}(w_t^A, h_{t-1}^A)$, $W_o \in \mathbb{R}^{n_V \times n_H}$，次元 n_V は語彙数となる．

連絡先: 宮西大樹，国際電気通信基礎技術研究所，京都府相楽郡精華町光台 2 丁目 2 − 2 ,miyanishi@atr.jp

2.2 エンコーダ：動的記憶ネットワーク

質問とストーリーをベクトル表現に変換するため，エンコーダとして，動的記憶ネットワーク（以下，DMN と略記）[Xiong 16] を用いる．DMN は，入力部，エピソード記憶部，質問，解答部の 4 つに分かれる．本稿では，解答部は前述のデコーダを用い，入力部は，ストーリー中の文のベクトル表現を事実情報 f_i とし，複数の事実間の関係を GRU を用いて保持し，質問文のベクトル表現と注意機構 [Bahdanau 15, Luong 15] を用いて質問に対応する内容を問い合わせる．

2.2.1 入力部と質問

DMN の入力部では，ストーリーの各文をベクトル表現に変換し，それらの前後関係を表現する．文のベクトル表現を得るため，文中の単語の位置を考慮した位置埋め込みを行う．具体的には，$s_i = \sum_J^{j=1} l_j \circ w_j^i$，ここで $\circ$ は要素ごとの積を表す．l_j は列ベクトルであり，$l_{jd} = (1 - j/J) - (d/n_V)(1 - 2j/J)$ とする．d は単語の文中での位置，J は文中の単語の数を表す．質問文のベクトル表現は $q = \sum_J^{j=1} l_j \circ w_j^Q$，ストーリー中の文のベクトル表現は，$d = \sum_J^{j=1} l_j \circ w_j^D$ になる．

次に，ストーリー中の文間の双方向の関係を捉えるため，双方向 GRU（Bidirectional GRU; bi-GRU）を用いる．

$$f_i = \text{bi-GRU}(f_{i-1}, f_{i+1}, d_i) \tag{1}$$

ここで，f_i は i 番目の双方向 GRU の状態ベクトル，f_{i-1} は前向き GRU の状態ベクトル，f_{i+1} は後ろ向き GRU の状態ベクトルになる．状態ベクトル f を事実とし，これがエピソード記憶部の入力となる．

2.2.2 エピソード記憶部

DMN のエピソード記憶部では，質問に応じて特定の事実に注目しながら $F = [f_1, ...f_N]$ の検索を行う．ここで，注意ゲート g_i^t は，質問 q が事実 f_i に対してどの程度注意するかを表す．

$$z_i^t = [f_i \circ q; f_i \circ m^{t-1}; |f_i - q|; |f_i - m^{t-1}|] \tag{2}$$

$$Z_i^t = W_\beta \tanh(W_\alpha z_i^t + b_\alpha) + b_\beta \tag{3}$$

$$g_i^t = \frac{\exp(Z_i^t)}{\sum_{k=1}^{J_i} \exp(Z_k^t)} \tag{4}$$

ここで，f_i は i 番目の事実で，m^{t-1} は以前のエピソード記憶，q は質問文のベクトル表現，$|\cdot|$ は要素ごとの絶対値，; はベクトルの結合を表す．

次に注意ゲート g_i^t と GRU を用いて，質問 q と事実 F の関係を表す文脈ベクトル c を求める．ここで，GRU の式 (1) の z_t を g_i に置き換えることで，質問 q に応じて事実 f_i の前後の文脈を重み付けした表現を獲得する．

$$h_i = (1 - g_i^t) \circ h_{i-1} + g_i^t \circ \tilde{h}_i \tag{5}$$

GRU の h_N を文脈ベクトル c とする．

記憶の更新は t 回行う．t 回目文脈のベクトル表現を c^t, t 回更新後のエピソード記憶を表すベクトル表現を m^t とする．記憶の更新は下式で行う．

$$m^t = \text{ReLU}(W^t[m^{t-1}; c^t; q] + b) \tag{6}$$

ここで，$W^t \in \mathcal{R}^{n_H \times n_H}$，活性化関数として ReLU（Rectified Linear Unit）関数 [Nair 10] を用いる．m^t がエンコーダ DMN の最終的な出力になる．

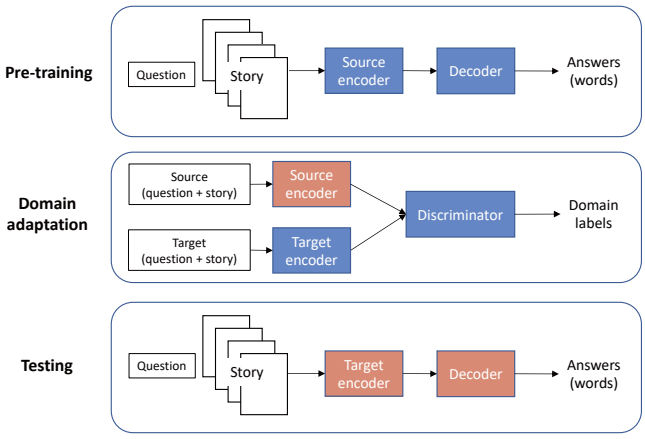


図 2: 敵対的ドメイン適応の模式図

2.3 ニューラル質問応答モデルのドメイン適応

ニューラル質問応答モデルの仮想世界のデータから現実世界のデータへのドメイン適応を行うため，敵対的ドメイン適応（Adversarial Discriminative Domain Adaptation; ADDA）[Tzeng 17] を用いる．ドメイン適応のため，ADDA では，下記目的関数を段階的に最適化する．

$$\min_{M_s, C} \mathcal{L}_{seq}(\mathbf{X}_s, Y_s) =$$
$$-\mathbb{E}_{(\mathbf{x}_s, y_s) \sim (\mathbf{X}_s, Y_s)} \sum_{k=1}^{K} \mathbf{1}_{[k=y_s]} \log C(M_s(\mathbf{x}_s))$$
$$\min_D \mathcal{L}_{adv_D}(\mathbf{X}_s, \mathbf{X}_t, E_s, E_t) =$$
$$-\mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s}[\log D(M_s(\mathbf{x}_s))] - \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log(1 - D(M_t(\mathbf{x}_t)))]$$
$$\min_{M_s, M_t} \mathcal{L}_{adv_M}(\mathbf{X}_s, \mathbf{X}_t, D) =$$
$$-\mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s}[\log D(M_t(\mathbf{x}_t))] \tag{7}$$

ここで，$\mathbf{x}_s$ はソース側のストーリーと質問 $\mathbf{D}_s, Q_s$, y_s は解答 A_s に相当する．$\mathbf{x}_t$ はターゲット側のデータを表す．まず，ソース側のエンコーダ M_s とデコーダ C をラベルありのデータ $\mathbf{X}_s$ と Y_s 用いて $\mathcal{L}_{seq}$ を最適化する．次に，ソース側のエンコーダ M_s の重みを固定して，ターゲット側のエンコーダと識別器 D を用いて $\mathcal{L}_{adv_D}$ を最適化する．その後，デコーダ M_t の $\mathcal{L}_{adv_M}$ を最適化する．図 2 にドメイン適応の模式図を示す．

3. データセット

3.1 仮想世界のデータ：The Sims のシミュレーション

日常生活のシミュレーションデータを取得するため，シミュレーションゲーム The Sims 4 を用いる．The Sims 4 では Sim と呼ばれる人間を模したキャラクターが自身の性格や感情や欲求に従って屋内を半自動的に行動する．例えば，空腹ゲージが下がるとキッチンの冷蔵庫から食べ物を取り出して，ダイニングの机に置いて食事を始めたり，膀胱ゲージが下がるとトイレに行く行動を取りやすくなる．Sim が生活する屋内環境はカスタマイズ可能であり，物体（コーヒーメーカー，食洗機，PC など）を追加することで Sims の行動が変化する．また，物体を選択することで，その物体に関連する行動を Sim にとらせることも可能である．

本研究では，日常生活行動のシミュレーションを行うため，異なる 3 つの住宅環境を用意した．各環境には，キッチン・ダ

図 3: The Sims 4 の屋内環境の例，左図は男性とペットの住環境，右図は夫婦の住環境．

イニング，リビング，バスルーム，ベッドルームの部屋と，日常生活に必要な物体（冷蔵庫やベッドや TV など）を用意した．各環境内では，男性 1 人とペット 1 匹（犬），男性 2 人，男性 1 人女性 1 人の夫婦とその子供一人（女性）の 3 つの組み合わせそれぞれ日常生活を送る．日常生活のシミュレーションでは，1 日の行動を 1 セッションとして，１０セッション分のデータを取得し，各個人にどこで・だれが・何をしたかといった日常生活行動を表す文のラベリングを行った．

3.2 現実世界のデータ：屋内での日常生活行動

提案手法の評価を行うため，日常生活行動データセットを用いる．このデータセットは日常生活行動をヘッドマウントカメラで撮影し，撮影した一人称映像をもとにして，いつ・どこで・誰が・何をしたかを表す文を人手でラベリングしたものである．このデータセットを作成するため，5 人の被験者がモーションセンサとウェアラブルカメラを装着して，19 個の日常生活行動を異なる場所ごと固有の行動をした．被験者はワークシートに書かれた行動を被験者の意思に従って任意のタイミングで行う．例えば，被験者はソファに座ってリビングルームでテレビを見ながらコーヒーを飲み，その後，台所に移動して皿を洗う，といった連続した日常生活行動を行う．この一連の行動を 1 セッションとして１０回繰り返し，１ 人につき１０セッション分のデータを収集した．このセミナチュナルな実験プロトコルによるデータ収集 [Bao 04] は，限られた環境の中で多様な行動についてのセンサデータを集めるために，行動認識の研究でよく用いられる．質問応答のデータ・セットを用意するため，質問のテンプレートを用意し，仮想世界と現実世界の文に対応する解答をルールベースで作成した．

4. おわりに

本稿では，ニューラル質問応答モデルのドメイン適応手法と当該手法の有効性を検証するためのデータセットを紹介した．実験設定の詳細や実験結果については，インタラクティブ発表にて紹介する．

5. 謝辞

本研究は，JSPS 科研費 16K21718，JST CREST JPMJCR15E2 の助成を受けたものです．

参考文献

[Bahdanau 15] Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, *ICLR* (2015)

[Bao 04] Bao, L. and Intille, S. S.: Activity recognition from user-annotated acceleration data, in *Pervasive*, pp. 1–17 (2004)

[Cho 14] Cho, K., Merriënboer, van B., Bahdanau, D., and Bengio, Y.: On the Properties of Neural Machine Translation: Encoder–Decoder Approaches, in *SSST-8* (2014)

[Chung 14] Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, in *Deep Learning and Representation Learning Workshop* (2014)

[Krishna 17] Krishna, R., Hata, K., Ren, F., Fei-Fei, L., and Niebles, J. C.: Dense-Captioning Events in Videos, in *ICCV*, pp. 706–715 (2017)

[Kumar 16] Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., Zhong, V., Paulus, R., and Socher, R.: Ask me anything: Dynamic memory networks for natural language processing, in *ICML* (2016)

[Luong 15] Luong, M.-T., Pham, H., and Manning, C. D.: Effective approaches to attention-based neural machine translation, *EMNLP* (2015)

[Miyanishi 18] Miyanishi, T., Hirayama, J.-i., Maekawa, T., and Kawanabe, M.: Generating an Event Timeline about Daily Activities from a Semantic Concept Stream, in *AAAI* (2018)

[Nair 10] Nair, V. and Hinton, G. E.: Rectified linear units improve restricted boltzmann machines, in *ICML*, pp. 807–814 (2010)

[Tzeng 17] Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T.: Adversarial discriminative domain adaptation, in *CVPR* (2017)

[Xiong 16] Xiong, C., Merity, S., and Socher, R.: Dynamic memory networks for visual and textual question answering, in *ICML*, pp. 2397–2406 (2016)

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

階層的な部分単語を入力としたニューラル機械翻訳

Improving Neural Machine Translation by Incorporating Hierarchical Subword Features

森下 睦　　鈴木 潤　　永田 昌明
Makoto Morishita　Jun Suzuki　Masaaki Nagata

NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation

This paper focuses on the subword-based neural machine translation (NMT). We hypothesize that the appropriate subword-units for three modules in the NMT model, namely, (1) encoder embedding layer, (2) decoder embedding layer, and (3) decoder output layer, can differ each other. We empirically investigate our assumption, and find that incorporating several different subword-units for input and output embedding layers can consistently improve the BLEU score on the IWSLT 2012, 2013 and 2014 evaluation datasets.

1. はじめに

ニューラル機械翻訳とは，符号化器/復号化器モデル (Encoder-decoder model) または系列変換モデル (Sequence-to-sequence model)[Sutskever 14] と呼ばれる方法を用いた機械翻訳手法の総称として用いられている．ニューラル機械翻訳の研究は，世界中の著名な研究機関が多く参入し，人工知能/自然言語処理分野の国際会議や論文誌で数多くの研究成果が報告されている非常に活発な研究トピックである．

多くの研究報告の中から，いくつかの方法論はデファクトスタンダードの技術となり，後続の研究やツールで広く利用されている．最も広く取り入れられた技術の典型例としては，注意機構 (attention mechanism) があげられる [Bahdanau 15, Luong 15]．近年では，部分単語 (subword) 単位を用いる方法 [Sennrich 16, Schuster 12] がニューラル機械翻訳の研究では広く用いられるようになりつつある．本稿では，部分単語単位を用いる方法の中でも特に byte-pair encoding (BPE) を用いる方法 [Sennrich 16] に着目する．以降，特に注意書きがない限り，本稿での部分単語単位とは，BPE に基づいて決定された単位を意味することとする．

ニューラル機械翻訳において．部分単語単位を用いる利点はいくつか考えられるが，部分単語単位を用いる最も大きな理由は，言語生成時の未知語問題への対応である，従来のニューラル機械翻訳では，単語そのものを生成する方式で行なっていたため，基本的に学習データに出現しない単語を生成することは不可能であった．一方，部分単語単位を用いると，部分単語単位で構成できる単語は理論的には全て生成可能となり，システムが理論的に生成可能な語彙数を爆発的に増やすことが可能となる．一般的に，従来の単語単位のニューラル機械翻訳システムでは，生成できる語彙数は多くても数十万ぐらいが妥当な数で，数百万単位の語彙数でシステムを構築するには，かなり多くの計算機リソースと計算時間が求められることになる．この時，入出力データを部分単語単位に変換して用いるだけで，数千から数万の部分単語で，見かけ上数百万以上の単語を表現可能となり，未知語問題をおおよそ解決したと言える程の効果が得られる．また，副次的効果として計算リソースや計算時間の大幅削減にも貢献する非常によい性質を持っている．

連絡先: 森下 睦，NTT コミュニケーション科学基礎研究所，京都府相楽郡精華町光台 2-4, morishita.makoto@lab.ntt.co.jp

本稿では，このように優れた結果を実験的に示している部分単語単位に関して，更に一歩考察を進めていきたい．まず，前述のように，部分単語単位を導入した主な理由は生成時の未知語対応にある．つまり，復号化器の出力層の語彙を部分単語単位にすることが主な目的と言える．また，復号化器の出力層の語彙が決まると，当然入力層の語彙も同じものを使うことが自然であり一般的な方法論なので，こちらも同様に決定される．更に，復号化器だけを部分単語単位にすると不整合が生じる恐れがあるため一般的には，符号化器の語彙も復号化器の部分単語単位を構築する設定で構築する場合が多い．あるいは，通称 joint BPE [Sennrich 16] と呼ばれる入力文と出力文を一括して部分単語単位を求める方法が用いられることも多い．このように，ニューラル翻訳では，部分単語単位の関わる部位が符号化器の入力，復号化器の入力と出力と合計 3 種類存在するが，基本的に単一の部分単語単位を決定する方法を用いている．本稿では，この点に焦点をあて，符号化の入力単位，復号化器の入力単位と出力単位の計 3 種類の部分単語単位は，それぞれ適した部分単語単位は同じではないのではないかという仮説を立てる．つまり，これら 3 種類の部位は，当然違った役割でモデル内に組み込まれているのであり，その役割に則って効果的に機能する部分単語単位は違う可能性は否定できないが，これまでは特にこの点は指摘されず慣習に基づいて単一の設定で部分単語単位を決定していた．この仮説を検証するため，まず複数の部分単語単位が扱えるようにモデルを拡張する．その後，様々な数/長さの部分単語単位を作成し，部分単語単位を変えた場合の効果の違いを実験にて調査する．

2. 部分単語単位に基づくニューラル機械翻訳

本稿のベースラインモデルは，文献 [Luong 15] で用いられている注意機構（attention mechanism）付きの RNN 符号化/復号化器モデル [Bahdanau 15] とする．本モデルは，ニューラルネットに基づく翻訳や要約（ヘッドライン生成）タスクにおける強力なベースライン法という位置付けで，多くの論文で利用されている [Chopra 16, Kikuchi 16]．既に多くの参考文献が存在するので，本稿では紙面スペースの都合上，本稿の議論に直接関係ない部位に関しては詳細説明は割愛する．$\boldsymbol{X} = (\boldsymbol{x}_i)_{i=1}^{I}$ を入力系列（入力文），$\boldsymbol{Y} = (\boldsymbol{y}_j)_{j=1}^{J}$ を出力系列

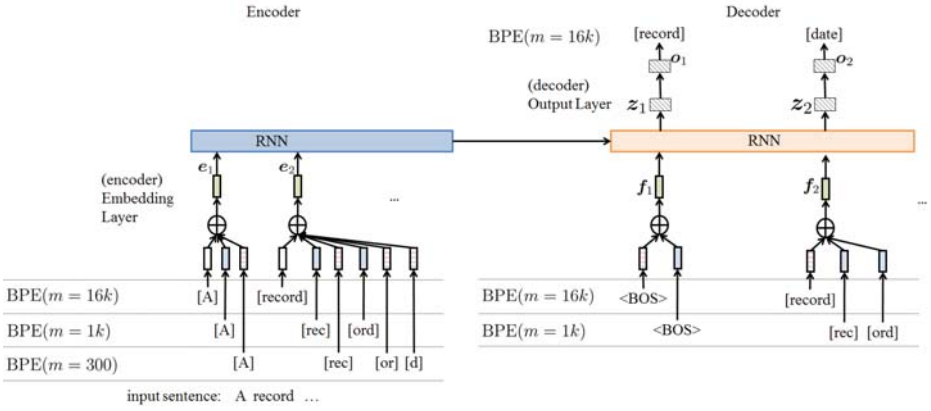


図 1: 提案法の簡単な構成図.

（出力文）とする [*1]．ただし，$\boldsymbol{x}_i$ を，i 番目の入力単語に対応する one-hot ベクトル表現，$\boldsymbol{y}_j$ を，j 番目の出力単語に対応する one-hot ベクトル表現とする [*2]．

2.1 定式化

符号化器（encoder）：$\Omega^{(\mathrm{s})}(\cdot)$ を RNN で構成される符号化器の全ての処理を表す関数とする．符号化器は，入力 $\boldsymbol{X} = (\boldsymbol{x}_i)_{i=1}^{I}$ を受け取って隠れ状態ベクトルのリスト $\boldsymbol{H}^{\mathrm{s}} = (\boldsymbol{h}_i^{\mathrm{s}})_{i=1}^{I}$ を返す処理となる．

$$\boldsymbol{e}_i = \boldsymbol{E}\boldsymbol{x}_i \tag{1}$$

$$\boldsymbol{H}^{(\mathrm{s})} = \Omega^{(\mathrm{s})}((\boldsymbol{e}_i)_{i=1}^{I}). \tag{2}$$

復号化器（decoder）＋注意機構 (attention)：本稿では，K ベストビーム探索（beam-search）を用いて，入力系列 $\boldsymbol{X}$ が与えられたときの出現確率が最大となる出力系列 $\hat{\boldsymbol{Y}}$（の近似解）を獲得する．ビーム探索では，各処理時刻 j で K 個の出力候補を保持しながら探索を行う．

本稿では，各処理時刻 j において，生成する語彙を選択する処理を述べる．まず，時刻 j における埋め込みベクトルの取得には，以下の式を用いて計算を行う．

$$\boldsymbol{f}_{j,k} = \boldsymbol{F}\tilde{\boldsymbol{y}}_{j-1,k}, \tag{3}$$

この時，$\tilde{\boldsymbol{y}}_{j-1,k}$ は，処理時刻 $j-1$ で予測された k 番目に確率が高い単語に対応する one-hot ベクトルとする．ただし，全ての k に対して $\tilde{\boldsymbol{y}}_{0,k}^{(k)}$ は，必ず特殊単語 BOS に対応する one-hot ベクトルとする．

次に得られた埋め込みベクトル $\boldsymbol{f}_j$ を使って，RNN と注意機構を用いて最終隠れ層のベクトル $\boldsymbol{z}_{j,k}$ を計算する．

$$\boldsymbol{z}_{j,k} = \texttt{RNN_Attn}(u_{(j-1,k)}, \boldsymbol{f}_{j,k}, \boldsymbol{H}^{(\mathrm{s})}) \tag{4}$$

ここでは，RNN_Attn を入力ベクトル $\boldsymbol{f}_{j,k}$ を受け取って，RNN と注意機構を用いて最終隠れ層のベクトル $\boldsymbol{z}_{j,k}$ を計算する処理全てを表す関数とする．ただし，$u_{(j,k)}$ は，処理時刻 j における k 番目の候補が，処理時刻 $j-1$ の時の 1 から K 番目のどの候補から生成されたかを示す値とする．よって，$u_{(j,k)} = \{1, \ldots, K\}$ である．この値は，$j-1$ 時刻の時にどの RNN を用いて処理が行われたのかを次の時刻 j に伝達する役割を持っている．次に，得られた最終隠れ層のベクトル $\boldsymbol{z}_j$ から，生成する単語を選択する基準となるスコアを以下の式を用いて計算する．

$$\boldsymbol{o}_{j,k} = \boldsymbol{W}^{(\mathrm{o})}\boldsymbol{z}_{j,k} + \boldsymbol{b}^{(\mathrm{o})}, \tag{5}$$

その後，K ベストビーム探索の処理を行い，処理時刻 j における上位 K 個の候補を得る．

$$\{(\tilde{\boldsymbol{y}}_{j,k}, u_{(j,k)})\}_{k=1}^{K} = \texttt{BeamSearch}_k\Big((\boldsymbol{o}_{j,k})_{k=1}^{K}\Big), \tag{6}$$

ここで，K 個の候補とともに，前述の $u_{(j,k)}$ の情報も取得する．

学習時は，$k = 1$ として予測結果 $\tilde{\boldsymbol{y}}_{j-1,k}$ の代わりに正解 $\boldsymbol{y}_{j-1,k}$ を利用することに相当する．

2.2 部分単語単位 (BPE)

統計量に基づく部分単語単位を決定する方法として，既にいくつか提案がなされているが，byte-pair encoding (BPE) を用いた方法が昨今のニューラル機械翻訳の研究でよく見られる．BPE は入力文を最も細かい部品（文字）まで分割し，各部分単語（文字を含む）を逐次的にマージ（結合）することで徐々に文字から単語へ部分単語を組み上げていく処理を行う．その組み上げの処理に置いて，事前に決めたマージ回数に到達したら処理を終了する．

一般的には，マージ回数（m）の値はハイパーパラメタであり，人手により経験的に良いと思われる値が用いられる．近年のニューラル機械翻訳では，数千から数万の値が使われることが多く，千以下，十万以上の値はあまり用いられない傾向にある．これは，マージ回数が少ない場合は，文字単位の処理に近く，それぞれの文字がもつ意味的な情報量は限定的になるため，あまり効果的ではないと予想され，また，マージ回数が多い場合は，単語単位の処理と近くなり，折角部分単語を導入した意味が薄れてしまうということが考えられる．このような理由から，一般ドメインの翻訳で必要になる語彙数が数百万語彙だと仮定した場合，経験的に数千から数万のマージ回数とするのは妥当な値と考えられる．

3. 提案法：階層的な部分単語の活用

BPE の特性として，マージ回数が 0 回の時は，文字単位の処理と一致し，マージ回数を無限大にすると単語単位と同じになる．よって，BPE の観点で部分単語単位を用いる方法論を整理すると，文字単位の処理から単語単位の処理までをマー

*1 簡単のため，列ベクトルのリスト $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I)$ を，$(\boldsymbol{x}_i)_{i=1}^{I}$ と記述する．

*2 処理の単位は必ずしも単語である必要はないが，説明を簡単にするために，ここでは処理の単位が単語であると仮定して議論をすすめる．

ジ回数という観点で，離散値で段階的に遷移する方法論と捉えることができる．つまり，BPE は，その性質上，文字単位の処理も単語単位の処理も包含する枠組みと捉えることができる．このことから，本稿では,「部分単語単位」という用語は，直感的に思い浮かぶ単語の一部という意味だけではなく，単語そのものや文字単位の状態も含む概念として用いる．また，本稿では，m を BPE のマージ回数を表す変数とし，特に，BPE($m = 0$) を文字単位を用いる方法，BPE($m = \infty$) を単語単位を用いる方法を表すこととし，以下の議論では，文字単位や単語単位の場合を区別せず全て部分単語の文脈で議論を行う．

図 1 に提案法の簡単な構成図を示す．提案法では，符号化器，及び，復号化器の入力層を拡張する．より具体的には，複数の部分単語単位を取り扱えるような拡張を行う．

まず，復号化器の出力に対して，複数の部分単語単位を出力するように修正することを考える．これはマルチタスク学習の設定と考えれば技術的には容易に対応可能であるが，符号化復号化器では，逐次的に単語予測を繰り返すという処理を行う性質上，複数の予測結果間の整合性を担保するには，制約付きの復号化処理などが必要となる．これは，学習と評価時の復号化（デコード）処理が煩雑になるため，本稿では取り扱わないこととする．よって本稿では，復号化器の出力部分は変更不要なことを担保した状態で符号化器，及び，復号化器の入力層の修正を行うという考えを基本方針とする．

この時，復号化器の入力部分の拡張は以下の通りである．

$$\boldsymbol{f}_{j,k} = \sum_r \boldsymbol{F}_r \psi_r(\tilde{\boldsymbol{y}}_{j-1,k}) \tag{7}$$

ただし，$\boldsymbol{F}_r$ は復号化器の埋め込み行列とする．前述のように，復号化器の予測は単一であることを仮定するため，$\psi_r(\tilde{\boldsymbol{y}}_{j-1,k})$ は，予測結果 $\tilde{\boldsymbol{y}}_{j-1,k}$ をキーとした事前に定義したマッピング関数を表しており，バイナリベクトルを返すこととする．また，$\psi_r(\cdot)$ によって返されるバイナリベクトルは，$\tilde{\boldsymbol{y}}_{j-1,k}$ の部分単語単位とする．例えば，BPE($m = 16k$) の `record` という部分単語が予測された場合，BPE($m = 1k$) で `record` の部分単語となる `rec` と `ord` がマッピング関数で引かれるといった処理となる．このように，より m が小さい部分単語単位は包含関係にあるため，一意に対象となる部分単語が決まり，容易に用いることができる．つまり，上記の復号化器の入力部分の拡張は，復号化器の予測結果が $\tilde{\boldsymbol{y}}_{j-1,k}$ 一つであるため，これからマッピング関数で一意に決定できる部分単語を，特徴として利用していることに相当する．

同様に，符号化器側の入力部分の修正は以下の通りである．

$$\boldsymbol{e}_i = \sum_q \boldsymbol{E}_q \phi_q(\boldsymbol{x}_i) \tag{8}$$

ただし，$\boldsymbol{E}_q$ は復符号化器の埋め込み行列とする．$\phi_q(\boldsymbol{x}_i)$ は，$\psi_r(\tilde{\boldsymbol{y}}_{j-1,k})$ と同様に，$\boldsymbol{x}_i$ から一意に導出可能なマッピング関数を表しており，要素が 0 または 1 をとるバイナリベクトルを返すこと仮定する．

4. 実験

ここでは，提案法の効果を検証するために，翻訳の実験を行う．

4.1 実験データ

本稿では，音声翻訳の評価型ワークショップである IWSLT[Cettolo 12] で利用されているデータを用いて評価実験を行った．また，複数ある言語対の中で，英語/フランス語，及び，英語/ドイツ語の両方向の翻訳，合計 4 種の実験設定を用いた．学習データには，IWSLT-2016 の学習データ (train)，開発データには，IWSLT-2012 の評価データ (tst2012)，評価データには，IWSLT-2013 と IWSLT-2014 の評価データ (tst2013, tst2014) をそれぞれ用いた．データの前処理として，Moses の tokenizer[*3] と truecaser[*4] の処理を行った．学習データに関しては，学習効率を考慮して 50 単語以上となる文を削除した [*5]．表 3 に用いたデータの規模を表す指標としてトークン数と文数を示す．これまでの議論の通り，部分単語単位への分割には，文献 [Sennrich 16] に対応する実装として公開されているコード [*6] を利用した．

実験に用いたモデルパラメタや学習の設定を表 4 に示す．学習時の学習率の調整方法として，最初の 30epoch は学習率の初期値 1.0 で学習を行い，その後 40epoch まで各 epoch 毎に学習率を 0.8 倍して減衰させて学習を行った．また評価時には，ビームサーチ（ビーム幅 20）を用いた．その際文長に対する正規化方法として，文献 [Cromieres 16] で紹介されているスコア（負の対数尤度）を単語数で割る方法を適用した．翻訳の評価としては，BLEU スコア [Papineni 02] を用いた [*7].

4.2 結果

表 1 にドイツ語から英語 (DE-EN) と英語からドイツ語 (EN-DE) の翻訳実験結果を示す．同様に，表 2 にフランス語から英語 (FR-EN) と英語からフランス語 (EN-FR) の翻訳実験結果を示す．表中の `(a)` は近年最も標準的に用いられる部分単語単位を利用したニューラル機械翻訳の実験結果であり，本稿におけるベースラインと考える．`(b)` から `(h)` に関しては，`(a)` のベースラインから符号化器の設定を変更した実験結果である．同様に，`(i)` から `(k)` に関しては，`(a)` のベースラインから復号化器の設定を変更した実験結果である．最後に `(l)`, `(m)` は，符号化器と復号化器の両方の設定を変更した実験結果である.「単位」と「素性」の列が表しているものは，用いた部分単語単位のマージ回数 (m) である．複数記載がある場合は，複数の部分単語単位を一括して扱っていることを意味する．また ∞ は単語単位を意味する．カッコ内の値は，ベースライン `(a)` と比較した際の BLEU スコアの増減を表している．各評価データ（列）で最も高い BLEU スコアだったものを太字で示している．

実験結果から，通常の部分単語を入力とする手法と比較して，階層的な部分単語列を適用した提案法の方がより高い精度が得られていることがわかる．特に，符号化器に対して階層的な部分単語列を入力とした場合，ベースラインと比較して顕著に精度が向上している．さらに符号化器，復号化器ともに提案法を適用した場合でも全体的に精度が向上している．

これらの実験により，入出力層で適切な部分単語列は異なる可能性や，複数の部分単語列を階層的に入力とすることで既存

*3 `https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl`

*4 `https://github.com/moses-smt/mosesdecoder/blob/master/scripts/recaser/truecase.perl`

*5 ここでの 50 単語とは，Moses tokenizer の結果に対する単語数（トークン数）であり，BPE をかけた後の単語数ではない．BPE をかけたことによって単語数が 50 単語以上になっても特に削除などの処理は行わない．

*6 `https://github.com/rsennrich/subword-nmt`

*7 より詳細には，正解参照訳は前述の truecaser をかけた評価データを用い case-sensitive の設定で評価を行った．実際の評価スクリプトには，Moses に付属の `multi-bleu.perl` (`https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl`) を用いた．

表 1: ドイツ語から英語 (DE-EN) と英語からドイツ語 (EN-DE) の翻訳実験結果

	Enc		Dec		DE-EN			EN-DE		
	単位	素性	単位	素性	tst2012	tst2013	tst2014	tst2012	tst2013	tst2014
(a)	16k	16k	16k	16k	31.70	34.38	28.81	26.00	28.67	24.05
(b)	∞	∞	16k	16k	29.99 (−1.71)	32.42 (−1.96)	27.39 (−1.42)	25.37 (−0.63)	27.65 (−1.02)	23.76 (−0.29)
(c)	∞	∞, 1k	16k	16k	32.20 (+0.50)	34.18 (−0.20)	29.72 (+0.91)	26.48 (+0.48)	29.09 (+0.42)	24.61 (+0.56)
(d)	∞	∞, 300	16k	16k	32.85 (+1.15)	34.84 (+0.46)	29.77 (+0.96)	26.93 (+0.93)	28.81 (+0.14)	25.18 (+1.13)
(e)	∞	∞, 1k, 300	16k	16k	32.60 (+0.90)	**35.22 (+0.84)**	30.08 (+1.27)	26.69 (+0.69)	29.03 (+0.36)	24.84 (+0.79)
(f)	16k	16k, 1k	16k	16k	31.78 (+0.08)	34.29 (−0.09)	29.17 (+0.36)	26.26 (+0.26)	28.85 (+0.18)	24.59 (+0.54)
(g)	16k	16k, 300	16k	16k	33.04 (+1.34)	34.90 (+0.52)	30.12 (+1.31)	26.72 (+0.72)	28.81 (+0.14)	24.29 (+0.24)
(h)	16k	16k, 1k, 300	16k	16k	**33.22 (+1.52)**	35.04 (+0.66)	**30.81 (+2.00)**	27.23 (+1.23)	29.47 (+0.80)	**25.35 (+1.30)**
(i)	16k	16k	16k	16k, 1k	31.59 (−0.11)	33.37 (−1.01)	28.87 (+0.06)	26.19 (+0.19)	28.81 (+0.14)	24.20 (+0.15)
(j)	16k	16k	16k	16k, 300	32.46 (+0.76)	34.43 (+0.05)	30.12 (+1.31)	26.13 (+0.13)	28.96 (+0.29)	24.21 (+0.16)
(k)	16k	16k	16k	16k, 1k, 300	31.47 (−0.23)	33.34 (−1.04)	29.10 (+0.29)	26.79 (+0.79)	28.78 (+0.11)	24.38 (+0.33)
(l)	∞	∞, 1k, 300	16k	16k, 1k, 300	32.69 (+0.99)	34.99 (+0.61)	29.47 (+0.66)	**27.41 (+1.41)**	**30.01 (+1.34)**	25.31 (+1.26)
(m)	16k	16k, 1k, 300	16k	16k, 1k, 300	32.86 (+1.16)	34.78 (+0.40)	30.38 (+1.57)	26.42 (+0.42)	29.19 (+0.52)	24.48 (+0.43)

表 2: フランス語から英語 (FR-EN) と英語からフランス語 (EN-FR) の翻訳実験結果

	Enc		Dec		FR-EN			EN-FR		
	単位	素性	単位	素性	tst2012	tst2013	tst2014	tst2012	tst2013	tst2014
(a)	16k	16k	16k	16k	42.24	39.56	36.67	43.99	40.00	36.83
(b)	∞	∞	16k	16k	40.54 (−1.70)	38.54 (−1.02)	36.22 (−0.45)	43.22 (−0.77)	39.43 (−0.57)	36.66 (−0.17)
(c)	∞	∞, 1k	16k	16k	42.63 (+0.39)	39.26 (−0.30)	36.93 (+0.26)	44.94 (+0.95)	40.68 (+0.68)	38.34 (+1.51)
(d)	∞	∞, 300	16k	16k	42.95 (+0.71)	39.79 (+0.23)	37.70 (+1.03)	**45.52 (+1.53)**	**41.79 (+1.79)**	38.23 (+1.40)
(e)	∞	∞, 1k, 300	16k	16k	43.11 (+0.87)	40.32 (+0.76)	**38.01 (+1.34)**	43.11 (−0.88)	40.32 (+0.32)	38.01 (+1.18)
(f)	16k	16k, 1k	16k	16k	43.27 (+1.03)	39.99 (+0.43)	37.28 (+0.61)	44.56 (+0.57)	40.94 (+0.94)	38.18 (+1.35)
(g)	16k	16k, 300	16k	16k	44.05 (+1.81)	40.20 (+0.64)	37.36 (+0.69)	44.94 (+0.95)	41.10 (+1.10)	38.50 (+1.67)
(h)	16k	16k, 1k, 300	16k	16k	**44.12 (+1.88)**	**40.40 (+0.84)**	37.42 (+0.75)	44.90 (+0.91)	40.68 (+0.68)	38.09 (+1.26)
(i)	16k	16k	16k	16k, 1k	42.37 (+0.13)	39.13 (−0.43)	37.23 (+0.56)	44.22 (+0.23)	40.21 (+0.21)	37.36 (+0.53)
(j)	16k	16k	16k	16k, 300	42.48 (+0.24)	38.78 (−0.78)	36.56 (−0.11)	43.45 (−0.54)	39.99 (−0.01)	37.13 (+0.30)
(k)	16k	16k	16k	16k, 1k, 300	42.50 (+0.26)	39.02 (−0.54)	26.62 (−0.05)	43.19 (−0.80)	40.29 (+0.29)	36.86 (+0.03)
(l)	∞	∞, 1k, 300	16k	16k, 1k, 300	42.98 (+0.74)	39.61 (+0.05)	37.43 (+0.76)	45.30 (+1.31)	40.88 (+0.88)	**38.54 (+1.71)**
(m)	16k	16k, 1k, 300	16k	16k, 1k, 300	43.71 (+1.47)	39.74 (+0.18)	37.53 (+0.86)	45.12 (+1.13)	41.32 (+1.32)	38.26 (+1.43)

表 3: 実験に用いたデータのトークン数と文数. (#token：トークン数. #sent：文数)

	DE-EN		FR-EN	
	#token	#sent	#token	#sent
train (学習データ)	3.2M	189.3K	3.8M	208.3K
tst2012 (開発データ)	30.9K	1.7K	21.7K	1.1K
tst2013 (評価データ)	21.0K	1.0K	21.9K	1.0K
tst2014 (評価データ)	25.0K	1.3K	25.0K	1.3K

表 4: 実験に用いたモデルと学習時のパラメタの値

ハイパーパラメタ	値
Embedding dimension	512
Hidden dimension	512
Attention dimension	512
Encoder layer	2
Decoder layer	2
Optimizer	SGD
Initial learning rate	1.0
Gradient clipping	5.0
Dropout rate	0.3
Mini-batch size	128

手法と比較してさらなる精度向上が見込めることが示せた.

5. おわりに

本稿では，近年デファクトスタンダードとなっている部分単語単位を用いたニューラル機械翻訳に着目し，符号化器の入力単位，復号化器の入力単位と出力単位の計 3 種類の部分単語単位は，それぞれ適した部分単語単位は同じではないのではないかという仮説を立て，それを実験的に検証することを主たる目的とした．この仮説を検証するために，本稿での提案法となる符号化器と復号化器の入力単位に複数の部分単語単位を同時に利用できるようなモデルの簡単な拡張を行った．IWSLT の実験データを用いた実験から，復号化器の出力単位から導出できる部分単語単位を利用することで，BLEU スコアによる翻訳性能が安定的に向上することが確認できた．同様に，符号化器は，元の単語の単位とマージ回数が少なめ ($m = 1k$ または $m = 300$) の部分単語単位を同時に利用することで，BLEU スコアによる翻訳性能が安定的に向上することが確認できた．また，これら二つを組み合わせることでも更に BLEU スコアによる翻訳性能が安定的に向上することが確認できた．

本稿での検証結果は，簡単な改良で安定的に翻訳性能を向上させることができることから，今後のデファクトスタンダードな方法論になりえることが期待できる．

参考文献

[Bahdanau 15] Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, in *Proceedings of ICLR* (2015)

[Cettolo 12] Cettolo, M., Girardi, C., and Federico, M.: WIT3: Web inventory of transcribed and translated talks, in *Proceedings of EAMT*, pp. 261–268 (2012)

[Chopra 16] Chopra, S., Auli, M., and Rush, A. M.: Abstractive Sentence Summarization with Attentive Recurrent Neural Networks, in *Proceedings of NAACL*, pp. 93–98 (2016)

[Cromieres 16] Cromieres, F., Chu, C., Nakazawa, T., and Kurohashi, S.: Kyoto University Participation to WAT 2016, in *Proceedings of WAT*, pp. 166–174 (2016)

[Kikuchi 16] Kikuchi, Y., Neubig, G., Sasano, R., Takamura, H., and Okumura, M.: Controlling Output Length in Neural Encoder-Decoders, in *Proceedings of EMNLP*, pp. 1328–1338 (2016)

[Luong 15] Luong, M.-T., Pham, H., and Manning, C. D.: Effective Approaches to Attention-based Neural Machine Translation, in *Proceedings of EMNLP* (2015)

[Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, in *Proceedings of ACL* (2002)

[Schuster 12] Schuster, M. and Nakajima, K.: Japanese and korean voice search, in *Proceedings of ICASSP*, pp. 5149–5152 (2012)

[Sennrich 16] Sennrich, R., Haddow, B., and Birch, A.: Neural Machine Translation of Rare Words with Subword Units, in *Proceedings of ACL*, pp. 1715–1725 (2016)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in *Proceedings of NIPS* (2014)

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

大量の Twitter 画像を用いた Conditional Cycle GAN による食事写真カテゴリ変換

堀田 大地 Daichi Horita　成冨 志優 Shu Naritomi　丹野 良介 Ryosuke Tanno　下田 和 Wataru Shimoda　柳井 啓司 Keiji Yanai

*1電気通信大学 情報理工学部
The University of Electro-Commnications, Tokyo

This paper proposes “Food Image Transformation” based on the Conditional Cycle GAN (cCycle GAN) with a large-scale food image data collected from the Twitter Stream for more than five years. By the experiments, we showed that two hundred and thirty thousand food images with cCycle GAN enabled very natural food photo transform among ten kinds of typical Japanese foods: ramen noodle, curry rice, fried rice, gyodon, cold Chinese noodle, spaghetti with meat source, white rice, eel bown, and yakisoba.

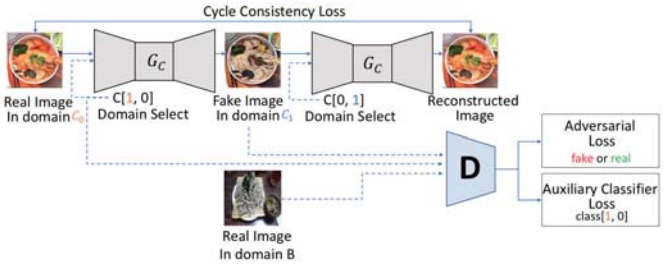


図 1: Conditional CycleGAN のネットワーク全体

1. はじめに

近年，生成モデルと深層学習を組合せた深層生成モデル Generative Adversarial Networks(GAN) が従来手法と比べてより本物らしい画像を生成できるとして注目を集めている．訓練データの分布に近似するよう最適化することで本物らしい画像の生成に成功している．GAN の研究において用いられるデータセットは CelebA データセットの顔画像や MNIST の数字文字画像，LSUN の居住画像など，ある程度パターンが限られる画像群が通常用いられる．また，最近では，[1] のように衣服画像へのデザイン転送タスクといった新しい課題を提案し，GAN や Neural Style Transfer のような深層学習技術を応用する研究がでてきている．一方で，本研究のような食事に限定した食事画像生成・変換に関する研究は未だ存在しないのが現状である．

本研究では，深層学習技術を用いて，自動的に食事画像を変換するという新しい問題に焦点を当てる．食事画像と変換先のカテゴリ情報を入力すると，リアルタイムに特定のカテゴリに変換された食事画像を生成することを目指す．深層学習による画像変換手法の 1 つである CycleGAN の手法を拡張し，1 つの変換ネットワークで複数のカテゴリへと変換可能とする conditional CycleGAN を用いた食事画像変換手法を提案し，10 種類 23 万枚の Twitter から収集した食事画像に適用することで，きわめて自然な 10 種類食事間での食事カテゴリの相互変換が可能であることを示す．

2. 関連研究

GAN は一様分布や正規分布などからノイズベクトル z をサンプリングするため，生成される画像のコントロールをすることができない．そこで，GAN の構造に条件付き信号 conditional vector を付与することで，条件付き確立分布を学習するモデルに拡張したものが cGAN である．一方で，cGAN には入力画像を潜在表現に落とし込む機構 (Encoder) が欠けているため，画像の変換は行うことができない．pix2pix は Adversarial Loss と ConvDeconvNet を組合せることで，画像のペア集合間の変換方法を学習することが可能となり，線画彩色や白黒画像のカラー化などの変換を学習させることができる．

[2] では学習データ間 X, Y の写像を学習する方法が提案された．通常の GAN で用いられる損失関数に再構築誤差である Cycle Consistency Loss を追加することで,「集合 X, Y に共通する構造を保って」変換する写像関数の学習に成功している．よって，本研究においても Cycle Consistency Loss による制約を設けることで,「集合 X, Y に共通する構造を保って」，つまりは，食事画像であるならば，食事の部分のみ別のカテゴリの食事に変換し，それ以外は，元の形状を保ったまま変換されることが可能になると考えた．

3. Conditional CycleGAN による画像変換

本節では，まず関連する画像変換技術である pip2pix [3], cycleGAN について説明し，その後，本研究で用いる Conditional CycleGAN (cCycleGAN) [2] について説明を行う．

3.0.1 pix2pix

pix2pix [3] は conditional GAN の一種であるが，通常の GAN では，一様分布や正規分布からサンプリングしたノイズベクトル z を Generator への入力とするが，pix2pix や後述する CycleGAN では，画像 x を Generator の入力とする点が大きく異なる点である．入力に用いていた乱数 z は直接サンプリングする代わりに Generator の複数の層に Dropout でノイズを加えるように代替されている．

pix2pix では，式 1 で表される cGAN の損失関数に加えて，より本物らしい画像を生成するために，式 2 の L1 正則化項の追加と Discriminator のベース構造に [4] で提案された PatchGAN を組合せた式 3 が最終的な pix2pix の損失関数と

なる．入力には変換前と変換後の画像のペアを必要とし，(変換元画像，変換先画像) or (変換元画像，Generator が生成した画像) のいずれのペアであるかを Discriminator に判断させるように学習する．

$$L_{cGAN}(G,D) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}[\log D(x,y)] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}}[\log(1-D(x,G(x,z)))] \quad (1)$$

$$L_{L_1}(G) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}}[||y-G(x,z)||_1] \quad (2)$$

$$G^* = arg\min_G \max_D L_{GAN}(G,D) + \lambda L_{L_1}(G) \quad (3)$$

3.0.2 CycleGAN

pix2pix[3] では，変換前と変換後の画像の 1 対 1 ペアを必要とする制限があったが，CycleGAN[2] ではドメイン間の写像を学習できるように拡張することで，1 対 1 に対応せずとも学習が行えるのが特徴である．ここで，ドメイン X とドメイン Y があるとして，$X \to Y$ への写像を G，その逆写像 $Y \to X$ を F と定義する．また，入力が G によって生成された偽物の X か元の X のデータかを判別する D_Y，入力が Y によって生成された偽物の Y か元の Y のデータかを判別する D_X をそれぞれ定義する．この G, F, D_X, D_Y を式 4 と式 5 の 3 つの損失の和で表される式 6 を用いて学習する．式 4 は Vanilla GAN で用いられる Adversarial Loss そのままであるが，式 5 は Cycle Consistency Loss と呼ばれるもので，ドメイン X に属する x から生成された $\hat{Y}$ を再度，ドメイン X に属する $\hat{x}$ に戻しても元のドメイン X に一致するように制約をかけるものである．この Cycle Consistency Loss を小さくすることは，$G(F(x))$ により変換した結果がそれぞれ元のデータを再構築できるだけの情報を保持することを意味する．よって，学習に成功した場合は，$G(F(x))$ とした場合,「ドメイン X とドメイン Y に共通する構造を保ったまま，一方のドメインに属するデータをもう一方のドメインのデータに変換する」写像関数が得られることになる．

$$L_{GAN}(G,D_Y,X,Y) = \mathbb{E}_{\boldsymbol{y}\sim p_{data}(\boldsymbol{y})}[\log D_Y(y)] + \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log(1-D_Y(G(x))] \quad (4)$$

$$L_{cyc}(G,F) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(x)}[||F(G(x))-x||_1] + \mathbb{E}_{\boldsymbol{y}\sim p_{data}(y)}[||G(F(y))-y||_1] \quad (5)$$

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y,X,Y) + L_{GAN}(F,D_X,Y,X) + \lambda L_{cyc}(G,F) \quad (6)$$

3.0.3 conditional CycleGAN

図 1 として conditional CycleGAN(cCycleGAN) の模式図を示す．CycleGAN [2] を conditional 化することで，1 つの Generator で複数のカテゴリへと変換可能とする conditional CycleGAN に拡張してある．CycleGAN の conditional 化には，[5] と同様に [6] で提案されている分類誤差項 Auxiliary Classifier Loss を Discriminator に追加することで実現する．本物か偽物かの判断をさせるだけでなく，Discriminator にどのカテゴリに属する画像かの識別も同時に学習させることで，複数のカテゴリに変換可能な Generator の学習を行った．こうすることで，Generator は単に Discriminator を欺くように画像を生成するだけでなく，Discriminator の識別エラーを最小限に抑えるように偽物のサンプルを生成できるようになる．つまり，各カテゴリのサンプルを生成できるように最適化されることを意味する．

よって，最終的な損失関数は，Adversarial Loss L_{adv} に式 7 で表される Cycle Consistency Loss と式 8，式 9 で表される Auxiliary Classifier Loss にそれぞれの重みバイアス項 λ_{ccl} 及び λ_{acl} を追加した式 10，式 11 を conditional CycleGAN の損失関数として用いた．

$$L_{ccl} = \mathbb{E}_{x,c,c'}[||x-G(G(x,c),c')||_1] \quad (7)$$

$$L_{acl}^{real} = \mathbb{E}[-\log D_{acl}(c'|x)] \quad (8)$$

$$L_{acl}^{fake} = \mathbb{E}_{x,c}[-\log D_{acl}(c|G(x,c))] \quad (9)$$

$$L_D = L_{adv} + \lambda_{acl} L_{acl}^{real} \quad (10)$$

$$L_G = L_{adv} + \lambda_{acl} L_{acl}^{fake} + \lambda_{ccl} L_{ccl} \quad (11)$$

4. 実験

4.1 学習データ

Cycle Consistency Loss を追加することで,「集合 X, Y に共通する構造を保って」変換することが可能である．そのため，学習データに「共通する構造」がある方が変換が上手くいくと推測される．よって今回は,「丼」という制約を設けて UECFOOD-100[7] の 100 カテゴリの食事の中から「丼」の構造をもつ 10 個のカテゴリを選出した．その 10 カテゴリについて高品質な食事画像の選別のために，2011 年より継続的に Twitter Stream より収集している食事画像データベース [8, 9] の中から UECFOOD-100[7] で学習した食事認識エンジンを用いて，各カテゴリ毎に認識精度が高い順にリランキングした結果から表 1 にある枚数分を学習データとした．この中で「ラーメン」のカテゴリに至ってはその種類の多様性が他のカテゴリと比べて高かったため（例えば,「二郎系のラーメン」は基本的に「丼」からはみ出るほどの具材が乗っているため，他のラーメンと比べて差が大きい．つまり,「共通する構造」が同カテゴリであるが，差が大きくなってしまい，学習が難しくなる恐れがある.），図 2 の処理を行った．8 万枚の「ラーメン」画像に対して，ImageNet で学習済みの VGG16 を特徴抽出器として用い，224x224x3(150,528 次元) を fc6 層 (4,096 次元) まで特徴量を圧縮して k-means により k 個のクラスタ (今回は，k=8 に設定した) に分割を行った．想定していた通り,「二郎系ラーメン」が大部分を占めるクラスタを得られたため，そのクラスタを除外した画像を「ラーメン」カテゴリの画像とした．全ての学習において，訓練 9 割，テスト 1 割となるように配分した．

4.2 学習モデル構造

conditional CycleGAN のネットワークは，基本的に CycleGAN [2] と同一である．変換ネットワーク (Generator) は [10] で提案された ConvDeconvNet の中間層に Residual Block を何層も積層する FastStyleNet の構造を用いて 256x256 の画像を学習に用いた．なお，conditional signal は one-hot vector で表現し，入力画像サイズにブロードキャストした後に，入力画像とチャネル方向に結合して，Generator に入力している．また，Discriminator には [4] で提案された PatchGAN を採用してある．重みの更新頻度は Discriminator を 5 回更新した後に Generator を 1 回更新するようにした．学習は NVIDIA Quadro P6000 を利用しバッチサイズ 32，最適化手法には Adam を用いて 20epoch 繰り返した．テスト時は 512x512 の画像を生成するようにした．

表 1: 学習データ

カテゴリ	学習枚数
冷やし中華	13,499
ミートスパゲティ	7,138
蕎麦	3,530
ラーメン	74,007
焼きそば	24,760
白飯	21,324
カレーライス	34,216
牛丼	18,396
うな重	5,329
炒飯	27,854
合計	230,053

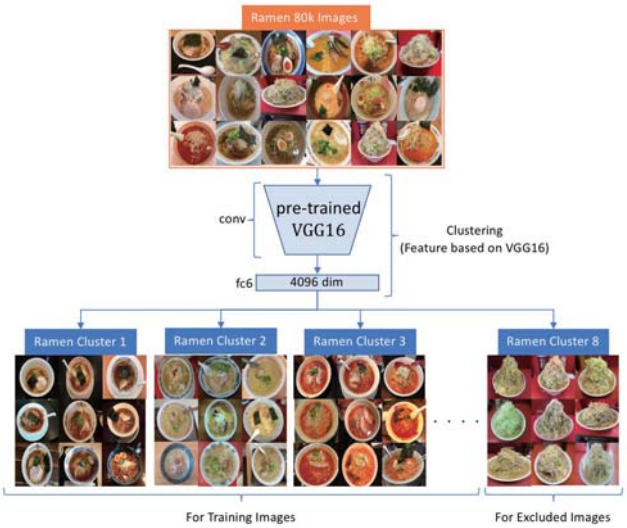


図 2: 多様性があるカテゴリに対するクラスタの構築

4.3 食事画像変換結果

本手法により変換した結果を図 3 に示す．最左列を入力画像として，最上部の 10 カテゴリのドメインへ同時に変換した例を示してある．食事が複数品目ある場合に対しても正確に食事領域のみ対象のドメインへと変換できていることがわかる．再構築誤差 Cycle Consistency Loss により「ドメイン X ドメイン Y に共通する構造を保ったまま，あるドメインに属するデータをもう一方のドメインのデータに変換する」写像関数の学習に成功し，「共通構造=丼，器，食器」の概念を Generator が獲得していることを意味する．また，分類誤差 Auxiliary Classifier Loss を導入することで Generator は単に Discriminator を欺くように画像を生成するだけでなく，Discriminator の分類エラーを最小限に抑えるように偽物のサンプルを生成できるようになり，各ドメインのサンプルを生成できるように最適化されることで，歪みや GAN に特有のブラーが掛かっていない高いクオリティで変換できていることがみてとれる．

4.4 学習に用いるデータ数の変化による変換結果のクオリティへの影響

学習に用いるデータ数が表 1 と比べて小規模な場合，変換結果のクオリティがどのように変化するかの考察を行った．学習に用いるデータセットは 3 種類あり，まとめると以下のようになる．「白飯」「冷やし中華」「牛丼」「カレー」の 4 カテゴリ

図 3: 食事画像変換結果

について，1 カテゴリあたりの画像枚数，総画像枚数がどのようなクオリティ変化を示すの考察を行った．

1. 1 カテゴリ 1 千枚．合計 1 万枚のデータセット．
2. 1 カテゴリ 1 万枚．合計 10 万枚のデータセット．
3. 表 1 の合計約 23 万枚のデータセット．

各条件により学習したモデルで変換した画像を総学習枚数が少ない順に左から並べたものを図 4，図 5 に示す．各カテゴリ千枚の比較的小規模なデータセットでも変換先ドメインの大域的特徴を捉えることには成功しているが，局所的にみると細かいディテールまでは再現できていないようにみえる．つまり，画像枚数が多ければ多いほど，大域的特徴に加えて局所的な特徴をもった細部の細かい部分まで正確に変換先のドメインに変換可能な写像関数の学習ができている結果となった．また，図 4 のカテゴリ「冷やし中華」の変換結果に着目すると，1 列目は 1 千枚，2 列目は 1 万枚，3 列目は表 1 にある通り，1.3 万枚と 2 列目と 3 列目で画像枚数は 3 千枚ほどしか変わらない．しかし，2 列目より 3 列目の方が細かい部分まで変換できていることがわかる．一方で，「冷やし中華」以外の画像枚数も考慮すると，2 列目は総学習枚数 10 万枚に対し，3 列目は 23 万枚の大規模データセットを用いている．他カテゴリの画像から得られた特徴も上手く変換結果に反映されていることがこの結果から伺えるが，これは，1 つの Generator で複数のカテゴリに変換可能にすることで，「食事変換」という共通特徴を Generator が獲得していることを意味する．つまり，1 つの生成器が複数のカテゴリへの変換を担うことで，画像枚数が少ない特定のカテゴリが存在した場合でも，どのカテゴリへも一定の質を保って変換することが可能としていることになる．

5. まとめ と 今後の課題

本研究では，深層学習技術を用いて，自動的に食事画像を生成・変換するという新しい問題に取り組み，CycleGAN の手法を拡張した conditional CycleGAN を用いることで，

1. 変換前と変換後で共通構造を保ったままの変換
2. 複数のカテゴリへの変換を行うことで，変換カテゴリの共通特徴の獲得による変換クオリティの向上

を実現し，実験により実際に高品質に食事画像が変換可能であることを示した．

今後の課題としては，現状，学習したモデルの有効性を示すために，主観的な定性評価しか行っていないため，他者による客観評価実験や変換した画像が期待するターゲットドメインへと変換できているかについて，食事画像分類問題を解くことで定量評価としたい．また，本研究で学習したモデルを用いてモバイルアプリとして実装する予定である．

謝辞 本研究は JSPS 科研費 15H05915, 17H01745, 17H05972, 17H06026, 17H06100 の助成を受けたものです．

参考文献

[1] S. Jiang and Y. Fu. Fashion Style Generator. In *Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.

[2] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proc. of IEEE International Conference on Computer Vision*, 2017.

[3] P. Isola, J. Zhu, T Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2017.

[4] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context Encoders: Feature Learning by Inpainting. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[5] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. arXiv:1711.09020, 2017.

[6] A. Odena, C. Olah, and J. Shlens. Conditional Image Synthesis With Auxiliary Classifier GANs. In *Proc. of the 34th International Conference on Machine Learning*, 2017.

[7] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2012.

[8] 柳井 啓司 河野 憲之. ラーメン vs カレー：2 年分のログデータと高速食事画像認識エンジンを用いた twitter 食事画像分析とデータセット自動構築. In **電子情報通信学会パターン認識・メディア理解研究会**（*PRMU*）, 2013.

[9] K. Yanai and Y. Kawano. Twitter food image mining and analysis for one hundred kinds of foods. In *Proc. of Pacifit-Rim Conference on Multimedia (PCM)*, 2014.

[10] J. Johnson, A. Alahi, and L.F. Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Proc. of European Conference on Computer Vision*, 2016.

図 4: 学習に用いるデータ枚数のクオリティへの影響結果 (1). 左から入力入力画像,「白飯」カテゴリへの 1 万枚学習モデルでの変換結果，10 万枚モデル結果, 23 枚万枚モデル結果,「冷やし中華」カテゴリへの 1 万枚モデル変換結果，10 万枚モデル結果, 23 枚万枚モデル結果.

図 5: 学習に用いるデータ枚数のクオリティへの影響結果 (2). 左から入力入力画像,「牛丼」の 1 万枚モデル結果，10 万枚モデル結果, 23 枚万枚モデル結果,「カレー」の 1 万枚結果，10 万枚モデル結果, 23 枚万枚モデル結果.

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

修飾節付与による複文のニューラル生成

Divide and Generate: Neural Generation of Complex Sentences

尾形朋哉 *1 Tomoya Ogata　小町守 *1 Mamoru Komachi　高谷智哉 *2 Tomoya Takatani

*1首都大学東京 Tokyo Metropolitan Univercity　*2トヨタ自動車株式会社 TOYOTA MOTOR CORPORATION

In recent years, research on chat dialogue systems has been attracted much attention. Typical chat dialogue system selects and outputs an appropriate response from a dialogue database for the input user utterance. However, it is not possible to output an appropriate response if the coverage of the database is limited. Therefore it is necessary to augment the database beforehand. In this research, in order to amplify various kinds of responses in the database, we propose a task to generate a complex sentence from a simple sentence. We first divide a complex sentence into a main clause and a subordinate clause to learn a generator model of modifiers, and then use the model to generate a modifier clause to create a complex sentence from a simple sentence. We present an automatic evaluation metric to estimate the quality of models.

1. はじめに

近年，コンピュータが人と会話を行うための雑談対話システムの研究が行われている．雑談対話システムには，入力されたユーザ発話に対して，データベースから適切な応答文を選択して出力する手法が広く使われている [Ji 14]．この手法では，データベース内に十分な応答文がない場合，ユーザの発話に対して適切な応答文を出力できないことがある．したがって，ユーザのあらゆる発話に対応するためには，事前に多様な応答文を生成し，データベースを拡張する必要がある．

応答文を生成するための研究の 1 つとして，テンプレートとして用意された応答ルールに単語を当てはめることで応答文を生成する手法がある [Stent 04]．この手法は人手で応答ルールを作成するため，文法的に違和感のない応答文を生成できるが，人手で作成できる応答ルールには限界があり生成できる文の種類は限られてしまう．

一方，ニューラルネットを用いて対話行為という構造化されたデータから応答文を生成する研究がある [Wen 15][Dušek 16]．これらの研究では入力の制約を満たす応答文の生成に成功しているが，対話行為がアノテーションされたコーパスが大量に必要になる．しかし，日本語において対話行為がアノテーションされた利用可能な大規模コーパスは存在しない．

そこで，本研究では単文に対して修飾節を付与することで，複文を生成することを目指す．複文の生成を人手で行う場合，修飾節を挿入する位置や主節に対してふさわしい修飾節を考えなくてはならず，大量に複文を生成するための手間が大きい．しかし，本研究のように自動的に複文を生成することで，生成された文が自然かどうかを人手で判断するだけでよく，複文の生成にかかる人的コストを減らすことができる．本研究はドメインごとに一定数利用可能な単文を入力として，それらを元にした複文を生成することでデータベースの拡充を目指す．

本研究の入力と出力は表 1 に示す通りである．本研究の入力は修飾可能な名詞を含む文であり，出力は入力された文に修飾節を加えた文となる．修飾節は入力の文より長くならず，出力の長さは入力の高々2 倍である．

連絡先: 尾形朋哉，首都大学東京，ogata-tomoya1@ed.tmu.ac.jp

表 1: 本研究における入力と出力の応答文の例

	入力	出力
応答文 1	車に乗りました	彼に借りた車に乗りました
応答文 2	方法を探しています	この先に進む方法を探しています

本研究の主な貢献は以下の通りである．

- 入力の応答文に対して修飾節を拡張した応答文を生成するタスクを提案した．
- 複文を含む応答文データベースから自動的に修飾節拡張のためのコーパスを生成し，ニューラルネットワークによって修飾節拡張生成モデルを学習する手法を提案した．
- 修飾節拡張に対する評価尺度を検討し，全てを End-to-end で行うベースラインと比較して，修飾節拡張をパイプラインで行う手法が流暢性と多様性の両方で改善することを示した．

2. 関連研究

文書要約や文圧縮が入力の文書や文から部分的な情報を生成する研究であるのに対して，部分的な情報から文を生成する研究がある [Wen 15][Dušek 16]．Wen らや Dušek らの研究は，入力として与えられた対話行為の制約を満たすような自然言語文を生成する．これらの研究において出力に含むべき内容は全て入力の対話行為に与えられているが，本研究では修飾節を指定せず，学習データに基づいた修飾節を生成している．

また，入力として情報を与えずに学習データに基づいた文生成を行う研究がある [Bowman 16]．この研究は事前にデータから学習した確率分布に従い文を生成することができ，学習データに基づき修飾節が自動で生成される点で本研究と類似しているが，本研究は入力を制約として修飾節を生成している．すなわち，本研究は生成される文を入力側に与える情報で操作することができるので，特定のドメインに含まれる文や特定のキーワードを含んだ文を用いることができる．

さらに，複文を生成することを目的とした研究がある [緒方 16]．緒方らは 2 つの単文が与えられたとき，従属節とする単文をルールベースで変形し，主節の指定された位置に挿

表 2: 修飾節の構成文節として抽出する文節の例

文節の条件	例
動詞を含む	飲む，連れてきた
名詞述語を含む	中枢である

入することで複文を生成する．緒方らの研究では従属節と主節に当たる単文を入力として与えなくてはならないが，本研究では学習データに基づいて主節にふさわしい従属節を自動で予測し，修飾節を生成する．また，緒方らの研究では入力の 2 つの文の他に，従属節の変形や挿入位置に関しても入力として与える必要がある．本研究ではルールで検出した位置に対し，学習データに基づいて自動で修飾節の生成を行う．

3. 修飾節付与による応答文生成

本研究では既存の応答文データベースに存在する単文の応答文を入力として，修飾節を挿入して複文の応答文を生成することを目指す．

本研究の生成モデルとして Attention 機構付きの Encoder-Decoder を用いる [Bahdanau 15]．Encoder-Decoder は任意長の入力列から任意長の出力列を出力するように重みを学習するモデルである．Attention 機構により，単語を生成する際に対応する入力側の情報を考慮することができる．本研究において入力列は文，出力列は入力の文に修飾節を付与した文からなるパラレルコーパスを用いて Encoder-Decoder を学習する．

3.1 節において修飾節付与を行うためのコーパスの作成方法を説明し，修飾節の拡張を行うモデルを 3.2 節で提案する．さらに，生成された文の評価尺度を 3.3 節にて説明する．

3.1 コーパス

修飾節拡張を学習するためには単文または複文に修飾節がアノテートされた複文からなる修飾節拡張パラレルコーパスが必要であるが，それを人手で作成するのは困難である．したがって，本研究では複文からなる生コーパスから修飾節を削除することで，疑似的に修飾節拡張パラレルコーパスを作成した．本研究の実際の入力は単文を想定しているが，単文を収集するためにはルールにより複文や重文を除く必要があるため，十分な量コーパスが集まらない可能性がある．そのため，本研究では学習データは単文に限定せず収集し，修飾節拡張の学習に用いた．本研究では以下のルールを用いて修飾節を抽出した．なお，係り受け解析には CaboCha を用い，形態素解析には MeCab+IPAdic を利用した．

1. 文を係り受け解析する．
2. 品詞細分類が “一般”または “固有名詞”の名詞を含む文節に係る文節を抽出する．
3. 抽出した文節において，表 2 に該当する文節を修飾節の構成文節として抽出する．なお，複数存在する場合はランダムに 1 つ選択する．
4. 修飾節の構成文節に含まれる文節に係る文節を抽出し，修飾節の構成文節に追加する．
5. 手順 4 を繰り返し行い，抽出されたすべての文節を 1 つの修飾節とする．

本研究では，単文と複文からなるコーパスでモデルの学習をし，テスト時には単文のみのコーパスに対して修飾節付与を行なうように，モデルの学習と実際の文生成で異なるコーパスを使える．なお，モデルの学習を行うコーパスと実際に文生成に使うコーパスのドメインがかけ離れている場合，不自然な修飾節が挿入されることがあるので学習コーパスのドメインには気をつける必要がある．

Algorithm 1 挿入位置の検出

```
chunks ← 文を係り受け解析
名詞 index リスト : noun_idx ← []
検出フラグ : d ← FALSE
for i = 0 to |chunks| − 1 do
  if chunks[i] に品詞が名詞の単語を含む then
    append i to noun_idx
    if chunks[i] に係る動詞または助動詞が存在しない then
      chunks[i] の名詞の前後に特殊記号を挿入
      d ← TRUE
      break
    end if
  end if
end for
if d ≠ TRUE then
  i ← min(noun_idx)
  chunks[i] 内の名詞の前後に特殊記号を挿入
end if
return chunks
```

表 3: パイプラインモデルにおける修飾節生成モデルの入力と出力の応答文

	入力	出力
応答文 1	<ins> 車 </ins> に乗りました	彼に借りた車に乗りました
応答文 2	<ins> 方法 </ins> を探しています	この先に進む方法を探しています

3.2 パイプラインモデル

本研究のベースラインは Encoder-Decoder を用いて，入力の文に対して修飾節を付与した文を End-to-end で生成するモデル (End-to-end モデル) である．End-to-end モデルは，入力された文に対して修飾節を挿入する位置の検出と，修飾節の生成を同時に行なっているため，問題を複雑にしている可能性がある．本研究では，挿入位置の検出と修飾節の生成を別々に行うことで，より頑健に修飾節を生成するパイプラインモデルを提案する．

パイプラインモデルは挿入位置の検出をルールベースにより行い，修飾節の生成を Encoder-Decoder を用いて行う．まず，挿入位置の検出のルールを Algorithm 1 に示す．

修飾節の生成の学習に使うコーパスは，表 3 で示すように，3.1 節で作成したコーパスの入力に対して，1 つの単語の前後に特殊記号を挿入したものとなる．このコーパスを用いて Encoder-Decoder を学習することで，入力側で修飾位置を指定した単語に対して，修飾節を生成することができる．

3.3 評価尺度

本研究は入力の単文に対して修飾節を付与することで複文を生成する．この時，生成される修飾節は修飾先の単語にふさわしい単語であれば何でもよいため特定の正解は存在しない．したがって，生成される文を評価するために BLEU などの参照訳を利用する評価尺度を用いるのは適切ではない．しかし，生成される文としては流暢であるものが望まれるため，テストデータにおいて作成した N-gram 言語モデルによる Perplexity

で流暢性を評価する．

また，修飾節を付与した後の文は，元の文より情報量が増えていると考えられる．ここでの情報量の増加は単語のタイプ数の増加に依存するものなので，文中の情報量を評価するための尺度として文内の単語タイプ数を用いることを提案する．本タスクの目的は，流暢性をできるだけ損ねることなく，多様性を向上させることである．

4. 単文に対する修飾節付与

4.1 実験設定

コーパス．会話文コーパスの作成のために，小説投稿サイトである「小説家になろう」に投稿された小説の文章データを利用する．この文章データにおいて「」に囲まれた部分を会話として抽出し，会話文コーパスを作成する．このとき，会話は複数の文を含む可能性があるので，文を正規化したあと，“。”や“?”などの記号を文の区切りとみなして文を作成している．ここで作成された会話文コーパスに対して，3.1 節で説明した手法を用いて修飾節を除いた文と元の文のパラレルコーパスを作成した．テストデータと開発データに関しては，入力として単文を想定しているので，3.1 節のルールに加え，修飾節を 1 つ抽出した後の文が動詞または助動詞を 1 つのみ含む文のみを抽出することで作成した．なお，拡張される修飾節が極端に短い場合や長い場合には，修飾節にほとんど意味がないことや，修飾節が長すぎるために不自然な文になることがあるので，修飾節の文節数が 2 以上のものと修飾節の文字数が元の文よりも短い文のみを利用している．さらに，生成される修飾節が特定の修飾節に偏るのを防ぐため，学習データにおいて拡張される修飾節が同じであるものを取り除いた．ここで作成したコーパスは学習データが 95,234 文で，テストデータと開発データはそれぞれ 1,000 文である．

また，「小説家になろう」から作成されたコーパスによって学習したモデルが，外部ドメインの文に対して正しく修飾節付与を行えるかを実験する．外部ドメインとして使用するデータは対話破綻チャレンジの雑談対話コーパス [東中 14] から単文のみを抽出したものを用いる．このコーパスはシステムとユーザの発話からなる対話コーパスで，本研究ではユーザの発話を抽出して修飾節の拡張を行う．

モデル．単文を入力した時に，適切な位置に適切な修飾節を挿入できるかを End-to-end モデルとパイプラインモデルでそれぞれ実験する．また，パイプラインモデルにおいて探索幅 10 でビームサーチをした時にどのような出力が生成されるかを実験する．ニューラルネットワークのハイパーパラメータは，語彙サイズを 10,000，埋め込み層を 512，隠れ層を 512，バッチサイズを 128 として実験を行なった．単語ベクトルの初期値は学習データで学習した word2vec，最適化アルゴリズムは Adagrad で学習率は 0.01 を用いた．モデル選択はエポックを 20 まで回し，各 dev セットで BLEU が最大になるエポック数を選択した．

評価．それぞれのモデルに対する自動評価は生成された文を N-gram (N=4) 言語モデルによる Perplexity と単語のタイプ数の平均により評価する．なお，N-gram 言語モデルは modified Kneser-Ney 法によるディスカウンティングと補間を用いている．また，人手での評価としてそれぞれのシステムからランダムにサンプルした文のうち，どちらの方が自然な文かを相対評価した．この時，それぞれのシステムの出力文の文としての自然さが同程度である場合も考慮し，優劣がつく文数が 100 文になるまでサンプルした．

表 4: Perplexity と単語タイプ数による評価

	Perplexity	単語タイプ数
End-to-end モデル	54.9	13.85
パイプラインモデル	**46.9**	**14.44**

4.2 実験結果

定量的評価．Perplexity と単語タイプ数による評価の結果を表 4 に示す．パイプラインモデルの方が End-to-end モデルよりも Perplexity が低くなっており，より流暢性の高い出力が出せていると言える．また，単語タイプ数はパイプラインモデルの方が大きく，より情報量の多い文を出せていると言える．

ランダムにサンプルした 210 文について人手で出力文のみを見て流暢性を相対評価した結果，End-to-end モデルが良い : パイプラインモデルが良い : 同程度＝ 32 : 68 : 110 となり，パイプラインモデルの方が流暢性の高い文を生成できていた．

出力例．End-to-end モデルとパイプラインモデルのそれぞれの出力を表 5 に示す．なお，パイプラインモデルの下線部は挿入された修飾節を表している．表 5 の 1 行目のように End-to-end モデルは修飾節内に修飾先の単語を出力してしまうことがあった．一方で，パイプラインモデルは修飾先の単語と同じものを修飾節内に出すことはなかったが，文の主語と重複する単語を出力することがあった．また，表 5 の 2 行目の元の文における “大手”が End-to-end モデルでは似た意味の “大型”に変わることで，不自然な日本語になってしまうことがある．さらに，表 5 の 3 行目や 5 行目のように End-to-end モデルは出力が，おかしな単語が出力され，文法的におかしく，意味の分からない文になる事例が多く見られた．パイプラインモデルは文法的におかしいものはほとんど出力されないが，表 5 の 4 行目や 5 行目の例のように，不自然な位置に修飾節を挿入したり，修飾先に対して適切でない修飾節を生成してしまうことがある．

最後に，外部ドメインの文に対して，パイプラインモデルが実際に生成した文を表 7 に示す．表 7 の 1 行目のように，修飾先の単語がモデルの語彙に含まれるものに関しては比較的正しい修飾節が生成できている．一方，修飾先の単語が未知語であるものに対しては，表 7 の 2 行目のように正しい修飾節が生成できた例もあるが，表 7 の 3 行目のように，正しい修飾節を挿入できないものが多く見られた．特に，未知語に対しては “俺が作った”のようにモデルが生成しやすい修飾節を挿入する例が多く見られた．

5. 考察

パイプラインモデル．End-to-end モデルとパイプラインモデルの両方に共通して，あらゆる入力文に対して汎用的な修飾節が付与されやすいという問題があった．この問題は，モデルを選択するための評価尺度と関係している．本研究では BLEU が最大のモデルを選択して用いたが，修飾節に関しては正解は 1 つに定まらないため，BLEU が最大のモデルが良い修飾節を生成できるとは限らない．また，エポック数が増えるにつれて，生成される修飾節は多様性を増す傾向にあるので，エポック数の高いモデルを何らかの尺度で選択することができれば，多様性のある修飾節を挿入することが可能である．本研究では開発データにおける Loss や Perplexity が上昇し続けたため，BLEU に基づくモデル選択をしたが，N-gram 言語モデルの Perplexy や単語のタイプ数に基づきモデルを選択することも考えられる．

表 5: End-to-end モデルとパイプラインモデルの出力の比較

End-to-end モデル	パイプラインモデル
流石 , この 国 の 英雄 と 呼ば れ た 英雄 と いう べき です	流石 , この 国 を 救っ て くれ た 英雄 と いう べき です な
大型 柄 , 私 が 見 た 分量 ござい ます	この 国 を 守る 大手 柄 , おめでとう ござい ます
ケイ ぅ , 何 か を する コース で ね	俺 が 持っ て き た 手作り チョコ って 名目 で ね
ルシエル 様 は この 街 に ある 飛行船 を 見 られ ませ ん でし た か	私 の よう な ルシエル 様 は 飛行船 を 見 られ ませ ん でし た か
商業 ランク なんて , 勝手 に ある 幻々 する だけ だ よ	俺 が 連れ て き た 商業 施設 なんて , 苛々 する だけ だ よ

表 6: ビームサーチによるパイプラインモデルの上位 5 件の出力

入力	犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , 術 を 完成 さ せる
beam 1	それ を 知っ て いる 犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , 魔力 を 使う 術 を 完成 さ せる
beam 2	私 達 を 襲っ て き た 犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , 俺 たち を 倒す 術 を 完成 さ せる
beam 3	私 達 を 倒し た 犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , 俺 たち を 守る べき 術 を 完成 さ せる
beam 4	私 達 を 含め た 犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , それ を 使う 術 を 完成 さ せる
beam 5	私 達 を 襲っ た 犯人 は , リフレイア という 娘 だ	俺 も できる だけ 早く , 俺 たち を 守る 術 を 完成 さ せる

表 7: 外部ドメインに対するパイプラインモデルの出力（<unk:>は入力側の未知語を表す）

入力	パイプラインモデル
海 は <unk:うきうき> し ま す ね	この 街 に ある 海 は うきうき し ま す ね
<unk:野菜> <unk:ジュース> だけ で す	俺 が 作っ た 野菜 ジュース だけ で す
<unk:カエル> が 好き な ん で す か	俺 が 作っ た カエル が 好き な ん で す か

End-to-end モデルでは似た意味の単語が代わりに出力されたり，おかしな単語が出力されることで不自然な文が生成されることがあった．これらは，End-to-end モデルが修飾節の挿入位置を予測するために隠れ層に保存する，挿入位置に関する情報が単語選択に影響したと考えられる．

パイプラインモデルにおいて不自然な位置に修飾節が挿入される問題に対しては，ルールではなく，RNN を用いて修飾節の挿入位置の検出を行うことで，柔軟な修飾位置の検出ができ，より自然な文が生成されると考えられる．

表 6 にパイプラインモデルにおけるビーム幅 10 のビームサーチにおける上位 5 文の結果を示す．それぞれ流暢性が高く，意味合いの異なる文を生成することができている．したがって，文内で同じ単語を複数回出力してしまう問題に対しては，Decoder でビームサーチをすることで複数候補文を保持し，同じ単語が出現するものに対してペナルティをかけてリランキングすることで，重複が発生する出力を避けることができると考えられる．

修飾先の単語が未知語となる場合に適切な修飾節が生成できない問題に対しては，単語のカテゴリを元にして，物や人物，場所などカテゴリに応じた未知語を表す記号を用意することで，カテゴリに対応した無難な修飾節を生成できる可能性がある．

評価方法. 今回，文の情報量の評価尺度として，文中の単語タイプ数を用いた．この評価尺度ではより多くの種類の単語を使った文が良い評価になるが，文長に対する制限を加えていないため，より長い文が有利になりやすいという問題点がある．また，本研究では助詞のように単体ではほとんど意味のない単語も，名詞や動詞などの文において重要度の高い単語と同列に扱っているため，助詞などで冗長になっている文の方が良い評価になってしまう可能性がある．したがって，名詞や動詞など文にとって重要度の高い品詞のタイプ数を情報量の尺度として用いることも考えられる．

6. おわりに

本研究では入力文に修飾節を付与し，複文を生成する手法を提案した．この手法は修飾節を含む文さえあればよいので，任意のドメインでコーパスを作成し，入力の文に対して修飾節を付与した文を得ることができる．今回の実験で，入力側の単語が未知語になった時にふさわしい修飾節を生成するのが難しいという問題が明らかになった．今後は，入力側での未知語を表す記号を複数用意するなど，入力側の未知語の問題に対して実験を行いたい．

参考文献

[Bahdanau 15] Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, in *ICLR* (2015)

[Bowman 16] Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A., Józefowicz, R., and Bengio, S.: Generating Sentences from a Continuous Space., in *CoNLL* (2016)

[Dušek 16] Dušek, O. and Jurcicek, F.: Sequence-to-Sequence Generation for Spoken Dialogue via Deep Syntax Trees and Strings, in *ACL* (2016)

[Ji 14] Ji, Z., Lu, Z., and Li, H.: An Information Retrieval Approach to Short Text Conversation, *arXiv* (2014)

[Stent 04] Stent, A., Prasad, R., and Walker, M.: Trainable Sentence Planning for Complex Information Presentation in Spoken Dialog Systems, in *ACL* (2004)

[Wen 15] Wen, T.-H., Gasic, M., Mrksic, N., Su, hao P., Vandyke, D., and Young, t. J.: Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems, in *EMNLP* (2015)

[緒方 16] 緒方 健人, 佐藤 理史, 松崎 拓也：日本語文生成器 Haori における複文合成, 言語処理学会 第 22 回年次大会 (2016)

[東中 14] 東中 竜一郎, 船越 孝太郎 : Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション, *SIG-SLUD*, Vol. B4, No. 02, pp. 45–50 (2014)

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

Smart Grid Optimization by Deep Reinforcement Learning over Discrete and Continuous Action Space

Tomah Sogabe[1,2,3], Dinesh Bahadur Malla[2], Shota Takayama[2], Katsuyoshi Sakamoto[2], Koichi Yamaguchi[2], Thakur Praveen Singh[3], Masaru Sogabe[3]

[1]Info-Powered Energy System Research Center, [2]Department of Engineering Science,
[1,2]The University of Electro-Communications, Chofu, Tokyo, 182-8585, Japan
[3]Technology Solution Group, Grid Inc., Kita Aoyama, Minato-ku, Tokyo, 107-0061, Japan

Abstract — Energy optimization in smart grid has gradually shifted to agent-based machine learning method represented by the state of art deep learning and deep reinforcement learning. Especially deep neural network based reinforcement learning methods are emerging and gain popularity to for smart grid application. In this work, we have applied the applied two deep reinforcement learning algorithms designed for both discrete and continuous action space. These algorithms were well embedded in a rigorous physical model using Simscape Power Systems™ (Matlab/Simulink™ Environment) for smart grid optimization. The results showed that the agent successfully captured the energy demand and supply feature in the training data and learnt to choose behavior leading to maximize its reward.

1 Introduction

Energy grid system containing renewable energy resources(RES) such as photovoltaic energy, wind power as well as hydropower have been considered as alternative power supply configuration. It is renovating conventional grid systems, aiming at reducing the emission of CO_2 while mitigating the global warming. A decentralized energy system is more robust and resilient against the unexpected natural disasters, which are frequently occur in countries such as Japan. However, due to the intermittent nature of RES, a mismatch between electricity supply and demand is often encountered and causes instability and limit of power output. As an effective approach to these challenges, smart grid has been proposed and has shown great technological innovation towards intelligent, robust and functional power grid [1][2].

Smart grid evolves energy transmission among different sub-smart grid utilities, which finally contribute to the efficient energy management ecosystem of energy storage, energy supply, balanced load demand over large scale grid configuration. Construction of efficient smart grid system is in principle a control optimization mathematical problem. A wide range of methods have been proposed to tackle this challenge including linear and dynamic programming as well as heuristic methods such as PSO, GA, game or fuzzy theory and so on[3]. In the recent years, studies on energy optimization in smart grid has gradually shifted to agent-based machine learning method represented by state of art deep learning and deep reinforcement learning. Especially deep neural network based reinforcement learning methods are emerging and gain popularity to for smart grid application[4][5].

In this work, we focus on the following issues and tasks:

(1) Different from previous reports, we have developed our deep reinforcement learning algorithm embedded in a rigorous physical model using Simscape Power Systems™ for smart power grid optimization. All the parameters used in smart grid represents the realistic electric circuits and detailed fluctuation regarding the voltage, frequency and phase can be therefore fully revealed, which are not available in previous reports where the constructed smart grid system could not output sufficient information.

(2) For RL, model-free off-policy deep Q-learning suing Matlab™ is developed. DQN is suited for addressing continuous state space and discrete action space. Here we have focused on the discrete action control designed for switching the grid power supply/sell and battery charge /discharge.

(3) For continuous state and continuous action space, we

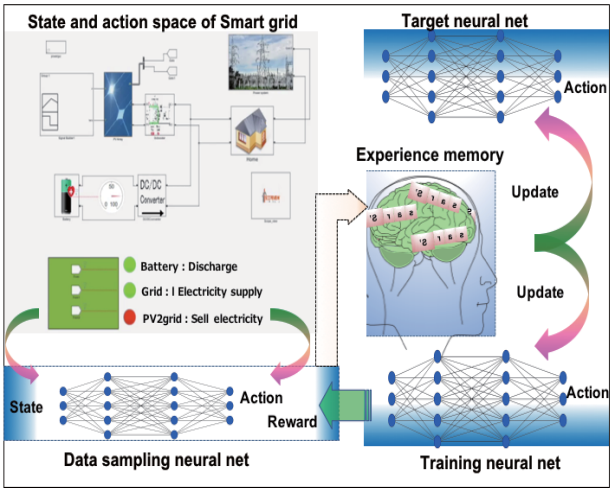


Fig.1. Sketch of smart grid optimization using deep neural network based reinforcement learning algorithm.

Contact: Tomah Sogabe, i-PERC, The University of Electro-communications, sogabe@uec.ac.jp

have self-developed a Gibbs deep policy gradient algorithm, in which we have hybridized the latest deep deterministic policy gradient with the deep actor-critic stochastic policy gradient.

2. ALGORITHM AND MODEL

(i) **Deep Q-Learning (DQN):** A general model which describes the main framework is given as follows. In this

Algorithm: Model-free off-policy Deep Q-learning with Experience Replay

Load the data from data directory for Matlab® simulation
Open Simulink® model
Initialize Simulink® model with initial parameter {first action}
Initialize replay memory D to capacity N
Initialize weight with random weights
Initialize action with zeros (in this simulation 7 actions)
for episode = 1:M do
  Initialise state s = {*starting Value*}
  for t = 1: T do
    set starting and end simulation time for Simulink®
    set update parameter
    start simulation for data which are output of simulation
    simulation out = [Photovoltaic power generation,
                      Battery State of charge(SOC)…
                      Load consumption,
                      DC power generator]
    State= simulation out
    forward state to Deep Q-learning network
    with probability select a random action a_t otherwise select $a_t = max_a Q^*(\varphi(s_t), a; \vartheta)$
    Deep Q-learning network return action index
    Where
    1 = switch from Photovoltaic array to DC power generator
    2 = switch from Photovoltaic array to Load
    3 = switch from DC power generator to Load
    4 = switch for battery charge
    5 = switch from Photovoltaic array to DC power generator
        & switch from Photovoltaic array to Load
    6 = switch from Photovoltaic array to DC power generator
        & switch for battery charge
    7 = switch from DC power generator to Load
        & switch for battery charge
    set action in the Simulink model
    1 = "ON"
    0 = "OFF"
    simulate Simulink Model
    get simulation data from model
    set s_{t+1} = {Next *state*} from getting simulation data from model
    execute action a_t in emulator and observe reward r_t
    store transition (s_t, a_t, r_t, s_{t+1}) in D
    set $s_t = s_{t+1}$
    until s_{t+1} is terminal
    sample random minibatch of transitions (s_t, a_t, r_t, s_{t+1}) from D

$$y_j = \begin{cases} r_j & for\ -\ and\ +\ reward \\ r_j + \gamma max_a Q(s_{j+1}, a^0; \vartheta) & for\ 0\ \ reward \end{cases}$$

    Perform a gradient descent step on $(y_j - Q(\varphi(s_t), a; \vartheta))^2$ for weight upda
  end for
end for

sketch, we adopted deep Q-learning algorithm as an example to illustrate the learning principle: physical model of smart grid simulation environment based on Simscape Power Systems™ was constructed. The state space is always continuous and action space is set either discrete or continuous for off-policy Q learning and deep policy gradient algorithm respectively. A detailed operation flow is given as follows by the form of pseudo-simulation code:

(ii) **Gibbs Deep Policy Gradient (GDPG)**: Deterministic policy is in theory efficient at the late stage of simulation because the policy distribution is less variant and more deterministic. Policy gradient is usually formulated as follows, where η is the policy object function; θ is the function approximation parameter (in neural network, it is the weight w); s and a correspond to the state and action $Q_\pi(s,a)$ is the state-action function under certain policy $\pi(a|s,\theta_p)$ and is the :

$$\nabla\eta(\theta_p) = E\left[Q_\pi(s,a)\,\nabla_{\theta_p} log\ \pi(a|s,\theta_p)\right]$$

policy distribution function. David et al has shown that if the policy is treated as deterministic, the above equation can be reformed as: [6]

$$Q_\pi(s,a)\,\nabla_{\theta_p} log\ \pi\left(a\middle|s,\theta_p\right) = \nabla_a Q(s,a) \qquad (2)$$

and if the action a is approximated as policy action function:

$$a = u(s|\theta^\mu) \qquad (3)$$

using the chain rule $\nabla_a(s,a)$ can be further extended as:

$$\nabla_{\theta^a} Q(s|\theta^a)\nabla_{\theta^\mu} u(s|\theta^\mu) \qquad (4)$$

and then policy parameter θ^μ is updated as the usual gradient decent:

$$\theta^\mu = \theta^\mu + \alpha \cdot \nabla_{\theta^a} Q(s|\theta^a)\nabla_{\theta^\mu} u(s|\theta^\mu) \qquad (5)$$

However, implementing the deterministic policy at the early simulation stage will inevitably cause high variance and slow convergence because the policy is far from optimal policy so the policy distribution is fairly stochastic and less deterministic with high bias. The hybridized algorithm is designed in such a way that both the advantage of deterministic and stochastic policy is assimilated thus a stable learning profile with fast convergence can be achieved.

(iii) **Neural Network Model**: In this work, we use multilayer neural network including four hidden layers to approximate the

$$\Delta\theta^\mu = \begin{cases} \alpha \cdot \nabla_{\theta^a} Q(s|\theta^a)\nabla_{\theta^\mu} u(s|\theta^\mu), & Step_t \\ \alpha \cdot \{r_{t+1} + \gamma V(s_{t+1}) - V(s_t)\,\nabla_{\theta^\mu} u(s|\theta^\mu), & Step_{t+1} \end{cases}$$

state-action value function. The activation function is fixed at hyperbolic-tangent function and epsilon-greedy algorithm is utilized to enhance the exploration in the case of DQN for discrete action and re-parameterization. These techniques were used when using GDPG for continuous action space.

3. RESULTS

Here we present one representative simulated results by employing DQN algorithm to optimize four discrete action controls: (1) Grid on/Battery off and (2) Grid

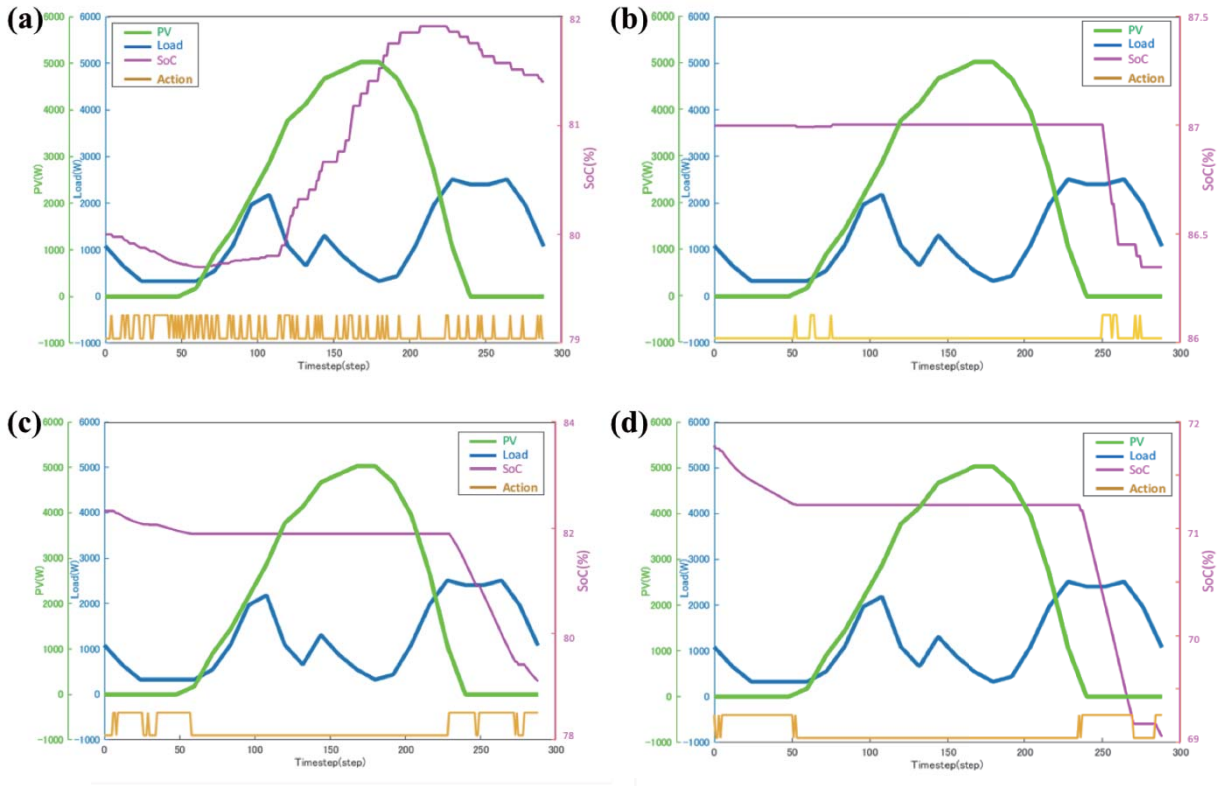


Fig.2. Agent training results using the DQN algorithm. (a) Early training stage; (b) Middle training stage; (c) and (d) latter training stage.

off/Battery on. Action pair (1) was designed to deal with situation when the PV power exceeds the load demand and is able to get the reward by selling the electricity to Grid. On the other hand, Action pair (2) was designed for agent to learn to charge the battery during daytime for the discharge when the PV power is off during night time. The load demand can be supplied through either the grid or the battery and thus the agent has to learn how to control these actions to maximize the reward designed in advance. As shown in Fig.2 the action profile varied during learning process under the preset reward function. At the early stage, the agent showed random actions, as shown in Fig.2(a), where a frequent switch on/off was observed during training. In the middle of training, the agent gradually grasped the inherent feature between demand and supply and learnt how to choose stable action to maximize its reward. This can be seen in Fig.2(b) where constant on or off period were extended and the action profile experiences less noise than Fig.2(a). The agent made a decision to sell its surplus PV power to grid during the day time when its SoC remains high. In the latter training stage as shown in Fig.2(c) and (d), the agent successfully learnt to discharge its battery power during night instead buying electricity from the grid. More detailed results regarding the comparison between discrete action and continuous action will be presented at the conference. Deep insights related how to design various reward functions to train the agent for the different targets posed by smart grid energy system would also be presented.

4. CONCLUSION

We present here a deep reinforcement learning method applied for smart grid optimization. From the preliminary simulation results, the agent was able to catch the feature involved in the balance of load demand, PV power surplus and battery discharge/charge as well as grid integrate. The agent successfully learnt how to tune its action profile to maximize the reward function during training. More detailed results regarding to the comparison between DQN and GDPG and the key role played by reward function will be given at the conference.

REFERENCES

[1] R. H. Khan and J. Y. Khan, "A comprehensive review of the application characteristics and traffic requirements of a smart grid communications network," *Computer Networks*, vol. 57, no. 3, pp. 825-845, 2013.

[2] H. E. Brown, S. Suryanarayanan, and G. T. Heydt, "Some characteristics of emerging distribution systems considering the smart grid initiative," *The Electricity Journal*, vol. 23,no. 5, pp. 64 -75, 2010.

[3] M. R. Alam, M. St-Hilaire, and T. Kunz, "Computational methods for residential energy cost optimization in smart grids: A survey," *ACM Comput. Surv.*, vol. 49, pp. 22-34,Apr. 2016.

[4] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91-99,

2016.

[5] V François-Lavet, Q Gemine, D Ernst, R Fonteneau, "Towards the minimization of the levelized energy costs of microgrids using both long-term and short-term storage devices, "*Smart Grid: Networking, Data Management, and Business Models*, P295-319,2016

[6] S. David, L. Guy, H. Nicolas, D. Thomas, W. Daan, and R.Martin. "Deterministic policy gradient algorithms," *ICML*, 2014

量子自己符号化器の開発

Development of quantum autoencoder

黄川田優太[1], 坂本克好[1], 山口浩一[1],横川慎二[2],曽我部東馬*[1,2]

[1] 電気通信大学 先進理工学科

[2] 電気通信大学 i-パワードエネルギーシステム研究センター

Classical autoencoder is used for compression dimension of the image data and take characteristics, called prior learning. Here we propose quantum autoencoder using quantum annealing. Quantum annealing is optimization technique by adding a wide magnetic field to spin system expressed in isingmodel. But, to solve the optimization problem using quantum annealing, it is necessary to convert the problem into a combination problem of two values(±1) as direction of the spin(up spin:+1, down spin:-1). In this paper, we explain learning method of the image using quantum autoencoder, then we show the result that took characteristics from original image data. Finally, we tried noise reduction from original image data using characteristics.

1. はじめに

近年, 組み合わせ最適化問題を解く量子力学的な手法として量子アニーリングが注目を浴びている. 特に, 量子アニーリングを用いて D-wave systems 社が世界で初めての商用量子コンピュータとして世に送り出した D-wave One が世間を賑わせた事は記憶に新しい. 量子アニーリングを用いた組み合わせ最適化問題では解きたい問題をイジングモデルで表されるスピン系に置き換え, 系のエネルギーを最小化する事によって解を得る. それ故, 問題自体を±1 の二値(+1: スピン上向き, -1: スピン下向き)で表現される組み合わせ問題に変換する必要がある. 2 次元のスピン系で置き換えられた問題のハミルトニアンは(1) 式のようになる. ただし, 簡易化のため外部磁場は 0 としている.

$$H = -J_{ij} \sum_{<i,j>} \sigma_i^z \sigma_j^z - \Gamma(t) \sum_i \sigma_i^x \qquad (1)$$

ここで, 右辺第一項の和は最隣接 4 つのスピン和を表し, J_{ij} はスピン間の相互作用定数, σ^z , σ^x は pauli の z スピンと x スピン, $\Gamma(t)$ は横磁場の強さを表すパラメータを表している. (1) 式の右辺第二項のように, 量子アニーリングでは系のハミルトニアンに横磁場項を導入する事で量子揺らぎによる状態ベクトルの重ね合わせを実現し, 横磁場を徐々に弱めていく事によって最適解を得る事を可能としている. しかし, 系のサイズ(スピン数)が大きくなると一般的な計算機上での計算は困難になる. そこで用いられるのが経路積分量子モンテカルロ法である. これは(1) 式のハミルトニアンから鈴木トロッタ分解を用いて量子系を次元を1つ増やした古典系にマッピングする事により近似的に系の自由エネルギーを求め, モンテカルロ法を用いてそれを最小化する事によって解を得る方法である. 2 次元のスピン系から近似的に得られるエネルギーは(2) 式のようになる.

$$E = 1/m \sum_{<i,j>}^{N} \sum_{k}^{m} J_{ijk} \sigma_{i,k}^z \sigma_{j,k}^z - 1/2 log coth(\beta\Gamma(t)/m) \sum_{i=1}^{N} \sum_{k=1}^{m} \sigma_{i,k}^z \sigma_{i,k+1}^z \quad (2)$$

ここで, m は追加したトロッタ次元, β は系の逆温度である. 本研究では量子アニーリングを用いて二値画像の特徴量を学習し, 元画像からノイズを除去する事を目的とした[鈴木 2018].

連絡先: 曽我部 東馬,電気通信大学 i-パワードエネルギー・システム研究センター, sogabe@uec.ac.jp

2. 量子自己符号化器学習モデル

画像データに対して, 従来の自己符号化器は出力画像が入力画像を再現するようにニューラルネットワークのウェイトを更新し, 画像データの特徴を学習する[岡谷 2017 年]. 本研究ではこの従来の自己符号化器のプロセスを量子アニーリングを用いて行う. 量子アニーリングを用いた最適化では(2) 式で表される系のエネルギーを最小化する事によって解を得た. 量子アニーリングを用いて画像データの特徴量を取るためには画像データを±1 の二値に変換し, +1 を上向きスピン, -1 を下向きスピンに対応させ, 初期状態を作る必要がある. その後, なるべく元画像を再現するように系のエネルギーを最小化していく. 本研究ではこのエネルギーを最小化する過程の中に, (3) 式のようなスピン間相互作用定数 J の更新を組み込む事によって画像データの特徴量を取るモデルを考案した. ただし, J の初期値は[0,1)の乱数である.

$$J_{ijk} = J_{ijk} - \varepsilon E' \qquad (3)$$

$$E' = 1/m \, \sigma_{i,k}^z \sigma_{j,k}^z \qquad (4)$$

ここで, ε は学習率であり, E' は(2) 式を J_{ijk} で微分したもので, (4) 式のようになる.

2.1 相互作用定数 J の設計

本研究では(3)式を用いて相互作用定数 J を更新するために J の行列表現を用いた. その際, 隣り合うスピン間で J の値に相違がないようにしなければならない. 以下の図 2-1 に J の行列表現の例を示す.

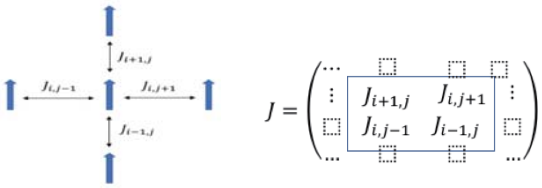


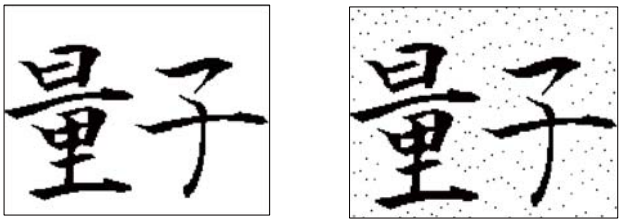
図 2-1　　J の行列表現の例

図 2-1 では中央のスピンに対する最隣接 4 つの相互作用を表しており, 行列 J は各要素を 2×2 のブロックとしたブロック行列と

し, 系のサイズ(縦方向のスピン数 N_Y, 横方向のスピン数 N_x)に対して周期境界条件を課した場合の 行列 J のサイズは $2N_Y \times 2N_x$ となるように設計した.

2.2 画像データ

本研究では[3]を参考に作成した図 2-2 の画像を用いた.



(a) (b)

図 2-2 画像サイズ 130×170

図 2-2 (a) は元画像, (b) は元画像にノイズを入れた画像である.

3. 結果

以下の図 3-1 に横磁場の減少グラフを示す.

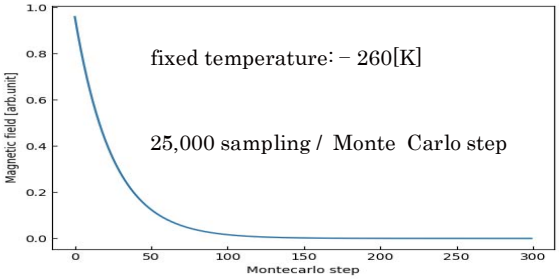
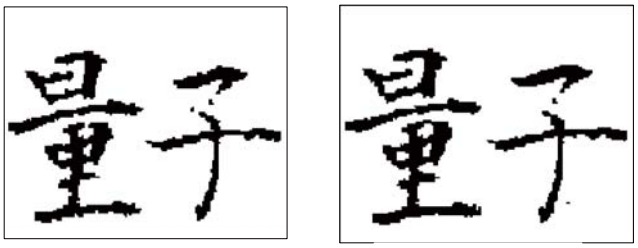


図 3-1 横磁場の減少スケール

図 3-1 は横磁場の減少を 1 モンテカルロステップあたりで表記している. 具体的なスケールの式は(5) 式のようになり, N は系のサイズ(スピン数), C は N におおよそ比例する定数である. 本研究では簡易化の為に $N = C$ とし, t はモンテカルロステップで 300 ステップとした.

$$\Gamma(t) \propto t^{-C/N} \qquad (8)$$

次に, 図 3-2 (a) , (b)に元画像の特徴を取った後, つまり相互作用定数 J を学習した後の画像と, 学習した J を使ってノイズの入った元画像(図 2-2 (b))からノイズを除去した画像を示した.



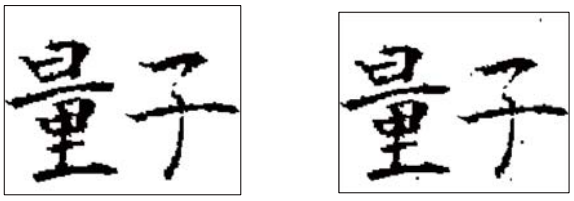
(a) 学習後 (b) ノイズ除去

図 3-2 トロッタ次元数 $m = 10$, 学習率 $\varepsilon = 0.07$

図 3-3 (a) を見ると画素値の境界付近に欠落している部分が多く見られた事からこの付近の特徴が取りきれていない事が分かる. 一方, 図 3-3 (b) を見るとノイズ除去に関しては元画像を概ね維持しつつ, ほとんどのノイズを除去できている事が分かる.

4. 議論

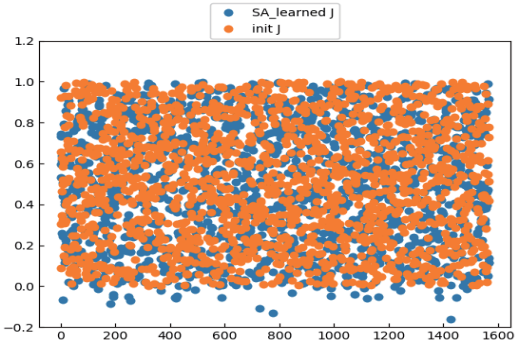
量子アニーリング(QA)は量子力学的な手法を用いて組み合わせ最適化問題を解く手法であったが, 従来の手法としてはシミュレーテッドアニーリング(SA)がある. 本研究において量子アニーリングを用いたことによる利点を評価するために, 以下の図 4-1 に SA を用いた場合の結果を示す.



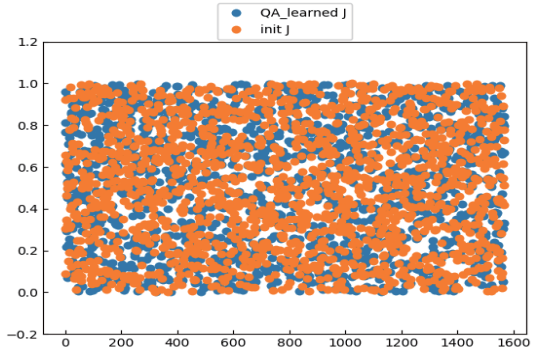
(a) 学習後 (b) ノイズ除去

図 4-1 SA を用いた結果. 学習率 $\varepsilon = 0.06$

図 4-1 (a) を見ると, 学習後の画像は量子アニーリングを用いた場合とほとんど違いが見られなかった. しかし, (b) を見ると, 量子アニーリングを用いた方に比べてノイズを除去できていないのが分かる. この事から SA に比べて量子アニーリングを用いた方がより画像の特徴を取れているからであると考えられた. そこで, 量子アニーリングと SA 双方において学習後の相互作用定数 J がどのようになっているかを画像データセットである MNIST の数字 0 (サイズ: 28×28)を用いた場合を以下の図 4-2 に示した.

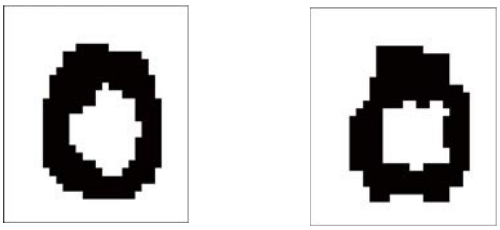


(a) SA を用いた時の J 値



(b) QA を用いた時の J 値

図 4-2 SA と QA による J 値の比較



(a) MNIST の元画像　　(b) QA での学習後

図 4-3 量子自己符号化器での学習結果

また, 図 4-3 に MNIST の元画像(2 値に変換済み)と SA, QA 双方での学習後の画像を示した.図 4-2 は縦軸に J の値, 横軸に J (行列)の要素数を取ったものである. MNIST 画像のサイズは 28 ×28 であるから J の要素数は 56 × 56 = 3136 となるが, 隣り合うスピン間で J の値は同じになる為, 重複を避けた結果 1568 となっている. 本研究では J の初期値を[0,1)の乱数として行なっている. それを踏まえて図 4-2 の(a) を見ると SA を用いた場合は学習後の J 値に負の値が出てきているのが分かる. 一方, (b) を見ると QA を用いた場合は負の値が出てきていない. この結果から, SA を用いて学習した方が QA を用いて学習した場合に比べて画像の境界付近の特徴を取れていると言えた. 図 4-3 からも QA は境界付近の特徴をうまく取れていないと言える. また, 図 3-2 (b) と図 4-1(b) のノイズ除去後の画像を見比べると QA を用いて学習した方が良い結果が得られているのが分かる. この結果から, SA を用いた場合は局所的な画像の特徴を, QA を用いた場合はより大域的な画像の特徴を取っているのではないかと考えられた.

参考文献

[鈴木　2008 年]　鈴木 正: 組み合わせ最適化問題と量子アニーリング: 量子断熱発展の理論と性能評価.

[田中　2009 年] 田中宗, 栗原賢一, 宮下精二 ：量子アニーリング法を用いたクラスタ分析.

[西森 2014 年] 西森 秀稔: 量子アニーリングの数理, 物性研究・電子版, vol.3, No.3,033203, 2014 年

[岡谷 2017 年] 岡谷 貴之: 深層学習, 講談社

[1] 物理のいらない量子アニーリング入門: http://blog.brainpad.co.jp/entry/2017/04/20/160000

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

深層学習による Facebook 広告の CTR 予測

Deep CTR Prediction in Facebook Ads

岩崎 祐貴
Yuki Iwazaki

株式会社サイバーエージェント
CyberAgent, Inc.

Click-through-rate(CTR) prediction of online advertising is important for better user experience and business effectiveness. However, In spite of abundant delivery log, we prone to focus on not accuracy but speed to predict. Especially high-dimensional features such as text and image is not used comparing with user behavior. In this paper, we introduce a CTR prediction model based on deep neural network by combined one-hot feature and huge matrix feature in end-to-end. As a result, our model improves the performance than the baseline model. We also show that visualizing pixel-based bias by using social media specific indicators.

1. はじめに

技術の発展に伴い表現力の高いインターネット広告が増えている中，ユーザーと広告主双方に利点のある広告のクリック率 (CTR) 予測が盛んに行われている．その広告に関連するあらゆる情報，例えば広告の出ているメディアやページ内の表示場所，広告を見たユーザーのデモグラフィック，広告主とその商品，クリエイティブと呼ばれる広告バナーの画像や文面等がクリックを引き起こす要因として考えられる．しかし，高速なレスポンスタイムが求められるインターネット広告では，情報量が多く処理や学習に時間を要してしまうことから，画像や文面を用いて予測性能を改善したり，知識を抽出する研究があまり進んでいない．その解決策として高次元タスクにおいて顕著な成果を上げている Deep Learning を用いて画像やテキストといった各要素を End-to-End に学習させ，得られた結果から配信影響について議論する．また，従来の手法との比較実験から考察とまとめを行う．本研究の目標は，新規の広告に対して CTR を見積もること，また学習されたモデルから CTR に寄与する特徴，配信影響を発見することである．

2. 関連研究

広告 Platform・Deep Learning・CTR 予測というキーワードにおいて様々な手法が提案されている．Microsoft の検索連動型広告の配信データから新規の広告の表示順を推定する研究 [Richardson 07] や，配信設定と検索クエリを Deep Learning で End-to-End に学習する研究 [Edizel 17]，EC サイトにおけるユーザーの行動を商品画像のリストと見立て，紐づく商品情報や広告情報と組み合わせる研究 [Ge 17] がある．また，予測結果だけでなく，学習によって圧縮された特徴から入力画像やテキストとの影響も分析する研究も盛んに行われている．[Mo 15][Zhai 16]．Chen らは広告の配信データを Basic(ユーザー行動，製品情報) と Image(画像) 特徴の 2 つに分け，それを入力に End-to-End で CTR 予測を行う Deep CTR Prediction を提案している [Chen 16] が，Text 特徴が加味されていない．そこで本稿では，この手法に Text 特徴を付与したモデルを提案し，後述する Facebook Ads に対して実験を行う．

連絡先: 岩崎 祐貴，株式会社サイバーエージェント，Email: iwazaki_yuki@cyberagent.co.jp

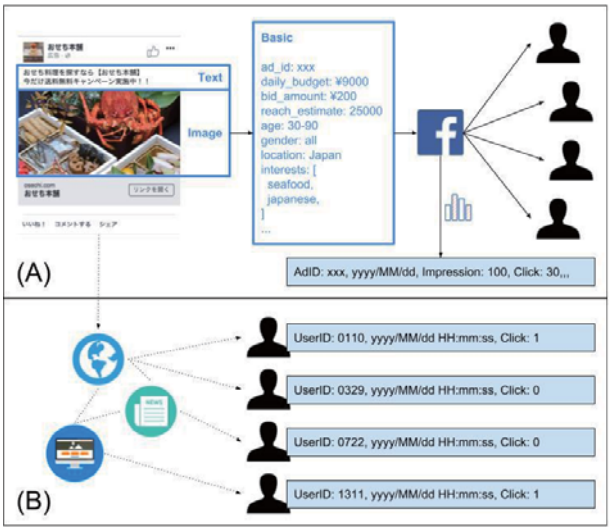


図 1: Facebook Ads(A) vs the others(B)

3. Facebook Ads

今回対象とする配信データの取得元の広告 Platform である Facebook Ads*1(図 1 A) と，その他の一般的な広告配信システム (図 1 B) のデータ構造の違いを説明する．B は 1 ユーザーあたりの広告表示単位でのログなため CTR は二値分類の確率値として表されるのに対し，A は Platform 側で提供されている API 経由で手に入るログであり，ユーザーグループ群に日時単位で何人に表示・クリックされたかの統計が広告 ID という概念単位で集計されるため CTR は回帰問題として数値を予測する．また，Facebook や Twitter に台頭される広告 Platform は各配信システムのログ設計に依存せずに，API により広告の情報が整備されているため，学習データとして構造化，共通化が容易という利点がある．特に本稿では広告バナーを示すクリエイティブの説明文を Text，画像を Image，予算やユーザー属性等，その他の配信設定情報を Basic と呼び，学習時の特徴量 (図 1 A 青枠内) として使用する．

*1 https://www.facebook.com/business/products/ads

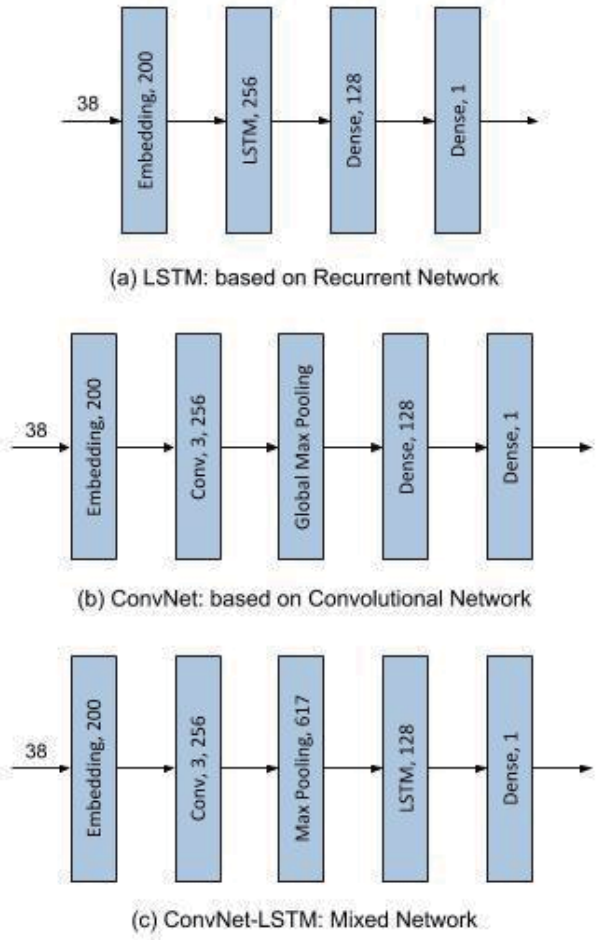


図 2: The architecture of text regression model

表 1: Comparison of text regression performance

Models	*MSE* *(x 1e-5)*	*KL Divrg.* *(x 1e-3)*
LSTM	3.99	2.29
CNN	2.28	1.31
CNN-LSTM	2.22	1.43

4. 予備実験

Text 情報を入力とした CTR 予測タスクに有効なモデルを検証するため，Long short-term memory(LSTM)・Convolutional Neural Network(CNN)・CNN-LSTM ベースの 3 つのモデル (図 2) を構築し，性能比較を行った (表 1)．どのモデルも入力 Text を形態素解析 [Sato 17] 後，Wikipedia の Entity を利用 [Suzuki 16] して分散表現とした．学習には SGD を用い，Chen ら [Chen 16] の設定に倣い，バッチサイズは 128，学習率は 0.01，momentum は 0.9，weight decay は 0.0005 とした．この実験から，性能と学習コストを加味した結果，提案手法では CNN ベースのモデル (図 2 b) を採用した．

5. 提案手法

概要を図 3 に示す．図 1 (A) の青枠にあるように広告の情報を Text, Image, Basic の 3 つの入力に分ける．Text は予備実験のモデル (図 2 b) から最終 Linear 層を取り除いた CNN へ，Image は 197x197 にリサイズし ImageNet で事前学習させた 50 層の Residual Network[He 16] の最終 Softmax 層を取り除いた CNN へ，Basic は one-hot 表現に変換後，3 層からなる Multi Layer Perceptron へそれぞれ入力する．前述した 3 つの Network の Scale の違いを吸収させるために Batch Normalization をかけて結合し，数値として CTR を出力する構造である．

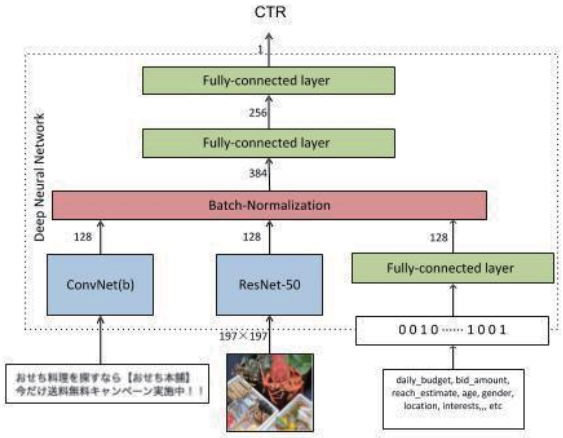


図 3: Deep CTR with Text

6. 実験

6.1 データセット

データ量や整備性の関係から，広告配信システムの配信データではなく，広告 Platform である Facebook Ads の広告情報とその配信実績 (表 2) を用いる．画像やテキストの本質的な特徴を学習させるため，訓練データとテストデータに同じ画像・テキストが存在しないよう無作為に間引いた後の 8 割を訓練データ，残りをテストデータとした．なお，関連研究で上げた Richardson らの研究 [Richardson 07] では広告主をベースに訓練データとテストデータを分け，かつ各広告主毎に 1000 件になるようにサンプリングしているが，彼らよりデータ量が 100 倍近く少なくボリュームが落ちすぎてしまうため，表示回数が 100 以上の広告を扱うという制約のみ採用している．

6.2 実験手法

Deep CTR Prediction[Chen 16] で提案されている特徴量の他，Text 特徴を学習する機構を新たに加えた提案手法を用い，CTR 予測タスクに対して性能比較を行う．具体的には Basic feature だけを使い CTR 予測を行ったモデルを baseline とし，Basic・Image・Text feature の組み合わせによって，Mean Squared Error(MSE)，Kullback Leibler Divergence(KLD)，KLD の改善率の 3 軸の評価指標を示す．なお，学習は 4. 節と同じ設定を使用する．

6.3 結果と考察

学習の進み具合を図 4，性能比較の結果を表 3 に示す．Image や Text 単体のモデルは Baseline よりも悪化するが，Feature を組み合わせると性能の向上が見られ，特に全ての Feature を利用した提案手法が最高性能となった．単体のモデルが悪化した原因については，表 2 で上げているように，複数の Ad で同じ画像やテキストが使われていて絶対数が少ないこと，予算やユーザー属性を含む Basic feature の影響が大きいことが考え

表 2: Statistical Information of Facebook Ads Datasets

Period	*Impressions*	*Clicks*	*CTR(%)*	*Ads*	*Images*	*Texts*
2015/01/01-2017/12/31	8,661,163,035	113,936,635	1.32	348,700	50,720	35,844

表 3: Comparison of model performance

Features	*MSE (x 1e-5)*	*KL Divrg. (x 1e-3)*	*% Imprv.*
Basic (Baseline)	1.95	2.23	-
Image	2.96	2.24	-0.44%
Text	3.99	2.29	-2.69%
Basic+Image	2.11	1.00	55.16%
Basic+Text	2.78	1.84	17.49%
Basic+Image+Text	2.15	0.98	**56.05%**

られる．また，画像と比べテキストの CTR への影響が比較的低いことがわかり，直感的な結果となった．

7. Saliency Mapping

前章の広告単位でのデータとは異なり，画像単位で実績を集計したデータを用い，CTR 予測問題を学習したモデルに対し，Saliency Mapping という手法 [Simonyan 13] を用い，入力画像で CTR 予測する際に寄与した部分を pixel 単位で可視化し，目的変数を Media 特有のいいね数，コメント数に変えた時のモデルとの結果と比較した (図 5)．一番左の入力画像に対し，baseline は Residual Network で ImageNet の分類タスクを学習させたモデル，続いて CTR, いいね数，コメント数である．人間の直感と異なり，一番目立つものというよりは各目的に対して異なる部分が局所的に Mapping されていることがわかる．これを Facebook Ads の他，Twitter や各 Media の配信データに適用すると，その Media のユーザーまたは Media で働いているアルゴリズムが良い広告と判断した基準が含まれている可能性があり，所謂，Media やユーザーはクリエイティブのどこを見ているかともいえる部分が描画でき，新規にクリエイティブを作る際に KPI を変化させた時の微妙なニュアンスの違いの指標としての応用が期待できる．

8. まとめ

広告の基本的な特徴に加え，高次元な画像・テキストを学習し，その性能と配信影響を分析した．その結果，画像やテキスト単体よりも，基本的な配信情報に付与することで性能改善が見られることがわかった．今回は Facebook のみで行った実験では共通特徴が多く使えるため，Twitter Ads や LINE Ads 等，他の Platform に適用できる．画像やテキストといった入力が増え，複雑な表現が学習できるようになったが，その分学習や予測データを集めるコストが増大した．この対策として，得られた学習済みモデルから各入力に対する重みだけを転移学習させることで各入力だけの少量なデータで適用させることが考えられる．Attention Model によるテキストの CTR 寄与部分の可視化，表現力の向上，動画広告への応用は今後の課題である．

参考文献

[Chen 16] Chen, J., Sun, B., Li, H., Lu, H., and Hua, X.-S.: Deep CTR Prediction in Display Advertising, in *Proceedings of the 2016 ACM on Multimedia Conference*, MM '16, pp. 811–820, New York, NY, USA (2016), ACM

[Edizel 17] Edizel, B., Mantrach, A., and Bai, X.: Deep Character-Level Click-Through Rate Prediction for Sponsored Search, in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pp. 305–314, New York, NY, USA (2017), ACM

[Ge 17] Ge, T., Zhao, L., Zhou, G., Chen, K., Liu, S., Yi, H., Hu, Z., Liu, B., Sun, P., Liu, H., Yi, P., Huang, S., Zhang, Z., Zhu, X., Zhang, Y., and Gai, K.: Image Matters: Visually modeling user behaviors using Advanced Model Server (2017)

[He 16] He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)

[Mo 15] Mo, K., Liu, B., Xiao, L., Li, Y., and Jiang, J.: Image Feature Learning for Cold Start Problem in Display Advertising, in *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pp. 3728–3734, AAAI Press (2015)

[Richardson 07] Richardson, M., Dominowska, E., and Ragno, R.: Predicting Clicks: Estimating the Click-Through Rate for New Ads, in *Proceedings of the 16th International World Wide Web Conference(WWW-2007)* (2007)

[Sato 17] Sato, T., Hashimoto, T., and Okumura, M.: Implementation of a word segmentation dictionary called mecab-ipadic-NEologd and study on how to use it effectively for information retrieval (in Japanese), in *Proceedings of the Twenty-three Annual Meeting of the Association for Natural Language Processing*, pp. NLP2017–B6–1, The Association for Natural Language Processing (2017)

[Simonyan 13] Simonyan, K., Vedaldi, A., and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps (2013)

[Suzuki 16] Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N., and Inui, K.: Neural Joint Learning for Classifying Wikipedia Articles into Fine-grained Named Entity Types, in *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pp. 535–544 (2016)

[Zhai 16] Zhai, S., Chang, K.-h., Zhang, R., and Zhang, Z. M.: DeepIntent: Learning Attentions for On-

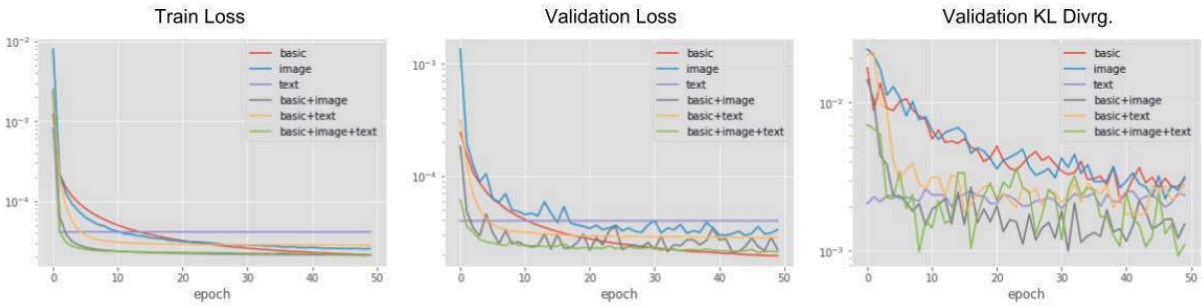


図 4: Epoch vs Metrics

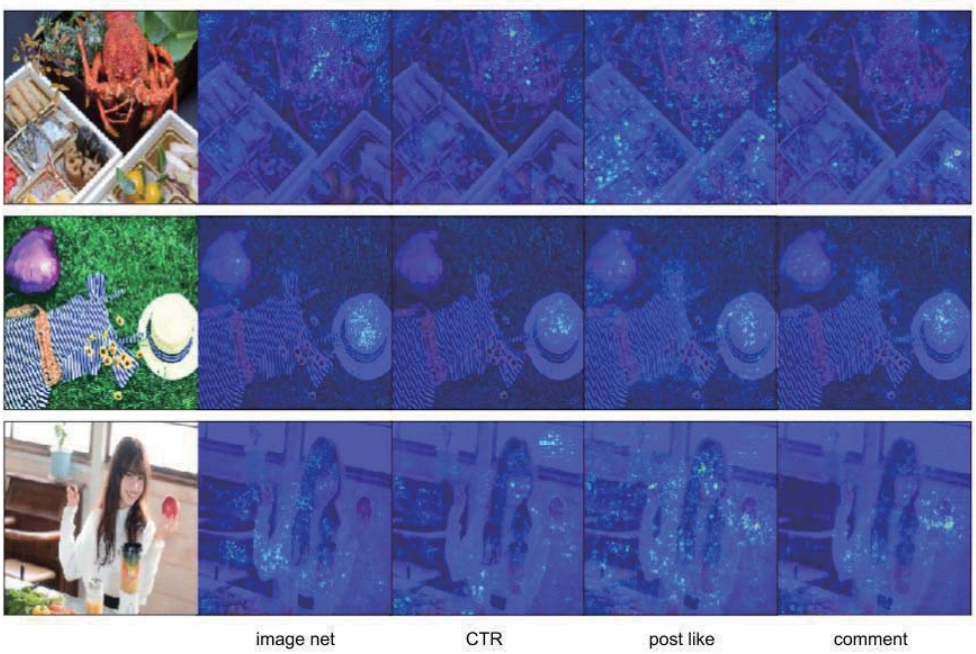


図 5: Why did Facebook say that?

line Advertising with Recurrent Neural Networks, in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 1295–1304, New York, NY, USA (2016), ACM

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

マルチラベル分類と知識ベースの埋め込みを用いたテーブル中の列の概念決定

Determination of Column Concepts in Tables using Multi-label Classification and Knowledge Base Embeddings

竹岡 邦紘 *1 Kunihiro Takeoka　小山田 昌史 *2 Masafumi Oyamada　中台 慎二 *2 Shinji Nakadai　岡留 剛 *1 Takeshi Okadome

*1関西学院大学大学院理工学研究科 Graduate School of Science and Technology, Kwansei Gakuin University

*2NEC データサイエンス研究所 NEC Data Science Research Laboratories

Because relational data represented in tables are not connected with a knowledge base, efficient usage of tables such as semantic search in the web requires the determination of column concepts, that is, annotating a column with an entity in the knowledge base. This paper proposes a new method for annotating columns with entities in the knowledge base using multi-label classification independently. The method also annotates columns in a table with entities using knowledge base embeddings. It uses multi-label classifiers for named-entity(NE) and literal columns. Assuming Markov random field, in which entities corresponding to columns are represented by latent variables, the method maximizes the joint distribution of the latent and observed variables to estimate the optimal entities. Experiments using 183 annotated tables in UCI Machine Learning Repository with WordNet show the superiority of the method.

1. はじめに

構造化された関係データの表現であるテーブルは，一般に体系化された知識ベースとは接続されておらず，テーブルの意味的な索引付けや検索といったテーブルの活用にはテーブルの概念決定が必要になる [Venetis 11, Deng 13, Mulwad 13]．一般に，テーブルの概念決定は，以下の 3 つの対応付けからなる [Zhang 17]．

- テーブル中の各要素に対する，知識ベースにおける要素の概念を表現するノードへの対応づけ．
- テーブル中の各列に対して，1 つの列中の各要素が同一の概念の「実現値」であると仮定し，知識ベースにおけるその概念を表現するノードへの対応づけ．
- テーブル中の複数の列に対して，それらの概念的なつながりの知識ベースに含まれる関係との対応づけ．

本稿では，知識ベースのノード（1 つのエントリ）のことを**エンティティ**と呼び，列に対応する概念の決定，すなわち列に対応するエンティティ推定を扱う．

本研究では，Zhang の手法 [Zhang 17] に従い，テーブル中の列を NE 列とリテラル列の 2 種類に分ける．**NE 列**は列内の各要素がエンティティと対応づけられる列と定義し，**リテラル列**は列内の各要素がエンティティと対応づけられない列と定義する．例えば，NE 列は国名や地域名を要素とする列であり，リテラル列は国や地域の GDP を要素とする列である．つまり，リテラル列は数値とその単位，記号などを要素とする列である．定義よりリテラル列については各要素とエンティティの対応づけを行わない．

本研究では，(1) マルチラベル分類に基づいて，列ごとに列に対応するエンティティを推定する手法と，(2) リテラル列を含む複数の列間の相互依存関係を利用した，複数列に対して同時に列に対応するエンティティを推定する手法を提案する．各列に対応するエンティティは複数存在することを仮定し，列に対応するエンティティ推定をマルチラベル分類によって定式化する．また，列間の相互依存関係をマルコフ確率場によって表現し，列に対応するエンティティを表す潜在変数，タイトルを表す潜在変数，テーブルの要素を表す観測変数の同時確率を最大化することで，テーブル中の複数の列に対応するエンティティを同時に推定する．

リテラル列に対するエンティティ推定の先行研究として，Neumaier らの手法 [Neumaier 16] や Pham らの手法 [Pham 16] がある．Neumaier らは数値のみで構成される列に対して，背景知識のグラフを用意することでエンティティ推定する手法を提案している．一方で，Pham らはリテラル列を含む一般の列に対して，類似度に基づいた 2 クラス分類器を用いてエンティティ推定する手法を提案している．リテラル列は数値のみで構成されると限定されないことから，本研究では，リテラル列に対するエンティティ推定の既存手法として Pham らの手法を比較対象とする．

また，Limaye らは，マルコフ確率場によってテーブル中の要素間や列間の関係をモデル化する方法を提案している [Limaye 10]．それは，類似度を用いてポテンシャル関数を表現する自由度の高いモデルである．しかし，NE 列にのみに対応しており，リテラル列を推定するためには拡張が必要である．本研究では，Limaye らの手法をリテラル列に対して拡張したモデルをベースラインとし，その上で，列に対応するエンティティ推定に対してマルチラベル分類を組み込んだ新たなモデルを提案する．提案手法では先行研究同様マルコフ確率場を用いる．

先行研究に対する本研究の優位性は以下である．(1) NE 列とリテラル列のそれぞれの列に対応するエンティティ推定をマルチラベル分類として定式化することで，各列の特徴ベクトルと複数のエンティティとの対応づけを学習データから学習でき，さらに列に複数のエンティティが対応するような曖昧性を考慮することができるため，他の列との依存関係によってその曖昧性を解消するモデルへと接続しやすい．(2) 列間の関係を全結合関係とし，新たにタイトルという潜在変数を用いたテーブルのモデル化をすることで，列間を独立にみることでは得られない列間依存関係を列のエンティティ推定に利用できる．

連絡先: 竹岡邦紘，関西学院大学大学院理工学研究科，kunihiro.takeoka.93@gmail.com

(3) 列間の相互依存関係を含んだモデルの同時確率を最適化することで，テーブル内の複数列に対応するエンティティを同時に求めることができ，さらに列ごとに推定するときよりも性能向上する．

本研究では，提案手法を，UCI Machine Learning Repository にある 183 の CSV データと，知識ベースとして WordNet を用いて検証する．CSV データを 1 つのテーブルとみなして人手によるアノテーションを行いそれを真値とした．

2. 関連研究

2.1 類似度に基づく列に対応するエンティティ推定手法

Pham らは，テーブルの列に対応するエンティティ推定のために，類似度に基づく分類手法を提案している [Pham 16]．Pham らの手法は，各列ごとに独立にエンティティ推定を行う．1 つの列を要素の集合として表現し，集合間に文字列的な類似度と数値的な類似度を定義する．例えば，文字列的な類似度に要素の 3-gram の jaccard 係数を用い，数値的な類似度にはコルモゴロフ-スミノフ検定の P 値を用いる．このように定義された複数の類似度の値で構成されるベクトルを入力とし，それらが同じエンティティを指すかどうかを出力する 2 クラス分類器を学習する．この手法では，推論時に，新たに入力された列と学習データ内の列との類似度ベクトルを求める必要がある．従って，学習データが多ければ推定の性能は向上するが，推論にかかる時間は学習データ数に比例する．

2.2 マルコフ確率場によるテーブルの変数間表現

マルコフ確率場を用いてテーブルの変数間の関係を表現したテーブルの概念決定の手法として，Limaye らの手法がある [Limaye 10]．Limaye らは，テーブルを複数の NE 列からなるものとして，テーブルの概念決定を行う手法を提案している．要素と列名，列間の関係についてそれぞれ潜在変数を定義し，マルコフ確率場によってテーブル中の変数間の関係を表現する．各変数間のポテンシャル関数には，テーブルの要素の値である文字列とエンティティの類似度と，エンティティ間の類似度，エンティティの組と関係の間の類似度を用いる．その上で，ポテンシャル関数の積で表現される同時確率の最大化によってテーブルの概念決定を行う．また，同時確率を最大とする潜在変数の値の組を推定することは NP 困難であるため，近似推論の手法も提案している．この手法は，リテラル列を無視することを前提としており，すべてのテーブルに用いることはできない．

3. マルチラベル分類に基づく列に対応するエンティティ推定

列ごとに，NE 列とリテラル列それぞれに対する列のエンティティを推定する手法を提案する．本手法では，列に対するエンティティ推定をマルチラベル分類 [Tsuoumakas 06] として定式化する．NE 列とリテラル列は異なる特徴を持つため，それぞれの特徴量を設計する．その特徴量を用いて，NE 列とリテラル列それぞれにマルチラベル分類器を学習する．学習したマルチラベル分類器を用いて，新たな列に対応するエンティティ推定を行う．ここでは，学習データおよび新たな列が NE 列とリテラル列のどちらかという分類は事前になされていると仮定する．

3.1 NE 列の特徴量

NE 列に対する特徴量は，各要素に対応づけられる可能性のあるエンティティを用いて定義する．定義より NE 列の各要素はエンティティに割り当てることが可能であり，それらによって特徴付けられる．知識ベースのベクトル空間への埋め込みを用いると，エンティティはベクトルで表現できる．ここでは，ユークリッド空間への埋め込みに比べて少ない次元でエンティティ間の構造を表現することが可能なポアンカレ埋め込み [Nickel 17] を用いて，知識ベースを埋め込む．NE 列の特徴量は，以下の手順で定める．(1) NE 列の各要素について，要素中の単語を同義語とするようなエンティティの集合を求め，それの和集合を列を特徴づける集合とする．(2) その集合の各要素について，知識ベースの埋め込みによって対応づけられたベクトルを求める．(3) 求めたベクトル集合について，それの平均と分散とをその NE 列の特徴量とする．

3.2 リテラル列の特徴量

リテラル列に対する特徴量は，Pham らの手法を参考にして定義する．リテラル列は，NE 列と異なり各要素がエンティティとは対応づけられない．従って，NE 列とは異なり，各要素の文字列の特徴を特徴量とする．また，リテラル列は数値や記号を含む文字列を要素とすることが多い．これは数値や記号はエンティティに対応づけることが難しいためであり，そのため日付や時刻，値段などが要素となる．リテラル列の 1 つの列を，文字列で表される要素の集合として解釈する．Pham らの手法は，2 つの列の間，つまり集合間に対して複数の類似度を用い，その類似度として数値的な類似度と文字列的な類似度を利用している．本提案手法では，類似度の代わりに，数値的な特徴量と文字列的な特徴量を用いる．数値的な特徴量は，列内の要素中の数値で構成される集合に対する最小値と最大値・中央値・平均・標準偏差・3 次モーメント・歪度・尖度とする．文字列的な特徴量は，今回の実験で用いたデータ中に頻出する記号である「:」,「,」,「.」,「/」の要素ごとの平均出現回数とする．

3.3 列に対応するエンティティ推定

本手法では，列の種類ごとにマルチラベル分類器を用いて，列に対応するエンティティを推定する．マルチラベル分類の最も簡単な手法として，Binary Relevance [Tsuoumakas 06] を用いる．Binary Relevance は，各ラベルごとに 2 クラス分類器を用意し，マルチラベル分類を実剣する方法である．この方法では，各分類器は入力に対してそのラベルが対応するかどうかを出力する．ここでは，ラベルをエンティティと読み替えて，マルチラベル分類器を学習する．また，2 クラス分類器にランダムフォレストを用いる [Breiman 01]．新たな入力列を特徴ベクトルに変換し，学習したマルチラベル分類器を用いて，各エンティティの出力確率を求める．これにより，新たに入力された列に対して，各エンティティがどの程度対応づけられるかを推定する．

4. 複数列に対する同時概念推定

列ごとに推定する手法を拡張して，テーブル中の複数列に対して同時に列のエンティティを推定する手法を提案する．テーブル中の各変数間の関係を表現するためにマルコフ確率場を用いる．提案手法では，テーブル中のリテラル列に対応するエンティティも同時に推定することができる．

4.1 テーブルの中の変数とポテンシャル関数

テーブルの行数を R とし，列数を C とする．r 行 c 列目の要素に対応する観測変数を x_{rc} とし，テーブルのタイトルに対応する潜在変数を t，テーブルの c 番目の列のエンティティに対応する潜在変数を y_c とする．観測変数は文字列を値とし，

潜在変数はエンティティを値とする．これらの変数を用いて，テーブルに対するマルコフ確率場を定義する．例えば，2 行 3 列のテーブルに対応するマルコフ確率場は図 1 のようになる．Limaye らの提案したモデルと異なり，観測変数 x_{rc} が直接 y_c に影響するモデルとなっている．

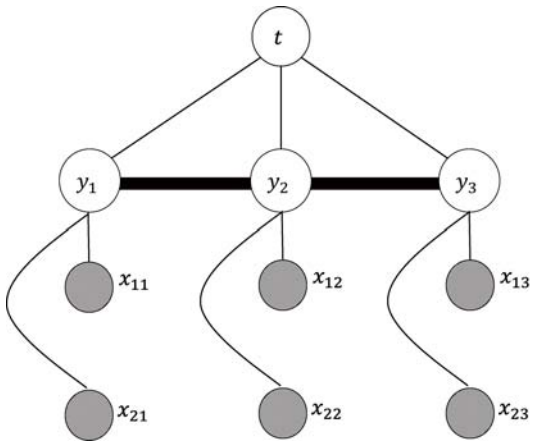


図 1: 2 行 3 列のテーブルに対するグラフィカルモデルの例．y_c を貫く太い線は y_c 間の全結合を表す．

各変数間のポテンシャル関数を以下で定義する．(1) 潜在変数 y_c と列 c の観測変数の集合 $\mathbf{x}_c = \{x_{rc}\}_{r=1}^{R}$ のポテンシャル関数 $\phi_{yx}(y_c, \mathbf{x}_c)$ は，学習データ中の $\mathbf{x}_c$ と似た列のエンティティが y_c であれば，大きな値をとる関数である．具体的には，$\phi_{yx}(y_c, \mathbf{x}_c) = f_{y_c}(\mathbf{x}_c)$ と定義する．ただし，$f_{y_c}(\mathbf{x}_c)$ は $\mathbf{x}_c$ を入力としたときのマルチラベル分類器の出力が y_c をラベルとして出力する確率値を表す．マルチラベル分類器によって ϕ_{yx} を定義することによって，学習データから対応づけ方が自動的に抽出される．

(2) 潜在変数 y_c と $\mathbf{y}_{-c}$ のポテンシャル関数 $\phi_{yy}(y_c, \mathbf{y}_{-c})$ は，y_c と $\mathbf{y}_{-c}$ が学習データ中で共起している場合大きな値をとる関数である．ここで $\mathbf{y}_{-c}$ は $\mathbf{y}$ から y_c を除いたものとする．このポテンシャル関数 ϕ_{yy} は，ϕ_{yx} と同様にマルチラベル分類器の出力を用いて，$\phi_{yy}(y_c, \mathbf{y}_{-c}) = g_{y_c}(\mathbf{y}_{-c})$ と定義される．$g_{y_c}(\mathbf{y}_{-c})$ は，$\mathbf{y}_{-c}$ を特徴ベクトルに変換したものを入力としたマルチラベル分類器の出力が，y_c をラベルとして出力するときの確率値を表す．知識ベースの埋め込みを用いて，以下の手順で $\mathbf{y}_{-c}$ を特徴ベクトルに変換する．$y_{c'} \in \mathbf{y}_{-c}$ に対応したベクトルを埋め込み空間から求める．得られたベクトルの集合を $\mathbf{V}_{-c}$ として，$\mathbf{V}_{-c}$ の平均と分散を $\mathbf{y}_{-c}$ の特徴量とする．学習データを用いて，このような特徴量を入力とし，c 列目に対応するエンティティ集合を出力とするマルチラベル分類器 $\mathbf{g}$ を学習する．

(3) 潜在変数 t と y_c のポテンシャル関数 $\phi_{ty}(t, y_c)$ は，y_c に対して意味的に近いタイトル t に対して大きな値をとる関数である．具体的には，ϕ_{ty} は埋め込み空間上での距離関数を $d(t, y_c)$ として，$\phi_{ty}(t, y_c) = \exp\{-wd(t, y_c)\}$ と定義される．ただし，w は重みパラメータである．

4.2 推論

これらのポテンシャル関数を用いて，マルコフ確率場上の同時確率は以下の式で定義される．

$$p(\mathbf{X}, \mathbf{y}, t) \propto \prod_{c=1}^{C} \phi_{yx}(y_c, \mathbf{x}_c)\phi_{yy}(y_c, \mathbf{y}_{-c})\phi_{ty}(t, y_c) \qquad (1)$$

観測変数の集合 $\mathbf{X} = \{x_{rc}\}$ が与えられた下で，式 1 を最大化する $\mathbf{y}$ を求める．

潜在変数はエンティティを値として取るため，式 1 を最大にする潜在変数の組を求めることは，組合せ最適化となり困難である．ここでは，初期値を適当に与え，ギブスサンプリングによってこの同時分布の最大化を近似する．具体的には，各列ごとに ϕ_{yx} の値が最大となる y_c を求めて初期値とする．それから，全ての y_c が固定された下で，$\phi_{ty}(t, y_c)$ が閾値以上になる全ての t を候補として擬似事後分布からサンプリングする．さらに，t と $\mathbf{y}_{-c}$ が固定された下で，y_c に対応するマルチラベル分類器の出力できる候補集合から，サンプリングする．これらを 300 回程度繰り返し行い，式 1 を最大にする $\mathbf{y}$ を求める．

5. 評価実験

5.1 実験設定

UCI Machine Learning Repository [Lichman 13] の 183 の CSV データを利用する．各 CSV データをテーブルとして，実験に用いる．これらのテーブルは，781 の NE 列と 4109 のリテラル列によって構成される．

知識ベースには，WordNet [Miller 95] を用いる．特に，列に対応するエンティティは名詞であることが多いことから，本実験ではエンティティを名詞に限る．WordNet を用いて UCI のデータセットを複数人によってアノテーションする．このアノテーション結果を真値とする．ただし，1 つの列に対して複数のエンティティが真値となることがある．

ポアンカレ埋め込み [Nickel 17] を用いて，WordNet のエンティティを上位下位概念の関係を保存するように 5 次元の双曲空間に埋め込む．ポアンカレ埋め込みを用いることで，ユークリッド空間にに比べて低次元であっても，うまく構造を保存する．

5.2 評価指標

MAP@k と nDCG@k・類似度に基づくスコア（類似度@k）を用いて評価する．MAP@k と nDCG@k は，ランキングに関する一般的な評価指標である [Clarke 08]．また，類似度@k は，真のエンティティ集合と推定結果がどれほど類似しているかを表現する．上位 k 位までの推定結果について，真のエンティティとの Wup 類似度 [Wu 94] を求め，その確信度に応じた重み付き平均で定義する．Wup 類似度は，0 から 1 の範囲の値をとり，2 つのエンティティが類似している場合大きな値をとる．本実験では $k = 5$ として評価する．

5.3 NE 列に対する推定

NE 列に対して，列に対応するエンティティ推定の実験結果を表 1 に示す．Limaye らの手法であった列間の関係を無視し，独立な列として列に対応するエンティティ推定を行う手法と，Pham らの手法の中で文字列的な類似度のみを用いた手法を比較手法とする．ただし，ここで用いる提案手法は，列独立に NE 列に対応するエンティティ推定を行う手法である．提案手法は，すべての指標で比較手法を上回る．

5.4 リテラル列に対する推定

リテラル列に対して，列に対応するエンティティ推定の実験結果を表 2 に示す．比較手法として，Pham らの類似度に基づくエンティティ推定手法を用いる．ただし，ここで用いる提案手法は，列独立にリテラル列に対応するエンティティ推定を行う手法である．提案手法は比較手法を上回る性能を示した．Pham らの手法と提案手法の違いを考えるために，学習データ中に，メートルとインチのように単位は異なるが列に対応す

表 1: NE 列に対する，列に対応するエンティティ推定の実験結果．Limaye らの手法に基づく比較手法（比較手法 1）と類似度に基づく比較手法（比較手法 2），マルチラベル分類による提案手法の比較．

手法	MAP@5	nDCG@5	類似度@5
比較手法 1	0.021	0.077	0.411
比較手法 2	0.196	0.328	0.470
提案手法	**0.314**	**0.365**	**0.526**

るエンティティが同じ場合を考察する．Pham らの手法では，この場合，類似度が低いにも関わらず同じエンティティを指すため，適切な重みを学習することが困難になる．一方で，提案手法は，この場合それぞれを特徴量に変換して考えるため，学習に大きな影響を与えることが少ないと考えられる．

表 2: リテラル列に対する，列に対応するエンティティ推定の実験結果．類似度に基づく比較手法とマルチラベル分類による提案手法の比較．

手法	MAP@5	nDCG@5	類似度@5
比較手法	0.270	0.330	0.467
提案手法	**0.397**	**0.430**	**0.590**

5.5 テーブル全体に対する推定

複数列のテーブルに対して，列に対応するエンティティ推定の実験結果を表 3 に示す．Limaye らの手法をリテラル列に対して拡張した手法と，列間の依存関係を無視した提案手法，さらに列間の関係を含んだ提案手法の 3 つの比較である．列間の関係を含んだ提案手法の評価は，提案手法において列独立に行う場合より優れている．これより，列間の関係をマルチラベル分類器を用いてモデル化し，タイトルを表現する潜在変数を導入したことによって，推定性能が向上するといえる．また，Limaye らの手法は，様々な類似度を用いてポテンシャル関数を定義するため，適切な類似度を設定する必要がある．提案手法は，類似度の決定の代わりに知識ベースの埋め込みと，マルチラベル分類器がポテンシャル関数を定義している．学習データがある環境では，それらに応じたポテンシャル関数を定義できる提案手法が優れていると考えられる．

表 3: 複数列のテーブルについて，列に対応するエンティティ推定を行った実験結果．Limaye らの手法を元にリテラル列に拡張した手法（比較手法）と，列間の依存関係を無視して実行した提案手法（列独立），列間の依存関係を考慮する提案手法（列依存）の 3 つの手法の比較．

手法	MAP@5	nDCG@5	類似度@5
比較手法	0.225	0.291	0.480
提案手法 (列独立)	0.351	0.413	0.537
提案手法 (列依存)	**0.464**	**0.741**	**0.635**

参考文献

[Breiman 01] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.

[Clarke 08] Clarke, C. L., M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon (2008). Novelty and diversity in information retrieval evaluation. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 659–666. ACM.

[Deng 13] Deng, D., Y. Jiang, G. Li, J. Li, and C. Yu (2013). Scalable column concept determination for web tables using large knowledge bases. *Proceedings of the VLDB Endowment*, 6(13), 1606–1617.

[Lichman 13] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[Limaye 10] Limaye, G., S. Sarawagi, and S. Chakrabarti (2010). Annotating and searching web tables using entities, types and relationships. *Proceedings of the VLDB Endowment*, 3, 1338–1347.

[Miller 95] Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39–41.

[Mulwad 13] Mulwad, V., T. Finin, and A. Joshi (2013). Semantic message passing for generating linked data from tables. *International Semantic Web Conference*, 363–378. Springer.

[Neumaier 16] Neumaier, S., J. Umbrich, J. X. Parreira, and A. Polleres (2016). Multi-level semantic labelling of numerical values. *International Semantic Web Conference*, 428–445. Springer.

[Nickel 17] Nickel, M., and D. Kiela (2017). Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems 30*, 6338–6347.

[Pham 16] Pham, M., S. Alse, C. A. Knoblock, and P. Szekely (2016). Semantic labeling: a domain-independent approach. *International Semantic Web Conference*, 446–462. Springer.

[Tsuoumakas 06] Tsoumakas, G., and I. Katakis (2006). Multi-label classification: an overview. *International Journal of Data Warehousing and Mining*, 3(3), 1–13.

[Venetis 11] Venetis, P., A. Halevy, J. Madhavan, M. Paşca, W. Shen, F. Wu, G. Miao, and C. Wu (2011). Recovering semantics of tables on the web. *Proceedings of the VLDB Endowment*, 4(9), 528-538.

[Wu 94] Wu, Z., and M. Palmer (1994). Verbs semantics and lexical selection. *Proceesings of the 32nd Annual Meeting on Association for Computational Linguistics(ACL '94)*, 6, 133–138.

[Zhang 17] Zhang, Z. (2017). Effective and efficient semantic table interpretation using tableminer+. *Semantic Web*, 8(6), 921-957.

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

創作者と人工知能: 共作実現に向けた創作過程とメタデータ付与 4 コマ漫画ストーリーデータセット構築

Creators and Artificial Intelligence: Enabling Collaboration with Creative Processes and Meta-data for Four-scene Comic Story Dataset

上野 未貴 *1
Miki Ueno

*1豊橋技術科学大学, 情報メディア基盤センター
Toyohashi University of Technology, Information and Media Center

Recently, researches in the field of Artificial Intelligence(AI) have become widely developed so that computational methods to analyze and build creative contents have attracted a notable following. Especially, computational methods as deep learning, which is one of the methods of machine learning for content creation has rapidly developed; i.e. caption generation, automatic colorization, and sketch drawings. On the other hand, creativity in humans is an intellectual activity meaning that it is really hard to simulate the whole creative process as computational methods. In this field, the number of data for machine learning is too small because creative processes often contain precious and sensitive data. In this research, we propose and construct a four-scene comics story dataset with creative process and meta-data toward collaboration between "Creators and Artificial Intelligence".

1. はじめに

人工知能技術の発展により，創作物を計算機が理解したり，自動生成する [1] ことに強い関心が集まっている．画像キャプション生成 [2] や自動彩色，スケッチ生成 [3] など，創作に関わる特に深層学習を利用した技術に近年大きな発展が見られる．一方で，人の創作は高次の知的作業であり，計算機上に創作過程全体を具現化することは極めて難しい．この分野において，創作物が貴重で繊細な情報をしばしば含むため，機械学習向けのデータが非常に少ない．本研究では,「創作者と人工知能」の共作に向け，創作過程とメタデータを有する 4 コマ漫画ストーリーデータセットを構築する．

2. 基本方針:研究姿勢と本研究の目的

2.1 研究姿勢

五感から入力を得て，感じ考えたことを絵や言葉で表して創作物として出力する全体機構 [4] は，現在，人にしかできない．人工知能分野の技術が発展しても，感情や経験，背景知識なしには，フレーム問題 [5] に代表されるように，機構の大部分の実現は計算機では困難と考えられてきた．問題を定式化するため，ストーリーを持つ創作物はある種，現実世界との関連を保って状況構築した有限世界の好題材と考えられる．広範には，工学は最終的には人の役に立つ技術手法を研究する学問とされる．ストーリーを有し，言語，画像，音などの多くの要素を含む創作物であるアニメの創作現場環境には課題が多く，人が一番時間をかけたい作業に時間をかけられないという問題が発生することがある．例えば「旅立つ」ストーリーを描く際，各種漫画を参考に，主人公の表情，台詞などについて，新たな表現を試みるため，類似資料が存在していないか調べるコストがかかる．小説，漫画，絵本，アニメなどの，創作場面ではアナログな資料管理が主流であり，デジタルデータを残す文化が少ない．また人同士の対面のコミュニケーションで進行する制作が多く，従来，計算機科学との親和性は高くない．一方，デジタル技術の発展により，積極的に技術活用する現場もある．工学系の研究者として，技術発展のために，創作物を貴重なデータとしてもみなす必要がある．一方で，創作物は長い時間をかけて創作者が創り上げた作品であり現場の意見を丁寧に拾い扱う必要がある．以上の考えから，著者はこれまでから，創作に関わる人へ意見を聞き，要素技術 [6] と応用手法 *1 の研究を進めてきた．人が自由に利用したい情報技術を選択して利用する，という姿勢で創作支援を目指す．

2.2 本研究の目的

1. 漫画のストーリーを機械学習で扱うためのパターン定義

創作の中でも着目する過程に異なりがある．著者はストーリー理解に特に着目している．ストーリーを計算機で扱う研究は小説や漫画などの形態を問わず困難とされる．

言語と画像のマルチモーダルな媒体である漫画は，芸術的，心理学的見地だけでなく，工学研究 [7] においても注目される．特に，著者は，漫画の中でも，形状と長さの統一，ストーリーが明瞭という観点から，日本で流通が多い 4 コマ漫画を対象に取る．n コマ漫画を計算機に生成 [8] する研究を進め，深層学習を 4 コマ漫画に適用し感情識別・順序識別する要素技術を提案した．一方，現状許諾を得られる既存作品のストーリー分析には制約が存在する．本研究で，計算機と親和性が高い 4 コマ漫画パターンを提案し，新たなデータセットの構築指針を示す．

2. 創作過程を有した創作物のデータ構築と応用方法

解決すべきタスクの細分化が困難，実現場での応用方法のイメージ不足，権利関係や既存作品から創作過程と作者の付与したメタデータを得ることが難しい．そのため，計算機上で機械学習に利用できるストーリーを有するデータが少なく，分野発展への壁がある．本研究では，創作者と計算機の共作を目指し，作者の創作過程と読者の読解過程のストーリーに関わるメタデータが付与された 4 コマ漫画ストーリーデータセットを構築する．また，想定タスクと現場応用方法を述べる．

連絡先: 上野 未貴, 豊橋技術科学大学，情報メディア基盤センター, 〒 441-8580 愛知県豊橋市 天伯町雲雀ヶ丘 1-1,
E-mail: ueno@imc.tut.ac.jp

*1 創作者と人工知能分野研究者のコラボレーション:
https://www.tut.ac.jp/docs/PR180222.pdf

3. 分野動向:創作物の工学系研究

本分野における，最終的な研究目的は，技術的に共通する部分もあるが，多くの場合，**自動生成と創作支援，解析と生成に大別される**．比較的ルールが厳密に定義でき，近年，深層学習の興隆から飛躍的な進歩を遂げるゲームプレイヤプログラム AlphaGo などとは状況が異なる．創作過程の要件を細分化して定義することは極めて困難であり，現状，全自動で明らかに人を凌駕する技術があるとは言えない．

3.1 創作物の種別と特徴

創作物に関する工学系の研究と特徴を述べる．

3.1.1 漫画

オブジェクト認識，自動コマ割レイアウト，レビュー分析，吹き出し推定，線画補正，ネタバレ解析，読解教材応用などの研究 [7] が知られる．

3.1.2 小説

ストーリーの扱いが難しく，全文自動生成は困難だと考えられている．しかし，深層学習による部分的な文章生成が発展し，創作支援の立場を取る研究も増加している．

3.1.3 アニメ

アニメは創作過程において，自然言語，静止画，動画，音楽と扱うデータが多く，各作品の創作に関わる人数が多い [1]．創作過程の**絵コンテ**は，形状が同一のコマ画像と台詞や状況説明の自然言語から成り，本研究で対象とするコマ漫画と類似する．カメラワークや尺などの情報がある．

3.1.4 絵本

対話的な絵本の自動生成 [9] が知られる．連続する絵でストーリーを表すだけでなく，各ページを鑑賞する時間が長く，一枚の絵の中に複数のストーリーが存在することが多い．ストーリーを解析する際にはストーリーの領域分割が必要と考えられる．見開きの 2 ページで，絵のページと，文字だけのページを独立させる場合と，見開きの 2 ページともに，絵の中に白抜きで文字を入れる場合が存在する．

3.2 漫画の従来データセット

漫画の研究のデータセットとして，国内ではプロの漫画 109 冊の全ページの画像を電子化したデータセット MANGA109[10] が知られる．ジャンルや年代が異なる漫画が広く含まれており，異なるタッチの漫画オブジェクト認識の研究等に向く．4 コマ漫画は 5 冊含まれるが，作者や話のテーマが大きく異なる．漫画の形態や流通状況は国別に異なる部分もあるが，海外でも漫画は工学研究の題材として着目されている．フランスでは，eBDtheque [11] というデータセットが知られる．漫画 100 ページに対し，台詞やコマ位置などのメタデータを付与したものだが，ジャンルに偏りが見られる．いずれも，ストーリーに関わるデータを作成することは主目的とされていない．

3.3 従来データセットの問題点:ストーリーの観点

ストーリーは，俳句，古典落語などに典型的に見られるように，型に嵌めること，型を破ることで生まれる．4 コマ漫画は決められたコマ内にストーリーを画像と言語で表す型があり，計算機科学の良質な題材である．一方，同一オブジェクトでもタッチが異なること，オリジナリティの観点から同一プロットを異なる作者が書くことは稀である．そのため，データを集められたとしても，共通部分を見つける難しさがあり，ストーリーの解析のための深層学習の適用に課題がある．一方，ストーリーの定量化が難しいことから，素性設計が困難である．

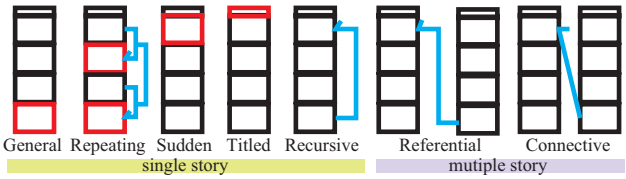


図 1: 4 コマ漫画ストーリーパターン 7 種
注:特にオチのあるコマを赤色で，読む順と関連を水色で表す．

4. 4 コマ漫画構造ストーリパターンの定義

問題解決のため著者は 4 コマ漫画のストーリーパターンを提案している．コマ構造について，アメリカで有名な 4 コマ漫画ピーナッツに対し認知心理学的な読解過程の研究 [12] が存在するが，本研究とは機械学習適用を主目的としない点で異なる．図 1 に，本研究で再整理した 4 コマ漫画のストーリーパターン 7 種を示し，以下，図の左から順に対応させ説明する．

1. 一般 標準的な起承転結．原則的に 4 コマ目にオチがある．
2. 繰返し 1-2 コマ間の類似事象が 3-4 コマにも起きる *2.
3. 出オチ 1 コマ目におかしな絵が描かれてオチがある *3
4. タイトルオチ 最後にタイトルを見返してオチがわかる．
5. 再帰 4 コマ目から 1 コマ目に戻り話として成立する．
6. 参照 1 つ以上前の話の続きの話．
7. 連続 連続した 4 コマ を 2 話 並べて 8 コマで話となる．

新聞や，広告の一部に掲載される 4 コマ漫画もあるが，流通する漫画雑誌や単行本では 1 ページに 2 話の 4 コマ漫画が並ぶことが通例である．日本では特に，連載の間隔が短い．ストーリー 4 コマ漫画と呼ばれる各話で話が短く完結しているが，複数話で全体の大きな話を持つ，連載などに向いた形態が多い．ストーリー 4 コマ漫画では，単一話と複数話という 2 種を定義し，構造を考える必要がある．

単一話とは，各話で完結する話である．複数話とは，2 話以上で関係を持つ話である．

ストーリーパターンを，大きく構造と内容に分けて定義する．網羅性に課題があり限定的ではあるが，広義には，構造はタイトルと各コマのオチの有無の組合せ 2^5 種のパターンで表せると考えた．中でも，使用頻度が高いか，視覚的に特徴的な組合せには，制作現場で呼称が存在する．

頻度が高いパターンは,「一般」であり，他に視覚的に特徴的なものは「出オチ」である．

5. 4 コマ漫画ストーリーデータセット構築

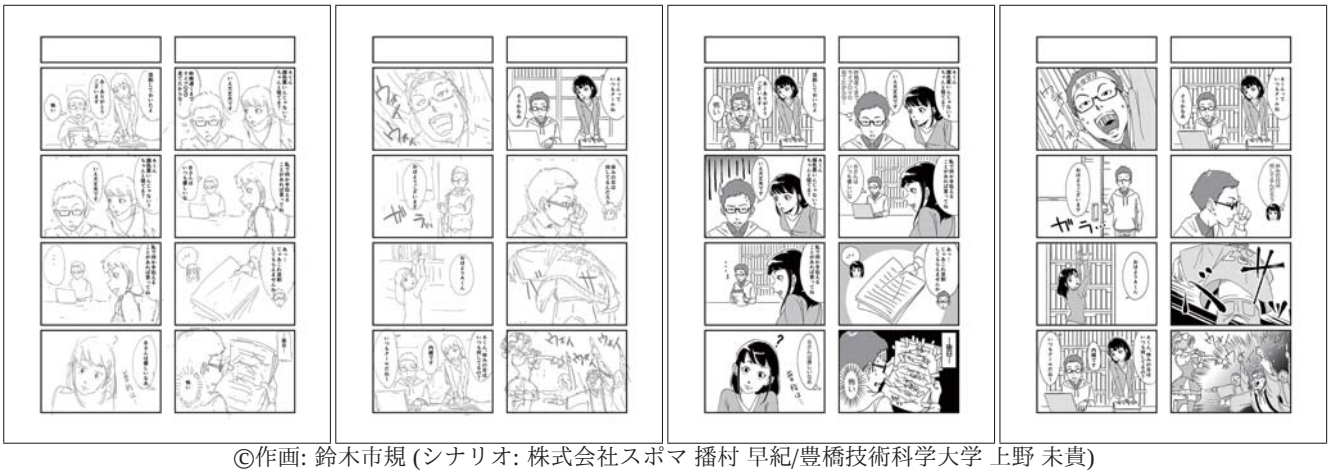
本研究では，上記 2 種の構造に係るストーリーパターン（以下，**構造ストーリーパターン**）に着目し，内容に関わるストーリーパターンを**メタプロット**として，なるべく固有名詞を含まない抽象化したテキストで表した．

5.1 構築方針

キャラクタ情報，舞台情報，10 個のメタプロットとプロットを作成し，2 つの構造ストーリーパターンを指定して，20 個のシナリオを作成し，5 人の異なるタッチ:ギャグ，少年，青年，少女，萌え，のプロ作者が計 100 作品の漫画を制作した．

*2 3-4 コマ目の方が強調された事象が起きる．
*3 4 コマ目にもオチがあることが多い．

©作画: 浦田カズヒロ (シナリオ: 株式会社スポマ 播村 早紀/豊橋技術科学大学 上野 未貴)



©作画: 鈴木市規 (シナリオ: 株式会社スポマ 播村 早紀/豊橋技術科学大学 上野 未貴)

図 2: 第 1，2 話，2 構造，2 作者のラフと清書
注:日本では原則として，漫画は各話に関して各コマ内で右から左，コマ間で上から下，ページ内で右から左へ読む．

■第一話-構造 1
[1] B「A くんっていつもクールね」A「そうかなあ」
[2] B　心の声（休みの日は何してるんだろう）
[3] A　法被を着る
[4] アイドルのコンサートで法被を着てサイリウムを振る A
■第一話-構造 2
[1] アイドルのコンサートで法被を着てサイリウムを振る A
[2] 場所：研究室挨拶をして研究室に入ってくる A「おはようございます」
[3] B「おはよう A くん」
[4] B「A くん、休みの日はいつも何してるの？」A「内緒です」B「いつもクールだね～」
■ 第二話-構造 1
[1] B「A くん顔色悪いんじゃない？ちゃんと寝てる？」A「いえ大丈夫です」A 心の声（昨晩遅くまでライブ DVD 見てたからな…）
[2] B「私で何か手伝えることがあれば言ってね」A（B さんはいつも優しいな）
[3] A「あっ…じゃあこれ添削してもらえませんか」B に原稿を渡す A
[4] ナレ：一後日一赤と付箋だらけの原稿が返却される A 心の声（怖い）
■第二話-構造 2
[1] B からびっしり赤と付箋だらけの原稿を渡される A
B「添削しておいたよ」A「あ…ありがとうございます」A 心の声（怖い）
[2] B「A くん顔色悪いんじゃない？ちゃんと寝てる？」A「いえ大丈夫です」
[3] B「私で何か手伝えることがあれば言ってね」

[4] A 心の声（B さんは優しいなあ）書き文字：普段は

図 3: 第 1，2 話のシナリオ
注:作者へ提示したものになるべく沿い，本稿用に調整した．

5.2　創作過程

本データセットの創作過程は以下のとおりである．合わせて作業の担当を示す．各過程から得られる各情報は自然言語情報もしくは画像情報として保持した．

1. キャラクタおよび舞台設定を決定 [著者]
   - 共通の主人公キャラクタ 2 人の情報共有
   - 共通の舞台設定の情報共有
2. メタプロットを作成 [著者]
3. メタプロットからプロットを作成 [著者]
4. プロットからシナリオを作成 [制作会社と著者]
5. シナリオからラフを作成 [各漫画作者]
6. ラフから清書 [各漫画作者]

5.3　レイヤ情報およびメタデータの付与

本データセットを作成するにあたり可能な限り創作に係る情報を保持してデータを付与した．

レイヤ情報: コマ枠/各キャラクタの表情に係る顔部位/各キャラクタの顔部位以外/各吹き出し/各台詞/各背景

メタデータ付与: 計算機上で扱いやすいフォーマットを考慮し，各作者と読者の双方が以下のメタデータを付与した．

- キャラ単位の感情 7 種: 基本 6 感情とニュートラル
- オチを説明するキャプション: 文字数を指定

6.　データ具体例と考察

全体の舞台とキャラ設定，第 1，2 話のデータを具体例として示す．図 2 に 2 作者のラフ画像と清書を示す．表 1 にメタプロットとプロット，図 3 にシナリオを示す．

表 1: 第 1，2 話のメタプロットとプロット

話数	メタプロット	プロット
1	見た目は <1> だが実は <2>	A キャラ紹介:見た目は < クール > だが実は < 熱中しているもの < アイドル > がある >
2		B キャラ紹介:見た目は < 可愛い > が実は < 怖い一面がある >

注:シナリオ制作用のデータ．漫画作者へは提示しない．本稿用に内容変更せず表記調整した．

舞台: 大学の研究室．主人公はキャラクタ A，B の学部生と大学院生の二人．人工知能分野の感性情報系の研究をしており異なるテーマを持つ．

キャラ特徴: 外見として，髪，目，身長，体系，服装，内面として，年齢，学年，兄弟構成，性格，等を設定．

共通シナリオには多様な表現方法がある．例えば 2 作者間において，第 1 話，構造 1, 1 コマ目, 1, 4 コマ目の構図が類似する．また，第 1 話, 構造 1, 1 コマ目では，構図は類似するがカメラワークが異なる．第 1 話, 構造 1, 2 コマ目ではキャラクタの視点が異なる．

異なる作者の共通クラスの単一イラスト画像の学習手法 [3] は発展している．一方で，特にプロが一般的に同じシナリオに関する漫画を書くことは稀である．本研究では，創作過程を保持し，共通シナリオに対する漫画データを構築することで，多様な表現，例えば，構図やカメラアングルなど，ストーリーの演出に関わる表現について，また，吹き出しや背景効果などの漫画特有表現の考察が可能である．

担当編集や演出が担うような計算機手法の構築が考えられる．創作時に経験豊富なプロに相談するのはもちろんだが，相談が難しい場合に，気軽に助言を求められるパートナーとしての計算機手法の創出を目指す．

7. 展望:4 コマ漫画ストーリーデータセットの実験タスク例と要素技術の応用例

本データセットを用いた，深層学習による 4 コマ漫画データセットの実験タスクの例を示す．以下，著者提案の 2 つの 4 コマ漫画ストーリー理解の要素技術を例に取る．

感情識別: ConvolutionalNeuralNetwork(CNN) を用いた 4 コマ漫画のコマの感情識別 [4]

順序識別: CNN を用いた 4 コマ漫画のコマの順序識別 [6]

本データセットを用いて構築した要素技術の応用例として，プロットの創作支援や，創作中の関連資料検索などが考えられる．4 コマ漫画以外の漫画やクリップアートなどの画像に対して提案された，オブジェクト識別やキャラクタ識別にも用いることを想定して構築した．

8. まとめ

本研究では，漫画の機械学習の研究において，研究初期段階に十分な共通データを揃えることが難しい，という従来の既存漫画を用いることによる問題に対し，あらかじめ，作者と読者の両面を考慮したメタデータを，創作過程の言語情報と画像情報，および，レイヤ情報を残し，権利関係に配慮して，4 コマ漫画データを構築する仕組みを提案した．4 コマ漫画ストーリーデータセットは「創作者と人工知能」の研究発展のため，順次公開を予定する．今後の課題は以下の通りである．

- データの増加
- 実応用タスクに関する意見聴衆
- 深層学習フレームワークの構築

謝辞

本研究は, JST, ACT-I の支援を受けたものである; 課題名「深層学習による 4 コマ漫画のストーリー解析用データセット及びフレームワークの開発」.

本稿で述べたデータセットの制作に対し，株式会社スポマ 代表取締役　臼井 英明 氏，同社 播村 早紀 氏，天野 聡彦 氏，本稿への掲載漫画を作成して下さった，浦田カズヒロ 氏，鈴木市規 氏へ謝意を示す．

参考文献

[1] Miki Ueno, Kiyohito Fukuda and Naoki Mori, *”Can Computers Create Comics and Animations?”*, *Computational and Cognitive Approaches to Narratology*, pp. 164–190, 2016.

[2] Kuniaki Saito, Andrew Shin, Yoshitaka Ushiku, and Tatsuya Harada, ”DUALNET: DOMAIN-INVARIANT NETWORK FOR VISUAL QUESTION ANSWERING”, The 18th IEEE International Conference on Multimedia and Expo , 2017

[3] David Ha and Douglas Eck. ”A Neural Representation of Sketch Drawings”. *ArXiv e-prints*, 2017.

[4] 上野 未貴, 井佐原 均. ”漫画内の特徴的要素が与えるストーリーの印象についての検討”. 人工知能学会, No. 2J5-OS-08b-4in2, 2016.

[5] John Mccarthy and Patrick J. Hayes, ”Some philosophical problems from the standpoint of artificial intelligence”, Machine Intelligence, vol.4, pp.463-502, 1969.

[6] Miki Ueno, Naoki Mori, Toshinori Suenaga, and Hitoshi Isahara. ”Estimation of structure of four-scene comics by convolutional neural networks”. In *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding, MANPU@ICPR 2016*, pp. 9:1–9:6, 2016.

[7] 松下 光範, 山西 良典, 松井 勇佑, 岩田 基, 上野 未貴, 西原 陽子, 中村 聡史. ”私のブックマーク「コミック工学」”. 人工知能, Vol. 32, 2017.

[8] Miki Ueno, Naoki Mori, and Keinosuke Matsumoto. ”2-scene comic creating system based on the distribution of picture state transition”. *Advances in Intelligent Systems and Computing*, Vol. 290, pp. 459–467. Springer, 2014.

[9] 藤野 紗耶, 安井 瑛彦, 上野 未貴, 森 直樹, 松本 敬之亮. ”Pictgent を用いた対話型絵本システムの提案”, 第 60 回システム制御情報学会研究発表講演会, 315-4, 2016.

[10] Yusuke Matsui, Kota Ito, Yuji Aramaki, Toshihiko Yamasaki, and Kiyoharu Aizawa. ”Sketch-based manga retrieval using manga109 dataset”. *CoRR*, Vol. abs/1510.04389, 2015.

[11] Clément Guérin , Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier, and Arnaud Revel. ”ebdtheque: a representative database of comics”. In *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*, 2013.

[12] Neil Cohn, ”You ’re a good structure, Charlie Brown: The distribution of narrative categories in comic strips”, *Cognitive Science*, Vol. 38, No. 7, pp. 1317-1359, 2014.

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

隠れマルコフモデルによる歴史テキストの人物移動のモデル化

Movement Modeling for History Documents with Hidden Markov Models

水谷 陽太 *1　　鶴岡 慶雅 *2
Yota Mizutani　　Yoshimasa Tsuruoka

*1東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

*2東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

It is difficult for computers to understand the "meaning" of natural language sentences. To tackle this problem, some existing methods use predicate logic. However, they cannot deal with quantitative data such as geographical distance which is important for understanding historical events. We introduce a new method to construct a simulatable world model from documents. Simulations with this model will help computers to understand the contexts, guess unwritten information, and realize some rules. We experiment with some documents about the Sengoku period in Japanese Wikipedia, and construct a hidden Markov model about people's movement.

1. はじめに

品詞タグ付けや構文解析と言った自然言語処理の基盤的技術は，近年の機械学習技術の発展に伴い大幅に精度が向上し，分野適応や学習データの作成コストの問題は依然として残るものの，テキスト処理のための実用的な技術として確立しつつある．しかし，自然言語の「意味」の理解を必要とする技術，すなわち（広い意味での）質疑応答や，(知的なエージェントを実現するための）自然言語理解などに関しては，現在の自然言語処理技術によってコンピュータで実現可能なことは極めて限られている．

近年,「ロボットは東大に入れるか?」プロジェクト [新井 12] を始めとして，計算機による自然言語理解の実現に対する注目は高まっている．しかし，現在の言語処理システムでは，例えば,「1467 年の応仁の乱で京都の大半が消失した」というテキストから,「15 世紀末に京都の人口が減少した」ということを推論することはほぼ不可能である．この種の世界知識を必要とする推論に関する問題は，人工知能研究の黎明期から研究されており，古くは Winograd による SHRDLU [Winograd 72] や，最近では，カーネギーメロン大学の Read the Web プロジェクト [Carlson 10] などによって実現可能なアプローチが模索されている．しかし，計算機による自然言語理解という問題に対する既存のアプローチのほとんどは述語論理をベースとした知識表現や推論機構を用いており,「世界」に関する定量的な情報や，時間発展の現象を扱うことが極めて難しいという問題がある．このことは，先に例としてあげたような歴史的事象の理解を目的とする場合には致命的な欠点であり，どれだけ大量に知識を積み重ねても，いわゆる「風が吹けば桶屋が儲かる」といったような，論理的にはつながっていても現実的には起こり得ない推論をしてしまう問題を避けることができない．また，歴史テキストを解釈するためには，歴史イベント同士の地理的な（遠近の）関係などもしばしば考慮する必要があるが，この種の定量的な情報はそもそも述語論理で表現することが適さない．

そこで本研究では，世界史や日本史といった，歴史的事象に関するテキストを理解・解釈することのできる計算機システムを実現するため，様々な歴史イベントを表現可能な，自律的に時間発展する世界モデルを用いることを提案する．テキストに記述された歴史イベントを読み取り，記述された内容に即した世界モデルを学習により獲得する．世界モデルは，歴史的事象を表現可能であるだけでなく，その時間発展が，史実，すなわちテキストの内容に合致することが求められる．学習により得られた世界モデルは，文脈の判断 [村上 16] や，言外の情報の推測，より本質的な原理の理解など,「意味」を理解した言語処理の助けになることが期待される．

将来的には人物，地理，時間等の多数のパラメータに依存して発生する様々な種類のイベントを扱うことが目標だが，本稿ではその前段階として，人物の移動のみに注目して実験を行った．EM アルゴリズムの一種である Baum-Welch アルゴリズム [Baum 70] を用いて特定の地点間の遷移確率を推定し，隠れマルコフモデルを構築する実験を行った．その結果，データ不足に起因する偏りは確認されたものの，テキストから読み取ったイベントを表現するモデルを獲得することに成功した．

連絡先: 水谷陽太，東京大学大学院工学系研究科電気系工学専攻，mizutani@logos.t.u-tokyo.ac.jp

2. 提案手法

本手法では人物の移動モデルを，移動先がその人物の現在位置のみに依存する単純マルコフ過程であると仮定する．時刻 t で特定の位置 p_i に存在するとき，$t+1$ で p_j に移動する確率を表す遷移行列 A_{ij} を求めることが目的となる．ある人物の様子が記述されたテキストを解釈することで，その人物が時刻 t において位置 p に存在したという記述 (t, p) を得る事ができる．ただし，全ての t について (t, p) がテキストから得られることは稀であり，記述が存在しない t については，位置の推定が必要となる．この問題に対し提案手法では，記述の有無を含めた移動モデルを隠れマルコフ過程で表現し，Baum-Welch アルゴリズム [Baum 70] を用いて遷移行列の推定を行う．より具体的には，時刻 t において人物が特定の位置 p_i に存在しているとき，観測として「(t, p_i) の記述が得られる」場合と,「t における記述が得られない」場合の二通りが存在すると考える．一般的な隠れマルコフモデルでは，この他に i と異なる j についても (t, p_j) が得られる可能性を考慮し，またこれらの観

測を得る確率は現在位置 p_i によって異なることを仮定するが，本稿では誤った記述 (t, p_j) は存在しないことを仮定し，また，記述が行われる確率は現在位置によらず一定であるとした．

2.1 アルゴリズム

N 人の人物の各 (t, p) のリストを入力とし，遷移行列 A_{ij} が出力となる．通常の Baum-Welch アルゴリズムとは異なり，観測シンボル確率分布は固定であることに注意する．

- 遷移行列 A_{ij} をランダムに初期化する．
- 特定の人物について，(t, p) のリストから観測シンボル列を生成する．
- 現在のパラメータ A_{ij} を用いて，通常の Baum-Welch アルゴリズムと同様に期待値計算を行う．
- 他の人物においても同様に観測シンボル列の生成及び期待値計算を行う．
- N 人全ての処理が終わったら，Baum-Welch アルゴリズムに従いパラメータ A_{ij} の更新を行う．

以上の操作を 1 epoch とし，複数回繰り返す．

3. 実験

日本の戦国時代の人物に関して，日本語 Wikipedia[*1] の記述から遷移行列を求める実験を行った．

3.1 データの作成

日本語 Wikipedia の「戦国時代の人物一覧 (日本)」ページに挙げられた人物について，リンク先の各人物について記述されたページ内の記述から，データ生成を行った．時間の単位は一ヶ月とした．文単位でパターン検索を行い，一つの文内に時刻のデータと地名のデータが両方存在する場合，それらの組をデータに加える．時刻は一ヶ月単位で保存し，もし年のみの記述が存在し，月が不明の場合は暫定的に 6 月であるとした．地名に関しては，日本語 Wikipedia の「令制国」ページ及び「日本の城一覧」ページから地名の一覧を作成した．ただし，データ取得後,「令制国」ページの「変遷」の章にあるデータから，取得した地名を戦国時代に使われていた国名へ変換した．同様に,「日本の城一覧」ページ内のデータから，城の名前を所在地のデータへと変換した．なお，本実験では人物の移動に着目するため，時刻と地名のデータ対が 2 組以上存在した人物のデータのみを用いた．取得されたデータ数を表 1 に示した．地名は実際に使用されるデータ中に存在するものの数である．

表 1: 実験に用いたデータ数

人物	地名	データ数
1189	68	4880

例えば「織田信長」のページでは,「元亀 2 年（1571 年）2 月、信長は浅井長政の配下の磯野員昌を降し、佐和山城を得た。」という文から，(1571 年 2 月, 佐和山城) というデータを得ることができる．この際，佐和山城は所在地である「近江国」へと変換され，(1571 年 2 月, 近江国) というデータとして扱われる．表 2 の太字部分がこのデータにあたる．

*1 https://ja.wikipedia.org/wiki/

3.2 結果

Baum-Welch アルゴリズムを 10000 epoch 実行し，遷移行列の推定を行った．推定された遷移行列を用いたシミュレーションの例として，織田信長のデータの欠損部分を最尤推定した結果の一部を，表 2 に示す．

3.3 考察

概ね隣接する国間を移動する様子が推定できている．例えば 1571 年 6 月の三河国から 1572 年 6 月の摂津国へと移動をする際は，地理的にそれらの二国間に存在する近江国，山城国を経由する推定を行っている．同様に，1577 年 2 月の紀伊の国から 1577 年 12 月の美作国へと移動する際も，大和国，摂津国，播磨国を経由する推定となった．

反面，1572 年 6 月の摂津国から 1573 年 2 月の陸奥国への移動の際は，途中経路が全く推定できていない．今回の手法では離れた場所間の移動経路を推定するのは難しいことが読み取れる．ただし，このデータは本来三河国の野田城を指すテキストを処理する際に，陸奥国にある同名の野田城であると誤認したものである．今回の実験におけるデータ生成は単純なパターンマッチにより行っているため，このように不正確なノイズが入りやすくなっている．

4. おわりに

本稿では，歴史上のイベントについて記述されたテキストから，シミュレーション可能なモデルを構築し，テキストの内容を理解，解釈する手法を提案し，実際に人物の移動というイベントに関して，Baum-Welch アルゴリズムを用いた遷移確率の推定を用いてモデル化を行った．

今後の課題としては，データ数の増加，データの正確性の向上が挙げられる．今回は単純なパターン検索によるデータ作成を行ったが，文構造の解析を組み合わせるなどの手法によって，より正確なデータを作成できる．また，文脈を考慮することにより，文をまたいだデータを追加することが可能となる．完全な構文解析や文脈判断は困難だが，本稿の手法により得られた世界モデルを用いることで，それらの精度向上が見込める．手法を組み合わせることで相互に精度を向上させることが期待できる．

また，本稿では人物の移動のみに注目したシミュレーションを行ったが，季節や地理，人物の家系，地位，など様々なパラメータの組み込みを検討し，扱えるイベントの種類を増やすことで，より高度かつ広範囲な応用が期待される．

参考文献

[Baum 70] Baum, L. E., Petrie, T., Soules, G., and Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *The annals of mathematical statistics*, Vol. 41, No. 1, pp. 164–171 (1970)

[Carlson 10] Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E. R., and Mitchell, T. M.: Toward an architecture for never-ending language learning., in *AAAI*, Vol. 5, p. 3Atlanta (2010)

[Winograd 72] Winograd, T.: Understanding natural language, *Cognitive psychology*, Vol. 3, No. 1, pp. 1–191 (1972)

表 2: 織田信長の移動経路推定

日時	既知	推定	推定確率	日時	既知	推定	推定確率
1570 年 9 月		越前国	0.504074	1574 年 11 月		越前国	0.988014
1570 年 10 月		近江国	0.525936	1574 年 12 月		越前国	0.993344
1570 年 11 月		近江国	0.629295	1575 年 1 月	越前国		1
1570 年 12 月		近江国	0.741078	1575 年 2 月		大和国	0.989675
1571 年 1 月		近江国	0.862448	1575 年 3 月	摂津国		1
1571 年 2 月	**近江国**		1	1575 年 4 月	三河国		1
1571 年 3 月		近江国	0.682636	1575 年 5 月		三河国	0.898059
1571 年 4 月		三河国	0.498432	1575 年 6 月		三河国	0.795449
1571 年 5 月		三河国	0.740088	1575 年 7 月		三河国	0.691781
1571 年 6 月	三河国		1	1575 年 8 月		三河国	0.586337
1571 年 7 月		三河国	0.838587	1575 年 9 月		三河国	0.478777
1571 年 8 月		三河国	0.687775	1575 年 10 月		近江国	0.551551
1571 年 9 月		三河国	0.548078	1575 年 11 月		近江国	0.681435
1571 年 10 月		三河国	0.420206	1575 年 12 月		近江国	0.828343
1571 年 11 月		三河国	0.305626	1576 年 1 月	近江国		1
1571 年 12 月		三河国	0.206029	1576 年 2 月		近江国	0.70717
1572 年 1 月		近江国	0.187306	1576 年 3 月		近江国	0.433539
1572 年 2 月		近江国	0.162148	1576 年 4 月		山城国	0.226821
1572 年 3 月		山城国	0.20815	1576 年 5 月		摂津国	0.382394
1572 年 4 月		摂津国	0.2867	1576 年 6 月	摂津国		1
1572 年 5 月		摂津国	0.418255	1576 年 7 月		摂津国	0.491541
1572 年 6 月	摂津国		1	1576 年 8 月		摂津国	0.406495
1572 年 7 月		摂津国	0.429042	1576 年 9 月		摂津国	0.363946
1572 年 8 月		摂津国	0.303165	1576 年 10 月		摂津国	0.365772
1572 年 9 月		陸奥国	0.375088	1576 年 11 月		摂津国	0.365958
1572 年 10 月		陸奥国	0.498288	1576 年 12 月		山城国	0.454871
1572 年 11 月		陸奥国	0.613336	1577 年 1 月		山城国	0.936523
1572 年 12 月		陸奥国	0.728341	1577 年 2 月	紀伊国		1
1573 年 1 月		陸奥国	0.84351	1577 年 3 月		大和国	0.572397
1573 年 2 月	陸奥国		1	1577 年 4 月		摂津国	0.428538
1573 年 3 月		陸奥国	0.849924	1577 年 5 月		摂津国	0.288138
1573 年 4 月		陸奥国	0.736186	1577 年 6 月		播磨国	0.293367
1573 年 5 月		陸奥国	0.599787	1577 年 7 月		播磨国	0.382155
1573 年 6 月		山城国	0.418539	1577 年 8 月		播磨国	0.432742
1573 年 7 月	山城国		1	1577 年 9 月		播磨国	0.519875
1573 年 8 月	山城国		1	1577 年 10 月		播磨国	0.510755
1573 年 9 月		紀伊国	0.512643	1577 年 11 月		播磨国	0.831622
1573 年 10 月		近江国	0.590877	1577 年 12 月	美作国		1
1573 年 11 月	近江国		1	1578 年 1 月		播磨国	0.999488
1573 年 12 月		越前国	0.504339	1578 年 2 月		播磨国	0.596002
1574 年 1 月	越前国		1	1578 年 3 月	播磨国		1
1574 年 2 月		越前国	0.993344	1578 年 4 月		播磨国	0.590693
1574 年 3 月		越前国	0.988014	1578 年 5 月		播磨国	0.577634
1574 年 4 月		越前国	0.983885	1578 年 6 月		播磨国	0.465685
1574 年 5 月		越前国	0.980956	1578 年 7 月		播磨国	0.389094
1574 年 6 月		越前国	0.979208	1578 年 8 月		播磨国	0.312321
1574 年 7 月		越前国	0.978625	1578 年 9 月		播磨国	0.245015
1574 年 8 月		越前国	0.979208	1578 年 10 月		摂津国	0.22699
1574 年 9 月		越前国	0.980956	1578 年 11 月		摂津国	0.223781
1574 年 10 月		越前国	0.983885	1578 年 12 月		摂津国	0.212737

[新井 12] 新井 紀子, 松崎 拓也 他：ロボットは東大に入れるか?, 人工知能学会誌, Vol. 27, No. 5, pp. 463–469 (2012)

[村上 16] 村上 優樹, 鶴岡 慶雅 他：現実世界の時間・空間制約を用いた共参照解析の精度向上, 言語処理学会第 22 回年次大会, pp. 250–253 (2016)

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

多段 Doc2Vec によるエンティティリンキングの応用

An entity linking method using multiple Doc2Vec

津々見誠[*1] 村上浩司[*2] 梅田卓志[*2]
Makoto Tsutsumi Koji Murakami Takashi Umeda

[*1] 楽天株式会社 Architecture & Core Technology Platform 部
Rakuten, Inc. Architecture & Core Technology Platform Department

[*2] 楽天技術研究所
Rakuten Institute of Technology

Appropriately linking a polysemous word in a text to its corresponding entity in a knowledge base is an essential part in producing structured knowledge. Linking becomes challenging as issues such as name variations, entity ambiguities or absence of entity in knowledge base occurs in the process. We present an entity linking method consisting of multiple Doc2Vec model, achieving high performance and cost effectiveness. In the experiments, our proposed method achieved 83.5% in mapping accuracy, improving 31.0 points higher than the simple string matching.

1. はじめに

近年，データ処理基盤の高性能化に伴ってデータ活用の必要性の高まりが著しい．さらに機械学習や自然言語処理分野における技術革新により，インターネット上に存在する Twitter や Facebook などの CGM (Consumer Generated Media)や，話し言葉の書き起こし，E-Commerce の商品タイトルなど，データを組み合わせて利用する試みが広くみられる．しかしながらこうした非構造化データを扱うことは容易ではなく，何らかの構造化が必要になることが多い．データの構造化は，その目的によって着目する情報の種類が異なる．典型的には大きさや材料などの属性抽出が考えられるが，話題により抽出対象が変化する場合もある．また類義語の抽出や部分全体関係抽出，上位下位関係抽出もデータ構造化の重要なタスクである．こうした構造化においての難しさの一つが，単語の多義性である．例えば「アップル」には「企業名」と「果物名」という2種類の側面があり，必要に応じて適宜構造化する必要がある．

多義性を持つ語と，その実体を正しくリンクさせるための自然言語処理分野の研究技術として，エンティティリンキング(Entity Linking)が挙げられる [1]．これは，テキスト中で何らかの実体（Entity）を指示する記述(Mention)を抽出して，外部の知識ベースに対応付ける技術である．これにより，表記のゆれや語義の曖昧性を解消してエンティティ間の関係性を正確に把握することが可能となる． データ構造化の集大成として整備が進められている外部知識ベースに Wikipedia[*3] がある．これは最も大規模なオンライン百科事典であり，多くのエンティティ間に関係のリンクが設定されており，言語資源としての価値は高い．

我々はデータの構造化の一環として，楽天市場の商品タイトルから必要な情報の抽出を行なっている．ブランドやメーカーは典型的な属性抽出の対象であるが，その名称そのものだけではなく，サブブランド，シリーズと言った上位下位関係の同定も必要となる．本研究では，楽天市場上の商品のメーカー名と，そのメーカーと一致する企業名とのマッピングの自動化に取り組む．本研究でマッピングの対象となる企業は，世間に広く認知されている公開企業とした．本研究における大きな課題は，多義性のあるメーカー名と企業名を高い精度でマッピングすることである．例えば，楽天市場に出現するメーカーに「パイオニア」があるが、これは異なる2企業で使われている．一方は電気機器を主力製品とする上場企業であり，他方は雑貨の企画及び製造販売を行う企業である．マッピングの自動化のためには，こうした企業を正確に区別し適切にマップする枠組みが必要である．

我々は，楽天市場の商品情報から予め抽出した約 70,000 種類のメーカーリストと，約 2,000 の企業リストの間でマッピングを行う．また，企業の連結子会社等もマッピングの対象とした．つまり，メーカー「NEC」や「日本電気通信システム」は，ともに「日本電気」に紐付けられることになる．同様にして，「フジテレビジョン」や「ポニーキャニオン」は，ともに「フジ・メディア・ホールディングス」を正解とする．

また使用する外部知識には，(1) 有償の企業リストなど外部データは使用しない，(2) 人手によるアノテーションはあくまで評価用データに限定し，必要な人手を最小限にする，の2点の制約を設け収集，利用した．我々は実験結果から，文書や単語の分散表現を多段に用いたエンティティリンキング手法により，従来の部分文字列一致等による手法を大きく上回る精度を得た．

2. 関連研究

2.1 エンティティリンキング

エンティティリンキングは一般に，「候補の生成」，「候補の順位づけ」，「リンク有無の予測」に分かれる [2]．候補の生成では，表記ゆれ等を列挙した辞書ベースの使用，Wikipedia のリンク構造等を利用した語の拡張，サーチエンジンの使用等の研究例がある．候補の順位づけにおいては，教師有りの手法，また教師無しの手法が見られる．教師有りにおいてはバイナリ分類器からグラフベースの手法まで様々であるが，教師データの作成に大きなコストがかかる上，これらの多くはデータセットやタス

連絡先：津々見誠，楽天株式会社 Architecture & Core Technology Platform 部，〒158-0094 東京都世田谷区玉川一丁目 14 番 1 号，050-5581-6910，ts-makoto.tsutsumi@rakuten.com

[*1] https://www.rakuten.co.jp/
[*2] https://rit.rakuten.co.jp/
[*3] https://ja.wikipedia.org/wiki/

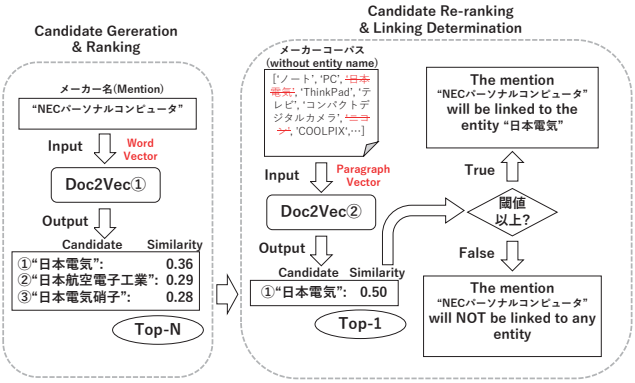


図 1　提案手法によるリンキング

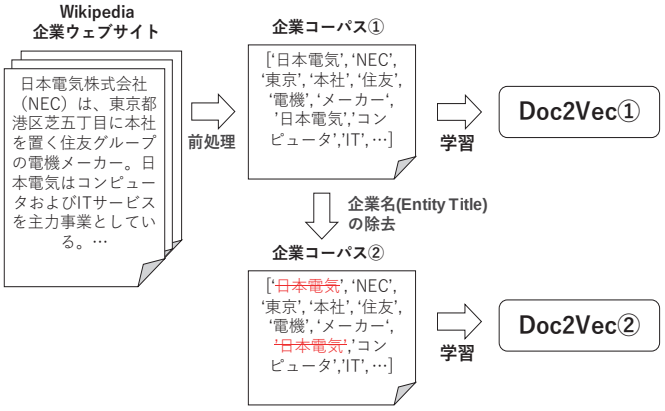


図 2　コーパスの準備と Doc2Vec モデルの学習

クに依存することから，異なるドメインへのモデルの転用が難しく，汎用性に欠ける [3].

日本語のエンティティリンキングの研究は英語圏と比較して少ない．古川らは学術分野の知識ベースが英語中心であることに着目し，日本語の用語を英語に対訳する手法を提案している [4]．また Zhou らの手法では Wikipedia のアンカーテキストを用いたエイリアス辞書により語を拡張し，Entity タイトルとの文字列間距離をとって候補を生成する．順位づけでは文字列距離，エンティティの被リンク数，周辺語の BoW(Bag-of-Words)，分散表現等を素性に用いた上で，SVM による教師有り学習を行うことで良好な結果が得られている [5]．これらの研究は候補生成と候補の順位づけにおいて，異なる手法またはモデルを採用している共通点がある．候補生成においては辞書の活用や語の拡張に留まり，分散表現を利用した研究は多くない．

我々の研究は，以下の 2 点で Zhou らの研究と大きく異なる．(1) 候補生成と候補順位づけで使用するモデル基盤を共通化する．すなわち，Doc2Vec を多段に使用する．この際，語の拡張やエイリアス辞書の作成を必要としない．(2) 候補生成，候補順位づけともに教師なし学習である．このため，教師データの作成を必要としない．これらの理由から，提案手法は従来手法と比較して大幅な省力化が可能である．

2.2 Doc2Vec

Doc2Vec[1]とは，単語を数百次元程度の実数ベクトルで表現する手法である Word2Vec を，文書単位へと拡張した手法である [6] [7]．Word2Vec はニューラルネットワークモデルであり，中心の単語から周辺の単語を予測するモデルである Skip-gram と，周辺の単語から中心の単語を予測するモデルである CBoW（Continuous Bag-of-Words）に分かれる．Doc2Vec においては，Skip-gram を拡張したものは PV-DBOW(Paragraph Vector with Distributed Bag-of-Words)，CBoW を拡張したものは PV-DM(Paragraph Vector with Distributed Memory)と呼ばれる．

Doc2Vec は分散表現として単語ベクトル(Word Vector)とパラグラフベクトル(Paragraph Vector)を保持し，これらはニューラルネットワークモデルの中間層の重みで表現される．パラグラフベクトルは局所的な文脈の中で失われた情報を代表するものであり，パラグラフのトピック等抽象度の高い情報を記憶する機能を持つ [7]．離散的で高次元，かつスパースな BoW による表現と比較して，分散表現は低次元で密な表現であり，単語や文書間の距離をより高い精度で計算可能である．

3. 提案手法

3.1 基本的な考え方

メーカー名と企業名とのリンキングを考えた場合，基本的には着目するメーカー名に近い名を持つ企業が候補として考えられる．しかし，1節の「フジテレビジョン」や「ポニーキャニオン」のように，複数のメーカーが単一の企業に結びつく場合や，「パイオニア」のように，同一のメーカー名が異なる企業を指す場合を考慮する必要がある．つまり語やエンティティの多義性が大きな問題となる．我々はこの問題に対して，(1)任意の文書には，用語・文脈・文書トピック等，抽象度の異なる情報が存在する，(2)語やエンティティの多義性の解消には，これら抽象度の異なる情報を組み合わせることが有効である，の2つの仮説を立てた．人間は多義性のある用語の解決を試みるとき，語そのものからは一旦距離をおき，より広い文脈から俯瞰して用語の意味を検討する．上記の仮説はこの流れに沿うものである．

上記の仮説に基づき，提案手法では Doc2Vec モデルを多段に組み合わせ，抽象度の異なるコーパスと分散表現を使用することで語とエンティティの多義性解消を試みる．1段目ではエンティティ名を含むコーパスと Word Vector を使用することで，局所的で具体性の高い用語レベルの情報を重視し，候補漏れを最小限に抑えた候補生成と順位づけを行う．2段目ではエンティティ名を除いたコーパスと Paragraph Vector を使用することで，大域的で抽象度の高い情報を重視し，語の多義性を解消しつつ，候補の再評価とリンキング有無の決定を行う．

3.2 手法の流れ

提案手法は大きく(1)コーパスの準備，(2)リンキング候補の生成と順位づけ (3)候補の再評価とリンキング有無の決定，の 3 つから構成される．全体の処理を図 1 に示す．

(1)　コーパスの準備

ナレッジベースとなる企業側の情報として，Wikipedia および企業のウェブサイトを使用した．

最初に Wikipedia 上に存在する企業のページを取得した．Wikipedia はエンティティ間に多数のリンクを内包し，さらにページ内の重要な用語には対応するエンティティへのリンクが付加されている．一方で，知名度の低い企業等はページやリンクが存在しないケースが多く，これらの情報を利用できない．こうした情

[1] https://radimrehurek.com/gensim/models/doc2vec.html

報の欠損に対処するため，Wikipedia に加えて各企業のウェブサイトから，会社情報や事業紹介等を中心に 3 ページほど取得した．さらに Wikipedia は id=content タグ内を，企業ウェブサイトは body タグ内を対象とし，HTML タグや javascript コード等を除去した．次に各文書に対して以下の前処理を行ったのち MeCab [8]＋ipadic-utf8（ユーザー辞書に企業名を追加）の組み合わせで形態素解析し，名詞を抽出してコーパス（企業コーパス①）とした．

- 英数字は，全角から半角に変換
- 頻度の高い用語のうち，「項目」「関連」等，具体性が低くノイズとなりそうな用語をストップワードとして除去

上記の前処理に加えて，2 段目の Doc2Vec モデル用に企業コーパス①から企業名を除去したコーパス（企業コーパス②）も用意する．

またモデルへの入力とするリンキング元の情報として，予め抽出したメーカー名リストに加えて，各メーカーを表現する文書を準備する．作成にあたり，「商品価格ナビ」のウェブページを利用した[1]．本ページでは任意のメーカーの主要製品をジャンル別の売上順に把握できる．このページ及び上位 4 製品の商品情報をメーカーの文書とし，企業コーパス②と同様に前処理，名詞抽出，企業名削除を行いメーカーコーパスとした．

(2) リンキング候補の生成と順位づけ(Doc2Vec 1 段目)

候補となるエンティティの生成と順位づけを行う．本段階での目標は，約 2,000 ある企業のうちリンキングすべき企業を TOP-N の範囲に含めることである．ここで TOP-N に含まれない場合，後述する段階での挽回が不可能となるため，高い精度で TOP-N に含める仕組みが求められる．さらにメーカー「ポニーキャニオン」と企業「フジ・メディア・ホールディングス」のように名称に類似性がないリンキングも少なくない．この問題の解決のため，表層的な文字列間距離ではなく単語の分散表現を使用した．具体的にはメーカー名の Word Vector と企業名の Word Vector との間のコサイン類似度を計算し，上位 N 件の企業を TOP-N として選び出す．

Doc2Vec モデルの学習に用いたパラメータは以下の通りである．これらは予備実験を通して最良な結果が得られた組み合わせである．図 2 にモデル学習の概要を示す．

- モデルとして，PV-DBOW を使用する．
- Word Vector の再学習を行う．
- 周辺単語の窓幅=5，中間層サイズ=300，エポック数=10

(3) 候補の再評価とリンキング有無の決定(Doc2Vec 2 段目)

生成された候補を TOP-N の範囲で再度順位づけし，TOP-1 を選び出してリンキング有無の決定を行う．本段階での目標は，候補を別の観点から再評価し，リンキングの誤りをできるだけ減らすことである．例えば，「パイオニア」のように企業名が同一で実体が異なる企業の判別は，メーカー名を入力とする 1 段目のモデルだけでは原理上不可能である．2 段目では，これら多義性を含む語の語義曖昧性の解消が求められる．

そのため 2 段目の Doc2Vec では，モデルとパラメータは1段目と共通であるが，コーパスとベクトルの種類を変更して学習を行う．具体的には企業コーパス②を用いて学習させた Doc2Vec モデルに対して，メーカーコーパスを入力とし，その Paragraph Vector と企業の Paragraph Vector との間のコサイン類似度を計算する．このとき類似度が最大となる候補を TOP-1 として決定する．最終的に，この類似度が閾値を上回ればリンキングし，下回る場合はリンキングなしとする．

4. 評価実験

4.1 データセット

提案手法の有効性を確認するため，表 1 に示すような評価用データを手作業で 200 件作成した．企業名が「-」の部分は，メーカー名に対応する公開企業データが存在しないことを示す．例えば 1 行目のメーカー「パイオニア」は上場企業である「パイオニア」が対応するが，2 行目の「パイオニア」はワッペンを主力製品とする非公開企業であり，対応する公開企業データは存在しない．同様にメーカー「アサヒビール」は企業「アサヒグループホールディングス」と対応するが，「アサヒシューズ」は靴メーカーであり企業「アサヒグループホールディングス」とは無関係である．テストデータ 200 件中，対応企業有りのメーカー数は 142 件，対応企業なしのメーカー数は 58 件であった．また，メーカー名と企業名が完全一致するペアは 113 件，部分一致は 63 件，不一致は 24 件であった．

4.2 実験方法

実験では，(1)「テキストマッチング」，(2)「Doc2Vec1 段」，(3)「Doc2Vec2 段」により構築したシステムのマッピング精度を比較する．

表 1 評価用データセット例

メーカー名	正解企業名	文字列一致度	関係性
パイオニア	パイオニア	完全一致	同一企業
パイオニア	－	(完全一致)	関係性無し
アサヒビール	アサヒグループホールディングス	部分一致	子会社
アサヒシューズ	－	(部分一致)	関係性無し
麒麟麦酒	キリンホールディングス	不一致	孫会社
ポニーキャニオン	フジ・メディア・ホールディングス	不一致	連結子会社

表 2 各手法による精度の比較

リンキング手法		精度(%)
テキストマッチング		52.5
Doc2Vec1 段	Paragraph Vector	46.0
	Word Vector	71.5
Doc2Vec2 段	Paragraph Vector（1 段目） Word Vector（2 段目）	61.5
	Word Vector（1 段目） Paragraph Vector（2 段目）	78.5
	Word Vector（1 段目） Paragraph Vector（2 段目） 2 段目企業名除去	**83.5**

[1] https://product.rakuten.co.jp/

表 3　誤り例の分析

メーカー名	正解企業名	予測企業名	類似度①	類似度②	備考
エスピーケイ	SPK	- (小林製薬)	0.41	0.16	1段目で小林製薬が候補に選ばれた. 2段目では類似度②が閾値 0.18 以下より, 「対応企業無し」と予測
ソニー・ ミュージックダイレクト	ソニー	- (ソニー)	0.91	0.09	1段目でソニーが候補に選ばれた. 2段目では類似度②が閾値 0.18 以下より, 「対応企業無し」と予測

ここで, 先頭 5 文字一致によるテキストマッチングをベースラインとした. すなわち, メーカー名の先頭 5 文字が, 企業名に含まれる場合にメーカーと企業をリンキングする. リンキング候補となる企業が複数存在する場合は, 最初に出現した候補をリンキングする. 閾値は各アルゴリズムで最良の精度となる閾値を探索的に求めた. 予備実験から, 2 段構成のシステムにおける 1 段目における最適な TOP-N 数は 1 とした.

4.3 実験結果

実験結果を表 2 に示す. まず Doc2Vec1 段では, Paragraph Vector による手法が 46.0%とベースラインの 52.5%を下回ったのに対して, Word Vector による手法は 71.5%であり, ベースラインより 19.0 ポイント高い精度を示した. また Doc2Vec2 段では, 1 段目に Word Vector, 2 段目に Paragraph Vector を用いた手法は 78.5%であり, Doc2Vec1 段と比較して更に 7.0 ポイント向上したことから, Doc2Vec を多段に使用することの有効性が確認された. 最も精度が高かったのは, Doc2Vec2 段, かつ2段目のモデルを企業コーパス②で学習させた場合の 83.5%であり, ベースラインと比較して精度が 31.0 ポイント向上した. この結果から, 仮説で示した抽象度の異なる情報を組み合わせることの有効性が確認された.

4.4 考察

実験結果において最も精度が高い手法について, 33 の誤り例を分析した. まず, 対応する企業が存在するにもかかわらず 1 段目で TOP-N 圏外となった誤り例が 10 例存在した. 残りの 23 例は 2 段目での順位づけまたは閾値による対応づけで失敗していた. それぞれの誤り例を表 3 に示す. ここで類似度①は Word Vector による類似度を, 類似度②は Paragraph Vector による類似度を示す.

まず表 3 の 1 段目の誤り例では, 正解は企業「SPK」であるのに対し, 1 段目では誤って「小林製薬」と予測した. Wikipedia で小林製薬のページを確認すると, 小林製薬には関連企業として「エスピー・プランニング株式会社」があり, 「エスピー」が重要な特徴として予測に影響を与えた可能性が考えられる. 実際, 「エスピーケイ」を本研究で用いた形態素辞書を用いて改めて分かち書きすると, 「エスピー」と「ケイ」に分かれることを確認した. 表記揺れに対しては逐次ユーザー辞書に追加することで回避可能だが, 辞書の構築と維持にはコストがかかる. 効率の良い辞書構築法や固有表現の抽出法の検討は今後の課題である.

2 段目の誤り例では, 正解は企業「ソニー」であり, 1段目では正しく TOP-1 に選出されたものの, 2 段目では Paragraph Vector の類似度が 0.08 と非常に低くなり, 閾値以下となったことから対応企業無しと誤判定された. これは, (1)ソニーは企業規模が大きく, 事業が多角化されていること, (2) ソニー・ミュージックダイレクトの商品は主に CD 等音楽関連商品に限定されること, (3)ドキュメントのソースがメーカー側と企業側で異なる(メーカー側は商品情報が主であり, 企業側は企業情報が主である), の大きく3点の理由から, 企業「ソニー」を代表するコーパスと, メーカー「ソニー・ミュージックダイレクト」のコーパスとの間の語彙分布に大きな差異が生じ, 結果として Paragraph Vector の類似度が低い値になったと推測される. このようにコーパスの性質の違い等に起因する語彙分布の差異をどう吸収するかも, 今後の課題である.

5. 終わりに

本研究では, 楽天市場の商品のメーカーとその企業との自動マッピングに取り組んだ. 役割の異なる Doc2Vec モデルを多段に重ねて複数の観点からリンキングの正当性を検証する提案手法は, 用語の多義性の解消に対して有効であることを示した. さらに, この手法は辞書による語の拡張, 教師データの作成といった人手を要する工程を必要とせず, コスト面の優位性が大きい. 今後は, 一般的なデータセットを使用した本手法の有効性の確認や, より適切な固有表現抽出, コーパスの差異にロバストな手法の開発に取り組む予定である.

参考文献

[1] R. Mihalcea , A. Csomai, “Wikify!: linking documents to encyclopedic knowledge,” Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, 2007.

[2] W. Shen, J. Wang , J. Han, “Entity linking with a knowledge base: Issues, techniques, and solutions,” IEEE Transactions on Knowledge and Data Engineering, 2015.

[3] X. Ling, S. Singh , D. S. Weld, “Design Challenges for Entity Linking,” Transactions of the Association for Computational Linguistics, 2015.

[4] 古川竜也, 相良毅 , 相澤彰子, “言語横断エンティティリンキングのための語義曖昧性解消,” 情報知識学会誌 24.2, 2014.

[5] S. Zhou, N. Okazaki, K. Matsuda, R. Tian , K. Inui, “Supervised Approaches for Japanese Wikification,” Journal of Information Processing 25, 2017.

[6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado , J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” Advances in neural information processing systems, 2013.

[7] Q. Le , T. Mikolov, “Distributed Representations of Sentences and Documents,” International Conference on Machine Learning, 2014.

[8] T. Kudo, “MeCab : Yet Another Part-of-speech and Morphological Analyzer,” IEICE technical report 110(85), 2006.

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

CycleGAN を用いた感情スタイル転送

Emotional Style Transfer using CycleGAN

大田 和寛 *1
Kazuhiro Ota

*1株式会社サイバーエージェント
CyberAgent, Inc.

Our emotions or impressions to the images largely depend on the visual features of each image. If people in a photography are smiling, we might perceive a happy impression. On the other hand, if violent aliens are drawn, we may feel scary. We consider an emotional style transfer model that transforms an impression of the given image to the desired one. We utilize adversarial training of pixelwise transformation to learn a model with Behance Artistic Media dataset and report initial results.

1. はじめに

近年，様々なストックフォトサービスの台頭により容易に高精細な画像を手に入れることが可能となった．画像に付与されているタグで検索するだけでそのタグに該当する画像を収集できるため，バナー広告のクリエイティブ制作現場などでも素材候補として頻繁に利用されている．しかしながら，ストックフォトサービスにおけるタグデータは基本的にユーザが付与するものであり，写真の被写体や画像の持つ印象に大きく依存するため，しばしば検索のノイズとなり易い [Yang 17]．Wilber らはこのような画像とメタデータとのミスマッチ問題に対応するために Behance Artistic Media（BAM）*1 というデータセットを提案している [Wilber 17]．これは画像の被写体，表現媒体，感情における全 20 種の属性について，クラウドソーシングによりそれぞれの属性に対するスコア付けを行いメタデータとして添付したデータセットとなっており，コンピュータビジョンによる写真に留まらない画像表現の理解・学習の促進を期待するものである．

一方で，コンピュータビジョンにおける Deep Learning の応用事例の中でも画像表現に着目した興味深い研究が報告されている．Gatys らはコンテンツ画像とスタイル画像を入力としてコンテンツ画像に対しスタイル画像の画風のみを転送するといった画風スタイル変換手法を提案している [Gatys 16]．Goodfellow らはノイズから画像を生成する生成器ネットワークとその生成画像に対し本物・偽物の判断を行う判別器ネットワークとを連結させて互いに競い合うように学習することで本物に近い画像を自動生成する Generative Adversarial Network（GAN）と呼ばれるネットワーク構造を提案している [Goodfellow 14]．Isola らが提案した手法である pix2pix は画像ペアごとの変換規則およびそれぞれのドメイン間で固有の損失関数を取得することで手書きエッジから写実的な物体を生成するような所謂 Image-to-Image と呼ばれる変換を行っている [Isola 17]．また，Zhu らはさらにこの pix2pix から学習データのペア制約を取り除くことにより，教師なし学習の枠組みで Image-to-Image を行う CycleGAN を提案している [Zhu 17]．

本稿では，BAM に含まれる感情属性のスコアをもとにデータセットを “happy”，“scary” に分割し，感情ドメインにおける Image-to-Image を CycleGAN にて行うことで感情表現の相互転送手法を提案する（図 1）．画像データから感情表現を抽出することができれば，幸福や恐怖といった画像が有する印象をもとに素材を加工することが可能となる．これは例えばバナー広告のデザイン元の素材画像に対して感情を付与できることを意味している．ストックフォトサービスの画像におけるタグやフリーワードなどのメタデータの信頼度を強固にできるなどの応用が考えられ，バナー広告の制作現場に対する支援の一手となり，これまで手作業で行っていた印象に基づく画像操作の指標となり得るものである．

図 1: 本稿では BAM に含まれる “happy” と “scary” という 2 つの感情ドメインの画像群に対し CycleGAN を用いた双方向の感情スタイル転送手法を提案する，笑顔の女性をまるでホラー映画のキャラクターのように変換し，骸骨や物寂しい風景画をカラフルに彩色している．

2. 提案手法

2.1 BAM

BAM に含まれているメタデータの中で本稿において重要なものを以下に簡潔にまとめておく．詳細は原著論文 [Wilber 17] を参照されたい．

連絡先: 大田和寛，株式会社サイバーエージェント，ota_kazuhiro@cyberagent.co.jp

*1 https://bam-dataset.org/

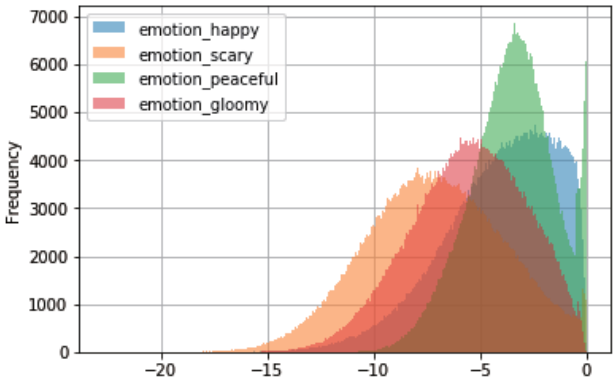


図 2: BAM における感情ドメインの属性 4 種 “emotion_happy”，“emotion_scary”，“emotion_peaceful”，“emotion_gloomy”のスコア分布．

属性
画像の被写体（“content_people”，“content_car”など），表現媒体（“media_pen-ink”，“media_comic”など），感情（“emotion_happy”，“emotion_scary”など）によって分類された全 20 種類の属性名である．

ラベル
クラウドソーシングとラベル付アルゴリズムとの組み合わせである Human-in-the-Loop 手法 [Yu 15] により，画像が上記 20 種のいずれかの属性として妥当性を持つか否かの判断結果が “positive”，“unsure”，“negative”の 3 値で与えられている．

スコア
ラベル同様 Human-in-the-Loop 手法によって得られた値だが，画像 1 件に対し全属性分の信頼度が負値の連続値で格納されている．図 2 に感情ドメインの属性におけるスコアの分布を示す．

図 3 に BAM における感情ドメインの属性である “emotion_happy”と “emotion_scary”におけるスコア上位の画像群を幾つか示す．いずれも双方の感情表現が画像によく現れており，スコアでのソート結果が画像特徴の理解を促進するものであることが伺える．

2.2 CycleGAN

ある画像集合 X および Y があったとき，CycleGAN は X，Y をそれぞれ互いに変換する $G^\star : X \to Y$，$F^\star : Y \to X$ を学習する．その際の損失関数は以下のように設計されており，

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) &= \mathcal{L}_{\mathrm{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\mathrm{GAN}}(F, D_X, X, Y) \\ &+ \mathcal{L}_{\mathrm{cyc}}(G, F), \end{aligned} \quad (1)$$

これを以下の mini-max 式で解くことで得られる変換が $G^\star$，$F^\star$ である．

$$G^\star, F^\star = \arg\min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y). \quad (2)$$

ここで D_X，D_Y はそれぞれ $F : Y \to X$，$G : X \to Y$ によって変換された画像に対する GAN 判別器 [Goodfellow 14] であり，式 (1) の第 1 項，第 2 項に示す $\mathcal{L}_{\mathrm{GAN}}$ によって F，G に

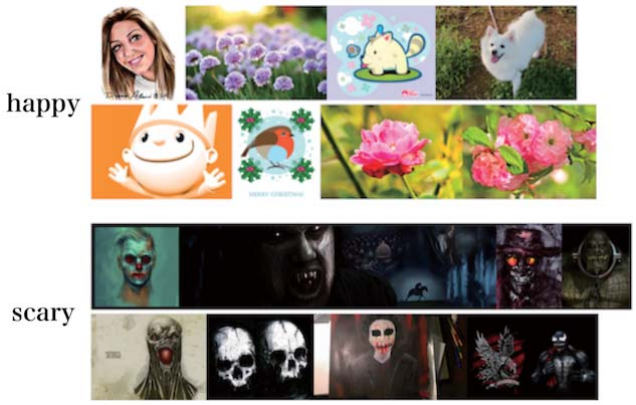


図 3: BAM における “emotion_happy”属性と “emotion_scary”属性のスコア上位に属する画像の例．“emotion_happy”には動物や色鮮やかな草花が多く含まれ，“emotion_scary”にはダークな色合いで凶暴性をもったキャラクターが見受けられる．

よる変換結果に対する敵対的損失を計上する．さらに式 (1) の第 3 項に示す $\mathcal{L}_{\mathrm{cyc}}$ は下記のように定式化されており，

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(G, F) &= \mathbb{E}_{x \sim p_{\mathrm{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\mathrm{data}}(y)}[\|G(F(y)) - y\|_1], \end{aligned} \quad (3)$$

X，Y 両ドメインの画像に対し F，G を交互に作用させた際のオリジナルデータとの差異の期待値を復元損失として計上するものである [Zhu 17]．近年の応用事例としては情報隠蔽への応用 [Chu 17] や，人間の顔をうまく交換した提案 [Jin 17] も報告されており，非ペア制約条件下におけるドメイン間のスタイル変換に対し印象的な結果をもたらしている．

2.3 概要

BAM データセットの属性種別のうち “emotion_happy”および “emotion_scary”のスコアでの上位 1,000 件分の画像群をそれぞれ “happy”ドメイン，“scary”ドメインとし，CycleGAN によりこのドメイン間で画風スタイルの変換学習を行う．図 4 に概要図，図 5 および図 6 に両ドメインの学習データにおける属性とラベルのペアの分布図を示す．“happy”ドメインを X，“scary”ドメインを Y として学習データと考えた時，$G : X \to Y$，$F : Y \to X$ を CycleGAN によって最適化することに相当し，それぞれのドメインが相互に教師ドメインの役割を果たすため，BAM のように明確な教師ラベルデータが付与されていないデータ群に対し確率的な学習効果を発揮すると考えることができる．

3. 実験

3.1 概要

画像のサイズは全て 286 × 286 ピクセルにリサイズし，最初の 100 エポックは学習係数 0.0002，次の 100 エポックは学習係数 0 まで線形減衰させる形で合計 200 エポックの学習を，PyTorch による実装 *2 を利用し CUDA を介した NVIDIA GTX 1080 Ti 上で実施した．

*2 https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

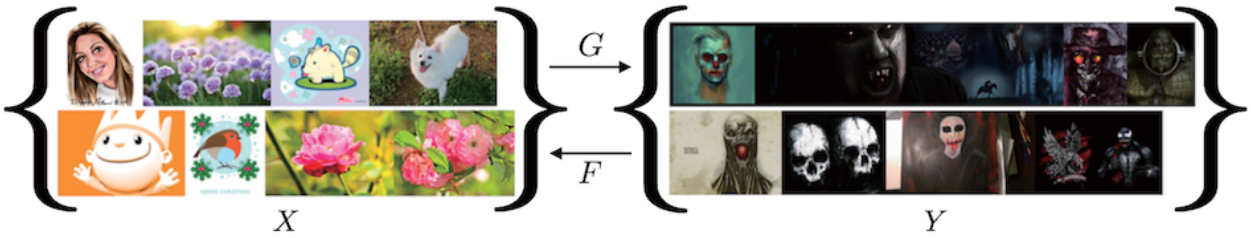


図 4: CycleGAN による感情ドメイン間での画風変換学習の概要図．BAM から抽出した “happy” ドメイン X と “scary” ドメイン Y の間のギャップを相互変換する関係 $G : X \to Y$ と $F : Y \to X$ をペア制約のない教師なし手法で学習する．

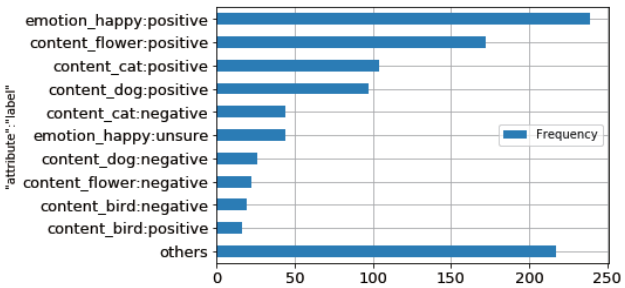


図 5: “happy” ドメイン学習データに含まれる属性値とラベルのペアの内訳．“emotion_happy:positive”，“content_flower:positive”，“content_cat:positive” などが上位を占めている．

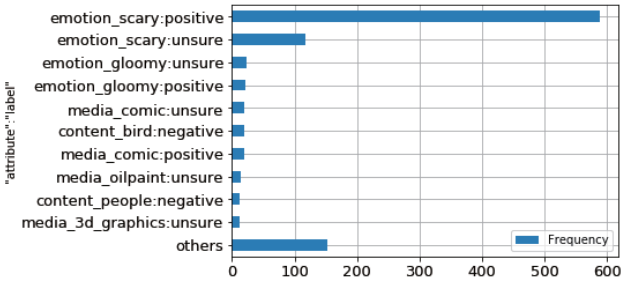


図 6: “scary” ドメイン学習データに含まれる属性値とラベルのペアの内訳．“emotion_scary:positive”，“emotion_scary:unsure”，“emotion_gloomy:positive” などが上位を占めている．

3.2 結果考察

“emotion_happy” および “emotion_scary” のスコアでの上位 1,001〜1,200 件の画像をそれぞれ学習された変換 $G^\star$，$F^\star$ にかけ，得られた画風スタイル変換結果に対し官能的評価を行った．

結果例を図 1 及び図 7 に示す．“happy” →“scary” に関してはカラフルで明るい色味を持った画像が赤黒色を保存しつつエッジをより際立たせるような形での減色がなされ，“scary” →“happy” に関してはモノクロや線画の画像を人間や植物の色に近い形で色付けされエッジを平滑化するような効果が見受けられる結果となった．これは図 5 と図 6 に示した学習データにおける属性値とラベルのペアの分布がうまく反映された結果だと考えられ，CycleGAN による明示的なラベルデータの準備を行うことのない感情表現の転送が達成できたと言える．また，今回の感情表現転送においてあまり視覚的な効果が見受けられなかった例を図 8 に幾つか示す．それぞれのドメインが片方のドメインの特徴を有している場合において期待する効果が得られなかったと考えられる．

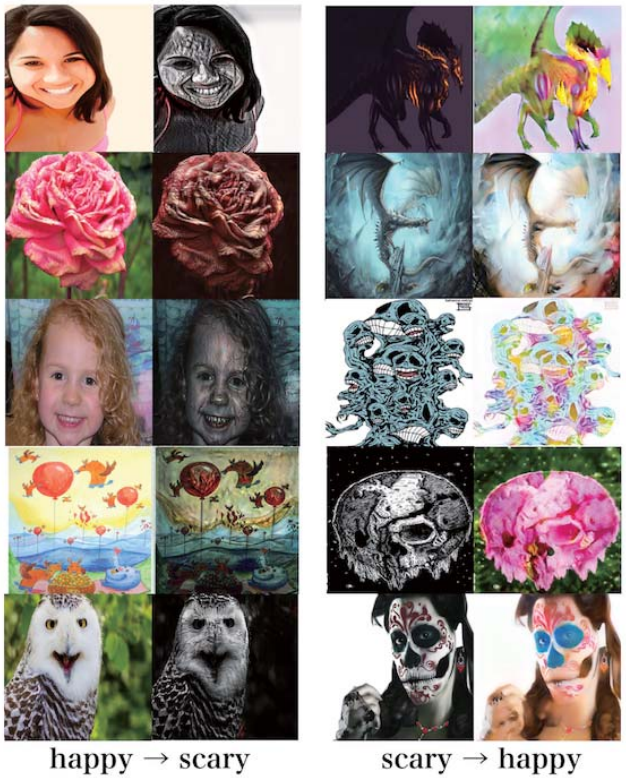


図 7: BAM 感情スタイル変換実験における結果例．“happy” → “scary” においては赤黒色を保存しつつ明度・彩度を下げエッジを際だたせるシャープネス処理がかかった効果が学習された．一方，“scary” → “happy” においては人間の肌色や髪色を考慮しつつ自動彩色し平滑化処理がかかりエッジがぼかされるような効果が確認できる．

4. まとめ

本稿では BAM データセットがもつ属性における 4 種類の感情属性のうち “happy” および “scary” の 2 ドメイン間に存在する画風表現の違いを把握し，CycleGAN を用いた実験によりそれぞれのドメイン間での画風スタイルの転送を行った．笑顔の人間の顔に痣が現れたり，異形のエイリアンが草花のように自動彩色されたりといった画風スタイル変換が行われ，感情ドメイン間における表現の違いがそれぞれの画像群に対して視覚的な変化を与えることが確認できた．これはバナー広告

図 8: BAM 感情スタイル変換実験において視覚的な効果が見受けられなかった結果．“happy”の画像に黒色が多かったり，“scary”の画像がカラフルであったりと，スタイル転送元の画像が片方のドメインの特徴を既にある程度有していることが伺える．

のクリエイティブに利用される素材のような高精細画像に対しても有効に活用できることが見込まれるため，素材データのバリエーションの拡張に繋がると考えられる．しかしながら，今回は 2 種類の感情のみの CycleGAN での画風スタイル転送結果しか得られていないため，定量的評価は今後の課題である．また，画風の転送は行えたものの，“happy”ドメインの女性の笑顔が笑顔のままであったり，“scary”ドメインの骸骨が骸骨のままであったりとコンテンツへの考慮の不足といった問題もある．今後，損失関数の設計など，改善点を見つけ何らかの対応を進めていきたい．

参考文献

[Chu 17] Chu, C., Zhmoginov, A., and Sandler, M.: CycleGAN, a Master of Steganography, in *NIPS* (2017)

[Gatys 16] Gatys, L. A., Ecker, A. S., and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, in *CVPR* (2016)

[Goodfellow 14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative Adversarial Nets, in *NIPS* (2014)

[Isola 17] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in *CVPR* (2017)

[Jin 17] Jin, X., Qi, Y., and Wu, S.: CycleGAN Face-off, *arXiv*, Vol. 1712.03451, (2017)

[Wilber 17] Wilber, M. J., Fang, C., Jin, H., Hertzmann, A., Collomosse, J., and Belongie, S.: BAM! The Behance Artistic Media Dataset for Recognition Beyond Photography, in *ICCV* (2017)

[Yang 17] Yang, F., Kale, A., Bubnov, Y., Stein, L., Wang, Q., Kiapour, M. H., and Piramuthu, R.: Visual Search at eBay, in *KDD* (2017)

[Yu 15] Yu, F., Zhang, Y., Song, S., Seff, A., and Xiao, J.: LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop, *arXiv*, Vol. 1506.03365, (2015)

[Zhu 17] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networkss, in *ICCV* (2017)

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

ナイーブベイズ法に基づく SNS を利用したペルソナ推定

Persona Estimation using SNS based on Naive Bayes Method

山岡 拓生 [1] 佐野 睦夫 [1]

Hiroki Yamaoka[1] Mutsuo Sano[1]

[1] 大阪工業大学, 情報科学部

[1]Media Information Department, Faculty of Information Science and Technology,

Osaka Institute of Technology University.

Abstract : High-capacity data with various types of information updated frequently is called big data and SNS is also one of them. In recent years, big data has improved research /product service quality and management strategy, etc. This research is aimed at estimating detailed items of persona in marketing, and we propose individual posting from big data in this paper. We propose an individual posting method from big data and hobby analysis method for users. The proposed methods can make it easy to detect the language features and analyze SNS. We apply a Naive Bayes method to create classifiers and estimate their hobbies. The result of estimating their hobbies was an average of about 93% correct answer rate. We discuss that the proposed methods could create their persona.

1. はじめに

本論文では, Twitter 上の個人アカウントの投稿情報より趣味の推定を行う. この推定を行う事によって, ターゲットの趣味や趣味となりうる趣味の推定を行う事を目的としている.

近年, 様々な形でビッグデータが盛んに利用されている. その利用範囲は経済や医療, その他のサービス等でデータを利用した解析や研究などが行われている. その中でもマーケティング分野での利用は大きいと考えられる. しかしながら, ビッグデータを利用した解析などにおいてはすでに確立されたメディアに対して行うものが多く, 新規事業での利用はあまり多くないといえる. そこで本研究では, マーケティング手法の 1 つであるペルソナマーケティングにおけるペルソナをビッグデータを利用し, 作成を行う事を目的とし, 今回はその中でも趣味の項目を作成した.

2. ペルソナ項目推定

本研究の目的にあるペルソナとその項目とは, ペルソナ/シナリオ法というマーケティング手法に利用し作成されるものであり, 架空のユーザやターゲットの代表像の事である. ペルソナは実際に存在するユーザのデータを基にして作成されており, その項目はパーソナルデータなど様々項目が存在している. 項目の例を次の図 1 に示す.

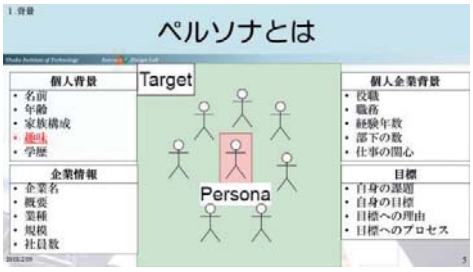


図 1 ペルソナ

本研究では, 趣味項目の分類に機械学習のナイーブベイズ法を利用し, 学習には Twitter 上の文章を利用した.

本研究では趣味の分析を行うが, 実際での趣味の数は膨大であり抽象的な趣味もあるため, 今回は統計局で公開されている平成 28 年社会生活基本調査の結果 [1]より, 年代を問わず総計での人口が多い趣味をいくつか対象として分析を行った.

また本研究での分析は, 次の 4 つの手順で分析を行う.

まず初めに 1 つ目は分析のためのデータ収集である. 2 つ目にデータ分析はほとんどが収集したデータそのままの形式では利用ができないため, 前処理が

必要となる.3 つ目に 2 つ目の前処理を行って成形されたデータを用いて分類器を作成する.4 つ目は三つ目で作成された分類器を用いて未知データの分類を行う事である．システム概要は以下の図 2 にまとめて記述している．

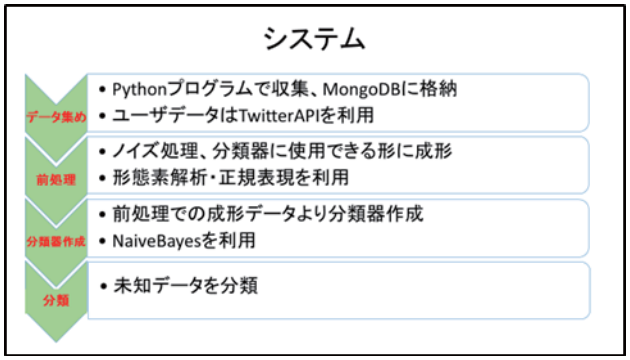


図 2　本研究システム概要

次項から，システムの各手順について詳述していく．

2. 1. データ収集

本手順では，TwitterDevelopper より提供されている TwitterAPI [2]と Python [3]を用いて，Twitter [4]から各趣味に適したユーザ情報を抽出し，データベースに蓄積を行う．また，本手順で Twitter [1]を利用した理由としては， TwitterAPI を利用することによってプログラムでの操作が容易となることや，リアルタイム性の高い情報を得られることなどがあり．その他にも Twitter には，自身の趣味に関しての投稿に制限して投稿を行う趣味アカウントなどが存在しており，それらは個人の趣味嗜好に沿った考えが近いデータを取り扱えることなどがあげられるためである．

2. 2. 前処理

前処理の手順では，収集されたデータを後述する本研究で用いられる分類器の作成に利用するために，データの成型を行う．詳細は後述するが，本研究では NaiveBayes 法を用いてデータの分類を行う．NaiveBayes 法を適用するためには，ユーザのテキスト情報を形態素に分解する必要がある．よってそれに伴いデータの成型を行うために形態素解析を行う．形態素解析には Python で利用しやすい形態素解析ライブラリである Janome [5]を利用する．

2. 3. 分類器作成

分類器作成では，前処理で形態素解析によって分解された文章の語彙の中から名詞と動詞を抽出して分類器を作成する．また今回の分類器作成には前述したように NaiveBayes 法を用いて作成する．NaiveBayes 法については以下に詳述する．

2. 3. 1. NaiveBayes 法

NaiveBayes 法とは教師ありの機械学習の 1 つであり，テキスト分類を行う手法の中でも単純な仕組みであるが高速で精度が高い手法である．この手法は，ベイズの定理を応用した次の式(1)で表すことができます．式中の C はカテゴリを表し，D は文書を表しています．ここでの文書とは単語の集合のことです．

$$P(C \mid D) = (P(C)P(D \mid C))/P(D) \propto P(C)P(D \mid C) \quad (1)$$

$$P(D|C) = P(W_i \wedge \cdots \wedge W_k|C) = \prod_i P(W_i|C) \quad (2)$$

この式(1)における P(C|D)は事後確率とよばれ,本手法ではテキストのカテゴリが未知の文書では，事後確率が最も高いカテゴリへと分類をします．この分類推定を MAP 推定と呼びます．

ここでの文書 D は bag-of-words であり，単語間に繋がりがない単語 W の集合であり，その単語間の独立性を仮定した時に以下の式(3)は成り立つ．

ただし，実際の文書では単語間で出現率に共起が起こるため単語間の独立性が成り立つことはない．そのためあくまで仮定を行った上で式(2)の様に文書の確立の積で表す式が NaiveBayes 法である．

また，式(2)における P(Wi |C)は単語の条件付き確率とも言い，カテゴリ中の単語の出現のしやすさを表している．これは訓練データ中の特定カテゴリ C におけるある単語 W の出現数をカテゴリ C 中の全単語数で割ることによって求めることができる．

以上の結果よりまとめた式が次の式(3)である．

$$M = argmax_C logP(C|D) = argmax_C logP(C) \prod_i P(W_i|C) \quad (3)$$

NaiveBayes 法での実装式は上記のとおりだが，実装に関してゼロ頻度問題が発生する．このゼロ頻度問題とは訓練時に存在しなかった未知の単語が含まれている文書を予測する場合に，計算式に席を用いているために結果がゼロになってしまうという問題がある．この問題には学習にない未知の単語の出現回数に $\alpha = 1$ を加算するラプラススムージング等が代表的な対策であり，本研究ではこちらのラプラススムージングと出現回数に $0 < \alpha \leqq 1$ の範囲で値を加算する Lidstone スムージングを採用した．今回の α には α = 0.1, 0,5, 1 の 3 つの値で検証を行った．

2.4. 分類

本手順では，前手順で作成した分類器を用いて未知の文書を分類した．分類結果には K 分割交差検証と混同行列を用いて分類器の性能評価をおこなった。

3. 結果

表 1　K 分割交差検証結果

α	0.1	0.5	1
平均正解率	0.8704	0.8600	0.8496
正解率の標準偏差	0.0657	0.0587	0.0636

表 2　混同行列 性能評価 α=0.1

α=0.1	Class0	Class1	Class2	Class3	Class4
正解率	0.9547	0.9342	0.9136	0.9259	0.9506
検出率	0.9167	0.8776	0.8140	0.8039	0.8125
精度	0.8049	0.8113	0.7292	0.8367	1.0000
真陰性	0.9614	0.9485	0.9350	0.9583	1.0000

表 2　混同行列 性能評価 α=0.5

α=0.5	Class0	Class1	Class2	Class3	Class4
正解率	0.9588	0.9300	0.8848	0.9300	0.9342
検出率	0.9429	0.8462	0.7174	0.8810	0.7647
精度	0.8049	0.8113	0.7292	0.8367	1.0000
真陰性	0.9614	0.9485	0.9350	0.9583	1.0000

表 3　混同行列 性能評価 α=1

α=1	Class0	Class1	Class2	Class3	Class4
正解率	0.9630	0.9218	0.8724	0.9300	0.9177
検出率	0.9706	0.8148	0.6889	0.9211	0.7222
精度	0.8049	0.8302	0.6458	0.7143	1.0000
真陰性	0.9617	0.9524	0.9141	0.9317	1.0000

4. 考察

前章の結果より，表 1 の結果から α=0.1 が最も高い平均正解率であり，最も低い α=1 と比較してもその差は 3%程度ということがわかる．このことより，α が小さいほど本実験での分類器は精度が高い傾向にあるということがわかる．α はテストデータに未知語が存在している場合に影響があるため，まだ今回のデータには未知語が多い．または未知語に対しての対応力がまだ低いということがわかる．これに関しては収集されるデータ量が増加した場合に今回より結果がよくなると考えられる．これは今後の課題になると考えられる．

5. まとめ

本研究では，背景として近年多様な利用範囲とニーズがあるビッグデータの利用方法の 1 つである小売分野のレコメンドに着目した．その中でもユーザ重視のマーケティング方法であるペルソナ/シナリオ法に注目し，その多様な項目の作成を行う事を目的とした．本研究ではその中でも趣味の推定を行う事を目的とした．

本研究の実現方法には，ビッグデータの Twitter を対象に投稿情報を利用して趣味の推定を行う分類器を教師あり機械学習の 1 種類であるナイーブベイズ法を用いて作成を行う事により実現させた．

実現させた分類器の性能評価を行うことによってその有効性を明らかにすることを目的として実験に K 分割交差検証と混同行列の 2 種類の手法を利用した．

結果として，表 3.2 より全体の分類は最も高い平均として約 93%を示したが，検出率と精度を確認すると正しい分類以外にも誤分類も同時に行っていることも考えられた．全体として分類性能は約 80%程度とやや低い性能であるといえる．この要因として，クラス毎のデータ数によってはデータ数が少ないため分類に必要な語彙を発見できていなかったのではないかと考えられる．

以上の事より今後の課題として，各クラスのデータの増量が必要であることが挙げられる．また，今回は 5 クラスの分類と実際の趣味で考えた場合ではかなり狭い範囲にのみ対応する形となっているためより多くのクラスを作成する必要がある．これによって増加したデータでの分類器の性能評価も重要であると考えられる．

また，本研究での分類器作成によって趣味の分類を行う事は可能であることが証明されたため，他のペルソナ項目にも応用が可能であるかの検証も今後の課題として挙げられる．

今後の展望としては，前述した課題の解決とその他の機械学習法との比較や他の項目での適用方法を考案していく．

謝辞

本研究を行うにあたって，ご指導をいただきました指導教員の佐野睦夫教授に深謝の意を表する．また，本研究を行うまでの中で知識や示唆をいただいた自然言語処理研究室の平博順准教授に深謝の意を表する．その他にも，日常の議論を通じて多くの知識や示唆をいただきましたインタラクションデザイン研究室の大井翔氏をはじめとした同研究室の皆様に深謝の意を表する．

参考文献

[1] “統計局ホームページ/平成 28 年社会生活基本調査の結果,” [オンライン]. Available: http://www.stat.go.jp/data/shakai/2016/kekka.htm. [アクセス日: 20 12 2017].

[2] Twitter, Inc., “Twitter Developer Platform - Twitter Developers,” Twitter, Inc., [オンライン]. Available: https://developer.twitter.com/. [アクセス日: 25 01 2018].

[3] Python ソフトウェア財団, “Welcome to Python.org,” [オンライン]. Available: https://www.python.org/. [アクセス日: 31 01 2018].

[4] Twitter Inc., “Twitter,” Twitter, Inc., 15 07 2006. [オンライン]. Available: https://twitter.com/. [アクセス日: 25 01 2018].

[5] T. Uchida, “Welcome to janome’s documentation! (Japanese) − Janome v0.3 documentation (ja),” 2015. [オンライン]. Available: http://mocobeta.github.io/janome/. [アクセス日: 31 01 2018].

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

研究内容の時間変化と所属情報を考慮した類似研究者検索に関する検討

A Note on Similar Researcher Retrieval Considering Temporal Changes of Research Content and Affiliations

西澤 浩之 *1 Hiroyuki Nishizawa　桂井 麻里衣 *2 Marie Katsurai　大向 一輝 *3 Ikki Ohmukai　武田 英明 *3 Hideaki Takeda

*1同志社大学大学院理工学研究科 Graduate School of Science and Engineering, Doshisha University

*2同志社大学理工学部 Faculty of Science and Engineering, Doshisha University

*3国立情報学研究所 National Institute of Informatics

As a research theme becomes more complicated, the range of knowledge necessary for research spreads wider. The collaborative research by multiple researchers is known to be effective for doing complicated research. Existing methods on research collaborator recommendation calculate the relevance between researchers using research content similarity and co-authorship relationships. This paper focuses on introducing new features, i.e., *temporal changes of research content* and *affiliations*, into researcher profiling for collaborator recommendation. In the proposed method, we use the titles and abstracts of academic papers as textual information for modeling research topics. Specifically, calculating a topic vector of a researcher's publications in each year, we represent the researcher's interests using a series of topic vectors. We also obtain each researcher's affiliation information from the Database of Grants-in-Aid for Scientific Researcher, named KAKEN. Based on the resulting profiles consisting of temporal changes of research content and affiliations, we present a novel similarity measure of researchers. In experiments, we present topic extraction results from a subset of CiNii Articles and show an interface example of similar researcher retrieval based on the proposed method.

1. はじめに

研究課題が複雑化するにつれて，研究に必要な知識範囲は広がる．複雑化した研究課題を円滑に進めるには，複数の研究者が共同で研究に取り組むことが有効といわれている．共同研究と生産性の関係や共同研究の推進の効果については多数の議論がある．例として，文献 [Abramo 09] は，共同研究と生産性の関係を実証するために，異分野共同研究や外部との共同研究，産学連携とその研究成果の相関を調べた．文献 [Wang 17] は，共同研究者の推薦が学術研究の発展を促進することを証明した．以上の共同研究の重要性から，効率的に共同研究者候補を検索する手法が種々提案されている．我々の過去の研究では，研究者の論文集合から多次元トピックベクトルを抽出し，トピックベクトル間の類似度に基づき研究者間の類似度を算出した [Araki 17]．この手法では，学術論文の出版日時を考慮しておらず，各研究者に単一のトピックベクトルのみを割り当てる．一方，学術情報に限定しない情報推薦の研究では，ユーザのプロファイルを構築する際に，ユーザ嗜好の時間変化を考慮すべきといわれている [Liang 12]．加えて，研究者間のコミュニケーションを促進するためには，研究内容の類似度の他にも共通点を発見しうる新たな特徴を追加すべきと考えられる．

そこで本研究では，研究内容の時間変化と所属情報を考慮した類似研究者検索手法を提案する．提案手法は，(i) 学術論文集合と所属情報に基づく研究者プロファイリングと (ii) 研究者プロファイルに基づく類似研究者検索から構成される．はじめに，年ごとに分割した学術論文集合のタイトル・概要から各研究者の研究内容特徴ベクトルを算出するとともに，研究者の所属とその位置情報をウェブから取得する（2 章）．構築したプロファイルに基づき，研究者間類似度を定義する（3 章）．これにより，研究内容の変遷が類似しており，かつ地理的に近い研究者が検索可能となる．本文の最後には，CiNii Articles*1 の論文を用いた研究トピック抽出結果と，提案手法に基づく研究者検索インタフェースの構築例を示す．

連絡先: 西澤浩之，同志社大学大学院理工学研究科，〒 610-0394 京田辺市多々羅都谷 1-3，nishizawa@mm.doshisha.ac.jp

2. 学術論文集合と所属情報に基づく研究者プロファイリング

本章では，学術論文集合と所属情報に基づく研究者プロファイリング手法を説明する．以降，学術論文集合から年ごとに研究内容を抽出する方法と所属の位置情報を算出する方法について説明する．

2.1 学術論文からの研究内容抽出

本節では，学術論文集合から年ごとに研究内容を抽出する方法を説明する．学術論文の本文から研究内容に直接関連する文章のみを抽出することは困難であるため，各論文のタイトルと要約を一つの文書とみなし，研究内容の分析に用いる．手法の概要を図 1 に示す．具体的には，各研究者の論文を出版年ごとに分割し，以下の研究内容抽出手順を適用する．

1. 論文テキストから名詞のみを抽出し，Bag-of-Words (BoW) 表現に変換する．
2. BoW 表現に対して Latent Dirichlet Allocation (LDA)[Blei 03] を適用し，トピックベクトルを算出する．
3. 閾値以下のベクトル要素を 0 にし，L2 正規化を行う．

以降，各手順の詳細を説明する．

2.1.1 テキストからの BoW 算出

まず，LDA の学習と推定の精度向上を目的とし，以下のように学術論文のタイトル・要約テキストを前処理する．

*1 https://ci.nii.ac.jp/

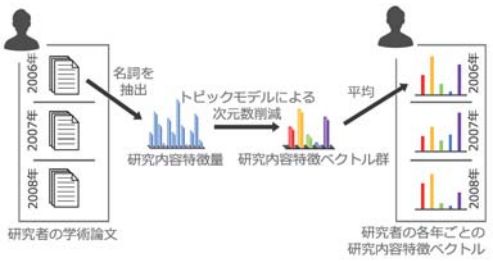


図 1: 研究者の研究内容特徴ベクトル群算出の概要図.

- 記号の除去.
- 数字の除去.
- 全角空白の除去.
- 括弧の除去.
- URL 文字列の除去.
- 半角カタカナを全角カタカナに変換.
- 全角英字を半角英字に変換.
- 英大文字を英小文字に変換.

次に，日本語形態素解析エンジン MeCab[*2] を用いてテキストから名詞のみを抽出する．得られるボキャブラリには，「研究」「提案」「一般」のように研究者の研究内容を特徴付ける効果が低いと考えられる単語が混在する．そこで，出現頻度が最上位または最下位となる語をストップワードとして除去する．以上により得られた単語リストを用いて，各論文に対し BoW ベクトルを算出する．

2.1.2 確率的トピックモデルによる次元削減

BoW の次元数はデータセット中に登場するユニークな単語数に相当する．語の関連性を考慮し，かつ類似度算出時の計算量を削減するために，LDA を適用して次元数を削減する．LDA は確率的トピックモデルの一つであり，文書は複数のトピックから構成されていると仮定する．文書集合をトピックに分解し，各文書をトピックの組み合わせで表す．一連の文書に用いる場合，トピックは集合のテーマとして解釈可能であり，文書のトピック分布は各文書がどのテーマに関するものかを示す．これにより，BoW の次元をトピック数 K まで削減する．

具体的には，まずデータセットの全論文を用いて LDA モデルを学習し，トピック分布を算出する．次に，研究者 a の論文集合を年 y ごとに分割する．分割された論文集合 $S_{a,y}$ ごとに，学習したモデルで各論文 $d \in S_{a,y}$ のトピックベクトル $\boldsymbol{\theta}_d$ を算出する．最後に，次式のように集合 $S_{a,y}$ のトピックベクトルの平均を算出する．

$$\boldsymbol{v}_{a,y} = \frac{1}{|S_{a,y}|} \sum_{d \in S_{a,y}} \boldsymbol{\theta}_d \tag{1}$$

得られた $\boldsymbol{v}_{a,y}$ を研究者 a の年 y における研究内容特徴ベクトルとみなす．

2.1.3 研究者の研究内容の抽出

算出した研究内容特徴ベクトル $\boldsymbol{v}_{a,y}$ の要素に着目したとき，あるトピックに非常に小さな値が割り当てられている可能性がある．そのようなトピックは研究者の研究内容を表現する際に重要ではないと考えられる．そこで，ベクトル $\boldsymbol{v}_{a,y}$ の k 番目のトピックに対する要素 $v_{a,y,k}$ に対し，次式の閾値処理を適用する．

$$v_{a,y,k} = \begin{cases} v_{a,y,k} & (v_{a,y,k} \geq Threshold) \\ 0 & (v_{a,y,k} < Threshold) \end{cases} \tag{2}$$

これにより，研究者の主要な研究内容を表すトピックのみが残される．最後に，ベクトル $\boldsymbol{v}_{a,y}$ を L2 正規化する．

2.2 研究者の所属情報の取得

本節では，各研究者の所属の位置情報を算出する方法を説明する．はじめに，KAKEN[*3] の研究者詳細画面から各研究者の所属文字列を抽出する．研究者詳細画面は一意の研究者番号を含む URL で管理されているため直接アクセス可能であり，研究者の氏名，研究者番号，所属，研究分野，キーワード，研究課題が記載されている．KAKEN を用いることで，各大学・研究所の研究者ページへ個々にアクセスすることなく研究者のメタデータが得られる．得られた所属情報から Google Maps API[*4] を用いて位置情報（緯度・経度）を得る．Google Maps API は地図の描画や緯度経度の検索，ルート検索などの機能があり，その一部を研究者検索インタフェースの構築に利用する．

3. 研究者プロファイルに基づく類似研究者検索

本章では，前章で構築した研究者プロファイルに基づく類似研究者検索手法を提案する．まず，研究者 a, b に対し，研究内容の時間変化の類似度 $ResearchContentSim(a, b)$ を次式により算出する．

$$ResearchContentSim\ (a, b) = \frac{1}{n} \sum_{y=y_0}^{y_n} CosSim\ (\boldsymbol{v}_{a,y}, \boldsymbol{v}_{b,y}) \times \log\ (y - y_0 + 2) \tag{3}$$

$$CosSim\ (\boldsymbol{v}_a, \boldsymbol{v}_b) = \frac{\sum_{t=1}^{k} v_{a,t} \times v_{b,t}}{\sqrt{\sum_{t=1}^{k} v_{a,t}^2} \times \sqrt{\sum_{t=1}^{k} v_{b,t}^2}} \tag{4}$$

上式において，n は研究内容特徴ベクトル群の分割年の総数を示す．log 関数を適用することで，最新の研究ほど類似度に寄与する．次に，位置情報による研究者間の類似度 $LocateSim(a, b)$ として，Google Maps API を用いて取得された緯度経度から計算された直線距離を正規化し類似度へ変換する．

以上の二つの類似度を組み合わせ，研究者 a，b 間の最終的な類似度を次式で算出する．

$$Sim\ (a, b) = \phi \times ResearchContentSim\ (a, b) + (1 - \phi) \times LocateSim\ (a, b) \tag{5}$$

上式において，ϕ は二つの類似度のバランスを調整するためのパラメータである．与えられた研究者 a に対し，式 (5) の類似度が上位となる研究者 b を提示することで，研究者 a の類似研究者を容易に把握できる．

4. 実験

4.1 データセット

CiNii Articles からの論文収集にあたり，2003 年から 2010 年の間に電子情報通信学会 [*5] で一度でも発表したことがあり，

*2 http://taku910.github.io/mecab/

*3 https://kaken.nii.ac.jp/

*4 https://developers.google.com/maps/

*5 http://www.ieice.org/jpn/

表 1: 実験用のデータセットの詳細.

研究者数	4,680 名
学術論文数	38,098 本
論文出版年	2003-2010

表 2: LDA によって推定された研究トピックの例．トピックの解釈を見出しで示す.

トピック 24 テキストマイニング		トピック 50 検索エンジン		トピック 96 画像処理	
単語	確率	単語	確率	単語	確率
単語	0.092	検索	0.254	画像	0.428
クラスタリング	0.045	分類	0.118	処理	0.037
辞書	0.041	類似	0.095	復元	0.026
出現	0.039	属性	0.040	エッジ	0.025
頻度	0.037	クエリ	0.030	局所	0.022
語彙	0.025	精度	0.020	画素	0.019
日本語	0.020	エンジン	0.018	テクスチャ	0.019
漢字	0.018	キーワード	0.018	カラー	0.014
対象	0.015	方法	0.014	解像度	0.014
共起	0.014	向上	0.012	特徴	0.014

トピック 112 推薦システム		トピック 133 VR・AR・MR		トピック 150 人工知能	
単語	確率	単語	確率	単語	確率
ユーザ	0.435	空間	0.395	学習	0.310
推薦	0.040	仮想	0.109	ルール	0.042
表情	0.038	現実	0.058	識別	0.039
興味	0.023	力	0.036	サンプル	0.028
嗜好	0.021	再現	0.034	問題	0.023
提示	0.021	世界	0.022	推論	0.020
サイト	0.017	覚	0.019	獲得	0.019
顔	0.016	環境	0.018	ニューラルネットワーク	0.019
意図	0.012	物理	0.017	入力	0.014
コンテキスト	0.010	複合	0.012	ファジィ	0.014

かつ KAKEN の ID をもつ研究者リストを構築した．実験用のデータセットの詳細を表 1 に示す．なお，所属の位置情報を取得する際，KAKEN のメタデータに明らかに誤った記載があったため，Google Maps API で位置情報が取得できる形式となるよう文字列を手動で修正した.

4.2 研究者の研究内容変化の特徴抽出

研究者の研究内容特徴ベクトルを抽出するために，全ての論文テキストに MeCab[*6] を適用して単語リストを生成した．MeCab の辞書には新語・固有名詞を考慮可能な mecab-ipadic-NEologd[*7] を使用した．その後ストップワードとしてデータセットの 20 %（7619 本）以上に出現する単語および 9 本以下にのみ出現する単語を除去した．LDA のトピック数は $K = 150$，ハイパーパラメータは $\alpha = \frac{50}{K}$，$\beta = 0.01$ と設定した．LDA によって推定された研究トピックとその単語分布の例を表 2 に示す．LDA により，6907 次元であった BoW が 150 次元まで削減され，低次元の研究内容特徴ベクトルが得られた．また研究内容特徴ベクトルから研究内容を抽出するにあたり，弱いトピックの判断基準として 0.1 を閾値とした.

4.3 類似研究者検索インタフェースの構築

提案手法に基づき，対象とする研究者の類似研究者を検索できる Web アプリケーションを構築した．本実験では，研究内容の類似性を重視して $\phi = 0.65$ とした．類似研究者検索インタフェースの画面表示例を図 2 に示す．対象研究者の氏名，研究者 ID を画面上部，所属と過去の論文リストを左部に示す．また，所属の位置情報は右部に地図表示する．画面右下には，提案手法において類似度が上位となった研究者を表示する．クリックされた研究者が新たに検索クエリとなるため，類似研究者を次々と検索可能となる.

*6 http://taku910.github.io/mecab/

*7 https://github.com/neologd/mecab-ipadic-neologd

図 2: 類似研究者検索インタフェースの画面例.

5. まとめと今後の展望

本研究では，共同研究者推薦の実現に向け，新たに研究内容の時間変化と所属情報を考慮した研究者のプロファイリングとそれに基づく類似度算出方法を提案した．実験では，提案手法の応用例として，類似研究者検索インタフェースを構築した．今後の課題として，研究者の所属の時間変化など，より多くの特徴をプロファイルに加えることが挙げられる．本稿では研究者の所属情報から Google Maps API を用いて緯度・経度を取得したが，同じ大学でもキャンパスごとに位置情報が異なる問題があった．KAKEN から取得した情報ではキャンパス情報が収録されていないため，研究内容と所属情報に基づく在籍キャンパスの推定も一つの課題となりうる．今後の展望として，複数分野の研究者に提案手法を適用することによる異分野共同研究者の推薦や，ユーザがより使いやすい研究者検索インタフェースの構築，各地方や学研都市の研究内容可視化インタフェースの構築を検討している.

参考文献

[Abramo 09] Abramo, G., D'Angelo, C. A., and Costa, Di F.: Research collaboration and productivity: is there correlation?, *Higher Education*, pp. 155–171 (2009)

[Araki 17] Araki, M., Katsurai, M., Ohmukai, I., and Takeda, H.: Interdisciplinary Collaborator Recommendation Based on Research Content Similarity, *IEICE Trans. Information and Systems*, pp. 785–792 (2017)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Leraning Research*, pp. 993–1022 (2003)

[Liang 12] Liang, H., Xu, Y., Tjondronegoro, D., and Christen, P.: "Time-aware Topic Recommendation Based on Micro-blogs, in *Proc. the 21st ACM Int. Conf. Information and Knowledge Management*, p. 16571661 (2012)

[Wang 17] Wang, W., Cui, Z., Gao, T., Yu, S., Kong, X., and Xia, F.: Is Scientific Collaboration Sustainability Predictable?, in *Proc. the 26th Int. Conf. World Wide Web Companion*, pp. 853–854 (2017)

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

解答解説機能付き微積分ソルバの開発

Development of *Calculus Solver* software which provides solving hints and answer steps

大塚　基広 *1　　平　博順 *1　　真貝　寿明 *1
Motohiro Otsuka　　Hirotoshi Taira　　Hisaaki Shinkai

大阪工業大学大学院 情報科学研究科
Faculty of Information Science & Technology, Osaka Institute of Technology

We develop a python-based software which provides hints and answers for freshmen-level calculus interactively. Our solver covers multi-variable differentiations, Taylor expansions, graph drawings, and single-variable definite/indefinite integrations. A user can input a problem using GUI interface, and the output comes as a LaTeX-typed pdf. We find that differentiation processes can be coded with four tasks, while integration requires more. In order to prepare "readable" answers, we have to adjust output expressions. Although this software cannot handle some problems, the provided answers pass tests as the grade B to D, so that it helps freshmen for their self-education.

1. はじめに

これまでまでに Mathematica，Maple，Matlab といった優れた数式処理ソフトウェアが開発され，利用されてきている．数式処理ソフトウェアは，人間が手計算で行なっていたものを自動で計算してくれることから，様々な研究開発の分野で，その研究開発スピードを高めている．また，多くの数式処理ソフトウェアでは，様々な可視化機能が付属しており，ユーザに処理結果を分かりやすく提示することができる．

このような可視化機能を利用して，大学における数学教育で数式処理ソフトウェアを導入することで，微積分や線形代数における基本的な概念を学生に理解させやすくする試みはこれまでいくつかの報告がある．例えば木村は，数式処理ソフト Sage を学部生の線形代数の授業内で利用し，ベクトルの線形変換の様子などを可視化することで，数式だけだとイメージしにくい線形変換などの線形代数の概念を学生に理解させる上で一定の効果をあげている [1].

また，中島らは，数式処理ソフト Maple を微積分の講義で利用し，2 変数関数のグラフ，接平面，グラフの極値と鞍点，2 重積分などを可視化することで学生の数式や数式処理について具体的なイメージを持たせる試みを行なっている [2].

以上は，概念の可視化を通じて，学生の理解を助ける取り組みである．他方，微積分や線形代数の計算自体に習熟させるために，効率的な e-learning 教材の利用も提案されている．例えば中村は，STACK と呼ばれる数学解答評価システムを開発し，問題の数式の係数などをランダムな値として類似問題を自動生成することで，教員側が大きな労力をかけなくても，学生が繰り返し計算問題に挑戦できる仕組みを作っている [3]. ただし，この取り組みでは扱える問題が係数が異なる程度の類題にとどまっている．

一方，先に述べた数式処理ソフトでは，入力された関数や関数に対する微積分などの処理や可視化は簡単に行なえるものの，途中の数式処理については提示されず，またその解き方について解説してくれるわけでもないので，入力されたものに対する結果しかユーザは得ることが出来ない．

また,「ロボットは東大に入れるか」プロジェクト（東ロボ）では，東大に合格できるような試験問題ソルバが開発されており，センター試験の数学の模擬試験では偏差値が 70 を越えるようなソルバが開発されている．しかしながら，これらのソルバの出力する解答は数学的に正しい解答であっても，普通の受験生が書くような答案とは大きくかけ離れている．問題文に対して形態素解析，係り受け解析，照応解析などを行なったあと，量化記号を含む論理式を立式し，限量記号消去法（Quantifier Elimination 法; QE 法）を用いて計算処理が行なわれている．この結果を用いて解答を生成しているため，受験生が授業や参考書などで学ぶ処理方法とはかけ離れており，学生にとって参考になる解答とはいえない．

受験生が解くような解き方で入力された数学の問題を解くソフトウェアとして，フリーウェア・フリーのアプリとして，Mathway や Microsoft 社の Math 3.0，Photomath などがある．これらは，中学校レベルの数学の問題では途中の解き方を出力できるが，大学レベルの微積分の問題では，基本公式に当てはまる問題程度しか解答を出力できず，置換積分のような問題については解き方を出力することはできない．

そこで，本研究では大学の数学教育において，学生が自ら微積分などの計算問題を入力し，置換積分や部分積分など高度な解き方を含む問題についても，自動的にその解答と解説を生成することで，学生が多様な計算について興味を持ち，自分の解き方も確認できるような数式処理ソフトウェアの開発を試みた．

開発したシステムについて，大学の微積分の講義における期末試験問題の過去問についてどの程度正しく，解答と解説が自動生成できるかについて評価したところ，3 年分の試験で試験の解答として 70%が正しいものが出力されることが確認できた．

2. 実装へのアプローチ

2.1 微分に関して

例えば関数 $f(x) = x\cos(2x)$ を微分するとき，人間の思考では次のような手順で考える．

1. 関数 $f(x)$ は 2 つ関数の積 $f(x) = g(x)\,h(x)$ からできている．

連絡先:　大塚　基広，大阪工業大学情報科学部,m18.aiandnip.m.o@gmail.com

2. 積の微分公式を用いることになる．

$$f'(x) = g'(x)\,h(x) + g(x)\,h'(x)$$

3. $g(x) = x$, $h(x) = \cos(2x)$ と置く．

4. $h(x)$ は合成関数である．$h_1(x) = 2x$ とおいて，$h(x) = \cos(h_1(x))$ となる．

5. 合成関数の微分公式を用いることになる．

$$h'(x) = \frac{\partial h(h_1(x))}{\partial (h_1(x))} \cdot \frac{\partial h_1(x)}{\partial x}$$

6. $h'(x) = 2\sin(2x)$ である．したがって，与式の微分は，

$$f'(x) = \cos(2x) - 2x\sin(2x).$$

以上の過程を，計算をするときに必要な部分と，解答を作るときに必要な部分をまとめて簡略化すると，次の 4 つのタスクにまとめることができる．

(A) 合成関数としての分離，ヒエラルキーの把握

(B) 定数・変数の把握

(C) 式全体の積の微分の演算結果の保存

(D) 1 つの演算単位での微分結果の保存

これらに加え，関数をすべて積の関数としてみなして扱うことにすれば，微分演算の手順をまとめることができる．

それぞれ 4 つのタスクを，上記の 6 つの手順と照らし合わせてみると，次のようになる．

1 は，認識のレベルでタスクの（B）に相当する．2 は，関数の積と解釈できればいいので，タスク（B）に相当する。3 は，公式を使用し，新たに関数を定義しなおしているので．タスクの（A），(C）に相当する。4 は，同様にして，(A），(C）に相当する．5 は，定数と変数が把握できれば良いので，タスクの（B）である．6 はすべての結果を合わせるのですべてのタスクに相当する．

このようにして，合成関数と積関数が入り混じった計算でも，この 4 つの情報を保持できれば解答を付けることが可能になる．

2.2 積分について

積分も微分と同様に上記 4 つのタスクを使用すれば，ある程度計算はできる．しかし，積分演算では，部分積分・置換積分・漸化式を用いる解法・部分分数分解を必要とする積分などさまざまな技法がある．また，解析的には計算できない積分もある．したがって，それらを個別に実装する必要があった．

本研究で実装できなかったものに関しては，解答のヒントを示すようにしている．なお, 部分積分 (漸化式解放を除く) と，部分分数分解を使用する計算に関しては実装ができた．部分分数分解を必要とする積分は sympy の apart を使用した．

積の関数を積分するときには，部分積分

$$\int g'(x)\,h(x)\,dx = g(x)\,h(x) - \int g(x)\,h'(x)\,dx$$

を用いることになるが，どちらを先に積分して $g'(x)$ としておくのか，という点については，必ずしも一意ではない．場合によってはできないこともあるし，一方に三角関数があれば，部分積分を 2 回行う手法もある．

以下に擬似コードを示す．

擬似コード

```
str = ’xCos_{x}’ #input
list = str.split(’*’) #list=[x,Cos_{x}]
if(list[0] == 三角関数):
    if(list[1] == 三角関数):
        print 漸化式解法（未実装）
    else:
        list[0] を h(x), list[1] を g(x) として部分積分
else:
    list[0] を g(x), list[1] を h(x) として部分積分
```

3. ソフトウェアの概要

GUI での操作ができるようにした．起動すると，まず，計算するカテゴリー（微分・積分・Taylor 展開）を選択する画面になる (図 1).

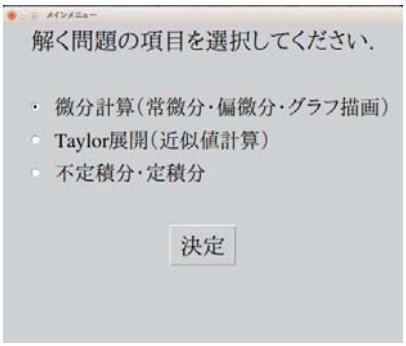


図 1: GUI のメイン画面

決定ボタンを押すと，式や条件などの入力画面になる．入力の方法は，入力例ボタンをクリックすると表示される．ヒントや解説を表示するかどうか，グラフを描く場合の範囲の指定，級数展開の場合は展開次数などの選択項目を指定すると，計算が実行される．処理のフローを図 2 に示す．

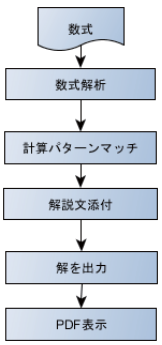


図 2: 処理のフロー

3.1 式の入力方法に関する工夫

ここで入力インタフェイスについて簡単に述べる．入力の仕様は以下のようにした．

- 四則演算子
  乗算を表す記号（∗）は不要
- べき関数
  ^ を使用し，べきの部分は { } で囲む．
  例 `x^{2}` $\rightarrow x^2$
- 三角関数
  最初の文字を大文字にして，_ (下線) を付与した関数名とし，引数部分を { } で囲む．
  例 `Cos_{x}` $\rightarrow \cos(x)$
- 対数関数
  最初の文字を大文字にして，_ (下線) を付与した関数名とし，底と真数部を { } で囲む．底を省略すると，底を e として扱わる．
  例 `Log_{x}` $\rightarrow \log(x)$

このインタフェイスにした理由は、手書きに近く，利用者の多い Tex や Mathematica に似た入力仕様にしたかったためである．

例えば，$x \cdot \cos(2x)$ を微分するときには「`xCos_{ 2x }`」と入力し，微分する変数と，微分階数を指定することになる．

3.2 アルゴリズム

処理のアルゴリズムを図 3 で示す．上記の入力例は

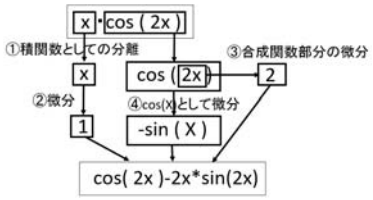


図 3: 処理のアルゴリズム

```
List [ [Cos,_,{,[2x],},],]
```

のように分解される．List の中身を探索することによって，[Cos,_,{,[2x],},] が出現する．この時点で入力された関数が，三角関数 cos と，その引数が $2x$ であることが分かり，合成関数であることがわかる．そこで，cos の処理を行うための関数に処理を渡す．その後，合成関数の処理をし計算結果を返す．結果を格納するリストを result と定義すると，

```
result[2,*,Sin,_,{,2x,}]
```

となる．Taylor 展開や偏微分の計算は，微分と同じものを使用し，その関数を繰り返し適応する計算を行うことになる．

3.3 出力に関する工夫

一般的に，PC で処理された数式は，∗ やがそのまま使用されていることが多い．また，整数で書くべきところが小数点付で表示することもある．出力される答案では，これらの記号を使わないように工夫した．

Taylor 展開を表示する場合，例えば，

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$$

のような表示をしたいとき，展開係数は階乗記号で書かれる方が好ましい．通常の処理をしてしまうと，分母の階乗記号が計算されてしまう（例えば，$4! = 24$）．しかし，本プログラムでは，$\frac{1}{4!}$ とった形で出力されるように，n 階の簡略化された微分結果と，階乗の文字列を繋げるようにし，階乗記号が計算されないように工夫した．実際の出力を示す．

[問題]
微分する関数を $f(x)$ と定義する．

$$f(x) = x\cos(2x) \tag{1}$$

[基礎知識] . 積の微分公式

$$f(x) = g(x)h(x) \rightarrow f'(x) = g'(x)h(x) + g(x)h'(x) \tag{2}$$

商の微分公式

$$f(x) = \frac{g(x)}{h(x)} \rightarrow f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2} \tag{3}$$

合成関数の微分公式

$$f(x) = g(h(x)) \rightarrow f'(x) = \frac{\partial h(x)}{\partial x}\frac{\partial g(h(x))}{\partial (h(x))} \tag{4}$$

[解答]
関数 $f = \cos(2 \cdot x)$ は $f(x) = \cos(g(x))$ の形を取っている．すなわち合成関数の微分を行う必要がある．
この関数の場合 $g(x) = 2 \cdot x$ である．
$g(x)$ を合成関数として考えると $g'(x) = 2$ となる．
したがって関数関数 $f = \cos(2 \cdot x)$ の微分結果は $f' = -2 \cdot \sin(2 \cdot x)$ となる
以上より $f(x)$ を x で微分させた結果は次の通りになる．

$$f(x)_x = -2x\sin(2x) + \cos(2x) \tag{5}$$

3.4 グラフ出力に関する工夫

グラフは，python の matplotlib モジュールを用いて表示した．微分計算の答案を作成するにあたっては，微分して極大・極小を与える場所を求め，増減表を作成する必要がある．これらの各ステップを表示するようにプログラムする必要がある．

また，グラフの描画域も整数値で指定できるようにして，ユーザの答案作成の視点に近づけた．また，解答にグラフが記述される前に，どのようなグラフになるか見られるプレビュー機能も作成した．この機能により，ユーザはグラフを何度も調整することができる．

4. ソルバの性能

4.1 対象のテストの結果

このソルバを使用し，本学の微積分学 I の期末試験を解かせた結果を示す．なお，過去問は Web[*1] 後悔されてている．

表 1: 過去問の大問ごとの採点結果

年度	大問 1 微分	大問 2 積分	大問 3 Taylor 展開	大問 4 偏微分	合計
配点	30 点	30 点	20 点	20 点	100 点
2017 年度	25 点	25 点	17 点	14 点	81 点
2016 年度	20 点	13 点	11 点	14 点	59 点
2015 年度	25 点	19 点	12 点	14 点	70 点

なお，この試験では 100 点スコアがつくとともに，A〜D（合格），F（不合格）の 5 段階評価がおこわなれる．5 段階評価の基準は以下の通りである．

D 判定　大問 1（微分とグラフ，ライプニッツ）と大問 2（積分と，求積問題）の得点の和が一定以上

C 判定　D 判定の判定と同条件（ボーダーラインが上がる）

∗1 `http://www.oit.ac.jp/is/~shinkai/lecture/`

B 判定　C 判定を取得できる者のうち，大問 3（Taylor 展開）の得点が一定以上

A 判定　B 判定を取得できる者のうち，大問 4（偏微分）の得点が一定以上

この基準を元に，システムの出力について，微積分の講義の担当教員に評価してもらった結果，2017 年度は B，2016 年度は D，2015 年度は A という判定であった．

このように，合格はできるものの一定以上の点数が取れないのは，文章題や積分計算の一部がまだ解けないことが原因である．なかには「$(x^2+x)\sin x$ の n 階微分を求めよ（ライプニッツの公式を用いる）」のようにまだプログラミングが対応できていない問題もある．置換積分でどの部分を置換したらよいのかが正しく判定できない場合や，

- 「○○の回転体の表面積を求めよ」
- 「$(1-x)^n$ の級数展開を用いて，$\sqrt[3]{0.9}$ を小数第 3 位まで求めよ」

のような，計算だけでは求まらない問題があり，今後の課題である．

4.2　汎用的な問題を解いた結果

実際に，どの程度の範囲であるならば正しく解答が得られるのかを評価した．講義で使用されている教科書 [8] をであげられている項目の内，実装が完了した計算を表 2 に挙げる．

表 2: 実装できた計算ツール

微分	べき関数，指数関数，対数関数，三角関数，商関数，積関数，Taylor 展開，偏微分
積分	一変数のべきの関数，指数関数，対数関数，周期関数を 1 つだけ含む積の関数，特殊な置換積分，分母の微分が分子になる商関数，一部の部分分数分解可能な分数，定積分（一部除く）

現時点で対応できていない問題としては，以下のようなものがある．

1. 二重指数を含む微積分　例　$\dfrac{d}{dx}x^x$
   これは対数微分法で解けるものであるが，対数を取って微分をした後，再び対数を取って変数を元にもどす処理が複雑で実装が完了していない．
2. 複数回の部分積分が必要な積分
   例　$\int e^x \cdot \sin mx\, dx$
   部分積分自体は実装できているので，計算をしてくれるが無限ループに入る．そのループを止める基準である，「与式」と「部分積分しているうちに出てくる関数」が一致しているという判定が難しい．

4.3　解けない問題への対応

積分の問題は，解析的にできないものも多い．たとえば

$$\int \frac{1}{1-x^2}dx$$

は置換積分で解けるが，

$$\int \frac{1}{\sqrt{(1-x^2)(1-k^2x^2)}}dx$$

は，楕円積分の知識がないと解けない．そして，一般には解析で解けないものは級数解法か数値積分を行うことになる．

学生にとってのヒントとしては，できる限り，解ける道筋を示すべきだが，ソルバとしては『お手上げ』状態になることも避けられない．そのようなときは，「このソルバでは解けません．○○か○○の方法で解けるかもしれません.」あるいは「入力された関数は解析的に積分できない可能性があります.」などと表示させることにした．Mathematica などでも，解析的にできない積分はそのまま返されるが，それよりは親切な表示といえる．

5.　まとめと展望

本研究では，解答解説付き微積分ソルバを開発した．本ソフトウェアの特徴は，学生の答案作成の見本を示すように，答案として途中経過を表示させることができることである．なるべく，人間として自然な答案を示すために，計算結果を表示するだけではなく，分数の形式で表示をする，階乗記号は計算させない，整数は整数の形を保ったまま表示するなどの工夫が必要であった．

実装する過程で，微分の計算は 4 つのタスクと，積の関数として微分することで大半の問題が解けることが分かった．しかし，二重指数の問題や，有理化が必要な計算に関しては，別ルーチンを組む必要があった．一方で，積分では，4 つのタスクだけでは解ける問題が少なく，部分積分，置換積分，漸化式解法，商の積分などを別ルーチンで組む必要があった．

量化記号が必要になるほどの論理が必要な問題には，まだ対応できていない．今後の課題である．

参考文献

[1] 木村　巌,「free な数式処理ソフト Sage の学部教育での活用事例」，数理解析研究所講究録第 1865 巻 p101-109（2013）

[2] 中島匠一，水谷明,「数式処理ソフト Maple による微積分の教育支援」計算機センター Vol.31(2010)

[3] 中村泰之,「数学 e ラーニング　数学解答評価システム STACK と Moodle による理工系教育」

[4] 伊藤巧，松崎拓也，佐藤理史，名古屋大学大学院工学研究科,「数学問題テキストにおけるさまざまな照応現象の解析」　人工知能学会全国大会（2017）

[5] 岩根秀直，穴井宏和「数式処理による入試問題への挑戦」，FUJITSU(2015)

[6] 岩根秀直，松崎拓也，穴井宏和，新井紀子，辻通研究所，国立情報学研究所，九州大学,「数式処理による入試問題の解法と言語処理との接合点における課題」　人工知能学会全国大会 (2013)

[7] 長岡亮介,「総合研究論理学で学ぶ数学-思考ツールとしてのロジック」,2017 年　旺文社,

[8] 真貝寿明,「徹底攻略　微分積分」　2013 年　共立出版

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

ニューラル機械翻訳モデルを用いたマルチソース文法誤り訂正

Multi-source neural grammatical error correction

曹 国林 *1　高村　大也 *2*3　奥村 学 *2*4
Cao Guolin　Hiroya Takamura　Manabu Okumura

*1東京工業大学工学院
School of Engineering, Tokyo Institute of Technology

*2東京工業大学科学技術創成研究院
Institute of Innovative Research, Tokyo Institute of Technology

*3産業技術総合研究所
AIST

*4理化学研究所 AIP
RIKEN AIP

This is the first attempt to use a multi-source encoder-decoder model for the grammatical error correction task (GEC). In addition to the possibly erroneous sentence written in a second language, our model uses the sentence written in the mother tongue of the learner. With our model, we achieved up to 1.13 GLEU score increases than the single source baseline model.

1. はじめに

文法誤り訂正は，特に第二言語学習者が書いた文における誤りを検出し正しい文に訂正する課題である [9]．高性能な文法誤り訂正システムは，語学学習者に正しい指導を与えることができ，語学学習の効率が上がることが期待される．

既存の文法誤り訂正課題では，規則による手法，統計的翻訳モデル (SMT) を使った手法と，ニューラル機械翻訳モデル (NMT) を使った手法が提案されている．提案手法に関わらず，従来の文法誤り訂正システムは，語学学習者の書いた文だけを入力とし，正しい文に訂正することを目指している．しかし，誤り文だけでは，正しい訂正ができない場合もある．例えば，表 1 の入力に示したように，英語を学習している日本人が “I am **leaving** in Tokyo.”と言う文法的誤りを含む文を書いたとする．このような文に対し，“I am **living** in Tokyo.”と “I am leaving Tokyo.”という少なくとも二つの可能な正解文が考えられる (表 1 の可能な出力)．どちらが正しいかは，学習者が伝えたい内容に依る．例えば，学習者が伝えようとする内容が母国語で “私は東京に住んでいる”(表 2 の入力) だと分かると，誤り文を正しく “I am living in Tokyo.”に訂正する有用な手がかりとなる．ここで，母国語文は正しく書けるとする仮定は妥当だと言えよう．

そこで本研究では，誤り文と共に同じ意味の母国語文を入力とし，文法誤り訂正を行うマルチソースの文法誤り訂正モデルを提案する．また，Allen の研究で [4] 同じデータセットに対し，NMT による手法が SMT による手法より高い GLEU スコアに達成した為，本研究では NMT による手法を選択している．その結果，提案モデルはベースラインとなる誤り文のみを入力とするニューラル文法誤り訂正モデルより，高い GLEU スコアを達成している．

表 1: 曖昧な誤り訂正の例

入力	可能な出力
I am leaving in Tokyo.	I am living in Tokyo.
	I am leaving Tokyo.

表 2: 母国語文の情報を用いた訂正の例

入力	可能な出力
I am leaving in Tokyo.	I am living in Tokyo.
私は東京に住んでいる.	

2. 関連研究

2.1 ニューラル文法誤り訂正

ニューラル機械翻訳モデルによる文法誤り訂正手法は，文法誤り訂正を，誤りのある文から正解文に「翻訳」する課題として捉えている．即ち，翻訳モデルを使い，入力した誤り文 $X = [x_1,..., x_n]$ に対し，正解文 $Y = [y_1,..., y_m]$ を直接生成することである．ここで x_n と y_m は，それぞれ誤り文と正解文の各単語に対応する．ニューラル文法誤り訂正モデルは Yuan[7] により初めて提案されている．

2.2 マルチソースニューラル翻訳モデル

Zoph はニューラル機械翻訳モデルをもとに，同じ意味を持つフランス語文とドイツ語文を当時に入力し，英語へ翻訳するマルチソースニューラル翻訳モデルを提案している [8]．マルチソースニューラル翻訳モデルは，encoder-decoder の encoder の部分を，二つに拡張した翻訳モデルである．

本研究では，Zoph の提案したモデルをもとに，母国語の情報を活用するマルチソース文法誤り訂正モデルを提案する．

2.3 和文英訳における文法誤り訂正

和文英訳における文法誤り訂正は，学習者が訳した学習文から，誤りを検出し正しい文に訂正する課題である [9]．その応用例として Duolingo*1 などのサービスがある．語学学習者が決められた文に対し翻訳を行い，システムは翻訳文の間違いを訂正し，語学のライティング支援を提供している．

このような研究は，母国語文からの翻訳文に対する訂正であり，母国語文の意味保存が訂正の軸である．本研究では言語学習者が書いた誤り文に対し，母国語を補助情報として訂正を行う．

連絡先: 曹国林，東京工業大学工学院情報通信系 奥村・高村研究室，横浜市緑区長津田町 4259 R2-728, 045-924-5295, kokurin@lr.pi.titech.ac.jp

*1 https://www.duolingo.com/

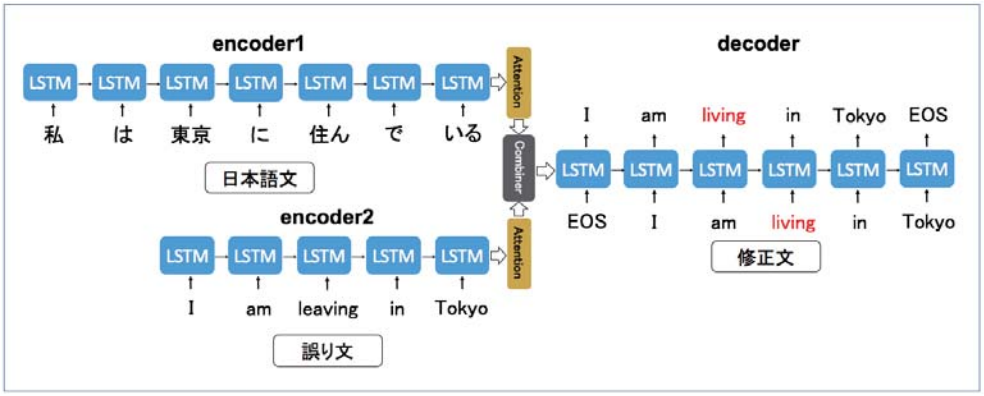


図 1: マルチソース文法誤り訂正モデル

3. ニューラル文法誤り訂正モデル

3.1 ベースラインモデル

ベースラインモデルである，注意機構付き両方向 LSTM 文法誤り訂正モデルについて説明する．

まず，encoder は英語文 $X = [x_1,..., x_n]$ の各要素 (単語) を単語ベクトルで表現した $[e_1^x,..., e_n^x]$ に対し，式 (1) と式 (2) により前向きの隠れ状態と逆向きの隠れ状態を計算し，式 (3) により連結させ入力単語 x_i の隠れ状態 h_i^x を計算する:

$$\overrightarrow{h_i^x} = f(\overrightarrow{h_{i-1}^x}, e_i^x), \quad (1)$$

$$\overleftarrow{h_i^x} = f(\overleftarrow{h_{i+1}^x}, e_i^x), \quad (2)$$

$$h_i^x = [\overrightarrow{h_i^x}; \overleftarrow{h_i^x}]. \quad (3)$$

次に，decoder では式 (4) により decoder の入力単語 y_j に対する隠れ状態ベクトルを計算する:

$$h_j^y = f(h_{j-1}^y, e_j^y). \quad (4)$$

ここで e_j^y は入力単語 y_j の単語ベクトルである．

本研究では Luong ら [1] が提案したモデルを適用し，decoder における注意機構の計算に用いる隠れ状態ベクトル $\tilde{h_j}$ を計算する．

注意機構とは，decoder の注意力の隠れ状態 $\tilde{h_j}$ を計算する際，encoder の隠れ状態を考慮する機構である．encoder の各隠れ状態に対し，式 (5) によって計算されるスコアの加重平均を計算し，式 (6) に示す文脈ベクトルを計算する:

$$a_i^j = \frac{exp(h_j^{yT} h_i^x)}{\sum_{i=1}^{n} exp(h_j^{yT} h_i^x)}, \quad (5)$$

$$c_j = [h_1^x ... h_n^x]^T [a_1^j ... a_n^j]. \quad (6)$$

注意機構によって計算された文脈ベクトルと，decoder の現在の隠れ状態 h_j^y を連結させたものに，重み W_a をかけ，式 (7) により decoder の最終隠れ状態 $\tilde{h_j}$ を計算する:

$$\tilde{h_j} = W_a[h_j^y; c_j]. \quad (7)$$

$\tilde{h_j}$ を利用し，式 (8) によって正解文中の各単語 y_j を予測させる:

$$p(y_j|y_{<j}, X) = softmax(\tilde{h_j}). \quad (8)$$

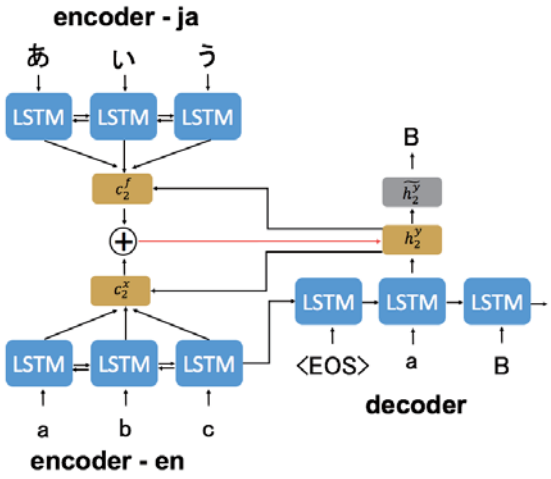


図 2: マルチ加算注意機構モデル．英語文「abc」と日本語文「あいう」を同時に入力した場合,「aBc」に訂正する decoder の出力一部である．ここで ja は日本語で en は英語を表す．

3.2 提案モデル

本研究では，Zoph のマルチソースニューラル翻訳モデルをもとに，二つの異なる注意機構をもったマルチソースニューラル文法誤り訂正モデルを提案する．2 つの提案モデルの概要は図 1 に示す．

3.2.1 マルチ加算注意機構モデル

図 2 にマルチ加算注意機構モデルの詳細を示している．マルチ加算注意機構モデルは，英語文 $X = [x_1,..., x_n]$ と，母国語文 $F = [f_1,..., f_l]$ を入力とする 2 つの encoder を持っている．それぞれの encoder は，入力文を隠れ状態 $h^x = [h_1^x,..., h_n^x]$ と $h^f = [h_1^f,..., h_l^f]$ に変換する．次に，英語文の最終の隠れ状態である h_n^x を decoder に渡し，decoder の隠れ状態の初期値とする．また，二つの encoder の隠れ状態に，式 (5)，式 (6) に示した注意機構を適用し，二つの文脈ベクトル c_j^x と c_j^f を計算する．

最後に式 (6) によって計算された英語文と母国語文の文脈ベクトル c_j^x と c_j^f に対し，式 (9) により decoder の最終隠れ状態の計算手法を提案する:

$$\tilde{h_j} = W_{sum}[(h_j^y); (c_j^x + c_j^f)]. \quad (9)$$

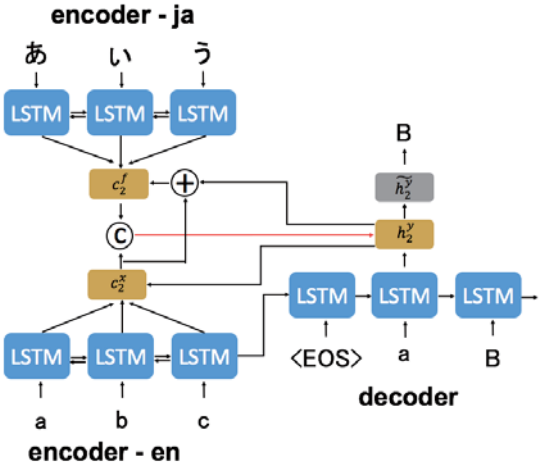


図 3: マルチ階層注意機構モデル．英語文「abc」と日本語文「あいう」を同時に入力した場合，「aBc」に訂正する decoder の出力一部である．ここで ja は日本語で en は英語を表す．ここで©は二つのベクトルを連結することを意味する．

式 (9) によって計算された $\tilde{h_j}$ を利用し，式 (8) によって正解文中の各単語ベクトル y_j を予測させる．

3.2.2 マルチ階層注意機構モデル

Yang ら [6] の研究で，階層注意機構モデルは文書分類を行う際，入力文からより分類結果と関連する単語情報を選択できることを示している．図 3 にマルチ階層注意機構モデルの詳細を示している．本研究では階層注意機構モデルのこのような特徴を利用し，decoder が各単語を予測する際，日本語入力文から，関連する単語情報をよりよく選択するように学習させるため，マルチ階層注意機構モデルを提案している．マルチ階層注意機構モデルは，decoder の現在の隠れ状態 h_j^y と英語文の文脈ベクトル c_j^x を足した結果を利用し (式 (10))，母国語文の文脈ベクトルを計算する仕組みである (式 (11))：

$$a_k^{f,j} = \frac{exp((h_j^y + c_j^x)^T h_k^f)}{\sum_{k=1}^{n} exp((h_j^y + c_j^x)^T h_k^f)}, \quad (10)$$

$$c_j^f = [h_1^f ... h_l^f]^T [a_1^{f,j} ... a_l^{f,j}]. \quad (11)$$

最後に英語文と母国語文の文脈ベクトル c_j^x と c_j^f に対し，式 (12) にもとづき $\tilde{h_j}$ を計算し，式 (8) によって正解文中の各単語ベクトル y_j を予測させる：

$$\tilde{h_j} = W_{hier}[h_j^y; c_j^x; c_j^f]. \quad (12)$$

4. 実験

本研究の全てのモデルは Github にある，PyTorch の seq2seq のチュートリアルコード *2 をもとに実装を行っている．

*2 https://github.com/spro/practical-pytorch/blob/master/seq2seq-translation/seq2seq-translation.ipynb

表 3: ベースラインと二つのマルチソースニューラル文法誤り訂正モデルの GLEU スコア (%)

モデル	開発データ	評価データ
入力文のまま	59.65	58.21
baseline	60.59	58.76
Multi-source-sum	61.41	59.54
Multi-source-hier	**61.92**	**59.89**
日本語文	34.33	33.42

表 4: ベースラインのモデルと提案モデルの訂正例

例 1	
誤り文	Please correct in writing .
日本語入力文	私の文章を訂正してください.
正解文	Please correct my writing .
ベースライン	Please correct in writing .
提案モデル:sum	Please correct my writing .
提案モデル:hier	Please correct my writing .
例 2	
誤り文	I am leaving in Tokyo .
日本語入力文	私は東京に住んでいる.
正解文	I am living in Tokyo .
ベースライン	I am leaving in Tokyo .
提案モデル:sum	I am living in Tokyo .
提案モデル:hier	I am living in Tokyo .

4.1 実験データ

まず，本研究では誤り文，正解文と母国語文の三つの文が揃ったデータが必要である．しかし，既存のデータは誤り文と正解文のペアになっているデータしか存在しない．そこで，本研究では既存の Lang-8 Learner Corpora[5] から，正解文を日本語に翻訳し，対応する日本語母国語文のデータを作成した．正解文の翻訳には Google Translation*3 を使用し，翻訳した日本語文は MeCab*4 で形態素解析をしている．実験では，学習用データ 1,070,000 文，開発データ 18,452 文と評価用データ文 10,227 文を使用した．また encoder と decoder は一層のニューラルネット構造を使い，隠れ状態は 300 次元に設定している．

4.2 実験結果

今回の実験では，GLEU[2] を用いて各モデルの評価を行う．GLEU は機械翻訳課題での評価手法である BLEU[3] を文法誤り訂正課題向けに拡張したものであり，その計算方法は式 (13) に示している． ここで p^* は計算された GLEU スコアを表し，A と B は文を表す．また $count_{A,B}(ngram)$ は，A における $ngram$ の出現頻度と，B における $ngram$ の出現頻度のうち小さい方を返す．．また $count_C(ngram)$ は，C における $ngram$ の出現頻度を返す．即ち GLEU では，入力文 S において訂正されるべき $(ngram)$ が，システムの出力文 C で訂

*3 https://cloud.google.com/translate/?hl=ja
*4 http://taku910.github.io/mecab/

$$p^* = \frac{\sum_{ngram \in C \cap R} count_{C,\ R}(ngram) - \sum_{ngram \in C \cap S} max[0,\ count_{C,\ S}(ngram) - count_{C,\ R}(ngram)]}{\sum_{ngram \in C} count_C(ngram)} \quad (13)$$

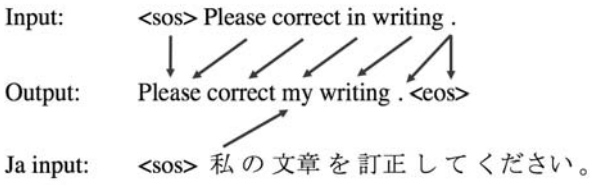


図 4: 各単語を出力する際の注意機構の働き．Ja は日本語を表す．<sos> と <eos> はそれぞれ文頭と文尾を意味する．矢印は，各単語を出力する際に，注意機構の重みが最大となる入力単語を表している．ただし，日本語文については，“my”以外はすべて <sos> が最大であったため，それらは図では省略している．

正されていない場合，減点することでモデルを評価する．ここで R は正解文を意味する．

実験結果は表 3 に表す．同じ実験設定で実験を行った結果，二つの提案モデルともベースラインのモデルを上回る精度を得ている．また，マルチ階層注意機構モデルはベースラインのモデルと比較して，GLEU スコアにおいて 1.13 ポイントの上昇を達成している．

また，今回の実験用データを使い，日本語文から正解文への翻訳を行なっている。その結果，開発データと評価データに対しそれぞれ 34.33 と 33.42 の GLEU スコアを達成している。

4.3 出力に対する考察

ここで，ベースラインモデルと提案モデルの出力結果に対して分析を行う．表 4 に二つの誤り文 “Please correct in writing” と “I am leaving in Tokyo .”の入力に対し，各モデルの出力結果を示す．ベースラインモデルは二つの入力文を正しく訂正出来ていないが，両提案モデルは正しく訂正を行っている．これは提案モデルが日本語入力の情報を利用し，訂正を行っているからだと考えられる．

図 4 にマルチ階層注意機構モデルが表 4 の例 1 の入力に対し，各単語を予測する際の注意機構の行動を示している．システムの各出力単語の母国語入力文における $a_k^{f,j}$ の値を確認した結果，単語 “my”を予測する時，日本語入力文の単語 “私”の情報がもっとも重要視されている．

5. おわりに

本研究は，文法誤り訂正課題において，初めて言語学習者の母国語の情報を利用した研究である．また，実験において，マルチソースニューラル文法誤り訂正のモデルの有効性を検証した．提案モデルは，GLEU スコアにおいてベースラインモデルを上回っていることを示した．

今後の課題として，水本 [10] により提案されたコーパスの構築手法を使い，機械翻訳によって生成された母国語ではなく，語学学習者が書いた母国語文の含むコーパスに対しモデルを適用することが挙げられる．

参考文献

[1] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *CoRR*, 2015.

[2] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel R. Tetreault. GLEU without tuning. *CoRR*, 2016.

[3] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002.

[4] Allen Schmaltz, Yoon Kim, Alexander M. Rush, and Stuart M. Shieber. Adapting sequence models for sentence correction. *CoRR*, 2017.

[5] Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. Tense and aspect error correction for esl learners using global context. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 198–202, Jeju Island, Korea, July 2012.

[6] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489, San Diego, California, June 2016.

[7] Zheng Yuan and Ted Briscoe. Grammatical error correction using neural machine translation. In *HLT-NAACL*, 2016.

[8] Barret Zoph and Kevin Knight. Multi-source neural translation. *CoRR*, 2016.

[9] 永田亮. 語学学習支援のための言語処理. 自然言語処理シリーズ 11. コロナ社, 2017.

[10] 水本智也. 語学学習 SNS の添削ログからの母語訳付き学習者コーパスの構築に向けて. 第 6 回コーパス日本語学ワークショップ予稿集, 2014.

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

単語の分散表現を用いた Earth Mover's Distance と文ベクトルによる対訳コーパスの自動生成

Automatic Generation of Bilingual Corpus by Earth Mover's Distance and Sentence Vectors Using Word Embedding

田上 諒[*1] Ryo Tanoue
越前谷 博[*1] Hiroshi Echizen'ya
荒木 健治[*2] Kenji Araki

[*1] 北海学園大学大学院工学研究科
Graduate School of Engineering, Hokkai-Gakuen 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

We propose a new method to automatically generate bilingual corpus using word embedding. The bilingual corpus is effective as the language resource for natural language processing. However, the generation of bilingual corpus by human needs high cost. In this paper, we propose a new method to automatically generate bilingual corpus using word embedding. Our proposed method uses two similarities: One is the similarity based on Earth Mover's Distance (EMD); another is the similarity based on the sentence vector. Through the evaluation experiments, we confirmed that the weighted mean between the similarity based on EMD and the similarity based on the sentence vector is effective to extract the correct bilingual sentence pair.

1. はじめに

対訳コーパスは，言語学や外国語教育などの言語分析の観点で大きな意義を持つことはもちろん，工学的な観点においても不可欠な言語リソースである．特に近年ではニューラル機械翻訳の進展に伴い，より大規模な対訳コーパスが求められている．しかし，人手により対訳コーパスを収集することはコストが大きな問題となる．そこで対訳コーパスを自動生成するための研究[内山 03]が従来より盛んに行われている．

対訳コーパスを自動生成する手法には，単語の文脈情報を利用した手法[Fung 98]や翻訳用辞書と単語の出現回数を利用した手法[Tamura 12]，そして，潜在的意味解析を利用した手法[Preiss 12, 江里口 14]などが挙げられる．しかし，これらの手法は単語の意味を十分に考慮したものとはなっていない．

そこで，本研究では，単語の分散表現に基づき対訳コーパスを自動生成する新たな手法を提案する．提案手法では，異言語間の文同士の対応関係を単語の分散表現に基づく類似度により決定する．その際，2 つの分布間の距離である Earth Mover's Distance(EMD)と文ベクトルに基づく距離の 2 つの距離を用いて類似度を求める．EMD [Yossi 00, 柳本 07]を用いるのは単語の分散表現を特徴量とすることで，単語の意味を考慮した文間の類似度を得ることができるためである．提案手法では対訳辞書などの高品質な対訳知識を用いることなく，単語の意味を考慮した対訳コーパスの自動生成が可能である．

性能評価実験では，日英 Wikipedia 京都関連文書対訳コーパス [1] にある 100 の対訳文を英語文と日本文に分離したうえで，提案手法により正しい対訳文をどの程度抽出できるかを評価した．その結果，EMD による類似度と文ベクトルによる類似度の加重平均を用いた提案手法に基づくシステムでは EMD による類似度のみと文ベクトルに基づく類似度のみのシステムよりも多くの正しい対訳文が得られた．

2. 提案手法

2.1 提案手法の概要

言語が異なる文を直接比較することはできないため，一方の言語の文の単語分散表現をもう一方の言語の文の単語分散表現に翻訳行列を用いてマッピングする．そして，2 つの分布間の距離を EMD を用いて求める．さらに，対応関係にある文間においては，構成単語数も近いとの仮定に基づき，単語の分散表現の平均同士でベクトル間類似度を求める．本稿ではこの単語の分散表現の平均を文ベクトルと呼ぶ．そして EMD と文ベクトルの 2 つの類似度の加重平均を最終的な類似度とする．

図1に英語から日本語へ向けた類似度計算の概要を示す．

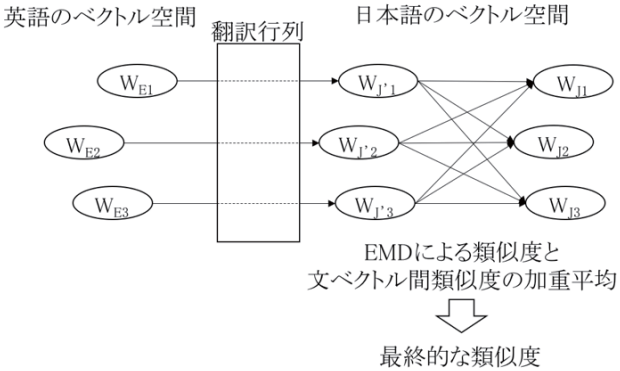


図1 英日方向における異言語文間の類似度計算

連絡先：田上 諒，北海学園大学大学院 工学研究科電子情報生命工学専攻，〒 064-0926 札幌市中央区南 26 条西 11-1-1，6717101r@hgu.jp

[1] https://alaginrc.nict.go.jp/WikiCorpus/

2.2 翻訳行列の生成

翻訳行列はある言語の単語ベクトル空間から他の言語の単語ベクトル空間へ任意の単語を写像するために用いる. 翻訳行列Wは訳語ペアの分散表現の集合$\{x_i, z_i\}_{i=1}^{n}$を用いて, 次の式(1)より得る.

$$\min_{W} \sum_{i=1}^{n} \|Wxi - zi\|^2 \quad (1)$$

2.3 Earth Mover's Distance による類似度計算

ある言語を他の言語のベクトル空間へマッピングした後, 同じベクトル空間上で EMD より類似度を求める. EMD は特徴量, 重み, 距離式を定義することで 2 つの分布間の距離を得ることができる. 2 分布P, Qの特徴量をそれぞれ$\{(p_1, w_{p_1}), \cdots, (p_m, w_{p_m})\}$と$\{(q_1, w_q), \cdots, (q_m, w_{q_m})\}$とする. 分布$P$は$m$個の特徴量で表現されており, p_iは特徴量, w_{pi}はその特徴量に対する重みである. 分布Qも同様である. また, p_iとq_jの距離をd_{ij}とし, 輸送量をf_{ij}と定義する. ${f_{ij}}^*$を最小化された輸送量とすると EMD は以下の式(2)より得られる.

$$\mathrm{EMD}(P, Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}{f_{ij}}^*}{\sum_{i=1}^{m}\sum_{j=1}^{n} {f_{ij}}^*} \quad (2)$$

提案手法では特徴量として単語の分散表現, 重みとして$tfidf$[徳永 99], 距離式にはコサインを用いる. $tfidf$は任意のドキュメントにおける出現単語の重要度を表す尺度である. また, 距離dを計算する際に, 翻訳行列によりマッピングされた単語ベクトルと最も近い単語を近似単語とし, それらが他言語の文に存在するか否かの結果を反映させる. 近似単語が存在する場合には, 単語間の対応関係の信頼性は高いと考えられる. 距離dは以下の式(3)より得る.

$$d = \begin{cases} 1.0 - \cos(\text{近似単語あり}) \\ 1.0 - cos^{2.0}(\text{近似単語なし}) \end{cases} \quad (3)$$

式(3)より, 近似単語が存在しない場合には, コサインを 2 乗することで, 近似単語が存在する場合よりも距離が大きくなるように制御する. さらに$tfidf$の利用においては文中の各単語の$tfidf$の合計が 1.0 になるように正規化することで, EMD の値を 0.0 から 1.0 の範囲に収める. また, 値が大きいほど類似度が大きくなるように, 1.0 から EMD の値を引く.

2.4 文ベクトルによる類似度計算

文の長さが大きく異なる 2 文間において, 短い文の単語の多くが長い文の単語と意味に近い場合でも, 文としての類似度は低いと考えられる. そこで提案手法では, 文の長さの違いをより反映した類似度として, 文を構成する単語の分散表現の平均を文ベクトルとして文ベクトル間のコサインを求める. そして, EMD による類似度と文ベクトルによる類似度との間の加重平均を求め, それを最終的なスコアとする. 加重平均を用いるのは抽出精度に対して EMD による類似度と文ベクトルによる類似度の影響が等しいとは限らないためである.

2.5 対応関係にある文の決定

得られた類似度に基づき対応関係にある文同士を決定する. 英日方向の場合, 1 つの英語文とすべての日本語文との類似度を求め, 類似度の高い順にソートする. その結果, 上位 1 位にランキングされた日本文を基準として, 今度はその日本語文と類似度が最も高い英語文を求める. その結果, 双方向で類似度が最も高い場合, 対応関係にあるとして, 抽出する. 図 2 にその具体例を示す. 図 2 では英日方向で E1 と J1, 日英方向でも J1 と E1 が対応したと見なされたため, (E1:J1)が対応関係にある文として決定される.

2.2 以降の処理をさらに日英方向でも行う. そして, 双方向で一致した組み合わせを最終的な対応関係にある文同士とする. 例えば, 図 2 で使用された英語文と日本文において, 日英方向で(J1: E1)が得られた場合は(E1:J1)と一致するため, E1 と J1 は対応関係にある文として決定される.

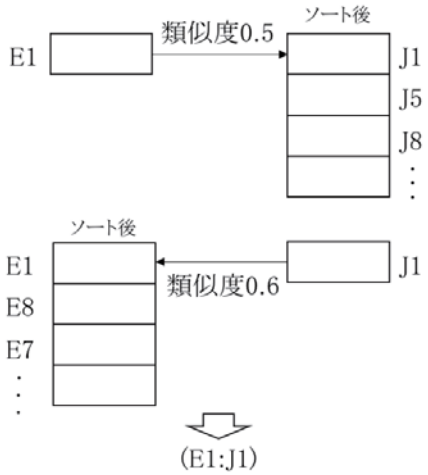


図 2 英日方向における対応関係にある文の決定

3. 性能評価実験

3.1 実験方法

性能評価実験を行うための実験データには Wikipedia 日英京都関連文書対訳コーパスに含まれているカテゴリ「仏教」から対訳文 100 文をランダムに抽出した. そして, 対訳文を英語文と日本語文に分離し, 提案手法により正しい対訳文をどの程度得られるのかを評価した.

提案手法を用いる際, 単語の分散表現は英語と日本語の Wikipedia ダンプデータから word2vec[Tomas 13 (a)]を用いて得た. ただし, サブサンプリングの閾値パラメタを 0.001 に設定することで, 機能語のような高頻度単語の影響を抑制する. 翻訳行列については, 文献[Tomas 13 (b), 石渡 16]により, 英日方向, 日英方向共に原言語の次元数は 800, マッピング先の目標言語の次元数は 200 で学習を行なった. また, 学習データに用いた対訳語ペアについては Wikipedia ダンプデータにおいて出現頻度が上位の単語と Google 翻訳を用いて得られた訳語のペアを用いた. 本研究では, 対訳語 10,000 ペアを翻訳行列の学習データとして用いた.

また, 提案手法の有効性を検証するために, EMD のみのシステムと文ベクトルのみのシステム, そして, 提案手法のシステムを用いた. ただし, EMD のみのシステムにおいては近似単語を用いていない.

3.2 実験結果

3.1 で述べたシステムの実験結果を表1に示す. 抽出精度は英日 100 文ずつから正しい対訳文を抽出できた割合を示す. また, 表 1 の提案手法は, EMD による類似度と文ベクトルによる類似度の重みが等しい相加平均によるシステムと文ベクトルによる類似度の重みを EMD による類似度に対して 2 倍とした加重平均によるシステムの結果とした.

表1 実験結果

手法	抽出精度
EMDのみ	19%
文ベクトルのみ	44%
提案手法(相加平均)	36%
提案手法(加重平均)	45%

3.3 考察

表1より提案手法における加重平均が最も高い抽出精度を示した. 加重平均では文ベクトルによる類似度の重みが EMD による類似度の重みの 2 倍であるため, 文ベクトルが効果的な役割を果たしていることがわかる. 文ベクトルのみの抽出精度が44%であったことからも文ベクトルの効果は大きいと考えられる. 提案手法の加重平均と文ベクトルのみで抽出された対訳文を比べると, 提案手法により新たに得られた対訳文は 16, 変化のなかった対訳文は 29 であった. 提案手法により新たに得られた対訳文の一つは英語文「by intellectual understanding, it is impossible to clap with one hand and make a sound.」と日本文「知的な理解では片手では拍手はできず音はしない。」からなる対訳文であった. 提案手法では, この対訳文においては英日方向, 日英方向のいずれも類似度が最も高かったため, 正しい対訳文として抽出された. それに対して, 文ベクトルのみでは, 英日方向では 100 番中 6 番に類似度が高く, 日英方向では 13 番目に類似度が高かったため, 正しい対訳文として抽出されなかった. したがって, 抽出精度としては提案手法の加重平均は文ベクトルに対してわずかに上回ったという結果ではあるが, 得られた対訳文の 3 割が異なるため, さらなる精度向上が期待できる.

4. おわりに

本稿では, 単語の分散表現を用いた対訳コーパスの自動生成のための新たな手法を提案した. 提案手法により EMD による類似度と文ベクトルによる類似度を組み合わせることの有効性を確認した. しかし, 抽出精度はまだ不十分である. 今後は, 2 つの類似度のそれぞれの利点をさらに引き出すことで抽出精度の向上を図る予定である.

参考文献

[内山 03] 内山将夫, 谷村緑: パラレルコーパスの自動生成技術, 情報通信研究機構季報. 53(3), 23-28 ,情報通信研究機構 (2007).

[Fung 98] Fung P. and Yee L. Y.: An IR approach for translating new words from nonparallel, comparable texts, COLING '98 Proceedings of the 17th international conference on Computational linguistics - Volume 1, 414-420 (1998).

[Tamura 12] Tamura A., Watanabe T. and Sumita E.: Bilingual lexicon extraction from comparable corpora using label propagation, EMNLP-CoNLL '12 Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 24-36 (2012).

[Preiss 12] Preiss J.: Identifying comparable corpora using LDA, NAACL HLT '12 Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 558-562 (2012).

[江里口 14] 江里口瑛子, 小林一郎: 潜在情報を利用したパラレルコーパス生成, 第 28 回人口知能学会全国大会論文集, 3I4-1, 人工知能学会 (2014).

[Yossi 00] Yossi Rubner, Carlo Tomasi, Leonidas J. Guibas: The Earth Mover's Distance as a Metric for Image Retrieval, International Journal of Computer Vision November 2000 Volume 40 Issue 2, 99–121 (2000).

[柳本 07] 柳本豪一, 大松繁: Earth Mover’s Distance を用いたテキスト分類, 第 21 回人口知能学会全国大会論文集, 1G3-4,人工知能学会 (2007).

[徳永 99] 徳永健伸:情報検索と言語処理, 東京大学出版会 (1999)

[Tomas 13 (a)] Tomas M., Kai C., Greg C. and Jeffrey D.: Efficient estimation of word representations in vector space, arXiv:1301.33781v3 (2013)

[Tomas 13 (b)] Tomas M., Quoc V. Le, and Ilya S.: Exploiting similarities among languages for machine translation, arXiv:1309.4168v1 (2013).

[石渡 16] 石渡祥之佑, 鍛治 伸裕, 吉永 直樹, 豊田 正史, 喜連川 優: 文脈語間の対訳関係を用いた単語の意味ベクトルの翻訳, 人工知能学会誌, 32(1), 20-29, 人工知能学会 (2016).

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

漢字分解したテキストによるニューラル機械翻訳

Neural Machine Translation with Kanji Decomposition

ビィシュウ グプタ Vishu Gupta　中村 亮裕 Akihiro Nakamura　福田 治輝 Haruki Fukuda　綱川 隆司 Takashi Tsunakawa　狩野 芳伸 Yoshinobu Kano
西田 昌史 Masafumi Nishida　西村 雅史 Masafumi Nishimura

静岡大学
Shizuoka University

This paper proposes a method for neural machine translation (NMT) with kanji decomposition of Japanese text. NMT models have restrictions of the vocabulary size, which can be solved by applying subword, character-level, or byte-based models. In Japanese text, the vocabulary size would not be minimized even in a character level because of kanji varieties. We report an experimental result of NMT model using Japanese text with kanji decomposition that is expected to satisfy both of decreasing vocabulary size and keeping kanji information.

1. はじめに

ニューラル機械翻訳モデルでは語彙は一定のサイズに限定される．この課題に対処する方法の一つとしてサブワード単位や文字単位で処理する方法が提案されている [Sennrich 16]．これらの方法の利点の一つとして未知語の問題がなくなることが挙げられる．文字単位であれば語彙サイズは原言語と目標言語のアルファベットの範囲内となり非常に小さくなる一方で，翻訳性能は維持される，または向上するとの報告もある [Lee 17]．さらに文字コードを用いてバイト単位にまで分割することで，言語に依存せず語彙サイズを 256 にまで小さくするモデルも提案されている [Costa-jussà 17]．

日本語では従来，形態素解析によりトークン化したものを機械翻訳の入出力としていた．これを文字単位とした場合，漢字のバリエーションが多いために語彙サイズは比較的大きくなる．語彙サイズを小さくするために漢字を読みに変換することはできるが，同音異義語の情報が失われてしまう．

本研究ではこれらの問題を抑えつつ，漢字よりも細かい単位への分割を適用するため，漢字をその構成要素に分解してからニューラル言語処理向けトークナイザ SentencePiece[*1] を適用したトークン列を生成し，ニューラル機械翻訳の入出力として扱う方法を提案し，その効果を検証した．

2. 提案手法

図 1 に提案手法の概要を示す．まず，コーパス中の日本語の各漢字を KRADFILE[*2] に基づく変換テーブルを参照して分解する．次に，SentencePiece により構成要素に分解した文字列を再度結合する．この日本語テキストを用いてニューラル機械翻訳モデルを学習させる．翻訳結果として得られたテキストは逆変換によりもとの漢字に戻して出力する．

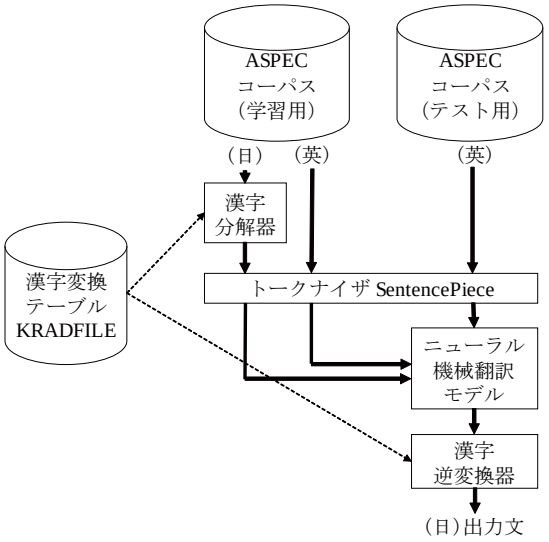


図 1: 提案手法概要

2.1 KRADFILE による漢字分解と逆変換

KRADFILE は，漢字の構成要素による検索を容易にするために提供されている，漢字とその構成要素の変換テーブルである．例えば，

- 哀：衣口𠂉
- 愛：心爪冖夂

のような漢字分解が示されている．構成要素はそれを表す代表の漢字である場合がある．例えば，“こざとへん” は “阡” で表される．

本稿では原則としてこのテーブルに従って漢字を置換する．ニューラル機械翻訳によって得られた文字列は，左方最長一致で合致するテーブルの項目に該当する漢字に逆変換する．しかし，この方法では組合せによる誤結合が発生するため，下記の処理を施している．

- 同じ分解結果となる異なる漢字を区別するため，分解結

連絡先: 綱川 隆司，静岡大学情報学部，〒 432-8011 静岡県浜松市中区城北 3 丁目 5-1，053-478-1487，tuna@inf.shizuoka.ac.jp

∗1 https://github.com/google/sentencepiece
∗2 http://www.edrdg.org/krad/kradinf.html

果の末尾に判別用の文字を追加する.（例　跟：口艮足止 α，踉：口艮足止 β）

- 構成要素として用いられる漢字自体は，前後に判別用の文字を追加したパターンに変換する.（例　耳：#耳#）
- 左からの最長一致により誤結合が起こる単語に対応するため，一部の漢字の構成要素列の順序を入れ替える.（例　需：雨而　のままだと，需要→雨而女西→嬬西，となり逆変換に失敗するため，需：而雨　とする）

2.2 SentencePiece の適用

提案方法では漢字分解した文字列をそのまま入力するのではなく，SentencePiece によって得られたトークン列として入力する．SentencePiece はサブワード分割アルゴリズムの一種を採用しており，バイトペアエンコーディング (BPE) により文字列から直接分割を学習する．これにより語彙サイズと分割単位のバランスをとることができる．

本研究では英語テキスト，および日本語テキストの漢字分解後の文字列に対して SentencePiece を適用し，得られた分割をトークンとしてニューラル機械翻訳の入力とする．出力結果のトークン列を単純に結合し，漢字への逆変換を施す．

下記は，日本語文 “現在，筋ジストロフィー患者の移動介助において文書マニュアルを使用している。” に対して単に SentencePiece を適用した場合と，漢字への分割を施してから SentencePiece を適用し，漢字への逆変換を施した場合の例を示す．トークンの区切り目はアンダースコア (_) で示している．

SentencePiece 現在_，_筋ジストロフィー_患者の_移動_介助_において_文書_マニュアル_を使用し_ている_。

漢字分割＋ SentencePiece ＋逆変換（区切り削除前） 現_|_丿一土_，_筋ジストロフィー_|_口口心_者_の移_|_一日力里丿_介助_において_文_書_マニュアル_をノ一化_口_用_している_。

SentencePiece による分割では，助詞が結合するケースはあるものの概ね従来の形態素解析器に近い結果となっているのに対し，漢字分割したケースではより細かく分割され，また漢字の構成部分の途中に区切りが来るケースが散見された．特に，“|” や “口” は単独のトークンとして扱われる傾向にあった．評価実験では翻訳結果の出力の際に区切りを削除し結合してから逆変換を施すため，逆変換に失敗するケースはほぼみられない．

3. 実験

科学技術論文抄録の対訳からなる ASPEC コーパス [Nakazawa 16] の日英対訳文対を用いて英日方向の翻訳実験を行った．学習用データ 100 万文対を訓練に用いてニューラル機械翻訳モデルを学習し，テスト用データ 1812 文対を用いて評価した．ニューラル機械翻訳モデルは OpenNMT[Klein 17] を用いて構築し，語彙サイズを 40000，トークン列長の最大数を 150 とした．漢字分解の効果を検証するため，日本語テキストに対して直接 SentencePiece による分割を施した場合，および SentencePiece を適用せずに漢字分解後の文字単位で分割した場合との比較を行った．評価時には分割方法による差異を解消するため，参照文および出力文に対して MeCab[*3] による形態素解析を施してから BLEU スコアを算出した．

*3　http://taku910.github.io/mecab/

表 1: 実験結果

適用手法	BLEU	BP
SentencePiece	0.2480	0.914
漢字分割+SentencePiece	0.2408	0.911
漢字分割	0.1260	0.670

表 1 に各実験設定における BLEU スコア，およびスコア算出の際に用いる BP (Brevity Penalty) を示した．BP は出力文のトークン数が参照文のトークン数より短い場合にかかる係数であり，小さいほど出力文長が短いことを表す．漢字分解を適用した場合のスコアは適用しない場合をやや下回っており，漢字分解の効果は確認できなかった．また，漢字分解を適用しても出力文長にはほぼ変化がみられなかった．SentencePiece 適用後の語彙サイズは漢字分解適用後もあまり変化していないと考えられる．一方，SentencePiece を適用しない場合は低い精度にとどまった．これは，漢字分解後の文字単位の入力トークン列が非常に長くなるために十分な学習ができず，出力文も短くなったことによるものと考えられる．

4. おわりに

本稿ではニューラル機械翻訳による英日翻訳において，語彙サイズの制約を軽減するために日本語の漢字分解を適用した文字列を入出力に用いる方法を提案した．実験結果からは SentencePiece によるトークン化を適用した場合，漢字分解によりトークンが比較的細かく分割されるという結果が得られるものの，漢字分解による精度向上はみられなかった．今後の課題として，漢字の部首と読みを用いるなどの漢字分解方法の検討や，文字ベースのニューラル機械翻訳 [Lee 17] の適用が挙げられる．

参考文献

[Costa-jussà 17] Costa-jussà, M. R., Escolano, C., and Fonollosa, J. A. R.: Byte-based Neural Machine Translation, in *Proc. of the 1st Workshop on Subword and Character Level Models in NLP*, pp. 154–158 (2017)

[Klein 17] Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. M.: OpenNMT: Open-Source Toolkit for Neural Machine Translation, *ArXiv:1701.02810* (2017)

[Lee 17] Lee, J., Cho, K., and Hofmann, T.: Fully Character-Level Neural Machine Translation without Explicit Segmentation, *Trans. of the Association for Computational Linguistics*, Vol. 5, pp. 365–378 (2017)

[Nakazawa 16] Nakazawa, T., Yaguchi, M., Uchimoto, K., Utiyama, M., Sumita, E., Kurohashi, S., and Isahara, H.: ASPEC: Asian Scientific Paper Excerpt Corpus, in *Proc. of the 9th International Conference on Language Resources and Evaluation*, pp. 2204–2208 (2016)

[Sennrich 16] Sennrich, R., Haddow, B., and Birch, A.: Neural Machine Translation of Rare Words with Subword Units, in *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1715–1725 (2016)

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

ファクトチェックのための要検証記事探索の支援

Support for Searching for Articles to be Verified for Fact Checking

内山 香 *1 Kaori Uchiyama
鈴木 海渡 *1 Kaito Suzuki
田上 翼 *1 Tsubasa Tagami
塙 一晃 *1 Kazuaki Hanawa
乾 健太郎 *1*2 Kentaro Inui
小宮 篤史 *3 Atsushi Komiya
藤村 厚夫 *3 Atsuo Fujimura
町野 明徳 Akinori Machino
楊井 人文 *4 Hitofumi Yanai
山下 亮 *4 Ryo Yamashita

*1東北大学 Tohoku University
*2理化学研究所 革新知能統合研究センター RIKEN Center for Advanced Intelligence Project
*3スマートニュース株式会社 SmartNews, Inc.
*4一般社団法人日本報道検証機構 Watchdog for Accuracy inNews-reporting, Japan

False information have become a social problem and it is necessary to verify huge information such as news articles on the Internet. In this paper, we constructed a support system to efficiently verify the authenticity of news articles. This system estimates news articles that describe incorrect information based on posts that refer to them on SNS. We found that it can be expected to improve efficiency of searching for articles that require human verification by using the created system.

1. はじめに

国内外を問わず誤情報の拡散が社会的な問題となっている．アメリカの大統領選挙に関してメディアの報道やソーシャルメディア（SNS）上で投票に影響を与えうる多くの誤情報が拡散されたことや，国内でも健康・美容情報サイトにおいて人の健康に影響しうる誤った医療情報の記事が掲載されるなど，誤情報への対策が急務となっている．このような誤情報の拡散を防ぐためには政治家の言説，メディアの報道，ウェブ上のコンテンツや SNS 上の投稿など，社会に影響を及ぼしうる情報の真偽や正確性を検証するファクトチェックが必要である．

ファクトチェックの対象は政治分野だけでなく多岐に渡る．対象をニュース記事に絞っても日々配信される膨大な量のニュース記事全てに対し，人手で個々の内容を調査をすることは不可能である．そこで一般社団法人日本報道検証機構 *1 では以下の 3 つのステップによって内容を調査をするニュース記事を選別している．第 1 のステップではニュース記事に言及している SNS の投稿から (1) のような検証の必要性を示唆する情報（同機構ではこれを「端緒情報」と呼んでいる）を人手で収集する．第 2 のステップでは収集された端緒情報の内容や記事の影響度・深刻度等に基づいて，より詳細な本調査を必要とする記事（以下「要検証記事」と呼ぶ）を拾い上げる．最後に個々の要検証記事についてファクトチェックの本調査を行い，必要に応じて検証結果を記事化する．

(1) a. ○○新聞では縮小といってる。他紙と違って**不可解**ですが両方共ちゃんと裏付けしてるのか？> http://xxx

b. この記事は**誤報**では？千代田区も路上喫煙はダメで過料が科されているはずです！> http://xxx

このファクトチェックプロセスは，SNS 上の情報をある種の集合知として活用することによって要検証記事をできる限り網羅的に同定することをねらうもので，その有効性は日本報道検証機構のこれまでの実績にも示されている．ただし，膨大な SNS 情報から端緒情報を人手で探索する第 1 のステップが膨大な時間を要する作業となっており，この部分がボトルネックとなって要検証記事の探索コストを押し上げていることが深刻な問題となってきた．端緒情報の探索では，SNS の投稿を「誤報」や「デマ」などのフレーズでフィルタリングしてはいるが，それでもフィルタを通過した投稿のほとんどは (2) のようにファクトチェックには役立たない投稿である．実際に，2017 年に日本報道検証機構とファクトチェック・イニシアティブ *2，賛同メディアが連携して実施した衆議院議員総選挙ファクトチェックでは，総選挙関連の Twitter の投稿を 1 日あたり約 16,000 件人手でチェックしたが，そのうち端緒情報と認められた投稿は平均わずか 12 件であった．

(2) a. 本当に信じられない。**嘘**であって欲しい。言葉が見つからないけどご冥福をお祈りします。! > http://xxx

b. ○○○○大統領との親密さをアピール？札束でアピールの**間違い**だ !! > http://xxx

そこで本研究では，この端緒情報の収集を自動化し，人手による要検証記事探索作業を技術的に支援する仕組みの構築を考える．具体的には，代表的な SNS である Twitter*3 の投稿を対象に，そこから端緒情報である可能性が高い投稿を自動抽出し，それらの情報に基づいてニュース記事を検証必要度の観点からランキングする．これによって検証必要度が高い記事をチェックするだけで要検証記事を網羅的に収集できるようになると考えられる．本稿では，要検証記事の探索支援に向けた研究開発の途中経過を報告する．なお我々が調べた限りでは，インターネット上の誤情報を検出する研究は行われているものの [1, 4]，Twitter の投稿から端緒情報を抽出し，要検証記事を収集する研究は本研究が初である．

連絡先: 田上 翼，東北大学，宮城県仙台市青葉区荒巻字青葉 6-6-05，022-795-7091，tagami@ecei.tohoku.ac.jp

*1 http://wanj.or.jp/

*2 ファクトチェック活動の支援を目的として 2017 年 6 月に設立され，2018 年 1 月に NPO 法人を取得． http://fij.info/

*3 https://twitter.com/

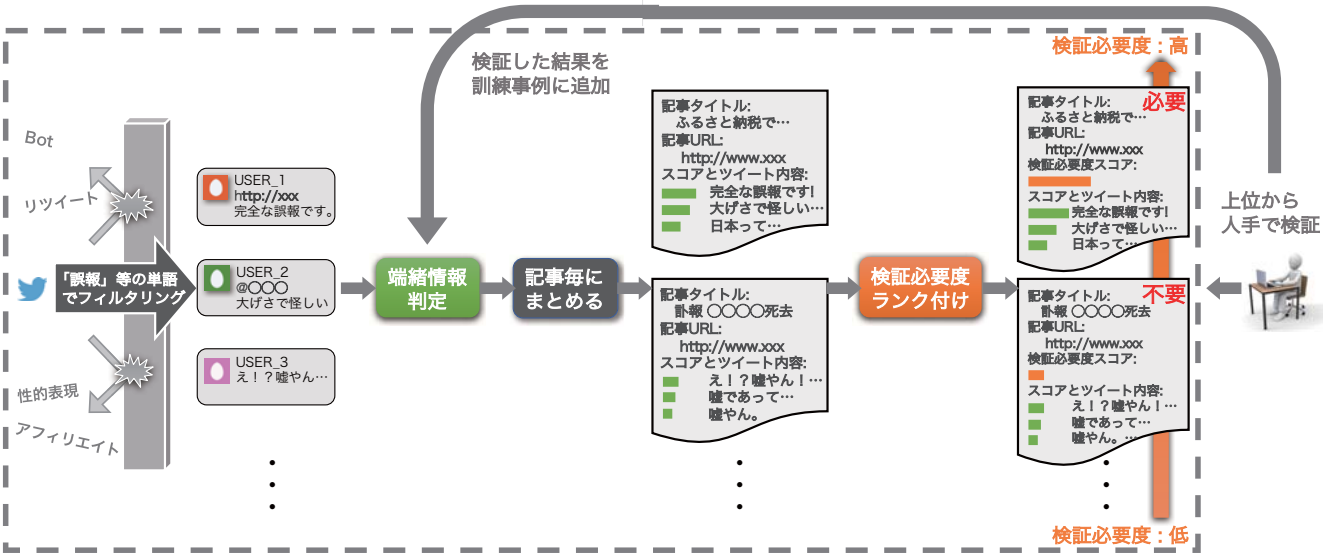


図 1: ファクトチェック支援システムの構成

2. 要検証記事の探索

要検証記事を探索するために用いる手法の概略を図 1 に示す．手順はツイートの収集，端緒情報の判定，記事の検証必要度のランク付け，人手による検証の 4 つに分けられる．

2.1 ツイートの収集

対象のニュースサイトで掲載された記事へ言及しているツイートを以下の 3 つの方法を用いて収集した．また収集対象は 2017 年 1 月から 6 月のツイートとした．

URL を含むツイート

ツイートが言及している対象の記事を明確にするため，対象のニュース記事の URL を含むツイートを収集した．

リプライツイート

Twitter では特定のツイートに対する返信をリプライという．いくつかのニュースサイトでは Twitter アカウントを所有しており，記事を配信すると同時にその記事についてツイートをして周知をすることがある（以下ではこのツイートを「ニュースツイート」と呼ぶ）．このツイートに対するリプライツイートは記事に対して言及していると考えられるため，ニュースツイートに対するリプライツイートを収集した．

リツイート直後のツイート

Twitter では他のユーザのツイートを共有するリツイートと呼ばれる仕組みがある．図 2 に示すように，あるユーザがリツイートした直後にそのツイートに関連する内容のツイートをすることがある．そこでニュースツイートをリツイートしたユーザのリツイート直後のツイートを収集した．収集したツイートには記事に関連していないものも含まれるため，1016 件のツイートを対象にリツイートとの関係の有無を調査したところ，リツイートとその直後にしたツイートまでの時間に図 3 に示す関係があることがわかった．図 3 より記事に関係する内容のツイートは時間が経つごとに減少することがわかる．また 240 秒を超えた付近で関係のあるツイートの割合が大きく下がるため，本研究では 240 秒以内に投稿されたツイートのみを収集した．

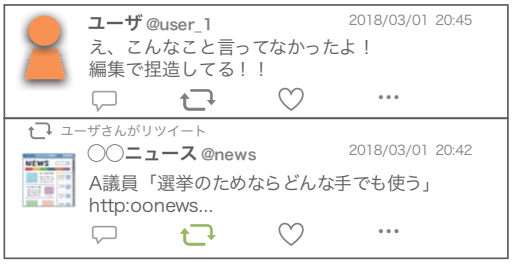


図 2: リツイート直後のツイートの例

2.2 端緒情報の判定

端緒情報を抽出するために，2.1 節で収集したツイートから日本報道検証機構が実際に用いている表 1 のフレーズでフィルタリングした．それぞれの手法で収集されたツイートのフィルタリング後の数は URL を含むツイートが 468,188 件，リプライツイートが 17,232 件，リツイート直後のツイートが 18,429 件であった．また，ニュース記事について言及しているツイートには，(3) のように記事のタイトルや本文の一部が引用されることが多い．解析の対象としたいのは記事のタイトルや本文を除いたコメント部分であるため，言及している記事のタイトルと本文を獲得しツイートとの重複部分を除去する．この他にツイート中の URL とハッシュタグ，記号および英数字を除去する．またリツイートや Bot によるツイートなどは除外する．次に，収集した各ツイートを 3. 章で述べる端緒情報の分類器で判定し，端緒情報である確率によりスコア付けする．本研究ではこの判定を行う分類器を作成したデータセットを用いて構築し，その性能を評価した．

(3) 〇〇でデマを流す男、〇〇。2018 年の ACL 鹿島はストレートインだろうが。＞サッカー次期代表監督、外国人に限定しないで: 〇〇〇〇新聞 http://xxx

2.3 記事の検証必要度ランク付け

端緒情報か判定されたツイートを記事ごとにまとめ，各ツイートのスコアを用いてそれぞれの記事の検証必要度をランク付けすることで，要検証記事の探索を支援する．本研究ではこ

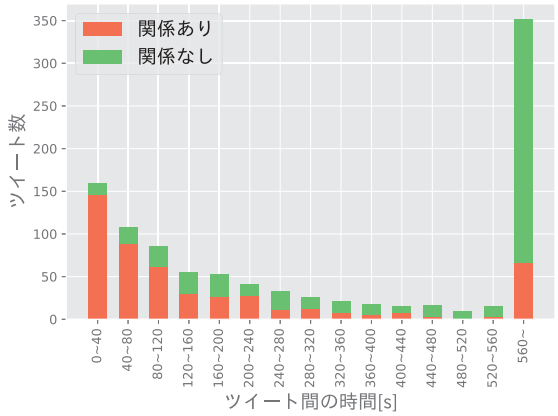


図 3: ツイート間の時間ごとのリツイートとその直後のツイートの関係の有無

表 1: フィルタリングに用いるフレーズ一覧

怪しい	偽	偽装	違和感	嘘	噂	改竄	ガセ	疑い	疑惑	記事について
誤り	誤解	誤認	誤報	誤用	根拠	信用しない	信用できない	ステマ	デマ	とんでもない
捏造	パクリ	否定	フェイク	不可解	偏向	変造	マスゴミ	間違い	ミスリード	無根

の記事の検証必要度のランク付けによる要検証記事の分類性能も評価した．

2.4 人手による検証

実際にファクトチェックを行う時にはシステムの出力を人手でチェックし，要検証記事であるかを判定する．このときシステムが判断した要検証記事と端緒情報を人手で修正することにより，学習のトレーニングデータを拡充することができる．この仕組みにより端緒情報の判定と記事の検証必要度をランク付けするモデルの性能改善が期待される．

3. 実験

人手でラベル付けを行ったデータセットを用いてツイートが端緒情報であるか判定する分類器を構築し，その性能を評価した．また構築した分類器を用いて記事の検証必要度をランク付けし，その性能についても評価した．

3.1 データセット

ツイートの端緒情報判定に用いるためのツイート単位データセットと，ニュース記事の要検証記事分類に用いるための記事単位データセットの 2 つを人手で作成した．

3.1.1 ツイート単位データセットの作成

2.1 節で述べた 3 つの方法で収集した各ツイートを人手で端緒情報であるか判定した．これらのツイートうち正例のツイートの数は 1036 件で URL を含むツイートで 50,000 件中の 606 件（1.21%），リプライツイートで 5173 件中の 423 件（8.17%），リツイート直後のツイートで 508 件中の 7 件（1.37%）であった．また負例については正例でないツイートのうち bot によるツイートや，記事の本文やタイトルを引用しているだけのツイート，リンク切れなどの要因で記事の本文やタイトルとの重複部分を自動で除去できないツイートを人手で取り除いたものを使用した．負例のツイートの数は 6739 件でその内訳は URL を含むツイートで 50,000 件中の 2914 件（5.82%），リプライツイートで 5173 件中の 3468 件（67.0%），リツイート直後のツイートで 508 件中の 357 件（70.3%）であった．

表 2: 端緒情報判定の性能

モデル	適合率	再現率	F 値
ロジスティック回帰	0.66	0.58	0.62
LinearSVM	0.66	0.58	0.62
決定木	0.49	0.48	0.48
ランダムフォレスト	0.67	0.39	0.49
LSTM	0.62	0.55	0.59

3.1.2 記事単位データセットの作成

ツイートが言及している記事 URL ごとにまとめたツイートの集合に，端緒情報であるツイートが 1 件でも含まれている記事を正例とし，ツイート単位データセットと同様に人手で判定した．正例である記事は 564 記事であり，また負例は正例でない記事のうち 1,271 記事を用いた．

3.2 手法

ツイートが端緒情報であるかを判定する分類器を，以下の 5 つのモデルで構築した．形態素解析には，形態素解析器 MeCab-0.996[*4] と，辞書として mecab-ipadic-neologd[*5] を使用した．モデルの実装には機械学習ライブラリ scikit-learn[*6] と深層学習ライブラリ Keras[*7] を使用した．

1. ロジスティック回帰
2. LinearSVM
3. 決定木
4. ランダムフォレスト
5. LSTM[2]

モデル 1 から 4 については素性に Bag-of-Words と Bag-of-Bigrams を使用した．また，前処理として助詞を全て取り除き，また形態素解析の結果から基本形の数を数え，ストップワードとして頻度の高いの単語を利用した．モデル 5 については，単語の分散表現は，表 1 のフレーズを含む約 450 万件の日本語ツイートを抽出し，word2vec[3] を用いて学習した．次元数は 300 とした．

3.3 実験 1：端緒情報判定

上述の 5 つのモデルを用いて端緒情報の分類器を構築し，それぞれの手法の適合率と再現率，F 値を評価するために，5 分割交差検証を行なった．またそれぞれの手法について，今後データを増やすことで分類器の性能向上を期待できるかを確認するために，トレーニングデータのサイズを 1/8，1/4，1/2 と変化させた場合の F 値の変化についても計測した．

表 2 に各手法で構築した端緒情報分類器の適合率と再現率，F 値を示し，図 4 にトレーニングデータのサイズを変化させた時の学習曲線を示す．表 2 より，ロジスティック回帰と LinearSVM が F 値で最も高いスコアを示した．LSTM 以外の 4 つの手法では素性として Bag-of-Trigtams も用いる場合や，Bag-of-Words のみを用いた場合でも同様の実験を行なったが，Bag-of-Words と Bag-of-Bigrams を用いた場合が最も優れたスコアを示した．次に図 4 より，各手法でトレーニングデータのサイズを大きくしていくと，F 値は緩やかに上昇しており，まだ飽和は生じていないことがわかる．このことから実際にシステムを利用してトレーニングデータを拡充することにより，いずれの手法でも端緒情報分類器の性能向上が期待できる．

*4 https://taku910.github.io/mecab/
*5 https://github.com/neologd/mecab-ipadic-neologd
*6 http://scikit-learn.org/stable/
*7 https://keras.io/

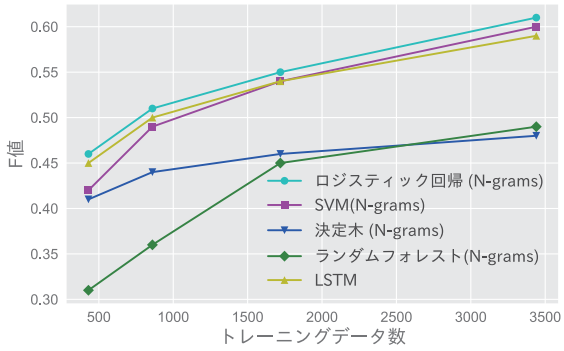


図 4: トレーニングデータサイズを変えた時の F 値

表 3: 端緒情報判定の誤り例：+1 は端緒情報であり，-1 は端緒情報でないことを表す．

ツイート例	予測	正解
(1) ○○の「県民の意向」は、聞き飽きた … 何処の県民？河北とか四川とか、○○省の間違いだろ	+1	-1
(2) 金の掛かる万博、本当にやりたいんだろうか。東京みたいにミスリードするな。	+1	-1
(3) 遼寧は普通に 30 ノット出してたし格別遅くはないぞ。嘘はよくないな。	-1	+1

いずれの手法でも正しく判定できなかったツイートを表 3 に示す．表中の (1) と (2) はいずれの手法でも誤って端緒情報であると判断したツイートの例であり，(3) はいずれの手法でも誤って端緒情報でないと判断したツイートの例である．例 (1) は単純に皮肉を言っているものであり，例 (2) はミスリードを指摘しているが，指摘の対象が記事ではなく行政などを指すものである．このようなツイートを正しく判定するには，ツイートが言及する記事などの外部知識を用いて，皮肉の検出や指示対象の特定などの深い処理が必要であると考えられる．次に，例 (3) のツイートに含まれる，「嘘」という単語はフィルタリングに用いた表 1 の他の単語に比べて日常的にツイートに使われやすく，データセットの負例中の約 24%のツイートに含まれている．このため「嘘」を含むツイートは単純な手法では誤って端緒情報でないと判断されやすいのだと考えられる．

3.4 実験 2：検証必要度に基づく記事のランク付け

3.3 節で用いた 5 つのモデルによる端緒情報分類器を利用し，3.1 節で作成したデータセット上で記事の検証必要度をランク付けする実験を行った．各記事の検証必要度のスコアは記事毎にまとめられた各ツイートの端緒情報分類器によるスコアのうち最も高い値を利用した．評価尺度には，要検証記事の分類性能を評価するために適合率と再現率，F 値を用いた．また，十分な量の要検証記事を獲得するには上位何件の記事を検証すれば良いかを調べるために Recall@K による評価も行った．

表 4 に適合率と再現率，F 値を示し，図 5 に Recall@K を示す．図 5 より，正例と負例が 3:7 の割合である記事単位データセットにおいて，端緒情報分類器としてロジスティック回帰と linearSVM と LSTM を用いた場合では，検証必要性スコア上位の 26%の記事を見れば 6 割，45%を見れば 8 割の正例を獲得できることがわかる．このように今回作成したデータセット上では有効なことは確かめられたが，実際は要検証記事は全体の数%であるため今後はデータセットを拡充し実際の条件に近づけたときの性能を調べる必要があると考えられる．また本研究では各記事の検証必要度のスコアを，記事毎にまとめられた各ツイートの端緒情報分類器によるスコアのうち最も高いものを利用したが，この方法では記事に対するツイート数などを考慮しておらず，評価指標を検討し直す必要があると考えられる．

表 4: 要検証記事の分類性能

モデル	適合率	再現率	F 値
ロジスティック回帰	0.73	0.62	0.67
LinearSVM	0.74	0.63	0.68
決定木	0.62	0.34	0.44
ランダムフォレスト	0.81	0.34	0.48
LSTM	0.64	0.75	0.69

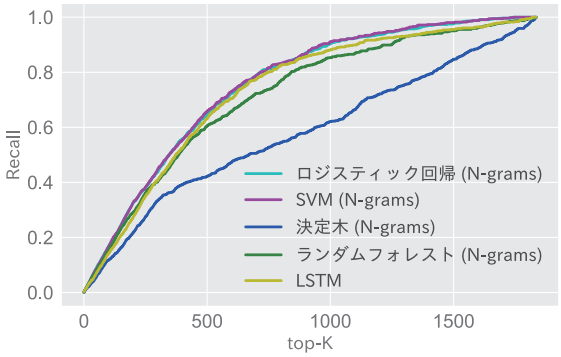


図 5: 検証必要度のランク付け性能

4. おわりに

本研究ではファクトチェックを支援するために，要検証記事の探索を支援するシステムの構築に取り組んだ．システム中では端緒情報の判定と記事の検証必要度をランク付けするために，端緒情報の分類器を構築した．この分類器を用いることでデータセット上においては要検証記事の探索作業の効率化を期待できることが確かめられた．今後は実際にこの仕組みによるシステムをファクトチェック業務に利用し，端緒情報と要検証記事のデータセットの拡充や学習のアルゴリズムの改良を進め，要検証記事の検出性能の向上を目指す．また収集されたツイートの言語的特徴の分析や実用上で生じる課題の検討に取り組む予定である．

謝辞

本研究の一部は NTT コミュニケーション科学基礎研究所，JSPS 科研費 15H01702，東北大学 Step-QI スクールの支援を受けた．

参考文献

[1] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1803–1812. ACM, 2017.

[2] Sepp Hochreiter and J Urgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[3] Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[4] 宮部真衣, 梅島彩奈, 灘本明代, 荒牧英治. 流言情報クラウド: 人間の発信した訂正情報の抽出による流言収集.

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

イベント系列からの有意性を考慮した菱形エピソードマイニング

Mining Significant Diamond Episodes from Event Sequences

谷 陽太 *1 Yota Tani　古谷 勇 *1 Isamu Furuya　平田 耕一 *2 Kouichi Hirata　有村 博紀 *1 Hiroki Arimura

*1北海道大学 大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

*2九州工業大学大学院 情報工学研究院
Department of Artificial Intelligence, Kyushu Institute of Technology

In this paper, we consider the framework of statistically significant ranking for frequent episodes, proposed by Tatti (DMKD, 2015), where an episode is a structured pattern in the shape of vertex labeled directed acyclic graphs. For the class of diamond episodes, we present an algorithm, called MineRankedDmd, that generates frequent episodes with their ranking in a collection of event sequences. In the experiments on synthetic data set, we evaluated the performance of the proposed algorithm.

1. はじめに

データマイニングの一種である系列マイニングは、系列データから頻出する部分系列を抽出する手法である [1]。一方、Mannila ら [4] によるエピソードマイニングは、イベント列に同時に出現するイベントの構造である**エピソード**のうち、頻出なものを抽出する手法である。

Tatti [5] は、エピソードの部分クラスである**厳密エピソード**に対して、**ランク**の定義を与えている。ランクは、イベント列の集合と厳密エピソードが与えられた場合に、イベント列の集合におけるエピソードの出現確率と、実際のエピソードの出現数から、そのエピソードの有意性を測る指標である。一方で、Tatti [5] は厳密エピソードを列挙する手法は与えていない。

Katoh ら [3] は、厳密エピソードの部分クラスである**菱形エピソード**に対して、イベント列からすべての頻出菱形エピソードを効率的に抽出するアルゴリズム PolyFreqDmd を提案している。本稿では、イベント列の集合からすべての頻出菱形エピソードを、Katoh ら [3] の手法にしたがって列挙すると同時に、解となる頻出菱形エピソードのランクを Tatti [5] の手法に基づいて計算するアルゴリズム MineRankedDmd を与える。最後に、人工的に生成したイベント列にアルゴリズム MineRankedDmd を適用することで、アルゴリズムの性能を評価する。

2. 準備

本節では、Tatti [5] にしたがって、厳密エピソードと呼ばれるエピソードの部分族におけるランク計算の枠組みを導入する。整数 $i \leq j$ に対して、$[i..j] = \{i, i+1, \ldots, j\}$ は i から j までの整数区間を表す。

2.1 イベント列とエピソード

アルファベット Σ に対して、その要素 $e \in \Sigma$ を**イベント**という。Σ 上の長さ k の**イベント列**は、$S = \langle s_1, \ldots, s_k \rangle \in \Sigma^*$ である。ここに、$s_i \in \Sigma$ はイベントである。アルゴリズムの入力は、イベント列集合 $\mathcal{S} = \{S_1, \ldots, S_{|\mathcal{S}|}\} \subseteq \Sigma^*$ である。*1

エピソード [4] は、頂点ラベルつき非巡回有向グラフ $G = (V_G, E_G, lab_G)$ である。ここに、V_G は頂点集合、E_G は辺集合、lab_G はラベル関数 $lab_G : V_G \to \Sigma$ である。エピソードは、出現の半順序関係が定義されたイベントの多重集合とみなせる。

エピソード G がイベント列 $S = \langle s_1, \ldots, s_n \rangle$ に**出現する**ことを $S \succeq G$ で表し、次の条件を満たす関数（マッチング関数と呼ぶ） $f : V_G \to [1..n]$ が存在することと定める：(i) f は一対一である: $(\forall u, v \in V_G)\ f(u) = f(v) \ \Rightarrow\ u = v$; (ii) f は順序を保存する: $(\forall u, v \in V_G)\ (u, v) \in E_G \ \Rightarrow\ f(u) < f(v)$; (iii) f は頂点ラベルを保存する: $(\forall v \in V_G)\ lab_G(v) = s_{f(v)}$. このとき、イベント列 S はエピソード G を**カバーする**ともいう。

イベント列の集合 $\mathcal{S}$ 上で、エピソード G を**カバーするイベント列の集合**を $\mathbf{C}_{\mathcal{S}}(G) = \{\, S \in \mathcal{S} \mid S \succeq G \,\}$ と書く。このとき、G の**観測支持度**（単に**支持度**とも呼ぶ）を $sup(G \mid \mathcal{S}) = |\mathbf{C}_{\mathcal{S}}(G)|$ と定義する。さらに、エピソード G と**最小頻度** $\alpha > 0$ に対して、$sup(G \mid \mathcal{S}) \geq \alpha$ となるとき、G は**頻出である**という。本稿では、$\mathbf{C}$ の添字 $\mathcal{S}$ は自明な場合省略する。

2.2 エピソードの部分族

すべての要素間に順序関係が定義されている場合、そのエピソードは**直列である**という [4]。イベントの集合 E とイベント a, b に対して、$a \mapsto E \mapsto b$ を**菱形エピソード** [3] という。これは、E 中のすべてのイベントは a の後、b の前に出現することを意味する。以降、a と b を先頭と末尾と呼び、E を本体と呼ぶ。単に $a \mapsto b$ で、b が a の後に出現することを表す。

一般に与えられたイベント列 S とエピソード G に対して、カバー $S \succeq G$ の判定は NP 困難である [5]。この困難性は主にマッチング関数の一対一性に起因する。そのため、Tatti [5] は、次の厳密エピソードを導入した。*2

連絡先: 谷 陽太，北海道大学大学院 情報科学研究科，〒 060-0814 札幌市北区北 14 条西 9 丁目，TEL/FAX: 011-706-7680，Email: tani@ist.hokudai.ac.jp

*1 Mannila ら [4] は、1 つのイベント系列 S と正整数 k を入力とし、長さ k の滑り窓モデルでのエピソード発見を議論している。Mannila らの枠組みは、S 中の長さ k のすべての部分文字列からなる系列集合を考えることで、本稿の枠組みで扱える。

*2 Katoh と Hirata [2] は、Tatti [5] と独立に厳密エピソードの族を提案し、線形エピソードによるそれらの特徴付けを与えている。

定義 1 (厳密 [5])**.** エピソード $G = (V_G, E_G, lab_G)$ が**厳密**であるとは、任意の異なる頂点 u と $v \in V_G$ に対して、$lab_G(u) = lab_G(v)$ ならば、u から v への有向パス、または v から u への有向パスが存在することをいう。

Tatti [5] の定義では、エピソードの DAG の推移閉包を考えて、二頂点が同じラベルをもつならば両者をつなぐ有向辺があることと定義している。これは本稿の定義と同値である。上記の直列エピソードと菱形エピソードは、いずれも厳密エピソードである。*3

2.3 エピソードのカバー確率とランク

本稿では、イベント列のランダム生成モデルとして、イベントの生起確率のベクトル $\mathbf{q} = (q_e)_{e\in\Sigma} \in [0,1]^{|\Sigma|}$ で定められる記憶のない情報源を仮定する。ここに、$\sum_e q_e = 1$ である。

エピソード G とイベント列生成モデル $\mathbf{q}$ に対して、G **のカバー確率**とは、$\mathbf{q}$ にしたがって生成された長さ k のランダムなイベント列 X が G をカバーする確率

$$p_k = p(G \mid \mathbf{q}, k) = p(X \succeq G \mid |X| = k, X \sim \mathbf{q})$$

である。このとき、G の**期待支持度**は $\mu = \sum_{S\in\mathcal{S}} p_{|S|}$ となる。

イベント列の集合 $\mathcal{S}$ に対して、$|X_i| = |S_i| (1 \le i \le |\mathcal{S}|)$ となるようなランダムなイベント列の集合 $\mathcal{X}$ を考える。非負整数 $n \ge 0$ に対して $sup(G \mid \mathcal{X}) = n$ となる確率は、

$$p(sup(G \mid \mathcal{X}) = n \mid \mathcal{X} \sim \mathbf{q}) = \sum_{\substack{\mathcal{T}\subseteq\mathcal{S} \\ |\mathcal{T}|=n}} \prod_{S\in\mathcal{T}} p_{|S|} \prod_{S\in\mathcal{S}\setminus\mathcal{T}} (1 - p_{|S|})$$

である。この値は n が十分に大きいとき、平均 μ で分散 $\sigma^2 = \sum_{S\in\mathcal{S}} p_{|S|}(1 - p_{|S|})$ の正規分布 $N(\mu, \sigma)$ から推定できる。以降、この値を求める場合、正規分布から推定することとする。

Tatti [5] は、イベント列の集合 $\mathcal{S}$ とイベント列生成モデル $\mathbf{q}$ において、エピソード G の**有意度ランク** (単に**ランク**とも呼ぶ) を次のように定義している。

定義 2 (有意度ランク [5])**.** イベント列の集合 $\mathcal{S}$ とイベント列生成モデル $\mathbf{q}$ において、エピソード G のランクとは、

$$\begin{aligned} r(G \mid \mathcal{S} \sim \mathbf{q}) &= -\log p\big(sup(G \mid \mathcal{X}) \ge n \mid X \sim \mathbf{q}\big) \\ &= -\log\Big(1 - \sum_{k=1}^{n-1} p\big(sup(G \mid \mathcal{X}) = k \mid X \sim \mathbf{q}\big)\Big) \end{aligned}$$

である。ここで、$0 \le r(G \mid \mathcal{S} \sim \mathbf{q}) \le \infty$ である。

2.4 エピソードからの有限オートマトンの構築

Tatti [5] は、イベント列がエピソード G をカバーするかを、有限オートマトンを用いて判定する方法を提案している。さらに、この有限オートマトンをカバー確率と G のランクの計算にも用いる。

ここに、エピソードは有向グラフであることに注意されたい。エピソード $G = (V, E)$ に対して、部分エピソード $H = (W, F)$ が G の**接頭辞部分グラフ**であるとは、それが条件 (i) $v \in W$ かつ $(w, v) \in E$ ならば $w \in W$、かつ (ii) $v, w \in W$ かつ $(v, w) \in E$ ならば $(v, w) \in F$ を満たすことをいう。G **の接頭辞部分グラフ全体の集合**を $Pre(G)$ で表す。例として、$G = a \mapsto \{b, c\} \mapsto d$ のとき、$Pre(G) = \{(), (a), (a \mapsto b), (a \mapsto c), (a \mapsto \{b, c\}), (a \mapsto \{b, c\} \mapsto d)\}$ である。

エピソード $G = (V, E, lab)$ に対して、次のように G **の有限オートマトン** $A(G) = (Pre(G), \Sigma, \delta, H_0, H_f)$ を構築する [5]。

1. 状態集合は、すべての接頭辞部分グラフのなす集合 $Pre(G)$ である。初期状態は空な頂点集合の誘導するグラフ $H_0 = (\emptyset, \emptyset, \emptyset)$ とし、受理状態は $H_f = G$ である。
2. 状態 $H_1, H_2 \in Pre(G)$ に対し、H_2 から頂点を 1 つ取り除くことで H_1 が導かれる場合、遷移関係 $\delta \subseteq Pre(G) \times \Sigma \times Pre(G)$ は、取り除いた頂点のラベル $e \in \Sigma$ をもつ辺 (H_1, e, H_2) をもつ。
3. 各状態 $H \in Pref(G)$ と任意の $e \in \Sigma$ に対して、e が H から出るいかなる有向辺にもラベルとして出現しない場合、ラベル e をもつ自己ループ辺 (H, e, H) を追加する。

以上の手順で構築した有限オートマトン $A(G)$ の領域は、G のサイズの指数関数である。Tatti [5] は、次の補題を示した。

補題 1. 任意のイベント系列 S と、任意のエピソード G、その有限オートマトン $A(G)$ に対して、S がエピソード G をカバーすることと、$A(G)$ が S を受理することは同値である。

補題 2. G が厳密エピソードならば、その有限オートマトン $A(G)$ は決定性有限オートマトンである。

2.5 カバー確率の計算

Tatti [5] は、2.4 節で導入した有限オートマトンを用いて、与えられた G と、各文字の生起確率 $\mathbf{q}$, 正整数 $k \ge 0$ に対して、独立生成モデルにおける長さ k のランダム文字列による G のカバー確率を効率良く計算する方法を与えている。

$A(G) = (Pre(G), \Sigma, \delta, H_0, H_f)$ の構成より、イベント列 S が $A(G)$ に受理されるならば、かつその場合に限り、S は G をカバーする。よって、イベント列の集合 $\mathcal{S}$ において、$A(G)$ に受理される $S \in \mathcal{S}$ の数が G の観測支持度となる。また、長さ k のイベント列が $A(G)$ の受理状態に到達できる確率を求めることで、カバー確率 p_k が計算できる。

イベント列の生起確率のベクトル $\mathbf{q} = (q_e)_{e\in\Sigma} \in [0,1]^{|\Sigma|}$ と有限オートマトン $A(G)$ において、$\mathbf{q}$ から生成された長さ k のランダムイベント列にしたがい、$A(G)$ を初期状態から遷移し

Algorithm 1 CalcProb($A = (Pre, \Sigma, \delta, H_o, H_f), \mathbf{q} = (q_e)_{e\in\Sigma}, n$)

Input: 有限オートマトン A,
イベント列の生起確率のベクトル $\mathbf{q}$,
非負整数 n.
Output: $p_k (k = 1, \ldots, n)$ の列.

```
p(H_0, 0) := 1;
p(H_0, 0) 以外の p(H, i) := 0 (i := 0 → n)
for all i := 1 → n do
  for all H ∈ Pre do
    p(H, i) := 1;
    for all d ∈ δ such that d(f, e) = H do
      p(H, i) := p(H, i) − q_e;
    end for
    p(H, i) := p(H, i) × p(H, i − 1);
    for all d ∈ δ such that d(H, e) = f do
      p(H, i) := p(H, i) + q_e × p(f, i − 1);
    end for
  end for
end for
output ⟨p(H_f, 1), ..., p(H_f, n)⟩;
```

---

*3 菱形エピソード $a \mapsto E \mapsto b$ は、定義より本体 E が集合であり、同じ要素を重複して含まないため、厳密エピソードとなる。

Algorithm 2 MineRankedDmd($\mathcal{S}, \mathbf{q}, \Sigma, \alpha$)

Input: イベント列の集合 $\mathcal{S}$,
イベント列の生起確率のベクトル $\mathbf{q}$,
アルファベット Σ,
最小頻度 $0 < \alpha \leq |\mathcal{S}|$.
Output: 頻出菱形エピソードとそのランク.

```
1: Σ0 := 頻出するイベントの集合 (Σ0 ⊆ Σ);
2: for all ( a ∈ Σ0 ) do
3:    CaclProb(A(a), q, max_{S∈S} |S|) でカバー確率を計算;
4:    output ⟨a, a のランク⟩;
5:    for all ( b ∈ Σ0 ) do
6:       G0 := (a ↦ ∅ ↦ b);
7:       A0 := G0から構築した有限オートマトン
8:       C0 := C_S(G0);
9:       RecMineRankedDmd(G0, C0, A0, S, q, Σ0, α);
10:   end for
11: end for
```

Algorithm 3 RecMineRankedDmd($G = (a \mapsto E \mapsto b), C, A, \mathcal{S}, \mathbf{q}, \Sigma, \alpha$)

Output: $a \mapsto E \mapsto b$ ($E \subseteq \Sigma$) という形の頻出菱形エピソードすべての集合.

```
1: if ( |C| ≥ α ) then
2:    CaclProb(A, q, max_{S∈C} |S|) でカバー確率を計算;
3:    output ⟨G, G のランク⟩;
4:    for all ( e ∈ Σ (e > max(E)) ) do
5:       G' := a ↦ (E ∪ {e}) ↦ b;
6:       B := A から初期状態を取り除いた有向グラフ ;
7:       B' := A から受理状態を取り除いた有向グラフ ;
8:       A' := B と B' の対応する頂点間に
               ラベル e を持つ辺を加えた有限オートマトン ;
9:       C' := C_C(G');   //カバーされるイベント系列集合
10:      RecMineRankedDmd(G', C', A', S, q, Σ, α)
11:   end for
12: end if
```

た結果の状態が $H \in Pre(G)$ である確率を $p(H,k)$ とする。このとき、$p(H,k)$ は以下の式で求められる [5]。

$$\begin{aligned} p(H,k) &= \big(1 - \sum_{(H,e,F)\in\delta} (q_e)\big) p(H,k-1) \\ &\quad + \sum_{(F,e,H)\in\delta} (q_e) p(F,k-1) \end{aligned}$$

カバー確率 p_k は $p(H_f,k)$ となる。この値は、動的計画法を用いて $\mathcal{O}(v^2 k)$ 時間で計算できる。ここに、$v = |Pre(G)|$ である。Algorithm 1 に、カバー確率 p_k の列 $(k = 1, \ldots, n)$ を計算するアルゴリズム CalcProb を示す。

3. 基本アルゴリズム

本節では、イベント列の集合からすべての頻出菱形エピソードとそのランクを重複なく出力するアルゴリズム MineRanked-Dmd を示す。$\mathcal{S}$ をイベント列の集合とし、$\alpha > 0$ を最小頻度とする。

3.1 概要

Algorithm 2 にアルゴリズム MineRankedDmd の概要を示す。MineRankedDmd では、はじめに初期エピソードを計算した後、再帰関数 RecMineRankedDmd を呼び出す。Algorithm 3 に、アルゴリズム RecMineRankedDmd の概要を示す。RecMineRankedDmd は、初期エピソードから開始して、自身を再帰的に呼び出しながら、深さ優先探索により、初期解を拡大して得られるすべての菱形エピソードを重複なく生成する。再帰の各繰り返しで、RecMineRankedDmd は受け取った本体のサイズが m の菱形エピソード（親エピソード）を拡張して、サイズ $m+1$ の菱形エピソード（子エピソード）を生成する。その後、観測支持度を求め、対応する有限オートマトン $A(G)$ を構築する。これを用いてランクを計算する。

3.2 初期エピソードの計算

アルゴリズム MineRankedDmd は、はじめに最もサイズの小さい解として、すべてのイベントの組 $(a,b) \in \Sigma$ に対して、最小の菱形エピソード $G_{ab} = (a \mapsto \emptyset \mapsto b)$ を生成する。さらに、G_{ab} と、G_{ab} をカバーするイベント列の集合 $C = \mathbf{C}(G_{ab})$、および G_{ab} から構築した有限オートマトン $A(G_{ab})$ を引数とし、再帰関数 RecMineRankedDmd を呼び出す。

3.3 エピソードの更新

このステップでは、Katoh ら [3] が提案した菱形エピソードの列挙手法を用いて、再帰的に菱形エピソードを生成する。

基本的なアイディアとして、親エピソード G に新しいイベントを追加して、サイズが一つ大きな子エピソードを生成する（エピソードの拡張）。しかし、このときに単に本体に新しいイベントを追加するだけでは、同じエピソードを異なる経路で複数回生成してしまうことになる。

RecMineRankedDmd では、指数サイズの中間メモリを用いることなく、すべて菱形エピソードを深さ優先探索する。親である菱形エピソード $G = (a \mapsto E \mapsto b)$ が与えられとき、$e > max(E)$ となる任意のイベント $e \in \Sigma$ に対して、本体 E に e を加えて得られる菱形エピソード $G' = (a \mapsto (E \cup \{e\}) \mapsto b)$ を子エピソードとして生成する。

3.4 有限オートマトンの更新

上記のエピソードの更新に伴う、決定性有限オートマトン $A(G)$ の更新は以下の手順で行う。

1. $A(G)$ のコピーを作成し、$A'(G)$ とする。
2. $A(G)$ の受理状態と $A'(G)$ の初期状態を削除する。
3. $A(G)$ 中の初期状態以外の任意の頂点を H とし、H に対応する $A'(G)$ 中の頂点を H' としたとき、ラベル e をもつ辺 (H, H') を追加する。

生成した 1 つの有限オートマトンを更新後の $A(G)$ とする。再帰が 1 つ深くなるごとに、構築する有限オートマトン $A(G)$ のサイズは倍になる。すなわち、エピソード本体 E のサイズが m ならば、$A(G)$ は $2^m + 2$ の状態をもつ。

3.5 カバー確率とランクの計算

上記で更新した決定性オートマトン $A(G)$ を用いて、補題 1 より、G をカバーするイベント列の集合 $\mathbf{C}(G)$ を計算できる。さらに、2.5 節に説明した方法を用いて、G のカバー確率を計算し、G のランクを求める。

以上の手続きを再帰的に繰り返すことで、任意の最小頻度 α に対して、アルゴリズム MineRankedDmd は、イベント列の集合 $\mathcal{S}$ から頻出菱形エピソードとそのランクを重複なく出

力する。生成する各オートマトンのサイズは対応する菱形エピソード $a \mapsto E \mapsto b$ の本体サイズ $|E|$ に指数的であるため、解 1 つあたりの計算時間は $|\Sigma|$ に対して指数的になる。

4. 実験

本節では、3. 節で提案したアルゴリズム MineRankedDmd の評価実験の結果を示す。

4.1 データ

データとして、次のようにランダム生成したイベント列集合を用いた。イベント列集合のサイズ $|\mathcal{S}|$ を 200 から 2000 まで 200 ずつ変化させた 10 種類のデータセットを生成した。アルファベットはサイズ 20 の $\Sigma = \{a, b, \ldots t\}$ であり、イベント列のサイズは $|S_i| = 20$（$i = 1, \ldots, |\mathcal{S}|$）である。さらに、生成したイベント列の約 5% の割合で、菱形エピソード $G_1 = a \mapsto \{b, c, d\} \mapsto e$ を、長さ 5 の部分イベント列として意図的に埋め込んだ。菱形エピソード G_1 が埋め込まれた部分以外のイベントは、Σ から一様ランダムに生成した。

4.2 方法

3. 節で提案したアルゴリズム MineRankedDmd を C++言語で実装し、各データセットに適用した。最小頻度は $\alpha = 100$ とした。アルゴリズムが仮定するイベント列の生成モデルは、入力データセット上の各イベントの出現割合を用いた。

実行環境として、PC(Intel Core i5 2.9GHz, 8GB memory, MacOSX 10.11.6) と g++コンパイラ (Apple LLVM ver8.0.0) を用いた。実験は 5 回繰り返して行い、計算時間を time コマンドで計測した。

4.3 結果

図 1 に、イベント列集合のサイズにおけるアルゴリズム MineRankedDmd の実行時間の平均と、出力解の個数を示す。図 1 より、MineRankedDmd は、解の個数の増加にしたがって、計算時間が増加することがわかる。

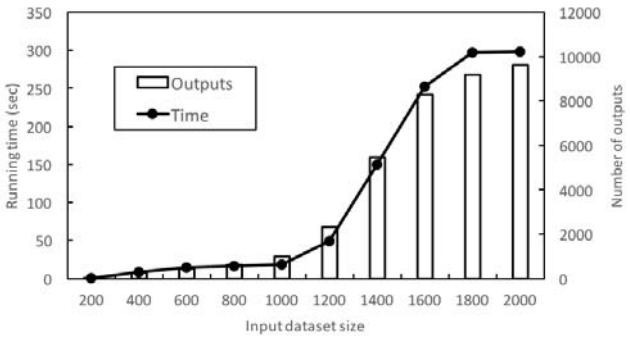


図 1: MineRankedDmd の実行時間と解の個数

$|\mathcal{S}| = 1000$ のデータセットにおいて、解の数は 1011 だった。表 1 に、$|\mathcal{S}| = 1000$ のデータセットにおける有意度ランク上位 5 つのエピソードと、そのランク及び観測支持度を示す。表 1 より、意図的に埋め込んだエピソード G_1 が高いランクを示したことがわかる。表 2 に、$|\mathcal{S}| = 1000$ のデータセットにおける観測支持度上位 5 つのエピソードと、そのランク及び観測支持度を示す。表 2 より、サイズの小さいエピソードが高い観測支持度を示したことがわかる。

表 1: 有意度ランク上位 5 つの解のランクと観測支持度

G	$r(G \mid \mathcal{S} \sim M)$	順位	$sup(G \mid \mathcal{S})$	順位
$a \mapsto \{b, c, d\} \mapsto e$	∞	1	228	409
$a \mapsto \{b, d\} \mapsto e$	513.640	2	245	314
$a \mapsto \{b, c\} \mapsto e$	490.149	3	246	305
$a \mapsto \{c, d\} \mapsto e$	443.951	4	241	338
$a \mapsto \{c, d, e\} \mapsto b$	190.435	5	104	926

表 2: 観測支持度上位 5 つの解のランクと観測支持度

G	$r(G \mid \mathcal{S} \sim M)$	順位	$sup(G \mid \mathcal{S})$	順位
c	1.8377	558	737	1
a	2.9334	492	722	2
d	2.3886	526	721	3
b	4.7831	383	719	4
e	1.5061	638	702	5

5. おわりに

本稿では、Tatti [5] の厳密エピソードのランク計算手法を用いて、イベント列からすべての頻出菱形エピソードを発見しながら、同時にそのランクを計算するアルゴリズム MineRankedDmd を提案した。

今後の課題として、カバー確率の計算を、親エピソードでの計算結果から効率的に計算する手法の開発が挙げられる。アルゴリズム MineRankedDmd では、エピソードの拡張と有限オートマトンの更新については、親エピソードでの計算結果を用いて効率的に行う一方で、カバー確率の計算ははじめから再計算している。この計算の効率化は重要な課題である。加えて、MineRankedDmd を一般のエピソードに拡張することと、実データに対してアルゴリズムを適用し、性能を評価することも今後の課題である。

参考文献

[1] Rakesh Agrawal, Ramakrishnan Srikant. “Mining sequential patterns.” Proc. 11th ICDE, 3-14, 1995.

[2] Takashi Katoh, Kouichi Hirata. “A Simple Characterization on Serially Constructible Episodes.” Proc. PAKDD 2008, 600-607, 2011.

[3] Takashi Katoh, Hiroki Arimura, Kouichi Hirata. “A Polynomial-Delay Polynomial-Space Algorithm for Extracting Frequent Diamond Episodes from Event Sequences.” PAKDD 2009: 172-183

[4] Heikki Mannila, Hannu Toivonen, A. Inkeri Verkamo. “Discovery of Frequent Episodes in Event Sequences.” Data Min. Knowl. Discov. 1(3): 259-289 (1997)

[5] Nikolaj Tatti. “Ranking episodes using a partition model.” Data Min. Knowl. Discov. 29(5): 1312-1342 (2015)

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

一般公開版「都道府県議会会議録検索システム」の概要

Public Release of a Web-based System for Visualizing Prefectural Assembly Minutes in Japan

乙武北斗[*1] Hokuto Ototake
高丸圭一[*2] Keiichi Takamaru
内田ゆず[*3] Yuzu Uchida
木村泰知[*4] Yasutomo Kimura

[*1] 福岡大学 Fukuoka University
[*2] 宇都宮共和大学 Utsunomiya Kyowa University
[*3] 北海学園大学 Hokkai-gakuen University
[*4] 小樽商科大学／AIP Otaru University of Commerce / AIP

This paper illustrates a web-based system for visualizing prefectural assembly minutes in Japan. It was public-released in February 2018. Although many local governments in Japan provide public access to political documents, researchers and scientists cannot take full advantage of them because they consist only of raw text. Therefore, there is a need for these to be converted into panel data and to visualize differences by region or over time. For ordinary citizens, visualizing differences between the topics discussed by assembly members can be useful in assessing their characteristics. The proposed system aims to provide a complete solution that meets all of these requirements. The system has five main functions: i) full-text search for political documents, ii) showing keywords in context, iii) map and time-series visualization, iv) cross-tabulation, and v) political keyword extraction. Although three existing systems offer some of these functions, the proposed system integrates all the five necessary functions and provides a complete solution that meets all of the requirements.

1. はじめに

筆者らは科研費研究課題「地方議会会議録コーパスの構築とその学際的応用研究」(2010～2013 年度)において，全国の自治体の地方議会会議録の収集・整理を進め，423 自治体の議会会議録を収録した「地方議会会議録コーパス」(約 1 億 3400 万文，約 80GB)を構築した(第Ⅰ期コーパス)．これを活用し，情報工学の分野における情報抽出手法の開発[木村 2013], [葦原 2015], [木村 2011]や，言語学分野における言語使用や言語変異の実証的研究[高丸 2015], [井上 2013], [二階堂 2015]が進められてきた．第Ⅰ期コーパスは情報工学分野，言語学分野においては，有用な大規模言語資源としての利用可能であることが明らかとなった．しかし，収集対象の自治体の範囲や期間が統制されていないコーパスであるため，地域間の比較を統計的に行うような研究には活用しづらいという問題があった．

現在，科研費研究課題「議論の背景・過程・結果を関連づける地方政治コーパスの構築とその学際的応用」(2016～2019 年度)において，新たに地方政治における議論の過程と結果，およびその背景を結びつけた「地方政治コーパス」の構築を進めている(第Ⅱ期コーパス)．このコーパスでは，比較分析を定量的に行うことができるように，地方議会会議録の収集条件(対象期間・対象自治体)を整えたコーパスの整備を行っている．新たに構築する会議録コーパスは，データ規模は小さくなるものの，時系列(時間に伴う変化)，および，横断面(ある時点の地域による変異)の分析が可能なものとなる．現在，第Ⅱ期コーパスの収集が進み，これを利用した統計的分析等が進められつつある．

収集したデータを学術研究だけで使用するのではなく，一般の市民が関心を持ったことばを検索できるようにすることで，地方政治への関心をさらに高めることができ，筆者らとしても，研究成果の社会還元を行うことができる．このため，収集した会議録の全文検索および検索結果の可視化が可能な Web システムの実装を進め，第Ⅱ期コーパスの一部を「都道府県議会会議録検索システム」(http://local-politics.jp/47pref/)として一般公開することとした．本稿では，2018 年 2 月に一般公開したシステムの概要について述べる．収集データについては 2 章で述べる．システムの概要について 3 章で述べる．検索結果の表示機能の特徴について 4 章に述べる．

連絡先：乙武北斗，福岡大学工学部，福岡県福岡市城南区七隈 8-19-1，ototake@fukuoka-u.ac.jp

2. 収録データ

2.1 データの収集と整理

「都道府県議会会議録検索システム」には，平成 23 年 4 月の統一地方選挙から 4 年分(平成 27 年 3 月まで)の全国 47 都道府県議会の本会議会議録が収録されている．

会議録は句点を基準に 1 発言ごとに分割し，それぞれの発言の発言者名を抽出した上で，会議の開催情報と併せて，表 1 に示す 17 項目のフィールドをもつテーブルに登録した．

表 1 データベースのフィールド

1.	**識別子**(主キー，発言文に固有に振られた ID)
2.	**自治体名**(「北海道」「青森県」「宮城県」…)
3.	**会議名**(「平成24年2月定例会」「第335回」など)
4.	**号**(各会議の会期における号数(日数))
5.	**年**(開催年(和暦))
6.	**月**(開催月)
7.	**日**(開催日)
8.	**会議種別**(定例／臨時)
9.	**会議名フルテキスト** (例：「平成24年　2月　定例会(第335回)-02 月 27 日－04 号」)
10.	**発言者フルテキスト**(「62 番岩本剛人君」)
11.	**発言者 ID**(発言者テーブルを参照する外部キー)
12.	**発言者名**(例：「岩本剛人」)
13.	**発言者の役職**(例：「62 番」)
14.	**発言文**(例：「次に，冬の節電対策について伺います。」)
15.	**発言以外の記録文**(「(拍手)」「[…君登壇]」「－－－－－」など)
16.	**原本 URL**
17.	**HTML ファイルのパス**

収集した会議録から，会議名や発言者名の抽出を行った．さらに，発言者を同定するために人手によって，名寄せを行った．特に，議員については発言者を識別する「発言者 ID」を付し，

議員の属性情報を格納したテーブルを外部参照できるようにした.（詳しくは[高丸 2017]参照）

なお，現在のところ，発言文と発言以外の記録文の分割は行っていない．このため，発言以外の記録文はすべて直前の発言者の発言の一部として，発言文に格納されている．この点は今後の課題である．

表 2　都道府県別の収録データ数

都道府県	発言レコード数	発言文字数	発言者数
北海道	115,091	6,690,886	142
青森	96,095	6,075,997	123
岩手	83,436	5,304,262	120
宮城	123,815	6,591,736	113
秋田	57,635	3,564,034	65
山形	33,075	2,219,422	98
福島	69,741	4,106,475	313
茨城	70,492	4,387,984	210
栃木	81,452	4,086,005	101
群馬	104,851	5,664,259	134
埼玉	139,901	6,216,179	189
千葉	133,345	6,966,264	150
東京	115,750	5,746,805	240
神奈川	101,895	5,991,067	148
新潟	90,476	5,008,553	104
富山	74,132	4,688,790	86
石川	72,677	4,413,817	110
福井	87,817	4,568,260	96
山梨	79,247	4,274,357	119
長野	136,055	7,597,746	133
岐阜	67,610	4,283,206	100
静岡	91,526	5,456,750	136
愛知	88,416	5,881,919	179
三重	75,893	5,047,765	123
滋賀	142,596	8,646,095	88
京都	72,824	5,031,745	166
大阪	95,929	4,645,928	181
兵庫	74,480	4,223,394	138
奈良	71,490	4,193,615	114
和歌山	62,970	3,473,289	111
鳥取	172,523	10,844,069	101
島根	106,708	6,010,462	95
岡山	111,268	6,392,762	111
広島	54,298	3,357,243	113
山口	82,116	4,995,372	95
徳島	69,015	3,797,876	94
香川	68,475	4,029,591	105
愛媛	69,609	3,999,989	108
高知	83,058	5,859,125	102
福岡	84,918	4,933,062	106
佐賀	97,474	5,760,078	92
長崎	135,966	6,585,947	112
熊本	82,164	4,771,578	123
大分	94,297	4,666,031	112
宮崎	103,344	6,546,640	88
鹿児島	113,984	7,266,893	100
沖縄	155,947	7,501,424	119
合計	**1,651,676**	**92,290,795**	**2,542**

2.2 収録データ数

表 1 に都道府県別の発言レコード数，発言文字数（発言文フィールドの文字数），発言者数を示す．総収録データは 1,651,676 レコードであり，総文字数は 92,290,795 文字である．また，名寄せの結果得られた発言者数は合計 2,514 名であった．

3. システム

3.1 概要

本システムは，2 章で述べた 47 都道府県議会の本会議会議録を検索・視覚化するウェブアプリケーションである．ユーザは一般的なウェブブラウザを用いて，会議録を検索し，結果を視覚化することができる．システムのアウトラインを図 1 に示す．①システムが検索クエリを受け取ると，②クエリは全文検索エンジンに送られる．③検索エンジンから検索結果を受け取った後，④システムはユーザのリクエストに応じて検索結果を可視化して出力する．全文検索エンジンには Elasticsearch[1]を用いた．

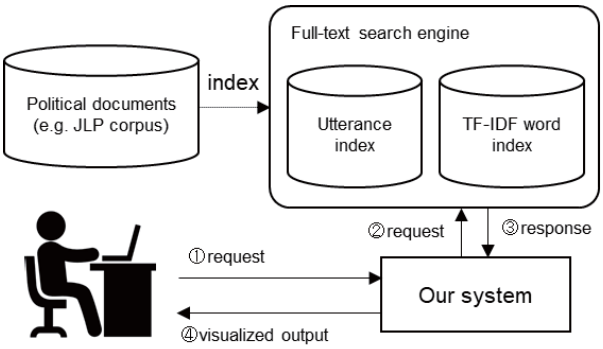


図 1　システムのアウトライン

図 2　システムのメニュー画面

一般公開システム[2]のトップページの様子を図 2 に示す．はじめて訪れたユーザが試用しやすいように，検索文字列に「オリンピック」をデフォルト値として入力している．右側の「検索」ボタン，または，下部のアイコンをクリックすることにより，検索を開始することができる．また，上部には，「発言検索」「マップ検索」「クロス

[1] https://www.elastic.co/
[2] http://local-politics.jp/47pref/

表検索」「時系列グラフ」統計」の 5 つのメニューがあり，ここから本システムが有するすべての機能へアクセスすることができる．

「発言検索」メニューには「発言検索」「KWIC」「TF･IDF キーワード検索」のサブ項目がある．「マップ検索」メニューには「発言マップ検索」「TF･IDF マップ検索」のサブ項目がある．各機能の概要は次節以降に述べる．

「統計」メニューをクリックすると，本システムが提供するデータに関する以下の統計情報を表示することができる．

- 都道府県発言･文字数
- 都道府県発言･文字数（年度別）
- TF･IDF 上位 100（自治体･発言者単位）
- TF･IDF 上位 100（自治体の発言者平均）

4. 機能

4.1 全文検索（発言検索）

「全文検索」機能は，検索文字列を含む発言文（レコード）を抽出する．図 3 に全文検索の例を示す．「検索文字列の扱い」で「N-gram 分割」を選択すると，検索文字列を 2-gram および 3-gram に分割し，これらを含む検索結果を表示する．「フレーズ完全一致」を選択すると，検索文字列全体をそのまま含む検索結果を表示する．「複数検索語の扱い」では，空白によって区切られた複数の検索文字列を AND 検索するか，OR 検索するかを選択することができる．「出力結果のソート」では，結果の表示順を「スコア降順」，「日付降順（昇順）」，「自治体コード降順（昇順）」，「発言文字数降順（昇順）」から選択することができる．さらに「詳細な条件を設定」をチェックすることで，検索対象の自治体および発言時期を限定した検索を行うことができる．

図 3　全文検索の例

(1)　文脈の展開機能

検索結果として表示される情報は，議会における発言の一部分（1 文）である．その発言の前後にどのような発言があるかを知りたいという要望に答えるために，全文検索の検索結果には「前 5 発言を取得する」「後 5 発言を取得する」ボタンを用意した．これをクリックすることで，5 文単位で発言を遡ったり，発言を辿ったりすることができる．

(2)　KWIC 表示

また，上部のメニューから「発言検索→KWIC」を選択すると，検索結果を KWIC 形式[Luhn 1960]で表示し，前文脈と後文脈を一覧することができる．この例を図 4 に示す．

処理時間: 54 ms　ヒット件数: 11,095

#	属性	前文脈	キーワード	後文脈
1.1	長崎県 24	ロンドン	オリンピック	ならぬ長崎オリンピック。
1.2	長崎県 24	ロンドンオリンピックならぬ長崎	オリンピック	。
2.1	新潟県 25	東京	オリンピック	、東京都オリンピックという議論もあります。
2.2	新潟県 25	東京オリンピック、東京都	オリンピック	という議論もあります。
3.1	東京都 24	ロンドン・	オリンピック	のことを何か失敗したかのようにおっしゃっていたが、ロンドン・
3.2	東京都 24	のことを何か失敗したかのようにおっしゃっていたが、ロンドン・	オリンピック	は成功したオリンピックです。
3.3	東京都 24	のようにおっしゃっていたが、ロンドン・オリンピックは成功した	オリンピック	です。
4.1	静岡県 25	まず要望ですが、	オリンピック	・パラリンピックに関連したものでありますが、東京オリンピック
4.2	静岡県 25	オリンピック・パラリンピックに関連したものでありますが、東京	オリンピック	は夏季オリンピックでありますね。
4.3	静岡県 25	リンピックに関連したものでありますが、東京オリンピックは夏季	オリンピック	でありますね。

図 4　KWIC 表示の例

4.2 マップ検索

「マップ検索」機能は，検索結果の件数を都道府県別に集計し，地図上で塗りつぶしの濃度によって可視化する機能である．地図をクリックすることで，当該地域での発言を抽出して，全文検索表示することができる．マップ検索の例を図 5 に示す．検索文字列は「離合」である．

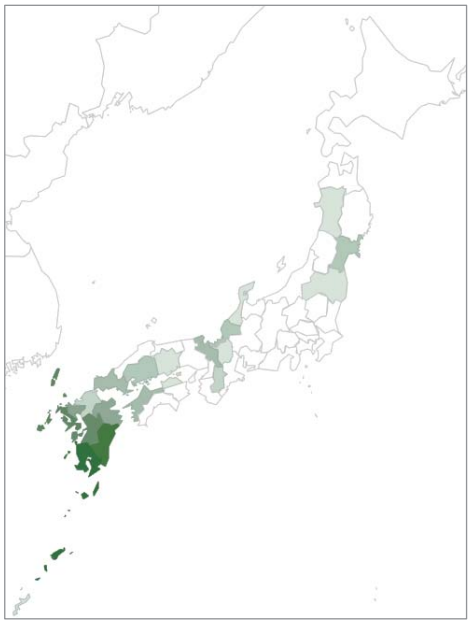


図 5　マップ検索の例

4.3 クロス表検索

「クロス表検索」機能は，「行要素」および「列要素」にそれぞれ項目名を指定することで，検索結果の件数のクロス集計結果を表形式で表示する機能である．行要素・列要素として指定できる項目は，発言地域「都道府県」，発言時期「年度」または「四半期」，発言者「発言者名」または「議員生年年代（10 年ごと）」の 5 種類である．クロス表内の数値を右クリックし「この条件で発言検索」をクリックすることで，行要素・列要素に指定した条件に合致する発言文を全文検索形式で表示できる．

クロス表検索の例を図 6 に示す．検索文字列に「オリンピック」，行要素に「都道府県」，列要素に「議員生年年代」を設定した検索結果である．

表を最大化する　すべてのセルを選択する　処理時間: 67 ms

	A	B	C	D	E	F	G
1	都道府県＼議員生年年代	1930	1940	1950	1960	1970	1980
2	北海道	0	12	34	27	27	0
3	青森	0	38	38	43	31	10
4	岩手	0	12	4	20	9	0
5	宮城	0	80	70	12	17	0
6	秋田	0	15	33	0	3	0
7	山形	0	5	32	51	1	0
8	福島	0	27	84	23	30	0
9	茨城	0	30	49	23	22	0
10	栃木	0	27	27	57	22	6
11	群馬	0	11	22	35	54	3
12	埼玉	6	63	49	51	75	16
13	千葉	0	104	69	105	47	8
14	東京	6	115	490	347	191	9
15	神奈川	9	16	88	70	83	11
16	新潟	1	25	50	53	34	0
17	富山	0	9	38	20	4	9
18	石川	0	26	30	29	6	0
19	福井	0	50	31	13	0	0
20	山梨	15	81	37	81	9	0
21	長野	1	16	88	49	35	0
22	岐阜	0	33	16	49	10	3
23	静岡	0	45	24	97	41	2
24	愛知	0	9	41	47	22	0
25	三重	0	16	26	18	21	0
26	滋賀	0	8	45	29	12	0
27	京都	0	7	15	35	14	0

図 6　クロス表検索の例

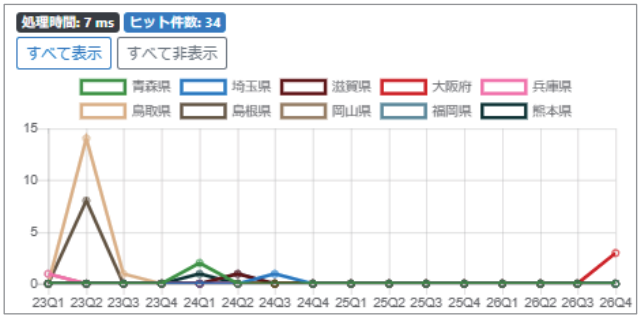


図 7　時系列グラフの例

4.4 時系列検索

「時系列検索」機能は，都道府県別に四半期単位で検索結果の件数を集計し，折れ線グラフで表示する機能である．都道府県名または「すべて表示」「すべて非表示」をクリックすることで，グラフ上に表示する都道府県を切り替えることができる．時系列検索の例を図 7 に示す．検索文字列は「生レバー」である．

4.5 TF･IDF 検索

全都道府県の議員の発言を対象に TF･IDF によって特徴語を求めた。議員特徴語を検索するために「発言検索」メニュー，「マップ検索」メニューに「TF･IDF キーワード検索」「TF･IDF マップ検索」を用意した。

「TF･IDF キーワード検索」では，検索文字列の TF･IDF 値が高い順に，都道府県名，議員名，TF･IDF 値を一覧表示する．また，「TF･IDF マップ検索」では，検索文字列の都道府県内の議員の TF･IDF 値の合計を塗りつぶしの濃度によって地図上に可視化する．

5. まとめ

地方議会会議録の横断検索を提供する Web システムは既に複数存在する（例えば「chiholog」（http://chiholog.net/）や「議事ロックス」（https://giji.rocks/））．これらのサービスでは，現在進行形で最新の地方議会会議録が収集・提供されている．しかしながら，名寄せ処理が行われていないため，検索結果が「誰の」発言であるかを条件に入れた検索や集計を行うことはできない．

筆者らが一般公開した「都道府県議会会議録検索システム」は，4 年間の限定された期間を対象とした横断検索システムであるが，「いつ」「どこで」「誰が」発言したかを検索し，検索結果を自動的に集計・可視化する機能を有する．「誰の」発言であるかが明らかであるため，TF･IDF に基づく議員別の特徴語を検索する機能を有する．本システムが，地方議会における議員の活動を明らかにし，市民の地方政治への参画意識が増進する手助けになるように，今後もデータの増強，システムの周知を進めていく予定である．

謝辞

本研究の一部は JSPS 科研費 JP16H02912，JP17K02739 による．

参考文献

[木村 2013] 木村泰知，関根聡:主辞に基づく政治問題抽出手法，人工知能学会論文誌，28(4)，370-378，2013.

[葦原 2015] 葦原史敏，木村泰知，荒木健治:地方議会会議録における節単位による議員の要望抽出，電子情報通信学会論文誌，J98-D(11)，1390-1401，2015.

[木村 2011] 木村泰知，渋木英潔，高丸圭一，乙武北斗，小林哲郎，森辰則:地方議員マッチングシステムにおける能動的質問のための質問生成手法，人工知能学会論文誌，26(5)，580-593，2011.

[高丸 2015] 高丸圭一，内田ゆず，乙武北斗，木村泰知:地方議会会議録コーパスにおけるオノマトペ―出現傾向と語義の分析―，人工知能学会論文誌，30(1)，SP2-K，306-318，2015.

[井上 2013] 井上史雄:去った〇日，ことばの散歩道，明治書院，154-155，2013.

[二階堂 2015] 二階堂整，川瀬卓，高丸圭一，田附敏尚，松田謙次郎:地方議会会議録による方言研究―セミフォーマルと気づかない方言―，方言の研究，1，299-324，2015.

[高丸 2017] 高丸圭一，内田ゆず，木村泰知:地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析，宇都宮共和大学シティライフ学論叢，18，136-155，2017.

[Luhn 1960] Luhn, H. P.:Key word-in-context index for technical literature (kwic index)，Journal of the Association for Information Science and Technology，11.4，288-295，1960.

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

会話によるニュース記事伝達のための発話意図理解

Utterance Intention Understanding for News Articles Transfer by Conversation

高津 弘明 *1 Hiroaki Takatsu　横山 勝矢 *1 Katsuya Yokoyama　本田 裕 *2 Hiroshi Honda　藤江 真也 *1*3 Shinya Fujie　林 良彦 *1 Yoshihiko Hayashi　小林 哲則 *1 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

We are developing a conversation system which efficiently transfers a massive amount of information like news articles by spoken dialogue. Here, "efficient" means that only the necessary information is transferred except unnecessary information for the user from target articles. In our system, feedbacks from the user are indispensable in order to realize high EoIT (Efficiency of Information Transfer). Therefore, we propose a utterance intention recognition method combining language information and prosodic information for the purpose of understanding diverse feedbacks from users. The feature of the proposed method is that it automatically extracts prosodic features with a high contribution ratio by using deep learning. We confirmed the effectiveness of the proposed method using a corpus with utterance intention tags designed based on dialogue data collected using our conversation system.

1. はじめに

発話意図理解のための深層学習に基づく韻律特徴量の自動抽出手法を提案する．我々が情報伝達のために開発した即応性に富む会話システム [高津 18] を用いて収集した対話データを分析し設計した発話意図タグ付きコーパス [横山 18] を用いて提案手法の有効性を示す．

我々はニュース記事のようなまとまった量の情報を音声対話によって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，ユーザー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上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ー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．このようにユーザー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ーからのフィードバックが必要不可欠である．フィードバックに関する技術課題として，フィードバックの誘発とフィードバックの理解という観点がある．本研究は後者に関するもので，ユーザーからの多様なフィードバックを理解することが目的である．

ユーザーのフィードバックは必ずしも言語的に明示された形で表れるとは限らない．場合によっては，抑揚で表現されるニュアンスなどにユーザーの意図が表れることもある．そこで，本研究では言語情報と韻律情報を組み合わせた発話意図認識手法を提案する．提案手法の特徴は，従来意図の種類ごとに観察に基づいて行われてきた特徴パラメータの設計を深層学習を用いることで，寄与率の高い韻律特徴量を自動で抽出しているところにある．実験ではデータセットとして我々の会話システムを用いて収集した対話データを分析し設計した発話意図タグ付きコーパス [横山 18] を使用し，提案手法の有効性を示す．

本稿の構成は次の通りである．2. 章で関連研究について述べる．3. 章で提案する発話意図認識モデルについて説明し，4. 章で発話意図タグ付きコーパスを用いた実験結果を報告する．

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. 関連研究

従来，発話意図の認識・理解は韻律情報を用いて行なわれてきた．例えば，[藤江 03] らは，システムに対する利用者の発話態度（肯定的か否定的か）を推定するために，第 1 モーラの基本周波数（F0）の傾き，発話全体の F0 レンジ，最終モーラの継続長からなる 3 次元の特徴量を使用している．[Ando 15] らは，対話データにおいてその発話が肯定的であるか否定的であるかを推定するために，音響特徴量として単語ごとに算出した F0，パワーの最大・最小，継続長，間などの情報を，言語特徴量としてバイグラム言語モデルのパープレキシティを使用している．[Nisimura 06] らは，収集した子供の音声に対して喜んでいるか嫌悪を示しているかの推定を行う際，F0 値，パワーから求められる 16 次元の特徴量の中から，因子分析により寄与の高い因子を選択することで，より識別率の高い特徴量を使用している．[林 14] らは，発話タグ（疑問文，平叙文，相槌，同意，笑い）の推定を行う際，F0 値やパワー，話速などの 23 次元の音響特徴量を使用している．

これらの研究は，韻律情報を詳細に観察・分析することで意図理解のための特徴パラメータを設計している．しかしながら，人手で設計可能な特徴パラメータは，推定対象の発話意図の違いが文章や音声を観察することで見いだせる場合に限られる．つまり，観察・分析だけでは容易に特徴が見つけられないような微妙なニュアンスの意図については特徴量の設計が困難になる．加えて，発話意図ごとに特徴パラメータの設計が必要となる．

本研究では，発話意図理解の韻律特徴量をニューラルネットワークを用いて自動で抽出する枠組みを提案する．具体的には，音響情報としてスペクトログラムを CNN を含む AutoEncoder に入力し，その中間層出力を韻律特徴量として用いる．これにより，発話意図の種類に依らず同一の仕組みで韻律特徴量の獲得が可能となる．

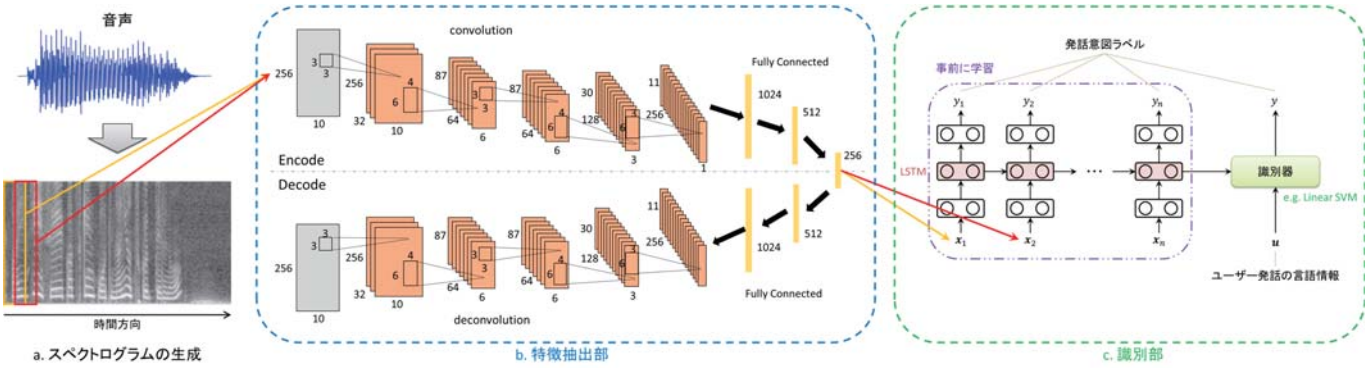


図 1: 発話意図認識モデル

3. 発話意図認識モデル

時間方向に可変な音声の入力を短い時間幅で逐次処理し，その都度過去の履歴を考慮して現時点で予測される発話意図ラベル（e.g. 質問，待機要求，反復要求）を出力するモデルを構築した．まず，短い時間幅で切り出した音声の断片からスペクトログラムを生成する．次に，得られたスペクトログラムを CNN を含む AutoEncoder（CNN-AutoEncder）に入力し，その中間層に圧縮された韻律特徴量を時系列に沿って LSTM に入力する．LSTM は逐次発話意図ラベルを出力するが，音声認識結果が得られた段階で，LSTM の最終層と音声認識結果から得られる発話の言語特徴量を識別器に与え，最終的な発話意図ラベルの推定結果を得る．モデルの全体像を図 1 に示す．以下，特徴抽出部と識別部について説明する．

3.1 特徴抽出部の設計

一般に発話意図の推定では，特徴量として基本周波数（F0）が用いられる．しかしながら，音声波形の準周期性や周辺雑音，有声音中の基本周波数の変化が広域に渡るなどの理由により，基本周波数を正確に抽出するのは難しい．そこで，F0 の推定を介さずに音声の時間・周波数スペクトルから直接特徴量を抽出する方法を提案する．

音韻や声の高さに関する特徴はスペクトログラムの模様として表れる．そこで，このスペクトログラムの模様を二次元の画像と見なして，5 層の畳み込み層と 3 層の全結合層からなるネットワークを折り返した全 16 層で構成される CNN-AutoEncoder（図 1.b）を学習する．そして，その中間層に圧縮された韻律特徴量を用いて発話意図の識別を行った．

3.2 識別部の設計

音声は時間方向に可変長であり，時間方向の長さに頑健なモデルであることが望まれる．同じ文字列の音声でも人の違いや発話する条件や状態によってその継続長は異なる．また，発話末のピッチが上昇すると「質問」と捉えやすくなるなど，発話意図認識において韻律の時間方向の変化は有用な情報である．

本研究では，RNN の中でも長期の依存性に長けた LSTM を用いた．さらに，音声認識結果が得られた段階で，LSTM の最終層と音声認識結果から得られる発話の言語特徴量を用いて再度識別を行うことで，発話内容に対しても頑健なモデルを構築した（図 1.c）．

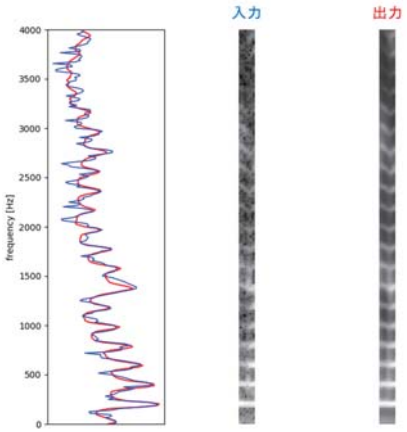


図 2: CNN-AutoEncoder の出力例（左:各スペクトログラムの 5 番目のスペクトルの周波数強度（青：入力，赤：出力），中：入力したスペクトログラム，右：出力したスペクトログラム）

4. 発話意図識別実験

4.1 特徴抽出部の学習と結果

特徴抽出部ではスペクトルの周期性を表すスペクトルパターンの情報を圧縮する．図 1.b の CNN-AutoEncder でスペクトログラムを復元するように学習を行うことで，その中間層にスペクトログラムの情報を圧縮する．この学習過程で期待することは，CNN におけるフィルターが音声情報，つまりスペクトログラムを扱うのに長けたものに学習されることである．そのため，学習で使うデータはドメインが異なる音声データであってもよい．そこで，本研究では公開されている日本語音声コーパスの中でも大規模な『日本語話し言葉コーパス（CSJ）』[*1] を用いて特徴抽出部の学習を行った．

CNN-AutoEncoder の入力は，フレームサイズ 800（50ms），フレームシフト 160（10ms），チャンクサイズ 1024 で切り出した音声から生成したスペクトログラムを時系列に並べたものとし，そのサイズは 10×256 とした．このデータをもとにネットワークの学習を行い，特徴抽出器を構築した．

学習した CNN-AutoEncoder の出力例を図 2 に示す．ここで，中央のスペクトログラムが入力として与えたスペクトログラムで，右側のスペクトログラムが CNN-AutoEncoder が出力したスペクトログラムである．左側のグラフは各スペクトロ

*1 http://pj.ninjal.ac.jp/corpus_center/csj/

表 1: 「質問」「非質問」データセットの統計

	総数	「質問」の数	「非質問」の数
訓練セット	2,000	925	1,075
テストセット	1,000	467	533
開発セット	257	138	119

グラムの 5 番目のスペクトルの周波数強度を表している．これらの結果からモデルが入力スペクトログラムの模様パターンを復元できていることが分かった．

4.2 識別部の学習と結果:発話意図「質問」の識別

横山らが構築したデータセット [横山 18] のうち，ラベラー間の一致率が高くデータ数が多い発話意図「質問」に関して識別実験を行った．

4.2.1 発話意図タグ付きコーパス

発話意図の情報が付与されたコーパスとして横山らが構築したデータセットを用いた [横山 18]．このデータセットは，我々が情報伝達のために開発した即応性に富む会話システム [高津 18] と 24 名の大学生が会話して得られた約 2,000 対話分の音声対話データに基づいて作られたデータセットである．収集したユーザー発話のうち，VAD で切り出した 1.5 秒以下の音声に対して 10 人のラベラーが発話意図ラベルを付与した．ここでは，その中でもラベラー間の一致率が高くデータ数が多い発話意図「質問」に関して識別実験を行った．本実験で用いた「質問」データセットの統計を表 1 に示す．ここで，「質問」以外の発話意図に分類された発話を「非質問」とする．

4.2.2 LSTM の学習と識別結果

まず，4.1 節で学習した CNN-AutoEncoder の中間層の値（256 次元）を入力として LSTM の学習を行った．ここで，CNN-AutoEncoder への入力には，100ms のスペクトログラムを 50ms ずつシフトさせながら与えた．つまり，現時刻の入力と次の時刻の入力で 50ms のオーバーラップが存在する．これは，LSTM に時間変化の情報を陽に学習させるためである．モデルのパラメータは，入力層のユニット数を 100，中間層のユニット数を 50, 出力層のユニット数を 50 に設定した．

LSTM の最終層の出力ラベルと正解ラベルの一致率を求めたところ，Accuracy は 0.896,「質問」を「質問」と識別する精度は 0.878,「非質問」を「非質問」と識別する精度は 0.912 であった．エラー分析を行ったところ，「なにそれ」や「どうして」「誰が」など，言語表記を見れば「質問」と判別できるような発話を「非質問」と判定してしまう事例が多々見られた．このような誤りを解消すべく，次節では発話の言語情報も組み合わせた発話意図の識別について検討した．

4.2.3 韻律情報と言語情報を用いた識別器の学習と識別結果

本実験では以下の観点について検証を行った．

- 有効な識別器
  scikit-learn[*2] の識別器を一通り比較し，有効な識別器について調査する．各識別器のパラメータには初期値を用いた．

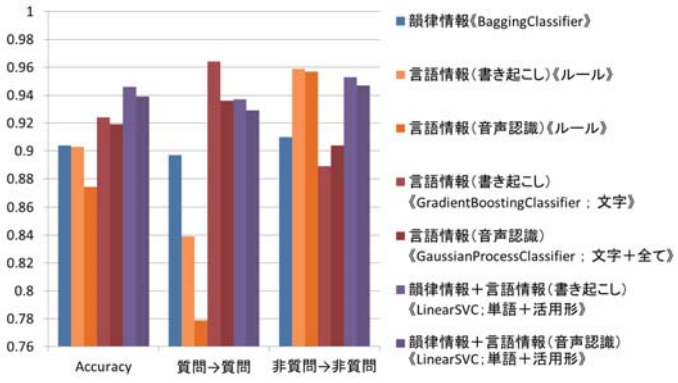


図 3: 発話意図「質問」の識別性能

表 2: 識別誤りの例

正解	出力	書き起こし	音声認識結果
質問	非質問	えっ	
		てなに	
		え何年	奈良根
		何ドルに	マンドリン
非質問	質問	低調	
		おー	
		んなんじゅ	なん j
		の	の

- 言語情報として有効な素性の組み合わせ
  言語情報には，基本素性としてユーザー発話の文字または単語の Bag-of-Words（BoW）を使用し，追加素性として JUMAN（Ver.7.01）[*3] を適用して得られる形態素情報 { 品詞（大分類と細分類）, カテゴリ, ドメイン, 活用形, 活用型 } を使用した．これらの素性に関して有効な組み合わせについて調査する．
- 言語情報のみ，韻律情報のみ，韻律情報と言語情報の 3 通りで識別器を学習したときの比較
  識別器を言語情報のみを用いて学習した場合と韻律情報のみを用いて学習した場合，韻律情報と言語情報を組み合わせて学習した場合で性能を比較し，各特徴量の有効性を確認する．韻律情報には 4.2.2 節で学習した LSTM の最後の隠れ層の値を用いた．
- 発話の表記として人手で書き起こした結果を用いたときと音声認識結果を用いたときの比較
  人手で書き起こした結果から抽出した言語特徴量を用いた場合と音声認識結果から抽出した言語特徴量を用いた場合でどの程度性能に差が見られるか確認する．音声認識には Google Cloud Speech API[*4] を使用した．
- ユーザー発話の言語表記に関してルールで識別したときと機械学習で識別したときの比較
  ベースラインとして基本的なルールでどの程度の識別性能を実現できるか確認する．実際に用いたルールは付録 A に示す．これらのルールは訓練セットの書き起こしデータに基づいて設計した．

*2 http://scikit-learn.org/stable/

*3 http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

*4 https://cloud.google.com/speech/?hl=ja

実験結果を図 3 に示す．まず，言語情報のみを用いてルールで識別した場合と機械学習で識別した場合について比較すると，ルールによる識別では「質問」を「質問」と識別する性能が低いことが分かった．次に韻律情報のみを用いたときと言語情報のみを用いたときの結果を比較すると，言語情報のみ（機械学習）の方が高い性能を示した．

最も良い結果を示したのは，韻律情報と言語情報（単語の表層形と活用形）を特徴量に用いて線形カーネルの SVM を学習し識別を行ったときであった．ここで，書き起こしではなく音声認識結果を用いたときの Accuracy の劣化は 0.7%にとどまっていた．誤り例を表 2 に示す．書き起こしを用いたときは正しかったが，音声認識結果を用いたときに誤った事例について確認したところ，音声認識できなかったもの（e.g. 「えっ」→「」）や音声認識誤りで疑問詞が消えてしまったもの（e.g. 「え何年」→「宇奈根」,「え何」→「えなり」）が誤りとして多く見られた．また，言語的に誤りを含むものだけでなく，言い淀みなどで音声的に不完全な発話も誤りやすいことが分かった．

5. おわりに

情報伝達効率の高い会話インタラクションを実現するためには，ユーザーからの多様なフィードバックを理解する必要がある．しかしながら，全てのフィードバックが言語的に明示された形で行われるとは限らない．場合によっては，抑揚で表現されるニュアンスなどにユーザーの意図が表れることもある．そこで，本研究では言語情報と韻律情報を組み合わせた発話意図認識手法を提案した．提案手法の特徴は，従来意図の種類ごとに観察に基づいて行われてきた特徴パラメータの設計を深層学習を用いることで，寄与率の高い韻律特徴量を自動で抽出しているところにある．

我々が情報伝達のために開発した即応性に富む会話システム [高津 18] を用いて収集した対話データを分析し設計した発話意図タグ付きコーパス [横山 18] を用いて評価実験を行ったところ，発話意図「質問」の識別に関して 94%程度の識別率を達成した．今後は，他の発話意図に関してもデータを増やし，識別実験を行う．

参考文献

[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)

[横山 18] 横山勝矢, 高津弘明, 本田裕, 藤江真也, 林良彦, 小林哲則, 会話によるニュース記事伝達のための発話意図分類とデータベースの構築, 第 32 回人工知能学会全国大会論文集 (2018)

[藤江 03] 藤江真也, 江尻康, 菊池英明, 小林哲則: パラ言語の理解能力を有する対話ロボット, 情報処理学会研究報告音声言語情報処理（SLP）, Vol. 2003, No. 104(2003-SLP-048), pp. 13-20 (2003)

[Ando 15] Ando, A., Asami, T., Okamoto, M., Masataki, H., and Sakauchi, S.: Agreement and disagreement utterance detection in conversational speech by extracting and integrating local features, in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, pp. 2494-2498 (2015)

[Nisimura 06] Nisimura, R., Omae, S., Kawahawa, H., and Irino, T.: Analyzing dialogue data for real-world emotional speech classification, in *Proceedings of the 7th Annual Conference of the International Speech Communication Association*, pp. 1822-1825 (2006)

[林 14] 林佑樹, 大佛駿介, 中野有紀子: 協調学習における韻律特徴を用いた発話タグ推定モデル, 教育システム情報学会第 39 回全国大会, pp. 441-442 (2014)

[田中 98] 田中真詞 , 川端豪: 文型と音調によるユーザの発話意図の推定, 情報処理学会研究報告音声言語情報処理（SLP）, Vol. 1998, No.68(1998-SLP-022), pp. 55-60 (1998)

[岩田 12] 岩田和彦, 小林哲則: 終助詞とその音調とによって聞き手に伝わる発話意図の分析, 電子情報通信学会技術研究報告, Vol. 112, No. 281, pp. 31-36 (2012)

[Schuller 04] Schuller, B., Rigoll, G., and Lang, M.: Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 577-580 (2004)

[Han 14] Han, K., Yu, C., and Tashev, I.: Speech emotion recognition using deep neural network and extreme learning machine, in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, pp. 223-227 (2014)

[Ghosh 16] Ghosh, S., Laksana, E., Morency, L.P., Scherer, S.: Representation learning for speech emotion recognition, in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, pp. 3603-3607 (2016)

[Satt 17] Satt, A., Rozenberg, S., and Hoory, R.: Efficient emotion recognition from speech using deep learning on spectrograms, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, pp. 1089-1093 (2017)

[Krizhevsky 12] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097-1105 (2012)

A 4.2.3 節の実験で用いた識別ルール

- 以下を含めば「質問」
  「何」「なに」「なーに」「どこ」「だれ」「誰」「なんで」「なんの」「なん日」「どう」「いつ」「いくら」「なぜ」「どんな」「どの」「どれ」「例えば」「たとえば」「なんて」
- 以下で終われば「質問」
  「か」「かね」「かな」「の」「って」「っけ」「は」「で」「が」「に」「ってこと」
- 以下に一致していれば「質問」
  「え」「ん」「えっ」
- 以下を含めば「非質問」
  「そっか」「そうか」「なんでもない」「何でもない」

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

ディープラーニングを用いた論理構造推定手法の提案

The Estimation of Argument Structure using Deep Learning

木下涼 Ryo Kinoshita　宇都雅輝 Uto Masaki　植野真臣 Maomi Ueno

電気通信大学情報理工学研究科
Graduate school of Informatics and Engineering, The University of Electro-Communications.

In this article, we present a novel approach for parsing argumentation structures. We classify argument components using Long-Short Term Memory(LSTM). Then, we use the result of component classification as the feature of the next task, identifying argumentative relations. Finally, the proposed model globally optimizes argumentative relations using integer linear programming whose objective function is the probability estimated in relation identification. Comparative studies show that each of the proposal are effective and our model significantly outperforms the most advanced model.

1. まえがき

近年，文章中の論理構造を自動で推定する論証マイニングと呼ばれる技術が注目されている [Peldszus 14, Stab 17]．論理構造の自動推定は自然言語処理分野の重要タスクの一つであり，多くの技術に応用されている [Ghosh 16]．

論証マイニングでは，次のサブタスクを順に解くことで文章中の論理構造を推定する．

要素抽出 文章中から，論理構造の構成要素となる文または節を論理要素として抽出する．

要素分類 各論理要素を「主張」あるいは「前提」に分類する．「主張」は，著者の意見を表す要素であり,「前提」は，データや例を示すことで，他の論理要素の説得力を高める要素である．

関係分類 段落内の全ての論理要素間について因果関係の有無を分類する．ここで，要素 i が要素 j を支持する因果関係について，要素 i を要素 j の原因，要素 j を要素 i の結果と呼ぶ．

構造推定 各論理要素をノードとみなし，要素間の因果関係を有向エッジで表現するグラフ構造を推定する．具体的には，上記のサブタスクの結果を用いて，任意の制約を満たすような構造を推定する．一般に，循環が存在せず，各ノードから引かれるエッジは一本以下という制約が用いられる．

これまで，これらの個々のサブタスクのみに着目した研究は多数なされてきた [Lippi 16]．しかし，全てのサブタスクを統合した手法は，[Stab 17] に留まる．

[Stab 17] は，要素抽出，要素分類，関係分類を機械学習手法を用いて解き，それらの結果を線形計画法を用いて統合することで構造推定を行う手法を提案している．しかし，この手法において，因果関係がある箇所を正しく推定できる割合は50%程度であり，十分に精度が高いとは言えない．この手法の課題として，次の点が挙げられる．

1) 論理要素が「主張」,「前提」のどちらであるかは前後の文脈に依存すると考えられるが，この手法では要素分類に文脈情報を活用していない．

2) 因果関係の有無は，要素分類の結果に依存すると考えられるが，この手法では，要素分類と関係分類を独立に解いている．

連絡先: 木下涼，電気通信大学情報理工学研究科，The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, 182-8585, Japan，kinoshita@ai.is.uec.ac.jp

これらの問題を解決するために，本研究では，文脈情報を活用し，要素分類と関係分類のサブタスクに依存関係を導入した新たな論証マイニング手法を開発する．具体的には，ディープラーニングの一種である Long-Short Term Memory(LSTM) を用いて，論理要素の前後の文脈を考慮した要素分類手法を提案する．また，要素分類によって推定される各論理要素が「主張」である確率推定値を関係分類の特徴量として用いることで，関係分類に要素分類の結果を反映する．

提案手法には，以下の利点が期待される．

1) LSTM を用いて文脈の情報を取り込むことで，要素分類の精度が向上する．

2) 要素分類の結果を関係分類に用いることで，関係分類の精度が向上する．

3) 要素分類，関係分類の精度が向上することにより，最終的な論理構造の推定精度が向上する．

本論文では，ベンチマークコーパスを用いて提案手法を [Stab 17] と比較した．結果として，提案手法を用いることで，要素分類の精度が約 7 %，関係分類の精度が約 8 %向上し，最終的な論理構造をより高精度に推定できることが示された．

2. 提案手法

2.1 提案手法の流れ

提案手法の流れを説明するために，n 個の論理要素で構成される段落に対する提案手法の概要を図 1 に示す．

提案手法における要素分類 (Classifying Argument Components) は，各論理要素が「主張」である確率をその要素と前後の文から推定することで行う．ここで，入力 $\boldsymbol{S}^{(i)} = (\boldsymbol{S}_1^{(i)}, \boldsymbol{S}_2^{(i)}, ..., \boldsymbol{S}_l^{(i)})$ は，段落中の i 番目の論理要素とその前後の文の単語を出現順に並べたベクトルであり，次元数はその単語数 l となる．従って，$\boldsymbol{S}_t^{(i)}$ は，論理要素 i とその前後の文章における t 番目の単語を表す．また，$\boldsymbol{H}_i$ は，$\boldsymbol{S}^{(i)}$ を入力とした LSTM の隠れ層の値を用いて，計算される変数である．$\boldsymbol{f}_i^c$ は i 番目の要素に対して，[Stab 17] で抽出している要素分類の特徴量ベクトルであり，次元数は 86 である．本研究では，要素 i が「主張」である確率推定値 q_i は，$\boldsymbol{H}_i$ と $\boldsymbol{f}_i^c$ を所与としたシグモイド関数で表現できると仮定する．

関係分類 (Identifying Argumentative Relation) は，[Stab 17] の特徴量に加えて，要素分類の結果を用いて，論理要素間に因果関係が存在する確率を推定することで行う．$\boldsymbol{f}_{ij}^r$

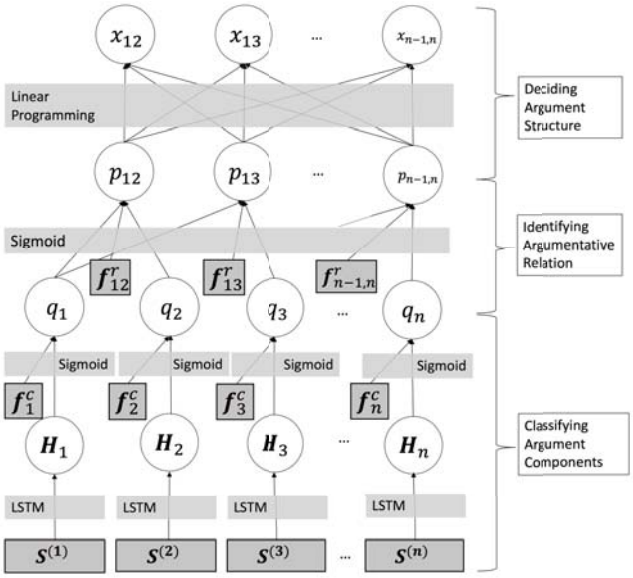


図 1: 提案手法の概要

は要素 i が要素 j の原因とした時，[Stab 17] で抽出している関係分類の特徴量ベクトルであり，次元数は 57 である．本研究では，要素 i が要素 j の原因である確率推定値 p_{ij} が，要素分類で推定された q_i，q_j と $\boldsymbol{f}_{ij}^r$ を所与としたシグモイド関数で表現できると仮定する．

構造推定 (Deciding Argument Structure) は，循環が存在せず，一つのノードから引かれるエッジは一本以下という制約のもと，関係分類で推定された p_{ij} を重みとする目的関数を線形計画法を用いて解くことで行う．

以降では，図 1 の各サブタスクについて詳細を説明する．

2.2 要素分類

提案手法では，要素分類に論理要素の前後の文脈を利用するために，LSTM を用いる．LSTM は，時系列データから長期の文脈を学習することのできるディープラーニング手法として，自然言語処理分野を中心に広く活用がなされている．

LSTM の隠れ層は，長期の依存関係を反映したものであり，文脈の情報が保持されているとみなせる．この特徴を利用し，提案手法では要素分類に LSTM の隠れ層に基づく特徴量を加える．

本研究では，要素 i が「主張」である確率推定値 q_i が，$\boldsymbol{S}^{(i)}$ を入力とした LSTM の隠れ層から算出される $\boldsymbol{H}_i$ と既存の特徴量 $\boldsymbol{f}_i^c$ を所与とした次のシグモイド関数で表現できると仮定する．

$$q_i = \frac{1}{1+\exp(-\boldsymbol{w^{(1)}}(\boldsymbol{f_i^c})^T - \boldsymbol{w^{(2)}}\boldsymbol{H_i^T})} \tag{1}$$

$\boldsymbol{w}^{(1)} = (w_1^{(1)}, w_2^{(1)}, ..., w_{|\boldsymbol{f}_i^c|}^{(1)})$，$\boldsymbol{w}^{(2)} = (w_1^{(2)}, w_2^{(2)}, ...w_{|\boldsymbol{H}_i|}^{(2)})$ は重みベクトルであり，訓練データから学習される．本研究では，この確率推定値を要素分類の出力とする．

LSTM の隠れ層に基づく特徴量として，全ての隠れ層の値の平均である Average Pooling や最後の隠れ層の値が一般に用いられている [Kim 14]．そのため，提案手法においても，Average Pooling を特徴量として用いる手法と最後の隠れ層を特徴量として用いる手法の二つを想定する．

Average Pooling を特徴量として用いる手法では，$\boldsymbol{S}^{(i)}$ を LSTM の入力とした時に出力される全ての隠れ層の平均値を特徴量 $\boldsymbol{H}_i^{ave}$ とする．ここで，$\boldsymbol{H}_i^{ave}$ の次元は，任意の値に設定することができる．この手法では，$\boldsymbol{H}_i^{ave}$ を $\boldsymbol{f}_i^c$ と組み合わせ，式 (1) で要素 i が「主張」である確率推定値 q_i を出力する．

一方，LSTM の最後の隠れ層を特徴量として用いる手法では，入力 $\boldsymbol{S}^{(i)}$ の最後の単語である $\boldsymbol{S}_l^{(i)}$ を入力とした時の LSTM の隠れ層ベクトルを特徴量 $\boldsymbol{H}_i^{last}$ とする．ここでも，$\boldsymbol{H}_i^{last}$ の次元は，任意の値に設定することができる．この手法では，$\boldsymbol{H}_i^{last}$ を $\boldsymbol{f}_i^c$ と組み合わせ，式 (1) で要素 i が「主張」である確率推定値 q_i を出力する．

2.3 関係分類

論理要素間の因果関係の有無は，個々の論理要素が「主張」であるか否かに依存すると考えられる．そこで，提案手法では，要素 i, j 間の関係分類を行う際に，要素分類の出力である要素 i が「主張」である確率推定値 q_i と要素 j が「主張」である確率推定値 q_j を特徴量に加える．

本研究では，要素 i が要素 j の原因である確率推定値 p_{ij} が，q_i，q_j，$\boldsymbol{f}_{ij}^r$ を所与とした次のシグモイド関数で表現できると仮定する．

$$p_{ij} = \frac{1}{1+\exp(-\boldsymbol{w^{(3)}}(\boldsymbol{f_{ij}^r})^T - w^{(4)}q_i - w^{(5)}q_j)} \tag{2}$$

ここで，$\boldsymbol{w}^{(3)} = (w_1^{(3)}, w_2^{(3)}, ..., w_{|\boldsymbol{f}_{ij}^r|}^{(3)})$ は重みベクトル，$w^{(4)}$，$w^{(5)}$ は重みスカラーであり，訓練データから学習される．本研究では，この確率推定値を関係分類の出力とする．

提案手法では，要素分類の結果を関係分類の特徴量として用いているため，ある要素が「主張」であるかを考慮した上で，因果関係の有無を分類することができる．

2.4 構造推定

提案手法では，要素分類の結果を関係分類で用いているため，[Stab 17] のように要素分類と関係分類の結果を統合する必要はない．従って，本研究では，構造推定を次のように，関係分類の結果のみを用いる線形計画問題として定義する．

$$\arg\max_{x\in(0,1)} \left(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}x_{ij} - \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} \right) \tag{3}$$

s.t.

$$\forall i : \sum_{j=1}^{n} x_{ij} \leq 1 \tag{4}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} \leq n-1 \tag{5}$$

$$\forall i : x_{ii} = 0 \tag{6}$$

$$\forall i \forall j : x_{ij} - a_{ij} \leq 1 \tag{7}$$

$$\forall i \forall j \forall k : a_{ik} - a_{ij} - a_{jk} \leq -1 \tag{8}$$

$$\forall i : a_{ii} = 0 \tag{9}$$

提案手法では式 (3) のように要素 i が要素 j の原因である確率推定値 p_{ij} を重みとする目的関数を設定し，線形計画法を用いて，制約を満たした上で，x_{ij} の値を決定することで最終的な論理構造を決定する．

式 (3) 中の $\lambda\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}$ は L1 正則化項であり，λ はチューニングパラメータである．コーパス中には，一つの段落内に複数の木構造が存在するような文章も含まれている．L1

正則化を用い，全体のエッジ数を制限することで，一つの段落内に複数の木構造が含まれる場合でも正しく推定することが可能となる．

制約である (4) から (9) 式は [Stab 17] と同様であり，循環がなく，一つのノードから引かれるエッジは一本以下である木構造を保証している．ここで，式 (7) から (9) で利用している a_{ij} は次式で値が決定され，循環構造の判定に用いられる．

$$a_{ij} = \begin{cases} 1 & (\text{要素iから要素jに可到達である}) \\ 0 & (\text{要素iから要素jに可到達でない}) \end{cases} \quad (10)$$

3. 評価実験

本章では，提案手法の有効性を評価する実験について述べる．評価実験では，小論文形式の文章に対して，論理要素の箇所と因果関係の有無がタグ付けされている"Argument Annotated Essays[*1]"に含まれる 400 文書を用い，10 交差検証を行って各サブタスクと最終的な構造推定の精度を求めた．精度の指標には，各ラベルに対する F 値とそのマクロ平均を用い，[Stab 17](以降では Stab2017 と表記する) と比較した．

3.1 要素分類の評価実験

本節では，要素分類に LSTM に基づく特徴量（以降では，LSTM 特徴量と表記する）を用いる効果を評価する．

本実験では，LSTM 特徴量を求める際に，LSTM の入力を論理要素（S）にする場合に加え，論理要素とその前文 1 文，(Pf1)，その前文 2 文 (Pf2)，論理要素以前に存在する段落内の文全て (PfAll)，論理要素とその次文 1 文（Pb1），その次文 2 文 (Pb2)，論理要素以降に存在する段落内の文全て (PbAll) の 6 パターンを用いた．ここで，次文を入力に含める際は，より文脈の情報を考慮できるように単語列の順序を逆転した上で入力とした．

また，それぞれの入力に対して，LSTM 特徴量を Average Pooling である $\boldsymbol{H}^{ave}$ とする手法と，最後の隠れ層である $\boldsymbol{H}^{last}$ とする手法を比較している．さらに，全ての手法に対して，隠れ層の次元数を 50・100・200 次元に設定し，比較した．ここでは，特徴量の差のみを評価するために，Stab2017 と同様に，Support Vector Machine(SVM) を用いて分類した結果を報告する．

Stab2017 と提案手法の要素分類の結果を表 1 に示す．F-Claim は「主張」の分類についての F 値，F-Premise は「前提」の分類についての F 値である．

表 1 より，論理要素のみを LSTM の入力とするだけでは，精度がほぼ変わらないことがわかる．一方で，論理要素とその前文 1 文を LSTM の入力とすることで,「主張」,「前提」のどちらの分類に対しても Stab2017 の精度を上回っていることから，LSTM を用いることで，文脈の情報を適切に活用できていることが示された．さらに，隠れ層を比較的高次元に設定した LSTM の最後の層の値を用いた際に，特に高い精度となることが明らかとなった．具体的には，論理要素と前文 1 文を入力とし，次元数 200 に設定した LSTM の最後の隠れ層を特徴量として用いた場合が最も高精度であった．これらの結果より，LSTM 特徴量を用いることで，要素分類の精度が大きく向上することが示された．

ただし，論理要素とその前後 2 文以上を入力とすると，精度は大きく低下する傾向にあることがわかった．これは，3 文以上の入力では入力列が極端に長くなるため，LSTM を用いても勾配消失問題が生じ，学習が難しくなることが原因と考えられる．

同様に，今回の結果からは，次文の文脈を考慮することによる分類精度の改善はみられなかった．この原因として，本実験で使用したコーパスでは，論理要素と次文 1 文の単語数の平均が，論理要素と前文 1 文の単語数の平均よりも 10 単語以上多く，前後 2 文以上を入力とした時と同様に，入力列が長くなり，適切に学習できなかった可能性が考えられる．

3.2 関係分類の評価実験

次に，関係分類の特徴量として，要素分類で推定した各要素が「主張」である確率推定値（以降では，主張特徴量と表記する）を用いる効果を評価する．

本実験では，主張特徴量を推定するのに既存の特徴量のみを用いた場合と，既存の特徴量と LSTM 特徴量を組み合わせて用いた場合の関係分類の精度を求め，関係分類を独立に行う Stab2017 の精度と比較した．また，要素分類の評価実験を踏まえ，LSTM 特徴量には，次元を 200 に設定し，論理要素と前文一文を入力とした LSTM の最後の隠れ層の値を用いている．ここでも，特徴量の差のみを評価するために，Stab2017 と同様に，SVM を用いて分類した結果を報告する．

これらの結果を表 2 に示す．F-Relation は「因果関係有」分類の F 値，F-noRelation は「因果関係無」分類の F 値，F-All は F-Relation と F-noRelation のマクロ平均である．

表 2 より，Stab2017 と LSTM 特徴量を用いない提案手法を比較すると，提案手法の「因果関係有」の分類精度が Stab2017 を大きく上回ることがわかる．同様に,「因果関係無」の分類精度も大幅に上回っており，関係分類の特徴量に主張特徴量を加えることで，関係分類の精度が大きく向上することが示された．

さらに，LSTM 特徴量を用いた提案手法が，比較手法の中で分類精度が最も高いことから，要素分類の精度向上が適切に反映されていることが示された．

3.3 構造推定精度の評価実験

本節では，提案手法の最終的な構造推定精度を評価する．ここでは，Stab2017 と提案手法のそれぞれについて，要素分類に LSTM 特徴量を用いる場合と用いない場合の構造推定精度を求める．本実験でも，要素分類の評価実験を踏まえて，LSTM 特徴量には，次元を 200 に設定し，要素と前文一文を入力とした LSTM の最後の隠れ層の値を使用する．また，式 (3) の L 1 正則化項に含まれるチューニングパラメータ λ の値は交差検証によって，刻み幅 0.01 で検証し，決定した．

これらの結果を表 3 に示す．F-Relation は「因果関係有」分類の F 値，F-noRelation は「因果関係無」分類の F 値，F-All は，F-Relation と F-noRelation のマクロ平均である．

表 3 から，LSTM 特徴量を用いた手法を比較した場合，提案手法の分類精度が Stab2017 を上回ることがわかり，LSTM 特徴量を用いる場合，提案手法の構造推定手法が有効であることが示された．

一方，LSTM 特徴量を用いない場合，提案手法よりも Stab2017 の分類精度の方が高い．この原因として，Stab2017 の構造推定は要素分類の影響を受けにくい定式化であり，要素分類の精度が低い場合でも，構造推定の精度が大きく変化しないことが考えられる．

評価実験の結果から，Stab2017 と比較して，LSTM 特徴量を用いる提案手法では，論理構造推定の精度が大幅に上回るこ

*1 https://www.ukp.tu-darmstadt.de/data/argumentation-mining/argument-annotated-essays/

表 1: LSTM を用いた比較実験結果

要素分類手法	$\boldsymbol{H}$	入力	次元	F-Claim	F-Premise
Stab2017	-	-	-	0.627	0.832
提案手法	$\boldsymbol{H}^{ave}$	S	50	0.623	0.836
			100	0.623	0.838
			200	0.622	0.833
		Pf1	50	0.663	0.853
			100	0.671	0.858
			200	0.679	0.860
		Pf2	50	0.610	0.823
			100	0.611	0.824
			200	0.605	0.825
		PfAll	50	0.617	0.824
			100	0.616	0.826
			200	0.613	0.817
		Pb1	50	0.613	0.835
			100	0.612	0.836
			200	0.599	0.831
		Pb2	50	0.609	0.830
			100	0.606	0.829
			200	0.606	0.827
		PbAll	50	0.623	0.833
			100	0.615	0.825
			200	0.608	0.828

要素分類手法	$\boldsymbol{H}$	入力	次元	F-Claim	F-Premise
提案手法	$\boldsymbol{H}^{last}$	S	50	0.622	0.838
			100	0.631	0.849
			200	0.648	0.855
		Pf1	50	0.608	0.832
			100	0.642	0.849
			200	**0.698**	**0.875**
		Pf2	50	0.618	0.823
			100	0.601	0.831
			200	0.601	0.834
		PfAll	50	0.614	0.831
			100	0.617	0.836
			200	0.606	0.833
		Pb1	50	0.615	0.834
			100	0.612	0.836
			200	0.598	0.834
		Pb2	50	0.613	0.835
			100	0.601	0.832
			200	0.596	0.833
		PbAll	50	0.613	0.835
			100	0.612	0.836
			200	0.599	0.831

表 2: 関係分類の評価実験結果

関係分類手法	LSTM	F-Relation	F-noRellation
Stab2017	-	0.428	0.731
提案手法	無	0.460	0.794
	有	**0.507**	**0.828**

表 3: 提案手法の評価実験結果

構造推定手法	LSTM	F-Relation	F-noRelation	F-All
Stab2017	有	0.506	0.890	0.698
	無	0.466	0.881	0.674
提案手法	有	**0.527**	**0.894**	**0.718**
	無	0.451	0.879	0.665

とがわかった．従って，提案手法では，Stab2017 よりも論理構造を高精度に推定できることが明らかとなった．

4. むすび

本論文では，LSTM を用いて文脈情報を活用し，サブタスク間に依存関係を導入した新たな論証マイニング手法を開発した．具体的には，以下の手順で論理構造を推定する．

1）論理に関わる文や節である論理要素が「主張」であるか分類する要素分類において，入力を論理要素とその前文 1 文にした LSTM の最後の隠れ層の値を既存の特徴量と組み合わせて用いる．本研究では，各要素が「主張」である確率推定値がこれらの特徴量を所与としたシグモイド関数によって表現されると仮定し，推定された値を関係分類の特徴量として採用する．

2）論理要素間の因果関係の有無を分類する関係分類では，要素分類で推定した各要素が「主張」である確率推定値を既存の特徴量と組み合わせて用いる．本研究では，因果関係の存在する確率推定値がこれらの特徴量を所与としたシグモイド関数によって表現されると仮定し，推定された値を，線形計画問題の目的関数の重みとして採用する．

3) 関係分類で推定された，各論理要素間に因果関係の存在する確率推定値を目的関数の重みとする線形計画問題を，木構造の制約のもと，線形計画法を用いて解くことで，最終的な論理構造の推定を行う．

コーパスを用いた評価実験により，提案手法では，要素分類と関係分類の精度が既存手法を大きく上回っていることがわかった．また，提案手法の構造推定精度が既存研究の推定精度を大きく上回ることが明らかとなった．

ただし，それでもまだ因果関係ありについての分類精度は 53%程度と十分に高いとはいえず，応用のためにはさらなる精度の向上が必要である．そのために，さらなるディープラーニング手法の活用や特徴量選択により精度を向上させていく．

謝辞 本研究は JSPS 科研費 JP17H04726，JP17K20024 の助成を受けたものです．

参考文献

[Ghosh 16] Ghosh, D., Khanam, A., and Muresan, S.: Course-grained Argumentation Features for Scoring Persuasive Essays, in Proceeding of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 549–554, Berlin, Germany (2016), Association for Computational Linguistics

[Kim 14] Kim, Y.: Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882 (2014)

[Lippi 16] Lippi, M. and Torroni, P.: Argumentation Mining: State of the Art and Emerging Trends, ACM Trans. Internet Technol., Vol. 16, No. 2, pp. 10:1–10:25 (2016)

[Peldszus 14] Peldszus, A.: Towards segment-based recognition of argumentation structure in short texts, in Proceedings of the First Workshop on Argumentation Mining, pp. 88–97, Baltimore, Maryland (2014), Association for Computational Linguistics

[Stab 17] Stab, C. and Gurevych, I.: Parsing Argumentation Structures in Persuasive Essays, Computational Linguistics, Vol. 43, No. 3, pp. 619–659 (2017)

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

Analogy comprehension between psychological experiments and word embedding models

浅川 伸一 *1 Shin Asakawa
岡 龍之介 *2 Ryunosuke Oka
楠見 孝 *3 Takashi Kusumi

*1東京女子大学 Tokyo womens' Christian university
*2京都大学教育学研究科 Kyoto University, School of Education
*3京都大学教育学部 Kyoto University

Word embedding models such as word2vec (Mikolov,2013) and GloVe (Pennington et al.2014) became to be widely recognized, while similarity judgment of analogy and comprehension could be described on a Euclidian space. We tried to compare between word embedding models and human performances. Despite that several models were proposed to deal with adjust each domain in order to improve their performance, it has still remained unsolved the relation between human judgment and understanding of analogy and their performance of word embedding models. We investivated the relationship between them. Our results suggest that human understanding of analogy might be understood in terms of word embedding spaces. We disscussed the further possibilities to understannd the semantic space in human in these models.

1. はじめに

語義の分散仮説は言語の基本課題の一つである。たとえば文献によれば "words that occur in the same contexts tend to have similar meanings" [Harris 54] "a word is characterized by the company it keeps" [Firth 57] などの記述が見られる。単語埋め込みモデルは自動翻訳，推薦，チャットボット，極性分析，など広範な言語課題で用いられている。Mikolov ら [Mikolov 13] は $\mathbb{R}^d$ に単語を埋め込むことで，テキストコーパス内の類似の文脈内で単語が共起するときにその内積が最大になるように学習を行った。分布仮説 [Harris 54, Firth 57] と呼ばれる古典的な概念を実装したとみなすことが可能である。

典型的には，埋め込み方法の目的は，埋め込み空間におけるそれらの類似性がその意味または機能を反映するような方法で記号オブジェクト（単語，物体，対象，概念）を編成することである。この目的のため，対象物の類似性は，通常それらの距離または埋め込み空間内の内積 (cosine 類似度) によって測定される。

Mikolov らが単語埋め込みモデルの性能評価に用いた意味類推課題の例を挙げれば，たとえば $\mathrm{vec}(\text{ロンドン}, \text{イギリス}) \simeq \mathrm{vec}(\text{バグダッド}, \text{イラク})$ のごとくである。a:ロンドン, a^*:イギリス, b:バグダッド, b^*:イラク (この場合の正解語) とすれば，以下のように定式化できる。Mikolov らは cosine 類似度 $(\mathrm{Sim}_{\cos}(\mu,\nu) = \cos(\mu,\nu)/(|\mu||\nu|))$ を用いた。これに対し Levy & Goldberg[Levy 14] は以下のような派生手法を提案した。

PAIRDIRECTION:

$$\arg\max_{b^* \in V} (\cos(b^*, b - a + a^*)) \tag{1}$$

3COSADD:

$$\arg\max_{b^* \in V} (\cos(b^* - b, a^* - a)) \tag{2}$$

3COSMUL:

$$\arg\max_{b^* \in V} \left(\frac{\cos(b^*, b)\cos(b^*, a^*)}{\cos(b^*, a) + \epsilon} \right) \tag{3}$$

Contact: Shin Asakawa, Tokyo women's Christian university, Zempukuji 2-chome, 6, 1, Suginami, 1678585, Tokyo, asakawa@ieee.org

一方，Che ら [Che 17] は Euclid 距離を用いることを提案した。

EUC:

$$\arg\max_{b^* \in V} \left(1 - \frac{|(a^* - a) - (b^* - b)|}{|a^* - a| + |b^* - b|} \right) \tag{4}$$

Che らは更に a, a^*, b, b^* の推論には 3 つの別問題を考えることが可能であり，cosine 類似に基づく推論が成り立たない場合があることを示した。*knowing:knew::selling:sold* の場合を表 1 に示した。

表 1: knowing:knew::selling:sold の推論課題結果 [Che 17]Tab.1 より

ターゲット語	knowing		sold	
予測語	thought	0.573	**sold**	**0.568**
	know	0.504	sell	0.535
	wanted	0.494	bought	0.528
	konwing	**0.489**	buying	0.486

表 1 中**太字**は正解語を示している。このように同一単語で推論方向を変化させると正解に至らない場合が存在する。

認知心理学の文献では，心理実験結果に基づく単語の意味空間は Euclid 的であることが示唆されている (図 1)。

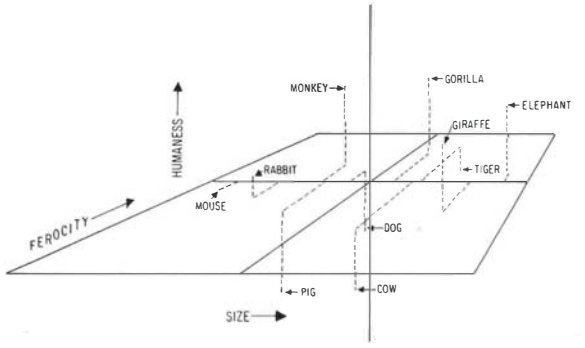


図 1: 心理実験による動物概念の布置 [Rumelhart 73] より

Rumelhart & Abrahamson は意味空間内の単語意味空間 (図 1) で人間の意味判断が Luce の選択公理 [Luce 59] に従うと仮定すれば心理実験結果を説明可能であることを示した。類推課題において単語 b^* が選択される確率は以下の式 5 に従う:

$$Pr\left(b^* \mid a, a^*, b\right) = \frac{v\left(d_i\right)}{\sum_j v\left(d_j\right)}, \tag{5}$$

ここで d_i は各単語間の距離である。意味判断過程に距離の応じた関数 $d \sim \exp\left(-\alpha \mathrm{Sim}\left(a, b\right)\right)$ を考えれば，式 5 はソフトマックス関数に等しくなる。Rumelhart & Abrahamson は式 5 が心理実験結果を説明可能であるとした。加えて Sternberg & Gardner[Sternberg 83] は，異なる単語類推課題である，系列補間課題，分類課題を提案した (図 2(b) および (c))。Sternberg & Gardner 課題を考慮に入れるとすれば，単語埋め込みモデルの評価に際して用いられるテスト課題に基いて性能を評価することの意味は再考する必要があるとも解釈できよう。

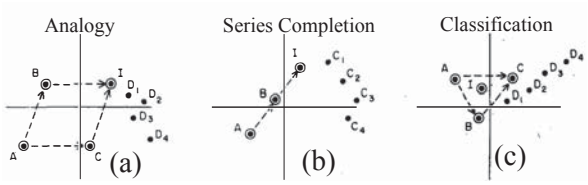


図 2: Stenberg [Sternberg 83]Fig.1 より

このような事実から，単語埋め込みモデルにおける推論について検討する必要があると考える。

2. 実験

2.1 中本・楠見の日本語の比喩理解

中本・楠見 [中本 04] の心理実験による日本語の比喩理解結果 (以下 NK04 と略記する) データと単語埋め込みモデルとを比較した。NK04 は 120 語対からなる日本語の比喩表現の心理評価データである。例えば「笑顔は花のようだ」のような比喩表現が用いられた。評定は，理解可能性，構成語類似性，独創性，面白さの 4 項目であり，9 段階評定である。単語埋め込みモデルとしては，2017 年 7 月時点の日本語ウィキペディア [*1] を mecab + NEologd[*2] によって分かち書きし，word2vec[*3] により単語埋め込みベクトル化した [*4]。Skip-gram を使用し，ベクトル化した際のパラメータは，埋め込みベクトル次元:200，ウィンドウ幅:20，負例サンプリング:20 とした。出現頻度 5 回以上の単語のみを考慮することとし，総語彙数 180,543 語を得た。NK04 に用いられる単語で「水晶細工」の 1 単語だけ単語埋め込みモデルの訓練に用いた日本語ウィキペディアの語群に該当単語が存在しなかった。そのため以下では「水晶細工」を含む比喩文データを削除した全 119 単語対について結果を示した。

NK04 による単語対の関係を図 3, 4 に示した。図中のシンボルは原論文でのクラスタを示している。図 3 は横軸が cosine 距離であり，縦軸が Euclid 距離である。図中には双方への回帰直線も示した。

*1 https://dumps.wikimedia.org/jawiki/latest/
*2 https://github.com/neologd/mecab-ipadic-neologd
*3 https://code.google.com/archive/p/word2vec/
*4 https://github.com/ShinAsakawa/2017jpa

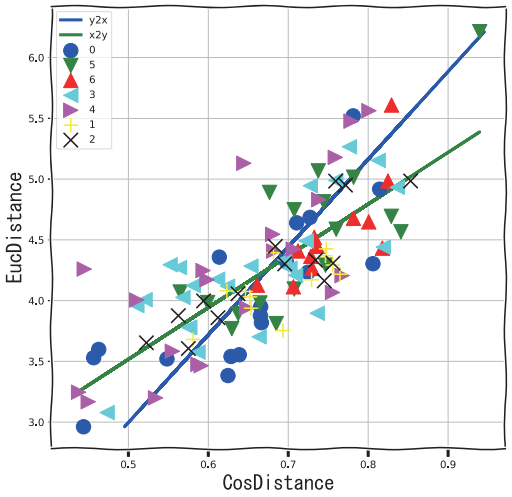


図 3: NK04 と cosine(x) and Euclid(y) distance

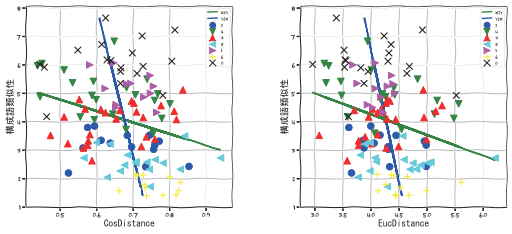


図 4: NK04:comparison human comprehension

図 4 は心理実験データが類似度評定であり単語埋め込みモデルでは距離を算出したため回帰直線も右下がりの傾向を示している。全体的な傾向は人間の判断と一致しているものの，単語埋め込みモデルと人間の類似性判断には尚乖離があることが推察される。

NK04 による心理実験結果と単語埋め込みモデルによる cosine 距離，及び Euclid 距離の相関係数を表 2 に示した。先の図 3 と図 4 とを考慮すれば，相関係数間の差異のみに基いて優劣を判断するのは難しいと考えられる。

表 2: NK04 の類似性評価と word2vec(wikipedia.ja) 間の相関係数

	NK04	cosine
cosine	-0.276	
Euclid	-0.279	0.768

日本語ウィキペディアは，英語版ウィキペディアと比較すると総量で 10GB 程度の差があり既存の単語埋め込みモデルが十分な性能を示すことができないという考え方もありうる。そこで次節以降では，英語版ウィキペディアを用いて人間の類似性判断との相違を検討した。

2.2 SimLex999

SimLex999[Hill 14][*5] は 999 単語対からなるデータセットである。NK04 と同様の手続きにより SimLex の 999 単語対を人間が類似性判断したデータセットを単語埋め込みモデルを用いて予測することを試みた。単語埋め込みベクトルには Google の公開している訓練済の単語埋め込みモデル `GoogleNews-vectors-negative300.bin.gz`[*6] を用いた。Google 版の単語埋め込みモデルは語彙数 3,000,000 であった。

NK04 の図 3 に対応する図が 5 であり，同じく図 4 のそれへの対応が図 5 である。NK04 の相関係数表 2 に対応する表が表 3 である。

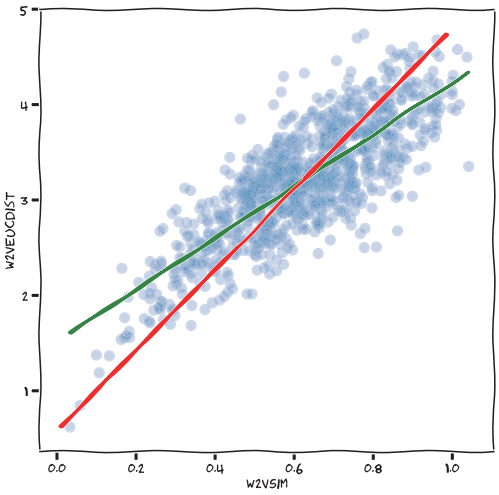


図 5: SimLex:cosine vs Euclid

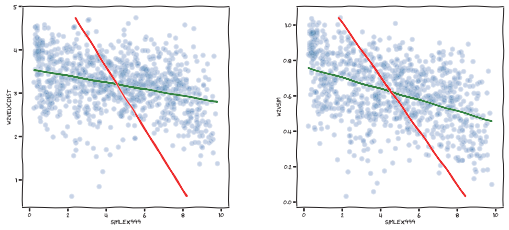


図 6: SimLex:comparison human comprehension

表 3: SimLex999 correlations

	SimLex	cosine
cosine	-0.453	
Euclid	-0.330	0.802

NK04 の結果と比較すれば，cosine 距離，Euclid 距離両者とも人間の類似性判断との相関係数が大きく，かつ，cosine, Euclid 距離測度間の相関係数も絶対値が大きくなった。データセットの語彙対が 999/119 と 8 倍以上，ウィキペディアの語彙数では 3,000,000/1,805,463 と 1.6 倍であった。評価データ数と訓練データ数と差異が精度向上に関与する可能性は考えられるものの，図 5 からは cosine 距離と Euclid 距離との乖離は保たれていた。このため人間が行う類似性判断との乖離は埋められずに残されていると見做すことが可能であろう。

2.3 MEN

MEN[*7] は人間による類似性判断データセットである [Bruni 12]。Amazon Mechanical Turk (MTurk)[*8] を通じて収集された。データ数は 3000 語対である。用いた単語埋め込みモデルは SimLex999 と同じく Google 版の訓練済データを用いることとした。先と同様の手法により図 7, 8, 表 4 を作成した。

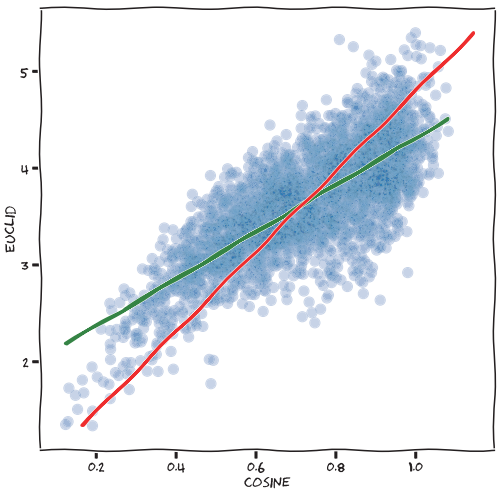


図 7: MEN:cosine vs Euclid

cosine, Euclid 距離間の相関係数は SimLex ほど高くないが，両距離とも人間の類似度評定との関連がか高くなった，単語埋め込みモデルは SimLex999 と MEN とで共通であるから，相違が生じたとすれば，データセットの持つ性質である可能性が考えられる。しかし，本データだけでは得られた結果の相違を十分に説明する証拠とは断言できないように思われる。

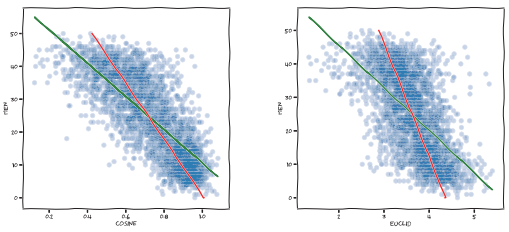


図 8: MEN:comparison human comprehension

*5 https://www.cl.cam.ac.uk/~fh295/simlex.html
*6 https://github.com/mmihaltz/word2vec-GoogleNews-vectors
*7 http://clic.cimec.unitn.it/~elia.bruni/MEN
*8 https://www.mturk.com/

表 4: MEN correlations

	MEN	cosine
cosine	-0.770	
Euclid	-0.612	0.766

3. 考察

人間による単語の類似性判断を説明するため，単語埋め込みモデルが用いることが可能か否かを検討するために 3 つのデータセットを用いて評価した。得られたデータの差異は，言語間の差異に帰すものであるか，訓練データ量であるのか，データセットの持つ性質によるのかを本研究だけから判断することはできない。しかし，単語埋め込みモデルが自動翻訳，文章理解，生成など自然言語処理技術に活用される基礎技術となった今日，単語埋め込みモデルを活用しない方が不自然であるとさえ感じられる。

本研究の目的の一つは，日本語比喩理解のための心理学的モデルとしての単語埋め込みモデルの可能性を検討することであった。比喩理解のためには Kintsch[Kintsch 01] の古典的モデルでも単語表象に対して，意味空間内での隣接語の相互作用が仮定されている。このことは可塑的フィッテイングと見做すことも可能であるが，一方で入力層の one-hot vector を単語埋め込みモデルを用いて次元圧縮し，上位層であるリカレントニューラルネットワーク層への入力とする意味では，言語モデルであれ翻訳モデルであれ，本研究で用いた比喩理解と同様に，入力表現を変形させることが言語理解過程であると見做すことも可能であろう。

4. まとめと展望

人間の意味理解，語彙の類似性判断と単語埋め込みモデルの関連を検討した。示した結果は初歩的な結果の域を出ていないが，人間が行っている文章理解の解明とその実装へ向けて意味モデルの果たす役割は大きいと考える。人間の示す意味理解を深く理解しなければ，実用的な知的システムを構築することはできない。例えば，mixup[Zhang 17] のようなデータ拡張を自然言語処理でも行うための基礎技術としても単語埋め込みモデルを精緻化していく必要があると考えられる。

文献

[Bruni 12] Bruni, E., Boleda, G., Baroni, M., and Tran, N.-K.: Distributional Semantics in Technicolor, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 136–145, Jeju, Republic of Korea (2012)

[Che 17] Che, X., Ring, N., Raschkowski, W., Yang, H., and Meinel, C.: Traversal-Fre eWord Vector Evaluation in Analogy Space, in Hwa, R. and Riedel, S. eds., *Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark (2017)

[Firth 57] Firth, J. R.: *A synopsis of linguistic theory 1930-55.*, Vol. 1952-59, The Philological Society, Oxford (1957)

[Harris 54] Harris, Z. S.: Distributional Structure, *Word*, Vol. 10, No. 2-3, pp. 146–162 (1954)

[Hill 14] Hill, F., Reichart, R., and Korhonen, A.: SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation, *arXiv preprint* (2014)

[中本 04] 中本 敬子, 楠見 孝比喩材料文の心理的特性と分類—基準表作成の試み—, 読書科学, Vol. 43, No. 1, pp. 1–10 (2004)

[Kintsch 01] Kintsch, W.: Predication, *Cognitive Science*, Vol. 25, pp. 173–202 (2001)

[Levy 14] Levy, O. and Goldberg, Y.: Linguistic Regularities in Sparse and Explicit Word Representations, in *Proceedings of the Eighteeth Conference on Computational Language Learning*, pp. 171–180, Baltimore, Maryland, USA (2014)

[Luce 59] Luce, R. D.: *Individual Choice Behavior*, Wiley, New York, USA (1959)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, in Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. eds., *Advances in Neural Information Processing Systems 26*, pp. 3111–3119, Curran Associates, Inc. (2013)

[Rumelhart 73] Rumelhart, D. E. and Abrahamson, A. A.: A Model for Analogical Reasoning, *Cognitive Psychology*, Vol. 5, pp. 1–28 (1973)

[Sternberg 83] Sternberg, R. J. and Gardner, M. K.: Unities in Inductive Reasoning, *Journal of Experimental Psychology: General*, Vol. 112, No. 1, pp. 80–116 (1983)

[Zhang 17] Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D.: *mixup*: Beyond Empirical Risk Minimization, *arXiv preprint* (2017)

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

比喩理解における意味構造の対応づけ：不定化した自然変換の探索として

Associating semantic structure in metaphor comprehension : as a finding indeterminate natural transformation

布山美慕 *1　西郷甲矢人 *2
Miho Fuyama　Hayato Saigo

*1玉川大学 Tamagawa University
*2長浜バイオ大学 Nagahama Institute of Bio-Science and Technology

In the present paper, we propose a new theory named “Theory of inderterminate natural transformation(TINT)” to investigate the processes of creating meanings as association relationships between images, focusing on the metaphor comprehension as an example. TINT models the meaning creation as a constructing indeterminate natural transformation which is defined by the category theory and stochastic processes. This theory allows us to represent structure-structure interactions between the meaning of target and source of metaphor.

1. 意味の創造過程としての比喩理解

私たちはイメージの世界に生きている．いま見えているもの，聞いていること，触れている感覚，味わい，感じる匂い，こういった五感や体性感覚に基づくイメージ、想像上の架空のイメージ，言語的あるいは非言語的なイメージ，そういった種々のイメージの中で私たちは生きている．たとえば，グラス，雪，月，母，ドラゴン，文字の「あ」，数字の「49」，積分，神様，あの彼，などのイメージがあるだろう．

私たちはこれらのイメージに対して種々の意味を感じる.「月」は私にとってはどのような意味を持つか，と考えると，月は私にとって夜空にあり，明るく，アルテミスのように賢く美しく，同時に狂気も思わせ，また小さい頃に家族と一緒に見た思い出が浮かび... といったように無数の意味が現れる．

人は，こういったイメージやその意味を新たに創造することができる．実際，日常的にこの創造をなしている．身近には “花” についての知識を得ることで “花” のイメージの意味が変化することもあれば，数学や芸術などの仮想的な世界，あるいは “現実世界” を変革しているような種々の科学技術や多様な文化を築き上げることもある．それらはこの豊かなイメージの意味の創造によるものと考えられる．

前述のイメージの意味として，様々に想起される内容はそれ自体また別のイメージである．これを踏まえ，著者らはイメージの意味を，そのイメージから他のイメージへの想起関係の総体として捉えたい．本論文では，この作業仮説を元に，イメージの意味の創造過程の端的な一例として比喩理解に注目し，比喩理解の新しい理論として不定自然変換理論（Theory of Indeterminate Natural Transformation）の提案を行う．比喩は「A は B のようだ」というようにあるものを別のもので言い表すことである（このとき，A を被喩辞，B を喩辞という）．A と B の間の想起関係がこの比喩を認識する以前に弱ければ，「A は B のようだ」という表現を知ることによって，A から B への新しい想起関係が生まれる．A からの想起関係の総体が A の意味だとすると，これは A の意味の変化，新規な A の意味の創造と見なせる．一つの想起関係だけが増えているため，新規な意味の創造過程としてもっとも端的な場合の一つであろう．したがって,（新規）比喩の理解過程を理論化することで，新規な意味の創造過程の理論の基礎の構築となることが期待できる．

連絡先: 布山美慕，玉川大学，東京都町田市玉川学園 6-1-1 玉川大学工学部機械情報システム学科　岡田研究室（133 室），042-739-8326，miho02@sj9.so-net.ne.jp.

2. 圏論の利用

「A は B のようだ」という比喩において，A と B は異なるイメージであるが，A と B はなんらかの意味で同じため，このような比喩が理解されうると考えられる．このように，自明に異なるものの間の同じさを扱う方法論として，数学の一分野である圏論の概念の応用が有用と考えられる [西郷 17]．加えて，圏論は種々の関係性を扱うことに優れており，こういった “同じさ” を関係性の関係として表現できる．これは，あるイメージの意味をあるイメージから別のイメージの想起関係として捉える本研究の思想と相性が良いと考えられる．こういった観点から，圏論の概念を利用して比喩の意味理解の過程の理論化を試みる．

本節ではごく簡単に必要となる圏論の概念を説明する．より詳しくは [MacLane 98] などを参照されたい．不定自然変換理論に必要となる圏論の主要な概念は “圏”，“関手”，“自然変換”，“コスライス圏” である．

圏とは，大まかに言えば,「対象」たちを連絡する合成可能な矢印=「射」たちのなすネットワークである．対象はなんらかの現象を表し，射は現象の間の変換や過程を表すと解釈できる．形式的には，**圏**とは，**対象**と**射**からなる次の条件を満たす体系である.「対象」と「射」の間の関係について，まず各射 f には，2 つの対象 dom(f)→ cod(f) とが対応づけられていて，それぞれ**域**，**余域**と呼ばれる．dom(f) と cod(f) は同じ対象であっても良い．

「射 f の域が X，余域が Y である」ということを

$$f : X \longrightarrow Y$$

あるいは

$$X \xrightarrow{f} Y$$

などと記し，こういった矢印を用いて組み上げられた表記を**図式**と呼ぶ．図式中の矢印の向きの表記は左から右に限らない．

各対象 X には，X を域とし，余域とするような射 1_x が対応づけられており，X の恒等射と呼ばれる．X の恒等射 1_X については，$1_X : X \longrightarrow X$ である．

次に，射 f, g で，cod(f) =dom(g) となるものがあったとき，つまり，

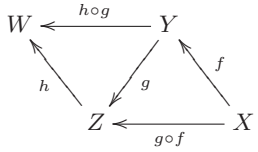
$$Z \xleftarrow{g} Y \xleftarrow{f} X$$

というような状況のとき，こういった f, g に対しては，これらの**合成**と呼ばれる射

$$Z \xleftarrow{g \circ f} X$$

が存在する．

これらの対象や射が満たすべき関係として，次の 2 つを要請する．

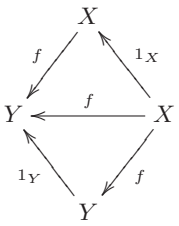


一つめは結合律で，上の図式ような状況のとき，X から W への経路は平行四辺形の上側を通る経路（$(h \circ g) \circ f$）と下側を通る経路（$h \circ (g \circ f)$）の二種類が考えられるが，これらが射として同じものでなければいけないという要請である．つまり，

$$(h \circ g) \circ f = h \circ (g \circ f)$$

となる．このように，図式内の対象から対象への道筋（つまり射の合成方法）が複数あるときに，結果が選び方によらない場合，その図式は**可換**であると呼ばれる．

この概念を用いれば，二つめの関係である**単位律**は，任意の射 $f : X \longrightarrow Y$ に対して，下記の図式が可換であることを要請するものとして述べられる．



つまり

$$f \circ 1_X = f = 1_Y \circ f$$

でなければならないというルールである．

まとめると，圏には「対象」と「射」とが備わっていて，これらは域，余域の概念を通じて関係しあっており，射には結合律をみたす合成という操作があり，対象には単位律をみたす恒等射が紐付けられている．

次に，二つの圏があったときに，これらの圏の間の同じさの一種として関手を次のように定義する．

定義 1：圏 $\mathcal{C}$ の対象，射から圏 $\mathcal{D}$ の対象，射への対応 F が**関手**（functor）であるとは，以下の 3 条件をみたすときにいう．

1. $\mathcal{C}$ の射 $f : X \longrightarrow Y$ を $\mathcal{D}$ の射 $F(f) : F(X) \longrightarrow F(Y)$ に対応させる．
2. $\mathcal{C}$ の各対象 X の恒等射 1_X について，$F(1_X) = 1_{F(X)}$ となる．
3. $\mathcal{C}$ の射 f, g の合成 $f \circ g$ について，$F(f \circ g) = F(f) \circ F(g)$ となる．

つまり，図式を図式に「きちんと」対応させるものが関手であると言える．ただし，この関手の定め方は一意とは限らない．

関手の定め方は一意とは限らないため，複数の関手を考えることができる．すると，これら複数の関手の間の変換を考えることができる．これが**自然変換**である．これは，ある圏と別の圏の一つの対応づけが自然変換を通じて他の対応づけに変換されると見ることができ，いわば関手自体を対象と見てそれらの間の射を考えることに相当する．

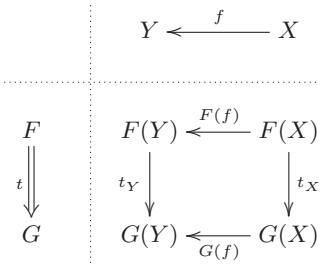
定義 2：F, G は圏 $\mathcal{C}$ から圏 $\mathcal{D}$ への関手とする．t が F から G への**自然変換**（natural transformation）であるとは，以下の 2 条件をみたすときにいう．

1. t は，$\mathcal{C}$ の各対象 X に対して $\mathcal{D}$ の射 $t_X : F(X) \longrightarrow G(X)$ を対応させる [*1]．
2. $\mathcal{C}$ の各射 $f : X \longrightarrow Y$ について，$F(X)$ から $G(Y)$ への射として

$$t_Y \circ F(f) = G(f) \circ t_X$$

が成り立つ．

この自然変換の図式は次のように描くことができる．



次に，イメージの意味の総体を表すのに用いる**コスライス圏**の概念を導入する．

定義 3：$\mathcal{C}$ を圏，X を $\mathcal{C}$ の対象とするとき，**コスライス圏** $X/\mathcal{C}$ を次のように定義する．

1. 対象は dom(f)= X となる全ての射 $f \in \mathcal{C}$.
2. 射は $f : X \longrightarrow a$ から $g : X \longrightarrow b$ への，$h \circ f = g$ を満たすような $h \in \mathcal{C}$.
3. 射の合成は $\mathcal{C}$ の合成で行う．
4. 恒等射は $\mathcal{C}$ の恒等射を引き継ぐ．

3. 不定自然変換理論の骨子

3.1 被喩辞と喩辞の間の意味ある関手の自然変換による探索

不定自然変換理論では，次のようにイメージの圏とイメージの意味の圏を定義する．

定義 4：イメージの圏 $\mathcal{C}$ の対象はイメージ（一節で述べたように，五感等に基づくものでも，言語的・非言語的なものでも，抽象的でも具体的でも良い種々のもの），$\mathcal{C}$ の射はそれら

*1 つまり自然変換は，そのそもそもの「身分」としては，対象から射への対応である．

の間の想起関係とする（対象間の射は複数あって良いが，この拡張は今後の研究にし，本論文ではこれを想起可能性として一つに限定して議論する）．

定義 5：A がイメージの圏 $\mathcal{C}$ の対象であるとする．このとき A の意味の圏をコスライス圏 $A\backslash\mathcal{C}$ で表現する．

1 節で述べたように，不定自然変換理論ではイメージの意味を他のイメージとの想起関係として捉える．たとえば，赤の意味は，赤から想起される他のイメージとの想起関係全体と考え，これを赤のコスライス圏として定義する．そして，不定自然変換理論では比喩理解の過程を次のようにコスライス圏の間の関手の自然変換の探索過程と見なす．

以上の定義のもとに，不定自然変換理論は比喩の意味理解を以下のように説明する．まず，「A は B のようだ」という比喩がなされたとする．すると，A から B への想起が起こる．これは図 1 のように表現されたイメージ A と B の意味を表すコスライス圏 $A\backslash\mathcal{C}$ と $B\backslash\mathcal{C}$ の間に，図 2 に示すように比喩によって一つの射 f が生まれることに相当する（この f によって，厳密には全体の圏が変化するため，$\mathcal{C}'$ と記している．ただし，あとで確率過程を導入して定式化しなおす際にこの変化は吸収される）．この f によって，自然に一意にコスライス圏 $B\backslash\mathcal{C}'$ から $A\backslash\mathcal{C}'$ への関手 $f\backslash\mathcal{C}'$ が生まれる．ここで，$f\backslash\mathcal{C} := (\cdot)\circ f : B\backslash\mathcal{C} \longrightarrow A\backslash\mathcal{C}$ であり，たとえば，図 2 の $b_1 \in B\backslash\mathcal{C}'$ を $b_1 \circ f \in A\backslash\mathcal{C}$ にうつすような関手である．この関手を *Base of Metaphor Functor* と呼ぶ．この Base of Metaphor Functor は，図 2 で見れば，「B にとっての y」を「A にとっての B にとっての y」というようなかたちで B を媒介として y と A との想起関係を作る．しかし，このように B を経由するかたちでは，直接的な想起関係でないため，比喩の解釈が不明確で，比喩理解がなされたとは考えにくい．そこで，$f\backslash C'$ の自然変換によってより自然な関手 F を見出し，意味を理解しようとする．これは，たとえば，$b_1 \circ f$ に対応するもとのコスライス圏 $A\backslash C$ の対象を探索することに相当する（すでに $b_1 \circ f \in A\backslash\mathcal{C}'$ なので，$A\backslash\mathcal{C}'$ の射の探索となっており，自然変換の探索に相当する）．この探索によって，図 3 のようにに対応する j が見つかるとき（ただし他の対象との関係性も含めて自然変換のかたちで対応することが条件），この対応づけによって新たな関手 F が見出される．この対応づけでは，「B にとっての y」は「A にとっての j」というように，B と A それぞれから直接に想起されるイメージ同士の間に対応づけがなされ，比喩の意味が解釈される（たとえば，「薔薇のような愛」という比喩なら，「薔薇にとってのトゲ」が「愛にとっての残酷さ」のように対応することで，この比喩の意味の理解が進む）．こういった一連の対応づけをなす関手 F が比喩の意味となると考える．ただし，自然変換は強い条件であり，コスライス圏 $B\backslash\mathcal{C}'$ から $A\backslash\mathcal{C}'$ への関手そのもの，つまりコスライス圏 $A\backslash\mathcal{C}'$ と $B\backslash\mathcal{C}'$ 全体に対して見出すことは通常難しいと考えられる（「薔薇」から想起されるイメージ全てに対して「愛」から想起されるイメージの全てを単一の解釈で結びつけることは難しい）．したがって，コスライス圏 $A\backslash\mathcal{C}'$ とコスライス圏 $B\backslash\mathcal{C}'$ の部分的な圏同士を結びつける関手に対する自然変換を探索すると考えるのが妥当であろう．部分的な圏は複数あり，自然変換も複数ありえるため，一つの比喩には複数の解釈がありえ，それらが並存する状態が比喩理解の状態だと考えられる．

3.2 動的な理解過程の記述のための圏の不定化

3.1 で記述したような圏の変化・発展過程（比喩の措定による射 f の発生や，自然変換の探索過程など）は圏論の枠組みの中では表現できない．そこで，これらの動的な過程を表現す

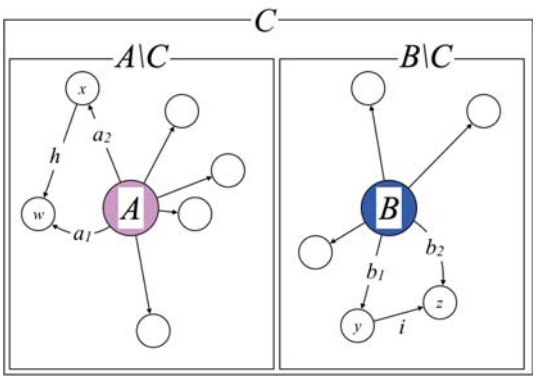


図 1: イメージ A の意味とイメージ B の意味を表現するコスライス圏 $A\backslash C$ と $B\backslash C$'.

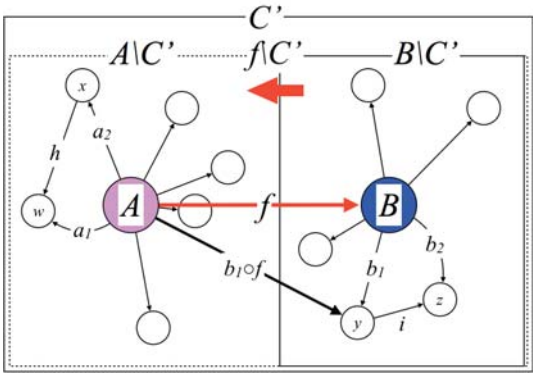


図 2: 比喩の措定．A は B のようだという比喩の措定によって f が生まれる．f を元に $f\backslash C$（Base of metaphor functor が生まれる．

るために，不定自然変換理論では確率的な過程を導入する．確率過程の入れ方については，現在研究を進めており，本論文では現在までのアイディアを記す．

基本的なアイディアとしては，イメージの圏 $\mathcal{C}$ の射である想起に，想起の確率的な重み μ を導入し，かつこの想起の連鎖過程のルールとして射の励起と緩和のルールを導入する．以下にこの不定化のアイディアを比喩理解の過程にそって羅列する．

- 全てのイメージの体系と全ての想起関係は圏 $\mathcal{C}$ としてモデル化される．想起確率 μ と励起緩和のルールは以下のように導入される．
- イメージ A の意味はコスライス圏と想起確率 μ によってモデル化される
- 「A は B のようだ」という比喩表現によって，射が励起される．f は f の μ の値に関わらず必ず励起される．この射 f の励起を契機に，励起と緩和の過程を経て Base of metaphor functor に対する自然変換を満たす関手が構築される．
- 励起の過程は以下のルールに従う．

  0 （Basic rule）: 励起した二つの射の合成によってできた射は励起する．域と余域の対象と同一視できる

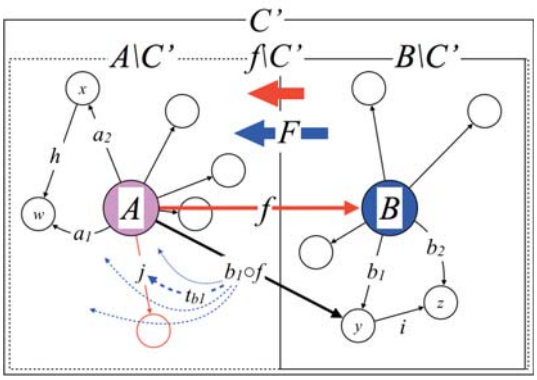


図 3: $f\backslash C$ に対する自然変換が探索され，比喩の関手 F が構築される．

恒等射が励起している射は励起する（μ に関わらず必ず励起する）．

1 （Neighboring rule）: 励起した射の余域を域に持つ射は μ に依存して決まるある確率で励起する．

2 （Fork rule）: 域を共有している射が励起しているとき，その余域の間の射（あるいは間の射に相当する合成射）が探索され確率 μ に従って励起する．– ルール 1 あるいは 2 の特別な場合として，逆射も確率 μ で励起する．

- 緩和の過程は以下のルールに従う（励起に比較して長い時間で緩和過程は起こる）．

1 (Anti-Fork rule): 域を共有する 2 つの射が互いの余域間に射を持たないとき，この 2 つの射の励起はおさまる（2 つの射はその時点の圏からなくなる）

- 以上の過程を経て，不定圏は定まり，励起した射の族は $f\backslash\mathcal{C}$（Base of metaphor functor）の自然変換を成す（ここで生まれる自然変換が不定自然変換である）
- ここまでの不定自然変換の過程によって，μ の値が変化する（学習が起こる）．この μ の変化はイメージの想起関係の変化であり，この変化によって私たちの世界の見え方が変化する．

4. 既存の比喩の理論との違いと今後の展望

本理論の特徴は，関手と自然変換の概念を用いて，被喩辞と喩辞のイメージの構造の対応づけの記述を既存の理論よりも精緻化した点である．被喩辞と喩辞の概念の構造間の対応づけの探索に注目した類似の理論として，隠喩履歴理論 [Bowdle05] がある．しかし，隠喩履歴理論では，概念が有する種々の特徴の一つずつの対応づけが主であり，本理論が記述を試みている全体の構造同士の対応づけ（たとえば可換図式等で記載される関係性の保存）ほど強い制約は明確ではない．現時点では，本理論の実証的な検証は行なっていないため，実際の人の認知にどちらがより妥当かは不明であるが，この点は理論の違いとして重要であろう．これ以外にも，多くの比喩理解の理論が存在するが（レビューとして [平 11]），こういった他の理論との詳細な比較は今後の課題である．

今後の展望・課題としては以下のような順での研究を検討している．

1. 簡単なトイモデルを作り比喩理解の表現として不定自然変換理論がどの程度妥当か確かめ，改善する
2. 比喩の理解過程の時間的な発展をシミュレーションできるように実装し，人が妥当だと思う比喩の解釈との類似性を調べ，理論の妥当性や改善を試みる．
3. より詳しい先行研究との比較を行う．将来的にはパフォーマンスの比較も行う
4. 人を使った検証実験を考え，試みる

これらのステップごとに本理論を改定し，より良い理論の構築を目指す．特に，自然変換という強い制約に対して，比較的自由度の高い確率過程の導入をどのように行うのか，試行を繰り返しながら検討したい．

また，ステップ 2 のシミュレーションでは，たとえば，コーパスを元にして構築された距離の入った意味空間上での距離の逆数が想起関係に相関すると仮定することで実際の比喩の理解過程をシミュレーションすることを検討している（たとえば，skip gram モデル [Mikolov13] などを用いたベクトル空間を利用できる可能性があるだろう）．こういったシミュレーションによって，定量的・定性的に既存の理論との比喩理解の説明のパフォーマンスの比較が可能となり，本理論の説明範囲が明確になることが期待される．

参考文献

[西郷 17] 西郷甲矢人：「同じさ」の措定：圏論的観点から，JCSS Japanese Cognitive Science Society, pp.334–342（2017）．

[MacLane 98] Mac Lane, S.: Categories for the Working Mathematician **5**(2). Springer (1998).

[Bowdle05] Bowdle,B. F. & Gentner, D.: The Career of Metaphor. Psychological Review, **122**(1), pp.193–216(2005).

[平 11] 平知宏・楠見孝：比喩研究の動向と展望，心理学研究，**82**(3)，pp.283–299(2011).

[Mikolov13] Efficient estimation of word representations in vector space. arXiv preprint, arXiv: 1301.3781(2013).

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

Towards Interpretation as Natural Logic Abduction

Naoya Inoue[*1] Pontus Stenetorp[*2] Sebastian Riedel[*2] Kentaro Inui[*1*3]

[*1] Tohoku University [*2] University College London
[*3] RIKEN Center for Advanced Intelligence Project

This paper studies an abductive reasoning framework on natural language representations. We show how to incorporate Natural Logic (McCartney et al. 2009) into Interpretation as Abduction (Hobbs et al. 1993). The resulting framework avoids the difficulty of natural language-formal representation conversion, while handling important linguistic expressions such as negations and quantifiers. For proof-of-concept, we demonstrate that missing, implicit premises of arguments can be recovered by the proposed framework by a manual example walkthrough.

1. Introduction

Abduction is inference to the best explanation. This paper explores abduction for natural language understanding. Given a textual observation, the task is to generate the best, textual explanation. For example, the task of finding a missing assumption of an argument [Habernal 17] is a special form of abduction. Suppose the following argument is given[*1]:

(1) *Claim: Immigration is really a problem.*
Reason: Illegal immigrants have overrun large portions of California.

Our task is to generate an additional, unstated assumption that is needed for this argument to be valid, for example, the following statement:

(2) *These portions of California are doing far worse than anywhere else in the state.*

Argument completion is useful for many applications where logic-based evaluation of texts is needed, such as automated essay scoring [Taghipour 16], and automated essay feedback. Additionally, an abductive framework in general has been considered as a promising, unified framework for text understanding [Charniak 91, Hobbs 93b].

Conventionally, abduction has been studied in the context of formal knowledge representation such as propositional logic [Santos 94], or first-order logic [Stickel 91, Hobbs 93b, Inoue 12a]. The advantage is that reasoning on formal knowledge representation provides an ability to handle information in a structured way by using special operators such as quantifiers, negations, etc. However, converting a natural language text into a formal representation, a.k.a. semantic parsing, itself is a difficult task to date [Krishnamurthy 16], which hampers applying abduction to real-world problems.

This paper explores an abductive inference framework that directly works on natural language texts, which is orthogonal to the conventional formal representation-based approaches. More specifically, we study *Natural Logic* [Maccartney 09, Bowman 13, Angeli 14] in the context of abductive reasoning. Natural Logic is a framework originally developed for entailment prediction. The advantage is that (i) it works on natural language expression directly, and (ii) it supports important linguistic expressions such as negations and quantifiers.

We show how Natural Logic-based inference can be integrated with Interpretation as Abduction [Hobbs 93a], one of the promising, abductive natural language understanding framework on formal representations (Sec. 3.). The key idea is that following the Natural Logic edit rules, we rewrite an input sentence (i.e. a textual observation) to generate its explanation, instead of applying logical inference on formal representations. We then demonstrate how this proposed framework solves an argument completion problem with a manual example walkthrough (Sec. 4.).

2. Preliminaries

2.1 Natural Logic

Given a premise p and a hypothesis h (in natural language representation), Natural Logic [Maccartney 09] provides a mechanism to determine whether p entails h or not.

Natural Logic first finds a sequence of atomic edits that transforms p into h. An atomic edit can be one of the following: (i) insertion (INS), (ii) deletion (DEL), or (iii) substitution (SUB). For example, given p = "Stimpy is a cat." and h = "Stimpy is not a dog.", the edit sequence will be the following: {SUB(cat, dog), INS(not)}. The corresponding mutated sentences are: p' = "Stimpy is a dog.", and p'' = "Stimpy is not a dog.", respectively.

To determine the relation between p and h, Natural Logic assigns a semantic operator to each edit. For example, the semantic operator of SUB(cat, dog), the first edit, is "alternation" (|) because cat and dog are mutually exclusive concepts. The semantic operator of INS(not) is "negation" (^). Finally, a semantic relation between p and each mutated sentence (e.g. p', p'') is estimated by these local semantic operators. For example, we first have $p|p'$ (by initialization). We then apply the second semantic operator (i.e. "negation") to this semantic relation, yielding "for-

Contact: Naoya Inoue, Graduate School of Information Sciences, Tohoku University, naoya-i@ecei.tohoku.ac.jp

*1 Taken from the Reasoning Comprehension Task: https://github.com/UKPLab/argument-reasoning-comprehension-task

4Pin1-33

ward entailment" (i.e. $p \sqsubset p''$), according to the Natural Logic theory. The result of semantic operator application depends on the context (e.g. whether an edited word is in a negation scope or not), but it is omitted here for the brevity. See the original paper[Maccartney 09] for further details.

2.2 Interpretation as Abduction

Hobbs et al. propose *Interpretation as Abduction* (IA), the framework of text understanding based on the idea that interpreting sentences is to prove the logical form of the sentence [Hobbs 93b]. They demonstrated that a process of natural language understanding, such as word sense disambiguation or reference resolution, can be described in the single framework based on abduction.

In IA, observations are given with costs, and background axioms are given with weights. It then performs backward-reasoning on each observation, propagates its cost to the assumed literals according to the weights on the applied axioms, and merges redundancies where possible (henceforth, *factoring*). A cost of interpretation is then the sum of all the costs on elemental hypotheses in the interpretation. Finally, it chooses the lowest cost interpretation as the best interpretation.

Formally, the task of IA can be described as follows:

Given: Background knowledge B, and observations O, where B is a set of first-order logical formulae, and O is a set of literals or substitutions,

Find: $\arg\min_{H \in \mathcal{H}} Cost(H)$, where $\mathcal{H}$ is a set of hypotheses H such that $H \cup B \models O$, $H \cup B \not\models \bot$,

where H is a set of literals or substitutions and $Cost(H)$ is a cost function of hypotheses.

Let us suppose $O = r(a)^{\$20} \wedge b(a)^{\$10}$. We use backward chaining to generate new hypotheses, increasing the cost of a literal back-chained on. For example, when $\forall x(p(x)^{0.3} \wedge b(x)^{0.9} \rightarrow r(x))$ is applied, we get $p(a)^{\$6} \wedge b(a)^{\$18} \wedge b(a)^{\$10}$. The cost of this hypothesis is $\$6+\$18+\$10 = \$34$. Because we have two literals with the same content, i.e. $b(a)$, we can apply factoring. This yield new hypothesis: $p(a)^{\$6} \wedge b(a)^{\$10}$, taking the least cost among those of two literals. This factoring mechanism encourages IA to find a small explanation that explains input observations.

3. Interpretation as Natural Logic Abduction

We now describe how Natural Logic can be applied to Interpretation as Abduction (IA). The basic idea is that we replace backward chaining in IA with atomic edits in Natural Logic to perform abductive reasoning on natural language representation. This allows us to perform abductive reasoning without the difficulty of natural language-formal representation conversion, while handling important linguistic expressions such as negations and quantifiers properly.

Analogously to IA, we define observations to be cost-assigned textual observations (e.g. O ={"Immigration is really a problem."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}).

To generate new hypotheses, we first initialize H to be O. Formally, let $H = \{s_1^{\$c_1}, s_2^{\$c_2}, ..., s_n^{\$c_n}\}$. We then apply the following procedure:

1. Choose a target sequence ID $i \in \{1, 2, ..., n\}$.
2. Mutate s_i by applying one of Natural Logic edits described in Sec. 2.1 (i.e. substitution, deletion, or insertion). Let the mutated sentence be s_i'.
3. Let $H' = \{..., s_{i-1}, s_i', s_{i+1}, ...\}$.
4. Estimate a semantic relation between s_i and s_i' according to the Natural Logic theory. If $s_i \sqsubset s_i'$, let $c(s_i') = 1.2 \cdot c(s_i)$; otherwise $c(s_i') = c(s_i)$. Intuitively, this means that abductive reasoning has to pay a cost, but deductive reasoning does not.
5. Factor H' if possible.

For substitution in Step 2, one can use a knowledge base of lexical relations such as hypernym/hyponym, cause-effect relations (e.g. WordNet [Fellbaum 98], ConceptNet [Speer 12]). For insertion in Step 2, one can add a modifier (e.g. adjective or adverb) to a noun phrase or verb phrase in s_i according to a language model-like knowledge (e.g. World Knowledge Proposition (WKP) [Clark 09, etc.]), or adding special linguistic operators such as *every*, *not*.

By applying this procedure repeatedly, we will have a set of hypotheses (i.e. $\mathcal{H}$ in Sec. 2.2). Finally, we pick a hypothesis with the minimum cost from $\mathcal{H}$ as an output.

4. Proof-of-concept

In this section, we show a walkthrough demonstrating how Natural Logic-based IA can recover an additional, implicit assumption of an argument.

We use Example (1) for a walkthrough. We assume the following knowledge base for Natural Logic atomic edits:

- Value judgement: problem $\sqsubset$ BAD (V1)
- Agentive: immigration $\rightsquigarrow$ illegal immigrants (A1)
- Ontology: exists in $\sqsupset$ overrun (O1)
- Generic: X is bad $\sqsupset$ {X exists in Y, Y is in a bad situation} (G1)

With this knowledge base, we perform the following Natural Logic-based abductive reasoning:

1. Initialization: {"Immigration is really a problem."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}
2. By V1: {"Immigration is really BAD."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}

3. By A1: {"Illegal immigrants are really BAD."[$10], "Illegal immigrants have overrun large portions of California."[$10]}.

Since substitution with an agentive lexical relation (substituting Immigration with Illegal immigration in Step 3) is not defined in the original Natural Logic theory [Maccartney 09], the semantic relation between its original sentence and the mutated sentence is not obvious. The question here is: does a statement S_1 about the thing X entail a statement S_2 about the thing Y that is created by X? S_1 is likely to entail S_2 because S_1 implicitly states that Y is involved in the statement (e.g. "*I love the guitar.*" is likely to entail "*I love music.*"). Hence, we assume that this reasoning is deductive; we will study reasoning with lexical relations not defined in the Natural Logic theory in future work.

The cost of the hypotheses unchanged in Step 1-3 because the reasoning is deductive. The reasoning continues.

4. By G1: {"Illegal immigrants exist in X."[$6], "X is in a bad situation."[$6], "Illegal immigrants have overrun large portions of California."[$10]}

The cost increased because the reasoning is abductive. Note that we distributed the increased cost \$12 to two statements.

5. By O1: {"Illegal immigrants overrun X."[$7.2], "X is in a bad situation."[$6], "Illegal immigrants have overrun large portions of California."[$10]}

6. By factoring with X = large portions of California: {"Large portions of California is in a bad situation."[$6], "Illegal immigrants have overrun large portions of California."[$7.2]}

The final result contains "Large portions of California is in a bad situation." with the cost \$6, which means that this statement is an implicit assumption of the original input argument.

5. Related work

In Natural Language Processing, commonsense reasoning has received much attention in recent years. One such example is the field of knowledge graph completion. Given an existing knowledge graph encoding entity relationships (e.g. Trump-isPresidentOf-US), the goal is to infer a missing relationship between entities in the graph. There are a wide variety of computational models for this task; see [Wang 17] for a good overview. The second example is the study on narrative schema [Chambers 08, Modi 14, Granroth-wilding 16, Pichotta 16, Modi 16, Weber 17]. Given an event mention (e.g. *John commited a crime*), the goal is to predict an event that will most likely happen next to the input (e.g. *Police arrest John*). These are not studied in the context of abduction; however, these are closedly related to substitution, one of the Natural Logic atomic edits. In future work, we will borrow these well-developed technologies for making abductive reasoning more robust.

In the machine learning community, the interpretability of models have recieved much attention. There are a wide variety of studies for explaining predictions of black box machine learning models (e.g. neural network) [Zhang 18]. On the other hand, our work aims at creating a transparent, white box model instead of finding an explanation to its own prediction.

6. Conclusions and future work

We have presented an abductive natural language understanding framework, integrating Interpretation as Abduction [Hobbs 93b] with Natural Logic [Maccartney 09]. For proof-of-concept, we have demonstrated that this framework explicates an implicit assumption underlying an argument on one concrete example.

Our future work includes several research directions. First, we plan to extract lexical knowledge from an existing, rich knowledge base such as WordNet [Fellbaum 98], FrameNet [Baker 98], or ConceptNet [Speer 12]. Second, we will extend our framework to support a wider variety of commonsense reasoning. We develop new Natural Logic semantic operators for lexical relations such as cause-effect relations, which are frequently used in natural language understanding. In order to make inference more robust, we also plan to use a knowledge graph embedding models [Wang 17], sequence-to-sequence models [Sutskever 14] for substituion, and distributed representation-based similarity calculation for factoring. We also try to learn the cost function of weighted abduction in a supervised manner, extending [Inoue 12b]. Finally, we will develop an efficient search mechanism for finding the least-cost explanation.

Acknowledgements

This work is supported by JSPS KAKENHI 16H06614, and JST CREST (JPMJCR1513).

References

[Angeli 14] Angeli, G. and Manning, C. D.: NaturalLI : Natural Logic Inference for Common Sense Reasoning, *Proceedings of EMNLP*, pp. 534–545 (2014)

[Baker 98] Baker, C. F., Fillmore, C. J., and Lowe, J. B.: The berkeley framenet project, in *Proc. of COLING*, pp. 86–90Association for Computational Linguistics (1998)

[Bowman 13] Bowman, S.: Can recursive neural tensor networks learn logical reasoning?, *arXiv preprint arXiv:1312.6192*, pp. 1–10 (2013)

[Chambers 08] Chambers, N. and Jurafsky, D.: Unsupervised Learning of Narrative Event Chains., in *ACL*, pp. 789–797 (2008)

[Charniak 91] Charniak, E. and Goldman, R.: A Probabilistic Model of Plan Recognition, in *Proceedings of the*

Ninth National Conference on Artificial Intelligence, pp. 160–165 (1991)

[Clark 09] Clark, P. and Harrison, P.: Large-scale extraction and use of knowledge from text, *Proceedings of the fifth international conference on Knowledge capture - K-CAP '09*, pp. 153–160 (2009)

[Fellbaum 98] Fellbaum, C.: WordNet: an electronic lexical database (1998)

[Granroth-wilding 16] Granroth-wilding, M. and Clark, S.: What Happens Next ? Event Prediction Using a Compositional Neural Network Model, in *AAAI* (2016)

[Habernal 17] Habernal, I., Wachsmuth, H., Gurevych, I., and Stein, B.: The Argument Reasoning Comprehension Task (2017)

[Hobbs 93a] Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P.: Interpretation as abduction, *Artificial Intelligence*, Vol. 63, No. 1-2, pp. 69–142 (1993)

[Hobbs 93b] Hobbs, J. R., Stickel, M., Martin, P., and Edwards, D.: Interpretation as Abduction, *Artificial Intelligence*, Vol. 63, pp. 69–142 (1993)

[Inoue 12a] Inoue, N., Ovchinnikova, E., Inui, K., and Hobbs, J.: Coreference Resolution with ILP-based Weighted Abduction, in *COLING*, pp. 1291–1308 (2012)

[Inoue 12b] Inoue, N., Yamamoto, K., Watanabe, Y., Okazaki, N., and Inui, K.: Online Large-margin Weight Learning for First-order Logic-based Abduction (2012)

[Krishnamurthy 16] Krishnamurthy, J. and Tafjord, O.: Semantic Parsing to Probabilistic Programs for Situated Question Answering, pp. 160–170 (2016)

[Maccartney 09] Maccartney, B. and Manning, C. D.: An extended model of natural logic, *Proceedings of the Eighth International Conference on Computational Semantics IWCS8 09*, No. January, pp. 1–12 (2009)

[Modi 14] Modi, A. and Titov, I.: Inducing neural models of script knowledge, in *CoNLL-2014*, p. 49 (2014)

[Modi 16] Modi, A., Anikina, T., Ostermann, S., and Pinkal, M.: Inscript: Narrative Texts Annotated with Script Information, *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016)*, pp. 3485–3493 (2016)

[Pichotta 16] Pichotta, K. and Mooney, R. J.: Using Sentence-Level LSTM Language Models for Script Inference, *Arxiv*, pp. 279–289 (2016)

[Santos 94] Santos, E.: A linear constraint satisfaction approach to cost-based abduction, *Artificial Intelligence*, Vol. 65, No. 1, pp. 1–27 (1994)

[Speer 12] Speer, R. and Havasi, C.: Representing General Relational Knowledge in ConceptNet 5, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3679—-3686 (2012)

[Stickel 91] Stickel, M. E.: A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation, *Annals of Mathematics and Artificial Intelligence*, Vol. 4, No. 1-2, pp. 89–105 (1991)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in *Advances in neural information processing systems*, pp. 3104–3112 (2014)

[Taghipour 16] Taghipour, K. and Ng, H. T.: A Neural Approach to Automated Essay Scoring, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP-16)*, pp. 1882–1891 (2016)

[Wang 17] Wang, Q., Mao, Z., Wang, B., and Guo, L.: Knowledge graph embedding: A survey of approaches and applications, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 12, pp. 2724–2743 (2017)

[Weber 17] Weber, N., Balasubramanian, N., and Chambers, N.: Event Representations with Tensor-based Compositions (2017)

[Zhang 18] Zhang, Q. and Zhu, S.-C.: Visual Interpretability for Deep Learning: a Survey (2018)

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

機械学習による局地気象予報の試み

Innovative approach for prediction of local weather using machine learning

吉兼 隆生[*1]　　芳村 圭[*1]
Takao Yoshikane　　Kei Yoshimura

[*1] 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo

Local weather forecasting is technically very difficult because the mechanisms are still unknown and there are a lot of errors with numerical model. In this study, we found some close relations between the observed precipitation and the simulated precipitation distribution patterns in a local area, and then we have applied the relations to predict the local precipitation using machine learning (A regression model of support vector machine). As the result, the model biases were drastically reduced by the innovative approach. It is expected that this approach is useful to improve the performance of the local weather forecast.

1. はじめに

1.1 局地気象予測の難しさ

気象観測や数値シミュレーションを行う以前は，各地域において経験的に知られていた天気の変化パターンにより天気を予測してきた[1].「観天望気」とも呼ばれており，現在も船舶の出航の判断などに利用されている[2][3][4].

現在は，気象衛星を含む気象観測データが充実し，スパコンを用いた数値予報の精度も向上している．しかし，いつ，どこで，どのくらいの降水があるのかなど，地域詳細な気象予報については十分な精度が得られていない．理由の1つとして，短時間で大きく変化するメソスケール現象(数百キロ以下の規模を持つ気象システム)の予測が難しいことが挙げられる．降水は個々の積雲対流によりもたらされるが，それらの発生発達は地形や風の影響を受けて複雑に変化し，さらには複数の積雲対流によるメソ対流システムを構成することもある[5]．このような個々の積雲対流の形成プロセスや細かい地形の影響を含めて予測を行うには，数値モデルの高解像度化が不可欠だが，様々な技術的要因から現状では実現が難しい．さらに初期値依存性（カオス現象）の影響も考慮する必要がある．また，メソスケール現象の形成メカニズムについては未解明な点も多く，現象の複雑さが予測を一層困難にしている．

船乗りにとっては，急激な天気の変化がいつ起こるのか把握することは安全面で極めて重要である[2]．観天望気は，現在の科学的知見や技術では対応できない局地の天気を予測するために必要とされている．

1.2 局地気象予測における機械学習の可能性

観天望気を重要視しているのは，前兆現象として限定された地域内で何らかのパターン（規則性）が存在しているためと考えられる．そのパターンがどのようなものなのか認識できれば，局地気象予測に活用出来るだろう．近年，機械学習を用いて現在の大気の状態から数時間先を予測する手法が開発されてきた[6][7][8]．観測された降水分布から数時間先の降水を予測するものや上空の気温状況など降水に関係する複数の要素を取り入れて学習させて降水予測を行うものなど様々だが，いずれも時間の経過とともに精度が大きく低下する．メソスケール現象は時空間スケールが小さく，時間の経過とともに大きく変化するため，同じような大気の状態であっても結果が大きく異なる．メソスケール気象の特性が，機械学習でのパターン認識を困難にしていると推測される．

数値予報は，上述の手法やレーダ降雨域の移動などから降水予測を行うナウキャストと比較して長時間の精度の高い予測が期待される[8][9][10]．数値モデルはメソスケール現象の発生・発達・衰退などの過程を扱える他，データ同化手法により観測値からの誤差を改善できる．一方で，モデルの不完全さ（各過程や解像度など）による誤差を避けることは難しい．図1に気象庁の数値予測情報の1つであるメソスケール客観解析値[11]とレーダーアメダス[12]による 2015 年1月の月平均降水量分布である．レーダーアメダスでは日本海沿岸地域の比較的標高の低い平野部で大きな降水が見られるが，数値モデルでは平野部で少なく山岳地域で多い傾向が見られる．冬季の北西季節風が山岳風上側で強制されて上昇流となり，日本海沿岸地域に沿って断続的に降水が形成される．大まかな降水特性は数値モデルでも再現されているが，顕著な降水分布の誤差（モデルバイアス）が確認できる．バイアス原因として，解像度が粗い（格子間隔 5km）こと，降水過程などの物理プロセスが適切でないことが推測される[13][14]．それ以外にもモデル座標系による計算誤差[15]も想定されるため，単に数値モデルを高解像度化しただけでバイアスが改善するとは考え難い．

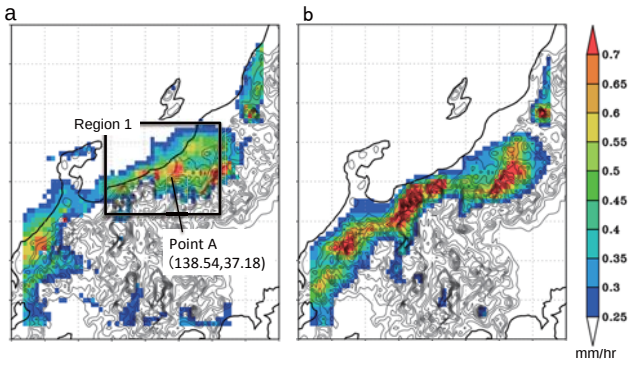


図1　2015 年1月の月平均降水量分布．レーダーアメダス(a)と気象庁メソスケール客観解析値(MSMGPV) (b) .

連絡先：吉兼隆生，東京大学生産技術研究所，277-8574 千葉県柏市柏の葉 5-1-5，E-mail takao-y@iis.u-tokyo.ac.jp

一方で，ある程度広い範囲の領域平均降水量の時間変化を見るとモデルと観測で大きな違いはなく，モデルは実際の降水特性を良く再現していると判断できる(図2)．モデルバイアスにより降水分布は実際と大きく異なるものの，冬季季節風に対応した降水特性(パターン)を再現しているのかもしれない．

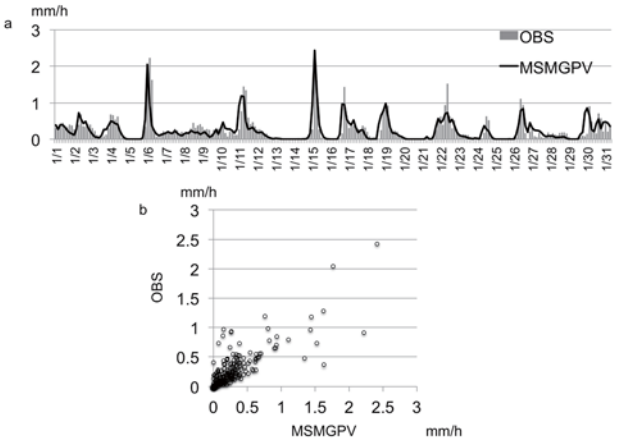


図2　Region 1(図1)での観測とモデル(MSMGPV)の領域平均降水量の時間変化(a)と時間降水量の関係(b)．

仮にモデルと観測の降水分布パターン特性が対応しているならば，パターン認識により局所的な降水予測が可能になると想定される．本研究では，モデルで再現された降水パターンを利用して機械学習による局所降水予測について調査した．

2.　方法

機械学習にはサポートベクターマシンの回帰モデル(SVR)を用いた．SVR は統計解析ソフト R のパッケージソフトウェアである e1071 を利用した[16].

入力にはレーダーアメダスの3時間ごとの観測降水(0.06 度四方に補正したもの)と気象庁メソスケール客観解析値(MSMGPV)のモデル降水を用いた．学習期間は，テスト期間(2015 年 1 月)を除く 2014 年～2017 年までの冬季(12 月~2 月)とした．0.06 度四方(約 5km 四方)の格子に対して，その格子での観測降水と格子を中心とした 40x30 格子の領域でのモデル降水分布の関係性を SVM により学習させた．(図3)

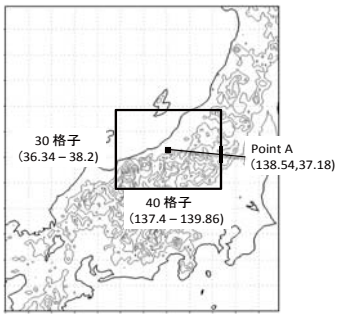


図3 SVM 入力データの設定．

3.　予測結果

SVM(SVR)による 2015 年 1 月の月平均降水量分布を図 4b に示す．降水分布特性は観測(図 4a)とよく似ており，定量的にもよく再現している．図 5 は Point A(図1)における降水量(3時間ごと)の時系列を示す．降水の起こるタイミングはよく一致しており(相関係数:0.75)，観測降水との関係性が高いことを示している．これらの結果から，モデル降水の分布パターンと実際の降水(観測値)との間に強い関係性があることが分かる．

地形に対応した降水は風速場に強く影響している．実際に日本海側山岳域で起こる降水は，冬季季節風が主要因である．モデルの分布パターンが風速場によるならば，モデルの降水分布パターンの代わりに風速場を特徴ベクトルとして採用しても同様に再現すると想定される．風速場の特徴ベクトルとして，40x30grid の風速分布(東西，南北成分)をそのまま与えた場合(Case 1)と，領域平均した風速場を与えた場合(Case 2)で，再現性を調査した(図 4d,e)．Case 1 では，降水分布パターンを与えた場合とほぼ同じ分布となったが(図 4d)，Case 2 では全体で過小評価であり，再現性は良くなかった(図 4e)．領域平均した風速場でも季節風の強弱が反映されているのだが，局所的な降水特性は大規模場との関係よりも地域詳細な風速分布パターンとの関係の方が強いことを示している．また，Case1 では降水量の時間変化(図 5b)が，特徴ベクトルにモデル降水分布を与えた場合(図 5a)と同様に再現されているが，Case 2 では，時間変動パターンの再現性が低く，変化のめりはりがあまり見られない(図 5c)．

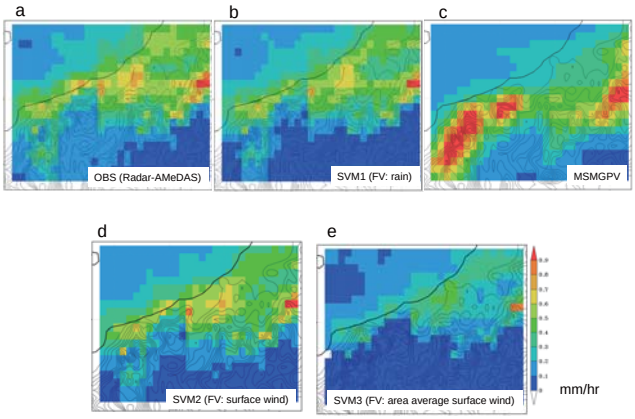


図4　2015 年1月の月平均降水量分布．観測(レーダーアメダス)(a)，モデル降水分布を特徴ベクトルとした SVR による予測値(b)，モデル(MSMGPV)(c)，地上風を特徴ベクトルとした予測(d)，領域平均地上風を特徴ベクトルとした予測(e)．

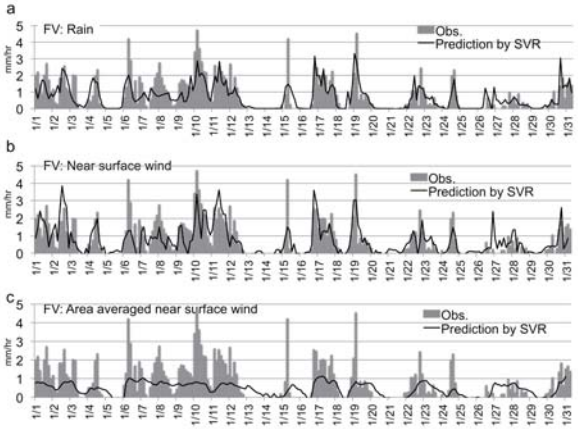


図5　Point A(図1)での降水量の時間変化．特徴ベクトルは、降水分布(a)，地上風分布(b)，領域平均地上風(c)．

4. 考察

4.1 観測とモデルの関係

SVR による予測結果は，観測された降水は，領域内部の分布の違い（パターン）により大きな影響を受けることを示している．つまり，局地気象予測，特に局地降水予測の場合は，季節風や高低気圧などの大規模スケールの気象現象との関係だけでなく，Region 1（図1）内部で形成される中小規模スケールの気象現象との関係性も無視できないことを意味する．複雑な地形の影響を強く受ける地域であれば，それに対応した気象現象との関係性を注意深く観測する必要があるだろう．従って，数値モデルバイアスの補正が可能であるのは，中小規模擾乱の再現が可能である高解像度のモデルで，降水分布や地上風分布の特性が十分に再現されている場合に限定されるだろう．図4と図5に示された SVR 予測値の高いパフォーマンスは，数値モデル（MSMGPV）による局地気象特性の再現精度が極めて優れていることを示している．

モデルで再現された気象要素と観測値との関係性を用いてバイアスを補正する手法として，「統計ダウンスケーリング」が一般に用いられる[17]．この手法は，解像度の粗い(格子間隔50km 以上)の温暖化予測モデルの結果を実用的な地域スケールの温暖化予測情報に補正するためによく使われる．しかし，基本的には大規模スケールの気象現象との関係のみで補正されているため，メソスケール気象現象と深い関係にある局地降水予測では大きな誤差が生じることも懸念されている[17]．本研究で用いた手法を適用することにより，従来のダウンスケーリング手法の問題点を改善し，高い精度での局地降水予測が可能になると期待される．

4.2 機械学習を用いた手法の問題点

第2章で示したように，局地気象予測において数値モデルと機械学習の組み合わせが有効である．しかし，機械学習では学習できない事象に対しては予測できない．また，出現頻度の低い極端降水現象に対して適用するのも難しい．稀な現象を予測する場合は，学習期間を長くするなどの必要があるかもしれない．対象とする地域や季節によって学習条件を含む特徴ベクトルの設定を変更する必要がある．

基本的には，観測結果と関係あると想定される要素を特徴ベクトルの候補として採用するが，そもそもどのような要素が深く関係しているのかさえ分からない場合も多い．局地気象の成因が複雑であり，メカニズムをほとんど解明できていないことが特徴ベクトルの選択を難しくしている．いずれにせよ，機械学習による局地気象予測のパフォーマンスを向上させるためには，特徴ベクトルの設定が極めて重要になるだろう．一つの方法として，特徴ベクトルを自動的に摘出可能なニューラルネットワーク，深層学習を適用することにより，人には複雑すぎて認識できないパターンを捉えて精度の高い予測が可能になるかもしれない[18].

4.3 機械学習への期待

局地気象については、防災面からも予測精度を改善することが求められている．また，農業など様々な産業や，太陽光，風力，小水力発電など再生可能エネルギーにも大きく影響するため，経済的な側面からも地域詳細かつ高精度での予測が必要とされている．現在，気象庁ではメソスケールモデル（MSMGPV）による 39 時間予報を実施しており，ほぼリアルタイムで予報値データが配信されている．機械学習を用いて数値予報値のバイアス補正を行うことにより局地降水予報の精度向上にも応用できる．既存の数値モデルの潜在能力を引き出して機械学習の手法をさらに発展させることにより，コストを大幅に抑制し，観天望気に加えてさらに精度の高い局地気象予報が可能になると期待される．

気象学を含む自然科学分野が扱うのは複雑系であり，起こった結果が偶然か必然かの判断ができないケースも多々ある．しかし，人には判断できなくても，機械学習により複雑な現象から規則性が見出されるかもしれない．規則性を見出すことができれば，それにより予測が可能になるだろう．従来の方法では検出できなかったパターンについても，様々な機械学習の手法を用いることにより発見できるかもしれない．さらに，未知のメカニズムを解明し自然現象への理解を深めるための糸口を提供してくれるかもしれない．現時点において，気象分野では機械学習，人工知能が一般に広く取り入れられているわけではないが，今後さらに理解が進めば有効な研究手法の1つになるだろう．気象分野と人工知能研究分野の多くの方に関心を持っていただき，自然災害や地球温暖化など気象に関わる様々な環境問題を解決するきっかけとなることを期待したい．

5. まとめ

- 冬季の日本海側降水特性について，観測降水とモデル降水にサポートベクターマシンの回帰モデルを適用することにより，モデル降水のバイアスを大きく低減できた．
- 地域詳細の降水分布や風向風速分布パターンを認識することにより，局地降水を高い精度で予測できる．
- 機械学習は，複雑な現象から規則性（パターン）を認識して予測を行うことが可能であり，数値モデルと組み合わせることにより，局地気象予報の精度向上に有効である．

参考文献

[1] Kimura R, Numerical weather prediction. Journal of Wind Engineering and Industrial Aerodynamics. 90.12-15, 1403-1414, 2002.

[2] 山田力: 小型船舶安全運航のための観天望気を入れた気象海象の話,関東小型船安全協会, 2007, http://www.shoankyo.or.jp/kisyou/pdf/khn2.pdf.

[3] Acharya, S., Presage biology Lessons from nature in weather forecasting. Indian Journal of Traditional Knowledge, 10, 114-124. 2011.

[4] Met Office: Red sky at night and other weather lore, 2017, https://www.metoffice.gov.uk/learning/learn-about-the-weather/how-weather-works/red-sky-at-night.

[5] 小倉義光: メソ気象の基礎理論, 東京大学出版会, 1997.

[6] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall: Rainfall forecasting in space and time using a neural network. Journal of hydrology 137.1-4, 1-31, 1992.

[7] Kuligowski, Robert J., and Ana P. Barros: Experiments in short-term precipitation forecasting using artificial neural networks. Monthly weather review 126.2, 470-482, 1998.

[8] 帰納的機械学習手法の天気予報への応用, 気象研究所技術報告, 第 39 号 2000. http://www.mri-jma.go.jp/Publish/Technical/DATA/VOL_39/39_005.pdf

[9] WMO: Nowcasting. 2018.
http://www.wmo.int/pages/prog/amp/pwsp/Nowcasting.htm.

[10] 気象庁: 解析雨量・降水短時間予報・ 降水ナウキャストの改善. 2018.
http://www.jma.go.jp/jma/kishou/books/yohkens/18/chapter3.pdf.

[11] 気象庁: メソモデル・局地モデル. 2018.
http://www.jma.go.jp/jma/kishou/know/whitep/1-3-6.html.

[12] 気象庁: 解析雨量. 2018.
http://www.jma.go.jp/jma/kishou/know/kurashi/kaiseki.html.

[13] Giorgi, Filippo, and Maria Rosaria Marinucci.: A investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. Monthly Weather Review 124.1: 148-166. 1996.

[14] Barros, Ana Paula, and Dennis P. Lettenmaier: Dynamic modeling of orographically induced precipitation. Reviews of geophysics 32.3. 265-284. 1994.

[15] Haney, Robert L.: On the pressure gradient force over steep topography in sigma coordinate ocean models. Journal of physical Oceanography 21.4: 610-619. 1991.

[16] David Meyer et al.: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2017.
https://cran.r-project.org/web/packages/e1071/e1071.pdf

[17] Wilby, Robert L., et al.: Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA 27, 2004.

[18] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton: Deep learning. nature 521.7553, 436. 2015.

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

ランダムフォレストを用いた地震動予測式の構築

Construction of ground motion prediction equation using random forest

久保久彦[*1] 功刀卓[*1] 鈴木進吾[*1] 鈴木亘[*1] 青井真[*1]
Hisahiko Kubo Takashi Kunugi Shingo Suzuki Wataru Suzuki Shin Aoi

[*1] 防災科学技術研究所
National Research Institute for Earth Science and Disaster Resilience

In this study, we construct a new ground motion prediction equation, which empirically predicts peak ground acceleration, using a random forest method and strong-motion big data recorded in Japan. The obtained predictor can reproduce the features of observed ground motions, although we still have some problems such as the effect of imbalanced data-set.

1. はじめに

地震動予測式は地震による強い揺れ(強震動)を過去の記録に基づいて経験的に予測する手法の一つである[日本建築学会 2005] [翠川 2009]. 地震動予測式では地震動の特性を地震規模(マグニチュード)・地震の震源からの距離・地盤の増幅率などによってモデル化しており, 地震動の強さが震源からの距離によって減衰していく様を表すことから距離減衰式とも呼ばれる. その簡便さから, 地震工学の分野では緊急地震速報や強震動予測などにおいて広く一般的に使われてきた.

本研究ではこの地震動予測式の構築を機械学習手法の一つであるランダムフォレストによって試みる. 先行研究では最小二乗法に基づく回帰によって地震動予測式を導出しており, 説明変数と目的変数の関係を表す関数の形を事前に仮定する必要があった[司 1999] [Morikawa 2013]. それに対して本研究ではランダムフォレストに代表されるノンパラメトリックな機械学習手法を用いることにより, 関数形を事前に仮定することなく, 関数形も含めて過去のデータから直接求めることができる. ニューラルネットワークなどの機械学習手法を地震動予測式の構築に用いた研究は既に行われているが[Derras 2012], 本研究では機械学習手法の利点を最大限に生かしながら, より高い予測性能を持つ地震動予測式の構築を目指す.

2. データ・手法

データセットには, 防災科学技術研究所が維持・管理する陸海統合地震津波火山観測網 MOWLAS の全国強震観測網 K-NET・基盤強震観測網 KiK-net [功刀 2009] [Aoi 2011]による強震ビックデータを使用する. これらの観測網は 1995 年兵庫県南部地震による阪神・淡路大震災を契機として日本全国を均質に網羅する形で整備され, 観測された強震動記録は web を通じて公開されている.

訓練データとしては 1997 年 4 月から 2015 年 12 月までに発生した地震 計 2177 イベントによる強震記録 計 288,644 記録を使用した. またテストデータには 2016 年 1 月から 10 月までに発生した地震 計 223 イベントによる強震記録 31,273 記録を使用した. ここでは, F-net モーメントテンソル解カタログ[福山 1998]および気象庁一元化震源カタログに存在し, 4.5 以上 7.5 以下のモーメントマグニチュード(Mw)を持ち, 深さ 200km より浅い深さで発生した地震の記録のみを用いた. 地震の規模に上限を用いているが, 巨大地震(Mw 8 以上)は発生頻度が少なく, それらを学習データセットに組み入れると, 構築する地震動予測式において巨大地震の地震動予測が過去の数少ない記録に大きく引きずられる可能性があるためである.

本研究の予測対象である目的変数は地表最大加速度(PGA)である. PGA を予測するための説明変数として, 震源と予測地点の水平距離(震央距離, L), 震源の深さ(H), Mw, 予測地点の地下深さ 30m までの平均 S 波速度(Vs30), 予測地点の地下において S 波速度が 1400m/s に達するまで深さ(D1400)の 5 つを用いる. L の計算における地震の水平位置, H, Mw は F-net モーメントテンソル解カタログの情報を参照した. Vs30 については防災科学技術研究所の地震ハザードステーション J-SHIS (http://www.j-shis.bosai.go.jp)で公開されている表層地盤情報[松岡 2008] [若松 2013]から各地点での値を参照した. D1400 については J-SHIS の深部地盤構造モデル[藤原 2009] [藤原 2012]から各地点での値を参照した. 先行研究[Morikawa 2013]では浅部地盤および深部地盤による増幅を Vs30 および D1400 を用いてモデル化した上で, 地盤増幅の補正を行っており, 本研究でも同様の補正効果を見込んで Vs30 および D1400 を説明変数に加えた. 解析におけるデータセットの前処理として, PGA と L の対数値をとった上で, 各目的変数・説明変数それぞれについて標準化を行った.

解析には Python のオープンソース機械学習ライブラリ scikit-learn において実装されているランダムフォレストを用いる. 最大深さ 100 で, 最小説明変数が 1 つである決定木が計 2000 で構成されるランダムフォレストを使用した. これらの最大深さ・最小説明変数・決定木数は交差検定の結果から決定した.

3. 結果

図 1 と図 2 に構築した地震動予測式による予測値の距離減衰の様子を, 三つ異なるマグニチュードの場合の予測距離減衰について示している. また図1は深さ 8km と浅い地震, 図 2 は深さ 44km とやや深めの地震の場合である. それぞれにおいて距離が大きくなるに伴い PGA が小さくなる傾向, また地震の規模が大きいほど PGA が大きくなる傾向が見て取れる. この傾向は先行研究で構築された地震動予測式と調和的であり, 地震動の物理的特性と整合している.

また異なるマグニチュードにおける予測距離減衰を比べると, 地震の規模によって異なっており, またその違いが線形で表すことができないことも分かる. これは地震規模(震源継続時間)の

連絡先:久保久彦, 防災科学技術研究所, 茨城県つくば天王台 3-1, 029-863-7304, hkubo@boai.go.jp

違いが実体波と表面波という距離減衰が異なる 2 種類の地震の励起に与える影響の違いを反映しているためだと考えられる. 回帰式を事前に仮定せずに, 機械学習によって直接的に地震動予測式を求めた利点が出ていると考える. ただし規模の大きな地震はデータセットの中で記録数があまり多くなく, 更なる検証が必要である.

図 1 と図 2 では予測対象としている地震の深さが異なるが, 深さの違いが地震動予測式に与える影響も異なっていることが分かる. 浅い地震(図 1, 深さ 8km)の方が震源直近で大きな PGA が予測されているが, やや深い地震(図 2, 深さ 44km)ではより遠方まで比較的強い PGA が予測される結果となっている. この違いは線形で表すことができるような関係でないことも見て取れ, この点にもノンパラメトリックな手法を用いた本研究のメリットが出ていると考えられる.

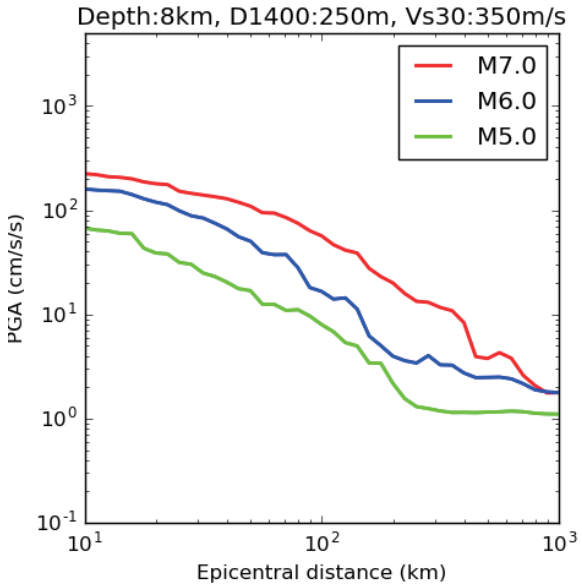


図 1　震源の深さ 8km・Vs30 350m/s・D1400 250m の場合における予測距離減衰. 赤線が Mw7.0, 青線が Mw 6.0, 緑線が Mw 5.0 の場合の予測を示す.

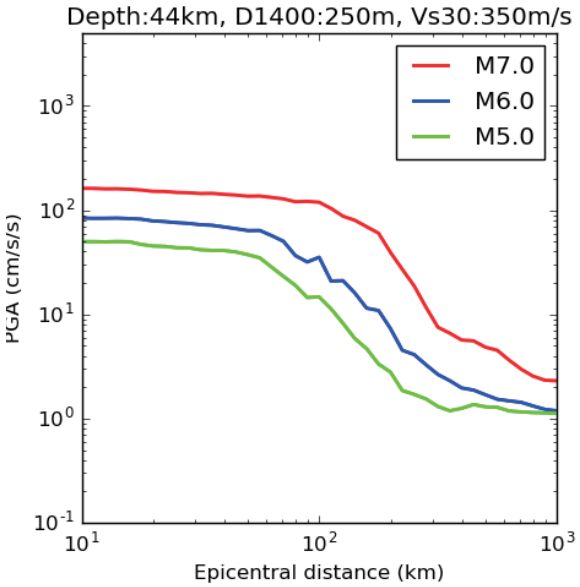


図 2　震源の深さ 44km・Vs30 350m/s・D1400 250m の場合における予測距離減衰. 赤線が Mw7.0, 青線が Mw 6.0, 緑線が Mw 5.0 の場合の予測を示す.

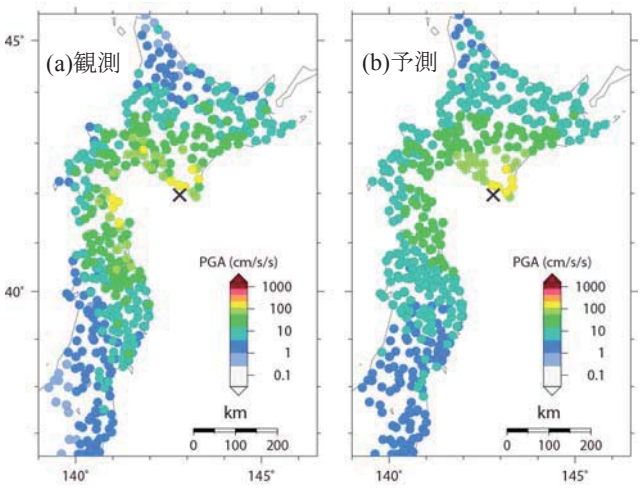


図 3 2016 年 1 月 14 日 12 時 25 分に浦河沖で発生した地震(Mw 6.6, 深さ 56km)における(a) 観測値と(b)予測値の分布. ×は震央(震源の水平位置)を表す.

また, これまでの地震動予測式は比較的大きな規模(M6 以上)の地震を対象とする場合が多かった. これは従来の手法では実体波(S 波)のみをターゲットとしており, 同じ距離減衰パラメータで異なる規模の地震全てに対応することが難しかったためである. 本研究はランダムフォレストを活用することにより, M5 程度の規模までをカバーする地震動予測式を構築することができた.

図 3 では 2016 年 1 月 14 日 12 時 25 分に浦河沖で発生した地震(Mw 6.6, 深さ 56km)における 観測と予測の比較を行っている. この地震は訓練データには含まれておらず、テストデータの一つである。距離に対する地震動強度の広がりを含め, 大まかな地震動の分布は再現できており, 実用レベルの地震動予測式が得られたと考える. しかしながら, 東日本では前弧(火山フロントの東側)と背弧(火山フロントの西側)で地下の減衰構造が異なることに起因して同じ距離でも前弧の方がより強い地震動が観測されるということが分かっており[宇津 2001] [森川 2006], この地震の観測記録(図 3a)でも見られるそのような傾向は予測結果(図 3b)では再現できていない. このような地域ごとの地震動特性を加味した地震動予測式へとするためには, 説明変数に地震の発生位置座標や観測地点の座標位置を加えていくことが考えられる.

構築したランダムフォレストの各決定木における説明変数へのアクセス数から算出した重要度は[L, Mw, H, Vs30, D1400]=[0.489, 0.190, 0.072, 0.136, 0.125]である. 震源距離 L と Mw の重要度がほかに比べて飛びぬけて大きく, 二つを合わせると全体の約 7 割を占める. それらに比べると他の説明変数の重要度は相対的に低いが, これらを考慮することによって地震動予測式の予測性能が向上することを確認している.

図 4 は訓練データとテストデータについて, 本研究で用意した観測値と予測値の比較を示しており, 観測と予測が完全に一致すれば対角線上に位置することとなる. 全体的には対角線上にのっているが, 観測値が大きいほど過小評価する傾向も見られた. これは, 学習データセットに偏りがある(小さな PGA の記録が多く, 大きな PGA の記録が少ない)ために生じていると考えられる. 今後, 気象庁や自治体などが持つ地震動記録を追加し, データセットをさらに充実させていくとともに, 不均衡データに対応した手法に改良していく必要がある.

4. 結論

本研究ではランダムフォレストと強震ビックデータを用いた地震動予測式の構築を図った．得られた地震動予測式は実用レベルの予測が可能であることを確認し，また機械学習手法を適用したことによるメリットも見られた．強震データの偏りに起因する問題や地震の地域特性の取り込み方については今後検討していく必要があると考える．

参考文献）

[Aoi 2011] Aoi, S., Kunugi, T., Nakamura, H., and Fujiwara, H.: Deployment of new strong motion seismographs of K-NET and KiK-net, in Earthquake Data in Engineering Seismology, Geotech. Geol. Earthquake Eng., Vol.14, pp.167–186, Springer, Dordrecht, Netherlands (2011)

[Derras 2012] Boumédiène Derras, Pierre-Yves Bard, Fabrice Cotton, and Abdelmalek Bekkouche: Adapting the neural network approach to PGA prediction: an example based on the KiK-net data, Bull. Seismol. Soc. Am., Vol. 102, pp. 1446–1461 (2012)

[福山 1998] 福山英一, 石田瑞穂, Douglas, S. Dreger, 川井啓廉: オンライン広帯域地震データを用いた完全自動メカニズム決定, 地震 2 (1998)

[藤原 2009] 藤原広行, 河合伸一, 青井真, 森川信之, 先名重樹, 工藤暢章, 大井昌弘, はお憲生, 早川譲, 遠山信彦, 松山尚典, 岩本鋼司, 鈴木晴彦, 劉瑛:強震動評価のための全国深部地盤構造モデル作成手法の検討, 防災科学技術研究所研究資料, 第 337 号 (2009)

[藤原 2013] 藤原広行, 河合伸一, 青井真, 森川信之, 先名重樹, 東宏樹, 大井昌弘, はお憲生, 長谷川信介, 前田宜浩, 岩城麻子, 若松加寿江, 井元政二郎, 奥村俊彦, 松山尚典, 成田章:東日本大震災を踏まえた地震ハザード評価の改良に向けた検討, 防災科学技術研究所研究資料, 第 379 号, pp. 197-200 (2012)

[功刀 2009] 功刀卓, 青井真, 藤原広行: 強震観測―歴史と展望―, 地震 2, S19-S24 (2009)

[松岡 2008] 松岡昌志・若松加寿江(2008) : 地形・地盤分類 250m メッシュマップ全国版に基づく地盤のゆれやすさデータ, 産業技術総合研究所, 知的財産管理番号 H20PRO-936 (2008)

[翠川 2009] 翠川三郎: 地震動強さの距離減衰式, 地震 2, S471-S478 (2009)

[森川 2006] 森川信之, 神野達夫, 成田章, 藤原広行, 福島美光:東北日本の異常震域に対応するための最大振幅および応答スペクトルの新たな距離減衰式補正係数, 第 6 巻第 1 号, pp. 23-3, (2006)

[Morikawa 2013] Morikawa, N. and Fujiwara, H.: A New Ground Motion Prediction Equation for Japan Applicable up to M9 Mega-Earthquake, Journal of Disaster Research, Vol.8, No.5, pp.878–888 (2013)

[日本建築学会 2005] 日本建築学会:地盤震動―(日本建築学会・編集), 丸善, 408pp (2005).

[司 1999] 司宏俊, 翠川三郎:断層タイプ及び地盤条件を考慮した最大加速度・最大速度の距離減衰式, 日本建築学会構造系論文集, No. 523, pp. 63-70(1999)

[宇津] 宇津徳治:地震学第 3 版, 共立出版, 2001, 310pp.

[若松 2013] 若松加寿江, 松岡昌志:全国統一基準による地形・地盤分類 250m メッシュマップの構築とその利用, 地震工学会誌, No.18, pp.35-38 (2013)

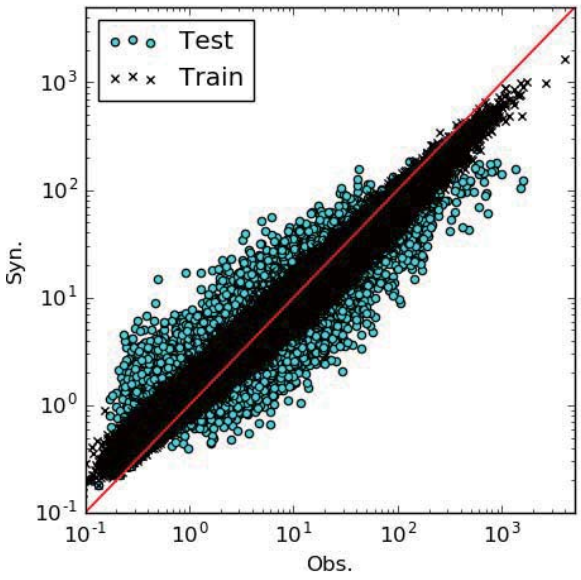


図 4 観測値と予測値の比較．○が訓練記録を，×がテスト記録を示す．

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

脳波を用いた両腕の運動方向分類器の開発

Development of the motion direction classifiers for bimanual movement using electroencephalogram

加藤正起[*1]
Masaki Kato

嶋田総太郎[*1]
Sotaro Shimada

[*1] 明治大学
Meiji University

Brain-machine interface (BMI) is a system that manipulates machines directly from the brain activity data. In daily life, there are many tasks that require cooperative movements of both arms, but most previous BMI employed unimanual movement, and few researches have assumed bimanual movement. The previous study in rhesus monkeys have shown that the BMI system developed for bimanual movement showed better performance than that combined two systems which were developed for left and right unimanual movements separately [Ifft 2013]. In this study, we developed classifiers based on electroencephalogram (EEG) data when bimanual or unimanual reaching movement was performed. The results showed that classifiers learned from EEG data with bimanual movement was more accurate than the classifier learned from EEG data with unimanual movement. Our results suggest that it is crucial to employ EEG data with bimanual movements to develop a BMI system for bimanual movements.

1. はじめに

ブレインマシンインターフェース（brain-machine interface : 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（electroencephalogram : EEG）や機能的核磁気共鳴計測（functional magnetic resonance imaging : fMRI）による研究では，脳活動からコンピュータ画面上のカーソルの操作[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．

Korik（2016）は，被験者に 3 次元空間上に配置したターゲットに対して左腕を動かさずに，右腕のみでリーチング運動を行わせた．測定した EEG から回帰分析によって運動軌跡予測器を作成したところ，EEG から再構成した運動軌跡と実際の運動軌跡の間に相関（r = 0.7）があり，運動軌跡を予測できたという結果を報告した．このように，機械学習を基にした回帰分析によって EEG から腕の連続的な運動軌跡を再構成した研究があるが，片腕の運動を想定した実験パラダイムになっている．一方，実生活においては，料理を作ったり，洗濯物を畳んだりと両腕の協調作業を必要とする場面は多いが，現在の多くの BMI の研究は片腕運動を想定した研究となっており，両腕の運動を想定したものは少ない．

サルを対象とした研究では，両腕運動時の脳活動というのは単純に片腕運動時の脳活動を組み合わせたものではないということが示唆されている[Rokni 2003, Steinberg 2002]．また両腕運動中は補足運動野および一次運動野などの皮質領域が特定の活動パターンを示すことが報告されている[Steinberg 2002]．アカゲザルを対象とした侵襲性の研究では両腕運動の軌跡予測において，片腕運動時の脳活動で学習した予測器を組み合わせたシステムよりも，両腕運動時の脳活動で学習し，両腕を同時に予測することの出来るシステムの方が，予測精度が良いということが示されている[Ifft 2013]．

本研究では，ヒトを対象としたシステムにおいても両腕運動を想定する必要があると考えた．ここでは，脳活動から腕の運動方向を分類する分類器を作成し，両腕運動時の脳活動で学習した分類器と片腕運動時の脳活動で学習した分類器ではどちらの方が両腕運動を想定した状況下における精度が良いか検討する．脳活動の計測には時間分解能が高い EEG を使用し，分類器には先行研究[Liao 2014]でも利用されているカーネルサポートベクトルマシン（kernel support vector machine : kSVM）を用いた．

2. 実験方法

2.1 被験者

3 名の健康な右利きの男性（22 ± 0.8 歳，平均 ± 標準偏差）が実験に参加した．ただし，1 名分のデータに関しては機器の不具合により，データが破損していたため，解析から除外し，2 名分のデータで解析を行った．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 のように液晶モニタ（LCD-MF276XDB-B,I-O DATA）にはゲーム開発ソフト（Unity, Unity Technologies）で作成したアバターアームが映し出されていた．このアバターアームは被験者に装着したモーションキャプチャー（Perception Neuron, Noitom Ltd.）によって被験者の動きと同期しており，被験者はアバターアームを操作することで，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の運動と画面上のアバターアームの運動の遅延時間は，ハイスピードカメラ（EX-F1, CASIO）で測定したところ約 107ms 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: 28ch
  (Fp1,Fp2,F7,F3,Fz,F4,F8,FC3,FCz,FC4,T7,C3,C1,Cz,C2,C4,T8,CP5,CP1,CP2,CP6,P7,P3,Pz,P4,P8,O1,O2)

連絡先：加藤正起，明治大学理工学部，神奈川県川崎市多摩区東三田 1-1-1，E-mail: ciakato07@gmail.com

- 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 : 1-100 Hz

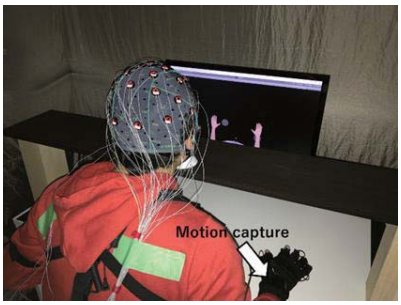


図 1　実験風景

2.3 実験手順

1 トライアルの流れは図 2 に示したように，被験者がアバターアームを一辺 8 cm のセンターターゲット上に置くことでトライアルの開始とした．センターターゲットの位置は毎回同じで，直径 8 cm の円のターゲットはセンターターゲットの位置から前後左右の 15 cm 離れた場所に出現させた．ターゲットは 1 つのアバターアームに対して，4 つの配置があり，16 通りの配置はランダムに出現させた．このトライアルを行う際には両腕でトライアルを行う bimanual 条件と左腕のみでトライアルを行う unimanual left 条件，右腕のみでトライアルを行う unimanual right 条件の 3 種類の条件があった．unimanual left 条件と unimanual right 条件時のトライアルは片方のアバターアームとセンターターゲット，ターゲットのみが画面上に映り，その間は使わない腕は動かさないように被験者に指示した．この実験は 2 セッションを行い，セッション内には 4 つのブロックがあった（図 3）．1 つ目と 4 つ目のブロックは bimanual 条件で行い，2 つ目のブロックは unimanual left 条件，3 つ目のブロックは unimanual right 条件で行った．それぞれのブロックは 150 トライアルが行われた．被験者は実験が行われる前に練習として 1 回分のセッションを，脳波測定器を装着しない状態で行った．「できるだけ早く，正確にリーチング課題を行う」という教示を与えた．

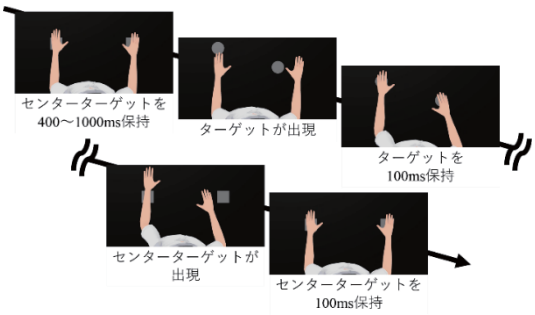


図 2　1 トライアルの流れ

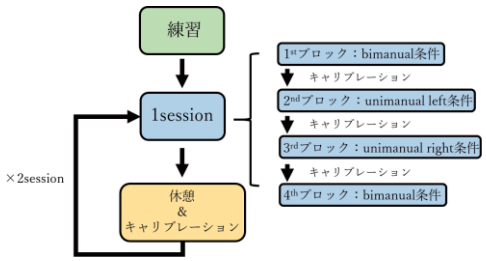


図 3　実験手順

3. 解析

3.1 前処理

測定した脳波データに対して，45 Hz までの低域通過フィルタをかけた．また脳波データに含まれる眼電成分を取り除くために独立成分分析を行った．眼電成分を取り除いた後，脳波データを 240 Hz までダウンサンプリングを行った．各電極のデータは Common Average Reference フィルタにかけた．その後，ウェーブレット変換によって時間-周波数解析を行い 1-3 Hz, 4-7 Hz, 8-11 Hz, 12-17 Hz, 18-27 Hz, 28-40 Hz の 6 つの周波数帯内でパワー値を平均化し，トライアルごとにデータを分割した．トライアル内でデータを運動前と運動後に分割した後，時間方向に平均化することで特徴量とした．subject 1 の Fp1,Fp2,F7,F8,T8,P7,P8 と subject 2 の F7,T7,T8,P8 に関してはノイズが多く含まれていたため，解析から除外した．

3.2 機械学習

脳活動から腕の運動方向を分類するためにデータを運動前と運動後に分類する 2 クラス分類 kSVM と運動後に分類されたデータのうち前後左右の 4 方向のどのクラスに属するかを分類する 4 クラス分類 kSVM の 2 種類の kSVM を組み合わせた．作成した運動方向分類器のモデルを図 4 に示した．それぞれの kSVM には 2 種類のハイパーパラメータがあり，最適な組み合わせを求めるためにグリッドサーチ[Dong 2017]を用いた．ハイパーパラメータの組み合わせの分類精度を確かめるために 10 分割交差確認法[Dong 2017]を使用して 10 回分の平均分類精度を求めた．最も良い平均分類精度を出力したハイパーパラメータの組み合わせを用いて，全ての訓練標本で kSVM を学習し，運動分類 kSVM と方向分類 kSVM を作成した．

kSVM に学習させる脳波データとクラスラベルを変えることで 4 種類の分類器を作成した．作成した分類器を表 1 に示した．最終的な分類精度は，bimanual 条件時の脳活動のテスト標本で検証することで，どの分類器が両腕運動に対応した分類器か検証した．

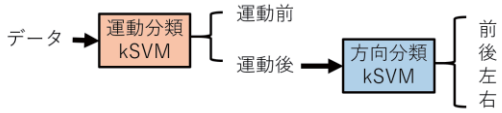


図 4　運動方向分類モデル図

表 1　分類器

分類器	学習データ	分類
bimanual left	bimanual 条件	左腕方向
bimanual right	bimanual 条件	右腕方向
unimanual left	unimanual left 条件	左腕方向
unimanual right	unimanual right 条件	右腕方向

4. 結果と考察

4.1 結果

subject 1 と subject 2 の運動分類 kSVM と方向分類 kSVM の学習時の分類精度の結果を図 5 と図 6 にそれぞれ示した．図の赤の点線はチャンスレベルを示す．ここで，運動分類 kSVM と方向分類 kSVM における学習時の分類精度とは，それぞれの

kSVM での最も高い平均分類精度のことを示す．運動分類 kSVM のチャンスレベルは 50%で，全ての運動分類 kSVM でチャンスレベルを超えていた．また，方向分類 kSVM のチャンスレベルは 25%で，全ての方向分類 kSVM でチャンスレベルを超えていた．以上のことから全ての kSVM で学習ができていると考えられる．

最終的な精度として各分類器に bimanual 条件時のテスト標本を入力したとき，入力したデータが運動前又は前後左右の 4 方向に正しく分類された精度をテスト時の分類精度として表 2 にまとめた．分類器のチャンスレベルは 31.25%で，全ての分類器でチャンスレベルを超えていた．両腕運動時の脳活動で学習した分類器の分類精度は subject 1 で 56.78 ± 0.71%（平均 ± 標準偏差），subject 2 で 57.21 ± 0.45%だった．片腕運動時の脳活動で学習した分類器の分類精度は subject 1 で 46.97 ± 1.64%，subject 2 で 49.78 ± 0.23%だった．以上のことからテスト時の分類精度は，両腕運動時の脳活動で学習したときの分類器の方が，片腕運動時の脳活動で学習した分類器よりも高かった．

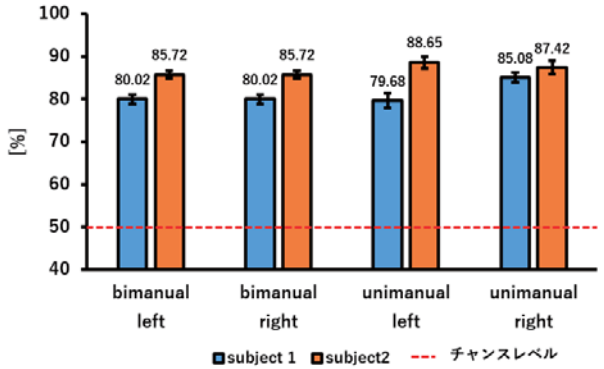


図 5　運動分類 kSVM における学習時の分類精度

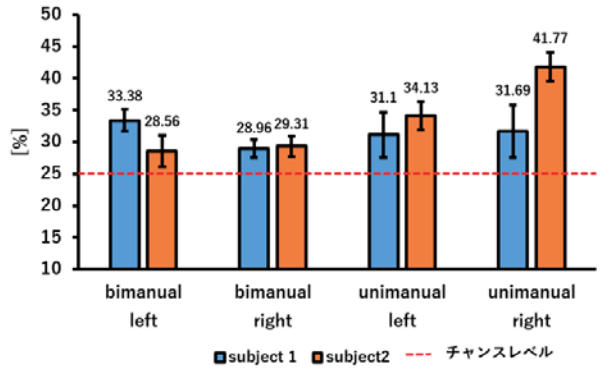


図 6　方向分類 kSVM における学習時の分類精度

表 2　テスト時の分類精度の結果

分類器	精度[%]		チャンスレベル[%]
	subject 1	subject 2	
bimanual left	57.48	56.76	31.25
bimanual right	56.07	57.66	31.25
unimanual left	45.33	49.55	31.25
unimanual right	48.60	50.00	31.25

4.2 考察

表 2 より，テスト時の分類精度において両腕運動時の脳活動で学習した分類器の分類精度の方が高いという結果になった．これは片腕運動時の脳活動で学習した分類器は両腕運動に対応したシステムとしては十分とはいえず，ヒトの EEG においても両腕運動に対応したシステムを作るには，両腕運動時の脳活動で学習する必要があることが示唆される．

5. 展望

本研究では，kSVM を学習させる際，グリッドサーチによってハイパーパラメータのチューニングを行ったが，特徴量の抽出・選択は行わなかった．今回使用したデータの特徴量は 28ch×6 つの周波数帯となっているため，最大で 168 個の特徴量を持つことになる．実際のデータはノイズが含まれる電極を取り除いているため，168 個よりも少ないが，機械学習においては不要な特徴量を出来るだけ減らすことが精度を上げる 1 つの手段である．今後は主成分分析や逐次後退選択などを行うことで特徴量の削減していく必要がある．また，時間方向に平均化することで特徴量としていたが，時系列情報を残した状態での分類や回帰分析によって腕の軌跡を再構成することで，より現実的な BMI の開発につながると考えられる．

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.

[Korik 2016] Korik, A., Sosnik, R., Siddique, N., & Coyle, D.: 3D hand motion trajectory prediction from EEG mu and beta bandpower, Progress in Brain Research, Vol.228, pp.71-105, 2016.

[Rokni 2003] Rokni, U., Steinberg, O., Vaadia, E., & Sompolinsky, H.: Cortical representation of bimanual movements, Vol.23(37), pp11577-11586, 2003.

[Steinberg 2002] Steinberg, O., Donchin, O., Gribova, A., Cardoso de Oliveira, S., Bergman, H., & Vaadia, E.: Neuronal populations in primary motor cortex encode bimanual arm movements, European Journal of Neuroscience, Vol.15(8), pp1371-1380, 2002.

[Ifft 2013] Ifft, P. J., Shokur, S., Li, Z., Lebedev, M. A., & Nicolelis, M. A. L.: A Brain-Machine Interface Enables Bimanual Arm Movements in Monkeys , Science Translational Medicine, Vol.5(210), 2013.

[Liao 2014] Liao, K., Xiao, R., Gonzalez, J., & Ding, L.: Decoding individual finger movements from one hand using human EEG signals, PLoS ONE, Vol.9(1), pp.1-12, 2014

[Dong 2017] Dong, E., Li, C., Li, L., du, S., Belkacem, A. N., & Chen, C.: Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain-computer interfaces , Medical and Biological Engineering and Computing, pp.1-10, 2017

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

脳表象モデルを用いた任意の視覚入力に対する知覚内容推定システム

A computational system estimating perception evoked by arbitrary visual inputs on the basis of modeling the perceptual representation in the brain

西田知史[*1,2]
Satoshi Nishida

西本伸志[*1,2]
Shinji Nishimoto

[*1] 情報通信研究機構
National Institute of Information and Communications Technology

[*2] 大阪大学
Osaka University

Although brain decoding techniques using functional magnetic resonance imaging (fMRI) have many potential real-world applications, the measuring cost of fMRI makes it difficult to realize many of such applications. Here, we propose a new decoding framework for estimating perceptual experiences evoked by visual materials with no additional fMRI measurement after model construction. Our framework consists of brain-activity prediction and perceptual decoding models constructed from individual brain data. Once the training of these models using movie-evoked fMRI data has been done, the framework combines these models and estimates each person's perceptual experiences regarding novel scenes without any additional fMRI measurements. Our results showed that our framework well estimated perceptual experiences that were evoked by novel scenes, and the estimated contents varied across the models for individual persons. Thus, our framework may provide a new computational system for estimating personal perceptual experiences, evoked by arbitrary visual inputs, via the perceptual representation in the brain.

1. はじめに

機能的磁気共鳴画像法（fMRI）で計測した脳活動から知覚内容や意図を解読する脳情報デコーディング技術では、近年、自然な映像により誘発される知覚内容が、映像や単語の形で解読可能になっている [Huth 16、Nishida 17、Nishimoto 11]。このような技術は、例えば CM 映像の印象評価サービスなどの形で、社会実装も進められている（参考：NeM sweets DONUTs®, http://www.nem-sweets.com/）。しかし、新たな視覚入力に対する脳活動解読を行うためには、多大な金銭的および時間的コストを必要とする fMRI 計測が必要であり、これが技術の社会実装を困難にする一因となっている。

一方で、深層学習 [LeCun 15]のような最新の人工知能システムは、任意の視覚入力から物体カテゴリのような知覚ラベルを、高い精度で推定することが可能で、様々な形での社会実装が進められている。しかし、人間並み、あるいはそれ以上に正確なレベルで画像や映像に客観的な知覚ラベルを付与することが可能な手法であっても、個人差を含むような主観的な知覚内容を推定することはまだ難しい。

そこで本研究では、任意の視覚入力により誘発される知覚内容を、少量の脳計測データからモデル化した脳内の知覚表象空間を介して推定する手法を提案する。これにより、追加の fMRI 計測を要さずに、任意の視覚入力を扱える脳情報デコーディング技術が実現できる。また提案手法を別の観点で捉えると、個人の脳内知覚表象空間のモデルを取り入れることで、個人差を含む人間らしい知覚推定を行うことが可能となりうる、人工知能システムとみなすこともできる。本研究では、この提案手法を、映像により誘発する知覚内容を単語の形で推定する解読モデル [Nishida 17] に適用し、従来のデコーディング技術である脳活動からの知覚推定と性能の比較を行ったうえで、提案手法の推定結果が知覚の個人差も反映しうることを検証した。

連絡先：西田知史，情報通信研究機構，s-nishida@nict.go.jp

2. 提案手法

2.1 脳活動予測モデル

脳活動予測モデルの構築（図 1A）は、符号化・復号化モデリング手法 [Naselaris 11] に準ずる。符号化モデリングは、任意の特徴空間を介した、知覚入力空間から脳活動空間への写像を学習する。fMRI の計測単位であるボクセルごとの脳活動の系列を **R**、知覚入力の系列を **S**、その特徴表現の系列を $f(\boldsymbol{S})$とすると、以下の数式により表現されるモデル重み $\mathbf{W_e}$ をリッジ線形回帰により推定する。

$$\mathbf{R} = f(\boldsymbol{S})\mathbf{W_e}$$

これにより、一旦 $\mathbf{W_e}$ が学習された後は、新たな $f(\boldsymbol{S})$が得られれば、誘起される脳活動 **R** の予測が可能なモデルとして機能する。

本研究では、特徴空間として、深層学習ネットワークの一種である 16 層 VGG [Simonyan 15] の中間層活性化パターンを利用した。深層学習ネットワークの中間層活性化パターンを利用すると、任意の視覚入力に対して活性化パターンが算出できる。また、同様の活性化パターンは、視覚入力に誘発されるヒト脳活動を精度良く予想することが報告されている [Güçlü 15]。知覚刺激として用いた映像の各フレームに対して、活性化パターンを計算したうえで 1 秒ごとに平均し、特徴表現の系列 $f(\boldsymbol{S})$としてモデル学習に用いた。モデル学習は被験者ごとのデータを別々に用いて行った。また、VGG の中間層のうち 8 層（5 つのプーリング層、3 つの全結合層）のそれぞれを用いて、別々にモデルの学習を行い、被験者一人あたり 8 個のモデルを作成した。

また、脳活動の予測精度向上のために、fMRI 信号が時間的に緩やかに変化する特性を考慮して、自己回帰成分（過去の脳活動系列）を説明変数に含む脳活動予測モデルも構築した。このモデルは、知覚入力の特徴表現系列に加え、過去の脳活動系列（主成分分析により 1000 次元に圧縮）も回帰項とするような線形回帰モデルとなる。このモデルも同様に、VGG の 8 つの中間層に対応する形で、被験者一人あたり 8 個のモデルを作成した。

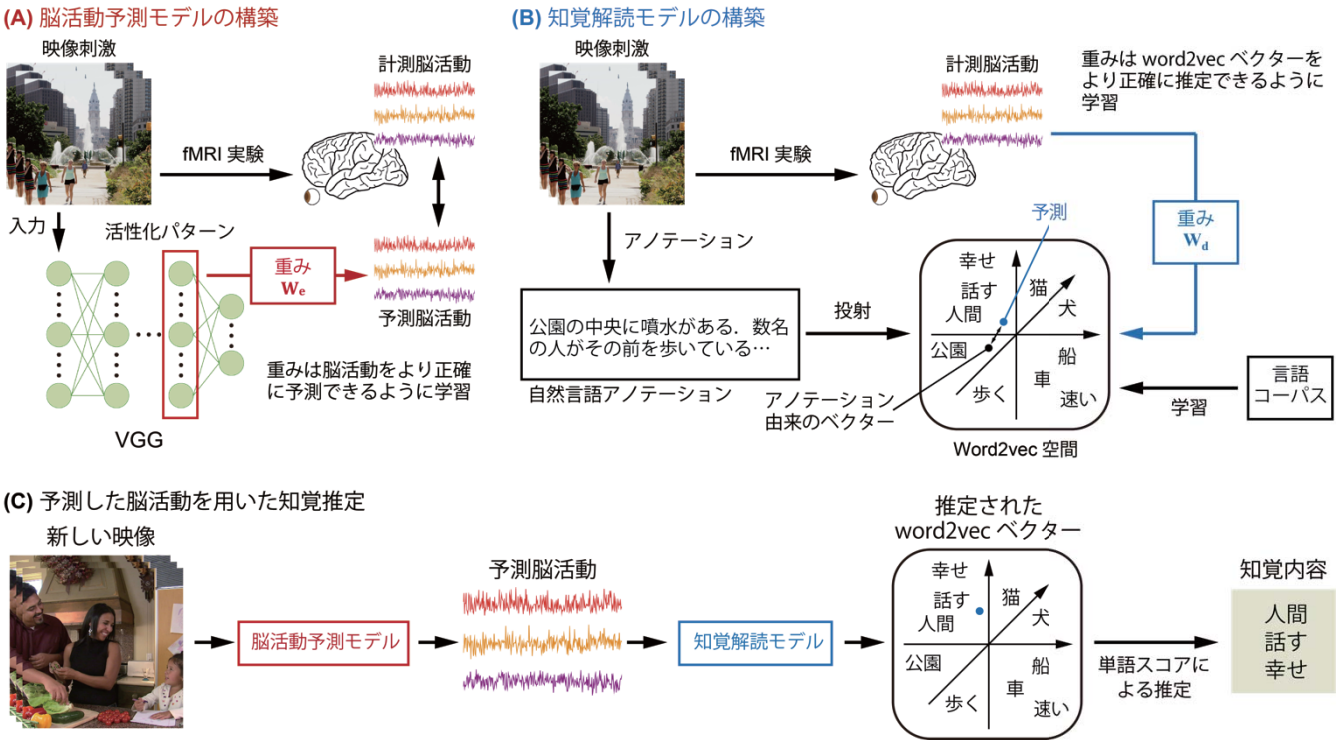


図 1：脳活動予測モデルと知覚解読モデルによる知覚推定の概要

2.2 知覚解読モデル

知覚解読モデルの構築（図 1B）は、復号化モデリングに相当する。このモデリングは、符号化モデリングとは逆に、脳活動空間から特徴表現空間の写像を学習する。つまり、以下の数式により表現されるモデル重み $\mathbf{W}_d$ をリッジ回帰により推定する。

$$g(\boldsymbol{S}) = \mathbf{R}\mathbf{W}_d$$

一旦 $\mathbf{W}_d$ が学習された後は、新たな $\mathbf{R}$ が得られれば、知覚内容を反映する特徴表現 $g(\boldsymbol{S})$ を推定可能なモデルとして機能する。

特に、本研究の知覚内容解読モデルでは、特徴空間に word2vec 空間 [Mikolov 13] を用いた。Word2vec 空間では、単語と単語の意味関係が、対応するベクター表現の空間内の位置関係として適切に表現されることが知られている。また、脳情報デコーディングに特徴空間として用いると、映像に誘発される知覚内容を高い精度で推定可能なことが報告されている [Nishida 17]。本研究では、100 次元の word2vec 空間を、日本語 Wikipedia コーパスを用いてあらかじめ学習しておいた。

知覚入力の word2vec 空間上の表現を得るために、知覚刺激として用いた映像の自然言語アノテーションを取得した。アノテーションは映像の 1 秒ごとのキャプチャ画像に対して、50 文字以上のシーン記述の形で収集した。また、1 枚の画像に対し 5 人以上からのアノテーションを得た。各アノテーションは形態素解析を行った後、word2vec の単語ベクター表現に変換し、1 枚の画像に対するアノテーションに含まれる全単語ベクターの平均として、映像の各シーンに対する特徴表現を算出し、モデル学習に用いた。

知覚内容の推定を単語の形で行う手順は次の通りとなる。新たな脳活動データが得られると、学習済みモデルの重みを利用して word2vec ベクターを算出した。そのベクターと、コーパスに頻出の 1 万単語のそれぞれに対応する word2vec ベクターの相関係数（単語スコア）を計算した。より高い単語スコアを持つ単語ほど知覚内容を反映する単語とみなし、単語スコアの上位単語を名詞・動詞・形容詞に分けてリストアップした。このようにして、単語による知覚推定を行った。

2.3 予測した脳活動を用いた知覚推定

新たな視覚入力に対する脳活動予測を行うために、入力となる映像をフレーム画像に分割し、VGG に入力して活性化パターンを算出した。その活性化パターンの値に、予め学習しておいた脳活動予測モデルの重みを適用し、脳活動を予測した。各ボクセルにおいて、脳活動予測値は 8 つの中間層のモデルに対応して 8 つの値が得られるが、学習データにおける交差検定により算出した予測精度（脳活動の予測系列と計測系列の相関）に基づき、そのボクセルにおいて最も精度の高いモデルの予測値を使用した。

続いて、その予測脳活動と活性化パターンを、自己回帰成分を持つ学習済み脳活動予測モデルに入力し、再度脳活動を予測した。同様に、学習データにおける交差検定により算出した予測精度に基づき、8 つの中間層に対応するモデルのうち、ボクセルごとに最も精度の高いモデルの予測値を使用した。

その後、全ボクセルについての脳活動予測値に対して知覚解読モデルを適用し、単語による知覚推定を行った（図 1C）。

3. MRI 実験

3.1 被験者

fMRI による脳計測実験には 8 名の被験者（男性 5 名、女性 3 名、23～40 歳）が参加した。全ての被験者から実験前に書面による同意を得た。また、実験プロトコルは情報通信研究機構の倫理審査委員会および安全審査委員会から承認を得ている。

3.2 MRI 計測

被験者の脳機能画像を Siemens 社の 3T MRI MAGNETOM Prisma を用いて取得した（64ch receiver coil、multiband gradient

	従来手法の結果（計測脳活動からの解読内容）			提案手法の結果（予測脳活動からの推定内容）		
	名詞	動詞	形容詞	名詞	動詞	形容詞
1	金髪	着る	優しい	金髪	着る	優しい
2	髪	憧れる	可愛い	女性	話す	親しい
3	髪型	かぶる	若い	男性	悩む	可愛い
4	服	嫌う	幼い	髪型	憧れる	幼い
5	黒髪	悩む	親しい	言動	嫌う	怖い
6	着用	慕う	悪い	独身	喋る	若い
7	女性	喋る	怖い	母親	殴る	悪い

図 2：1 つの映像シーンにおける知覚推定結果の比較

echo-EPI sequence [multiband factor = 6]、TR = 1000 ms、TE = 30 ms、flip angle = 60°、voxel size = 2×2×2 mm、matrix size = 96×96、number of slices = 72）。また同じ装置を用いて、脳構造画像の取得も行った。

3.3 映像視聴課題

被験者には、MRI スキャナー内のスクリーン（視角 28.0°×15.5°）に写し出される 1 スキャンあたり 10 分 10 秒間（最初 10 秒分の脳活動データは使用しない）の映像刺激を、16 スキャンに分けて呈示した。映像視聴中、スクリーン中央に呈示される小さな点を固視するよう被験者に教示した。

映像は動画共有サイトの Vimeo（https://vimeo.com）から、投稿者が明示的にダウンロードを許可しているもののみ入手した。映像には主に、映画の予告編、自主制作映画、プロモーションビデオ、風景映像などが含まれる。それらを 10 秒から 20 秒のランダムな長さで切り抜き、ランダムな順番で再びつなげることにより、10 分 10 秒×16 本の映像刺激を作成した。

16 スキャンで取得した脳活動データのうち、12 スキャン分はモデル学習用に用いるデータ（学習データ）であり、それらのスキャンにおいては、全て異なる映像を呈示した。残りの 4 スキャン分はモデル評価用に用いるデータ（評価データ）であり、データの SN 比を上げるため、各スキャンで 4 回同じ映像を繰り返し呈示し、それに対する活動平均を評価データとして利用した。

4. 結果

4.1 知覚推定における提案手法と従来手法の比較

評価データ取得時の 10 分間の映像に対して、提案手法を用いて、脳活動予測および知覚推定を行った。また一方で、同じく評価データの脳活動を使用して、知覚解読モデルのみを用いた場合の、従来の手法における知覚解読も行った。図 2 に、1 つの映像シーンにおいて、2 つの手法のそれぞれを用いて行った推定結果の例を示す。図中左の画像は、被験者が閲覧した、または脳活動予測モデルに入力した映像シーンを表す。図中右の単語リストは従来手法（左）および提案手法（右）による知覚推定結果を名詞・動詞・形容詞に分けて表示したものである。1 万単語の中から、それぞれ最も単語スコアが高かった（最も知覚内容を反映しているであろう）上位 7 単語を上から順に並べている。提案手法は、従来手法と比較しても遜色ない程度に、シーンを適切に記述する単語を推定結果として出力していることが見て取れる。

続いて、映像の全シーンの推定結果を利用して精度を計算した。精度は解読モデルが算出した word2vec ベクターと、シーンに対するアノテーションから算出した word2vec ベクターの間の相関係数で定量化した。そして、その精度について従来手法と提案手法で比較を行った（図 3）。その結果、8 名の被験者全員において、提案手法はむしろ従来手法に比べて高い推定精度を示すことが分かった（Wilcoxon test、$p < 10^{-14}$）。

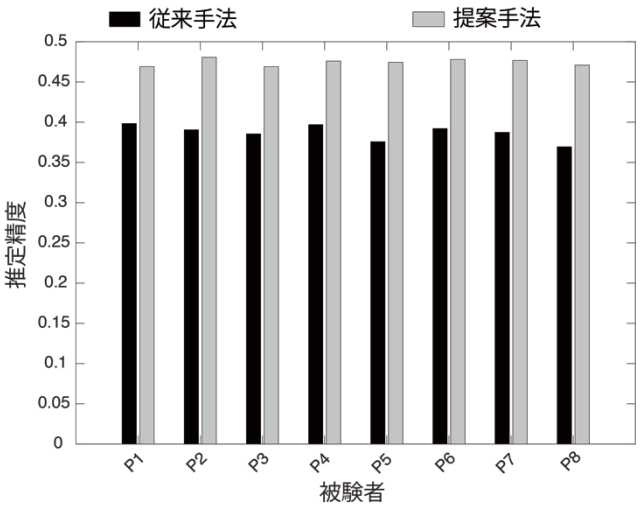


図 3：推定精度の比較

4.2 知覚の個人差との関係

過去の研究において、word2vec による知覚解読モデルを用いた際の、各シーンにおける知覚推定結果の被験者間のバラつきが、同じシーンにおけるアノテーションのアノテーター間のバラつきと相関することが報告されている [Nishida 17]。これは、映像によって誘起される知覚の個人間のバラつきが、知覚解読結果に反映されることを示唆している。まずその再検証のため、本研究の評価データにおいて、アノテーションのバラつきをアノテーション由来の word2vec ベクターにおけるアノテーター間距離の総和として評価し、知覚推定のバラつきを脳活動から解読した word2vec ベクターにおける被験者距離の総和として評価した。そして、それらの相関を計算したところ、確かに有意な相関が認められた（$r = 0.20$、$p < 10^{-6}$）。次に、提案手法の推定結果においても、同様の相関関係が成り立っているか検証するため、脳活動から解読した word2vec ベクターの代わりに、提案手法により推定された word2vec ベクターの被験者間距離の総和を用いて、相関を計算した。その結果、有意な相関が認められ（$r = 0.20$、$p < 10^{-5}$）、提案手法による知覚推定の結果も、映像によって誘起される知覚の個人間のバラつきを反映することが示唆された。

5. 考察

本研究では、符号化・復号化モデリングの枠組みを応用して、任意の映像に対する知覚内容の推定を、新たな fMRI 実験を要さずに行う手法について提案を行い、その性能を検証した。現段階では、性能を厳密に評価するため、少量のデータに対して手法を適用し評価を行ったが、従来手法に劣らない性能を示し、またその推定結果は知覚の個人差を反映しうることが示唆された。今後、さらに大規模なデータに対しても提案手法を適用し、有効性を確認することで、提案手法が fMRI 計測コストを劇的に削減する、画期的な知覚推定手法として脳情報デコーディング技術の応用可能性を大幅に広げることが期待される。

知覚推定精度の比較では、提案手法は従来手法より高い精度を示した（図 3）。脳活動予測では、計測脳活動を完全に予測できるわけではなく、この予測のための誤差が知覚推定の精度も下げると直観されるが、なぜこの直観と相反する結果となったのだろうか？一つの理由として考えられるのは、脳活動予測によって、計測脳活動に内在する計測ノイズや課題と無関係な変

動が除去された可能性である。脳活動予測モデルは 2 時間分の学習データに基づいて構築されるため、ノイズや変動を無視した純粋な脳応答信号をモデル化したと考えられる。本提案手法で採用した符号化モデリングは、過去の研究においても、時間周波数視覚特徴の脳内受容野マップ推定 [Nishimoto 11] や、意味情報の脳内表現マップの可視化 [Huth 12]などで有用であったように、脳内情報表現の定量化とそれに基づく脳活動予測を行ううえで効果的に機能するといえる。

提案手法を、脳表象を取り入れた知覚推定システムとして、既存の人工知能技術の発展とみなしたとき、どのような利点があるだろうか？一つの利点は、現状の人工知能技術では性能を上げにくい、人間らしい主観にもとづく識別などの課題への適用である。実際、過去の研究では、既存の識別モデルや自然言語処理モデルに脳表象を組み込み、人間の主観が強く影響するような課題に対して、モデルの精度向上を確認した例がある [Fong 17、Fyshe 14、Ruan 16]。本提案手法でも、今後様々な課題において、既存の人工知能技術との性能比較を行っていく予定である。もう一つの利点は個人差の反映である。提案手法では、個々人の脳を基にシステムが構築できるため、本研究でも示したように、知覚の個人差をうまく取り込める可能性がある。今後さらなる検証が必要ではあるが、知覚の個人差つまり個性を人工知能に実装することができ、特定の人物の代役をつとめるコピーロボットの開発や、個性のデジタル・アーカイブ化につながる技術になることが期待される。

以上のように、提案手法は、脳活動計測のコストを劇的に削減しつつも高い精度で知覚推定を行うデコーディング技術として利用可能であると同時に、個人の脳表象を組み入れることで主観や個人差を反映可能な新しい人工知能システムとしても利用可能となる。また、今回は単語で知覚推定を行う解読モデルを利用したが、提案手法の枠組みでは、任意の解読モデルに置き換えることができる。したがって、例えば感性や意思決定、経済的行動などを解読するモデルを適用することで、個性を考慮した知覚・認知推定システムとして、幅広い分野で提案手法が重要な役割を果たす可能性を秘めている。

参考文献

[Fong 17] Fong R, Scheirer W, Cox D (2017) Using Human Brain Activity to Guide Machine Learning. arXiv:1703.05463.

[Fyshe 14] Fyshe A, Talukdar PP, Murphy B, Mitchell TM (2014) Interpretable semantic vectors from a joint model of brain- and text-based meaning. In: ACL, pp 489–499.

[Huth 12] Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224.

[Huth 16] Huth AG, Lee T, Nishimoto S, Bilenko NY, Vu AT, Gallant JL (2016) Decoding the semantic content of natural movies from human brain activity. Front Syst Neurosci 10:1–16.

[Güçlü 15] Güçlü U, van Gerven MAJ (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J Neurosci 35:10005–10014.

[LeCun 15] LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444.

[Mikolov 13] Mikolov T, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 26:3111–3119.

[Naselaris 11] Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410.

[Nishida 17] Nishida S, Nishimoto S (2017) Decoding naturalistic experiences from human brain activity via distributed representations of words. Neuroimage, in press.

[Nishimoto 11] Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646.

[Ruan 16] Ruan Y-P, Ling Z-H, Hu Y (2016) Exploring semantic representation in brain activity using word embeddings. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 669–679.

[Simonyan 14] Simonyan K, Zisserman A (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556.

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

ガスセンサアレイによる人工嗅覚実現へ向けた試み

Metal-Oxide-Semiconductor Gas Sensor Array Toward Artificial Olfaction

斎藤 雄太[*1] 横山 誉宗[*1] 田中 貴久[*2] 内田 建[*2]
Yuta Saito Takamune Yokoyama Takahisa Tanaka Ken Uchida

[*1] 慶應義塾大学 理工学研究科 総合デザイン工学専攻
Faculty of Science and Technology, Keio University

[*2] 慶應義塾大学 理工学部 電子工学科
Department of Electronics and Electrical Engineering, Keio University

As key components of artificial olfaction, small, stable, low-power, and highly selective gas sensors are demanded. Metal oxide semiconductor (MOS) gas sensors satisfy some of those requirements. However, sensors in this category do not have high selectivity against target molecules. On the other hand, it is considered that human olfaction is realized by a number of various low-selectivity acceptors and processing of acceptor signals in brain. In this work, we prepared a sensor array consisting of three MOS-based gas sensors with different target molecules. Sensor array responses to three kinds of Sake were analyzed using machine-learning techniques. Accuracy of Sake indefication was achieved to be greater than 90%.

1. はじめに

人工嗅覚を実現させる要素となるガスセンサは，小型・低消費電力，高いガス選択性や安定性が必要であるといわれている．現在開発されているガスセンサには酸化物半導体式ガスセンサや膜型表面応力センサ，導電性ポリマーセンサなどがあり，ナノワイヤやグラフェンなどの新素材を用いた研究も行われている[Kumar 15]．しかし，これらの要求をすべて満たすようなガスセンサはまだ実現されていない．

一般に流通している酸化物半導体式ガスセンサは，主に二酸化スズや酸化亜鉛などの酸化物半導体が用いられ，200-400℃程度まで加熱して使用される．ガス分子が半導体センサ表面に吸着すると表面で酸化還元反応が生じ，半導体中のキャリア密度や移動度が変化し，センサ抵抗が変化する．このため，センサの小型化が容易であり，小型化により感度上昇と低消費電力を実現することができる．一方で，酸化還元反応を利用しているためにガス選択性が低く，対象となるガス以外にも大きな応答を示してしまう．これにより，応答が任意の対象ガスであるのか正確な判定が難しく，汎用なにおいの識別は難しい．[北村 17]

また，ヒトの嗅覚の構造は対象となる化学物質が異なる複数の受容体で構成されており，それぞれの受容体も複数の化学物質に応答する．これらの受容体からの電気信号を脳で処理することによってヒトはにおいを識別しているといわれている．最近では深層学習によって画像内の物体認識を行う研究が広く行われており，同様の手法によってにおいの識別を行える可能性がある．

本稿ではヒトの嗅覚を参考にし，異なる種類の酸化物半導体式ガスセンサを複数組み合わせることでセンサアレイを作成し，データ解析によるにおい分析を報告する．においの対象としてヒトがある程度識別可能である，日本酒の銘柄識別とした．データ解析には前処理として正規化処理およびデータ拡張を行い，判別器には深層学習を用いた．センサアレイ及びデータ解析のシステムによって日本酒の各銘柄および大気状態のリアルタイムでの判別を検証する．また，他種センサにおける前処理手法について考案を行う．

連絡先：斎藤 雄太，慶應義塾大学 理工学研究科 内田研究室，
神奈川県横浜市港北区 3-14-1 14-306，
e-mail:saito@ssn.elec.keio.ac.jp

2. 実験

2.1 センサアレイ

ユニバーサル基板上にセンサを固定し，センサアレイを作成した．使用したガスセンサは市販の酸化物半導体式ガスセンサ3 種類を用いた．センサ番号とセンサ種類，型番，サンプリングレートを表 1 に示す．

表 1 センサアレイで使用したセンサ

センサ番号	型番	対象ガス	サンプリングレート
Sensor 1	TGS2620	エタノール	1000 Hz
Sensor 2	TGS2602	VOC	1000 Hz
Sensor 3	TGS8100	水素	1 Hz

それぞれのセンサ回路（図 1）へ接続されている負荷抵抗の電圧値（V_{OUT}）を測定し，電圧値をセンサ応答とした．

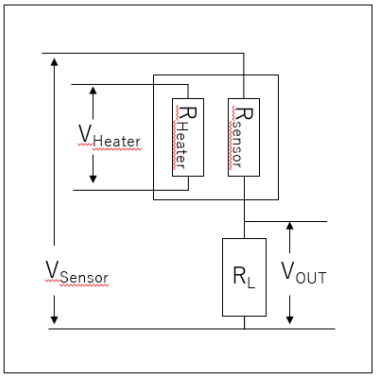


図 1 センサ部分回路概要

このセンサアレイを試験槽の蓋へ取り付け，試験槽内へ日本酒を入れた容器を設置し，蓋の開け閉めを繰り返すことにより日本酒のにおいの有無の切り替えを行った．一度の測定でおよそ 15 分間測定した．これをすべての日本酒にて複数回行った．反応概要を図 2 に示す．

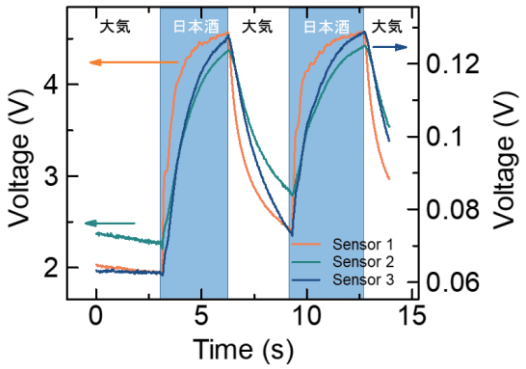


図 2　センサアレイ反応概要（日本酒：八海山）

2.2　データセット

においの状態をリアルタイムで判別するため，15 分間すべての測定を 1 データとするのではなく，30 秒前から現在時刻までの 30 秒間を 1 データと定義した．データセットのパラメータをまとめたものを表 2 に示す．また，1 データの概要を図 3 に示す．

表 2　データセット詳細

ラベル	データ個数
大気	5,065 個
久保田（日本酒）	754 個
八海山（日本酒）	1,055 個
越乃寒梅（日本酒）	2,111 個

1 データあたりのデータ点数はサンプリングレート×30 秒間より，Sensor 1 , 2 は 30,000 点，Sensor 3 は 30 点となっている．

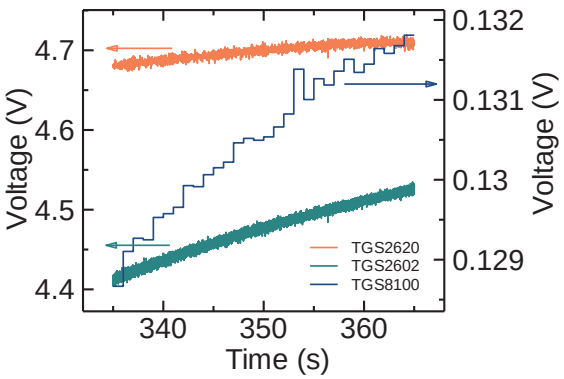


図 3　1 データ例（日本酒：八海山，334-364 秒）

3.　解析手法

3.1　前処理

深層学習で学習するに当たって，データセットへいくつかの前処理を行った．センサの仕様によりセンサ経路へ入力している電圧(図 1 V_{Sensor})が異なっているため，センサ応答のレンジが異なっている．これを入力電圧で割ることでレンジを 0~1 となるように正規化処理を行った．

センサ数に依存しない配列処理として，各センサ毎にレイヤーを作成し，それぞれのセンサへ 227x227 の配列変換を行った．また，データ点の不足分は 0 パディングを行った．（図 4）

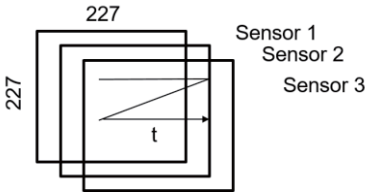


図 4　前処理概要

Sensor 3 においては，他のセンサよりデータ点数が少ないため，30 点で学習したモデルと，それぞれのデータ点を 1000 個に複製し，データ点数を 30,000 点へ拡張したデータセットも用意し，それぞれの比較を行った．また，センサ毎にデータを削減し，単一センサとみたてた場合の比較も行った．

3.2　深層学習

本研究では分類器として深層学習を用いた．ネットワーク部分は AlexNet をベースとした CNN を用い，最適化手法は Adam，反復回数は 1000 回とした．前処理を行ったデータセットは 8:2 に分割し，それぞれを訓練，テストセットとした．訓練セットによって学習したモデルをテストセットで検証した精度を表 3 に示す．

表 3　各データセットの日本酒識別精度

データセット	精度
単一センサ（平均）	81.73%
全センサ（データ拡張前）	80.97%
Sensor 3 データ拡張後	90.72%

データの拡張することによって 90%以上の精度得られた．単一センサとした場合と，データ拡張していない全センサの精度が同様であることから，各センサの特徴を生かせていないと考えられる．データ点数をそろえることによって各センサの特徴が現れるようになり，センサアレイの効果を得たにおいの識別が可能になった．

4.　おわりに

本稿では市販の酸化物半導体式ガスセンサによるセンサアレイとデータ解析によって日本酒の銘柄を識別した．判別精度として 90%程度の精度を得ることができた．また，データ点数が異なる場合，データ点数を拡張することによって精度向上の効果が得られることが示唆された．

今回の判別器では各センサデータを 2 次元配列にし，時系列情報へ周期性を持たせてしまっていたが，1 次元配列にし，より時系列情報の効果を得られるように改善をしていきたい．

参考文献

[北村　17]　北村雅季,:酸化物半導体ガスセンサのための材料作製技術：最近の研究動向, J. Vac. Soc. Jpn，2017.

[Kumar　15] Rakesh Kumar, Shweta Malik, B.R. Mehta: Interface induced hydrogen sensing in Pd nanoparticle/graphene composite layers, Sensors and Actuators, B: Chemical，2015.

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

IoTセンサデータの統合による商店街の賑わいの推定モデル

Estimation model of bustle of shopping streets by integration of IoT sensor data

吉野 碧 *2　家入 祐也 *2　菱山 玲子 *2
Aoi Yoshino　Yuya Ieiri　Reiko Hishiyama

*1早稲田大学経営システム工学専攻
Graduate School of Creative Science and Engineering, Waseda University

In recent years, the decline of shopping districts is regarded as a problem. However, the situation that is said to be declining, bustling is not clear. IoT is used to quantitatively show this situation. In this research, IoT sensors are installed according to the characteristics of shops in shopping districts. By analyzing the collected data, we clarify the data combination that can explain the bustle of the shopping street. As a result, it became possible to quantitatively express bustle of shopping streets by using Iot sensor. And I was able to discover the periodicity of bustling. By increasing the number of stores to be installed, it is expected that more sophisticated analysis on bustling will be possible.

1. はじめに

近年，商店街の衰退が問題視されている．これは大型店の進出による来店数の減少など様々な要素から説明されているが，明確な衰退，賑わいの状況は表せていないため，IoT(Internet of Things) に注目し，活気ある商店街づくりを検討するための情報を，商店街に設置したセンサデータからどのように獲得できるかを検討する．これらのデータは，商店街の様々な店舗の業態に合わせてセンサを設置することで収集し，分析する．

2. 関連研究

本節では商店街の賑わいに関する研究について触れる．濱名ら [1] は，歩行者空間の面積当たりの歩行者数を賑わいを表す指標として用い，現地調査を通じて商店街の賑わいに与える影響を定量的に明らかにした．また，朴ら [2] は道路の面積当たりの各道路の歩行者の数を賑わいを表す指標とし，都心商業地区内における歩行者整備状況等が賑わいに与える影響を定量的に明らかにした．これらの研究の特徴は，歩行者密度を賑わいとしている点にある．これらには，商店街の個々の店舗における賑わいを表現することができないという問題がある．

そのため，活気ある商店街づくりを検討するための情報とするため，本研究の特徴 2 点を示す．1 点目は，賑わいについて，商店街を構成する店舗，および商店街全体の賑わいを説明するためのデータ要素とその組み合わせを明らかにするという点．2 点目は，賑わいを，過大な現地調査のコストをかけず，IoT センサを用いたデータ収集と分析から明らかにする点である．なお，今回，賑わいはセンサより推定された入場人数と購買人数の和と定義した．

3. 研究目的

前節で示したとおり，本研究は，商店街の賑わいの状況がどのようなデータ要素から説明されるかを明らかにすることに意義がある．そして，賑わいの評価をするためには，これらを定量的に探る方法と指標を示す必要があり，そのためのデータを IoT デバイスの活用により取得する．しかし，商店街における IoT デバイスの利活用法も明らかではないため，一定の指針を示す．このことにより，従来の現地調査とは異なる新たな分析手法の可能性を示す．また，これを方法論として構築することで，店舗の特性に応じたセンサリングの提案が可能となり，そこから店舗間の関係や，商店街全体としての傾向も把握できる可能性がある．

連絡先: 吉野碧，早稲田大学経営システム工学専攻，〒 169-0072 東京都新宿区大久保 3 － 4 － 1，03-5286-3000，blue.yoshino@fuji.waseda.jp

4. 研究手法

図 1 は，本研究の IoT ネットワークを示している．IoT データは広域でコストをかけずデータ収集を行える LPWA 経由でクラウド上のデータベースに蓄積される．そして，クラウド上に保存したセンサデータを統合し，賑わいを推定するためのデータとする．これらのデータを分析することで大まかなデータの傾向や周期性を発見することができる．実際にセンサを設置する際には，商店街を構成する店舗のうち，業態の異なる店舗をいくつか選択し，仮説として「賑わい」の検出に役立ちそうなセンサを設置し，データを収集し蓄積する．また，本研究においては，賑わいという指標を IoT データの統合により表現する過程自体がモデルとしている．

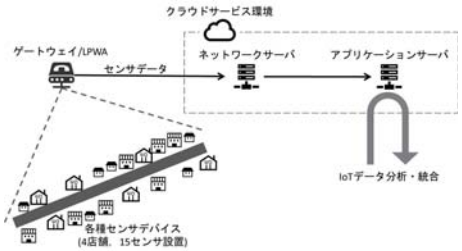


図 1: IoT ネットワークの概要

5. 実験概要

実験は新宿区内の商店街にて実施した．商店街の業態の構成比を基に，飲食業 2 店舗 (イートイン業態，テイクアウト業態 1 店舗ずつ)，小売業 1 店舗，サービス業 1 店舗にセンサを設置した．また，場所制約や店舗の特性に基づいて設置するセンサの種類，数，設置個所を決定した．実際に設置したセンサを表 1 に示す．実際に設置したセンサは以下のとおりである．

赤外線センサ 人感センサに用いられ，人が近づくと反応するため，入退場人数，購買人数を推定する．

リードスイッチ 磁気型近接センサとも呼ばれ，つがいとなる磁石が離れると電流が流れる．ドアの開閉回数やレジのドロアの開閉回数を推定する．

温度センサ 温度を測定する．賑わいと店舗内の温度の関係をみることができる．

音センサ 音の大きさを測定する．店内の騒がしさと賑わいの関係をみることができる．

表 1: 各店舗におけるセンサ設置個数

	赤外線センサ	リードスイッチ	温度センサ	音センサ
飲食業（イートイン）	0	1	1	1
飲食業 (テイクアウト)	1	1	0	1
小売業	2	0	1	1
サービス業	2	1	1	1

6. 結果と考察

今回，賑わいは 2 節で示したように，センサより推定された入場人数と購買人数の和と定義した．また，賑わい推定用のデータは，センサデータを 20 分を基準に統合している．データの統合過程を以下に示す．

1. センサデータに付与されているタイムスタンプを基に，基準となる時刻 (2017/12/23 00:00) からの経過分数を算出する．
2. 経過分数を 20 で除し，小数点以下を切り捨てた値を timeid としてデータに付与する．
3. データをセンサデバイスごとに分割する．
4. 各センサデバイスのデータに対して，timeid ごとにデータを合計し，統合する．この時，出入口に設置したセンサに関しては値を 2 で除す．
5. 各店舗に対して，各センサデバイスのデータを timeid ごとに入場者数と購買者数を合計する．これを，各店舗における賑わいとする．
6. 各店舗の賑わいを timeid ごとに合計する．これを商店街全体の賑わいとする．

図 2 はある一日における，個々の店舗を合計した商店街の賑わいの推移である．グラフから，店舗が営業していない深夜は賑わいが 0 に近く，店舗が営業している時間のうち，16 時から 18 時にかけて賑わいが高い状態にあることがわかる．また，サービス業の店舗の入口に同じ役割として赤外線センサとリードスイッチの両方を設置し，それぞれの検知数を比較したものを図 3 に示す．この図をみると赤外線センサのほうが全体を通して検知数が多く，センサの性質がここに表れている．リードスイッチは扉の開閉状況を検知しているのに対し，赤外線センサは周囲に人がいるかいないかを検知しているため，実際に出入りしていなくても検知範囲に人がいると検知されてしまうというようにノイズが混入しやすくなるためである．

今後の IoT センサの利活用の方法論を明らかにするため，今回使用したセンサの総評を以下にまとめる．より良い精度を得るためにはより長期間のデータ収集と適切な個所にセンサを設置することが必要不可欠である．更に，より多くのセンサを用いることでより詳細な賑わい情報が検出可能となる．

- 赤外線センサは設置可否が店舗の特徴によらないが誤検知は多くなる傾向にある．設置個所の工夫によって精度の向上が期待できる．
- リードスイッチは扉やレジのドロアがあるような特徴を持つ店舗にしか設置できない．その代わりに誤検知が少なく高い精度が期待できる．
- 音センサの感度は高くなく，小さな音では反応しないことが多い．細かい音を検知するには設置個所や感度を上げる工夫が必要となる．
- 温度センサは現状妥当な結果が得られている．ケースの通気性等を考慮することでより高い精度を得ることが可能である．

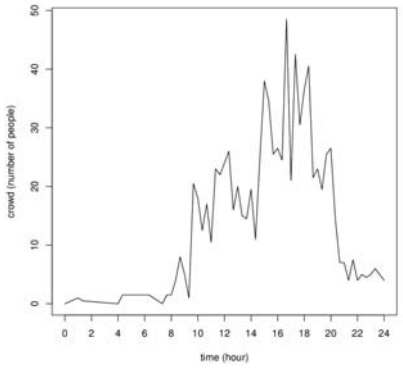


図 2: ある一日における商店街の賑わいの推移

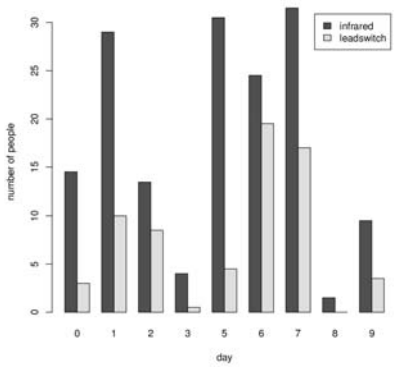


図 3: サービス業でのセンサごとの各日の検知人数比較

7. まとめと今後の課題

今回の研究では新宿区の商店街を対象に IoT センサを用いて賑わいを推定法，商店街における IoT センサ利用の知見が得られた．高い精度を得るためには，設置箇所への配慮が必要不可欠であるが，そのためにはデバイスのサイズによる設置個所の制約を克服し，より小型なデバイスを用いることによってより設置個所の自由度が増し，さらに，設置可能な個数が増えるためより精度よく，かつ詳細な賑わい情報を得ることが可能となることが予想される．

謝辞 本研究は JSPS 科研費 16K12553 の助成を受けたものである．

参考文献

[1] 濱名 智，中川 大，松中亮治，大庭哲治：歩行者に対する道路空間配分状況が商店街の賑わいに及ぼす影響に関する研究，都市計画論文集，Vol. 44, No. 3, pp. 85-90, 2009.

[2] 朴 東旭，中川 大，松中 亮治，大庭 哲治:現地調査に基づく都心商業地区の賑わいの要因構造に関する研究， 土木学会論文集 D3(土木計画学)， Vol. 68, No. 5, pp.513-521, 2012.

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

大規模展示ホール施設を対象とした段階的避難の有効性に関する検討

Examination of effectiveness of phased evacuation strategies in large-scale exhibition hall facility

西川 憲明[*1] Noriaki Nishikawa　廣川 雄一[*1] Yuichi Hirokawa　山田 武志[*2] Takeshi Yamada

印南 潤二[*2] Junji Innami　浅野 俊幸[*1] Toshiyuki Asano

[*1] 国立研究開発法人海洋研究開発機構 地球情報基盤センター
Center for Earth Information Science and Technology, Japan Agency for Marine-Earth Science and Technology (JAMSTEC)

[*2] 株式会社ベクトル総研
Vector Research Institute, Inc.

In recent years, crowd simulation has received a surge in attention. One of the reasons for investigating crowd flow is to optimize evacuation procedures from densely occupied buildings and urban centers in case of an unpredicted event. Such evacuation planning can save many lives at sudden catastrophic events like tsunami. In this paper, the effectiveness of phased evacuation strategies is examined in a large-scale exhibition hall facility. It is clearly demonstrated by numerical analyses that phased evacuation has positive effects on the mitigation of high-density pedestrian flows and heavily congested state in comparison with traditional simultaneous evacuation. Our approach has potential for contribution toward investigating disaster risk reduction strategies in complex urban space.

1. はじめに

大規模集客施設や高層ビルなどの極めて多くの人々が集まっている場所で災害が発生した場合，危険な状況から人々を迅速に遠ざけるとともに，避難途上における群集事故を発生させないような的確な避難誘導が必要とされる．そのためには，対象施設の空間特性や利用実態に応じた現実的な避難誘導対策を計画し，避難誘導計画が多数の災害状況に対して機能するかを事前に評価，検証しておくことが求められる[山下 11].

本研究では，群集避難シミュレーションを実施することで，大規模展示ホール施設を対象とした災害発生時の群集避難行動において，歩行者の安全性が危惧される高密度歩行者流あるいは歩行者の滞留現象の発生を低減させる方法として，避難開始時間を区域ごとに調整する段階的避難の有効性について定量的に評価することを目的とする．

超高層建築物に代表される多層の建築物においては，大規模な火災や非常事態の発生により，在館者の全館避難が強いられる場合の避難計画として，段階的避難の有用性が指摘されている．一方，本論文で対象とする大規模展示ホール施設は高層建築物というわけではなく，段階的避難の有用性についての理解は十分とは言えず，安全かつ効率の良い避難誘導計画を段階的避難の観点から検討することには大きな社会的意義があると考える．

2. 対象施設の概要

シミュレーションの対象施設としては，図 1 に示すような大規模展示ホール施設とする．施設 1 階に位置する展示ホールは総面積約 20,000 m^2，4 分割可能で，無柱構造となっている．施設図面のうち，黄色の破線で囲んだ範囲をシミュレーション対象領域とする．

本論文では，上記対象施設の来場者が展示ホールから 2 階に移動し，最寄り駅方面に向かう避難行動を研究対象とする．シミュレーション対象領域の空間構成については，1 階空間は展示ホール，施設屋内コンコース，屋外コンコースから構成される．2 階空間は施設屋内コンコース，屋外コンコース（デッキ）から構成される．

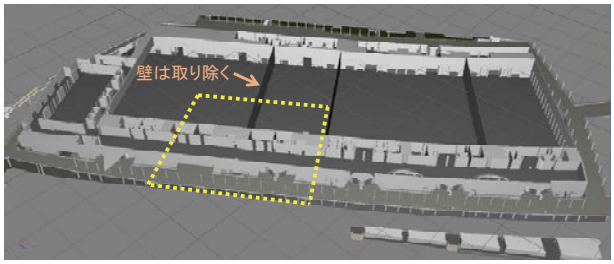


図 1 シミュレーション対象領域（黄破線領域内）．

展示ホール内には 6 つの進入不可領域を設定し，各進入不可領域の周辺に 6 か所の避難者エージェント発生領域を定義した（図 2）．各発生領域の大きさは1 m×1 m 程度とする．

歩行者発生領域は 1 階の展示ホール内部に計 36 か所設置されており，各領域から発生する歩行者エージェントの総数は同一とした．各発生領域からのエージェント発生速度は 1 人/秒とした．また，図 2 に示すように，各発生領域は B1 から B6 までの 6 ブロックに割り付けられているとした．

避難者エージェントの動線は図 2，図 3 に示すとおりである．本検討では避難者の最終目的地は 2 階のデッキ上とする．展示ホール内で発生した避難者エージェントは，展示ホール出入口から施設屋内コンコースに移動する．その後，屋内エスカレータ（ESC）もしくは外階段を利用して 2 階に移動し，最終目的地を目指すことになる．

連絡先：西川 憲明，国立研究開発法人海洋研究開発機構 地球情報基盤センター，〒236-0001 神奈川県横浜市金沢区昭和町 3173 番 25，nnishikawa@jamstec.go.jp

現在，エスカレータは建築基準法上，避難施設に該当しないため，災害時にどのように利用すべきかほとんど検討がなされていない．一般的には，避難経路としてエスカレータは認められておらず，非常時には停止させることが前提とされている．避難活用を検討する場合も，停止した状態で階段の代わりに使うことを想定することが多いが，災害時に避難すべき方向に走行させて安全に利用することができれば，避難時間の短縮に大きな効果があるだけでなく，階段での避難に困難がある高齢者や身体障害者の避難にも有効であると考えられる[岡田 11]．

以上の観点から，本研究では 2 階への通行手段としてエスカレータと階段の両方を利用可能とした場合を想定するものとする．

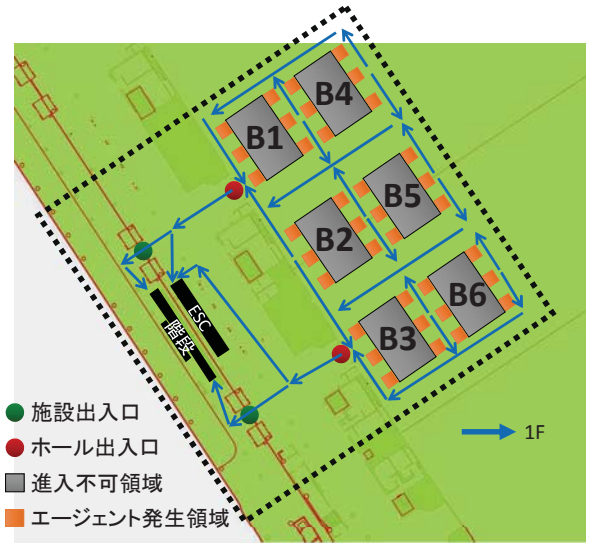


図 2 1 階の動線（黒破線が対象領域）．

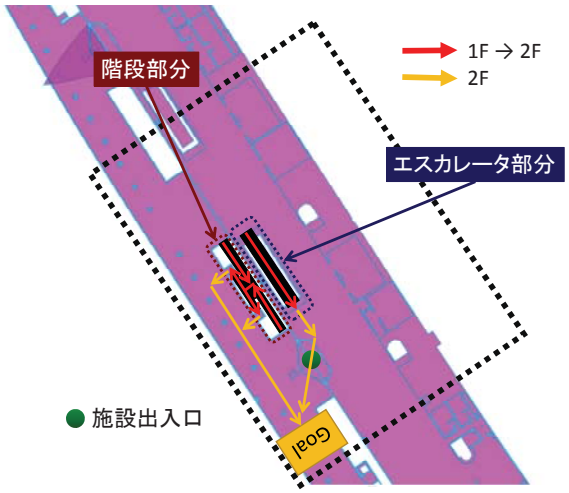


図 3 2 階の動線（黒破線が対象領域）．

3. シミュレーション結果と考察

大規模展示ホール施設の群集避難性状を明らかにするために，群集避難シミュレーションを実施した．ホール内在館者の避難行動を表現するシミュレーションモデルとしては既往論文[西川 17]を踏襲した．モデル詳細については当該文献を参照されたい．本研究では，展示ホール内にいる来場者が館内放送を受けて一斉に避難する場合と段階的に避難する場合の群集避難性状を比較することで，段階的避難の有用性について検討する．

段階的避難については，2 段階避難と 6 段階避難の 2 種類の仮想的な避難指示を想定した．ここで，2 段階避難とは，まずはブロック B1～B3 内の歩行者が避難を開始し，一定時間経過後，ブロック B4～B6 内の歩行者が遅れて避難を開始するものである．また，6 段階避難とは，ブロック B1 から B6 の番号順で，ブロック内の歩行者が一定時間遅れで順次避難を開始するものである．

歩行者の発生総数は 288 人，900 人，1512 人とした 3 ケースを想定した．また，各発生総数における計算結果は，擬似乱数のシードを変えて得られた 10 回のシミュレーション結果の平均値として算出した．

段階的避難における群集避難性状を概括するために，平均旅行時間と平均避難完了時間から構成されるパラメータ空間に，一斉避難，2 段階避難および 6 段階避難の計算結果をまとめてプロットしたものを図 4 から図 6 に示す．図中の秒数は段階的避難における避難開始遅れ時間を表している．一斉避難の計算結果は，避難開始遅れ時間を 0 秒としたものであることに注意されたい．

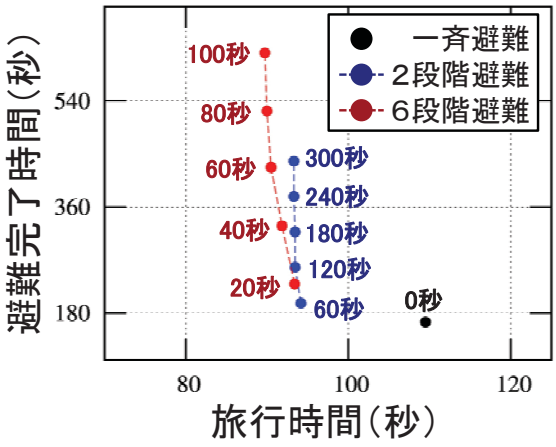


図 4 旅行時間と避難時間により表現される群集避難性状．歩行者数 288 人．

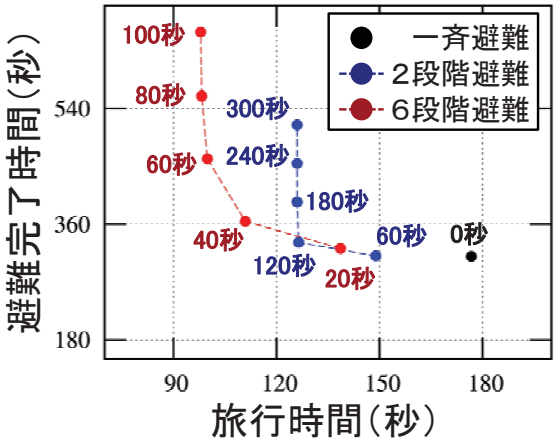


図 5 旅行時間と避難時間により表現される群集避難性状．歩行者数 900 人．

歩行者の発生総数が 288 人の場合，一斉避難に対する段階的避難の優位性は見られない（図 4）．一斉避難の場合のシミュレーション結果を見てみると，混雑のないスムーズな歩行者交通流が実現されていることから妥当な結果であると考える（図は省略する）．しかしながら，歩行者の発生総数を 900 人，1512 人と増大していくと，言い換えると，避難途上における歩行者の滞留現象の顕在化にともない，一斉避難に対する段階的避難の有用性は向上していく傾向が確認される（図 5, 図 6）．

歩行者の発生総数が 900 人，1512 人の場合には，段階的避難における区域数および避難開始遅れ時間を適切に設定することで，避難完了時間の増加を抑え込みながら，旅行時間を大きく減少させる，すなわち群集滞留の発生を効果的に抑制できることが観察される．

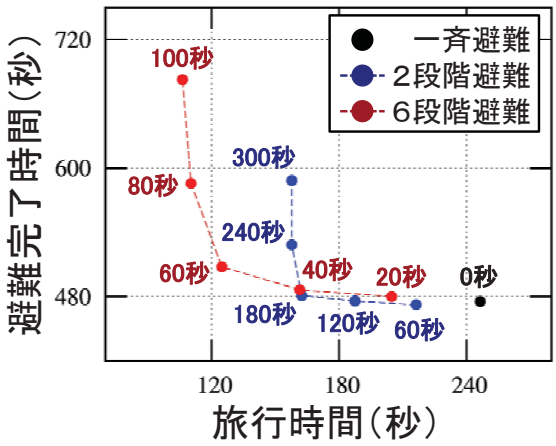


図 6 旅行時間と避難時間により表現される群集避難性状．歩行者数 1512 人．

最後に，避難途上における群集の滞留パターンを把握するために，歩行者の発生総数を 1512 人とした場合の群集避難行動のシミュレーション結果を図 7，図 8 に示す．

青線は 1 階の障害物線分，青丸は 1 階を移動中の避難者，赤線は 2 階の障害物線分，赤丸は 2 階を移動中の避難者を表現している．

一斉避難の場合のシミュレーション結果（図 7）を見てみると，シミュレーション開始から 1 分経過後には，ホール出入口およびエスカレータの乗り口付近において避難者の滞留が確認できる．2〜3 分経過後には，ホール出入口の滞留は完全に解消されるが，階段の上り口付近に新たな滞留が発生するとともに，エスカレータの乗り口付近における滞留が顕著化する．その後，避難者の滞留規模は徐々に縮小していき，7 分経過後には大多数の避難者が最終目的地に到着することとなる．

一方，段階的避難の場合のシミュレーション結果（図 8）を見てみると，エスカレータ乗り口および 2 階施設出入口付近で小規模な滞留は発生するものの，概ねスムーズな歩行者交通流が実現されていることが確認できる．

4. おわりに

本稿では，大規模展示ホール施設を対象とした災害発生時の群集避難行動において，歩行者の安全性が危惧される高密度歩行者流あるいは歩行者の滞留現象の発生を低減させる方法として，避難開始時間を区域ごとに調整する段階的避難の有効性を定量的に明らかにした．

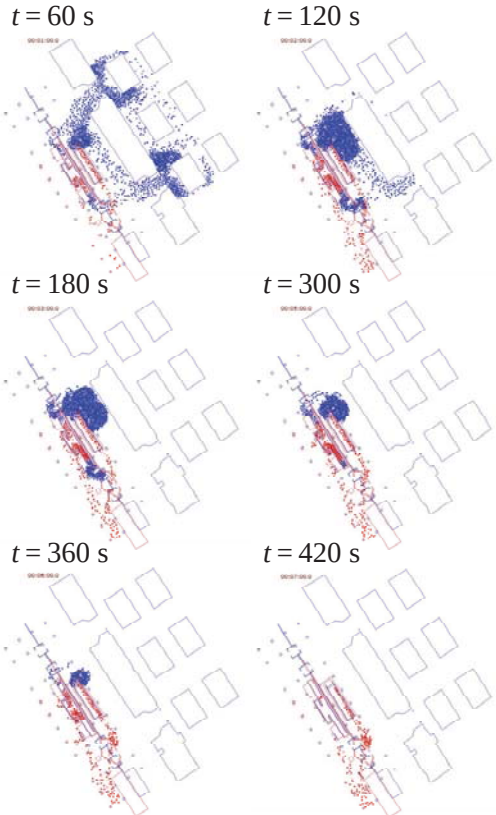


図 7 一斉避難における群集流動．歩行者数 1512 人．

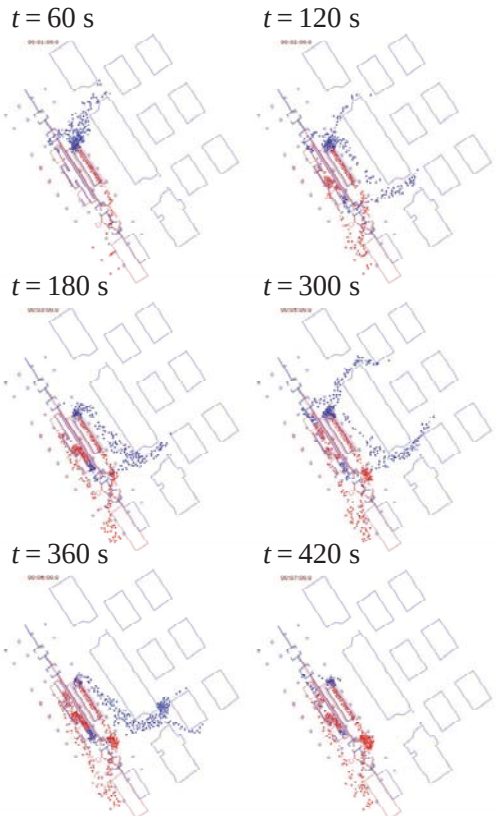


図 8 6 段階避難における群集流動．避難開始遅れ時間 60 秒，歩行者数 1512 人．

今後は，シミュレーション対象領域を展示ホール施設全体に拡張し，展示ホール内の在館者約 2 万人を対象とした群集避難シミュレーションを実施する計画である．シミュレーション結果を用いて，避難誘導計画をより良いものにするための施設管理者との協働的な検討を進め，安全・安心による施設価値の一層の向上を目指せればと考えている．

謝辞

本研究は，文部科学省ポスト「京」萌芽的課題 2「複数の社会経済現象の相互作用のモデル構築とその応用研究（多層マルチ時空間スケール社会・経済シミュレーション技術の研究・開発）」の元で実施したものです．また，本研究は JSPS 科研費 JP17K00328 の助成を受けたものです．

参考文献

[山下 11] 山下倫央，副田俊介，大西正輝，依田育士，野田五十樹：“センサデータマイニングを活用した安全安心な避難誘導への取組み”，電子情報通信学会誌，vol.94, no.4, pp.294-298, Apr., 2011.

[岡田 11] 岡田尚子，長谷見雄二，森山修治，岡本衣未：“エスカレータを用いた上方避難に関する実験研究”，日本建築学会環境系論文集，vol.76, no.668, pp.855-862, Oct., 2011.

[西川 17] 西川憲明，廣川雄一，山田武志，印南潤二，浅野俊幸：“大規模展示ホール施設を対象とした群集避難シミュレーション”，合同エージェントワークショップ＆シンポジウム 2017 (JAWS2017) 予稿集，2017.

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

乗換案内ログとSNSの融合による未来に発生する混雑原因の特定

Why does the station get crowded irregularly in the near future?

山下 達雄[*1] 坪内 孝太[*1] 丸山 三喜也[*1] 山浦 優樹[*1] 岡田 宏一朗[*1]
Tatsuo Yamashita Kota Tsubouchi Mikiya Maruyama Yuuki Yamaura Koichiro Okada

[*1] ヤフー株式会社
Yahoo Japan Corporation

In previous work of future congestion prediction using transit search logs, whether target station will get crowded on target time or not is predictable, but we cannot specify the cause of irregular congestion. In order to specify the irregular congestion cause, we analyzed posts containing information on future events from SNS. By combining congestion prediction information from the transit search log and event information from SNS, it was possible to extract the cause of congestion accurately.

1. 序論

世の中においての人の動きを予測することは重要である。特に、いつどこで混雑するのか、何が原因で混雑するのかを知ることができれば、様々な応用が考えられる。その応用の一つとして、乗換案内サービスがある。乗換案内サービスは、移動元と目的地の 2 つ駅と移動日時を入力として路線を検索し最適な経路を提示する。その際、該当経路で異常な混雑が予測される場合、注意喚起を行うことでユーザの利便性が向上する。

このような観点から我々は、路線検索ログから未来の異常混雑予測を行う手法を提案した[坪内 17]。これにより、「いつどこで(どの駅で)」予期せぬ混雑が発生しそうかの予測できるようになった。しかし、この方法では「何が原因で」混雑しそうなのかは分からない。実用上の観点から、ユーザに原因も提示することで安心感を与えたい。そこで、本研究では、この未来に発生すると予測される異常混雑の原因を特定する手法の提案を行う。

混雑原因を特定するには、イベント開催予定情報やイベントについて述べられている SNS 投稿などの言語リソースが必要となる。我々は、SNS である Twitter に着目し、未来のイベントについての内容が含まれる投稿(ツイート)を収集・整理して、注目度の高いイベントを抽出した。そして、これを既存の路線検索ログからの異常混雑予測と融合することにより、未来の日時に混雑が予想される駅に対してその混雑原因を提示することが可能となった。

本稿では、まず、融合要素の一つである、路線検索ログから未来の異常混雑予測を行う手法を説明する。次に、もう一つの融合要素である、Twitter から混雑場所と混雑原因を取り出す手法を説明する。そして、この二つの要素を結びつけることによる、混雑原因イベントを抽出する方法について述べ、その有効性を確認するための実験と評価について述べる。

2. 路線検索ログによる混雑駅の予測

本研究は、乗換案内の検索ログから異常混雑予測を行う坪内らの手法[坪内 17]をベースとしている。

坪内らの研究では、まず過去の路線検索データを集め、時間ごとの検索量の通常パタンを導出した。たとえば「A 駅から B 駅まで、火曜日の 10 時 00 分から 10 時 10 分の間はいつもであれば 80 回検索される」というパタンである。このパタンの導出にはバイリニアポアソン回帰モデルを用いた。

連絡先:山下達雄, ヤフー株式会社,
tayamash@yahoo-corp.jp

次に、近い未来の検索量を自己回帰モデルで推測した。つまり、「このままの様子で検索量が増えた場合に、最終的に何件になるか」という推定である。ユーザは移動の前に予めスケジュールを検索することが多い。その検索量を集め、数日後の検索量を予測した。

最後に導出された「いつものパタン」から推定される検索量と、「このまま増えたらどうなるか」という観点から推定された検索量を比較し、両者に乖離がある場合は異常混雑が発生する可能性が高いとした。

このように検索ログから未来に発生する異常混雑を予知することができる。しかし、この手法では混雑する駅と日時は予測できるが、混雑原因は分からないという問題がある。

3. Twitter による混雑スポットの予測

本節では、SNS (Twitter)からのイベント情報抽出について述べる。これは前節の手法では不可能である「未来の異常混雑の原因」を取得するための前段処理である。

未来のイベントについて言及しているツイートには、イベントの開催日(例:1 月 15 日)や開催場所(例:東京ドーム)が含まれているものがある。そのイベントに言及しているツイートすべてにこれらの情報が含まれているわけではないが、ツイート数の多い人気のイベントではそれらを含むものが一定数存在する。

異常混雑予測のためには人気のイベントを補足できればよいので、我々は開催日や開催場所に着目して情報抽出を行う。

まず、第 1 段階の処理として、未来の日付を含むツイートを収集する。例えば、今日が 7 月 24 日だとして、未来の日付である「7 月 25 日」について言及しているツイートを収集するには、本文中に「7 月 25 日」または「7/25」という文字列が含まれているもののみを選別する。これにより、日付をラベルとしたツイート集合ができる。

次に、第 2 段階の処理として、各日付を含むツイート集合からランドマーク(イベントの開催場所となりうるもの)を含むツイートを選別する。ツイートからのランドマークの抽出には、エンティティリンキングツールを用いる。具体的には、ランドマーク辞書のエントリ文字列にマッチさせ、曖昧性解消処理を経て、最終的に抽出結果を得る。ランドマーク辞書は、ランドマーク名と最寄り駅(複数)情報を持つ。本研究は、乗換案内の利便性向上を目的としているため、どの駅からも遠いランドマークは使用しない。

これら 2 段階の処理により、特定日付と特定ランドマークを同時に含むツイートの集合ができる。ツイート集合中のツイート数

が多ければ、その日時にそのランドマークで開催されると思われるイベントの人気が高い、つまり、当日そのランドマークが混雑すると考えられる。

しかし、当然、ノイズが多く含まれる。特定日付と特定ランドマークが含まれたツイートが多いという事実だけでは、イベントへの言及であるとは限らない。また、イベントへの言及であったとしても、必ずしも参加者が多いことを意味するわけではない。

4. 路線検索ログによる予測と Twitter による予測の融合

前節で説明した、ツイッターからの未来イベント抽出は、そのままで利用するにはノイズが多いという問題がある。しかし、本来の用途である、路線検索ログからの異常混雑予測に対する原因抽出という観点であれば、つまり、路線検索ログによる予測と融合すれば有効に利用できる。

前述の通り、ランドマーク辞書の各エントリには最寄り駅情報が含まれている。これを用いて、ツイッターによる未来日付イベント予測と路線検索ログによる未来日付駅異常混雑予測とを結びつける。未来日付イベント予測はツイート数を閾値として、ある一定の値を超えるもののみ利用する。

これらの処理によって、未来の日時に異常混雑が予想される駅に対してその混雑原因を予測することが可能となる。

4.1 評価

まず、Twitter からの未来日付イベント予測結果について、混雑の有無の判定を人手で行い、精度を出した(twitter only)。次に、駅名と日付を介して検索ログによる未来日付駅異常混雑予測と結びつくものに絞って、精度を出した(proposed)。また、未来日付イベント予測のツイート数を閾値として、精度の変化を見た(Threshold)。

データの期間は 2016 年 12 年 29 日～2017 年 1 月 18 日。路線検索ログによる混雑予測 209 件のうち、ツイッターによるイベント予測(ツイート数 5 以上)に紐づくものは 129 件であった。

評価は 3 名の判定者により行った。判定基準は、未来日付イベント予測のツイート集合から混雑要因となるイベント情報が得られるか、である。ただし、イベントが実際に混雑したかについては、実地での観測データ(正解データ)がないため、判定者各自がインターネット上の情報で確認を行った。最終的な判定は、3 名の判定の多数決とした。

結果を図 1 に示す。未来日付イベント予測の際のツイート数が多いほど、混雑イベント抽出能力が高い傾向にあるが、検索ログによる異常混雑予測と結びつけることにより精度が上がるこ

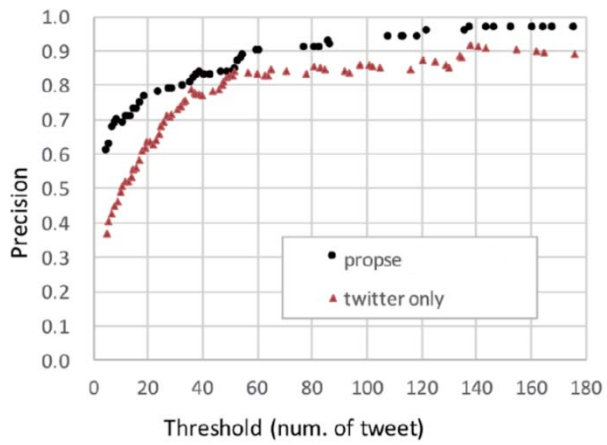


図 1

図 2

https://blog-transit.yahoo.co.jp/congestion/

とが分かる。つまり、融合により、高い Precision で混雑要因が抽出されていると言える。

4.2 サービスへの適用

融合により十分な精度が得られることが分かったが、実際の応用(サービス適用)に際しては、混雑要因をユーザに提示する必要がある。つまり、ツイート集合からイベント名やアーティスト名などの情報を抽出する必要がある。

我々はツイート集合に多く含まれるハッシュタグに着目した。Twitter による未来日付イベント予測のツイート集合(検索ログによる予測と紐づいたもの)17 件から取り出したハッシュタグの頻度上位 3 件を、判定者 1 名がイベント内容を表しているか否かを評価した結果、精度は 0.94 (16/17)であった。人手による NG ワードフィルタ等を併用すれば実用上十分な精度である。

以上の結果を踏まえ、乗換検索アプリに新機能として実装し、2018 年 2 月にリリースした(図 2)。

5. 関連研究

Twitter のテキスト(ツイート)から直接イベント情報を抽出する研究として[山田ら 16]がある。ランドマークで検索して集めたツイートを用い、パターンマッチングと機械学習で、イベント日付とイベント名を抽出している。これは近い未来に発生するイベントの予知ではあるが、混雑状況は正確に推測できない。

6. 結論

本研究では、路線検索ログからの異常混雑予測だけでは分からなかった混雑原因を、SNS (Twitter) データを解析して得たイベント情報と組み合わせることにより、抽出可能とした。その成果を踏まえて実際のサービスとしてリリースした。今後は、主にイベント情報の抽出精度向上を行っていく。

参考文献

[坪内ら 17] 坪内 孝太, 下坂 正倫, 小西 達也, 丸山 三喜也, 山下 達雄. 乗換案内データを用いた未来の混雑予測の研究, 2017 年度 人工知能学会全国大会講演集, 4I1-4in1, 2017.

[山田ら 16] 山田悠, 落合桂一, 菊地悠. Twitter を用いた地域イベント発見技術, NTT ドコモ R&D テクニカル・ジャーナル, Vol.23, No.4, 2016.

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

家庭電力データからのライフパターン分析に関する検討

Proposal of Lifestyle Analysis from Household Electricity Consumption Data

白井 佑 *1　服部 俊一 *2　高間 康史 *1
Yu Shirai　Shunichi Hattori　Yasufumi Takama

*1首都大学東京大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

*2一般財団法人電力中央研究所
Central Research Institute of Electric Power Industry

This paper aims to analyze the lifestyle of residents from household electricity consumption data. Improvement of QOL(Quality Of Life) of elderlies has attracted attention in a super-aging society. It is known that the lifestyle of a person directly affects his / her health and QOL. Therefore, understanding a lifestyle is expected to be useful for providing various support for improving QOL, such as recommending adequate actions and daily habit. As a means for understanding residents' lifestyle, this paper focuses on household electricity consumption data, which gets to be available with the spread of smart meters. The result of analyzing lifestyles from actual electricity consumption data by focusing on essential actions such as awakening meal is reported.

1. はじめに

本稿では QOL 向上や健康維持のための行動推薦への応用を目的として，家庭から得られる電力からライフパターン分析を行う．近年高齢者の孤立化・孤独死問題から，高齢者の QOL(Quality Of Life) の向上が重要視されている．代表的な QOL の定義の 1 つとして Bert は身体的状態，心理的状態，社会的交流，経済的・職業的状態，宗教的・霊的状態を提唱している [Bert 96]．このうち， 本稿では，身体的状態，心理的状態，社会的交流に着目する．3 要素の向上に伴い，睡眠促進，心理的ストレスの軽減，地域に関する魅力の再発見・再認識等が期待できるとされている [総務省 06]．また，ヘルスプロモーション活動展開に関するプリシード・プロシードモデル [Lawrence et al. 97] では，行動・ライフスタイルが健康，QOL に直接影響するとしており，ライフパターン分析に基づき，居住者へその改善につながる行動推薦を行うことで，QOL の向上が期待できる．

本稿では分析対象として，家庭電力データに着目する．近年，スマートメータの普及に伴い家庭電力データの活用の 1 つとして，QOL の向上や健康維持の観点から居住者の行動推定や，ライフパターン分析に利用する研究が行われており，エネルギー消費は生活行動に強い依存傾向があることが指摘されている [Robert et al. 53]．居住者の行動，ライフスタイルに関する新たな知見を得るために，生活行動を抽出し，これを特徴量としてクラスタリングを行う．狛江市の独居高齢者の負荷電流のデータセットを分析した結果について示す．

2. 関連研究

Wilhelm らはを目的とした電力データを用いた居住者の在・不在判定手法を提案している [Wilhelm et al. 15]． 1 日を 15 分間隔で区切り，各区間における消費電力の最小，最大，平均，標準偏差，絶対差の和，自己相関，オンオフ変化の検知数，最大と最小の差などを特徴量として用いている．これらを元に在不在判定を行った結果，最大で 94%の判定精度が得られたとしている．

連絡先: 高間康史，首都大学東京大学院システムデザイン研究科，〒 191-0065 東京都日野市旭が丘 6-6，ytakama@tmu.ac.jp

富越らは省 CO2 効果を目的とした省エネ行動提案ツールを提案している [富越ら 13]．冷暖房・給湯・照明他の 3 つのエネルギー使用量に応じて居住者への省エネに最適な行動を推薦する他，用途ごとのエネルギー使用量の推計を行う．50 世帯に対してケーススタディを行った結果，ツール導入により室内環境への満足度低下を抑えつつ，省エネ効果が得られたとしている．

3. 分析手法

人間の行動は必需行動，拘束行動，自由行動の 3 つに大別できるとされている [NHK 放送文化研究所 11]．必需行動は睡眠，食事，療養・静養といった必要不可欠性の高い行動であり，2011 年の社会生活基本調査 [総務省 11] では全国の起床・就寝の行為者率は 99%，朝食・夕食は 80%を超えている．起床時にはテレビを付ける，冷暖房を付ける等といった人為的な電力消費を行い，就寝時にはテレビを消す，照明の電源を切る等の行為で電力消費が変化することが想定される．また，食事に関しても，調理に伴いレンジ，ポット等の高消費電力の家電が複数使われることが想定される．従って特定の時間での総負荷電流の変化に着目する事で，必需行動である睡眠と食事の時刻を推定することは可能と考える．

食事は原則朝・昼・夜にとることを想定し，睡眠は起床，就寝のみとし昼寝などは対象としない．これらの条件から，起床，朝食，昼食，夕食，就寝の 5 種類の行動を，総負荷電流の変化から読み取りやすい必需行動として推定し，これらの時刻を特徴とみなしてライフパターン分析を行う．

3.1 データセット

本稿では狛江市に住む 1 人暮らしの高齢者 11 世帯の総負荷電流データ [中野ら 10] を用いた．分電盤に取り付けた端末によって世帯毎の総負荷電流を 1 分毎に計測したものであり，2008 年 10 月 1 日から 2009 年 9 月 30 日までの 1 年間の総負荷電流を取得している．

3.2 起床・就寝の特徴量抽出

Chen らの手法 [Chen et al. 13] を用いて，1 日毎の在・不在判定を行う．得られた在・不在判定結果を元に起床・就寝の時刻を推定する．社会生活基本調査 [総務省 11] を参考に，起床および就寝の想定時刻を定め，その範囲内で検出する．午

前 4 時から午前 9:59 時までの間で，不在から在宅に変化した時刻の中で，1 番早い時刻を起床時刻とする．また，午後 8 時から午後 23:59 時の間で在宅から不在に変化した時刻の中で，1 番遅い時刻を就寝時刻とする．上記を満たす時刻が存在しなかった場合，起床・就寝時刻は未検出とする．

3.3 朝食・昼食・夕食の特徴量抽出

Chacko らの手法 [Chacko et al. 12] を参考にして抽出したピーク検出回数を元に，食事時刻の推定を行う．

世帯番号 j の，2008 年 10 月 1 日から x 日目における，午前 0:00 を起点として m 分目における 1 分間の総負荷電流を i_{jxm} とする．各 i_{jxm} について，前の時刻との差分 Δi_{jxm} を求める．

世帯 j の x 日目における差分 Δi_{jxm} の集合を Δi_{jx}，世帯 j の差分集合を Δi_j とし，Δi_j の正の値である要素の平均を閾値 thd_j として求める．サイズ w[分] のスライド窓を Δi_{ix} に適用し，l 番目のスライド窓において，閾値 thd_j を超えた Δi_{jxm} の数をピーク検出回数 PC_{jxl} とする．朝食は午前 9:59 まで，昼食は午前 10:00 から午後 3:59 まで，夕食は午後 4:00 以降に行われると仮定し，朝食・昼食・夕食の時間範囲における PC_{jxl} が 5 回以上の集合 M_{jx}，L_{jx}，D_{jx} を求め，各集合における最大値に対応した時刻をそれぞれ朝食，昼食，夕食時刻とする．条件を満たす窓が存在しない場合は該当する時刻を 0 とする．

3.4 クラスタリング

3.2，3.3 節で求めた起床・朝食・昼食・夕食・就寝時刻の特徴量を元に，k-means++[David 07] を用いて分析を行う．全世帯のデータを対象としてクラスタリングを行い，Elbow 法によりクラスタ数を 10 と決定した．このクラスタ数を用いて世帯ごとに求めたクラスタリング結果に基づきライフパターン分析を行う．

4. 分析結果

全世帯に共通する性質として，就寝時刻が全国よりも早い傾向があった．これは，居住者が高齢者であるためと考える．また，世帯間の比較では，ほとんどのクラスタで同じ主行動が抽出されている世帯と，クラスタによって抽出された主行動に違いがみられる世帯に大別された．前者に属する世帯の例を表 1，後者に属する世帯の例を表 2 に示す．表において，2010 年の食事のピーク時刻 [NHK 放送文化研究所 11] と 2011 年の起床・就寝の平均時刻 [総務省 11] の差が 30 分以内を緑，30-60 分以内を黄，60-90 分以内を赤，それ以上を無色としている．表 1 に示すタイプの世帯では，いつも自宅で朝食・昼食・夕食をとる傾向があることがわかる．一方，表 2 に示したタイプの世帯では，食事の時刻が抽出されていないクラスタも多く，外出を含めた多様なライフパターンが見られることがわかる．

5. おわりに

本稿では QOL 向上や健康維持のための行動推薦への応用を目的として，家庭電力データのライフパターン分析を行った．特徴量として各日の起床・朝食・昼食・夕食・就寝の時刻を推定し，その結果を用いてクラスタリングを行った結果， 自宅で 3 食とる傾向にある世帯と，多様なライフパターンが見られる世帯に大別できることを示した． 今後の計画として，現状では起床・就寝時刻の抽出の欠損，朝食と起床時刻の矛盾などの問題があり，主行動推定精度の向上が必要と考える．

表 1: 世帯番号 2 のクラスタ別平均行動時刻

	起床	朝食	昼食	夕食	就寝
C21(30)		7:13	12:19	18:26	
C22(10)		5:30	13:20	21:59	21:09
C23(21)					
C24(33)	5:35	5:57	13:34	22:21	
C25(25)	5:33	5:51	12:34	18:12	
C26(22)	5:40	6:56	11:49	21:30	22:16
C27(42)	6:27	8:25	11:29	22:25	
C28(33)		7:20	12:08	22:11	
C29(22)			13:12	20:09	
C210(35)	6:48	8:52	12:41	18:23	

表 2: 世帯番号 3 のクラスタ別平均行動時刻

	起床	朝食	昼食	夕食	就寝
C31(121)				19:56	
C32(73)					
C33(67)			14:14	19:14	
C34(20)			13:15		
C35(6)	4:22		9:16	18:39	21:55
C36(25)		8:43		19:54	
C37(31)		9:20	12:06	19:39	
C38(11)		8:54			
C39(9)		9:42	11:20		
C310(2)		5:04	12:43		22:52

参考文献

[Bert 96] Bert Spilker, Quality of Life and Pharmacoeconomics in Clinical Trials, Lippincott Williams & Wilkins; Subsequent, 1996

[総務省 06] 総務省自治行政局過疎対策室, 田舎と都会の「縁」づくり 地域力を高める自治体間交流のすすめ 過疎地域・都市間における自治体間交流に関する調査 報告書概要版, http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/2001/kaso/pdf/chosa_10.pdf, 2006 (2018/2/3 現在)

[Lawrence et al. 97] Lawrence W. Green, Marshall W. Kreuter, 神馬征峰 (訳), ヘルスプロモーション—PRECEDE-PROCEED モデルによる活動の展開, 医学書院, 1997

[Robert et al. 53] Robert L. Thorndike, Who belongs in the family?, PSYCHOMETRIKA, Vol.18, Issue 4, pp. 267-276, 1953

[Wilhelm et al. 15] Wilhelm Kleiminger, Christian Beckel, Silvia Santini, Household Occupancy Monitoring Using Electricity Meters, UbiComp'15, pp. 975-986, 2015

[富越ら 13] 富越大介, 伊香賀俊治, 川久保俊, 藤崎浩太, 居住者行動とエネルギー使用量 の分析に基づく省エネ行動提案 ツールの構築, 日本建築学会技術報告書, 第 19 巻, 第 42 号, pp. 655-660, 2013

[中野ら 10] 中野幸夫, 上野剛, 小泉一夫, 秋元勇雄, 石田吉晴, 独居高齢者見守りシステムの実用化-狛江市における実証試験, 電力中央研究所報告 R 研究報告 (09014), pp. 1-16, 巻頭 1-3, 2010

[Chen et al. 13] Dong Chen, Sean Barker, Adarsh Subbaswamy, David Irwin, Prashant Shenoy, Non-Intrusive Occupancy Monitoring using Smart Meters, BuildSys'13, pp. 1-8, 2013

[総務省 11] 総務省統計局, 平成 23 年社会生活基本調査, http://www.stat.go.jp/data/shakai/2011/, 2011 (2018/2/3 現在)

[Chacko et al. 12] Chacko J. Deepu, Zhihao Chen, Ju T. Teo, Soon H. Ng, Xiefeng Yang, Yong Lian, A Smart Cushion for Real-Time Heart Rate Monitoring, BioCAS2012, pp. 53-56, 2012

[NHK 放送文化研究所 11] NHK 放送文化研究所, 2010 年国民生活時間調査, https://www.nhk.or.jp/bunken/summary/yoron/lifetime/pdf/110223.pdf, 2011 (2018/2/3 現在)

[David 07] David Arthur, k-means++: The advantages of careful seeding, Proc. of the eighteeenth annual ACM-SIAM symposium on Discrete algorithm, pp. 1027-1035, 2007.

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

事例紹介:24 時間周期データに対する教師無し学習の適用

A pilot study: Application of machine learning to 24 hour period data – three case studies

後藤　勲[*1]
Isao Goto

[*1] 住友電気工業株式会社
Sumitomo Electric Industries, LTD

In this article, I'd like to introduce three case studies for which I tackled for three years as a pilot study. This pilot study was applied to three fields. 1. Energy consumption in houses. 2. Anomaly detection of photovoltaic power system at mega-solar system. 3. Sunlight combination type plant factory. In each case 24 hour period data was treated as a frame (but three data formats and types are different), and was analyzed in unsupervised learning. These analyses found out beneficial knowledge for each field. **Key words: k-means, Probabilistic Latent Semantic Analysis(PLSA), Bayesian network, Home energy management system, Photovoltaics, Sunlight combination type plant factory**

1. はじめに

本稿の目的は, 24 時間周期データに対して機械学習を適用した事例と, その結果の一部を報告することである. そのため, 背景, 方法論や結果の詳細については割愛している.

24時間周期データは世の中に多い. 太陽の動きと関連した事象, 例えば日射量, 気温や人の活動などもそれにあたる. これらのデータは, 一日の総量や平均値(±標準偏差), または連続的な時系列データとして取り扱われる場合が多いと思われるが, 今回は 24 時間周期を一つのフレームとして捉え, データを分析した. それぞれ紹介する事例の分析の目的は異なるが, このフレームに落とし込むことで, 同様の考え方で分析が可能となった.

本稿では, 実際に社内で取り扱った 3 つの事例を報告する. 一つは, Home Energy Management System(HEMS)から得られた世帯毎の消費電力データと属性データ, 太陽光発電(PV)の発電データ, そして, 太陽光利用型植物工場内の環境データと植物の吸水量データの 3 種類である.

今回, 主に使用した手法は, 教師無し学習の k-means と確率的潜在意味解析(PLSA)で, これらをデータと目的に応じて使い分けた. その他, 必要に応じてベイジアンネットワークを含む統計解析等を用いて結果を評価した.

2. 方法

2. 1 k-means

k-means はクラスタリング手法の一つで, 最も広く使用される手法のである[Lloyd 1982][Wu 2008]. 今回はその概要のみを記載する. k-means とは分析対象のデータ X を任意の k 個のクラスタに分割するアルゴリズムで, 目的関数である式(1)を最小化するクラスタ中心を見つける.

$$\varphi = \sum_{x_j \in X} \min_{i \in k} \|x_j - c_i\|^2 \cdots (1)$$

ここで, $x_j (j = 1, \dots, N)$は各データ, N はレコード数, $c_i (i = 1, \dots, k)$はクラスの重心を示す. アルゴリズムは以下の通りである.

1. 任意の k 個のクラスタ中心をランダムに選択
2. 全てのデータをx_jからもっとも近いクラスタに割当
3. 式(2)を用いてクラスタ毎に重心を算出

$$c_i = \frac{1}{|C_i| \sum_{x_j \in C_i} x_j} \cdots (2)$$

ここで, C_iはクラスタ i のデータ集合で$|C_i|$はクラスタC_iに含まれるデータ数である.

4. クラスタの変化がなくなるまで, 2, 3 を繰返し

計算は R とパッケージ stats 3.4.2 を用いた.

2. 2 確率的潜在意味解析(PLSA)

PLSA は自然言語処理における文書と単語の行列から文書の潜在的な意味を推定するために提案された分析手法の一つである[Hofmann 1999]. 本稿では, N 個の文書d_i(i = 1, ... N)と M 個の単語$w_j (j = 1, \dots M)$を目的に応じて読み替えて, PLSA を適用した[石垣 2011]. また潜在クラスタ数 k と仮定し, その変数をz_kと表し, その関係を以下の同時確率としてモデル化する.

$$p(x_i, y_j, z_k) = p(z_k)p(x_i|z_k)p(y_j|z_k) \cdots (3)$$

文章 i の単語 j の出現数をN_{ij}とすると, その尤度対数は,

$$L = \sum_i^n \sum_j^m \log \sum_k^K p(z_k)p(x_i|z_k)p(y_j|z_k) \cdots (4)$$

となる. このモデルを EM アルゴリズムで対数尤度を最大化する条件付き確率を推定する. 各条件付き確率に対して, 初期値を乱数で与えると, 式(3)の変形から潜在変数の条件付き確率は以下の式で計算できる.

$$p(z_k|x_i, y_i) = \frac{p(z_k)p(x_i|z_k)p(y_j|z_k)}{\sum_k^K p(z_k)p(x_i|z_k)p(y_j|z_k)} \cdots (5)$$

上記の計算後, それぞれd_iとw_jの最も所属確率が高いカテゴリに所属するとした.

PLSA の計算には, 国立研究開発法人産業技術総合研究所人工知能研究センターの開発した APOSTOOL3.0 を用いた.

2. 3 ベイジアンネットワーク(BN)

BN は対象とする確率変数のノードと変数同士の依存関係を確率的なネットワークとしてモデル化したものである. その確率ネットワークはグラフ構造として表現することが可能で, 視覚的に表

※本研究内容は住友電気工業株式会社の公式見解を示すものではありません.

連絡先:後藤　勲, 住友電気工業株式会社, 大阪市此花区島屋 1-1-3, 06-6466-8231, gotou-isao@sei.co.jp

現・理解し易く，グラフィカルモデルによる確率推論の手法を用いることができる．

BN の構築には，BayoNet6.2(現 BayoLink, NTT 数理システム)を用いた．

3. 結果

3.1 事例 1: HEMS データの活用-大規模 HEMS 情報基盤整備事業（経済産業省：平成 26 年度から平成 27 年度までの事業）

弊社は上記整備事業に参画し，事業内の企業コンソーシアム(i エネコンソーシアム)の枠組みで，平成 27 年 12 月から二か月間，三重県桑名市と四日市市周辺の 457 世帯にご協力いただき，一般家庭向けの節電実証を実施した．当整備事業内で匿名化された世帯毎の電力・属性データから，消費電力パターンと家族構成や節電機器の関係を分析した．

今回は PLSA を用いて世帯と消費電力量を同時に自動カテゴリ化する．同様の試みは存在する[Verdu 2006]が，同時カテゴリ化できないことや，結果をヒトが理解しづらいことなど課題があった．また一般的なクラスタリングでは，次元の呪いのため高次元データを取り扱えないことなども課題であった．そこで今回は PLSA を用いることで上記の課題を解決可能か検証し，その後，BN によりカテゴリと属性データ間の構造モデルを構築し，その関係性を検証した．

電力データは 30 分 1 コマのデータを平日 10 日分，属性データは世帯毎の家族/建屋デモグラ，設置機器等である．

今回は 24 時間周期をもつ消費電力量の時系列データを PLSA により世帯単位で時間帯毎の消費電力量を同時カテゴリ化した．潜在カテゴリ数(k)は赤池情報量基準(AIC)から決定することはできるが，本稿では k = 3 に固定し，10 回初期値を変えて計算した．本来ならばカテゴリ化後，その妥当性を検証すべきだが，正解データが存在しないため，カテゴリ化の結果と実際のカテゴリ毎の消費電力パターンの平均値を比較した．その後，潜在カテゴリと各世帯の属性情報との関係をベイジアンネットワークでモデル化し,カテゴリをエビデンスとしたときの変数（属性データ）の条件付き確率を比較した．

PLSA 適用結果を図 1 に示す．図 1 は各カテゴリの時間帯毎の電気を使うカテゴリ間と自身の時間帯毎の相対的な確率を示している．各カテゴリが異なる時間帯に電気を使用する確率が高いことが示されており，例えば，C03 は朝と夕方から夜にかけて電気を使う確率が高くなっている．図 2 は各カテゴリの時間帯毎の平均的な消費電力量を示している．それぞれが図 1 で示されたパターンと同様の特徴を示しており，カテゴリ毎の特徴を明確に分類できている．図 3 は BayoNet を用いて構造学習した結果で，この構造モデルからカテゴリをエビデンスとしたときの給湯

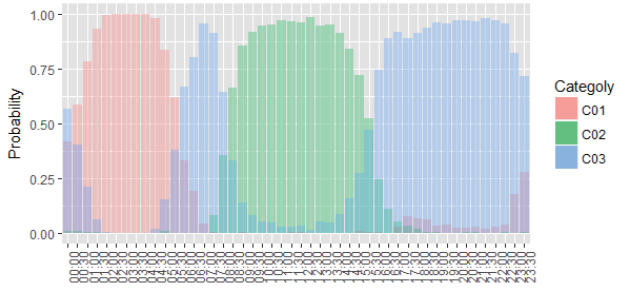


図 1 カテゴリ毎の時間帯別電気使用確率

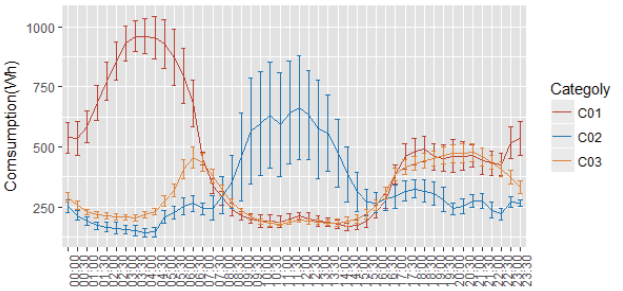


図 2 カテゴリ別の時間帯別平均消費電力量

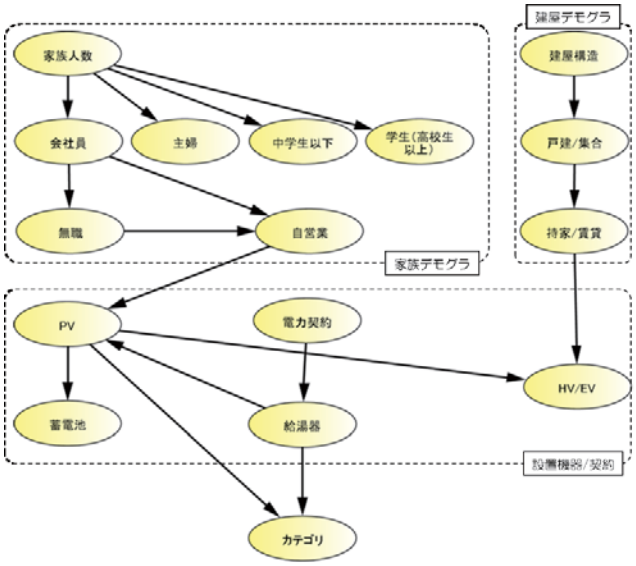


図 3 構築されたベイジアンネットワークモデル

器と PV の所有確率(事後確率)を表 1 に示す．C01 は，深夜から早朝に電力を使用する確率が高いが，その原因がエコキュートによる夜間電力を用いた電力の消費であることが予想できる．また，C2 は他のカテゴリ PV 設置率が高いことは日中に発電した電力を積極的に使用していることと一致すると考えられる．

表 1 カテゴリ毎の省エネ機器設置確率

	カテゴリ		事後確率	
エビデンス	所属世帯数（世帯）	電力消費パターン	エコキュート設置率(%)	PV 所有率 (%)
C01 = 1	83	深夜-早朝	76	52
C02 = 1	17	日中	31	60
C03 = 1	346	朝,夕方-夜	30	17
全体	446	-	38	25

3.2 事例 2: 太陽光発電の異常検知

今後のエネルギーミックスの問題からも PV の有効活用は今後重要であるが，その課題の一つとして異常検知が挙げられる．異常の検知のため，まず「異常」を定義することが必要である．そして，さらに分析のための長期間のデータや別途照度計等のデータが必要になってくる．上記の課題を解決するため，ストリング単位のデータを 1 フレーム 24 時間周期と捉え，ストリング毎の発電パターンをクラスタリングした．その後，現地調査によりクラスタ毎の異常原因を特定した．

データは，当社のストリング監視システムから取得したストリング単位の発電データで，国内のあるメガソーラーサイトの 456 ストリングから得た．データのサンプリングは，1 分単位で計測可能

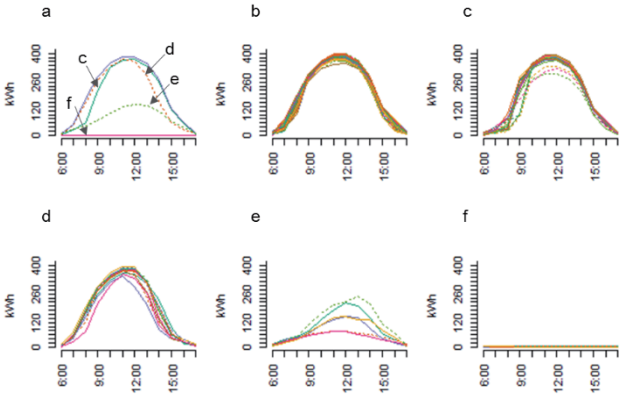


図 4 クラスタリング結果と各クラスタのストリング発電データ

だが, 雲等の影響を除去するため, データを 60 分の平均値とした. さらに, データを晴天時の一日分, そして発電が期待される時間帯を 6-18 時と限定すると 12 次元データとなり, 適用範囲が ~8 次元程度とされている k-means の採用は難しいが, 全ストリングの発電パターンは太陽の動きに依存し, 全てが基本的にはベル型(同じ形)となるため k-means でのクラスタリングが十分可能と考えた. k の決定については, 変化させ AIC 等で決定する方法はあるが, 今回は,実務者の意見を取り入れ, k=5 とした. これによりヒトの感覚に近い異常の定義となった(k は 2-10 まで変化させた). これらの手続きで各クラスタの特徴から異常を定義した.

図4a は結果の各クラスタの重心である. 図 4 b-f は各クラスタの所属ストリング毎の発電データで, b が最も総発電量が大きく異常はないと判断した. c と d はそれぞれ午前, 午後の発電量が低くなっているおり, e は全体的に発電量が低く, f はデータが 0 に近かった. これらの結果を現場調査結果と突合すると, c と d はそれぞれ位置的な問題で物体の影となりこれらの異常が発生していることが分かった. また f はデータ取得エラーとなっており, e は日中, 影の影響がみられる型とシステムによる異常が見つかった. 本手法により異常有りと判断されたストリングは実務者の目視でも異常と判断された. また, 同様の手法で別の二つのメガソーラーサイトでも同様の結果が得られている(Data not shown).

3.3 事例 3: 温室におけるトマトの給水量予測モデル

作物(今回はトマト)の吸水量は, その生長状態や健康状態の定量化に重要な情報であるが, 環境によってもその量が変動するため, その予測は難しい. これまでの研究から一日の平均室温の積算値(積算温度)が一日の吸水量と相関することはわかっているが, 社内の複数の検証実験では精度は, 平均絶対パーセント誤差(MAPE) ≒ 0.2-0.3 で精度的には十分ではなかった (Data not shown).

本稿で用いるデータは, 千葉大学柏の葉キャンパスの太陽光利用型植物工場(以下, 温室とする)でトマト栽培から得られた環境データ(湿度: Hum., 光粒子: Ill., 飽差: Sat. Def., 室温: Temp.)と, 当社の製品ニューサンドポニックスより得られた給水量データである. 今回は給水量 ≒ 吸水量と考える. データの粒度は環境データが 1 時間単位, 吸水量データが 24 時間単位であり, 単位を吸水量に合わせる必要がある. しかし, 環境データを 1 日の統計値としてしまうと, 環境の変化パターンを無視することとなり, 重要な現象を見逃す可能性もある. 今回はこの課題を解決するために, 以下の手法(図 5)を検討し, その有効性を検証した.

まず, 24 時間(フレーム)毎に分割した各環境データを k-means によりクラスタリングし, パターンを抽出した. 今回は k を

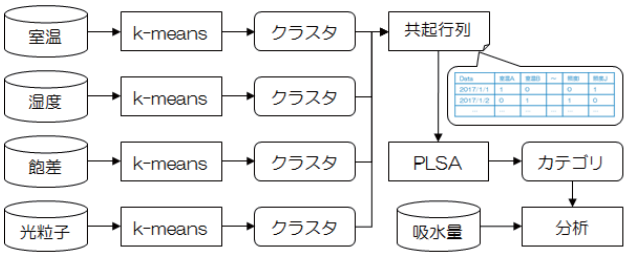


図 5 データ分析の流れ

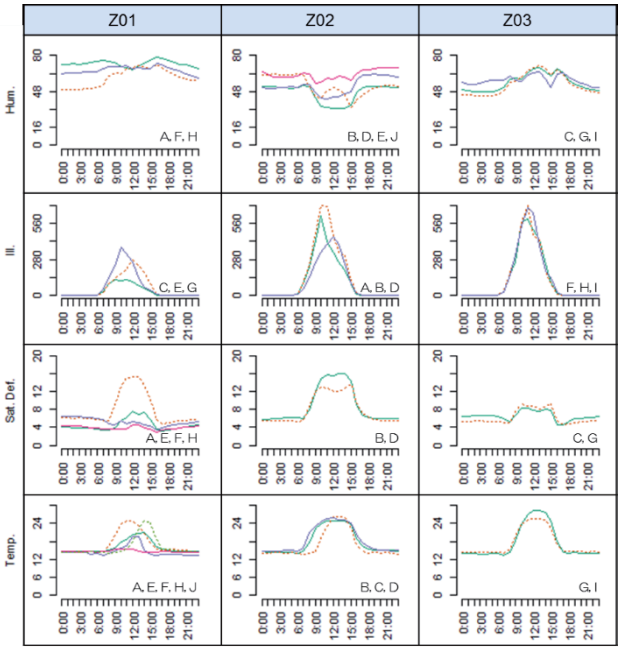


図 6 カテゴリ毎の所属環境パターン

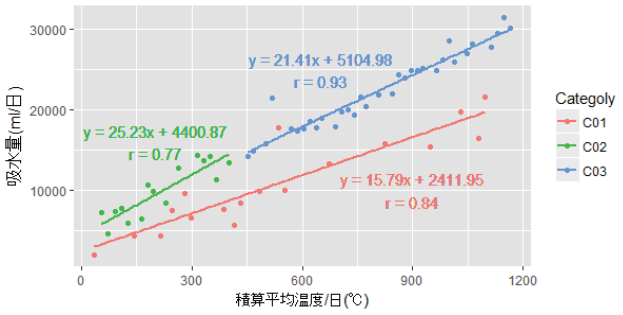


図 7 カテゴリ別の積算平均温度と吸水量

Gap 統計量とクラスタ内分散により決定した. またクラスタは各環境データに A から順にアルファベットでクラスタ名を付けた. その後, クラスタリング結果から共起行列を作成し, PLSA により潜在カテゴリを抽出した. PLSA のカテゴリ数に関しては, 温室の作業従事者と共に検討し 3 とした. この手続きにより, 文章と単語と同様に, 日付と環境パターンで同時カテゴリ化が可能になった.

図 6 は各環境データのパターン毎(図中の A-J)の所属カテゴリを示している. 各カテゴリで異なるパターンを持つことがわかる. さらに, 一日の平均温度の積算値と吸水量の散布図をカテゴリ毎に分類したものを図 7 に示す. この図から, カテゴリ毎にモデルが異なっていることが予想される. さらに, 全データを用いた場合とカテゴリ毎に分け, 単回帰/重回帰による給水量予測モデル

を検討した. 交差検証(5-fold, 1000 回実施)時の平均絶対パーセント誤差(MAPE)を表 2 に示す. カテゴリ化と平均温度の積算値との回帰が最も MAPE が低くなった.

表 2 吸水量予測精度の比較

MAPE ± SD	単回帰	重回帰
全データ	0.22±0.07	0.22±0.06
C03 のみ	0.04±0.02	0.05±0.02

4. まとめと今後の課題

4.1 まとめ

本稿では 3 つの事例における高次元データを 24 時間というフレームに落とし込みデータ分析を進めた. 分析方法を主として教師無し学習を用い, データから特徴を抽出し, 抽出されたデータを他のデータと組み合わせることでビジネス活動に有用な知見を得ることができた.

4.2 高次元データから 24 時間周期データへ

今回, 時間と共に変化する時系列データを, 目的に合わせデータを変形/修正/圧縮し, 24 時間というフレームに当てはめて分析した. このフレームに当てはめることで, データの取り扱いやすくなり, その後の分析が容易になった. 例えば, 事例 1 では, 480 次元×約 450 世帯のデータを BN に適用でき, 事例 3 でも無数にある環境パターンの組み合わせも, データからあり得る組合せを抽出することで, 回帰分析による精度の高いモデリングが可能になった.

4.3 得られた知見

本手法によりビジネスに有用となる知見を得ることができた. 例えば, 事例 1 からは, 消費電力量のパターンから世帯の設置機器等を推定できる可能性を示した. これは今後普及が進むスマートメーターデータの活用することで 新たなサービスが拡がる可能性を示している. 一方でこれらのデータの管理(特にセキュリティー)面の重要性も示している.

また, 事例 2 は, ある条件のもとでは, 簡単なクラスタリング手法で異常検知に有用な知見が得られた. 今回想定したユースケースでは(例えば, 一日一回の異常検知), 本手法も十分に有効であった. また, 今回得られた結果を教師データとして用いれば, 教師有り学習による, より賢い異常検知も可能かもしれない. さらに今回の手法は, 別途照度計の設置や教師データを必要としないこと, 晴天時一日分のデータで良いことが 利点として挙げられる. しかし, 本手法を用いても, 時定数の短い瞬時的な発電異常の発見は難しいと考えられるため, 網羅的な異常検知にはさらに研究が必要となる.

IoT(Internet of Things)技術の進歩で, 農業ビッグデータも蓄積されつつある. 例えば, 今回用いた植物工場内の環境データもその一つである. 事例 3 では環境データを分析し, 温室内の環境状態を離散化(カテゴリ化)できた. さらに給水量データと組み合わせることで, 精度の高い給水量予測モデルを構築できた. 精度の高さから, このモデルはある程度環境要因による給水量への影響を相殺できたと考えられる. 植物の生長状態が「吸水量」と関係すると考えられるため, 「給水量 ≒ 吸水量」と環境データから, 植物の生長状態や健康状態を推定可能になるかもしれない. また, 本稿における事例は特定の温室の一つの作物についての結果だが, 同様枠組みにより, 温室/作物に限定せず給水量モデルの構築は可能と考えられる.

4.4 今後の課題

3 つの事例の共通した課題は, k の決定である. 今回はヒトの感覚やわかりやすさを重視し決定した. 実際, 結果は作業従事者には理解しやすかったが, このような主観的な方法を用いることで, 見逃してしまう現象もあるかもしれない. 今後は k を変化させ, それぞれどのようなパターンが生成されるのかを検証していく必要がありかもしれない.

参考文献

[Lloyd 1982] S. Lloyd: Least squares quantization in PCM, IEEE Trans Inform Theory, IT-28, pp. 129-137, 1982

[Wu 2008] Xindong Wu, et al.: Top 10 algorithms in data mining, Knowl. Inf. Syst., Vol. 14, pp. 1-37, 2008

[Hofmann 1999] Hofmann, T.: Probabilistic latent semantic analysis , Proc. of Uncertainty in Artificial Intelligence , pp289-296, 1999

[石垣 2011] 石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, Vol. 56, No. 2, pp.77-83, 2011

[本村 2003] 本村陽一: ベイジアンネットソフトウエア BayoNet, 計測と制御, Vol. 42, pp. 693-694, 2003

[Verdu 2006] S.V. Verdu, M.O. Garcia, C. Senabre, A.G. Marin, F.J.G. Franco: Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps, Vol. 21, pp. 1672 – 1682, 2006

謝辞

経済産業省:大規模 HEMS 情報基盤整備事業にご協力いただいた桑名市, 四日市市の 450 世帯の皆様, KDDI 株式会社をはじめとする i エネコンソーシアム参加企業, 特に CCC マーケティング株式会社, 株式会社ヤマダ電機に感謝いたします.

本成果の一部は経済産業省の委託業務「平成29年度新エネルギー等の保安規制高度化事業(電気施設保安技術高度化の評価・検証事業)」である.

本成果の一部は国立大学法人千葉大学との共同研究の一部である.

最後に, 本研究にあたりご指導いただいた, 産業技術総合研究所本村陽一首席研究員に感謝いたします.

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

発想支援システム「ひらめきエンジン」の紹介

Introduction of idea generation support system "Hirameki engine"

上山　雅樹[*1]
Masaki UEYAMA

[*1] (株)本田技術研究所
Honda R&D Co.,Ltd

AI is soon substituted for memorizing or judgment based on data. However, since AI is hard to be substituted for "idea Yes and flashing", importance increases more. Honda aims at realization of "robotics society where people's wonderfulness is conspicuous." Keywords are three "E", Empower which expands people's possibility, Experience which grows with people, and Empathy which empathizes with people. Introduce idea generation support system "Hirameki engine" which is working as a system which supports people's idea.

1. はじめに

単に記憶することやデータにもとづく判断の多くは，やがてAI に代替されるようになる．一方「発想すること・ひらめくこと」はAI では代替されにくいため，今後は，より重要性が高まると考えられる．Honda は「人の素晴らしさが際立つロボティクス社会」の実現をめざしている．その中で大切にしているキーワードは，人の可能性を拡大する Empower，人と共に成長する Experience，人と共感する Empathy の3つの「E」である．その一環として，すでに人の身体機能支援として歩行アシストの商品開発を行っているが，本稿では人の知的機能支援として取り組んでいる発想支援システム「ひらめきエンジン」を紹介する．

2. 背景

近年，AI の技術進化はめざましいものがある．すでに記憶や判断は，一部の領域においては人のレベルを凌駕しており，単なる記憶やデータにもとづく判断は AI に代替される事が多くなると考えられる．

そのような状況の中では，「発想すること」や「ひらめくこと」はAI では代替されにくいため，より重要性が高まってゆくと考えられる．

Honda は，AI を単なる人の代替ではなく，また人と対立するものでもなく，人と共存し，よりよい人のくらしのために役立つものにしてゆきたいと考え，「人の素晴らしさが際立つロボティクス社会」の実現を目指している．

その考えに基づいて，すでに身体機能のアシストとして，歩行訓練機である歩行アシストを法人向けにリース販売している．

本稿では，知的機能のアシストとして開発に取り組んでいる発想支援システム「ひらめきエンジン」を紹介する．

目指す方向性を実現するために，Honda では3つの E で始まる Empower，Experience，Empathy をキーワードとしている．

・Empower　：人の可能性を拡大する
・Experience　：人と共に成長する
・Empathy　：人と共感する

連絡先：上山雅樹，株式会社本田技術研究所，東京都港区赤坂 5-3-1 赤坂 Biz タワー27 階 Honda イノベーションラボ Tokyo，03-5545-4661，03-3585-2411，masaki_ueyama@n.w.rd.honda.co.jp

発想支援システムは，この Empower の一部ととらえており，ひとが発想することを支援するシステムである．

3. ひらめきエンジンの概要

3.1 ひらめきが起きる状況

ひらめきは三上に在りという(1)．三上とは，馬上・枕上・厠上，つまり乗り物にのっているとき，寝床，お手洗いといった，集中から解放されているときにひらめきが得られるとの説である．何かに没頭しているときは，その事で頭がいっぱいになっている．その状態で散歩するなど，集中して考え続けることから意識を解放すると，ひらめきが訪れることがある．ひらめきのメカニズムは，それまで考え続けていた事と頭の中にある様々な概念を無意識にランダムに組み合わせを行っていて，良いアイデアが思いついたと無意識が判断した場合にだけ意識にのぼるのではないかと考えた(2)．

3.2 ひらめきエンジン詳細

ひらめきエンジンは，いつも頭の片隅にある事柄を示すワードを一つのグループとしておき，それとは無関係なワードをランダムにぶつけて一つの言葉として表示するもので，偶発的な組み合わせによって自分では思いつかないような「驚きのある組み合わせ」として発想のきっかけを得られる仕組みである．

この無関係なワードは自分の頭の中にあるものや散歩をしているときなどに見聞きしたことに限定されているが，ひらめきエンジンはワードをネット検索などの手法を用いて，より幅広くランダムに抽出することで，個人ではアクセスできない幅広い種類のワードを用いることで発想の幅を広げるものである．ひらめきエンジンをながめているときにひらめきが訪れることを期待している．

具体的には，あらかじめ設定したワード群からひとつのワードを選び出し，それとは別のワード群からひとつ選び，これをランダムに組み合わせてひとつのワードとして表示するものである．

それぞれのワードはそれぞれ異なった背景色もって左右からぶつかり，ひとつのワードが合成されて別の背景色が表示されるカラーアニメーションとして作成した．

3.3 コラボレーション

ひらめきエンジンは，Honda イノベーションラボ Tokyo を 2017 年に設立するにあたって，株式会社乃村工藝社と株式会社カヤック，株式会社本田技術研究所の 3 社が共同で開発したものである．

もともとは，発想支援システムとは似ても似つかないデジタルサイネージとして検討を開始したものであった．しかしこの 3 社の担当者が議論を重ねるうちに，単なる情報掲示板としてのデジタルサイネージではコンテンツとして魅力に欠ける事，かりに最先端の技術を投入したディスプレイをつくったとしても数年で陳腐化がさけられないことなどの理由から，研究開発の現場で役にたつものを開発するという方向に変更し，アイデア支援システムの検討をおこなった．

開発にあたって基本コンセプトは 3 社の検討メンバーの徹底した議論から生み出し仕様を決定した．視覚的な工夫を織り込んだプログラミングはソフトウェア製作の豊富な経験を生かしてカヤックが行った．また様々な研究開発施設の建築設計の経験をもとに，世の中で受け入れていただけそうかどうかの観点ではたした乃村工藝社の役割は非常に大きい．つくりあげたひらめきエンジンは，国内外の Honda の事業所 5 か所でモニター使用を行い，ユーザビリティを含めた改善を重ねている．現在，ご興味をもっていただいている会社で利用していただいてる．

このひらめきエンジンは，良い物を創り上げようと徹底的に議論し，試作や確認を繰り返すことで信頼関係が築かれ，関わった全ての者が「御社，弊社」の関係ではなく，個人名で呼び合い意見をぶつけ高めあう関係となって推進したことにより，初期の構想からは想像しえないほどの進化を遂げたものを創り上げることができた．

4. 適用のイメージ

適用のイメージを移動，仕事，生活の3つのパターンで示す．

4.1 移動

Honda の商品である「移動」を軸に考えると，旅行で行く場所や旅行先でのアクティビティを決める際に使える可能性がある．連休の旅行先を家族と決める際に，地域をひとつのワード群に，アクティビティをもうひとつのワード群に設定しておき，これらのランダムな組み合わせによって一つのワードをつくる事で，旅行計画を立案する支援となる．たとえば，

信州×祭り
東北×グルメ
京都×絶叫
八ヶ岳×縄文遺跡

などである．これらのなかから、「これは」と思うところがあれば、しめたものである．

4.2 仕事

仕事で解決すべき課題としていつも考えている事に関連したいくつかのワードをワード群に登録しておき，いっぽうでインターネット等をもちいて様々なワードを抽出して，もうひとつのワード群に登録する．これらから一つづつワードをランダムに抽出し左右からぶつけて一つのワードとして表示する．これを会議が始まる前や休憩時間に特に強く意識せずに眺めていると，課題解決のヒントにつながるワードに出会う事がある．

スピーカー×エンジン振動
ふたりでひとつ×前処理

などである．これらを元に実際に特許になった過去の例であり，このひらめきエンジンをつかったのと同じような思考パターンでひらめいたものである．

また，企画会議等でアイデア出しが求められる場合がある．良いアイデアを出すために，ともかくたくさんのアイデアを出す方法がある．そのなかから良さそうなアイデアを選び出し，磨きをかけるわけである．

会議の会話をテキスト化し，そこからワードを抽出し，そのワードとあらかじめテーマに沿って設定しておいたワードをぶつけ，ひとつのワードとして表示する．多くはノイズにしか過ぎないが，なかには思いがけないアイデアのきっかけとなる可能性がある．

4.3 生活

毎晩の夕食の献立を考えるのは結構大変である．しかしひらめきエンジンで一つ目のワード群に食材，二つ目のワード群に料理法を設定しておき，それらをランダムに組み合わせることで，食べてみたい料理ができる可能性がある．もちろん，食材は冷蔵庫の中にあるものや，その日の特売品をいれておくのも家事を預かる身としては有効な使い方である．

じゃがいも×カレー
じゃがいも×肉じゃが
じゃがいも×シチュー

これらはありふれた料理であるが，
レンコン×ステーキ
ハクサイ×おさしみ
サトイモ×アヒージョ

これらは著者自身もためしてみたが，思いもかけないおいしい料理を作り出すことができた．

まとめ

Honda では人の役に立つ道具を創りだしたいと考えて努力しており，人の発想を支援するシステムとして，移動や仕事だけでなく，生活の様々なシーンで活用できる可能性を秘めた発想支援システム「ひらめきエンジン」を開発している．AI が社会に実装される時代であるからこそ，人のすばらしさを際立たせるための道具として，この発想支援システムを発展させてゆきたい．

参考文献

（1） https://kotobank.jp/word/%E4%B8%89%E4%B8%8A-513767

（2） https://lineblog.me/mogikenichiro/archives/8329218.html

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

個人の記憶想起支援に向けて:タグクラウドとネットワーク

Toward Personal Memory Recall Support: Tag Cloud and Network

村上 晴美 Harumi Murakami
村上 龍太郎 Ryutaro Murakami
上南 佑太 Yuta Ueminami

大阪市立大学大学院創造都市研究科
Graduate School for Creative Cities, Osaka City University

We report the progress of our research on “constructing externalized-memory,” conducted since 2013. We have focused on systems to support personal memory recall using an information structure called history structure. Before 2013, we developed a system called knowledge-space browser, which displays a network called “knowledge-space” from various information activities. Since 2013, we have been developing a system, called tag browser, that displays a tag cloud based on the usage of Calendar and Twitter, as well as continuing our development of the knowledge-space browser. The main feature of our current research is to use the number of photos to weight a system’s keywords. In this paper, we (a) describe the tag browser and (b) compare the tag cloud and the network as approaches to support human memory recollection.

1. はじめに

筆者の研究室では「個人の人生の記録」と題する研究を行っており，[村上 13]において進捗を報告した．本稿では，「個人の人生の記録」の中で「個人の外化記憶構築」に関する研究について，[村上 13]以降の進捗を報告する．

2. 先行研究の概要

個人的な記憶や情報を蓄積するために，個人の記憶を外化してコンピュータ上に疑似的に蓄積する外化記憶という概念を提案してきた．見る，聞く，考えるなどの人間の情報処理過程をコンピュータ上に疑似的に蓄積することにより，自己の理解と，記憶の想起や個人的な情報へのアクセスを支援する．プロトタイプとして Memory-Organizer を試作してきた．

Memory-Organizer は，思考内容などから外化記憶の作成，編集，探索を支援する「思考空間ブラウザ」，Web 閲覧時の外化記憶の作成，編集，探索を支援する「オーバレイ Web ブラウザ」，ユーザの時系列の興味空間を作成して外化記憶の探索を支援する「興味空間ブラウザ」，外化記憶を地理空間上に配置して探索を支援する「地理空間ブラウザ」，外化記憶をカレンダー上に配置して探索を支援する「カレンダーブラウザ」，ユーザの知識空間を作成して外化記憶の想起を支援する「知識空間ブラウザ」を試作した[村上 01, Murakami 04, Murakami 12 など]．

3. 履歴構造を用いたシステム

近年は Memory-Organizer で扱う外化記憶の中でも，「日時，キーワード群，URI 等」からなる「履歴構造」と呼ぶ情報構造に履歴を蓄積して記憶の想起を支援する研究に焦点をあてている[Murakami 10]．履歴構造を用いたシステムの概要を図 1 に示す．

2013 年以前は，Web 閲覧，メール，Twitter，カレンダー，購買履歴等の多様な情報利用行動から履歴構造を作成し，興味空間ブラウザと知識空間ブラウザの開発を行った．

2013 年以降は，多様な情報利用行動の中でも，特に記憶想起に有効な情報源として Twitter とカレンダーに着目した．また，「写真を多く撮る日に印象的な出来事がある」という考えに基づき，写真枚数を用いた重み付け手法を導入した．知識空間ブラウザの改良を行いつつ，知識空間としてタグクラウドを表示するタグブラウザの開発を行った．

以下では，(a) タグブラウザと，(b) 記憶想起支援におけるタグクラウドとネットワークの比較について述べる．

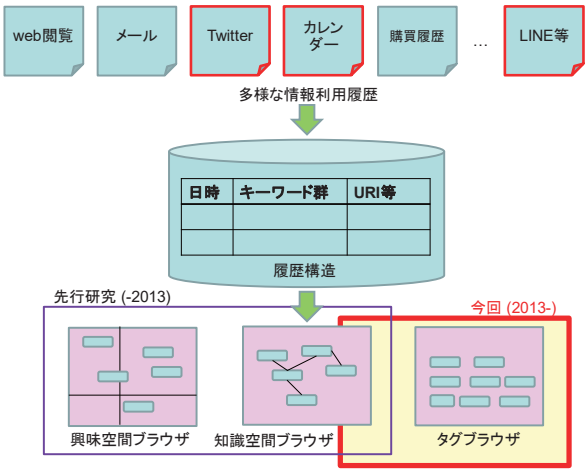


図 1：履歴構造を用いたシステム

4. タグブラウザ

タグブラウザはユーザの知識空間をタグクラウドの形式で表示して記憶の想起を支援する．

松本らは，多様な情報源の中でも Twitter とカレンダーに焦点をあててキーワード群を抽出し，「写真を多く撮る日に印象的な出来事がある」という考え，ユーザの撮影した写真の枚数に基づきキーワード(タグ)の重み付けを行い，タグクラウドを表示するタグブラウザを開発した[松本 14, Matsumoto14]．情報源の違いは色で表示する．ユーザによって有用な情報源が異なるため，松浦らは，Twitter とカレンダーの重みをユーザが手動で変更できるように改良した[松浦 14]．渡部らは写真だけでなく「地名」や「人名」等に重みを付加した実験を行い，複数の情報源(Twitter とカレンダーの両方)に含まれるキーワードに追加の重みを付与することが有効であることを示した[渡部 16]．

連絡先:村上 晴美，大阪市立大学大学院創造都市研究科，
大阪市住吉区杉本 3-3-138，harumi@osaka-cu.ac.jp

村上らは，複数の情報源を容易に取り込めるよう，キーワードの重み付け手法を汎用化した[村上 18]．情報源としてカレンダーと Twitter に加えて LINE を利用した．Twitter は青，カレンダーはオレンジ，LINE は緑とし，2 種類以上の情報源で出現するキーワードは赤とした．また，情報源毎のキーワード量によってログ部の位置を自動的に設定するようにした．さらに，タグ検索とキーワード検索，写真表示機能を付加した（図 2）．

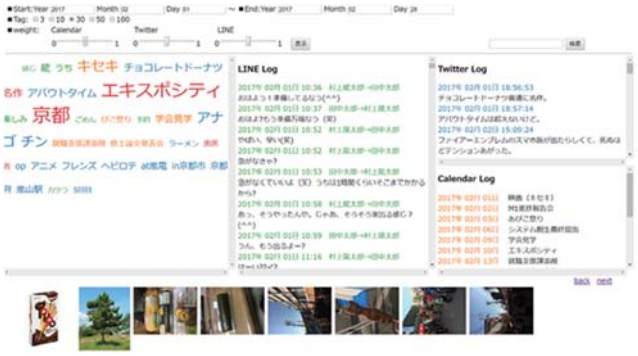
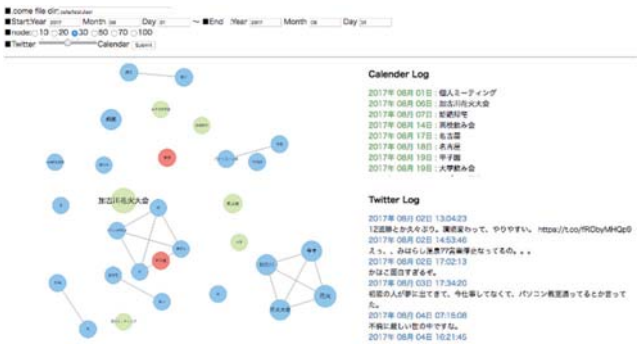


図 2：タグブラウザ

5. タグクラウドとネットワークの比較

タグクラウドとネットワークの相違点を調べるため可視化以外の部分は同じシステムを試作した．

(a) タグクラウド



(b) ネットワーク

図 3：タグクラウドとネットワークの比較

図 3 に試作した 2 システムを示す．情報源として Twitter とカレンダーを利用し，表示するキーワードの重みは先行研究[松本 14, Matsumoto14]と同じ，色は Twitter は青，カレンダーは緑，両方に出現するものは赤とした．タグクラウドはキーワードを出現日時順に表示している．ネットワークは履歴構造中で共出現するキーワード間を線でつないでおり，位置はランダムである．ネットワークは拡大縮小できる．

被験者は 10 人である．システムを操作して指定した期間（1 か月）の記憶を想起させ，質問紙に答えさせた．

システム単体の実験（5 段階評価）では，「記憶想起」で 4.20（タグクラウド）と 3.90（ネットワーク），「デザイン」で 3.40（タグクラウド）と 3.80（ネットワーク）となる等，デザイン以外の質問においてタグクラウドの結果が良かった．2 システムのどちらが良いと思うかという直接比較においては，「記憶想起」で 7 人，「デザイン」で 8 人となる等，ネットワークの方が良かった．

タグクラウドの良さはキーワードが一目でわかること，ネットワークの良さはキーワード間の関連がわかることであることが，被験者のコメントから伺える．

6. おわりに

「個人の外化記憶構築」研究に関して，履歴構造を用いた記憶支援に関する 2013 年以降の進捗を述べた．今後も開発と実験を継続するとともに，記憶支援以外の効果についても検討する予定である．

参考文献

[村上 13] 村上 晴美：個人の人生の記録，2013 年度人工知能学会全国大会（第 15 回）論文集 (2013)

[村上 01] 村上 晴美，平田 高志: Memory-Organizer：個人の外化記憶構築システム，2001 年度人工知能学会全国大会（第 15 回）論文集 (2001)

[Murakami 04] Murakami, H., Hirata, T.: A System for Generating User's Chronological Interest Space from Web Browsing History, *International Journal of Knowledge-Based and Intelligent Engineering Systems*, Vol.8, No.3, pp.149-160 (2004)

[Murakami 12] Murakami, H., Mitsuhashi, K.: A System for Creating User's Knowledge Space from Various Information Usages to Support Human Recollection, *International Journal of Advancements in Computing Technology*, Vol. 4, No. 22, pp. 496-508 (2012)

[Murakami 10] Murakami, H.: History Structure for Exploring Desktop Data, in *Proceedings of the SIGIR 2010 Workshop on Desktop Search (Understanding, Supporting and Evaluating Personal Data Search)*, pp.25-26 (2010)

[松本 14] 松本 若樹，松浦 翔，村上 晴美：写真の枚数を用いた記憶の想起支援，情報処理学会第 76 回全国大会講演論文集，東京電機大学 (2014)

[Matsumoto 14] Matsumoto, M., Sho Matsuura, Kenta Mitsuhashi and Harumi Murakami: Supporting Human Recollection of the Impressive Events Using the Number of Photos, in *ICAART 2014 - Proceedings of the 6th International Conference on Agents and Artificial Intelligence*, Vol.1, pp.538-543 (2014)

[松浦 15] 松浦 翔，松本 若樹，村上 晴美：タグクラウドを用いた記憶の想起支援，情報処理学会第 77 回全国大会講演論文集 (2015)

[渡部 16] 渡部 渡部 秀太，橋本 悠佑，村上 晴美：記憶想起支援のためのキーワード抽出手法の改善，情報処理学会第 78 回全国大会講演論文集 (2016)

[村上 18] 村上 龍太郎，村上 晴美：複数の SNS を用いた記憶想起の支援，情報処理学会第 80 回全国大会講演論文集 (2018)

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

社員食堂の利用ログの活用を目指して
- グループ分析とメニュー分類 -

Toward Utilization of Use Log of Employee Cafeteria
- Group Analysis and Menu Categorization -

山下 達雄[*1] 笹谷 奈翁美[*1] 鍛治 伸裕[*1] 石下 美保[*1] 沼田 瑞木[*1] 嶋 隆宏[*1] 清水 伸幸[*1]
Tatsuo Yamashita Naomi Sasaya Nobuhiro Kaji Miho Ishige Mizuki Numata Takahiro Shima Nobuyuki Shimizu

[*1] ヤフー株式会社
Yahoo Japan Corporation

To improve employee satisfaction and engagement, we analyze dietary patterns in the usage log of Yahoo Japan Corporation employee cafeteria, which includes consumed menu items along with their time stamps. Employee cluster analysis and menu classification are conducted.

1. 序論

昨今、人々の働き方の改善に関して官民で施策がなされ、さまざまな調査も行われている[1]。

社員の働き方の改善に向けた分析対象として、我々はヤフー株式会社(以下ヤフー)の社員食堂の利用ログに着目した。

ヤフーの社員食堂では、社員 ID と飲食したメニューと日時のログが蓄積されている。これらを分析し、「食べたもの」による社員の健康やパフォーマンスへの影響が分かれば、社員食堂のメニュー企画やユーザへのレコメンド等さまざまな応用が可能になり、社員の働き方の質の向上につなげることができる。我々は、その流れの第一段階として、食堂利用ログ自体で完結する分析を行った。

本稿では、まず、対象となる社員食堂のデータの概要、および、集計によって見えてくる傾向について述べる。次に、今回我々が行ったグループ分析とメニュー分類の 2 つの分析についてそれぞれ説明する。最後に、社員食堂ログと他のデータを結びつけることで広がる活用の可能性について考察する。

2. 分析対象データ

本研究の分析対象であるヤフーの社員食堂とそのログデータについて述べる。

2.1 ヤフーの社員食堂

ヤフーの本社がある東京の紀尾井町オフィスはグループ会社もあわせて 7,000 名超の従業員が勤務している。そのオフィス内にある社員食堂「BASE11」は 2016 年 10 月の本社移転と同時にグランドオープンした(図 1)。座席数 822 席、広さ 3,300 平米で、朝・昼・夜の 3 つの時間帯で食事を提供している。

ランチメニューは定食 4 種類、麺類 3 種類(ラーメン・うどん・そば)、丼、カレーのほか、量り売り「グラムビュッフェ」など 10 種類。有名店とのコラボや拠点がある地域の特産品を使ったメニュー等の特別メニューも定期的にある。1 日約 2,600 食が出る。

連絡先:山下達雄, ヤフー株式会社,
tayamash@yahoo-corp.jp

会計はセルフサービスでフロアに 8 つあるオートレジで行う(図 2)。食器をのせたトレーを置くと食器に貼り付けられたチップの情報が読み取られ合計金額がレジ画面に表示される。利用者は社員証または交通系 IC カードを専用端末にタッチし精算をする。社員証決済したものはイントラの専用ページで履歴が閲覧でき、金額以外にもメニュー名、カロリー、3 大栄養素等が表示される。

図 1: BASE11

図 2: オートレジ

2.2 社食ログ

BASE11 では、メニューの会計ごとに下記の要素を含むログが蓄積される。

- 会計日時
- 社員 ID（匿名化されている）
- 会計したオートレジ番号（1～8）
- メニュー名

ログの例を下記に示す。

会計日時, 社員 ID , レジ番号, メニュー名
2017/02/22 13:08:06, ACC110BF, 1, カツオのたたきの漬け丼
2017/02/22 13:10:52, BF0E7899, 3, かけそば
2017/02/22 13:14:01, BADE7368, 2, 麻婆うどん

本研究での分析対象データは、BASE11 の昼食のみで、期間は 2016 年 9 月 26 日から 2017 年 3 月 15 日である。ログ数は約 40 万件、利用者は約 5000 人である。

簡単な集計により、社食利用の傾向を見ていく。図 3 は曜日ごとの利用者数のグラフである。金曜日が少なめな傾向があるのは、休日の前に在宅勤務を行う社員が多いためだと思われる。図 4 は社員ごとの期間中の社食利用日数の分布である。週に 2,3 回利用する利用者が多く、定常的に利用されていることがわかる。

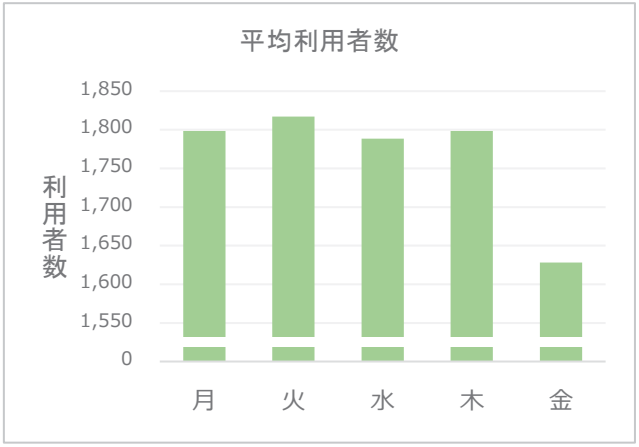


図 3

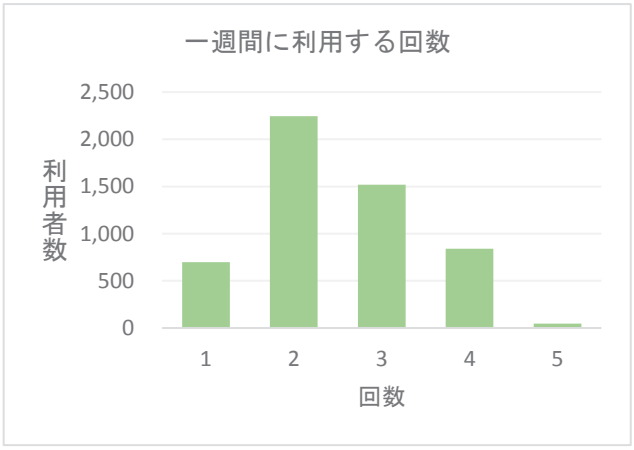


図 4

3. ログの分析

我々は最初のステップとして、食事ログのみを用いて、グループ分析とメニュー分類の 2 つの分析を行った。グループ分析では、何人のグループで食事をしているかという行動情報を食事ログから推定した。メニュー分類では、食事ログからメニューの分散表現を学習し、メニューのクラスタリングとその可視化を行った。

3.1 グループ分析

会社生活における食事（ランチ）には、一人で食べるか、誰かと食べるか、また、いつも同じ人と食べるか、といったコミュニケーションの側面がある。この切り口で、将来的には食べるものや食事以外の行動の差異の有無を分析したい。そのための前段階として、社食ログを用いて食事グループの分析を行った。

まず、食事行動の概念として「ひとり飯」「一緒飯」を定義する。ひとり飯は、社食でほぼ毎回一人で食事をすること、一緒飯は、社食でほぼ毎回複数人で一緒に食事をすることとする。

現在の BASE11 では複数の社員が一緒に食事をしたかどうかを直接知る方法はない。そこで、代替手段としてオートレジの会計時間を利用した。社食利用者は、グループでの食事を終えると一斉に会計に向かうことがほとんどである。つまり、ある短い時間内に会計を済ました人々は一緒に食事をしていた可能性が高いと言える。それが一度だけならば、たまたま会計時間が近い別のテーブルの人のこともあるが、頻度が大きければ確実性は増す。この考えをもとにして一緒飯・ひとり飯の判定方法を考えた。

1. 一定時間内（S 秒以内）にレジ会計を済ます集団（全組み合わせ）を検知
2. 1 のうち一定以上頻度（N 回以上）のある集団を一緒飯集団とする
3. その他をひとり飯とする

時間 S を 60、頻度 N を 5 として、利用者 5349 人に対して分析を行った結果、一緒飯該当者数（1 回でも一緒飯をしたことある人）は 2181 人（40.8%）、一人飯該当者数は 3168 人（59.2%）となった。

図 5 に一緒飯ネットワークの一部を示す。ノードは社食利用者を、エッジは一緒飯の関係にあることを表す。また、エッジの太さは一緒飯の頻度（一緒に食べた）回数を、ノードの色は濃いほど多くの人と一緒飯をしたことを表す。例えば、右下の集団は高頻度で一緒飯をしており、同じチームのメンバーであると推

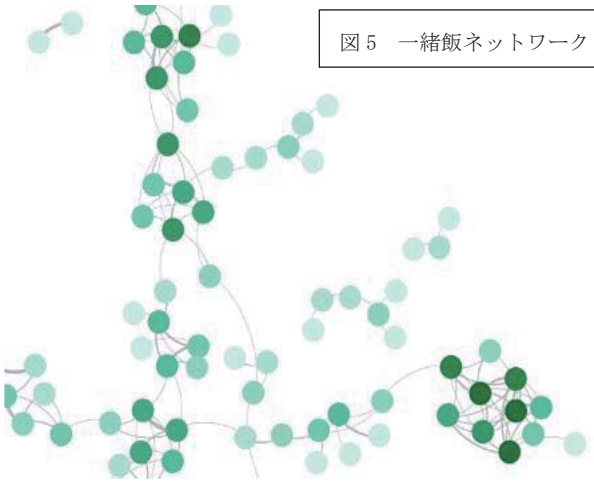


図 5　一緒飯ネットワーク

測される。また、左上の 2 人は一緒飯の頻度が高く、他の利用者とのエッジもないことから、ほぼ常に一緒に昼食をとっていると考えられる。

なお、現段階では、実際に一緒飯をしたかどうかの検証は行えていない。今後、アンケートなどでの評価する予定である。

3.2 メニュー分類

どのユーザ（利用者）がどのメニューを注文したかというログデータからメニューの分散表現を学習し、メニューのクラスタリングおよび可視化を行い、その有効性の検証を行った。

社食ログデータにおけるユーザのメニュー注文ログを、ユーザとメニューの共起データであるとみなし（あるユーザがあるメニューを注文したときそれらのユーザとメニューが 1 回共起したと考える）、skip-gram model with negative sampling [2] を用いてメニューの分散表現を求めた。学習処理には我々の独自実装のライブラリを用いた[3]。分散表現の次元数は 100 次元とした。その結果得られたメニュー分散表現に k-means を適用してクラスタリングを行い、tSNE [4]による結果の可視化を行った。

図 6 に可視化の結果を示す。麺類、小鉢、デザートなど、直感的に妥当なメニューのクラスタが得られていることが確認できる。この結果より、今後のログ分析調査において、例えば類似したメニューを自動的に集約したのちに種々の統計情報を算出するなど、メニューの分散表現の活用の可能性を確認できた。

4. 社食ログ分析の今後

本稿では社員食堂ログのみを使った分析として、グループ分析とメニュー分類を行った。グループ分析では食後の会計時間を用いて一緒に食事をとる「一緒飯」を抽出し、メニュー分類では利用者とメニューの共起を元にメニューの分散表現を求めクラスタリングを行い、今後の分析においての利用可能性を確認した。

今後は、他のデータと結びつけることにより活用範囲を広げていく予定である。例えば、社員の健康診断データやパフォーマンスなどと食べたものの関係を調べることで、働く環境の快適化をサポートできる。また、飲食パターンによるクラスタリング（ラーメンクラスタ、ローカーボクラスタなど）や状況や属性に応じた食事のレコメンデーションなどの応用も考えられる。

参考文献

[1] 働き方改革の実現，首相官邸ホームページ https://www.kantei.go.jp/jp/headline/ichiokusoukatsuyaku/hatarakikata.html

[2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13), 2013.

[3] Nobuhiro Kaji, Hayato Kobayashi. Incremental Skip-gram Model with Negative Sampling. In Proceedings of EMNLP. 2017.

[4] Laurens van der Maaten, Geoffrey Hinton. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9(Nov):2579-2605, 2008.

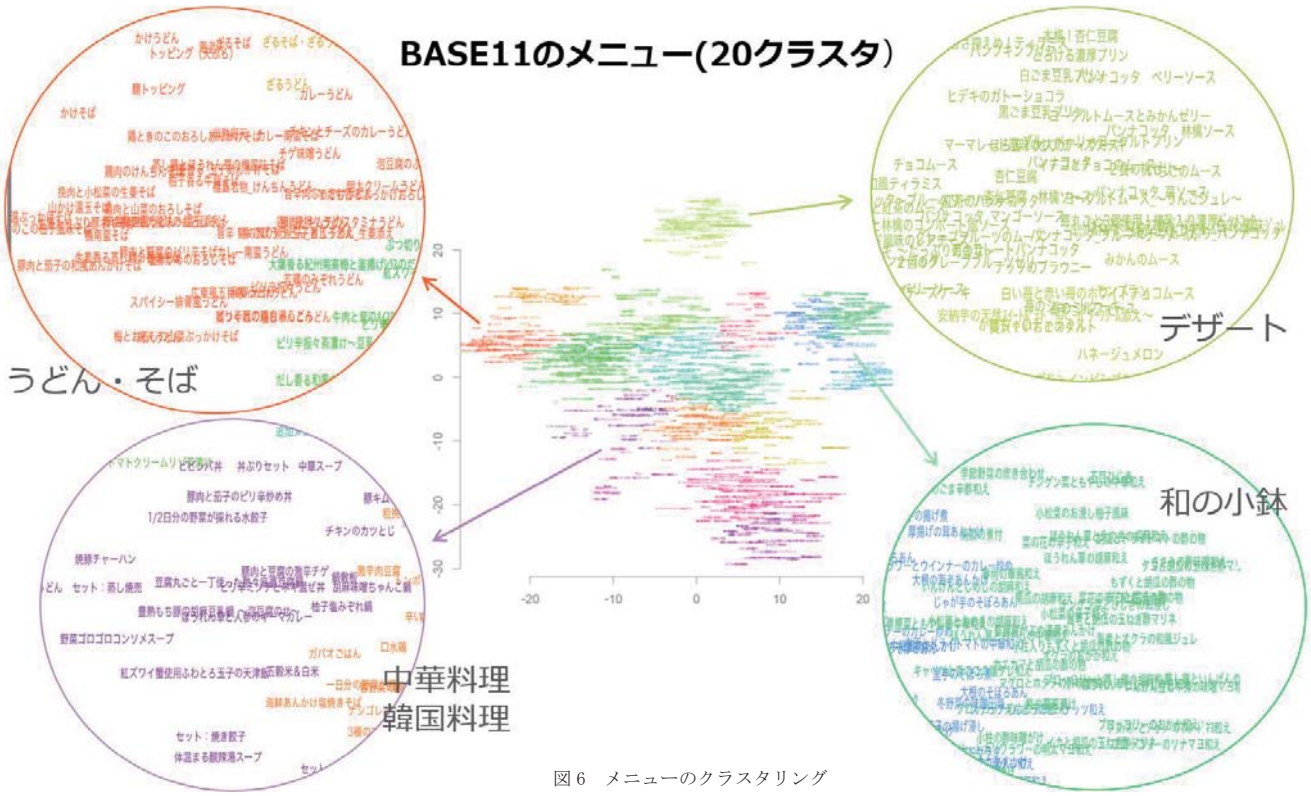


図 6　メニューのクラスタリング

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

ビットパターンカーネルフィルタによる細胞判定重複除去

Bit-pattern kernel filtering for duplication removing in cell annotation

長坂 暢[*1]
Toru Nagasaka

[*1] 株式会社エーアイ・アンド・アイ
AI & Eye Co., Ltd.

In recent years, deep learning architectures have been shown to achieve good classification performance and a digital histopathology image analysis has been used in clinical settings. A partial image which are sequentially cropped from same cell by grid search may result in double counting. In this work, we constructed bit-pattern kernel filtering algorithm to remove duplication in cell annotation. Using unique combination of bit-pattern kernel (3x3) filter which contract particular ON-pixel pattern, the duplication remove in cell annotation was achieved.

1. 背景

近年，深層学習法による画像認識精度は飛躍的に向上しており，医療現場への応用も始まっている．今回，病理組織および細胞検体へ画像解析を応用する際の細胞判定重複の問題を解決するアルゴリズムを開発したので報告する．通常，細胞画像から格子状に各座標の部分画像について細胞判定が行われ，各種の細胞の個数が計数される．しかしながら，例えば 2 個以上の部分画像が同じ細胞から切り抜かれると，実際には 1 個と数えられるべきにも関わらず，2 個以上と計数されてしまう（図 1）．

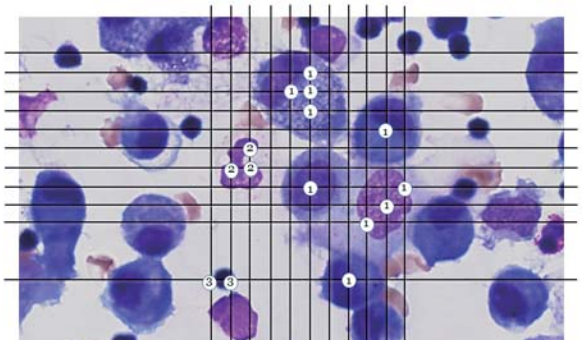


図 1: 細胞判定の実際．格子状走査(grid search)により同一の細胞上で複数の同一判定が起きうる．

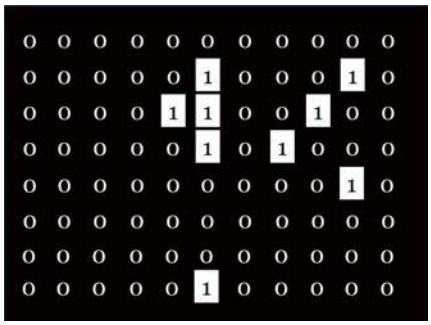


図 2: 細胞判定を ON 画素と OFF 画素のビット形式へ変換する．ON 画素の隣接を重複と定義し，削除の対象とする．

連絡先: 株式会社エーアイ・アンド・アイ
長坂 暢 toru.nagasaka@ai-and-eye.co.jp

ここで，細胞判定のグリッドを ON 画素と OFF 画素へ変換したマトリックス（図 2）における ON 画素の上下左右・斜め方向の隣接状態を判定重複と定義し，図 3 のように ON 画素を 1 つにまとめるアルゴリズムを報告する．

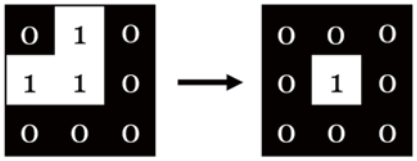


図 3: 細胞判定重複の例．ON 画素が 3 個隣接しているものを 1 個の ON 画素へまとめる．

2. 方法

2.1 深層学習による細胞判定

顕微鏡画像に対し， 18.3×18.3μm 部分画像を格子状に順次切り抜き，畳み込みニューラルネットワーク(CNN)または Residual networks (ResNet) [Kaiming 2015] を部分構造として用いた深層学習法による判定を行った．

2.2 ビットパターンカーネル

3×3 ピクセルの中で ON 画素が 1 つのもの(Kernel 1)を 1 種類(図 4), Kernel_2 を 8 種類(図 5), Kernel_3 を 20 種類(図 6), Kernel_4 を 32 種類（figure not shown）, Kernel_5 を 38 種類(figure not shown)準備した．

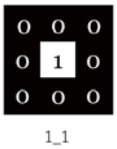


図 4: Kernel 1. ON 画素が 1 個.

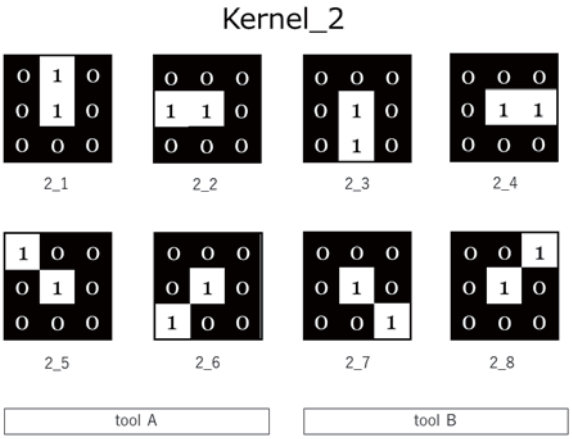


図 5: Kernel 2. ON 画素が 2 個の組み合わせ.

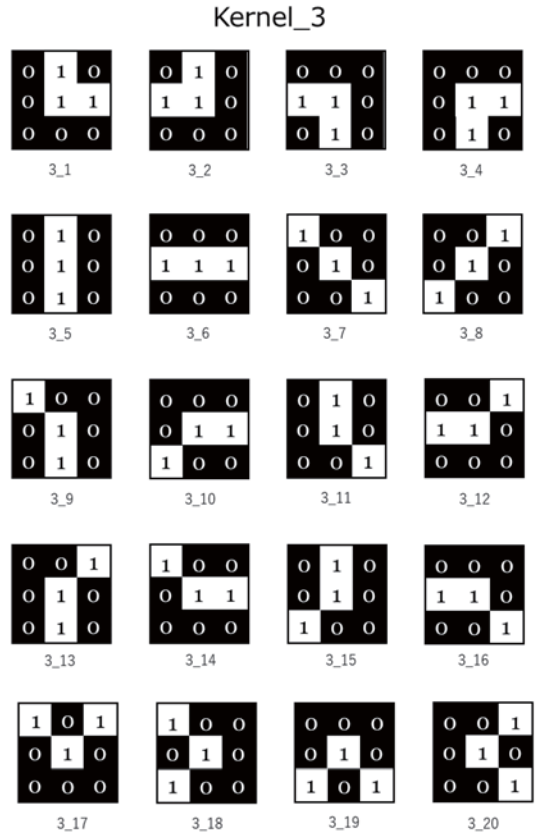


図 6: Kernel 3. ON 画素が 3 個の組み合わせ.

2.3 カーネルによる走査とカーネルセット

入力マトリックスに対し複数のカーネルによる走査を行い，一致したパターンがあるとその部位の中心のピクセルを ON 画素として記録し，それをマトリックスとして出力する.

ON 画素が横に 3 つ隣接する判定重複を例に説明する. Kernel 3_6 による走査で 2 行 3 列目（座標(2,3)）に ON 画素が記録される(図 7). また Kernel 2_4 による走査では座標(2,2)に ON 画素が記録される(図 8). 2 つのカーネルの走査により出力されたマトリックスを合成し，次の Kernel 2_2 への入力マトリックスとする（図 9）.

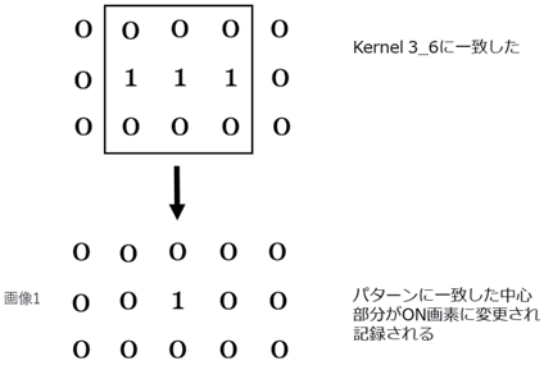


図 7: Kernel 3_6 による走査.

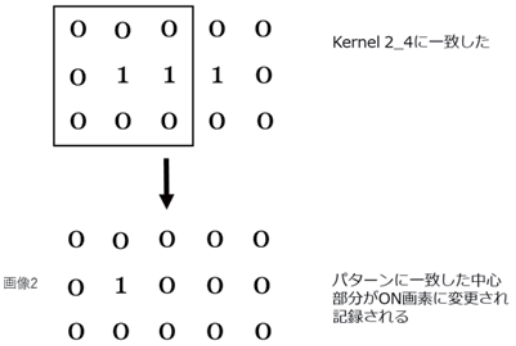


図 8: Kernel 2_4 による走査.

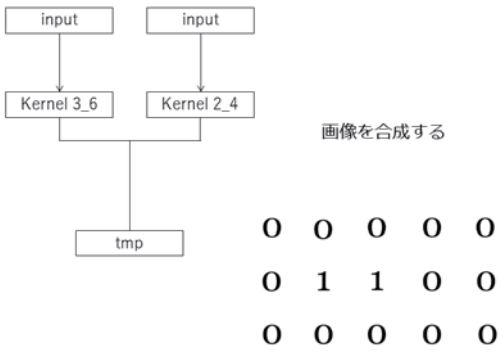


図 9: 合成マトリックスの生成.

合成されたマトリックスを Kernel 2_2 で走査すると座標(2,3)のみが ON 画素となったマトリックスが得られる(図 10). このように 3 種類のカーネルを組み合わせて順次処理することで, 重複除去が得られる. この組み合わせを Kernel 3_6 セットと呼ぶ.

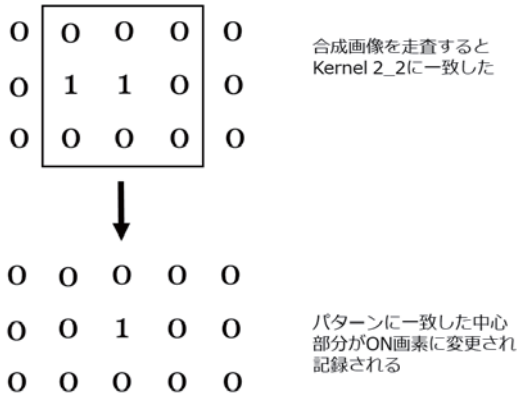


図 10: 合成マトリックスに対する Kernel 2_2 による走査.

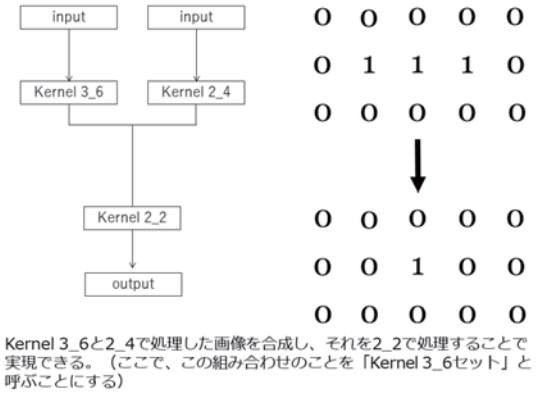


図 11: Kernel 3_6 セット. ON 画素が横に 3 つ隣接する判定重複に対する処理方法のまとめ.

2.4 走査ステップ

前述したカーネルセットと単独のカーネルによる走査を並列で行う処理を 1 つのステップとして構成する. 例として Step A で使用されるカーネルセットとカーネルを図 12 に示す.

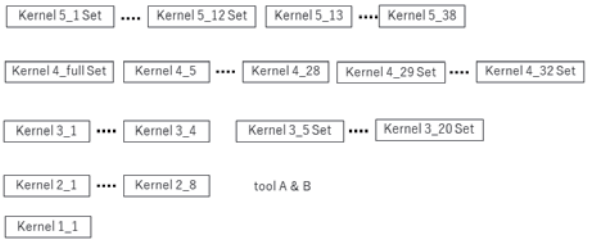


図 12: Step A. 使用されるカーネルとカーネルセット.

さらに使用するカーネルを順次減らして Step B, Step C, Step D を構成する(data not shown). 入力マトリックスを Step A から順に Step D まで処理することで, 重複判定が段階的にまとまっていく(図 13).

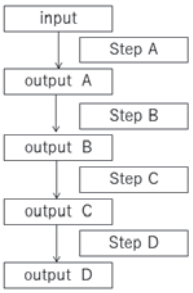
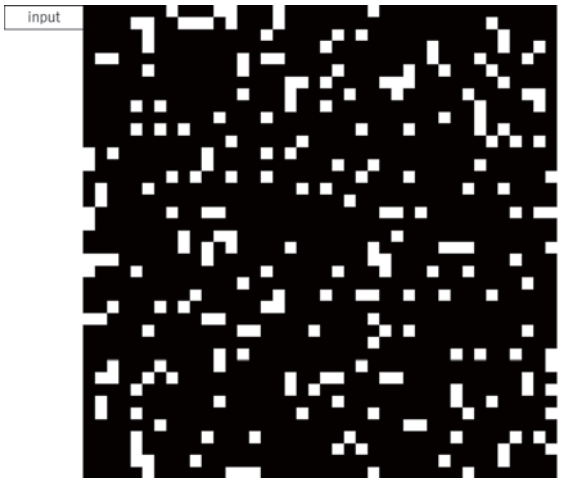
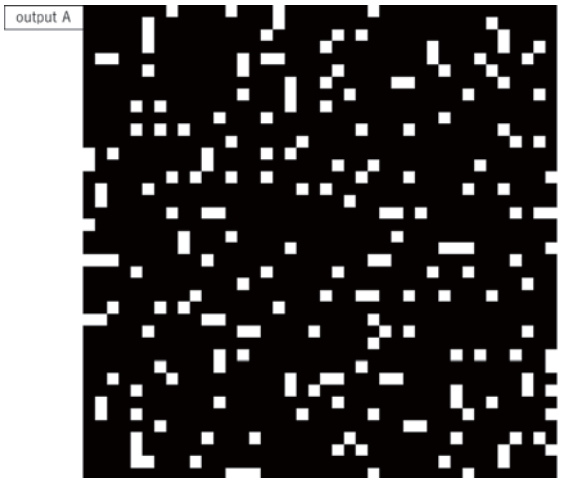


図 13: 走査ステップの流れ. Step A からの出力を Step B に入力し, 順次, マトリックスを処理していく.

3. 結果

インプットおよび各ステップから出力されたマトリックスを示す(図 14).





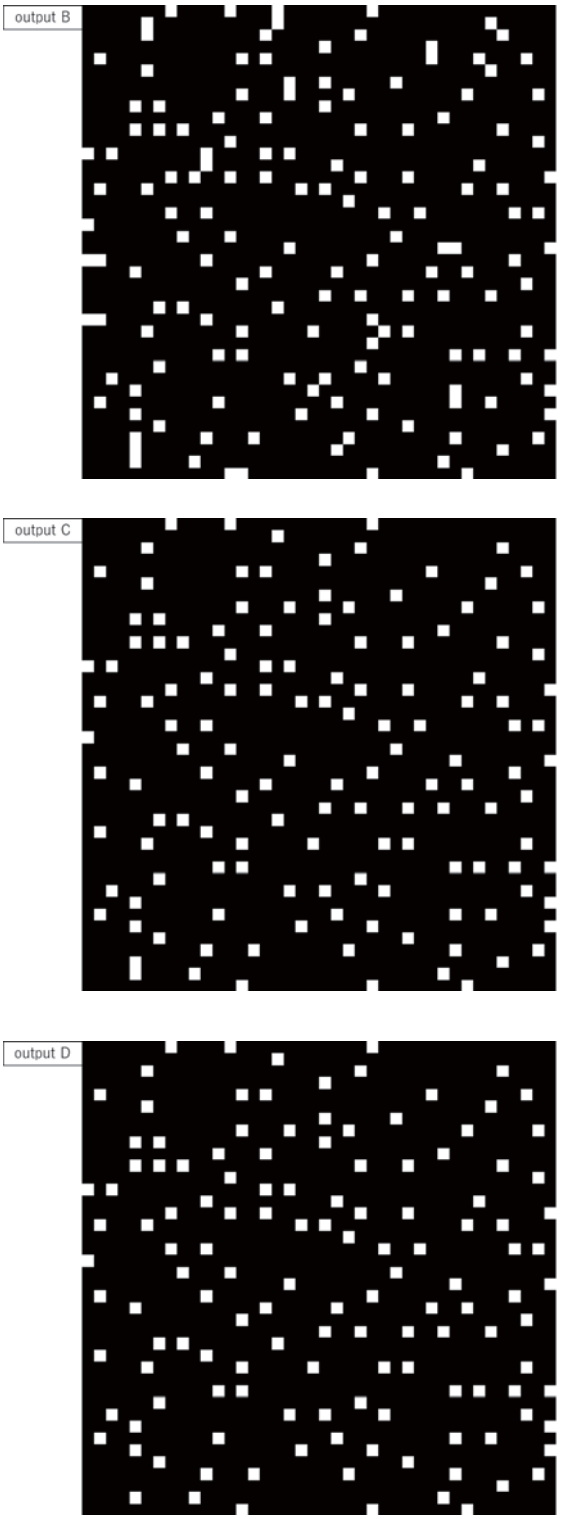


図 14: 各ステップからの出力マトリックス. Step A から順に処理することで, 重複部分が除去され, 隣接のない ON 画素のみとなっている.

実際の細胞判定への適用例を示す(図 15).

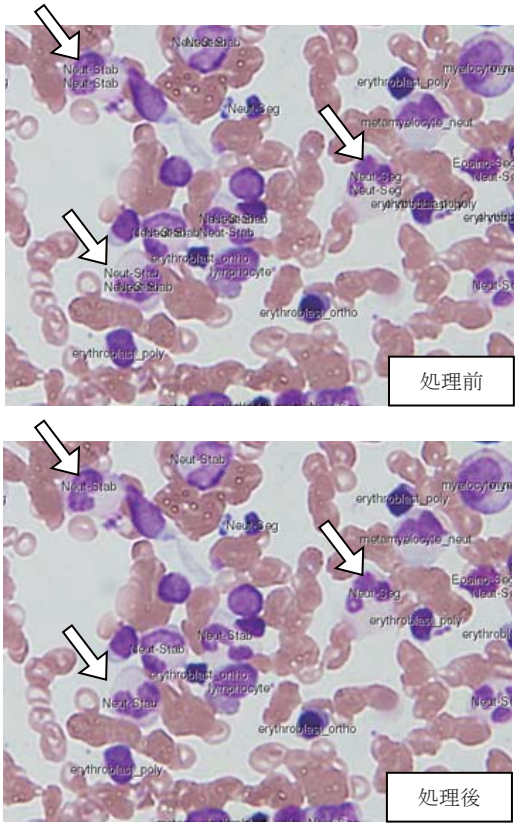


図 15: 細胞判定の重複除去例. 処理前には同一の細胞で重複して判定がでている部分があるが(上図矢印), 処理後には判定重複がなくなっている(下図矢印).

4. 考察

深層学習法の画像認識分野での進歩により, 細胞判定自体の精度が飛躍的に向上している. 臨床応用が期待される中で, 人工知能の完成度を上げ, 実際の運用に耐えられるような背景技術が求められている.

具底的な問題点として, 判定重複を挙げた. 関連技術として画像処理の分野では, 膨張処理, 収縮処理およびその組み合わせのオープニング, クロージングが広く用いられている. しかしながら細胞判定マトリックスに収縮処理を適用すると残すべき隣接のない ON 画素も消失してしまうため使用することができない.

今回, 判定重複を除去して計数の精度を上げるアルゴリズムを開発した. 細胞カウントだけでなく, 同様な集計結果の精度が重視されるタスクに応用が可能と考えられる.

参考文献

[Kaiming 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CoRR, Vol. abs/1512.03385, 2015.

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

Twitter URL Paraphrase Corpus に基づく要約データセットの構築

Construction of dataset for summarization based on the Twitter URL Paraphrase Corpus

永塚 光一[*1]　　渥美 雅保[*1]
Koichi Nagatsuka　　Masayasu Atsumi

[*1]創価大学理工学部情報システム工学科
Information Systems Science, Faculty of Science and Engineering, Soka University

The purpose of text summarization is to produce a condensed version of an input text which has a core meaning of the original. Most of the summarization systems are built on dataset using news articles. While Social Networking Service(SNS) such as Twitter is increasingly becoming an important information resource, the lack of dataset for SNS is one of the challenges in extending the range of summarization systems. In this paper, we address the problem by transforming the Twitter URL Paraphrase Corpus into summarization dataset. In order to extract important paraphrases for summary and increase the number of high quality paraphrases, we created the paraphrase classifier and paraphrase generator using supervised learning based on the corpus. In experiments, we evaluate paraphrase classifier by quantitative evaluation and paraphrase generator by qualitative evaluation respectively.

1. はじめに

近年, アテンションを取り入れたエンコーダデコーダモデル[Bahdanau 14]による機械翻訳における性能向上を受けて, ニューラルネットワークを用いた自動要約研究が盛んに行なわれている.自動要約研究に用いられるデータセットとして, DUC[Over 07]や, Gigaword[Graff 03]などがある. しかし, こうした要約データセットの多くが, 要約対象のデータとしてニュース記事を採用している一方で, 現在ニュース記事と並んで Twitter に代表される SNS が大きなテキスト情報源となりつつある.

本論では, URL を共有する tweet に基づいて自動収集されたパラフレーズのデータセットである Twitter URL Paraphrase Corpus[Lan 2017]を要約データセットの構築へ応用することを提案する. Twitter URL Paraphrase Corpus の一部には, 人間の評価者により各派生 tweet に対する元 tweet との意味的類似度ラベルが付与されており, これを教師データとして 2 つの tweet が paraphrase とみなせるか否かを判定する paraphrase 分類器を作成し, 良質な paraphrase ペアを収集している. しかしながら, twitter 要約データセットには, 入力としてより多くの paraphrase ペアを収集することが望ましい. そこで, 本手法では, paraphrase 分類器に加え, 元 tweet から派生 tweet を生成する paraphrase 生成器を学習により作成する. 要約データセットの構築は, 分類器と生成器により, 元 tweet に対して派生 tweet を選別, もしくは生成することによりなされる. 本論では, 分類器と生成器の性能を実験により評価することを通じて, これらを用いて, 要約データセットを構築することの実現性を示す.

2. データセット

2.1 Twitter URL Paraphrase Corpus

Twitter URL Paraphrase Corpus は,教師ツイート 1 文とそれに対する複数の派生ツイートを 1 ペアとして,51524 ペアから構成される.教師ツイートは,CNN や New York Times, BBC などのニュース機関の公式アカウントから投稿されたもので, 派生ツイートには,教師ツイートとのパラフレーズとしての意味的類似度がラベル付けされている.この意味的類似度の決定においては, 6 人の評価者により,各派生ツイートが教師ツイートに対するパラフレーズと言えるかどうかの投票が行われ,最小で 0,最大で 6 の 7 段階でラベルが付与される. これに加え,ラベル付けがされていないサブデータセットが 114,025 ペア公開されている. このサブデータセットは,人手によりラベル付けされたデータセットから作成された paraphrase 分類器により収集されている[Lan 17]. 本手法では,こうした paraphrase 分類器の作成に加え, ラベル付けのないデータセットから paraphrase 生成器も作成し.要約データセットの構築に適用させることを提案する. 図 1 にラベル付けされたデータセットにおける派生ツイートのラベルの内訳を示す.

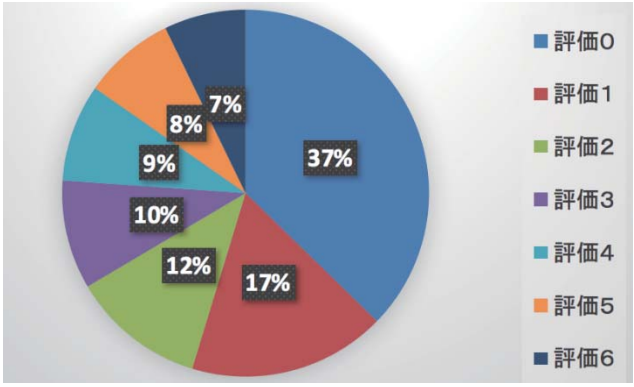


図 1 Twitter URL Paraphrase Corpus の
意味的類似度ラベルの内訳

2.2 Twitter 要約データセット

あるトピックに対する要約タスクを考える時, ニュース記事と比べて Twitter が異なる点として, 書き手が複数であり, 表現がまばらであるということが挙げられる. 特に, Twitter などにおいては,表現形式として短文が主体であり, 同じ内容の言い換え表現であることが多い. これらのことから, 本手法では, Twitter の要約の一形態を, トピックを共有する paraphrase 群からの要約文生成タス

連絡先: 永塚 光一, 創価大学理工学部情報システム工学科
〒108-0074 東京都港区高輪 1-1-11-206 080-6576-6124
e1452206@soka-u.jp

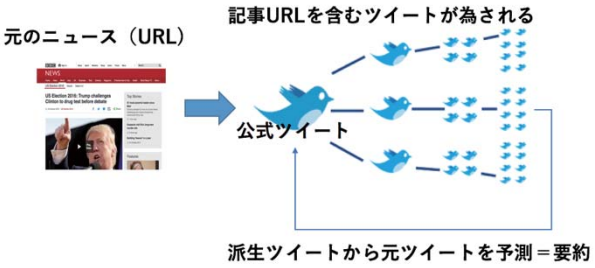


図 2 tweet 要約の流れ

クとして定義する(図 2). Twitter URL Paraphrase Corpus において, tweet に同じ URL を共有することと, 同じトピックを共有することは, 同義となっており, この前提に従うことで, 多くの paraphrase tweet を, URL のみに基づいて, paraphrase 分類器により機械的かつ持続的に収集することが可能となる[Lan 17]. 実際に, Twitter URL Paraphrase Corpus は, 分類器を用いて, 月平均で, 約 30000 の paraphrase を収集することに成功している. 提案する要約データセット収集システムもこの手法に則り, 独自に paraphrase 分類器を作成した. 加えて, 本手法では,paraphrase 生成器により, データセットを増加させることを提案する. これは, 要約モデルの学習が, 入力データとして複数文の paraphrase を必要とするため, 元 tweet に対して,通常よりも多くの paraphrase を含むことが望ましいからである.

3. 提案手法

paraphrase 分類器と paraphrase 生成器を組み合わせた要約データセット構築の流れを図 3 に示す.

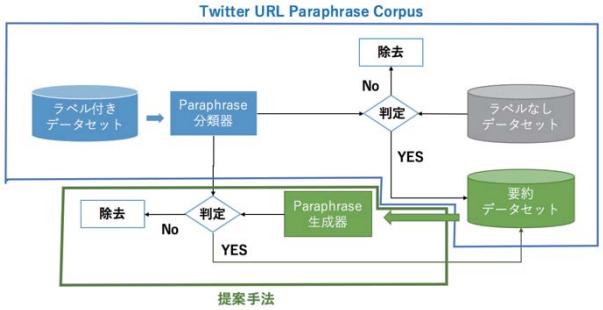


図 3. 要約データセットの構築手法

まず, ラベル付きデータセットを訓練データとして, paraphrase 分類器を作成する. その後, ラベルなしデータセットに対し, この分類器を適応させることで, paraphrase である可能性の高い tweet を自動的に選別する. 加えて, 構築したデータセットより, paraphrase 生成器を学習する. 最終的に生成された paraphrase をもう一度, paraphrase 分類器に判別させることで, 要約データセットの総量を増加させることが出来る.

3.1 Paraphrase 分類器

paraphrase 分類器学習においては, データセット構成によるバイアスを防ぐため, 全体における割合が高い 0~3 のラベルを 0.5 の確率で除去した.結果的に, 36532 ペアのトレーニングデータ, 1000 ペアのテストデータを獲得した. 図 4 に分類器ネットワークの簡略図を示す. 各 tweet を, 重みを共有した LSTM に入力したのち, 最終の隠れ層を結合し, フィードフォワードニューラルネットにかけて, sigmoid 関数により, paraphrase である確率を生成

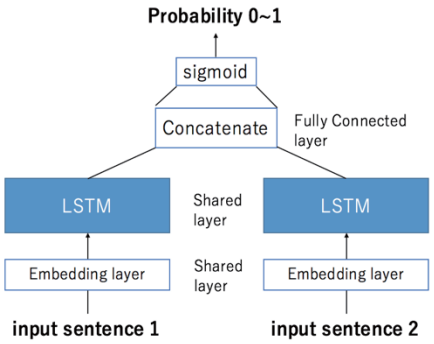


図 4. Paraphrase 分類器

する. この時, 教師データとして, 0~6 のラベル値を 6 で割り, 教師信号となる確率として正規化して与えている.

3.2 Paraphrase 生成器

生成器の学習には, attention 付きエンコーダデコーダモデル [Bahadanu 14]を使用している. ここでは, 学習速度を速めるため, LSTM よりパラメータ数の少ない GRU を用いている. また, 学習では, [Lan 17]による paraphrase 分類器によって, 事前に処理されたデータセットを用いて, トレーニングを試みている. 文の最大長は 25 であり, 最適化手法を Adam として, 1,000,000 イテレーションの学習を行った.

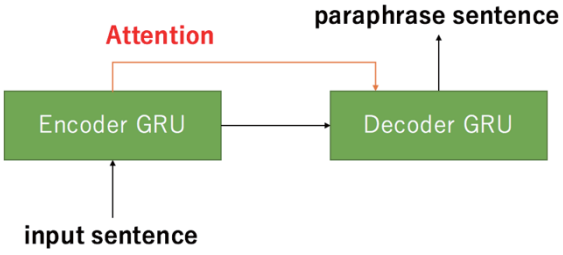


図 5 paraphrase 生成器

4. 実験

4.1 実験概要

実験では, paraphrase 分類器に対して, recall と accuracy による分類精度評価を, paraphrase 生成器に対して, 質的評価をそれぞれ行う.

4.2 分類器の実験結果

図 6, 7 にトレーニングデータとテストデータにおける分類性能を示す. 分類器のトレーニングでは, 学習高速化及び汎化性能向上のために, ミニバッチを導入し, その値を 100 としている. 文の最大長を 25 で統一し, ドロップアウト率を 0.5 とした上で, 3epoch 学習させた. 実験では, paraphrase と判断する出力確率の閾値を 0.1~0.9 まで 0.1 毎に変えて, トレーニング時とテスト時における正例と負例の recall, 及び accuracy の性能を評価する.

図 6, 7 から分かるように, トレーニングとテスト時において, 閾値を上げる程, 正例の recall は下がり, 負例の recall は増加している. また, 図 8 が示すように, トレーニング時とテスト時のどちらも, accuracy は中央値である閾値 0.5 付近において, 最も高い値であった. (トレーニング時で 85.8%, テスト時で, 71.6%). 閾値の設定は, これらの結果を踏まえた上で, 収集したい paraphrase の質に応じて, 変える必要がある.

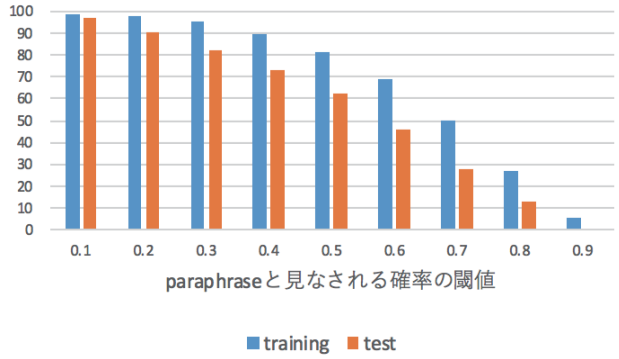


図 6 閾値の変化による正例の recall(%)

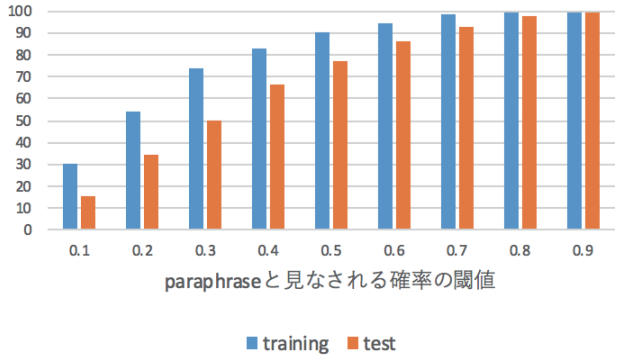


図 7 閾値の変化による負例の recall(%)

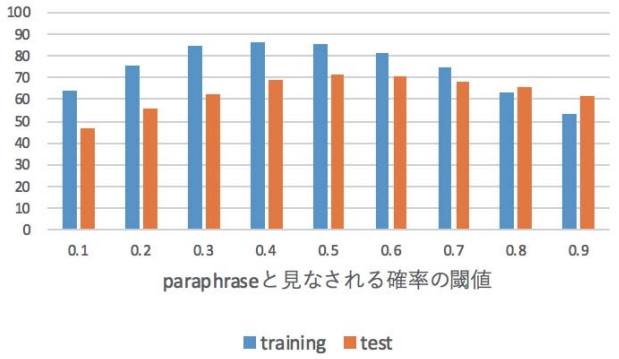


図 8 閾値の変化による accuracy(%)

4.3 生成器の実験結果

表 1 は, 生成した paraphrase の一部を示したものである. 生成 paraphrase を見ると, 中心的な意味を保ち, 語彙や付属語を変化させて文法的に正しい paraphrase を生成出来ていることがわかる. しかしながら, 中には, 単語の繰り返し表現や, 文法的かつ意味的に成り立たない paraphrase も多く含まれていた. 生成した paraphrase の末尾には, 作成した分類器によって出力した元 tweet に対する paraphrase である確率を示している. 表 1 に示したサンプルに関しては, どの文も 7 割を上回っている. これは,前述した閾値の設定において, 最大 0.7 を選択した場合においても, paraphrase としてデータセットに採用されることを意味する.

表 1 生成した paraphrase 例

example1
input
Donald Trump Agrees to Pay $25 Million in Trump University Settlement
reference
donald trump agreed to pay million to settle fraud lawsuits . about students will share in the settlement .
genetated paraphrase examples
donald trump agreed to pay million to settle fraud lawsuits . <EOS>
example2
input
Be careful what you say of others . Donald Trump Agrees to Pay $25 Million in Trump University Settlement
reference
donald trump agreed to pay million to settle fraud lawsuits . about students will share in the settlement .
genetated paraphrase examples
donald trump has agreed to pay million to settle three lawsuits against trump university . <EOS>
example3
input
Trump in March I dont settle cases . I win in court . Agrees to pay $25M to settle Trump U case
reference
donald trump agreed to pay million to settle fraud lawsuits . about students will share in the settlement .
genetated paraphrase examples
donald trump agreed to pay million to settle three lawsuits against trump university . <EOS>

5. むすび

本論文では, Twitter URL Paraphrase Corpus を活用した SNS 要約データセットの構築システムを提案した.その上で, データセット構築に用いる paraphrase 分類器と paraphrase 生成器の性能評価を行った. 分類器と生成器は共に改善の余地はあるものの, システムの実現可能性を示すことが出来た. 今後の課題として, 本手法により構築したデータセットと, 処理を加えないデータセットに対して, 要約学習を行い, 結果を比較する必要がある.

参考文献

[Over 07]Paul Over, Hoa Dang, and Donna Harman. 2007. Duc in context. Information Processing & Management, 43(6):1506–1520.

[Graff 03]David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2003. English gigaword. Linguistic Data Consortium, Philadelphia.

[Bahdanau 14]Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio: Neural machine translation by jointly learning to align and translate, *arXiv preprint arXiv:1409.0473*, 2014.

[Lan 2017]Wuwei Lan, Siyu Qiu, Hua He, Wei Xu: A Continuously Growing Dataset of Sentential Paraphrases, *arXiv preprint arXiv:1708.00391*, 2017.

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

スマートフォンアプリの GUI ユーザビリティ評価方法の提案

Proposal of GUI usability evaluation method of smartphone application

田中　貴之[*1]
Takayuki Tanaka

井上聡[*1*2]
Satoru Inoue

[*1] 埼玉工業大学大学院
Graduate School of Engineering, Saitama Institute of Technology

[*2] 埼玉工業大学
Saitama Institute of Technology

Although the size of the hand, how to hold the smartphone, how to use it is different depending on the person, the application can not respond to such diversity. As a result, users using smartphones will be dissatisfied with the design of the interface. In order to eliminate such problems it is also necessary to ascertain what kind of interface design to satisfy the all smartphone users. Therefore, we propose an interface evaluation method to deal with various usage.

1. はじめに

今、スマートフォンは私たちの生活にはなくてはならないものとなりつつある。スマートフォンを使っていく中で、多くの人が様々なアプリケーションをインストールしている。アプリケーションによりデザインや機能は全く異なる。そしてアプリケーションの多機能化により、インターフェイスのデザインが複雑になりつつある。そのことにより上手く操作できない場合や、不便に感じる場面もある。そして人により手の大きさ、スマートフォンの持ち方は異なり、このことも不便に感じる要因になると考えられる。しかし、上記で述べたように人により使い方は違うのにも関わらず、それぞれのスマートフォンの使い方に合わせようとするアプリケーションは数少ない。

そこで本研究では、上記のような問題を解決するために実験用アプリケーションを作成し使用してもらう。その過程での指の動きやスマートフォン本体の動きなどを取得し、そのデータを元にどのようなインターフェイスのデザインがユーザーを不満に感じさせるのかを検出する。この検出されたデータを解析することにより、多くのユーザーが操作しづらいと感じるデザインが明らかになると考えられる。

2. 本研究の目的

本研究では多くの人にテスト用アプリケーションを使ってもらう。指示された一連の操作をしてもらい座標値、時間、加速度センサーのログを取得する。取得したログを元にアプリケーションを使用した多くのユーザーが操作しづらいと感じた操作や配置を検出する。指示された操作が他の指示よりも時間が掛かかる、スマートフォン本体の動きが大きい場合、ユーザーがインターフェイスデザインに不満を感じていると考えられる。画面上の異なる座標でボタンを配置し一連の操作をしてもらうことにより、有益なデータを取得できると考えられる。

持ち方や手の大きさの違いによって、どのようなインターフェイスデザインが人々に不満の違いが出るのか。そして検出されたデータを元に、アプリケーションを使うユーザーが操作しづらいと感じるインターフェイスのデザインを明らかにすることを目的とする。そのことによりインターフェイスデザインの改善を行うことで、ユーザビリティの向上が可能に繋がると考えられる。

連絡先：田中　貴之，埼玉工業大学大学院，埼玉県深谷市普済寺 1690 番地，E-mail：schoo938mt@gmail.com

3. 方法

ユーザーに行ってもらう操作はタップ、スワイプ、フリックの3つの操作を指示された通りに複数回行ってもらう。

操作している最中に指示されてからの時間、画面上に指が置かれた座標、操作している最中のスマートフォン本体の傾きを取得する。

手の大きさやスマートフォンの持ち方によってユーザーが不満を感じるデザインは異なると思われる。実験者には年齢（20代後半、30代前半など）、性別、手の大きさ、スマートフォンの持ち方（右手片手持ち操作、右手持ち左操作など）を入力してもらう（図3）。入力されたデータを元にデータを分類する。

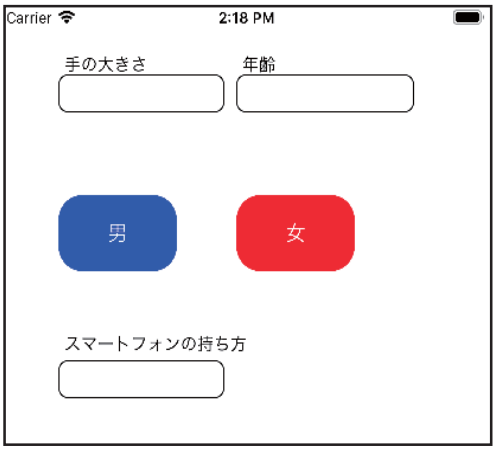


図 3 実験者に提示する入力画面

3.1 タップ操作

タップとは、画面を一回突く操作方法である。画面上の一点を指定する場合、ボタンを押す場合に使用する。

タップ操作での検出では指示された番号順にボタンにタッチしてもらう(図 3.1 左)。表示されている全てのボタンをタップすると異なる配置のボタンが表示され同じようにタップし続ける。これらの操作を複数回行ってもらい、データを取得する。

複数ボタンを配置することによって、画面上でどのようにボタンを配置することによりユーザーが不満を感じているかがわかると考えられる。

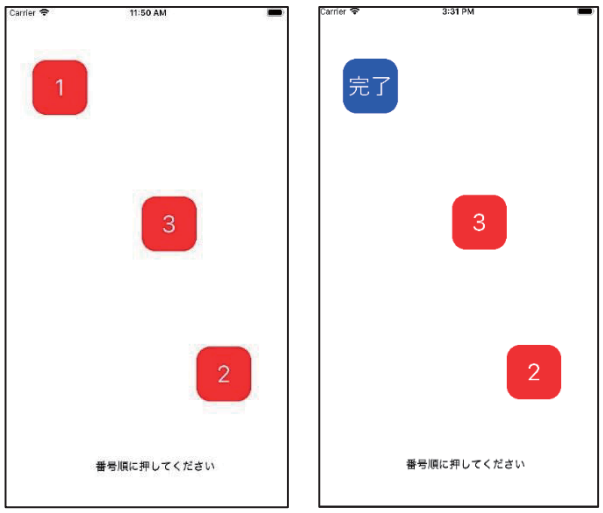


図 3.1 タップ操作での検出(左)　タップ後の画面(右)

3.2 スワイプ操作

スワイプとは、画面を指で押して一定方向へ指を動かす操作方法である。画面の切り替え時、ロック画面の解除などに使用する。

スワイプ操作では1から2へスワイプしてもら(図 3.2 左)う。指示された動作を終えるとタップ操作と同様、異なる配置で表示された1と2のボタンを表示する。指示が終わるまで、それらを繰り返す。

あらゆる場所から画面をスワイプすることによりどの場所からか、どちらの方向へ、どの場所への操作に不満を感じているかがわかると考えられる。

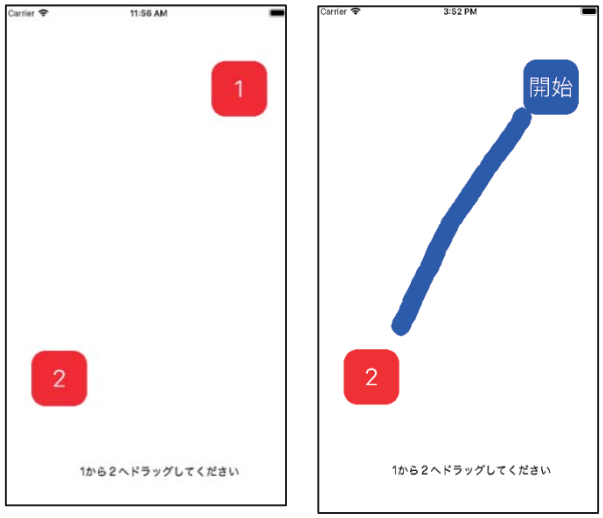


図 3.2 スワイプ操作での検出(左)　スワイプ中の画面(右)

3.3 フリック操作

フリックとは、画面を指で押してから上下左右に指を弾く操作方法。画面の切り替えや日本では文字の入力の際に使用する。

フリック操作では右、左、上、下のうちの一つが書かれたボタンが表示される(図 3.3 左)。指示された方向にフリックすると上記の動作と同じように異なる場所でボタンが表示され、ボタン上に書かれている方向にフリックする(図 3.3 右)。指示が終わるまで、それらを繰り返す。

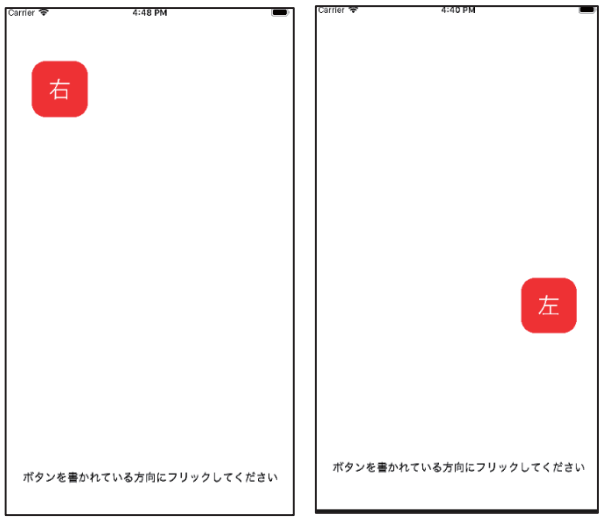


図 3.3 フリック操作での検出(左)　フリック後の画面(右)

4. 今後

まず、データを取得するための実験用アプリケーションを完成させ、また実際に多くの被験者にこのアプリケーションを使ってもらいデータを収集する必要がある。

データを集め、どのようなインターフェイスデザインの配置がアプリケーションを使うユーザーを不便にしているかがわかると思われる。得られたデータをもとに、インターフェイスの配置を行うことによりユーザーにとって使いやすいアプリケーションになると考えられる。

この方法を応用した場合どのような機能を頻繁に使うか、癖や手の大きさなど個々の使い方に合わせたインターフェイス(適応型インターフェイス)を行うことができより多くの人が使いやすく、わかりやすいアプリケーションに進化していく事が可能になると考えられる。これはスマートフォンのみならずパソコンなどの GUI ならば個々に合わせたインターフェイスの変更により、ユーザーのユーザビリティの向上につながると思われる。

参考文献

[山岡　12]　山岡俊樹:GUI デザインのための評価方法の提案, 日本デザイン学会, 2012

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

自動運転の駐車指示を対象にした空間的意味記述の生成への取り組み

Generating Spatial Semantic Descriptions of Parking Instructions for Self-driving Cars

稲子 明里 *1　塚原 裕史 *2　小林 一郎 *1
Akari Inago　Hiroshi Tsukahara　Ichiro Kobayashi

*1お茶の水女子大学　*2株式会社デンソーアイティーラボラトリ
Ochanomizu University　Denso IT Laboratory, Inc.

This paper describes a method to transform a combinatory categorial grammar(CCG) parse tree into a spatial description clause(SDC) to be grounded onto the objects or the phenomena in the real environment. We especially designed the CCG for handling spatial meanings contained in parking instructions for self-driving cars. In previous studies, we developed a parser for generating parse trees of parking instructions according to the CCG we defined. We propose a rule-based converter to perform a conversion from these parse trees into SDCs using heuristic knowledge gained from our analysis on a careful investigation on individual examples. We built the converter and evaluated it on the results of our parser. The result of conversion shows fairly high accuracy for parse trees generated by our parser from parking instructions in the corpus used in this study.

1. はじめに

近年, 車の自動運転の実用化に向けた動きが活発化している. 車を運転できない人でも自動運転車を操作するために, より操作を容易にする手法の開発が期待されている. なかでも, 口頭指示による対話的な操作は, 車への意思疎通を行う手段として今後必要になることが予想される. そこで我々は, 自然言語で表現された操作指示内容と実世界との対応づけ (グラウンディング) を目的とし, 駐車指示内容を車の操作へと結びつける空間意味記述の生成を目指す. 先行研究において, 我々は駐車指示を空間意味記述の前段階となる構文木へ変換する手法を提案し, そのパーザを実装した [Inago 17]. その後パーザの n-best 解のリランキングによって, 精度を向上する手法を提案し, リランカーを実装した [Inago 17-2]. これを踏まえて本研究では, 生成した構文木を空間意味記述に変換する手法を提案する.

2. 関連研究

ロボットへの自然言語の経路指示からロボットが行うべき行動を推定する研究は盛んに行われてきた. Kollar らは自然言語から実世界の空間意味関係を抽出するために Spatial Description Clause(SDC) という意味構造を提案した [Kollar 10]. SDC の構造は, Jackendoff (1983) [Jackendoff 83], Landau and Jackendoff (1993)[Landau 93], Talmy (2005)[Talmy 05] らによって提案された意味構造に基づいている. 統計的機械翻訳に始まり, 言葉と実世界を結ぶグラウンディングの問題を言葉の構文構造に従ったグラフィカルモデルとして扱う取り組みが増えている. Tellex ら [Tellex 11] は SDC の構造に従って生成される Generalized Grounding Graphs(G^3) と呼ばれる確率的グラフィカルモデルを提案した. 近年では, Paul ら [Paul 16] が抽象概念を階層構造に反映させた Adaptive Distributed Correspondence Graph というモデルを提案した. Arkin ら [Arkin 16] は複数の連なった指示を分割して, 個々の文に意味を割り当てる Monologic Distributed Correspondence Graph を提案した. 一方, 最近の自然言語を論理式に変換するための取り組みにおいては, Guu らが周辺尤度最大化を用いた系統的な探索に強化学習のランダムな探索を組み合わせた学習アルゴリズム (RANDOMER) を提案し, ニューラル意味解析器に適用した. この手法は, indirect supervision のみの学習が可能な場合において間違ったプログラムから偶然正しい結果が出力されるという事例の防止を課題として提案された.

連絡先: 稲子明里, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース, 〒 112-8610 東京都文京区大塚 2-1-1 小林研究室, g1320504@is.ocha.ac.jp

3. 提案手法

我々の手法は, Kollar ら [Kollar 10] と Tellex ら [Tellex 11] の研究に基づいている. 駐車指示に空間的な背景知識を組み込むための意味構造として, Kollar によって提案された SDC を用いた. SDC の生成に関して, 先行研究で Kollar らは条件付き確率場を用いてラベル付けをして直接変換していたが, この手法では文全体の要素の依存関係を捉えることができなかった. また Tellex らは Stanford parser[Marneffe 06] を用いて直接 SDC に変換していたが, 統語カテゴリを用いた構文解析を行っていたため, 空間意味構造への変換には, 制約を伴った入力文しか受け付けられず, ほぼテンプレートマッチングと同様の処理になってしまっていた. そこで我々は, Combinatory Categorial Grammar (CCG) [Steedman 00] において統語範疇のカテゴリを空間意味のカテゴリに変更した文法を新しく導入することでより複雑な入力文から正確な空間意味の依存木を抽出し, それを SDC に変換する手法を提案する. CCG を用いることで, 空間的な意味をラムダ式で表現することが可能になる. ラムダ式を用いることによって, 文章が空間的な意味を正しく捉えられているか否かに対する確認を行うことができる. 本研究では, Kollar らによって定義された SDC における 4 つのタイプに加え, STATE と VIEW の 2 つが追加されている [Inago 17]. これにより, SDC には EVENT, OBJECT, PLACE, PATH, STATE, VIEW の 6 つのタイプが存在し, それらに対応して E(EVENT), O(OBJECT), L(PLACE), P(PATH), S(STATE), V(VIEW) を CCG における基底カテゴリとして扱う. また指示内容と観測した物体とのグラウンディングには, Tellex らが提案した G^3 を拡張して用いる. 図 1 に提案手法の概要を示す. 本研究では, 点線で囲まれた言語,

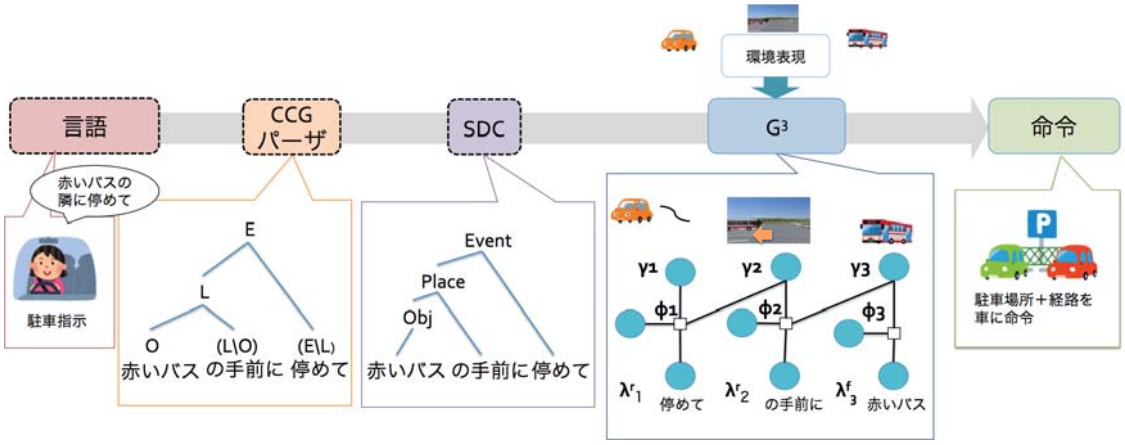


図 1: 提案手法の概要

CCG パーザ, SDC の部分に焦点を当て, 構文木から SDC への変換手法の提案を行う.

4. 構文木から空間意味記述への変換

4.1 変換規則

このセクションではパーザによって解析された木を SDC に変換する方法について説明する. 変換は深さ 1 の部分木ごとに行い, 日本語の駐車指示コーパス [Inago 17] を参照して得たヒューリスティックな知識に基づいた変換規則を用いる. この変換規則は条件に合致するかどうかに従って, 49 個存在する. 条件は大きく 6 個に分かれる. ノードがどのカテゴリをもつか, 子が葉かどうか, 全体の木の深さはいくらか, 親のカテゴリと右の子のカテゴリが同じか, 子がスペースを表す言葉（例えば「駐車スペース」,「場所」,「そこ」など）を持つどうか, 左の子が relation フィールドのテキストを持つかどうかがある. 構文木から SDC への変換は, 以下の 4 ステップに分けられる.

ステップ 1. 深さ 1 の部分木を選択する.

ステップ 2. 部分木の情報に従って, 適切な変換規則を選択する.

ステップ 3. 変換規則に従って, 部分木を SDC に変換する.

ステップ 4. 全ての部分木を変換するまでステップ 1 から 3 を繰り返す.

図 2 は深さ 1 の構文木 L (O (赤い車 the red car) (L\O) (の隣に next to)) を SDC PLACE(r=の隣に next to l=OBJ(f=赤い車 the red car)) に変換する例である. このプロセスは, ステップ 2 と 3 に相当する.

4.2 省略された SDC の補完

駐車指示には, グラウンディングに必要な情報が暗黙的に省略されており, そのまま SDC をへ変換した際に, グラウンディンググラフを生成するために必要な部分が抜けている場合が存在する. そこで, 我々は省略された SDC の部分を補完するための変換規則も定義した. 省略された SDC は,

1. 根の SDC が PLACE である場合
2. 根の SDC が STATE である場合

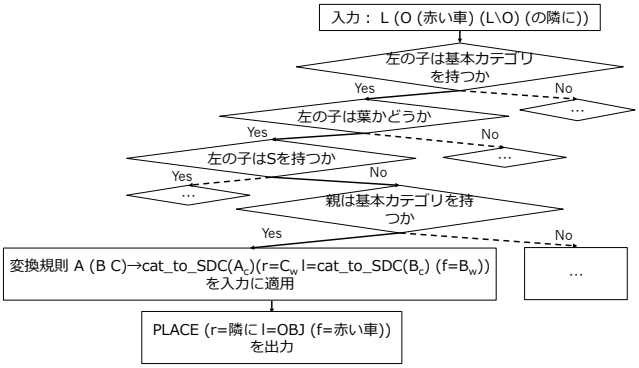


図 2: 深さ 1 の部分木から SDC への変換の例

3. PLACE が relation フィールドに属するテキストを持っているが, landmark フィールドの SDC は持っていない場合
4. PLACE が relation フィールドに属するテキストと condition フィールドに属する SDC を持っているが, landmark フィールドの SDC は持っていない場合

の 4 種類に分類される. 表 1 に省略された SDC の補完規則を示す. *は簡略化した任意の SDC またはテキストである. 式の左辺は省略がある SDC を表し, 右辺は補完後の SDC を表す. 図中では, A が親カテゴリを持つノードを表し, B と C が単語とカテゴリまたは SDC とカテゴリを持つ子ノードを表す. 変換規則の→の左辺が変換前の深さ 1 の部分木, 右辺が変換後の SDC を表す. N_w はノード N が持つ単語 w, N_c はノード N が持つカテゴリ c を表す. cat_to_SDC(C) は基底カテゴリ C から変換された SDC を表す.

5. 実験

先行研究で提案したパーザから生成された構文木から SDC への変換に対する評価実験について述べる.

5.1 実験設定

対象データには, クラウドソーシングによって集められたコンビニエンスストアの駐車場での駐車シーンにおける 6,019 文

表 1: 省略された SDC の補完のための変換規則

Type	Conversion rule	Example
1	PLACE(*) → EVENT(r=停めて l=PLACE(*))	PLACE(f=スペース r=前の l=OBJ(f=歩行者)) → EVENT(r=停めて l=PLACE(f=スペース r=前の l=OBJ(f=歩行者)))
2	STATE(*) → EVENT(r=停めて l=PLACE(f=スペース c=STATE(*)))	STATE(r=が空いてる l=PLACE(r=右端 l=OBJ(f=駐車場))) → EVENT(r=停めて l= PLACE(f=スペース c=STATE(r=が空いてる l=PLACE(r=右端 l=OBJ(f=駐車場)))))
3	(*PLACE(r=*)*) → (*PLACE(r=* l=PLACE(f=スペース c=STATE(r=空いてる)))*)	EVENT(r=停めて l=PLACE(r=真ん中に c=STATE(r=空いてる))) → EVENT(r=停めて l=PLACE(r=真ん中に l=PLACE(f=スペース c=STATE(r=空いてる))))
4	(*PLACE(r=* c=STATE(*))*) → (*PLACE(r=* l=PLAE(f=スペース space c=STATE(*)))*)	EVENT(r=停めて Park l=PLACE(r=右側 c=STATE(r=空いてる))) → EVENT(r=停めて l=PLACE(r=右側 l=PLCAE(f=スペース c=STATE(r=空いてる))))

表 2: 駐車指示コーパスにおける構文木のタイプと事例数

NO.	CCG ベースの文法による構文解析パターン	例文	事例数
1	(E (L (O (L\O)) (E\L)))	入り口の近くに停めて	586
2	(L (O (L\O)))	コンビニ前	562
3	(E (L (O (((O/O) (O (O/O)\O)) O) (L\O)) (E\L)))	黒い車の右の車の右側に停めて	451
4	(L (O (((O/O) (O (O/O)\O)) O) (L\O)))	コンビニの前の白い車の右隣	417
5	(E (L ((L/L) (O (L/L)\O) L) (E\L)))	黒い車から一番離れた場所に停めて	295
6	(L ((L/L) (O (L/L)\O) L))	トイレから反対側の右側	292
7	(E (L (L (O (L\O)) (L\L))) (E\L))	黒いクルマからずっと奥の一番右に停めて	137
8	(L (L (O (L\O)) (L\L)))	コンビニとは反対側の真ん中	131
9	(L ((L/L) (O (L/L)\O) L (O (L\O))))	白い車の横でコンビニ前	128
10	(L (O (S (L (O (L\O)) (S\L)) (O\S)) (L\O)))	歩行者の右側にある白いクルマのさらに右側	117
...	...	...	...
372	(L (S (L\S) ((L\S)/L L)))	空いてる右端のスペース	1
373	(L ((L/L) (S (L (O (L\O)) (S\L)) (L/L)\S) (L (S (L\S)))))	トイレの前ではなくて右側の空いてるスペース	1
374	(E (O ((O/O) (O (S (L (O (L\O)) (L\S)) (O\S)) (O/O)\O) O) (E\O)))	トイレの隣にあるコンビニの入口に停めて	1
375	(L ((L/L) (O (L/L)\O) (L (S (O (S\O)) (L\S)) ((L\L) (O (L\L)\O)))))	白いクルマの歩行者がいない方, コンビニ側	1
376	(L (S (O (S\O)) ((L\S) (V (O (V\O)) (L\S)\V))))	歩行者がいない白いクルマの向かって右側	1
377	(E (S (O (S\O)) ((E\S) (L (O ((O/O) (O (O/O)\O) O) (L\O)) (E\S)\L))))	歩行者は避けてコンビニの入口の前に停めて	1
378	(L (S (O ((O/O) (O ((O/O) O) (O/O)\O) O) (S\O))) (L\S))	一番右側のコンビニ入口の前の歩行者がいないところ	1
379	(E (L ((L/L) (S (L (S\L)) (L/L)\S) (L (S (L\S)))) (E\L)))	真ん中にある目の前の空きスペースに停めて	1
380	(S (L ((L/L) L) (S\L)))	真ん中の駐車場が空いています	1
381	(E (L (S (O ((O/O) (O ((O/O) (O (O/O)\O) O) (O/O)\O) O) (S\O)) (L\S)) (E\L)))	黒いクルマと白いクルマの間の歩行者がいるところに停めて	1

の日本語の駐車指示コーパスを用いた [Inago 17]. クラウドソーシングのワーカー数は 168 で, 自車から見える駐車シーンの画像とその画像上に示された駐車位置のマークを提示し, ランドマークとして, コンビニ, コンビニの入り口, コンビニに隣接する屋外トイレ, 駐車している白色と黒色の自動車, 歩行者 1 人 (2 台の自動車の間) を利用して, 2 台の車の右, それらの間にある 2 つの駐車スペースについて, それぞれ以下の条件の時における, 駐車指示を入力するように依頼した.

1. 自車と駐車位置との関係のみを利用
2. 各ランドマーク 1 つと駐車位置との関係を利用
3. 2 つのランドマークと駐車位置との関係を利用（3 種類の異なる組合せを選択）

このようにして収集された駐車指示文には, 人手で CCG の構文木と SDC がアノテーション付けられており, 構文木の種類は全部で 381 種類存在する (表 2) . コーパスのアノテーションは一人で行い, もう一人が内容確認を行った. これらの構文木中で出現する意味カテゴリの数は, 45 となった (表 3). また, 駐車指示の文の長さの平均は 9.54, 分散は 3.09 となった. 駐車シーンの中の物体の数が少ないため, 形態素の語彙数は 348 と少なくなっているが, 空間的な意味を表す CCG Bank の語彙数は形態素とは必ずしも一致せず, それらの組み合わせから成り, 文脈に依存して幾つかの空間意味カテゴリを表しうるため, 全部で 1262 となった.

5.2 実験結果と考察

まず, コーパスの駐車指示を [Inago 17-2] のパーザを用いて生成した構文木について, SDC への変換を行う実験結果について述べる. 表 4 はパーザから出力された構文木から SDC への変換精度を表している. 「全ての木」は, パーザから出力された全ての構文木を SDC に変換した場合の精度であり, 「正

表 3: 駐車指示コーパスの構文木アノテーションに含まれる意味カテゴリ

カテゴリ
S, (S/S), (L\S)\O, (L/L)\S, (S\L), (L\S)\S, (L\L)\S, (S/S)\O, V, (L\S)\V, (L\V)\O, (V\O), (V\L), (L\V), (L\V)/L, (L\S), (L\S)/L, (S\L)\L, (O/O)\S, (E\S)\L, (O\S), (S\O), (E\S), (O\V), (L\V)\L, (O/O)\V, (L/L)\V, (E\L)/S, (E/S)\S, (S/E), (L\L)\O, (S\L)\O

解木」はパーザから出力された構文木の中で正解だった構文木のみを SDC に変換した場合の精度である．精度においては，SDC の構造が正解データと完全に一致していたら正しいとみなす．括弧内は実験データの数である．パーザから出力された構文木が正しい場合には，SDC に正しく変換できたことを示している．運転指示文から直接的に複雑な分岐構造を持つ SDC へ変換するのではなく，その中間表現として，シンプルな分岐から成る CCG の構文木を得ることの有効性を示していると考えられる．今後は駐車指示データの拡充し，より多様な表現への対応についても評価を行っていきたい．

表 4: 構文木から SDC への変換精度

全ての木	正解木
0.847(6019)	1.0(5098)

6. まとめと今後の課題

本研究では，自動運転車に向けた駐車指示から空間意味記述を生成することを目的とし，構文木から SDC への変換手法についての提案と，実装した変換器の日本語の駐車指示のコーパスによる評価を行った．ヒューリスティックな知識に基づき定義した変換規則を用いることにより，今回使用した実験データについては，パーザから出力された全ての構文木が正しい場合には，正しい SDC に変換することができた．これは，中間表現として CCG の構文木を得ることの有効性を示していると考えられる．高い精度で運転指示を SDC に変換できたのは，固定された駐車シーンに含まれるランドマークの数が少なく，表現が多様でなかったためだと考えられる．そこで今後は，多様な駐車シーンでかつ様々なランドマークが存在する状況を含めて，より多くの駐車指示文に柔軟に対応するために，言語データの拡充や未知語処理の実装を目指し，駐車指示以外への運転指示への拡張を考えていきたい．さらにはラムダ式の導入により SDC に背景知識の組み込みを行う予定である．

参考文献

[Inago 17] 稲子明里, 塚原裕史, 小林一郎,「自動運転の言葉による指示を対象にした空間的意味表現の構造化への取り組み」, 2B3-OS-07a-5in2, 愛知, 2017 年度 人工知能学会全国大会 (第 31 回), 2017.

[Inago 17-2] 稲子明里, 塚原裕史, 小林一郎,「自動運転のための駐車指示から空間意味記述表現への変換パーザの精度向上に関する取り組み」, P1-11, 岡山, 言語処理学会第 24 回年次大会, 2018.

[Kollar 10] T.Kollar, S.Tellex, D.Roy, N.Roy: Toward Understanding Natural Language Directions, ACM/IEEE Int ’l Conf. on Human-Robot Interaction (HRI), pp.259-266, 2010.

[Steedman 00] T.Kollar, M.Steedman: The Syntactic Process, MIT Press, 2000.

[Marneffe 06] M.Marneffe, B.MacCartney, C.D.Manning: Generating Typed Dependency Parses from Phrase Structure Parses, Proceedings of the International conference on Language Resources and Evaluation (LREC), 2006.

[Jackendoff 83] R.Jackendoff: Semantics and Cognition, MIT Press, 1983.

[Talmy 05] L.Talmy: The fundamental system of spatial schemas in language, Hamp, B., ed., From Perception to Meaning: Image Schemas in Cognitive Linguistics, 2005.

[Landau 93] B.Landau and R.Jackendoff: “What” and “where” in spatial language and spatial cognition, Behavioral and Brain Sciences 16, pp.217-265, 1993.

[Tellex 11] S.Tellex, T.Kollar, S.Dickerson, M.R.Walter, A.G.Banerjee, S.Teller, N.Roy: Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation, AAAI, 2011.

[Paul 16] Rohan Paul, Jacob Arkin, Nicholas Roy, and Thomas M. Howard: Efficient Grounding of Abstract Spatial Concepts for Natural Language Interaction with Robot Manipulators, RSS, 2016.

[Arkin 16] Jacob Arkin, Matthew R. Walter, Adrian Boteanu, Michael E. Napoli, Harel Biggie, Hadas Kress-Gazit, and Thomas M. Howard: Contextual Awareness: Understanding Monologic Natural Language Instructions for Autonomous Robots, IEEE, 2017.

[Guu 17] Kelvin Guu, Panupong Pasupat, Evan Liu, Percy Liang : From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood, ACL, 2017.

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

2次元コード用電子透かしと符号化開口の共進化的同時設計に関する基礎検討

A Preliminary Study on Coevolutionary Simultaneous Design of Digital Watermark for Two-dimensional Codes and Coded Aperture

竹下 真悟 *1 濱崎 弘樹 *1 前原 武 *1 中居 謙太郎 *1 園田 聡葵 *2 三鴨 道弘 *1
Shingo Takeshita Hiroki Hamasaki Takeru Maehara Kentaro Nakai Toshiki Sonoda Michihiro Mikamo
川崎 洋 *2 長原 一 *3 小野 智司 *1
Hiroshi Kawasaki Hajime Nagahara Satoshi Ono

*1鹿児島大学大学院 理工学研究科 情報生体システム工学専攻
Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University
*2九州大学大学院 システム情報科学研究院 情報知能工学専攻
Department of Advanced Information Technology, Graduate School of Information Science and Electrical Engineering, Kyushu University
*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 recent years, authentication using two-dimensional codes has become widespread. However, it is easy to duplicate two-dimensional code, and inappropriate duplication of 2D codes is apprehended. Therefore, a method of detecting duplication by embedding a digital watermark in a two-dimensional code has been proposed. However, when the watermark is embedded in high-frequency components of a cover image, the watermark components may be lost due to photographing. A coded aperture is a technique that maintains frequency components included in the captured image. In this research, we propose a simultaneous design method of a digital watermark and a coded aperture, which allows extracting watermark with wide photographing depth range and rotation. Experimental results show that the proposed aperture could extract watermarks from blurred images regardless of with or without deconvolution.

1. はじめに

近年，航空機の搭乗券やオンライン決済の認証手段として2次元コードが普及している．特に，2次元コードをスマートフォンの画面に表示して認証を行う手法は，利便性と共にペーパレス化が図れるため一般的になっている．しかしながら，2次元コードは悪意ある複製や偽造によって，容易に複製が可能という問題がある．図1のようなスマートフォンのカメラを用いた撮影を用いた撮影でも複製が可能である．このような複製は2次元コードを用いた決済が普及するなかで，より深刻な問題を引き起こすことが懸念される．本研究では図1のように，スマートフォンの画面に表示した2次元コードをほかのスマートフォンのカメラで撮影する行為を不正な複製とする．

一方，電子透かしは著作権の保護，画像の修正や変更を検出するために広く用いられている．電子透かしは頑健，脆弱，半脆弱の大きく3つに分類することができ，半脆弱な電子透かしは特定の処理を受けた場合にのみ破壊される特徴がある [Rey 02]．2次元コードに半脆弱な電子透かしを埋め込むことで，上記のような複製を検知する手法が提案されている [前原 15]．正規の2次元コードからは透かしを抽出できるが，複製が行われた2次元コードでは透かしが破壊されるため透かしが抽出されず，真贋判定が可能である．上記方式の電子透かしは離散ウェーブレット変換（DWT）[Vahedi 12] を用いて画像の高周波成分に透かしを埋め込む方式である．しかし，原本の電子透かしの抽出の撮影において焦点ボケが発生すると画像の高周波成分を失うこととなり安定的に抽出を行えない問題がある．

連絡先: 竹下真悟，鹿児島大学大学院 理工学研究科 情報生体システム工学専攻，〒 890-0065 鹿児島市郡元 1-21-40，sc113035@ibe.kagoshima-u.ac.jp

図 1: スマートフォンのカメラを用いた不正な複製

また近年，画像処理によって画像を生成することを前提とした撮影手法のコンピュテーショナルフォトグラフィ技術が注目されている．その代表的な技術の一つである符号化開口は，カメラの絞りの形状を2次元のマスクパターンに変更する技術で，撮影画像に含まれる周波数成分を制御する特徴を持つ．

本研究では，抽出における問題の解決策として抽出用カメラに符号化開口を用いる．先行研究では特定の電子透かしに対して有効な符号化開口の進化的設計を単一進化の条件で行った [濱崎 17]．しかし，この手法では解析を容易にするために単純な透かしを用いていた．複数の条件で安定的な電子透かしの抽出を行うための最適な電子透かしと符号化開口の組み合わせは未知である．このため，本研究では電子透かしの設計と符号化開口の設計を進化型最適化を用いて同時に行う方式を提案する．電子透かしと符号化開口を同時に設計することで，広い撮影距離や撮影角度で透かしの抽出が行えることが期待される．実験により，従来のシステムに比べ安定的な透かしの抽出を行えるような電子透かし及び符号開口の設計が行われたことを示す．

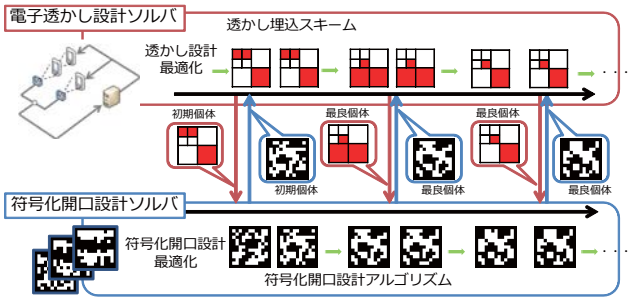


図 2: 提案する共進化的同時設計方式

2. 関連研究

電子透かしはデジタルデータに対し秘匿情報を埋め込む技術であり，著作権保護，画像の修正や改ざん検出などに広く利用されている．Vahedi らはカラー画像を対象とし遺伝的アルゴリズム（GA）を用いて，DWT を用いて得られた周波数帯への埋め込む強度を自動的に設計する方式を提案した [Vahedi 12]．Mingzhi らは，遺伝的アルゴリズムを用いて，DWT と離散コサイン変換により得られる周波数帯に対し電子透かしを埋め込む際の強度を自動的に決定する方式を提案している [Mingzhi 13]．

符号化開口はカメラの絞りを 2 次元のマスクパターンで符号化する技術であり，ボケ除去，超解像，全焦点画像生成等に利用されている．Zhou らはボケ除去用の符号化開口を GA を用いて設計する方式を提案した [Zhou 09]．符号化開口を用いた手法は逆畳み込み演算を後処理として含んでいる．Zhou らの手法は，深さごとに推定されたカーネルでの 1 回の逆畳み込み演算で動作する．しかしこれらの方法は不安定であり，人手による援助が必要となることがある [Zhou 09]．

3. 提案手法

本研究は電子透かしと符号化開口の同時設計手法を提案する．本章では提案する方式の詳細について述べる．

3.1 協調型共進化アルゴリズム

本方式は電子透かし設計問題および符号化開口設計問題の同時最適化を実現するために協調型共進化アルゴリズムを用いる．図 2 に本方式の模式図を示す．本方式では各ソルバで設計された部分解をそれぞれのソルバに伝達し，伝達された解を用いて最適化を行う．すなわち，電子透かし設計ソルバは，符号化開口設計ソルバにより設計された符号化開口を用いて電子透かしの評価を行い，符号化開口設計ソルバでは，電子透かし設計ソルバにより設計された電子透かしを用いて符号化開口の評価を行う．

3.2 電子透かしの設計

提案手法の電子透かしと符号化開口の設計の処理手順を図 3 に示す．提案手法は GA に基づいており，解候補を評価する際はシミュレーションにより透かしの抽出性能を評価する．シミュレーションに撮影を再現し，解候補の評価値に基づく新たな解候補の生成とを繰り返すことで最適化を行う．GA における解候補の表現および目的関数については 3.2 節に，シミュレーションの詳細については 3.4 節に示す．

提案方式の電子透かしの設計は，[前原 15] の方式を参考に透かしを埋め込む領域ごとの周波数帯の選択と，埋め込み強度の調整とを同時に実行する．図 4 に透かしを埋め込む対象となる 2 次元コードの例を示す．提案方式は周波数解析方式とし

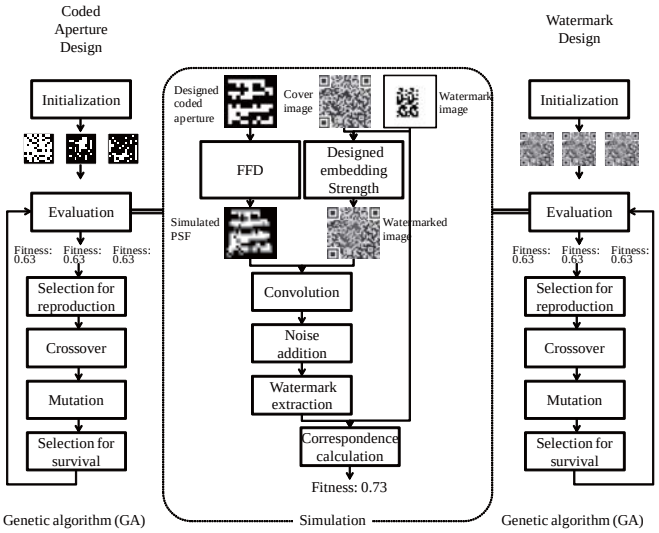


図 3: 提案する方式の処理手順

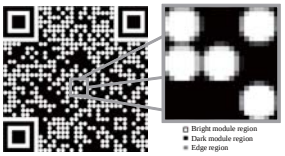


図 4: 2 次元コードの領域

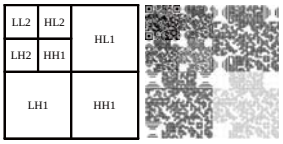


図 5: DWT で得る周波数帯

て Haar 関数を基底とする 2 段階の DWT を用いる．また，カバー画像となる 2 次元コードの輝度値に応じ，暗モジュール領域 D，明モジュール領域 B，および，エッジ領域 E の 3 つの領域に分割し，領域毎に異なる周波数帯に透かしを埋め込むこととする．図 5 にカバー画像に対して 2 段階の DWT を適用することで得られる 7 つの周波数帯 {HH1, HL1, LH1, HH2, HL2, LH2, LL2} を示す．提案方式では得られた周波数帯の埋め込み強度を，カバー画像の領域毎に決定する．設計変数の個数は画像領域と周波数帯の組み合わせの合計 21 個となる．

画像領域 r，周波数帯 b に対応する設計変数は $v_{r,b}$ と表し，0 から 1 までの実数値をとる．解候補 $\boldsymbol{v}$ は，$v_{r,b}$ を各要素として持つ 21 次元のベクトルとして表現される．提案方式では $v_{r,b}$ の値が 0.5 を上回る場合にのみ，該当領域に強度 $L_{r,b}$ で透かしの埋め込みを行う．

$$L_{r,b} = \begin{cases} 2 \times (v_{r,b} - 0.5) \times L_{max} & \text{if } v_{r,b} > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$v_{r,b}$ が 0.5 を下回る場合は透かしを埋め込まない．また，L_{max} は透かしを埋め込む際の最大強度を示している．透かしの埋め込みの有無を判断する閾値を 0.5 とする理由は，設計変数が初期集団生成や突然変異などでランダムな値を取った場合に，透かし埋め込みの有無を 50 %の確率で決定するためである．

本研究の目的関数は，撮影画像からの透かしの抽出率とする．透かしの抽出率を画素単位の一致率（Bit Correct Ratio: BCR）により表現し，目的関数を以下のように定義する．

$$f(I) = \sum_{d \in S_d} \sum_{a \in S_a} \mathrm{BCR}(W, W_{d,a}^{\mathrm{cap}}) \quad (2)$$

ここで，I は解候補，W は透かし画像の原本，$W_{d,a}^{cap}$ は a 度回転された透かし入り 2 次元コードを距離 d で撮影した画像

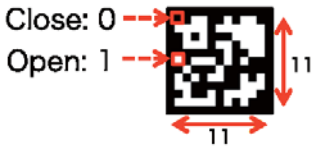


図 6: 符号化開口設計における設計変数

(a) カバー 2 次元コード　(b) 透かし画像

図 7: 実験に用いた 2 次元コードと透かし画像

(a) 設計された透かし　(b) 設計された符号化開口

図 8: 提案手法で設計した解

(a) Prop　(b) Circ　(c) Zhou

図 9: 提案手法と比較対象の符号化開口.

$Y_{d,a}^{cap}$ から抽出した透かし画像である．BCR は下記の式で算出する．

$$\mathrm{BCR}(W, W') = 1 - \frac{\sum_{w=1}^{w_W} \sum_{h=1}^{h_W} \left(W_{w,h} \oplus W'_{w,h}\right)}{w_W h_W} \quad (3)$$

ここで，$W_{w,h}$，$W'_{w,h}$ は，埋め込んだ透かし画像および抽出した透かし画像の (w, h) 座標に位置する画素の値である．w_W，h_W は透かし画像の幅および高さ（画素数）を表し，$\oplus$ は排他的論理和である．

3.3 符号化開口の設計

符号化開口の設計は [濱崎 17] を参考に行う．開口形状を 11×11 の 2 次元パターンとする．図 6 に示すように光の遮断，透過をそれぞれ 0，1 で表すことで，符号化開口の設計を 11×11 次元の組み合わせ最適化問題として表現する．符号化開口設計の目的関数は 3.2 節と同様である．

3.4 シミュレーション

一般的に，撮影画像 Y' は以下の式によって表現される．

$$Y' = Y \otimes PSF + N \quad (4)$$

ここで，Y，Y' は原画像および疑似撮影画像を示し，$\otimes$ は畳み込み積分を示す．N はノイズを表し，本研究では輝度値のガウシアンノイズを付与する．PSF はボケカーネルを表す．理想的な環境下であれば符号化絞りの形状自体を撮影画像に生じる PSF として用いることができるが，実際には光学系由来の歪みなどによって完全に対応しない．このため本論文では，シミュレーションのモデルを以下のように定義する．

$$Y' = D(PSF) \otimes Y + N \quad (5)$$

ここで $D(\cdot)$ は光学系によって生じる，符号化絞り形状と PSF の非線形な応答を表し，本研究では Free Form Deformation（FFD）[Sederberg 86] によって表現する．本研究では，事前に校正を行うことで $D(\cdot)$ を得るものとし，能動絞りカメラに 3×3 の格子状のパターンを投影して，符号化開口の歪み量を取得する．

4. 評価実験

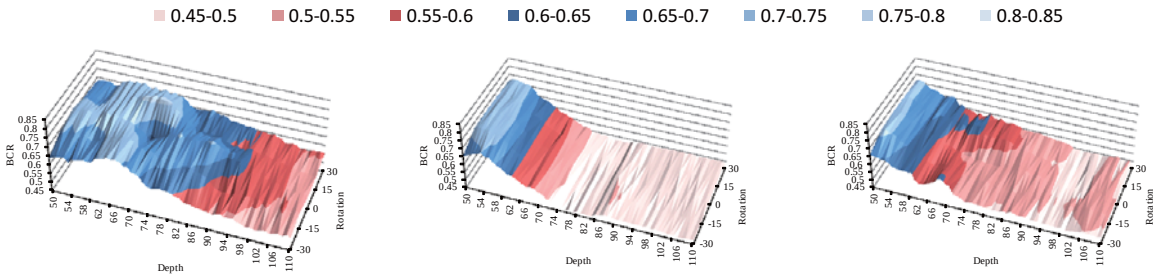
4.1 透かしの埋め込み強度と符号化開口の同時設計

提案手法を用いて電子透かしと電子透かし用符号化開口の形状の設計を試みた．本研究では，複数の深度（S_d）および回転角度（S_a）の組み合わせの条件下で撮影を行い BCR の平均値を最大化することで，撮影時の距離と角度の変化に頑健な透かしと符号化開口の設計を試みる．今回の実験で（S_d,S_a）の評価の対象は $\{(70, -30), (80, -30), (90, -30), (80, 0), (70, 30), (80, 30), (90, 30)\}$．の合計 7 組とした．図 7 に実験に用いたカバー 2 次元コードおよび透かし画像を示す．最適化に用いた GA はそれぞれ [前原 15, 濱崎 17] を参考とし，透かし設計では，交叉および突然変異はそれぞれ blend crossover（BLX-α）[Eshelman 93] および Uniform Mutation を用いた．個体数，世代数の上限，交叉率，突然変異率はそれぞれ，100，600，0.5，0.05 とした．符号化開口設計では，交叉および突然変異はそれぞれ 2 次元の 2 点交叉 [Kawasaki 13] およびビット反転を用いた．個体数，世代数の上限，交叉率，突然変異率はそれぞれ，100，600，0.97，0.01 とした．シミュレーションでは，ノイズレベル σ と透かし強度 α をそれぞれ 0.0002 および 192 に設定した．変形モデル $D(\cdot)$ は深度 80mm において生成することとした．

図 8 に設計された解を示す．図 8(a) に示す透かしは，カバー画像の 3 つの領域すべてで HH1 成分と HH2 成分に透かしが埋め込まれた．符号化開口は図 8(b) に示す形状が設計され，埋め込まれた透かしの成分に似た形状をしていることが確認できる．

4.2 設計された解の実機を用いた事後評価

提案手法で得られた開口形状（Prop）についての事後評価を行い，一般的な円形開口 (Circ)，ボケ除去のための符号化開口 (Zhou)[Zhou 09] と比較した．比較を行った開口の形状を図 9 に示す．このとき，Circ と Zhou に関しては，ブラインド逆畳み込み [Richardson 72] を行った後に透かし抽出を行う場合 ($\mathrm{Circ_d}$ と $\mathrm{Zhou_d}$) と比較した．上記の事後評価においては，符号化開口を用いた撮影を能動絞りカメラ [Nagahara 10] を用いて行った．能動絞りカメラは絞りの交換に一切の物理的な操作が必要なく，画像をカメラにアップロードするだけの電子的な操作のみで行えるため，効率的な実機環境実験が可能となる．異なる絞り形状を用いて撮影を行う場合であっても，使用する絞り形状以外の条件を全く同じに揃えることができる．本論文で使用する能動絞りカメラは，反射型液晶ディスプレイ素子（Liquid Crystal on Silicon: LCoS）を絞りとして用いる．透過型液晶ディスプレイ (Liquid Crystal Display: LCD) を用いる方式と比較すると，LCD では透過する液晶素子間に



(a) Prop (b) $\mathrm{Circ_d}$ (c) $\mathrm{Zhou_d}$

図 10: 提案手法と比較対象の事後評価の結果.

駆動回路があるために生じてしまう光の回折の影響を緩和できる利点がある.

また，撮影対象となる 2 次元コードは偏向による影響が生じないようにスマートフォンのスクリーンに表示し，電動ステージを使用して撮影距離を調整した．合焦距離を 50mm とし，カメラのレンズから 50mm と 110mm の間の距離を 1mm の間隔で移動させながら撮影を行い，各開口による透かしの抽出率を評価した．この比較では，最適化で使用されたものとは異なる 3 つの透かし画像を使用し，平均 BCR を求めた．

図 10 に事後評価の結果を示す．提案された透かしと符号化開口の組み合わせでは，BCR が 0.6 以上の透かしを ±30 度の回転内で 85mm の深さまで抽出することができる一方で，$\mathrm{Zhou_d}$ および $\mathrm{Circ_d}$ の開口ではそれぞれ 62mm および 61mm までであった．これは被写界深度を 11 から 12mm だったものを 35mm まで拡大することに成功したことを示している．また，提案手法の組み合わせは後処理デコンボリューションを必要としないため，200.4 m s 程度で透かしの抽出を行うことができ，これはデコンボリューションを用いる手法に比べて約 30 倍速く真贋判定を実行できる利点がある.

5. まとめ

本研究では，再撮，ボケ，スケールの 3 つの要素の変化に対して頑健な電子透かしを実現することを目指し，電子透かしの埋め込み方法と透かし抽出用符号化開口の共進化による同時設計方式を提案した．実験により，通常の開口やボケ除去用の符号化開口では焦点ボケにより透かし情報が消失してしまうような状況であっても，提案方式により作成した透かしと符号化開口を用いることで焦点ボケを含む撮影画像から透かしを抽出できることを確認した．今後，協調型共進化の解の交換方法や評価方法を変更することで，より安定的な真贋判定に寄与する透かしと符号化開口の設計を試みる.

謝辞

本研究の一部は，科学研究費補助金（課題番号 : 15H02758）の支援を受けて実施した．ここに記して感謝の意を表する.

参考文献

[Eshelman 93] Eshelman, L. J.: Chapter real-coded genetic algorithms and interval-schemata, Foundations of genetic algorithms, Vol. 2, pp. 187–202 (1993)

[Kawasaki 13] Kawasaki, H., Horita, Y., Masuyama, H., Ono, S., Kimura, M., and Takane, Y.: Optimized aperture for estimating depth from projector's defocus, in Int'l Conf 3D Vision-3DV, pp. 135–142 (2013)

[Mingzhi 13] Mingzhi, C., Yan, L., Yajian, Z., and Min, L.: A combined dwt and dct watermarking scheme optimized using genetic algorithm, J.multimedia, Vol. 8, No. 3, pp. 299–305 (2013)

[Nagahara 10] Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H., and Nayar, S. K.: Programmable aperture camera using LCoS, in Computer Vision–ECCV 2010, pp. 337–350, (2010)

[Rey 02] Rey, C. and Dugelay, J.-L.: A survey of watermarking algorithms for image authentication, EURASIP J. on Advances in Signal Processing, Vol. 2002, No. 6, p. 218932 (2002)

[Richardson 72] Richardson, W. H.: Bayesian-based iterative method of image restoration, JOSA, Vol. 62, No. 1, pp. 55–59 (1972)

[Sederberg 86] Sederberg, T. W. and Parry, S. R.: Free-form Deformation of Solid Geometric Models, SIGGRAPH Comput. Graph., Vol. 20, No. 4, pp. 151–160 (1986)

[Vahedi 12] Vahedi, E., Zoroofi, R., and Shiva, M.: Toward a new wavelet-based watermarking approach for color images using bio-inspired optimization principles, Digital Signal Processing, Vol. 22, No. 1, pp. 153–162 (2012)

[Zhou 09] Zhou, C. and Nayar, S.: What are good apertures for defocus deblurring?, in Int'l Conf Computational Photography (ICCP), pp. 1–8 (2009)

[前原 15] 前原武, 中居謙太郎, 池田亮, 谷口康太郎, 小野智司 : 進化型多目的最適化を用いた 2 次元コード真贋判定用電子透かしの設計, 信学論 D, Vol. 98, No. 5, pp. 835–846 (2015)

[濱崎 17] 濱崎弘樹, 中居謙太郎, 竹下真悟, 前原武, 園田聡葵, 三鴨道弘, 川崎洋, 長原一, 小野智司 : 電子透かしの抽出を目的とした符号化開口の設計に関する実験的検討, 信学技報 (マルチメディア情報ハイディング・エンリッチメント), Vol. 117, No. 282, pp. 1–6 (2017)

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

読み曖昧性解消のためのデータセット構築手法

Dataset Construction Method for Word Reading Disambiguation

西山 浩気[*1] Koki Nishiyama
山本 和英[*1] Kazuhide Yamamoto
中嶋 秀治[*2] Hideharu Nakajima

[*1] 長岡技術科学大学 Nagaoka University of Technology
[*2] NTT メディアインテリジェンス研究所 NTT Media Intelligence Laboratories

We propose a data construction method for word reading disambiguation. The method gives unique reading word to each reading of reading ambiguous word, collects sentences including the unique word, replaces the unique word in sentences to the original ambiguous word and tags readings of reading ambiguous words to the reading corresponding to the unique word. Through experiments, we confirmed the method collects data numerically balanced between readings.

1. はじめに

音声合成において，文章の区切りや読み・アクセントを決める言語解析は重要な技術である．単語の連接情報に基づいた形態素解析処理の結果によって，あるいは，単語の連接に関する統計情報に基づいて，単語に正しい読み・アクセントを付与することが概ね可能である [山森 99, 長野 06, 森 10 など]．しかし，それらの単語の連接情報は当該の単語の前後の数単語の情報を利用する程度に留まっており，誤る場合が有る．その誤りの中でも，音声合成の目的において目立つ誤りとして，同形異音異義語間での読み付与の誤りがある．例えば，“方”という単語には，選択肢の一つや方角を意味する“ホウ”という読みと人を意味する“カタ”という読みの曖昧性がある．このような読み曖昧性の解消が必要である．

このような曖昧性解消の研究には，できるだけ大規模なコーパスが必要である．一般に入手可能な現代日本語書き言葉均衡コーパス(BCCWJ)では，人手によって約 2,000 文への読み付与が行われているが，我々が必要とする同形異音異義語間での読み多義解消用のデータとして用いるには規模が十分ではない．意味の多義性解消における誤りの約 7 割が訓練データ不足による誤りであるとの報告[新納 15]から考えると，既存のコーパスのみでは十分な精度で多義性解消することは期待できず，読み・アクセント情報が付与されたデータセットを構築する必要がある．

多義語の語義毎の例文収集を目的として，言い換えを利用したデータ拡充手法が報告されている[戸田 16]．この手法に倣って，本研究では読み曖昧性解消用のデータセット構築を行なう．同形異音異義語は同じ字面でありながら，意味によって発音が異なる単語である(以後，多音語と呼ぶことにする)．多音語の読みごとのデータセットを構築するため，読みを 1 つしか持たない同義語(以後，一音語と呼ぶことにする)に置き換え，例文を収集することで，読みの定まった例文収集を行なう．同義語は文脈が似ていると期待できるから，一音語へ置き換え検索収集する．本手法では読みと語義ごとにデータの収集が可能であるため，多音語を検索キーとして例文収集を行い，後から読みを対応付ける手法に比べ，それぞれの読みに対応する例文を効率よく収集する手法であることを示す．

西山浩気 長岡技術科学大学 自然言語処理研究室
Email: nishiyama@jnlp.org

2. 提案手法

提案手法の概略を図 1 に示す．初めに，語義と読みの両方が異なる異音異義語，すなわち，多音語をそれぞれの同義語で読みが一通りである一音語へ置換する．置換した一音語を検索語として，クローリングや大規模コーパスから，その検索語が含まれる文を収集する．この検索語を元の多音語に直し，多音語の複数の読みの中で前記の一音語に対応する読みを収集した文内での多音語の読みとして対応付ける．最後に，収集した文に付与された読みに対して人手による評価を行う．

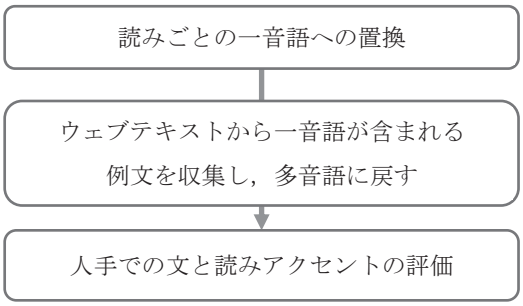


図 1 提案手法

2.1 読みごとの一音語への変換

多音語を読み毎に一音語へ置き換える．各単語の日本語 WordNet [Isahara 08]のエントリから各語義と等価な単語であり，かつ，読みの曖昧性の無い語を選定する．WordNet に検索語が存在しない語，あるいは，検索結果に適切な単語や複合語がない場合は人手で選定する．例として，“今日(コンニチ/キョー)”の読み毎に一音語へ選定する例を示す． 初めに，日本語 WordNet で“今日”を検索すると以下のような結果が得られる．

- WordNet における”今日”の検索結果の例
  today ：{方今，輓近，頃来，近頃，近ごろ，今，今日，今時，今日このごろ，最近，当節，現代，現在，今節，今どき，今日此の頃，今日此頃}

この検索結果から語義をひとつしか持たない単語をひとつ選択する．表 1 に“今日”に対して選定した一音語を示す．

表 1 一音語への変換の例「今日」

読み	一音語
コンニチ	最近
キョー	本日

2.2 例文収集

一音語を検索語として文の検索を行う. そのあと, 収集した文に含まれる一音語を元の多音語の表記に置きなおすことで, 読みや意味毎に対応づいた例文とする.

「今日」を例として読みごとの例文収集を説明する. 2.1 節において「今日」の読み「コンニチ」「キョー」に対して一音語「最近」「本日」への置き換えをそれぞれ行った. 置き換えた一音語を検索語として Web からクローリングを行なうなどして文を収集できる. 次に, 収集した例文の一音語を元の多音語に置き換えることで, 読みと対応づいた例文を収集する. 一音語で Web テキストを検索した結果と, その置換後の文について例を示す.

- ”最近”が含まれる例文(コンニチに対応)

  検索結果:最近は寒の戻りの特異日ですが,
  春の陽光たっぷりの一日になりそうです・・・

  置換後 :今日は寒の戻りの特異日ですが,
  春の陽光たっぷりの一日になりそうです・・・

- ”本日”が含まれる例文(キョウに対応)

  検索結果:私も本日気になったので本屋を覗いて見ても
  それらしい本はありませんでした.

  置換後:私も今日気になったので本屋を覗いて見ても
  それらしい本はありませんでした.

2.3 文と読みの妥当性に関する人手評価

前ステップで収集した文の一音語を多音語に置き換えた. このことによって文として成り立たなくなってはいないか, 読みが正しいか否かを人手で評価する. 評価は以下の 2 段階に分かれる.

評価 1. 文中の多音語の用法が正しいか否かの評価

評価 2. 用法が正しい場合は, 文中の多音語の読みが正しいか否かを評価し, 読みが誤っている場合は正しい読みを付与する.

多音語の複数の読みのうち取りうる読みをあらかじめ収集しておき, 作業者へ提示する. 例えば, “最近”を検索語として例文を収集した. “最近”を元の“今日(コンニチ)”に置き換えた文を挙げて以下に示す.

- 用例が“正しい”, 読みが「コンニチ」の例

  ちなみに今日の研究結果からは 3000 年前には既に弥生時代に入っていた、て結果だったよね.

上記の例においては文中での“今日”の用法は“正しい”, 読みは“コンニチ”とラベル付けする.

3. 多音語の置き換えによるデータセットの構築実験

提案手法を用いて同形異音多音語の一音語への置換を行い, 語義ごとの文収集を行ない, 人手による評価を行った. 本手法を用いることで読みごとにほぼ均一に例文を収集することができ, 多音語の訓練データ不足を補う手法として有効であることを示す.

3.1 実験設定

今回の実験では, 形態素辞書 UniDic の中の多音語を対象とした. 同一の表記と品詞を持つ形態素のうち, 読みの異なる形態素 522 組 2,128 語を抽出した. これらの単語を以下のように分類し, 分類名 1) の異義語に含まれる 43 組 88 語の多音語を対象とした.

表 2 辞書から抽出した語の分類

分類名	事例
1) 異義語	今日:キョー/コンニチ
2) 清濁の違い	会社:カイシャ/ガイシャ
3) 音訓の違い	縁:エン/フチ
4) 同義語	下腹:シタハラ/カフク
5) 音の異なり	行く:イク/ユク

読みごとに置換する一音語選定には前記の日本語 WordNet を用いる. 一音語を含む例文の収集にはクローリングも可能であるが, 今回は, 前もって収集されたウェブテキスト[1]内での検索によって例文の収集を行なった. このコーパスには IPAdic-2.7.0 の見出し語を検索語として Yahoo!WebAPI による検索結果に含まれるウェブページをクローリングした 396GB(数十億文)に相当するテキストが含まれる. 収集した文の中から最大 100 文をその読みのデータセットとした. 1 文に対して 2 名の作業者で評価を行い, 前記の 2 つの観点からの評価を行なった. 1 つの文に対して異なる 2 名の作業者を割り当て, 作業者には互いに相談することなく作業を行わせた.

3.2 収集例文数と作業者の信頼性評価

一音語を多音語に置き換えたことで日本語テキストとして意味の通らない文が存在するかもしれない. これは, 多義性解消機能自体で評価もできるが, 集めた文をより詳しく評価するため, 人手評価における「用法は正しいか否か」のラベルを通して妥当な置き換えが行われたか否かを判別した. 2 名の作業者がいずれも正しいと判断した文であり、かつ同一の読みを付与した文であれば訓練データとして利用可能であると考えた. Web コーパスから収集した例文数のうち, 2 名の作業者がいずれも正しいと判断した例文数(評価 1)と, 評価 1 かつ同一の読みに編集した例文数(評価 1&2)を集計し, 表 3 に示す. 両作業者が同一の読み仮名を選択した場合に限り、正解とした.

表 3 総例文と収集したデータの総数

例文数	評価 1	評価 1&2
11, 653	5, 790	5, 389

表 3 より例文数と評価 1&2 を比較すると, 収集された例文は Web コーパスから収集した例文数の 4 割である. また,2 人の作業者が正しいと判断した文に関してはほぼ一致した読み情報が付与されていることがわかる. しかしながら, 作業者の付与ラベルが正確であるとは限らないため, 次にラベルの信頼性について述べる. 作業者が“正しい”とラベル付けした文に対して, 人手で文中の多音語の用法が正しいか否かを再度評価し, 作業者との一致率を確認した. “作業者 2 名のうち、いずれかの作業者 1 名が正しいと判断した文”と”作業者 2 名が正しいと判断した文”のそれぞれをランダムに 100 文抽出し, 人手で再度評価したラベルと比較した一致率を表 4 に示す.

表 4 作業者が”正しい”と判断した文と人手で再評価した結果との一致率

作業者 1 名	作業者 2 名
42%	98%

[1] テキストアーカイブ-日本語ウェブコーパス2010, http://syata.jp/corpus/nwc2010/texts/

表 6 平均採用率 10%以上の多音語における語義ごとに獲得した文の例

異義語	語義/一音語	例文
今日	キョー/本日	8 月に入った途端、暑い夏が続いていますが、暦では秋‥今日は立秋です
	コンニチ/最近	ちなみに今日の研究結果からは 3000 年前には既に弥生時代に入っていた、て結果だったよね.
寒気	サムケ/悪寒	思っていたとしても、一歩間違えると私と姑の関係のようになってしまう寒気がします
	カンキ/冷たい風	午前には寒気が強くなっていたんですが午後には止み、JSB 決勝の頃はほぼカンペキなドライコンディション

作業者 1 名のみでは再評価したラベルとの一致率は低く，十分な信頼性は得られないが，2 名の作業者がいずれも正しいとする例文での一致率は非常に高く，十分な信頼性が得られていることがわかる．従って，収集した 11,653 文のうち評価 1,2 を満たす 5,389 文，割合で示すと 46.2%の文がデータセットとして利用可能であると言える.

3.3 単語ごとの例文収集効率について

前項では，データセットとして利用可能な例文数を示した一方で，収集した例文数に比べ利用可能な例文の割合が 4 割程度と効率の悪い収集方法であるように見える. そのため本項では多音語ごとに評価 1&2 を満たす例文数を評価することで，例文の収集効率について言及する．読みごとに Web テキストから収集した例文数と作業 1&2 の比率で採用率を計算し，その平均した結果を 表 5 に表す. この結果からは，データセット全体での採用率は 6 割程度であるように見えるが，これは採用率が 10%未満の著しく採用率の低い一音語が，対象とする多音語 82 語のうち 15 語存在しているためである. この採用率 10%未満の語を除いた例文の採用数を同様に表 5 に示す. この結果から本稿で対象とした 82 語のうち，15 語を除いた 67 語に対しては平均 8 割程度の採用率で訓練データを収集することができる.

表 5 データの語義ごとの収集率

対象	数量	平均採用率[%]
全語義	82 語	60. 4
採用率 10%以上の語	67 語	80. 4

次に，読みごとの例文収集効率について，単純に多音語を検索キーと比較した場合と比較する. “今日”を検索キーとして Web コーパス内を検索し，検索結果からランダムに 200 文を抽出，人手で”今日”の読み仮名を評価した. その結果，“キョー”が 197 件，“コンニチ”が 3 件となり，使用頻度に非常に大きな偏りがあることがわかった. その一方で，本手法では両方の読みを 200 文ずつ集めることができており，読み毎に例文を収集する手法として効率よく収集ができていたといえる.

3.4 採用率 10%以上の語の誤り分析

採用率の高い語とその例文について述べる. Web コーパスからの例文収集数が 100 文，かつ作業者 2 名にその全ての例文が正しいと判断された「今日(キョー/コンニチ)」「寒気(サムケ/カンキ)」について，各語義とその例文について表 6 に示す.「今日」について収集した例文では，その日一日を表す語義である(キョー)とその日だけでなくそれまでの日も表す語義(コンニチ)の例文が収集できていることがわかる.「寒気」でも同様に不快な寒気を表す語義と単に冷たい風を表す語義の例文を収集することができた. これらの語においては単語を置き換えたことにより，例文の内容が変わってしまうことは無く，適切な一音語への置き換えができていたといえる. 一方で誤りとなる場合についてランダムに 100 文を抽出し，以下の表 7 のように 3 種類の誤りに分類した.

表 7 収集率 10%以上の語の誤りの分類と割合

誤り要因	該当語数
1) 一音語ではない語への置換	55 語
2) 複合名詞の置換	43 語
3) その他	2 語

(1) 一音語ではない語への変換

[多音語”避け(ヨケ)”を”よけ”に置き換えた誤り]

原文：定点観測が 5/28 以降更新されていませんが、現在の状況をご存知の方、よければ情報を教えて下さい

置換後：定点観測が 5/28 以降更新されていませんが、現在の状況をご存知の方、避ければ情報を教えて下さい

“よけ”には”良け”, “除け”など曖昧性があるため，“除け”など曖昧性を回避する表現に置換するなどして，より適切な一音語へ置き換えることで改善できる.

(2) 複合名詞の変換

[多音語”頭(カシラ)”を”上部”に置き換えた誤り]

原文：赤石トンネルを上村側に抜けると、坑口上部に「赤石隧道」の表札が 掲げられている

置換後：赤石トンネルを上村側に抜けると、坑口頭に「赤石隧道」の表札が 掲げられている

複合名詞の多くはその組み合わせが異なると違和感のある文になりやすいのではないかと考えられる. 複合名詞になる例文はなるべく収集しないように改良する必要がある.

(3) その他

[多音語”何人(ナニジン)”を“どこの国の人”に置き換えた例]

原文：「私もこの人(資料請求した)どこの国の人と聞いたことがある」と話していたことである

置換後：「私もこの人(資料請求した)何人ナニジンと聞いたことがある」と話していたことである

プログラムの誤り(多音語の直後に読みが挿入される)，作業者への指導の誤り，数は少ないので無視できる.

3.5 採用率 10%未満の語の誤り分析

採用率が 10%未満の 15 語についても誤り分析を行った. 誤りを 3 つに分類し，読みごとに該当する誤りを分類した. 誤りの要因とその該当語数について表 8 に示す.

表 8 収集率 10%未満の語の誤りの分類と割合

誤り要因	該当語数
1) 不適切な一音語への置換	10 語
2) 一音語ではない語への置換	3 語
3) その他	2 語

次に、誤りの要因の例とその誤りについての今後の対策について挙げる.

(1) 不適切な一音語への変換

[多音語”面(ツラ)”を”顔面”に置き換えた誤り]

原文：ただ普通に勝負したら人間が勝つのは目に見えるので、顔面に洗濯ばさみというスタイルで挑みます.

置換後：ただ普通に勝負したら人間が勝つのは目に見えるので、面に洗濯ばさみというスタイルで挑みます.

このカテゴリは一音語を多音語の置き換え語として選択したが, 誤りとなった語義が含まれる. 例においては”面”は不適切な置換であるが, “相手の面に蹴りを入れる”など俗語的な使われ方をする場合には適切な場合がある. 従って, 多音語の出現する文脈を考慮した置換が必要であるが, 多音語と同一の文脈で出現する適切な一音語が見つからない場合には, 収集効率は低下することを踏まえた上で, 必要数より多く見積もって例文を収集しておく必要がある.

(2) 一音語ではない語への置換

[多音語”人気(ヒトケ)”を“人の気配”に置き換えた誤り]

原文：蠍座田中支配人の気配りで、お客様がより時間に制約されないようにと、午前と午後の 1 日 2 回のプログラムを組んでくれました.

置換後：蠍座田中支配人気で、お客様がより時間に制約されないようにと、午前と午後の 1 日 2 回のプログラムを組んでくれました.

“気配”には”ケハイ”と”キクバリ”の読み曖昧性があるため, 表記の一致をとる場合に”キクバリ”と読む文を誤って収集してしまっている. “気配り”と書く場合は”キクバリ”という読みを予め原文に割り当てることでこの曖昧性を回避し, 文の収集ができる.

(3) その他

3.4 項同様の誤り, 無視できる範囲で事例が少ないため記述を割愛した.

以上を踏まえて, 適切な一音語への置換を行うためには, 異義語の語義に近い単語かつ多義を持たない場合であれば有効であると考える.

面(ヅラ)のような同義語を人手で考えることが難しい単語に対して, 類似語検索等の補助を行うことで, より適切な一音語への変換が可能になると考えられる. また、置き換え先の語は一定以上の出現頻度が必要である. “中日(ナカビ)”を”中間の日”に置き換えを行ったが, Web コーパスにおいての頻度が非常に低く 3 文程度のデータ収集に留まった. 実験に用いた Web コーパスの量は十分であることから, クロール量を増やす以前にある程度の頻度を持つ単語への置き換えが重要である.

最後に例文の収集にかかる時間ついて考察する. 1 人当たり約 4, 000 文の評価に費やした時間は, 納品までに約 2 日間である. 例文が正しいか否かを判断するために, 例文を全て読む必要があるが, 例文が正しいか否かを判断した後の読みの読み判別は日本語母語話者であれば比較的迅速に作業を行うことができていた. 新たな例文を収集するために本手法では既に Web テキストとして存在している文の一部を改変する形で拡充しているが, もうひとつの手法としてクラウドソーシングを用いた人手での例文生成が考えられる. しかし, 本稿の手法と比較すると 1 文を生成するために必要なコストが高く, アノテート済みのデータを大量に作ることには向かないと考える.

作業が簡単であるため, 1 人あたりの金額を抑えで募集することができた. 更に多くのデータを収集する際に低コストで大量に収集するための手法として有効である.

4. おわりに

本稿では, 既存のコーパスでは不足している同形異音異義語(本稿では多音語と定義)の曖昧性解消のためのデータセットの構築を行った. 多音語の語義ごとに語義を一つしか持たない語へ人手で置換を行い, 置換した語を検索キーとして Web コーパスから置換した語を含む例文を収集した. 40 組 82 語の異義語が含まれる計 11, 653 文の例文に対して, 1 文あたり 2 名の作業者が人手でのアノテート作業を行い, 例文が利用可能か否かと読みアクセント情報を付与した. その結果, 訓練データとして利用可能であると判断された 5, 790 のデータセットを構築した. 本手法では異義語を一音語へ置換することで, “今日(コンニチ)”のような語義によって出現頻度に大きな偏りがある語であっても, 例文を十分数収集できることを示した. また, 15 語の著しく例文の収集率の低い異義語を除いた 67 語の異義語に対して, Web コーパスから収集した例文のうち約 80. 4%を訓練データとして利用可能であることを示し, 語義ごとの例文収集手法として有効な手法であることを示した.

参考文献

[山森 99] 山森和彦: 未来ねっと技術シリーズ 4 メディア処理技術. オーム社, 1999.

[長野 06] 長野徹, 森信介, 西村雅史: N-gram モデルを用いた音声合成のための読みおよびアクセントの同時推定. 情報処理学会論文誌, Vol. 47, No. 6, pp. 1793‐1801, 2006.

[森 10] 森信介, Graham Neubig: 仮名漢字変換ログの活用による言語処理精度の自動向上. 言語処理学会第 16 回年次大会, 2010.

[新納 15] 新納浩幸, 白井清昭, 村田真樹, 福本文代, 藤田早苗, 佐々木稔, 古宮嘉那子, 乾孝司: クラスタリングを利用した語義曖昧性解消の誤り原因のタイプ分け, 自然言語処理, Vol. 22, No. 5, pp. 319-362, 2015.

[Okumura 11] Okumura, M. , Shirai, K. , Komiya, K. , and Yokono, H. : “On SemEval-2010 Japanese WSD Task. ”, 自然言語処理, Vol. 18, No. 3, pp. 293-307, 2011.

[西尾 94] 西尾実, 岩淵悦太郎, 水谷静夫, 岩波国語辞典第 5 版, 岩波書店, 1994.

[戸田 16] 戸田勇馬, 村田真樹, 馬青: 言い換えと機械学習を用いた日本語単語の多義性解消, 言語処理 学会 第 22 回年次大会, Vol. 18, No. 3, pp. 172-175, 2016.

[Isahara 08] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama and Kyoko Kanzaki: Development of Japanese WordNet, LREC-2008 , Marrakech , 2008 .

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

複数オブジェクトのイベントの組み合わせによる行動記述方式

An Activity Description Method by Multiple Events of Multiple Objects

瀬光 孝之 Takayuki Semitsu　南本 高志 Takashi Nammoto　毬山 利貞 Toshisada Mariyama

三菱電機株式会社
Mitsubishi Electric Corporation

Events of single object is not sufficient for describing highly semantic activities. We propose a description method for person activity recognition, especially abnormal activity detection. We discussed the need of appropriete granularity of both spatial and temporal event descriptions. Using 3 examples from Sports activities, we evaluated the effectiveness of our description method.

1. はじめに

行動認識の研究で認識対象となっている「行動」は，体の動きに関する具体的な記述から，状況や相手の反応まで描写する抽象的な記述まで，様々である．これを記述の意味的な階層ととらえる．厳密な区別は難しいが，大まかに以下の 3 つの層に分けられる．

1. 低層:「手を動かす」など，体の動きに関する具体的な記述 [Kwapisz 11].
2. 中層:「皿を持つ」など人物やモノに対する直接的なインタラクションを含んだ記述 [中川 16].
3. 高層:「談笑する」・「〜するため，ボタンを押す」など，状況・未来・状態変化を含む記述．低・中層で記述できない．

高層の行動記述では，相手の反応を含んだ行動の記述を行う．例えば，「話す (発話する)」という動作でも，相手が答えて発話すれば「会話する」であるが，相手が不審に思って「離れ」れば「不審行動」である (図 1a，図 1b)．また，人物・モノに関する属性が加わることで行動の意味が変わる場合がある．例えば，シナリオ 2 と同じ「発話する」,「離れる」という動作でも，発話したのが警察であれば「不審行動」をしているのは「離れた」側の人である (図 1c)．このように，高層の行動を記述するためには低層・中層の記述では不十分で，相手のインタラクションまで含めて記述しないと状況がわからない場合がある．先行研究では，物との直接的または同時のインタラクションを含んだ中層の記述が主な認識対象とされているが，状況・未来・状態変化を含む高層の記述についての認識はこれまで事例が少なかった．

オブジェクト (人物やモノ) が画面に登場したり，何らかの行動をとることをイベントと呼ぶことにする．高層の行動記述を低層・中層の記述のみを用いて実現するため，複数オブジェクトのイベントの組み合わせにより行動を記述する方式を提案する．これはオブジェクトのイベントとそれに対する周囲の人・モノのイベントをまとめたものである．

連絡先: 瀬光孝之，三菱電機株式会社 情報技術総合研究所，〒 247-8501 神奈川県鎌倉市大船 5-1-1 ，Semitsu.Takayuki@dw.MitsubishiElectric.co.jp

以降では，2. 節で提案方式の説明を行い，3. 節でスポーツプレイを例にとった行動記述粒度の必要性を議論する．

2. 複数オブジェクトのイベントの組み合わせによる行動記述方式

イベントは，オブジェクトの見え方に関するものと，動きに関するものに分けられる．見え方のイベントとは，ある属性を持った人物が画面に登場した際に発生するイベントである．推定には人物検出結果および属性の識別結果を用いる．今回は CNN ベースの人物検出器を用い人物を検出する．検出された人物に対し，制服を着ているなどの属性を識別する．画面に現れた時間を登場イベントの起こった時間とし，人物を追跡した軌跡をまとめて 1 つのイベントとした．

動きに関するイベントは，人物が画面上でどう動いたか記述するものである．人物追跡結果から矩形の移動を検出し，静止状態から動きを開始する動き始めのイベントを記録した．

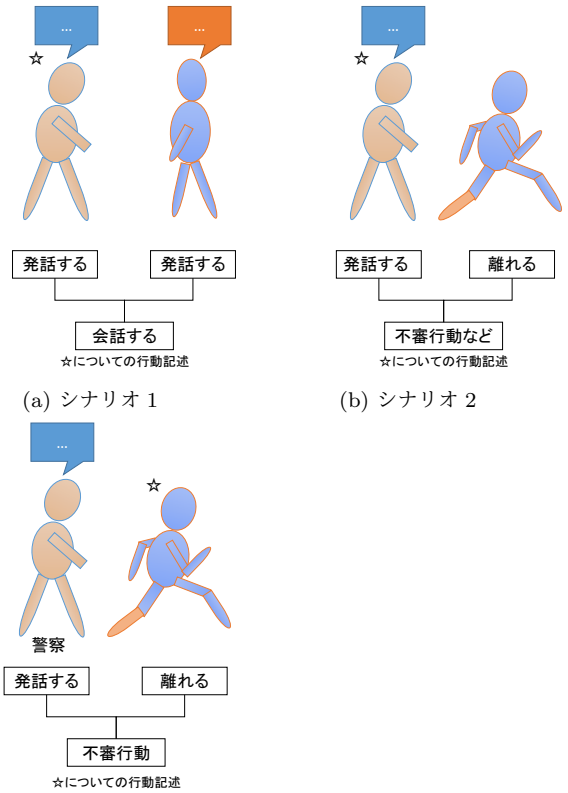
これらのイベントをオブジェクトごとに検出し，それらを組み合わせるルールを定めることで，高層の行動記述を実現する．組み合わせの方法はさまざま考えられる．

1. 一定時間内にあるイベントの組み合わせが共起する．
2. 一定時間の間にあるイベントの組み合わせがその順番で起こった．
3. 上記の組み合わせで，あるイベントは順序関係を持ち，他のイベントは順不同に起こった．

身体的特徴・行動的特徴を用いる場合，前述の曖昧性により不審行動の検出が難しかった．複数のイベントの組み合わせに注目することで，前記のような曖昧性のない不審行動の検出が可能となる．

3. 行動記述の粒度

高層の行動記述を複数のイベントの組み合わせにより記述する例として，アメリカンフットボール (アメフト) のセットプレイを取り上げる．アメフトは 1 チーム 11 人のプレイヤーにより行うが、セットプレーの種類によって各選手の動き方が細かく決められている。プレー始めの選手配置 (フォーメーション) ごとに 5-10 通りの動き方のセットプレイが設定され



(a) シナリオ 1　(b) シナリオ 2

(c) シナリオ 3

図 1: 行動認識のシナリオ

表 1: アメフトの行動記述例: セットプレイ A

X 座標	Y 座標	X 速度	Y 速度	方向	ID
0	2	0	1	前	1
5	0	-0.8	-0.6	左前	2
10	1	0.9	0.2	右	3
11	2	0.5	0.5	右前	4
12	1	-0.3	-0.4	左前	5

表 2: アメフトの行動記述例: セットプレイ B

X 座標	Y 座標	X 速度	Y 速度	方向	ID
0	2	0	1	前	1
5	0	0	1	前	2
10	1	0.9	0.2	右前	3
11	2	0.6	0.8	左前	4
12	1	0.9	0.2	右前	5

表 3: アメフトの行動記述例: セットプレイ C

X 座標	Y 座標	X 速度	Y 速度	方向	ID
0	2	0.7	0.7	右前	1
0	5	-0.8	0.6	左前	2
10	1	0.9	0.2	右	3
11	2	-0.8	-0.6	左前	4
12	1	0	1	前	5

ている例が多い．アメフトのあるチームのセットプレイの例を 3 つ，表 1〜表 3 に示す．各表で，セットプレイにおいて，ID 列で識別される各プレイヤーの位置と速度，方向を記述している．ただし，座標値・速度は説明のため，フィールド上で攻撃方向を上としたとき左下を原点とするグリッドを設定し，セットプレイ A の ID:1 のプレイヤーの最初の行動の速度を 1 と定めて記述している．

セットプレイ同士を区別するためにこれらの記述のうちどれが必要かは自明でない．例えば，速度については速さ，方向のそれぞれの値が比較対象となるが，どこまで量子化するかが記述の粒度に対応する．方向については，8 方向 (前，右前，右，右後，後，左後，左，左前) の表記を”方向”列に記載した．例に挙げたセットプレイの例 3 つでは 8 方向による量子化で十分であり，速さは比較する必要がない．しかし，位置情報を除いた場合，セットプレイ A と C はともに (前，右前，右，左前，左前) の組み合わせで記述されるため，区別ができない．このように，行動記述の粒度は適切に設定する必要がある．時間についても同じことが言え，時間を考慮せず順不同に並べて扱ってよいのか，順番を考慮すべきなのか，時間を量子化して時間間隔を考慮すべきなのかを検討する必要がある．

4. おわりに

本研究では複数オブジェクトのイベントの組み合わせによる行動記述方式を提案した．アプリケーションの一例として不審行動を取り上げ，提案方式の必要性・有効性を説明した．複数のイベントの組み合わせにより一連の行動を記述することで，1 人の行動記述では曖昧性の残っていた場面で行動を識別することが可能になる．高層の行動を識別するため，どの程度の行動記述粒度が必要であるかは問題ごとに異なる事を，スポーツのセットプレイを例として説明し，行動記述の意味的・時間的粒度を設定する必要がある事を説明した．

今後の課題として，識別対象である高層の行動 (不審行動や) を提案方式により識別する実験評価を行う．加えて，システムを実運用していくにあたってはプライバシー問題への配慮すること，不審行動パターンを自動で抽出するような運用面の効率化・自動化が必要である．また，複数オブジェクトのイベントの組み合わせとして実際にはインタラクションのなかった (たまたま同時に起こった) イベントも同時に記録される可能性がある．インタラクションの検出方法についても検討する必要がある．

参考文献

[Kwapisz 11] Kwapisz, J. R., Weiss, G. M., and Moore, S. A.: Activity recognition using cell phone accelerometers, *ACM SigKDD Explorations Newsletter*, Vol. 12, No. 2, pp. 74–82 (2011)

[中川 16] 中川愛梨, 守谷一希, 諏訪博彦, 藤本まなと, 荒川豊, 八田俊之, 三輪祥太郎, 安本慶一 他：位置・電力情報を基とした宅内行動認識システムへの居住者の加速度データの追加とその改善効果の調査, マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016, pp. 1442–1448 (2016)

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

情報科学論文における問題解決手法と評価表現の付与仕様の検討

Towards Annotating Problem Solutions and Evaluation Expressions in Information Science Papers

白井 穂乃 *1　井之上 直也 *1　乾 健太郎 *1*2
Hono Shirai　Naoya Inoue　Kentaro Inui

*1東北大学 Tohoku University
*2理化学研究所 AIP センター RIKEN Center for Advanced Intelligence Project (AIP)

The ultimate goal of our research is to extract problem solutions and its evaluation expressions from Information Science papers. This paper presents the preliminary annotation scheme for this task and conducts an annotation study in Natural Language Processing papers. By conducting the annotation study, we discuss the remaining issues of the present annotation scheme and its future direction.

1. はじめに

近年、学術論文の出版数が急増している。米 American Journal Experts の 2016 年次報告書 *1 によると、薬学の分野では 560,000 本、工学では 280,000 本、生物分野では 260,000 本を超える論文が 1 年のうちに出版されたと言われている。我々研究者は、こうした膨大な学術文献の中から関連分野の情報を適切に取捨選択し、把握することが年々難しくなってきている。

このような状況を打開すべく、学術論文から有用な情報を自動抽出するための様々な研究が盛んに行われている。例えば、Teufl ら [9] は、学術論文の各文を「背景に関する記述」「先行研究に関する記述」などに分類する Argumentative Zoning というタスクに取り組んでいる。また、文献からの情報抽出を関係抽出の問題として定式化し、評価型ワークショップを開催する動きもある。例えば、BioNLP [3] では、生物医学分野の文献を対象として、タンパク質等の専門用語の認識タスク、たんぱく質間の関係や、物質とその副作用などを抽出する関係抽出タスクの研究が行われている。また、ScienceIE [1] では、物理学、材料科学、計算機科学を対象として、フレーズの抽出や同義語・下位語の関係抽出などの研究が行われている。また、引用評価極性解析 (Citation Sentiment Analysis) の分野では、文献内で引用している文献に対する著者の感情極性を解析する研究が盛んに行われている [6]。最後に、文献間の引用関係に基いて、文献の重要度や、技術のトレンドなどを自動的に解析する取り組み (Citation Network Analysis) も盛んに行われている [7]。

このような背景のもと、我々は、計算機科学分野の学術論文の自動解析の研究に取り組んでいる。計算機科学の論文の中で主に展開される議論のうちの一つは、ある問題（タスク）に対する解決手法を提案し、提案手法と既存手法のメリット・デメリットを述べるものである。そこで我々は、まずこうした情報の自動抽出を実現するために、文献内の**あるタスクを解くための手法**（**問題解決手法**と呼ぶ）、とその手法に対する**特定の観点に基づく評価**を抽出することを目指している。例えば、次のような一文が論文内にあったとしよう。

(1) *The results indicate that the whole-sentence-based classifier performs the best.*

例 (1) では、*the whole-sentence-based classifier* という手法に対して、*performs the best*、すなわち「性能という観点」において、「良い」という評価がなされていることを自動認識するのが、我々の目的である。このように個々の論文における問題解決手法とそれに対する評価を自動認識し、その結果を集約することにより、分野全体の技術の特徴を手早く俯瞰できるようになると期待できる。

自然言語処理の分野では、このような観点付き評価の抽出問題は、評価分析 (Sentiment Analysis)、または意見マイニング (Opinion Mining) として、古くから取り組まれてきた [5]。しかしながら、こうした解析を学術論文のドメインに適用する試みはこれまでになく、タスクの具体的な設計方針（何を評価対象とみなすか、何を観点とみなすかなど）、コーパス等の言語資源も整備されていない。

そこで本稿では、学術論文ドメインにおける観点付き評価の抽出タスクの設計方針を明らかにするための予備調査について報告する。より具体的には、ある一つのタスクについての自然言語処理の学術論文を対象とし（2. 節）、計算機科学ドメインに特化した観点付き評価分析のタスクを設計する（3. 節）。また、この仕様に基いて、実際にコーパスアノテーションを行った結果を報告する（4. 節）。

2. データ

まず、本稿が研究対象とするデータについて説明する。本稿が対象とする論文は、ACL Anthology Corpus [2] に収録されている自然言語処理の論文である。自然言語処理では、様々な問題（またはタスク）が定義されており（例えば、形態素解析、構文解析）、それらに関してさまざまな問題解決手段が提案されている。本稿では、観点付き評価抽出タスクの設計に関する知見を得るために、ある一つのタスク、感情分析 (Sentiment Analysis) についての論文を収集した。具体的には、まず Google Scholar*2 を用いて、“Sentiment Analysis” をクエリとして検索し、引用関係にある合計 36 本の論文を収集した。このうち、ACL Anthology Corpus に含まれる 28 本の論文からテキストデータを抽出した。本稿では、問題解決手法の特徴や、既存研究との差異について端的に書かれている Introduction の節を用いて後述のアノテーションを行った。今後、これを Related Work の節に拡張することも予定している。

連絡先: 白井穂乃，東北大学大学院情報科学研究科，乾研究室，hshirai@ecei.tohoku.ac.jp

*1 https://www.aje.com/jp/arc/dist/docs/International-scholarly-publishing-report-2016.pdf

*2 http://scholar.google.com

3. アノテーションスキーム

自然言語処理のような技術論文は、問題を解決するための技術の提案について書かれている。

本研究では解決する手法 (TERM) とその評価極性、TERM の評価の理由 (REASON) という 2 つの要素についてアノテーションすることとした。

3.1 アノテーションの仕様

アノテーションするラベルを 2 つ定義する。問題を解決する手法・しくみである TERM, その TERM に対する評価を判断する理由・根拠 REASON である。以降の節では、それぞれの定義について説明する。

3.1.1 TERM

問題解決の手法に関する記述をアノテーションするために、TERM というラベルを導入する。具体的には、モデル・アルゴリズムといった仕組み・仕組みの持つ機能・仕組みが動作する方法を表す名詞句を TERM として付与する。例えば、下記の文を見てみよう。

(2) *We employ a novel adaptive multi-compositionality layer in recursive neural network, which is named as AdaRNN (Dong et al., 2014).*

例 (2) では、*recursive neural network* と *AdaRNN* はそれぞれニューラルネットワークというモデルとその一種であるため、TERM として付与する。

なお、既存研究では問題設定 (e.g. Sentiment Classification, Classification Task) もアノテーション対象としているが、本研究は分野全体の技術の把握を目的としているため、問題設定についてはアノテーション対象としない。

3.1.2 REASON

次に、前述の TERM ラベルが付与された問題解決の手法に対する著者の観点付き評価を捉えるために、REASON というラベルを導入する。REASON は、論文の著者が TERM というターゲットに対する評価を述べているフレーズ（評価表現）である。例えば、下記の例文を見てみよう。

(3) *The results indicate that the whole-sentence-based classifier performs the best.*

例 (3) では、まず *the whole-sentence-based classifier* が TERM として付与される。著者は、*the whole-sentence-based classifier* に対して、*performs the best* (その性能が最高である) と評価しているため、これを REASON と付与する。なお本稿では、REASON は TERM と同一文内のフレーズに限って付与し、文をまたいだ TERM と REASON の関係は、後処理において共参照関係の解消を行った後に集約することを想定する。また、REASON は必ず単一の TERM に対応しているため、TERM と REASON の間に Link という関係を定義する。

3.1.3 TERM-POSITIVE, TERM-NEGATIVE

さらに本稿では、評価分析 (Sentiment Analysis) の先行研究に倣い、評価をポジティブ・ネガティブの二値で表現し、この評価を TERM に対する属性として付与する（以後、これを TERM-POSITIVE, TERM-NEGATIVE と示す）。例 (3) では、*performs the best* は TERM である *the whole-sentence-based classifier* に対するポジティブな評価であると考えられるため、TERM に「ポジティブ」の属性を付与する。

なお、REASON を付与しない事例として、単に問題解決手法の特徴、性質を述べているだけのフレーズがある。例えば、

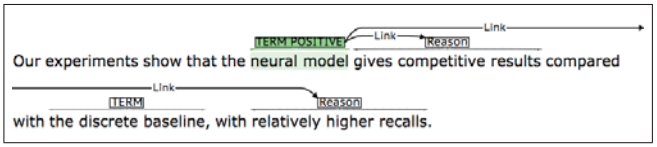


図 1: brat によるアノテーション

表 1: アノテーション結果

ラベル		数
TERM		520
	TERM-POSITIVE	107
	TERM-NEGATIVE	31
REASON		147

例 (2) では、著者は TERM である *recursive neural network* に対する評価を示していないため、REASON ラベルは付与しない。

3.2 アノテーションの手順

前節の仕様に基いて、下記のような手順でアノテーションを行う。

1. Introduction 内の問題解決手法を表す表現に対して TERM ラベルを付与する。
2. TERM と同一文内において、TERM に対する評価表現が存在するか判断する。あると判断した場合、下記 (a), (b) を実施する。
   (a) 評価表現に基いて、TERM ラベルを TERM-POSITIVE,TERM-NEGATIVE ラベルに変更する。
   (b) 評価表現に REASON ラベルを付与する。
   (c) TERM-POSITIVE, TERM-NEGATIVE から REASON に Link ラベルを付与する。

4. 結果と考察

3. 節で定義した仕様の妥当性を検証するために、2. 節のデータに対して実際にアノテーションを行った。本稿では、アノテーション結果の考察を通して、アノテーションをさらにスケールさせる際の課題について議論する。

なお、アノテーションの際には、brat[8] を用いて図 1 のようなユーザインターフェースを作成し、第一著者 1 名がアノテーションを行った。複数アノテータによるアノテータ間一致率の分析については、今後の課題とする。

アノテーション結果は表 1 のようになった。

以下、アノテーション結果から分かったことをさらに詳しく報告する。

4.1 TERM の種類

TERM のラベルが付与される単語にどのような傾向があるか考察する。まず、model, approach, method など TERM の定義そのものを表す単語を含む名詞句には、TERM のラベルが付与されていることが分かった。

(4) *... our model outperforms state-of-the-art methods in both the supervised and semi-supervised settings*

例 (4) では、model という単語が含まれる *our model* と method という単語が含まれる *state-of-the-art methods* に TERM のラベルが付与された。

一方で、TERM の定義を表す単語が含まれていない名詞句に TERM が付与された事例もある。これらの事例には、文脈によって TERM であるか否かが変わるような名詞句が含まれていた。例えば、以下の 2 つの事例を見てみよう。

(5) *... we do not require gold standard sentence-level labels for training.*

(6) *... we propose a joint two-level model to address the aforementioned concerns.*

例 (5) において *we* は*"our model"*と同じ意味を持つため TERM であるが、例 (6) では *we* はモデルを提案する著者であるので TERM ではない。このほか、共参照関係を解消する必要がある *it* などの代名詞についても、同様の傾向がみられた。

4.2 TERM-POSITIVE, TERM-NEGATIVE

TERM-POSITIVE, TERM-NEGATIVE が付与された事例には、大きく (1) ドメイン非依存なポジティブ・ネガティブな評価表現を含んでいる事例、(2) ドメイン依存な評価表現を含んでいる事例があった。

ドメイン非依存な評価表現の事例には、例えば次のような事例がある。

(7) *... our system can generate high-quality labeled data.*

例 (7) では、*our system* に対して TERM-POSITIVE のラベルが付与され、このときの評価表現として *can generate high-quality labeled data* が付与されていた。この評価表現には、*can* や *high-quality* といった、ドメイン非依存と考えられるポジティブな単語が含まれている。高い質のラベル付きデータを生成できるというフレーズは *our system* をポジティブに評価する表現であると判断できる。このようなドメイン非依存な単語が含まれている場合は感情極性辞書を利用することで自動的に解析することが可能だと考えられる。

一方で、ドメイン依存な評価表現の事例には、例えば次のような事例がある。

(8) *This approach requires no manually-specified information about the meaning of the connectors, just the connectors themselves*

例 (8) では *This approach* に TERM-POSITIVE のラベルが付与されており、*requires no manually-specified information* に REASON のラベルが付与されていた。ここでは、REASON の部分にはポジティブな単語が含まれていないが、*manually-specified information* (手作業で指定された情報) を必要としないという表現は「タスクを解決する上で必要とする情報が少ない」という観点において TERM にとってポジティブであることを示すため、*requires no manually-specified information* は *This approach* の評価表現である、というドメインの知識に基づくある種の推論が起きている。このようなドメイン依存な知識に基づく評価表現は、アノテータのドメインに関する知識に依存するため、将来アノテータ間の揺れが起こりやすいと予想される。

4.3 REASON における観点の種類

最後に、REASON のラベルを付与された事例について、評価表現にどのような観点が含まれているか考察する。REASON のラベルを付与された事例には、(1) TERM の性能、 (2) TERM の前提条件、 (3) TERM の持つ機能、という観点があった。

まずはじめに、性能の観点における評価表現として、例えば次のような事例がある。

(9) *... it is a strong and robust performer*

例 (9) は *it* に TERM-POSITIVE のラベルが付与されており、*a strong and robust performer* という評価表現に REASON が付与されていた。ここで、*performer* は性能という観点を想起させる単語であり、この観点に対して、*strong, robust* という修飾句で、ポジティブな評価を表明している。同様に、例 (4) は、*outperform* が性能という観点を表す単語であり、かつ「上回る」というポジティブな意味を持つ単語であるためポジティブな評価を表明している。

次に、TERM の前提条件における観点の事例を挙げる。例 (8) では、*requires no manually-specified information* に REASON のラベルが付与されていた。ここでは、*This approach* が動作するためには、*manually-specified information* が必要ないこと、すなわち前提条件に関する観点が示されている。*manually-specified information* は、ドメイン知識によりネガティブなもの（準備にコストがかかるもの）であると読み手には分かるため、これが必要ないと表明することによって、この前提条件が暗に「良い」ことを表明している。

最後に、TERM の機能という観点における評価表現として、次のような事例がある。

(10) *... neural models use real-valued hidden layers to automatically learn feature combinations, which can capture complex semantic information that are difficult to express using traditional discrete manual features.*

例 (10) では *neural models* に TERM-POSITIVE のラベルが付与されており、それに対する評価表現として *can capture complex semantic information* に REASON のラベルが付与されている。ここでは、*can capture* は TERM の持つ機能という観点を想起させ、*complex semantic information* という難しい情報を捉える、と表明することで機能がポジティブであると評価している。

(11) *The model can also avoid overfitting to features derived from neutral or objective sentences.*

例 (11) では *The model* に TERM-POSITIVE のラベルが付与されており、その評価表現である *can also avoid overfitting* に REASON のラベルが付与されている。*can also avoid* は TERM の機能という観点を想起させる単語であり、*overfitting* (過学習) というネガティブなものを防ぐことができる、と表明することで機能がポジティブであると評価している。

5. 関連研究

建石ら [10] の研究は、情報科学論文に出現する用語間の関係を構造化するためのタグ付けスキーマを提案している。これは論文中に存在する用語すべてに対して意味クラスを付与し、用語間にも関係のクラスを付与することで論文内容の構造化を行うことを目的としている。また、自然言語処理分野の評

価型ワークショップ SemEval では、論文ドメインでの情報抽出タスクが提案されている。SemEval-2017 の評価タスクである ScienceIE[1] は物理学、材料科学、計算機科学の論文からフレーズと関係の抽出を行うタスクである。フレーズは Task, Process, Material の 3 つのクラスに対応している。フレーズのクラスを分類するタスク、フレーズ同士が類義語・下位語の抽出を試みるタスクが提案されている。SemEval-2018 Task 7 [4] は ACL Anthology Corpus[2] の Abstract について、エンティティ同士の関係を分類するタスクを提案している。エンティティは概念を表現する名詞句と規定されており、比較・結果などの 5 つの関係を抽出・分類するタスクである。本研究では、これら一般的な論文解析タスクとは異なり、文章中の用語を全てを対象にはせず、問題解決手法である TERM のみを抽出している。また、TERM だけでなく TERM に対する評価（極性）とその評価表現 REASON を付与している。

冒頭でも触れたが、自然言語処理の分野では、レビュー文章のドメインにおける観点付き感情極性分析が行われている。例えば、SemEval-2015 Task 12[5] は、ホテルやレストランの料理の価格やサービスの質などの定義した観点に基づいて評価分析を行うタスクである。しかし、我々の知る限り、論文ドメインにおいては観点に基づいた評価分析は行われていない。

6. おわりに

本稿では、情報科学論文から問題解決手法とその評価表現を自動抽出することを目標に見据え、そのアノテーションスキームの設計・検討を行った。アノテーションの結果より、アノテーションをさらにスケールさせる際の注意点、すなわち、問題解決手法ラベル TERM の文脈依存性や、評価表現ラベル REASON 付与時におけるドメイン特化知識の必要性が明らかになった。今後の課題として、付与対象の大規模化、複数アノテータによる付与、アノテータ間一致率の分析、相対的な評価表現（既存手法に比べて「相対的に」ポジティブ等）への対応、自動解析モデルの構築が挙げられる。

謝辞

本研究は、JSPS 科研費 16H06614、JST CREST (課題番号: JPMJCR1513) の支援を受けて行った。

参考文献

[1] Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum. Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 546–555, 2017.

[2] Steven Bird, Robert Dale, Bonnie J Dorr, Bryan Gibson, Mark T Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. *Proc. Sixth Int. Conf. Lang. Resour. Eval. (LREC 2008)*, pp. 1755–1759, 2008.

[3] Louise Delėger, Robert Bossy, Estelle Chaix, Mouhamadou Ba, Arnaud Ferrė, Philippe Bessieres, and Claire Nėdellec. Overview of the bacteria biotope task at bionlp shared task 2016. In *Proceedings of the 4th BioNLP Shared Task Workshop*, pp. 12–22, 2016.

[4] Kata Gábor, Davide Buscaldi, Anne-Kathrin Schumann, Behrang QasemiZadeh, Haïfa Zargayouna, and Thierry Charnois. Semeval-2018 Task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.

[5] Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, and Michal Konkol. UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proc. 10th Int. Work. Semant. Eval.*, pp. 342–349, 2016.

[6] Myriam Hernández-Alvarez and José M. Gomez. Survey about citation context analysis: Tasks, techniques, and resources. *Nat. Lang. Eng.*, Vol. 22, No. 3, pp. 327–349, 2016.

[7] Yuya Kajikawa, Junko Ohno, Yoshiyuki Takeda, Katsumori Matsushima, and Hiroshi Komiyama. Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, Vol. 2, No. 2, p. 221, Jul 2007.

[8] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 102–107, Avignon, France, April 2012. Association for Computational Linguistics.

[9] Simone Teufel, et al. *Argumentative zoning: Information extraction from scientific text.* PhD thesis, University of Edinburgh, 2000.

[10] 建石由佳, 仕田原容, 宮尾祐介, 相澤彰子. 情報科学論文からの意味関係抽出に向けたタグ付けスキーマ. 言語処理学会第 19 回年次大会, 2013.