Naoki Otani

Language Technologies Institute   Email: notani@cs.cmu.edu
School of Computer Science
Carnegie Mellon University

EDUCATION

2019	Ph.D. in Language and Information Technologies, Carnegie Mellon University
	Advisor: Dr. Eduard Hovy
2017–2019	Master of Science in Language Technologies, Carnegie Mellon University
	Advisor: Dr. Eduard Hovy (2018-), Dr. Yiming Yang (2017-2018)
2015–2017	Master of Informatics, Kyoto University, Japan
	Thesis: "Commonsense Knowledge Acquisition via a GWAP on a Dialogue System"
	Advisor: Dr. Sadao Kurohashi
2011–2015	Bachelor of Engineering, Kyoto University, Japan
	Thesis: "Quality Control of Hierarchical Crowdsourced Classification"
	Advisor: Dr. Hisashi Kashima

WORK EXPERIENCE

May–Aug. 2021 **Research Intern**   *Microsoft Research*, Remote

- Developed a method for automatically inducing intent-based representations of to-do task texts
- Exploited external knowledge resources to automatically acquire training data of the model
- Presented the outcome as a long paper at the top-tier academic conference

Jun.–Aug. 2020 **Research Intern**   *Robert Bosch LLC*, Remote

- Developed a response generation model based on social common-sense knowledge
- Filed a patent on conversational AI technology

Aug.–Oct. 2016 **Research Intern**   *Microsoft Research Asia*, Beijing, China

- Developed a paraphrasing method for Japanese chat-bots
- Mined a paraphrasing table from Japanese microblog data

Feb. 2016 **Research Intern (part time)**   *Yahoo Japan Corporation*, Tokyo, Japan
–Feb. 2017
- Developed a quiz game in C++ for acquiring common-sense knowledge from users
- Closely collaborated with the development team and deployed the game on a widely-used conversational assistant app, *Yahoo! Onsei Assist (Yahoo! Voice Assistant)*

Sep. 2015 **Research Intern**   *IBM Research - Tokyo*, Tokyo, Japan

- Developed a method for acronym disambiguation in biomedical papers based on citation networks

RESEARCH PROJECTS

Common-sense acquisition and reasoning for language understanding

- Integration of event knowledge for situated response selection/generation (ongoing work)
- Text representation learning based on event knowledge for short and under-specified To-Do descriptions [2]
- A dataset and a method to identify human motives in restaurant and movie review texts for explaining the reason of sentiment [4]
- Low-cost acquisition of Japanese common-sense by translating English text data (Contributed to ConceptNet (https://conceptnet.io/)) [6]

Cross-lingual natural language processing

- Analysis of multi-word expressions in cross-lingual word embeddings [3]
- An unsupervised method to align word embeddings [5]
- Empirical analysis of methods for cross-lingual event type detection in a low-resource setting [7]

Crowdsourcing for data collection and system evaluation

- Statistical model based on item response theory for aggregating annotation from crowd workers to perform multi-class classification [1][9] and evaluate machine translation systems [8]
- A quiz game on a conversational assistant app in C++ to acquire large-scale common-sense data from users [10]

AWARDS

2017–2019	Funai Overseas Scholarship (Full amount of tuition for two years. 9/109 applicants were awarded.)
2016	Award of Excellence, Internship program at Microsoft Research Asia
2016	Student Scholarship, EMNLP2016

SKILLS

Computing	Python, C/C++, HTML/CSS, Shell script
Language	Japanese (native), English (fluent), Chinese-Mandarin (conversant)

PUBLICATIONS

Journal

[1] **Naoki Otani**, Yukino Baba, and Hisashi Kashima. 2016. Quality Control of Crowdsourced Classification Using Hierarchical Class Structures. *Expert Systems with Applications (ESWA)*, 58:155–63.

Conference

[2] **Naoki Otani**, Michael Gamon, Sujay Kumar Jauhar, Mei Yang, Sri Raghu Malireddi, and Oriana Riva. 2022. LITE: Intent-based Task Representation Learning Using Weak Supervision. In *Proceedings of 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 2410–2424, Seattle, Washington, USA, July. Association for Computational Linguistics.

[3] **Naoki Otani**, Satoru Ozaki, Xingyuan Zhao, Yucen Li, Micaelah St Johns and Lori Levin. 2020. Pre-tokenization of Multi-word Expressions in Cross-lingual Word Embeddings. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 4451–4464, Online, November. Association for Computational Linguistics.

[4] **Naoki Otani** and Eduard Hovy. 2019. Toward Comprehensive Understanding of a Sentiment Based on Human Motives. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4672–4677, Florence, Italy, July. Association for Computational Linguistics.

[5] Ruochen Xu, Yiming Yang, **Naoki Otani**, and Yuexin Wu. 2018. Unsupervised Cross-lingual Transfer of Word Embedding Spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2465–2474, Brussels, Belgium, November. Association for Computational Linguistics.

[6] **Naoki Otani**, Hirokazu Kiyomaru, Daisuke Kawahara, and Sadao Kurohashi. 2018. Cross-lingual Knowledge Projection Using Machine Translation and Target-side Knowledge Base Completion. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1508–1520, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

[7] Aldrian Obaja Muis, **Naoki Otani**, Nidhi Vyas, Ruochen Xu, Yiming Yang, Teruko Mitamura and Eduard Hovy. 2018. Low-resource Cross-lingual Event Type Detection in Documents via Distant Supervision with Minimal Effort In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 70–82, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

[8] **Naoki Otani**, Toshiaki Nakazawa, Daisuke Kawahara, and Sadao Kurohashi. 2016. IRT-based Aggregation Model of Crowdsourced Pairwise Comparison for Evaluating Machine Translations. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 511–20, Austin, Texas, USA, November. Association for Computational Linguistics.

[9] **Naoki Otani**, Yukino Baba, and Hisashi Kashima. 2015. Quality control for crowdsourced hierarchical classification. In *Proceedings of 2015 IEEE International Conference on Data Mining*, pages 937–42, Atlantic City, New Jersey, USA, November. IEEE.

Workshop

[10] **Naoki Otani**, Daisuke Kawahara, Sadao Kurohashi, Nobuhiro Kaji, and Manabu Sassano. 2016. Large-Scale Acquisition of Commonsense Knowledge via a Quiz Game on a Dialogue System. In *Proceedings of Open Knowledge Base and Question Answering Workshop*, pages 11–20, Osaka, Japan, December. The COLING 2016 Organizing Committee.