Hideo Kobayashi

Website: https://hideodeo.github.io
Email: hideo.kobayashi@utdallas.edu
Work Authorization: F1 Visa
Phone: 469-660-8010

PROFILE

I'm a Ph.D. student, researching natural language processing (NLP) at Erik Jonsson School of Engineering and Computer Science at the University of Texas at Dallas. I'm advised by Vincent Ng and closely collaborate with Yufang Hou at IBM Research. My research interests have been centered around the tasks of language understanding at the discourse level, such as indirect anaphora resolution (bridging resolution) and coreference resolution, and applying them to downstream tasks.

ACADEMIC EXPERIENCE

Research with Vincent Ng at The University of Texas at Dallas August 2018 - Present

- Survey in Bridging Resolution (**COLING 2020**)
  Provide an overview of major milestones made in bridging resolution and future directions.
- Making Sense in Bridging Resolution (**NAACL 2021**)
  Provide suggestive evidence that rule-based and ML approaches are complementary rather than competing. Outperforms state-of-the-art results by 1.4-4.8% F-scores with a hybrid approach.
- Neural Anaphora Resolution in Dialogue (**Winner in CODI-CRAC Shared Task 2021, 2022**)
  Provide a multi-pass sieve approach consisting of neural and feature-based learning sieves.
- Large Dataset for Decontextualization (**Current Project**)
  Create a silver dataset for decontextualization using ChatGPT. Aim at improving T5-based decontextualizer and study its applications on tasks such as question answering.

Research with Yufang Hou at IBM Research, Ireland (collaboration) May 2021 - Present

- Constrained Multi-Task Learning for Bridging Anaphora Resolution (**ACL 2022**)
  Provide a multi-task learning framework with cross-task consistency constraints that guide the learning of the complex model. Outperforms state-of-the-art results by 1.4-4.6% F-scores.
- End-to-End Neural Bridging Resolution (**COLING 2022**)
  Evaluate resolvers in an end-to-end setting, strengthen them with better encoders, and gain a better understanding via perturbation experiments.
- PairSpanBERT: Enhanced Language Models for Bridging Resolution (**Current Project**)
  Pre-train a language model enhanced for bridging resolution by exploring the syntactic structure of noun phrases. Outperforms state-of-the-art results by 1.5-2.3% F-scores. [under review]

SELECTED PUBLICATIONS

- **Hideo Kobayashi**, Yufang Hou, Vincent Ng, "End-to-End Neural Bridging Resolution", International Conference on Computational Linguistics (COLING 2022). [slides][poster]
- **Hideo Kobayashi**, Yufang Hou, Vincent Ng, "Constrained Multi-Task Learning for Bridging Anaphora Resolution", Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022). [slides][poster]
- **Hideo Kobayashi***, Shengjie Li*, Vincent Ng, "Neural Anaphora Resolution in Dialogue", Workshop on Computational Approaches to Discourse & Workshop on Computational Models of Reference, Anaphora and Coreference (CODI-CRAC@EMNLP 2021). [slides]
- **Hideo Kobayashi**, Vincent Ng, "Bridging Resolution: Making Sense of the State of the Art", North American Chapter of the Association for Computational Linguistics (NAACL 2021). [slides][poster]
- **Hideo Kobayashi**, Vincent Ng, "Bridging Resolution: A Survey of the State of the Art", International Conference on Computational Linguistics (COLING 2020). [slides]

OTHER PUBLICATIONS

- **Hideo Kobayashi**, Christopher Malon, "Analyzing Coreference and Bridging in Product Reviews",

Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC@COLING 2022).
- Shengjie Li, **Hideo Kobayashi**, Vincent Ng, "Neural Anaphora Resolution in Dialogue Revisited", Workshop on Computational Approaches to Discourse & Workshop on Computational Models of Reference, Anaphora and Coreference (CODI-CRAC@COLING 2022).
- Shengjie Li*, **Hideo Kobayashi***, Vincent Ng, "The CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis Resolution in Dialogue: A Cross-Team Analysis", Workshop on Computational Approaches to Discourse & Workshop on Computational Models of Reference, Anaphora and Coreference (CODI-CRAC@EMNLP 2021). [slides]

PROFESSIONAL EXPERIENCE

Applied Scientist Internship at AWS AI, New York, NY — May 2023 - Aug 2023
- Mentor: Patrick Ng
- LLM for business intelligence.

Research Internship at NEC Laboratories America, Princeton, NJ — May 2022 - Aug 2022
- Mentor: Christopher Malon
- Crowdsourced dataset creation and analysis of coreference and bridging for natural language inference in online product reviews. (**CRAC@COLING 2022**)

Software Engineering Internship at Fixstars Corporation, Tokyo, Japan — Jan 2018 - Jul 2018
- Implemented fast parallel algorithms for continuous collision detection.

EDUCATION

The University of Texas at Dallas, Richardson, TX — August 2018 - December 2023 (expected)
Ph.D. in Computer Science. Advisor: Prof. Vincent Ng, GPA 3.74/4.00

Soka University, Tokyo, Japan — April 2015 - March 2017
Master of Science in Engineering. Advisor: Prof. Norihiko Shinomiya, GPA 4.00/4.00

Soka University, Tokyo, Japan — April 2012 - March 2015
Bachelor of Arts in Economics. GPA 3.90/4.00

AWARDS & HONORS

Best Teaching Assistant Award at The University of Texas at Dallas — 2022

Overall Winner in CODI-CRAC Shared Task 2021, 2022 — 2021, 2022
- Coreference task, bridging task, discourse deixis task

Ph.D. Fellowship from JASSO funded by Japanese govt. (acc. rate: 24%) — 2018-2021

SERVICE & COURSES

Service: ACL Rolling Review 2021, 2022

Teaching Assistant: Natural Language Processing, Machine Learning, Algorithm Design and Analysis

Selected Graduate Coursework: Natural Language Processing, Machine Learning, Database Design, Advanced Natural Language Processing, Artificial Intelligence, Convex Optimization, Theory of Computation.

REFERENCES

Vincent Ng, Professor at the University of Texas at Dallas
- vince@hlt.utdallas.edu. Relationship: Ph.D. advisor.

Yufang Hou, Research Scientist at IBM Research
- bnuxiaofang@gmail.com. Relationship: Research collaborator.

MISC.

Computer skills: Python, Java, TensorFlow, PyTorch, Git, Amazon Mechanical Turk

Natural Languages: English (business), Japanese (native)