Aaron Chun Hei LO

✉ aaronlo326@gmail.com
🌐 https://aaronlolo326.github.io/

SUMMARY

- NLP researcher and engineer with over 6 years of experience, ranging from parsing, distributional semantics to information retrieval and natural language generation
- Demonstrated abilities in publishing papers in top conferences as well as developing practical applications
- Team player and mindful independent learner

EDUCATION

Ph.D. (Research Area: Natural Language Processing) Aug 2019–Jan 2024
The Chinese University of Hong Kong (CUHK)
Thesis: *Learning Semantics from Meaning Representations: From Distributional and Graph-Grammatical Perspectives*

B.Sc. (Hons) with First Class Honors in Computer Science Sept 2015–July 2019
The Chinese University of Hong Kong

- **Dean's List**: Awarded for academic excellence in the Faculty of Engineering 2016, 2017, 2018, 2019
- **Master's List**: Awarded to the top student of each major of each class in Wu Yee Sun College 2017, 2018, 2019
- **ELITE Stream Student Scholarship**: Awarded to recognize excellence in advanced-level courses 2017, 2019
- **Computer Science Scholarship**: Awarded to top students admitted to the computer science program 2016

WORK EXPERIENCE

Junior Research Assistant July 2019
Department of Systems Engineering and Engineering Management, CUHK

- Conducted collaborative research on cross-framework meaning representations (MRs) parsing
- Developed a transition-based text-to-MRs parser
- Published a paper in CoNLL 2019 Shared Task[1]

Software Engineer Intern June 2018–Aug 2018
Set Sail Software

- Collaborated with digital marketing agencies in developing chatbots for multiple clients tailored to their needs
- Developed backends of chatbots using Node.js and Firebase Cloud Functions
- Created tools for automating performance analyses of chatbots

NLP/ML PROJECTS

Word to Function: Functional Distributional Semantics 2021–2023

- Studied Functional Distributional Semantics (FDS), a computational semantic framework that provides a truth-conditional interpretation over distributional representations of words
- Enhanced linguistic preciseness and computational efficiency of FDS for complex sentence structures
- Developed a variational autoencoder from scratch using PyTorch sped up with distributed data parallelism
- Outperformed >20 models on the tasks of semantic composition and verb disambiguation, including BERT that uses $12\times$ more data and $2\times$ more parameters
- Revealed that FDS captures hypernymy if trained on a corpus that follows the Distributional Inclusion Hypothesis (DIH) and proposed an alternative training objective that improves hypernymy detection from corpora
- Published papers at **SEM 2023*[4] and *ACL 2024*[5], and was invited to present at the 19th DELPH-IN Summit

Task-Oriented Dialogue System with Unstructured Knowledge Access 2021

- Collaboratively developed a dialogue system with the CUHK team that performs RAG of natural language response based on dialogue contexts
- Coordinated a sub-team that works on ranked retrieval from an FAQ knowledge base, where we trained a BERT model that ranks the relevance of knowledge according to users' queries and dialog contexts using Huggingface and PyTorch
- Participated in the Ninth Dialog System Technology Challenge (DSTC9) and our team ranked 12 out of 24[2]

Text Generation via Semantic Graph Parsing 2019–2021

- Demonstrated the use of a synchronous graph grammar in approximating the syntax–semantics interface of English
- Developed a probabilistic graph parser from scratch that reconstructs syntactic derivations from semantic graphs, with devised adaptations that improve accuracy, efficiency, and coverage of graph parsing
- Achieved better graph-to-text translation than a neural sequence-to-sequence method under out-of-domain settings, while additionally providing derivations for syntactic disambiguation
- Published in *ACL 2022*[3], and was invited to present at the Foundations of Language Processing of Umeå University

CV–JD Recommendation System 2019

- Researched hypernymy detection and automatic taxonomy induction from corpus using distributional and pattern-based approaches
- Wrote a web crawler to scrape over 40,000 publicly available CVs and 20,000 job descriptions (JDs)
- Devised a CV–JD matching algorithm using doc2vec and Latent Dirichlet Allocation (LDA)

INVITED TALKS

Functional Distributional Semantics (FDS) at Scale and *Probing for Hypernymy in FDS*, 19th DELPH-IN Summit, Language and Information Society of University of A Coruña   27 June 2023

Semantic Composition with PSHRG for Derivation Tree Reconstruction from Graph-Based Meaning Representations, Seminar at Foundations of Language Processing of Umeå University, Virtual   16 Sept 2022

ADDITIONAL EXPERIENCE

Teaching Assistant, Faculty of Engineering, CUHK
- CSCI2100: Data Structures   2020–2023
- SEEM3550: Fundamentals in Information Systems   2021–2023

Resident Tutor, Wen Lin Tang, Chung Chi College, CUHK   2021–2023
- Provided pastoral care and any necessary support to undergraduate residents, including check-in/out, orientation to new residents, integration of international students, and conflict resolution
- Served as an intermediary between the undergraduate residents, the warden, the Student Residents' Association the hostel staff, and the Student Hostels Committee of the College
- Reviewed and assisted the Student Residents' Association with the annual financial report

CORE SKILLS

Natural Languages Cantonese (*native*), English (*proficient*), Mandarin (*proficient*)

Programming Languages Python, C, SQL

ML/DS Libraries PyTorch, Tensorflow, scikit-learn, Pandas, Numpy

NLP Libraries PyDelphin, Hugging Face, NLTK, Spacy, Gensim, WordNet,

Others Git, Google Cloud

PUBLICATIONS AND PREPRINTS

[5] Chun Hei Lo, Wai Lam, Hong Cheng, and Guy Emerson. 2024. Distributional Inclusion Hypothesis and Quantifications: Probing for Hypernymy in Functional Distributional Semantics. To appear in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Bangkok, Thailand

[4] Chun Hei Lo, Hong Cheng, Wai Lam, and Guy Emerson. 2023. Functional Distributional Semantics at Scale. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 423–436, Toronto, Canada

[3] Chun Hei Lo, Wai Lam, and Hong Cheng. 2022. Semantic Composition with PSHRG for Derivation Tree Reconstruction from Graph-Based Meaning Representations. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5425–5439, Dublin, Ireland

[2] Mudit Chaudhary, Borislav Dzodzo, Sida Huang, Chun Hei Lo, Mingzhi Lyu, Lun Yiu Nie, Jinbo Xing, Tianhua Zhang, Xiaoying Zhang, Jingyan Zhou, Hong Cheng, Wai Lam, and Helen Meng. 2021. Unstructured Knowledge Access in Task-oriented Dialog Modeling using Language Inference, Knowledge Retrieval and Knowledge-Integrative Response Generation. arXiv:2101.06066

[1] Sunny Lai, Chun Hei Lo, Kwong Sak Leung, and Yee Leung. 2019. CUHK at MRP 2019: Transition-Based Parser with Cross-Framework Variable-Arity Resolve Action. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning*, pages 104–113, Hong Kong