Aaron Chun Hei LO

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

Summary

NLP/ML researcher and engineer with over 6 years of experience, ranging from information retrieval and large language models (LLMs) to parsing and distributional semantics. Demonstrated abilities in publishing award-winning papers in top conferences as well as developing NLP systems.

Education

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

- **Area Chair Award**: Awarded to 21 out of 1742 papers accepted to ACL 2024 2024

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
- **Admission Scholarship**: Awarded to high achievers newly admitted to the engineering program 2015

Work Experience

Senior Algorithm Engineer (NLP/LLM Team) Mar 2024–Present
TCL Corporate Research (Hong Kong) Co., Limited

- Design, develop and deploy chatbots based on retrieval-augmented generation (RAG) using open-source LLMs that work on domain-specific bilingual (Chinese and English) data
- Devise a hybrid retrieval strategy and implemented the data pipeline which combines vectorial and keyword-based queries using MySQL, Milvus, and Elasticsearch, thereby replacing team's previously developed method
- Tailor summarization algorithms for clients for more structured and explainable outputs than pure-LLM approaches
- Apply large vision-language models for parsing PDF documents to structured data

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

- Researched and developed a transition-based parser for meaning representation graphs

Software Engineer Intern June 2018–Aug 2018
Set Sail Software

- Developed the backends of chatbots and performance analyses tools using Node.js and Firebase Cloud Functions

Selected NLP/ML Projects

Word Representation Learning (*SEM 2023; ACL 2024 (Area Chair Award)) 2021–2023

- Developed a VAE from scratch for word representation learning using PyTorch with distributed data parallelism
- Outperformed >20 models on the tasks of semantic composition and verb disambiguation, including BERT which uses 12× more data and 2× more parameters
- Hypothesized and confirmed lexical entailment can be learnt by our models from restricted classes of corpora

Dialogue System with Unstructured Knowledge Access (DSTC9 @ AAAI 2021) 2021

- Led a team that worked on ranked retrieval from an FAQ knowledge base, where we supervised–fine-tuned a BERT model as a reranker 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 (ACL 2022) 2019–2021

- 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

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

Core Skills

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

Programming Languages Python, C, SQL

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

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

Others Git, Docker, Google Cloud

Selected Publications

[3] Chun Hei Lo, Wai Lam, Hong Cheng, and Guy Emerson. 2024. Distributional Inclusion Hypothesis and Quantifications: Probing for Hypernymy in Functional Distributional Semantics. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14625–14637, Bangkok, Thailand. **(Oral and Poster; Area Chair Award)**

[2] 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. **(Oral and Poster)**

[1] 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. **(Oral and Poster)**