

Haofei Yu

5704 Darlington Road, Pittsburgh, PA 15217

☎ (+1) (412)-537-0110 | ✉ haofei@cs.cmu.edu | 🌐 www.haofeiyu.me | 📷 lwaekfjlk

Education

Carnegie Mellon University, Language Technologies Institute

Pittsburgh, US

MASTER OF SCIENCE IN INTELLIGENT INFORMATION SYSTEMS

Aug. 2022 - May. 2024 (Expected)

- Enrolled and Seeking for Research Advisor

Zhejiang University, ChuKochen Honors College

Hangzhou, PRC

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND TECHNOLOGY

Sep. 2018 - Jul. 2022

- Overall GPA: 3.96/4.00 | Ranking: 7/134
- Provincial Scholarship due to excellent academic performance (2018-2019)

Research Experiences

Encoder as Partial Re-ranker for Open-Domain Question Answering

Hangzhou, PRC

WESTLAKE UNIVERSITY, SUPERVISED BY PROF. YUE ZHANG

Sep. 2021 - Feb. 2022

- Considered the encoder part in the Fusion-in-Decoder architecture to be partially responsible for re-ranking and enhanced its re-ranking ability by modifying the training objective
- Conceptually modeled a new latent variable to describe the correlation between retrieved candidate passages and answers and used labeled golden passages as supervision signals for training
- Achieved more than 1.5 points EM improvements on both NaturalQuestions and TriviaQA datasets

Empirical Study on Personalized Dialog Generation

Atlanta, US (online)

GEORGIA INSTITUTE OF TECHNOLOGY, SUPERVISED BY PROF. DIYI YANG

Jul. 2021 - Sep. 2021

- Reviewed personalized response generation papers from speaker model (using persona embedding) to P^2 Bot (using RL methods to model mutual persona perception).
- Utilized tensor factorization to explicitly learn the latent representation of users' embedding in order to tackle the data sparsity problem of explicit and high-quality persona or demographic information.
- Proposed to add user-user-generation signals to the existed Transformer-based Dialog Generation framework to achieve mutual persona perception in the dialog.

Uni-Encoder: A Fast and Accurate Response Selection Paradigm for Generation-Based Dialogue Systems

Hangzhou, PRC

WESTLAKE UNIVERSITY, SUPERVISED BY PROF. ZHENZHONG LAN

Mar. 2021 - Jul. 2021

- Developed a new paradigm called *Uni-Encoder*, that keeps the full attention over each pair as in *Cross-Encoder* while only encoding the context once, as in *Poly-Encoder*
- Designed in-batch negatives mechanism that makes other responses in one batch as negative samples for one history-response pair
- Reached state-of-the-art performance on 4 benchmark datasets (Ubuntu v1, Ubuntu v2, ConvAI, and Douban)

Internship & Work Experiences

Tencent AI Lab, NLP Center

Shenzhen, PRC

RESEARCH INTERN, SUPERVISED BY DR. WEI BI

Feb. 2022 - Jul. 2022

- Responsible for a research project related to Long-range Language Modeling
- Proposed an unified framework for Long-range Language Modeling which considers Transformer-XL, Routing Transformer, ExpireSpan and other long-range transformer variants to be special cases
- Made modifications based on K-means clustering and improved Transformer-XL to become a general form
- Achieved 0.3 Perplexity drop on Wikitext-103 dataset with a constant extra time cost

AWARDS & EXTRA CURRICULAR ACTIVITIES

2020 **MCM/ICM Interdisciplinary Contest in Modeling**, Honorable Mention

Hangzhou, PRC

2019 **Micrisoft Student Club (MSC)**, Member of PG Group

Hangzhou, PRC