Peixuan Han

+86 18910493775   hanpx20@mails.tsinghua.edu.cn

EDUCATION

Tsinghua University   Sep 2020 - Jun 2024

Department of Computer Science and Technology

Attended college at 15.

Overall GPA: 3.91/4.0

Programming Skills: Proficient in C++, Python, SQL, Pytorch, Huggingface Transformers.

Highlighted Courses: Object-Oriented Programming, Software Engineering, Data Structure, Introduction to Artificial Intelligence, Artificial Neural Networks, Operating System, Theory of Computer Network, Computer Architecture.

Awards: Won Scholarship for Academic Excellence three times.

PUBLICATIONS & PREPRINTS

· *Chinese People's Passage of Time Judgment During COVID-19 and Its Association with Emotion: Evidence from the Weibo Dataset*: submitted to American Psychologist, under review.

· *Distributionally Robust Unsupervised Dense Retrieval Optimization on Web Graphs*: submitted to The Web Conference 2024, under review. Code of this work can be found at https://github.com/Hanpx20/GroupDRO_Dense_Retrieval.

RESEARCH EXPERIENCE

Distributionally Robust Unsupervised Dense Retrieval Optimization on Web Graphs   Apr 2023 - Oct 2023

· **Mentor**: Professor Chenyan Xiong. Part of this work was done during an internship at Carnegie Mellon University.

· **Objective:** Propose a group-level data clustering and reweighting strategy for retrieval model training. Train a model that achieves state-of-the-art in unsupervised setting.

· Designed the overall algorithm: train an embedding model on the web graph for clustering and use group distributionally robust optimization during training to update group weights.

· Implemented the code for data extraction and filtering, cluster generating, data loading with clusters, dynamic group weights updating as well as evaluation on MsMarco and BEIR. The model outperformed unsupervised baseline for 1.2%.

Training Bilingual Large Language Models   Jan 2023 - Sep 2023

· **Mentor**: Professor Maosong Sun.

· **Objective**: Train a conversational large language model in both Chinese and English.

· Implemented the code(based on Megatron-LM) for inference and evalution on several datasets: MMLU(knowledge), C-Eval(knowledge in chinese), GSM8k(maths ability) and HELM-Reasoning.

· Chain of thought and majority voting are also implemented and used for training and inference.

· Launched a 30B language model named *Instruct-DLM-v2*. The model achieves an accuracy of 56.8% on C-Eval.

Causal Relationship between Individual's Sentiment and Perception of Time   May 2022 - Oct 2022

· **Mentor**: Professor Min Zhang.

· **Objective**: Discover the causal relationship between people's sentiments, time expectation, and their time perception.

· Used a neural network to classify sentiments. The model is jointly trained on three-way classification of valence and six-way classification of detailed sentiments.

· data augmentation by synonym replacement and modified weight of loss function are applied.

HONORS & AWARDS

National Olympiad in Informatics in Provinces, First Prize	2018.11
Asia-Pacific Informatics Olympiad, First Prize	2019.5
Scholarship for Academic Excellence	2021.9, 2022.9, 2023.10
Scholarship for Social Work	2022.9