Xiaomeng Hu

Shenyang, China

greghxm@foxmail.com 18540310722

https://gregxmhu.github.io/ https://github.com/GregxmHu

Education Experience

Northeastern University **Shenyang, China**

Artificial Intelligence, overall GPA: 91.6/100, rank: 1/61 2019-2023(expected)

coursework: Data Structure/Algorithm, Optimization Methods, Machine Learning, Discrete Mathematics, Probability Theory, Calculus, Linear Algebra

Activities: Machine Learning Club in NEU, Math-Fans Club in NEU

Research Experience

Tsinghua NLP Lab **Beijing, China**

Research Intern Aug 2021 - Present

- Research Area: Natural Language Processing, Information Retrieval, and their intersections.
- Achievements: A short paper has been submitted to SIGIR 2022 (under review), two derived open source projects has been released on Github.
- Work Overview: We applied prompt-based learning method to the reranking phase of the information retrieval pipelines. According to the experimental results, we conclude that there are two gaps in the matching of pre-trained language models(PLM) and downstream reranking task: training schema gap and task knowledge gap. Through a series of comparative experiments and ablation experiments, we can find that the training schema gap comes from the different optimization goals in the pre-training stage and the adapting(to reranking task) phase, and the task knowledge gap comes from the lack of knowledge related to ranking task in the pre-training stage. We proposed a new neural ranker to alleviate the two gaps. The training schema gap can be mitigated through prompting, and by injecting the downstream task-related knowledge into the PLM through pre-finetuning, the task knowledge gap can be significantly alleviated. Our model achieves the PLM's efficient matching to downstream reranking task, especially in few-shot scenarios.

Publications and Preprints

* indicates equal contribution.

- *P^3 Ranker: Mitigating the Gaps between Pre-training and Ranking Fine-tuning with Prompt-based Learning and Pre-finetuning*, in **SIGIR 2022**(under review)

  Xiaomeng Hu*, Shi Yu*, Chenyan Xiong, Zhenghao Liu, Zhiyuan Liu, Ge Yu

Open-Source Projects

P^3 Ranker Aug 2021–Jan 2022

- - The project is based on the **OpenMatch** toolkit. We proposed a two-stage ranker for the reranking task of Neural Information Retrieval in the project. This model is based on a pre-trained text-to-text language model (T5). In the first stage, the language model is fine-tuned on extra supervised tasks to acquire the knowledge required for the next stage. In the second stage , the prompting methods grounding with the model pre-finetuned in the first stage will be used to complete the reranking. Both the discrete prompt and the continuous prompt is supported. Hope it will provide significant convenience to the future researchers.

Multi-task trainer Nov 2021–Dec 2021

– - This is a unified prompt-based training framework for multi-task(NLP tasks) learning. The framework is grounding with the text-to-text language models: T5. This framework can fit every NLP tasks as long as it's dataset can be converted into text-to-text format. Text from different tasks uses different person-specified prompts.

🏆 Honors and Awards (Selected)

Outstanding Student pacesetter(top 1%)	Oct 2021
First Class Scholarship(top 3%)	Oct 2021
First Prize in The Chinese Mathematics Competitions.	Dec 2020