Feiyu Duan

duanfeiyu@buaa.edu.cn · (+86) 180-8056-2978 · simon_dfy31415926

Education

Beihang University, Beijing, China 2022 – Present

Master student in Computer Science
Advisor: Wenge Rong

Beihang University, Beijing, China 2018 – 2022

B.S. at Sino-French Engineering School
GPA: 3.8/4 (Top 5%)

Publication

[1] PositionID: LLMs can Control Lengths, Copy and Paste with Explicit Positional Awareness

Zekun Wang*, **Feiyu Duan***, Yibo Zhang*, Wangchunshu Zhou, Ke Xu, Wenhao Huang, Jie Fu
Findings of EMNLP 2024

[2] D-CPT Law: Domain-specific Continual Pre-Training Scaling Law for Large Language Models

Haoran Que*, Jiaheng Liu*, Ge Zhang*, Chenchen Zhang, Xingwei Qu, Yinghao Ma, **Feiyu Duan**, Zhiqi Bai, Jiakai Wang, Yuanxing Zhang, Xu Tan, Jie Fu, Wenbo Su, Jiamang Wang, Lin Qu, Bo Zheng
NeurIPS 2024

[3] Generative Spoken Language Modeling with Quantized Feature Enhancement

Feiyu Duan, Chen Li, Keheng Wang, Si Wu, Chuantao Yin, Wenge Rong
IJCNN 2024

Research Project

[4] HelloBench: Evaluating Long Text Generation Capabilities of Large Language Models

Haoran Que*, **Feiyu Duan***, Liqun He*, Yutao Mou, Wangchunshu Zhou, Jiaheng Liu, Wenge Rong, Zekun Moore Wang, Jian Yang, Ge Zhang, Junran Peng, Zhaoxiang Zhang, Songyang Zhang, Kai Chen
Under Review at ICLR 2025

[5] LLMs Know What They Need: Leveraging a Missing Information Guided Framework to Empower Retrieval-Augmented Generation

Keheng Wang, **Feiyu Duan**, PeiguangLi, Sirui Wang, Xunliang Cai
Under Review at COLING 2025

[6] FIRE: Flexible Integration of Data Quality Ratings for Effective Pre-Training

Xu Liangyu*, Xuemiao Zhang*, **Feiyu Duan***, Sirui Wang, Xunliang Cai
Under Review at AAAI 2025

[7] Enhancing LLMs via High-Knowledge Data Selection

Feiyu Duan*, Xuemiao Zhang*, Sirui Wang, Haoran Que, Yuqi Liu, Wenge Rong, Xunliang Cai
Under Review at AAAI 2025

[8] Knowledge-Driven CoT: Exploring Faithful Reasoning in LLMs for Knowledge-intensive Question Answering

Keheng Wang*, **Feiyu Duan***, Sirui Wang, Peiguang Li, Yunsen Xian, Chuantao Yin, Wenge Rong, Zhang Xiong
Under Review at IEEE Transactions on Emerging Topics in Computational Intelligence

[9] Language Models Can Remember: Improving Memory and Knowledge Retrieval Ability for Closed-book QA

Feiyu Duan, Keheng Wang, Rumei Li, Sirui Wang, Chuantao Yin, Wenge Rong
Under Review at Neurocomputing

Internship Experience

Meituan, Beijing June 2023 – August 2024

LLM Algorithm Intern Basic Research and Development Platform

Application of CoT in Gastronomy Knowledge Graph Construction June 2023 – September 2023

This project aim to explore whether LLMs could be prompted to use their reasoning abilities to contribute to the construction of a gastronomy knowledge graph.

- Design model reasoning prompts
- Distill reasoning data from ChatGPT to construct gastronomy-specific dataset
- Conduct supervised fine-tuning (SFT) on 7B-scale models, exploring models such as ChatGLM2-6B and InternLM-7B

Construction of Encyclopedia Knowledge Evaluation System September 2023 – November 2023

The goal was to build an evaluation system capable of comprehensively assessing the factual knowledge memory abilities of LLMs, covering a variety of encyclopedic knowledge.

- Encyclopedia knowledge passage cleaning
- Encyclopedia knowledge question generation
- Extract entities to construct subgraphs and generate complex questions

High-Knowledge Data Selecetion Algorithm Research November 2023 – May 2024

Construct a global knowledge repository with high coverage, guiding the construction of knowledge-enhanced data. This knowledge repository is categorized into different domains to enable domain-specific knowledge enhancement and long-tail knowledge enhancement.

- Extraction of knowledge elements from Wikipedia passages and used GPT-4 to label knowledge categories, building a knowledge repository.
- Annotate samples in CC, BaiduBaike, and other pre-training datasets with the knowledge repository.
- Conducted 100B annealing experiments on a 7B model, achieving around 1.16 pp improvement on knowledge-intensive tasks like MMLU

Multi-Rater Fusion Algorithms Research May 2024 – August 2024

Directly applying GPT-4 as scorer based on multiple rules is not ideal due to its inability to perfectly follow each rating dimension. Therefore, we explore multi-rater fusion methods to select data from diverse perspectives.

- Fit win rates based on knowledge scorer and quality scorer and adopt a simple multiplicative fusion
- Experiments on an 8B model + 2T training data, achieving improvements in knowledge-intensive task compared to direct quality scorer selection.

♡ Honors and Awards

- Received the First-Class Academic Scholarship for Master's Students at Beihang University in 2023.
- Received the Freshman Scholarship for Master's Students at Beihang University in 2022.
- Awarded the title of Outstanding Graduate at Beihang University in 2022.
- Won the First Prize in the International Track of the Feng Ru Cup Science and Technology Competition at Beihang University in 2021, with the project "Wearable Devices for Gesture Recognition."
- Received the China National Scholarship for the academic year 2018-2019.

Others

- English - Fluent (CET-4: Score 617, CET-6: Score 568)
- French - Fluent (DELF B2)