Wenzheng Zhang

Email: vincentwzustc@gmail.com | Tel: +1 732 790 6623 | Homepage: wenzhengzhang.github.io

Education

Rutgers University, PhD in Computer Science	Sept 2021 – Present
Rutgers University, MS in Computer Science	Jan 2020 - June 2021
University of Science and Technology of China, BS in Applied Physics	Sept 2015 - May 2019

Industry Experience

Microsoft　　　　　　　　　　　　　　　　　　　　　　　　　　　June 2022 – Aug 2022
Applied Scientist Intern
Hosts: Prof. Chenyan Xiong and Arnold Overwijk

- Designed a novel multitask learning framework to enhance task specialization in multitask information retrieval.
- Leveraged optimized pretrained models, compatible prompting, and a novel adaptive learning method to ensure parameter specialization for individual tasks.
- Achieved state-of-the-art performance, surpassing task-specific retrievers on the KILT benchmark. Results published in **TACL 2023**.

Meta　　　　　　　　　　　　　　　　　　　　　　　　　　　　　May 2024 – Dec 2024
Research Scientist Intern at FAIR
Hosts: Mingda Chen, Victoria Lin and Scott Yih

- Designed a novel model architecture for a Retrieval-Augmented Generation (RAG) system based on the Llama-3 model.
- Utilized the attention mechanism of the large language model (LLM) for retrieval and enhanced it to improve retrieval performance.
- Unified the retriever and generator components within a single LLM for RAG system.
- Explored various key-value state compression techniques to reduce disk usage and accelerate inference.
- Achieved strong performance on knowledge-intensive tasks.

Publications

Seq2seq is All You Need for Coreference Resolution	EMNLP 2023
Wenzheng Zhang, Sam Wiseman, Karl Stratos	
Improving Multitask Retrieval by Promoting Task Specialization	TACL 2023
Wenzheng Zhang, Chenyan Xiong, Karl Stratos, Arnold Overwijk	
EntQA: Entity Linking as Question Answering.	ICLR 2022
Wenzheng Zhang, Wenyue Hua, Karl Stratos	
Understanding Hard Negatives in Noise Contrastive Estimation	NAACL 2021
Wenzheng Zhang, Karl Stratos	

Honors and Awards

ICLR Spotlight Paper	2022
Outstanding Publications Award (Rutgers University)	2021

Professional Services

Reviewer of EMNLP 2023	2023
Reviewer of NAACL 2021	2021

Research Interests

Natural Language Processing (NLP), Information Retrieval, Retrieval Augmented Generation (RAG), Large Language Models (LLMs)

Code repositories at GitHub

Teaching Experience

Teaching Assistant for Natural Language Processing(CS533), Rutgers University	Spring 2023
Teaching Assistant for Natural Language Processing(CS533), Rutgers University	Spring 2022
Teaching Assistant for Machine Learning (CS461), Rutgers University	Fall 2021

Skills and Techniques

Programming Languages: Python, Java, R, C

Skills: Pytorch, Huggingface, Git, Bash, Latex

Relevant Course Work: Natural Language Processing, Machine Learning, Computer Vision, Probability and Statistics, Artificial Intelligence, Algorithms