Zheng Huang

Phone: 434-328-0876 · Email: zheng.huang.gr@dartmouth.edu · Homepage · Google Scholar

Education

Dartmouth College *Sept 2023 - Present*
PhD of Computer Science, advisor: Prof. Yujun Yan *Hanover, NH, United States*

University of Virginia *Aug 2019 - Dec 2021*
Master of Computer Science, advisor: Prof. Jundong Li *Charlottesville, VA, United States*

University of British Columbia *June 2018 - August 2018*
UBC Visiting Student, Department of Computer Engineering *Vancouver, BC, Canada*

Hebei University of Technology *Sept 2015 - June 2019*
Bachelor of Computer Science, major GPA: 87/100 *Tianjin, China*

Research Interest

I am interested in **Machine Learning (ML)** and **Large Language Models (LLMs)**, including:

- **Graph ML**: Designing frameworks for diverse graph data to enhance model generalizability and applying graph ML models to real-world applications, such as recommender systems
- **Multimodal LLMs**: Developing multimodal LLM frameworks to bridge neuroscience and computer vision & natural language processing to understand how the human brain processes and represents knowledge
- **Federated Learning (FL)**: Building FL systems to make life-long learning possible and mitigate privacy risks & training costs resulting from traditional, centralized ML

Selected Publication

- **Zheng Huang**, Qihui Yang, Dawei Zhou, Yujun Yan, "Enhancing Size Generalization in Graph Neural Networks through Disentangled Representation Learning ", ICML 2024, Arxiv
  - Researched the generalization of Graph Neural Networks (GNNs) through disentangled representation learning
  - Proposed a novel and model-agnostic framework designed to disentangle size factors from graph representations
  - Employed size- and task-invariant augmentations, introducing a decoupling loss to minimize shared information in hidden representations
  - Conducted in-depth research on out-of-distribution generalization, explainable GNN models and disentangled representation learning
- **Zheng Huang**, Jing Ma, Yushun Dong, Natasha Zhang Foutz and Jundong Li, "Empowering Next POI Recommendation with Multi-Relational Modeling", SIGIR 2022, Arxiv
  - Studied on Points of Interests (POI) recommendation by capturing the influence of multiple relations
  - Utilized multiple Graph Convolutional Networks (GCNs) with Self-Attention mechanism to capture multiple user-user social relations (family or colleague) and user-location check-in relations
  - Adopted coupled Recurrent Neural Networks (RNNs) to capture the mutual influence between users and POIs, improving recommendation performance
  - Conducted in-depth research on recommender systems, sequential recommendation and GCNs
- Full Publication List (Google Scholar)

Industry Experience

Alexa Speech Recognition, Amazon.com, Inc. *Mar 2022 - Jan 2023*

Machine Learning Engineer *Seattle, WA, United States*

- Developed Federated Learning (FL) systems to preserve users' privacy and improve the quality of speech recognition
- Worked on a team and delivered an on-device FL Recurrent Neural Network Transducer prototype that is capable of learning from the audio without relying on sending users voice recordings to the cloud
- Implemented FL on-device trainer that was constructed during the device idle time and deconstructed after the completion of training tasks to minimize the footprint of training in memory to avoid customer friction
- Conducted in-depth investigation on decentralized machine learning, privacy-preserving and AWS infrastructure

Research Experience

Neural Decoding of Visual Stimuli with Multimodal LLMs *Jan 2024 - Present*

Research Assistant, advised by Professor Yujun Yan

- Designed a multimodal LLM framework for reconstructing visual stimuli (images) from functional MRI (fMRI) data, bridging neuroscience and computer vision
- Developing a training-free framework that integrates a diffusion model with an LLM for enhanced prompt understanding through in-context learning, enabling the generation of hierarchical descriptions of visual stimuli
- Implementing a graph-based architecture that maps whole-brain fMRI patterns to Language model latent space, enabling hierarchical semantic interpretation of neural responses and reconstructing the viewed images through diffusion models
- Conducting in-depth research on multimodal LLMs, diffusion models, and neuroscience

Technical Skills

- **Programming**: Python, Java, C++, C, R, Javascript, HTML, CSS, SQL, MATLAB
- **Tools**: PyTorch, TensorFlow, Scikit-learn, PySpark, AWS, Linux, Numpy, Pandas, Latex

Services

- **Conference reviewer:** ICLR 2024, NeurIPS 2024, ICLR 2023, NeurIPS 2024, ICLR 2023, JMLR 2022, ECML 2022, PAKDD 2021, WSDM 2021
- **Industry reviewer:** Amazon Machine Learning Conference 2022

Awards

- ICML Travel Grant *2024*
- SIGIR Student Grant *2022*
- Compuster Science Department 2nd-level Academic Fellowship *2018*
- Compuster Science Department 1st-level Academic Fellowship *2017*
- Compuster Science Department 2nd-level Academic Fellowship *2016*

Teaching

Teaching Assistant, Network Science and Complex Systems, Winter 2023