Donghong Cai

@ donghongcai98@gmail.com
https://ilikevegetable.github.io

Education

Washington University in St. Louis — St. Louis, Missouri
Ph.D. in Computer Science — *Aug 2024 – present*
- **Advisor**: Dr. Yixin Chen

Zhejiang University — Hangzhou, China
M.S. in Software Engineering; ***GPA: 3.84/4.00*** — *Sep 2020 – Mar 2023*
- **Research Directions:** Time Series and Graph Data Mining
- **Advisor:** Dr. Yang Yang

Huazhong University of Science and Technology — Wuhan, China
B.S. in Software Engineering; ***GPA: 3.67/4.00*** — *Sep 2016 – Jun 2020*

Professional Experience

Alibaba Group — Shenzhen, China
Machine Learning Engineer*, International E-commerce Seller Team* — *Apr 2023 – Jul 2024*
- Conducted research and developed multiple recommendation algorithms for sellers' promotional tools, including Flexi Combo (buy more, save more), New Buyer Voucher, Regular Voucher, etc. These algorithms boosted stores' GMV and Units Ordered by an average of **8.28%** and **6.55%**.
- Designed a novel same-store identification algorithm to search potential identical sellers on competing e-commerce platform, resulting in a cost savings of **$73,000** for the BD team during the 3rd quarter of 2023.
- Presented a transformer-based time series forecasting model for predicting monthly GMV and order volume for stores on an e-commerce platform. Performance increased by **21.7%** and **36.8%** respectively, compared with the experience-based manual settings.
- Based on prefix-tuning approach, performed data cleaning and designed a model on an intent classification dataset with over 350 classes across six languages, achieving a classification accuracy of **90%**↑.
- Built a seller chatbot for an e-commerce platform using an LLM Agent, incorporating a two-step retrieval method. This involved QA-pair matching in common query and Retrieval-Augmented Generation (RAG) in mass multimodal data which are database (product, order, etc.) and documents.
- Designed a framework to extract the hot questions by customers on an e-commerce platform. Converted customer service call recordings to text, anonymized, and processed with an LLM to extract keywords. A custom clustering method then groups and ranks hot topics.

Zhejiang University — Hangzhou, China
Graduate Research Assistant*, AINet Lab* — *Mar 2021 – Sep 2022*
- Conducted research in the fields of **time series classification**, **self-supervised learning**, and **domain adaptation** on multi-channel brain signals data, resulting in a conference paper in **KDD'23** research track.
- Researched on a pre-training framework that more effectively integrates and aligns **image**, **text**, and **graph** modalities in multi-modal learning, resulting in a paper for submission.
- Built an experimental pipeline for studying large-scale SEEG brain signal data, implemented the integration of various specialized models (classification model, pre-training model, and domain generalization model), and used it for clinical trial applications of epilepsy detection.

Publications

1. **Donghong Cai**, Junru Chen, Yang Yang, Teng Liu, and Yafeng Li. 2023. MBrain: A Multi-channel Self-Supervised Learning Framework for Brain Signals. *In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23).*

Research Experience

Epileptic Wave Detection Using Hierarchical Graph Diffusion Learning Zhejiang University

Graduate Research Assistant *Mar 2021 – May 2021*

- Formulated and studied the epileptic wave detection problem for SEEG data using an automatic end-to-end data-driven method.
- Proposed ***BrainNet*** to simultaneously learn the dynamic diffusion graphs and model the brain wave diffusion patterns thereon in a hierarchical fashion to achieve accurate epileptic wave detection under conditions of imbalanced labels and severe noise.
- Conducted comprehensive experiments on a large-scale real-world SEEG dataset across multiple patients. The experimental results validated the effectiveness of *BrainNet* on epileptic wave detection and its superiority in capturing the diffusion process. In the channel-level epileptic wave detection task, *BrainNet* outperforms all baselines on F_2-score with an increase of **36.66%**.

Pre-training Framework for Brain Signals Zhejiang University

Graduate Research Assistant *Jun 2021 – Feb 2022*

- Designed a generalized self-supervised learning framework ***MBrain*** consisting of three well-designed tasks, which can be applied to pre-train both EEG and SEEG brain signals.
- Proposed multi-channel CPC (Contrastive Predictive Coding) and theoretically proved that optimizing the goal of multi-channel CPC can lead to a better predictive representation. Based on the multi-channel CPC, three self-supervised learning tasks were designed to explicitly capture the spatial and temporal correlations of brain signals to learn informative representations for downstream tasks.
- Validated the superior effectiveness and clinical value of the proposed framework through extensive experiments of seizure detection on large-scale real-world EEG and SEEG datasets. In subject independent seizure detection experiment which meet practical clinical needs, *MBrain* outperforms all baselines on F_2-score with an increase of **9.23%** and **27.83%** in EEG and SEEG datasets.

Multimodal Learning with Graph Alignment Zhejiang University

Graduate Research Assistant *Mar 2022 – Sep 2022*

- Introduced the graph modality into the realm of multi-modal fusion and proposed a graph alignment task to effectively combine multiple modalities.
- Proposed **MMGA** (**M**ulti-**M**odal learning with **G**raph **A**lignment), an innovative pre-training framework designed to unify information from graph (social network), image, and text modalities on social media platforms to enhance user representation learning.
- Constructed the first multi-modal social media dataset containing image, text, and graph modalities, including over 2 million posts, a million-scale graph, and labeled user/post tags.

Honors & Awards

Award of Honor for Graduate	*Dec 2021*
Outstanding Graduate of Huazhong University of Science and Technology	*Jun 2020*
Academic Excellence Scholarship	*Sep 2018*
Individual Scholarship	*Sep 2018, Sep 2019*

Languages & Skills

Languages: Chinese (Native), English (Fluent)

Programming: C, C++, Java, JavaScript, Python, MATLAB, MySQL

Professional Skills: SKlearn, PyTorch, Markdown, LaTeX, Shell Script

Interests: Basketball, Football, Travelling, Photography