Liwen Sun

Website: dominicslw.github.io Email: liwens@andrew.cmu.edu Addr: 4742 Centre Avenue

EDUCATION

Carnegie Mellon University *Aug. 2023–Present*

Master of Science in Intelligent Information Systems (Language Technologies, Computer Science)

Research Interest: Text Mining, Graph Mining, Multimodal Information Retrieval

University of Illinois at Urbana-Champaign *Aug. 2020–May 2023*

Bachelor of Science in Computer Science and Mathematics GPA: 4.0/4.0

Honor: Summa Cum Laude and Bronze Tablet ***(highest undergraduate honor***, top 3% in college, final year), James Scholar (top 5% in department, every semester), Dean's List (top 5% in college, every semester)

PUBLICATION

1. *ED-Copilot: Reduce ED Wait Time with Language Model Diagnostic Assistance,* ICML 2024
   Liwen Sun, Abhineet Agarwal, Aaron Kornblith, Bin Yu, and Chenyan Xiong
2. *Citation Prediction for Text-rich Network,* submitted to ACL 2024
   Liwen Sun, Wei Hu, Xinyi He, Qi Zhu, and Jiawei Han
3. *Few-shot Text Classification with Dual Contrastive Consistency,* submitted to ICLR 2023,
   Liwen Sun, and Jiawei Han
4. *Causal Fusion for Recommender System*, CONF-CDS, 2022
   Liwen Sun, Chi Zhang, Zilin Su, Feiran Zhang, and Yuen Zhou

WORK EXPERIENCE

National Center for Supercomputing Applications Champaign, IL

Research Intern, advised by Prof. Yuxiong Wang *Oct. 2022 – Jan. 2023*

- Worked on early detection and prediction of parkinsonism powered by multi-modal few-shot learning.
- Explored time-series models to identify Parkinsonism via frame-level geometrical keypoint features.

RESEARCH EXPERIENCE

Information Retrieval Group Pittsburgh, PA

Research Assistant, advised by Prof. Chenyan Xiong *Aug. 2023 -Present*

- Utilize publicly available patient records and collaborate with real clinicians in the emergency department to curate a MIMIC-ED-Assist benchmark to advance research in the AI healthcare domain.
- Developed an ED-Copilot agent to offer cost-effective diagnostic assistance by using BioGPT to encode patient information and reinforcement learning from human feedback (RLHF) to minimize laboratory test time and maximize prediction accuracy of critical outcomes simultaneously.
- Developed a novel multimodal retriever driven by medical knowledge to augment chest report generation from the biomedical vision-language foundation model.

Data Mining Group Champaign, IL

Research Assistant, advised by Prof. Jiawei Han *Mar. 2022 – Aug. 2023*

- Proposed a novel fine-grained taxonomy construction method by GPT-4 to guide topic classification for paper-reviewer matching and author identification tasks. Explored parameter-efficient large language model architectures to optimize model fine-tuning.
- Proposed a novel citation prediction framework of joint modeling graph structure and textual signals in the text-rich heterogeneous bibliographic network by designing embedding propagation strategies with graph neural network to aggregate neighbor paper's textual attributes into query paper's representation from multi-view graphs and retrieving high-quality target papers.
- Proposed a novel semi-supervised framework to perform text classification in few-shot settings by leveraging noisy unlabeled data from back-translation and integrating supervised contrastive learning on few-labeled data.

TECHNICAL STRENGTHS

Programming Languages: Proficient in Python, C/C++, Java,
Machine Learning Package: PyTorch, TensorFlow, Scikit-learn, PyG