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)

Nanjing University of Information Science and Technology *Aug. 2018–July 2020*
Bachelor of Science in Computer Science (Transferred out) GPA: 3.8/4.0

PUBLICATION

1. *Few-shot Text Classification with Dual Contrastive Consistency,* preprint, 2022
   Liwen Sun, and Jiawei Han
2. *Citation Prediction for Text-rich Network,* preprint, 2022
   Liwen Sun, Wei Hu, Xinyi He, Qi Zhu, and Jiawei Han
3. *Causal Fusion for Recommender System*, CONF-CDS, 2022
   Liwen Sun, Chi Zhang, Zilin Su, Feiran Zhang, and Yuen Zhou

RESEARCH EXPERIENCE

Taxonomy-guided Reviewer Match Champaign, IL
Research Assistant (Data Mining Group) *Mar. 2023 – Aug. 2023*
Supervisor: **Jiawei Han**, University of Illinois at Urbana-Champaign

- Proposed a novel fine-grained taxonomy construction method by GPT-4 to guide topic classification for paper-reviewer matching and author identification tasks.
- Explored different parameter-efficient large language model architectures to optimize model fine-tuning.

Early Detection and Prediction of Parkinsonism Powered by Multi-Modal Few-Shot Learning Champaign, IL
Research Intern (National Center for Supercomputing Applications) *Oct. 2022 – Jan. 2023*
Supervisor: **Yuxiong Wang**, University of Illinois at Urbana-Champaign

- Incorporated video modality into pre-trained vision backbone by aggregating spatiotemporal region attention.
- Explored different time-series models to identify Parkinsonism via frame-level geometrical keypoint features.

Citation Prediction in Text-rich Network Champaign, IL
Research Assistant (Data Mining Group) *Sep. 2022 – Jan. 2023*
Supervisor: **Jiawei Han**, University of Illinois at Urbana-Champaign

- Proposed a novel citation prediction framework based on multiple types of relations and entities in the text-rich network by clustering paper embeddings from multi-view graphs and retrieving quality target papers.
- Implemented approximate personalized PageRank to reduce computation cost from graph neighbors and retrieve query paper's top K ones in a heterogeneous bibliographic network.
- Designed embedding propagation strategies to aggregate neighbor paper's textual information into query paper.

Few-Shot Text Classification with Dual Contrastive Consistency Training Champaign, IL
Research Assistant (Data Mining Group) *May 2022 – Oct. 2022*
Supervisor: **Jiawei Han**, University of Illinois at Urbana-Champaign

- 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.
- Devised a novel contrastive consistency schema that can generate soft pseudo-labels for propagating feature structure from labeled examples to unlabeled ones dynamically.

Causal Fusion for Recommender System Pittsburgh, PA
Remote Research *Dec. 2021 – Mar. 2022*
Supervisor: **Pradeep Ravikumar**, Carnegie Mellon University

- Adapted matrix factorization models and unbiased estimation techniques by causal inference to handle selection bias from user-item rating in recommender system. Integrated causal inference into neural collaborative filtering framework to boost rating model robustness and scalability.
- Proposed data fusion algorithm and incorporated unbiased data into biased training dataset to jointly learn doubly robust estimator for matrix factorization called MF-DRJL to debias and optimize prediction and imputation error model.

TECHNICAL STRENGTHS

Programming Languages: Proficient in Python, C/C++, Java, Haskell
Machine Learning Package: PyTorch, TensorFlow, Scikit-learn, PyG
Web Development: HTML, CSS, Bootstrap, Python Flask, JavaScript
Database Management Systems: MySQL, MongoDB, Neo4j