JUANHUI LI

📞 517-303-7625 ✉ lijuanh1@msu.edu 🏠 https://juanhui28.github.io/ ⭘ https://github.com/Juanhui28/

SUMMARY

A third-year PhD student at Computer Science and Engineering, Michigan State University. Her research interests include graph mining, information retrieval, recommendation systems, and language models.

EDUCATION

Michigan State University (MSU)   Sep. 2021 – Present
PhD in Computer Science and Engineering   *Michigan, U.S.*

Sun Yat-sen University (SYSU)   Sep. 2017 - Jun. 2020
Master of Computer Science and Technology   *Guangdong, China*

Sun Yat-sen University (SYSU)   Sep. 2013 - Jun. 2017
Bachelor of Software Engineering   *Guangdong, China*

REPRESENTATIVE PUBLICATIONS

[1] **Juanhui Li**, Harry Shomer, Haitao Mao, Shenglai Zeng, Yao Ma, Neil Shah, Jiliang Tang, Dawei Yin. "Evaluating Graph Neural Networks for Link Prediction: Current Pitfalls and New Benchmarking", *The 37th Conference on Neural Information Processing Systems (**NeurIPS 2023**)*

- Reveal several pitfalls in **link prediction** and pioneered a benchmarking system for **fair and consistent evaluation** across a diverse range of models and datasets.
- Conducted insightful analyses on the performance and features of various models, enhancing our comprehension of these models and datasets.
- Introduced a more **practical setting** called HeaRT, which integrates multiple heuristic methods to generate customized negative samples. By closely mirroring real-world applications, HeaRT could guide the field towards more effective models.

[2] **Juanhui Li**, Harry Shomer, Jiayuan Ding, Yiqi Wang, Yao Ma, Neil Shah, Jiliang Tang, Dawei Yin. "Are Message Passing Neural Networks Really Helpful for Knowledge Graph Completion?", *The 61st Annual Meeting of the Association for Computational Linguistics (**ACL 2023**)*

- We found that the message passing component, once pivotal in **Message Passing Neural Networks (MPNNs)**, has limited influence on **knowledge graph completion**. This finding has reshaped the understanding of MPNN's integral components.
- We conducted detailed experiments on scoring functions and loss functions. Our findings emphasized their significant impact on performance, offering a fresh perspective on where to prioritize model design.
- We developed a **MLP-based ensembling method**, which not only proved to be more efficient but also consistently surpassed the performance of conventional prevailing MPNN-based models.

[3] **Juanhui Li**, Yao Ma, Wei Zeng, Suqi Cheng, Jiliang Tang, Shuaiqiang Wang, Dawei Yin. "Graph Enhanced BERT for Query Understanding", *The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2023**)*

- Designed a pretraining framework that transfers graph semantic information from the **query-query graph** (constructed using the search log from the Baidu search engine) to the **transformer module**. This approach ensures the pretrained transformer captures both the depth of query content semantics and the intricacies of the query-query graph relationships.
- Our presented model excels in offline tasks (like query classification and matching) as well as online medical query classification. A noteworthy 3.2% improvement was observed during **A/B testing**, emphasizing the model's capability to better comprehend queries and boost **search performance**.

[4] Harry Shomer, Wei Jin, **Juanhui Li**, and Yao Ma, Jiliang Tang. "Learning Representations for Hyper-Relational Knowledge Graphs". *The 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Marrakesh, Morocco, 6-9 November 2023* (**ASONAM 2023**)

[5] **Juanhui Li**, Yao Ma, Yiqi Wang, Charu Aggarwal, Chang-Dong Wang, Jiliang Tang. "Graph Pooling with Representativeness", *2020 IEEE International Conference on Data Mining (**ICDM 2020**)*

- Propose a novel **graph pooling** operator that prioritizes both node importance and node representativeness. To this end, we design a **greedy algorithm** to select important and representative nodes one by one.
- The proposed pooling module RepPool stacked over the **graph convolutional layers** forms a hierarchical architecture to conduct graph classification task. Comprehensive experiments demonstrate the superiority of the RepPool.

Diverse Experiences

Homedepot   June 2023 – present
U.S

- Worked on the session recommendation to predict the next item based on the users' historical sessions. We built a framework that combines text and user interaction information to make more accurate predictions. We used **language models** to understand and process the text information and **Graph Neural Networks (GNN)** to understand user interactions. This method made our predictions more reliable and helped create recommendations that are more tailored to individual users.

Baidu Inc.   May 2020 – August 2021
Research Intern   *Beijing, China*

- Developed an advanced pretraining framework, leveraging the BERT model, to augment its comprehension of queries originating from the Baidu search engine. This enhancement improves performance across offline and online tasks, evidencing a 3.2% improvement in A/B testing. This work has been accepted by SIGIR 2023.

Technical Skills

Languages: Python (PyTorch, Torch-Geometric, torch_scatter, Scikit-Learn, Transformers,etc.), C++, Matlab
Developer Tools: VS Code, Eclipse, Pycharm
Technologies/Frameworks: Linux, GitHub