Yu Xia

+1 (734)-834-8538 ◇ Ann Arbor, Michigan
xiayuu@umich.edu ◇ linkedin.com/in/xiayu9 ◇ github.com/Andree-9

EDUCATION

University of Michigan — Ann Arbor, MI, US
Master of Science in Information — *Sep. 2022 - Now*
Shanghai Jiao Tong University — Shanghai, China
Bachelor of Electrical and Computer Engineering — *Sep. 2019 - May 2023*

RESEARCH EXPERIENCE

Large Language Models with Different Learning Paradigms
*Work with Prof. Shuai Li (**Shanghai Jiao Tong University**) and Dr. Tong Yu (**Adobe Research**)* *Mar. 2023 - Sep. 2023*

- Proposed an active learning framework to efficiently alleviate factual inconsistency problem of LLMs in text summarization. Developed a diversity-based sampling to select diverse hallucinations for annotation and finetuning. [C2] *(Under Review)*
- Studied LLMs' in-context learning ability from weight shifting perspective. Conducted experiments validating the similarity between transformer's attention computation and gradient descent on softmax regression. [C3] *(Under Review)*
- Developed an increasing bandit algorithm for online selection of LLMs considering the increasing-then-converging performance trend in LLMs' finetuning process to efficiently choose the optimal model for long-term applications. [C4] *(Under Review)*
- Viewed LLM alignment with human preference from a causal perspective. Developed a causality-enhanced reinforcement learning framework for language model alignment with human feedback. *(Work in Progress)*
- Extensive experience in finetuning LLMs such as GPT-2 and Flan-T5 for natural language generation tasks.

Conversational Recommender System with Bandit Feedback
*Work with Prof. Shuai Li (**Shanghai Jiao Tong University**) and Dr. Tong Yu (**Adobe Research**)* *May 2021 - Feb. 2023*

- Developed a deconfounded conversational recommender system to alleviate harmful biases via causal inference. Proposed user regulation to utilize useful biases based on user feedback signals to both items and attributes. [C1] *(Published in KDD'23)*
- Designed a hybrid conversational recommender system that seamlessly asking users rating or comparison questions over item attributes. Proposed a bandit algorithm to utilize both absolute and relative feedback. [J1] *(Minor Revision in UMUAI)*

PUBLICATIONS

- [C1] **Yu Xia**, Junda Wu, Tong Yu, Sungchul Kim, Ryan A. Rossi, and Shuai Li. 2023. User-Regulation Deconfounded Conversational Recommender System with Bandit Feedback. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
- [J1] **Yu Xia***, Zhihui Xie*, Tong Yu, Canzhe Zhao, and Shuai Li. 2023. Towards Joint Utilization of Absolute and Relative Bandit Feedback for Conversational Recommendation. User Modeling and User-Adapted Interaction. *(Minor Revision)*
- [C2] **Yu Xia**, Xu Liu, Tong Yu, Sungchul Kim, Ryan A. Rossi, Anup Rao, Tung Mai, and Shuai Li. 2023. Hallucination Diversity-Aware Active Learning for Text Summarization. *(Under Review)*
- [C3] Shuai Li, Zhao Song, **Yu Xia**, Tong Yu, and Tianyi Zhou *(Alphabetical Order)*. 2023. The Closeness of In-Context Learning and Weight Shifting for Softmax Regression. *(Under Review)*
- [C4] **Yu Xia***, Fang Kong*, Tong Yu, Liya Guo, Ryan A. Rossi, Sungchul Kim, and Shuai Li. 2023. Convergence-Aware Online Model Selection with Time-Increasing Bandits. *(Under Review)*
- DecAl: Deconfounded Language Model Alignment from Causal-Reweighted Feedback. *(Work in Progress)*

COURSE PROJECTS

Science Fiction Text Generation [Link]

- Finetuned a GPT-2 model on a science fiction corpus with over 2M sentences for text generation.
- Improved the base GPT-2 model by 32.6% in generated texts' fluency and 20% in human evaluation scores.

Beer Recommender System [Link]

- Built a beer recommender system with collaborative filtering using over 1M beer review records.
- Achieved recommendation improvements over random baseline of 18.9% in F1 Score and 40.8% in RMSE.

AWARDS

- 2020 & 2021 UM-SJTU Joint Institute Yu Liming Scholarship (Top 5%).
- 2020 & 2021 Shanghai Jiao Tong University Outstanding Undergraduate Award (Top 20%).

SKILLS

Programming: Python, C/C++, SQL. **Web Dev:** HTML, CSS, Django. **ML/NLP:** PyTorch, TensorFlow.