Xinyi Fang

774-737-4561 | xfang1@wpi.edu | shineef.github.io | huggingface.co/xinyifang

Education

Brandeis University — Waltham, MA
Ph.D. in Computer Science, advised by Prof. Constantine Lignos — *Starting soon*

Worcester Polytechnic Institute — Worcester, MA
Master of Science in Computer Science — *Aug. 2023 – May 2025(Expected)*
SIGIR, WSDM reviewer
WPI Infolaber, advised by Prof. Kyumin Lee

Guangzhou University — Guangzhou, Guangdong
Bachelor of Science in Data Science — *Aug. 2019 – May 2023*
Honoree of Learning Star

Research Interests

I am broadly interested in machine learning and artificial intelligence. My recent research primarily focuses on natural language processing, particularly enhancing the text generation abilities of the Large Language Models and improving downstream task performance on text-attributed graphs such as node classification and link prediction.

Experience

Teaching Assistant — Aug. 2024 – Present
Worcester Polytechnic Institute — *Worcester, MA*
- CS573 - S25: Data Visualization
- CS5084 - F24: Introduction to Algorithms: Design and Analysis

Projects

Intelligent NLP Framework | *PyTorch, scikit-learn, nltk, LangChain, Matplotlib, seaborn* Feb. 2024 – Apr. 2024
- Developed a citation recommender system that utilizes GraphSAGE for node classification, MLP regressor to map text embeddings, and similarity-based retrieval to suggest the most relevant and appropriate citations.
- Implemented a question-answering system based on BERT, explored unanswerable question detection methods, and designed adaptive sequential chain structures using LangChain, demonstrating dynamic response generation.
- Conducted sentiment analysis on the Amazon Fine Food Reviews dataset by implementing and comparing various text classification methods, including Word2Vec, BERT (with and without fine-tuning), and BERT with LoRA.
- Performed fake news detection by developing and comparing various machine learning models, including naive Bayes, SVM, CNNs, RNNs, and LSTM, using features such as TF-IDF, sentiment scores, and POS-tags.

Advanced Information Retrieval | *PyTorch, TensorFlow/Keras, scikit-learn, pandas* Aug. 2023 – Nov. 2023
- Engineered a multi-faceted book recommender system utilizing collaborative filtering, deep learning, transformer architectures, and support vector machines, combined with an intuitive user interaction component.
- Implemented a web crawler and search engine utilizing BeautifulSoup for HTML parsing and the PageRank algorithm to index, rank, and retrieve relevant web pages from a corpus of HTML documents.
- Developed an advanced search engine using binary trees, permuterm indexing, wildcard queries, tokenization, and crawling techniques to efficiently index and search through large corpora of text documents.

Deep RL Strategy Optimization | *PyTorch, Matplotlib, gymnasium, NumPy* Aug. 2023 – Nov. 2023
- Applied advanced Dueling Double Deep Q-Network techniques in PyTorch to master Breakout, achieving a top score of 367 by optimizing neural network architectures and enhance the replay memory mechanism.
- Developed and applied Monte Carlo and Temporal Difference learning methods to improve decision-making in Blackjack and Cliff Walking simulations, deepening knowledge of model-free RL techniques.
- Implemented and tested dynamic programming techniques including policy evaluation as well as value iteration for Markov Decision Processes to optimize decision-making strategies in game environment.

Technical Skills

Languages: Python, C/C++, JavaScript, HTML/CSS, Matlab, R, SQL
Libraries: PyTorch, TensorFlow/Keras, pandas, scikit-learn, NumPy, Matplotlib
Developer Tools: VS Code, Jupyter Notebook, Google Cloud Platform, PyCharm