Moshi Wei

Email: wmswms938@gmail.com

Linkedin: https://www.linkedin.com/in/moshi-wei/

NLP, Python, recommendation system

Technical Skills

- Language: Python JS C++ Java PHP Solidity Golang
- Tool: Huggingface NLTK Pytorch Tensorflow Keras BS4 Scikit-learn SpaCy StanfordNLP Seaborn Flask FastAPI VueJS Neo4J PostgreSQL MongoDB
- Knowledge: NLP AST Deep learning Machine learning RNN LSTM BERT RoBERTa Prompt Learning Contrastive Learning Transformer Code search BIKER CodeBERT DeepAPI

Experience

NLP Researcher, Ph.D. Candidate, York University   Sep 2020.09 – Now

- Trained and Deployed Deep learning models on Colab, GCP, AWS, Vast.ai, and Azure.
- Creator of the current API recommendation SOTA tool(CLEAR paper)
- Extracted and formatted data from Stackoverflow, official Java document with BS4 then tokenize and perform EDA analysis with Scikit-learn, StanfordNLP, NLTK, and visualize with Seaborn and Pyplot
- Extract APIs with a heuristic parser and paring AST tree with PythonJST parser from code snippet and tokenized with a customized tokenizer
- Trained MLP models such as Seq2seq LSTM, BERT, CodeBERT, Pegasus, Text-CNN with Huggingface, Pytorch or Tensorflow
- Experimented with multiple learning architectures for performance optimization including contrastive learning, prototypical network, siamese network
- Trained ML models such as SVM, NB, KNN, and XGBoost as experiment baseline with scikit-learn
- Evaluate the result using multiple metrics including MRR, MAP, precision, recall, and F1
- Compare SOTA papers and perform ablation studies to identify the performance bottleneck and experiment alternatives correspondingly
- Presented 2 conference talks (one online, one in-person) at international conferences
- Top 1.2% leaderboard on Kaggle House Prices Competition

Software Engineer – Business Intelligence, Achievers   May. 2019 – Sep 2020

- Initiated and developed an awarded collaborative filtering-based user recommendation system with Spacy, Scikit-learn, and pandas. Run the system on 200 users for testing and feedback.
- Developed and maintained daily ETL tasks with Centerprise ETL and PostgreSQL
- Migrated existing ETL tasks to low-code ETL tool Exago
- Speedup TB-level ETL data warehouse restoration time by 24 times compared with the previous version by restructuring workflow using PostgreSQL

Education

- York University, Ph.D. of Software Engineering, API recommendation Research
- University of Waterloo, Master of Software Engineering, Program Repair Research

Publication

- Wei, Moshi, et al. "CLEAR: contrastive learning for API recommendation." *Proceedings of the 44th International Conference on Software Engineering.* 2022.
- Wang, Song, et al. "Automatic Unit Test Generation for Machine Learning Libraries: How Far Are We?." *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE).* IEEE, 2021.
- Lutellier, Thibaud, et al. "Coconut: combining context-aware neural translation models using ensemble for program repair." *Proceedings of the 29th ACM SIGSOFT international symposium on software testing and analysis.* 2020.