Xiang Liu

Email: xliu29@gmu.edu

Mobile: +852-97479004

Education

- **The University of Hong Kong** — HK, CHINA
  Master of Science(MSc) - Computer Science — *Sep 2022 - Dec 2023*
- **George Mason University** — VA, USA
  Bachelor of Science - Computer Science; GPA: 3.71/4.0 — *Aug 2018 - Aug 2022*
  Honors/Awards: *Dean's List (2018-2020)*
  Courses: *Natural Language Processing, Visual Computing, Data Structure, Database Concepts*
- **Henan University** — Henan, CHINA
  Bachelor of Engineering - Networking Engineering; GPA: 88/100; Rank: 4/292 — *Sep 2017 - Jun 2022*

Experience

- **Baidu Research Cognitive Computing Lab** — Beijing,CHINA
  Intern (Full-time) — *Dec 2021 - June 2022*
  - Worked on dependency parsing, using Open Information Annotation (OIA) method to parse sentences.The OIA method will convert sentence to a DAG graph and the graph was generated by the dependency information in this sentence.
  - Investigate Chinese sentence and apply OIA to Chinese sentence. Propose the node type and edge type specific to Chinese OIA. Improve the performance of Chinese OIA close to English OIA.
  - During the internship, my essay reading skills and coding skills improved substantially, also have research experience in OIE and logical reasoning.
  - Worked with Dr.Xin Wang and Dr.Mingming Sun.

Competitions

- **Kaggle — Feedback Prize - Predicting Effective Arguments Competition, Sliver Medal Top 2%** :
  Team Leader, Jun 2022 - Aug 2022
  - Writing the code baseline for teammates, design different data preprocessing strategies, model structures. Arrange teammates workload and planning meeting.
  - Master the idea of transforming problems. Using token classification instead of the sequence classification in this competition. This new solution boosting our ranking and saving lot time on training and inference. Acquire the ability to use the Transformers training API.
- **Kaggle — U.S. Patent Phrase to Phrase Matching, Bronze Medal Top 9%** :
  Solo, Mar 2022 - Jun 2022
  - Writing whole pipeline code, try different combine with pre-train language models and model structures.
  - Master the different fine-tuning structures, like mean pooling, multi dropout, layer normalization. Using cross validation to prevent overfitting and find best ensemble weights.
- **CCF Collegiate Computer System & Programming Contest Top 5%**:
  Team Member, Sep 2021 - Nov 2021
  - The goal of this competition is to predict default or non-default using user personal information and consumption information.
  - Master the using of lightGBM and XGBoost models, and the model stacking method. Mastered real-world data science challenge with machine learning pipeline and team collaboration.

Projects

- **Rumor Prediction Model for Microblog Epidemic**:
  - A Covid-19 rumor prediction model was built based on the pre-trained model BERT, which completes the deep learning model complete process.
  - In this project, I went through the whole process of idea generation, data search, reference papers, code practice, model building and article writing, and gained a very deep understanding of the NLP project process.
- **Emotional analysis case**:
  - Sentiment analysis is performed on the movie review text dataset, and the bag-of-words model, N-gram model, loss function, and gradient descent function are implemented manually.
  - Through this project, I have gained a deeper understanding of the principles of deep learning, the real usage of gradient descent, and a deeper understanding of statistical-based language models.

Personal Summary

- **Languages**: Python, SQL, JAVA,C++
- **Frameworks**: PyTorch, Scikit, NLTK, SpaC