Haorui He

E-mail: hehaorui11@gmail.com : HarryHe11

EDUCATION

Nanjing University of Posts and Telecommunications, Nanjing, China *Sept. 2019 - June 2023(expected)*

BE in Software Engineering **GPA:** 4.02/5, equivalent to 3.88/4 **IELTS:** 7.5 **GRE:** 327

RESEARCH EXPERIENCE

Software Vulnerability Inspection System Based on Active Learning *Aug. 2022 - Oct. 2022*

MITACS Globalink Research Intern, advised by Prof. Jack Jiang (YorkU)

- Implemented an active machine learning-based software vulnerability inspection system, which can detect software's vulnerable source code files based on software metrics, source codes, and crash dump stack trace features.
- Simulated code inspections on C and C++ files from the Mozilla Firefox project to evaluate the system. The system found 99% of the vulnerabilities by inspecting 35% of the source code files.

Improved Target-specific Stance Detection by Delving into Conversation Threads *June 2021 - Sept. 2022*

Research Intern, advised by Prof. Francis C.M. Lau (HKU) & Prof. Yupeng Li (HKBU)

- Developed a distributed Web crawler system with Requests, Scrapy, Scrapy-Redis, and XPath libraries, which collects streaming data from six social platforms, performed data wrangling, and stored the data in a Mysql database;
- Designed and implemented a Branch-BERT model with Pytorch, which facilitates target-specific stance detection with contextual information in conversation threads, outperforming several context-free baselines, including the state-of-the-art method, by 10.3% in the F1 score on a newly proposed Conversational Stance Detection dataset.

Adaptive Knowledge Distillation for Efficient Relation Classification *Mar. 2021 - Mar. 2022*

Leader of a Distinct National University Students' Science and Technology Innovation Project

- Designed a knowledge distillation algorithm to avoid excessive uncertainty in training relation classification models by adaptively adjusting the temperature parameter according to the smallest margin uncertainty of the models;
- Designed a knowledge adjustment method to avoid genetic errors in knowledge distillation;
- Improved the performance of a distilled LSTM student model by 3.83% in the F1 score. The model's performance is comparable with its BERT-based teacher, but the student uses 95% less time in inference and 67% less memory.

PUBLICATION & MANUSCRIPT

1. Haorui He, Yuanzhe Ren, Zheng Li, and Jing Xue. **Adaptive Knowledge Distillation for Efficient Relation Classification**, *ICANN 2022*. [pdf]

2. Yupeng Li, Haorui He (Corresponding), Shaonan Wang, Francis C.M. Lau, and Yunya Song. **Improved Target-specific Stance Detection on Social Media Platform by Delving into Conversation Threads**, *IEEE Transactions on Computational Social Systems under review.* [arXiv]

MACHINE LEARNING COMPETITIONS

Purchase and Redemption Forecast : **https://github.com/HarryHe11/Tianchi-PRF**

-- Alibaba Cloud Tianchi Big Data Competition (Top 5% worldwide)

- Performed exploratory data analysis and preprocessing, including encoding and standardization;
- Constructed 46 features based on date information and historical trading records;
- Selected best feature combination by complex collinearity, correlation, permutation importance, and SHAP value;
- Implement an XGBoost regression model with Tensorflow to forecast Alipay's daily capital inflow and outflow amounts (i.e., purchase and redemption) in the forthcoming month and achieve a final online score of 132.89.

Bank Loan Risk Prediction Based on Machine Learning

-- Bank of Jiangsu Financial Big data Modeling Challenge 2020 (Top 10% and the only winning undergraduate)

- Performed exploratory data analysis and preprocessing, including standardization and missing data imputation;
- Constructed 51 features based on loan characteristics, customer information, and their in-loan behaviour;
- Implement an XGBoost model to predict customers who will have bad debts in the forthcoming six months;
- Finetuned model hyperparameters by grid search, and the model achieved an F1 score of 73%.

CORE SKILLS

- ✓ **Programing Language:** Proficient in Python, experienced in C++/Java/Javascript/HTML/PHP/Linux Shell;
- ✓ **Machine Learning:** Proficient in PyTorch/Tensorflow/Scikit-learn/Numpy/Pandas/Xgboost/Scipy;