Haorui He

Email: hehaorui11@gmail.com
Portfolio: harryhe11.github.io
Mobile: 86-139-8336-3231
Github: github.com/HarryHe11

Education

- **Nanjing University of Posts and Telecommunications** Nanjing, China
  B.Eng. in Software Engineering; GPA: 90.07/100 *September 2019 - June 2023*

Research Experience

- **Improved Target-specific Stance Detection by Delving into Conversation Threads** HKU & HKBU, HK
  Research Intern, supervised by Prof. Francis C.M. Lau & Prof. Yupeng Li *June 2021 - Sep. 2022*
  - **Dataset Construction**: Constructed the first conversational stance detection dataset by organizing conversation threads gathered from social media platforms.
  - **Model Design & Implementation**: Designed and implemented an Branch-BERT model using PyTorch to enhance target-specific stance detection by leveraging contextual information within conversation threads.
  - **Evaluation**: Conducted experiments to evaluate the proposed model. The results showed that it outperformed several context-free baselines, including the state-of-the-art method, by achieving a 10.3% improvement in the F1 score.
- **Active Learning Based Software Vulnerability Inspection** YorkU, CA
  Mitacs Globalink Research Intern, supervised by Prof. Jack Jiang *Sep. 2022 - Dec. 2022*
  - **Model Implementation**: Implemented an active learning-based software vulnerability inspection model, which classifies software's vulnerable code files based on software metrics, source codes, and crash dump stack trace features.
  - **Evaluation**: Simulated code inspections on C and C++ files obtained from the Mozilla Firefox project to evaluate the system's effectiveness. The model achieved a remarkable success rate of 99% success rate in identifying vulnerabilities by inspecting only 35% of the source code files.

Publications

[1] Yupeng Li, **Haorui He (Corresponding)**, Shaonan Wang, Francis C.M. Lau, and Yunya Song. Improved Target-specific Stance Detection on Social Media Platforms by Delving into Conversation Threads. IEEE Transactions on Computational Social Systems. (under the 2nd round of revision, [pre-print].)

[2] Yupeng Li, Dacheng Wen, **Haorui He**, and Francis C. M. Lau. Contextual Target-Specific Stance Detection on Twitter: New Dataset and Method. Proc. of ACM MM, 2023. (under review)

Services

- Reviewer: ACM MM 2023

Projects

- **Stance.ai Web App** [Github]
  Full-Stack Developer
  - **Web Development**: Built a full-stack AI-powered stance detection web application with Flask, React, and MongoDB.
  - **Model Design & Implementation**: Designed and implemented an BERTA model which utilized the attention mechanism to extract target-specific keyword features for stance detection, improving the performance by 1.61%.
  - **Model Deployment**: Deployed the model as a RESTful API to provide stance detection service. Implemented RESTful APIs for user management, data processing pipelines, model selection, and result demonstration.
- **The Purchase and Redemption Forecasets** [Github]
  – *Alibaba Cloud Tianchi Big Data Competition*
  - **Data Analysis & Preprocessing**: Conducted exploratory data analysis and preprocessing, including encoding and standardization techniques, to prepare the data for further analysis.
  - **Feature Engineering & Selection**: Constructed 46 features using date information and historical trading records, leveraging domain knowledge to capture relevant patterns and trends. Employed techniques such as correlation analysis, permutation importance, and SHAP value analysis to identify the best combination of features for modeling.
  - **Modeling**: Implemented an XGBoost regression model with TensorFlow to predict future cash flows (i.e., purchase and redemption) of Ant Financial Services Group. Achieved a final online score of 132.89, placing in the top 5% worldwide.
- **Loan Risk Prediction Based on Machine Learning**
  Bank of Jiangsu Financial Big Data Modeling Challenge 2020
  - **Data Analysis & Preprocessing**: Conducted exploratory data analysis and preprocessing to ensure data quality, including standardization and missing data imputation techniques.
  - **Feature Engineering**: Constructed 51 comprehensive features by leveraging customer information, loan characteristics, and in-loan behaviors, providing a holistic view of loan risk factors.
  - **Modeling & Optimization**: Implemented an XGBoost classification model to detect customers who were likely to default on their loans within the forthcoming six months. Utilized weighted loss and over-sampling techniques to address data imbalance. Achieved remarkable results with an F1 score of 0.73, placing in the Top 10%.

Skills

- **Languages**: Python, JavaScript, Java, HTML, CSS, SQL
- **Machine Learning**: PyTorch, TensorFlow, Keras, Numpy, NLTK, Scikit, SpaCy
- **Web Development**: Flask, React, Express.js, Node.js, MySQL, SQLite, MongoDB, Git, Heroku