Yuxiang Zhu

(+86)18938889050 | zyx@smail.nju.edu.cn

Education

Software Institute, Nanjing University — Nanjing, China
Master in Software Engineering; GPA: 8.84/10.0 — *Expected Jun. 2020*

Software Institute, Nanjing University — Nanjing, China
Bachelor in Software Engineering; GPA: 8.56/10.0 — *Sept. 2014 - Jun. 2018*

Academic Experience and Internship

Research Assistant — Nanjing, China
Software Engineering Group, State Key Laboratory for Novel Software Technology, Nanjing University — *Oct. 2018 - Present*

- **Issue Classification**: Developed a model to classify issues from Issue Tracking Systems such as GitHub and Jira
  - Collected and preprocessed over 1.2M issues in GitHub and Jira to prepare a clean dataset for training
  - Altered and implemented the k-NN algorithm to detect issue that is wrongly classified by its submitter, resulting in 1.3% improvement in F-score of final model
  - Implemented and trained an attention-based bi-directional LSTM, resulting in 11.1% improvement in F1-score compared to SVM model, and 5%-8% improvement in F1-score compared to state-of-the-art approaches
  - A first-authored article is in preparation
- **Pull Request Summarization**: Implemented and trained a sequence-to-sequence model to summarize diff files in pull requests to short informative messages. Achieved BLEU score of 17.3 on test set of 30k pull requests
- **Program Comprehension**: Conducted a systematic literature review in the field of automatic source code summarization
  - Examined 41 relevant studies from 2010 to 2019. Compared different approaches and discussed different techniques used in the state-of-the-art approaches, including 4 kinds of data extraction method, 4 kinds of description generation method, 5 kinds of evaluation method, and 7 kinds of code artifact
  - A first-authored article is under review

Research Assistant — Nanjing, China
Software Engineering Group, State Key Laboratory for Novel Software Technology, Nanjing University — *Nov. 2016 - May 2018*

- **Requirement Extraction**: Developed a tool to automatically collect and analyze user reviews in Google Play
  - Collected and preprocessed 394k user reviews from 100 popular apps
  - Used information entropy to filter out 41.3% user reviews that didn't contain enough information for requirements
  - Designed and implemented the K-Means++ algorithm to cluster reviews by tf-idf word vector. Then we selected words with highest tf-idf weight in each cluster. For every app, we formed 60 clusters and display them with keywords lists, ordered by Silhouette Coefficient
  - Formed a completely automatic tool and developed a Java Web System to display the result

Software Engineer Intern — Shanghai, China
Meituan-dianping Inc. — *Jul. 2017 - Jan. 2018*

- **GroupJoy**: As the project manager of a team of six newcomers, designed and developed a platform for internal staff to launch and attend activities, to socialize and find common interest. We spent 4 weeks on collecting requirements, designing architecture and coding. Our system contains more than 4k lines of code, serving hundreds of staffs
- **CustomerInfoService**: Maintained and iterated a Java Web System full-stack-ly, which helps the sales staff register, manage, and contact with merchants.
- **Data Analysis**: Analyzed and visualized data of merchant relationship, involving over 1M merchants. Independently set up a small system which emails every day statistics graphics about our merchant data to all(15) colleagues in our group

Projects

- **Stars' Microblog Analyzer**: A model to analyze Chinese entertainment stars' behavior by microblog content
  - Collected 60k microblogs of 500 stars in Weibo. Designed and implemented a cluster algorithm to group stars that send similar content. Used the Pagerank Algorithm to identify the center star of each group
- **Stock Analysis**: A website that comprehensively and visually displays statistics and graphics of stock data
  - Crawled historical records of stocks from famous stock website and generated statistics and graphics
  - Built a sub-system that enables user to design static buy-and-sell strategy and simulate it to see its profitability

Skills

- **Languages**: Python, Java, SQL, C++, PHP, HTML, CSS, Javascript
- **Technologies**: Maven, Git, Linux, Microsoft Office, Spark, LaTeX
- **Development**: Object Oriented, Relational databases, Machine Learning, NLP
- **English**: TOEFL 107; GRE V160+Q170+AW3.5