Tianpei Xia

• 360-808-3650 • xiatianpei@gmail.com • LinkedIn

EDUCATION

North Carolina State University, Raleigh, NC — Aug. 2016 - Dec. 2021
Ph.D. in Computer Science | Advisor: Dr. Tim Menzies

The University of Texas at Dallas, Richardson, TX — Aug. 2013 - Dec. 2015
M.S. in Computer Science

Nanjing University of Posts and Telecom., Nanjing, China — Sep. 2009 - Jul. 2013
B.S. in Electrical Engineering

SKILLS

- Programming tools: Python, Java, AWS, Docker, MySQL, MongoDB, Redis, Grafana, Presto, Jenkins, Spark
- ML frameworks: Scikit-learn, Pytorch, Tensorflow/Keras
- ML development: data analysis/processing, algorithm research, model iteration, profiling and deployment.

SELECTED PROJECTS

Video Chaining Recommendation System — Aug, 2022 - Present
Software Engineer, NewsBreak, Mountain View, USA

- Video Chaining Evaluation System
  - In charge of engineering efforts to design new business metrics for the engineering team to evaluate the NewsBreak video chaining product.
  - Built and maintained the daily and hourly dashboards for monitoring video chaining key metrics.
- Video Chaining Retrievers
  - Built geo-location based retrievers for the video chaining recommendation system to promote local UGC videos, which results in 2.5% user engagement rate increased.
  - Built dedicated exploration retrievers to promote those new UGC videos with low impression rates.

D2D (Document to Document) Recommendation System — Feb, 2022 - Jul, 2022
Software Engineer, NewsBreak, Mountain View, USA

- D2D Filtering Pipeline
  - Designed and implemented a category-based document filter to improve the freshness of the articles in D2D. The average freshness improvements are 25% (p50), 25% (p90) and 4% (p99), respectively with slight user engagement improvements.
  - Designed an implemented a location based filter for local D2D system, which is to remove candidate articles far away from the user and improve local relevancy.
- D2D Blending Strategy
  - Designed and implemented a comprehensive article blending strategy in D2D recommendation list. In A/B test, the performance of new blending experiment is improved in click-per-user (9.5%), check-per-user (0.5%) and pv-time-per-user (8.7%), compared to the control group.

Personalized Local Sports — May, 2021 - Aug, 2021
Software Engineer Intern, NewsBreak, Mountain View, USA

- Designed and implemented geo-information mapping service to tag sports news/articles with potential professional and college sports teams. Those tagged articles will be retrieved by downstream local recommendation systems.
- Designed and Implemented doc indexing services of tending sports and Tokyo2020 Olympic for local news recommendation system in NewsBreak App.

NSF Funded: Search-based Software Engineering Research — Aug, 2017 - Dec, 2021
Research Assistant Under Dr. Tim Menzies, North Carolina State University, Raleigh, USA

- **Evolutionary Algorithms for Hyperparameter Optimization:** Proposed and developed a hyperparameter optimization framework called **OIL** (Optimized Inductive Learning), where evolutionary algorithms are integrated to supercharge software analytic tasks. Experimental results show that OIL improved the performance of effort estimation in terms of accuracy (won 16 out of 18 cases) and efficiency (reduced runtime from days to hours), respectively.

- **Sequential Model Optimization for Software Effort Estimation:** Designed a sequential model based method (a.k.a active learning method) named **FLASH** for the first time in software effort estimation domain to improve software effort estimators. FLASH can efficiently find good configurations of machine learning methods (e.g. CART) for effort estimations. Overall it can improve the performance of software effort estimation tasks by **11%** on average in terms of accuracy.
- **Project Health Prediction for Open-Source Software:** Studied and investigated how predictive methods could help project health prediction. In the study, **78,455 months** of data from **1,628 GitHub projects** has been collected. A group of health indicators is defined based on project developing process and industrial domain knowledge. The preliminary results show that the process action on project level can be predicted to a high level of accuracy (**10% error rate**) with hyperparameter tuning on predicting methods.

SELECTED PUBLICATIONS

- **Tianpei Xia**, Rui Shu, Xipeng Shen, Tim Menzies, *Sequential Model Optimization for Software Effort Estimation*. **Transactions on Software Engineering**, 2020. 📕.
- **Tianpei Xia**, Wei Fu, Rui Shu, Rishabh Agrawal, Tim Menzies, *Predicting health indicators for open source projects (using hyperparameter optimization)*. **Empirical Software Engineering**, 2022. 📕.
- Rui Shu, **Tianpei Xia**, Jianfeng Chen, Laurie Williams, Tim Menzies, *How to Better Distinguish Security Bug Reports (using Dual Hyperparameter Optimization)*. **Empirical Software Engineering**, 2021. 📕.
- Rui Shu, **Tianpei Xia**, Laurie Williams, Tim Menzies, *Omni: automated ensemble with unexpected models against adversarial evasion attack*. **Empirical Software Engineering**, 2022. 📕.
- Suvodeep Majumder, **Tianpei Xia**, Rahul Krishna, Tim Menzies, *Methods for Stabilizing Models Across Large Samples of Projects (with case studies on Predicting Defect and Project Health)*. **Mining Software Repositories**, 2022. 📕.