Tianpei Xia

• 360-808-3650 • txia4@ncsu.edu • 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 - Jun. 2013
B.S. in Electrical Engineering

SKILLS AND INTERESTS

• Programming tools: Python, Java, JavaScript, AWS, Docker, MySQL, NoSQL, Linux, Git, Jenkins, Spark
• ML frameworks: Scikit-learn, Pytorch, Tensorflow/Keras
• ML experience: Data analytics, Visualization, Supervised/unsupervised modeling, ML optimization
• Interests: Apply and develop state of the art machine learning models, productionalize and ship those models; work on large-scale software projects.

SELECTED PROJECTS

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

- **Evolutionary Algorithms for Hyperparameter Optimization:** Proposed and developed a hyperparameter optimization framework called **OIL** (Optimized Inductive Learning), where evolutionary algorithms (e.g. Differential Evolution and NSGA-II) are integrated to supercharge software analytic tasks. OIL was tested on a wide range of optimizers with **945 software projects** data. 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. With the constraints of specific computation costs, 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. Furthermore, predictive models based on random forests, SVM, and CART have been proposed for the project health prediction. 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.

BOTS: Computer Programming Application — Jan, 2017 - May, 2017
Research Project, Game2Learn Lab, North Carolina State University, USA

- BOTS is a programming puzzle game designed to teach fundamental ideas of programming to novice users.
- Planned and implemented the migration of BOTS from “unity4” to “unity5”, enabling more powerful features for the future extention of the whole platform.
- Designed and developed some key features to the system to improve user experience.

Landscape Change Detection from Satellite Image — Jan, 2017 - May, 2017
Graduate Course Project, North Carolina State University, USA

- Drove the technical discussion with teammates and broke down the problem into executable sub tasks.
- Proposed and implemented Gaussian mixture model as the key algorithm to identify landscape changes.
- This grid-based method has competitive performance in Bi-temporal change detection. Given two very high resolution satellite images from the same landscape area, it can achieve similar performance as humans in terms of landscape change detection.

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, Tim Menzies, *Predicting Project Health for Open Source Projects (using the DECART Hyperparameter Optimizer)*. **Empirical Software Engineering (Under Review)**, 2020.
- Rui Shu, **Tianpei Xia**, Laurie Williams, Tim Menzies, *How to Better Distinguish Security Bug Reports (using Dual Hyperparameter Optimization)*. **Empirical Software Engineering**, 2020.