

XUEQI YANG

[Homepage](#), [Github](#), [Linkedin](#), [ResearchGate](#)

+1-571-392-0734 [◇ xyang37@ncsu.edu](mailto:xyang37@ncsu.edu)

EDUCATION

PhD in Computer Science

Aug 2018 - Present

North Carolina State University, GPA: 3.61/4.0

Advisor: Dr. Tim Menzies

Research interests: **Software Engineering, Static Code Analysis, Data Mining and Deep Learning**

Coursework: Automated SE | Algorithm | Networking | Spatial Temporal Data Mining | Natural Language Processing

Bachelor in Information Management and Information System

Sep 2014 - July 2018

Dongbei University of Finance and Economics, China, GPA: 90/100

Coursework: C | Java | Data Structure | Data Mining | Database | Web Design | Operation Research

SKILLS AND STRENGTHS

Languages	Python, C, Java, Bash, JavaScript, MATLAB, SQL, ASP.Net
Tools	PyTorch, Keras, TensorFlow, Scikit-learn, LaTeX
Others	Operation Research, Statistics

SELETED PROJECTS

Simpler Hyperparameter Optimization for Software Analytics

May 2020 - Sep 2020

NSF funded project in the RAISE lab

Raleigh, NC

- Apply a simpler hyperparameter optimization (DODGE, using a technique called ϵ -domination) to 120 SE datasets to find the optimal control settings for data miners.
- Avoid the high training overhead by evaluating and ranking the parameter space in comparison with traditional optimizers, either Differential evolution (DE) or Genetic algorithm (GA).
- Implement Box-counting methods to estimate the intrinsic dimensionality of SE datasets and standard Machine learning datasets (UCI).

Detection for Static Defects with Incrementally Active Learning

July 2020 - Present

NSF funded project in the RAISE lab

Raleigh, NC

- Test the linux mainline at source tree level with coccinelle, a program matching and transformation engine which provides the language SmPL (Semantic Patch Language) for specifying desired matches and transformations in C code.
- Implement feature extractors from warning messages and patches generated from coccinelle with TF-IDF and code2vec embedding methods.
- Utilize Incrementally Active Learning to predict actionable warnings and help Linux maintainers avoid the false positives reported by static analysis tools.

Detection for Static Defects with DNN Models

Sep 2019 - Jan 2020

NSF funded project in the RAISE lab

Raleigh, NC

- Implement deep neural networks in Keras and PyTorch with static defect artifacts to predict real defects to act on.
- Utilize regularisers to avoid DNN models from overfitting and lower the runnning overhead.
- Use Box-counting methods to explore the intrinsic dimension of SE data and match the complexity of machine learning algorithms with the datasets it handles.

Static Warnings Analysis using active learning

Jan 2019 - Aug 2019

NSF funded project in the RAISE lab

Raleigh, NC

- Identify actionable static warnings of nine Java projects generated by FindBugs with incrementally active learning and machine learning algorithms to achieve higher recall with lower cost by reducing false alarm.
- And utilize different sampling approaches (random sampling, uncertainty sampling and certainty sampling) to query warnings suggested by active learning algorithm.

- Interact the system with human oracle to update the system.

Multi-task Learning for Evaluating Peer Assessments

Sep 2020 - Present

Coursework project

Raleigh, NC

- Leverage a state-of-the-art language representation model (BERT, Deep Bidirectional Transformers) in multi-task learning to automatically evaluate peer feedback comments. Utilize oversampling method (in data-level and algorithm-level) to avoid the data imbalance issue. Use Subword Tokenization method, WordPiece which splits a text into subwords, to address the out-of-vocabulary (OOV) problem in NLP. And compare the performance of multi-task model with baseline method, single-task model.
- Also, implement word2vec (CBOW and Skip-grams) and doc2vec (Doc2vec and Part-of-speech tagging) models in Python 3 on Sentimental Analysis Dataset and Question Answering Dataset. And compare performance of proposed methods with baseline methods (TF-IDF and BOW) in individual projects.

Spatial Temporal Object Change Detection and Localization

Jan 2020 - May 2020

Coursework project

Raleigh, NC

- Utilize Mask R-CNN with PyTorch for satellite images change detection and localization. Assess building damage from satellite imagery with a variety of disaster events and different damage extents.

SmartWeather App with Agile Development

Jan 2020 - May 2020

Coursework project

Raleigh, NC

- Implement SmartWeather App in C# with Xamarin and Visual Studio. Use Architecture Diagram, Context Diagram and Quality Attribute Scenarios in software design. Utilize Fuzzy logic controller to converts a crisp input value into a fuzzy set with a predetermined lower and upper bound of impreciseness. And follow the Scrum process to iterate and manage software development.

Quadratic Surface Support Vector Regression for Electric Load Forecasting

Aug 2017 - Aug 2018

Undergraduate Research Project

China

- Use LS-SVR and QSSVR models with the interior point algorithm for electric load forecasting.

PUBLICATIONS

- [1] Xueqi Yang, Jianfeng Chen, Rahul Yedida, Zhe Yu and Tim Menzies, Learning to Recognize Actionable Static Code Warnings (is Intrinsically Easy), **Empirical Software Engineering (Under second round review)**, 2020.
- [2] Xueqi Yang, Zhe Yu, Junjie Wang and Tim Menzies, An Expert System for Learning Software Engineering Knowledge (with Case Study in Understanding Static Code Warning), **Expert Systems with Applications (accepted and to appear)**, 2020.
- [3] Amritanshu Agrawal, Xueqi Yang, Rishabh Agrawal, Xipeng Shen and Tim Menzies, Simpler Hyperparameter Optimization for Software Analytics: Why, How, When?, **Transactions on Software Engineering (Under review)**, 2020.
- [4] Jian Luo, Tao Hong, and Xueqi Yang, Fuzzy Support Vector Regression Models for Short-term Load Forecasting, **IEEE Transactions on Cybernetics (Under review)**, 2020.
- [5] Jian Luo, Xueqi Yang, Ye Tian and Wenwen Yu, Corporate and Personal Credit Scoring via Fuzzy Non-kernal SVM with Fuzzy within-class Scatter, **Journal of Industrial and Management Optimization, accepted**, 2018.

HONORS

- Research Assistant (Jan 2019 - Present), North Carolina State University.
- Teaching Assistant - C and Software Tools (CSC 230, 601) (2018 Fall), North Carolina State University.
- Undergraduate First Prize, Jan 2018.
- Honorable Mention in Interdisciplinary Contest in Modeling, Jan 2016.
- National Second Prize in China Undergraduate Mathematical Contest in Modeling, Sep 2015.