Wei Fu

http://weifu.us • 919-534-5251 • wfu@ncsu.edu • weifoo • weifoo

EDUCATION

North Carolina State University, Raleigh, NC — May. 2013 - May. 2018
Ph.D. Computer Science | Adviser: Dr. Tim Menzies

Beijing University of Posts and Telecommunications, Beijing, China — Sep. 2009 - Mar. 2012
M.S. Electrical Engineering

Nanjing University of Technology, Nanjing, China — Sep. 2005 - Jun. 2009
B.S. Electrical Engineering

ABOUT

- Worked on several end-to-end computer vision based systems with deep learning and traditional CV methods.
- Experienced in industry level machine learning systems from data labeling to model iteration (Python, C++).
- Experienced in machine learning and deep learning frameworks (TensorFlow, PyTorch, etc).

INDUSTRY EXPERIENCE

Applied Scientist | **Landing AI, Palo Alto, CA**
Data Pipeline, ML Algorithm Exploration, Model Iteration/Optimization, Error Analysis — May. 2018 - Now

- Build and implement a computer vision-based software solution for a multinational electronics contract manufacturer to inspect IC chips.
- Key contributor to design and implement for computer vision components of AI-harvesters to enable automatic harvesting.
- Drive to design, build and implement an end to end solution for a multinational biopharmaceutical company to do syringe inspections.

Research Intern | **ABB, Raleigh, NC**
Software Project Data Visualization and Exploration — May. 2016 - Aug. 2016

- Design mapping rules and building a web interface to visualize and manage data with asp.net and database.
- Based on ABB industrial data, building predictive models to improve quality of software development process.
- Compare and analyze the different characteristics of open source software data and proprietary data for building analytics models.

RESEARCH PROJECTS

NSF Funded: Search-based Software Engineering Research — Sep, 2014 - May, 2018
Research Assistant Under Dr. Tim Menzies, North Carolina State University, USA

- **Hyper-parameter Tuning for Software Analytics:** Software researchers and practitioners routinely use machine learning to explore software project data. However, they rarely tune hyper-parameter of their learners. By applying *Differential Evolution* on defect predictors, we find it (DE) can quickly find tunings that alter detection precision from 0% to 60%; DE can dramatically reduce clustering instability for LDA.
- **Differential Evolution v.s. Grid Search:** Grid search has become a de-facto hyper-parameter tuning technique for machine learning algorithms over years. However, for software analytics, we show that Differential Evolution as a tuner has better performance than grid search and also $210X$ faster.
- **Supervised v.s. Unsupervised Learning:** There's a debate about choices of defect predictors. In this project, we show that, in practice, unsupervised learning is not stable for deploying defect prediction, some supervised data is required to prune weaker models when building effort-aware just-in-time defect predictors.
- **Simple Techniques for Software Analytics:** *Deep Learning* has become a buzzword in both academia and industry. It seems that every single task should be solved by deep learning. In this project, we revisited a SE task recently solved by deep learning. However, after applying differential evolution-based parameter tuning on SVM, our results outperform the deep learning method in terms of performance metrics and also *84X* faster.

NSF Funded: Transfer Knowledge between Software Projects — Aug, 2015 - Dec, 2017
Research Assistant Under Dr. Tim Menzies, North Carolina State University, USA

- **Heterogeneous Defect Prediction:** For new software projects, historical data is missing. To build defect predictors, we proposed that historical data with different metrics from different projects can be used to build software quality models to predict quality of the target project. By using the mathematical models, we identify categories of data sets as few as 50 instances are enough to build a defect prediction model.
- **Bellwether Effect in Software Analytics** We find a "bellwether" effect in software analytics. Given N data sets, we find there always one data set produces the best predictions on all the others. This "bellwether" data set then can be used for all subsequent predictions.

SELECTED PUBLICATIONS

- A. Agrawal, **W. Fu**, D. Chen, X. Shen, and T. Menzies. *How to" DODGE" Complex Software Analytics?.* **TSE**, 2019. →http://tiny.cc/Dodge.
- D. Chen, **W. Fu**, R. Krishna, and T. Menzies. *Applications of psychological science for actionable analytics.* **FSE'18**. →tiny.cc/wfufft.
- **W. Fu** and T. Menzies. *Easy over Hard: A Case Study on Deep Learning.* **FSE'17**. →tiny.cc/wfuDL.
- **W. Fu** and T. Menzies. *Revisiting Unsupervised Learning for Defect Prediction.* **FSE'17**. →tiny.cc/wfuOneWay.
- JC Nam, **W. Fu**, S. Kim, T. Menzies, and L. Tan. *Heterogeneous Defect Prediction*. **TSE**, 2017. →tiny.cc/wfuHDP.
- **W. Fu**, T. Menzies, X. Shen, *Tuning for Software Analytics: is it Really Necessary*. **IST**, 2016. → tiny.cc/wfuTuning.
- R. Krishna, T. Menzies, and **W. Fu**. *Too Much Automation? the Bellwether Effect and Its Implications for Transfer Learning.* **ASE'16**. →tiny.cc/wfuBellwether.