Title: A Partial Reproduction of Hyperparameter Optimization in Defect Prediction

Original Authors: Wei Fu, Tim Menzies, Xipeng Shen, titled, “Tuning for Software Analytics: is it Really Necessary?” IST 2016

Reproduced Paper Authors: Huy Tu, Vivek Nair, titled, “Is One Hyperparameter Optimizer Enough?”, SWAN 2018

What: The black art hyperparameter tuning was used by Fu for the defect prediction problem to simply, automatically, and effectively configure the data miners’ optimal settings (for Random Forest and Decision Tree). In our paper, we reproduced their work to compare against Grid Search, Random, and Bayesian Optimization hyperparameter optimization techniques (for also k-Nearest Neighbors and Support Vector Machine). Note that these works satisfy the ACM definition of “reproduction” since we coded up their algorithms from scratch.

Why: Fu et al. demonstrates how to automate the optimal configuration with DE search improved the performance of defect predictors against the default configuration. It was important in our goal of surveying and contrasting all state-of-the-art hyperparameter optimizers to extend the extensive definition of tuning, especially in defect prediction field. Therefore, Fu’s work was reproduced.

How and Where: The code of Fu et al. work is not open source. We reproduced the code from the pseudocode given in the paper. Reproduced package of DE for defect prediction problem is now available at https://github.com/ai-se/hyperall/tree/master/DE. Our whole work package is already open-sourced and could be found at https://github.com/ai-se/hyperall.

Discussion: The authors did propose the necessary pseudocode for DE which made it easier for us to take an effort to reproduce the package. In the future, we would like to recommend that every researcher should be open source including the code, and the results. This will reduce the time and effort to reproduce the implementation and verify the results. State of the art progress would be more accessible for other researchers to know, use, and reproduce the results by having a pre-print version of the paper available digitally (especially before the work is being accepted).