Software Defect Prediction Using Support Vector Machine

By

Haneen Ghaleb Abu Al-Hija'a

Supervisor

Dr. Mohammed Al-Azzeh

Co-Supervisor

Dr. Fadi Al-Masalha

Master Thesis

Submitted in Partial Fulfilment of the Requirements for the Master's Degree in Computer Science

Deanship of Research and Graduate Studies

Applied Science Private University

Amman – Jordan

2019-2020

Abstract

Software defect prediction is an important activity during software development lifecycle as it can help managers to identify the defect-proneness modules. Thus, it can reduce the test cost and the risk, assign testing resources efficiently and enhance the work plan. To make sure if the software is defective or not, there are many classification methods that can be used such as Decision Tree, Recognition, Support Vector Machine, Neural Network (ANN), Naive Bayes. However, Support Vector Machine (SVM) has not been used extensively for such problem, because when used different dataset shows none stable, needs too many parameters tuning to adjust its configuration, including kernel functions and there are many versions of SVM have been proposed such as c-SVM, epsilon SVM and nu-SVM which affects the outcome of proposed models. Therefore, this research attempts to study the performance of SVM with different kernels. Kernel functions have a great impact on the accuracy of SVM. Various public datasets from PROMISE project will be used to run our experiments and test our hypothesis. The results demonstrate that there is no kernel function that can give stable performance across different experimental settings. In addition, the use of feature subset selection using PCA did improve performance of kernel functions over some datasets.

Keywords: Software Defect Prediction, Support