

Machine learning techniques for software bug prediction: a systematic review

ABSTRACT

The goal of software bug prediction is to identify the software modules that will have the likelihood to get bugs by using some fundamental project resources before the real testing starts. Due to high cost in correcting the detected bugs, it is advisable to start predicting bugs at the early stage of development instead of at the testing phase. There are many techniques and approaches that can be used to build the prediction models, such as machine learning. This technique is widely used nowadays because it can give accurate results and analysis. Therefore, we decided to perform a review of past literature on software bug prediction and machine learning so that we can understand better about the process of constructing the prediction model. Not only we want to see the machine learning techniques that past researchers used, we also assess the datasets, metrics and performance measures that are used during the development of the models. In this study, we have narrowed down to 31 main studies and six types of machine learning techniques have been identified. Two public datasets are found to be frequently used and object-oriented metrics are the highly chosen metrics for the prediction model. As for the performance measure, both graphical and numerical measures are often used to evaluate the performance of the models. From the results, we conclude that the machine learning technique can predict the bug, but there are not many applications in this area that exist nowadays. There are a few challenges in constructing the prediction model. Thus, more studies need to be carried out so that a well-formed result is obtained. We also provide a recommendation for future research based on the results we got from this study.

Keyword: Software bug prediction; Machine learning techniques; Literature review