

TEST CASE PRIORITIZATION USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

Andrado Mudiyanseilage Don Nisal Suranaka

A dissertation submitted in partial fulfillment of the requirement for
Bachelor of Science (Honours) degree in Computer Science

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2021

Abstract

Regression testing is the method of retesting software that has been updated or extended during software development. Usually, the execution of a series of test cases is used to conduct regression testing. However, it can be expensive and time-consuming to keep the regression suite up to date. Furthermore, executing all of the test cases in a regression suite could take a long time. Quality engineers in the industry are constantly confronted with this problem, and they must find bugs as soon as possible.

To overcome this issue, test case prioritization is one technique that is used to execute regression testing efficiently. Test case prioritization can greatly improve regression testing through organizing test cases according to their goals and optimizing their execution. Although there are numerous methods for prioritizing test cases, they all need knowledge of the internal source code (white-box testing). In industry, black box testing is an important requirement that typically does not allow access to source code.

This research project introduces a novel supervised machine learning and natural language processing-based technique in order to prioritizing the test cases which is currently a manual process of most industrial companies. This proposed system prioritizes the test cases into several priority levels as High, Medium and Low, which will be a perfect companion for quality engineers to make their life easier.

Keywords: Test Case Prioritization, Machine Learning, Natural Language Processing, SVM, TF-IDF, Tokenization