To Design a Framework for Test Suite Optimization in Regression Testing

ABSTRACT
of
THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

(A Central University)

Lucknow

FOR THE DEGREE OF

Doctor of Philosophy

In

INFORMATION TECHNOLOGY

By

Shilpi Singh
(Enrolment No. 1377/15)

Under the Supervision of

Dr. Raj Shree

Department of Information Technology

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
LUCKNOW-226025, INDIA

2018

ABSTRACT

Leading-edge technologies demand software to be built on a solid foundation of requirements, development processes, project management methodologies and quality practices. Such software-intensive systems demand high-quality product. However, it is estimated that software developers approximately spend half of the development cost and time on testing to release a quality product. Better software testing practices can cut down the cost incurred to develop quality products and also avert software failures. Regression testing is an important testing procedure used in validating modifications introduced in a system during software maintenance. It is expensive, yet an important process. As the test suite size is very large, system retesting consumes a large amount of time and computing resources. Unfortunately, there may not be sufficient resources to allow for the re-execution of all test cases during regression testing. This leads to focusing on proper test optimization techniques. Test suite reduction and test case prioritization are such techniques that address the problem of optimizing test suites. Test case optimization using reduction and prioritization techniques aim to improve the effectiveness of regression testing. Test suite reduction minimizes the test suite size by discarding duplicate and obsolete test cases according to some selected test criteria. Whereas, prioritization re-orders the test cases so that the most beneficial test cases are executed at the earliest, with higher priority.

In regression testing, software testing and retesting is done frequently. Growing of software in regression testing add a new test case to test new functionality, which is added to the existing software. These evolutions of software, which is denoted as version, create a redundant test case in the test suite. Test case becomes redundant due to more test cases satisfying the same requirement or covering same code. Availability of limited resources and time force to detect those redundant test cases which exercise the same requirement. The process of removing these redundant test cases is called as test suite minimization. The reduced test case which is derived after removal of a redundant test case is called as a representative set.

Test case prioritization techniques organize the test cases in a test suite, allowing for an increase in the effectiveness of testing. A primary performance goal of the system, the fault detection rate, is a measure of how quickly faults are determined during the testing process. An improved rate of fault detection can provide faster and productive feedback about the quality of the software application under test.

The details of the literature study reveal that the most of the reduction and prioritization strategies proposed in the literature are coverage based which do not always give a satisfactory result in terms of fault detection rate. Moreover, few reductions and prioritization strategies consider the similarity-based approach to optimize the test cases. The study also suggests that the use of multiple coverage criteria is very beneficial for the optimization process. This recommends that a combination of more than one testing criteria should be used for determining the optimal representative set, thus promoting the multi-objective heuristics to solve the optimization problem effectively.

In this sense, the main objective in this thesis is to improve the process of test suite optimization by proposing a strategy based on similarity and multi-coverage criteria in the context of code based testing aiming to maximize the fault coverage. In the context of minimization or reduction, the idea is to identify the degree of similarity among the test cases and keep in the suite the most different ones that together can meet a set of test requirements, and at the same time maintaining some redundancy in the reduced suite with the applicability of the multiple criteria. Whereas, similarity-based test case prioritization techniques primarily concentrated on test case diversity that helps to detect more faults. The technique evaluates the similarity degree for each test case pair and accordingly assigns the execution order to the test cases in a test suite.

This thesis presents a new similarity-based greedy approach, which involves the combination of regression testing techniques: minimization, and prioritization both. The main focus is on multiple regression activities with multiple criteria rather than using only a single activity to produce an optimal solution. The clustering method is also incorporated in this work, which could simplify and enhance the minimization and prioritization task. To evaluate the effectiveness of the strategy, we performed an experimental investigation together with an eminent heuristic Harrold Gupta and Soffa (HGS), considering the testing measures of the minimized test suite size and fault

coverage. The results show that similarity-based greedy approach with multiple coverage criteria can be quite effective in terms of fault detection loss of reduced test suite without much affecting the percentage of suite size reduction.

A new similarity-based test suite optimization (SB-TSO) algorithm using multiple coverage criteria is proposed in this thesis. The approach identifies the diverse test cases by comparing the distances between the test case pair with the help of three important coverage criteria i.e. statement, branch, and MC/DC. A cluster of similar test cases is generated using agglomerative clustering method. Within each cluster, optimal test cases are identified and to make the representative test suite more effective the test cases are ranked according to their assigned weight. To assess the effectiveness of the proposed approach, performance is evaluated and compared to the existing approaches. The result shows that the proposed method generates more optimal test cases that satisfy maximum coverage, minimum FDE loss, and overall size is also reduced.

This thesis presents several similarity-based prioritization techniques based on pair-wise selection strategy. Pair-wise comparison of test cases is a fundamental strategy to inspect test cases and choose an association between them in a finite test set. The proposed approach evaluates the similarity degree for each test case pair in three levels and accordingly assigns the execution order to the test cases in a test suite. The results of the experimental study show the improvement in fault detection rate that let developers initiate debugging and amending faults before the release of any software product. The comparative analysis reveals that some of the proposed techniques can attain higher fault detection rate, in terms of APFD, in comparison with the other existing techniques and random ordering.

The proposed approaches have been validated using ten different software programs. The applications were developed in C++ and necessary test cases were designed and executed for getting test related data such as code coverage in terms of statement, branch and, MC/DC. To collect the coverage information of test cases for each selected criterion, the subject program and the source code were instrumented. The proposed approaches also use hand-seeded faults for the subject programs to measure the fault detection rate of the optimized test suite.

The results of the experimental study show that the proposed approaches efficiently optimize the test suite by identifying the optimal subset of test cases from a large pool of test cases and reordering the test cases to improve the FDE thereby enhancing the quality of software under test.