DYNAMIC SMOKE TESTING
DYNAMIC REGRESSION TEST CASE SELECTION AND PRIORITIZATION

Yasitha Nuwan Mallawaarachchi

(168245A)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2020

DYNAMIC SMOKE TESTING
DYNAMIC REGRESSION TEST CASE SELECTION AND PRIORITIZATION

Yasitha Nuwan Mallawaarachchi

(168245A)

Thesis submitted in partial fulfilment of the requirements for the degree Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2020

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other media. I retain the right to use this content in whole or part in future works.

..... 22/05/2020.....

Yasitha Nuwan Mallawaarachchi Date

The above candidate has carried out research for the Masters/MPhil/PhD thesis/ dissertation under my supervision.

....

Dr. Charith Chitraranjan Date

ABSTRACT

With the advancement and increasing popularity of agile software development practices in large scale software development projects, frequent product releases are encouraged so that clients can actively participate in the software development life cycle (SDLC) by providing early feedback on developed features. This approach leads to iterative shorter cycles of development and continuous integration. So, the importance of regression testing and regression test suite is well emphasised in such methodologies. Regressions have become the most widely used approach in maintaining the quality of continuously changing software systems.

Even though the agile SDLC requires faster regression feedback given the shorter length of the release cycles, size and the complexity of the regression test suites increases over time; hence execution time keeps on growing. Therefore, it is not practical to run the regression test suite on every code change. In turn, it has become a significant dilemma in current regression testing. Therefore, it is essential to implement a regression testing strategy which is highly selective but accurate, to ensure the committed code changes does not inflict any ill behaviour on the current working software before it is merged and released for client feedback. To achieve this objective, it is critical to find out the distinct effects on behaviour that have impacted the software at the earliest during the continuous integration (CI) cycle. This research is focused on selecting and prioritizing the most suitable test cases from the regression test suite to detect any behaviour that is no longer intact due to the code change. Also, the capability of employing machine learning principles to learn and identify the most impactful characteristics of test cases is considered as another key objective of this study.

Keywords: Regression Test, Selection, Prioritization, Machine Learning, Clustering

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr Charith Chitraranjan for the continuous support of my MSc study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my MSc study.

I am grateful for the support and advice given by Dr Indika Perera, by encouraging continuing this research till the end. Further, I would like to thank all my colleagues for their help in finding relevant research material, sharing knowledge and experience and for their encouragement.

I am as ever, especially indebted to my parents for their love and support throughout my life. Finally, I wish to express my gratitude to all my colleagues at MillenniumIT Pvt. Ltd, for the support given me to manage my MSc research work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
BBD	Block Branch Diagram
CI	Continuous Integration
CR	Change Request
DB	Database
E2E	End to End
IR	Information Retrieval
ML	Machine Learning
OOP	Object Oriented Programming
ORD	Object Relational Diagram
OSD	Object State Diagram
PCA	Principle Component Analysis

QA	Quality Assurance
SDLC	Software Development Life Cycle
TF-IDF	Term Frequency – Inverse Document Frequency
WSS	Within-cluster Sum of Square