A FEATURE RANKING ALGORITHM IN PRAGMATIC QUALITY FACTOR MODEL FOR SOFTWARE QUALITY ASSESSMENT

RUZITA AHMAD

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)
UNIVERSITI UTARA MALAYSIA
2013
Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Abstrak

Kualiti perisian adalah satu bidang penyelidikan yang penting dan telah mendapat perhatian dikalangan komuniti kejuruteraan perisian terutama dalam mengenal pasti atribut penting dalam proses pembangunan perisian. Tesis ini menerangkan penyelidikan asli dalam bidang model kualiti perisian dengan memperkenalkan algoritma Feature Ranking Algorithm (FRA) untuk model Pragmatic Quality Factor (PQF). Algoritma yang dicadangkan mampu memperbaiki kelemahan model sedia ada dalam mengemaskini dan mempelajari kombinasi atribut untuk penaksiran kualiti perisian. Teknik penaksiran sedia ada kurang keupayaan untuk menyenaraikan atribut mengikut keutamaan dan keupayaan pembelajaran data yang boleh meningkatkan proses penaksiran kualiti. Tujuan kajian ini adalah untuk mengenal pasti dan mencadangkan penggunaan teknik dalam bidang Kepintaran Buatan ke arah meningkatkan proses penaksiran kualiti dalam model PQF. Oleh itu, algoritma FRA yang menggunakan Feature Ranking Technique (FRT) telah dibina dan prestasi algoritma FRA telah dinilai. Metodologi yang digunakan terdiri daripada kajian teori, reka bentuk rangka kerja formal untuk kualiti perisian pintar, mengenal pasti kesesuaian ciri-ciri FRT untuk penyenaraian atribut, pembangunan dan penilaian algoritma FRA. Penaksiran atribut telah bertambah baik dengan menggunakan algoritma FRA yang mengandungi formula untuk mengira keutamaan atribut dan diikuti oleh adaptasi pembelajaran melalui aplikasi Java Library for Multi Label Learning (MULAN). Hasil kajian menunjukkan bahawa prestasi algoritma FRA mempunyai kolerasi yang sangat kuat dengan model pakar iaitu model PQF. Ujian statistik menunjukkan bahawa FRA telah menghasilkan keputusan ketepatan yang lebih baik berbanding algoritma Kolmogorov-Smirnov Correlation Based Filter (KSCBF) iaitu 98% berbanding 83% masing-masing. Ujian statistik juga menghasilkan keputusan bagi algoritma FRA iaitu 0.052 adalah lebih baik berbanding dengan algoritma KSCBF iaitu 0.048. Ini menunjukkan bahawa keputusan FRA adalah lebih signifikan berbanding algoritma yang digunakan. Sumbangan utama kajian ini adalah dalam pelaksanaan teknik FRT yang memperkenalkan pengiraan Most Priority of Features (MPF) dalam algoritma FRA untuk teknik penaksiran tersebut. Kesimpulannya, penemuan kajian ini menyumbang kepada usaha penyelidikan baru dalam bidang pemilihan atribut dalam kualiti perisian.

Kata Kunci: Perisian kualiti, Algoritma FRA, Teknik Kepintaran Buatan, dan Mesin Pembelajaran
Abstract

Software quality is an important research area and has gained considerable attention from software engineering community in identification of priority quality attributes in software development process. This thesis describes original research in the field of software quality model by presenting a Feature Ranking Algorithm (FRA) for Pragmatic Quality Factor (PQF) model. The proposed algorithm is able to improve the weaknesses in PQF model in updating and learning the important attributes for software quality assessment. The existing assessment techniques lack the capability to rank the quality attributes and data learning which can enhance the quality assessment process. The aim of the study is to identify and propose the application of Artificial Intelligence (AI) technique for improving quality assessment technique in PQF model. Therefore, FRA using FRT was constructed and the performance of the FRA was evaluated. The methodology used consists of theoretical study, design of formal framework on intelligent software quality, identification of Feature Ranking Technique (FRT), construction and evaluation of FRA algorithm. The assessment of quality attributes has been improved using FRA algorithm enriched with a formula to calculate the priority of attributes and followed by learning adaptation through Java Library for Multi Label Learning (MULAN) application. The result shows that the performance of FRA correlates strongly to PQF model with 98% correlation compared to the Kolmogorov-Smirnov Correlation Based Filter (KSCBF) algorithm with 83% correlation. Statistical significance test was also performed with score of 0.052 compared to the KSCBF algorithm with score of 0.048. The result shows that the FRA was more significant than KSCBF algorithm. The main contribution of this research is on the implementation of FRT with proposed Most Priority of Features (MPF) calculation in FRA for attributes assessment. Overall, the findings and contributions can be regarded as a novel effort in software quality for attributes selection.

Keywords: Software Quality, FRA Algorithm, Artificial Intelligence (AI) Technique, and Machine Learning
Acknowledgement

In the name of Allah, the Beneficent, the Merciful. Alhamdulillah, grateful to Almighty Allah (SWT) for all the blessing He has bestowed upon me, lastly this thesis is finally completed.

A lot of thanks to Universiti Utara Malaysia for the supports and the facilities provided. My deepest gratitude and appreciation goes to my respected supervisors, Associate Professor Dr. Jamaiah Hj Yahaya and Dr. Siti Sakira Kamaruddin for never ending support, courage and helps along to complete this thesis.

Last but not least, I wish to express my deepest appreciation goes to my supportive husband, my loving mother, and my dear children, for believing in me, their continuous prayers and unconditional support. Lastly I would like to dedicate this work to the memory of my father, Allahyarham Ahmad bin Sulaiman (1938-2010) whom I unexpected lost during this study. Thank you all.
## Table of Contents

Permission to Use.................................................................................................................... i
Abstrak................................................................................................................................ ii
Abstract ................................................................................................................................... ii
Acknowledgement ................................................................................................................ iv
Table of Contents .................................................................................................................... v
List of Tables ........................................................................................................................... ix
List of Figures .......................................................................................................................... x
List of Appendices .................................................................................................................... xi
List of Abbreviations ............................................................................................................... xii

### CHAPTER ONE INTRODUCTION ........................................................................................................1

1.1 Overview ........................................................................................................................................ 1
1.2 Research Background ...................................................................................................................... 1
1.3 Research Problem Statement ........................................................................................................ 3
1.4 Research Motivation ...................................................................................................................... 5
1.5 Research Objectives ...................................................................................................................... 5
1.6 Research Scope .............................................................................................................................. 6
  1.6.1 Research Methodology ............................................................................................................. 6
  1.6.2 Theoretical Study ...................................................................................................................... 7
  1.6.3 Design of formal framework on intelligent software quality ................................................. 7
  1.6.4 Identify and proposed the Feature Ranking Technique (FRT) for an intelligence software quality model ................................................................. 7
  1.6.5 Construction of an Feature Ranking Algorithm (FRA) algorithm ........................................ 7
  1.6.6 Evaluation of study ................................................................................................................. 8
  1.6.7 Research Contribution ............................................................................................................ 8
1.7 Thesis Outline .............................................................................................................................. 8

### CHAPTER TWO LITERATURE REVIEW ..........................................................................................10

2.1 Introduction ................................................................................................................................... 10
2.2 Definition of Software Quality ..................................................................................................... 10
2.3 Software Quality Model .............................................................................................................. 12
Chapter Three Theoretical Study

3.1 Theoretical Study

3.1.1 Theoretical Study .................................................................63
3.1.2 Design of theoretical framework on intelligent software quality .......65
3.1.3 Identify and proposes the Feature Ranking Technique (FRT) for intelligent software quality model .................................................................68
3.1.4 Construction of an Feature Ranking Algorithm (FRA) Algorithm........69
3.1.5 Evaluation of Study ................................................................73

3.2 Summary of Research Methodology ...........................................74

Chapter Four Proposed Feature Ranking Algorithm (FRA) Algorithm

4.1 Introduction ................................................................................76
4.2 Background Issues .......................................................................76
4.3 Proposed FRA Algorithm ............................................................77

4.3.1 The Development of Most Priority of Features (MPF) Formula ........81
4.3.2 The Example of MPF calculation .............................................84
4.3.3 The Application of Classifiers .................................................85
4.3.4 The Step of FRA Algorithm ......................................................89

4.4 The Development of Kolmogorov-Smirnov Correlation Based Filter (KSCBF) Algorithm .................................................................91

4.5 Summary of Chapter Four ............................................................94

Chapter Five Data Analysis and Result

5.1 Introduction ................................................................................95
5.2 Performance Results .....................................................................96

5.2.1 Result of Experiment: Feature Ranking Algorithm (FRA) Algorithm ....96
5.2.2 Result of Experiment: Kolmogorov-Smirnov Correlation Based Filter (KSCBF) Algorithm .................................................................100

5.3 Evaluation Measurement ............................................................102

5.3.1 Human Expert Evaluation ......................................................102
5.3.2 Normalization of Data Performance ........................................103
5.3.3 Correlation Coefficient ..........................................................107
5.3.4 Statistical Significance Test (t-test) ..........................................109

5.4 Summary ....................................................................................111
CHAPTER SIX DISCUSSION AND CONCLUSION ....................................... 113
6.1 Overview ............................................................................................................ 113
6.2 Research Summary ............................................................................................ 113
6.3 Research Contribution ...................................................................................... 114
6.4 Limitation of the Research .............................................................................. 116
6.5 Future Work ...................................................................................................... 117
6.6 Summary .......................................................................................................... 118

REFERENCES ....................................................................................................... 119
Appendix A .............................................................................................................. 135
Appendix B .............................................................................................................. 136
Appendix C .............................................................................................................. 148
List of Tables

Table 2.1: Quality Characteristics in Previous Software Quality Models 6.1 ..........18
Table 2.2: Classification of Attributes and Weight Factor ...........................................23
Table 2.3: Comparison of Quality Score Obtained by Case X, Y and Z ....................26
Table 2.4: Summary of Static Quality Model and Dynamic Quality Model ...............35
Table 2.5: KSCBF Algorithm ....................................................................................57
Table 2.6: The Selected Elements in FRA Algorithm ...............................................61
Table 4.1: Efficiency of Database ..............................................................................84
Table 4.2: Algorithm of FRA .....................................................................................90
Table 4.3: Kolmogorov-Smirnov Correlation Based Filter (KSCBF) .........................91
Table 5.1: Example of Software Quality Attributes with Assigned Weight ..........96
Table 5.2: Result in Averaged Score and Standard Deviation ...................................97
Table 5.3: Result of FRA Algorithm .........................................................................98
Table 5.4: Result of Classification Accuracy for Two Redundant Attributes ..........99
Table 5.5: Final Ranking of FRA Algorithm ...........................................................100
Table 5.6: Result of Symmetrical Uncertainty (SU) Value .....................................101
Table 5.7: Final Ranking Result of Attributes Scores in KSCBF Algorithm ............101
Table 5.8: Final Result of PQF model, FRA and KSCBF Algorithm .......................103
Table 5.9: Normalization of Data .............................................................................104
Table 5.10: Analyzed Results of Correlation Coefficient of FRA to PQF model ... 108
Table 5.11: Analyzed Results of Correlation Coefficient of KSCBF to PQF model .....108
Table 5.12: Analyzed Results of Statistical Significant Test of FRA and KSCBF algorithm ...........................................................................................................111
Table 5.13: Final Results of Correlation Coefficient and Statistical Significant Test for FRA and KSCBF to PQF Model .................................................................111
List of Figures

Figure 2.1: Component of Pragmatic Quality Factor (PQF) Model ..................21
Figure 2.2: Decomposition of Functionality .................................................28
Figure 3.1: Inputs, Activities and Deliverables of Theoretical Study ..............64
Figure 3.2: Inputs, Activities and Deliverables of Design Framework ..........66
Figure 3.3: Theoretical Framework of Feature Ranking Algorithm (FRA) .......67
Figure 3.4: Inputs, Activities and Deliverables of Phase Three .....................69
Figure 3.5: Inputs, Activities and Deliverables of Construction of Feature Ranking Algorithm (FRA) ...............................................................71
Figure 3.6: Experimental Design .................................................................72
Figure 3.7: Inputs, Activities and Deliverables of Evaluation .........................74
Figure 4.1: Assessment Technique in FRA algorithm .....................................80
Figure 4.2: Executing Process of Unit Tests for MLkNN Classifier .................86
Figure 4.3: Executed of Unit Tests for MLkNN Classifier ...............................87
Figure 5.1: The Quality Attributes Ranking of FRA, KSCBF and PQF Model ....106
List of Appendices

Appendix A: Sample Data ....................................................................................... 135
Appendix B: Result of Experiment .......................................................................... 136
Appendix C: List of Publication .............................................................................. 148
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHS</td>
<td>Automatic Hybrid Search</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>ARFF</td>
<td>Attribute Relation File Format</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>BNS</td>
<td>Bi-Normal Separation</td>
</tr>
<tr>
<td>CBFS</td>
<td>Correlation Based Feature Selection</td>
</tr>
<tr>
<td>CBFSS</td>
<td>Consistency Based Feature Subset Selection</td>
</tr>
<tr>
<td>CBR</td>
<td>Case-Based Reasoning</td>
</tr>
<tr>
<td>CS</td>
<td>Chi-Square</td>
</tr>
<tr>
<td>DF</td>
<td>Document Frequency</td>
</tr>
<tr>
<td>DFM</td>
<td>Default F-Measure</td>
</tr>
<tr>
<td>DGM</td>
<td>Default Geometric Mean</td>
</tr>
<tr>
<td>ESD</td>
<td>Airforce Electronic System Division</td>
</tr>
<tr>
<td>FAS</td>
<td>Filter Attribute Selection</td>
</tr>
<tr>
<td>FCBF</td>
<td>Fast Correlation Based Filter</td>
</tr>
<tr>
<td>FRA</td>
<td>Feature Ranking Algorithm</td>
</tr>
<tr>
<td>FRT</td>
<td>Feature Ranking Technique</td>
</tr>
<tr>
<td>FS</td>
<td>Feature Selection</td>
</tr>
<tr>
<td>FSST</td>
<td>Feature Subset Selection Technique</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GR</td>
<td>Gain Ratio</td>
</tr>
<tr>
<td>GRNN</td>
<td>Generalized Regression Neural Network</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HFS</td>
<td>Hybrid Feature Selection</td>
</tr>
<tr>
<td>IBL</td>
<td>Instance Based Learning</td>
</tr>
<tr>
<td>IEEE</td>
<td>International Symposium on Requirement Engineering</td>
</tr>
<tr>
<td>IG</td>
<td>Information Gain</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization Standard</td>
</tr>
<tr>
<td>JRE</td>
<td>Java Runtime Environment</td>
</tr>
<tr>
<td>KNN</td>
<td>K-Nearest Neighbour</td>
</tr>
<tr>
<td>KS</td>
<td>Kolmogorov Smirnov</td>
</tr>
<tr>
<td>KSCBF</td>
<td>Kolmogorov-Smirnov Correlation Based Filter</td>
</tr>
<tr>
<td>K-S TEST</td>
<td>Kolmogorov Smirnov Two Sample Test</td>
</tr>
<tr>
<td>LEET</td>
<td>Large Experiment and Evaluation Tool</td>
</tr>
<tr>
<td>LOC</td>
<td>Lines of Code</td>
</tr>
<tr>
<td>LR</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>MATLAB</td>
<td>Matrix Laboratory</td>
</tr>
<tr>
<td>MI</td>
<td>Mutual Information</td>
</tr>
<tr>
<td>MLKNN</td>
<td>Multi Label K-Nearest Neighbour</td>
</tr>
<tr>
<td>MLOSS</td>
<td>Machine Learning Open Source Software</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi Layer Perceptron</td>
</tr>
<tr>
<td>MPF</td>
<td>Most Priority of Attribute</td>
</tr>
<tr>
<td>MULAN</td>
<td>Java Library for Multi Label Learning</td>
</tr>
<tr>
<td>NB</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>OA</td>
<td>Overall Accuracy</td>
</tr>
<tr>
<td>PQF</td>
<td>Pragmatic Quality Factor</td>
</tr>
<tr>
<td>PS</td>
<td>Probabilistic Search</td>
</tr>
<tr>
<td>QFD</td>
<td>Quality Function Deployment</td>
</tr>
<tr>
<td>RADC</td>
<td>Rome Air Development Centre</td>
</tr>
<tr>
<td>RAKEL</td>
<td>Random $k$-Labelstes</td>
</tr>
<tr>
<td>RS</td>
<td>Rough Sets</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Sciences</td>
</tr>
<tr>
<td>SQA</td>
<td>Software Quality Assurance</td>
</tr>
<tr>
<td>SQuaRE</td>
<td>Software Product Quality Requirement and Evaluation</td>
</tr>
<tr>
<td>STS</td>
<td>Spring Source Tool Suite</td>
</tr>
<tr>
<td>SU</td>
<td>Symmetrical Uncertainty</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
</tr>
<tr>
<td>WEKA</td>
<td>Waikato Environment Knowledge Analysis</td>
</tr>
<tr>
<td>WLLR</td>
<td>Weighted Log Likelihood Ratio</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

1.1 Overview

Chapter One presents the overall study and briefly explains the aims of the research. Several sections have been defined to classify and identify the purpose of this study. These include research background, problem statement of the research, research motivation, research objectives, scope of study and methodology.

1.2 Research Background

Nowadays, rapid development and diffusion of software quality is related to technologies in several industries. Statistics shows on insufficiently understood requirements accounted to 50% of errors. This was followed by design incorrectly understood from requirements, which accounted to 30% of errors. Hence, programming errors of system design contributed to 20% of errors (Humphrey et al., 1989). In fact, the organization has outlined the exactly errors in perfectly before they starts to develop a software product. Thus, Software Quality Assurance (SQA) is a very important domain in software development and its purpose is to find ways to reduce the rate and associated cost of failure from poor product and services (Humphrey et al., 1989).

In order to reduce errors in systems design and to fulfill user needs and requirements, the quality of systems development should be highlighted as an important goal. Normally, the standard level of quality is recommended by the International Organization for Standardization (ISO) and IEEE as well. ISO defines quality as the
The contents of the thesis is for internal user only
REFERENCES


Duch, W., Winiarski, T., Biesiada, J., & Kachel, A. (2003, June). Feature selection and ranking filters. *International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP)*. 251-254.


