Analogy-Based Software Project Effort Estimation

Mohammad Y. A. Azzeh

PhD

2010

Analogy-Based Software Project Effort Estimation

Contributions to Projects Similarity Measurement, Attribute Selection & Attribute Weighting algorithms for Analogy-Based Effort Estimation

Mohammad Y. A. Azzeh

A thesis submitted for the degree of
Doctor of Philosophy

Department of Computing
School of Computing, Informatics & Media
University of Bradford

2010

DECLRATION

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Signature_____

Abstract

Software effort estimation by analogy is a viable alternative method to other estimation techniques, and in many cases, researchers found it outperformed other estimation methods in terms of accuracy and practitioners' acceptance. However, the overall performance of analogy based estimation depends on two major factors: similarity measure and attribute selection & weighting. Current similarity measures such as nearest neighborhood techniques have been criticized that have some inadequacies related to attributes relevancy, noise and uncertainty in addition to the problem of using categorical attributes.

This research focuses on improving the efficiency and flexibility of analogy-based estimation to overcome the abovementioned inadequacies. Particularly, this thesis proposes two new approaches to model and handle uncertainty in similarity measurement method and most importantly to reflect the structure of dataset on similarity measurement using Fuzzy modeling based Fuzzy C-means algorithm. The first proposed approach called Fuzzy Grey Relational Analysis method employs combined techniques of Fuzzy set theory and Grey Relational Analysis to improve local and global similarity measure and tolerate imprecision associated with using different data types (Continuous and Categorical). The second proposed approach presents the use of Fuzzy numbers and its concepts to develop a practical yet efficient approach to support analogy-based systems especially at early phase of software development. Specifically, we propose a new similarity measure and adaptation technique based on Fuzzy numbers.

We also propose a new attribute subset selection algorithm and attribute weighting technique based on the hypothesis of analogy-based estimation that assumes projects that are similar in terms of attribute value are also similar in terms of effort values, using row-wise Kendall rank correlation between similarity matrix based project effort values and similarity matrix based project attribute values. A literature review of related software engineering studies revealed that the existing attribute selection techniques (such as brute-force, heuristic algorithms) are restricted to the choice of performance indicators such as (*Mean of Magnitude Relative Error* and *Prediction Performance Indicator*) and computationally far more intensive. The proposed algorithms provide sound statistical basis and justification for their procedures.

The performance figures of the proposed approaches have been evaluated using real industrial datasets. Results and conclusions from a series of comparative studies with conventional estimation by analogy approach using the available datasets are presented. The studies were also carried out to statistically investigate the significant differences between predictions generated by our approaches and those generated by the most popular techniques such as: conventional analogy estimation, neural network and stepwise regression. The results and conclusions indicate that the two proposed approaches have potential to deliver comparable, if not better, accuracy than the compared techniques. The results also found that Grey Relational Analysis tolerates the uncertainty associated with using different data types. As well as the original contributions within the thesis, a number of directions for further research are presented.

Most chapters in this thesis have been disseminated in international journals and highly refereed conference proceedings.

Acknowledgments

I would first and foremost like to thank Dr. Daniel Neagu and Prof. Peter Cowling for their dedicated supervision and unwavering support, without which this thesis would never have got off the ground.

I would especially like to express my gratitude to my lovely wife Tamara for her love, patience, and support. I also dedicate this work to my parents who have been supportive and understanding throughout the years of my PhD study. This work is also dedicated to my lovely daughter Toleen, who provided an additional and joyful dimension to my life mission.

I would also like to record my thanks to my friends at University of Bradford who have made contributions to this thesis. I would also like to acknowledge the support and funding of Applied Science University, Jordan.

Publications

Published papers:

[1] Azzeh, M., Cowling, P. & Neagu, D. (2010), Software Stage-Effort Estimation Based on Association Rule Mining and Fuzzy Set Theory, in proceedings of the 10th IEEE international conference on computer and information technology (Accepted).

[2] Azzeh, M., Neagu, D. & Cowling, P. (2010), ***Fuzzy grey relational analysis for software effort estimation***, Empirical Software Engineering, Springer, Vol. 15, Issue 1, 60-90.

[3] Azzeh, M., Neagu, D. & Cowling, P. (2009), Software Effort Estimation Based on Weighted Fuzzy Grey Relational Analysis, in proceedings of 5th international conference on software predictor models, Co-located with ICSE'31, Article number 8.

[4] Azzeh, M., Neagu, D. & Cowling, P. (2008c), Adjusting Analogy Software Effort Estimation Based on Fuzzy Logic, in Proceedings of 3rd International conference on software and data technology ICSOFT (SE/MUSE/GSDCA), 127-132.

[5] Azzeh, M., Neagu, D. & Cowling, P. (2008b), Software Project Similarity Measurement Based On Fuzzy C-Means, in proceedings of 2nd International Conference on software process, Leipzig, Germany, 123-134.

[6] Azzeh, M., Neagu, D. & Cowling, P. (2008a), Fuzzy Feature Subset Selection for Software Effort Estimation, in proceedings of 4th international workshop on software predictors PROMISE'08 (part of ICSE'08), Leipzig, Germany, 71-78.

[7] Azzeh, M., Neagu, D. & Cowling, P. (2007), An Overview of Web Cost Estimation, in proceedings of the 8th Informatics Workshop, University of Bradford, 96-99.

Papers under review:

[8] Azzeh, M., Neagu, D. & Cowling, P., (2010), ***FGRA+: An Improved Method For Software Effort Estimation, Journal of Information & Software Technology***, Elsevier, Accepted subject to corrections.

[9] Azzeh, M., Neagu, D. & Cowling, P., ***Analogy-Based Software Effort Estimation based on Fuzzy Numbers***, Journal of System & Software, Elsevier, Submitted (under review)

[10] Azzeh, M., Neagu, D. & Cowling, P. (2010), Model Tree-Based Adaptation Strategy for Analogy Based Effort Estimation, 6th international conference on software predictor models (under review).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANGEL	Analogy based effort estimation tool
ANN	Artificial Neural Network
CBR	Case Based Reasoning
EBA	Estimation by Analogy
FCM	Fuzzy C- Mean
FN	Fuzzy Number
FGRA	Fuzzy Grey Relational Analysis
GFNSE	Generalized Fuzzy Number Software Estimation
FST	Fuzzy Set Theory
GRA	Grey Relational Analysis
ISBSG	International Software Benchmarking Standard Group
KFS	Kendall based Feature Selection
MMRE	Mean of Magnitude Relative Error
MdMRE	Median of Magnitude Relative Error
PRED	Prediction Performance indicator
SR	Stepwise Regression