Increasing the accuracy of software development effort estimation using projects clustering

V. Khatibi Bardsiri\(^1\) D.N.A. Jawawi\(^1\) S.Z.M. Hashim\(^1\) E. Khatibi\(^2\)

\(^1\)Department of Software Engineering, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia (UTM), Skudai 81310, Johor Bahru, Malaysia
\(^2\)Bardsir Branch, Islamic Azad University, Kerman, Iran

E-mail: dayang@utm.my

Abstract: Software development effort is one of the most important metrics that must be correctly estimated in software projects. Analogy-based estimation (ABE) and artificial neural networks (ANN) are the most popular methods used widely in this field. These methods suffer from inconsistent and irrelevant projects that exist in the software project datasets. In this paper, a new hybrid method is proposed to increase the accuracy of development effort estimation based on the combination of fuzzy clustering, ABE and ANN methods. In the proposed method, the effect of irrelevant and inconsistent projects on estimates is decreased by designing a new framework, in which all the projects are clustered. The quality of training in ANN and the consistency of historical data in ABE are improved using the proposed framework. Two large and real datasets are utilised in order to evaluate the performance of the proposed method and the obtained results are compared to eight other estimation methods. The promising results showed that the proposed method outperformed the other methods on both datasets. The performance metrics of mean magnitude of relative error (MMRE) and the percentage of the prediction (PRED) (0.25) have been improved by average of 51 and 127% in the first dataset, as well as 52 and 94% in the second dataset.

1 Introduction

Since the role of software in today’s business market is undeniable, accurate estimation of software development size and effort is very important. Planning, development, construction and all aspects of software projects are affected by relevant estimations. During the recent decades, various methods have been proposed to increase the accuracy of estimates in this field. Selecting the best method seems to be impossible because the performance of each method depends on numerous factors, such as available information, development techniques, application type, project features and so on. An estimation method may be suitable only for a special type of software projects. Basically, during the first stages of a software project, available information about the project’s features is incomplete, which may lead to inaccurate estimates. In addition, some specifications of software projects, such as unstable customer requirements, confusing relationship among project features, lack of standards, high diversity of development techniques as well as rapid progress of hardware platforms, make the estimation process to be more difficult than the other projects. These barriers have inspired researchers to propose reliable and flexible estimation methods in terms of software development effort. Numerous researches have been conducted to improve the accuracy of development effort estimation in the recent years and various methods have been proposed in this area. Mathematical and statistical techniques, data mining methods and soft computing techniques have been widely used to estimate software development effort. In general, when a software project starts, requirements, scope and the real situation of the project may be ambiguous and hard to determine. Therefore relying on fixed and algorithmic estimation models may be useless in this field. On the other hand, comparison of new project with past similar completed projects through non-algorithmic methods is the most common solution considered in the previous studies. Although, this may seem so simple, prior studies show that employing non-algorithmic methods based on comparison and analysis of past projects are consistent with complex and dynamic nature of software projects. Among non-algorithmic methods, neural networks and analogy-based estimation (ABE) have been extensively employed in this area, because they could deal with software project features, and estimate development effort accurately. In this paper, a combination of these two methods is presented using fuzzy clustering technique to improve the accuracy of software development effort estimates.

This paper is organised in eight sections. In Section 2, motivation of the research is explained. Next, Section 3 describes the previous research works in the field of current study, whereas fuzzy clustering and analogy methods are explained in Sections 4 and 5, respectively. Then, the proposed method is stated in Section 6. Section 7 includes evaluation procedure, whereas the experimental results are presented in Section 8. Eventually, Section 9 presents the conclusion for this research.