Decision-Tree Models for Predicting Time Performance in Software-Intensive Projects

Nermin Sökmen, TÜBİTAK, Kocaeli, Turkey
Ferhan Çebi, Faculty of Management, Istanbul Technical University, Istanbul, Turkey

ABSTRACT

Initial requirements, new requirements and technical issues are the main factors that have a great effect over the software product development process. Difficulties resulting from incomprehensibility of initial requirements indicate two sub-factors: Deviations determined during analysis of initial requirements and deviations resulting from interpretation of requirements inaccurate and incomplete. New requirements being received from customers or end users during the development process affect the project performance. There can be problems during the implementation of product specifications, inaccurate formation of architectural design and technical solutions, incorrect coding of functions, or wrong realization of interfaces. The general technical problems cover all problems arising from technical reasons and the negative situations they create on the project. During the design and implementation activities of software intensive projects, these three main factors can be affected by other sub-factors. The aim of this study is to examine the factor classes which influence these three problem domains with CHAID (Chi-squared Automatic Interaction Detection) technique. Time deviations caused by initial requirements, new requirements and general technical problems are selected as target variables. In this research, 75 projects that develop software intensive products are studied to formalize the most accurate decision mechanism.

KEYWORDS

Decision-Tree, Project Management, Project Performance, Project Risks, Software Development, Software-Intensive Projects

1. INTRODUCTION

The quality of project planning affects the project success (Dvir and Lechler, 2004). According to Sarigiannidis and Chatzoglou (2014), the negative relationship exists between project quality and risk. Quantifying impact of risks has a positive association with the overall project schedule targets of new product development program (Oehmen et al., 2014). It is very important to identify risks and assess their likelihood and potential effect on project schedule before the project started.

Software development risks have been widely discussed in the literature (Barki et al., 1993; Jiang and Klein, 2000; Schmidt et al., 2001; Wallace and Keil, 2004; Wallace et al., 2004b; Han and Huang, 2007). Difficulties resulting from initial and new requirements and problems related to technical issues have always been important risk items in software development projects (Sökmen et al., 2014). Liu et al. (2011) indicate that the performance of software development process is affected by the requirements that are not defined completely, correctly and comprehensively. Incorrect, unclear, inadequate, ambiguous or unusable requirements increase problems or risks in software development projects (Wallace et al., 2004a). Goal changes that can be due to changes in requirements or lack of ability to meet existing requirements may cause change in project plans in order to meet the
new requirements (Dvir and Lechler, 2004). Requirements instability negatively affects the project performance with extending the project time, causing remaking of the substantial part of work and affecting the other project features (Ferreira et. al., 2009).

Technical sub-system risks determined by uncertainty and technical complexity have strong impact on project management risks (Wallace et al., 2004b). In fact, functional development risk is directly associated with increase in system development risk (Na et al., 2007). According to Carvalho et al. (2015), project complexity has a significant effect on schedule variation. The inaccurate formation of architectures, interfaces and data structures, incorrect coding and environment uncertainty are some of the negative situations that affect the performance of the entire project unless they are managed well. These negative situations are evaluated together under a single factor called as “general technical problems”.

There are only few studies that examined the schedule performance of the sub-processes or tasks in development projects in the literature. In this study, the difficulties resulting from analysis of initial requirements are examined with the problems arising from new requirements and general technical difficulties. Sökmen et al. (2014) used multiple linear regression analysis to examine the effects of these three factors (or sub-processes) on project schedule in software-intensive projects. This study is expected to go one step further by analyzing the impact of sub-factors. The aim of this study is to find out the effect of the sub-factors influencing these three main factor classes. Schedule delays or time deviations in the project due to initial requirements, new requirements and general technical problems are the target variables. The study aims to build three predictive models to help reveal the most effective factors on the success of the software development processes.

In the study, the factors including the status of requirements, product innovation, project manager competency, team competency, preliminary preparation studies, reuse, customer support and product type are selected as the factors that could be effective on these destructive factors, namely, problems caused by initial requirements, problems caused by new requirements and general technical problems. Here, the most effective factors on the target variable are determined by using CHAID (Chi-squared Automatic Interaction Detection) algorithm. With the help of the hierarchical tree structure, the effects of the factors on the target variable are found out.

The paper is organized as follows. Section 2 introduces related works on effort and time estimation models. It also discusses problems related to initial requirements, new requirements and general technical issues and their possible effects on development process. This is followed by presentation of the research methodology and model variables. Section 4 presents three decision tree models which reveal the effect of factors acting on these main factors. Section 5 provides a discussion of the results. Finally, Section 6 presents the future work.

2. LITERATURE REVIEW

2.1. Effort and Time Estimation Models

The literature review results summarized in Table 1 in the Appendix show that a wide variety of techniques have been used to estimate software effort and cost. Finnie at al. (1997) took Function Point (FP) and Unadjusted FP (UFP) as input parameters and used regression analysis and neural network techniques to estimate software development effort. While Fenton and Neil (1999) applied Bayesian Belief Networks (BBN) to model effort estimation, Pendharkar (2005) benchmarked the performance of Bayesian point software development forecasts with popular nonparametric neural network and the CART (Classification and Regression Tree) approaches. Heiat (2002) compared the neural network estimation method to regression analysis for software effort estimation.
A Taxonomy of Stakeholders: Human Roles in System Development
www.igi-global.com/chapter/taxonomy-stakeholders-human-roles-system/22695?camid=4v1a

Knowledge Flow Networks and Communities of Practice in Business Process Management
www.igi-global.com/chapter/knowledge-flow-networks-communities-practice/54534?camid=4v1a