

Characterization of Runaway Software Projects Using Association Rule Mining

Sousuke Amasaki, Yasuhiro Hamano, Osamu Mizuno, and Tohru Kikuno

Graduate School of Information Science and Technology
Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
amasaki@computer.org,
{y-hamano, o-mizuno, kikuno}@ist.osaka-u.ac.jp

Abstract. In this paper, characteristics of a runaway project are revealed based on combinations of risk factors which appear in the project. Concretely, an association rule mining technique is applied with an actual questionnaire data to induce rules that associate combinations of risk factors with runaway status of software projects. Furthermore, the induced rules are integrated and reduced based on a certain rule obtained from experts' perception to simplify the representation of characteristics of a runaway project. Then, for confirming the effectiveness of this characterization, it is evaluated how many runaway projects in distinct data set were identified by the reduced rules. The result of the experiment suggested that the induced rules are effective to characterize runaway projects.

Keywords: association rule mining, risk factors, project characterization.

1 Introduction

Recently, software development projects have often been put in a very risky situation because of increasing demand for high quality, short period, and low cost. Thus, detecting signs of problems at an early stage of the software project is important. So, much research has been carried out on the detection of problem signs of a software development project [1, 2]. So far, we have proposed methods to predict runaway status of projects using responses of questionnaire [3, 4].

During past works, the following question is recognized: "How can the runaway-prone projects avoid the runaway status?" One answer to this question is a set of parameters in a logistic regression model. Although these parameters can indicate dominant risk factors, the number of parameters included in a model is restricted by the property of the regression analysis [5]. On the other hand, some risk factors are simultaneously found in the same runaway project frequently. As the result, we consider that there exist combinations of risk factors which significantly affect to the runaway status of projects.

In this paper, we propose an approach using a data-mining technique to induce combinations of risk factors related to runaway status of software projects significantly. The proposed approach consists of the following two phases:

Phase 1: To characterize runaway projects by combinations of risk factors, the association rule mining technique [6] was applied for response data of a questionnaire checking potential risks, collected in a certain company.

Phase 2: To evaluate the effectiveness of the rules induced in **Phase 1**, the rules were applied to distinct response data collected in the same company.

From the result of experiment, we can conclude that the induced combinations of risk factors can characterize runaway projects.

2 Preliminaries

2.1 Definition of Runaway Projects

In the company cooperating with us, the Software Engineering Process Group (SEPG) has tried to analyze characteristics of runaway projects. In this paper, a definition established by the SEPG is used to judge a “runaway project”. The definition is as follows: 1) Cost and duration is out of a certain range, and, 2) A project fell into an uncontrollable situation during development. The SEPG in the company judges whether these conditions are satisfied.

1. Requirements		Eval.
Q1.1	Ambiguous requirements.	
Q1.2	Insufficient explanation of the requirements.	
Q1.3	Misunderstanding of the requirements.	
Q1.4	Lack of commitment regarding requirements between the customer and the project members.	
Q1.5	Frequent requirements or specification changes.	
2. Estimations		Eval.
Q2.1	Insufficient awareness of the importance of estimation.	
Q2.2	Insufficient skills or knowledge of estimation methods.	
Q2.3	Insufficient estimation for the implicit requirements.	
Q2.4	Insufficient estimation for the technical issues.	
Q2.5	Lack of stakeholders commitment for estimation.	
3. Planning		Eval.
Q3.1	Lack of management review for the project plan.	
Q3.2	Lack of assignment of responsibility.	
Q3.3	Lack of breakdown of the work products.	
Q3.4	Unspecified project review milestones.	
Q3.5	Insufficient planning of project monitoring and controlling.	
Q3.6	Lack of project members' commitment for the project plan.	
4. Team Organization		Eval.
Q4.1	Lack of skills and experience.	
Q4.2	Insufficient allocation of the resources.	
Q4.3	Low morale.	
5. Project Management Activities		Eval.
Q5.1	Lack of resource management of project managers throughout a project.	
Q5.2	Inadequate project monitoring and controlling.	
Q5.3	Lack of data needed to keep track of a project objectively.	

For each item, please answer with one of "Strongly agree", "Agree", "Neither agree nor disagree", and "Disagree".

Fig. 1. Questionnaire for risk identification

2.2 Questionnaire to Identify Problems

For early identification of runaway status, a questionnaire was constructed that includes possible risk factors in the software development projects. The questionnaire is distributed to and filled out by a project manager. All questions in the questionnaire and applicable answer types are shown in Figure 1. The questionnaire consists of five viewpoints: requirements, estimations, planning capability, team organization, and project management activities. Each sub-item regarding risk factors in the questionnaire must be filled in according to the Likert scale.