Implementation of using classification Data Mining Techniques for Software Cost Estimation

Snehal A. Deshmukh  
ME  
Department of Computer Science & Engg.  
PRMIT&R, Badnera

Prof. S. W. Ahmad  
Assistant Professor  
Department of Computer Science &  
Engg. PRMIT&R, Badnera

Abstract:-Cost estimation is one of the biggest problems faced by software industry. It is necessary to have accurate cost estimation in order to conduct well budget. Under-estimation may lead to unexpected increase in budget, delay of project completion or its low quality, while over-estimation may lead to losing business opportunities. In this dissertation the idea of building data mining techniques into existing software cost estimation model such as COCOMO II  model is implemented. Data mining allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.

1. INTRODUCTION

Software Effort Estimation

Cost estimation is prediction of percentage and number of hours for the effort invested during a software project. Estimating the cost is very necessary and most analysed variable in recent years. It is used basically in project management. Software engineers were facing problem of effort predictions since 1950. Estimation overrun was occurring even for small projects. SLIM- Software Life Cycle Management and COCOMO- Constructive Cost Estimation are the basic models for effort estimation. Tremendous growth of software system trade resulted in new technologies. Actual estimation is often a difficult task. Cost estimation techniques are generally classified into algorithmic and non-algorithmic techniques. Using data mining on projects of organization for cost estimation has its benefits, such as ease of understanding and controlling collected data.
low’ to ‘extra high’) that express the impact of the driver and a corresponding set of effort multiplier. The nominal level always has an effort multiplier (EM) of 1.00, which does not change the estimated effort. So a cost driver's qualitative rating is translated into a quantitative one for use in the model. The COCOMO II model can be used to estimate effort and schedule for the whole project or for a project that consists of multiple modules. The size and cost driver ratings can be different for each module.

Fig 1: - functional diagram of existinmethodology

4. CONCLUSION

Data mining has proven itself as a valuable tool in many areas, however, current data mining techniques are often far better suited to some problem areas than to others, therefore it is recommend to use data mining in most companies to help managers to make accurate cost estimation. The results in this dissertation suggest that building data mining classification techniques into existing software estimation techniques such as COCOMO II can effectively improve the performance of a proven method. COCOMO II model is used to estimate the cost, effort and schedule when planning a new software development. The effort Equation is used to estimate the number of person / month. The degree of effort and duration increase with increase in the size of project, cost drivers and scale drivers, then the average staffing increase linearly. The COCOMO II model is a good guide to estimate the requirements and maintenance of software. It takes project, product, personnel and hardware attributes into account when predicting effort required. The time to complete a project is not proportional to the number of people working on the project. On the basis of results it can be said that cost drivers and scale factors perform important role in this estimation. Best structure of data mining methods with meticulous data sets will be utilized for accessing the accurate SCE models.

5. REFERENCES


