_____

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.

1.2 Objectives

- To identify the cost drivers and cost factors of Cocomo II model.
- To apply data mining classification techniques for software cost estimation.
- To determine what resources to commit to the project and how well these resources will be used as projects can be easier to manage and control when resources are better matched to real needs.
- To improve the overall business plan of a software organization
- To classify and prioritize development projects with respect to an overall business plan.
- To provide more accurate estimated software project to the customers.

2. SYSTEM ANALYSIS AND DESIGN FOR PROPOSED SYSTEM

2.1 Problem Statement

Effort estimation is necessary for many people and different departments in an organization. At various point of project lifecycle well-defined effort estimation is essential. The aim of this dissertation work is to identify the important cost drivers in the past project data with the help of data mining classificaton techniques .Cost drivers are multiplicative factors of cost estimation model that determine the effort required to complete software project. In the analogy estimation models, the cost drivers are the base of cost estimation models. They estimate the new project with compare the past project data or cost drivers and set the value of cost drivers in the new projects.

2.2 Data Mining Algorithms

The k-means Algorithm:

The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. Here's how the algorithm works [9]

1. The algorithm arbitrarily selects k points as the initial cluster centres ("means").

2. Each point in the dataset is assigned to the closed cluster, based upon the Euclidean distance between each point and each cluster centre.

3. Each cluster centre is recomputed as the average of the points in that cluster.

4. Steps 2 and 3 repeat until the clusters converge. Convergence may be defined differently depending upon the implementation, but it normally means that either no observations change clusters when steps 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters.

3. SYSTEM IMPLEMENTATION

The methodology is simple .Two different-different fields data mining and the software engineering are combined and accurate cost of the project with the help of past project data whose cost or effort is known is generated. Data mining classification techniques and cocomo II model is used for software estimation. Cocomo II model is the analogy model. In this model the new project cost estimation by comparing the past project data is done. The feature of new project and past project is very similar. In his application , the past project data whose efforts are already given is used. With the help of clustering similar group of cost drivers is formed . The cost drivers are very helpful to predict the cost of the new projects. COCOMO II has 7 to 17 multiplicative factors that determine the effort required to complete a software project. All cost drivers have qualitative rating levels ('extra

_____

_____

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.
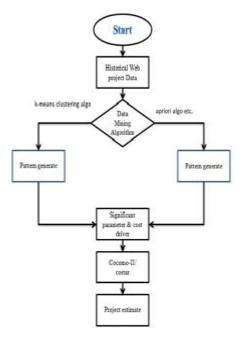

Fig 1:- functional diagram of existinmethodology

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

[1] Sonam Bhatia, Varinder Kaur Attri, "Implementing Decision Tree for Software Development Effort Estimation of Software Project", *International Journal of Innovative Research in Computer and Communication Engineering Vol.* 3, *Issue 5, May 2015.*

[2] Saiqa Aleem1 , Luiz Fernando Capretz1 ,and Faheem Ahmed, BENCHMARKING MACHINE LEARNING TECHNIQUES FOR SOFTWARE DEFECT DETECTION *International Journal of Software Engineering & Applications (IJSEA), Vol.6, No.3, May 2015.*

[3] Jyoti Shivhare, "Effectiveness of Feature Selection and Machine Learning Techniques for Software Effort Estimation" June 2014.

[4] Mohita Sharma1, Neha Fotedar2, Software Effort Estimation with Data Mining Techniques- A Review *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY 3(3): March, 2014].*

[5] Hasan Al –Sakran, "Efficient approach to develop software Cost estimation model using case-based reasoning and agent technology",*Journal of Theoretical and Applied Information Technology 31st May 2014. Vol. 63 No.3.*

[6] Ms. K. Gayathiri 1, Dr. T. Nalini 2 ,Dr. V. Khanaa , " Data mining techniques for software effort estimation to improve cost efficiency", *International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 4 April, 2013 Page No. 1215-1220.*

[7] Sweta Kumari, "Comparison and Analysis of Different Software Cost Estimation Methods",(IJACSA) *International Journal of Advanced Computer Science and Applications, Vol. 4, No.1, 2013.*

[8] T. C. Sharma & M. Jain ,(2013) "WEKA approach for comparative study of classification algorithm", *International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 4, 7 pages.*

[9] Narendra Sharma, Ratnesh Litoriya , "Incorporating Data Mining Techniques on Software Cost Estimation: Validation and Improvement", *International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 3, March 2012).*

[10] Ruchika Malhotra, "Software Effort Prediction using Statistical and Machine Learning Methods", (IJACSA) *International Journal of Advanced Computer Science and Applications, Vol. 2, No.1, January 2011.*

[11] Sumeet Kaur Sehra1, Yadwinder Singh Brar2, and Navdeep Kaur3 , "Soft Computing Techniques For Software Project Effort Estimation", *International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624.Vol 2, Issue 3, 2011, pp 160-167.*

[12] Prasad Reddy P.V.G.D, Hari CH.V.M.K and Srinivasa Rao, ―Multi Objective Particle Swarm Optimization for Software Cost Estimation,‖ *International Journal of Computer Applications, Vol.-32, 2011 .*

[13] J. S. Pahariya, V. Ravi, M. Carr, M. Vasu, ―Computational Intelligence Hybrids Applied to Software Cost Estimation‖, *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM),Vol. 2, pp. 104-112, 2010.*

[14] C. S. Reddy, K. Raju, ― An Improved Fuzzy Approach for COCOMO's Effort Estimation using Gaussian Membership Function‖, *Journal of Software, VOL. 4, NO. 5, pp. 452-459, 2009 .*

[15] Vinaykumar, K., Ravi, V., Carr, M. and Rajkiran, N. "Software cost estimation using wavelet neural networks," *Journal of Systems and Software, 2008, pp. 1853-1867.*

_____