ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

Intelligent Software Effort Estimation through a Multiple Comparisons Algorithm

R.Manimegalai[1], J.Selvakumar[2], M.Rajaram[3]

Software Engineering, Sri Ramakrishna Engineering College Coimbatore, India[1, 2, 3]

Abstract: Software Cost Estimation (SCE) can be related as the process of estimating the most realistic effort necessary to accomplish a software project. The rapidly improved demand of large-scaled and complex software systems leads managers to settle SCE as one of the most vital actions that is closely associated with the success or failure of the whole development procedure. Propose an analytical framework based on a multiple comparisons algorithm in order to rank several cost estimation methods, determining those which have important dissimilarity in accuracy, and clustering them in nonoverlapping groups. To overcome this problem proposed an improved cost effort estimation methods and compared using appropriate statistical procedures. In this paper we develop an intelligent Expert System that supports all type of software development regardless of their type - either using conventional computer languages or component based visual languages. Classification is most common method used for finding the mine rule from the large database. We also extend our work to C5.0 algorithms applied on customer database for classification. The proposed framework is applied in a large-scale setup of comparing 12 prediction models over three datasets.

Keywords: Software Cost Estimation, C5.0, Expert System, Scott-Knott, prediction model.

I. INTRODUCTION

The significant and the importance role of Software Cost Estimation (SCE) to the well-equalized management of an upcoming project are certainly represented through the introduction and utilization of a large number of methodologies during the past decades [1].

The rapidly improved demands of large-scaled and complex software systems leads managers to settle SCE as one of the most essential actions that is closely associated with the success or failure of the whole development progress. Defective estimates can be proved catastrophic to both customers and developers since they can cause the delay of the product deliverables or, even worse, the cancellation of a contract. Due to the above-specified are necessity, importance has been concentrated on the open research issue of the selection of the "best" estimation method. According to an extended systematic survey of studies [1], the more familiar research field of SCE is the introduction and assessment of estimation approaches. On the other hand, the several of prediction techniques are also correlated with incompatible and inconsistent findings regarding the superiority of one method over others. The most determining factor for these controversial outcomes seems to be an inherent fundamental of prediction systems, i.e., their strong dependency on the kind of available data (types and number of project attributes and sample size) used in method fitting [2]. The complexity of building an accurate method swiftly improves if we assumed the alternative difference of a generic estimation model (e.g., regression analysis). In several studies researchers have based their inferences on a small number of datasets, so generalization of findings may be quite misleading.

Furthermore, there is a continuous consideration and lack of convergence concerning the appropriateness of the error measures used for the comparison of different techniques [3]. Although Mean Magnitude of Relative Error (MMRE) has been criticized as a problematic accuracy measure to elect the "best" method [4], it continues to be assumed as the main indicator for the performance of SCE models. A certain lacks of several past studies is comparison without using applicable analytical hypothesis testing. This can start to incorrect outputs and groundless generalizations regarding the predictive accuracy of estimation approaches [5]. Although comparison of models without analytical tests may lead to unsound outcomes [6], many current papers still base their findings solely on single indicators [7].

Another source of bias can also be the analytical method that is used when comparing multiple prediction approaches. In the case of a simple comparison between two competitive methods, the null hypothesis is examined via a classical analytical test (i.e., paired t-test or Wilcoxon signed rank test). With more than two comparative methods, the meaning of "important difference" becomes more difficulties, and the issues

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

correlated with it are known in statistics as the "multiple comparisons methods". Due to the huge number of expected cost estimation approaches, it is mandatory for project managers to methodically base their choice of the more accurate method on well-established analytical techniques in order to diminish the uncertainty threatening the estimation progress [8]. However, to the best of our knowledge, the issue of simultaneous comparisons among multiple prediction methods has not been studied yet in the sense that there is no analytical method which can determined the important differences between a number of cost estimation models and at the same time be able to rank and cluster them, nominated the best ones.

All of the problems examined above lead us to conclude that there is an imperative demand to assume what the state of the art in analysis is before trying to derive conclusions and unstable results regarding the superiority of a prediction approach over others for a particular dataset. The answer to this issue cannot establish a unique solution since the notion of "best" is quite subjective. In fact, an expert can always rank the prediction methods according to a predefined accuracy measure, but the critical problem is to determine how many of them are apparently best, in the sense that their various from all the others is analytically significant. Hence, the research question of finding the "best" prediction procedures can be restated as an issue of determining a subset or a group of best methods.

The goal of the paper [9] is to propose a statistical framework for comparative SCE experiments determining multiple prediction methods. It is worth remarking that the setup of the recent study was also stimulated by an analogous try to dealing with the issue of comparing classification methods in Software Defect Prediction, a research area that is also closely associated to the improvement of software quality [10].

The proposed methodology of this paper [9] is based on the analysis of a Design of Experiment (DOE) or Experimental Design, a basic statistical tool in many applied research areas such as engineering, financial, and medical sciences. In the field of SCE it has not yet been used in a systematic manner. Generally, DOE refers to the process of planning, designing, and analyzing an experiment in order to derive valid and objective conclusions effectively and efficiently by taking into account, in a balanced and systematic manner, the sources of variation [11]. In the present study, DOE analysis is used to compare different cost prediction models by taking into account the blocking effect, i.e., the fact that they are applied repeatedly on the same training-test datasets.

Our aim of this paper is the estimation is done accurately, it results in error decrease. Estimation process reflects the reality of project's progress. It avoids cost/budget or schedule overruns. This process is quite simple which takes a few inputs. This assessment framework helps inexperienced team improve project tracking and estimation. C5.0 algorithm is an extension of C4.5 algorithm. C5.0 is the classification algorithm which applies in big data set.

II. RELATED WORKS

Miyazaki et al. [12] demanded that the "de facto" MMRE accuracy determined tends to advance methods that underestimate the actual effort, while Kitchenham et al. [3] indicated the several of accuracy measures as a primary source of inconclusive studies. Toward this direction, Foss et al. [4] investigated the basis of this criticism through a simulation study, proposed alternative accuracy indicators and concluded that there was a need for applying well established analytical techniques when conducting SCE experiments.

Myrtveit et al. [8] extended the above-mentioned findings and pointed out that inconsistent results were not caused only by accuracy measures but also by unreliable research methods. Through a simulation study, they studied the consequences of three main ingredients of the comparison progress: the single data sample, the accuracy indicator, and the cross-validation technique. Furthermore, they supported possible explanations for the lack of convergence, such as the small sample size of many software studies and the splitting of training and test sets in the validation method that affected the comparison, even for samples drawn from the same populations. The researchers also inferred that the conclusions on "which model is best" to a large degree depend on the chosen accuracy indicator and that different indicators can lead to contradicting results.

Recently, Menzies et al. [13] studied the issue of "conclusion instability" through the COSEEKMO toolkit that supported 15 parametric learners with row and column preprocessors based on two various sets of tuning parameters. In order to decide the predictive power of the alternative models, they used performance ranks, whereas the selection of the best method was based on a heuristic metric. Their experiments on COCOMO-style datasets concluded that there were four best approaches and not just a single option.

The Scott-Knott test presented here was used in another context in [14], for combining classifiers applied to large databases. Specifically, the Scott-Knott test and other statistical tests were used for the selection of the best subgroup among various classification algorithms

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

and the subsequent fusion of the methods' decisions in this subgroup via simple models, like weighted voting. In that study extensive experiments with very large datasets showed that the Scott-Knott test provided the highest accuracy in difficult classification issues. Hence, the choice of the test for the present paper was motivated by former results obtained by one of the authors.

In [15], Demsar discussed the problem of analytical tests for comparisons of various machine learning classifiers on multiple datasets reviewing various analytical techniques. The model proposed as more suitable is the nonparametric analogue of ANOVA, i.e., the Friedman test, along with the corresponding Nemenyi post hoc test. The Friedman test ranks all the classifiers separately for each dataset and then uses the average ranks of procedure to test whether all classifiers are equivalent. In case of differences, the Nemenyi test performed all the pairwise comparisons between classifiers to determine the significant differences. This model was used by Lessmann et al. [10] for the comparison of classifiers for prediction of defected modules.

The technology described in this paper [9], apart from the fact that was applied to a several issues, i.e., the SCE where cost and prediction errors were continuous variables.

Specifically, the methodology presented in [9] ranks and clusters the cost prediction methods based on the errors measured for a particular dataset. Therefore, each dataset had its own set of "best" methods. That had been more realistic in SCE practice software implemented in the organization had its own dataset and wanted to find the methods that best fit its data rather than trying to find a globally best model which was unfeasible. Furthermore, the clustering as an output was various from the output of pairwise comparisons tests, like the Nemenyi test. A pairwise test, for instance, can possibly indicate that models A and B are equivalent, models B and C were also equivalent, but models A and C were different. The grouping of model B was therefore questionable. For larger numbers of models the overlapping homogeneous groups resulting from pairwise tests were ambiguous and problematic in interpretation. On the other hand, a ranking and clustering algorithm provided clear groupings of models, designating the group of best models for a particular dataset.

III. PROPOSED SYSTEM

A. Expertise Based Techniques

Delphi technique was derived from them. Under this model, project specifications are given to a few experts and their opinion taken. Steps:

1. Selection of Experts.
2. Briefing to the Experts
3. Collation of estimates from experts
4. Convergence of estimates and finalization

Selection of Experts: Experts are selected who have software development experience, which have worked and possess knowledge in application domain at hand; they may be from within or without the organization.

Briefing the Experts: The experts need to be briefed about the project. They need to know the objectives of estimation, explanation of project scope, completion and its nature in project bidding.

Collation of estimates received from experts: The experts are expected to give one figure for the development effort and optionally software size. Each oracle gives the opinion.

Convergence of estimates and finalization: Now the estimates are converged using either the statistical mode from opinions offered by experts or extreme estimates are interchanged i.e. higher estimate is given to expert who gave lowest figure estimate, lower estimate is given to expert who gave highest figure estimate, average estimate can be derived using arithmetical average.

$$T(e)=\{t(o)+4t(m)+t(p)\}/6 \quad (1)$$

$$Var2=\{t(p)-t(o)\}2/36 \quad (2)$$

The Scott-Knott test [16] is a multiple comparison methods based on principles of cluster analysis. The clustering refers to the treatments (methods or in our case models) being compared and not to the individual cases, while the criterion for clustering together treatments is the statistical significance of differences between their mean values. Our preference for the Scott-Knott test relies on a specific desirable characteristic of the method, i.e., that it is able to separate the models into nonoverlapping groups. In our case, the values of the response variable that is affected by the models are translated to expressions of the prediction errors derived from the models being compared. The algorithm we describe next is therefore able to rank and cluster prediction models according to their accuracy.

B.C5.0 Algorithm

It is an extension of C4.5 algorithm. C5.0 is the classification algorithm which applies in big data set. C5.0 is better than C4.5 on the efficiency and the memory. C5.0 model works by splitting the sample based on the field that provides the maximum information gain. The C5.0 model can split samples on basis of the biggest information gain field..The sample subset that is get from the former split will be split afterward. The process will

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

continue until the sample subset cannot be split and is usually according to another field. Finally, examine the lowest level split, those sample subsets that don't have remarkable contribution to the model will be rejected.

Information Gain:

Gain is computed to estimate the gain produced by a split over an attribute

Let S be the sample:

- C_i is Class I; i = 1,2,…,m
  $I(s_1,s_2,..,s_m)= - \Sigma\ p_i \log2\ (p_i)$
- Si is the no. of samples in class i
  Pi = Si /S, log2 is the binary logarithm
- Let Attribute A have v distinct values.
- Entropy = E(A) is
  $\Sigma\{(S1j+S2j+..+Smj)/S\}*I(s1j,..smj)$ j=1
- Where Sij is samples in Class i and subset j of Attribute A. $I(S1j,S2j,..Smj)= - \Sigma\ p_{ij} \log2\ (p_{ij})$
- Gain(A)=I(s1,s2,..,sm) - E(A)

Gain ratio then chooses, from among the tests with at least average gain,

The Gain Ratio= P(A)

$$\sum_{i}^{t} \frac{S_i}{S} \log\left(\frac{S_i}{S}\right). \quad (3)$$

Gain Ratio (A) = Gain (A)/P (A)

In customer membership card model C5.0 algorithm is used to split up data set & find out the result in the form of decision tree or rule set.

1) Splitting criteria used in c5.0 algorithm is information gain. The C5.0 model can split samples on basis of the biggest information gain.

2) Test criteria is decision tree have any number of branches available not fixed branches like CART.

3) Pruning method performed after creating decision tree i.e. post pruning single pass based on binomial confidence limits.

4) Speed of c5.0 algorithm is significantly faster & more accurate.

IV. EXPERIMENTAL RESULTS

This section provides details concerning the setup of the framework and the experimental design of the study. The basic idea of the experimental setup was to take into account: 1) different cost prediction methods covering a major part of the variety of the proposed methodologies that have appeared so far in the SCE literature and which are governed by a diversity of principles, 2) different datasets, and 3) different measures of error. Moreover, the experiment was designed to take into account the effect of training-test splitting of each dataset.

A. Comparative Prediction Models

The 12 selected methods can be grouped into three main categories that are regression-based models, analogy-based techniques, and machine learning methods.

B. Results

In order to address the disagreement on the performance measures, we apply the whole analysis on three functions of error that measure different important aspects of prediction techniques: accuracy and error.

1. Prediction Accuracy

More precisely, Absolute Error (AE) is used in order to evaluate the accuracy of models, whereas error ratio z has been adopted as a measure of bias accounting for underestimations (z < 1) or overestimations (z > 1) with an optimum value of 1. The most widely known MRE indicator was also used since, it provides a measure of the spread of the error ratio z.

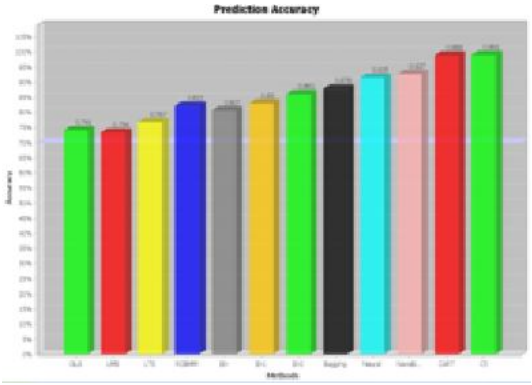

Fig. 1 Prediction Accuracy

In that graph Fig 1. Prediction Accuracy is compared for existing and proposed system. Methods are represented in x-axis and accuracy is represented in y-axis. Compare with all the methods our proposed systems method C.5 has better results than other methods.

2. Prediction Error

In our case, the values of the response variable that is affected by the models are translated to expressions of the prediction errors derived from the models being compared. The algorithm we describe next is therefore able to rank and cluster prediction models according to their accuracy.

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

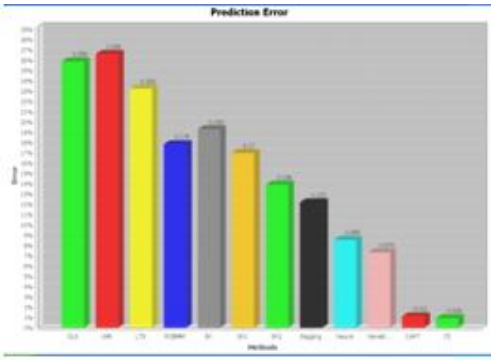

Fig. 2 Prediction Error

The diagram plots comparative models (x-axis) against the transformed mean errors (y-axis), whereby all methods are sorted according to their ranks. Compare with all the methods our proposed systems method C.5 has better results than other methods.

3. Scott-Knott Error

Generally, the application of Scott-Knott tests for three datasets reveals one of the most appealing findings of this study: Despite the large divergences of error functions among alternative prediction models, there is no statistical evidence that some methods differ significantly. Hence, the notion of the "best" estimation technique should be revised, whereas at the same time it is probably more proper to refer to the "best group of estimation techniques."

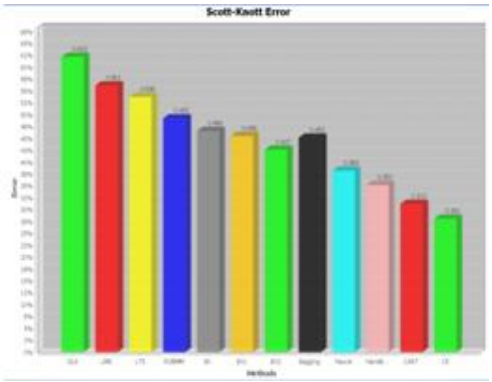

Fig. 3 Scott-Knott Error Graph

The results of the Scott-Knott procedure are also presented in a graphical manner for two cases (Fig. 3). The diagram plots comparative models (x-axis) against the transformed mean errors (y-axis), whereby all methods are sorted according to their ranks. Compare with all the methods our proposed systems method C.5 has better results than other methods.

V. CONCLUSION AND FUTURE WORK

In this paper, we deal with a critical research problem in SCE concerning the simultaneous comparison of alternative prediction methods. We checked the predictive power of 12 methods over three public domain datasets. The whole technique is settled on well-established analytical technologies taking into examination the multiple comparison issues. Keeping in mind the critical role of the adoption of reliable practices in the implement process for both project managers and customers, we proposed when quantified or empirical data is absent, then expertise based techniques are needed. The opinion of experts is taken, but the drawback with this technique is that the estimate is as good as the expert's opinion only. We also extend our work to provide the way for Decision making process of Customer for recommended the membership card. Here C5.0 & CART algorithms applied on customer database for classification. Both algorithms first applied on training dataset & created the decision tree, pruning method used for reducing the complexity then rule set are derived from decision tree. Same rules then applied on evaluation data set.

In future work, another interesting finding concerns the utilization of complicated and more sophisticated models. It seems that very often a linear model is adequate enough to catch the trend between effort and other cost drivers of projects. Therefore, in certain cases it may be useless to strive to introduce new, highly complicated algorithms which, in practice, just cannot provide any further improvement. Finally, it is our strong belief that new estimation techniques should be tested and compared using appropriate statistical procedures.

REFERENCES

[1] M. Jorgensen and M. Shepperd, "A Systematic Review of Software Development Cost Estimation Studies," IEEE Trans. Software Eng., vol. 33, no. 1, pp. 33-53, Jan. 2007.

[2] M. Shepperd and G. Kadoda, "Comparing Software Prediction Techniques Using Simulation," IEEE Trans. Software Eng., vol. 27, no. 11, pp. 1014-1022, Nov. 2001.

[3] B. Kitchenham, S. MacDonell, L. Pickard, and M. Shepperd, "What Accuracy Statistics Really Measure," IEE Proc. Software Eng., vol. 148, pp. 81-85, June 2001.

[4] T. Foss, E. Stensrud, B. Kitchenham, and I. Myrtveit, "A Simulation Study of the Model Evaluation Criterion MMRE," IEEE Trans. Software Eng., vol. 29, no. 11, pp. 985-995, Nov. 2003.

[5] N. Mittas and L. Angelis, "Comparing Cost Prediction Models by Resampling Techniques," J. Systems and Software, vol. 81, no. 5, pp. 616-632, May 2008.

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

[6] E. Stensrud and I. Myrtveit, "Human Performance Estimating with Analogy and Regression Models: An Empirical Validation," Proc. IEEE Fifth Int'l Software Metrics Symp., pp. 205-213, Nov.1998.

[7] B. Kitchenham and E. Mendes, "Why Comparative Effort Prediction Studies May Be Invalid," Proc. ACM Fifth Int'l Conf. Predictor Models in Software Eng., pp. 1-5, May 2009.

[8] I. Myrtveit, E. Stensrud, and M. Shepperd, "Reliability and Validity in Comparative Studies of Software Prediction Models," IEEE Trans. Software Eng., vol. 31, no. 5, pp. 380-391, May 2005.

[9] Nikolaos Mittas and Lefteris Angelis," Ranking and Clustering Software Cost Estimation Models through a Multiple Comparisons Algorithm", IEEE Trans. Software Eng, vol. 39, no. 4, April 2013.

[10] S. Lessmann, B. Baesens, C. Mues, and S. Pietsch, "Benchmarking Classification Models for Software Defect Prediction: A Proposed Framework and Novel Findings," IEEE Trans. Software Eng., vol. 34, no. 4, pp. 485-496, July/Aug. 2008.

[11] J. Antony, Design of Experiments for Engineers and Scientists. Butterworth-Heinenmann, 2003.

[12] Y. Miyazaki, M. Terakado, K. Ozaki, and H. Nozaki, "Robust Regression for Developing Software Estimation Models," J. Systems and Software, vol. 27, pp. 3-16, 1994.

[13] T. Menzies, O. Jalali, J. Hihn, D. Baker, and K. Lum, "Stable Rankings for Different Effort Models," Automated Software Eng.,vol. 17, no. 4, pp. 409-437, Dec. 2010.

[14] G. Tsoumakas, L. Angelis, and I. Vlahavas, "Selective Fusion of Heterogeneous Classifiers," Intelligent Data Analysis, vol. 9, no. 6, pp. 511-525, Dec. 2005.

[15] J. Dem_sar, "Statistical Comparisons of Classifiers over Multiple Data Sets," J. Machine Learning Research, vol. 7, pp. 1-30, 2006.

[16] A. Scott and M. Knott, "A Cluster Analysis Method for Grouping Means in the Analysis of Variance," Biometrics, vol. 30, no. 3, pp. 507-512, Sept. 1974.