Prospects and Challenges of using Machine Learning Algorithms for Software Quality Assessment and Prediction

Ananthi Sheshasaayee
Research Supervisor
Department of Computer Science
Quaid E Millath College for Women, Chennai

Roby Jose
Research Scholar
Department of Computer Science
Quaid E Millath College for Women, Chennai

ABSTRACT
Maintainability of the software is one of the key quality while evaluating software product. Of the overall software development cost, major stake is employed at the maintenance phase. Maintenance time of software is always greater than its development time, so it becomes essential to measure the maintainability of software so that maintenance operational time can be brought down.While going over the prevailing knowledge of literature it is understood that maintenance cost can be regulated by using software metrics at the design phase. There is substantial works in proving that machine learning algorithms is a suitable alternative for many domains of computational sciences including software engineering. This paper is aimed at carrying out a detailed study on the usage of machine learning approaches in the prediction, assessment and evaluation of software maintainability

General Terms
Software Engineering, Literature Study, Software Quality

Keywords
Machine learning algorithms, Maintainability Prediction, Software Maintainability Prediction Models and Metrics, Software metrics.

1. INTRODUCTION
Software maintainability refers to the activities carried out for the system being operational, once the system is installed. These activities arouse of the fact that the software undergoes changes after the installation due to many reasons. Software system or component can be modified to correct faults, improve performance or other attributes or adapt to a changed environment [1]. Maintainability in a system should be considered as one of the vital quality characteristic. A lot of research work has been carried out to measure the maintainability. To measure maintainability, it is necessary to identify the set of attributes or factors that bear on the efforts needed to make specified modification. Maintainability definition according to IEEE is the ease with which a software system or component can be modified to correct faults, improve performance or other attributes, or adapt to a changed environment is maintainability. Maintainability is the capability of a system to evolve from its current state to its future desired state. It is termed as the most difficult and costliest activity due to its inherent involvement in making predictions about the future [1]. Lientz and Swanson [2] identified four major categories of maintenance based on the context and purposes on which changes are made to the software.

Almost sixty percent of the total software development cost is shelled out for maintenance activities. The amount of resource, effort and time used up for software maintenance is a lot higher than what is being spent on development of the software. If companies focus on the development of easily maintainable software's potentially huge cost and effort can be resuscitated. In this context software metrics, is the support. With the help of software metrics at the design, maintainability prediction [3] can be made using various tools and processes. Accurate prediction of maintainability can provide guidance of maintenance process efficiency, help managers in various business decision activities and staff allocation et al..

1.1 Evaluating Maintainability
Maintainability is an external quality attribute. The external quality attribute which is not quantifiable is to be calculated using some surrogate measure which are quantifiable. No fixed criteria exist to measure or evaluate maintainability. Some studies make use of number of changes [3] [4] made to the software code as the surrogate measure of software maintainability.

1.2 Literature Review for Maintainability Models
Models to review the maintainability for the software applications had been designed by many researchers. Metrics based maintainability prediction has been divided into two sections based on the approach followed for developing the model. The first section assesses the literature for statistical approach and the second section evaluates the literature based on the machine learning approaches.

1.2.1 Statistical Approach
Li [5] used Multiple Linear Regression (MLR) model which successfully predicted maintenance effort and put aside those metrics which can strongly influence the maintainability calculation of object oriented systems.Niessink and Vliet [6] used standard statistical techniques viz. Principal Component analysis and multiple regression analysis to explain the variance in the maintenance effort. These statistical techniques are used to explore relationships between characteristics of change to identify the metrics which highly impacted prediction of adaptive maintenance for object oriented system. Fioravanti and Nesi [7] presented a metric analysis. The metrics and model was validated against real time data and the validation identified that metrics can be used forcalculating maintainability. Dagpinar et al [8] conducted study on

empirical data to establish the relation between object oriented software metrics and maintainability. The study recorded significant impact of direct coupling metric and size on software maintainability whereas other parameters like cohesion, inheritance and indirect coupling were less significant for software maintainability. Misra [9] applied linear regression to a suite of twenty design and code measures to obtain the indicators of software maintainability.

1.2.2 Machine Learning Approach

Thwin and Quah [10] build the object oriented maintainability prediction model using neural networks. The study concluded that neural network based prediction models can be used to forecast quality of systems developed in object oriented environment..Stamelos et al [11] propose the use of Bayesian belief networks (BBNs), already applied in other software engineering areas, to support expert judgment in software cost estimation.. Koten and Gray [12] used Bayesian Belief Network (BBN) to predict object oriented software maintainability. The results suggest that the Bayesian network model can predict maintainability more accurately than the regression-based models for one of the two systems that were studied. Zhou and Leung [13] adopted multiple adaptive regression splines (MARS), to build software maintainability prediction models using the metric data collected from two different object-oriented systems. The study suggested that the MARS is the best modeling technique in terms of accuracy when compared to other widely used modeling approaches. Prasanth et al [14] employed the fuzzy repertory table (FRT) technique to acquire the necessary domain knowledge of testers from which the software complexity analysis is made. Regression analysis is then used to predict maintainability from the product's code complexity. Elish [15] used Tree Net, a novel advance in data mining for building model of prediction for software maintainability and proved that Tree Net based model of maintainability provided competitive results when compared with other models. Kaur et al [16] suggested the use of soft computing approaches like Fuzzy Inference System (FIS) and Adaptive Neuro Fuzzy Inference System (ANFIS) for the calculation of software maintenance.Ping [17]used Hidden Markov Model (HMM) to simulate maintenance of software using metrics to measure the maintenance behavior.Jin and Liu [18]presents the application of support vector machine(SVM) and unsupervised learning in software maintainability prediction using object-oriented metrics and had shown that SVM and clustering technique were useful in constructing software maintainability predictor. Malhotra et al [19] proposed maintainability models based on machine learning algorithms like Probabilistic Neural Network, Genetic algorithms and Group Method of Data Handling (GMDH). The empirical study concluded GMDH as a sound alternative among machine learning algorithms for the prediction of software maintenance.

2. ANALYSIS

On the basis of the study performed in the previous section, upon the work done by different researchers in predicting maintainability for various application development domains, following tables (Table I and Table II) can be modeled. An analysis of the previous section indicates that many studies were dedicated to the usage of machine learning algorithms in the prediction, assessment and evaluation of maintainability of object oriented systems. While the traditional approaches rely heavily on the accuracy of the data used for modeling, there are approaches like fuzzy logic in machine learning paradigm that do not depend on the data. This approach uses less data or almost no data for the modeling purpose. So this study concludes that there is the possibility of utilizing machine learning approaches for the assessment of software quality factors such as maintainability, reliability et al.

Table 1. Maintainability modeling using machine learning approaches

Serial No	Authors	Method utilized
1	Stamelos,Ioannis	Bayesian Belief Networks
2	MMT Thwin, Quah	General Regression Neural Network
3	Koten, Gray	Bayesian Networks
4	Zhou, Leung	MARS(Multi Adaptive Regression Spline)
5	M O Elish, K O Elish	Tree nets
6	Kaur et al	ANN, Fuzzy Inference System(FIS), Adaptive Neuro Fuzzy Inference System(ANFIS)
7	L Ping	Hidden Markov Model(HMM)
8	Jin , Liu	Support Vector Machine
9	Prasanth et al	Fuzzy Repertory Table and Regression Analysis
10	Malhotra, Chug	Genetic Algorithm and Probabilistic Neural Networks(PNN)

Table 2. Maintainability modelling using traditional approaches

Serial No	Authors	Method utilized
1	Li, Henry	Regression based
2	Niessink, Vliet	Regression Models
3	Fioravanti,	Multiple Linear Regression

	P Nesi	
4	Dagpinar, Jhanke	Best subset Regression model
5	SC Misra	Correlation and Linear Regression

3. LIMITATIONS

The study was concentrated on maintainability modeling of software. The other quality attributes were not considered in the study. There may be studies that involved modeling of more than one quality attribute such as maintainability and reusability. Another limitation can be attributed as selection of a few international journals and conference proceedings. This may have led to the omission of certain studies which are relevant to the subject studied.

4. CONCLUSION

A topic much to the interest of software engineering researchers is the maintainability model. Traditionally the maintainability model takes in software metrics and statistical approaches. The advanced machine learning approaches and bio inspired computing techniques like Artificial Neural networks and Genetic algorithms, now find their own place in software engineering. These approaches that possess better prediction accuracy as compared with traditional analytic approach make them the much preferred ones in the research. Much of the discussed works are on object oriented metrics. In software engineering domain, Object Oriented approach has been extended to Aspect Oriented paradigm to achieve modularity. The modularity in this approach is achieved by the separation of concerns to a module termed as aspect. So as future direction of research, applicability of these approaches to aspect oriented metrics is planned and the prediction accuracy of the model based on aspect oriented dynamic metrics are to be undertaken.

5. REFERENCES

[1] Edelstein, D. Vera. "Report on the IEEE STD 1219–1993—standard for software maintenance." ACM SIGSOFT Software Engineering Notes 18.4 (1993): 94-95.

[2] Lientz, B. P., Swanson, E. B., 2000. Software Maintenance Management, Addison - Wesley Reading, MA.

[3] Li, Wei, and Sallie Henry. "Object-oriented metrics that predict maintainability."Journal of systems and software 23.2 (1993): 111-122

[4] S. Henry and S. Wake, "Predicting Maintainability with Software Quality Metrics," Software Maintenance: Research and Practice, vol. 3, pp. 129-143, 1991.

[5] Li, Wei. "Another metric suite for object-oriented programming." Journal of Systems and Software 44.2 (1998): 155-162

[6] Niessink, Frank, and Hans Van Vliet. "Two case studies in measuring software maintenance effort." Software Maintenance, 1998. Proceedings., International Conference on. IEEE, 1998.

[7] Fioravanti, Fabrizio, and Paolo Nesi. "Estimation and prediction metrics for adaptive maintenance effort of object-oriented systems." Software Engineering, IEEE Transactions on 27.12 (2001): 1062-1084.

[8] Dagpinar, Melis, and Jens H. Jahnke. "Predicting maintainability with object-oriented metrics-an empirical comparison." 20th Working Conference on Reverse Engineering (WCRE). IEEE Computer Society, 2003.

[9] Misra, Subhas Chandra. "Modeling design/coding factors that drive maintainability of software systems." Software Quality Journal 13.3 (2005): 297-320.

[10] Thwin, Mie Thet, and Tong-Seng Quah. "Application of neural networks for software quality prediction using object-oriented metrics." Journal of systems and software 76.2 (2005): 147-156.

[11] Stamelos, Ioannis, et al. "On the use of Bayesian belief networks for the prediction of software productivity." Information and Software Technology 45.1 (2003): 51-60.

[12] Van Koten, Chikako, and A. R. Gray. "An application of Bayesian network for predicting object-oriented software maintainability." Information and Software Technology 48.1 (2006): 59-67.

[13] Zhou, Yuming, and Hareton Leung. "Predicting object-oriented software maintainability using multivariate adaptive regression splines." Journal of Systems and Software 80.8 (2007): 1349-1361.

[14] Prasanth, N. Narayanan, S. Ganesh, and G. A. Dalton. "Prediction of maintainability using software complexity analysis: An extended FRT."Computing, Communication and Networking, 2008. ICCC 2008. International Conference on. IEEE, 2008

[15] Elish, Mahmoud O., and Karim O. Elish. "Application of tree net in predicting object-oriented software maintainability: A comparative study." Software Maintenance and Reengineering, 2009. CSMR'09. 13th European Conference on. IEEE, 2009.

[16] Kaur, Arvinder, Kamaldeep Kaur, and Ruchika Malhotra. "Soft computing approaches for prediction of software maintenance effort." International Journal of Computer Applications 1.16 (2010).

[17] Ping, Liang. "A Quantitative Approach to Software Maintainability Prediction."Information Technology and Applications (IFITA), 2010 International Forum on. Vol. 1. IEEE, 2010.

[18] Jin, Cong, and Jin-An Liu. "Applications of Support Vector Machine and Unsupervised Learning for Predicting Maintainability Using Object-Oriented Metrics." Multimedia and Information Technology (MMIT), 2010 Second International Conference on. Vol. 1. IEEE, 2010.

[19] Malhotra Ruchika, and Anuradha Chug. "Software maintainability prediction using machine learning algorithms." (2012).