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

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 for calculating 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. Stamolos 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 maintainability. 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 maintainability.

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

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Authors</th>
<th>Method utilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Li, Henry</td>
<td>Regression based</td>
</tr>
<tr>
<td>2</td>
<td>Niessink, Vliet</td>
<td>Regression Models</td>
</tr>
<tr>
<td>3</td>
<td>Fioravanti,</td>
<td>Multiple Linear Regression</td>
</tr>
</tbody>
</table>

Table 1. Maintainability modeling using machine learning approaches

Table 2. Maintainability modelling using traditional approaches

34
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 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