Maintainability Model with Object-Oriented Software Metrics

Mr. N.V. Syma Kumar Dasari and Dr. Satya Prasad Raavi
Research Scholar, Dept. of Computer Science, Krishna University, Machilipatnam, A.P., India.
Syam.researchscholar@gmail.com
Assoc. Professor, Dept. of CSE, Acharya Nagarjuna University, Guntur. A.P., India.
Prof.rsp@gmail.com

Abstract-- Software maintainability is an effective measure for the system quality in the stage of software development. Understandability, modifiability and analyzability are the important factors of the software maintainability that can judge the maintainability of the system. These three factors depend mainly on the 3 size and 8 structural metrics of various UML specifications. We proposed the different models to measure the maintainability and its important three factors in this paper. In the process of developing the models we used exponential non-linear representation. We utilized regression approach along with two-tailed t-test to find out the significance of the variables. The proposed methods are compared to the other non-linear models also made in this research paper.

Index Term-- Analyzability Model, Modifiability Model, Maintainability model, Software Quality, t-test, Understandability Model.

I. INTRODUCTION

The software maintainability plays the significant role in judging the software quality of the system in the development stage of the software life cycle. Software maintainability is the major attribute compared to other factors of the software quality. Maintainability may depend on the several non-functional requirements [1], [10], [26] of the system like as Understandability, Analyzability, Modifiability, Durability, Flexibility…etc. Software maintainability is a challenging task and cost effective because there are no proven measures for this factor [11].

In the object-oriented system design class diagrams plays the major role to estimate the maintainability of the system in the form of various models. Here we focused on the three quality attributes of the software maintainability viz., as Understandability, Modifiability and Analyzability. The well-known standard ISO/IEC-9126-1[25] describes the Understandability Modifiability factors and ISO/IEC-9126-3 [5] reveals the information about the Analyzability phase of the Software Maintainability. In this paper we consider the size and structural metrics for establishing the models for maintainability and its factors.

Regression approach is utilized by us to identify the independent variables for the dependant variables. The t-test is used for the purpose of finding the significant factors which reside in the proposed models and remaining non-significant factors are simply discarded from the final model. Weighted-sum method [4] is also used in the maintainability values prediction with help of its factors. In the non-linear representation of the maintainability and its factors we used the sample data as per the requirements of the exponential notation.

The road map of the paper is as follows. We discussed the Importance of the software maintainability and its factors in Section-I. Section-II shows the literature survey regarding the research topic. Metrics section and sample data for the purpose of estimating models was done in Section-III and Section-IV and. Section-V reveals useful information regarding the research topic and statistical importance of the various estimated models respectively. Comparison of the proposed maintainability model to the other non-linear models is done in the Section -VI. Section-VII gives the conclusion of the paper. The remaining future work was discussed in Section-VIII.

II. LITERATURE SURVEY

Depending upon the various factors like completion time of the variables, Lines of code…..etc. several authors[1],[2],[3],[10],[18],[19],[33],[34],[35] measured the software maintainability in good manner. Even though no one model was not giving the correct methodology toward software maintainability. In the prediction of the software maintainability estimation models the regression approach plays the crucial role. The maintainability and its factors are non-functional requirements for this reason regression approach is the best fitted factor for the purpose of finding the models for maintainability.

With the help of polynomial regression Muthanna et al.[20] developed the software maintainability models. The developed models were not applicable for object oriented designs. Kiewkanya et al. [36] developed the maintainability
models with the help of regression approach. In these models several independent factors were taken into account of the maintainability model. This model took much time to calculate the maintainability model and its factors as Understandability and Modifiability models. With the help of backward step wise regression approach Rizvi et al.[21] developed the linear model for the maintainability and its factors understanability and modifiability with the help of sample data[22],[36] and every model takes the more number of significant attributes for calculating the different models. This model requires more number regression steps until got the zero significant factors.

Satyaprasad et al.[14],[24] proposed the models for Understandability, Modifiability, Analyzability and Maintainability factors with the help of log-normal non-linear standards. In these papers used method was two tailed t-test for finding the significant factors for the dependent variable. Here only one step regression process was applied and two more new levels were introduced (i.e.trivial and consequential) for finding the factors of maintainability in easy manner. Here less number of significant factors were identified in every model. In this paper we are trying to come with a new approach in the non-linear model by using same regression technique. The new approach used here is exponential notation instead of log-normal for getting the new models for maintainability and its factors.

III. METRICS SELECTION

In the OOD with UML specification some of the authors [7],[13],[28] produced the metrics towards software maintainability with the help of various class diagrams. Along with these metrics many more OOD metrics [6], [8], [9], [10], [12], [15], [16], [17], [23], [27], [29], [30], [31], [32], [37] were developed by the different authors to be utilized in various applications for the good system maintenance. In this paper we use total 11 metrics for finding the different models of maintainability factors. These 11 metrics are correlated, empirically validated [40] and measured with the factors of maintainability.

The total 11 metrics are divided as 8 structural metrics and 3 size metrics. These structural and size metrics are developed on the specified class designs with UML specifications. The 8 structural metrics are NAssoc. (Number of Associations), NDep (Number of Dependencies), NGen( Number of Generalizations), NGenH (Number of Generalization Hierarchies), MaxDIT (Maximum DIT), NAgg(Number of Aggregations), NAggH(Number of aggregation Hierarchies), MaxHagg(Maximum Hagg.). Remaining 3 size metrics are NC (Number of Classes), NA( Number of Attributes), NM(Number of Methods).

IV. SAMPLE DATA

Sample data identification plays the major role in the process of selection of the model for the factor. In this paper we proposed the models for maintainability and its factors. The maintainability factors (i.e., understandability, modifiability and analyzability) depend upon the Size and structural complexity metrics of the given system. Data sets utilized in this are paper taken from Satyaprasad et al.[24]. Which contains two datasets,. one is for the purpose of identifying the models of Understandability, Modifiability and Analyzability. Second one utilizes the weighted-sum method for finding the maintainability model with help of Understandability, Modifiability and Analyzability factors. With the help of these two data sets we developed various non-linear models in the exponential notation.

V. PROPOSED METHODOLOGY FOR ESTIMATING MODELS

In this paper we used two tailed t-test methodology to find out the significant factors from variables with 0.05(95%) level of confidence. Here Degrees of Freedom (DF) variable should be taken as the difference between from total number of samples (n) to number of independent variables (t) with adding 1. The formula for the degrees of freedom (df) can be calculated as follows.

\[ DF = n - t - 1 \]

In this paper we applied only single attempt of regression to find the significant factors (important) only .For model estimation the remaining in-significant factors are discarded from the model consideration in single step.

Exponential notation was utilized for the purpose of estimating the non-linear multiplicative models in this paper. The form of multiplicative non-linear exponential representation is as

\[ Y = e^{(1+a + b * X_1 + c * X_2 + d * X_3 + e) \cdots} \]  

On both sides apply the log-normal transformation for the above multiplicative representation

\[ \ln(Y) = \ln(e^{(1+a + b * X_1 + c * X_2 + d * X_3 + e) \cdots}) \]

The above equation is simplified as follows

\[ \ln(Y) = 1+a*X_1+b*X_2+c*X_3+d*X_4, \cdots \]

By utilizing the above equation we proposed the various models for the datasets [20]. The datasets have total 45 samples of data. Among the 45 samples 35 samples are used for the purpose of model estimation. The remaining 10 samples are used to find out the calculated values for the dependent factors.

A. Understandability Model

From the total 11 structural and size metrics were behaved as the independent variables for the dependant factor understandability. The value of the dependant factor here is 23. The two-tail test value for the dependant factor value 23 with 95% of level of confidence is 2.07. Among the 11
metrics only two of the metrics were significant means that they have greater t-value than the 2.07. The resultant significant factors namely NAgg and MaxDIT were displayed in the below table-I shows the understandability model with other useful information.

Table I
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>St.Error</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.614</td>
<td>0.07</td>
<td>8.788</td>
</tr>
<tr>
<td>NAgg</td>
<td>0.045</td>
<td>0.02</td>
<td>2.27</td>
</tr>
<tr>
<td>MaxDIT</td>
<td>0.107</td>
<td>0.032</td>
<td>0.314</td>
</tr>
</tbody>
</table>

The exponential notation for the Understandability model is of the form.

\[ \ln(U) = 0.614 + 0.045 \times NAgg + 0.107 \times \text{MaxDIT} \]  \( \rightarrow (5) \)

Table II
<table>
<thead>
<tr>
<th>R</th>
<th>R²</th>
<th>AdjR²</th>
<th>Std.Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>UND</td>
<td>0.981</td>
<td>0.961</td>
<td>0.943</td>
</tr>
</tbody>
</table>

The value of the Coefficient of Determination (R²) and Adjusted R² show how much effect of total variance on the estimated understandability mode in table-II. The ANOVA model for the Understandability with the statistical significance was shown in Table-III.

Table III
<table>
<thead>
<tr>
<th>Source of validation</th>
<th>Sum Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>f</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>5.745</td>
<td>11</td>
<td>0.522</td>
<td>34.833</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>0.345</td>
<td>23</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6.09</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table-IV performs the comparison made between the calculated and actual values of understandability. Understandability values are calculated with the help of understandability model estimated with non-linear exponential notation. Here UA means Actual Understandability and UC means calculated Understandability values. Table-V shows the Pearson’s correlation between the calculated and actual values of understandability model. The correlation values of the understandability were taken at 0.01(99%) level of significance.

Table V
<table>
<thead>
<tr>
<th>Correlation between UA &amp; UC values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA</td>
</tr>
<tr>
<td>UA</td>
</tr>
<tr>
<td>UC</td>
</tr>
</tbody>
</table>

B. Modifiability Model

We had also taken the 8 structural and 3 size metrics of the different applications in the modifiability model estimation. Here the value of the two-tailed t-test with degrees of freedom of 23 is 2.07. Here three factors namely NA(Number of attributes), NGen(Number of Generalizations) and MaxDIT (Maximum DIT) had the greater t-value rather than the 2.07. The constant value and other information associated with model available in below table-VI.

Table VI
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>St. Error</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.627</td>
<td>0.072</td>
<td>8.733</td>
</tr>
<tr>
<td>NA</td>
<td>0.011</td>
<td>0.005</td>
<td>2.212</td>
</tr>
<tr>
<td>NGen</td>
<td>-0.015</td>
<td>0.006</td>
<td>-2.36</td>
</tr>
<tr>
<td>MaxDIT</td>
<td>0.129</td>
<td>0.033</td>
<td>3.864</td>
</tr>
</tbody>
</table>

Consider the coefficient values of the above table the modifiability model is of the form

\[ \ln(M) = 0.627 + 0.011 \times NA - 0.015 \times NGen + 0.129 \times \text{MaxDIT} \]  \( \rightarrow (6) \)

Table VII
<table>
<thead>
<tr>
<th>R</th>
<th>R²</th>
<th>AdjR²</th>
<th>Std.Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>0.0972</td>
<td>0.944</td>
<td>0.917</td>
</tr>
</tbody>
</table>

Table-VII Model Summary

The R² and Adjusted R² values were displayed in the table-VII shows the considerable effect on the Modifiability model. Table-VIII shows the ANOVA model for the modifiability with 0.001 level of confidence.

Table VIII
<table>
<thead>
<tr>
<th>Source of validation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>f</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>6.127</td>
<td>11</td>
<td>0.557</td>
<td>35.194</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>0.364</td>
<td>23</td>
<td>0.016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6.491</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In table-IX and table-X shows MA means Actual Modifiability values and MC means calculated values with the estimated modifiability model in exponential notation. Table-IX is the comparison between the actual and calculated values of modifiability. The estimated and calculated values are also nearer to the actual values. Table-10 shows the Pearson correlation between the actual and calculated values of modifiability. Correlation values were taken with 0.001(99%) level of confidence.

Table IX
<table>
<thead>
<tr>
<th>Actual and Calculated values for Modifiability Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>MA</td>
</tr>
<tr>
<td>MC</td>
</tr>
</tbody>
</table>
C. Analyzability Model

In the Analyzability model significant factors selection process is done with the help of 3 size metrics and 8 structural metrics. Here the depth of freedom (df) value is 23. Hence the 2-tail T-test value for the df value 23 is again 2.07. in this model 3 factors are become significant and remaining 8 factors are insignificant for the model identification. The significant factors for this model is NA(Number of Attributes), NAgg(Number of Aggregations) and NGen(Number of Generalizations) have greater value than 2.07. The associated value of the model and constant were displayed in the below table-XI.

The analyzability model with the above table data is of the form is

\[ \text{LN}(A) = 0.67 + 0.01 \times \text{NA} + 0.045 \times \text{NAgg} - 0.011 \times \text{NGen} \rightarrow (7) \]

D. Maintainability Model

In the maintainability model estimation process understandability, analyzability and modifiability are taken as independent factors for the maintainability. The second dataset [20] was utilized in the process of estimating the maintainability model. Here also 35 samples of data was taken for the model estimation and remaining 10 samples were used for the purpose of comparison between the actual and calculated values of maintainability. The statistical significance and constant associated with maintainability model was displayed below table-XVI.

The comparison between the actual and calculated values of analyzability is shown in the table-14. The calculated values of analyzability are taken with the model(7).

<table>
<thead>
<tr>
<th>Table X</th>
<th>Correlation between MA &amp; MC values.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MA</td>
<td>1</td>
</tr>
<tr>
<td>MC</td>
<td>0.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table XI</th>
<th>t-values of the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
</tr>
<tr>
<td>constant</td>
<td>0.67</td>
</tr>
<tr>
<td>NA</td>
<td>0.01</td>
</tr>
<tr>
<td>NAgg</td>
<td>0.045</td>
</tr>
<tr>
<td>NGen</td>
<td>-0.011</td>
</tr>
</tbody>
</table>

The maintainability model with the all three factors by utilizing the information of the above table-16 is of the form is as

\[ \text{LN}(M) = 0.489 - 0.078 \times \text{U} + 0.215 \times \text{A} + 0.071 \times \text{M} \rightarrow (8) \]

The \( R^2 \) and Adjusted \( R^2 \) value of the Table-XVII show how much effect by total variance on the maintainability model. The ANOVA model on the maintainability with 99% (0.001) level of confidence was taken in table-XVIII.
In the validation process of the maintainability actual maintainability (MAIN-A) and calculated maintainability (MAIN-C) values were taken. In table-XIX the comparison performed between the actual and calculated values of maintainability. The resultant values are much nearer to the actual values of the maintainability. In the calculation of the maintainability values model (8) was taken into account. Table-XX shows the Pearson correlation with 99% level of significance between the actual and calculated values of maintainability.

The table-XXI shows the information regarding the correlation among the maintainability and its factors. The below table shows the understandability, modifiability and analyzability factors were strongly correlated with the dependent factor maintainability. The values of the correlation analysis were taken at the 0.01(99%) level of confidence.

VI. COMPARISON WITH OTHER NON-LINEAR MODELS

In this paper we made the comparison between this exponential non-linear model with other models [14] [24] which were based on the log-normal non-linear models. Here our first observation is the significant variables are different from one non-linear model to another model. In the log-normal two types of models(1-7 level)(1-9 level) understandability and modifiability depend upon the same significant factors. Coming to the exponential notation the model was changed and also the significant factors are also changed for understandability, modifiability and analyzability.

The second observation is the validation process also the log-normal non-linear model gives best results rather than exponential notation. The compared values so much nearer to the actual values in the log-normal model than exponential model. In the correlation (Pearson-correlation) process also log-normal shows effective results rather than exponential model in all the cases (understandability, modifiability, analyzability and maintainability).

The third observation regarding the statistical importance of the both models. Here log-normal models are simpler than the exponential models. The Coefficient of Determination (R²) and Adjusted R² are higher values possible in the log-normal non-linear model than exponential notation. We observed that the ANOVA models are also better in log-normal models.

The fourth observation is the maintainability model of the log-normal model shows more and effective results displayed whether in the case of not only in statistical significance but also in the validation of the maintainability model. Hence we recommend the log-normal non-linear model for estimating the maintainability and its factors (i.e., understandability, modifiability and analyzability).

VII. CONCLUSION & FUTURE SCOPE

Software maintainability is a typical task because no proper metric was existed for measuring this property. So more number of research persons depend on various model representations of maintainability and its factors. Here in this paper we represented one non-linear model representation named as exponential notation for the maintainability and its factors (i.e. Understandability, modifiability and analyzability). In each and every model representation we took the significant factors for model preparation and insignificant factors are discarded from model with the help of single step regression process. The statistical importance and validation of all the models was done in this paper. The comparison made between this exponential model and other log-normal non-linear model representation here. In the process of comparison we gained better results in log-normal non-linear model than exponential non-linear model.

In future we want to focus on other non-linear models like polynomial-notation, trigonometric notation,…etc, for getting different models for maintainability and its factors. The maintainability has the various attributes like portability, usability, testability,…etc., rather than understandability, modifiability and analyzability. So we want to focus on those factors and find the better system maintainability forever.
REFERENCES


Author’s Profile

**Mr. D.N.V. Syma Kumar** received B.Tech Degree from JNTU, Hyderabad and M.Tech from BHARAT University, Chennai. He is currently pursuing his Ph.D. from Department of Computer Science, Krishna University, Andhra Pradesh. Presently he is working as an Associate Professor in the Department of Computer Science & Engineering, S.S.N. Engg. College, Ongole, Andhra Pradesh.

**R. Satya Prasad** received Ph.D. Degree in computer science in the faculty of Engineering in 2007 from Acharya Nagarjuna University, Andhra Pradesh. He received gold medal from Acharya Nagarjuna University for his outstanding performance in master’s degree. He is currently working as Associate Professor in the department of Computer Science & Engineering, Acharya Nagarjuna University. His current research is focused on Software engineering. He has published several papers in National & International Journals.