Abstract

In this paper, we provide a detailed comparison between various models that have been provided in literature for predicting faults in the software testing process. They are commonly known as software reliability models.

2. Software reliability model selections

Selection of a particular model is a challenging problem for software reliability prediction. There are two reasons for that. They are the selection of the release time and the value of resource allocation decision. These factors can affect the accuracy of the prediction. In the past, several solutions have been proposed to address the solution for the above-described problems. They are [6]:

- Use several software reliability models and select the one that gives the highest confidence.
- Use the recalibration method to compensate for the bias of a model.
- Use an adaptive model as an alternate approach

3. Evaluation Criteria

The model performance was measured in terms of the NRMSE.

\[
NRMSE = \frac{1}{n-1} \sqrt{\frac{\sum_{k=1}^{n} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{n} (y(k))^2}}
\]

Where \(y(k)\) is the actual accumulated faults and \(\hat{y}(k)\) is the estimated accumulated fault using various software reliability models.
4. Comparison among the models

In this section we cover the analysis part of the experiments. We have used three different data sets as a Benchmark, collected from different projects [1]. The data type is stochastic. Data sets range from military application projects, real time and control applications and finally, operating systems applications. This data are intensively used in literature to evaluate software reliability model [5].

4.1. Power model

The Power model seems to be a poor estimator in many cases. Its NRMSE values are: 11.5220, 6.5423 and 4.4730, for the following projects: Military application # 40, Real time control #1, and Operating system # ss1c. The Power model consistently performs poorly in most cases. The observations show that the Power model is the worst predictor and the highest value of NRMSE among the parametric models in most cases.

4.2. Exponential model

The Exponential model's behavior is somewhat similar to that of S-shaped model, except for project operating system # ss1c, where it has the lowest value of 2.3925. Its NRMSE values are 9.0910 for military applications # 40, 2.8991 for the real time control #1, and 2.3925 for operating system # ss1c, which is the lowest predictor when it is compared to other predictions by other parametric models. This finding suggests that the Exponential model has the best predictability compared to the Power model and S-shaped model.

4.3. S-shaped model

The S-shaped model seems to perform relatively well. Though it is a best predictor in most cases, it has projected the remaining faults most accurately in two out of three projects. The NRMSE of the S-shaped model measures of 8.0388, 5.4161, and 2.4177 are the lowest for the project military application # 40, operating system # ss1c, and real time control #1. This finding suggests that the S-shaped model has good predictability. The results show that the Exponential model is superior to the other parametric models, and the S-shaped is close to the Exponential models.

4.4. A Parsimonious auto-regression model

The Auto-regression models magnitude of order 4 seems to perform poorly among the developed models. For example, the NRMSE in the military application project is 3.1413, real time control project # 1 is 1.7086, and the operating system # ss1c is 1.0659, which has the highest NRMSE value among non-parametric models compared to the fuzzy logic and neural network.

4.5. Neural Networks model

From the observation in Table 1, The Neural Networks (NN) seems to perform relatively well. Although the neural networks are the best predictors in most cases, they have projected the lowest NRMSE in two of three projects. For example, in the military application projects they are 0.5644, which is the lowest NRMSE among the entire project under study. In real time and control project # 1, the NRMSE is 1.0755, and for the operating system # ss1c, the NRMSE is 0.7714, mean while neural networks perform 66% better than fuzzy logic model.

4.6. Fuzzy logic model

The Fuzzy logic model performs relatively well throughout all the experiments. Also, fuzzy logic has the best predictive capability in all the projects. For example, in the following, project military application # 40, and operating system # ss1c, the NRMSE are 1.1081, 1.2901, but in the real time control project # 1, the NRMSE is 0.9358, which is the lowest when compared between the two projects. As we have seen from this observation of the results, neural networks model gives the best or close to the best predictions in most of the cases. Also, the neural networks model is superior to the parametric models, regression model, and fuzzy logic.
From the observation in Table 1, we found out that the real time and control project NRMSE is the lowest compared to other models. This is why we decided to investigate this model by implementing an ANOVA test with a significance level, $\alpha = 0.1$.

In Table 2, we show the result of one-way ANOVA test, the $F$ value is $F_{(2,405)} = 2.83$ where 0.1 is the significant level, and 2 and 405 represent the degrees of freedom. The observed significance level is $p$-value is equal to 0.06. We can assert that there is a significant difference among groups since $p < 0.1$.

### Table 1: The performance of these models in term of normalized root mean square error \textit{NRMSE}

<table>
<thead>
<tr>
<th>Model</th>
<th>Military # 40</th>
<th>Real-time Control # 1</th>
<th>Operating Sys SSIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>11.5220</td>
<td>6.5423</td>
<td>4.4730</td>
</tr>
<tr>
<td>Exponential</td>
<td>9.0910</td>
<td>2.9991</td>
<td>2.3925</td>
</tr>
<tr>
<td>S-Shaped</td>
<td>8.0388</td>
<td>2.4177</td>
<td>5.4161</td>
</tr>
<tr>
<td>Auto Regression (4)</td>
<td>3.1434</td>
<td>1.7086</td>
<td>1.659</td>
</tr>
<tr>
<td>Fuzzy Logic</td>
<td>1.1081</td>
<td>0.9358</td>
<td>1.2901</td>
</tr>
<tr>
<td>Neural Network</td>
<td>0.5644</td>
<td>1.0755</td>
<td>0.7714</td>
</tr>
</tbody>
</table>

5. One-Way Analysis of Variance (ANOVA)

The purpose of one-way ANOVA is to find out whether data from several groups have a Common mean. That is to determine whether the groups are actually different in the Measured Characteristic.

![Comparison among models](image)

Figure 1: A comparison between various software reliability models with respect to NRMSE
Data sets for both large and small projects from diverse sources have been analyzed. Results presented here indicate that some perform better than others in most cases. This research shows evidence that a non-parametric approach provides the lowest normalized root mean square error and accurate result for the range of values in the experiments in most cases. The presented results show also that a fuzzy model and neural networks can be used to build prediction models for software reliability.

The results of the fuzzy logic and neural networks models were very promising. The error difference between the actual and estimated response was small. This finding gives a good indication of prediction capabilities of the developed fuzzy model and neural networks.

6. Conclusions and Future work

In this paper, we presented a comparison between various software reliability models. They include the power, exponential, S-Shape parametric, regression, neural network and fuzzy logic models. The experiments show that the non-parametric models are superior to the parametric models in their ability to provide an accurate estimate of reliability when historical data is missing. Currently we are investigating the use of evolutionary computations to solve the software reliability-modeling problem.

Bibliography


