A Brief Review of Software Reliability Prediction Models

Boby John¹, Rajeshwar S Kadadevaramath², Immanuel A Edinbarough³

¹SQC & OR Unit, Indian Statistical Institute, Bangalore
²Department of Industrial Engineering & Management, Siddaganga Institute of Technology, Tumkur
³Department of Manufacturing and Industrial Engineering, University of Texas, USA

Abstract: Software plays an important role in every field of human activity today varying from medical diagnosis to remote controlling spacecraft. Hence it is important for the software to provide failure-free performance whenever needed. The Information technology industry has witnessed rapid growth in the recent past. The competition among the firms also increased. The software organization in the developing countries like India can no longer survive on cost advantage alone. The software companies need to deliver reliable and quality software on time. A lot of research has been carried out on software quality management and reliability estimation. The objective of this paper is to provide a brief review of the major research contribution in the field of software reliability and identify the future research areas in software reliability estimation and prediction

Keywords: software reliability growth models, nonhomogeneous Poisson process models, s-shaped models, imperfect debugging

I. INTRODUCTION

Many organizations utilize information technology (IT) to improve productivity, enhance operational efficiency, responsiveness, etc [1] As a result, the IT industry has witnessed tremendous growth in the past few decades. As the number of information technology companies increased, the competition among them also increased. The software organization in the developing countries like India can no longer survive or grow based on cost advantage alone. But delivering reliable and quality software on time within budgeted cost is a challenge for many organizations [2], [3]. Many times the companies would compromise on software testing and release the software with residual defects. This would make the software unreliable.

The software reliability is defined as the probability of failure-free operation of a software system for a specified time in a specified environment [4]. The failure of the software during operations can lead to customer dissatisfaction, loss of market share, etc. The failure of a software used in the medical device or that used in air traffic control system can have a disastrous effect on the individual as well as society. Hence it is imperative for the software firms to ensure their product is sufficiently reliable before releasing the software for usage. This paper is a brief review of the important developments happened in the field of software reliability and identifies the future research areas.

The remaining part of this article is arranged as follows: the session II describes the literature review methodology, the literature review analysis is given in session III and the conclusion are discussed in session IV.

II. LITERATURE REVIEW METHODOLOGY

A lot of articles have been presented at conferences, published in journals and books have been written in the last few decades on software reliability estimation and prediction. The aim of this paper is to provide a brief review of the important researches carried on developing software reliability models. The process started with searching for relevant published articles. The scope of the review is limited to the published books and papers published in journals and important conference proceedings. The databases searched are IEEE explore, Science direct, Google scholar and research gate. Two hundred and nine papers are identified for review. After reading the abstract, ninety-seven papers are shortlisted for review. Another twenty-nine papers are later dropped as the content is not directly related to the focus area of the review. Finally, sixty-eight papers are included in the review. The details are given in fig 1.
III. LITERATURE REVIEW ANALYSIS

Most of these research resulted in developing software reliability prediction models. These models are based on the data from the software testing phase. Most of the aforementioned models are mathematical functions expressing the relationship between the number of faults detected and the effort spent on testing. The testing effort can be measured using calendar time, execution time or the number of test cases executed. The models can be used for predicting the software reliability as well as the number of faults remaining undetected (residual defects) in the software. The models can thus help on deciding when to stop testing and release the software. The basic assumption underlying software reliability estimation is that as testing progresses, more and more faults get detected. As a result, the number of residual defects decreases and the reliability increases. Hence these models are called software reliability growth models (SRGM).

The widely popular class of software reliability growth models are the nonhomogeneous Poisson process (NHPP) models. The NHPP models try to identify a nonhomogeneous Poisson process model which best fits the fault detection pattern as software testing progresses and use the best fit model to estimate the software reliability or residual defects. Most of these models have a parameter indicating the expected total number of faults in the software [5]. The reliability growth models provides a mathematical equation to estimate the number of defects detected $m(t)$ at time $t$

$$m(t) = aF(t)$$  \hspace{1cm} (1)

where time $t$ is often measured in terms of testing effort and $a$ is the parameter representing the expected total number of faults. Whenever the cumulative number of faults detected during testing is equal or close to the expected total number of faults, it is an indication that more or less all the faults injected to the software are detected and the software can be released. The difference between the number of faults detected and the number of faults predicted by the model at the time of software release gives the number of residual faults still present in the software. Hence the software reliability models can also be used for determining when to stop testing and release the software.

The software reliability growth models are broadly classified into two groups namely nonhomogeneous Poisson process models and stochastic process models. Out of the two classes of models, the NHPP models are more popular and widely used. The NHPP models are further classified based on a variety of criteria. The important among them are

A. The classification based on test effort measurement as testing time based and the number of test cases based models.

B. Models based on the shape of the reliability or mean value function $m(t)$ as exponential and $s$-shaped models. The $s$-shaped models can be further divided into delayed $s$-shaped and inflation $s$-shaped models

C. Models based on the type of assumption on debugging as models incorporating perfect debugging and those with imperfect debugging.
The classification of software reliability growth models is given in Fig. 2.

**Fig 2: Classification of Software Reliability Growth Models (SRGMs)**

- **Software Reliability Growth Models**
  - Stochastic process models
  - NHPP models
  - Others

- **Shape-based models**
  - Exponential models
  - S-shaped models

- **Debugging based models**
  - Perfect debugging
  - Imperfect debugging

The NHPP model assumes that the cumulative number of failures detected till time \( t \), \( [N(t), \ t \geq 0] \) as a nonhomogeneous Poisson process with failure rate or intensity function \( \lambda(t) \). Then the probability or chance that \( N(t) = k \), where \( k \) is an integer value is [6], [7] given by

\[
P[N(t) = k] = \frac{[m(t)]^k}{k!} e^{-m(t)}, \ k = 0, 1, 2, \ldots
\]  

(2)

where \( m(t) \) is the mean value function representing the expected number of failures till time \( t \) and is given by

\[
m(t) = \int_0^t \lambda(s) ds
\]  

(3)

Similarly, the intensity function \( \lambda(t) \) can be computed from \( m(t) \) as

\[
\lambda(t) = \frac{dm(t)}{dt}
\]  

(4)

In the case of stochastic process models, the assumption is that a fixed unknown number \( n_0 \) faults are injected to the software during development, which need to be detected and fixed to make the software perfectly reliable. As testing progresses, these faults will get detected and fixed. According to Jelinski and Moranda [8], the expected number of faults at time \( t \) or mean value function \( m(t) \) is given by

\[
m(t) = n_0[1 - e^{-\phi t}]
\]  

(5)

where \( \phi \) is the constant hazard rate. The corresponding failure intensity function is given by
The important papers discussing the stochastic process models are given in table 1.

<table>
<thead>
<tr>
<th>SL No.</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jelinski and Moranda [8]</td>
</tr>
<tr>
<td>2</td>
<td>Schick and Wolverton [9]</td>
</tr>
<tr>
<td>3</td>
<td>Shanthi Kumar [10]</td>
</tr>
<tr>
<td>5</td>
<td>Littlewood [12]</td>
</tr>
</tbody>
</table>

The NHPP models develop functions to estimate the expected number of faults at time $t$ or mean value function $m(t)$ in terms of testing effort. The test effort can be measured using calendar time, execution time or the number of test cases executed. The models based on the number of test cases executed are called discrete time models. The important articles discussing the discrete time models are given in table 2.

<table>
<thead>
<tr>
<th>SL No.</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brooks and Motle [13]</td>
</tr>
<tr>
<td>2</td>
<td>Yamada and Osaki [14]</td>
</tr>
<tr>
<td>3</td>
<td>Kapur et al. [15] - [17]</td>
</tr>
</tbody>
</table>

One of the ways used to group the NHPP model is based on the shape of the mean value function or fault detection curve. Many researchers assumed the fault detection curve will be an exponential curve and developed exponential NHPP models to predict software reliability. Later on, researchers found that during the bug fixing phase of software development life cycle, the programer's knowledge of the software product improves and as a result the fault detection curve will be most likely an s-shaped rather than exponential. Those researchers developed a group of NHPP models called s-shaped models. The important research work on exponential and s-shaped NHPP models are given in table 3.

<table>
<thead>
<tr>
<th>SL No</th>
<th>Shape</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exponential</td>
<td>Musa [18], Schneidewind [19], Goel and Okumoto [20], Yamada et al. [21], Hossain [22], Xie and Zhao [23]</td>
</tr>
<tr>
<td>2</td>
<td>s-shaped</td>
<td>Yamada et al. [24], Ohba [25], Ohba and Yamada [26], Ohba [27], Bittanti et al. [28], Kareer et al. [29], Kapur and Garg [30], Yamada et al. [31], Kimura et al. [32], Pham [33], Lee et al [34]</td>
</tr>
</tbody>
</table>

Another popular way of classifying the NHPP models is based on the assumption on debugging. Many researchers assume that as soon as a fault is detected, the work on fixing it will commence. The fix will be perfect and no new fault will be introduced while fixing the bugs. The models with such assumption on perfect debugging are classified as perfect debugging models. But many situations the debugging will not be perfect. There is the possibility of introducing new faults while fixing the bugs. Many models incorporate this scenario also. Such models are classified as imperfect debugging models. The important research papers published on perfect and imperfect debugging assumption are given in table 4.
TABLE 4: Models classified on debugging assumption

<table>
<thead>
<tr>
<th>SL No</th>
<th>Debugging Assumption</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Perfect Debugging</td>
<td>Dalal and Mallows [35], Schneidewind [36] - [39]</td>
</tr>
<tr>
<td>2</td>
<td>Imperfect Debugging</td>
<td>Ohba and Chou [40], Kapur and Garg [41], Pham [42], Zeephongseku [43], Gokhale et al. [44], Pham et al. [45], Yamada and Sera [46], Gokhale et al. [47], [48]</td>
</tr>
</tbody>
</table>

A lot of research work on NHPP models which will not fall strictly under the aforementioned classifications. The important among them are given in table 5.

TABLE 5: Other Important articles on NHPP models

<table>
<thead>
<tr>
<th>SL No.</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Musa and Okumoto [49]</td>
</tr>
<tr>
<td>2</td>
<td>Xie [50]</td>
</tr>
<tr>
<td>3</td>
<td>Kenney [51]</td>
</tr>
<tr>
<td>4</td>
<td>Leemis [52]</td>
</tr>
<tr>
<td>5</td>
<td>Farr [53]</td>
</tr>
<tr>
<td>6</td>
<td>Xie and Wohlin [54]</td>
</tr>
<tr>
<td>7</td>
<td>Mullen [55]</td>
</tr>
<tr>
<td>8</td>
<td>Gokhale and Trivedi [56]</td>
</tr>
<tr>
<td>9</td>
<td>Bishop and Bloomfield [57]</td>
</tr>
<tr>
<td>10</td>
<td>Kapur et al [58]</td>
</tr>
</tbody>
</table>

Recently many researchers have attempted to develop software reliability prediction models using machine learning techniques also. The important among them are given in table 6.

TABLE 6: Software reliability models using machine learning techniques

<table>
<thead>
<tr>
<th>SL No.</th>
<th>Technique Used</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fuzzy logic</td>
<td>Utkin et al. [59]</td>
</tr>
<tr>
<td>2</td>
<td>Artificial Neural Networks</td>
<td>Cai et al. [60], Tian and Noore [61], Kapur et al [62], Zheng [63].</td>
</tr>
<tr>
<td>3</td>
<td>Genetic Programming</td>
<td>Afzal and Torkar [64]</td>
</tr>
</tbody>
</table>

Since a large number of software reliability models have been proposed, the research on comparison and unification of the models started from 1980’s itself. The major attempts on comparing the different software models are given in table 7.

TABLE 7: Articles on comparison and unification of the software reliability models

<table>
<thead>
<tr>
<th>SL No.</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yamada and Osaki [65]</td>
</tr>
<tr>
<td>2</td>
<td>Langberg and Singapurwatta [66]</td>
</tr>
<tr>
<td>3</td>
<td>Abdel_Ghaly et al. [67]</td>
</tr>
<tr>
<td>4</td>
<td>Miller [68]</td>
</tr>
<tr>
<td>5</td>
<td>Huang et al [69]</td>
</tr>
<tr>
<td>6</td>
<td>Gokhale [70]</td>
</tr>
<tr>
<td>7</td>
<td>Kapur et al [71]</td>
</tr>
</tbody>
</table>

Even though a lot of models are proposed, no single model is suitable for predicting the reliability for all types of software cases. Moreover, identification of the best-suited model from the available models is also a challenge. Even different models may come up
as the best model based on the criteria used to select the best model.

IV. CONCLUSION

This paper provided a brief review of the important research happened in the field of software reliability estimation and prediction. Since software development is a human activity, it is almost impossible to eliminate the injection of faults during software development. Hence it is required to detect and fix as many faults as possible before releasing the software for use. The purpose of software testing is to detect as many faults as possible. As more and more faults are getting detected and fixed through testing, the number of faults remaining in the software decreases and software reliability increases. A lot of software reliability growth models have been developed in the past to estimate the software reliability. These models can also be used for deciding when to stop testing and release the software.

The development of a large number of software reliability models itself is an indication that there is no single model suitable for estimating the reliability of all types software. So one of the future areas of research can be developing a single model suitable for all kinds of scenarios or development of a unified or common model which can take different forms suitable for different types of the software.

Another important research area can be identifying the best-suited model for a given situation. Multiple evaluation criteria have been suggested for identifying the best fit model. Many times, different criteria may suggest different models as the best-fit model for a given situation. Hence the development of a methodology to identify the best model suited for a given case based on multiple criteria can be another area of future research.

REFERENCES

