SRGM with Imperfect Debugging using Capability Analysis of Log-Logistic Model

Dr.O.Naga Raju
Asst.professor
Dept. of C.S.E,
Acharya Nagarjuna
University

D.Rakesh Nag
M.Tech, Research Scholar
Dept. of C.S.E,
Acharya Nagarjuna
University

K.V.SubbaReddy
Research Scholar
Dept. of C.S.E,
Acharya Nagarjuna
University

ABSTRACT

Analyze the predict capability of some SRGMs to understand the different parameters to facilitate the estimate process. The predict validity analysis will be on two key factors, one pertaining to the degree of fitment on available failure data and the other for its prediction capability. The validity analysis will be to arrive at trade off in choosing a simple model as compared to complex model by determining their performances across multiple data sets. Data for the predict validity analysis has been taken from different time periods to understand the impact of these models across various technologies and process during the time frame.

Keywords

Software reliability, Imperfect-Debugging, time complexity, Mean Value Function

1 INTRODUCTION

There are many Software Reliability Models available. Each model use two are three parameters to get the reliability estimation from the actual failure data. These models are designed based on the expectation of the trend in the failure data .To identify the trends some assumptions are assumed and they are:

- Fault identification is rapid in initial stage and reaches a steady state.
- Fault identification is slow initially and rapidly increases, finally reaches to a steady state.
- Fault identification at the initial stage follows steady growth and reaches to steady state.

Each trend depends on the following issues completely or partially

- Performance capability of testing team
- Complexity of application domain
- Kind of technology used
- Size of the application
- The kind of software development process

The number of faults identified by an experienced tester is more than the less experience tester. The complexity of the application and the time constraint to identify the faults may create new faults or ignore existing faults.

New technologies introduced changes in the software development process SSAD (Structured System analysis and Design) technique was used in early times .With the advent of distributed objects OOAD (Object Oriented Analysis and Design) technique used. With new changes, there is change in testing and debugging process as compared to SSAD development process. With internet technology new systems and legacy systems can be integrated, studying such systems where both old and new system coexists is interesting.

Different software reliability growth models have been developed across various time periods. The complexity of models with two or more parameters increased to cater to change in need with change in time. In this paper we focus on validating predictions of two or more parameters software reliability growth models. Along with predictions done by Prince[1](2006), we present the Log -Logistic model for the both ungrouped and grouped data inputs taken from the early 80's and 90's.The degree of the fitment is studied for the log-logistic data using a specific percentage available failure data and validate the predictive performance of the model along with the models validated by Prince[1](2006).

2 Notation

a : Expected number of faults in the software when the testing begins

b : Fault detection rate per remaining fault in the software

m(t) : Expected number of observed failures during the time interval(0,t)

R(x/s) : Software reliability

c : Inflection parameters

$\lambda(t)$: Failure intensity function

z_i : Cumulative number of failures up to time t_i

3 Software Reliability Growth Models (SRGMS)

Along with six different software reliability growth models Prince[1](2006). , log-logistic model, a total of seven was selected for performing the predict capability analysis .The common assumptions made in these models are:

- The software system is subject to failure are random time caused by software faults.
- No failures at time t = 0
- Failure intensity is proportional to the residual fault content
- Faults identified are immediately corrected and will not appear again

3.1 Two parameter models: The following three SRGM's each having a specific characteristic and designed using two parameters were selected for the prediction capability analysis:

- Delayed S-shaped Growth Model

- Exponential Model
- Logarithmic Poisson Model

3.1.1 Delayed s-shaped growth model:

The mean value function for this model (Yamada et al., 1983) [2] is given below:

$$m(t) = a * (1 - (1 + bt)e^{-bt})$$

The advantage of this model is that it is designed for fault isolation data analysis. Fault isolation means some of the failures can be intentionally reproduced, so the fault identification and its removal can be achieved.

3.1.2 Exponential growth model:

The model proposed by Goel and Okumoto (1979) [3] has been taken for the study. The failure data is assumed to take an exponential curve. The mean value function for this model is:

$$m(t) = a * (1 - e^{-bt})$$

3.1.3 Logarithmic Poisson model:

This model proposed by Musa and Okumoto (1984) [4], has its mean value function as:

$$m(t) = a * \ln(1 + bt)$$

This model is classified as infinite failure model as it is assumed that there is no upper bound to the number of failures.

3.2 Three parameter models:

The following SRGM's designed using three parameters were selected for the performance analysis:

- Imperfect Debugging Model
- Inflection S-shaped Growth Model
- Logistic Model
- Log-Logistic Model

3.2.1 Imperfect debugging model:

Imperfect debugging occurs when the error debugging process does not lead to the removal of the error. The model proposed by Kapur et al., (1980) [5] has been taken.

The mean value function for this model is:

$$m(t) = \frac{a}{p} * (1 - e^{-pbt})$$

Where p is the probability of perfect debugging.

3.2.2 Inflection s-shaped growth model:

According to this model the observed software reliability growth becomes S-shaped if faults in a program are mutually dependent .The mean value function for this model (Obha 1984a) is:

$$m(t) = a * \frac{1 - e^{-bt}}{1 + ce^{-bt}}$$

where c is the inflection parameter and is given by $\frac{1-r}{r}$, (r>0)

where r is inflection rate.

3.2.3 Logistic growth model:

In this popular model (Yamada et al., 1983) the software failures are assumed to follow a logistic curve. The mean value function for this model is[6] :

$$m(t) = \frac{a}{1 + ce^{-bt}}$$

3.2.4 Log-logistic model:

In this model the software failures are assumed to follow a Log-logistic curve .The mean value function of this model is given [7] by:

$$m(t) = \frac{a(bt)^k}{1 + (bt)^k}$$

Where k is the inflection parameter.

4 Prediction Capability Approach

The prediction capability of each model is analyzed using two data sets. Each data is of different size and from different time periods.

The predict validity process consists of the following steps (Prince 2006):

- Estimation of the parameters of each model using 80% of the failure data
- Model analysis:
  1. Goodness of Fit for all models using the first 80% of the failure data.
  2. Model prediction capability is compared by validating against the last 20% of the available failure data.

4.1 Parameter Estimation

The model parameters for each of the selected model were estimated by maximizing the log likelihood function (Obha 1984a) and it is given by:

$$\ln L = \sum_{i=1}^{n} [(z_i - z_{i-1}) \ln[m(t_i) - m(t_{i-1})]] - m(t_n)$$

where z_i is the cumulated number of faults before time t_i, n is the interval domain size.

The respective mean value function of each model is substituted in the above equations. Then the result is differentiated with respect to the number of parameters to obtain the specific parameter based equations. These equations are set to zero and solved to arrive at the respective parameter values [8].

The details of the parameters estimated for each Model using published failure data sets are given in the tables 1 and 2

Data set 1: Kapur et al (1983):
Number of Failures: 19
Number of Failures identified for estimated =0.08 * 19 = 15

Data set 2: Obha (1984b):
Number of Failures: 109
Number of Failures identified for estimated =0.08 * 109 = 87

Table 1 Parameters estimation for data set1

Model	a	b	3rd Parameter
Delayed S-shape	105.461	0.270	
Exponential	138.09	0.077	
Logarithmic passion	79.066	0.153	
Imperfect debugging	103.567	0.103	p = 0.75
Inflection S-shape	137.203	0.079	r = 0.979
Logistic	101.741	0.348	p = 9.985
Log-logistic	99.125	0.276	k = 11.246

Table 2 Parameters estimation for data set2

Model	a	b	3rd Parameter
Delayed S-shape	561.864	0.05	
Exponential	735.825	0.014	
Logarithmic passion	543.818	0.019	
Imperfect debugging	548.869	0.019	p = 0.75
Inflection S-shape	545.205	0.058	r = 0.164
Logistic	531.819	0.08	m = 15.966
Log-logistic	528.567	0.064	k = 17.876

4.2 Model analysis

The performance of the each model was analyzed by Goodness Fit and Predictive capability of each model. For deriving the degree of fitness first 80% of the failure data was used .The remaining 20% of failure data was then predicted using the estimated parameters. The validation fitness and prediction capability of each model was measured by their SSE (Sum of Square Errors) calculations.

Table 3 SSE of the models using data set 1

SSE	Del's'pe	Exo'ial	Log'P'on	Impl'D'ug	In'S'pe	Log'ic	Lo'log'
Fit	551	147	165	148	146	175	180
Predict	0.70	17.9	29.3	19.18	17.94	0.31	0.28

Table 4 SSE of the models using data set 2

SSE	Del's'pe	Exo'ial	Log'P'on	Impl'D'ug	In'S'pe	Log'ic	Lo'log'
Fit	23189	84630	81837	76690	10732	15941	17543
Predict	1513	10705	50677	14449	435	81	76

4.2.1 Goodness Fit

- Three Parameters Models
  - The inflection s-shape model has the best fit amongst all the models analyzed.
  - Log-logistic model fitment improves as the failure data volume increases.
  - Logistic model fitment also improves as the failure data volume increases.
  - Implicit debug model fitment decreases as the volume increases
- Two Parameters Models
  - The delayed s-shape model has the best fit amongst all the two parameter models analyzed.
  - Exponential model fitment decreases as the failure data volume increases
  - Logarithmic Poisson model fitment decreases as the data volume increases

4.2.2 Prediction Capabilities

- Three Parameters Models
  - The log-logistic model shows the best overall prediction capability across the models
  - Next to the log-logistic is the logistic model.
  - Inflection s-shaped model prediction trends to improve as the failure data volume increases.
  - Imperfect debug model prediction capability is about average

- Two Parameters Models
  - The delayed s-shape model has the best prediction capability amongst all the two parameters models
  - Exponential models prediction capability has the tendency improve as the failure data volume increases.
  - Logarithmic Poisson model prediction capability decreases as the data volume increases

5 Conclusions

In this paper, we have presented different models, based on the two or three parameters for expectation of the trend in the failure data. Three parameter inflection s-shape model has the best fit amongst all the models analyzed and Log-logistic model shows the best overall failure data prediction capability performance. Among two parameters delayed s-shaped can be considered as best.

6 References

[1]D.R. Prince Williams (2006), "Prediction Capability Analysis of Two and Three Parameters Software Reliability Growth Models", ITJ 5(6): 1048-1052

[2]Kapur , P.K *et al.,* (1990), " Optimal software release policies for software reliability growth modles under imperfect debugging. RAIRO:Operations Res..24 295-305

[3]Obha, M., (1984a), "Inflection S-Shaped Software Reliability Growth Model Stochastic Models in Reliability Theory", Springer-Verlag, Berlin, pp 144-162

[4]Obha, M., (1984b), "Software Reliability Analysis models." IBM. J. Res. Deve., 28: 428-443

[5]Yamada, S., *et al.,* (1983) "S-shaped reliability growth modeling for software error detection" IEEE Trans .Reliability, 12:475-478

[6]Goel, A.L and K.Okumoto,(1970). "Time dependent error detection rate model for software reliability and other performance measure". IEEE Trans. Reliability 28:206-211

[7]Swapna S.Gokhale and Kishore S.Trivedi (2008). " Log-Logistic Software Reliabilty Growth Model" *IEEE Trans. On Software Engineering*, SE-11(12):1411–1423, Dec.1985.

[8]Musa, J.D and K.Okumoto(1984). "A logarithmic Poisson execution time model for software reliability measurement." Proceedings of 7th International Conf. on SE, pp:230-238

Authors:

Dr. O.NagaRaju received the Doctorate in Computer Science & Engineering from Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. He have a satisfactory consistent academic track of record and received gold medal from Acharya Nagarjuna University for his out standing performance in a first rank in Masters Degree. His research interests include Software Engineering, Network Computing, and Image Processing.

Mr. D.Rakesh Nag received Masters Degree in computer Science & Engineering from Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. His research areas include Software Engineering, Image Processing, Dataminig.