Journal of Engineering and Applied Sciences 13 (12): 4376-4380, 2018
ISSN: 1816-949X

The Analysis and Predict of Software Failure Time Based on Nonlinear Regression

Tae-Jin Yang
Department of Electronic Engineering, Namseoul University, Seoul, Republic of Korea

Abstract: Software reliability is an important issue in the software development process. The software development process with considerations of cost and failure time are essential. Software failure time have been proposed in the literature exhibit either constant, monotonic increasing or monotonic decreasing. For data analysis of software reliability model, trend analysis already been developed. The methods of trend analysis are arithmetic mean test and Laplace trend test. Trend analysis only offers information of outline content. In this study, we were discussed censoring failure time and predicted failure time using nonlinear regression models that is growth, quadratic and S-curve type which error terms, each other are different model. Model selection using the coefficient of determination and the mean square error were presented for effective comparison. In result of analysis, relatively, growth regression model than any models in terms of goodness of fit is effective model.

Key words: Software reliability, time censoring, nonlinear regression, model selection, determination, growth regression model

INTRODUCTION

Computer system failure due to software failures can cause tremendous loss to our society and related industries. Therefore, software reliability is an important issue in the software development process. This problem should satisfy the user's requirements and testing costs. Efficiently in order to reduce the costs of variability and reliability of the software must know in advance the cost of testing in software testing (debugging). Therefore, the reliability, the software development process with considerations of cost and failure time is essential. There is a need to develop a model so as to estimate the defect content of software products to improve the quality. Until now, many software reliability models have been proposed.

Some of these models for describing the software failure phenomenon are based on the Non-Homogeneous Poisson Process (NHPP). In fact, these models are fairly effective in describing the error-detection process with a time-dependent error-detection rate. However, the assumption is that each time an error occurs the fault that caused it can be immediately removed, leading to no new problems which is usually called perfect debugging (Gokhale and Trivedi, 1999). More recently, Huang (2005) incorporated both a generalized logistic testing-effort function and the change-point parameter into software reliability modeling. In attempting to predict software reliability, computer algorithm has been an effective technique in assisting the prediction.

Chiu *et al.* (2008) can explain the learning process that software managers to become familiar with the software and test tools for S-type model. In addition, Kim (2013a, b) studied to the comparative study of NHPP delayed S-shaped and extreme value distribution software reliability model using the perspective of learning effects and Shin and Kim (2013) was studied about the comparative study of software optimal release time based on NHPP software reliability model using exponential and log shaped type for the perspective of learning effect.

In this study, we were discussed censoring failure time and predicted time, using nonlinear regression models which error terms are different model was studied. For predict the future failure time (Ra and Kim, 2013), the S-curve type model, growth model and quadratic-type curve regression model were used.

Literature review: If the arithmetic mean increase pattern, results of the arithmetic mean test represents reliability growth. The Laplace trend test results from the Laplace factor value has varying between -2 and 2 indicate stable reliability (Yang and Kim, 2014; Tae-Jin, 2014). Also, the S-curve regression model, growth model and quadratic model were compared (Shin and Kim, 2013) in order to predict result of the future failure time.

MATERIALS AND METHODS

Nonlinear regression model: If possible, a linear relationship between the explanatory variable x and dependent variable y, the linear model represents the relationship as follows (Vamsidhar *et al.*, 2013):

$$y_i = b_0 + b_1 x_i + \varepsilon_i, (i = 1, 2, \cdots, n) \quad (1)$$

where, b_0 and b_1 are regression coefficients and $\varepsilon_i \sim N(0, \sigma^2)$. However, the nonlinear case (curve linear model) by taking the log or the weight can is converted into the linear model (Ra and Kim, 2013).

Growth regression model: The growth regression model was used in this study to predict the future failure time. The growth model is known as follows (Ra and Kim, 2013):

$$\hat{\ln}(y_i) = \beta_0 + (\beta_1 \times x_i), (i = 1, 2, \cdots, n) \quad (2)$$

Note that, β_0 and β_1 are estimation of b_0 and b_1, x_i denote explanatory variable and y_i is dependent variable.

Quadratic regression model: The quadratic regression model was used in this study to predict the future failure time. The quadratic model is known as follows (Isaac *et al*., 2012):

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2, (i = 1, 2, \cdots, n) \quad (3)$$

Note that, β_0 and β_1, β_2 are estimation of b_0 and b_1, b_2 and x_i denote explanatory variable and y_i is dependent variable.

S-curve regression model: The S-curve regression model was used in this study to predict the future failure time. The S-curve model is known as follows (Hyun-Cheul and Hee-Cheul, 2012):

$$\hat{\ln}(y_i) = \beta_0 + (\beta_1 / x_i), (i = 1, 2, \cdots, n) \quad (4)$$

Note that, β_0 and β_1 are estimation of b_0 and b_1 and x_i denote explanatory variable and y_i is dependent variable.

Criteria to determine the optimal model: How the predicted value and the actual value of the material to determine whether the error measure is the criterion for better prediction techniques (Ra and Kim, 2013). In general, the coefficient of determination, modify the coefficient of determination and the sum mean squared criterion are possible (Yang, 2017).

Coefficient of determination: The coefficient of determination (R^2) is defined as the Sum of the Square by the Regression line (SSR), the Sum of Squared Errors (SSE) and called the Sum Total Squared (SST). R^2 can measure how successful the fit is in explaining the variation of the data. It is defined as follows (Tae-Jin, 2014; Yang, 2017):

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (5)$$

Where:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2, SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2$ and $\hat{y}$ is estimation of y.

Mean square error: As the basis for determining the efficient model, the MSE (Mean Square Error) is defined as follows (Ra and Kim, 2013; Yang, 2017):

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \quad (6)$$

Using difference the predicted and the actual value, the minimum of MSE model is efficient.

RESULTS AND DISCUSSION

The proposed analysis and prediction algorithm: In this study, we were used to the following steps-algorithm for prediction using nonlinear regression:

Nonlinear regression algorithm:

Step 1: Validation analysis of fault data through Laplace trend test analysis

Step 2: Calculate parameter estimates for the proposed model

Step 3: Calculate coefficient of determination (R^2) and Mean Square Error (MSE) to determine the optimal model

Step 4: Analyze future failure time prediction ability for true values

Step 5: Determine the optimal model after analyzing the predicted failure data.

Using all of the above steps, we would like to determine the effective model. In order to realize this process, we was assayed as follows:

The failures time data (Prasad *et al*., 2011) summarized in Table 1 for the prediction. For his data, the first, trend data should be preceded (Kim, 2013). Laplace trend test analysis typically performed. Showing in Fig. 1, the

Table 1: Failure time data

Failure number	Failure time (h)	Failure number	Failure time (h)
1	30.02	16	151.78
2	31.46	17	177.50
3	53.93	18	180.29
4	55.29	19	182.21
5	58.72	20	186.34
6	71.92	21	256.81
7	77.07	22	273.88
8	80.90	23	277.87
9	101.90	24	453.93
10	114.87	25	535.00
11	115.34	26	537.27
12	121.57	27	552.90
13	124.97	28	673.68
14	134.07	29	704.49
15	136.25	30	738.68

Fig. 1: Laplace trend test

Table 2: Parameter estimation of each model

Models	Parameter estimation β_0	β_1	β_2
Quadratic	106.911	-17.152	1.265
S-curve	5.519	-3.198	-
Growth	3.514	0.102	-

Table 3: Model summary

Models	Model summary R^2	MSE
Quadratic	0.959	1877.633
S-curve	0.460	38091.780
Growth	0.970	1726.382

Laplace factor between -2 and 2 appears confidence growing by reliability growth properties. Therefore, using this data it is possible to estimate the time to failure time. Parameter estimates are summarized in Table 2.

In Table 3, coefficient determination and mean square error is summarized. This table because growth regression model than any models value of coefficient of determination is high and mean square error is low, growth regression model than any models in terms of goodness of fit is effective model.

Explanatory notes

Explanatory variable: Failure number, dependent variable: failure time, β_0 , β_1 and β_2: estimation of b_0, b_1 and b_2.

Explanatory notes

Explanatory variable: Failure number, dependent variable: failure time, R^2: Coefficient of determination, MSE: Mean Square Error.

Summarized in Fig. 2, from the point 1 (failure number) to the point 30 (failure number), prediction estimation value using S-curve model the growth model and the quadratic model are drawing. All models is similar to the true value in this figure, the growth model shows relatively better prediction and S-curve shows lower (underestimate to the true value), the quadratic model appears higher (overestimate to the true value).

In Table 4, basic statistic information failure times were summarized from the point 1 (failure number) to the point 30 (failure number) using S-curve model, the growth model and the quadratic model. In this table because estimated values of the maximum and the minimum of S-curve model appears from 10.181-224.132 in terms of range, S-curve model than any models shows better range. In addition, for the average, the quadratic model is the highest and the S-curve model is lowest. Because quadratic model is the highest for the dispersion (standard deviation and variance) and S-curve model is relatively small. If want to use quadratic model, the security about dispersion measures is necessary.

In Table 4, basic statistic information failure times were in Fig. 3 and Table 5 were summarized, failure times of future were predicted from the point 31 (failure number) to the point 40 (failure number) and statistic information using S-curve model, the growth model and the quadratic model.

In Fig. 3, predicted from of all models appearing in non-decreasing properties. In addition,

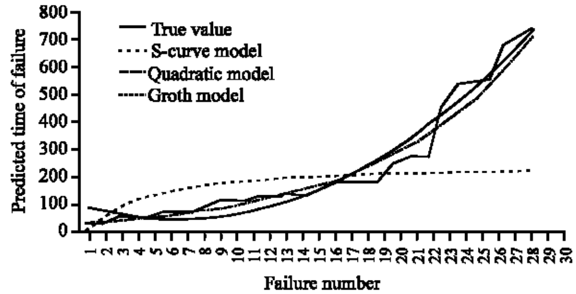

Fig. 2: Predictive failure from one point time to 30 point time

Table 4: Basic statistic information about predictive failure time data (1-30)

Statistics	Quadratic model	S-curve model	Growth model
Mean	239.697	180.950	233.940
SE	38.780	9.847	35.752
Median	145.254	202.818	163.079
SD	212.407	53.933	195.823
Variance	45116.858	2908.786	38346.568
Range	681.901	213.951	676.003
Minimum value	48.825	10.181	37.193
Maximum value	730.726	224.132	713.196
Observation number	30.000	30.000	30.000

Table 5: Basic statistic information about predictive failure time data (31-40)

Statistics	Quadratic model	S-curve model	Growth model
Mean	1102.492	227.735	1302.906
SE	69.628	0.561	125.910
Median	1092.373	227.861	1250.409
SD	220.182	1.774	398.161
Variance	48479.970	3.146	158532.100
Range	653.879	5.282	1185.198
Minimum value	790.731	224.904	789.663
Maximum value	1444.610	230.186	1974.861
Observation number	10.000	10.000	10.000

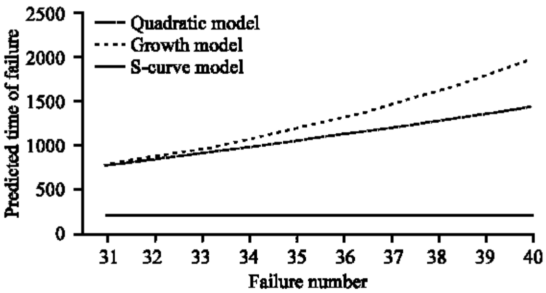

Fig. 3: Predictive failure time from 31-40 point time

growth model appearing rise is relatively high and S-curve model appearing gains are relatively low (Yang, 2017).

In Table 5 because range estimation of the S-curve model appears relatively low and growth model is the highest. In terms of the dispersion (standard deviation and variance), S-curve model is relatively small. Thus, if want to use growth model, the security about dispersion measures is necessary.

CONCLUSION

As software systems play an increasingly important role in computer systems, intensive studies have been carried out to ensure the software reliability. Evaluation can be modeled as software reliability growth test time or run-time failure time, more realistic and relatively efficient model in this area is unknown but model selection using the coefficient of determination and the mean square error was presented for effective comparison.

Therefore, fluctuations in the quality of the products that are produced in the manufacturing process which despite more than continue to run without any action or improvement process, so, the quality of the product will fall dramatically.

This study because growth regression model than any models value of coefficient of determination is high and mean square error is low, growth regression model than any models in terms of goodness of fit is effective model.

For systems design and related industries, the researchers anticipate and consider the growth of the information to censoring data can be considered, using growth prediction method, utilizing the software quality could be helpful to the improvement of software reliability. Therefore, in this study, the proposed prediction method can be used as an alternative model in this field in terms of failure time. From this study, software developers have to consider the nonlinear regression model by prior knowledge of the software to identify failure modes which can was helped.

ACKNOWLEDGEMENT

Funding for this study was provided by Namseoul University.

REFERENCES

Chiu, K.C., Y.S. Huang and T.Z. Lee, 2008. A study of software reliability growth from the perspective of learning effects. Reliab. Eng. Syst. Safety, 93: 1410-1421.

Gokhale, S.S. and K.S. Trivedi, 1999. A time/structure based software reliability model. Ann. Software Eng., 8: 85-121.

Huang, C.Y., 2005. Performance analysis of software reliability growth models with testing-effort and change-point. J. Syst. Software, 76: 181-194.

Hyun-Cheul, S. and K. Hee-Cheul, 2012. The study for software future forecasting failure time using curve regression analysis. J. Inf. Secur. Korea Convergence Secur. Assoc., 12: 115-121.

Isaac, O.A., A.A. Adedeji and I.R. Ismail, 2012. Polynomial regression model of making cost prediction in mixed cost analysis. Math. Theory Model., 2: 14-23.

Kim, H.C., 2013a. Assessing software reliability based on NHPP using SPC. Int. J. Software Eng. Applic., 7: 61-70.

Kim, H.C., 2013b. The comparative study of NHPP delayed S-shaped and extreme value distribution software reliability model using the perspective of learning effects. Int. J. Adv. Comput. Technol., 5: 1210-1218.

Prasad, R.S., K.R.H. Rao and R.R.L. Kantha, 2011. Software reliability measuring using modified maximum likelihood estimation and SPC. Int. J. Comput. Applic., 21: 1-5.

Ra, Y.S. and H.C. Kim, 2013. Study for Predict of the Future Software Failure Time using Nonlinear Regression. In: IT Convergence and Security 2012, Kim, K. and K.Y. Chung (Eds.). Springer, Dordrecht, Netherlands, ISBN:978-94-007-5859-9, pp: 1117-1124.

Shin, H.D. and H.C. Kim, 2013. The comparative study of software optimal release time based on NHPP software reliability model using exponential and log shaped type for the perspective of learning effects. Int. J. Adv. Comput. Technol., 5: 120-129.

Tae-Jin, Y., 2014. The comparative analysis of software failure time based on software reliability model and nonlinear regression model. Korea Knowl. Inf. Technol. Soc., 9: 723-731.

Vamsidhar, Y., Y. Srinivas and P.R. Srinivasa, 2013. Software reliability enhancement by using effective NHPP models. Intl. J. Comput. Sci. Technol., 4: 312-314.

Yang, T.J. and H.C. Kim, 2014. Comparative software infinite NHPP reliability growth model based on the intensity function of log-linear and log-power forms. J. Smart Comput. Rev., 4: 451-458.

Yang, T.J., 2017. The performance analysis comparative study depend on software reliability model and curve regression model. Intl. J. Soft Comput., 12: 313-317.