Reliability Analysis Using Weighted Combinational Models for Web-based Software

Chao-Jung Hsu
Department of Computer Science
National Tsing Hua University, Hsinchu, Taiwan
d948330@oz.nthu.edu.tw

Chin-Yu Huang
Department of Computer Science
National Tsing Hua University, Hsinchu, Taiwan
cyhuang@cs.nthu.edu.tw

ABSTRACT
In the past, some researches suggested that engineers can use combined software reliability growth models (SRGMs) to obtain more accurate reliability prediction during testing. In this paper, three weighted combinational models, namely, equal, linear, and nonlinear weight, are proposed for reliability estimation of web-based software. We further investigate the estimation accuracy of using genetic algorithm to determine the weight assignment for the proposed models. Preliminary result shows that the linearly and nonlinearly weighted combinational models have better prediction capability than single SRGM and equally weighted combinational model for web-based software.

Categories and Subject Descriptors
D.2.4 [Software/Program Verification]: Reliability.

General Terms: Reliability, Measurement, Performance, Management.

Keywords: Software Reliability Growth Model, Weighted Combination, Genetic Algorithm.

1. INTRODUCTION
Since 1970s, many software reliability growth models (SRGMs) have been proposed for the reliability estimation of software products during software development processes [2]. However, application of SRGMs has shown that there is commonly great disagreement in software reliability predictions, while none of them can be trusted to provide consistently accurate results across different applications. Consequently, it is hard for managers to select an adequate model to predict software reliability for their developed software. Therefore, some researches suggested that engineers can use combined SRGMs to obtain more accurate software reliability prediction during testing [2, 3].

Actually, the development of web-based software is very different from the non-web commercial off-the-shelf (COTS) software. Web-based applications have the characteristics of cross-platform support, massive user population, data security requirement, and complicated network environment. In recent years, Tian *et al.* [6] tried to evaluate software reliability based on web workload and failure data. Tsai *et al.* [7] also proposed a service-oriented software reliability model that dynamically evaluates the reliability of web services. In this paper, we will propose three simple and workable combinational models by combining different SRGMs to measure software reliability for web-based software, and study how to integrate genetic algorithm (GA) for the related weight assignments. Finally, results from applications to a real failure data obtained from web-based ERP system are presented.

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

2. PROPOSED MODELS
2.1 Weighted Combinations
In general, SRGMs depend on failure data to estimate failure pattern and reflect different testing conditions [2]. Sometimes, SRGMs may result in prediction bias from the collected data. One common way to eliminate this bias is suggested by averaging or combining the estimated results of models [1, 3]. On the other hand, software development processes involve many intricate human activities and environmental factors that may influence the applicability of SRGMs. Hence, different SRGMs have different underlying assumptions which are used to describe testing process, and they may have a certain relationship between SRGMs. However, because such relationship is unclear and still an open issue, here we formalize three weighted combinational models based on the calibration of *arithmetic mean* and *geometric mean* [1], which respectively represent linear and nonlinear relationships between models. These weighted combinational models are *equally weighted* (EW), *linearly weighted* (LW) and *nonlinearly weighted* (NLW) combinational models.

If the linear relationship exists in models, LW can be expressed as follows [3]:

$$\hat{E}(t) = \sum_{i=1}^{n} w_i \times E_i(t),\quad \sum_{i=1}^{n} w_i = 1, \tag{1}$$

where n is the number of models, $E_i(t)$ is the estimated result (i.e., cumulative number of faults) of the ith model by time t, and w_i is a stationary weight given to the ith model. LW is relevant to the arithmetic summation of estimated results with weights [1]. In the simplest combination, an equal weight could be assigned to each model for deriving a final prediction.

Similarly, if the relationship between models is nonlinear, NLW can be defined as follows:

$$\hat{E}(t) = \sqrt[n]{\prod_{i=1}^{n} E_i(t)^{w_i}},\quad \sum_{i=1}^{n} w_i = n. \tag{2}$$

The identity of NLW is based on the nth root of the geometric product of estimated results with exponential weights [1]. Clearly, if we take a logarithm on Eq.(2), it is given as

$$\log(\hat{E}(t)) = \frac{1}{n}\sum_{i=1}^{n} w_i \times \log(E_i(t)). \tag{3}$$

We can see that NLW will become a special case of LW. Likewise, when w_i are all equal to 1 in Eq.(2) or Eq.(3), NLW can have an equal weight.

SRGMs could be divided into several types based on their reliability growth curves [2]. In this paper, some SRGMs, such as the Goel-Okumoto (GO), the delayed S-shaped (DeS), the inflected S-shaped (InfS), and the generalized Goel-Okumoto (GGO) models, are considered as the candidate models for our proposed weighted combinations. It is noted that the parameters of these candidate models will be estimated by applying Maximum Likelihood Estimation (MLE) in advance [2].

2.2 Genetic Algorithms for the Weight Assignment

From Eqs.(1) and (2), it appears an annoying problem in terms of determining appropriate weight for each model. Fundamentally, this problem can be treated as a searching problem. Here we suggest the use of GA to conduct weight assignment. GA is a kind of evolutionary algorithms that uses bit-level computations for finding approximate solutions [4]. Its searching process includes several steps: *chromosome representation*, *selection scheme*, *crossover operator*, and *mutation operator*. This searching process will run these steps iteratively and evolve solutions until a best weight assignment is found or termination condition is reached. The evaluation of this searching process is determined by the value of fitness function [4]. Fitness function is an indicator of the fitness of every chromosome that shows how close this chromosome is to the desired solution. We can define the fitness function as follows:

$$fitness = 1/-\log(ML), \quad (4)$$

where log(*ML*) is the logarithm of maximum likelihood function (LML) [3]. Our goal is to maximize the fitness function for finding the best weighted assignment of our proposed models.

3. PRELIMINARY RESULTS

A real failure data based on a web-based ERP system (WebERP) [5] is used for illustration. This system was written by PHP language with about 220K lines of code. Over the course of 60 months (from Aug. 2003 to Jul. 2008), 146 failures were detected. Related parametric settings for GA are given as below: population size = 100 chromosomes, crossover rate = 0.6, and mutation rate = 0.2 [4].

To validate the proposed models with EW and traditional SRGMs, three criteria, named LML (see Eq.(4)), Akaike Information Criterion (AIC) [2], and relative error (RE) curve [2], are used for model comparisons. Results of parameter estimation of all selected models and related weight assignments are given in Table 1. Comparison results for all models and the plot of RE curves are presented in Table 2 and Figure 1, respectively.

Table 1. Parameter Estimation and Weight Assignment.

GO	a=598.162, r=4.46E-3
DeS	a=534.070, r=1.71E-2
InfS	a=814.140, r=8.40E-3, ϕ=2.001
GGO	a=998.314, r=7.86E-4, c=1.291
EW	w_{GO}=0.25, w_{DeS}=0.25, w_{InfS}=0.25, w_{GGO}=0.25
LW	w_{GO}=0, w_{DeS}=0.331, w_{InfS}=0.669, w_{GGO}=0
NLW	w_{GO}=0, w_{DeS}=0, w_{InfS}=1.595, w_{GGO}=2.405

a is the total number of faults; r is fault detection rate; ϕ is inflection factor; c is test quality.

Table 2. Comparison Results.

Models	LML	AIC	Models	LML	AIC
EW	178.417	372.018	GO	185.078	374.156
LW	177.762	361.524	DeS	**190.807**	**385.615**
NLW	178.094	362.188	InfS	179.097	364.195
			GGO	178.453	362.906

The shadow indicates the most accurate result in the LML or AIC criterion;
The bold indicates the least accurate result in the LML or AIC criterion.

From Table 1, it is noted that LW has two weights for DeS and InfS, and NLW also has two weights for InfS and GGO. According to the classification of SRGMs, DeS and InfS are S-shaped models, and GGO could approach S-shaped model as $c > 1$ [2]. On the other hand, from Table 1, we have $\phi = 2.001$. That is, the inflection rate (i.e., ratio of the number of detectable faults to the total number of faults in the program) is 0.333, indicating the growth curve of this data is S-shaped. This corresponds with the finding provided by LW and NLW, i.e., w_{GO}=0.

From Table 2, it is worth noting that LW and NLW give better fit for this dataset. Besides, LW and NLW are more accurate than EW in both LML and AIC criteria. Figure 1 also depicts that the LW and NLW still give a better fit to the observed data in most data points, and also predict the future behavior well in the tail area. On the other hand, since software reliability is related to the cumulative number of faults, we can easily use the estimated results of proposed LW and NLW models to compute software reliability. Let conditional reliability be $R(\Delta t|t)$, which is defined as the probability that a software will be successful in the time interval $(t, t+\Delta t]$ [2]. If we choose LW and $\Delta T = 0.01$ (month) as an example, then we have $R = 0.972$. During testing, this information would be very useful and important for project manager (PM) to determine the total amount of software testing time, resource allocation, and quality improvement, etc.

Finally, from above experiments and discussions, if PM plans to employ SRGM for estimation of reliability growth of products, engineers should select some SRGMs, and apply them in parallel. Although SRGMs sometimes give good results, there is no single model that can be trusted to give accurate results in all circumstances. In this case, we can conclude that the proposed weighted combinational models may be a good approach to provide a more accurate prediction of software reliability for web-based software than using SRGMs or EW.

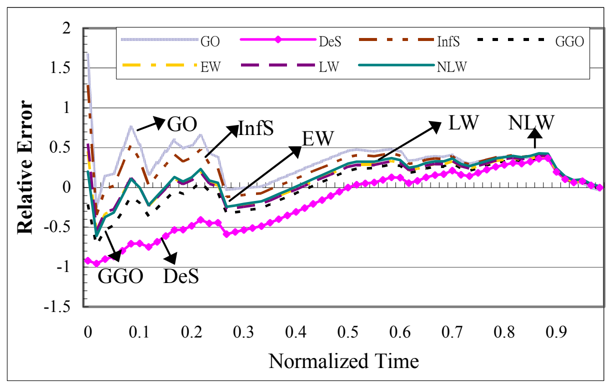

Figure 1. Relative Error Curves.

4. CONCLUSIONS

In this paper, three weighted combinational models, namely, EW, LW, and NLW, are proposed and can be used for reliability estimation of web-based software. Results from applying the proposed models to a web-based ERP system are compared with traditional SRGMs and show that the proposed models can give better accuracy in software reliability prediction.

5. ACKNOWLEDGEMENTS

The work described in this paper was supported by the National Science Council, Taiwan, under Grant NSC 97-2221-E-007-052-MY3.

6. REFERENCES

[1] Bullen, P.S., Mitrinovic, D.S., and Vasic, M., *Means and Their Inequalities*, Springer, 1988.

[2] Lyu, M.R., *Handbook of Software Reliability Engineering*, McGraw-Hill, 1996.

[3] Lyu, M.R. and Nikora, A., "Applying Reliability Models More Effectively," *IEEE Software*, vol. 9, no. 4, pp. 43-52, 1992.

[4] Mitchell, M., *An Introduction to Genetic Algorithms*, The MIT Press, 1998.

[5] SourceForge.net, http://sourceforge.net, Last Access Date: 2008 Dec. 11.

[6] Tian, J., Rudraraju, S., and Li, Z., "Evaluating Web Software Reliability Based on Workload and Failure Data Extracted from Server Logs," *IEEE Transactions on Software Engineering*, vol. 30, no. 11, pp. 754-769, 2004.

[7] Tsai, W.T., Zhang, D., Chen, Y., Huang, H., Paul, R., and Liao, N., "A Software Reliability Model for Web Services," *Proceedings of IASTED International Conference on Software Engineering and Applications*, pp. 144-149, MIT Cambridge, USA, 2004.