한국정보전자통신기술학회논문지(jkiiect)'16-04, Vol.9 No.2
논문 16-09-02-141 http://dx.doi.org/10.17661/jkiiect.2016.9.2.141

유연한 와이블 확장분포의 형상모수를 이용한 NHPP 소프트웨어 신뢰성 모형에 관한 비교연구

김희철* , 신현철**

A Comparative Study for NHPP Software Reliability Model based on the Shape Parameter of Flexible Weibull Extension Distribution

Hee-Cheul Kim*,Hyun-Cheul Shin**

요 약 소프트웨어 고장분석을 위한 비동질적인 포아송과정에서 결함당 고장발생률이 상수이거나, 단조 증가 또는 단조 감소하는 패턴을 가질 수 있다. 본 논문에서는 수리시점에서도 고장이 발생할 상황을 반영하는 무한고장 NHPP 모형들을 비교 제시하였다. 소프트웨어 신뢰성분야에서 많이 사용되는 유연한 와이블 확장분포에 근거한 무한고장 소프트웨어 신뢰성모형에 대한 비교문제를 제시하였다. 그 결과 형상모수가 비교적 작은 경우가 효율적으로 나타났다. 그리고 모수 추정법은 최우추정법을 이용하였고 모형선택은 평균제곱오차와 결정계수를 이용하였다. 이 연구를 통하여 소프트웨어 개발자들은 형상모수에 따른 소프트웨어 고장현상을 파악하는데 어느 정도 도움을 줄 수 있을 것으로 사료된다.

Abstract NHPP software reliability models for failure analysis can have, in the literature, exhibit either constant, monotonic increasing or monotonic decreasing failure occurrence rates per fault. In this paper, infinite failures NHPP models that repairing software failure point in time reflects the situation, was presented for comparing property. Commonly used in the field of software reliability based on Flexible Weibull extension distribution software reliability of infinite failures was presented for comparison problem. The result is that a relatively small shaping parameter was effectively. The parameters estimation using maximum likelihood estimation was conducted and model selection was performed using the mean square error and the coefficient of determination.. In this research, software developers to identify software failure property follows shape parameter, some extent be able to help is considered.

Key Words : Flexible Weibull Extension Distribution, Infinite Failure Model, Mission Time, Software failure point, Maximum likelihood estimation

1. 서론

소프트웨어 신뢰성은 일정한 환경조건에서 일정기간동안 고장이 나지 않고 운영 할 수 있는 확률이다.

따라서 소프트웨어 개발과정에서 소프트웨어 신뢰성은 중요한 문제이다. 이 문제는 사용자의 요구조건과 테스팅 비용을 만족시켜야 한다. 소프트웨어 테스팅(디버깅)면에서 비용을 줄이기 위해서는 소프트웨어의 신뢰성의 변동과 테스팅 비용을 사전에 알고 있어야 효율적이다. 따라서 신뢰도, 비용 및 방출시간의 고려사항을 가진 소프트웨어 개발 과정은 필수 불가결 하다.

지금까지 많은 소프트웨어 신뢰성모형이 제안

*Corresponding Author : Department of Industrial & Management Engineering, Namseoul University(kim1458@nsu.ac.kr)
Received April 10, 2016 Revised April 21, 2016 Accepted April 21, 2016

되었다. 이 중에서 비동질적 포아송 과정(Non-homogeneous Poisson process; NHPP)에 의존한 모형[1]은 에러탐색 과정측면에서는 우수한 모형이고 이러한 모형은 결함이 발생하면 즉시 제거되고 디버깅과정에서 새로운 결함이 발생되지 않는다는 가정을 하고 있다.

이 분야에서 Gokhale과Trivedi [2]은 고양된 비동질적인 포아송과정 모형(Enhanced NHPP model)을 제시하였고 Goel 과 Okumoto [3]은 결함의 누적수가 S-형태나 지수적인 형태(S-shaped or exponential-shaped)를 가진 평균값함수(Mean value function)를 이용한 지수적인 소프트웨어 신뢰성모형(Exponential software reliability growth model)을 제안 하였다. 이모형에 의존한 일반화 모형은 Yamada 와 Ohba [4]에 의해 지연된 S-형태 신뢰성장모형(Delayed S-shaped reliability growth model)과 변곡된 S-형태 신뢰성장모형(Inflection S-shaped reliability growth model)이 제안되었다. Zhao [5]는 소프트웨어 신뢰도에서 변환점 문제를 제시하였고 Shyur [6]는 변환점을 이용한 일반화한 신뢰도 성장모형을 제안하였다. Pham와 Zhang[7]는 테스팅 커버리지(Coverage)를 측정하여 소프트웨어 안정도를 평가 할 수 있는 소프트웨어 안정도모형을 제시했다. 그리고 Huang [8]은 일반화 로지스틱 테스팅 노력 함수(Generalized logistic testing-effort function)와 변환점 모수(Change-point parameter)를 통합하여 효율적인 소프트웨어 신뢰성 예측기술을 제시하기도 하였다. 그리고 또 다른 측면에서 S-형태 모형은 소프트웨어 관리자들이 소프트웨어 및 검사도구에 익숙해지는 학습과정을 설명할 수 있다고 하였다[9]. 또한 kim[10]은 반-로지스틱분포를 이용한 학습과정에 대한 연구도 이루어졌다.

본 논문에서는 유연한 와이블 확장모형 [11]을 이용한 형상모수의 효율성과 그 특징을 알아보고자 한다.

2. 관련연구

2.1 유연한 와이블 확장분포

유연한 와이블 확장분포(Flexible Weibull extension distribution)은 M. Bebbington, C.D. Lai, 와 R. Zitikis[12]에 의해 제안된 분포로서 위험함수의 패턴이 감소형태 혹은 감소형태를 따르는 분포로서 확률밀도함수와 분포함수는 다음과 같이 알려져 있다 [11].

$$f(t)=\left(\alpha+\frac{\beta}{t^2}\right)e^{\alpha t-\frac{\beta}{t}}\ e^{-e^{\left(\alpha t-\frac{\beta}{t}\right)}} \quad (1)$$

$$F(t)=1-e^{-e^{\left(\alpha t-\frac{\beta}{t}\right)}} \quad (2)$$

단, $\beta, \alpha, t>0$.

그리고 위험함수는 다음과 같이 유도된다.

$$h(t)=\ f(t)/(1-F(t))=\left(\alpha+\frac{\beta}{t^2}\right)e^{\alpha t-\frac{\beta}{t}} \quad (3)$$

2.2 효율적 모형을 위한 모형의 비교기준

일반적으로 모형에 대한 효율성을 조사하기 위한 기준으로서 평균제곱오차(Mean square error; MSE)와 결정계수(Coefficient of determination; R^2)를 사용한다[9].

평균제곱오차는 실제 관찰 값과 예측 값에 대한 차이를 측정하는 도구로서 다음과 같이 정의된다.

$$MSE=\frac{\sum_{i=1}^{n}\left(m(x_i)-\hat{m}(x_i)\right)^2}{n-k} \quad (4)$$

단, $m(x_i)$은 시간$(0,\ x_i]$까지 나타난 고장들의 누적개수를 의미하고 $\hat{m}(x_i)$는 x_i 시점까지 평균값 함수로부터 추정된 고장의 누적개수를 의미한다. 그리고 n은 관찰 값의 수이고 k는 모수의 수를 의미한다. 즉, 작은 평균제곱오차 값을 가진 모형은 효율적인 모형이 된다.

또한, 결정계수는 예측 값의 차이에 대한 설명력을 나타내는 도구로서 다음과 같이 정의된다. 따라서 보다 큰 결정계수를 가진 모형은 효율적인 모형으로 간주된다.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(m(x_i) - \hat{m}(x_i)\right)^2}{\sum_{i=1}^{n}\left(m(x_i) - \sum_{j=1}^{n} m(x_j)/n\right)^2} \tag{5}$$

3. 제안된 유연한 와이블 확장분포을 이용한 NHPP 모형

(1)식과 (2)식을 이용하여 유연한 와이블 확장분포을 유한고장 NHPP로 접근하면 강도함수(위험함수)와 평균값함수는 다음과 같이 표현 할 수 있다[11, 12].

$$\lambda(t) = h(t) = f(t)/(1-F(t)) \tag{6}$$
$$= \left(\alpha + \frac{\beta}{t^2}\right) e^{\alpha t - \frac{\beta}{t}}$$

$$m(t) = -\ln(1-F(t)) = e^{\alpha t - \frac{\beta}{t}} \tag{7}$$

단, $\beta, \alpha, t > 0$.

한편, 시간 $(0, t]$ 까지 조사하기 위한 시간절단(Time truncated)모형은 n 번째까지 고장시점 자료를

$$x_k = \sum_{i=1}^{k} t_k \quad (k = 1,2,\cdots,n\,;\, 0 \le x_1 \le x_2 \le \cdots \le x_n) \tag{8}$$

이라고 하면 데이터 집합 D_t 는 $\{n, x_1, x_2 \cdots x_n\,;t\}$ 와 같이 구성된다. n 번째까지 고장시점이 관찰된 고장절단모형일 경우에 데이터 집합 D_{x_n} 은 $\{x_1, x_2 \cdots x_n\}$ 으로 구성되며, 이 시간절단 모형에서 Θ 을 모수공간이라고 표시하면 무한고장 우도함수는 다음과 같이 알려져 있다[1].

$$L_{NHPP}(\Theta|\,\underline{x}\,) = \left(\prod_{i=1}^{n} \lambda(x_i)\right) \exp(-m(x_n)) \tag{9}$$

단, $\underline{x} = (x_1, x_2, x_3, \cdots, x_n]$ 이고 Θ은 미지의 모수집합을 의미하고 우도함수 (9)식에 (6)과 (7)식을 적용하고 최우추정법(Maximum Likelihood Estimation; MLE)을 이용한 유연한 와이블 확장 모형에 대한 로그우도함수는 다음과 같이 유도된다.

$$\ln L(\alpha,\beta\,|\,\underline{x}) = \sum_{i=1}^{n} \ln\left(\alpha + \frac{\beta}{x_i^2}\right) + \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} \frac{1}{x_i} - \left(e^{\alpha x_n - \frac{\beta}{x_n}}\right) \tag{10}$$

단, $\underline{x} = (x_1, x_2, x_3, \cdots, x_n]$.

본 연구에서는 형상모수 β를 고정된 상수 $(0.5, 1, 1.5)$로 간주하여 모수추정을 하고자 한다.

따라서 최우추정값 $\hat{\alpha}_{MLE}$ 는 다음 식을 만족한다.

$$\frac{\partial L(\alpha,\beta|\,\underline{x})}{\partial \alpha} = \sum_{i=1}^{n} \ln\left(\frac{x_i^2}{\alpha x_i^2 + \beta}\right) + \sum_{i=1}^{n} x_i - x_n\left(e^{\alpha x_n - \frac{\beta}{x_n}}\right) = 0 \tag{11}$$

또한, 임무시간에 따른 신뢰도는 다음과 같이 유도 된다[1].

$$\hat{R}(\tau|x_n) = \exp\left[-m(\tau + x_n) + m(x_n)\right] \tag{12}$$

단, τ는 임무시간(Mission time)을 의미하고

$$m(\tau + x_n) = e^{\alpha(\tau + x_n) - \frac{\beta}{(\tau + x_n)}},$$

$$m(x_n) = e^{\alpha x_n - \frac{\beta}{x_n}}.$$

4. 소프트웨어 고장시간 자료 분석

이 절에서는 소프트웨어 고장 시간자료[13](Failure time data)를 이용하여 본 논문에서 제시하는 소프트웨어 신뢰모형들을 이용한 소프트웨어 고장시간을 분석하고자 한다. 이 자료의 고장시간은 18.735 시간단위에 30번의 고장이 발생된 자료이며 [표 1]에 나열 되어 있다.

표 1. 소프트웨어의 고장시간자료
Table 1. Failure time data of software

Failure number	Failure time (hours)	Failure number	Failure time (hours)
1	0.479	16	10.771
2	0.745	17	10.906
3	1.022	18	11.183
4	1.576	19	11.779
5	2.61	20	12.536
6	3.559	21	12.973
7	4.252	22	15.203
8	4.849	23	15.64
9	4.966	24	15.98
10	5.136	25	16.385
11	5.253	26	16.96
12	6.527	27	17.237
13	6.996	28	17.6
14	8.17	29	18.122
15	8.863	30	18.735

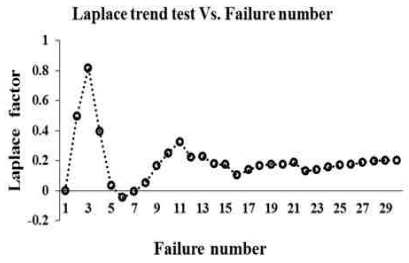

그림 1 라플라스 추세검정
Fig. 1. Laplace trend test

또한 제시하는 신뢰모형들을 분석하기 위하여 우선 자료에 대한 추세검정이 선행 되어야 한다[14]. 추세분석에는 일반적으로 라플라스 추세검정(Laplace trend test)을 사용한다. 이 검정을 실시한 결과 [그림 1]에서 라플라스 추세 검정의 결과는 라플라스요인(Factor)이 -2와 2사이에 존재함으로서 즉, 극단값(Extreme value)이 존재하지 않으므로 이 자료를 이용하여 신뢰성장모형을 제시하는 것이 효율적임을 시사하고 있다[15].

모수추정은 최우추정법을 이용하고 모수추정을 용이하게 하기 위하여 원래의 고장시간 데이터를 변수변환($Failure\ time \times 10^{-1}$)하여 적용하였다. 비선형 방정식의 계산방법은 수치 해석적 기본방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기 값을 0.01과 5.0을, 허용한계(Tolerance for width of interval)는 10^{-5} 을 주고 수렴성을 확인하면서 충분한 반복횟수인 100번을 C-언어를 이용하여 모수추정을 수행하였다.

표 2. 각 모형에 대한 모수추정값 및 MSE , R^2
Table 2. Parameter estimation of each model MSE , R^2

Parameter	MLE	Model Comparison	
		MSE	R^2
$\beta=0.1$	$\hat{\alpha}_{MLE}=1.6987$	**83.170**	**0.8836**
$\beta=0.5$	$\hat{\alpha}_{MLE}=1.7670$	111.066	0.8594
$\beta=1.0$	$\hat{\alpha}_{MLE}=1.8794$	134.659	0.8199

Note. MLE : Maximum likelihood estimation;
MSE: Mean square error;
R^2 : Coefficient of determination.

최우추정법의 결과와 모형에 대한 효율성을 조사하기 위한 기준으로서 MSE(평균제곱오차)와 R^2(결정계수)가 [표 2]에 요약되었다. 이 표에서 유연한와이블 확장모형에서 형상모수가 $\beta=0.1$인 경우가 $\beta=0.5$ 와 $\beta=1.0$인 경우보다 실제 값과 예측 값에 대한 차이를 측정하는 평균제곱오차가 낮다 그리고 형상모수 $\beta=0.1, 0.5, 1.0$인 경우 모두 예측 값의 차이에 대한 설명력을 의미하는 설명력도 모두 80% 보다 높게 나타나기 때문에 형상모수가 $\beta=0.1$인 경우가 $\beta=0.5$ 와 $\beta=1.0$인 경우보다 효율적인 모형으로 간주할 수 있다.

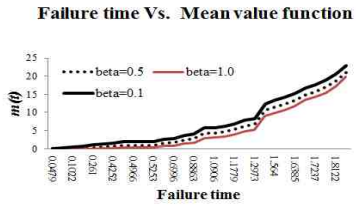

그림 2. 평균값함수
Fig. 2. Mean value function

[그림 2]에서는 평균값함수의 패턴을 요약 하였다. 이 그림에서도 유연한 와이블 확장모형에서 형상모수가 $\beta=0.1$인 경우가 $\beta=0.5$ 와 $\beta=1.0$인 경우보다 보다 참값에 더 가깝게 추정되었다. 그리고 [그림 3]에서는 위험함수의 패턴을 요약하였다. 이 그림에서도 와이블 확장모수가 $\beta=1.0$인 경우는 감소형태를 따르는 패턴이고 $\beta=0.5$인 경우와 $\beta=1.0$인 경우는 비-감소 속성을 나타내고 있다.

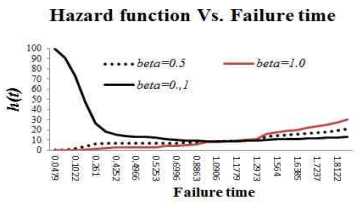

그림 3. 위험함수
Fig. 3. Hazard function

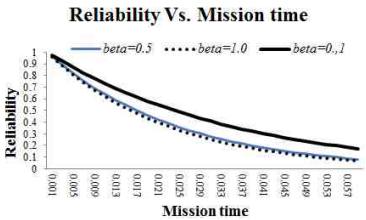

그림 4. 신뢰도
Fig. 4. Reliability

[그림 4]에서 보여 주듯이 임무시간에 대한 신뢰도 그림에서는 와이블 확장모수가 $\beta=0.1, 0.5, 1.0$인 경우에 대하여 신뢰도를 비교 했을 때 예상된 바와 같이 임무시간이 증가 할수록 $\beta=0.1$인 경우가 $\beta=0.5$ 와 $\beta=1.0$인 경우보다 신뢰도의 상승으로 나타나고 있다.

5. 결론

대용량 소프트웨어가 수정과 변경하는 과정에서 결점의 발생을 거의 피 할 수 없는 상황이 현실이다.

소프트웨어 신뢰성은 개발의 최종단계에 있는 테스트 공정이나 실제 사용단계에 있어서 소프트웨어 내에 존재하는 고장 수나 고장 발생시간에 의해서 효과적으로 평가할 수 있는 상황으로 그 평가기술이 중요하게 된다. 따라서 소프트웨어개발의 테스트공정이나 실제사용단계에 있어서 고장 발생환경이나 고장 발생현상을 수리적으로 모형화가 가능하면 평가를 할 수 있다. 테스트시간이나 혹은 실행시간, 발생된 고장 수와 고장 발생시간과의 관계를 효율적으로 관리함으로서 소프트웨어 신뢰도를 성장 시 킬 수 있다. 이러한 과정을 소프트웨어 성장과정이라고 볼 수 있다.

본 연구에서는 수리시점에서도 고장이 발생할 상황을 반영하는 무한고장 NHPP모형들을 비교 제시하였다.

따라서 본 연구에 제안된 유연한 와이블 확장분포를 따르는 형상모수에 근거한 무한고장 NHPP모형들을 비교 제시하였다. 그 결과 형상모수가 비교적 작은 경우에 효율적으로 나타났다. 이 연구를 통하여 소프트웨어 개발자들은 형상모수에 근거한 소프트웨어 고장형태를 고려함으로서 소프트웨어 고장형태에 대한 사전지식을 파악하는데 어느 정도 도움을 줄 수 있으리라 사료된다.

REFERENCES

[1] Hee-Cheul KIM, "The Comparative Study of NHPP Delayed S-Shaped and Extreme Value Distribution Software Reliability Model using the Perspective of Learning

Effects", International Journal of Advancements in Computing Technology, Vol. 5, No.9, pp. 1210 -1218, 2013.

[2] Gokhale, S. S. and Trivedi, K. S. A, "time/structure based software reliability model", Annals of Software Engineering. 8, pp. 85-121. 1999.

[3] Goel A L, Okumoto K, "Time-dependent fault detection rate model for software and other performance measures", IEEE Trans. Reliab. 28, pp.206-11, 1978.

[4] Yamada S, Ohba H, "S-shaped software reliability modeling for software error detection", IEEE Trans. Reliab, 32, pp.475-484, 1983.

[5] Zhao M, "Change-point problems in software and hardware reliability", Commun. Stat. Theory Methods, 22(3), pp.757-768, 1993.

[6] Shyur H-J, "A stochastic software reliability model with imperfect debugging and change-point", J. Syst. Software 66, pp.135-141, 2003.

[7] Pham H, Zhang X., "NHPP software reliability and cost models with testing coverage", Eur. J. Oper. Res, 145, pp.445-454, 2003.

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

[9] Kuei-Chen, C., Yeu-Shiang, H., and Tzai-Zang, L., "A study of software reliability growth from the perspective of learning effects", Reliability Engineering and System Safety 93, pp. 1410 - 1421, 2008.

[10]Hee-Cheul KIM, "The Comparative Study of NHPP Half-Logistic Distribution Software Reliability Model using the Perspective of Learning Effects", Journal of Next Generation Information Technology, Vol. 4, No. 8, pp. 132-139, 2013.

[11]M.A.E-Damcese1,Abdelfattah Mustafa, B.S.El-Desouky and M.E.Mustafa, "The Kumaraswamy Flexible Weibull Extension",International Journal of Mathematics And its Applications Vol. 4, Issue 1-A, pp. 1-14, 2006.

[12] M.Bebbington, C.D. Lai and R. Zitikis, "A exible Weibull extension", Reliability Engineering & System Safety", Vol. 92, No. 6, pp. 719-726, 2007.

[13] Y. HAYAKAWA and G. TELFAR, "Mixed Poisson-Type Processes with Application in Software Reliability", Mathematical and Computer Modelling, 31, pp. 151-156, 2000.

[14] Kim, Hee Cheul, "The Comparative Study of Software Optimal Release Time of Finite NHPP Model Considering Half-Logistic and Log-logistic Distribution Property", The Journal of Korea Society of Digital Industry and Information, 9(2): 1-10, June 2013.

[15] K. Kanoun and J. C. Laprie, "Handbook of Software Reliability Engineering", M.R.Lyu, Editor, chapter Trend Analysis. McGraw-Hill New York, NY, pp. 401-437, 1996.

저자약력

김 희 철 (Hee-Cheul Kim) **[정회원]**

- 1992년 2월 : 동국대학교 통계학과 (이학석사)
- 1998년 8월 : 동국대학교 통계학과 (이학박사)
- 2005년 3월 ~ 현재 : 남서울대학교 산업경영공학과 교수

<관심분야> 소프트웨어신뢰성 공학, 전산 통계, 웹프로그래밍

신 현 철(Hyun-Cheul Shin) **[정회원]**

- 2002년 2월 : 원광대학교 컴퓨터공학과 (공학박사)
- 2005년 3월 : 한국정보처리학회 이사
- 1994년 3월 ~ 현재 : 백석문화대학교 컴퓨터공학부 교수

<관심분야> 통신공학, 컴퓨터과학, 소프트웨어 공학, 소프트웨어공학