Journal of the Society of Korea Industrial and Systems Engineering
Vol. 25, No. 6, pp.7－11, December 2002.

마스크 데이터를 이용한 컴포넌트의 고장발생확률 추정

박 창 규

성균관대학교 기술혁신센터

Estimating Outbreak Probabilities of Systems and Components with Masked Data

Chang-Gyu Park

Technology Innovation Center

This paper estimates defect and outbreak probabilities of each individual component from some subset of masked data where the exact component causing system failure might be unknown. A system consists of k components that fails whenever there is a defect in at least one of the components. Due to cost and time constraints it is not feasible to learn exactly which components are defective. Because, test procedures ascertain that the defective components belong to some subset of the k components. This phenomenon is termed masking. We describe a, b, c type in which a sample of masked subsets is subjected to intensive failure analysis. This recorded data of a, b, c type enables maximum likelihood estimation of defect probability of each individual component and leads to outbreak of the defective components in future masked failures.

Keywords : masked data, MLE, defect probability

1. 서 론

컴퓨터와 같은 다양한 컴포넌트들로 구성된 시스템의 고장은 고장원인을 정확하게 관측하기 어렵다. 시스템의 고장원인 분석과 고장원인별 진단을 정확하게 조사하기 위하서는 고장분석장비, 인력, 그리고 시설 등의 비용의 많이 들기 때문에 사용현장에서는 제대로 이루어지지 못하고 있다. 그래서, 시스템의 고장기록은 쉽게 밝혀진 고장원인과 단지, 시스템이 고장이 발생했다는 정보로 기록되고 있다. 이와 같이 정확한 고장원인은 모르지만 고장이라는 정보를 가지고 있는 데이터를 마스크 데이터(masked data)라 한다.

마스크 데이터를 이용한 신뢰성 분석에 관한 연구들은 최우 추정량의 추정치를 구하는 연구로 진행되어 왔다. Miyakawa는 컴포넌트의 수명 데이터는 지수분포를 따르고 2개의 컴포넌트가 직렬로 연결되어 있는 시스템의 모수 추정을 연구했고[5], Usher와 Hodgson은 Miyakawa의 연구를 3개의 컴포넌트로 구성된 시스템까지 확장하고, 간단한 수치해석을 제안하였다[7].

Lin, Usher 그리고 Guess는 3개의 컴포넌트로 구성된 시스템의 정확한 최우 추정량을 찾는 절차를 연구하였고[4], Guess, Usher 그리고 Hodgson는 컴포넌트의 수명 데이터는 지수분포를 따르는 3개의 컴포넌트가 직렬로 구성된 시스템에서 마스크의 효과와 수명 데이터가 마스크 데이터일 때 일반적인 우도함수의 유도를 연구하였다[2]. 이들 연구에서 기본 가정은 마스크 된 고장원인들이 서로 독립이라는 것이다. Lin과 Guess는 2개의 컴포넌트로 이루어진 시스템의 모수를 추정하는데 있어 고장원인이 종속되어 있을 경우, 고장원인인 컴포넌트들의 마스크 발생확률에 대한 고장률의 변화를 연구하였다[3]. Usher와 Guess는 마스크 데이터가 와이블 분포를 따를 때 컴포넌트 모수 추정의 그래픽적 접근방법을 제

시하였다[9].

지금까지 연구들은 마스크 데이터가 컴포넌트들의 실제적인 고장에 기준을 둔 것이 아니라 컴포넌트들의 마스크 발생확률이 같다는 가정에 기준을 두고 연구한 것이다. 이러한 가정은 우도함수가 마스크가 발생하는 확률과 컴포넌트의 신뢰도로만 구성된다는 것이다. 즉, 컴포넌트 신뢰도의 최우 추정치만을 찾는데 목적이 있는 것이다. Lin과 Guess는 마스크의 발생확률이 같지 않을 때, 컴포넌트들의 고장원인에 따라 즉, 고장원인이 종속되어 있는 경우로 좀 더 현실적인 문제에 대해 연구를 하였다[3].

시스템이 고장나는데 있어서 모든 컴포넌트들이 고장나는 것은 아니다. 즉, 하나나 그 이상의 컴포넌트들이 고장남으로 인해 시스템은 정지하게 된다. 고장인 시스템들이 동작하게 되는 경우는 고장난 컴포넌트나 동작이 불확실한 컴포넌트를 교체하는 경우이다. 이러한 상황은 시스템을 가동시키는데 필요한 컴포넌트 이상으로 많이 보유해야하는 문제가 발생한다. 따라서 컴포넌트들의 정확한 고장발생확률을 추정할 수 있으면, 추정된 고장발생확률을 이용하여 시스템의 동작을 유지시킬 수 있는 컴포넌트들을 보유함으로써 추가적인 재고비용은 감소하게 된다.

본 연구는 마스크가 발생하는 경우, 시스템의 마스크 데이터를 이용하여 컴포넌트들의 고장 발생확률을 추정하는 방법을 제시한다.

2. 연구내용

본 연구에서는 마스크 되어 있는 시스템의 고장원인은 대개 다음 세 가지 경우로 기록되어 있다.

① 컴포넌트의 고장이 발생한 것과 하지 않은 것을 파악 후 기록
② 고장이 발생한 것 중 어느 컴포넌트에서 고장이 발생했는지 파악 후 기록
③ 어떤 컴포넌트가 고장원인인지 모르는 경우 기록

시스템의 고장원인이 어떤 컴포넌트에 의해 고장이 발생했는지, 그리고 고장이 발생했지만 고장원인을 정확히 모르는 경우와 고장 기록원의 문제로 인한 고장원인이 기록이 되지 못하는 경우로 구분 할 수 있다. 따라서 시스템의 수명 데이터가 기록되어진 경우에 따라 시스템의 분석 형태를 다음과 같이 구분한다.

a 형태 : 컴포넌트의 정확한 고장원인을 아는 형태
b 형태 : 컴포넌트의 고장원인을 알지만 정확히는 모르는 형태
c 형태 : 컴포넌트의 고장원인을 모르는 형태

본 연구에서 사용되는 기호는 다음과 같다.

N	시스템 수
k	컴포넌트 수
F_i	컴포넌트 i의 고장확률
R_i	$1-F_i$: 컴포넌트 i의 신뢰도
L	우도함수
a, b, c	시스템의 형태에 따른 구분
A, B, C	각 시스템 상태의 우도함수
$\mathfrak{I}(\cdot)$	지시함수 : $\mathfrak{I}$(비고장)=0, $\mathfrak{I}$(고장)=1, $\mathfrak{I}$(모르는 경우)=2
δ_i	$\mathfrak{I}$(컴포넌트 i가 고장)
$\boldsymbol{\delta}$	$(\delta_1, \cdots, \delta_k)$: 시스템을 구성하는 컴포넌트 형태 벡터
γ_i	$\mathfrak{I}$(컴포넌트 i의 관측)
$\boldsymbol{\gamma}$	$(\gamma_1, \cdots, \gamma_k)$: 관측된 시스템 벡터
N^A_δ	$\boldsymbol{\delta}$ 에서 a형태의 수
$N^B_{\gamma,\delta}$	$\boldsymbol{\delta}$ 에서 $\boldsymbol{\gamma}$인 b형태의 수
N^C_γ	$\boldsymbol{\gamma}$ 에서 c형태의 수
$\Pr\{\boldsymbol{\delta} \mid \boldsymbol{\delta}\}$	$\boldsymbol{\delta}$ 에서 정확한 고장원인을 아는 확률 (고장확률)
$\Pr\{\boldsymbol{\gamma} \mid \boldsymbol{\delta}\}$	$\boldsymbol{\delta}$ 에서 $\boldsymbol{\gamma}$ 를 갖는 마스크 확률
$\prod\{\boldsymbol{\delta} \mid \boldsymbol{\gamma}\}$	$\boldsymbol{\gamma}$ 에서 실제 고장원인인 $\boldsymbol{\delta}$의 확률 (고장 발생확률)

마스크가 발생하는 경우 시스템의 마스크 데이터를 이용하여 컴포넌트들의 고장 발생확률을 추정하는 방법의 기본가정으로는 다음과 같다.

① 각 독립인 시스템은 k개로 구성되고, 데이터는 go/no-go로 기록된다.
② 컴포넌트의 고장은 독립이다.
③ 시스템의 고장은 컴포넌트 고장에 기인하고, 정확히 컴포넌트의 고장은 기록된다.

시스템의 형태는 a, b, c의 세 형태로 구성된다. 시스템의 세 형태를 이용한 우도함수는 다음과 같다.

$$L = A \cdot B \cdot C \quad (1)$$

A, B, C ≡ 시스템의 우도함수 기여 (a형태, b형태, c형태)

2.1 a 형태 시스템

$\boldsymbol{\delta} \equiv (\delta_1, \cdots, \delta_k)$에서 a 형태 시스템의 우도함수 기여는 다음과 같다.

$$A_\delta = \left[\Pr\{\delta \mid \boldsymbol{\delta}\} \cdot \prod_{i=1}^{k} F_i^{\delta_i} \cdot R_i^{1-\delta_i} \right]^{N_\delta^A} \quad \cdots\cdots (2)$$

따라서 모든 a 형태의 시스템에 대한 우도함수 기여는

$$A = \prod_{\Omega_a} A_\delta \quad \cdots\cdots (3)$$

$\Omega_a \equiv a$ 형태 시스템의 모든 $\boldsymbol{\delta}$
이다.

2.2 b 형태 시스템

$(\boldsymbol{\gamma}, \boldsymbol{\delta}) \equiv (\gamma_1, \cdots, \gamma_k, \delta_1, \cdots, \delta_k)$에서 b 형태 시스템의 우도함수 기여는 다음과 같다.

$$B_{\gamma,\delta} = \left[\Pr\{\gamma \mid \boldsymbol{\delta}\} \cdot \prod_{i=1}^{k} F_i^{\delta_i} \cdot R_i^{1-\delta_i} \right]^{N_{\gamma,\delta}^B} \quad \cdots\cdots (4)$$

모든 b 형태의 시스템에 대한 우도함수 기여는

$$B = \prod_{\Omega_b} B_{\gamma,\delta} \quad \cdots\cdots (5)$$

$\Omega_b \equiv b$ 형태 시스템의 모든 $\boldsymbol{\gamma}, \boldsymbol{\delta}$
이다.

2.3 c 형태 시스템

$\boldsymbol{\gamma} \equiv (\gamma_1, \cdots, \gamma_k)$에서 c 형태 시스템의 우도함수 기여는 다음과 같다.

$$C_\gamma = \left[\sum_{\delta_1=0}^{1} \cdots \sum_{\delta_k=0}^{1} \Pr\{\gamma \mid \boldsymbol{\delta}\} \cdot \prod_{i=1}^{k} F_i^{\delta_i} \cdot R_i^{1-\delta_i} \right]^{N_\gamma^C} \quad \cdots\cdots (6)$$

모든 c 형태의 시스템에 대한 우도함수 기여는

$$C = \prod_{\Omega_c} C_\gamma$$

$\Omega_c \equiv c$ 형태 시스템의 모든 $\boldsymbol{\gamma}$

이다

2.4 고장발생확률

부정확한 고장원인이 포함되어 있는 마스크 데이터를 이용하여 세 가지 형태의 시스템 전체에서 발생하는 컴포넌트의 고장 확률을 고장 발생확률 또는 고장 진단확률이라 한다. 고장 발생확률을 추정하는 목적은 마스크되어 있는 데이터를 이용하여 각 컴포넌트의 고장 발생확률 즉, 시스템을 구성하고 있는 전체 컴포넌트 중에서 컴포넌트 각각의 고장 발생확률을 분석하는데 있다. 고장 발생확률은 시스템의 고장분석시 비용이 많이 들거나 시간이 많이 소요될 경우, 고장 발생확률이 높은 컴포넌트를 선택하는 의사결정 기준이 된다.

고장원인의 기록 형태에 따른 시스템의 우도함수와 베이지안 정리를 이용하면, 고장 발생확률은 다음과 같이 주어진다.

$$\Pi\{\boldsymbol{\delta} \mid \boldsymbol{\gamma}\} = \frac{\Pr\{\boldsymbol{\gamma} \mid \boldsymbol{\delta}\} \cdot \prod_{i=1}^{k} F_i^{\delta_i} \cdot R_i^{1-\delta_i}}{\sum_{\delta_1', \cdots, \delta_k'} \Pr\{\boldsymbol{\gamma} \mid \boldsymbol{\delta}'\} \cdot \prod_{i=1}^{k} F_i^{\delta_i'} \cdot R_i^{1-\delta_i'}} \quad \cdots\cdots (7)$$

3. 추정방법

식(7)의 추정치는 식(1)의 최우추정치와 동일한 성질을 갖는다. 식(7)을 이용한 고장 발생확률의 최우추정량을 추정하기는 어렵기 때문에 수치적 방법(numerical method)을 이용한다. 고장 발생확률을 추정하는 절차는 다음과 같다.

(1) b 형태로 정리된 마스크 데이터를 이용하여 초기 고장 발생확률을 설정한다.

$$\tilde{\Pi}^{(0)}\{\delta \mid \gamma\} = \frac{N_{\gamma,\delta}^B}{\sum_{\delta_1', \cdots, \delta_k'} N_{\gamma,\delta'}^B} \quad \cdots\cdots (8)$$

(2) c 형태의 데이터 수를 결정한다.

$$N_{\gamma,\delta}^{C(1)} = N_\gamma^C \cdot \tilde{\Pi}^{(0)}\{\delta \mid \gamma\} \quad \cdots\cdots (9)$$

(3) 컴포넌트 i의 고장확률 F_i을 구한다.

컴포넌트의 고장확률 F_i과 $\Pr\{\boldsymbol{\gamma} \mid \boldsymbol{\delta}\}$는 시스템의 각 형태로 정리된 데이터에 의해 계산할 수 있다. 고장이 발생한 컴포넌트 i와 시스템 구분에 대한 고장 확률 F_i는 다음과 같다.

$$\widetilde{F}_i^{(1)} = N^{-1} \cdot \left[\sum_{\delta_i = 1} N_\delta^A + \sum_{\delta_i = 1} N_{\gamma,\delta}^B + \sum_{\delta_i = 1} N_{\gamma,\delta}^{C(1)} \right] \cdots\cdots (10)$$

(4) δ의 형태에 대한 컴포넌트의 고장확률을 계산한다.

δ 의 형태에 대한 컴포넌트의 고장확률은 다음과 같다.

$$\widetilde{\Pr}^{(1)}\{\delta \mid \delta\} = \frac{N_\delta^A}{N_\delta^A + \sum_\gamma (N_{\gamma,\delta}^B + N_{\gamma,\delta}^{C(1)})} \cdots\cdots (11)$$

(5) $\gamma \neq \delta$ 인 경우의 마스크 확률을 구한다.

따라서 같은 방법으로 $\gamma \neq \delta$인 경우 마스크 확률은

$$\widetilde{\Pr}^{(1)}\{\gamma \mid \delta\} = \frac{N_{\gamma,\delta}^B + N_{\gamma,\delta}^C}{N_\delta^A + \sum_\gamma (N_{\gamma,\delta}^B + N_{\gamma,\delta}^{C(1)})} \cdots\cdots (12)$$

으로 주어진다.

만약, 컴포넌트가 하나인 경우는 위 식들의 추정치가 최우추정치가 된다. 반복적인 과정으로 위 식들을 이용하여 추정치를 추정하는 방법은 EM 알고리즘과 비슷하다[1].

4. 예 제

두 개의 컴포넌트로 구성된 45개의 시스템을 조사하여 3절의 추정방법을 통해 컴포넌트의 고장확률과 고장발생확률을 구하여 본다.

<표 1> 두 개의 컴포넌트를 갖는 시스템

(상태 : 비고장=0, 고장=1, 모르는 경우=2)

구 분	형 태	관 측 수
0,0	a	10
1,0		3
0,1		7
1,1		5
2,2;1,0	b	1
2,2;0,1		3
2,2;1,1		6
2,2	c	10

위 데이터를 정리하면 다음과 같은 의미를 갖는다.

① 45개 중 35개 시스템이 고장이다.
② 35개 중 15개는 정확한 고장원인을 알고 있는 것이다.
③ 35개 중 10개는 고장원인을 정확히 모르는 경우이다.
④ 35개 중 나머지 10개는 전혀 모르는 경우이다.

위 데이터를 가지고 3절에서 제시된 추정방법을 이용하여 각각의 확률을 구하는 절차는 다음과 같다.

(1) b형태의 데이터를 기준으로 초기 고장 발생확률을 설정한다.

$\widetilde{\Pi}^{(0)}\{1,0 \mid 2,2\} = 1/10$
$\widetilde{\Pi}^{(0)}\{0,1 \mid 2,2\} = 3/10$
$\widetilde{\Pi}^{(0)}\{1,1 \mid 2,2\} = 6/10$

(2) 식(9)를 이용하여 c 형태의 데이터 수를 결정한다.

$N_{2,2;1,0}^{C(1)} = 1$
$N_{2,2;0,1}^{C(1)} = 3$
$N_{2,2;1,1}^{C(1)} = 6$

(3) 식(10), (11), (12)를 이용하여 각각의 확률을 구한다.

i) 컴포넌트 1의 고장확률
$\widetilde{F}_1^{(1)} = (3 + 5 + 1 + 6 + 1 + 6)/45 = 0.4889$

ii) 컴포넌트 2의 고장확률
$\widetilde{F}_2^{(1)} = (7 + 5 + 3 + 6 + 3 + 6)/45 = 0.6667$

iii) 고장원인과 조사, 관측한 고장원인과 동일한 고장확률
$\widetilde{\Pr}^{(1)}\{1,0 \mid 1,0\} = 0.6000$
$\widetilde{\Pr}^{(1)}\{0,1 \mid 0,1\} = 0.5385$
$\widetilde{\Pr}^{(1)}\{1,1 \mid 1,1\} = 0.2941$

iv) 각 컴포넌트의 마스크 확률
$\widetilde{\Pr}^{(1)}\{2,2 \mid 1,0\} = 0.4000$
$\widetilde{\Pr}^{(1)}\{2,2 \mid 0,1\} = 0.4615$
$\widetilde{\Pr}^{(1)}\{2,2 \mid 1,1\} = 0.7059$

(4) 지금까지 구한 값들을 이용하여 식(7)의 고장발생확률을 구한다.

<표 2> 컴포넌트로 구성된 시스템에서의 고장확률, 마스크 확률, 고장발생확률

항 목	컴포넌트 1	컴포넌트 2	컴포넌트 1,2
고장 확률	$\hat{F} = 0.47$	$\hat{F}_2 = 0.65$	
정확한 고장원인별 확률	$\widehat{Pr}\{1,0 \mid 1,0\} = 0.52$	$\widehat{Pr}\{0,1 \mid 0,1\} = 0.51$	$\widehat{Pr}\{1,1 \mid 1,1\} = 0.32$
마스크 된 고장원인별 확률	$\widehat{Pr}\{2,2 \mid 1,0\} = 0.48$	$\widehat{Pr}\{2,2 \mid 0,1\} = 0.49$	$\widehat{Pr}\{2,2 \mid 1,1\} = 0.68$
고장발생확률	$\hat{\Pi}\{1,0 \mid 2,2\} = 0.17$	$\hat{\Pi}\{0,1 \mid 2,2\} = 0.37$	$\hat{\Pi}\{1,1 \mid 2,2\} = 0.46$

$$\tilde{\Pi}^{(0)}\{1,0 \mid 2,2\} = 0.1440$$
$$\tilde{\Pi}^{(0)}\{0,1 \mid 2,2\} = 0.3476$$
$$\tilde{\Pi}^{(0)}\{1,1 \mid 2,2\} = 0.5084$$

이와 같은 방법으로 고장 발생 확률이 수렴될 때까지 반복적으로 진행한다. 컴포넌트 고장 발생확률의 최우추정치는 <표 2>에 주어져 있다. 컴포넌트 1과 컴포넌트 2의 고장발생확률은 컴포넌트 2가 컴포넌트 1에 비해서 더 크다. 따라서 시스템의 설계 변경 시 각 컴포넌트의 개선여부를 고려할 경우 컴포넌트 1보다는 컴포넌트 2의 개선이 먼저 이루어져야하며, 서비스 재고관리 등에서는 컴포넌트의 고장발생확률에 따른 수명을 예측하여 각 컴포넌트의 보유 수량을 결정할 수 있다.

5. 결 론

시스템의 고장이 정확한 분석을 통해 고장인 컴포넌트를 파악할 수 있다면 시스템의 신뢰도 추정량은 좋은 추정치를 얻을 수 있다. 하지만, 현장에서 시험을 할 수 없는 서비스 기록 정보나 시험 장비, 시험조건 그리고 분석방법 등의 문제 때문에 시스템의 고장 기록 정보는 정확한 컴포넌트들의 고장으로 기록되지 않고 있는 것이 현실이다. 따라서 마스크 데이터가 존재하는 경우 고장확률, 마스크 확률과 고장발생확률을 추정함으로서 컴포넌트의 수리정책과 수리비용, 서비스 정책, 서비스 재고관리 등에 관한 정확한 시스템 정보를 제공 할 수 있다.

참고문헌

[1] Dempster A., Laird N., Rubin D., "Maximum Likelihood from Incomplete Data via the EM Algorithm", *J. Royal Satistical Soc, ser B*, Vol. 39, pp1-38, 1977.

[2] Guess F., Usher J., Hodgson T., "Estimating system and component reliabilities under partial information on the cause of failure", *J. Statistical Planning & Inference*, Vol 29, pp75-85, 1991.

[3] Lin D., Guess F., "System life data analysis with dependent partial knowledge on the exact cause of system failure", *Microelectronics & Reliability*, Vol. 34, pp535-544, 1994

[4] Lin, D., Usher, J. and Guess, F., "Exact Maximum Likelihood Estimation Using Masked System Data", *IEEE Transactions in Reliability*, Vol.42, No.4. pp631-635, 1993.

[5] Miyakawa, M., "Analysis of Incomplete Data in Competing Risks Model", *IEEE Transactions on Reliability*, Vol. R-33, No. 4, pp293-296, 1984.

[6] Usher, J., "Weibull Component Reliability-Prediction in the Presence of Masked Data", *IEEE Transaction on Reliability*, Vol. 45, No. 2, pp229-232, 1996.

[7] Usher, J. and Hodgson, T., "Maximum Likelihood Analysis of Component Reliability Using Masked System Life-Test Data", *IEEE Transaction on Reliability*, Vol. 37, No. 5, pp550-555, 1988.

[8] Usher J., Alexander S., Thompson J., "System reliability prediction based on historical data", *Quality and Reliability Eng'g Int'l*, Vol. 6, No. 3, pp209-218, 1990.

[9] Usher J., Guess F., "An iterative approach for estimating component reliability from masked system life data", *Quality & Reliability Eng'g Int'l*, Vol.5, pp257-261, 1989.