

# 베이저안 기법을 이용한 안전사고 예측기법

양 희 중\*

\*청주대학교 산업정보시스템공학과

## Safety Analysis using bayesian approach

Hee Joong Yang\*

\*Dept. of Industrial & Information Systems Engineering, Cheongju University

### Abstract

We construct the procedure to predict safety accidents following Bayesian approach. We make a model that can utilize the data to predict other levels of accidents. An event tree model which is a frequently used graphical tool in describing accident initiation and escalation to more severe accident is transformed into an influence diagram model.

Prior distributions for accident occurrence rate and probabilities to escalating to more severe accidents are assumed and likelihood of number of accidents in a given period of time is assessed. And then posterior distributions are obtained based on observed data. We also points out the advantages of the bayesian approach that estimates the whole distribution of accident rate over the classical point estimation.

Keywords : Safety Accident, Bayesian Forecasting, Influence Diagram

### 1. 서 론

국내외에는 늘 안전사고가 발생하고 있으며 이로 인한 인명의 피해도 심각한 실정이다. 안전사고의 발생 및 확산을 현실 상황에 가깝게 모델링하고 이를 바탕으로 효과적으로 예측할 수 있는 기법의 개발은 실로 절실하다. 이와 관련된 연구는 지속적으로 이루어졌고 다양한 방법이 다루어졌다. 대표적인 것이 사건 가지 및 고장 가지를 이용한 분석 기법이다[5],[6].

고장 가지 분석을 통해 모델 파라미터의 추정치를 찾고 이 값이 사건 가지의 모델 파라미터로 이용된다.

이 후 사건 가지상의 전체 파라미터 값을 추정하여 사고 발생률을 추정하고 이 값의 역수로서 다음번 사고까지의 시간을 예측하는 방법을 흔히 쓰고 있다. 이러한 방법을 통해 안전사고 예측에 많은 기여를 해왔지만 이 방법 자체에는 이론적, 그리고 응용상의 결점이 있는 것이 사실이다.

첫째로, 상이한 치명도의 자료를 활용해 다른 수준의

사고 예측을 하는 기법은 전무한 상황이다. 대다수 연구들은 안전사고 변수들과 안전사고 발생 간의 상관관계를 규명하는데 중점을 두고 있다[2],[3],[4].

그러나 고려해야 할 변수의 수는 상당히 많고 변수들 간에는 교호작용도 존재하며 변수 값 확인을 위해 경우에 따라서는 대형 시스템의 세부 분해까지 진행되어야 하기 때문에 이를 규명하는 것은 실제로 불가능에 가까우리 만큼 어려운 작업이다.

또한 흔히 발생하지 않는 대형사고의 경우에는 극히 일부 입수된 자료의 상관분석을 통해 이루어진 사고 발생예측은 그 정확성에서 의심 받지 않을 수 없다. 따라서 변수와의 상관분석 기법을 넘어 데이터의 변동을 파악해 일정기간 내 안전사고 발생 가능성 예측 및 강도, 재해 등의 예측에 중점을 두는 것이 실효성 있는 접근 방법이다. 이를 위해서는 서로 다른 유형의 자료가 유기적으로 연결되어 상호 예측에 기여할 수 있는 형태의 모델이 구축되어야한다.

둘째로, 자료가 입수 되는대로 실시간 모델 파라미터가 보정되는 모델 구축이 이루어져야한다. 사고의 관측 뿐 아니라 사고가 없이 진행되고 있는 그 시각 시각도 새로운 자료로서 입력되고 이를 통해 예측치가 보정될 수 있는 기법이 필요하다.

셋째로, 사고까지의 시간 예측이 모델 파라미터 값의 추정과 분명히 구분되어 이루어 질 수 있어야 한다.

본 논문에서는 이러한 점에 중점을 두어 보다 효과적이고 신뢰성 있는 예측을 할 수 있는 모델 구축 및 방법론 개발에 중점을 둔다.

## 2. 새로운 모형의 제시

### 2.1 점추정 방식의 탈피

안전사고 예측에서 가장 널리 사용되고 있는 방법 중의 하나가 그간 빈번히 사용되어 오던 모델을 설정하고 그 파라미터를 과거자료를 통해 점추정 하는 것이다[7]. 안전사고 예측에서도 흔히 과거자료로부터 안전사고 발생률  $\lambda$ 를 추정하는 방법을 사용하고 있다.

일정기간의 안전사고 자료를 수집하여 안전사고 간 평균시간을 구한 후 그 역수로서 발생률을 추정하고 있는 것이다. 그러나  $X$ 를 안전사고 발생까지의 시간,  $E\langle X \rangle$ 를 그 기대값,  $\lambda$ 를 발생률이라 할 경우, 발생률의 추정치를 다음 식에 의해 구한다면,

$$\hat{\lambda} = \frac{1}{E[X]}$$

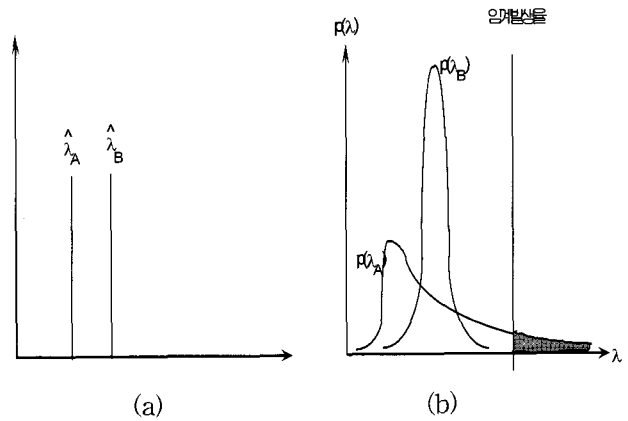
이는  $X$ 가 지수분포를 따르는 특수한 경우에만 성립되며 일반적으로 이 식이 옳다고 볼 수 없다.

이와 같이 발생률을 점추정 하는데서 비롯되는 또 다른 문제점은 추정치 만으로는 안전사고위험성으로부터 충분히 안전한지의 여부를 결정하기 어렵다는 데 있다. 두 안전 시스템의 비교 상황을 보자.

<그림 1(a)>에서 보면 시스템 A에서의 평균 발생률이 시스템 B에서의 평균 발생률보다 낮으므로 두 시스템 중 A가 더 안전하다는 결론을 내릴 수 있다. 그러나 각 시스템의 발생률에 대한 불확실성을 고려하여 A와 B에서의 발생률의 분포를 구하였을 때 <그림 1(b)>와 같은 결과가 얻어진다면 시스템 B에서는 피부로 느끼는 안전사고의 임계 발생률을 넘을 가능성이 없으나 지역 A에서는 이러한 확률이 존재하므로 시스템 A는 시스템 B에 비해 더 안전하다고 받아들일 수 없는 상반된 결론에 도달하게 된다.

따라서 평균 안전사고 발생률 혹은 안전사고까지의 평균시간 등의 한 수치로서 안전성을 대표할 수 없으

며 안전성에 중요한 영향을 미치고 있는 전체 분포가 효과적인 방법을 통해 추정 될 수 있어야 한다.



<그림 1> 점추정과 분포추정의 비교

### 2.2 베이지안 기법의 적용

위에 언급했듯이 안전사고 발생률  $\lambda$ 의 점추정치 뿐 아니라 그 분포 모양도 파악해야하는 필요성이 있으며 이를 위해서는 베이지안 접근방식이 타당하다. 베이지안 방식을 따르면 당연히 분포 모양에 대한 정보를 누적시켜 가는 장점이 있으며 또한 충분한 자료 획득이 어려운 사업장의 안전사고에 관한 예측의 경우에는 전문가의 지식, 유사분야의 자료, 실험을 통해 입수된 자료 등도 가치 있게 활용할 수 있다는 장점이 있다.

이러한 형태의 대표적 예로서 원자력 발전소 사고예측, 항공기 사고 예측 등을 들 수 있다.

### 2.3 Influence Diagram 활용

Influence Diagram은 베이지안 사고 방식에 그 기초를 두고 있으며 모델구축, 계수보정, 예측분포 결정 등을 효율적으로 보좌할 수 있는 도구이다[1].

Influence Diagram은 변수(Random Quantity)들 간의 상호독립 혹은 관련 여부를 표시하는 그래픽 도구로서 원(Circle, node)은 변수(Random Variable)를 나타내며 화살표(Directed Arc)는 서로 영향을 미칠 수 있음 (Possible Statistical Dependency)을 뜻하고 화살표 머리에 있는 원에 대한 조건부 확률은 그에 직접 선행하는 원들만으로서 구해질 수 있음을 의미하고 있다. 어둡게 채워진 원은 관측 가능한 변수를 뜻하며 채워지지 않은 원은 관측 가능한 변수에 영향을 미치는 관측 불가능한 모델 파라미터를 의미한다.

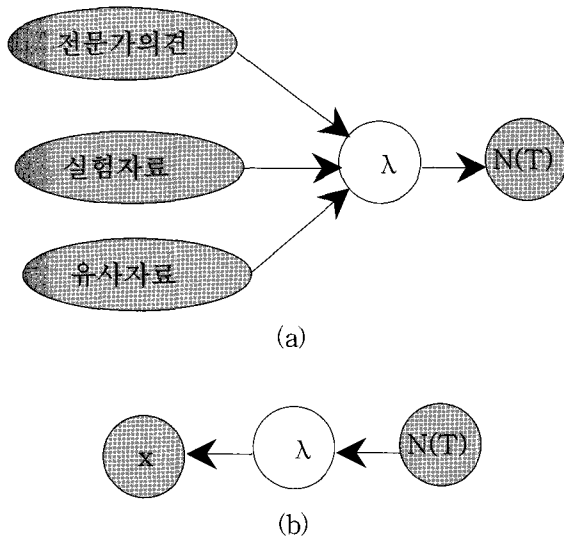
Influence Diagram은 베이지안 사고방식에 그 기초

를 두고 있으며 모델구축, 계수 보정, 예측분포 결정 등을 효율적으로 보좌할 수 있는 유용한 그래픽 도구이다[8],[9].

<그림 2>는 안전사고발생률과 자료간의 통계적 의존관계를 나타낸 Influence Diagram 이며 자료 입수 후 사후 분포를 획득하는 과정이 도해 되어있다.

<그림 2(a)>에서 보듯이  $\lambda$ 의 사전분포는 아무 현장 자료가 없는 상태에서 전문가의 의견, 가능한 실험자료, 유사한 분야에서 수집된 자료 등에 기초를 두고 설정된다. 베이저안 기법은 인간의 논리적인 사고 변화를 그대로 수식 및 수치로 변환시킨다는 장점도 지니고 있다. 이  $\lambda$ 는 <그림 2(a)>에서와 같이 (0, T) 시간 동안의 안전사고 수,  $N(T)$ , 에 영향을 미치게 된다.

이것이 현장자료 입수 전의 모델이다. 이후 T 시간이 흘러  $N(T)=n(T)$  의 자료를 입수했다면 우리는 기존에 갖고 있던  $\lambda$ 에 대한 생각을 관측된 자료를 통해 보정하게 될 것이다. 이 의존성이 <그림 2(b)>의 화살표( $N(T)$ 에서  $\lambda$ 로 향하는)에 나타나 있으며 이제  $P(\lambda | N(T)=n(T))$ , 즉  $\lambda$ 의 사후 분포를 구하게 된다.



<그림 2> 안전사고 발생률과 자료간의 통계적 의존관계를 나타낸 Influence Diagram

안전사고까지의 시간 X 에 대한 예측분포  $P(x|n(T))$  는 아래와 같이 관측 불가능한 파라미터  $\lambda$ 를 적분 소거함으로써 구해지며 이는 influence diagram 에서는  $\lambda$  node 를 흡수 (absorb) 하는 과정에 해당한다.

$$P(x | n(T)) = \int P(x | \lambda)P(\lambda | n(T))d\lambda$$

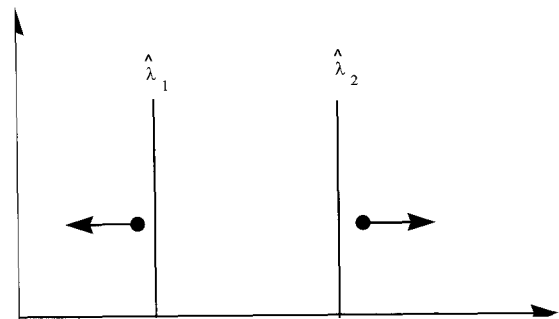
위 식에서 볼 수 있듯이 파라미터를 추정된 값으로서 나타내는 것이 아니라 안전사고 발생 시간에 대한 분포 전체를 관측된 자료에 입각하여 나타내게 된다.

## 2.4 서로 다른 유형의 고장자료 이용

기존의 연구에서 지적되어야 할 또 하나의 문제점은 각 유형별 안전사고 자료를 서로 독립적으로 사용하고 있다는 점이다.

만일 일정한 치명도 이상의 안전사고에 대해 관심이 있다면 이러한 유형의 안전사고자료를 수집하여 이와 관련된 발생률만을 추정하고 있다. 일반적으로 치명적 안전사고의 발생률  $\lambda_1$  은 경미한 안전사고의 발생률  $\lambda_2$  보다 더 작은 값으로 추정되며 일정기간 T가 지난 후 경미한 안전사고는 상당수 관측되나 치명적 안전사고는 발생치 않을 경우도 종종 있을 수 있다.

그렇다면 관측 후 추정치가 <그림 3>에서와 같이  $\lambda_1$  은 좌로  $\lambda_2$  는 우로 이동되어 경미한 안전사고 발생 가능성은 증가하나 치명적 안전사고 발생 가능성은 감소하고 있다는 모순된 결론에 달하게 된다.

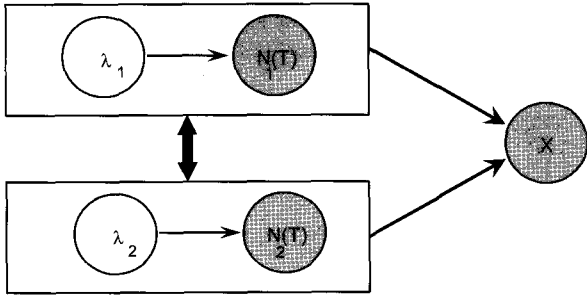


<그림 3> 독립적 분석에서 고장율의 변동모습

이러한 모순은 서로 다른 유형의 안전사고를 함께 고려해 볼 때는 위와 같이 분명히 확인할 수 있으나 대부분의 연구가 치명적 유형의 안전사고에 초점을 맞추어 진행되고 있기 때문에 이점을 파악하지 못하는 경우가 많이 있다. 이를 보완하는 방법으로서 아래 <그림 4>의 Influence Diagram prototype을 제시 할 수 있다.

여기서  $\lambda_1$  은 치명적 안전사고 수  $N_1(T)$ 에,  $\lambda_2$  는 경미한 안전사고 수  $N_2(T)$ 에 각각 영향을 주고 있으나 이 둘 간에는 분명히 존재하는 상호관계가 모델에 명시 되어야하고, 하나의 모델 속에서 파라미터 간, 변수 간 서로 연계되고 데이터 및 정보가 교류 될 수 있어야 한다.

그리하여 치명적 안전사고 발생까지의 시간  $X_2$  에 대한 예측이 각각 서로 다른 유형의 안전사고 수를 종합하여 효율적 이루어져서 예측 정확도가 높여 저야 한다.



<그림 4> 상이한 유형의 자료를 혼합하는 모형

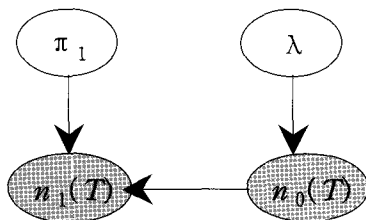
### 3. 예측모델의 개발

위의 내용에 기초하여 사고의 유형을 간단히 두개로 간주하여 사고를 예측하는 모델을 구축한다. 사고의 유형이 셋 이상이 되는 경우도 이를 기초로 어렵지 않게 확장 할 수 있다.

사고의 유형을 두 종류, 즉, 경미한 사고와 치명적 사고로 나눌 수 있는 경우를 보자. 일정 기간 (0, T) 사이에 발생한 경미한 사고의 수를  $m_0(T)$ , 이 기간 중 발생한 치명적 사고의 수를  $m_1(T)$  라고 하자. 여기에 새로운 개념의 사고 수를 추가한다.

즉, 수준 0 의 사고 수,  $n_0(T)$ 를 경미한 사고 수와 치명적 사고 수를 합친 모든 유형의 사고 수, 수준 1 이 사고 수,  $n_1(T)$ , 를 치명적 사고의 수라고 하자. 즉,  $n_0(T) = m_0(T) + m_1(T)$ ,  $n_1(T) = m_1(T)$ 가 된다.

$\lambda$ 를 모든 유형의 사고 발생률,  $\pi_1$ 을 모든 유형의 사고 중 수준 1의 사고로 진행되는 확률이라고 하자. 그러면 이 모델은 <그림 5>의 상호연관도로 표현될 수 있다.



<그림 5> 2 수준 모델

모델 파라미터의 초기분포  $p(\lambda)$ 와  $p(\pi_1)$ 이 추정되고 파라미터가 주어졌을 경우의 우도  $p(n_0(T) | \lambda)$ 와  $p(n_1(T) | n_0(T), \pi_1)$ 이 가정된다. 여기서 수준 0의 사고 수는 모든 유형의 사고 발생률  $\lambda$ 에만 영향을 받고 수준 1의 사고 수는 수준 1의 사고 수와 확산 확률  $\pi_1$ 에 영향을 받고 있음을 명확히 볼 수 있다.

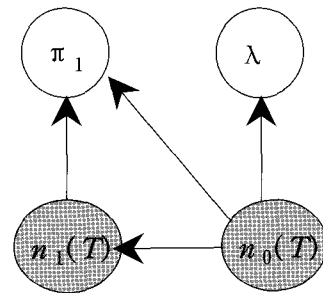
따라서 이 모델의 결합 확률분포는 아래와 같이 표현 될 수 있다:

$$p(\lambda, \pi_1, n_0(T), n_1(T)) = p(\lambda)p(\pi_1)p(n_0(T)|\lambda)p(n_1(T)|n_0(T),p(\lambda, \pi_1))$$

자료가 입수되면 입수된 자료에 포함되어 있는 정보를 이용해서 모델 파라미터를 보정하여 사후분포를 구한다:

$$p(\lambda, \pi_1 | n_0(T), n_1(T)) = p(\lambda|n_0(T))p(\pi_1|n_0(T), n_1(T))$$

이 과정은 상호연관도에서 화살표 방향전환의 과정에 해당되며 그 과정은 <그림 6>에 있다.

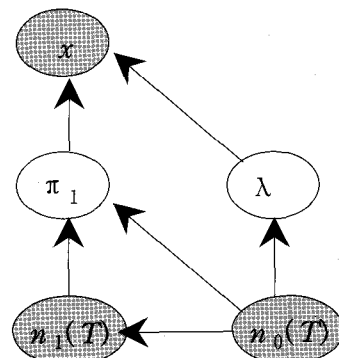


<그림 6> 자료에 입각한 파라미터 보정

다음 번 치명적 사고까지의 시간을  $x$  라고 하면 이는 모델 파라미터  $\lambda$  와  $\pi_1$  에 영향을 받게 되며 데이터에 의해 보정된 관측 불가능한 모델 파라미터는 적분 소거함으로써 데이터에 입각한  $x$  의 예측 분포를 구하게 된다.

$$p(x|n_0(T), n_1(T)) = \iint p(x, \lambda, \pi_1 | n_0(T), n_1(T)) d\lambda d\pi_1 = \iint p(x|\lambda, \pi_1) p(\lambda|n_0(T)) p(\pi_1|n_0(T), n_1(T)) d\lambda d\pi_1$$

이 과정이 그림 7의 상호연관도에 나와 있다.



<그림 7> 파라미터 적분 소거에 의한 예측분포

#### 4. 결론

안전사고 분석 시 종종 이용되는 사건 가치와 결합 가치에 포함되어 있는 정보들을 상호 연관도로 옮길 수 있다. 상호 연관도에서는 사건 가치와 결합가치에 실려 있는 정보 외에도 파라미터간의 통계적 의존성 혹은 독립성 여부까지도 설명 될 수 있다는 장점이 있다. 상호 연관도를 이용하여 서로 다른 유형의 사고 자료까지도 활용할 수 있는 모델이 개발 되었다.

이를 통해 흔히 발생하지 않는 유형의 사고도 종종 입수 할 수 있는 자료를 통해 더욱 효율적으로 분석할 수 있게 된다. 사고 분석을 사고 발생률의 추정에 그치는 것이 아니고 예측 분포를 구하는 것 까지 진행된다.

자료가 입수되는 대로 모델 파라미터가 보정되며 이렇게 보정된 파라미터 값에 따라 발생 하는 사고까지의 사건에 대한 분포는 이러한 파라미터를 적분 소거함으로써 구해진다.

이렇게 하여 예측 값이 관측 불가능한 모델 파라미터값에 의존하는 것이 아니고 관측된 자료에만 입각하여 구해지게 된다.

#### 5. 참고 문헌

[1] Aitchison, I.R., and Dunsmore, "Statistical Prediction Analysis", Cambridge University Press, (1975)

[2] Buske, S.Z. and Holland, D.F., "Risk Assessment Technology for the Evaluation of Tritium Accident Mitigation", Nuclear Technology Fusion, 4 (1983) : 539-543

[3] Holdren, J.P. et.al., "Exploiting the Competitive Potential of Magnetic Fusion Energy: The Interaction of Economics with Safet and Environmental Characteristics", Fusion Technology, 13(1988) : 7-56

[4] Khan, Faisal et al. "A New Methodology for Safety Management Based on Feedback from Credible Accident-Probabilistic Fault Tree Analysis System", Journal of Hazardous Materials, 87, No.1 (2002) : 23-36

[5] Piet,S.J. "Implication of Probabilistic Risk Assessment for Fusion Decision Making", Fusion Technology, 10 (1986) : 31-48

[6] Rasmussen,N.C., "Method of Hazard Analysis and Nuclear Safety Engineering" Annals New York Academy of Science, (1981) : 20-36

[7] U.S. Nuclear Regulatory Commission, "Reactor

Risk Reference Document", Office of Nuclear Regulatory Research, NUREG-1150, 1-3 (1987)

[8] Yang, Heejoong, "Approximation Method in Bayesian Prediction of Nuclear Power Plant Accidents", 한국산업공학회지,16, No.21, (1990) : 135-147

[9] Yang, Heejoong, "Forecasting Accidents by Transforming Event Trees into Influence Diagrams", 산업경영시스템학회지, 29, No.1 (2006) : 72-75

#### 저 자 소 개

양 희 중



서울대학교를 졸업하고 Univ. of Texas at Arlington 에서 산업공학 석사, University of California, Berkeley에서 산업공학박사를 받았다. 1996-1997에 Naval Postgraduate School에서 방문교수로 활동하였으며 현재 청주대학교 산업공학과 교수로 재직 중이다. 주요 관심분야는 품질경영과 안전 사고 예측이다.

주소: 충북 청주시 상당구 내덕동 36 청주대학교 산업공학과