



전통적 사고예측모형의 한계 및 개선방안

: Hauer 사고예측모형의 소개 및 적용

What goes problematic in the Existing Accident Prediction Models and How to Make it Better

한 상 진*	김 근 정**	오 순 미***
Han, Sangjin	Kim, Kewnjung	Oh, sunmi

Abstract

The main purpose of this study is to introduce Hauer's(2004) approach that overcomes current accident prediction models' limitation and to apply this approach to Korean situation using fatal accident data on motorways. After developing accident prediction models according to this approach, it is found that AADT and vertical grade could improve fitness of the model, whereas a radius of roads is not related to the number of accidents. The advantage of Hauer's approach is to reduce possibility to eliminate critical variables and to keep uncritical variables when we consider many variables to develop accident prediction models.

Keywords : *accident predictor model, negative binomial distribution, maximum likelihood estimation, good of fitness*

요 지

사고예측모형은 도로에서 발생한 교통사고자료를 통계적으로 모형화한 것으로 종속변수는 과거의 사고건수가 되고 설명 변수로는 주로 사고가 일어난 장소의 도로 기하구조 조건, 교통조건, 운영조건 등 도로의 속성자료가 이용된다. 기존의 사고예측모형의 한계를 극복하고자 새로운 방안인 Hauer의 연구를 구체적으로 소개하고 이를 국내 고속도로 사망사고자료를 통해 적용하였다. Hauer의 방법론에 의한 사고예측모형을 구축한 결과 AADT와 종단구배를 통해 사고예측모형의 적합도를 상당히 높일 수 있었으나, 곡선반경은 사고건수와 직접적인 관련이 있는 것으로 파악되지 않았다. 이러한 사고예측모형은 기존의 모형과 비교시 여러 설명변수 중 어떤 변수가 모형에 도입되어야 하는지를 결정할 때 분명한 근거를 지니기 때문에 중요한 변수가 누락되거나 혹은 중요하지 않는 변수가 도입될 가능성이 낮아지는 장점을 지니고 있다.

핵심용어: 사고예측모형, 음이항분포, 최우추정법, 적합도

* 정회원 · 한국교통연구원 도로교통연구실 연구위원
 ** 정회원 · 한국교통연구원 도로교통연구실 연구원
 *** 정회원 · Univ. of California Berkeley 박사과정



1. 서론

사고예측모형은 도로에서 발생한 교통사고자료를 통계적으로 모형화한 것으로 종속변수는 과거의 사고건수가 되고 설명변수로는 주로 사고가 일어난 장소의 도로 기하구조 조건, 교통조건, 운영조건 등 도로의 속성자료가 이용된다. 이러한 사고예측모형은 우리나라에서도 이기영·이용택·장명순(1999), 강정규·이성관(2002), 강민욱·도철웅·손봉수(2002), 조완기(2006), 박재범·이승준·강정규·김일환(2007), 윤병조(2007) 등 많은 연구가 수행되어 왔다.

하지만 앞선 연구에서 제시한 사고예측모형들은 모형의 적합도(Goodness-of-Fit)가 그다지 높지 않으며 설명변수의 선택기준이 불분명하여 중요하지 않은 설명변수가 모형에 포함될 가능성도 있다. 또한 시계열(time-series)자료가 아닌 횡단면(cross-section)자료를 기반으로 추정된 모형의 모수(parameter)가 실제 사고건수 변화에 미치는 영향(Accident Modification Factor)을 잘 대변하고 있다고 보기 곤란한 경우도 많다. 현재 계산된 모수 값은 모형식 혹은 변수의 조합에 따라 크게 달라질 가능성이 높기 때문이다.

이러한 사고예측모형의 한계를 넘어설 수 있는 새로운 방안이 최근 Hauer(2004)의 연구를 통해 소개된 바 있다. 본 연구에서는 이 방안을 구체적으로 소개하고 이를 국내 사고자료를 통해 적용할 수 있는지의 여부를 판단하고자 한다.

이를 위해 우선 기존의 전통적 사고예측모형에 대한 문헌고찰을 통해 사고예측모형 정립방법론을 정리하고 이러한 방법론이 지닌 한계에 대해 설명한다. 그 후 Hauer(2004)에서 소개된 새로운 사고예측모형 방법론을 구체적으로 설명한다. 새로운 사고예측모형의 국내 적용가능성은 고속도로 사망사고자료를 통한 모형 구축을 통해 판단하여 본다.

2. 전통적 사고예측모형

2.1 기존 사고예측모형 방법론

기존의 사고예측모형은 회귀분석을 통해 구축되며 일반적으로 사용되는 모형식은 다음과 같다.

$$Y = f(X_1, X_2, \dots, X_n, \beta_1, \beta_2, \dots, \beta_n) \quad (1)$$

여기서, Y 는 사고건수, X_1, X_2, \dots, X_n 은 사고건수에 영향을 주는 변수, $\beta_1, \beta_2, \dots, \beta_n$ 는 각 변수의 계수이다.

이러한 회귀분석모형은 대개 4단계를 따른다. 첫 번째 단계는 변수선택으로 수집된 자료 중에서 종속변수(사고건수 등)를 잘 설명할 수 있는 가능한 설명변수를 선택한다. 두 번째 단계에서는 종속변수와 설명변수 혹은 설명변수 사이에 상관성을 분석한다. 만약 여러 설명변수사이에 상관도가 높아 다중공선성이 존재한다면, 그 중 하나의 변수만 모형식에 포함되어야 한다. 세 번째 단계는 모형추정단계로 종속변수와 설명변수의 관계를 가장 잘 설명할 수 있는 모형식을 선정(model specification)한 후 관련 모수(parameters)를 추정하여 모형식을 완성한다. 모수 추정방법에는 최소자승법(LSE, Least Squared Estimation)과 최우추정법(MLE, Maximum Likelihood Estimation)이 있다. 이 때 모수의 적정성은 t 값을 통해 판단한다. 네 번째 단계는 적합도(Goodness-of-Fit) 판정단계로 추정된 모형식에서 도출된 사고건수가 조사된 사고건수를 얼마나 잘 설명하고 있는지를 검정하는 단계이다. 이 단계에서는 결정계수(R^2)가 1에 가까운지의 여부와 잔차가 특정한 경향성을 띄지 않는지를 파악한다. 만약 결정계수도 낮고 잔차가 경향성을 띄면 모형식을 다시 선정해야 한다. 이러한 과정으로 그림으로 정리하면 그림 1과 같다.

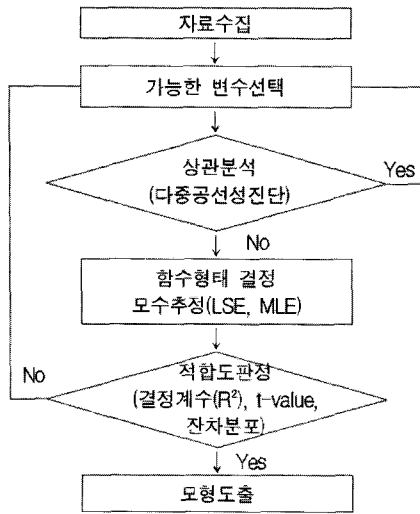


그림 1. 기존 사고예측모형 회귀분석 흐름도

2.2 기존연구사례

회귀분석에 기반한 교통사고 예측모형은 다양하게 연구되어져 왔다. 우선 이기영·이용택·장명순(1999)은 버스사고로 인한 사망자, 중상, 경상 및 이를 종합한 사상자수를 종속변수로 하고 운전자 요인, 도로기하구조, 환경요인 등을 설명변수로 하는 모형을 개발하였다. 이 때 모형식은 포아송회귀식 및 음이항회귀식을 이용하였다.

강정규·이성관(2002)은 호남고속도로를 직선부, 곡선부, 완화곡선부로 분류한 후 중회귀분석을 통해 교통량, 곡선길이, 곡선반경, 종단경사, 편경사 등의 변수를 사용한 사고예측모형을 도출하였다. 도출된 사고예측모형의 결정계수(R^2)는 0.432~0.696으로 나타났다.

강민욱·도철용·손봉수(2002)는 음이항분포가 사고건수를 설명하기에 가장 적합한 확률분포임을 밝히고, 이를 통해 고속도로 곡선구간에서 구간분할법과 최우추정법을 이용한 음이항회귀모형을 개발하였다. 모형에 적용된 변수로는 교통량을 이용한 차량노출계수, 곡선반경, 단위거리 당 편경사 변화값 등이다. 도출된 사고예측모형의 결정계수(R^2)는

0.277~0.526으로 나타났다.

조완기(2006)은 고속도로 엇갈림구간 자료로 선형회귀모형을 구축하고, 설명변수로는 본선 설계속도, 연결로 설계속도, 유출부 곡선반경, 유출부 접속각, 엇갈림 구간길이를 사용하였다. 도출된 사고예측모형의 결정계수(R^2)는 0.25로 나타났다.

박재범·이승준·강정규·김일환(2007)은 감속구간 길이, 감소차로 후 완화곡선길이를 설명변수로 IC 분류부의 사고예측모형을 음이항회귀식으로 도출하였다.

윤병조(2007)는 곡선반경, AADT(Average Annual Daily Traffic, 연평균일교통량)와 사고율과의 관계를 다항선형회귀모형으로 밝혔으며, 도출된 사고예측모형의 결정계수(R^2)는 0.59~0.90으로 나타났다.

이들 선행연구에서는 교통사고의 발생분포가 대체로 포아송 혹은 음이항 분포를 따를 것으로 가정하고 모형식을 추정하고 있으며 모수 추정을 위해서는 최소자승법(LSE) 혹은 최우추정법(MLE)을 사용하고 있는 것으로 나타났다. 교통사고건수를 설명하기 위해 주로 활용된 변수로는 교통량, 곡선반경, 편경사 등이 있다.

2.3 기존 방법론의 한계

기존 방법론의 한계는 다음과 같이 정리할 수 있다. 첫째, 상관분석을 통해 다중공선성이 발생하는 변수들을 제거하는 과정에서 중요한 변수가 누락되거나 혹은 중요하지 않은 변수가 도입될 가능성이 있다. 둘째, 각 설명변수들과 사고율 사이의 인과관계를 구하는 것이 쉽지 않다. 사고예측모형은 주어진 변수들로 사고위험도를 예측할 수 있어야 한다. 다시 말해 어떤 변수(교통량, 곡선반경 등)가 바뀌었을 때 그에 따른 사고위험도의 변화 또한 예측할 수 있어야 한다. 그러나 어떤 변수의 값이 바뀌었을 때 그에 따른 사고위험도의 변화는 그 변수로 인한 변화뿐만 아니라 모형에서 누락된 다른 변수들의 영향 또한 포함



하고 있기 때문에 각 변수들과 사고위험도와의 인과 관계를 규명하는 것이 쉽지 않다. 셋째, 변수별로 함수식을 달리하는 경우가 거의 없어 변수의 특성이 모형식에 적절히 반영되기 어려운 측면이 있다.

3. 새로운 사고예측모형(Hauer의 방법론)

3.1 모형의 일반구조

Hauer(2004)는 전통적 사고예측모형의 한계를 극복할 수 있는 새로운 사고예측모형을 제시하였다. 우선, 도로를 적절한 구간별로 나눈 후 차로수, 편경사와 같이 구간 전체에서 연속적으로 영향을 주는 변수들은 곱의 함수 형태로, 교량 혹은 합류지점의 수와 같이 점의 형태로 영향을 주는 변수들은 합의 함수 형태로 포함시키는 식으로 모형을 구축할 수 있음을 보였다. 이러한 모형의 일반식은 다음과 같다.

$$Y = (\text{조정계수}) \times [(\text{구간길이}) \times (\text{곱의 관계식}) + (\text{합의 관계식})]$$

$$\text{곱의 관계식} = f(\text{교통량}) \times g(\text{길이개폭}) \times \dots$$

$$\text{합의 관계식} = h(\text{교통량, 합류지점수}) + i(\text{교통량, 교량수}) + \dots \quad (2)$$

여기서, Y 는 특정 도로구간의 연간 사고빈도수이다. 조정계수는 연도, 날씨, 인구, 사고기록 등 모형에 포함되어 있지 않은 다른 변수들에 의해 바뀔 수 있으며, 구간길이는 사고빈도수를 구한 도로구간의 길이로서, 교차로의 경우 1로 설정한다.

3.2 변수의 선택

전체 모형식은 종속변수(사고건수)에 가장 영향을 많이 미치는 변수로 우선 구축한 후 다른 변수를 이용하여 모형의 설명력을 높일 수 있는지를 판단하여 다른 변수의 도입 여부를 결정한다. Hauer(2004)는 교통사고에 가장 큰 영향을 미치는 변수인 연평균

일교통량(AADT)으로 최초의 모형을 설정한 후 다른 변수들의 도입여부를 검토한다. 이러한 과정에서 변수별로 함수식을 달리 적용할 수도 있다.

예를 들어 화물차 비율을 새로운 변수로 고려한다면, $V_1 = 0-2\%$, $V_2 = 3-4\%$, ... 와 같이 나눈다. 그 후 각 구간에 속한 모든 자료들에 대하여 기존에 추정된 모형에 의해 사고건수를 추정한 후 구간별로 더하면 각 구간의 예상사고건수가 도출된다. 이를 $N_{\text{predicted}}(V_i)$ 이라 하고, 실제 자료에서 구할 수 있는 관측사고건수를 $N_{\text{recorded}}(V_i)$ 라고 하자. 이 두 값을 비교했을 때 그 관계에 경향성이 발견된다면 변수 V 는 모형에 포함될 필요가 있다. 또한 그 관계에 따라 V 의 함수 형태를 결정할 수도 있다. 단, 모형의 일반식이 변수의 특성에 따라 곱의 관계식과 합의 관계식으로 나뉘기 때문에 두 가지 경우를 따로 고려해야 한다.

3.2.1 Case 1

이전 변수가 곱의 관계식으로 추정된 경우, 각 그룹에 대한 관측사고건수와 예상사고건수의 관계는 다음과 같이 나타낼 수 있다.

$$R(V_i) = \frac{N_{\text{recorded}}(V_i)}{N_{\text{predicted}}(V_i)} \quad (3)$$

$R(V_i)$ 이 1보다 크면 모형이 과소추정된 것이고, 1보다 작으면 과다추정된 것이다.

3.2.2 Case 2

이전 변수가 합의 관계식으로 추정된 경우, 각 그룹에 대한 관측사고건수와 예상사고건수의 관계는 다음과 같이 나타낼 수 있다.

$$R(V_i) = N_{\text{recorded}}(V_i) - N_{\text{predicted}}(V_i) \quad (4)$$

$R(V_i)$ 이 0보다 크면 모형이 과소추정된 것이고, 0보다 작으면 과다추정된 것이다.

위에서 구해진 $R(V_i)$ 과 V_i 를 그래프에 나타낸 후



두 값 사이에 분명한 경향(증가/감소 등)이 발견되지 않으면 변수 V 를 기각하고, 경향이 발견된다면 V 를 모형에 포함시킨다. 함수의 형태는 그래프의 모양을 보고 극대점, 극소점, 변곡점 등을 잘 설명할 수 있도록 결정한다.

3.3 모수 추정

모형식에 새로운 변수가 추가될 때마다 모수는 재추정되어야 한다. 모수의 추정은 최우추정법(maximum likelihood estimation)으로 한다. 최우추정법은 관측된 사고건수의 확률을 최대로 하는 모수 값을 추정하는 방법이며, 이 때 사용하는 확률식을 우도함수(maximum likelihood function)라 한다. 이러한 우도함수를 구축하기 위해서는 주어진 설명변수(속성자료)를 이용한 사고예측모형의 함수식과 추정되어야 할 모수를 가정하여야 한다. 이러한 함수식의 값은 사고발생의 분포를 어떻게 가정하느냐에 따라 달라진다. 한 가지 방법은 사고발생건수가 평균이 Y 인 포아송분포(식(5) 참조)를 따른다고 가정하는 것이다. 이 분포에서는 측정 가능한 속성자료가 같으면 평균도 같아야 한다. 하지만 실제로 그런 경우는 드물다. 따라서 사고발생건수가 평균이 $Y\theta$ 인 포아송분포를 따른다고 가정할 수 있다. 여기서 θ 는 감마분포를 따른다. θ 는 자료에서 같은 속성값을 지니는 구간의 사고평균과 특정구간의 평균 사고건수가 어느 정도 차이나는지를 설명하는 승수(multiplier)이다. 이러한 가정을 따르는 분포를 음이항분포(Negative Binomial Distribution), 혹은 음다항분포(Negative Multinomial Distribution)라고 한다.

음다항우도함수는 다음과 같이 유도된다. i 구간, j 년도의 평균사고빈도를 μ_{ij} 라 하면,

$$\mu_{ij} = Y_{ij}\theta_i \quad (5)$$

여기서, Y_{ij} 는 j 년도에 대해 측정 가능한 속성은 i 와 모두 같지만 그 밖의 속성은 다른 값을 지니는 가상의 모집단에 대한 평균이다.

θ_i 의 분포는 평균이 1이고, 분산이 $\frac{1}{\phi}$ 를 갖는 감마분포라 가정하며 확률밀도함수는 식(6)과 같다.

$$f(\theta) = \frac{\theta^{\phi-1} e^{-\theta\phi} \phi^\phi}{\Gamma(\phi)} \quad (6)$$

년도 1, 2, ..., j , ..., n , 동안 관찰된 i 구간의 사고건수 $a_{i1}, a_{i2}, a_{i3}, \dots, a_{in_i}$ 는 평균 $Y_{ij}\theta_i$ 인 포아송분포를 따른다고 가정하면, θ_i 가 주어진 경우 사고건수는 식(7)로 표현할 수 있다.

$$P(a_{i1}, a_{i2}, a_{i3}, \dots, a_{in_i} | \theta_i) = \prod_{j=1}^{n_i} \frac{(Y_{ij}\theta_i)^{a_{ij}} e^{-Y_{ij}\theta_i}}{a_{ij}!} \quad (7)$$

그러나, θ_i 가 주어지지 않았다면, 분포(식(6))만 알게 되며 이 때 우도함수는 식(8)과 같다.

$$P(a_{i1}, a_{i2}, a_{i3}, \dots, a_{in_i}) = \int_0^\infty \prod_{j=1}^{n_i} \frac{(Y_{ij}\theta_i)^{a_{ij}} e^{-Y_{ij}\theta_i}}{a_{ij}!} * \frac{\theta^{\phi-1} e^{-\theta\phi} \phi^\phi}{\Gamma(\phi)} d\theta_i \quad (8)$$

여기서 기호를 간단히 하면

$$a_i = \sum_{j=1}^{n_i} a_{ij}, Y_i = \sum_{j=1}^{n_i} Y_{ij}, \prod_{j=1}^{n_i} \theta_i^{a_{ij}} = \theta_i^{\sum_{j=1}^{n_i} a_{ij}} = \theta_i^{a_i} \quad (9)$$

$$P(a_{i1}, a_{i2}, a_{i3}, \dots, a_{in_i}) = \int_0^\infty \prod_{j=1}^{n_i} \frac{(Y_{ij}\theta_i)^{a_{ij}} e^{-Y_{ij}\theta_i}}{a_{ij}!} * \frac{\theta^{\phi-1} e^{-\theta\phi} \phi^\phi}{\Gamma(\phi)} d\theta_i$$

$$= \frac{\phi^\phi \left(\prod_{j=1}^{n_i} Y_{ij}^{a_{ij}} \right) \Gamma(a_i + \phi)}{\left[\prod_{j=1}^{n_i} a_{ij}! \right] \Gamma(\phi) (Y_i + \phi)^{a_i + \phi}} \int_0^\infty \theta^{a_i + \phi - 1} e^{-\theta(Y_i + \phi)} d\theta_i \quad (10)$$



이런 함수식은 로그식으로도 나타낼 수 있다. 이렇게 되면, 상수를 포함하는 $a_{ij}!$ 는 생략할 수 있으며, 식(11)처럼 간단히 표현할 수 있다.

$$\ln(l_i) = \phi_i \ln(\phi_i) + \left[\sum_{j=1}^{n_i} a_{ij} \ln(Y_{ij}) \right] + \ln[\Gamma(a_i + \phi_i)] - \ln[\Gamma(\phi_i)] - (a_i + \phi_i) \ln(Y_i + \phi_i) \quad (11)$$

결국, 최대우도모수추정(Maximum likelihood parameter estimates)은 식(12)를 최대화하는 것이다.

$$\ln(l) = \sum_{i=1}^n \ln(l_i) \quad (12)$$

3.4 적합도 검토

CURE(CUmulative REsiduals)는 실제사고건수와 추정사고건수의 차이, 즉 잔차를 변수값의 증가에 따라 누적시킨 것을 의미한다. 모형의 적합도가 좋으면 CURE는 숫자 0 사이를 진동하고, 그렇지 않다면 CURE는 0과 멀리 떨어져 그려진다. 또한 만약 변수의 범위가 CURE 산점도에서 계속 증가 또는 감소한다면, 그 적합은 좋지 않으며, 모형식이 수정되어야 함을 암시한다.

잔차 제곱의 합을 $\hat{\sigma}_{(m)}^2$ 라 할 때 CURE산점도의 95% 신뢰구간 경계선 $\pm 2\sigma^*(n)$ 은 식(13)으로 구할 수 있다.

$$\sigma^2 = \hat{\sigma}_{(n)}^2 \sqrt{1 - \frac{\hat{\sigma}_{(n)}^2}{\hat{\sigma}_{(N)}^2}} \quad (13)$$

4. 고속도로 사고예측모형 개발

4.1 자료구축

4.1.1 자료수집

사고모형 구축을 위하여 2004년 전국 고속도로의 사망사고자료를 수집하였다. 수집범위의 총연장은 2968.1km이며 사망사고건수 총 427건이다. 이 중 조사위치가 정확한 자료 총 400건만을 활용하였다.

4.1.2 교통량 자료

각각의 사고에 대하여 사고 발생지점의 연평균 일교통량(AADT) 및 제한속도 자료를 설명변수로 활용하였다. 연평균 일교통량은 5,869대/일에서 203,143대/일까지의 범위를 가지며 평균값은 58,756대/일이다.

4.1.3 도로기하구조 및 운영 관련 변수

모형 구축에 이용 가능한 지형 관련 변수는 평면곡선반경(R , m), 편경사(i , %), 종단곡선구배(s , %), 평면곡선길이(m), 종단곡선길이(m), 차로수 등이 구축되었다. 제한속도는 80, 100, 110km/h의 세 범위로 나뉘지며 80km/h인 구간에서 총 28건, 100km/h인 구간에서 총 273건, 110km/h인 구간에서 99건의 사망사고가 발생하였다.

4.2 모형 구축과정

본 연구에서는 고속도로의 구간별 사고자료를 구하기 어려워 도로구간별로 사고건수를 분류하는 대신 AADT에 따라 사고건수를 분류하고 이를 설명할 수 있는 모형을 우선 구축하고자 한다. 또한, Hauer(2004)가 합의 관계로 설명한 지점 특성 변수들은 고려하지 않고 있으므로, 모형식은 곱의 관계식으로만 구성된다. 즉, 모형의 일반식은 다음과 같다.

$$Y = (\text{조정계수}) \times (\text{곱의 관계식})$$

$$\text{곱의 관계식} = f(\text{AADT}) \times g(\text{제한속도}) \times \dots$$

(14)

R 값 또한 곱의 관계식만이 존재하므로 Hauer의 모형 중 곱의 관계식에 해당하는 Case 1만 고려하였다. 전체적인 모형의 구축과정을 종합하여 나타내면

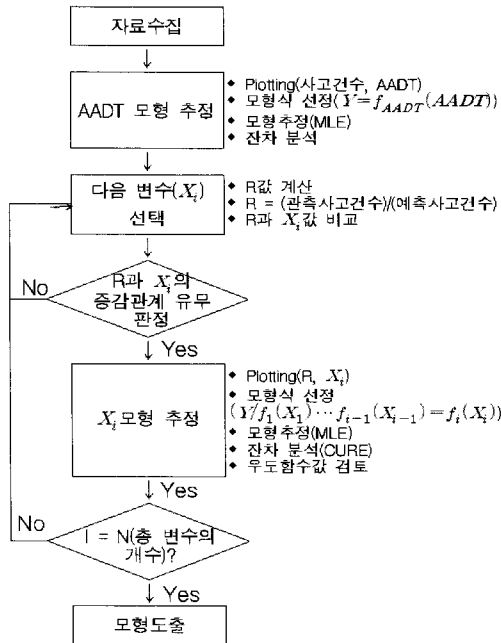


그림 2. 고속도로 사망사고예측모형 흐름도

그림 2와 같다. 사고건수와 가장 관련이 있을 것으로 예상되는 AADT를 첫 번째 변수로 설정하고, AADT와 사고건수를 그래프에 나타내어 AADT에 대한 함수를 선정하였다. 모수 추정에는 최우추정법(Maximum likelihood Method)를 사용하였고, 사고건수가 음이항분포(Negative Binomial Distribution)를 따른다고 가정하였다. 그 후 누적잔차분석(CURE)을 통하여 모형이 적합하다고 생각되면 다음 변수를 고려한다. 이때는 앞서 추정된 모형에 의한 예측사고건수에 대한 관측사고건수의 비인 R 값과 고려 대상 변수의 관계를 살펴 분명한 경향성이 없으면 그 변수를 기각하고, 경향성이 있으면 모형을 추정한다. 즉, R 값과 고려 대상 변수를 그래프로 나타내어 함수를 선정 한 후, 최우추정법에 의해 모형을 재추정한다. 마찬가지로 누적잔차분석을 통하여 모형이 적합하다고 생각되면 다음 변수를 고려한다. 고려 가능한 모든 변수에 대하여 이와 같은 과정을 반복하거나 혹은 모수의 값이 새로운 변수의 도입으로 크게 변화하지 않으면, 혹은 더 이상 우도함수의 값이 커지지 않는다면 모형이 완성된다.

4.3 AADT와 사고

고속도로 사망사고수에 가장 큰 영향을 끼치는 변수로 AADT를 가정하고, AADT와 사고건수를 설명할 수 있는 모형을 추정한다. 그러나 AADT와 사고건수 자체로는 정확한 경향성을 파악하기 어려워 AADT별 사고건수면적을 Plotting하고 이를 기반으로 모형을 추정하여 본다. 그 결과 모형의 관계식은 다음과 같이 추정된다.

$$Y = 15.57(AADT)^{1.12} e^{-0.06(AADT)} \quad (15)$$

여기서, Y 는 사고건수면적을 의미한다. 이 식을 AADT에 대하여 미분하면 실제사고건수를 설명하는 모형을 구할 수 있다. 즉, 실제사고건수는 다음과 같이 나타낼 수 있다.

$$y = 17.44e^{-0.06(AADT)} \times [1.12(AADT)^{0.12} - 0.06(AADT)] \quad (16)$$

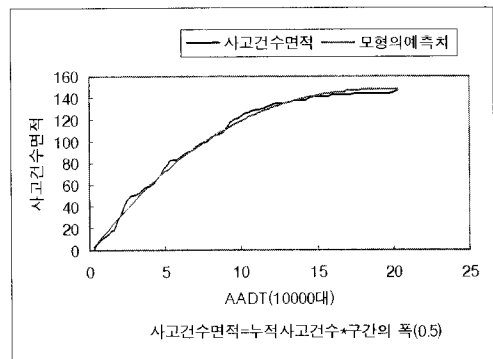


그림 3. 사고건수면적 모형

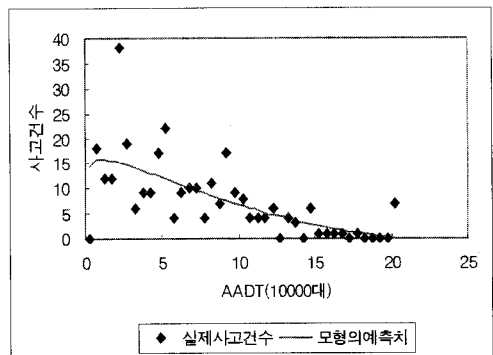


그림 4. 사고건수 모형



이러한 방식을 Integral-Differentiation 방식이라 한다. 그림 3은 사고건수면적과 AADT의 관계를, 그림 4는 사고건수와 AADT의 관계를 나타낸다.

구해진 모형을 바탕으로 잔차를 plotting한 후 CURE 방법을 적용하였다. 그림 5에서 잔차가 특별한 경향성을 띠지 않고, 그림 6에서 누적잔차가 대부분 $\pm 2\sigma$ 안에 위치하고 있는 것으로 보아 모형이 적절한 것으로 판단할 수 있다.

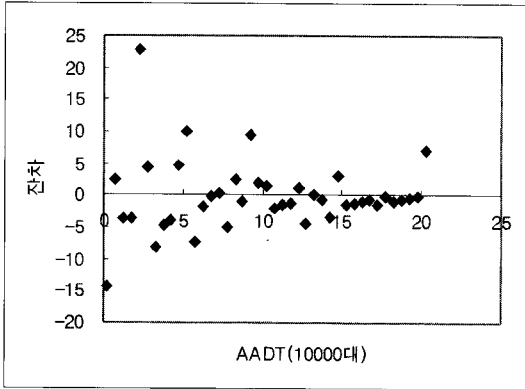


그림 5. AADT 모형의 잔차

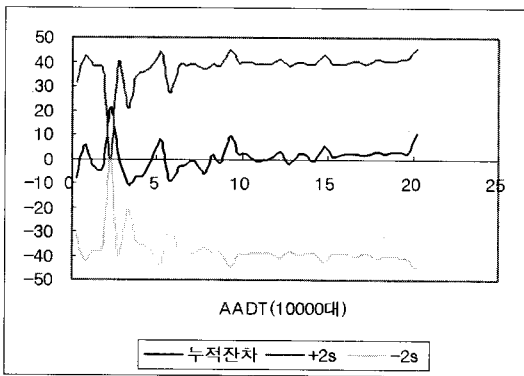


그림 6. $\pm 2\sigma$ 범위와 누적잔차

4.4. AADT, 종단구배와 사고

다음으로 추가할 변수를 선정하기 위하여 앞서 구한 AADT 모형을 대상으로 제한속도, 곡선반경, 종단구배별 R 값을 구하였다. 그림 7, 8, 9의 그래프에서 볼 수 있듯이, 제한속도의 경우 전고속도로 범위에서 80, 100, 110km/h의 세 값만이 존재하기 때문

에 모형으로 나타내기가 어렵고, 곡선반경 역시 특별한 경향성이 있다고 판단하기 애매한 측면이 있다. 따라서 비교적 경향성이 강한 종단구배를 다음에 추가할 변수로 선정하였다.

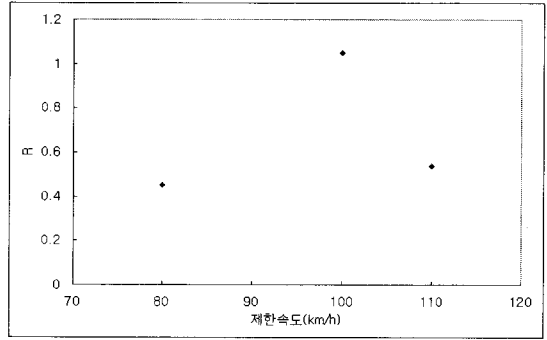


그림 7. 제한속도에 따른 R값

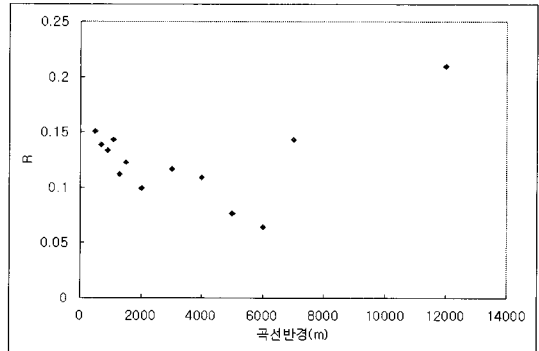


그림 8. 곡선반경에 따른 R값

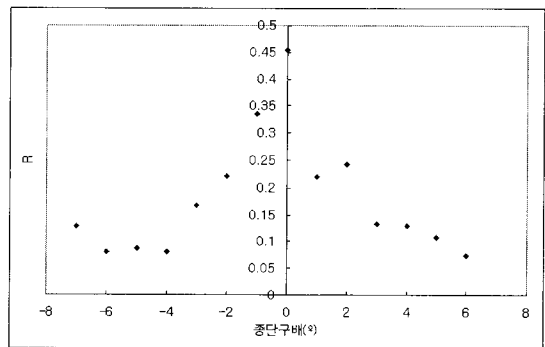


그림 9. 종단구배에 따른 R값

종단경사의 경우 종단경사 값이 0보다 큰 경우인 오르막 경사와 0보다 작은 내리막 경사에서 종단경



사의 영향이 서로 다르기 때문에, 종단경사가 0보다 큰 경우와 0보다 작은 경우로 나누어서 모형을 구축하였다. 구축된 모형식은 다음과 같다.

$$y = \begin{cases} 17.44e^{-0.06(AADT)}[1.12(AADT)^{0.12} - 0.06(AADT)] \\ \left(\frac{0.19}{g} + 0.07\right) & \text{if } g > 0 \\ 17.44e^{-0.06(AADT)}[1.12(AADT)^{0.12} - 0.06(AADT)] \\ (0.02g^2 + 0.18g + 0.47) & \text{if } g \leq 0 \end{cases} \quad (17)$$

여기서, y는 사고건수, g는 종단구배이다.

4.5 AADT, 종단구배, 곡선반경과 사고

다음으로 고려할 수 있는 변수인 곡선반경에 대하여 R값을 plotting 하였다.

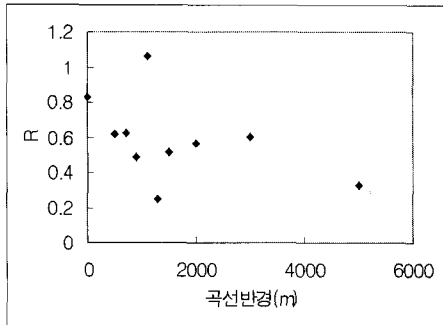


그림 10. 곡선반경에 따른 R값($g > 0$ 인 경우)

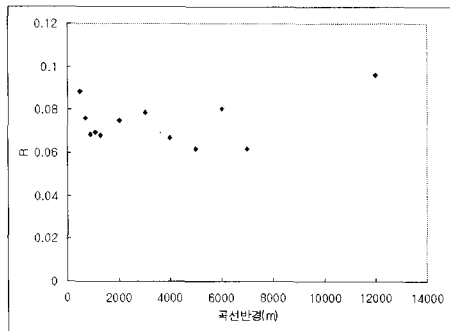


그림 11. 곡선반경에 따른 R값($g \leq 0$ 인 경우)

그래프에서 볼 수 있듯이 곡선반경과 R값은 뚜렷한 경향성을 보이지 않으므로 모형에 포함시키지 않

아도 무방한 것으로 판단할 수 있다. 따라서 최종 모형은 AADT와 종단구배를 변수로 하는 모형으로 선정하였다.

5. 결론 및 향후과제

본 연구에서는 기존의 전통적 사고예측모형이 지닌 한계를 극복하는 Hauer(2004)의 새로운 사고예측모형을 소개하고 이를 국내 고속도로 사망사고 자료를 통해 적용하여 보았다.

새로운 사고예측모형은 우선 AADT와 사고건수의 관계를 그림으로 확인하고 이를 가장 잘 설명할 수 있는 모형식을 선정한 후 모형의 모수를 최우추정법(Maximum Likelihood Method)를 통해 추정한다. 이 때 우도함수는 사고자료가 음이항분포를 따른다는 가정을 기반으로 구축된다. 만약 AADT와 사고건수의 관계가 그림으로 잘 나타나지 않으면 사고건수면적과 AADT의 관계를 살펴보는 Integral-Differentiation 방법론을 활용할 수 있다. AADT를 통한 사고예측모형이 구축된 후에는 새로운 변수를 통해 사고예측모형의 설명력을 높일 수 있는지를 검토한다. 이를 위해 모형을 통한 예측치와 관측치의 비율(R)이 새로운 설명변수와 일정한 관계를 지니는지의 여부를 검토한다. 만약 관계가 있다면 이를 설명할 수 있는 모형식을 선택하고 다시 모수를 추정하게 된다. 그 후 새로운 설명변수의 도입으로 모형의 로그 우도함수(log likelihood function)가 더 커지는지의 여부도 확인한다. 모형의 적합도는 누적오차가 0에 가까워지는지 혹은 이상치(outlier)는 어디서 나타나는지 등을 파악함으로써 판단한다. 이러한 방법으로 고속도로 사망사고 자료를 설명하는 모형을 구축한 결과, 자료의 한계로 많은 설명변수를 검토하지는 못했지만, AADT와 종단구배를 통해 사고예측모형의 적합도를 상당히 높일 수 있었다. 하지만 곡선반경은 사고건수와 직접적인 관련이 있는 것으로 파악되지 않았다.



이러한 사고예측모형은 기존의 모형에 비해 몇 가지 장점을 지닌다. 우선, 여러 설명변수 중 어떤 변수가 모형에 도입되어야 하는지를 결정할 때 분명한 근거를 지니기 때문에 중요한 변수가 누락되거나 혹은 중요하지 않은 변수가 도입될 가능성이 낮아진다. 가령, 본 연구에서는 R 값을 기준으로 종단구배는 설명변수로 활용하였으나 곡선반경은 제외하였다. 또한, 새로운 변수의 도입과정에서 변수의 특성에 맞는 적절한 함수식을 고안해 내기 편리하다. 예를 들어, 종단구배를 모형화할 때 값이 0보다 작은 경우와 큰 경우로 구분하였다. 한편, 새로운 설명변수 도입 기준이 분명하고, 변수 특성에 맞추어 모형식을 개발하기 때문에 각 변수가 사고에 미치는 영향(AMF, Accident Modification Factor)도 비교적 안정적으로 추정할 수 있다. 즉, 새로운 변수가 도입된다고 해서 AMF 값이 크게 바뀌지 않는다.

본 연구에서는 이러한 사고예측모형을 일반적으로 사용되는 spreadsheet 소프트웨어를 통해 수행하였다. 새로운 방법론은 기존의 고급 통계 소프트웨어보다는 오히려 이 방법이 더 적용하기에 편리하다. 복잡한 모형식도 '해 찾기' 기능을 통해 최우추정법에 기반을 둔 모수추정이 가능하기 때문이다.

하지만 새로운 사고예측모형 방법론은 몇 가지 측면에서 문제를 갖고 있다. 우선 모형식이 변수의 도입순서에 영향을 받는다는 한계를 지니고 있다. 물론 우도함수를 더 크게 증가시키는 변수를 먼저 도입해야 한다는 원칙이 있기는 하지만 변수의 도입순서가 사고수정인자(Accident Modification Factor, AMF)에 영향을 미치는 문제가 있다. 또한 과분산계수(overdispersion)의 의미를 아직 명쾌하게 정리하지 못한 아쉬움도 남아있다. 국내의 고속도로사망 사고 자료에 적용한 결과 과분산계수는 235147~ -0.5로 그 범위가 매우 크게 나타났으나 이 값이 모형의 적합도에 미치는 영향에 대해서는 향후 더 연구가 필요할 것으로 보인다. 또한 새로운 방법론은 모형식의 선정이나 적합도의 판단 등의 과정이 한 번으로 끝나는 것이 아니라 변수가 새로 선정될 때마다

반복적으로 이루어지기 때문에 다소 시간이 오래 걸리고 복잡한 과정을 거친다는 한계도 있다.

한편, 본 연구는 자료수집의 한계로 여러 해에 걸친 사고자료를 활용하지 못했다. 또한 AADT를 5,000대/일 단위로 구분하여 사고자료를 정리하였기 때문에 모형의 정확도가 높다고 보기 어려운 문제도 있다. 따라서 향후 연구에서는 정확도를 높이는 측면에서 이 값을 낮추어 분석할 필요가 있다. 또한 특정한 고속도로를 대상으로 구간을 구분하고 이들 구간에서 여러 해에 걸친 사고자료를 이용하여 사고예측모형을 수립할 필요도 있다. 한편, 고속도로 기본구간 이외에 유출입구간에서의 사고예측모형을 별도로 개발하는 것도 중요할 것으로 보인다.

감사의 글

본 연구는 건설교통부 건설기술혁신사업의 연구비지원(05 기반구축 D02)에 의해 수행되었습니다.

참고문헌

- Hauer, E·Council, F M·Mohammedshah, Y (2004), "Safety Models For Urban Four-Lane Undivided Road Segments", *TRR No. 1897*
- Hauer, E(2004), "Statistical Road Safety Modeling", *TRR No. 1897*
- E. Hauer·J. Bamfo(1997), "Two tools for finding what function links the dependent variable to the explanatory variables", *Proceedings of ICTCT 97 Conference, November 5-7 1997, Lund, Sweden*
- Anderson, I. B·Krammes, R. A.(2000), "Speed Reduction as a Surrogate for Accident Experience at Horizontal Curves on Rural Two-Lane Highways", *TRR No. 1701*
- 강정규·이성관(2002), "고속도로 선형조건별 교통사고 위험도 평가모형 개발", *대한교통학회지 제20권 제4호*
- 강민욱·도철용·손봉수(2002), "고속도로 평명선형상 사



고빈도분포 추정을 통한 음이향회귀모형 개발”, 대한
교통학회지 제20권 제7호
박재범·이승준·강정규·김일환(2007), 고속도로 분류부
설계기법 개선 연구, 대한교통학회지, 제25권 제1호
이기영·이용택·장명순(1999), “고속도로 사고모형의 개
발과 적용”, 대한교통학회 제 36회 학술발표회

윤병조(2007), 고속도로 트럼펫형 IC 연결로 유형별 교통
사고 예측모형 개발, 대한토목학회지 제27권, 제1호
조완기(2006), “연속류 엇갈림(weaving)구간에서의 사
고예측 및 형태모형개발에 관한 연구, 한양대학교

접 수 일: 2007. 9. 20
심 사 일: 2007. 10. 9
심사완료일: 2007. 10. 25