

# 비선형 회귀분석, 인공신경망, 구조방정식을 이용한 지방부 4지 신호교차로 교통사고 예측모형 성능 비교 연구

오주택<sup>1\*</sup> · 윤일수<sup>2</sup> · 황정원<sup>1</sup> · 한음<sup>2</sup>

<sup>1</sup> 한국교통대학교 도시공학과, <sup>2</sup> 아주대학교 교통시스템공학과

## A Comparative Study On Accident Prediction Model Using Nonlinear Regression And Artificial Neural Network, Structural Equation for Rural 4-Legged Intersection

OH, Ju Taek<sup>1\*</sup> · YUN, Ilsoo<sup>2</sup> · HWANG, Jeong Won<sup>1</sup> · HAN, Eum<sup>2</sup>

<sup>1</sup> Department of Urban Engineering, Korea National University of Transportation, Chungbuk 380-702, Korea

<sup>2</sup> Department of Transportation Systems Engineering, Ajou University, Seoul 443-749, Korea

### Abstract

For the evaluation of roadway safety, diverse methods, including before-after studies, simple comparison using historic traffic accident data, methods based on experts' opinion or literature, have been applied. Especially, many research efforts have developed traffic accident prediction models in order to identify critical elements causing accidents and evaluate the level of safety. A traffic accident prediction model must secure predictability and transferability. By acquiring the predictability, the model can increase the accuracy in predicting the frequency of accidents qualitatively and quantitatively. By guaranteeing the transferability, the model can be used for other locations with acceptable accuracy. To this end, traffic accident prediction models using non-linear regression, artificial neural network, and structural equation were developed in this study. The predictability and transferability of three models were compared using a model development data set collected from 90 signalized intersections and a model validation data set from other 33 signalized intersections based on mean absolute deviation and mean squared prediction error. As a result of the comparison using the model development data set, the artificial neural network showed the highest predictability. However, the non-linear regression model was found out to be most appropriate in the comparison using the model validation data set. Conclusively, the artificial neural network has a strong ability in representing the relationship between the frequency of traffic accidents and traffic and road design elements. However, the predictability of the artificial neural network significantly decreased when the artificial neural network was applied to a new data which was not used in the model developing.

도로의 안전성을 평가하기 위한 방법으로서 교통사고 자료를 이용하는 방법, 사전-사후평가를 통한 방법 또는 전문가 의견이나 기존 문헌을 통한 방법 등 다양한 방법들이 존재한다. 특히, 교차로 교통 안전성을 평가하는 경우 많은 연구들이 교통사고예측모형 개발을 통하여 교통사고와 관련한 원인과 안전성을 평가하고 있다. 교통사고예측모형 개발에 있어서 모형의 예측력과 전용성을 확보하는 것이 중요하다. 즉, 예측력을 확보함으로써 교통사고 건수나 교통 안전성 판단의 지표를 예측하는데 오차를 줄일 수 있고, 전용성을 확보함으로써 개발된 모형이 다른 지점이나 구간에 적용하더라도 문제없이 적용될 수 있는 대표성을 가질 수 있다. 따라서 본 연구에서는 교통사고예측모형 개발에 주로 사용되는 회귀모형과 인공신경망, 구조방정식을 이용하여 교통사고예측모형을 각각 개발하였으며, 개발된 모형의 예측력과 전용성을 평균절대오차와 평균제곱예측오차를 기준으로 확인하였다. 90개소 신호교차로의 모형개발자료를 이용하여 세 가지 방법으로 교통사고예측모형을 개발 후 개발데이터를 통해 예측력을 비교한 결과 인공신경망이 가장 높은 예측력을 보였다. 또한 모형의 전용성 검증을 위하여 별도로 수집한 33개소 신호교차로의 모형검증자료를 이용하여 개발된 모형을 검증한 결과 비선형 회귀모형이 가장 적합한 것으로 나타났다. 모형개발 과정에서 가장 높은 예측력을 보인 인공신경망의 경우 다른 대상지에서 수집된 모형검증 자료를 적용하였을 때 예측력에 큰 변화를 보여 전용성이 떨어진 것으로 분석되었다.

### Keywords

artificial neural network, nonlinear regression, intersections, structural equation, traffic accident  
인공신경망, 비선형 회귀모형, 교차로, 구조방정식, 사고분석

\* : Corresponding Author

jutaek@ut.ac.kr, Phone: +82-43-841-5158, Fax: +82-43-841-5410

Received 5 November 2013, Accepted 3 March 2014

## 서론

### 1. 연구의 목적 및 수행방법

기존의 교통안전 관련 논문들에서 살펴 볼 수 있듯이, 교통사고의 요인 중 제어가 불가피한 차량 요인, 인적 요인을 고려하는 것보다 도로 기하구조와 같은 제어 가능한 변수를 고려하여 교통안전제고를 해야 한다. 이를 통해 실질적 안전을 위한 도로설계 과정에서 대책을 마련할 수 있다. 또한 교통 운전자, 도로 설계자 등 전문가의 입장에서는 인적 요인보다는 상대적으로 실제적 측정이 가능하며 자료수집이 용이한 도로조건, 교통조건, 환경조건에 관한 요인에 초점을 두고 접근하고 있다.

도로의 안전성을 평가하기 위한 방법은 교통사고 자료를 이용한 단순비교, 사전-사후평가를 통한 방법, 전문가나 기존 문헌을 통한 방법 등 다양한 방법들이 있지만 도시부 및 지방부 4지 신호교차로에서의 교통사고에 관한 많은 선행 연구들이 사고예측모형을 통하여 교통사고와 관련이 있는 다양한 원인분석을 하고 있다(Kim et al, 2008; Oh et al, 2005; Park et al, 2008; Hong and Doh, 2002; Park et al, 2008; Ha et al, 2001). 이들 연구의 대부분은 비선형회귀모형들을 이용하여 분석을 하고 있다. 물론 이들 비선형 Econometrics 분석기법들이 교통사고의 임의성과 비선형성을 분석하는데 가장 기본이 되는 통계적 기법들이기는 하지만, 도로에서 발생하는 교통사고의 원인을 분석하는 측면에서는 통계적 방법들이 사고의 예측력을 높이는 데 중점을 두고, 이를 위해 소수의 유효한 설명변수들만을 모형식에 포함시키는 경향이 있기 때문에 사고발생의 설명변수들을 규명하는데 구조적인 한계가 발생할 수 있다. 이를 해결하기 위해 비선형 회귀분석 외 예측력이 높은 인공신경망, 여러 변수를 포괄적으로 설명가능한 구조방정식 등 다양한 방법들을 통하여 사고예측모형을 개발하고 있다. 하지만 인공신경망의 경우 산출과정이 명확하지 않으며 구조방정식은 인과관계 추론의 열악성과 자의성이 발생할 수 있다는 한계점을 가지고도 있다.

교통사고예측모형 개발에 있어, 교통사고예측모형 개발에 있어서 모형의 예측력(predictability)과 전용성(transferability)을 확보하는 것이 중요하다. 즉, 예측력을 확보함으로써 교통사고 건수나 교통 안전성 판단의 지표를 예측하는데 오차를 줄일 수 있고, 전용성을 확보함으로써 개발된 모형이 다른 지점이나 구간에 적용하

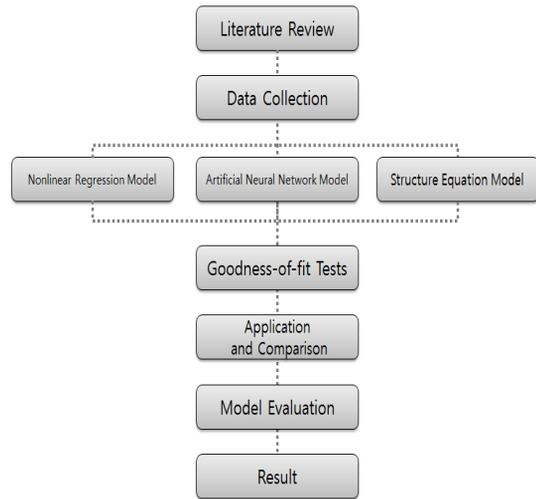


Figure 1. Research procedure flow

라도 문제없이 적용될 수 있는 대표성을 가질 수 있어야 한다. 이 중 교통사고에 대한 전용성은 모형을 개발한 자료에만 국한되지 않고 유사한 대상지에 개발된 사고모형을 적용하더라도 모형의 예측력이 변하지 않고 충분히 설명할 수 있는 모형이어야 한다는 것이다. 본 연구에서는 전용성을 확보한 사고예측모형에 적합한 방법을 확인하기 위하여 먼저, 모형개발에 사용될 자료를 90개소 신호교차로를 통해 사고 예측 모형 연구에 주를 이루는 비선형 회귀분석모형(regression Model)과 인공신경망 모형(Artificial Neural Network Model), 그리고 구조방정식 모형(Structural Equation Modeling)을 이용하여 각 방법별 모형을 개발하였다. 다음으로, 모형개발에 사용된 자료를 통한 각 방법별 모형 적합성을 검증하는 것과 더불어 유사한 조건의 33개소 교차로를 별도로 수집된 자료를 통해 개발된 모형이 33개의 다른 대상지 교통사고자료의 실제사고건수를 예측하는데도 그 적합성을 유지하는지 확인하였다. 연구의 수행절차는 Figure 1과 같다.

### 2. 관련 문헌 고찰

교통사고예측모형 개발에 앞서 선행연구 고찰을 통하여 모형개발에 필요한 자료의 종류와 방법에 대하여 고찰하였다. Kim et al.(2010)은 지방부 신호·비신호 3지 교차로의 안전성을 향상시키기 위하여 247개 지점을 대상으로 사고예측모형을 개발하였다. 기존 모형들이 특정지역의 자료만 이용함으로 개발된 모형들이 범용적

로 활용되지 못하고 있다. 이를 해결하기 위하여 대도시 외곽지역, 중소도시 외곽지역, 그리고 평지 및 구릉지 지역의 다양한 조건의 교차로 사고자료를 조사하여 사고예측모형을 개발하였다. 그 결과, 신호와 비신호 교차로에 공통적으로 영향을 미치는 변수는 주도로 좌회전 전용차로 유무와 ADT인 것으로 분석되었다.

Park et al.(2008)은 도시부 신호교차로의 안전성을 향상시키기 위하여 총 171개 지점을 대상으로 교통사고 예측모형을 개발하였다. 교통사고예측모형으로서 포아송 회귀모형을 사용하였고, 사고심각도 모형으로서 프로빗 모형을 이용하였다. 개발된 모형을 우도비와 MPB(Mean Prediction Bias), MAD(Mean Absolute Deviation)를 이용하여 검증하였다. 분석 결과, 사고건수에는 교통량 및 차로수, 버스정류장 및 딜레마구간 길이 등이 영향을 미치는 것으로 나타났으며 사고 심각도에는 차량량 비율, 우회전비율, 교차로 지형, 차로 당 평균폭 등이 영향을 미치는 것으로 분석되었다.

Kang et al.(2011)은 기존의 교통사고예측모형 개발방법의 제한된 변수선정 등의 문제점을 개선하고 자료 조사 수집과정에서 발생하는 자료의 불확실성을 최소화하는 방법으로서 퍼지이론과 인공신경망을 제안하였다. 이를 통해 도시부 신호교차로 인공신경망 교통사고예측모형을 개발하였다. 퍼지이론과의 성능평가를 위하여 실제 데이터를 이용하여 통계적인 검증이 이루어졌으며 중회귀모형, 포아송 회귀모형과 비교하였다. 퍼지이론과 인공신경망이론이 중회귀모형, 포아송 회귀모형에 비해 사고건수에 대한 설명력이 높은 것으로 나타났다.

Choi et al.(2004)는 신호교차로의 교통사고 심각도를 예측하는 방법으로 인공신경망을 사용하였다. 인공신경망 개발을 위한 자료는 수집과정에서 장기간을 요구한다는 점과 사고발생 기간동안 교차로운영, 기하구조 등이 수시로 변할 수 있다는 점을 들어 3-5일 정도의 자료만을 이용하였다. 상관분석을 통해 교통사고 심각도와 독립변수간의 상관성을 분석하고 추가로 다중회귀 사고 심각도 모형과 인공신경망의 예측 성능을 비교·분석하였다. 그 결과, 인공신경망을 이용한 교통사고예측모형의 예측력이 우수한 것으로 나타났다.

Oh et al.(2006)는 보행자-차량 간 충돌사고의 경우에서 보행자 사망여부를 확률적으로 예측하는 모형을 개발하였다. 베이지안 신경망과 로지스틱 회귀분석 기법을 이용해 두 모형을 개발하고 두 모형의 성능을 비교·분석하였다. 그 결과, 평균 CCR(Correct Classification

Rate)이 베이지안 신경망 모형은 68.23%, 로지스틱 회귀분석 모형은 63.24%로 베이지안 신경망 모형이 정확한 CCR을 나타내는 것으로 분석되었고 이는 모형의 학습(training) 및 테스트에 사용된 자료에 영향을 받았기 때문이라 판단하고 있다.

Lee et al.(2008)는 구조방정식을 이용한 고속도로의 교통사고 심각도에 작용하는 요인들을 분석하였다. 크게 도로 요인 및 운전자 요인, 환경 요인을 외생 잠재변수로 설정하고 교통사고 심각도를 내생 잠재변수로 설정하여 모형을 추정하였다. 모형 추정결과 도로요인과 환경요인은 사고 심각도와 음의 관계에 나타났고, 운전자 요인은 사고 심각도와 양의 관계에 있는 것으로 나타났다.

Jo(2008)는 보행자의 교통사고 심각도에 영향을 주는 요인을 구조방정식을 이용하여 분석하였다. 그의 연구에서는 세 가지 주요 요인을 사용하였으나, 차량 요인만이 영향을 주는 것으로 나타났다. 특히 차량 요인 인 범퍼높이, 차량무게, 충돌속도 중에서 충돌속도가 가장 큰 영향을 미치는 것으로 나타났다.

Kim et al.(2011)의 연구에서는 구조방정식을 활용하여 단속류 시설의 교통사고 유형에 따라 유발요인을 분석하였다. 사고의 요인을 차대차 사고와 차대사람 사고로 나누었으며 유형별 심각도와 외생 변수들 간의 관계를 추정하였다. 그 결과 차대차 사고에서는 도로 요인이, 차대사람 사고에서는 환경 요인이 크게 영향을 미치는 것으로 나타났다.

그 외 사고예측모형 개발에 관련한 해외 논문을 살펴보면, McCoy and Malone(1989)은 방향별 좌회전 차로 설치에 따른 교통사고 감소에 대한 효과분석을 하였다. 그 결과, 추돌사고, 접촉사고, 그리고 좌회전사고가 6-77%까지 감소하였으며 주도로 일방향에서는 대략 20% 내외, 주도로 양방향에서는 대략 40% 내외의 교통사고 감소가 나타난 것으로 분석하였다.

Harwood et al.(2002)는 방향별 좌회전 차로 설치에 따른 교통사고 감소에 대한 효과분석을 연구하였다. 좌회전 차로를 주도로 한 쪽에만 설치하였을 경우 대략 20%정도가 감소하였고 주도로 양방향 설치한 경우 대략 40%의 감소를 보이는 것으로 분석되었다.

Vogt(1999)는 3년간의 자료를 이용하여 총 205개소를 대상으로 모형을 개발하였다. 그 결과, 지방부 4지 교차로에서 주도로의 좌회전 차로의 설치가 교통사고를 38% 줄이는 것으로 분석하였다.

Bauer and Harwood(1997)는 미국 워싱턴주를 대상으로 기하구조와의 관계를 모형화하였다. 2년간 발생한 교통사고를 기반으로 포아송 회귀모형 및 음이항 회귀모형을 이용하여 사고등급 별, 지역 별로 교통사고예측모형을 개발하였으며 그 결과 교통량, 연결로의 형식, 가감속차로 길이, 연결로 길이, 유·출입여부가 연결로 교통사고에 영향을 주는 것으로 나타났다.

선행연구들을 살펴본 바, 교통사고 예측모형을 개발하기 위해 대부분 연구들이 비선형 회귀분석을 이용하여 모형을 개발하였으며 인공신경망과, 구조방정식 등 다양한 방법이 사용된 것을 확인하였다. 하지만 일부 연구에서 수집한 원자료를 통해 모형의 개발 후 원자료의 교통사고빈도나 심각도와의 예측력을 비교하는 것에 그쳤다. 이를 위해 본 연구에서는 다른 지점이나 구간에 적용하더라도 모형 간 비교우위를 절대적인 측면이 아닌 상대적으로 두고 좋은 모형 개발방법을 판단하는 것을 목적으로 한다. 이는 모형개발에 사용된 개발데이터와 별도로 수집된 검증데이터를 이용하여 모형의 전용성(transferability)를 확인하기 위함이다.

## 모형개발 이론 고찰

### 1. 비선형 회귀모형

신호교차로에서의 교통사고건수  $\hat{Y}_i$ 가 포아송 분포를 따른다는 가정하에  $i$ 번째 신호교차로에서  $m$ 개의 변수에 의해 교통사고가 발생할 확률의 일반식은 Table 1의 (1)과 같다. 다시 말해서, 포아송 회귀모형의 종속변수는 일정 기간 동안 주어진 교통사고의 발생 횟수를 나타낸 것이며, 모형계수인  $\beta_j$ 는 Maximum-Likelihood 평가를 통해 계산된 것이다.

교통사고는 불연속적이며 임의적, 산발적으로 발생하는 특징을 가지고 있기 때문에 이러한 특징을 반영할 수 있는 모형으로 포아송 회귀모형이 대표적으로 사용된다. 과분산은 조사되지 않은 다른 성질의 구간이 다양한 결과를 나타내기 때문이며 분산이 포아송 분포보다 더 크거나 더 작은 경우는 포아송 회귀모형이 적합하지 않음을 나타낸다. 따라서 과분산 문제를 해결하기 위해 과분산 계수가 1보다 작으면 음이항 회귀모형이 적합하며, 과분산 계수가 0에 가까우면 평균과 분산이 같아져 포아송 회귀모형이 적합하다. 음이항 회귀분석은 분산을 나타내는 과분산이 포함된 2차식의 형태이므로 Table 1의

**Table 1. Accident Frequency Prediction Methods**

	models	
(1)	$\hat{Y}_i = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m)$ $\alpha = 1$ (equidispersion) $x_i = \text{independ variable}$  where, $\beta_j$ : regression coefficient $\alpha$ : dispersion parameter	Poisson
(2)	$\hat{Y}_i = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon_i)$ $\alpha < 1$ (overdispersion)  where, $\beta_j$ : regression coefficient $\alpha$ : dispersion parameter	Negative Binomial
(3)	$\hat{Y}_i = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon_i)$ $\alpha > 1$ (underdispersion)  where, $\beta_j$ : regression coefficient $\alpha$ : dispersion parameter	Gamma

식(2)과 같은 형태를 가진다. 이와는 다르게 과분산 계수( $\alpha$ )가 1보다 크면 감마분포가 적합한 결과를 얻을 수 있다. Table 1은 포아송 회귀모형식, 음이항 회귀모형식, 감마 회귀모형식을 나타낸다.

모형의 실제사고건수와 예측건수를 비교하기 위해서 일반적으로 예측평균오차(Mean Prediction Bias,MPB)와 절대평균오차(Mean Absolute Deviation,MAD)가 사용된다. MPB는 모형을 만들기 위하여 종속변수로 사용된 자료에 대해서 모형에 의한 결과 값의 치우침 정도를 판단할 수 있는 기준을 제공해준다. 이 방법에 의한 결과 값이 작을수록 모형의 예측 값은 정확한 것을 의미한다. 또한 MAD는 모형의 예측 값이 평균적으로 얼마만큼의 오차가 포함되었는지를 판단할 수 있는 척도를 제공해준다. 이 방법이 MPB와 다른 점은 각 수치의 음과 양의 차이로 인해 상쇄되지 않는다는 점이다. 결과 값이 0에 가까울수록 실제로 관측된 자료에 부합되는 결과를 나타냄을 의미한다.

### 2. 인공신경망 모형

인공신경망은 인간의 뇌의 수많은 신경세포가 복잡하게 연결되어 있는 형태를 착안하여 모형화한 방법으로 인간의 인식과정이나 신경생태를 수학적으로 모형화한 것이다. 인공신경망은 어떤 문제에 대한 훌륭한 해결방안을 제시하고 기존의 방법으로 해결하지 못했던 문제들을 해결 가능성을 보여주고 있는 방법이다. 최근에는 패턴인식, 의사결정 등에 많이 이용되고 있다. 인공신경망은 변수의 연결강도(Weight)를 통해서 결과값과의 영

항을 결정하게 되는데 이러한 구조를 간단하게 수학적식으로 표현하면 식(1)과 같다.

$$NET = x_1w_1 + x_2w_2 + \dots + x_nw_n \quad (1)$$

$$= \sum_{i=1}^n x_iw_i$$

$x_1$  : 입력변수  
 $w_1$  : 연결강도(weight)

인공신경망은 다른 모형들과 가장 다른 점은 학습이라는 점이다. 변수들 간의 관계를 의미하는 연결강도는 바로 학습의 반복을 통한 조정으로 그 강도가 결정된다. 학습 방법에도 가장 일반적으로 지도학습 방법에 속하는 오류 역전파(Backpropagation) 알고리즘을 사용하고 있는데 이는 일반화된 델타 규칙이라고도 한다. 오류 역전파 알고리즘은 다중 퍼셉트론을 학습시키기 위한 체계적인 방법으로 모형의 비선형성을 해결하여 출력값과 목표값의 차이를 최소화한다.

인공신경망의 구조는 입력층(Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer)으로 구성된다. 입력층은 외부로부터 데이터를 받아들이는 계층이며 출력층은 입력층의 데이터를 처리하여 나오게 되는 결과를 도출하는 계층이다. 은닉층은 입력층과 출력층 두 계층의 사이에서 보이지 않는 과정(Lee, 2009)으로 각 층은 여러 개의 노드로 구성되어 있으며 각 층의 노드는 서로 다른 연결강도를 가지고 연결되어 있다. 이러한 과정에서 활성화 함수를 거쳐 결과를 도출하게 것이다. 일반적인 오류역전파 인공신경망의 구조는 Figure 2와 같다.

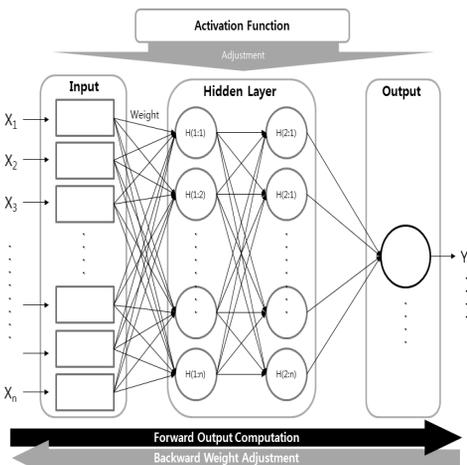


Figure 2. Artificial neural network

인공신경망을 통한 예측값의 검증은 위한 방법은 다양하다. MPB, MAD,  $R^2$ , RMSE, %RMSE 등이 있는데 MPB, MAD, RMSE, %RMSE 검증방법은 모두 인공신경망을 통한 출력값과 실제값과의 비교를 통해 오차를 표현하는 방법으로 값이 0에 가까울수록 오차가 작음을 나타내고  $R^2$ 은 설명력을 나타내는 값으로 그 %가 높을수록 설명력이 높다고 할 수 있다.

### 3. 구조방정식 모형

많은 변수를 사용하여 요약된 정보는 얻어내고자 할 때 요인분석을 사용한다. 요인분석(Factor Analysis)은 여러 변수들 사이의 상관관계를 바탕으로 하여 정보의 손실을 최소화하면서 변수의 개수보다 적은 수의 요인(factor)으로 자료 변동을 설명하는 다변량 분석기법이다. 요인으로 나누어진 변수들을 구조방정식을 이용하여 모형의 해석 및 적합성을 알아내는 것이다.

구조방정식은 SEM(Structural Equation Model)이라고 하는데 구성개념 간의 이론적인 인과관계와 상관성의 측정지표를 통한 경험적 인과관계를 분석할 수 있도록 개발된 통계기법이다. 다시 말해 확인요인분석을 통해서 측정오차가 없는 잠재요인을 발견하고 회귀분석으로 잠재요인 간을 연결하는 방법으로 이해하면 된다. 연구자는 이를 통해 다중변수관계를 포괄적으로 측정하고 탐색적인 분석에서 확인적인 분석까지 할 수 있다(Kim, 2010).

모형과 연구가설 개발 다음 절차로 자료를 수집 후 분석이 진행된다. KMO(Kaiser-Meyer-Olkin)와 Bartlett's

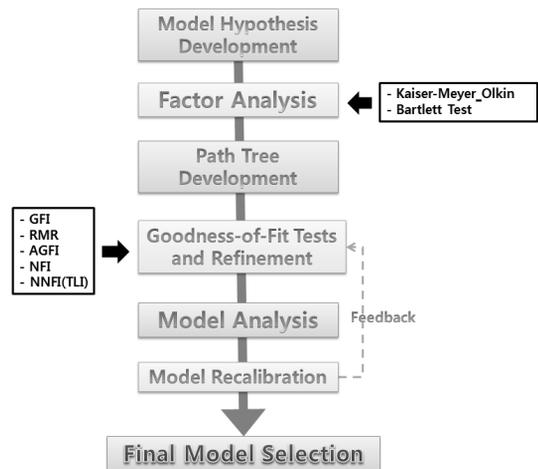


Figure 3. Structural equation modeling procedure flow

Test는 수집된 자료가 요인분석에 적합한지 여부를 판단하기 위한 통계적인 과정이라고 할 수 있으며, KMO의 MSA(Measure of Sampling Adequacy)는 0.5이상이면 요인 분석을 계속 진행 할 수 있다는 것을 나타내며, Bartlett-검정은 0.05이하이면 요인 분석을 계속 진행 할 수 있다는 것을 나타낸다. 또한 연구 데이터의 회전후 요인부하량(Rotation Component Matrix)(a)를 이용하여 연구모형을 개발할 수 있으며 그에 따른 연구가설을 설정할 수 있다. 요인분석의 과정을 거친 후 분석된 요인으로 자료파일을 불러온 상태에서 구조방정식모형 분석 패키지를 이용하여 경로모형을 구축한다.

다음, 모형의 분석 및 인정(Model Identification) 평가 단계에서는 모형이 가치 있는 모형으로 받아들여질지 여부를 평가하는 것을 말한다. 다시 말해서 모형을 수집된 자료에 적용해 볼 수 있는지를 심사하는 과정이라고 할 수 있다. 모형의 적합도 지수에 의해서 모형의 적합도가 만족하는 모형을 선정 한 후 연구자는 각 추정치에 대한 해석을 하고 결론을 논리적으로 도출한다.

인정평가를 통해 낮은 적합도가 명시될 경우 모델의 적합도를 개선하는 설명 가능한 또는 타당한 근거가 바로 수정지수(Modification Indices,MI)이다. 모형의 수정지수의 공분산 자료를 검토하여 연구모형을 수정할 수 있지만, 공분산이 존재할 개연성이 있을 때 처리해야 하며, 연결하는 것이 비논리적이거나 연결하지 않도록 하여야 한다. 앞에서 기술한 단계를 거쳐 연구자는 최종 모형을 선정할 수 있다.

앞서 언급한 모형의 적합성 인정평가방법은 학자마다 다소 의견차이는 있지만, 절대적합지수는 카이제곱(Chi-Square,  $\chi^2$ ), 적합도지수(Goodness of Fit Index, GFI), 수정된 적합도지수(Adjusted GFI, AGFI), 평균제곱잔차제곱근(Root Mean-Square Residual, RMR)이 있으며, 상대적합지수는 비표준 적합지수(Non-Nonmed Fit Index,NNFI) 등이 있다.

### 교통사고 자료의 수집 및 선정

본 연구에서는 2007년도 전라북도 전 지역과 공주시 외곽 지방부 국도변 4지 신호교차로를 중심으로 교통량 및 도로시설, 도로주변 환경 등의 현황자료와 물피사고를 포함한 경상, 중상, 사망 등 모든 사고를 포함한 교통사고자료를 통하여 모형 개발을 위한 90개 교차로 데이터와 모형의 전이성 검증을 위한 33개소 교차로 데이터

를 수집하였다. 총 123개 교차로 데이터는 도로의 안전성의 기준이 될 교통사고빈도를 포함하여 주·부도로 차로수, 주·부도로 진출입로 수, 주·부도로 좌회전 전용 차로수, 주·부도로 우회전 전용 차로수, 주·부도로 횡단보도 유무, 주·부도로 중앙분리대 유무, 주·부도로 속도, 주·부도로 가로등 유무, 주·부도로 교통섬 유무,

Table 2. Main explanatory variables

	Variables		Min.	Max.	Avg.
Frequency	Accident Frequency for 4_lagged intersections		0	11	1.83
Lane	Number of lanes for each approach	Major	1	4	2.85
		Minor	1	4	1.90
Entrances	Number of driveway at intersection	Major	0	9	1.71
		Minor	0	7	1.59
Lane_Left	Number of exclusive left_turn lanes	Major	0	2	1.24
		Minor	0	2	0.67
Lane_Right	Number of exclusive right_turn lanes	Major	0	2	0.82
		Minor	0	2	0.63
Crosswalk	Presence of crosswalk (none=0, yes=1)	Major	0	1	0.95
		Minor	0	1	0.85
Median	Presence of median (none=0, yes=1)	Major	0	1	0.60
		Minor	0	1	0.50
Speed	Speed for each approach	Major	50	80	73.0
		Minor	30	80	65.0
Lighting	Presence of intersection lighting (none=0, yes=1)	Major	0	1	0.88
		Minor	0	1	0.76
Traffic Islands	Presence of traffic islands (none=0, yes=1)	Major	0	1	0.21
		Minor	0	1	0.49
Bus Stops	Presence of bus stop (none=0, yes=1)	Major	0	1	0.37
		Minor	0	1	0.13
Bus_bay	Presence of bus_bay (none=0, yes=1)	Major	0	1	0.09
		Minor	0	1	0.02
Channelization	Presence of Channelization at intersection (none=0, yes=1)	Major	0	1	0.40
		Minor	0	1	0.39
Residence	Indicators for residential land use around intersection		0	1	0.12
Commerce	Indicators for commercial land use around intersection		0	1	0.36
Mixed-Use Building	Indicators for Mixed land use around intersection		0	1	0.29
Industry	Indicators for industrial land use around intersection		0	1	0.02
Width of Shoulder	Width of Shoulder	Major	0	2.20	0.79
		Minor	0	2.10	0.74
ADT	LN[Major·Minor ADT](vph)	Major	7.93	10.4	9.06
		Minor	7.10	10.1	8.27
HV	LN[Major·Minor Heavy Vehicle](vph)	Major	7.02	9.50	8.11
		Minor	5.31	9.00	7.49

주·부도로 버스정류장 유무, 주·부도로 버스베이 유무, 주·부도로 도류화 유무, 교차로 주변의 토지이용 현황, 주·부도로 길어깨 폭, 주·부도로 교통량, 주·부도로 중차량 교통량 등 총 35개 변수로 구성되어 있다. 지점별 교통사고빈도를 종속변수로 이용하였으며, 변수의 수 단위의 차이로 인해 가중치의 크기가 상이하게 나오는 것을 우려해 교통량과 중차량 교통량은 자연로그를 사용하여 변수의 자리수를 맞춰 분석에 이용하였다.

## 교통사고예측모형의 개발

### 1. 비선형 회귀분석 모형 개발

지방부 4지 신호교차로의 교통사고예측모형을 개발하기 위하여 통계패키지인 LIMDEP 8.0을 이용하여 분석하였다. 선정되는 독립변수에 따라 종속변수에 영향이 달라지기 때문에 모형의 설명력이 큰 차이가 날 수 있다. 상관성이 과도하게 떨어질 경우 변수의 신뢰성이 낮아지기 때문에 독립변수의 중복을 방지하고 각 변수의 독립성을 위해서 상관관계를 사전분석하고, Type I Error와 Type II Error를 고려하여 신뢰수준 90%( $\alpha=0.1$ )에서 모형개발을 하였다. 또한 다중공선성 분석을 통하여 설명변수의 독립성을 검증하였으며 분산팽창계수(Variance Inflation Factor, VIF)의 기준인 10이상보다 매우 낮은 1.042-1.585로 분석되었고 공차한계(Tolerance) 기준인 0.1이하보다 높은 0.631-0.960으로 분석되어 이상이 없는 것으로 분석되었다.

다양한 영향의 교통사고를 하나의 모형으로는 교통사고에 영향을 주는 주요변수만을 찾기가 어렵기 때문에 다수의 모형을 개발하였으며 개발한 모형 중 가장 적합한 모형을 선정하였다. 특히 모형은 교통량변수를 포함한 모형과 포함하지 않은 모형을 MPB와 MAD 기준을 분석한 결과 모형 1과 모형 2를 선정하였다. 교통량 변수를 포함하지 않은 모형은 거의 0에 가까워 높은 설명력

Table 3. Multicollinearity analysis

Variables	VIF	Tolerance
Minor Lane( $X_1$ )	0.635	1.575
Major Lane_Left( $X_2$ )	0.899	1.113
Major Crosswalk( $X_3$ )	0.944	1.060
Major Speed( $X_4$ )	0.960	1.042
Minor Lighting( $X_5$ )	0.865	1.156
Major Bus_bay( $X_6$ )	0.931	1.074
ADT( $X_7$ ) (Major+Minor)	0.631	1.585

은 나타내었으며 교통량 변수가 포함되었을 때가 오차가 큰 것으로 나타났다. 하지만 이는 결정권자의 해석에 따라 다르게 적용될 수 있는 부분이라 판단된다.

모형 1과 모형 2를 통해서 도로의 기하구조와 관련된 다수의 변수 중 부도로 차로수, 주도로 좌회전 전용 차로수, 주도로 횡단보도 유무, 주도로 속도, 부도로 가로등 유무, 주도로 버스베이 유무, 주도로·부도로 교통량이 교통사고에 영향을 미치는 것으로 분석되었다. 모형식 결정과 관련해서는 과분산(Overdispersion)검증을 수행한 결과, 평균과 분산의 차이가 0과 가까워 음이항 모형보다 포아송 모형이 더 적합한 것으로 나타났다. 또한 사고빈도가 0인 지점이 많은 경우 영과잉(Zero Inflation) 포아송 모형이 적합할 수 있지만 신호교차로의 경우 상층의 주요지점으로 교통량이 적은 지방부더라도 사고빈도가 0인 지점이 매우 적어 영과잉 포아송의 경우에도 적합하지 않는 것으로 분석되었다. Table 4는 지방부 4지 신호교차로 수집 자료를 통해 나온 포아송 회귀모형의 결과이다.

모형 1을 통해서 부도로의 차로수와 주도로의 속도, 주도로의 버스베이가 사고를 증가시키는 변수로 분석되었으며, 주도로 좌회전 전용 차로수, 주도로 횡단보도, 부도로 가로등이 교통사고를 감소시키는 변수로 분석되었다. 분석에 의한 모형 1을 식으로 정리하면 식(2)와 같다.

$$\hat{Y}_i = \exp(-0.803 + 0.341X_1 - 0.277X_2 - 0.724X_3 + 0.028X_4 - 0.470X_5 + 0.571X_6) \quad (2)$$

Table 4. Poisson regression results

Variables	Model1	Model2
Minor Lane( $X_1$ )	Coef.	0.341
	p-value	0.000
Major Lane_Left( $X_2$ )	Coef.	-0.277
	p-value	0.012
Major Crosswalk( $X_3$ )	Coef.	-0.724
	p-value	0.005
Major Speed( $X_4$ )	Coef.	0.028
	p-value	0.004
Minor Lighting( $X_5$ )	Coef.	-0.470
	p-value	0.005
Major Bus_bay( $X_6$ )	Coef.	0.571
	p-value	0.020
ADT( $X_7$ ) (Major+Minor)	Coef.	0.359
	p-value	0.022
Goodness of Test	MPB	-0.00000021
	MAD	0.95

교통사고에 영향을 미치는 중요한 변수 중 하나로 교통량과 관련한 변수를 배제할 수 없다. 모형 2는 예측력을 위한 목적이라기보다 설명력을 우선시 하는 모형이다. 사고예측모형을 통해서 안전성 평가 시 결정권자에 따라 좋은 예측력을 가진 모형을 사용할 수 있지만 교통량과 같은 중요한 변수가 포함된 모형을 사용하고자 하는 경우도 존재할 것이다. 본 연구에서도 모형 1을 통해서 예측력이 가장 좋은 모형은 개발하였지만 모형의 설명력을 더 높이기 위하여 모형 2 개발을 통해 모형의 오차를 감안하고도 필요에 따라 사용할 수 있도록 하였다. 모형 2에서는 모형 1과 같이 부도로의 차로수와 주도로의 속도, 주도로의 버스베이가 사고를 증가시키는 변수로 분석되었으며, 주도로 좌회전 전용 차로수, 주도로 횡단보도, 부도로 가로등이 교통사고를 감소시키는 변수로 분석되었고, 주도로·부도로의 교통량이 많을수록 사고를 증가시키는 것으로 분석되었다. 분석에 의한 모형 2를 식으로 정리하면 식(3)와 같다.

$$\hat{Y}_i = \exp(-3.486 - 0.311X_2 - 0.081X_3 + 0.028X_4 - 0.467X_5 + 0.597X_6 + 0.359X_7) \quad (3)$$

## 2. 인공지능망 개발

자료 수집을 통하여 취득한 변수 중에서 인공지능망 개발을 위하여 사고빈도를 종속변수로 인공지능망을 개발하였다. 설명변수는 본 연구가 방법별 모형의 예측력을 비교하기 위한 것을 목적으로 하고 있기 때문에 교통사고에 영향을 미치는 변수가 다른 방법의 모형과 변수가 유사하여야 설명하는데 의미가 있다고 판단하였다. 따라서 설명변수는 비선형회귀분석을 통하여 개발된 유의미하다고 판단된 모든 변수를 사용하여 인공지능망을 개발하였으며 사용된 변수는 부도로 차로수, 주도로 좌회전 전용 차로수, 주도로 우회전 전용 차로수, 주도로 횡단보도 유무, 주도로 속도, 부도로 가로등 유무, 주도로 버스베이 유무, 주도로·부도로의 교통량, 총 8개 변수이다. 사용된 8개 변수 간에 정확한 선형관계가 존재할 경우 좋지 않은 예측결과를 나타낼 수 있는데 이를 위하여 다중공선성을 분석한 결과에서 VIF의 기준인 10 이상보다 매우 낮은 1.064-1.589로 분석되었고 공차한계 기준인 0.1이하보다 높은 0.629-0.953으로 분석되어 다중공선성의 문제가 없음을 확인하였다.

인공지능망은 구조가 은닉층의 수가 많아질수록 많은

Table 5. Scenario of neural network modeling

Iteration	Learning Rates	The Number of Hidden Node
1,000,000	0.0005	16
		18
		20
		24
	0.001	16
		18
		20
		24

Table 6. Verification of neural network modeling

Scenario No.	Learning Rates	The Number of Hidden Node	MAD	MAPE
1	0.0005	16	0.0474	0.0306
2	0.0005	18	0.0472	0.0334
3	0.0005	20	0.0488	0.0349
4	0.0005	24	0.0437	0.0299
5	0.001	16	0.0498	0.0340
6	0.001	18	0.0332	0.0228
7	0.001	20	0.0442	0.0313
8	0.001	24	0.0408	0.0285

시간이 투자되어야 한다. Lee(2002)는 한 개의 은닉층으로도 함수를 학습하여 원하는 정도의 정확도에 도달할 수 있다는 것을 밝혔다. 따라서 분석시간의 단축을 위하여 본 연구에서는 1개의 은닉층을 사용하였다. 또한 인공지능망의 경우 은닉노드 수와 학습률, 반복횟수의 설정에 따라 결과가 달라질 수 있다. 기초분석을 통하여 은닉 노드 수는 16-24개, 학습률은 0.0005-0.001 수준, 학습반복횟수는 1,000,000번에서 최적 인공지능망을 찾기 위하여 다양한 학습률과 은닉노드 수를 갖는 인공지능망 시나리오를 Table 5과 같이 구축하여 실험하였다.

그 결과, 인공지능망의 특징상 예측과정을 정확하게 알 수 없으나 Table 6을 통해서 각 시나리오별 설명력을 살펴보았다. 총 8개의 시나리오에서 인공지능망은 대체적으로 큰 오차가 발생하지 않고 높은 예측력을 보여 주었으며, MAD와 절대 백분율 오차의 평균(Mean Absolute Percentage Error, MAPE)을 기준으로 가장 오차율이 적은 6번 시나리오(학습률:0.001, 은닉노드 수:18개)의 인공지능망을 최종 모형으로 선정하였다. 오차율은 인공지능망 특성상 교통사고건수를 음수(-)로 예측하지만 교통사고의 특성상 음수는 존재하지 않기 때문에 0건으로 치환하여 계산하였다.

### 3. 구조방정식 모형 개발

데이터 수집을 통해 입력한 자료를 분석을 위해 SPSS 20과 AMOS를 이용하였다. 먼저 구조방정식 모형을 구성하기에 앞서 수집된 변수 중 서로 구성되는 요인이 무엇인지 알아보기 위하여 요인분석(Factor Analysis)을 실시하였다.

자료가 요인분석에 적합할 경우 KMO and Bartlett's Test는 0.5 이상이면, Bartlett-검정 분석 결과가 0.05 이하이면 요인분석을 실시할 수 있다. 본 연구에서는 KMO의 MSA는 0.701로 분석되어  $MSA > \alpha = 0.500$ 를 만족하므로 요인분석을 계속 진행할 수 있다는 결론이 도출되었으며, Bartlett-검정 또한  $Sig.(p) = 0.000$ 으로  $Sig.(p) < \alpha = 0.05$ 이므로 단위행렬이 아니라는 충분한 증거를 보여주고 있다. 따라서 본 연구의 데이터는 KMO의 MSA와 Bartlett-검정 모두를 만족하므로 요인분석을 계속하여 진행 할 수 있다는 결과가 도출되었다.

다음으로 연구모형 및 연구가설을 개발하기 위하여 회진 후 요인부하량 분석을 시행하였다. 분석결과 주도로의 차로수, 중차량, 교통량이 하나의 요인, 부도로의 차로수, 중차량, 교통량이 하나의 요인을 이루었고 주·부도로 좌회전 전용 차로수, 주·부도로 우회전 전용 차로수가, 주·부도로의 제한속도가 각각의 요인으로 구성되었다. 분석결과를 통해 각 요인이 사고발생정도에 영향을 미칠 것이라는 연구모형 가설을 설정할 수 있으며 수정지수의 공분산데이터를 이용하여 총 2번의 수정을 통해 본 연구의 최종 모형이 설정되었으며, 그 결과는 Figure 4와 같다. 구조방정식의 경우 모형의 적합성 결정하기 위한 기준이 매우 다양하다. 따라서 학자마다 어떤 기준을 사용할지 의견이 다르지만 구조방정식 모형 기준 중 다른 적합도 기준보다 가장 중요하다고 판단되는 모형의 간명성 표현과 표본의 크기에 영향을 받지 않는

fit index, CFI), 모집단원소간 평균오차(Root Mean Square Error of Approximation, RMSEA)를 활용하여 모형을 평가하는 것이 바람직한 방법임을 제시하고 있다(Hong, 2000).

CFI는 0.9 이상, NNFI(TLI)는 0.9 이상, RMSEA는 0.05이하일 때 가장 바람직한 수준이라고 판단하고 있는데 본 연구의 구조방정식 모형은 CFI는 0.965, NNFI(TLI)는 0.899, RMSEA는 0.085로 모든 기준이 가장 매우 좋은 수준은 아니지만 CFI가 기준보다 크고 나머지 적합도 기준이 바람직한 수준으로 분석되어 본 모형을 최종모형으로 선정하였다. 선정된 최종 모형에 대한 Standardized estimates 분석 결과의 잠재요인을 수식화하여 표현하면 Table 8과 같이 나타낼 수 있다.

구조방정식 모형 결과, 총 4개의 Factor로 구분되었으며 Factor 1은 0.587, Factor 2는 0.555, Factor 3은 -0.318, Factor 4는 0.069로 사고발생에 영향을 주는 것으로 분석되었다. Factor 1은 주도로와 연관되는 변수로써 주도로 차로수, 주도로 교통량, 주도로 중차

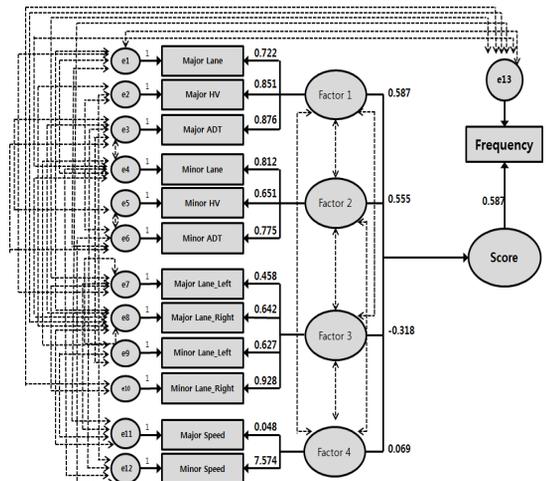


Figure 4. Standardized estimates for the final model

Table 7. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.701
Bartlett's Test of Sphericity	Approx. Chi-Square	519.804
	df	66
	Sig.	0.000

KMO Measure Criteria(Marija J. Norusis, 1993), 0.90-1.00 : Marvelous; 0.80-0.89 : Meritorious; 0.70-0.79 : Middling; 0.60-0.69 : Mediocre; 0.50-0.59 : Miserable; 0.00-0.49 : Unacceptable.

Table 8. Weights for the latent factors

	formula
Score	$0.587*Factor1 + 0.555*Factor2 + (-0.318)*Factor3 + 0.069*Factor4$
Factor1	$Major Lane*0.722 + Major HV*0.851 + Major ADT*0.876$
Factor2	$Minor Lane*0.812 + Minor HV*0.651 + Minor ADT*0.775$
Factor3	$Major Lane\_Left*0.458 + Minor Lane\_Left*0.642 + Major Lane\_Right*0.627 + Minor Lane\_Right*0.928$
Factor4	$Major Speed*0.048 + Minor Speed*7.574$

량 교통량으로 구성되었다. 이와 유사하게 Factor 2는 부도로와 연관하는 변수가 구성되었는데 구성된 변수는 부도로 차로수, 부도로 교통량, 부도로 중차량 교통량이다. Factor 1과 Factor 2를 통해서 주도로나 부도로의 차로수가 많으면 교통사고확률이 증가함으로 나타났다. 차로수가 많다는 것은 교통량이 많기 때문에 계획이나 신설에 의해 차로수가 많아진 것이므로 이는 교통량과 직접적으로 관련이 된다. 또한 중차량 교통량이 많을수록 주변 교통에 영향을 미쳐 사고가 증가한 것으로 판단된다. Factor 1과 Factor 2의 변수인 차로수와 교통량, 중차량 교통량은 노출 관련 변수로 주도도로와 부도로, 즉 교차로 교통사고에 노출(Exposure)이 중요한 영향을 미치는 변수임을 보여준다.

Factor 3은 -0.318로 차량의 회전과 관련된 변수들로 주도도로 좌회전 차로수, 주도도로 우회전 차로수, 부도로 좌회전 차로수, 부도로 우회전 차로수가 선정되었다. 이는 사고가 많은 교차로의 경우 적정한 좌·우회전 차로수에 관한 대책으로 사고를 줄일 수 있음을 의미한다. 회전을 위한 차량에게 명확하게 노선을 제공함으로 직진 교통량과의 상충을 최대한 줄이기 때문이라고 판단된다. 마지막으로 Factor 4는 0.069로 주도도로와 부도로의 제한속도로 구성되었다. 제한속도는 주도도로 제한속도가 0.048, 부도로 제한속도가 7.574로 가중치의 차이가 크게 나타났는데 주도도로의 속도가 높을 때에 비해서 부도로의 높은 속도로 인해 교차로 사고 발생확률이 높은 것을 나타낸다. 속도 역시 차량의 노출과 관련한 변수들과 깊은 관계를 가지고 있는 변수로 속도가 높을수록 운전자가 돌발 상황 발생시 제동거리가 증가하여 사고 발생확률이 높고 대응을 위한 시간이 짧기 때문에 사고의 확률이 증가할 수밖에 없다. 이를 해결하기 위한 과도한 차량 제한속도로 정해진 교차로를 수정 등 적절한 대책을 요구한다.

### 모형의 적합성 비교

대부분 많은 연구에서 사고예측모형의 적합성을 판단할 때 수집된 자료를 통해 사고모형을 개발하고 개발된 모형의 예측력과 사고모형개발에 사용된 자료의 실제사고건수와 비교로 적합성을 검증하였다. 따라서 본 연구에서는 모형개발에 사용된 교차로 교통사고자료에 대한 예측력보다는 유사한 특징을 가진 다른 장소의 33개 교차로 교통사고자료를 별도로 수집하여 각 방법별 개발모형을 적용하였을 때도 유사한 예측력을 나타내는

지 분석하였다.

비선형 회귀모형, 인공신경망 모형, 구조방정식 모형의 전용성을 검증하기 위하여 절대 평균 편차(MAD)와 평균 제곱 예측 오차(Mean Squared Prediction Error, MSPE)를 기준으로 적합성을 검증하였다. MSPE는 모형의 예측값과 관측된 자료의 차이를 절대값이 아닌 제곱을 사용하지만 각 수치의 음과 양의 차이로 인한 상쇄를 없앤다는 점에서 MAD와 유사하며 0에 가까울수록 실제의 관측된 자료와 부합되는 결과를 나타냄을 의미한다. 최초 비선형 모형의 설명력을 나타내는데 사용되는 MPB, MAD MSPE를 사용하였으나 MPB를 통하여 적합성 검증한 결과, 모형의 MPB가 모두 소수점 형태로 나타나 우위를 판단하기 어려워 적합성 검증기준에서 제외하였다. 분석한 내용은 Table 9와 같다.

각 방법별 모형 개발을 통해 개발사이트에 적용한 결과, 모형개발에 사용한 데이터를 통한 예측력은 인공신경망 모형, 비선형 회귀모형, 구조방정식 모형 순으로 좋았지만 검증을 위해 따로 수집된 대상지 자료에 적용 결과는 비선형 회귀모형, 구조방정식 모형, 인공신경망 모형 순으로 나타났다.

개발사이트에서는 인공신경망이 MAD: 0.03, MSPE: 0.01로 가장 좋은 예측력을 보여주었고 다음으로 비선형 회귀모형의 모형1이 MAD: 0.95, MSPE: 1.80으로 분석되었으며 구조방정식이 MAD: 1.33, MSPE: 2.80로 낮은 설명력을 보여주었다. 또한 ADT변수를 포함하고 있는 비선형 회귀모형의 모형2는 MAD: 2.08, MSPE: 6.24로 가장 낮은 설명력을 보였는데 ADT가 교통사고 설명에 중요한 변수지만 사고를 예측하는데 예측력을 저해하는 것으로 분석되었다. 세 가지 모형 중 특히 인공신경망은 <인공신경망 개발> 부분에서도 확인할 수 있듯이 인공신경망을 통해 개발된 다수의 모형 모두 개발사이트의 실제사고건수와 예측치를 비교하였을 때 실제사고건수와 매우 근사하게 분석됨을 확인할 수 있었다. 이로써 인공신경망의 예측력이 매우 높은 모형임을 확인하였다.

Table 9. Comparison

	Regression Model		Neural Network Model		Structural Equation Model			
	Model 1		Model 2					
	MAD	MSPE	MAD	MSPE	MAD	MSPE		
development	0.95	1.80	2.08	6.24	0.03	0.01	1.33	2.80
verification	0.94	1.88	2.28	7.60	1.78	4.92	1.48	2.75
variation(%)	-1.05	4.47	9.41	21.9	5276	52964	10.93	-1.99

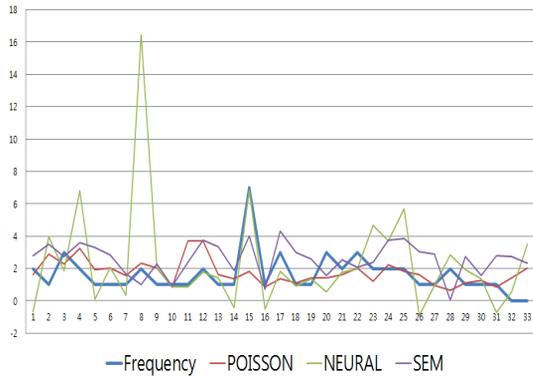


Figure 5. Graph of prediction result for each methods

하지만 별도로 수집된 33개의 검증사이트에 적용한 결과에서는 다른 결과를 확인할 수 있었다. 개발사이트에서 가장 예측력이 좋은 인공신경망 모형이 MAD: 1.78, MSPE: 4.92로 비교적 높은 수치는 아니지만 개발사이트에 적용하였을 때보다 MAD는 약 52배, MSPE는 약 530배 높아지는 것을 확인 할 수 있다. 이는 개발된 모형을 다른 사이트에 적용했을 때 큰 오차가 발생함을 보여주는 대목이다. 사고예측모형의 전용성 면에서는 적합하지 않음을 나타내었다. 반면 비선형 회귀 모형의 경우 모형1과 모형2 모두 다른 모형보다 전용성이 뛰어난 것으로 분석되었다. 개발 1의 경우 개발데이터에 적용했을 때의 예측력을 검증데이터에 적용한 후에도 변화가 거의 없었으며 모형 2의 경우도 개발데이터에서 ADT를 포함하여 낮은 예측력을 보였음에도 불구하고 전용성 면에서는 MAD는 약 10%, MSPE는 20% 정도의 오차 증가를 보여주어 다른 모형에 비해 대표성, 전용성 면에서는 대체로 양호한 수준을 나타내었다. 경우에 따라 결정권자가 오차를 감안하고 교통량 변수를 포함한 설명력 높은 사고예측모형을 이용하여 안전성 분석에 필요하다고 판단된다면 비선형 회귀모형의 모형 2를 사용하는 것이 적합할 것이라 판단된다.

각 방법으로 개발된 모형을 검증사이트에 적용하여 도출된 예측치를 그래프로 표현하면 Figure 5와 같다. 비선형 회귀모형은 예측력이 높은 모형 1을 사용하였다. Figure 5에서도 확인 가능하듯이 인공신경망 모형의 예측치가 실제사고건수를 크게 벗어나며 세 모형 중에서 비선형 회귀모형의 예측치가 실제사고건수와 가장 가까운 예측력을 보여주었다.

세 가지 모형 중 특히 인공신경망의 경우, 전용성을 확인하기 위하여 검증사이트에 적용하였을 때 변화량이

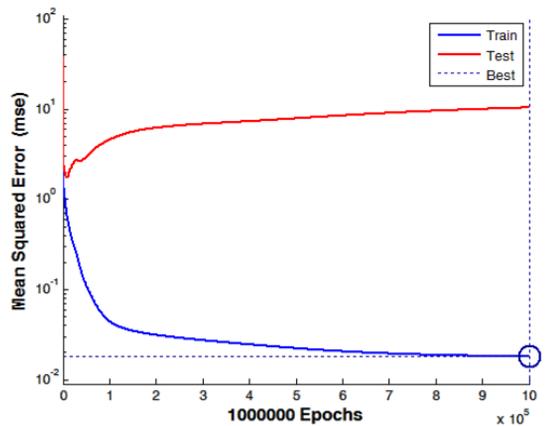


Figure 6. MSE of neural network model

가장 큰 것으로 분석되었는데 이는 Figure 6에서 확인할 수 있다. 인공신경망의 경우 입력된 정보에 대하여 가중치를 실제값에 맞도록 반복적인 학습 및 훈련을 통해 오차를 줄인다. 이는 모형을 개발하는데 사용되는 자료에만 적합한 가중치를 적용하여 계산되기 때문에 그 외의 자료에 개발모형을 적용하였을 때는 오히려 좋지 않은 예측력을 나타내는 것으로 판단된다. Figure 6는 인공신경망의 반복적인 학습의 특징을 잘 나타내는 그림이라 할 수 있다. 파란색 선은 Training 값으로 반복적 학습, 훈련을 통해 나타나는 결과이고 빨간색 선은 Test 값으로 반복적 학습 없이 개발된 모형에 적용하였을 때 나타나는 결과이다. Training 값은 학습반복횟수가 많아질수록 실제 결과값에 근접하는 반면, Test 값과의 오차는 조금씩 더 커지는 것을 확인할 수 있다. 인공신경망에 의해 개발된 모형의 경우 다른 대상지에 적용하여 사고를 예측하기에는 전용성이 낮은 모형이라 판단되며 자료가 수집된 해당 교차로에 대한 분석에 적합하다고 판단된다.

결과적으로 사고예측모형을 통해 여러 교차로의 안전성을 분석하는 역할을 하는데에는 비선형 회귀모형을 통한 방법이 가장 적합하다고 분석된 것이다. 하지만 목적에 따라 주요 변수만이 아닌 다수의 변수를 통해 예측모형을 개발하고자 하는 경우에는 구조방정식 모형이 적합하다고 판단되며, 해당 교차로의 사후 환경 개선을 통한 안전성을 예측할 경우는 인공신경망 모형이 적합하다고 판단된다.

## 결론 및 향후 연구과제

사고예측모형을 통한 신호교차로 안전성 평가를 위하여 사고예측모형이 가져야 할 요소는 교통사고에 영향을

미치는 주요한 변수를 포함하는 설명력과 교통사고자료의 실제사고를 예측하는 정도를 나타내는 예측력이라 할 수 있다. 연구자의 목적에 따라 두 가지 요소 중에서 중요순위가 달라질 수 있겠지만 설명력과 예측력은 불가분한 관계이다. 사고예측모형의 예측력에 대하여 논하기 위해 많은 선행연구들에서는 다양한 방법론을 통한 모형 개발과 검증이 이루어졌다. 하지만 기존 연구에서는 대부분 수집된 자료를 통해 사고예측모형을 개발 후 모형 개발에 사용한 자료로 개발된 모형을 검증하는 것이 대부분이다. 사고예측모형의 경우 개발된 모형이 다른 지점이나 구간에 적용하더라도 합당한 예측력이 유지되어야만 좋은 모형이라 판단할 수 있을 것이다.

따라서 본 연구에서는 사고예측모형 개발에 주로 사용되고 있는 비선형 회귀모형과 인공신경망 모형, 구조방정식 모형의 개발을 통해 모형 간에 비교우위를 절대적인 측면보다도 상대적인 측면에 두고 결정해 보았다. 우위를 비교하기에 앞서 모형개발에 사용될 90개소의 신호교차로 교통사고자료인 개발데이터와 별도로 그와 유사한 특성을 가지고 있는 33개소 신호교차로, 검증데이터로 구분하여 자료를 수집하였다.

먼저, 비선형 회귀모형 개발결과 Type I Error와 Type II Error를 고려하여 신뢰수준 90%( $\alpha=0.1$ )에서 부도로 차로수, 주도로 좌회전 전용 차로수, 주도로 횡단보도 유무, 주도로 속도, 부도로 가로등 유무, 주도로 버스베이 유무, 주도로·부도로 교통량이 4지 신호교차로 교통사고에 영향을 미치는 것으로 분석되었다. 과분산 분석을 통해서 분산과 평균이 0에 가까워 음이항 모형보다 Poisson 모형이 적합한 것으로 나타났다. 인공신경망의 경우 변수선정을 위해 다수 개발된 비선형 회귀모형에서 선정된 변수를 종합하여 다중공선성 분석을 거쳐 분석에 문제가 없다고 판단되어 부도로 차로수, 주도로 좌회전 전용 차로수, 주도로 우회전 전용 차로수, 주도로 횡단보도 유무, 주도로 속도, 부도로 가로등 유무, 주도로 버스베이 유무, 주도로·부도로의 교통량으로 분석을 하였다. 인공신경망에 영향이 큰 학습률과, 은닉 노드수를 달리한 8개의 시나리오 중 MAD와 MAPE를 기준으로 가장 적합한 모형을 선정하였으며 학습률 0.001, 은닉 노드수 18개, 반복횟수 1,000,000번으로 설정한 시나리오 6번이 선정되었다. 구조방정식은 요인분석을 통하여 4개의 요인으로 나뉘었으며 분석결과 주도로의 차로수, 중차량, 교통량이 하나의 요인, 부도로의 차로수, 중차량, 교통량이 하나의

요인을 이루었고 주·부도로 좌회전 전용 차로수, 주·부도로 우회전 전용 차로수가, 주·부도로의 제한속도가 각각의 요인으로 구성되었다. 구성된 요인을 통하여 구조방정식 모형을 구축하였고 NNFI(TLI), CFI, RMSEA를 기준으로 적합성을 검증하여 적합한 모형임을 확인하였다.

개발된 모형을 모형 개발에 이용한 개발데이터에 적용하여 MAD와 MSPE를 기준으로 확인한 결과, 인공신경망 모형, 비선형 회귀모형, 구조방정식 모형 순으로 예측력이 뛰어나게 나타났다. 하지만 개발된 모형의 대표성, 전용성을 확인하기 위해 별도로 수집된 33개소 신호교차로 자료에 적용한 결과는 비선형 회귀모형, 구조방정식 모형, 인공신경망 순으로 예측력이 뛰어나게 나타났다. 특히 인공신경망의 경우에는 개발데이터에 적용하였을 때보다 MAD는 약 52배, MSPE는 530배 차이가 나타나는 결과를 보였다. 높은 예측력을 가지게 하는 인공신경망의 반복적인 학습방법이 다른 대상지에 적용하므로 오히려 약점으로 작용하는 것을 확인하였다. 본 연구를 통해 각 모형개발의 방법이 연구자가 자신의 기호에 맞추어 무작위로 선정하여 개발할 것이 아니라 목적에 따라 그에 적합한 모형을 사용해야 함을 확인하였다.

향후 연구과제로는 지방부 4지 신호교차로 뿐만 아니라 지방부의 다양한 도로 유형에서 일어나는 사고예측모형에 대한 전용성을 확인할 필요성이 있으며 목적에 따라 연구자가 선정해야 할 개발방법에 대하여 구체적인 접근이 필요할 것이라 판단된다. 그리고 지방부 뿐만 아니라 도시부의 다양한 도로에 대한 연구도 필요하다고 판단되며 앞으로 사고발생에 영향을 주는 인적, 기후적 요인들까지 고려하는 연구가 이루어져야 한다고 판단된다. 또한 본 연구는 전체 4지 신호교차로를 대상으로 하여 어떠한 방법론이 적절한지 살펴본 연구로 더 나아가 교차로는 각 교차로마다 다양한 특성을 가지고 있으므로 각 교차로의 특성에 따른 적절한 분석 방법을 제시하므로 신호교차로의 안전성에 대한 기여를 높이도록 해야 할 것이다.

## ACKNOWLEDGEMENT

This work was supported by National Research Foundation of Korea Grant funded by the Korean Government(Ministry of Education, Science and Technology(NRF-2012R1A1A4A01010102)).

## REFERENCES

- Bauer K. M., Harwood D. W. (1997), Statistical Models of At-grade Intersection Accidents, Federal Highway Administration, FHWA-RD-96-125.
- Choi J. W., Kim S. H., Cho J. H., Kim W. C. (2004), A Study to Predict the Traffic Accident Severity Level Applying Neural Network at the Signalized Intersections, *J. Korean Soc. Transp.*, 22(3), Korean Society of Transportation, 127-135.
- Ha O. G. (2005), Development of Accident Prediction Models and Accident Injury Severity for Rural Signalized Intersections, Hanyang University graduate school, A master Dissertation.
- Ha T. J., Kang J. K., Park J. J. (2001), Development and Application of Traffic Accident Forecasting Model for Signalized Intersections (Four-Legged Signalized Intersections in Kwang-Ju), *J. Korean Soc. Transp.*, 19(6), Korean Society of Transportation, 207-218.
- Harwood D. W., Bauer K. M., Potts I. B., Torbic D. J., Richard K. R., Kohlman Rabbani E. R. et al. (2002), Safety Effectiveness of intersection Left and Right-Turn Lanes, Federal Highway Administration, FHWA-RD-02-089.
- Hong J. Y., Doh C. W. (2002), Development of a Traffic Accident Prediction Model and Determination of the Risk Level at Signalized Intersection, *J. Korean Soc. Transp.*, 20(7), Korean Society of Transportation, 155-166.
- Hong S. H. (2000), The Criteria for Selecting Appropriate Fit Indices in Structural Equation Modeling and Their Rationales, *The Korean Journal of Clinical Psychological*, 19(1), 161-177.
- Jo J. I. (2008), Analysis of factors affecting pedestrian injury severity, Hanyang University graduate school, A master Dissertation.
- Kang Y. K., Kim J. W., Lee S. I., Lee S. B. (2011), Development of Traffic Accident Frequency Prediction Model in Urban Signalized Intersections with Fuzzy Reasoning and Neural Network Theories, *J. Korea Soc. Road Eng.*, 13(1), 69-77.
- Kim D. H., Lee D. M., Sung N. M. (2010), A Development of Traffic Crash Frequency Prediction Models for Rural 3-Lagged Intersections, *Journal of transport Research*, 17(1), 37-48.
- Kim E. C., Lee D. M., Kim D. H. (2008), Development of Traffic Accident Frequency Model for Evaluating Safety at Rural Signalized Intersections, *J. Korea Soc. Road Eng.*, 10(4), 53-63.
- Kim K. S. (2010), AMOS 18.0, Hannarae Academy, Korea.
- Kim S. R., Bae Y. K., Chung J. H., Kim H. J. (2011), Factor Analysis of Accident Types on Urban Street using Structural Equation Modeling(SEM), *J. Korean Soc. Transp.*, 29(3), Korean Society of Transportation, 93-101.
- Kim W. C., Lee S. B., Namgung M., Hirofumi I. (2001), Constructing Method of Traffic Accidents Prediction Model for Safety Evaluation at Intersections, *KSCE Journal of Civil Engineering*, 21(4-D), korean society of civil engineers, 427-435.
- Lee H. S., Lim J. H. (2011), SPSS 18.0 Manual, Korea.
- Lee J. K. (2009), A Study on Prediction of the Early-Age Strength of Concrete using Artificial Neural Network Theory, Joongbu University graduate school, A master Dissertation.
- Lee J. P. (2001), Application of Artificial Neural Network to Predict Speed on Two-lane Rural Highway, Hanyang University graduate school, A master Dissertation.
- Lee J. Y., Chung J. H., Son B. S. (2008), Analysis of Traffic Accident Severity for Korean Highway Using Structural Equations Model, *J. Korean Soc. Transp.*, 26(2), Korean Society of Transportation, 17-24.
- Lee S. W. (2002), The Study on the Selecting Optimal Artificial Neural Networks Model Prior to Forecasting Stock. Inje University graduate school, A master Dissertation.
- McCoy P. T., Malone M. S. (1989), Safety Effects of Left-Turn Lanes on Urban Four-Lane Roadways, *Transportation Research Board, TRB*, 1239, 17-22.
- Neter J., Kutner M., Wasserman W., Nachtsheim C.

