

■ 論 文 ■

퍼지 및 신경망 이론을 이용한 교통사고예측모형 개발에 관한 연구

Development of Traffic Accidents Prediction Model With Fuzzy
and Neural Network Theory

김 장 욱

(서울시립대학교 교통공학과 박사수료)

남 공 문

(원광대학교 토목환경도시공학부 교수)

김 정 현

(한국철도기술연구원 미래기술실용화센터 책임연구원)

이 수 범

(서울시립대학교 교통공학과 교수)

목 차

I. 서론

- 1. 연구의 배경 및 목적
- 2. 연구의 방법 및 범위

II. 문헌고찰

III. 기본이론

- 1. 다중선형회귀이론
- 2. 수량화이론
- 3. 퍼지추론이론
- 4. 신경망이론

IV. 교통사고 요인 설명

V. 교통사고예측모형 개발

- 1. 다중선형회귀를 이용한 교통사고예측모형
- 2. 수량화 I 류를 이용한 교통사고예측모형
- 3. 퍼지추론을 이용한 교통사고예측모형
- 4. 신경망을 이용한 교통사고예측모형

VI. 사례분석

VII. 결론

참고문헌

Key Words : 교통사고, 다중선형회귀, 수량화 I 류, 퍼지추론, 신경망

Traffic Accidents, Multiple Linear Regression, Quantification I Theory, Fuzzy Reasoning, Neural Network

요 약

교통사고를 줄이기 위한 방안으로써, 교통사고와 다양한 요인과의 관계를 규명하는 것이 시급한 현실의 과제일 것이다. 본 연구에서는 전북권의 교통사고가 가장 많고, 치사율이 가장 높은 국도 17호선(전주~남원)를 대상으로 교통사고의 원인이 되는 다양한 요인들이 교통사고에 어느 정도 영향을 미치고 있는지에 대하여 교통안전분야에서 자주 사용되어오던 다중회귀이론, 수량화이론을 적용하여 교통사고예측모형을 구축하였다. 또한, 데이터의 불확실성 상태를 합리적으로 처리할 수 있는 퍼지추론이론 및 인간의 신경계를 수학적으로 모형화하여 학습에 의한 예측에 있어 뛰어난 것으로 알려져 있는 신경망이론을 적용한 교통사고예측모형을 구축하였다. 이를 통해, 퍼지추론이론 및 신경망 이론의 유효성을 입증하고 교통사고분석 분야의 적용 타당성을 확인하는데 초점을 맞추고 있다.

It is important to clarify the relationship between traffic accidents and various influencing factors in order to reduce the number of traffic accidents. This study developed a traffic accident frequency prediction model using by multi-linear regression and qualification theories which are commonly applied in the field of traffic safety to verify the influences of various factors into the traffic accident frequency.

The data were collected on the Korean National Highway 17 which shows the highest accident frequencies and fatality rates in Chonbuk Province. In order to minimize the uncertainty of the data, the fuzzy theory and neural network theory were applied. The neural network theory can provide fair learning performance by modeling the human neural system mathematically.

In conclusion, this study focused on the practicability of the fuzzy reasoning theory and the neural network theory for traffic safety analysis.

1. 서론

1. 연구의 배경 및 목적

오늘날에 도로교통은 우리의 생활에 직접적으로 영향을 미치는 매우 중요한 요소 중의 하나로 대두되고 있다. 사회가 발전하면서 사람과 물자의 이동성이 더욱 절실해지면서 자동차의 중요성 및 필요성이 더욱 절실해지고 있는 실정이다. 현대 사회에서 자동차는 필수품으로서 긍정적인 측면도 많지만, 교통사고라는 부정적인 측면이 있는 것도 사실이다. 현대사회에서 교통사고란 중요한 사회적인 문제로 대두되고 있는 실정이다. 도로교통체계란 사람, 도로, 및 차량이 복합적으로 어우러져서 운영된다고 볼 수 있다. 교통사고란 위의 3가지 요소 중에 한 가지 또는 여러 가지가 불완전할 경우 발생한다고 볼 수 있다.

이제까지 전 세계적으로 교통사고 예측 모형은 많이 개발되었다. 그러나 이제까지의 모형은 주로 도로기차 구조적 특성과 교통특성 등을 설명 변수로 이용하여 모형을 개발하였으며, 데이터 수집과정에서의 오류에 대한 보정에 관해서는 크게 신경을 쓰지 못한 것이 사실이다. GIGO(Garbage In Garbage Out) 원리에 의하면 수집 자료의 신뢰성이 떨어지면 그 자료를 이용한 모형의 결과에 대한 신뢰성도 떨어지게 된다. 즉, 수집 자료의 신뢰성 문제를 해결하지 못하면 교통사고의 원인을 정확히 규명하는 일과, 정확한 사고 예측에 한계가 있다는 의미이다.

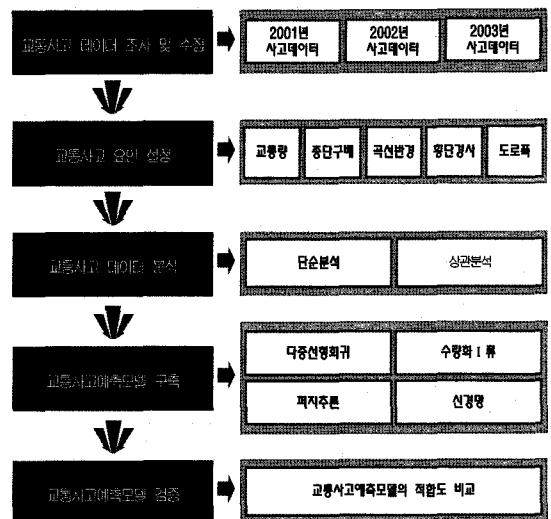
교통사고 관련 자료 수집 시 조사원의 숙련도 및 조사원간의 차이로 인하여 객관적이며 정확한 자료 수집이 한계가 있는 것이 현실이다. 이러한 경우에 교통사고의 주원인을 명확히 찾아내기가 어렵기 때문에 데이터의 불확실한 상태를 인정하면서 데이터의 불확실성을 최소화하여 이용할 수 있는 방법의 개발이 절실하다. 기존의 방법들은 데이터 수집 과정 속에 내포되어 있는 불확실성과 비선형성 및 시공간적 다양성에 의해 발생할 수 있는 오차를 규명하는데 한계가 있다. 이에 본 연구에서는 데이터의 불확실성 상태를 합리적으로 처리할 수 있는 퍼지추론이론 및 인간의 신경계를 수학적으로 모형화하여 학습에 의한 예측이 뛰어난 것으로 알려져 있는 신경망 이론을 적용한 교통사고예측 모형을 구축하였다. 또한 교통사고에 미치는 요인의 다양성을 고

려하여 기존의 다중회귀이론과 수량화 이론을 적용하여 요인의 다양성을 최대한 반영한 모형을 구축하였다. 이렇게 구축된 모형들의 적합성을 비교하여 어떤 이론이 적합성 측면에서 사고예측 시 정확도가 높은지를 제시하였다.

2. 연구의 방법 및 범위

본 연구는 초기단계로서 지방부 국도 상에 신호 교차로에서 차량과 보행자간의 사고 자료를 이용하여 모형을 구축하였다. 기존 모형에서 반영하지 못한 교통사고 원인의 다양성과 자료 조사 시의 오차를 최소화 할 수 있도록, 여러 가지 이론을 교통사고예측 모형에 적용하고 각각의 특성을 제시하는 것에 본 연구의 의의가 있다고 할 수 있다. 교통사고에 영향을 미치는 요인들을 명확하게 규명하기 위하여 본 연구에서는 전북권의 국도 17호선(전주~남원)를 연구대상으로 선정하였다. 본 연구에 필요한 자료를 수집하기 위해 경찰청에서 제공하는 교통사고자료와 도로교통 안전관리공단에서 제공하는 교통사고 잦은 곳 기본개선계획 보고서의 기하구조 자료를 수집하였다. 수집된 데이터를 기초로 하여 다중회귀이론, 수량화이론, 퍼지추론이론, 신경망이론을 이용하여 교통사고예측모형을 구축한다. 이를 통해 교통사고에 영향을 미치는 요인들을 규명한다.

교통사고는 인적요인, 차량적요인, 도로환경적요인 등이 복합적으로 작용하여 발생한다고 볼 수 있다. 하



<그림 1> 연구 흐름도

지만 본 연구에서는 교통사고에 있어 공통적인 요인이 될 수 있는 교통안전시설, 도로기하구조, 교통통계, 교통특성 등과 같은 도로환경적인 요인만을 고려하여 교통사고예측모델을 구축한다. 본 연구의 수행과정은 <그림 1>과 같다.

II. 문헌고찰

門田博知, 今田寛典(1976)는 수량화이론 제 I 류를 적용하여 “교차점에서의 사람 대 차량 사고대책에 관한 기초연구”를 수행한바 있다. 강경우(1997)는 공간자기회귀모형을 이용한 고속도로 교통사고 분석에 관한 그의 연구에서 공간적 상관관계를 고려한 교통사고분석을 위하여 교통사고에 영향을 미치는 변수로 단위구간별 교통량, I.C.유무 및 화물차량비율을 이용하여 공간자기회귀분석을 시도한바 있다. Zeeger(1998)는 미국의 7개주 5000마일에 이르는 2차선도로에 대한 자료를 기초로 사고유형별 사고율을 종속변수로 도로종단선형, 일평균교통량, 차선폭, 전체 노면 폭을 독립변수로 하는 사고모형을 개발하였다. 또한, Wei Lien Liang (1998)은 사고발생요인과 교통사고와의 관계를 규명하기 위한 기초 연구로써 도로 환경적 요인인 바람의 속도와 Visibility Level의 변화가 차량속도에 미치는 영향에 관한 연구를 수행하였다.

Kay Fitzpatrick (2001) 등은 사고분석에 있어서 기하구조변수로써 차로폭, 중앙분리대의 유/무, 곡선반경, 편향각(Deflection Angle) 등을 적용하여 사고와의 관계를 규명하였다. 특히 차로폭이 동일하지 않은 도로구간에 있어서는 차로폭이 중요변수로 작용함을 모형을 통해 제시하였다. 김원철(2001)은 교통사고 잦은 교차로를 연구지로 선정하여 다중회귀이론, 수량화이론, 구조방정식이론, 퍼지추론이론 등을 통하여 이론의 적용성을 평가하고 교통사고건수에측모형을 구축한 후 각각 모델간의 재연성을 확인하였다.

III. 기본이론

본 장에서는 교통사고예측모형화에 사용된 다중선형회귀이론, 수량화이론, 퍼지추론이론, 신경망이론에 대한 기본개념을 개략적으로 기술한다.

1. 다중선형회귀이론(Multiple Linear Regression Theory)

다중선형회귀분석법은 독립변수와 종속변수의 선형 (Linearity)관계를 파악하는 분석기법이다. 이는, 독립변수와 종속변수의 관련성의 강도와 독립변수 값의 변화에 따른 종속변수 값의 변화를 예측하는데 사용되는 분석수법이다. 다중회귀분석의 목적은 선형회귀방정식을 도출하여 종속변수를 예측하는데 있다. 다중회귀방정식은 최소자승법에 의해 추정되는데 최소자승법은 종속변수의 실제 값과 회귀방정식으로 예측한 값 즉, 오차의 제곱의 합을 최소화시킨 방법이다. 회귀방정식 모형은 식(1)과 같이 독립변수가 하나인 단순회귀(Simple Regression)와 식(2)와 같이 독립변수가 두 개 이상인 다중선형회귀(Multiple Linear Regression)로 구분된다.

$$Y = \alpha + \beta x + e \tag{1}$$

$$Y = \alpha + \beta_1 x_1 + \dots + \beta_i x_i + e \tag{2}$$

- 여기서, Y : 종속변수
- x : 독립변수
- α : 회귀식의 상수
- β : 추정해야 할 회귀계수
- e : 실제 관찰값과 회귀식의 예측값과의 오차

2. 수량화이론(Quantification Theory)

수량화 이론에서 외측기준 Y 의 발생을 예측하기 위해서 그 현상에 관련된 m 개의 항목(Item)을 몇 개의 카테고리(Category)로 구분한다. 이에 해당하는 설명특성(X_1, X_2, \dots, X_m)들이 측정될 때, 이 설명특성 X_j 에 기초하여 목적의 특성 수량 Y 를 예측한다. 또는, Item · Category의 요인분석을 행하기 위한 방법으로 이산적 데이터를 활용한 분석방법이 수량화이론 제 I 류이다. 이 분석방법은 설명특성 X_j 가 수량 즉, 연속적인 값을 취할 때의 다중회귀분석과 호응된다고 말할 수 있다. 이를 식으로 표현하면, 교통사고 지점 i 의 외생변수 y_i 를 교통사고건수, 설명변수를 교통사고 요인 x_{jki} 라 가정하면,

$$y_i = \sum \sum a_{jk} x_{jki} + \epsilon_i \quad (3)$$

여기서,

$$a_{jk} = \begin{cases} 1 : \text{카테고리}(C_{jk}) \text{에 해당} \\ 0 : \text{카테고리}(C_{jk}) \text{에 비해당} \end{cases}$$

ϵ : 오차항

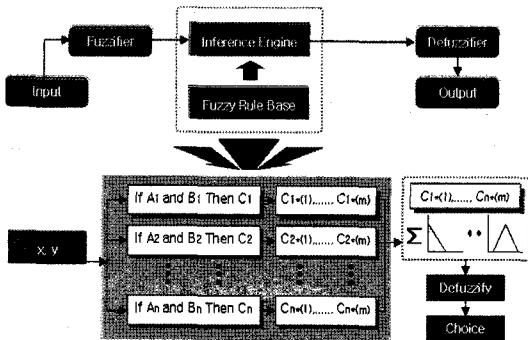
의 선형관계가 성립된다.

3. 퍼지추론이론(Fuzzy Reasoning Theory)

퍼지이론은 1965년 미국 버클리 대학의 자데(Lofti A. Zadeh)교수에 의해 처음 소개되었으며 일본 및 유럽에서 활발하게 연구되어 응용되고 있는 학문이다. 확실한 경계가 규정되지 않은 개념의 불확실성(Vagueness)을 다루는 퍼지추론은 몇 가지 퍼지명제로부터 어느 한 가지 명제를 끌어내는 추론법이다.

인간이 행하고 있는 추론과 상당히 비슷하다는 점에서 많은 관심을 받아왔다. 또한, 퍼지제어, 엑스퍼트시스템, 의사결정 등의 분야에서도 중요한 역할을 하고 있다. 퍼지추론의 개념은 <그림 2>와 같이 표현할 수 있다. 퍼지추론을 위한 일반적인 규칙의 형식은 "IF A_i , THEN B_i "으로 구성된다. 전건부의 A_i 를 가지고 있는 모든 규칙 i 와 입력값 A^* 와 합치하는 부분을 통하여 결과인 B^* 로 나타난다.

추론구조는 각 규칙에 대해 현재의 입력 값이 규칙에서 요구하는 표준 입력 값과 정확하게 맞지 않더라도 이 입력 값을 적용할 수 있도록 도와주고 근사적인 결과를 유도하게 된다. 여기에서 결과 값은 모든 B_i^* 의 조합에 의하여 만들어지며 최종적으로 모든 규칙의 결과 값은 비퍼지화(Defuzzification)되어 하나의 실수



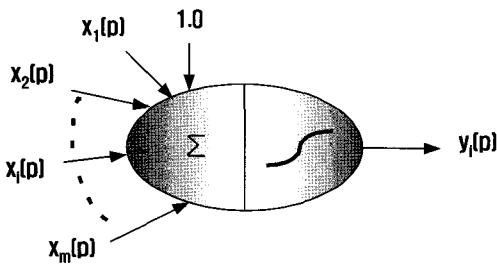
(그림 2) 퍼지추론에 의한 의사결정과정

(Real Value)로 최종적인 의사결정이 도출된다. 또한 추론시에 근사추론 구조를 사용하여 의사결정시 정보처리에 있어 유연성을 두었다.

4. 신경망이론(Neural Network)

인공 신경망은 인간의 두뇌가 특별한 기능이나 임무를 수행하도록 하는 방법을 모형화하도록 구성되어진 구조이다. 학습과정을 수행하는 절차는 학습 알고리즘이라 일컬어지며 원하는 목적을 성취하기 위해 신경망의 연결강도를 조정하는 기능을 한다. 이러한 신경망의 여러 장점이 알려져 있으나, 본 연구에 부합될 수 있는 면에서 야기될 수 있는 장점은 비선형성, 입력과 출력간의 사상(Mapping), 그리고 적응성(Adaptivity)이라 할 수 있다. 첫째, 신경망은 선형적 또는 비선형적일수도 있다. 그러나 비선형적인 뉴런들 간의 상호연결에 의한 신경망은 그 자체로 비선형적이며, 이 비선형성은 신경망 전체에 걸쳐 분포되어진다. 교통사고 데이터와 같이 현상 자체에 비선형성이 포함되어 있을 경우 비선형성을 내포할 수 있는 교통사고예측모형의 개발은 더욱 중요하다고 할 수 있다. 둘째, 신경망은 입력과 출력간의 사상에 의해 학습을 한다. 즉, 교사 학습이라 알려진 학습 방법은 훈련 자료를 신경망에 적용함으로써 그 자체의 연결강도를 조절한다. 적용된 각 자료는 독특한 입력 신호와 그에 상응하는 응답을 포함한다. 신경망의 훈련은 목표로 하는 응답과 신경망에 의해 발생된 응답 사이의 차이가 최소가 되어 연결 강도의 변화가 더 이상 일어나지 않는 안정된 상태에 도달될 때까지 반복되어 수행된다. 셋째, 신경망은 주위 환경에 적응되어지도록 연결 강도를 조절 할 수 있는 능력을 가지고 있다. 특히 특정한 환경에서 수행되어지도록 훈련된 신경망은 그 작용하는 환경 조건들의 조그만 변화를 다루기 위해 쉽게 재훈련(Retrain)되어질 수 있다.

신경망은 수학적 연산능력을 가진 노드들이 상호 연결된 것으로 적절한 학습규칙에 의하여 작동한다. 즉, 각 노드들은 결합 함수와 전달 함수(활성화 함수)에 의해 수학적 연산을 수행한다. <그림 3>은 각 노드에서의 신호의 흐름을 나타내고 있다. <그림 3>에서와 같이 실제 노드로 입력되는 신호는 가중치가 부여된 값들의 합이 되며 식(4)와 같이 나타낼 수 있다.



〈그림 3〉 신경망의 노드에서의 신호의 흐름

$$s_j = \sum_{i=0}^m (w_{ji} \cdot x_i) \quad (4)$$

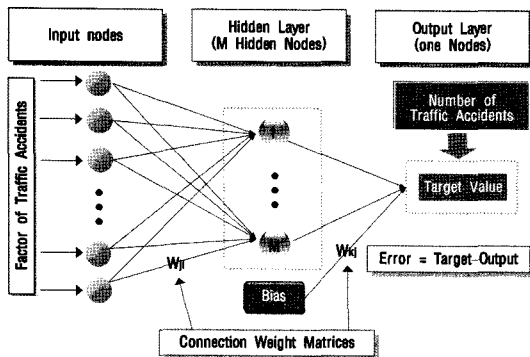
여기서, w_{ji} : 연결가중치, x_i : 입력값

식(4)에서 실제 입력 신호인 s_j 는 전달 함수 또는 활성화 함수라 불리는 비선형 함수를 통과하게 된다. 가장 일반적으로 사용되는 비선형 함수로는 시그모이드(Sigmoid) 함수가 있다. 본 연구에서는 단극성 시그모이드 함수를 사용하였으며, s_j 가 입력된 후의 결과인 y_j 는 식(5)로 나타낼 수 있다.

$$y_j = f(s_j) = \frac{1}{1 + \exp(-s_j)} \quad (5)$$

〈그림 4〉는 본 연구에 사용된 다층(Multi-Layer) 신경망이며, 입력층과 은닉층, 그리고 출력층으로 구성된다.

또한 사용된 학습 알고리즘은 모멘텀 상수와 적응식 학습율이 적용된 최급 강하법을 이용한 역전파 학습 알고리즘이다. 이에 포함된 활성화 함수는 단극성 시그모



〈그림 4〉 다층신경망의 구조

이드 함수와 선형(Linear) 함수를 사용하였다. 그리고, 훈련과정에서 효율성을 높이기 위해 Early Stopped Training Approach 기법을 사용하였다.

N. 교통사고 요인 설명

본 연구에서는 교통사고 분석에 대해 미시적인 접근을 시도하였다. 따라서, 교통사고의 요인으로써 교통량, 종단경사, 횡단경사, 곡선반경, 도로폭 등을 고려하여 분석을 하였다. 이러한 사고요인들을 선정하기 위해 교통사고에 영향을 많이 주는 요인 순으로 구분하였다. 또한, 요인간의 상관관계를 고려하여 상관성이 큰 변수들에 대해서는 상호간의 상관계수 값을 비교하여 큰 변수를 요인으로 채택하였다.

1) 교통량

교통량은 도시의 발전경향, 시가화의 상황, 도시 이용과 교통량 증감에 따른 교통밀도, 주행속도에 따른 운전자 심리에 영향을 주는 요소이다. 또한, 교통량이 증가하여 교통상황이 복잡하게 되면 운전자 상호간의 의사소통에 무리가 따르게 된다. 운전자의 감정을 자극하여 공격적 감정표출 및 공격적 행동으로 인한 사고를 유도하는 유발요인이 되기도 한다(이순철, 2000). 본 연구에서는 사고요인인 교통량에 대해 년 평균 일교통량(AADT : Annual Average Daily Traffic)을 이용하였다.

2) 종단경사

TRB(1987a)의 보고에 의하면 종단경사에 의한 시거확보의 부족으로써 발생한 사고가 경사가 고려되지 않는 지역에서보다 52%정도 높게 발생했다고 보고하였다. 이렇듯 사고에 많은 영향을 미치는 요인으로써 도로의 종단경사를 교통사고분석에 고려하기 위해서는 도로의 경사도나 Curve를 동시에 고려하는 방안이 강구되어야 한다.

3) 횡단경사

차도에서 배수를 위하여 노면의 중심을 정점으로 하고 양쪽으로 향하여 경사진 횡단경사를 붙인다. 노면에 내린 우수는 횡단경사로 측구 또는 가거로 유도된다.

횡단경사가 크면 클수록 배수상 유리하지만 운전상의 안전과 쾌적면으로 보면 수평에 가까울수록 좋다. 따라서 횡단경사는 배수에 지장이 없는 범위에서 될 수 있는대로 완만하게 하는 것이 좋다.

4) 곡선반경

자동차가 도로의 곡선부에서도 직선부와 마찬가지로 안정되고 쾌적하게 주행할 수 있도록 곡선부의 곡선반경은 어느 값 이상으로 유지되어야 한다. 자동차가 곡선부를 주행할 때 생기는 위험은 원심력에 의해 곡선부 외측으로 미끄러지거나 뒤집어 지는 것이다. 그 위험성은 자동차의 주행속도와 도로의 곡선반경, 편경사 및 노면의 마찰계수에 따라 달라진다.

5) 도로폭

도로의 폭은 자동차 주행에 주행 안전성을 확보할 수 있는 중요한 요소이다. 따라서 도로의 최소 폭은 안전한 주행을 보장할 수 있는 최소의 폭으로 설치되어야 하며 이 값은 설계속도가 커짐에 따라 증가한다. 도로의 폭이 규정보다 작게 설치되어 있을 경우 옆의 진행 차량, 혹은 반대편의 진행차량에 의해서 운전자는 불안감을 느낄 수 있고 이에 따라 운전상태에 영향을 미칠 수 있다.

V. 교통사고예측모델 개발

1. 다중선형회귀를 이용한 교통사고예측모델

교통량과 기하구조(종단경사, 곡선반경, 횡단경사, 도로폭)가 교통사고에 영향을 미치는 정도를 알아보기 위해 다중선형회귀이론을 이용해 모형을 구축하였다. 여기서, 모델의 적합도 검정으로 결정계수(R^2)값과 분산분석의 검정통계량 F값을 사용하였다. 결정계수 값은 모형을 설명하는 설명변수가 많으면 많을수록 그 값이 증가한다. 이러한 경향을 보완하기 위한 방법으로 수정결정계수(Adjusted R^2)값을 사용해 모형의 검정에 사용하였다. 또한 설명변수의 모형에 대한 설명력을 판단하기 위해 t 통계 값을 이용하였다. <표 2>를 보면, 다중선형회귀모델의 결정계수(R^2)는 0.239로써 그다지 높은 설명력을 보여주고 있지는 않다. 하지만,

<표 1> ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	2354.55	5	471.510	11.103	0.000
잔차	7517.33	177	42.465		
합계	9872.88	182			

<표 2> 다중선형회귀이론을 이용한 분석결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	48.095	9.342		5.325	0.000
교통량	0.146	0.104	0.075	1.172	0.194
종단경사	1.321	0.401	0.215	3.327	0.001
곡선반경	-0.130	0.085	-0.102	-1.532	0.130
횡단경사	-48.927	8.081	-0.357	-5.542	0.000
도로폭	-1.547	0.501	-0.213	-3.153	0.002
설명력	$R^2 = 0.238$ (Adjusted $R^2 = 0.216$)				

교통사고의 회귀분석시 보여주는 기존의 교통사고에 대한 회귀분석의 결정계수 값과 비교했을 때 어느 정도 설명력을 보여주고 있는 것으로 분석되었다.

다중선형회귀모델의 설명력이 높지 않은 이유는 교통사고의 주요 요인에는 인적요인, 차량적요인, 도로환경요인으로 구분할 수 있는데 이 중에서 인적요인이 교통사고에 있어서 가장 큰 비중을 차지하고 있기 때문이다. 또한, 사고데이터 수집자체의 문제 때문에 모델의 적합도는 그다지 높지 않을 것이라고 판단된다. 이에, 본 연구에서 구축한 모델의 결정계수(R^2)값은 그다지 낮지 않다고 사료된다. 사고데이터 분산분석의 검정통계량 F의 값 또한 유의수준 5%안에 들어 양호한 것으로 검정되었다.

2. 수량화 I 류를 이용한 교통사고예측모델

다중선형회귀이론을 적용시 사용했던 설명변수를 대상으로 수량화 I 류를 이용하여 교통사고예측모델을 구축하였다. 수량화 I 류를 이용하여 교통사고예측모델을 구축함에 있어서 각 요인들의 상대적 중요도, Item과 Category의 수량, Item의 범위, 중앙관계수, 평균에 측오차 등의 평가지수를 구할 수 있었다.

설명요인은 각각의 경우를 조합하여 각 Item별로 Category화 한 후 각 데이터 set을 만들어 모델을 구축하였다.

수량화 I 류 이론에 의한 교통사고예측모델을 구축한 결과, 사고요인들이 분석된 모델을 얼마나 잘 설명하고 있는지의 여부를 판단할 수 있는 범위를 구할 수 있었

〈표 3〉 수량화 I 류를 이용한 모델 구축결과

순위	Item	Category	Parameter	범위
1	교통량	18,000미만 18,000이상	-3.28090 3.07259	6.35349
2	종단구배	1%미만 1%이상	-0.91310 0.50900	1.42211
3	곡선반경	650m미만 650m이상	-1.74912 1.71130	3.46041
4	횡단구배	0.3%미만 0.3%이상	1.45299 -2.26585	3.71885
5	도로폭	17.70m미만 17.70m이상	0.13223 -0.24294	0.37518
Correlation Coefficient : 0.6356				
$R^2 = 0.400$				

다. 그런데, 주의할 점은 모델의 분석결과에서 Item의 범위의 값이 크면 클수록 요인의 중요도는 높아지게 되는 것이다. 수량화 I 류 이론에 의한 사고요인에 대한 Category 분석결과 교통량의 범위가 가장 큰 값을 보이고 있어 교통량이 사고요인으로 가장 큰 영향을 미치고 있음을 알 수 있다.

〈표 3〉를 보면 모델 설명력을 나타내는 중상관계수 값은 63.56%이었고, 수량화 I 류 이론을 이용한 모델의 결정계수(R^2) 값은 0.400으로 다중선형회귀이론을 이용하였을 때보다 높게 나타났다.

3. 퍼지추론을 이용한 교통사고예측모델

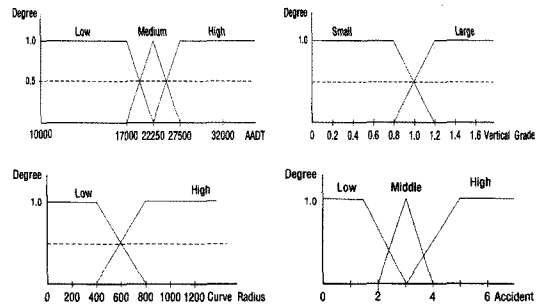
1) 퍼지추론 적용을 위한 멤버십 함수 작성

퍼지추론을 적용하기 위해서 입력변수로는 교통량, 종단구배, 곡선반경, 횡단구배, 도로폭을 사용하였다.

출력변수로는 교통사고건수를 이용하여 멤버십 함수를 작성하였다. 교통량은 연평균 일교통량을 이용하여 {적다, 보통이다, 많다}인 3개의 멤버십 함수로, 종단구배, 횡단구배와 곡선반경, 도로폭은 {높다, 낮다}, {넓다, 좁다}인 2개의 멤버십 함수로 작성하였다.

〈표 4〉 퍼지추론을 위한 IF-THEN 규칙

IF(Traffic Volume is High) and (Vertical Grade is High) and (Horizontal Curvature is Small) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is High)
IF(Traffic Volume is High) and (Vertical Grade is Low) and (Horizontal Curvature is Small) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is High)
IF(Traffic Volume is High) and (Vertical Grade is High) and (Horizontal Curvature is Large) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is High)
IF(Traffic Volume is High) and (Vertical Grade is Low) and (Horizontal Curvature is Large) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is Medium)
IF(Traffic Volume is Medium) and (Vertical Grade is High) and (Horizontal Curvature is Small) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is High)
IF(Traffic Volume is Medium) and (Vertical Grade is High) and (Horizontal Curvature is Small) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is Medium)
IF(Traffic Volume is Medium) and (Vertical Grade is Low) and (Horizontal Curvature is Small) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is Medium)
IF(Traffic Volume is Medium) and (Vertical Grade is High) and (Horizontal Curvature is Large) and (Cross Slope is High) and (Roadway Width is Small) Then (Traffic Accident is Medium)
IF(Traffic Volume is Low) and (Vertical Grade is High) and (Horizontal Curvature is Large) and (Cross Slope is High) and (Roadway Width is Large) Then (Traffic Accident is Low)
IF(Traffic Volume is Low) and (Vertical Grade is Low) and (Horizontal Curvature is Large) and (Cross Slope is High) and (Roadway Width is Large) Then (Traffic Accident is Low)
IF(Traffic Volume is Low) and (Vertical Grade is High) and (Horizontal Curvature is Small) and (Cross Slope is Low) and (Roadway Width is Large) Then (Traffic Accident is Low)
IF(Traffic Volume is Low) and (Vertical Grade is Low) and (Horizontal Curvature is Small) and (Cross Slope is Low) and (Roadway Width is Large) Then (Traffic Accident is Low)



〈그림 5〉 멤버십 함수

2) 연산방법

〈표 4〉는 퍼지추론의 연산을 위해 작성된 48가지의 퍼지추론 규칙의 예를 나타낸 것이다. 퍼지추론 규칙은 일반적인 상식에 대응되는 IF-THEN규칙을 사용하였다. 즉, “교통량은 많고 종단구배는 높으면서 곡선반경은 좁고 횡단구배는 높으며 도로폭이 좁다면 교통사고 발생 건수는 높아진다” 와 같은 퍼지추론 규칙을 작성하였다.

퍼지추론 방법으로는 Mamdani의 Min-Max 중심법(Centroid)을 적용하여 계산하였다. 또한, 각 변수에 대한 멤버십함수의 범위와 중복도는 전문가의 경험을 이용하여 각 데이터의 평균값과 50%중복도를 사용하였다. 이와 같은 방법에 의하여 퍼지추론에 의한 교통사고예측모델을 추론하였다. 퍼지추론모형의 평균 P_i 값은 적합성을 나타내는 것으로, 식(6)과 같다.

$$\Pi(\text{average}) = \frac{\sum_{i=1}^n |R_i - P_i|}{n} \quad (6)$$

여기서, n : 자료의 수, R_i : 실측값, P_i : 예측값

〈표 5〉을 보면 모델 설명력을 나타내는 P_i 값은 0.353이었고, MSE 는 0.491로 분석되었다. 그리고 퍼

〈표 5〉 퍼지추론을 이용한 모델 구축결과

설명변수	
교통량 (적다, 보통이다, 많다) 종단구배, 횡단구배 (높다, 낮다) 종단곡선, 도로폭 (넓다, 좁다)	
Pi (평균)	0.353
MSE	0.491
적중률(%)	65.23
R ²	0.692

지추론 이론을 이용한 모델의 결정계수(R²)값은 0.692로 다중선형회귀이론 및 수량화 I 류를 이용하였을 때보다 높게 나타났다.

4. 신경망을 이용한 교통사고예측모델

본 연구에서 신경망 모형을 학습시키기 위한 학습자료(Learning Data)로서 교통량, 종단구배, 곡선반경, 횡단구배, 도로폭을 사용하였다. 모든 학습 자료는 전처리과정(Pre-process)인 정규화(Normalization)를 한다. 즉, 평균과 표준편차가 (0,1)인 정규화된 자료로 변환하였다. 이는 후에 후처리 과정(Post-process)을 거쳐 실제 값으로 다시 변환되어진다.

신경망 모형을 학습시키기 위한 방법으로 일반적으로 널리 응용되고 있는 오차 역전파 알고리즘을 적용하여 모형에 따른 특성을 분석하였다. 예측치와 실측치를 비교하여 적중률을 알아보았다. 신경망 모형은 각각의 변수들을 적용하여 구성하였다. 입력층의 노드(Node)의 수를 n 이라 하면, 은닉층의 노드수는 n 부터 6n 까지의 변화를 시키면서 최적의 신경망모형을 선별하였다. 또한 은닉층 처리소자의 개수에 따른 학습효과를 분석하기 위해 동일한 반복횟수 10,000회에 대해 각 모델별로 모의하였다. 이는 10,000회 이상의 기간으로 설정할 경우, 최대 반복횟수까지 훈련이 반복되어 모델의 일반화 및 효율성이 떨어지는 것으로 나타났으나 10,000회 정도의 기간에서 가장 좋은 결과를 얻을 수 있었다. 따라서 본 연구에서는 모형에 대한 선정기간을 10,000회의 기간으로 설정하였다. 신경망의 보정을 위한 훈련에 사용된 학습방법은 알고리즘 적용시 오차가 지역 최소값(Local Minima)에 수렴하는 것을 방지하고 학습의 효율을 높이기 위해 모멘텀-적응식 학습을 방법을 사용하였다. 모멘텀 상수와 초기 학습률은 민감도 분석을 통해 얻었고 모든 모형에서 각각 0.1를 일괄적으로 사용하였다. 일련의 과정을 거쳐 얻어진 모형

〈표 6〉 신경망을 이용한 모델 구축결과

설명변수	
교통량 (적다, 보통이다, 많다) 종단구배, 횡단구배 (높다, 낮다) 종단곡선, 도로폭 (넓다, 좁다)	
Pi (평균)	0.285
MSE	0.371
적중률(%)	71.23
R ²	0.753

들의 평가기준은 수치적 기준을 적용하여 살펴보았다. 수치적기준인 MSE(평균제곱오차)는 모델 적합도의 전체적인 우수성을 보여주기 위한 것으로 식(7)과 같다. MSE값이 0이라면 관측치와 예측치가 완전히 일치하는 경우라고 할 수 있다.

$$MSE = \frac{\sum_{i=1}^n (obs_i - exp_i)^2}{n} \tag{7}$$

여기서, n : 자료의 수,
obs_i : 관측값, exp_i : 출력값

〈표 6〉를 보면 모델 설명력을 나타내는 Pi값은 0.285이었고, MSE은 0.371로 분석되었다. 그리고 신경망을 이용한 모델의 결정계수(R²)값은 0.753으로 앞의 3가지 모델을 이용하였을 때보다 높게 나타났다.

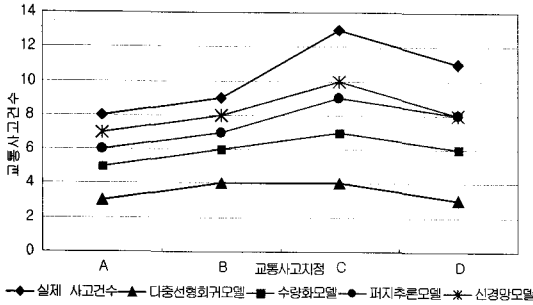
VI. 사례분석

V장에서 구축된 교통사고예측모델을 이용하여 사례 분석을 하기 위해서, 2001~2003년도에 보고된 국도 17호선(전주~남원)의 교통사고 데이터 중에서 교통사고가 잦은 4지점을 선정하였다. 실제 교통사고건수와 각 예측모델별 추론결과를 〈표 7〉과 〈그림 6〉에 나타내었다.

〈표 7〉 교차로별 실제 교통사고건수와 교통사고 예측결과

지점	실제 사고건수	다중선형 회귀모델	수량화 모델	퍼지추론 모델	신경망 모델
A	8	3	5	6	7
B	9	4	6	7	8
C	13	4	7	9	10
D	11	3	6	8	8

* A : 죽림온천, B : 슬치커브, C : 대원주유소, D : 임실역



〈그림 6〉 실제 교통사고건수와 추론결과와의 비교

실제 교통사고건수와 4가지 이론에 의해 구축된 예측모델의 구현 결과를 통해서, 퍼지추론 및 신경망에 의해 구축된 사고건수예측모델이 사고 재현성면에서 기존에 사용되어온 다중회귀이론과 수량화이론에 의해 구축된 예측모델보다 그 적용성이 유용하다는 것으로 판단할 수 있다.

Ⅶ. 결론

본 연구에서는 교통사고예측모형을 다중회귀이론, 수량화이론, 퍼지추론 및 신경망의 4가지 이론을 적용하여 구축하였다. 교통사고예측모형에 교통사고에 영향을 미치는 요인으로 사용된 변수들은 주로 도로환경적인 요인들로서 종단구배, 곡선반경, 횡단구배, 도로폭 및 교통량이 사용되었다.

우선 기존의 다중선행회귀분석 방법론을 이용하여 교통사고예측을 수행하였고, 적합성을 높이기 위해서 수량화 I류를 이용하여 모형을 정립하였다. 수량화 I류를 이용한 모형의 적합성이 기존의 다중선행회귀분석을 이용한 모형보다는 적합성이 높은 것으로 나타났다. 다중회귀분석 및 수량화 I류의 경우에는 조사 자료의 신뢰성을 고려할 수 없는 모형이라고 할 수 있다. 이를 해결하기 위해서 퍼지추론이론 및 신경망 이론을 고려하여 모형을 구축하였다. 그 결과 이 두 가지 모형의 적합도가 기존의 모형들의 적합도보다 훨씬 우수한 것으로 나타났으며 추정표준오차도 더 적은 값을 보이는 것으로 나타났다. 또한 국도 17호선(전주~남원)의 교통사고 데이터 중에서 교통사고가 잦은 4지점에 대한 사례분석 결과, 기존에 행해져 오던 다중회귀이론과 수량화 이론보다 퍼지추론 및 신경망을 적용하여 구축한 교통사고예측모델이 사고재현성 측면에서 매우 유용하다는 것을 알았다.

결론적으로 기존의 모형보다는 본 연구에서 제시한 퍼지추론이론 및 신경망 이론을 이용한 사고예측모형이 조사자의 오류를 감안하여 실제 값과의 적합도를 더 높일 수 있는 모형이라는 것을 제시하였다. 본 연구에서 사용한 자료가 지방부 도로의 신호교차로에 한정되어 있으므로, 향후에는 고속도로와 같은 연속류구간에 대한 사고데이터를 구축하여 사고예측모형을 구축할 것이다. 또한 기존의 직선부와 곡선부를 위주로 한 연구에서는 고려하지 못했던 직선부와 곡선부, 곡선부와 직선부, 배향곡선부로 세분화하여 사고예측모형을 구축할 필요가 있다고 판단된다. 마지막으로 교통사고예측모형을 GIS와 접목하여, 과거와 현재 그리고 미래의 사고 경향을 공간적으로 분석하고, 이에 대한 적절한 대안을 제시하는 과정도 필요하다고 사료된다.

참고문헌

1. 강경우(1997), 공간자기회귀모형을 이용한 고속도로 교통사고 분석, 대한교통학회지, 제15권 제1호, 대한교통학회, pp.5~15.
2. 도로교통안전협회(2000~2003), 전라북도 교통사고 잦은곳 기본개선계획, 도로교통안전협회.
3. 이순철(2000), 교통심리학, 학지사.
4. Brüde, U. and Larsson, J(1988), The use of prediction models for eliminating effects due to regression-to-the mean in road accident data, Accident Analysis and Prevention, Vol.20, No.4, pp.299-310.
5. K. W. OGDEN(1996), Safer Roads : A Guide to Road Safety Engineering, Avebury Technical.
6. K. M. Bauer & D. W. Harwood(1996), Statistical Models of At-Grade Intersection Accidents, Report No. FHWA-RD-96-125, Federal Highway Administration, Washington, D. C.
7. Laurence Capus & Nicole Tourigny(1998), Road Safety Analysis: A Case-Based Reasoning Approach, Transportation Research Board, January.
8. Miaou, S. & Lum. H, Modeling vehicle accident and highway geometric design relationship, Accidents Analysis and Prevention, Vol. 25, No 6, pp.689~709.

9. Wei Lien Liang & Fred Kitchener(1998),
The Effect of Environmental Factors on
Driver Speed : A Case Study, Transportation
Research Board, January.
10. William R. Dillon & Matthew Goldstein(1984),
Multivariate Analysis Method And Applications,
Research Foundation The City University of
New York.
11. 門田博知, 今田寛典(1976), 交叉點の人對車 兩事
故對策に關する基礎研究.

- ♣ 주 작성자 : 김장욱
♣ 교신저자 : 이수범
♣ 논문투고일 : 2006. 2. 24
♣ 논문심사일 : 2006. 3. 28 (1차)
2006. 9. 25 (2차)
2006. 11. 23 (3차)
♣ 심사관정일 : 2006. 11. 23
♣ 반론접수기한 : 2007. 4. 30