



## 데이터 마이닝을 이용한 차량 사고자 사망확률 모형

Development of Car Accidents Person Fatality Model using Data Mining

---

저자 (Authors)	김천식, 홍유식, 정명희 Cheonshik Kim, YouShik Hong, Myunghee Jung
출처 (Source)	<a href="#">전자공학회논문지-TC 43(9)</a> , 2006.9, 25-31(7 pages) <a href="#">The Institute of Electronics Engineers of Korea - Telecommunications 43(9)</a> , 2006.9, 25-31(7 pages)
발행처 (Publisher)	<a href="#">대한전자공학회</a> The Institute of Electronics and Information Engineers
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00757189">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00757189</a>
APA Style	김천식, 홍유식, 정명희 (2006). 데이터 마이닝을 이용한 차량 사고자 사망확률 모형. 전자공학회논문지-TC, 43(9), 25-31
이용정보 (Accessed)	한양대학교 166.***.140.13 2019/11/22 08:26 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

논문 2006-43TC-9-4

# 데이터 마이닝을 이용한 차량 사고자 사망확률 모형

## ( Development of Car Accidents Person Fatality Model using Data Mining )

김 천 식\*, 홍 유 식\*\*, 정 명 희\*

( Cheonshik Kim, YouShik Hong, and Myunghee Jung )

### 요 약

본 논문에서는 데이터 마이닝을 이용한 차량 사고 사망확률 모형을 제안하였다. 본 논문의 목표는 제안된 모델을 이용하여 기술적, 환경적 개선을 통해 교통사고를 줄이는 것으로 교통사고 데이터를 수집하여 데이터 마이닝 알고리즘을 적용하여 사망 원인에 대한 분석을 통해 사망확률 모형을 개발하였다. 이를 위해서 훈련 데이터와 테스트 데이터를 이용하여 최적의 모형을 개발하였다. 이모형을 통해 교통 사망사고에 대한 사망확률을 구할 수 있고 사망 사고의 원인이 되는 중요한 요인을 알아낼 수 있다. 교통사고의 원인이 되는 요인을 기술 개발과 교통 환경 개선에 활용한다면 향후 교통사고를 줄이는데 기여할 것이다.

### Abstract

In this paper, a fatality model of car accident using data mining is proposed with the goal of reducing fatality of traffic accident. The analysis results with a proposed fatality model are utilized to improve a technology and environment for driving. For this, traffic accident data are collected, a data mining algorithm is applied to this data, and then, a fatality model of car accident is developed based on the analysis. The training data as well as test data are utilized to develop the fatality model. The important factors to cause fatality in traffic accidents can be investigated using the model. If these factors are taken into account in traffic policies and driving environment, it is expected that the fatality rate of traffic accident can be reduced hereafter.

**Keywords :** Car Accident, Neural Network, Data Mining, Fatality Model

## I. 서 론

교통사고의 원인으로 선진국의 경우 전체 교통사고의 30~40%가 과속에 의해 발생하고 있다고 분석하고 있다. 그러나, 우리나라의 경우 과속으로 인한 사고가 전체의 1% 이다<sup>[1]</sup>. 이와 같은 이유는 교통사고 통계가 사고 발생의 근본 원인이나 가중 요인 분석 보다는 사고당시의 법규위반 종류를 분석하고 기록하기 때문이다. 실제로 음주운전, 중앙선침범, 신호위반 등 치명적인 인명 피해를 내는 사고의 대부분은 과속과 밀접한

관계가 있으므로 속도위반은 교통사고의 ‘숨은 주범(主犯)’이라 해도 과언이 아니다<sup>[2]</sup>.

우리나라의 경우 교통사고 사망자 수가 줄고 있지만 선진국에 비하여 사망자수의 비율이 높은 편이다. 이와 같은 이유로 교통사고에 대한 요인을 분석하는 많은 논문이 있었다. 그러나 차량 간에 발생하는 교통 사망자 모형에 관한 논문이 없었다. 따라서, 본 논문에서는 차량 사망사고와 관련하여 운전자의 심리적 혹은 육체적 원인, 도로의 상태나 날씨와 같은 환경적인 원인, 그리고 차량결합에 의한 원인 등을 분석하고자한다.

본 논문에서 제안한 모형은 신경망과 C4.5 알고리즘을 이용하였다. 제안한 알고리즘의 정확도를 분석하기 위해서 원주 교통방송국에서 제공하는 교통사고 데이터를 활용하였다. 데이터 100건 중 50건은 학습데이터로 활용했고, 학습결과를 실험하기 위해서 50건의 데이터

\* 정회원, 안양대학교 디지털미디어공학  
(Major in Digital Media Engineering, Anyang University)

\*\* 평생회원, 상지대학교 컴퓨터공학과  
(Dept., Computer Science, Sangji University)  
접수일자: 2006년8월14일, 수정완료일: 2006년9월11일

를 알고리즘에 적용하여 실험하였다.

본 연구에서는 도로에서 차량간의 충돌로 인해서 교통사고가 발생할 경우 운전자의 사망을 예측하는 모형을 제안하였다. 제안한 모형은 향후 운전자의 운전 습관과 태도, 차량관리, 도로의 안전한 관리 등의 개선을 유도하여 인명사고를 줄일 것으로 기대한다.

## II. 관련연구

Kim [Kim et al., 1985]은 심각한 사고를 초래하는 원인이 될 수 있는 운전자의 심리(부부문제, 동료 혹은 상사와의 갈등)와 행동을 명확히 하기위해서 로그-선형(log-linear) 모델을 개발했다<sup>[3]</sup>.

Ossenbruggen [Ossenbruggen et al., 2001]은 logistic regression 모델을 사용하여 도로의 위험도 분석을 목적으로 차량의 충돌 가능성을 예측하였다. 이 모형은 도로의 설계와 ITS등에 사용될 수 있다<sup>[4]</sup>.

Sohn [Sohn & Lee, 2003]은 교통사고 심각도 분류 분석을 함에 있어서 전형적인 데이터 마이닝 기법인 신경망, Decision-Tree, 로지스틱 회귀분석(logistic regression)을 이용하였다. 이들은 교통사고통계 원표에 기록된 여러 가지 범주형 설명변수들을 고려하여 사고 심각도를 3가지 범주(치명적 상해, 경미한 상해, 물적피해)와 2가지 범주(신체상해, 물적피해)로 분류하고 기법간의 분류정확도를 비교 분석하였다<sup>[5]</sup>.

Addelwahab [Abdelwahab and Addel-Aty, 2001]은 중앙 플로리다 지역의 1997년 사고데이터를 연구하였다. 이 분석은 교차로에서 발생하는 차량 사고에 중점을 두었다. 사고의 심각도는 세가지로 분류된다: 부상 없음, 부상가능, 부상 회피. 여기서는 MLP(Multi-layered Perceptron)과 Fuzzy ARTMAP의 성능이 비교되었고, MLP 분류 정확도가 Fuzzy ARTMAP보다 더 정확하다고 보고하였다. Levenberg-Marquardt 알고리즘은 MLP 트레이닝을 위해서 사용되었고 트레이닝과 테스트를 위한 분류 정확성이 각각 65.6%와 60.4% 이었다. 퍼지 ARTMAP는 56.1%의 정확성이었다<sup>[6]</sup>.

지금까지 교통 부상자에 대한 사고 원인과 관련한 많은 연구가 있어 왔다. 하지만 교통사고로 인한 사망 확률 및 사고요인을 예측하는 모형은 없었다. 따라서 본 논문에서는 교통사고 사망자 및 사망확률 예측 모형을 제안하였다. 제안한 모형은 향후 교통사고 사망자를 줄이는데 도움을 줄 것으로 기대한다.

## III. 데이터마이닝 수행 절차

SPSS, NCR, Daimler-Benz 등 여러 업체의 선도 회사들이 3년여 데이터 마이닝 작업의 표준화를 연구하여 1999년 발표한 것이 데이터 마이닝 표준 실행과정인 CRISP-DM(Cross Industry Standard Process for Data Mining)이다. 현재 전 세계 데이터 마이닝 프로젝트의 50% 이상은 CRISP-DM을 따라 수행되고 있다<sup>[7]</sup>. CRISP-DM은 다섯 단계로 구성된다. 모든 과학적 방법론이 그렇듯이 작업 단계들은 항상 전진하는 것이 아니라 때로는 후진하기도 하고 전체가 사이클이 되어 돌기도 한다.

### ① 비즈니스 이해

해당하는 업무를 이해하는 단계이다. 예를 들어 보험업, 신용카드 업, 유통업 등 그 분야의 기본 지식을 각종 참고 자료와 현업 책임자와의 커뮤니케이션을 통하여 이해해야 한다. 그 중에서 데이터 마이닝으로 접근할 수 있는 문제를 파악한다.

### ② 데이터 이해

현업이 보유, 관리하고 있는 데이터를 이해하는 단계이다. 레코드의 수, 변수의 종류, 자료의 질, 데이터 관리 체계 등을 파악한다. 흔히 변수의 종류가 수백 가지에 이르고 데이터가 여러 개의 컴퓨터 서버에 분산되어 있기 때문에 한 조직의 정보체계를 정확히 이해하는 데는 시간이 걸린다.

### ③ 데이터 준비

자료를 컴퓨터 서버로부터 내려 받고나서 분석 가능한 상태로 만들기 위하여 데이터 정제작업을 한다. 예컨대 고객이름과 주소, 전화번호를 1개의 표준 형태로 정리한다. 이 단계에 많은 작업이 필요로 하므로 전체 프로젝트 일정의 50% 이상을 차지하는 것이 보통이다. 또한 우리가 필요로 하는 변수를 만들어내는 일도 필요하다. 예컨대 고객별 총 거래액이 필요하다면 고객별로 개별 거래기록을 묶는 작업을 해야 한다.

### ④ 모델링

자료기술과 탐색을 포함하여 필요로 한 각종 모델링을 한다. 여기에는 신경망, 나무형 모델(decision tree) 등의 지도학습(supervised learning), 군집화(clustering), 연관성(association)분석 등의 비지도학습(unsupervised learning)이 포함된다.

### ⑤ 평가

앞 단계에서 생성된 모형이 잘 해석되는지, 독립적인 새 자료에 적용되는 경우도 재현가능한지를 검토한다.

## ⑥ 전개

검토가 끝난 모형을 실제 현업에 적용하는 단계이다. 예컨대 모든 고객에 대하여 이탈 가능성 점수를 산출하여 고객 관리자에게 전달하여 필요한 조치를 취하는 작업 등이다.

## IV. 제안한 모델 방법론

## 4.1 신경망을 이용한 차량사고 사망자 예측

이 절에서는 차량사고 시 사망자 예측을 위해서 신경망 알고리즘을 이용한다. X축은 차량 사고 요인이고, Y축은 변수의 값(결과 값)을 의미 한다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots \beta_n X_n + \epsilon \quad (1)$$

단, Y : 차량사고 사망자 예측(live, death)

$X_1$  : 종속변수에 영향을 주는 요인 1

$X_2$  : 종속변수에 영향을 주는 요인 2

$X_3$  : 종속변수에 영향을 주는 요인 3

⋮

$X_n$ : 종속변수에 영향을 주는 요인 n

학습 신경망 모형을 사용하기 위하여 다음과 같은 단계를 수행한다<sup>[8]</sup>.

S1. 네트워크의 상태를 결정하는 연결강도  $w_{ji}, w_{kj}$ 와 오프셋(offsets)  $\theta_j, \theta_k$ 를 각각 아주 작은 임의수로 초기화한다. 일반적으로 -0.5 ~ 0.5 사이의 값을 사용한다.

S2. 학습패턴을 설정한다.

S3. 학습 패턴의 값을 입력층 유니트에 제시하여 출력되는 값  $O_{pi}$  입력층과 중간층 사이의 연결강도  $W_{ji}$ 와 중간층 유니트 j의  $\theta_j$ 를 이용하여 유니트 j의 입력  $net_{pj}$ 를 구한다. 다음으로  $net_{pj}$ 와 시그모이드 함수 f를 이용하여 중간층 유니트 j의 출력  $O_{pj}$ 를 구한다.

$$net_{pj} = \sum_{i=0} w_{ji} O_{pi} + \theta_j \quad (1)$$

$$O_{pj} = f_j(net_{pj}) \quad (2)$$

S4. 중간층 유니트의 출력  $O_{pi}$  중간층과 출력층 사이의 연결 강도  $W_{kj}$ 와 출력층 유니트 k의 오프셋  $\theta_k$ 를 이용

하여 출력층 유니트 k의 입력  $net_{pk}$ 를 구한다. 다음  $net_{pk}$ 와 시그모드 함수 f를 이용하여 출력층 유니트 k의 출력  $O_{pk}$ 를 구한다.

$$net_{pk} = \sum_{j=0} w_{kj} O_{pj} + \theta_k \quad (3)$$

$$O_{pk} = f_k(net_{pk}) \quad (4)$$

S5. 학습패턴의 목표출력  $t_{pk}$ 와 실제 출력  $O_{pk}$ 와의 차이로부터 출력층 유니트 k에 연결된 연결강도와 출력층 유니트 k의 오프셋에 대한 오차  $\delta_{pk}$ 를 구한다.

$$\delta_{pk} = (t_{pk} - o_{pk}) f'_k(\neq t_{pk}) = (t_{pk} - o_{pk}) o_{pk} (1 - o_{pk}) \quad (5)$$

S6. 오차  $\delta_{pk}$ 와 중간층과 출력층간의 연결강도  $W_{kj}$ 와 중간층의 출력  $net_{pi}$ 로부터 중간층 유니트 j에 연결된 연결강도와 중간층 유니트의 오프셋에 대한 오차  $\delta_{pj}$ 를 구한다.

$$\delta_{pj} = f'_j(\neq t_{pj}) \sum_{k=0} \delta_{pk} W_{pk} = \sum_{k=0} \delta_{pk} w_{kj} o_{pj} (1 - o_{pj}) \quad (6)$$

S7. (S5)에서 구한 출력층 유니트 k에서의 오차  $\delta_{pk}$  중간층 유니트 j의 출력  $O_{pj}$ 정수  $\alpha$ 와의 곱을 더하여 중간층 유니트 j와 출력층 유니트 k에 연결된 연결강도  $w_{kj}$ 를 수정한다. 또한 오차  $\delta_{pk}$ 와 정수  $\beta$ 와의 곱을 더하여 출력층 유니트 k의  $\theta_k$ 를 수정한다.

$$w_{kj} = w_{kj} + \alpha \cdot \delta_{pk} O_{pj} \quad (7)$$

$$\theta_k = \theta_k + \beta \cdot \delta_{pk} \quad (8)$$

S8. 중간층 유니트 j의 오차  $\delta_{pj}$  입력층 유니트 i의 출력  $O_{pi}$  정수  $\alpha$ 와의 곱을 더하여 입력층 유니트와 중간층 유니트 j에 연결된 연결강도  $W_{ji}$ 를 수정한다. 또 오차  $\delta_{pj}$ 와 정수  $\beta$ 와의 곱을 더하여 중간층 유니트 j의 오프셋  $\theta_j$ 를 수정한다.

$$w_{ji} = w_{ji} + \alpha \cdot \delta_{pj} O_{pi} \quad (9)$$

$$\theta_j = \theta_j + \beta \cdot \delta_{pj} \quad (10)$$

S9. 다음 패턴을 학습시킨다.

S10. 모든 학습 패턴에 대하여 전부 학습할 때까지 S2로 되돌아간다.

S11. 학습의 반복 횟수를 센다.

S12. 학습의 반복 회수가 제한 횟수보다 작으면 S2로 되돌아간다.

<표 1>은 차량 사고에서 생존 결정을 예측하기 위한 12가지 서로 다른 조건을 입력 하였을 때 생존 결정을 예측하는 과정을 나타내고 있다.

신경망 학습의 초기값을 설정하는 것은 중요한 문제다. 초기값을 적절하게 선택함으로써 학습오차가 작고 학습과정이 빠르게 수렴될 수 있기 때문이다. 일반적으로 신경망의 학습은 특정 초기값에서 시작한다.

또한 학습률은 모수 값들을 어떻게 선택하느냐에 따라서 학습오차가 작으면서 학습과정이 빠르게 수렴 할 수도 있고 초기 포화점에 빠질 수도 있다. 그렇기 때문에 분석하고자 하는 자료에 적당한 모수를 선정하여 오차가 최소 값이면서 학습과정이 빠르게 수렴될 수 있게 학습하도록 하는 것은 매우 중요한 문제다.

이를 위해 제한적이지만  $\kappa$ ,  $\Theta$ ,  $\Phi$ ,  $\mu$  (kappa, theta, phi, mu)만을 가지고 각 범위 0.1, 0.3, 0.5, 0.7, 0.9에 따라 모든 경우를 고려해서 임의의 경우로 실험을 해보았다. 그리고 학습시간을 각각 500회로 제한하였다.

표 1. 신경망을 이용한 교통사고 사망자 결정 요소

Table 1. Traffic fatalities determinant that use neural network.

Input data	
변수1	운전자나이 : 20세 이하, 20-29, 30-39, 40-49, 50-59, 60세 이상
변수2	성별 (남, 여)
변수3	운전 경력 : 1년 이하, 1년, 3년, 5년, 10년 이상
변수4	차량의 종류 : 승용차, RV, SUV, 트럭
변수5	도로의 상태 : 상(1), 중(2), 하(3)
변수6	차량의 LIGHT 상태 : yes(정상) / no(이상)
변수7	날씨 : 맑음, 비
변수8	사고 유형 : 정면충돌, 추돌, 측면 충돌
변수9	사고 원인 : 속도, 추월, 신호위반, 졸음, 알콜,
변수10	차량속도 : 40이하, 60, 80, 100이상
변수11	안전벨트 착용 : yes / no
변수12	결과 (사망: no, 생존: yes)

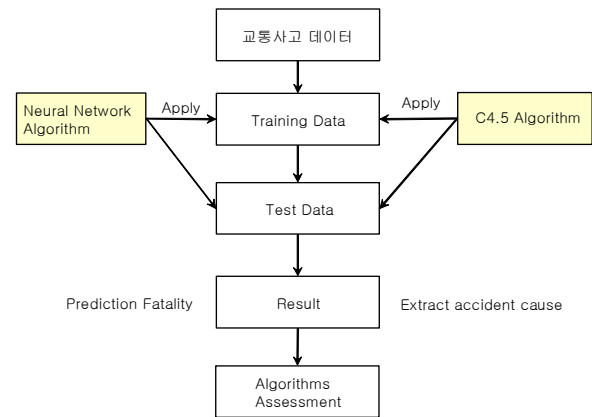


그림 1. 교통사고 사망자 예측 과정

Fig. 1. Traffic fatalities estimate process.

그림 1은 다음과 같은 처리를 수행한다.

P1. 신경망을 이용하여 11개의 서로 다른 조건 테스트 데이터를 학습시킨다.

P2. 50개의 테스트 데이터에 대하여 예측을 한 뒤 테스트 데이터와 예측 데이터의 오차를 계산한다.

$Z_1, Z_2, Z_3, \dots, Z_n$  : 테스트 데이터

$\hat{Z}_1, \hat{Z}_2, \hat{Z}_3, \dots, \hat{Z}_n$  : 예측 값

$$e_i = Z_i - \hat{Z}_i \quad (11)$$

$i$ 시점 시계열 테스트 자료와 예측 값에 대한 차이

$$Z'_j = Z_j + W(Z_j) \quad (12)$$

여기서,  $Z'_j$ 는  $j$ 번째 특이 값으로 식별된 테스트 데이터  $Z_j$ 의 수정된 값을 의미한다.

P3. 후처리로 데이터 마이닝의 C4.5 알고리즘을 이용한다.

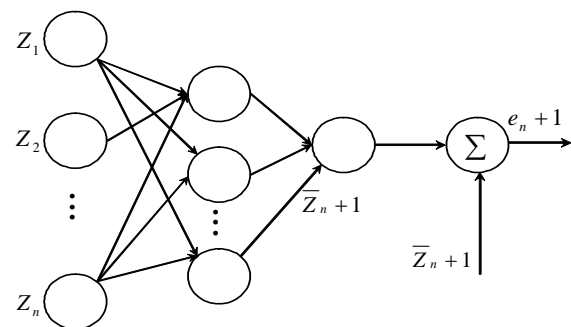


그림 2. 신경망을 이용한 교통사고 사망자 결정 모델

Fig. 2. Traffic fatalities deterministic model that use neural network.

## 4.2 C4.5 알고리즘

본 논문에서는 신경망 알고리즘을 이용해서 최적의 교통사고 시 사망자를 예측하는 모형을 개발하였다. 신경망 알고리즘은 자료 분석 분야에서 복잡한 구조를 가지고 있는 자료에 대하여 예측 문제를 해결하기 위한 유연한 비선형 모형의 하나로 분류될 수 있다. 인간의 신경생리학과 유사성 때문에 일반적으로 다른 통계적 예측모형에 비해 보다 흥미롭게 연구되어지고 있다. 특히, 예측 기법으로써 로지스틱 회귀분석 보다 신경망의 우수함을 비교한 연구들이 고려되고 있다. 그러나 동시에 신경망은 미래의 목표 값을 예측하는데 있어 입력벡터의 값의 수나 형태를 결정할 수 있는 체계적인 방법의 결여와 모델의 분류가 어떻게 이루어지는지 명확하게 이해 할 수 없는 단점이 제시되고 있다. 이러한 단점을 해결하기 위하여 신경망에서 상징적 분류 규칙을 찾거나 의사결정 나무를 통하여 이해 할 수 있는 해석을 얻고자 하는 연구 등이 진행 되어져 왔다.

C4.5의사결정 나무를 형성하기 위하여 처음 수행하는 작업이 분할정복이다. 입력되는 훈련 집합이 성공적으로 분할 되도록 모든 하부 집합에 하나의 클래스가 속하는 경우들로 구성될 때까지 나무를 형성한다. 노드를 분리하는 기준으로 정보이익비율(Information gain ratio)이 사용된다. 나무구조의 결정 규칙을 생성하기 위하여 각 단계에서  $p$ 개의 설명변수 중 어느 것에 의하여 가지분리를 할 것인가를 선택해야 한다. 이 때 결정 규칙들은 각기 다른 기준을 쓰는데, C4.5는 엔트로피 기준을 사용한다. 엔트로피(entropy)는 열역학에서 쓰는 개념으로 무질서도에 대한 측도이다. 자료집합  $T$ 가  $Y$ 에 의하여  $k$ 개의 범주로 분할되고 범주 비율이  $p_1, \dots, p_k$ 라고 하자.  $T$ 의 엔트로피는

$$Entropy(T) = - \sum_{i=1}^k p_i \log p_i \quad (10)$$

으로 정의 된다<sup>[9]</sup>.

C4.5 모형은 엔트로피 기준에서 가장 엔트로피를 낮추는 분리 변수를 찾고자 한다. 분리변수를 찾음으로서 가장 성취도가 좋은 변수 및 수준을 찾는 것이 나무규칙 생성 알고리즘이다. 본 논문에서는 Weka 3.5<sup>[10]</sup>를 이용하여 C4.5 알고리즘을 적용하여 신경망 알고리즘에서 설명할 수 없는 차량사고 사망 요인을 알아낼 수 있다. 실험 데이터는 <표 1>의 데이터를 이용하였고, 그림 1의 과정을 거쳐 그림 3의 결과를 얻었다.

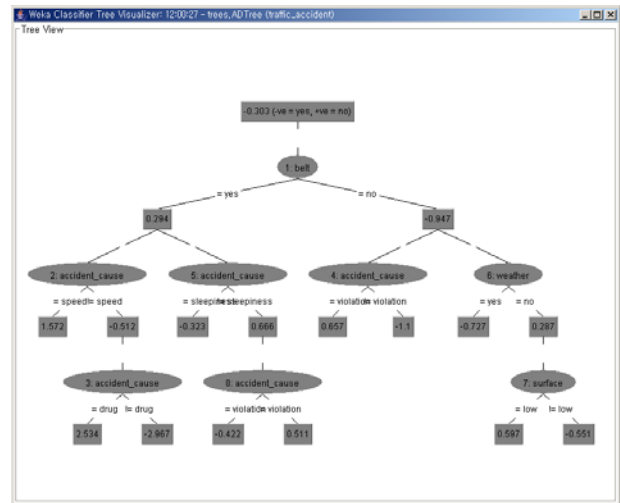


그림 3. C4.5 알고리즘의 실행 결과

Fig. 3. C4.5 result of algorithm.

## V. 제안한 방법의 모델 및 평가

### 5.1 모형 개발

실험의 정확도를 확인하기 위해서 CCR(correct classification rate)을 적용하였다. 전체 사고건수 중 모형에 의해 ‘사망’은 정확히 구분되는 사고수의 비율로서 나타낼 수 있다.

$$CCR(\%) = \frac{\text{'사망'이 정확히 구분된 사고건수}}{\text{전체 사고건수}} \times 100$$

### 5.2 평가 결과

교통사고 데이터를 입력으로 하여 신경망 알고리즘을 이용하여 실험데이터를 평가한 내용은 표 2와 같다. 같은 데이터를 C4.5 알고리즘을 적용한 결과는 표 3과 같다.

표 2. 신경망 실험결과  
Table 2. Experiment result of neural net.

Precision	Recall	F-Measure	ROC Area	Class
0.97	1	0.985	0.946	yes
1	0.943	0.971	0.946	no

평가 데이터를 이용하여 신경망 알고리즘을 적용한 결과 실제 사망자가 41명이었고, 예측 사망자는 35명이었다.

=== Confusion Matrix ===

a b <-- classified as  
55 4 | a = yes  
6 35 | b = no

표 3. C4.5 알고리즘 실험 결과

Table 3. C4.5 experiment result of algorithm.

Precision	Recall	F-Measure	ROC Area	Class
0.97	0.985	0.977	0.983	yes
0.971	0.943	0.957	0.983	no

평가 데이터를 이용하여 C4.5 알고리즘을 적용한 결과 실제 사망자가 41명이었고, 예측 사망자는 33명으로 8명의 차이가 남을 알 수 있다.

```

=== Confusion Matrix ===
 a  b  <-- classified as
52  7 | a = yes
 8 33 | b = no

```

그림 4는 C4.5를 이용하여 차량 사고의 요인을 구한 것이다.

```

If (belt = "yes") And (car_speed = "top") And (accident_cause =
"drug") Then
    RESULT = 사망
Elseif (belt = "no") And (road = "low") And (weather = "no") And
(accident_cause = "outrun" Or accident_cause = "violation") Then
    RESULT = 사망
Else
    RESULT = 생존
End If

```

그림 4. C4.5에 의해서 추출한 룰(RULE)

Fig. 4. Rule that extract that use C4.5.

## VI. 결 론

해마다 교통사고로 인한 인적 혹은 물적인 피해액은 감소하지 않고 있다. 교통사고는 가정이나 국가측면에서 많은 부담이 되고 있다. 이러한 피해를 줄이기 위해서 각 시 및 도에서는 도시별 ITS 프로젝트를 수행하여 효율적인 도로와 신호처리를 이용하여 교통사고를 줄이는 방안을 내놓고 있다.

교통사고의 원인은 운전자의 운전실수가 대부분이다. 그중에서 음주, 약물, 졸음, 과속, 신호위반등은 교통사고의 주된 원인이 되고 있다. 따라서 본 논문에서는 교통사고로 인한 사망자를 예측하는 요인을 알아내는 모델링을 개발하여 교통사고를 줄일 수 있는 기술적이고 환경적인 혹은 운전자에게 지침이 될 만한 요건을 알아내는 것을 목표로 하였다.

이와 같은 목적을 위해서 본 논문에서는 신경망 알고리즘과 C4.5 알고리즘을 이용하여 교통사고로 인한 사

망자를 정확히 예측하고 예측한 결과에 따른 교통사고의 원인을 정확하게 알아내는 모델링을 제안하였다.

## 참 고 문 헌

- [1] <http://www.foxnews.com>, 1998. 5. 15
- [2] 이홍로, 2004년도 교통문화지수 조사 보고서, 교통안전공단, 2004.12.
- [3] Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. Accident Analysis and Prevention, Vol. 27, 40. 4, 1985, pp. 469-481.
- [4] Ossenbruggen, P.J., Pendharkar, J. and Ivan, J. 2001, "Roadway safety in rural and small urbanized areas". Accidents Analysis and Prevention, 33 (4), pp.485 - 498.
- [5] Sohn, S. Y., and Lee, S. H. 2003, "Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity Road Traffic Accidents in Korea". Safety Science, 41(1), pp. 1-14.
- [6] Abdelwahab, H. T. and Abdel-Aty, M. A. 2001, "Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections". Transportation Research Record 1746, Paper No. 01-2234.
- [7] Pieter Adriaans, Dolf Zantinge, Data Mining, Addison wesley longman, 1996.
- [8] 김대수, 신경망 이론과 응용, 진한엠앤비 2005.
- [9] 허명희, 이용구, 데이터마이닝 모델링과 사례, 아카데미 출판사, 2003.7.
- [10] weka, <http://www.cs.waikato.ac.nz/ml/weka/>

## 저 자 소 개



김 천 식(정회원)  
1995년 안양대학교 전자계산학과  
(공학사)  
1997년 한국외국어대학교 컴퓨터  
및 정보통신공학과  
(공학석사)  
2003년 한국외국어대학교 컴퓨터  
및 정보통신공학과  
(공학박사)

2000년~2003년 경동대학교 정보통신공학부 교수  
2004년~현재 안양대학교 디지털미디어학부 교수  
<주관심분야 : 데이터베이스, 데이터마이닝, 유비  
쿼터스, 텔레매틱스, TPEG, DMB, 홈네트워크,  
e-Learning>



정 명 희(정회원)  
1989년 서울대학교 계산통계학과  
졸업.  
1991년 U. of Texas, Austin  
1997년 U. of Texas, Austin  
산업공학과 박사학위  
2006년 현재 안양대학교  
디지털미디어공학과 교수.

<주관심분야 : e-learning, 영상, 멀티미디어>



홍 유 식(평생회원)  
1984년 경희대학교 전자공학과  
(학사)  
1989년 뉴욕공과대학교 전산학과  
(석사)  
1997년 경희대학교 전자공학과  
(박사)

1985년~1987년 대한항공(N.Y.지점 근무)  
1989년~1990년 삼성전자 종합기술원 연구원  
1991년~현재 상지대학교 컴퓨터공학부 교수  
2000년~현재 한국 퍼지 및 지능시스템학회 이사  
2004년~현재 대한 전자 공학회 ITS 분과위원장  
2001년~2003 한국 정보과학회 편집위원  
2001년~2003 한국 컴퓨터 교육산업학회 이사,  
편집위원  
2004년~현재 건설교통부 ITS 전문심사위원  
2004년~현재 원주 시 인공지능신호등 심사위원  
2005년~현재 정보처리학회 이사  
2005년~현재 인터넷 정보학회 이사  
2005년~현재 정보처리학회 강원지부 부회장  
2006년~현재 인터넷 방송통신 TV학회 상임이사  
<주관심분야 : 퍼지 시스템, 전문가시스템, 신경  
망, 교통제어>