



CART 분석을 이용한 교통사고예측모형의 개발

Developing the Traffic Accident Prediction Model using Classification And Regression Tree Analysis

이 재 명* 김 태 호** 이 용 택*** 원 제 무****
 Lee, Jae-Myung Kim, Tae-Ho Lee, Yong-Taeck Won, Jai-Mu

Abstract

Preventing the traffic accident by accurately predicting it in advance can greatly improve road traffic safety. The accurate traffic accident prediction model requires not only understanding of the factors that cause the accident but also having the transferability of the model. So, this paper suggest the traffic accident diagram using CART(Classification And Regression Tree) analysis. developed Model is compared with the existing accident prediction models in order to test the goodness of fit. The results of this study are summarized below. First, traffic accident prediction model using CART analysis is developed. Second, distance(km), pedestrian shoulder(m) and traffic volume among the geometrical factors are the most influential to the traffic accident. Third, CART analysis model show high predictability in comparative analysis between models. This study suggest the basic ideas to evaluate the investment priority for the road design and improvement projects of the traffic accident blackspots.

Keywords : *traffic accident prediction model, classification and regression tree analysis, traffic accident diagram*

요 지

본 연구는 도로기하구조 요인과 교통사고간의 관계를 규명하기 위하여 CART분석을 이용하여 전국의 4차로 국도를 대상으로 교통사고예측모형을 개발하고, 다중회귀모형, 확률회귀모형과 CART분석모형을 비교분석하여 개발한 모형의 적합도를 검증하였다. 연구결과로는 첫째, 변수간의 복합적인 상호관계를 설명할 수 있는 CART분석을 이용하여 국도의 교통사고예측모형을 개발하고 도로기하구조 요인에 따라 표준교통사고율을 의미하는 교통사고발생도표를 제시하였다. 둘째, CART 분석모형에 근거하여 교통사고 발생률에 큰 영향을 미치는 도로기하구조 요인이 구간거리(km), 횡단보도폭(m), 횡단길이(m), 교통량 순으로 나타났다. 셋째, CART분석모형의 적합도 검증결과, CART분석모형이 실제교통사고율을 타 모형에 비해 전반적으로 잘 묘사하고 있었으나, 각 모형별로 교통사고율의 크기에 따라 교통사고율이 비교적 낮은 구간에서는 다중회귀모형이, 평균이상의 교통사고율을 나타내는 구간에서는 포아송 회귀모형의 예측력이 높았으며, CART분석모형은 교통사고율의 크기와 상관없이 우수한 예측력을 보였다. 넷째, 도출된 교통사고발생도표는 도로기하구조 조건에 따른 표준교통사고율을 제시해주기 때문에 도로설계 시에 안전한 기하구조 설계요소 선정기준을 제시할 뿐만 아니라, 교통사고 잦은 지점 개선사업추진 시 사업의 우선순위를 판단할 수 있는 기준을 제시하는 등 정책적 활용도가 매우 높을 것으로 판단된다.

핵심용어 : *교통사고예측모형, CART분석, 교통사고 발생도표, 적합도검증*

* 비회원 · 한양대학교 도시대학원 박사과정 · 주저자
 ** 비회원 · 한양대학교 도시대학원 박사과정
 *** 정회원 · 대한민국 감사원 감사관 · 공학박사 · 교신저자
 **** 비회원 · 한양대학교 도시대학원 원장 · 공학박사



1. 서론

1.1 연구배경 및 목적

2006년 기준 우리나라의 도로교통사고건수는 214,867건으로 이중 부상 340,578명, 사망 6,651명이 발생하여 OECD 29개국 중 27위의 낮은 교통안전수준을 나타내고 있다(건설교통부, 2007). 이러한 교통사고를 획기적으로 절감하기 위해서는 교통사고의 원인을 과학적으로 규명하여 이를 도로상에서 제거하여야 한다.

교통사고의 주요 요인으로는 사회·경제적 요인과 도로환경적 요인, 인적 요인이 존재하고 있다.(Peter D. Loeb, 1994) 그러나 사회·경제적 요인과 인적 요인은 사고와 직접적인 메카니즘을 밝혀내기 어렵고, 이를 찾더라도 개선까지는 많은 비용, 시간 등이 소요된다. 또한 인적 요인관련 자료는 운전자의 개인신상 및 생체정보를 수집하기 어려워 사고분석에 널리 사용되지 않고 있다. 따라서 본 연구에서는 실제 자료수집과 단기 도로설계 및 개선사업에 적용이 용이한 도로환경적 요인을 연구의 범위로 한정하였다.

본 연구에서는 전국의 4차로 국도상의 교통사고 자료와 사고에 영향을 미치는 도로환경적 요인변수를 수집하여, 개별 요인간 상호연관성을 복합적으로 고려할 수 있는 CART(Classification And Regression Tree)분석을 이용한 교통사고예측모형을 개발하고, 도로교통사고 관련 안전성평가를 수행할 수 있는 사고발생도표(Traffic Accident Diagram)를 제시하고자 한다. 아울러 기존 다중회귀모형, 확률회귀모형식과 CART분석모형을 비교분석(Comparative Analysis)하여 개발한 모형의 적합도(Goodness of fit)를 검증하였다.

1.2 연구의 내용 및 방법

연구의 주요내용 및 분석방법은 그림 1과 같다. (1단계) 4차로 국도를 대상으로 교통사고율 및 도로

환경요인관련 자료를 수집하고 기초통계분석을 수행한다.

(2단계) 기존의 교통사고예측모형 및 CART분석기법에 대하여 문헌고찰을 수행한다.

(3단계) CART분석을 이용한 교통사고예측모형을 개발하고 교통사고발생도표를 제시한다.

(4단계) 기존 연구에서 널리 사용되던 선형회귀분석, 확률회귀분석을 이용하여 교통사고예측모형을 개발한다.

(5단계) 기존 교통사고예측분석 모형과 CART 분석모형을 비교분석하여 개발한 교통사고예측모형의 적합도를 검증한다.

(6단계) 모형을 해석하고 정책적 시사점을 도출한다.

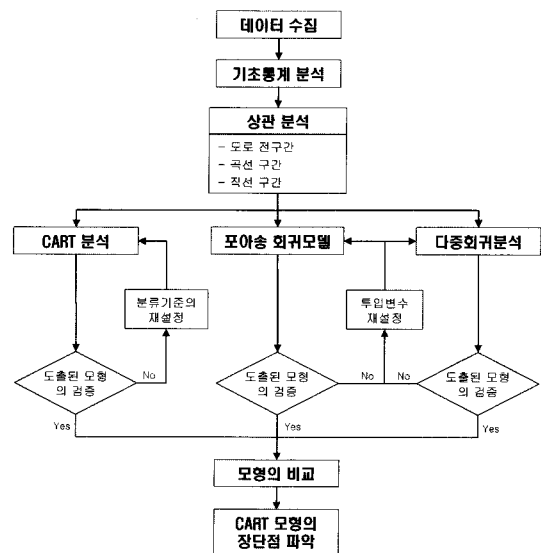


그림 1. 연구의 흐름도

2. 문헌고찰 및 분석방법론 검토

2.1 교통사고관련 문헌고찰

교통사고는 하나의 독립적인 요인으로 발생하기보다는 복합적인 요인이 상호작용을 하여 발생하는 것으로 발생요인을 제거함으로써 교통사고발생률을 현



저히 낮추거나 사전에 방지할 수 있다고 보고되고 있다(H.W. Heinrich, 1990 ; 이창수, 2003).

교통사고 유발변수에 관한 연구는 표 1과 같이 도로기하구조 요인, 차량결함요인이 교통사고에 미치는 영향을 해석하는 연구가 많으며, 교통사고예측모형에 관한 연구는 다양한 분석기법과 사고발생요인을 변수를 이용하여 모형화하고자 노력하고 있다. 분석기법을 살펴보면, 초기 연구에서는 지역적 특성자료를 집합화하여 분석하는 거시적 분석을 도입하고 있으며 분석기법도 주로 선형회귀분석을 이용하고 있다. 선형회귀분석의 경우 교통사고와 같이 산발적으로 발생하는 사건을 설명하기에는 많은 어려움이 있다고 보고된다. 그러나 최근 연구에서는 도로구간의 개별특성 자료를 분석하는 미시적 분석을 도입하고 있으며 분석기법도 확률회귀분석모형(포아송, 음이항 회귀분석) 또는 변수의 비선형 특성을 설명하는데 우수한 인공신경망모형 등을 다양하게 이용하고 있다.

표 1. 선행연구고찰 종합표

구분	저자	년도	종류	내용
분석모형별 관련 연구	Smeed	1949	거시적 분석	사망자와 인구, 자동차대수로 교통사고특성을 규명하려함.
	진교남	1983	미시적 분석	평균속도, 차종간 속도, 경사 비율, 곡선비율 등의 변수를 가지고 교통사고율에 대한 교 통사고 특성을 규명하려함.
	강정규	1985	미시적 분석	고속도로의 기하구조를 중심 으로 교통사고를 규명하기위 해 회귀분석모형을 개발.
	Jovanis & Chang	1986	미시적 분석	차량간의 추돌유형별로 통행 시간과 교통사고수와의 관계 를 규명하기위해 포아송회귀 모형을 개발.
	Adams	1987	거시적 분석	Smeed모형 발전된 것으로 인 구, 자동차대수로 교통사고특성 을 규명하려함.

구분	저자	년도	종류	내용
분석모형별 관련 연구	Miau	1992	미시적 분석	고속도로의 기하구조와 트럭 과 교통사고 사이의 관계를 규 명하기위해 포아송회귀모형을 이용하여 모형을 개발.
	권기중	1993	거시적 분석	한국의 8개 구역의 자료를 이 용 인구밀도, 자동차 보유대수 등의 변수를 가지고 교통사고 특성을 규명하려함.
	공선희 남두희 이용택	2005	미시적 분석	호남고속도의 도로기하구조, 교통운영관련 변수와 교통사 고 사이의 관계를 규명하기위 해 인공신경망 모형을 개발.
변수선 정에 관련 연구	John C	1983	미시적 분석	Illinois, Florida, Ohio와 Texas의 3,557개의 지점에 대하여 곡률반경에 대한 교통 사고특성을 규명하려함.
	Timothy	1983	미시적 분석	평균일교통량(ADT)에 따른 교통사고 특성을 규명하려함.
	Zegeer	1988	미시적 분석	차선폭, 노면폭 및 지형의 영 향에 따른 교통사고 특성을 규 명하려함.
	Bitzel	1988	미시적 분석	독일의 도로를 대상으로 곡선 반경 및 구배별 교통사고율에 대한 교통사고특성을 규명하 려함.
	한국건설 기술원	1990	미시적 분석	전국 국도상의 교통사고다발 지점을 대상으로 곡선반경별 교통사고율에 대한 교통사고 특성을 규명하려함.

그러나 기존 교통사고예측모형은 그림 2와 같이 모형의 설명력을 높이기 위하여 교통사고와의 상관관계가 높은 변수로만 선형회귀모형을 개발하여 다양한 요인을 모형 내에 반영하지 못하거나, 교통사고의 비선형적 특성을 반영하기 위해 비선형모형을 이용하더라도 교통사고관련 자료의 분산이 매우 커서 모형의 적용성에는 한계를 안고 있다. 따라서 본 연구에서는 수집자료의 특성에 따라 분류하고 변수의 비선형성을 모형개발전에 확인이 가능하며, 변수간의 상호교호작용을 고려할 수 있는 CART 분석을 이용하여 교통사고예측모형을 개발하고자 한다.

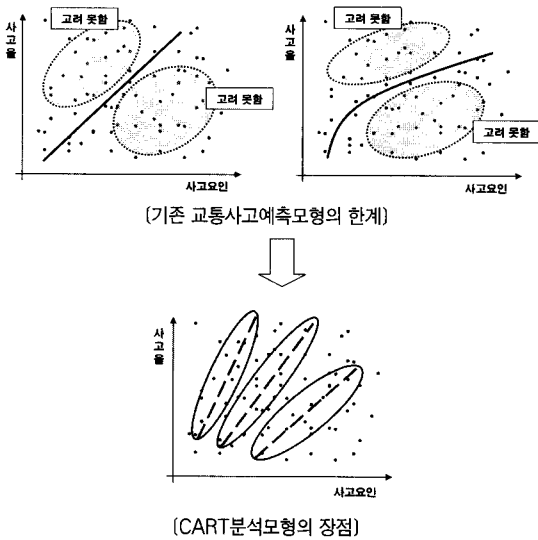


그림 2. 교통사고예측모형 개발을 위한 개념도

2.2 분석방법론 검토 : CART분석

CART분석이란 데이터마이닝 기법(의사결정나무법: Decision Tree Analysis)의 하나로서 그룹 내 동질성(Homogeneity)이 높은 그룹을 판별하기위해서 불순도함수(Impurity function)를 사용한다. 동질성이란 데이터의 특성 및 형태가 유사한 것들을 뜻하며, 불순도함수란 데이터들간의 분산 등과 같은 흩어짐의 정도를 의미한다. 이러한 CART분석의 기본 개념을 이용하여 사고발생유형이 유사한 데이터로 그룹을 형성하여 그룹 특성을 분석한다.

2.2.1 이지분리(Binary Split)기준

지니지수(Gini index), 투잉지수(Twoing Index), 엔트로피지수(Entropy Index) 또는 분산(Variance) 감소량을 그룹의 분리기준으로 사용한다. (Breiman et al, 1984)

1) 지니지수(Gini Index)

지니지수는 각 마디(Node)에서 불순도(impurity)를 재는 측도 중의 하나로써 동질성이 높은 변수를 그

룹핑(Grouping)할 때 가장 보편적으로 사용되는 기준으로 산출식은 식 (1)과 같다.

$$G = \sum_{j=1}^c P(j)(1 - P(j)) = 1 - \sum_{j=1}^c P(j)^2 = 1 - \sum_{j=1}^c (n_j/n)^2 \quad \text{식 (1)}$$

2) 투잉지수(Twoing Index)

투잉지수는 주로 종속변수의 범주수가 큰 경우에 사용하는 기준으로 지니지수와 유사한 나무모양으로 최종마디(terminal node)를 도출하며 산출식은 식 (2)와 같다.

$$T = \frac{P_L P_R}{4} \left[\sum_{i=1}^c |P_L(i) - P_R(i)| \right]^2 \quad \text{식 (2)}$$

$$= \frac{1}{4} \left(\frac{n_L}{n} \right) \left(\frac{n_R}{n} \right) \left[\sum_{i=1}^c \left| \frac{n_{iL}}{n_L} - \frac{n_{iR}}{n_R} \right| \right]^2$$

3) 엔트로피지수(Entropy Index)

엔트로피지수는 비교적 최근에 알려졌으며, 다항분포(Multi-nomial distribution)에서 우도비 검정통계량(Likelihood ratio test statistic)의 극대화를 추구하는 것으로 산출식은 식(3)과 같다.

$$\text{Min}(P(1), P(2)) - \sum_{j=1}^c P(j) \log_2 P(i) \quad \text{식 (3)}$$

4) 분산의 감소량

종속변수가 연속형일 때는 각 마디의 불순도를 측정하는 기준으로 산출식은 식 (4)와 같다.

$$V = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{식 (4)}$$

2.2.2 가지치기(Pruning)

나무의 구조에서 적당하지 않은 마디를 제거하는 방법으로 최종적인 의사결정나무의 설명력을 향상시키는 과정이다.

1) 표준오차(Standard Error)

표준오차규칙은 T번째 마디에서의 분리실패율이



$R(T)$ 라 하고, 이때의 표준오차를 $SE(R(T))$ 라 할 때, T_k 번째 마디의 가지치기 여부를 결정하는 조건식은 식 (5)와 같다.¹⁾ 즉 가지치기를 하더라도 더 이상 마디 분리의 실패율이 감소하지 않을 때 그 노드를 제거하는 방식이다.

$$R(T_k) \leq R(T_0) + \alpha SE(R(T)) \quad \text{식 (5)}$$

2) 최소위험(Minimum Risk)

최소의 예측위험방법은 앞서 나무 전체의 표준오차를 이용한 마디의 $R(T)$ 값을 이용한 것과 달리 단 순 최종마디의 $R(T)$ 값을 이용하여 제거여부를 판단하는 방법이다.

표 2. 수집자료에 대한 기술통계분석 결과

구 분	조 사 항 목	통계결과값	
		평균값	표준편차
기하구조	설계속도 (60~80 km/h)	78.37	5.44
	길어깨 (0~6 m)	1.91	0.84
	보도폭 (0~4.5 m)	0.26	0.89
	교차로개수 (0~10 개)	1.56	2.39
	구간거리 (0.5~ 6.1km),	3.2	1.23
	횡단보도(유: 1, 무: 0)	-	-
교통운영	교통량 (2,211~76,018 대/일)	23,307	13,174
교통사고	100만대당 교통사고율 (0.06~4.50)	0.86	0.75

3. 자료수집 및 모형의 개발

3.1 자료의 수집 및 정리

전국의 4차로 국도상에서 2003년도 건설교통부의 교통사고 잦은 지점으로 선정된 190개소의 100만 대당(Million Entering Vehicle : MEV) 교통사고율²⁾과 도로기하구조³⁾ 및 교통량⁴⁾자료를 표 2와 같이 수집하였다.

수집자료에 대한 기술통계분석결과, 표 2에서 나타난 바와 같이 총교통사고율의 경우 평균값(Mean)이 0.86으로 나타났다. 주요 기하구조별로는 구간거리는 3.2km, 길어깨 폭은 1.91m, 교통량은 23,307 대/일 인 것으로 분석되었다.

또한 이렇게 수집된 190개 자료 중 무작위로 178개를 추출하여 모형의 개발에 사용하고, 나머지 12개 자료는 모형의 적합도 검증에 사용한다.

3.2 모형의 개발

3.2.1 CART분석모형의 개발

CART분석을 실시한 결과, 각각의 노드에서 분석한 자료의 분산에 대한 증분(Improvement)이 감소하는 이진분리기준을 적용하여, 그림 3과 같이 최종 5개 노드 형태로 분할되는 「교통사고발생도표」를 도출하였다. 그림 3의 교통사고발생도표의 좌측 하단에 제시된 위험도표를 살펴보면, 전체 최종마디의 분산(within terminal node variance)을 나타내는 것으로 개발모형이 하나의 관찰치를 오분류(Misclassification)할 확률이 35.18%로, 이는 CART분석을 통하여 도출되어진 회귀분류나무의 분류능력이 64.82%(100-35.18%)라는 것을 의미한다. 5개 최종 노드의 요약결과를 나타내는 표 3을 살펴보면 4번 노드(구간거리가 1.91km보다 작고, 횡단보도폭이 3.25m보다 큰 곳)가 가장 높은 사고율을 가지는 것으로 나타났다. 이러한 최종노드의 의사결정 규칙은 표 4와 같다.

1) 식 (5)에서 α 값은 분리실패율과 표준오차를 보정하기 위한 가중계수로서 연구자의 판단에 따라 0값을 지정할 수도 있다.

2) 교차로 교통사고 잦은 지점 분석에서는 100만대당 교통사고율 $\left(\frac{\text{교통사고건수} \times 1,000,000}{365 \times \text{연평균 일평균 교통량}} \right)$ 을 널리 사용하고 있다(도철용, 1998).

3) 도로기하구조자료는 「중앙분리대 종합기본계획 수립연구, 2003. 6. 건설교통부」에 제시한 수집자료이며, 각 구간별 사고 데이터의 경우 전국 국도유지관리 사무소, 도로교통안전관리공단, 도로교통량통계연보에서 수집한 전국 4차로 사고데이터를 이용하였다. 이렇게 수집되어진 자료의 범주 내에서 문헌교찰을 통하여 교통사고에 중요한 영향을 미치는 변수를 1차 선별하고 이중 유의한 통계치를 나타내는 설명변수를 선정하여 모형을 개발하였다.

4) 교통량은 연평균일교통량(Annually Average Daily Traffic : AADT)로 2003년도 도로교통량통계연보(한국건설기술연구원, 2004)자료를 적용하였다.

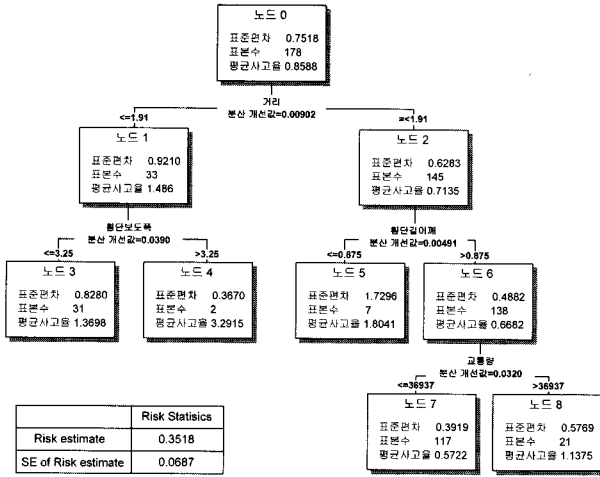


그림 3. CART분석 결과 (교통사고발생도표)

표 3. 선정된 노드의 주요 통계치

노드번호	교통사고율	표준편차	자료개수	자료비율 (단위:%)
3	1.3698	0.8180	31	17.42
4	3.2915	0.3670	2	1.12
5	1.8041	1.7296	7	3.93
7	0.5722	0.3919	117	65.73
8	1.1375	0.5769	21	11.80
합 계	-	-	178	100.00

표 4. 노드별 마디의 의사결정규칙

노드 번호	의사결정 규칙
1	IF (구간거리 ≤ 1.91)
2	IF (구간거리 > 1.91)
3	IF ((구간거리 ≤ 1.91) AND (횡단 보도폭(m) ≤ 3.25))
4	IF ((구간거리 ≤ 1.91) AND (횡단 보도폭(m) > 3.25))
5	IF ((구간거리 > 1.91) AND (횡단 길어깨(m) ≤ 0.875))
6	IF ((구간거리 > 1.91) AND (횡단 길어깨(m) > 0.875))
7	IF ((구간거리 > 1.91) AND (횡단 길어깨(m) > 0.875) AND (교통량(대/일, 양방향) > 36936.5))
8	IF ((구간거리 > 1.91) AND (횡단 길어깨(m) > 0.875) AND (교통량(대/일, 양방향) > 36936.5))

3.2.2 다중회귀모형

CART분석모형과 비교분석을 위하여 위와 동일한 172개소의 교통사고율과 도로기하구조 요인변수를 이용하여 다중회귀모형을 개발하였다.

회귀분석방법으로는 Stepwise 방식을 사용하였으며, 산출된 다중회귀모형식은 식 (5)와 같다. 추정된 모형식은 R²값이 26%로 비교적 낮게 나타났으며 독립변수의 t값은 횡단보도폭(m)을 제외한 모든 변수가 유의수준(α) 5%내에서 통계적으로 유의한 것으로 나타났다.

$$Y = 4.12 + 0.08X_1 + 1.1799E^{-6}X_2 - 0.04X_3 - 0.18X_4 - 0.05X_5 + 0.04X_6 \quad \text{식 (5)}$$

(4.16) (1.17) (3.02) (-3.03) (-2.92) (-2.93) (2.62)

여기서, X₁: 횡단보도폭(m), X₄: 횡단 길어깨(m)
 X₂: 교통량(대/일, 양방향), X₅: 구간거리(km)
 X₃: 설계속도(km/h), X₆: 교차로 개수
 Y: 100만 대당(MEV) 교통사고율

3.2.3 확률회귀모형 (포아송회귀모형)

CART분석 모형과 비교분석을 위하여 위와 동일한 172개소의 교통사고율과 도로기하구조 요인변수를 이용하여 확률회귀모형식을 개발하였다.

확률회귀모형을 포아송과 음이항모형 형태로 분석한 결과, 과분산계수(K) 값이 0.02로 0에 가까운 값을 나타내어 포아송회귀모형이 적합한 것으로 나타났다. 또한 산출된 포아송회귀식은 식 (6)과 같으며, ρ²값이 0.21로 적정하고, 독립변수 t값 또한 통계적으로 모두 유의한 것으로 나타났다.

$$Y = \exp(5.48 - 0.03X_1 - 0.09X_2 - 0.14X_3 + 0.09X_4 + 0.42X_5) \quad \text{식 (6)}$$

(29.89) (-17.04) (-15.94) (-13.36) (8.69) (12.69)

여기서, Y: 100만 대당(MEV) 교통사고율, Y/100 형태로 적용
 X₁: 설계속도(km/h), X₂: 구간거리(km)



X_3 : 횡단길어깨(m), X_4 : 횡단보도폭(m)
 X_5 : 일교통량(대/일, 양방향), $\log(X_6)$ 형태로 적용

표 5. 교통사고예측모형의 적합도 검증 결과

군집구분	검증자료	①	②	③	④	추정교통사고율 -실제교통사고율		
						①-④ (오차)	②-④ (오차)	③-④ (오차)
1	1	0.575	1.008	0.572	0.588	0.013	-0.420	0.016
	7	0.415	0.789	1.804	0.546	0.131	-0.243	-1.259
	11	3.345	4.032	1.804	3.317	-0.028	-0.715	1.513
2	6	0.595	0.654	1.369	0.900	0.305	0.246	-0.469
	10	0.540	0.869	0.572	1.179	0.639	0.310	0.606
	12	0.572	1.031	1.138	0.887	0.315	-0.144	-0.251
3	2	0.540	0.869	1.369	1.179	0.639	0.310	-0.190
	3	0.514	0.941	1.369	1.193	0.679	0.252	-0.176
	4	0.225	0.725	1.138	1.255	1.030	0.531	0.118
	5	0.345	0.662	0.572	0.593	0.248	-0.069	0.021
	8	1.965	2.318	0.572	0.876	-1.090	-1.442	0.303
	9	0.720	1.110	1.804	2.872	2.152	1.762	1.068

주) ① 다중회귀모형 예측교통사고율(MEV)
 ② 포아송회귀모형 예측교통사고율(MEV)
 ③ 사고발생도표 예측교통사고율(MEV)
 ④ 실제교통사고율(MEV)

4. 모형의 검증 및 결과해석

4.1 모형의 검증

CART분석모형의 적합도를 검증하기 위해서, 12개의 검증자료를 다중회귀모형, 확률회귀모형, CART분석모형에 각각 입력하여 얻은 예측교통사고율과 실제교통사고율의 오차를 비교분석하였다. 이때 모형의 장단점을 파악하기 위하여 모형검증데이터를 군집분석하여 총 3개의 그룹으로 구분하였다. 3개의 군집은 곡선반경, 교차로개수, 교통량 및 총사고율을 기준으로 구분되어 졌다.

분석결과 교통사고율이 낮고 교차로개수가 작은 형태의 1군집에서는 다중회귀모형의 설명력이 높고, 사고율이 높고 곡선반경이 짧은 형태의 2군집에서는 포아송 회귀모형의 예측력이 높은 것으로 나타났다. 그리고 일반적인 평균사고율과 일반적인 도로기하구조 형태를 가지는 3군집에서는 CART분석모형이 오차가 작은 것으로 나타났다. 결과적으로 다중회귀모형식은 교통사고율이 비교적 낮은 구간에서, 포아송 회귀모형식은 평균보다 높은 교통사고율을 보이는 구간에서 예측력이 높은 것으로 나타난 반면, CART 분석모형은 교통사고율의 높고 낮음에 상관없이 일반적인 도로구간에서의 예측력이 높은 것으로 나타났다.

그러나 CART분석모형은 함수식에 의해 하나의 예측교통사고율을 제시하지 않고 의사결정규칙에 의한 특정범주(range)내에서 동일한 예측교통사고율을 나타내기 때문에 다중회귀모형, 확률회귀모형처럼 변수에 민감한(sensitive) 예측교통사고율을 제시하지는 못하는 단점을 내포하고 있는 것으로 나타났다.

4.2 모형의 결과해석

CART분석의 모형에서 도출된 교통사고발생도표를 토대로 교통사고에 큰 영향을 미치는 도로의 기하구조요인을 선정하면, 나무의 분류 순서를 토대로 구간거리(km), 횡단보도폭(m), 횡단길어깨(m), 교통량 순으로 나타났다.

이는 앞서 도출되어진 주요 사고요인항목은 강정규(1986), Miao(1992), 공선희 외(2005)의 선행 연구에서 제시한 연구결과와 일치하는 것으로 나타났다. 그러나 교통사고요인의 영향력을 살펴보면, 교통사고에 영향을 미치는 중요요인에는 교통사고예측 모형별로 다소 차이가 있는 것으로 나타났는데, 다중회귀모형에서는 횡단길어깨(m)가, 포아송회귀모형에서는 횡단길어깨(m), 교통량이 중요요인으로 선정되었다.



이는 총사고율을 설명하는 요인이 독립적이라는 가정하에서 도출된 회귀모형식과 달리 교통사고요인 간의 상호관계성을 고려하여 주요 요인을 선정하는 CART분석모형의 특성 때문에 다른 결과가 도출되었다고 판단된다.

5. 결론 및 향후 연구과제

5.1 결론

지금까지 교통사고가 유발요인의 상호작용으로 인하여 발생한다는 기본가정하에 도로환경(도로기하구조)요인과 교통사고간의 관계를 규명하기 위하여, CART분석을 이용하여 전국의 4차로 국도를 대상으로 교통사고예측모형을 개발하고, 이에 근거하여 교통사고발생도표를 제시하였다. 또한 CART분석모형의 적합도를 검증하기 위하여 다중회귀모형, 확률회귀모형을 개발하고 동일한 검증자료를 이용하여 예측교통사고율과 실제교통사고율의 오차를 비교분석하였다.

이상의 연구결과를 요약하면 다음과 같다.

첫째, 변수간의 복합적인 상호관계를 설명할 수 있는 CART분석을 이용하여 국도의 교통사고예측모형을 개발하였다. CART분석을 통하여 도로기하구조요인 선정에 따라 표준교통사고율을 제시할 수 있는 5개 마디를 가진 교통사고발생도표가 산출되었다.

둘째, CART분석모형을 이용하여 교통사고발생에 큰 영향을 미치는 도로기하구조요인을 이해할 수 있는데, 이러한 요인으로는 구간거리(km), 횡단보도폭(m), 횡단길어깨(m), 교통량 순으로 영향을 주는 것으로 나타났다.

셋째, CART분석모형의 적합도 검증결과, CART분석모형이 타 모형에 비해 전반적으로 실제교통사고율을 잘 묘사하고 있었으나, 각 모형별로 교통사고율의 크기에 따라 예측력이 다르게 나타났다. 즉 교통사고율이 비교적 낮은 구간에서는 다중회귀모형

이, 평균이상의 교통사고율을 나타내는 구간에서는 포아송 회귀모형의 예측력이 높았으며, CART분석모형은 교통사고율의 크기와 상관없이 우수한 예측력을 보였으며 도로기하구조별로 표준교통사고율을 제시하기 때문에 설계지침(Design Guideline) 상에 범용적으로 사용이 가능하다.

넷째, CART분석모형을 통해 도출된 교통사고발생도표는 도로기하구조 조건에 따라 표준적인 장래 도로교통사고율을 제시해주기 때문에, 도로설계 시에 안전한 기하구조 설계요소 선정기준을 객관적으로 제시할 뿐만 아니라, 교통사고 잦은 지점 개선사업 추진 시 사업의 우선순위를 판단할 수 있는 기준을 제시하는 등 정책적 활용도가 매우 높을 것으로 판단된다.

5.2 향후 연구과제

향후 연구과제는 다음과 같다.

첫째, 4차로 국도 외에도 도시부도로, 다차선도로, 고속도로 등 다양한 도로시설에 대한 다양한 교통사고자료(Traffic Accident Database)를 수집하여 모형을 개발할 필요가 있다. 이를 위해서는 국내 도로에 대한 과학적인 도로교통 데이터베이스 구축이 선행되어야 할 것으로 판단된다.

둘째, CART분석모형의 경우, 도로기하구조별로 표준교통사고율을 제시할 수 있어, 정책적인 활용도가 높으며, 국내 다양한 실증연구가 지속적으로 이루어져야 한다.

셋째, 교통사고예측과정에서 요인간의 상호작용에 관한 이해는 매우 중요한 부분임으로 도로기하구조요인외에도 교통사고에 영향을 미치는 타 요인에 대한 고려가 필요하다.

마지막으로 복합적인 인과관계를 설명하기 위해서는 본 연구에서 검토된 모형외에도 구조방정식 등 새로운 모형의 개발과 적용에 대한 추가적인 연구가 필요하다.



참고문헌

공선희, 남두희, 이용택 (2003), 유전감독-인공신경망이론을 이용한 고속도로 교통사고모형, 대한토목학회 논문집 제23권 제4D호

도철웅 (1998) 교통공학원론(하), pp724

이수범, 김정현, 홍다희, 유창남 (2003), 도로등급 및 특성에 따른 교통사고예측모형개발, 대한교통학회지

이창수, 이승일, 김찬호, 조우현 (2003), 국가재해관리와 도시계획, 국토도시계획학회지.

이영섭 (2003), 회귀의사결정에서의 관심노드 찾는 분류 기준법, 응용통계연구학회지.

이기영, 이용택 (2004), 확률회귀모형을 이용한 고속도로의 사고요인분석, 도로교통 제94호.

임용빈, 오만숙 (2000), 분류와 회귀나무분석에 관한 소고, 품질경영학회지.

Oh, J., Washington, S.P., and Choi, K (2004), Development of accident prediction models for rural highway intersections, *Proceedings of the TRB Annual Meeting*.

Poul Greibe (2003), 『Accident prediction model for urban roads』, *Accident Analysis and Prevention*.

So Young Sohn, Sung Ho Lee (2003), Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, *Safety Science*.

Mohammed El-Sadig, J. Nelson Norman, Owen L. Lloyd, Peter Romilly, Abdulbari Bener (2002), Road traffic accidents in the United Arab Emirates: trends of morbidity and mortality during 1977-1998, *Accident Analysis and Prevention*, 2002.

Matthew G. Karlaftis, Ioannis Golias (2002), Effects of road geometry and traffic volumes on rural roadway accident rates, *Accident Analysis and Prevention*.

Mohamed A., Abdel-Aty, A. Essam Radwan (2000), Modeling traffic accident occurrence and involvement, *Accident Analysis and Prevention*.

Peter D. Loeb (1994), *Cause and Deterrants of Transportation Accidents*, Quorum book.

접 수 일: 2007. 10. 20
 심 사 일: 2007. 10. 29
 심사완료일: 2008. 3. 6