

다수의 결과를 고려한 한국의 도로교통사고 연관규칙분석

손소영[†] · 오기열 · 신형원

연세대학교 컴퓨터과학 · 산업시스템공학과

Association Rules for Road Traffic Accident in Korea with Multiple Outcomes

So Young Sohn · Ki Yeol Oh · Hyoung Won Shin

Department of Computer Science and Industrial Systems Engineering, Yonsei University, Seoul, 120-749

In many cases, the result of a road traffic accident can be described with more than one response variables. Nonetheless, most of the existing road accident data analysis deal with only one response variable and try to explain why it occurs. In this paper, we train association rules for a set of more than two response variables conditional on personal, environmental and vehicular/behavioral aspects of accident. Association rules are derived at 8% support and 70% confidence from the 1996 data of three police stations in Korea. We expect that these rules can contribute to effective safety practice in Korea.

Keywords: data mining, association rule, road traffic accident data

1. 서론

1978년에 한국의 자동차 등록대수는 38만대 수준이었으나, 1985년에 백 만대를 돌파한 이후, 1997년에는 천만대의 자동차가 등록되었고, 현재 1,200만대의 자동차가 등록되어 운행되고 있다. 그러나, 교통사고로 인해 발생하는 인명피해의 규모는 세계적인 수준이다.

OECD에 가입한 29개국 중 자동차 1만대 당 사망자수는 8명으로 세계에서 가장 높고, 인구 10만 명당 사망자 수는 19.8명으로 세계에서 세 번째로 높으며 OECD 가입국들의 평균 교통사고율의 3배를 넘고 있다. 이런 상황에서 교통사고 예방을 위한 다방면의 대책이 필요하다.

교통사고 예방을 위한 연구들이 다양한 측면에서 시도되고 있다. Lee & Lim(1990)은 한국의 1982-1989년의 전국의 교통사고자료에서 전체 인구수, 자동차 보유 대수, 운전면허 소지자 수, 도로 연장거리, 교통 경찰관 수, 국민 총생산 등의 자료를 회귀분석에 이용하여 교통사고의 위험도를 예측하였다. Oh

and Ko(1992)은 대형 교통사고의 요인들 중 인적 요인에 속하는 운전자의 법규 위반 유형과 가해 운전자의 사망여부를 인적, 차량적, 그리고 도로 환경 요인들에 대한 판별함수를 이용해 분석하였다. Kim(1993)은 교통사고 통계원표의 항목 중 인적 요인으로 음주, 과속운전 여부, 보행자의 무단 횡단에 따른 사망정도가 차이가 있는지를 대수선형 분석(log-linear analysis)으로 비교하였다. Sohn and Shin(1998)은 교통사고 심각도 분류 분석을 함에 있어 전형적인 데이터 마이닝 기법인 신경망(neural network)과 Decision-Tree, 로지스틱 회귀분석(logistic regression)을 이용하였다. 이들은 교통사고통계원표에 기록된 여러 가지 범주형 설명변수들을 고려하여 사고 심각도를 3가지 범주(치명적 상해, 경미한 상해, 물적 피해)와 2가지 범주(신체상해, 물적 피해)로 분류하고 기법간의 분류정확도를 비교 분석하였다. Sohn and Lee(2001)는 위의 자료를 이용하여 분류 정확도를 개선하고자 신경망, Decision-Tree를 이용하여 데이터 융합(Data Fusion), 앙상블(Ensemble), 그리고 클러스터링(Clustering)기법을 이용하였다.

[†] 연락처: 손소영 교수, 120-749 서울시 서대문구 신촌동 134 연세대학교 컴퓨터과학 · 산업시스템공학과, Fax: 02-2123-4014, e-mail: sohns@yonsei.ac.kr

2002년 1월 접수, 2회 수정 후 2002년 8월 게재 확정.

위의 연구에서 볼 수 있듯이 교통사고의 사고심각도 분류분석을 위해 판별분석, 신경망, Decision-Tree, 로지스틱 회귀분석, 회귀분석등의 분류 기법들이 적용되고 있다. 그러나, 위의 분류 기법들은 교통사고자료를 분석함에 있어서 여러 개의 입력 변수들을 이용하여 하나의 목표변수를 예측하거나 그와의 상관관계를 보고있다. 이러한 기존 연구들에는 두 개 이상의 목표변수와와의 연관성을 살펴볼 수 없다는 단점이 존재한다. 따라서, 본 논문에서는 연관규칙(Association rule)을 활용하여 교통사고 후 피해 심각도와 피해부위와 상태를 나타내는 두 개 이상의 변수들과 인적, 환경요인, 차량/행동유형 변수들간의 연관규칙을 살펴보았다.

본 논문에 사용된 자료는 1996년 서울시내의 3개 경찰서의 관할구역에서 발생한 11,564건의 교통사고자료로 각 자료는 79개의 변수로 구성되어 있다. 논문의 구성은 다음과 같다. 2장에서는 연관규칙의 이론을 소개하였고 3장에서는 교통사고 자료의 79개의 변수를 인적, 환경적 요인, 차량/행동유형으로 구분하여 네 개의 목적변수의 관점에서 도로교통사고자료에 연관규칙을 적용한 결과를 보였다. 4장에서는 결론과 연구방향을 제시하였다.

2. 연관규칙(Association rule)

컴퓨터의 보급 이후 데이터베이스는 의학, 공학, 행정, 금융, 증권 및 공장 등의 여러 응용분야에서 빠른 속도로 발전 및 확산되고 있다. 정보가 다양해지고 대용량화됨에 따라 현실세계에서 유용하게 사용될 수 있는 많은 정보들이 있으나 분석되지 못하고 사장되는 경우가 많다 (Kim *et al.* 1997). 이를 해결하기 위해서 대용량의 사건들이 기록되어 있는 데이터베이스에서 자주 발생하는 항목집합 간의 상호연관성을 발견하는 연관규칙(association rule) 탐사 방법이 사용될 수 있다.

연관규칙 분석은 하나의 거대나 사건에 포함되어 있는 둘 이상 항목들의 상호 관련성을 발견하는 것이다. 일반적으로 연관성 분석은 조건부확률에 대한 개념을 기반으로 하고 있다. 데이터 마이닝에서 연관규칙을 분석하는 주요 목적은 항목간에 연관성을 파악하는데 있으며 이를 위해 2가지의 중요한 평가개념인 지지도와 신뢰도에 대해서 알아보면 다음과 같다.

$I = \{i_1, i_2, \dots, i_k\}$ 를 항목들의 전체집합이라 할 때, 항목집합은 I 의 부분집합을 말한다. X 와 Y 가 항목집합일 때 연관규칙은 $X \Rightarrow Y$ 의 형태로 표현하며, 한 트랜잭션 안에 X 가 발생할 때 Y 가 발생한다는 것을 의미한다. 여기서 $X \subseteq I, Y \subseteq I$, 그리고 $X \cap Y = \emptyset$ 이다.

연관규칙 $X \Rightarrow Y$ 의 지지도는(support) 전체 트랜잭션의 개수 L 에서 X 와 Y 가 동시에 포함된 개수의 비율을 의미하며 다음과 같이 계산된다.

$$S = \frac{freq(X \cap Y)}{L}$$

신뢰도(confidence)는 항목 X 를 포함하는 트랜잭션 중 Y 가 포함된 개수의 비율을 나타낸다.

$$C = \frac{freq(X \cap Y)}{freq(X)}$$

이와 같이 계산되는 지지도와 신뢰도를 교통사고 분석에 사용하기 위하여 교통사고 한 건은 한 트랜잭션으로 볼 수 있고, 교통사고 한 건을 설명하는 변수들은 한 트랜잭션 내의 각각의 항목이 된다. 교통사고의 원인을 항목집합 X 라 할 때, $X = \{X_1, X_2, \dots, X_N\}$ 는 N 개의 항목을 가지고, 사고결과를 Y 라 할 때 $Y = \{Y_1, Y_2, \dots, Y_M\}$ 는 M 개의 항목을 가진다. $Pr(X_1 \cap X_2 \cap \dots \cap X_N \cap Y_1 \cap Y_2 \cap \dots \cap Y_M)$ 는 사고원인 X_1, X_2, \dots, X_N 과 결과 Y_1, Y_2, \dots, Y_M 가 동시에 일어날 확률을 말한다. 교통사고 통계원표의 자료에서 $X \Rightarrow Y$ 의 지지도는 $(X_1, X_2, \dots, X_N$ 의 상황하에 결과 Y_1, Y_2, \dots, Y_M 이 동시에 발생한 사고의 수)/(전체 교통사고 수)로 정의된다.

한편 신뢰도는 사고 원인 X_1, X_2, \dots, X_N 의 상황에서 교통사고가 발생했을 때 사고결과 Y_1, Y_2, \dots, Y_M 가 발생할 확률 $Pr(X_1, X_2, \dots, X_N | Y_1, Y_2, \dots, Y_M)$ 로 $(X_1, X_2, \dots, X_N$ 과 같은 상황하에 결과 Y_1, Y_2, \dots, Y_M 이 동시에 발생한 사고의 수)/(사고원인 X_1, X_2, \dots, X_N 을 포함하는 사고의 수)로 계산된다.

이 지지도와 신뢰도의 개념은 연관규칙 탐사에서 매우 중요한 개념으로 빈번한 항목집합(frequent item set)을 발견하는 데 큰 역할을 한다.

3. 교통사고 연관규칙 추출

본 연구에서는 교통사고 통계원표상의 변수 79개중 목적변수(target variable)로 사고상태를 나타낼 수 있는 사고정도, 신체상해 주부위, 신체상해 부위상태, 신체가해 주부위로 4개 변수를 선택하였다. 사고의 인적요인을 설명하는 변수로서 성별, 학력, 연령, 용도별, 당사자 종별, 면허유무의 6개 변수를 선정하였다. 환경요인으로 계절, 요일, 주야, 기상, 지형, 신호기, 도로선형, 노면상태, 도로형태, 차도폭 등 10개 변수를 선정하였다. 차량/행동유형으로 보호장구, 배기량, 사고직전속도, 행동유형 등 4개의 변수를 선택하였다.

교통사고 통계원표에 나타난 자료는 사망 109건, 중상 3,098건, 경상 3,659건, 물적 피해사고 4,629건이 발생한 것으로 확인이 되었으나, 통계원표상에 사고피해부위, 사고피해부위상태들이 세부적으로 표시된 자료는 그 중 90건(0.78%)이 해당되었다. 본 논문에서는 사고요인별 (인적요인, 환경요인, 차량/행동유형)로 4개의 목적변수(사고정도, 신체상해 주부위, 신체상해 부위상태, 신체가해 주부위)의 연관규칙을 도출하였다.

연관규칙 분석을 위해 사용된 90건의 사고는 사망사고 21건, 중상사고 38건, 경상사고 31건으로 나뉜다. 사고 발생일을 기준으로 평일이 57건, 주말이 33건 발생하였다. 주간에 발생한

사고가 43건, 야간에 발생한 사고는 47건으로 주간과 발생빈도가 비슷하다. 사고자의 연령은 30대가 38건으로 42%의 비율을 차지하였으며, 남자가 전체사고 90건 중 89건을 차지하였다. 학력은 고졸이 58건으로 64%를 차지하였다. 날씨는 맑은 날이 68건으로 76%를 차지하였고, 사고지역은 인구밀집한 시가지가 77건으로 전체사고의 86%를 차지하였다. 신호기가 없는 곳은 57건으로 63%를 차지하였으며, 포장되고 건조한 지역은 74건으로 82%, 직선평지도로는 83건으로 92%를 차지하였다. 차량의 배기량은 경차가 35건(38%), 소형차가 50건(56%)을 차지하였다. 보호장구를 착용한 비율은 56건으로 62%를 차지하였다. 머리부상은 28건(31%), 목 25건(28%), 다리 19건(21%)이 발생하였다. 골절이 27건(30%), 탈구염좌가 22건(24%)을 차지하였고, 신체충격부위는 차량내부가 61건으로 68%를 차지하였다.

교통사고 원인과 결과 사이의 효과적인 연관성 분석을 위하여 사고원인을 인적요인, 환경요인, 차량/행동유형으로 나누고 4개의 목적변수(사고정도, 신체상해 주부위, 신체상해 부위상대, 신체가해 주부위)와의 연관규칙을 도출하였다. 도출된 규칙 중 4개의 목적변수 모두에 관련이 있는 연관규칙을 채택하였으며, 임계치(threshold)는 지지도 8%, 신뢰도 70%로 지정하였다. 신뢰도의 한계를 지지도보다 상대적으로 높은 70%로 정한 이유는, 4개의 목적 변수가 모두 기재된 피해심각도가 높은 사고만을 추출하였으므로 사고 요인에 공통된 내용이 많기 때문이다.

전체 352,338개의 연관규칙 중 임계치를 만족하는 연관규칙은 5,993개로 나타났다. 이들 중 본 연구에서 관심을 갖는 사고정도, 신체상해 주 부위, 신체상해 부위상대, 신체가해 주부위를 목적변수로 하는 연관규칙을 추출한 결과 머리골절 사망사고, 다리골절 중상사고, 머리충격 중상사고, 목 탈구염좌 경상사고의 4가지 경우에서 연관규칙이 도출되었다.

다음은 인적요인, 환경요인, 차량/행동유형으로 나누어 도출된 각 경우의 대표적인 연관규칙을 살펴보면 다음과 같다.

3.1 인적요인에 의해 도출된 연관규칙

인적요인은 연령, 성별, 학력, 당사자종별, 면허여부의 5개의 변수로 연관규칙을 도출하였으며 총 1,100개의 연관규칙이 나타났다. 인적요인 중에서 30대 고졸학력을 가진 운전자들의 연관규칙이 자주 나타나는데 이는 운전자중 30대, 고졸학력의 남자들이 차지하는 비율이 상대적으로 높음을 알 수 있다.

차량내부충격시(901) 머리(861)골절(881) 사망사고(641)는 면허소지(721), 30대(233), 고졸(273), 남자(251), 승용차(411), 비사업용(452)의 항목들이 관련있는 것으로 볼 수 있다. 이는 머리골절 사망사고는 1) 면허를 소지하고 비사업용 승용차를 운전하는 고졸학력의 남자들이나 2) 면허를 소지한 30대 고졸학력 소지자에 의해 주로 발생하였다(<표 1> 참고).

<표 1> 에서 신뢰도 100%를 가지는 두 번째 규칙은 면허를

소지한 30대 고졸학력이 인명피해 사고와 관련이 있을 시에는 머리골절 사망사고가 주로 발생한다는 것으로 해석할 수 있다. 이는 분석에 사용된 90건의 자료가 인명피해가 발생한 사고로 제한되었으며, 그 이외의 사고정도, 신체상해 주부위, 신체상해 부위상대, 신체가해 주부위 중 한가지의 항목이라도 확인할 수 없는 사고와 인명피해가 발생하지 않은 사고는 분석대상에서 제외되었으므로 100%의 신뢰도가 발생한 것으로 보인다. 머리골절 사망사고는 23개의 연관규칙이 추출되었다. 표.1에서 2 가지의 연관규칙만을 기술한 이유는 나머지 21가지의 연관규칙, 즉 면허를 소지한 남자의 항목보다 면허를 소지한 30대 남자의 항목이 더 자세하고, 특별한 상황을 설명하므로 더 범주가 큰 면허를 소지한 남자의 항목은 제시하지 않았다.

표 1. 머리골절사망사고의 인적요인 연관규칙

신뢰도	지지도	연 관 규 칩
83.33	8.89	721, 452, 411, 273, 251 → 901, 881, 861, 641
100.0	8.89	721, 273, 233 → 901, 881, 861, 641

차량외부충격(902)시 다리(867)골절(881) 중상사고(642)와 관련된 중상사고는 15개의 연관규칙이 도출되었다. 고졸(273)학력의 무면허(722)의 남자(251)가 이륜 및 기타차량(414)과 관련된 사고가 발생할 때 차량외부 충격시 다리골절 중상사고가 주로 발생한다. 이륜차는 사고 시에 충격이나 넘어지면서 신체가 보호를 받지 못하고, 지면이나 구조물에 마찰 및 충격에 의한 피해를 입는 것이 추정된다.

표 2. 다리골절중상사고의 인적요인 연관규칙

신뢰도	지지도	연 관 규 칩
72.73	8.89	722, 414, 273, 251 → 902, 881, 867, 642

차량외부충격시(902) 머리(861)충격(883) 중상사고(642)는 7개의 연관규칙이 도출되었고, 고졸(273)학력의 무면허(722)의 남자(251)가 운전시에 주로 발생한다.

표 3. 머리충격중상사고의 인적요인 연관규칙

신뢰도	지지도	연 관 규 칩
71.43	8.89	722, 273, 251 → 902, 883, 861, 642

차량내부충격시(901) 목(863) 탈구염좌(884) 경상사고(643)는 15건의 연관규칙이 도출되었으며, 운전면허를 가진(721), 30대(233), 고졸(273), 남자(251)의 항목이 관련이 있다. 운전면허를 가진 30대, 고졸의 남자에 의한 인명피해 사고가 발생했을 경우에 차량내부충격시 목 탈구염좌 경상사고가 주로 발생한다

고 볼 수 있으며 이는 급정거나 충돌에 의한 머리충격으로 해석할 수 있다.

표 4. 목 탈구염좌 경상사고의 인적요인 연관규칙

신뢰도	지지도	연 관 규 칙
71.43	8.89	721, 273, 251, 233 → 901, 884, 863, 643

3.2 환경요인에 의해 도출된 연관규칙

환경요인은 계절, 요일, 주야, 기상, 지형, 신호기, 도로선형, 노면상태, 도로형태, 차도폭등 모두 10개의 변수의 연관관계를 살펴보았다. 그 결과 3,381개의 연관규칙이 도출되었다. 요일의 변수는 평일과 주말로 나누어 연관규칙을 도출하였다.

차량내부충격시 머리골절 사망사고는 30건의 연관규칙이 도출되었으며, 맑음(291), 신호기 없음(365), 직선평지(371), 포장건조(381), 단일로(332)의 항목들이 관련 있는 것으로 나타났다. 머리골절 사망사고는 맑은 날에 포장되고 건조한 직선도로의 단일로 중 신호기가 없는 곳에서 발생할 가능성이 높다. 그러므로 신호기가 사망사고의 발생에 큰 연관성을 가짐을 알 수 있다.

표 5. 머리골절사망사고의 환경요인 연관규칙

신뢰도	지지도	연 관 규 칙
78.57	12.22	381, 371, 365, 332 → 901, 881, 861, 641
78.57	12.22	381, 365, 332, 291 → 901, 881, 861, 641
83.33	11.11	381, 371, 365, 291 → 901, 881, 861, 641
76.92	11.11	381, 371, 332, 291 → 901, 881, 861, 641
83.33	11.11	371, 365, 332, 291 → 901, 881, 861, 641

차량의외부충격시 다리골절 중상사고는 여름(113), 주중(122), 맑음(291), 인구밀집한 시가지(351), 신호기가 있음(361), 건조한 도로(381), 직선평지(371)의 단일로(332)의 규칙들과 연관이 있다. 머리골절에 의한 사망사고와 비교해 볼 때에 신호기 설치 사망 사고를 감소시킨다고 볼 수 있다.

표 6. 다리골절중상사고의 환경요인 연관규칙

신뢰도	지지도	연 관 규 칙
71.43	8.89	381, 371, 361, 332, 113 → 902, 881, 867, 642
71.43	8.89	381, 371, 332, 291, 122 → 902, 881, 867, 642
71.43	8.89	381, 361, 351, 291, 113 → 902, 881, 867, 642
71.43	8.89	381, 351, 332, 291, 122 → 902, 881, 867, 642

차량의외부충격시 머리충격 중상사고는 맑은(291) 날씨에 신호기가 있음(361) 인구밀집한 시가지(351)의 직선평지(371)의 포장된 차도 폭 9미터(342) 미만의 도로에서 사고발생 시에 주로 발생하였다.

표 7. 머리충격중상사고의 환경요인 연관규칙

신뢰도	지지도	연 관 규 칙
83.33	10.00	371, 361, 351, 342, 291 → 902, 883, 861, 642

차량내부충격시 목 탈구염좌 경상사고는 주중(122), 인구밀집한 시가지(351)에 있는 포장되고 건조한(381), 차도 폭 9미터 미만(342)의 교차로(331)에서 발생하거나, 야간(132)의 맑은(291) 날씨에 직선평지(371)의 포장되고 건조한(381), 차도 폭 9미터 미만(342)의 도로에서 주로 발생하였다.

표 8. 목 탈구염좌 중상사고의 환경요인 연관규칙

신뢰도	지지도	연 관 규 칙
70.00	10.00	381, 351, 342, 331, 122 → 901, 884, 863, 643
70.00	10.00	381, 371, 342, 291, 132 → 901, 884, 863, 643

3.3 차량/ 행동유형에 의해 도출된 연관규칙

행동유형은 보호장구 착용여부, 용도별, 배기량, 사고직전 속도, 행동유형 등 5개의 변수로 분석을 하였다. 그 결과 1,512개의 연관규칙이 도출되었다.

차량내부충격시 머리골절 사망사고는 난폭운전(562)이나 음주운전(584)을 하지 않으나, 보호장구를 착용하지 않고(802), 직진(998)중에 차량단독사고(493)를 일으킬 때 주로 발생하였다. 그러므로, 보호장구를 착용하지 않고 발생한 사고의 피해가 보호장구를 착용한 경우의 사고보다 더 심각한 결과가 발생될 수 있다.

표 9. 머리골절사망사고의 행동유형 연관규칙

신뢰도	지지도	연 관 규 칙
71.43	8.89	998, 802, 584, 562, 493 → 901, 881, 861, 641

차량의외부충격시 다리골절 중상사고는 1000CC 미만(471)의 차량을 보호장구를 착용(801)한 운전자가 차대차 사고(492)의 항목에 대한 연관규칙이 나왔다.

표 10. 다리골절중상사고의 행동유형 연관규칙

신뢰도	지지도	연 관 규 칙
71.43	8.89	801, 492, 471 → 902, 881, 867, 642

차량의부충격시 머리충격 중상사고는 보호장구를 착용(801)한 운전자가 1000CC 미만(471)의 차량으로 차대차 사고(492)를 일으킬 때 주로 발생한다.

표 11. 머리충격중상사고의 행동유형 연관규칙

신뢰도	지지도	연 관 규 칙
71.43	8.89	801, 492, 471 → 902, 883, 861, 642

차량내부충격시 목 탈구염좌 경상사고는 1500CC 미만(472)의 소형차를 운전하는 운전자가 보호장구를 착용한(801) 상태에서 사고직전속도가 50 km/h(501)로 차대차 사고(492)를 일으킬 때 주로 발생한다. 차량의 배기량이 1000CC미만일 때에는 주로 중상이 발생하고, 1500CC 미만일 때는 경상사고가 발생함을 볼 수 있다.

표 12. 목 탈구염좌 경상사고의 행동유형 연관규칙

신뢰도	지지도	연 관 규 칙
75.00	10.00	801, 501, 492, 472 → 901, 884, 863, 643

4. 결론

본 연구에서는 교통사고 자료를 이용하여 네 개의 목적변수(사고정도, 신체상해 주부위, 신체상해 부위상태, 신체가해 주부위)와 인적, 환경적 요인, 차량/행동유형을 나타내는 각각 6개, 10개, 4개 변수간의 연관규칙을 사고 요인별로 분석해 보았다. 분석 결과, 차량내부 충격시 머리골절 사망사고는 맑은 날에 고졸학력의 30대 남자가 포장되고 건조한, 구배가 없는 단일 직선 도로를 운전중 신호기가 없는 곳에서 발생하거나, 보호장구를 착용하지 않고 비사업용 승용차 운전시 주로 발생하였다. 차량 외부 충격시 다리골절 중상사고는 무면허인 고졸학력의 남자가 운전 중에 발생하거나, 여름날 주중의 맑은 날에 보호장구를 착용한 채로 인구밀집한 시가지의 신호기가 없는 포장되고 건조한, 차도 폭 9미터의 직선 단일도로에서 발생되었거나, 1000CC 미만의 차량에 의한 차대차 사고가 발생할 때에 주로 발생하였다. 차량외부충격시 머리충격 중상사고는 보호장구를 착용(801)한 운전자가 1000CC 미만(471)의 차량으로 차대차 사고(492)를 일으킬 때 주로 발생한다. 차량내부충격시 목 탈구염좌 경상사고는 1500CC미만(472)의 소형차를 운

전하는 운전자가 보호장구를 착용한(801) 상태에서 사고직전 속도가 50 km/h(501)로 차대차 사고(492)를 일으킬 때에 주로 발생한다. 차량의 배기량이 1000CC미만일 때에는 주로 중상이 발생하고, 1500CC 미만일 때는 경상사고가 발생함을 볼 수 있다.

사망사고는 차량단독사고, 보호장구를 착용하지 않은 것과 신호기가 없는 곳에서 주로 발생하였다. 중상사고는 차대차 사고이거나, 보호장구는 착용을 하였거나, 무면허운전의 경우와 연관되는 것을 알 수 있다. 경상사고는 야간에 소형차에 의한 교차로에서의 차대차 사고로 패턴을 추출할 수 있다.

사고가 발생한 장소를 고려해 보면 경상사고가 교차로에서, 직선 단일로에서 중상사고와 사망사고가 주로 발생하였다. 이는 직선도로에서의 속도가 교차로의 속도보다 대체로 빠르기 때문에 사고 발생시, 중상사고나 사망사고와 같은 치명적인 명피해가 발생하는 것으로 해석할 수 있다. 목 부상은 야간에 급정거에 의한 차대차 충돌사고로 추정할 수 있다. 이는 야간 운전시에 안전거리를 충분히 유지하지 않은 원인을 고려할 수 있고, 주간에 비해 야간의 시야가 제한되기 때문에 운전 간에 안전과 시야확보 등에 필요한 조치가 강구되어야겠다.

목 부상은 야간에 급정거에 의한 차대차 충돌사고로 추정할 수 있다. 이는 야간운전 시에 안전거리를 충분히 유지하지 않은 원인을 고려할 수 있고, 주간에 비해 야간의 시야가 제한되기 때문에 운전간에 안전과 시야확보 등에 필요한 조치가 강구되어야겠다.

본 연구의 결과 사망사고, 중·경상사고를 일으키는 인적, 환경적, 차량/행동유형의 항목들이 연관규칙을 통해 도출되었으나 교통사고 통계원표 자료에서 인명피해 자료를 자세히 조사한 자료의 수가 적은 관계로 연구에 제한이 있었다. 정확하게 작성된 교통사고 통계원표의 자료는 교통사고 데이터로서 뿐만 아니라 사고예방과 인명피해의 감소, 도로의 효율적인 관리에 좋은 자료로 사용될 것이라 판단된다.

향후 연구는 도로선형, 도로형태 등과 같은 도로 환경적 요인이 사고정도, 사고 피해액에 미치는 영향을 분석해 보고자 한다.

참고문헌

- Agrawal, R., Imielinski, T., and Swami, A. (1993), "Database Mining: A Performance Perspective", in IEEE trans. on knowledge and data engineering., 5(6), 914-925.
- Kim, Eui Kyoung., Lee, Do Heon., Kim, Myoung Ho. and Lee, Yoon Joon. (1997), Mining Unidirectional Quantized Association Rules in Large Database, *Journal of KISS(B): Software and Applications*, 24(4), 373-385.
- Kim, Mi Young. (1993), Review of Road Traffic Accident Analysis Techniques and its Case Study Using Korean Data, A master's thesis of Seoul National University.
- Lee, Il Byoung and Lim, Hyon Yeon (1990), A Study on Development of Forecasting Model for Traffic Accident in Korea, *Journal of Korean Society of*

- Transportation*, 8(1), 73-88.
- Mannila, H. and Raiha, K.-J. (1987), "Dependency Inference," Proc. of 3rd Intl. Conf. on Very Large Data Bases., 155-158.
- Oh, Yun Pyo and Ko, Sang Sean (1992), A Study about Establishment of Discrimination Model of Impact Factors of Big Traffic Accident - with Laws Violation Type -, *Journal of Korean Society of Transportation*, 10(3), 173-180.
- Sohn, S. Y. and Lee, S. H. (2001), Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accident in Korea. accepted to Safety Science 2001.
- Sohn, So Young. and Shin, Hyung Won. (1998), Data Mining for Road Traffic Accident Type Classification, *Journal of Korean Society of Transportation*, 16(4), 187-194.