

의사결정나무를 활용한 업종별 버스 교통사고 특성 연구

Study on the Characteristics of Bus Traffic Accidents by Types Using the Decision Tree

박 원 일	Park, Wonil	한국운수산업연구원 연구원 (E-mail : wonil@kriti.re.kr)
김 경 현	Kim, Kyung Hyun	아주대학교 건설교통공학과 박사과정 (E-mail : kk6661@ajou.ac.kr)
한 음	Han, Eum	아주대학교 건설교통공학과 박사과정 (E-mail : hano3106@ajou.ac.kr)
박 상 민	Park, Sangmin	아주대학교 건설교통공학과 박사과정 (E-mail : stylecap@ajou.ac.kr)
윤 일 수	Yun, Ilsoo	정회원 · 아주대학교 교통시스템공학과 부교수 · 교신저자 (E-mail : ilsooyun@ajou.ac.kr)

ABSTRACT

PURPOSES : This study was initiated to analyze the characteristics of bus traffic accidents, by bus types, using the decision tree in order to establish customized safety alternatives by bus types, including the intra-city bus, rural area bus, and inter-city bus.

METHODS : In this study, the major elements involved in bus traffic accidents were identified using decision trees and CHAID algorithm. The decision tree was used to identify the characteristics of major elements influencing bus traffic accidents. In addition, the CHAID algorithm was applied to branch the decision trees.

RESULTS : The number of casualties and severe injuries are high in bus accidents involving pedestrians, bicycles, motorcycles, etc. In the case of light injury caused by bus accidents, different results are found. In the case of intra-city bus accidents, the probability of light injury is of 77.2% when boarding a non-owned car and breaching of duty to drive safely are involved. In the case of rural area bus accidents, the elements showing the highest probability of light injury are boarding an owned car, vehicle-to-vehicle accidents, and breaching of duty to drive safely. In the case of intra-city bus accidents, boarding owned car, streets, and vehicle-to-vehicle accidents work as the critical elements.

CONCLUSIONS : In this study, the bus accident data were categorized by bus types, and then the influential elements were identified using decision trees. As a result, the characteristics of bus accidents were found to be different depending on bus types. The findings in this study are expected to be utilized in establishing effective alternatives to reduce bus accidents.

Keywords

Decision Tree Method, CHAID Algorithm, Traffic Accidents, Bus Types

Corresponding Author : Yun, Ilsoo, Associated Professor
Department of Transportation System Engineering, Ajou University
206 World Cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16499, Korea
Tel : +82.31.219.3610 Fax : +82.31.215.7604
E-mail : ilsooyun@ajou.ac.kr

International Journal of Highway Engineering
http://www.ksre.or.kr/
ISSN 1738-7159 (print)
ISSN 2287-3678 (Online)
Received May, 13, 2016 Revised Sep, 05, 2016 Accepted Sep, 06, 2016

1. 서론

1.1. 연구의 필요성 및 목적

경찰청에서 발간하는 교통사고 통계에 따르면 우리나라 교통사고 발생건수와 사망자수는 1988년에 처음으로 각각 20만 건과 1만 명을 넘어 225,062건의 교통사고가

발생하였고 이로 인해 11,563명이 사망하였다. 2014년에 발생한 교통사고 발생건수와 사망자수는 각각 223,552 건과 4,762명으로서 교통사고 발생건수 규모는 유사하나, 사망자수는 크게 줄었다(The National Police Agency, 2015). 이를 자동차 1만대당 교통사고 발생건

수와 사망자수로 따져보면, 1988년에는 725건/1만대와 37.2명/1만대이고, 2014년에는 94건/1만대와 2.0명/1만대로서 각고의 노력을 통해 교통안전 측면에서 큰 성과를 이루었다고 할 수 있다(The National Police Agency, 2015). 또한 2000년부터 작성된 교통사고통계에서는 교통사고 사망 기준이 기존 교통사고 발생 시로부터 3일 이내에서 30일 이내로 변경된 것을 감안하면 큰 발전이라고 할 수 있다(The National Police Agency, 2015). 또한 최근 3년간 연평균 교통사고 감소율이 5~6%의 꾸준한 감소 추세를 보이고 있는 것도 고무적이라 할 수 있다(The National Police Agency, 2015).

버스 교통사고 사망자수의 경우에도 큰 성과가 있었다. 버스 교통사고가 작성된 이래로 버스 교통사고가 가장 많았고 또한 버스형태가 시내버스, 시외버스, 고속버스 등으로 세분화된 교통사고 정보를 제공하기 시작한 1992년에 696명의 사망자가 발생하였다. 버스 교통사고 사망자수도 1992년 696명에서 2014년 149명으로 약 80% 감소하는 성과를 거두었지만, 2006년 이후 교통사고 사망자수 감소율이 둔화되고 있다. 특히 2011년부터 4년 동안 버스 교통사고 사망자수는 143명~173명 수준으로 정체되고 있는 실정이다(The National Police Agency, 2015). 최근 들어 나타나는 이러한 사망자수 감소율 정체는 버스 속도제한장치 설치, 고속도로 전좌석 안전띠 착용 의무화, 과속 및 졸음운전 근절 캠페인 등 기존 버스 관련 교통사고 예방대책의 실효성이 떨어지는 결과라고 할 수 있다. 또한 버스 교통사고 예방을 위해서는 시내버스, 농어촌버스, 시외버스 등 버스 업종별로 교통사고 원인을 분석하고, 분석된 원인을 해결할 수 있는 맞춤형 교통사고 예방대책 수립이 필요하다. 하지만, 이러한 업종별 버스 교통사고 원인 분석과 관련된 연구가 국내에서는 부족한 실정이다.

이에 본 연구에서는 먼저 버스 교통사고의 업종별 사고원인 특성에 관한 자료를 수집하여 기초통계분석을 실시하였다. 또한 질적 변수(qualitative variable) 또는 범주형 변수(categorical variable)로 구성된 자료에 대한 분석이 가능하고 자료 간 관련성을 분류할 수 있는 의사결정나무(decision tree)를 사용하여 업종별 버스 교통사고에 영향을 미치는 주요 요인들의 특성을 살펴보았다.

본 연구 결과는 중앙정부, 교통안전공단, 경찰청, 지자체, 그리고 운수회사 등이 업종별 버스 교통사고 원인 분석 결과를 기반으로 하는 맞춤형 버스 안전대책을 수립하는데 근거자료로 활용될 수 있을 것으로 기대된다.

1.2. 연구의 범위

본 연구의 시간적 범위는 최근 3년(2011~2013년)이며, 공간적 범위는 전국을 대상으로 한다. 조사대상은 버스공제에 가입된 노선버스(시내버스, 농어촌버스, 시외버스) 교통사고 중 버스공제에 신고된 버스 교통사고이다. 참고로 고속버스는 시외버스에 포함되어 있고, 전세버스의 경우는 노선버스와 상당히 다른 통행패턴을 보이고 있기 때문에 본 연구에서는 포함되지 않았음을 밝힌다.

1.3. 연구의 방법

버스 교통사고에 영향을 미치는 주요 요인을 도출하기 위해서 질적 변수로 구성된 종속변수(dependent variable, 또는 목표변수라고 불리기도 함)와 독립변수(independent variable, 또는 예측변수라고 불리기도 함)를 사용하였다. 의사결정나무 모형 개발은 통계 프로그램인 SPSS 18.0을 이용하였다. 이를 위한 본 연구의 수행과정은 Fig. 1과 같다.

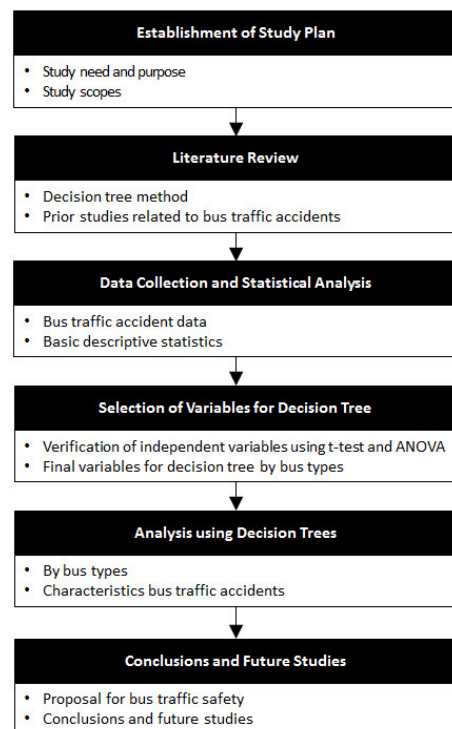


Fig. 1 Study Process

2. 관련 이론 및 문헌 고찰

2.1. 의사결정나무

2.1.1. 개념

의사결정나무는 나무의 구조(예, 뿌리, 줄기 그리고 가

지)를 활용하여 자료 간 관계를 모형화한다. 즉, 의사결정 나무 모형의 구조는 나무를 거꾸로 세워 놓은 형태로서 각 부분은 기능에 따라 뿌리마디(root node), 중간마디(internal node), 끝마디(terminal node), 가지(branch)로 분류된다(Fig. 2 참조). 또한 부모마디(parent node)는 분리 전 마디이고 자식마디(child node)는 분리 후 마디가 된다. 따라서 뿌리마디는 무조건 부모마디이고, 중간마디는 부모마디 또는 자식마디 둘 다 가능하며, 끝마디는 무조건 자식마디가 된다. 다시 말하면, 의사결정 나무 방법은 분석데이터 간의 관계를 바탕으로 뿌리에서 가지로 분류한 후, 분류된 가지 중에서 불필요한 가지(설명력이 약한 관계)를 제거함으로써 유용한 자료를 분류·예측 나무 구조로 만들어 자료 간 관계성을 파악하는 데이터 마이닝(data mining)의 분류(classification) 분석 방법 중의 한 가지이다.

데이터 마이닝은 대규모로 저장된 자료 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 보편적인 방법이다. 분류 분석이라 함은 기본적인 자료 분석 방법 중 하나로서 자료를 특성별로 집단화하여 여러 개의 집단으로 분류하거나 예측하는 것을 말한다. 이때 분류하는 대상은 양적 변수(quantitative variable) 중 1, 2, 3과 같이 셀 수 있는 이산형 변수(discrete variable)나 코카콜라, 펩시와 같은 명목형 변수(nominal variable) 등의 질적 변수를 포함하는 자료에 적용되고, 예측은 양적 변수 중 연속형 변수(continuous variable)를 포함하는 자료를 대상으로 한다(Lee, 2013).

의사결정나무 분석에는 Chi-squared Automatic Interaction Detection(CHAIID), Classification and Regression Trees(CRT), Quick, Unbiased, Efficient, Statistical Tree(QUEST) 등의 알고리즘이 부모마디에서 자식마디를 분류하는데 사용된다.

2.1.2. CHAID

CHAID 알고리즘은 카이제곱-검정이나 F-검정을 활용하여 한 개의 부모마디에서 3개 이상의 자식마디로 분리할 수 있는 다지 분리(multiway split)를 수행할 수 있는 알고리즘이다. 종속변수가 질적 변수인 경우 카이제곱 통계량을 이용하고, 양적 변수인 경우 F-검정을 이용하여 노드를 분리한다(Kang and Kim, 2013). 구체적인 사례는 다음 장에서 설명될 예정이다.

카이제곱 통계량은 관측도수(f_{ij})로 이루어진 $r \times c$ 분할표로부터 계산되며, Pearson의 카이제곱 통계량은 Eq. (1)과 같다(Won and Lee, 2009). 카이제곱 통계

량이 크고 이에 대한 유의확률(p)이 $\alpha = 0.05$ 보다 작은 경우 부모마디는 자식마디를 생성하게 된다.

$$x^2 = \sum_{ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad (1)$$

여기서, f_{ij} : 관측도수
 e_{ij} : 기대도수

2.1.3. CRT

CRT 알고리즘은 지니 계수나 분산의 감소량을 이용하여, 2개의 자식노드만이 형성되는 이진분리(binary split)를 수행할 수 있다. 종속변수가 질적 변수인 경우 불순도의 가중합을 나타내는 지니 계수를 최소화하여 부모마디와 자식마디를 구분한다(Kang and Kim, 2013).

이때 종속변수가 c 개의 범주를 가질 경우, 어떤 관측치가 종속변수의 i 번째 범주에 속해 있음에도 불구하고 j 번째 범주에 속한다고 오분류(misclassification)할 확률은 $P(i)P(j)$ 이 된다. 전체 오분류 확률의 합이 불순도의 가중합에 해당하고, Eq. (2)와 같이 계산할 수 있다(Lee, 2013).

$$G = \sum_{i=1}^c \sum_{j=1, j \neq i}^c P(i)P(j) \quad (2)$$

여기서, G : 오분류 확률 추정치의 합
 $P(i)$: 각 마디에서 한 개체가 목표변수의 i 번째 범주에 속할 확률
 c : 목표변수 범주의 수

지니 계수(Gini index)는 n 개의 원소 중에서 임의로 2개를 추출하였을 때 추출된 2개가 서로 다른 그룹에 속해 있을 확률을 의미하며, Simpson의 다양성 지수(Diversity Index)로도 알려져 있다(Lee, 2013). 실질적으로 지니 계수는 Eq. (3)을 이용하여 산출한다.

$$G = \sum_{j=1}^c P(j)(1 - P(j)) = 1 - \sum_{j=1}^c P(j)^2 = 1 - \sum_{j=1}^c (n_j/n)^2 \quad (3)$$

여기서, G : 지니 계수
 n : 그 마디에 포함되어 있는 관측치 수

n_j : 목표변수의 j 번째 범주에 속하는 관측치 수

2.1.4. QUEST

QUEST 알고리즘은 명목형 종속변수에 대하여 카이제곱-검정 또는 F-검정의 p값을 분리기준으로 하여 이치분리를 수행하는 알고리즘이다. 이때 사용되는 분리기준은 예측변수의 척도에 따라 결정되며, 명목형인 경우 Pearson 카이제곱-검정, 순서형 또는 연속형인 경우에는 F-검정이나 Levene-검정을 이용한다(Kang and Kim, 2013).

2.2. 관련 연구 고찰

Ha et al.(2010)은 고속도로에서 발생하는 비반복적인 정체(non-recurring congestion) 중 교통사고로 인하여 발생하는 돌발상황 처리시간 예측모형을 개발하기 위해 CHAID 알고리즘을 이용하여 의사결정나무 모형을 구축하였다. 종속변수인 사고처리시간을 명목 변수인 사고등급 A, B, C등급으로 구분하고, 독립변수로는 교통량, 교통사고 차량수, 사고 발생 시간대 등 총 15개 변수를 적용하였다. 돌발상황 처리시간 예측모형에서 영향력이 있는 변수로 도출된 변수들은 교통량, 중차량 포함 여부, 사고 발생 시간대로 나타났다. 그 결과 1차적으로 심각한 사고인 A, B등급과 경미한 사고인 C 등급으로 구분되었고, 2차적으로 대상도로의 교통량으로 분리되었다. 이는 사고심각도에 따라 사고현장을 처리하는 시간이 구분되고, 교통량이 많을수록 돌발상황을 처리하는데 소요되는 시간이 오래 걸리는 것으로 해석될 수 있다.

Son(2010)은 교통안전공단인 운전자 운전적성 정밀검사 항목을 활용하여 교통사고 발생에 대한 연령대별 분류 특성 규명을 위하여 의사결정나무 모형을 사용하였다. 의사결정나무 모형을 가지고 CRT 분석을 수행한 결과, 교통사고 발생 유형이 51세 이상 또는 50세 이하 2개의 집단으로 뚜렷하게 구분되었다. 연령을 기준으로 교통사고 발생여부에 영향을 미치는 특성을 파악하기 위해 의사결정나무 모형을 구축한 결과, 연령대가 높아질수록 심리·생리적인 속성과 운전 적성은 쇠퇴하는 것으로 나타났다. 다만 보상적 운전 행위로 인하여 일부 안전성을 담보하고 있지만, 시력 저하 등 운전에 절대적인 신체기능은 감소하여 교통사고의 위험성은 높아진다고 분석하였다.

Jeong(2013)는 부산시 내 도시고속도로의 교통사고 자료를 이용하여 지니 계수에 의한 이치분리를 수행하

는 의사결정나무 모형의 CRT 분석법을 이용하여 분류 분석을 시행하였다. 중상 교통사고 발생 유무를 종속변수로 설정한 후, 도시고속도로별·연도별 교통사고 발생 유형, 기상 상태, 도로 형태, 도로 선형, 운전자 연령, 차량 용도, 행동 유형, 교통법규 위반 내용을 포함한 총 8개의 독립변수에 대해서 종속변수에 미치는 영향정도를 분석하였다.

Lee(2013)는 중상 이상 교통사고 중 시도별·주행거리별 교통사고 표본을 추출하여, 교통사고 위험성에 영향을 미치는 범주형 독립변수(운전자 요인, 자동차 요인, 도로·환경적 요인)에 대한 분석모형 개발을 위해 의사결정나무 모형을 이용하였다. 중상 이상 교통사고에 영향을 끼치는 공통 요인은 사고 유형, 차량 종별, 도로 종류인 것으로 나타났다. 특히, 사고 유형 중에는 차대 사람 사고에서, 차량 종별에서는 승합자동차 및 화물자동차에서, 도로 종류에서는 일반국도 및 고속국도에서 사망사고의 위험성이 증가하고 있는 것으로 나타났다.

2.3. 연구의 차별성

기존의 의사결정나무 모형을 활용한 교통사고 관련 연구들은 중상 이상 교통사고 및 사고처리시간을 종속변수로 이에 영향을 미치는 독립변수들인 인적요인, 차량요인, 도로·환경요인의 영향 정도를 주로 분석하였다. 관련 연구에서는 시도별이나 도로별로 교통사고에 영향을 미치는 사고요인, 연령별 교통사고에 영향을 미치는 운전자 운전적성 정밀검사 항목, 사고처리시간에 영향을 미치는 도로환경특성 등을 검토하였다. 이로 미루어 볼 때 아직까지 사업용 자동차의 업종별 사고특성에 관한 연구는 미흡한 실정이다. 본 연구에서는 사업용 자동차 중에서도 노선버스의 업종별 교통사고 발생에 영향을 미치는 사고요인의 정도를 의사결정나무 모형을 이용하여 분석하였다.

3. 자료수집 및 기초분석

3.1. 자료수집

본 연구에서는 2011년부터 2013년까지 최근 3년간 버스공제조합에 접수된 교통사고를 대상으로 하였다. 시간적 및 공간적 범위 내에서 신고된 버스 교통사고는 64,316건이다. 이 중 기타 버스 교통사고 36건을 제외하고 업종별로 나누면 시내버스 57,085건, 농어촌버스 1,958건, 시외버스 5,237건으로서 총 64,280건을 대상으로 교통사고 분석을 실시하였다.

3.2. 업종별 버스 교통사고 특성

상해정도에 따른 업종별 버스 교통사고 특성은 Table 1과 같이 농어촌버스와 시외버스는 사망사고 비중이 1.5~1.6%로 시내버스의 0.7%에 비해 높다. 또한 농어촌버스의 중상사고 비중은 43.9%로 다른 업종에 비해 높은 것으로 나타났다.

Table 1. Classification of Bus Traffic Accidents by Accidents Severities

Accident types	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Fatal accidents	518 (0.8%)	402 (0.7%)	30 (1.5%)	86 (1.6%)
Serious injury accidents	23,145 (36.0%)	20,581 (36.1%)	859 (43.9%)	1,705 (32.6%)
Light injury accidents	40,617 (63.2%)	36,102 (63.2%)	1,069 (54.6%)	3,446 (65.8%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

주야별 버스 교통사고 특성은 Table 2와 같다. 야간에 발생한 시내버스와 시외버스의 교통사고 비율은 각각 27.3%, 24.7%로서 농어촌버스의 야간 교통사고 비율인 15.3%에 비해 높게 나타났다. 이는 농어촌버스의 야간 운행 빈도가 타 버스에 비해 적기 때문인 것으로 판단된다.

Table 2. Classification of Bus Traffic Accidents by Times

Times	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Day time	47,080 (73.2%)	41,478 (72.7%)	1,659 (84.7%)	3,943 (75.3%)
Night	17,200 (26.8%)	15,607 (27.3%)	299 (15.3%)	1,294 (24.7%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Note: the day time is between 6 am, and 6 pm.

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

계절별 버스 교통사고 특성은 Table 3과 같이 모든 업종이 가을에 가장 많이 발생하는 것으로 나타났다.

Table 3. Classification of Bus Traffic Accidents by Seasons

Seasons	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Spring	15,885 (24.7%)	14,173 (24.8%)	485 (24.8%)	1,227 (23.4%)
Summer	15,295 (23.8%)	13,560 (23.8%)	458 (23.4%)	1,277 (24.4%)
Autumn	16,998 (26.4%)	15,100 (26.5%)	515 (26.3%)	1,383 (26.4%)
Winter	16,102 (25.0%)	14,252 (25.0%)	500 (25.5%)	1,350 (25.8%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

날씨별 버스 교통사고 특성은 Table 4와 같이 시내버스는 맑음일 때 88.1%이고, 농어촌버스는 흐림일 때 8.2%이며, 시외버스는 비·눈일 때 각각 7.7%, 2.6%로 나타났다.

Table 4. Classification of Bus Traffic Accidents by Weather

Weather	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Sunny	56,231 (87.5%)	50,293 (88.1%)	1,639 (83.7%)	4,299 (82.1%)
Cloudy	3,363 (5.2%)	2,808 (4.9%)	161 (8.2%)	394 (7.5%)
Rainy	3,847 (6.0%)	3,334 (5.8%)	112 (5.7%)	401 (7.7%)
Snowy	799 (1.2%)	616 (1.1%)	46 (2.3%)	137 (2.6%)
Foggy	40 (0.1%)	34 (0.1%)	-	6 (0.1%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

도로형태별 버스 교통사고 특성은 Table 5와 같다. 시외버스는 고속도로에서 15.3%로 가장 높은 교통사고 발생비율을 보였다.

Table 5. Classification of Bus Traffic Accidents by Road Types

Road types	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Expressways	934 (1.5%)	132 (0.2%)	2 (0.1%)	800 (15.3%)
Motorways	126 (0.2%)	71 (0.1%)	1 (0.1%)	54 (1.0%)
Arterial highways	8,038 (12.5%)	7,430 (13.0%)	81 (4.1%)	527 (10.1%)
Streets	52,261 (81.3%)	47,017 (82.4%)	1,750 (89.4%)	3,494 (66.7%)
Side streets	603 (0.9%)	513 (0.9%)	39 (2.0%)	51 (1.0%)
Busways	692 (1.1%)	679 (1.2%)	4 (0.2%)	9 (0.2%)
Others	1,626 (2.5%)	1,243 (2.2%)	81 (4.1%)	302 (5.8%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

사고종류별 버스 교통사고 특성은 Table 6과 같다. 시내버스는 차 대 사람이 51.4%로 절반 이상을 차지하고, 농어촌버스는 차량 단독이 8.1%로 다른 업종의 두 배 비중을 차지하며, 시외버스는 차 대 차가 74.7%로 3/4을 차지하고 있는 것으로 나타났다. Table 6에서 보인 바와 같이 버스 업종별로 교통사고 발생 유형이 상이하다. 사고종류별 세부항목도 살펴보면 Table 6에 제시되어 있다. 시내버스는 차 대 사람 중 차내 전도가 34.2%이고, 농어촌버스는 차량 단독 중 차량 단독기타가 5.3%이며, 시외버스는 차 대 차 중 추돌이 39.3%로 많은 비중을 차지하는 것으로 나타났다.

Table 6. Classification of Bus Traffic Accidents by Accidents Types

Accident types	No. of Bus traffic accidents				
	Total	Intra-city buses	Rural area buses	Inter-city buses	
Vehicle only	Rollover	408(0.6%)	357(0.6%)	32(1.6%)	19(0.4%)
	Run-off road	60(0.1%)	42(0.1%)	9(0.5%)	9(0.2%)
	Others	2,289(3.6%)	2,048(3.6%)	103(5.3%)	138(2.6%)
	Crash to objectives	239(0.4%)	173(0.3%)	15(0.8%)	51(1.0%)
	Sub-total	2,996(4.7%)	2,620(4.6%)	159(8.1%)	217(4.1%)
Vehicle-to-vehicle	Head-on crash	1,290(2.0%)	1,079(1.9%)	68(3.5%)	143(2.7%)
	Angle crash	7,053(11.0%)	5,933(10.4%)	296(15.1%)	824(15.7%)

Vehicle-to-vehicle	Rear-end crash	11,305 (17.6%)	9,007 (15.8%)	242(12.4%)	2,056 (39.3%)
	Fender-bender	5,427(8.4%)	4,793(8.4%)	105(5.4%)	529(10.1%)
	Bycle	661(1.0%)	592(1.0%)	26(1.3%)	43(0.8%)
	Others	2,914(4.5%)	2,654(4.6%)	102(5.2%)	158(3.0%)
	Motorcycle	972(1.5%)	840(1.5%)	43(2.2%)	89(1.7%)
	Backward crash	328(0.5%)	239(0.4%)	17(0.9%)	72(1.4%)
Sub-total	29,950(46.6%)	25,137(44.0%)	899(45.9%)	3,914(74.7%)	
Vehicle-to-person	Jaywalking	1,870(2.9%)	1,702(3.0%)	50(2.6%)	118(2.3%)
	Crosswalk	1,202(1.9%)	1,111(1.9%)	19(1.0%)	72(1.4%)
	Walking	352(0.5%)	263(0.5%)	30(1.5%)	59(1.1%)
	Boarding	1,773(2.8%)	1,701(3.0%)	33(1.7%)	39(0.7%)
	Encroaching on sidewalk	83(0.1%)	79(0.1%)	-	4(0.1%)
	Others	2,536(3.9%)	2,205(3.9%)	108(5.5%)	223(4.3%)
	Alight	2,928(4.6%)	2,760(4.8%)	118(6.0%)	50(1.0%)
	Fallen in the car	20,590 (32.0%)	19,507 (34.2%)	542 (27.7%)	541 (10.3%)
	Sub-total	31,334 (48.7%)	29,328 (51.4%)	900 (46.0%)	1,106 (21.1%)
	Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

버스 중대·일반 교통사고 특성은 Table 7과 같이 시내버스와 농어촌버스의 중대사고가 8.3%로 시외버스에 비해 사고점유율이 높은 것으로 나타났다. 여기서, 중대사고란 신호위반, 중앙선침범, 개문발차 등 중대한 과실이 있는 사고이다. 그리고 일반사고란 중대사고 외에 안전거리미확보, 전방주시태만 등의 부주의한 사고를 말한다.

Table 7. Classification of Bus Traffic Accidents by Violation Types

Severity types	No. of Bus traffic accidents			
	Total	Intra-city buses	Rural area buses	Inter-city buses
Major	5,206 (8.1%)	4,753 (8.3%)	162 (8.3%)	291 (5.6%)
Minor	59,074 (91.9%)	52,332 (91.7%)	1,796 (91.7%)	4,946 (94.4%)
Total	64,280 (100.0%)	57,085 (100.0%)	1,958 (100.0%)	5,237 (100.0%)

Source: Database of the National Bus Mutual-Aid Association between 2011 and 2013

그 외에 피해자 탑승유형별 버스 교통사고 특성은 시내·농어촌버스는 자차탑승이 각각 55.8%, 53.3%로 절반 이상을 차지하고, 시외버스는 타차운전이 52.8%로

타 업종의 두 배 이상의 비중을 차지하였다. 여기서 자차 탑승이란 자차에 타고 있는 운전자와 승객을 말하고 타차운전이란 타차 운전자를 말한다.

4. 의사결정나무 모형을 이용한 분석

4.1. 독립변수 선정

전국버스공제조합의 교통사고 자료 분석을 통하여 서로 다른 패턴을 보이는 유형에 대하여 업종별로 구분된 집단의 독립변수로 설정하였다. 교통사고 요인 모형의 개발을 위하여 통계프로그램인 SPSS 18.0을 이용하였다. 또한 의사결정나무의 3가지 알고리즘 중 다지분리를 수행하는 CHAID 알고리즘을 적용하였다.

모형에 사용된 종속변수는 범주변수인 상해정도(사망, 중상, 경상)로 결정하였다.

버스공제조합에서 발행하는 교통사고 통계에는 다양하고 매우 세분화된 교통사고 요인들이 제시되어 있다. 따라서 모든 교통사고 요인들을 독립변수로서 모형에서 고려하기에는 모형이 과대하게 복잡해지는 문제가 발생

Table 8. Independent Variables for the Decision Tree Model

Types		Independent variables	
Human elements	Types of accidents	Vehicle only, vehicle-to-vehicle, vehicle-to-person	
	Boarding types	Boarding owned car, walking, boarding non-owned car, others	
	Violation	Major violation	Major violation, minor violation, safe driving duty breach
		Minor violation	
Others			
Road and environment elements	Times	Day time, night	
	Seasons	Spring, Summer, Autumn, Winter	
	Weather	Sunny, cloudy, inclement weather	
	Road types	Expressways, Arterial highways, Streets, Others	

Table 9. Significance of Independent Variables in the Decision Tree Models by Bus Types

Bus types	No. of samples	Significance probability	Student t-test	one-way ANOVA					
			Times	Seasons	Weather	Road types	Accident types	Boarding types	Violations
Intra-city buses	57,085	p<0.05	0.007	0.000	0.388	0.000	0.000	0.000	0.000
Rural area buses	1,958	p<0.05	0.002	0.233	0.809	0.008	0.000	0.000	0.003
Inter-city buses	5,237	p<0.05	0.155	0.215	0.228	0.000	0.000	0.000	0.000

할 수 있다. 따라서 본 연구에서는 문헌 검토 등을 통해서 크게 인적요인과 도로·환경적 요인 두 가지 항목으로 분류한 후, 각 항목별로 모형에서 고려될 독립변수를 선정하였다(Jeong, 2013; Lee, 2013). 모형에 사용된 독립변수 중 인적요인으로는 사고종류, 피해자탑승 유형, 법규위반 유형으로 구분하였다. 도로·환경적 요인으로는 주야, 계절, 날씨, 도로형태가 있다. 본 연구에서 선택된 독립변수 세부내용은 Table 8에서 제시되어 있다.

4.2. 독립변수 검증

독립변수 검증에는 종속변수에 대한 두 집단 간의 평균을 비교하는 독립표본 t-검정과 두 개 이상 집단의 평균 차이를 검증하는 일원배치 분산분석(one-way ANOVA)을 이용하였다.

업종별 독립변수 검증결과 유의확률은 Table 9와 같으며, 95% 신뢰도에 따라 유의확률(p)이 0.05 미만인 독립변수를 선별하였다.

업종별 독립변수 검증 결과 Table 10과 같이, 시내버스의 경우는 독립변수로서 주야, 계절, 도로형태, 사고종류, 피해자탑승 유형, 법규위반 유형이 유의한 것으로 나타났다. 농어촌버스의 경우는 주야, 도로형태, 사고종류, 피해자탑승 유형, 법규위반 유형이 유의하게 분류되었다. 시외버스는 도로형태, 사고종류, 피해자탑승 유형, 법규위반 유형이 유의하였다. 이 중에서 업종별 공통적으로 유의한 독립변수는 도로형태, 사고종류, 피해

Table 10. Accepted Independence Variables for the Decision Tree Models

Bus types	Accepted variables	Common variables
Intra-city buses	• Times, season, weather, road types, accident types, boarding types, violations	• Road types • Accident types
Rural area buses	• Times, road types, accident types, boarding types, violations	• Boarding types
Inter-city buses	• Road types, accident types, boarding types, violations	• Violation types

자탑승 유형, 법규위반 유형으로서 이들 4개 변수를 업종별 의사결정나무 모형의 독립변수로 선정하였다.

4.3. 모형결과 해석

4.3.1. 시내버스

시내버스 교통사고는 의사결정나무 모형 중 다지분류가 가능한 CHAID 알고리즘을 이용하여 분류 분석을 시행하였고, 분석 결과 도출된 의사결정나무는 Fig. 2와 같다.

예를 들어, CHAID 알고리즘을 이용한 경우 카이제곱 통계량이 크고 이에 대한 유의확률(p)이 $\alpha=0.05$ 보다 작은 경우 부모마디는 자식마디를 생성하게 된다. 또한 탑승유형에 따라 4개의 다지분리(자차탑승, 보행, 타차탑승, 기타)로 나뉘고, 자차탑승은 다시 사고종류에 따라 3개의 다지(차 대 사람, 차 대 차, 차량 단독)로 분류된다. 이처럼 CHAID 알고리즘의 다지분리는 3개 이상의 자식마디를 형성한다.

시내버스 교통사고에서 가장 중요한 변수는 탑승유형으로 나타났고, 다음으로 사고종류와 법규위반이 영향을 미치는 것으로 나타났다.

사망사고는 탑승유형이 보행인 경우 7.0%로 가장 많이 발생하는 것으로 분석되었다. 다음으로 탑승유형이 기타(3.8%)인 경우 자전거, 이륜차, 경운기 탑승자가 포

함되므로 사망사고가 많으며, 기타 중 일반·중대사고 시 5.4%로 높게 나타났다.

중상사고는 탑승유형이 보행(53.7%)인 경우 발생비중이 높다. 탑승유형이 기타(49.9%)인 경우 일반·중대사고 시 중상사고가 52.2%로 높다. 또한 탑승유형이 자차탑승자(39.1%)인 경우 차 대 사람(42.1%) 중 중대사고(53.5%)일 때 중상위험이 높다.

경상사고는 탑승유형이 타차탑승(76.0%)인 경우, 안전운전불이행·일반사고 시 77.2%가 경상사고이다. 또한 자차탑승(60.8%)인 경우, 차 대 차(70.7%)일 때와 차량 단독사고(62.7%) 중 일반사고(68.2%)일 때 경상사고가 많다.

4.3.2. 농어촌버스

농어촌버스 교통사고에 대한 의사결정나무 모형 분석 결과는 Fig. 3과 같다. 농어촌버스 교통사고에서 가장 중요한 변수는 탑승유형으로 나타났고, 다음으로 사고종류와 법규위반이 영향을 미치는 것으로 나타났다. 사망사고는 탑승유형이 보행·기타인 경우 9.1%로 분석되었다.

중상사고 역시 사망사고와 마찬가지로 탑승유형이 보행·기타인 경우 53.9%로 높게 나타났다. 다음으로 자차탑승자(47.7%)인 경우, 차량 단독·차 대 사람일 때 중상위험이 51.6%로 높다. 경상사고는 탑승유형이 타차탑승

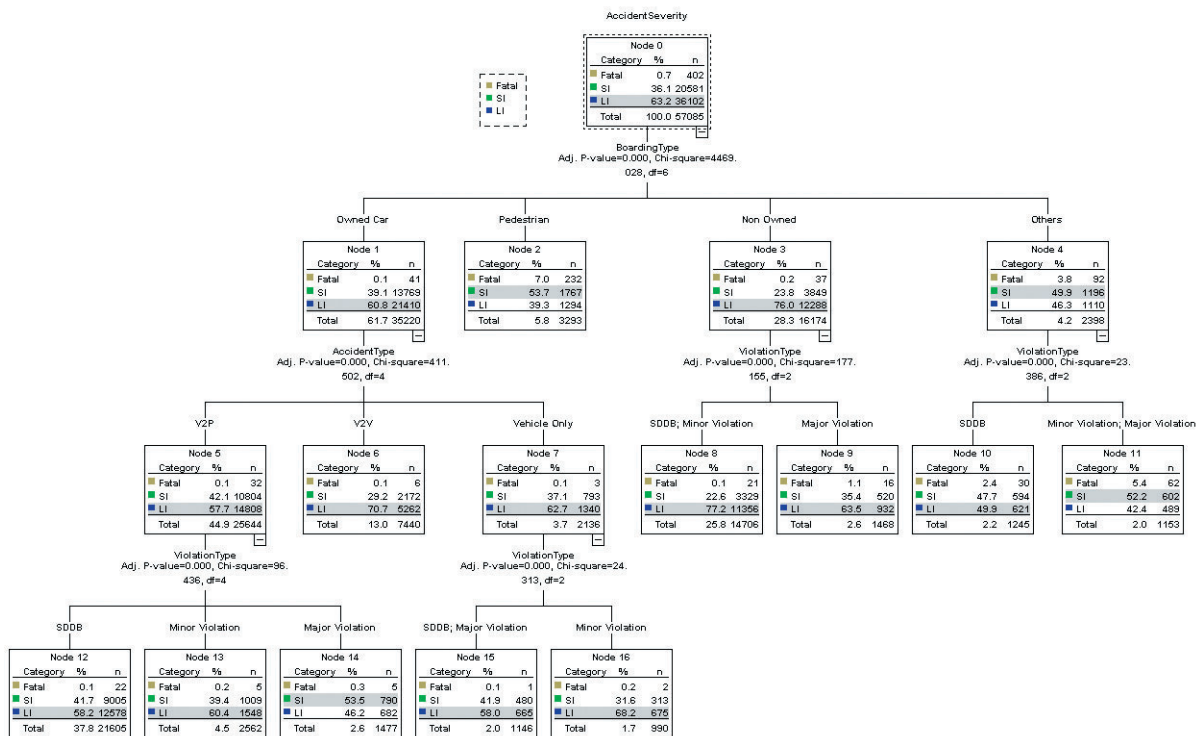


Fig. 2 Decision Tree for Traffic Accidents of Intra-city Buses

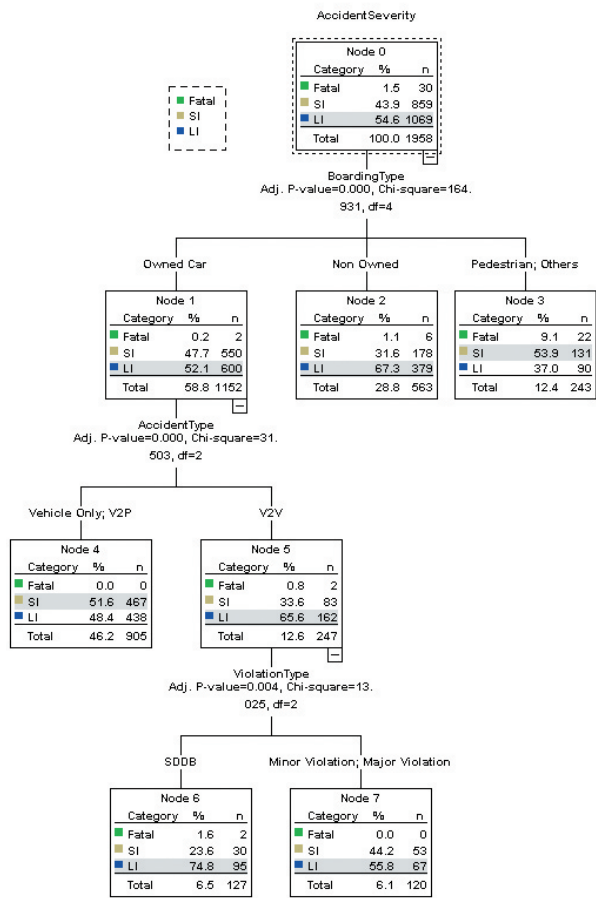


Fig. 3 Decision Tree for Traffic Accidents of Rural Area Buses

인 경우 67.3%로 높다. 다음으로 자차탑승(52.1%)인 경우, 차 대 차(65.6%)에서 안전운전불이행으로 74.8%의 경상사고가 발생한다.

4.3.3. 시외버스

시외버스 교통사고에 대한 의사결정나무 모형 분석 결과는 Fig. 4와 같다. 시외버스 교통사고에서 가장 중요한 변수는 탑승유형으로 나타났고, 다음으로 도로형태와 사고종류가 영향을 미치는 것으로 나타났다.

사망사고는 탑승유형이 보행·기타인 경우 비중이 10.6%로 높다. 중상사고 역시 사망사고와 마찬가지로 탑승유형이 보행·기타인 경우 51.8%로 높다. 다음으로 자차탑승(34.1%)인 경우, 고속도로·간선도로·기타(40.3%)에서 높은 편이다. 또한 타차탑승(28.3%)인 경우 고속도로에서 39.0%가 중상사고로 나타났다. 경상사고는 탑승유형이 타차탑승(70.9%)인 경우 고속도로를 제외한 도로에서 73.6%이다. 다음으로 자차탑승(65.5%)인 경우 일반도로(68.7%)에서 차 대 차일 때 경상사고가 75.8%로 높게 나타났다.

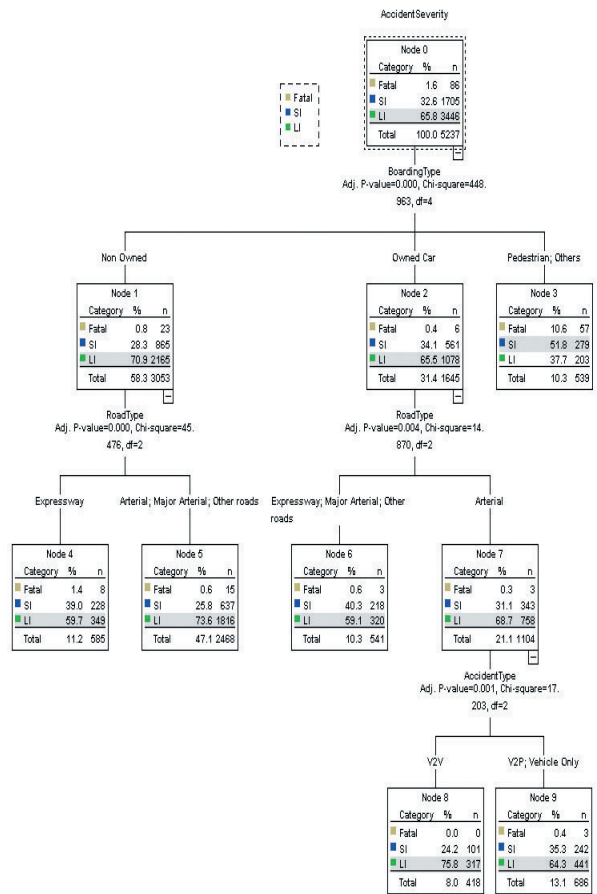


Fig. 4 Decision Tree for Traffic Accidents of Inter-city Buses

4.4. 위험도 및 분류 확률

Table 11에서 제시된 구축된 모형이 관찰치를 오분류할 위험 추정값을 살펴보면, 시내버스 0.355, 농어촌버스 0.418, 시외버스 0.327로 다소 낮은 것으로 나타났다. 추정값에 대한 표준편차는 시내버스 0.002, 농어촌버스, 0.011, 시외버스 0.006임을 알 수 있다.

Table 11. Risk Rates of Decision Trees by Bus Types

Bus types	Estimated risk rate of misclassification	Standard error
Intra-city buses	.355	.002
Rural area buses	.418	.011
Inter-city buses	.327	.006

4.5. 의사결정나무 모형 분석 결과

본 연구에서 의사결정나무 모형을 이용한 버스 교통사고 분석결과를 토대로 업종별 안전대책을 살펴보면, 사망 및 중상사고의 경우 모든 업종에서 보행, 기타(경운기·자전거·이륜차)사고가 많은 것으로 나타나 보행자 안전대

책과 함께 경운기·자전거·이륜차 안전대책이 마련되어야 한다. 이를 위해 무엇보다 감속운행이 가장 중요하고, 버스 특성상 사각지대가 많으므로 이에 대한 승무원의 주의와 보조 미러의 활용이 요구되며, 노인 인구가 증가하는 가운데 보행속도가 느린 노인 특성을 이해한 운전이 몸에 배도록 관련 교육을 강화할 필요가 있다. 야간보행자 식별을 위해 횡단보도 조명시설을 확대하고, 차량 내 센서를 통해 보행자를 감지하여 운전자에게 경고정보를 제공하는 통합보행자보호시스템(Integrated Pedestrian Protection Systems, IPPS) 장착이 필요하다.

업종별 중상사고 특성 및 예방대책을 보다 면밀히 살펴보면, 첫째로 시내버스는 중대사고 시 차내승객이 다치는 사고가 많다. 때문에 신호위반·개문발차 등의 중대사고를 예방해야 하며, 승무원의 과속·급가속·급정차·급차선변경·급커브 등 난폭운전 교정을 위한 체험교육을 활성화할 필요가 있다. 또한 버스회사에서는 디지털 운행기록계 데이터를 활용하여 승무원의 운전습관을 교정하도록 한다. 둘째로 농어촌버스는 시내버스 사고특성과 유사하지만 이에 더해 차량 단독 사고 시 중상위험이 높다. 이는 농어촌지역 도로환경이 열악한데 기인하므로 지자체에서는 도로 포장, 확장 등 개선이 추가로 요구된다. 셋째로 시외버스는 고속도로를 주로 운행하므로 고속주행 시 승객의 전도위험이 높고, 사고 시 다른 차량에 심대한 타격을 가할 수 있다. 때문에 관련기관에서는 승객의 안전띠 착용률을 높이기 위한 캠페인을 실시하고, 승무원은 방어운전을 통해 차 대 차 사고를 최대한 줄이고, 악천후 시 50% 감속과 차간거리를 길게 유지하여야 한다. 특히 정책적으로 대형사고 예방을 위해 시외버스에 첨단차량안전장치인 전방충돌회피시스템(Forward Vehicle Collision Warning System, FVCWS), 차선이탈방지시스템(Lane Departure Warning System, LDWS) 등을 설치하여 고속도로 안전을 최우선으로 해야 한다. 또한 졸음운전 예방장치나 HUD(Head-up Display)를 활용하여 승무원의 졸음 운전과 전방주시 태만 등을 예방할 수 있다.

5. 결론 및 향후 연구과제

5.1. 결론

본 연구에서는 업종별 버스 교통사고의 위험성에 영향을 미치는 주요 요인 특성을 분석하기 위해 버스공제의 최근 3년간(2011~2013년) 버스 교통사고 자료를 활용하였다. 버스 교통사고 안전대책은 업종별 사고특성

에 맞게 서로 다른 방향으로 추진되어야 한다. 이를 위해 범주형 독립변수인 버스 교통사고 요인의 분석 모형 개발을 위해 의사결정나무 모형을 이용하였다.

의사결정나무 모형을 활용한 업종별 버스 교통사고 요인분석 결과는 다음과 같다. 시내버스 사망사고는 탑승유형이 보행(7.0%), 기타(3.8%)인 경우 많다. 중상사고도 탑승유형이 보행(53.7%), 기타(49.9%)인 경우 많고, 탑승유형이 자차탑승(39.1%)인 경우 차 대 사람(42.1%) 사고 중 중대사고(53.5%)일 때 많다.

농어촌버스 사망사고는 탑승유형이 보행·기타(9.1%)인 경우 많다. 중상사고 역시 탑승유형이 보행·기타(53.9%)인 경우 많고, 탑승유형이 자차탑승(47.7%)인 경우 차량 단독·차 사람(51.6%)일 때 많다.

시외버스 사망사고는 탑승유형이 보행·기타(10.6%)인 경우 많다. 중상사고 역시 탑승유형이 보행·기타(51.8%)인 경우 많고, 탑승유형이 자차탑승(34.1%)인 경우 도로형태가 고속도로·간선도로·기타(40.3%)일 때 중상위험이 높다. 또한 탑승유형이 타차탑승(28.3%)인 경우 고속도로(39.0%)에서 많다.

업종별 버스 교통사고 요인분석이 시사하는 점은 다음과 같다. 첫째로 모든 업종에서 가장 중요한 변수는 탑승유형(자차탑승, 보행, 타차탑승, 기타)으로 나타났다. 둘째로 모든 업종에서 사망 및 중상사고는 탑승유형 중 보행이나 기타(자전거, 이륜차, 경운기)인 경우 많이 발생하고 있다. 셋째로 업종별 특성으로 시내버스 중상사고는 차내 승객이 다치는 중대한 범규위반 사고 예방대책이 필요하고, 농어촌버스는 차내 승객이 다치는 차량 단독·차 대 사람 유형의 사고를 줄여야 하며, 시외버스는 차내 승객·타차 탑승자가 모두 다치는 고속도로 사고유형의 사고예방대책이 필요하다.

5.2. 향후 연구과제

본 연구는 버스 교통사고의 업종별 요인 특성을 파악하여 각각의 사고대책을 수립하여 활용하는데 목적이 있다. 의사결정모형의 경우 범주형 변수를 분석하는데 용이하고, 분석과정을 쉽게 이해하고 설명할 수 있는 장점이 있지만, 범주가 많아질 경우 분리수가 많아지거나 서로 다른 특성이 합쳐지기 때문에 모형 결과를 해석하기 어려운 단점이 있다. 따라서 향후 교통사고의 세부특성을 분석하기 위한 다양한 통계분석방법을 연구할 필요가 있다.

또한 이번 연구는 버스공제에 가입된 노선버스(시내버스, 농어촌버스, 시외버스)의 교통사고 자료를 가지고 분

석이 이루어졌다. 물론 지난 7월 영동고속도로 봉평터널에서 발생한 전세버스 5중 추돌사고 등 전세버스 대형사고가 빈번하게 발생하고 있어, 이에 대한 교통사고 분석 및 대책마련이 시급한 실정이다. 하지만 현실적으로 전세버스와 노선버스 공제조합은 서로 다른 기관으로 교통사고 기본통계 양식이 달라 사고유형을 통일하기 힘든 부분이 있다. 향후 노선버스, 전세버스 등 사업용자동차의 기본통계 양식과 사고유형을 통합하여 사업용자동차 교통사고 특성 연구가 이루어질 필요가 있다.

감사의 글

이 논문은 2015학년도 아주대학교 일반연구비와 2013년도 및 2015년도 교육과학기술부의 재원으로 한국연구재단(NRF)의 지원을 받아 수행된 것임(NRF-2010-0028693, 2013K1A1A2A02078326).

REFERENCES

Ha, Oh-Keun, Park, Dong-Joo, Won, Jai-Mu, Jung, Chul-Ho, "The prediction Models for Clearance Times for the unexpected

Incidences According to Traffic Accident Classifications in Highway", The Journal of the Korea Institute of Intelligent Transport Systems, Vol. 9, No. 1, pp. 101-110, 2010.

Jeong, Y. "A Study on the Establishing Improvement Measures by Building discriminant and Classification Models for Impact Factors about Traffic Accidents on Motorways", Youngsan University, 2013.

Kang, B. and Kim, G., Statistical Analysis of Social Science, "Hannarae Academy", 2013.

Lee, M. "Analysis of Traffic Accident Factor Using Decision Tree Model", Myongji University, 2013.

The National Bus Mutual-Aid Association, "Statistics of Bus Traffic Accidents", 2011~2013.

The National Police Agency, "Statistics of Traffic Accidents", 2015.

Son, H., "Development of the Discriminant Traffic Accident Models Considering Human Factors for the Various Age Groups", Hanyang University, 2010.

Won, J. and Lee, S. "Statistical Analysis Learning through Figures", Parkyoungsa, 2009.