

보행자-차량 충돌사고 특성분석 방법론 비교 연구

최새로니* · 정은비 · 오철
한양대학교 교통·물류공학과

Comparison of Methodologies for Characterizing Pedestrian-Vehicle Collisions

CHOI, Saerona* · JEONG, Eunbi · OH, Cheol

Department of Transportation and Logistics Engineering, Hanyang University, Gyeonggi 426-791, Korea

Abstract

The major purpose of this study is to evaluate methodologies to predict the injury severity of pedestrian-vehicle collisions. Methodologies to be evaluated and compared in this study include Binary Logistic Regression(BLR), Ordered Probit Model(OPM), Support Vector Machine(SVM) and Decision Tree(DT) method. Valuable insights into applying methodologies to analyze the characteristics of pedestrian injury severity are derived. For the purpose of identifying causal factors affecting the injury severity, statistical approaches such as BLR and OPM are recommended. On the other hand, to achieve better prediction performance, heuristic approaches such as SVM and DT are recommended. It is expected that the outcome of this study would be useful in developing various countermeasures for enhancing pedestrian safety.

최근 운전자의 보행자-차량 충돌사고 감소를 목적으로 한 다양한 연구가 수행되고 있으며, 본 연구에서는 보행자-차량 사고 발생 특성 및 심각도 영향요인 분석을 위하여 다양한 분석방법론을 활용한 보행자 교통사고 분석을 수행하였다. 통계모형과 휴리스틱모형 적용시 각 기법에 따른 결과를 도출함으로써 보행자 사고분석시 분석목적에 적합한 방법론을 제시하는 것을 목적으로 하였다. 이를 위하여 최근 3년간 발생한 경기도 교통사고자료(2008-2010년)를 활용하여 보행자 교통사고의 발생특성을 분석하였다. 또한, 통계모형인 이항 로지스틱 회귀분석, 순서형 프로빗 모형을 이용하여 보행자 교통사고 심각도 증가에 통계적으로 유의한 영향을 미치는 변수를 도출하였으며, 휴리스틱모형인 서포트 벡터 머신, 의사결정나무를 적용하여 교통사고 심각도 분류를 위한 모형을 개발하고 그 결과를 비교분석하였다. 본 연구의 분석결과는 보행자 교통안전분석의 기초자료로 활용할 수 있으며 향후 국내 보행자-차량 충돌사고 분석시 유용하게 활용될 수 있을 것으로 기대된다.

Keywords

binary logistic regression, decision tree, ordered probit model, pedestrian-vehicle accident, support vector machine
이항 로지스틱 회귀분석, 의사결정나무, 순서형 프로빗 모형, 보행자-차량 사고, 서포트 벡터 머신

* : Corresponding Author
saerona@hanyang.ac.kr, Phone: +82-31-400-4503, Fax: +82-31-436-8147

Received 8 July 2013, Accepted 28 November 2013

I. 서론

보행자-차량 교통사고는 차량과 보행자가 충돌한 사고로, 2008-2010년 발생한 전체 교통사고 중 그 발생 비율이 30.7%(총 72,550건 중 22,295건)로 높게 나타나며 사망자는 무려 56.8%(사상자 총 1,770명 중 사망 1,006명)에 달한다. 사고 유형을 차대차, 차대사람, 차량단독 사고만을 고려할 때, 총 사고 중 보행자 사고(차대사람) 비율은 31%로 차대차 사고(66%)의 절반 수준으로 나타났으나 치사율(사망자/사고건수,%)은 4.51로 차량단독 4.49, 차대차 1.36과 비교할 때 가장 높게 나타나 이에 대한 대책이 시급하다(경기도 교통사고자료, 2008-2010년).

국내 관련 연구에서는 꾸준히 보행자 사고의 심각성에 대한 문제제기를 해왔음에도 불구하고 보행환경 개선, 교통안전 교육 등의 대책만을 제시하고 있다. 보다 체계적인 보행자 사고 분석방법론의 개발 및 적용을 통하여 보행자 사고 특성, 심각도 영향요인 분석을 수행하여 효과적인 보행자 사고감소 대책을 제시할 필요성이 있다.

본 연구는 차량과 사람이 충돌한 사고를 보행자 사고로 정의하고 보행자 사고특성 및 사고심각도 영향요인을 분석하였다. 통계분석과 데이터마이닝 적용시 각 기법에 따른 분석결과를 도출함으로써 보행자 교통사고 심각도 분석시 분석목적에 적합한 방법론을 제시하는 것을 목적으로 하였다. 먼저 국내외 보행자 교통사고 및 보행자 보호시스템 관련 연구를 종합적으로 검토하였다. 최근 3년간 발생한 경기도 사고자료(2008-2010년)를 활용하여 보행자 교통사고의 발생특성을 분석하였으며, 통계분석기법인 이항 로지스틱 회귀분석, 순서형 프로빗 모형을 이용하여 보행자 교통사고 심각도 증가에 통계적으로 유의한 영향을 미치는 변수를 도출하였다. 서포트 벡터 머신, 의사결정나무를 적용하여 교통사고 심각도 분류모형을 개발하였다. 또한, 각 기법에 따른 결과를 비교함으로써 보행자 사고분석시 분석목적에 적합한 방법론을 제시하였다.

II. 기존문헌고찰

1. 보행자 교통사고분석 관련 연구

기존 보행자 사고분석은 도로 환경, 차량특성, 보행자 특성에 주로 초점이 맞춰져 있다. 도로 환경 측면에서는

도로 종류, 즉 보행자 사고가 발생한 위치와 사고심각도 간의 관계를 도출하는 연구가 수행되었고(AI-Ghamdi, 2002), 차량 특성과 관련하여 보행자 사고는 차량과 충돌시 차량의 무게가 클수록 사고 심각도 증가에 영향을 준다는 결과가 출되었다(Ballesteros, 2004; Zajac et al., 2003). 보행자 특성과 관련해서는 보행자 중에서도 교통약자로 구분되는 고령자 또는 어린이 사고에 집중하여 사고심각도 모형 개발 및 사고특성 분석을 수행하였다(Park et al., 2010).

2. 분석방법론 관련 연구

1) 이항 로지스틱 회귀분석(BLR : Binary logistic regression)

이항 로지스틱 회귀분석을 활용하여 교통사고 영향요인을 도출하기 위한 연구가 수행되었다. 다양한 교통사고 중 화물차 교통사고(kim et al., 2012), 사고위험도로 분류된 사고(Park et al., 2009), 차량 종류별 사고(AI-Ghamdi, 2002) 등에 집중하여 교통사고 심각도의 확률적 증가에 영향을 미치는 요인을 도출하였다. 이와 유사하게 BLR을 이용하여 운전자특성, 도로특성, 환경특성, 사고유형특성 등이 사고심각도에 미치는 영향에 대해 종합적으로 검토하는 분석도 수행되었다(Sze et al., 2007).

일부 연구에서는 연령 및 성별로 분류된 집단에 대해 사고심각도를 분석하였으며(Roudsari et al., 2004), 보행자 음주여부, 속도제한 등과 사고심각도의 관계도출을 위한 분석을 수행하기도 하였다(Ballesteros et al., 2003). 계층 BLR을 이용하여 인명피해 및 차량피해정도를 예측하는 모형을 개발하기 위한 연구도 수행되었다(Moon et al., 2011). 회전교차로에서 자전거사고에 관하여 포아송 회귀분석과 로지스틱 회귀분석을 통한 통계적인 상관관계를 도출하기 위한 연구(Hels et al., 2007), 보행자 상해위험도에 영향을 미치는 인구 및 도로 환경의 영향 등 잠재적 요인을 도출하기 위한 연구 등이 수행되었다(Sze et al., 2007).

2) 순서형 프로빗 모형(OPM : Ordered probit model)

교통사고분석에의 순서형 프로빗 모형을 적용한 사례 연구를 살펴보면, 주로 사고심각도 영향요인 분석이 그 주를 이루고 있는 것을 볼 수 있다. 기상별 교통사고 심각도 영향요인 분석(Choi et al., 2012), 보행자 교

통사고 심각도 영향요인 분석(Tarko et al., 2011), 대형트럭사고의 부상심각도 영향요인 분석(Lemp et al., 2011), 공사장 교통안전 영향요인 분석(Hong et al., 2011) 등에 OPM이 활용되었다. 또한, 속도 및 교통량과 관련된 교통특성들과 교통사고 심각도와와의 관계분석(Christoforou et al., 2010), 다차로도로, 교차로 교통사고 등 특정 사고유형에 집중된 사고심각도 영향요인 분석(Kockelman et al., 2002) 등이 수행되었다.

3) 서포트 벡터 머신(Support Vector Machine)

서포트 벡터 머신을 활용한 기존 연구를 살펴보면, Eriksson(2001)은 운전자의 피로도를 감지하기 위해 운전자의 눈동자를 추적하는 시스템을 개발을 위한 연구에서 SVM을 이용하여 졸음운전을 분류하였다. 차량 번호판 인식, 보행자 검출 등에 SVM을 이용한 연구사례도 있었다(Wang et al., 2009). 타 분석기법과 분류성능 비교를 목적 SVM을 활용하기도 하였는데, 한 연구에서는 차량의 충돌을 예측하고 평가시 SVM 모형과 음이항 회귀분석을 이용하여 정확도 및 효율성을 비교하였다(Li et al., 2007). 교통사고심각도 분류를 위한 연구는 실시간 교통사고 위험도를 평가하기 위한 방법론으로 SVM을 활용(Yu et al., 2013)하였으며, 다항의 교통사고심각도 분류시 OPM과 비교하여 보다 향상된 분류성능을 가진 SVM 모형을 개발하기도 하였다(Li et al., 2012)

4) 의사결정나무(Decision Tree)

교통안전분석에서 의사결정나무를 활용한 연구사례는 교통사고 심각도 분류모형을 개발 및 타 기법과의 비교 분석으로 구분된다. 사고 심각도 분류모형 개발 연구는 분류트리를 적용하여 보행자의 교통사고 안전 분석기법의 향상 및 개발을 위해 충돌패턴의 상호의존성과 비유사성을 찾아내기 위한 연구(Montella et al., 2011)와 Classification and regression tree(CART)를 이용하여 편도 1차로의 교외도로에서 사고 발생시 사고 심각도에 영향을 미치는 위험요소를 분석(Kashani et al., 2009)한 연구가 있었다. 또한, CART를 이용하여 교통사고시 적절한 사고의 대응을 위해 성인과 미성년자의 구분 및 사고 심각도를 구분하는 연구도 수행되었다(Scheetz et al., 2009).

3. 시사점

국내외 보행자 교통사고 및 보행자 보호시스템 관련 연구를 검토한 결과, 국내 연구의 경우 보행자 사고에 대한 다양한 접근방법을 통해 보행자 사고 감소 대책을 마련하기 위한 연구를 수행한 바 있다. 보행자 교통안전 증진을 위해서는 인프라 개선, 교통류관리전략의 시행과 더불어 보다 능동적인 보행자 안전개선 방안 도출을 위한 연구가 필요한 실정이다.

국외에서는 다각적인 방법으로 보행자 사고특성에 대한 연구가 꾸준히 진행되어오고 있다. 국내에서도 일부 연구에서는 보행자 사고에 집중하여 사고심각도 영향요인 분석을 수행한 연구가 수행되었다. 또한 방법론 비교 관련 연구는 국내외에서 일부 진행되어 왔으나 단순히 모형의 분류성능을 평가하는데 그 목적이 있었다.

본 연구에서는 통계분석과 데이터마이닝 기법을 적용한 보행자 사고특성 및 사고심각도 영향요인 분석을 수행하였으며, 각 모형의 활용도 및 특성을 고려하여 분석 결과를 상세히 검토하였다. 이를 통하여 보행자 교통사고 분석시 분석 목적에 적합한 방법론을 제시하고자 하였다.

III. 보행자 사고특성 및 심각도 영향요인 분석

1. 보행자 교통사고 특성 분석

본 연구는 최근 3년간 발생한 경기도 사고자료(2008-2010년)를 활용하여 보행자 교통사고의 발생특성을 분석하였다. 보행자-차량 교통사고에서 보행자의 경우 머리, 다리 등의 신체에 충돌로 인한 충격을 직접적으로 받아 운전자에 비해 상해 정도가 매우 높다. Table 1과 같이 보행자-차량 사고 발생시 운전자의 경우 상해 없음이 97.93%로 나타났지만 보행자의 경우 사망, 중상, 경상이 차지하는 비율이 97.89%로 나타났다. 사고발생시 운전자의 경우 안전띠, 에어백 등의 보호장비를 통하여 사고심각도를 경감시킬 수 있으나, 보행자의 경우 사고 발생시 어떠한 보호장치도 없어 충격량이 직접적으로 신체에 전달되기 때문에 그 심각도가 매우 높다. 특히, 차량-보행자 충돌시 운전자가 안전속도를 준수하지 않고 과속할 경우 Table 2와 같이 치사율이 매우 높게 나타남을 알 수 있다. Table 3의 도로 위계별 보행자의 치사율은 일반국도(6.72), 지방도(5.30) 등 상대적으로 도로 규격이 커 주행속도가 높을 개연성이 있는 도로에서

발생한 사고가 높은 값을 보였다. 이는 차량의 주행속도 측면에서 볼 때, 보행자-차량 사고같은 구간에서의 속도 감소를 위한 인프라 개선, 안전시설 보완으로도 보행자 안전 효과가 있을 것으로 판단된다.

그러나 야간, 안개발생 등 운전자 시인성 저하와 밀접한 관련이 있는 불리한 도로환경 조건에서 발생한 보행자 사고의 경우 차량의 속도관리만으로는 안전성 향상은 어려울 수 있다. Table 4와 Figure 1에서 보는 바와 같이 안개의 경우 22.03%로 맑음(4.09%), 흐림(8.78%)보다 매우 높게 나타나며 야간의 경우 치사율은 5.82%로 주간 3.60%보다 높다.

2. 보행자 사고심각도 영향요인 분석

보행자 사고특성 분석결과에서 제시한 단순한 사고건수 비율, 치사율 등으로는 한 요인이 사고심각도에 미치는 영향의 정도를 상세히 분석하기 어렵다. 또한, 교통사고는 단일 요인이 아닌 복합적인 요인의 작용으로 발생하기 때문에 통계모형 및 데이터마이닝 기법 적용을 통한 변수간 상호작용 또한 검토할 필요성이 있다.

본 연구에서는 분석방법론으로 이항 로지스틱 회귀분

석, 순서형 프로빗 모형의 통계분석기법과 서포트 벡터 머신, 의사결정나무의 데이터마이닝 기법을 활용하여 분석을 수행하였다.

위의 분석방법은 지도학습(Supervised learning)이면서 다변량분석(Multivariate procedure)방법이라는 측면에서는 유사하나 BLR과 OPM은 각 독립변수가 종속변수의 확률적 증가에 미치는 영향에 대한 예측시 활용한다. BLR은 종속변수가 이분형(예: 사망유/무)일 때 적용하며 OPM은 종속변수가 순서형(예: 사망, 중상, 경상, 부상)일 때 적용한다. 반면 DT와 SVM은 주어진 자료에 대해 그 자료를 분리하는 기준을 구축함으로써 새로운 자료가 주어졌을 때 분리 기준에 따라 해당 데이터를 판별하는 데이터마이닝 기법, 즉 기계학습 알고리즘(Machine learning algorithm)의 한 종류이다. 각 방법론에 대한 자세한 설명은 본 장의 '4) 분석방법론'에 소개하였다.

1) 분석절차

본 연구의 분석절차는 Figure 2와 같다. 먼저 BLR 분석을 통하여 사고심각도 영향요인을 도출한다. 통계

Table 1. Comparison of injury severity in pedestrian and driver

Accident rate by pedestrian injury severity(%)								
Accident rate by driver injury severity (%)	Injury Severity	Fatal	Severe	Injured	Slight	Non injured	Others	Total
	Fatal	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Severe	0.04	0.13	0.03	0.00	0.02	0.00	0.00	0.22
Injured	0.02	0.06	0.05	0.00	0.00	0.00	0.00	0.14
Slight	0.00	0.07	0.03	0.00	0.01	0.00	0.00	0.11
Non injured	4.40	54.85	36.64	2.04	0.00	0.00	0.00	97.93
Others	0.04	0.52	1.00	0.04	0.00	0.00	0.00	1.60
Total	4.51	55.63	37.75	2.07	0.04	0.00	0.00	100.00

Table 2. Characteristics of pedestrian accidents by violation of a regulation

Contents of violation of a regulation	%	Fatality rate(%)
Speeding	0.17	45.95
Others	2.89	2.81
Violation of duty to protect pedestrians	11.84	2.25
illegal U-turn	0.65	1.38
Violation of the signal	9.28	3.69
Do not assure a safe distance	0.20	11.11
Do not safe driving	73.91	4.99
Violation of central line	1.06	3.83

Table 3. Pedestrian injury severity by level of road

Level of road	Fatality rate(%)
Freeway	36.11
County Highway	3.93
Others	1.68
Urban Highway	3.09
National Roadways	7.46
Rural Highway	5.78

Table 4. Pedestrian injury severity by weather condition

Weather condition	Fatality rate(%)
Sunny	4.09
Cloudy	8.78
Rain	5.71
Snow	3.86
Fog	22.03
Others	4.76

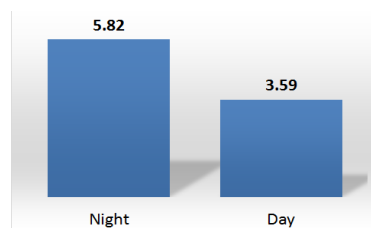


Figure 1. Pedestrian injury severity by day/night

적으로 유의하게 도출된 변수의 계수, 계수의 부호, Odds-ratio 등을 상세히 검토하여 각 요인이 보행자 사고 심각도에 미치는 영향에 대한 상세한 분석을 수행한다.

BLR 분석에서 유의하게 도출된 변수를 활용하여 DT, SVM 분석을 수행한다. DT나 SVM의 경우 기계 학습알고리즘이기 때문에 입력변수의 설정은 일반적으로 분석자의 주관적 판단에 따른다. 따라서 본 연구에서는 이러한 입력변수 선정 기준으로 통계적 유의성을 가지는지 여부를 적용하기 위하여 BLR 분석결과를 활용하였다. DT와 SVM 분석결과로 도출되는 분류정확도(CCR : Correct classification rate)는 분류기 성능을 비교시 활용하였다. 동일 데이터에 대한 분류성능을 비교해야 하므로 분류성능 비교시 DT와 SVM 모두 SVM의 테스트 데이터를 활용하여 CCR을 도출했다.

OPM은 BLR과 비교했을 때 보다 다양한 사고심각도를 고려한 분석이 가능하다. OPM의 Pseudo 탄력성과 한계효과를 검토하여 각 독립변수가 종속변수의 확률적 증가에 미치는 영향을 파악한다.

각 모형별 분석결과를 통하여 보행자 사고 분석시 분석목적에 따른 적절한 방법론을 제시하였다.

2) 데이터베이스 설정

본 연구에서 사용한 데이터베이스는 2008-2010년의 3

년간 경기도에서 발생한 사고자료이다. 차대차, 차대사람, 차량단독 사고 중 사고유형이 차대사람인 사고만을 분류하였으며, 피의자 및 피해자가 각각 1명인 사고만을 고려하였는데 여러대의 차량 또는 여러명의 보행자가 관련된 사고의 경우 본 분석에서 제외하였다.

데이터 중 주로 운전경력, 면허종별 등에 결측치가 발생하였는데, 분석의 신뢰성을 위하여 해당 케이스는 제거하였다. 최종 데이터베이스의 사고케이스 총 개수는 22,300개로 나타났다.

3) 변수설정

분석을 위하여 종속변수(또는 출력변수) 및 독립변수(또는 입력변수)를 설정하였다. 종속변수의 경우 사고심각도를 설정하였는데, BLR, DT, SVM의 경우 이분형 종속변수를 사용해야 하므로 사망, 중상의 심각한사고를 1로, 경상, 부상, 상해없음의 상대적으로 경미한사고를 0으로 설정하였다. 종속변수의 형태가 순서형으로 나타나는 OPM의 경우는 변수의 형태를 사망, 중상 및 경상, 부상이하로 집합화된 순서형으로 코딩하였다.

독립변수의 경우 도로요인, 환경요인, 운전자요인, 차량요인과 관련된 변수들을 추출하여 코딩하였다. 도로요인은 도로유형 및 사고위치이다. 도로유형(Road type)은 고속국도, 일반국도, 시도, 지방도, 군도로 나타나며 각 도로의 설계속도를 코딩하였다. 사고위치 변수는 차량

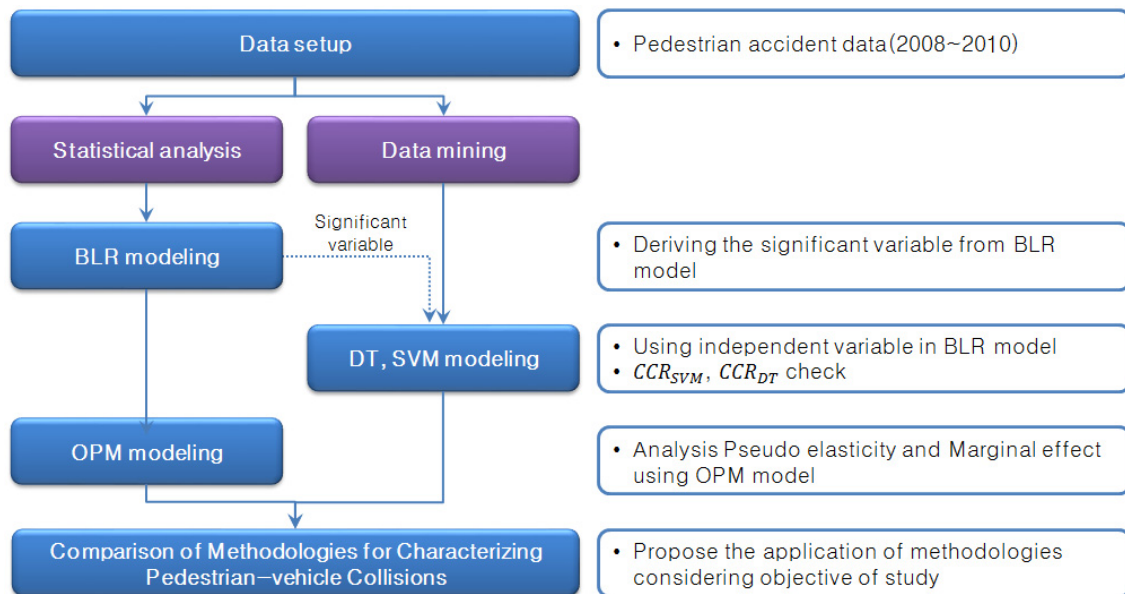


Figure 2. Procedure of analysis

Table 5. Description of variables

		BLR/DT/SVM		OPM			Total
		Minor	Severe	Minor	Severe	Fatal	
Day/ night	Day	5,153	8,006	270	12,416	473	13,159
	Night	3,736	5,405	201	8,408	532	9,141
Location	sidewalk	4,158	4,551	221	8,221	267	8,709
	driveway	4,731	8,860	250	12,603	738	13,591
Weather	Sunny	8,158	11,984	430	18,834	878	20,142
	Others	731	1,427	41	1,990	127	2,158
Vehicle type	Car	6,988	9,751	372	16,076	699	16,739
	Heavy vehicle	1,901	3,660	99	4,748	306	5,561
Alcohol	Non-drinking	8,391	12,843	440	19,839	955	21,234
	Drinking	498	568	31	985	50	1,066
Total		8,889	13,411	471	20,824	1,005	22,300

Table 6. Set up the variable

Variable		Description
Dependent variable (Output variable)	Injury severity	BLR/DT/SVM - Severe=1 : Fatal, Severe - Minor=0 : Slight, non-injured
		OPM: Fatal=2, Severe=1, Minor=0
Independent variable (Input variable)	Day/night	Day=0, Night=1
	Location	In driveway=1, In sidewalk=0
	Weather	Others=1, Sunny and cloudy=0 - Others : rain, snow, fog
	Vehicle type	Heavy vehicle=1, Car=0
	Driving experience	
	Road type	Application of design speed by road type
	Alcohol	Drinking=1, Non-drinking=0

의 속도가 높아 상대적으로 사고심각도가 높게 나타날 것으로 판단되는 차도를 1로, 그렇지 않은 보도는 0으로 코딩하였다. 환경요인은 주야, 기상조건을 고려하였다. 주간은 0으로, 야간은 1로 코딩하였으며 맑음과 흐림을 0으로, 비, 눈, 안개와 같은 기상악화 조건을 1로 코딩하였다. 운전자요인은 연속형 변수인 운전경력과 음주시 1, 비음주시 0으로 코딩된 음주여부를 변수화하였다. 차량요인은 트럭, 버스와 같은 중차량을 1로, 승용차, 소형승합차를 0으로 코딩하였다. Table 5에는 전체 데이터베이스의 명목형 변수에 대한 기술통계량을, Table 6에는 종속변수 및 독립변수에 대한 설정방법을 제시하였다.

4) 분석방법론

(1) 이항 로지스틱 회귀분석

이항 로지스틱 회귀분석(BLR)은 종속변수가 이분형으로 관측된 데이터를 적합 시킬 때 유용하게 사용되는 기법으로, 모형의 종속변수에서 심각한사고일 경우를 1로, 경미한사고일 경우를 0으로 설정하는 Binary classification 문제에 효과적으로 응용가능하다. 또한 승산비(Odds-ratio), Wald 통계량 등을 이용하여 종속변수의 확률적 증가에 통계적으로 유의한 영향을 미치는 독립변수 판단시 유용하게 사용할 수 있으며, 이러한 통계량을 활용하여 독립변수간 중요도를 평가하는데 활용 가능하다. 이항 로지스틱 회귀분석의 기본 수식은 식 (1)과 같다.

$$P(y = 1|x_1, \dots, x_i) = \frac{\exp[f(x_i, \beta_i)]}{1 + \exp[f(x_i, \beta_i)]} \quad (1)$$

$P(y = 1|x_1, \dots, x_i)$ = Fatal accident occurrence probability

x_i : Factors affecting pedestrian accident

β_i : Estimated parameters

모형내 변수의 계수의 부호가 양(+)인 경우 해당 변수는 심각한 사고발생의 확률적 증가에 영향을 미친다고 해석할 수 있으며, 부호가 음(-)인 경우 심각한 사고발생 확률을 감소시키는 것을 의미한다. 로지스틱 회귀분석의 결과물 중 하나인 독립변수의 승산비(Odds-ratio)는 다른 변수가 고정되었을 때 해당 독립변수의 위험도를 설명할 수 있다.

(2) 서포트 벡터 머신

서포트 벡터 머신(Support Vector Machine, SVM)은 패턴인식 분야에서 이진분류 문제를 해결하는 분류기술로서, 집단이 두 개인 경우 통계학적 이론을 배경으로 오분류율(Misclassification rate)을 최소화 시키는 최적분리 초평면(Hyperplane)을 제공함으로써 다양한 응용분야의 분류 및 회귀문제에서 우수한 성능이 입증되고 있다.

SVM의 종류에는 선형적으로 분리할 수 있는 학습 집단에 대해서 최대 마진 분류기를 구축하는 선형 SVM과, 선형적으로 분리할 수 없는 경우에 커널 함수에 의해 만들어진 비선형 결정함수를 이용하여 최적의 초평면을 구축하는 비선형 SVM이 있다. 최적의 경계면은 분류할 두 집단으로부터 가장 멀리 떨어진 초평면으로 정의되며,

경계면에 가장 가까이 있는 데이터를 서포트 벡터(Support vector)라 한다.

학습 데이터를 비선형 매핑함수 $\phi(x)$ 를 이용하여 입력 샘플 공간 X 로부터 선형 초평면이 만들어지는 고차원 특정 공간 F 로 매핑하고 인식 에러가 최소가 되게 하는 최적 초평면을 찾는다. 최적 초평면에 대한 결정 함수식은 식(2)와 같다.

$$f(x) = \sum_{i=1}^n \alpha_i t_i K(x, x_i) + b \tag{2}$$

식(2)에서 α_i 는 목적함수를 최소화 하는 계수이며, α_i 가 0이 아닌 경우의 x_i 를 초평면을 구성하는 샘플(Support vector)로 정의한다. t_i 는 입력변수의 그룹을 나타내며, b 는 Bias를 나타낸다. 커널함수 $K(x, x_i)$ 는 비선형 매핑 함수의 내적 계산을 함수 형식으로 치환하여 계산을 간단하게 해주며, 선형(Linear), 다항식(Polynomial), RBF(Radial basis function) 등의 3가지 형태의 커널함수가 주로 사용된다.

본 연구에서는 선형 커널함수를 사용하여 보행자교통사고 심각도 분류 알고리즘에 적용하였다. 식(2)을 통해 계산된 $f(x)$ 의 값이 0보다 작은 값을 가지면 집단 1(심각한 사고), 0보다 큰 값을 가지면 집단 2(경미한 사고)로 분류한다. SVM은 명백한 이론적 근거에 기반을 두므로 결과 해석이 용이하고, 실제 응용에 있어서 인공신경망 보다 더 좋은 성과를 내주며 적은 학습 자료만으로 신속하게 분류 및 회귀학습을 수행할 수 있어 다방면의 연구분야에서 활용되고 있다.

(3) 의사결정나무

의사결정나무(DT)는 변수가 범주형으로 관측된 데이터를 분석하는데 적합하며 독립변수들 중 설정된 클래스(종속변수)에 가장 많은 영향을 미치는 변수를 선정하여 선정된 변수의 특성에 따라 입력데이터를 분류함으로써 최종 클래스를 결정하는 분석방법이다. 계량 데이터만 다룰 수 있는 신경망 또는 SVM과는 달리 계량데이터뿐 아니라 다양한 특징을 지닌 데이터도 다룰 수 있다. 또한 데이터가 루트노드부터 최종노드까지 거처온 경로가 명확하게 나타나기 때문에 어떤 부류로 분류가 되었는지 해석이 가능하며, 인식시간이 빠르다는 장점을 갖는다.

DT 기법은 일정한 알고리즘을 통해 유사한 특성을 가진 데이터를 작은 집합으로 분류하여 더 이상 분류가 이루어지지 않는 마지막 노드(terminal node)에는 최종

적으로 그 데이터가 속하는 분류결과(Class)가 결정된다. DT 구축 시 최적의 분류트리는 최종노드에 각각 동일한 데이터가 담기며 최소의 노드로 구성된 트리이다. 이때 이러한 동질성을 측정해주는 기준이 불순도인데 불순도는 엔트로피, 지니불순도, 오분류불순도를 이용하여 산출할 수 있다. 본 연구에서는 엔트로피를 이용하였으며, 이때 엔트로피 값이 낮을수록 분류의 확실성이 높다는 것을 나타내고 불순도는 낮으므로 좋은 부류라 볼 수 있다. 엔트로피를 이용한 불순도 산출식은 식(3)과 같다.

$$im(T) = - \sum_{i=1}^M P(\omega_i|T) \log_2 P(\omega_i|T) \tag{3}$$

ω_i : M개의 부류 중 i 번째
 $P(\omega_i|T)$: 노드 T에서 ω_i 가 발생할 확률

불순도는 하나의 노드를 대상으로 측정하며 분기된 각 노드에 속한 데이터는 모두 같은 부류에 속하는 것이 좋으며 가급적 불순도는 낮은 것이 좋다. 불순도 감소량은 식(4)와 같이 나타낼 수 있다.

$$\Delta im(T) = \left[im(T) - \frac{|X_{Tleft}|}{|X_T|} im(X_{Tleft}) \right] - \left[\frac{|X_{Tright}|}{|X_T|} im(X_{Tright}) \right] \tag{4}$$

X_T : 상위노드(root node)
 X_{Tleft} : 노드 X_T 에서 좌측으로 분기된 노드
 X_{Tright} : 노드 X_T 에서 우측으로 분기된 노드

본 연구에서 사용한 ID3은 DT 구축시 사용되는 알고리즘의 한 종류로써, 비교적 간단하고 잘 알려진 시스템으로 명목형 이진 속성의 데이터에 활용이 유리하며 노드를 분기할 때 엔트로피를 측정한 후 정보의 이득(Information gain)에 근거하여 노드를 분기한다. 엔트로피는 불확실성을 계량화하기 위한 척도이며 값이 낮을수록 확실성은 증가한다. 정보이득은 이러한 엔트로피의 감소정도를 나타내며 데이터 분류 시 해당 속성의 효율성을 계량화하기 위한 척도로 값이 높을수록 확실성은 증가한다. 엔트로피를 이용한 정보이득을 산출하는 방법은 식(5)과 같다.

$$Gain(D, A) = Entropy(D) - \sum \left\{ \frac{|D_v|}{|D|} Entropy(D_v) \right\} \tag{5}$$

D : 데이터 d_i 로 이루어진 전체 데이터 집합
 A : 속성 a_i 로 이루어진 속성 집합
 D_v : D 의 부분집합

정보이득을 통하여 산출된 속성 값 중 가장 큰 값을 선정하여 데이터를 분류한다. 이러한 분기는 멈춤조건 임계값 또는 데이터의 수에 도달하여 더 이상 분류가 되지 않을 때까지 반복하여 분류가 수행되며 최종노드에서는 분류정확도를 산출한다.

(4) 순서형 프로빗 모형

순서형 프로빗 모형(OPM)은 순서화된 다항의 선택을 다룰 수 있도록 고안된 확률선택모형 중 ϵ_i (오차항)의 분포를 정규분포로 가정하는 모형이다. OPM은 종속변수가 교통사고 심각도와 같이 순위가 있는 이산형 변수이고 다양한 요인이 작용하는 교통현상에 대한 분석시 유용하게 사용할 수 있으며 한계효과 및 탄력성 검토를 통하여 각 독립변수들이 종속변수의 확률적 증가에 미치는 영향의 정도를 분석할 수 있다.

OPM의 기본적인 형태는 식(6)과 같다. 응답 가능한 항이 J 개라고 한다면 관찰 불가능한 변수 y^* 와 관찰 가능한 변수 y_i 과의 관계는 식(7)와 같이 가정한다. μ 는 각 설명변수 x 의 추정계수 β 를 취하여 추정할 수 있는 한계값이 된다. 또한 종속변수의 척도에 따른 각 선택대안에 대한 선택 확률 계산식은 식(8)과 같으며, 이를 통해 개개의 확률 값을 추정할 수 있다. ϕ 와 Φ 는 각각 표준정규분포의 확률밀도함수이다.

$$y^* = \beta'x + \epsilon \tag{6}$$

$$y = \begin{cases} 0, & \text{if } y_i^* \leq 0 \\ 1, & \text{if } 0 < y_i^* \leq \mu_1 \\ 2, & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ \vdots \\ J, & \text{if } \mu_{J-1} < y_i^* \end{cases} \tag{7}$$

$$P(y_i = 0) = \int_{-\infty}^{-\beta'_i x_i} \phi(\epsilon_i) d\epsilon_i = \Phi(-\beta'_i x_i) \tag{8}$$

$$P(y_i = 1) = \int_{-\beta'_i x_i}^{\mu_1 - \beta'_i x_i} \phi(\epsilon_i) d\epsilon_i = \Phi(\mu_1 - \beta'_i x_i) - \Phi(-\beta'_i x_i)$$

$$P(y_i = 2) = \int_{\mu_2 - \beta'_i x_i}^{\mu_2 - \beta'_i x_i} \phi(\epsilon_i) d\epsilon_i = \Phi(\mu_2 - \beta'_i x_i) - \Phi(\mu_1 - \beta'_i x_i)$$

$$\vdots$$

$$P(y_i = J) = \int_{\mu_{J-1} - \beta'_i x_i}^{\infty} \phi(\epsilon_i) d\epsilon_i = 1 - \Phi(\mu_{J-1} - \beta'_i x_i)$$

OPM은 종속변수가 연속형이 아닌 리커트척도와 같은 순차적 다분변수일 경우 독립변수의 크기를 잘못 평가할 수 있는 선형회귀분석의 단점을 보완해 줄 수 있을 뿐만 아니라 복잡한 원인의 교통사고 분석에 유용하게 활용할 수 있다. 또한 OPM의 한계효과 및 Pseudo 탄력도 검토결과는 설명변수들이 종속변수의 발생확률 증가에 미치는 영향의 정도를 나타낸다. 주로 탄력성은 명목형 변수일 때, 한계효과는 연속형 변수일 때 사용한다.

3. 보행자 사고심각도 영향요인 분석 결과

1) 이항 로지스틱 회귀분석 결과

이항 로지스틱 회귀분석(BLR)은 상용화된 통계패키지인 SPSS를 활용하여 분석을 수행하였다. BLR 분석 결과는 Table 7과 같다. 유의한 독립변수는 사고위치, 기상조건, 도로유형, 운전경력, 차종으로 나타났다. 주야, 음주여부는 변수의 계수가 비상식적으로 도출되는 결과가 나타나 모형에서 제외되었다.

Table 7. BLR results

*** : p<0.001, ** : p<0.01

Model summary			Hosmer and Lemeshow test			Classification Table					
-2Log likelihood	Cox and Snell R ²	Nagelkerke R ²	χ ²	d.f	유의확률	Predicted \ Observed		Minor	Severe	CCR(%)	
						Minor	Severe				
29514.184	0.021	0.029	3.019	8	0.933	821	8068	9.24			
							754	12657	94.38		
							Total			60.44	
		Coef.	Std. Err	z	d.f	P> z	Odds.	95% Conf. Interval			
								Lower	Upper		
Location***		0.510	0.028	326.619	1	0.000	1.666	1.576	1.761		
Weather***		0.245	0.048	25.977	1	0.000	1.278	1.163	1.405		
Road type***		0.014	0.002	55.906	1	0.000	1.014	1.010	1.018		
Driving experience***		0.079	0.034	5.312	1	0.021	1.082	1.012	1.157		
Vehicle type**		0.009	0.003	10.353	1	0.001	1.009	1.003	1.014		
Constant		-1.003	0.134	56.194	1	0.000	0.367	-	-		

분석결과를 살펴보면, 보행자가 차도통행 중 발생한 사고가 보도통행 중 발생한 사고보다 심각한 사고발생 확률이 1.7배 큰 것으로 나타났다. 차도에서 발생한 사고의 심각도는 차량의 주행속도를 감안할 때, 보도에 비해 사고심각도 증가에 대한 영향이 크다고 볼 수 있다.

기상조건 변수는 비, 눈, 안개 등 운전자 시인성이 저하되는 조건이 맑음, 흐림보다 심각한 사고발생 확률이 1.3배 증가하였다. 도로유형의 경우 설계속도가 높은, 즉 도로의 규격이 큰 도로에서 발생한 사고일수록 보행자 사고발생시 심각도가 큰 것으로 나타났으며 차종의 경우 승합, 화물 등 중차량일 경우 승용차보다 보행자 사고심각도가 높았다. BLR 모형식은 식(9)와 같다.

$$P = \frac{\exp(f(x))}{1 + \exp(f(x))} \tag{9}$$

$$f(x) = -1.003 + 0.510 * X_{\text{사고위치}} + 0.245 * X_{\text{기상조건}} + 0.014 * X_{\text{도로유형}} + 0.009 * X_{\text{차종}} + 0.079 * X_{\text{운전경력}}$$

2) 의사결정나무 분석결과

본 연구에서는 엔트로피 및 정보이득을 이용한 의사결정나무(DT)를 구축하기 위하여 MATLAB을 이용하였다. 입력변수는 BLR 모형에서 통계적 유의성이 있다고 도출된 사고위치, 도로유형, 차종, 운전경력, 기상조건 변수를 적용하였다.

DT 알고리즘 설정시 가지치기(Branching)의 임계값 설정시 분류정확도가 가장 높게 나타날 수 있도록 임계값을 설정해야하며, 본 연구에서는 휴리스틱한 기법을

이용하여 분류정확도를 최대한 하도록 임계값을 0.93으로 설정하였다. 이 때, 가지치기란 하위노드로의 데이터 분류를 의미한다. 또한, 각 노드에서 엔트로피 값이 0.93 이하로 산출될 경우 가지치기를 멈추도록 하였다. DT 알고리즘에서 사용한 5개(사고위치, 도로유형, 차종, 운전경력, 기상조건)의 입력변수의 수에 따라 트리의 깊이를 최대 5회로 설정할 수 있으나 데이터가 과적합(Overfitting)되는 문제로 인하여 트리의 깊이는 n-1값인 최대 4회로 설정하였다.

Figure 3의 DT 구축결과를 살펴보면, 의사결정나무 구축 시 사용된 변수 즉, 사고심각도에 영향을 미치는 중요변수로는 사고위치(Node 1), 도로유형(Node 2), 차종(Node 3)으로 나타났다. 운전경력, 기상조건은 분류 트리의 입력변수에서 제외되었다. 가장 상위노드로 사고 위치가 선정되었는데, 이는 사고심각도 결정에 가장 큰 영향을 미치는 것을 의미한다. 각 데이터가 최종적으로 분류된 노드를 의미하는 Terminal node에서 Class 1은 경미한사고, Class 2는 심각한 사고를 의미한다.

먼저 사고위치(Location, Node 1)가 보도(Sidewalk)인 경우를 살펴보자. 도로유형(Road type, Node 2)이 상대적으로 규격이 작은 시도, 군도, 지방도 일 때 차종(Vehicle type, Node 3)에서 상대적으로 규격이 큰 승합차, 화물차 등의 중차량이 발생시킨 사고의 경우 심각한 사고의 Class 2로 구분되는 것을 볼 수 있다(Terminal node 2, CCR 52.60%). 차종(Node 3)이 승용차 사고의 경우 경미한사고인 Class 1로 구분되었다(Terminal node 1, CCR 51.18%). 이는 고속도로, 일반국도 보

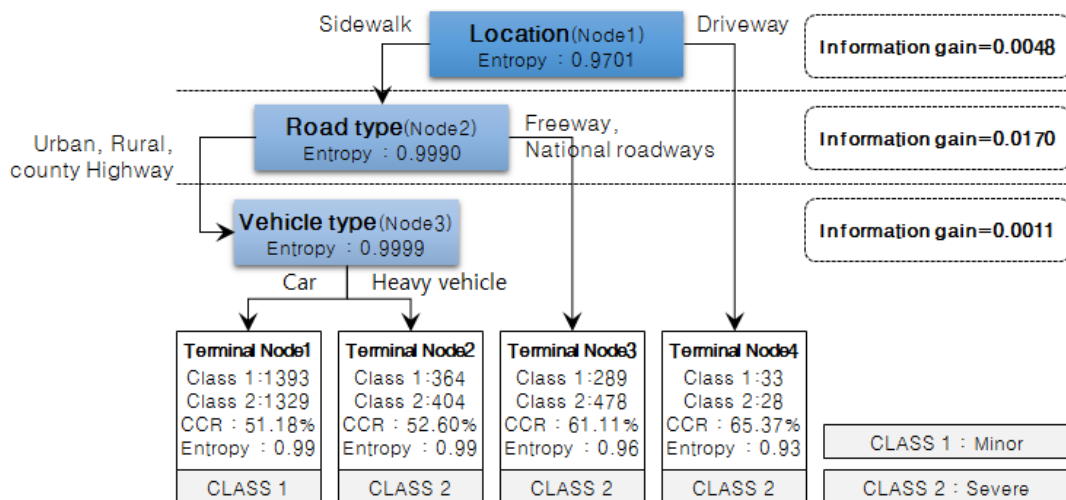


Figure 3. The result of decision tree

Table 8. DT results

Terminal Node	N of class 1	N of class 2	Entropy	Classified class	CCR
1	1393	1329	0.9996	1	51.18
2	364	404	0.9980	2	52.60
3	322	506	0.9641	2	61.11
4	2366	4467	0.9307	2	65.37

다 상대적으로 주행속도가 낮은 개연성이 있는 시도, 군도 등에서 발생한 사고라 할지라도 보행자와 충돌하는 차량의 규격이 큰 중차량일 경우 사고의 심각도가 높아 짐을 의미하는 결과라고 할 수 있다.

반면, 도로유형(Node 2)에서 고속도로, 일반국도 등 상대적으로 규격이 큰 도로에서는 차종에 관계없이 심각한 사고인 Class 2로 구분되는 것을 볼 수 있다(Terminal node 3, CCR 61.11%). 이는 BLR 분석결과와 마찬가지로 규격이 큰 도로의 경우 차량의 주행속도가 높을 개연성이 있으므로 이에 따라 사고심각도가 증가함을 알 수 있는 결과이다.

사고위치(Node 1)가 차도(Driveway)일 때를 살펴 보면, 케이스가 CCR 65.37%로 Class 2로 분류되는 것을 볼 수 있다(Terminal node 4). 결국, 보행자가 차도 통행 또는 횡단중에 발생한 사고는 그 심각도가 높다는 것을 알 수 있다.

Figure 3. 우측에 제시된 정보 습득(Information gain)은 노드(1-3)가 추가됨으로 인하여 엔트로피가 감소하는 양을 나타낸 것이다. 엔트로피가 감소했다는 것은 분류 불확실성에 대한 감소를 의미한다. 사고위치(Node 1) 변수가 추가되어 엔트로피가 0.0048 감소하였으며, 도로유형(Node 2)이 추가되어 0.0170 감소, 차종(Node 3)이 추가되어 엔트로피는 0.0011이 감소하는 결과가 도출되었다.

Table 8에는 각 Terminal node에서 Class 1 또는

Table 9. Correct classification rate of DT results

Observed \ Predicted	Minor	Severe	CCR(%)
	Minor	1391	
Severe	1381	5325	79.41
Total average CCR			55.35

2로 분류된 케이스의 개수(N of class 1, N of class 2), 엔트로피, 최종 분류된 클래스, CCR을 제시하였다. 사고심각도에 대한 DT 모형의 최종 분류정확도는 60.23%로 도출되었다.(Table 9).

3) 서포트 벡터 머신 분석결과

본 연구에서는 사고심각도 분류를 위한 SVM 분류기를 구축하기 위하여 MATLAB을 이용하였다. 입력변수는 DT모형과 마찬가지로 BLR 모형에서 통계적 유의성이 있다고 도출된 사고위치, 도로유형, 차종, 운전경력, 기상조건 변수를 적용하였다

Figure 4에 사고심각도 분류를 위한 SVM 분류기 구축 과정을 제시하였다. 전처리 단계에서는 SVM 분류기 구축을 위해 BLR을 통해 도출한 입력변수의 Training set을 준비한다. 준비된 Training set을 이용한 학습과정을 통해 사고심각도 분류를 최적화 하는 α_i (라그랑제 승수), x_i (Support vector), b(bias)를 도출하여 최종적인 SVM 분류기를 구축하게 된다. 최종적으로 생성된 SVM 분류기에 Test set을 적용하여 SVM 분류기의 성

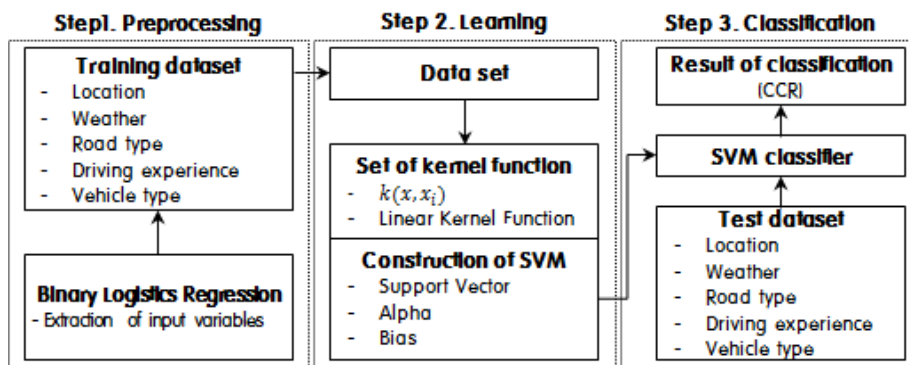


Figure 4. Procedure of constructing SVM classifier

Table 10. Correct classification rate of SVM results

Predicted \ Observed	Minor	Severe	CCR(%)
Minor	2079	2365	46.78
Severe	2312	4393	65.52
Total average CCR			56.15

능을 검증한다.

구축된 SVM 분류기는 9742개의 support vector와 α 를 가지며, bias는 -0.2191로 도출되었다. Test set의 입력변수를 SVM의 입력값으로 대입하여 산출된 $f(x)$ 값의 부호에 따라 사고심각도를 판단하게 되며, 사고심각도 분류결과는 Table 9에 제시하였다. 분류결과 CCR이 56.15%로 도출되었다. DT CCR인 55.35%와 비교했을 때 분류성능이 보다 높게 나타나는 것을 알 수 있다.

4) 순서형 프로빗 모형 분석결과

순서형 프로빗 모형(OPM) 분석은 상용화된 통계패키지인 STATA를 활용하여 분석을 수행하였다. OPM 분석결과는 Table 10과 같다. 음주여부 변수는 BLR에서와 같이 변수의 부호가 비상식적으로 도출되어 모형에서 제외하였다.

주야, 사고위치, 기상조건, 차종의 명목형 변수에 대해서는 Pseudo 탄력도를 검토하였다. 분석결과 운전자의 시인성이 저하되는 야간, 악천후시 사고심각도가 증가하는 것으로 나타났다. 이러한 도로환경은 경미한 사고 발생확률을 증가시키지는 않지만 사망사고의 발생확률을 야간의 경우 약 30%, 악천후의 경우 약 16% 증가시키는 것으로 도출되었다. 사고위치의 경우에도 보행자-차

량 사고발생시 보행자 위치가 차도일 경우 사망사고 발생확률이 약 50% 증가하는 것으로 도출되었다. 차종이 승합, 화물차 등의 중차량일 경우 보행자의 사망확률은 약 24% 증가하였다. 도로유형, 운전경력의 연속형 변수에서는 한계효과를 검토하였다. 도로의 위계, 즉 도로의 규격이 클수록, 운전경력이 증가할수록 보행자사고의 심각도별 발생확률이 모두 증가하는 것으로 도출되었다.

BLR 분석결과에서는 통계적으로 유의한 독립변수에 따른 심각한 사고, 경미한 사고의 발생확률만을 분석할 수 있었으나 보다 다양한 종속변수를 적용한 OPM 분석 결과, 보행자의 사망사고 발생확률에 대한 분석이 가능하였다.

IV. 결론

본 연구에서는 BLR, OPM, DT SVM의 분석방법을 활용하여 보행자 사고심각도 분석을 수행하였으며, 각 기법에 따른 결과를 비교함으로써 보행자 사고분석시 분석목적에 적합한 방법론을 제시하였다.

첫째, 통계모형의 분석결과를 요약하면 다음과 같이 정리할 수 있다. 불리한 기상조건, 야간 등의 운전자의 시인성 저하 조건에서 심각한 보행자 사고의 발생 가능성이 증가하는 것으로 나타났다. 이러한 도로환경에서는 운전자의 보행자 인지 저하로 인하여 회피행동이 어렵게 되므로 효과적으로 보행자 검출 및 충돌회피가 가능한 능동적 보행자 보호 시스템의 적용이 필요할 것이다. 또한, 사고 발생시 보행자가 차도 통행, 횡단 중 발생한 사고일 때, 즉 차량의 속도가 높은 개연성이 있는 위치에서는 심각한 사고발생 확률이 증가하는 결과가 도출되었

Table 11. OPM results

*** : p<0.001, * : p<0.1

	Coef.	Std. Err	z	P> z	95% Conf. Interval		Pseudo elasticity/Marginal effects		
					Lower	Upper	Minor	Severer	Fatal
Day/night***	0.128	0.025	5.170	0.000	0.079	0.176	-0.267	-0.006	0.312
Location***	0.187	0.025	7.480	0.000	0.138	0.236	-0.363	-0.007	0.497
Weather*	0.069	0.040	1.740	0.082	-0.009	0.147	-0.156	-0.004	0.158
Road type***	0.015	0.002	9.640	0.000	0.012	0.018	0.00008	0.00008	0.00014
Vehicle type***	0.100	0.029	3.440	0.001	0.043	0.157	-0.217	-0.005	0.235
Driving experience***	0.009	0.002	3.750	0.000	0.004	0.013	0.00011	0.00010	0.00021
/cut1	-0.729	0.110	-	-	-0.945	-0.513	-	-	-
/cut2	3.055	0.114	-	-	2.832	3.279	-	-	-
Log likelihood	-6236.159								
Number of obs	22300.000								
LR $\chi^2(7)$	243.700								
Prob > χ^2	0.000								
Pseudo R ²	0.019								

다. 이러한 경우 차량의 속도관리를 통해 사고심각도 감소가 가능하며 Active hood, 보행자 에어백 등 수동형 보행자 보호시스템의 적용은 보행자 사고 발생시 심각도의 감소에 매우 효과적인 역할을 수행할 수 있다. 또한, 승합차, 화물차 등 중차량의 보행자 사고 발생시 승용차보다 그 사고심각도가 높게 도출되었다. 이 경우 향후 중차량에 대한 보행자 보호시스템의 우선 적용도 고려할 필요성이 있을 것이다.

둘째, 보행자사고의 분석 목적에 따른 적절한 방법론은 다음과 같이 정리할 수 있다.

1. 이항 로지스틱 회귀분석의 활용

이항 로지스틱 회귀분석은 이분형으로 구분된 보행자 교통사고 심각도를 활용하여 사고심각도와 영향요인 사이의 인과관계를 규명하고자 할 경우 활용 가능하다. 심각도 영향요인에 대한 통계적 유의성, 각 요인에 대한 상세한 분석결과를 제시할 수 있다. 영향요인 계수의 부호를 검토하여 심각도 증가에 대해 긍정적 영향을 미치는지, 부정적 영향을 미치는지 알 수 있으며, Odd-ratio를 검토하여 각 요인이 사고심각도의 확률적 증가에 미치는 영향의 정도를 파악할 수 있다.

2. 의사결정나무의 활용

의사결정나무는 거시적 관점에서 노드간 상하관계에 따라 보행자 교통사고 영향요인간 우선순위를 파악하여 정책결정의 방향을 결정하는데 분석결과를 활용하고자 할 경우 사용할 수 있다. 분석을 통하여 도출된 결정트리를 활용하여 분석결과를 시각적으로 표현할 수 있는 장점이 있으며, 본 연구에서 활용한 엔트로피 개념을 적용한다면 결정트리에서 정보의 이득, 즉 변수의 추가로 발생하는 불확실성의 감소 정도를 파악할 수 있다.

3. 서포트 벡터 머신의 활용

서포트 벡터 머신은 서포트 벡터 개념을 적용, 보행자 교통사고자료를 활용하여 사고심각도를 분류에 최적화된 알고리즘을 개발하고자 할 경우 활용 가능하다. 본 연구의 결과와 같이 의사결정나무 보다는 분류능력이 높게 나타나므로 분석목적이 사고심각도의 분류정확도 향상인 경우 유용하게 활용 할 수 있다.

4. 순서형 프로빗 모형의 활용

순서형 프로빗 모형은 다양한 사고심각도를 고려한 보행자 교통사고 분석을 목적으로 심각도와 영향요인 사이의 인과관계를 규명하고자 할 경우 활용가능하다. 이항 로지스틱 회귀분석과 마찬가지로 영향요인 계수의 부호를 검토하여 독립변수가 종속변수에 미치는 긍정적 또는 부정적 영향을 파악할 수 있다. 또한, Pseudo 탄력도나 한계효과를 검토하여 각각의 사고심각도, 즉 경상사고, 중상사고, 사망사고의 확률적 증가에 미치는 유의변수의 영향의 정도를 세밀하게 파악할 수 있다.

향후 추가적인 연구를 수행하여 보다 가치있는 분석 결과 도출이 가능할 것으로 판단된다. 먼저, 보행자 사고특성 분석의 경우 운전자의 범규위반 내용, 차량의 용도 등 보다 상세한 분석을 통해 추가적인 사고특성 도출이 가능할 것이다. 또한, SVM의 경우 국외에서는 종속변수가 3가지인 분석에 대한 연구도 활발히 진행 중에 있다. 향후 이러한 분석기법은 OPM과 유사하게 보다 다양한 심각도에 대한 교통사고 분류모형 구축이 가능할 것으로 판단된다. 또한, 본 연구에서는 DT 알고리즘에 대한 임계값을 분석가에 의한 수동적인 방법으로 설정하여 분석하였지만 임계값 설정에 대한 정확한 기준설정 및 분류정확도의 향상을 위하여 유전 알고리즘 등의 최적화 기법을 이용한 추가분석이 필요할 것이다. 또한, 통계모형의 적합도나 분류기의 분류정확도를 향상시키기 위한 추가적인 분석을 수행한다면 향후 보행자 사고모형으로의 활용이 가능할 것으로 판단된다.

본 연구의 분석결과는 보행자 교통사고자료 분석시 분석자에게 유용한 정보를 제공할 수 있을 것으로 기대되며, 보행자 교통안전 증진을 위한 다양한 의사결정시 기초자료로 활용될 수 있을 것으로 판단된다.

ACKNOWLEDGEMENT

This research was supported by the Korea Ministry of Land, Transport and Maritime Affairs. It was also supported by the Korea Institute of Construction and Transportation Evaluation and Planning (Project No.: 11PTSI-C054118-03).

REFERENCES

- Al-Ghamdi A. S. (2002), Pedestrian-vehicle Crashes and Analytical for Stratified Contingency Tables, *Accident Analysis & Prevention*, 34, 205-214.
- Ballesteros M. F., Dischinger P. C., Langenberg P. (2003), Pedestrian Injuries and Vehicle Type in Maryland, 1995-1999, *Accident Analysis and Prevention*, 36, 73-81.
- Boonmee S., Tangamchit P. (2009), Portable Reckless Driving Detection System, *Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2009*, 6th International Conference on 01.
- Choi S., Lee G. Y., Oh C. (2012), Prioritizing Freeway Segments for Variable Speed Limit (VSL) Operations based on Traffic Safety, *J. Korean Soc. Transp.*, 30(3), Korean Society of Transportation, 57-70.
- Christoforou Z., Cohen S., Karlaftis M. G. (2010), Vehicle Occupant Injury Severity on Highways: An Empirical Investigation Original Research Article, *Accident Analysis & Prevention*, 42(6), 1606-1620.
- De Oña J., López G., Abellán J. (2012), Extracting Decision Rules From Police Accident Reports Through Decision Trees, *Accident Analysis and Prevention*, 50, 1151-1160.
- Eriksson M., Papanikolopoulos N. P. (2001), Driver fatigue: a Vision-based Approach to Automatic Diagnosis, *Transportation Research Part C*, 9(6), 399-413.
- EU, Considers Introducing Airbags for Pedestrians (2006), Bonn, Germany: Deutsche Welle, <http://www.dwworld.de/dw/article>.
- Froming R., Schindler V., Kuhn M. (2006), Requirement Engineering for Active Safety Pedestrian Protection Systems Based on Accident Research, *Advanced Microsystems for Automotive Applications 2006*, New York: Springer-Verlag, Apr. 25-26, 79-106.
- Han I. H., Yang K. S. (2007), Recognition of Dangerous Driving Using Automobile Black Boxes, *J. Korean Soc. Transp.*, 25(5), Korean Society of Transportation, 149-160.
- Hels T., Orozova-Bekkevold I. (2007), The Effect of Roundabout Design Features on Cyclist Accident Rate, *Accident Analysis and Prevention*, 39, 300-307.
- Hong J. Y., Kim K. T., Lee S. B. (2011), Developing the Accident Injury Severity on a Field of Construction Work Using Ordered Probit Model, *Journal of the Korean Society of Safety*, 26(2), 89-98.
- Hua Z., Zhang B. (2006), A Hybrid Support Vector Machines and Logistic Regression Approach for Forecasting Intermittent Demand of Spare Parts, *Applied Mathematics and Computation*, 181, 1035-1048.
- Jeong C. W. (2007), A Study on Impactive Factors about Major Causes In Freeway Accidents, *Police Journal*, 7(1), 220-247.
- Joksch H. C. (2000), Vehicle Design Versus Aggressivity, USDOT, HS 809, 194.
- Kashani A. T., Shariat-Mohaymany A., Ranjbari A. (2011), Analysis of the Traffic Injury Severity on Two-lane, Two Way Rural Roads Based on Classification Tree Models, *Safety Science*, 49, 1314-1320.
- Kim J. H., Lee S. B., Kim D. H., Hong J. Y. (2012), The Relationship Between Violation of Designated Lane Usage and Accident Severity on Freeways, *J. Korean Soc. Transp.*, 30(3), Korean Society of Transportation, 119-127.
- Kockelman K. M., Kweon Y. J. (2001), Driver Injury Severity : An Application of Ordered Probit Models, *Accident Analysis and Prevention*, 28, 313-321.
- Lemp J. D., Kockelman K. M., Unnikrishnan A. (2011), Analysis of Large Truck Crash Severity Using Heteroskedastic Ordered Probit Models Original Research Article, *Accident Analysis and Prevention*, 43, 370-380.
- Li X., Lord D., Zhang Y., Xie Y. (2007), Predicting Motor Vehicle Crashes Using Support Vector Machine Models, *TRB 87th Annual Meeting*.
- Li Z., Liu P., Wang W., Xu C. (2012), Using Support Vector Machine Models for Crash Injury Severity Analysis, *Accident Analysis and Prevention*, 45, 478-486.
- Lim H. S., Jeong E. B., Oh C., Gang G. P. (2011), Detection of Unsafe Zigzag Driving Maneuvers Using

- a Gyro Sensor, The Journal of The Korea Institute of Intelligent Transport Systems, 10(2), 42-54.
- Mallick D. (2009), Marginal and Interaction Effects in Ordered Response Models, School of Accounting, Deakin University, Economics and Finance, Burwood (Technical report).
- Montella A., Aria M., D'Ambrosio A., Mauriello F. (2011), Data-Mining Techniques for Exploratory Analysis of Pedestrian Crashes, Journal of the Transportation Research Board, 2237, 107-116.
- Moon S. R., Lee Y. I. (2011), Analysis of Traffic Crash Severity on Freeway Using Hierarchical Binomial Logistic Model, Journal of Korean Society of Road Engineers, 13(4), 199-209.
- Oh J. T., Jo J. H., Lee S. Y., Kim Y. S. (2008), Development of a Data-logger Classifying Dangerous Drive Behaviors, The Journal of The Korea Institute of Intelligent Transport Systems, 7(3), 15-28.
- Park B. H., Yang J. M., Kim J. Y. (2009), Logistic Regression Accident Models by Location in the Case of Cheong-ju 4-Legged Signalized Intersections, Journal of Korean Society of Road Engineers, 11(2), 17-25.
- Park J. H., Oh C., Kang G. P. (2011), Assessment of Freeway Crash Risk Using Probe Vehicle Accelerometer, Journal of Korean Society of Road Engineers, 13(2), 49-56.
- Park J. H., Oh C., Namkoong S. (2008), Estimation of Freeway Accident Likelihood Using Real-time Traffic Data, J. Korean Soc. Transp., 26(2), Korean Society of Transportation, 157-166.
- Park J. T., Choi B. B., Lee S. B. (2010), A Study on the Characteristics of Traffic Accidents for the Elderly Pedestrians on Rural Highways, J. Korean Soc. Transp., 28(5), Korean Society of Transportation, 155-162.
- Roudsari B. S., Mock C. N., Kaufman R., Grossman D., Henary B. Y., Crandall J. (2004), Pedestrian Crashes: Higher Injury Severity and Mortality Rate for Light Truck Vehicles Compared With Passenger Vehicles, Injury Prevention, 10, 154-158.
- Scheetz L. J., Zhangb J., Kolassa J. (2009), Classification Tree Modeling to Identify Severe and Moderate Vehicular Injuries in Young and Middle-aged Adults, Artificial Intelligence in Medicine, 45, 1-10.
- Sut N., Simsek O. (2011), Comparison of Regression Tree Data Mining Methods for Prediction of Mortality in Head Injury, Expert Systems with Applications, 38, 15534-15539.
- Sze N. N., Wong S. C. (2007), Diagnostic Analysis of the Logistic Model for Pedestrian Injury Severity in Traffic Crashes, Accident Analysis and Prevention, 39, 1267-1278.
- Tarko A., Azam M. S. (2011), Pedestrian Injury Analysis With Consideration of the Selectivity Bias in Linked Police-hospital Data, Accident Analysis and Prevention, 43(5), 1689-1695.
- Wang K. C. P., Hou Z., Gong W. (2009), Road Sign Classification based on Support Vector Machine (SVM), TRB 2009 Annual Meeting.
- Yoon P., Kang H., Jung H. G. (2006), Next Generation Pedestrian Protection Systems, Auto Journal, 28(4), 28-39.
- Yu R., Abdel-Aty M. (2013), Utilizing support vector machine in real-time crash risk evaluation, Accident Analysis and Prevention, 51, 252-259.
- Zajac S. S., Ivan J. N. (2003), Factors Influencing Injury Severity of Motor Vehicle-crossing Pedestrian Crashes in Rural Connecticut, Accident Analysis and Prevention, 35, 369-379.
- 알림 : 본 논문은 대한교통학회 제68회 학술발표회(2013. 2.23)에서 발표된 내용을 수정·보완하여 작성된 것입니다.
- ☞ 주 작성자 : 최세로나
 ☞ 교신저자 : 최세로나
 ☞ 논문투고일 : 2013. 7. 8
 ☞ 논문심사일 : 2013. 8. 21 (1차)
 2013. 10. 29 (2차)
 2013. 11. 28 (3차)
 ☞ 심사판정일 : 2013. 11. 28
 ☞ 반론접수기한 : 2014. 4. 30
 ☞ 3인 익명 심사필
 ☞ 1인 abstract 교정필