

블랙박스 영상 기반 고속도로 사고유형 분류 및 사고 심각도 예측 평가

Classification and Prediction of Highway Accident Characteristics Using Vehicle Black Box Data

조 준 한* · 이 성 준** · 박 성 민*** · 박 준 영****

* 주저자 : 삼성교통안전문화연구소 수석연구원
 ** 교신저자 : 한양대학교 교통물류공학과 석사과정
 *** 공저자 : 한양대학교 교통물류공학과 박사과정
 **** 공저자 : 한양대학교 교통물류공학·스마트시티공학과 교수

Junhan Cho* · Sungjun Lee** · Seongmin Park*** · Juneyoung Park****

* Samsung Traffic Safety Research Institute
 ** Dept. of Transportation and Logistics Eng., Hanyang University
 *** Dept. of Transportation and Logistics Eng., Hanyang University
 **** Dept. of Transportation and Logistics Eng. Smart City Engineering., Hanyang University

† Corresponding author : sug18@hanyang.ac.kr

Vol. 21 No.6(2022)
 December, 2022
 pp.132~145

pISSN 1738-0774
 eISSN 2384-1729
<https://doi.org/10.12815/kits.2022.21.6.132>

Received 4 October 2022
 Revised 21 October 2022
 Accepted 16 November 2022

© 2022. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요 약

본 연구는 고속도로에서 발생한 교통사고 블랙박스 영상을 기반으로 군집분석과 예측모형 비교를 수행하였다. 분석자료로 사고 직전의 도로 및 교통 상황을 파악할 수 있는 차량 주행행태, 노면 상태 등 사고 영상에서 추출이 가능한 항목을 설명변수로 활용하였다. 여러 요소에 의해 영향을 받는 교통사고 데이터의 특징을 고려하여 데이터의 이질성을 반영하는 군집분석을 활용하였다. 군집분석으로 분류된 각 군집을 사고 심각도 수준의 비율을 기준으로 나누고, 종속변수인 인명피해 수준을 반영하여 사고 예측 평가를 수행하였다. 사고 예측모형은 로짓모형(Logit model)을 적용한 결과, 전체 데이터를 분석한 경우보다 군집분석에 의해 두 개의 사고 심각도 그룹을 분류하여 예측했을 때 우수한 예측 능력을 보여주었다. 이는 군집분석을 통한 그룹별 사고 특성과 사고 심각도를 반영하여 사고위험을 예측하는 것이 더 효과적인 것으로 판단된다. 또한 2차 사고와 같은 정차 중 추돌사고, 차로변경 중 측면 추돌사고 등이 중요한 주행행태변수로 작용하는 것으로 나타났다.

핵심어 : 블랙박스 영상, 고속도로, 군집분석, 사고 심각도, 변수 이질성

ABSTRACT

This study was based on the black box images of traffic accidents on highways, cluster analysis and prediction model comparisons were carried out. As analysis data, vehicle driving behavior and road surface conditions that can grasp road and traffic conditions just before the accident were used as explanatory variables. Considering that traffic accident data is affected by many factors, cluster analysis reflecting data heterogeneity is used. Each cluster classified by cluster analysis was divided based on the ratio of the severity level of the accident, and then an accident prediction evaluation was performed. As a result of applying the Logit model, the accident prediction model showed

excellent predictive ability when classifying groups by cluster analysis and predicting them rather than analyzing the entire data. It is judged that it is more effective to predict accidents by reflecting the characteristics of accidents by group and the severity of accidents. In addition, it was found that a collision accident during stopping such as a secondary accident and a side collision accident during lane change act as important driving behavior variables.

Key words : Black box, Highway, Accident severity, Cluster analysis, Heterogeneity

I. 서 론

1. 개요

고속도로는 도시부 도로와 달리 차량 주행 속도가 상대적으로 높으므로 교통사고 발생 시 사고 피해 심각도가 크고 2차 사고나 다중사고가 발생할 가능성이 크다. 최근 5년간 (2017~2021년) 교통사고를 살펴보면, 고속도로의 5년간 평균 100건당 사망자 수는 5.29명으로 모든 도로의 평균 100건당 사망자 수인 1.61명에 비해 3.3배 높게 나타났다(taas.koroad.or.kr). 이처럼 고속도로는 주행 속도가 높고 정지거리가 길어지기 때문에 사고 발생 시 치명적인 인명피해가 발생하게 된다. 고속도로 특성상 주행 중 정면 추돌, 차로변경 중 측면추돌 등 다양한 교통 상황에서의 사고원인과 유형이 발생하기 때문에 세밀한 추돌사고 특성 분석이 필요하다. 또한, 고속도로는 상습 반복 정체, 차로변경 구간, 버스전용차로 등 다양한 차량 행태(Driving behavior)가 발생하므로 차량 간의 상호작용과 사고 심각도의 영향요인에 대한 깊이 있는 분석이 필요하다.

교통사고 심각도에 영향을 미치는 요인은 운전자, 도로 환경, 주행 상황, 기상 등 다양한 영향 요소가 존재한다. 이러한 영향 요소로 구성된 교통사고 데이터는 이산형/연속형 수치 변수, 정성적 지표 등 기본적으로 이질성(Heterogeneous)을 포함하게 된다(Depire et al., 2008). 즉 변수의 단위도 다르고, 표현하는 정도의 차이도 상이하므로 이질성을 고려하지 않으면 분석과정에서 현실을 제대로 반영하지 못하고 왜곡된 결과를 초래할 수 있다. 따라서 본 연구는 블랙박스 영상분석에서 추출되는 이질적인 사고 데이터를 토대로 고속도로 사고유형을 분류하고, 그 군집 결과를 기반으로 사고 심각도 예측모형 분석을 수행하였다.

본 연구는 고속도로 사고, 교통사고 심각도, 사고 예측모형 관련하여 기존 문헌을 고찰하여 이질적인 교통 변수 선정과 연구 방법론을 정립하였다. 분석자료는 블랙박스 사고 영상자료를 활용하여 사고지점의 도로 기하구조, 사고 직전의 주행 상황 등을 알 수 있는 변수를 추출하였다. 분석 방법은 사고유형 분류를 위한 군집 별 사고 특성을 분석하기 위해 잠재 계층분석(Latent class analysis)을 적용하였고, 예측모형은 이항로짓 모형(Logit model)을 활용하였다. 실제 사고 영상에서 도출된 자료를 바탕으로 사고 특성을 분류하고, 사고 심각도에 영향을 주는 요인을 규명하는 데 본 연구의 목적이 있다.

II. 기존 문헌 고찰

1. 고속도로 교통사고 연구

Yu and Abdel-Aty(2013)는 산악 고속도로 구간 5년간의 충돌 데이터를 사용하여 다단계 로지스틱 회귀분석, 베이지안 추론을 적용해 분석했다. 연구 결과, 단일 사고와 다중사고가 교통사고 발생 기여 요인이 다르

다는 것을 도출하였다. Hong et al.(2019)는 고속도로에서 발생한 7년간의 차량 유형별 충돌사고 데이터를 활용하여 로짓 모형을 통해 사고 특성을 수행하였다. 그 결과 터널과 본선 구간 내 다중 충돌사고는 주행 차량 대수와 관계가 있지만 램프 구간과 요금소(Toll plaza)는 관계가 적은 것으로 나타났다. Jeon and Kim(2020)는 고속도로의 교통사고 차량 대수에 따른 요인의 영향분석을 위해 3년간 사고 데이터를 다항 로짓모형을 이용하여 분석을 수행하였다. 분석 결과, 야간사고와 고속도로 본선, 터널, 교량에서 발생한 사고, 운전자 과실, 도로 장애물, 노면 문제로 인한 사고, 내리막 경사, 중차량 연관 사고가 다중 충돌사고로 이어질 가능성이 큰 것으로 분석되었다. Ma et al.(2009)는 2년간 고속도로 터널 교통사고를 사용하여 시간적·공간적 분포 특성을 분석하였다. 그 결과 터널구간은 다중 차량사고가 단일 사고보다 많았으며, 터널 내 구간은 노면이 건조한 상태인데도 강우 시 사고 발생이 46.7%로 비교적 높게 나타났다.

선행연구 검토 결과, 고속도로 사고조사 자료를 토대로 교통사고 심각도에 영향을 미치는 요인을 도출하였고, 주요한 영향변수는 운전자 과실, 다중사고, 야간, 교량·터널, 이상기후 등으로 나타났다. 실질적인 사고 직전의 도로·교통환경, 주행 상황 등을 반영한 연구 수행은 자료수집 특성상 한계가 있는 것으로 나타났다.

2. 교통사고 심각도 연구

Abdel-Aty(2003)는 여러 도로의 운전자 부상 심각도 수준을 분석하기 위해 2년간 미국 플로리다 교통사고 데이터를 토대로 순서형 프로빗 분석을 수행하였다. 신호 교차로, 기본구간 및 통행료 징수구간(Toll plaza)에 대해 각각 독립된 모형을 개발하였다. 노인과 남성 운전자, 안전벨트 미착용 운전자, 승용차 운전자(차량 종류), 운전석 측면(사고 특성)에 부딪힌 차량 및 속도위반(과속) 운전자는 모든 모형에서 심각한 부상 가능성이 큰 것으로 나타났다. Gray et al.(2008)은 영국에서 젊은 남성 운전자와 관련된 사고 특성을 분석하기 위해 13년간 교통사고 데이터를 활용해 순서형 프로빗 분석을 수행하였다. 주말(금요일~일요일) 야간 운전, 주변 차량 추월 운전행태, 제한속도 60 mph 상의 일방통행 운전 등에 속하는 경우 교통사고 시 중상 또는 치명적 부상 발생 가능성이 큰 것으로 나타났다. Garrido et al.(2014)은 5년간 교통사고 데이터를 토대로 순서형 프로빗 모형을 적용하여 차량 탑승자의 부상 심각도 영향요인을 도출하였다. 분석 결과 양방향 도로 내 건조한 노면을 주행하는 경차 탑승자가 일방통행 내 젖은 노면을 주행하는 중차량 탑승자보다 더 심각한 부상이 발생하였다. Wu et al.(2014)는 단일 사고와 다중사고의 운전자 부상 심각도를 분석하기 위해 2년간 뉴멕시코의 고속도로 사고자료를 활용하였다. 혼합로짓모형을 적용하여 분석한 결과, 시간 변수 중 야간일 경우와 날씨 변수 중 먼지와 모래가 지면 위로 솟구치는 경우에서 다중 충돌사고 심각도가 증가하였다. Eboli et al.(2020)는 2016년에 발생한 이탈리아 교통사고의 도로요인, 환경요인, 운전자 요인, 사고 상황 요인이 교통사고 유형에 미치는 영향을 이항 로지스틱 회귀모형을 적용해 분석하였다. 연구 결과, 교차로 유무 및 일방통행 유무와 같은 도로 조건과 노면 상태, 운전자 나이가 중대 사고와 관련이 있는 것으로 나타났다. Aiash and Robuste(2021)은 교통사고 심각도와 사망자에 따른 분류요인의 상관관계를 파악하였다. 4년간 고속도로 데이터를 기반으로 이항 프로빗 모형을 적용하였는데, 주말 또는 야간 시간대는 사고 심각도가 높게 나타났다. Mohamed et al.(2017)은 차량 충돌사고가 사고 심각도 분석에 영향을 미치는 요인을 파악하기 위해 아부다비 에미리트(AD) 2014년 데이터를 이항 로짓모형을 사용하여 분석하였다. 분석 결과 도로 유형 및 차선 수가 사고와 관련이 있는 것으로 나타났다.

선행연구 검토 결과, 운전자 나이, 야간, 종단선형, 노면 상태 등이 교통사고 심각도 영향요인으로 나타났다. 기존 문헌은 전통적인 교통 사고조사 지표를 토대로 사고 심각도 분석을 수행하였으며, 차량 주행행태 등 블랙박스 영상을 통해 확인할 수 있는 변수를 활용하는데 제약이 있는 것으로 분석되었다.

3. 군집 및 머신러닝 관련 연구

Iranitalab and Khattak(2017)은 4년간 미국 내 차량 충돌 데이터를 토대로 전통적인 통계기법과 머신러닝을 활용해 교통사고 심각도 예측모형의 성능 비교 연구를 수행하였다. 또한, K-means 군집분석과 잠재 계층분석(Latent class analysis)을 개별 적용했을 때 사고 심각도 예측모형의 성능을 비교하였다. 분석 결과, 두 군집기법 모두 심각도에 대해 예측성능이 향상하는 것으로 나타났다. Li et al.(2019)은 운전자 부상 심각도 영향요인을 분석하기 위해 차량 간의 충돌사고를 포함하는 2년간 데이터를 활용했다. 잠재 계층분석으로 전체 데이터 세트를 분류한 후 혼합로짓모형을 추가로 분석했다. 군집분석 결과, 농촌지역과 젖은 노면 상태 구간에서 운전자의 부상 심각도가 높은 것으로 나타났다. Lei et al.(2021)은 6년간 고속도로 교통사고 데이터를 기반으로 머신러닝 기법으로 사고 예측모형을 개발하고 교통류 변수 선정을 위해 랜덤 포레스트(Random forest) 기법을 사용하였다. SVM(Support vector machine) 기법을 적용한 사고 예측모형은 90%의 정확도를 보였다. Sun et al.(2021)은 3년간 중국 교통사고 데이터에 대해 잠재 계층분석을 적용하여 군집 특성 분석과 군집 별 사고 심각도 영향요인을 도출하였다. 선행연구 검토 결과, 교통사고 유형분류를 위한 군집기법은 K-means Clustering과 잠재 계층분석을 가장 많이 사용하였고, 예측모형은 이질적인 데이터 성격을 고려하여 SVM과 로짓모형 등을 적용하였다.

4. 기존 연구와의 차별성

고속도로 교통사고 영향요인 관련 연구는 기본적인 교통사고 조사자료를 토대로 인적요인, 도로요인, 주행요인 등 다양한 요소를 반영하여 군집분석과 사고 예측모형을 적용하였다. 본 연구는 사고 직전의 기하구조, 차량 거동 등 다양한 교통환경변수 추출이 가능한 블랙박스 영상자료를 활용하였다. 기존 논문과의 차별성은 기본적인 교통사고 관련 변수 외에 사고 영상에서 추출되는 차량 행태변수(차로변경, 추돌유형 등), 도로 환경변수(곡선부 유무, 노면 상태 등)를 설명변수로 반영하여 군집분석과 사고 심각도 예측모형을 적용했다는 점이다. 기존 문헌에서 언급하지 않은 새로운 변수와 사고 직전의 정확한 교통환경 변수를 반영한다는 것은 사고 특성 그룹화와 예측모형 성능 향상에 도움이 되고, 궁극적으로 사고 심각도에 영향을 미치는 요소를 도출하는 데 기여할 것으로 판단된다.

Ⅲ. 연구 방법론

1. 분석자료의 설명

본 연구는 최근 5년간(2016~2020년) 전국 고속도로에서 발생한 블랙박스 사고 영상을 통해 분석자료를 수집하였다. 본 연구는 블랙박스 사고자료를 수작업으로 영상분석을 수행하였다. 세부적으로 사고유형, 사고지점, 기하구조, 교통 상황, 사고 직전 주행 환경 등을 도출하기 위해 체크리스트를 작성하여 교통사고 영향요인을 추출하였다. 기존 경찰 교통사고 자료는 사고경위서를 통해 일정부분 사고 원인 규명이 가능할 수 있으나, 사고 직전의 운전행태(주행 중/정차 중 추돌, 추월차로/주행차로 진입을 위한 차로변경 등), 주변 차량 교통 상황, 사고지점의 정확한 기하구조, 차량간 충돌 각도 등을 정확히 추출하는 데 한계가 존재한다. 블랙박스 사고 영상자료는 이러한 상세한 사고 영향요인을 객관적으로 추출할 수 있다. 분석자료는 고속도로 기본

(본선)구간 총 783건을 추출하였고, 교통사고 인명피해 심각도에 영향을 미치는 요인을 크게 인적요인, 도로 요인, 사고요인, 기타요인(시간대, 기상)으로 구분하였다. 요인별 교통사고에 유의한 영향을 미치는 변수를 선정하였으며, 세부 변수와 관측 수(표본)는 <Table 1>과 같다. 인적요인은 가해 운전자 나이와 성별을 선정하였고, 도로요인은 노면 상태, 평면선형, 종단선형을 추출하였다. 사고요인은 차량 주행행태, 상충유형, 사고 구간 유형, 사고 형태로 구분하였고, 기타요인으로 시간대(주간/야간), 기상 상태를 도출하였다. 나이에 따른 결과를 비교하기 위해 가해 운전자 나이 변수를 구분하여 분석을 진행하였다. 나이 변수는 한국보건사회연구원(<https://www.kihasa.re.kr>)의 자료를 활용하여 청년, 장년, 중년, 노년으로 구분하였다. 사고 영상분석에서 도출할 수 있는 차량 주행행태 변수 중 ‘차량 단독사고’는 중앙분리대, 가드레일 등 도로시설물과 충돌하거나 이에 따라 전복되는 사고를 의미한다. ‘추월차로 진입을 위한 차로변경 사고’는 주행차로를 주행하다가 추월차로로 차로 변경하다가 발생한 사고이며, ‘주행차로 진입을 위한 차로변경 사고’는 추월차로 또는 다른 주행차로에서 사고가 발생한 주행차로로 차로 변경하다가 발생한 사고이다. ‘정차 중 추돌사고’는 전방에 정차 중인 차량과 동일 방향에서 주행하는 차량간 발생한 사고이고, ‘주행 중 추돌사고’는 주행하는 차량과 동일 방향에서 주행하는 차량간 사고이다.

<Table 1> Descriptive Statistics of the Explanatory Variables

Variable category	Variable type	Variable title	Count	Percentage	Description
Human factor	Age of driver	Age (~35)	249	31.80%	Ratio scale
		Age (36-50)	302	38.57%	
		Age (51-65)	192	24.52%	
		Age (66~)	40	5.11%	
	Driver gender	Male	646	82.50%	Driver gender (male =1, female =0)
		Female	137	17.50%	
Accident factor	Accident characteristics	Single vehicle accident	7	0.89%	Single vehicle accident (yes =1, no =0)
		Lane-changing accident in overtaking lane	71	9.07%	Lane-changing accident in overtaking lane (yes =1, no =0)
		Lane-changing accident in driving lane	178	22.73%	Lane-changing accident in driving lane (yes =1, no =0)
		Accident while parked	387	49.43%	Accident while parked (yes =1, no =0)
		Accident while driving	140	17.88%	Accident while driving (yes =1, no =0)
	Collision type	Rear-end collision	645	82.38%	Rear-end collision (yes =1, no =0)
		Head-on collision	7	0.89%	Head-on collision (yes =1, no =0)
		Side collision	131	16.73%	Side collision (yes =1, no =0)
	Accident section	Normal section	669	85.44%	Normal (yes =1, no =0)
		Construction section	4	0.51%	Construction (yes =1, no =0)
		Bridge section	8	1.02%	Bridge (yes =1, no =0)
		Tunnel section	102	13.03%	Tunnel (yes =1, no =0)
	Accident type	Multi-vehicle accident	206	26.31%	Accident type (Multi-vehicle =1, Vehicle-vehicle =0)
		Vehicle-vehicle accident	577	73.69%	
Road factor	Road surface	Dry	733	93.61%	Road surface (Dry =0, Abnormal condition =1)
		Abnormal condition	50	6.39%	
	Horizontal	Right curve	29	3.70%	Right curve (yes =1, no =0)

Variable category	Variable type	Variable title	Count	Percentage	Description
	alignment	Left curve	34	4.34%	Left curve (yes =1, no =0)
		Straight	720	91.95%	Straight (yes =1, no =0)
	Road grade	Downhill	6	0.77%	Downhill(yes=1, no=0)
		Uphill	21	2.68%	Uphill(yes=1, no=0)
		Flat	756	96.55%	Flat(yes=1, no=0)
Other factors	Time of the day	Day time	595	75.99%	Time of the day(night=1, day=0)
		Night time	188	24.01%	
	Weather	Clear	703	89.78%	Weather (clear =0, abnormal condition =1)
		Abnormal condition	80	10.22%	

2. 잠재 계층 분석(Latent Class Analysis)

잠재 계층분석(Latent class analysis)은 데이터에서 숨겨진 그룹을 찾아내는 군집분석 방법론으로 일련의 범주형 변수에 대한 상호 배타적인 잠재 군집으로 분류할 수 있는 통계기법이다. 잠재 계층분석을 통해 변수별 군집에 속할 확률값을 확인하여 군집의 특성을 파악할 수 있다. 범주형 변인의 잠재 계층분석에 대한 기본식은 아래와 같으며, Y_i 는 종속변수의 벡터이고, K 는 각각의 계층을 나타낸다. 변수별 클러스터에 속할 확률값을 통해 클러스터의 특성을 파악할 수 있으며 식(1)과 같이 표현할 수 있다(Vermunt and Magidson, 2002).

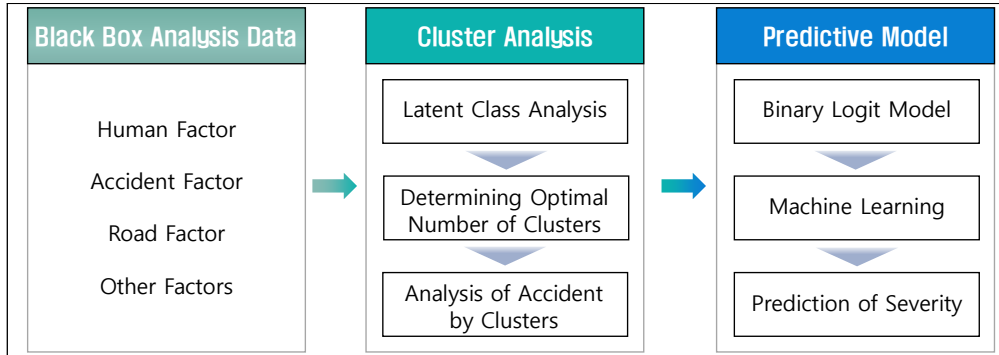
$$P_{Y_i} = \sum_{k=1}^K P_{X_k} P_{Y_i|X_k} \dots\dots\dots (1)$$

3. 이항 로짓 모형(Binary Logit Model)

로지스틱 회귀는 연결 함수가 선형 함수를 0~1 범위의 연속확률함수로 전달할 수 있으므로 사고 심각도가 두 가지로 나타나는 이항 분포를 측정하는 데 효과적이고 널리 사용되는 것으로 입증되었다. Y_i 가 사고 심각도 i 의 결과를 나타내도록 한다. $Y_i = 1$ 은 충돌 i 가 심각한 사고임을 의미하며, $Y_i = 0$ 은 충돌 i 가 비충돌을 의미한다. 이항 로지스틱 함수는 $Y_i = 1$ (β_i 로 표시)의 확률을 다음과 같이 독립변수와 연결하기 위해 사용된다. 본 연구는 사고 심각도를 부상 사고, 중상 및 사망사고에 대해 각각 y 를 0, 1로 부여하여 분석을 진행하였고 식 (2) 와 같이 나타낼 수 있다.

$$\log\bar{y}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{k=1}^K \beta_k X_{ik} + \epsilon_i \dots\dots\dots (2)$$

여기서 X_{ik} 은 충돌 i 에 대한 k 번째 독립변수의 값이고, β_0 은 모델의 절편이며, β_k 는 X_{ik} 에 대한 추정 계수이며, β_i 는 로지스틱 분포에 이은 랜덤 오차 항이다(Nagler, 1994).



<Fig. 1> Flow chart of the study

분석 절차는 블랙박스 사고 영상을 통해 데이터를 수집하여 분석자료를 인적요인, 도로요인, 사고요인, 기타요인으로 구분하여 데이터를 설정하였다. 여러 요소에 의해 영향을 받는 교통사고의 데이터의 특징을 고려하여 데이터 내에서 숨겨진 그룹을 찾아내는 군집분석 방법론을 활용하였다. 사용하는 데이터 세트가 범주형 변수이므로 일련의 범주형 변수에 대해 잠재 군집으로 분류할 수 있는 잠재 계층분석(Latent Class Analysis)을 활용하였다. 이때, 적합도 지표를 사용하여 최적 군집 수를 도출하여 각 군집 별 사고유형에 따라 분석하였다. 잠재 계층분석을 통해 도출된 군집을 사고 심각도에 따라 분류 후 로짓 모형(Logit model)을 수행하였다. 이를 통해 사고 심각도를 예측하고 비교하였다.

IV. 연구 결과 분석

1. 최적 군집 수 결정

본 연구는 고속도로를 대상으로 교통사고 심각도에 영향을 주는 요인을 군집별로 분류하기 위해 잠재 계층분석을 적용하였다. 이때 종속변수인 사고 심각도는 크게 부상 사고, 중상·사망사고로 구분하였다. <Table 2>는 최적 군집(Cluster) 수를 도출을 위한 군집 별 설명 지표 값이다. 군집 수를 도출하는 데 사용하는 Latent class의 적합성은 베이지안 정보 기준(BIC) (Raftery, 1986), Akaike의 정보 기준(AIC) (Akaike, 1987) 값을 사용하여 측정되었다. AIC, BIC 값이 낮을수록 잠재 군집모형에 더 적합하다는 것을 나타낸다. AIC와 BIC는 식 (3), 식 (4)로 나타낼 수 있다.

$$AIC = -2LL + 2K, \dots \dots \dots (3)$$

$$BIC = -2LL + K \ln(N) \dots \dots \dots (4)$$

여기서 LL은 수렴 시 로그 우도, K는 모수의 수, N은 관측치의 수이며 <Table 2>는 모형 추정에 대한 Parameter, LL, BIC, AIC 결과를 군집(Cluster) 수별로 나타냈다. 등급 수가 증가할수록 Parameter 수는 증가하지만, AIC와 BIC는 비례하지 않는다. Akaike(1974)가 제안한 AIC지수는 각 모형을 비교하는 데 사용된다. AIC값의 절대량은 의미가 없고, AIC값이 상대적으로 낮으면 더 나은 모형으로 평가하는데 일반적으로 AIC값의 차이가 4보다 작은 경우 사실상 모형의 차이가 없는 것으로 간주한다(Charlton et al., 2009). Nylund et

al.(2007)은 최적 군집 수를 결정하는데 BIC가 샘플수(N)에 대한 AIC 값보다 상대적으로 더 나은 지표로 제안했다. 또한 최근 많은 연구에서 잠재 군집 수를 결정하는 데 BIC를 사용함에 따라(Li et al., 2019) 해당 지표를 기초로 적정 군집 수를 결정하였다. 잠재 계층분석 결과, BIC 값 13291.11, AIC 값 12428.43으로 도출된 Cluster 6을 최적 군집 수로 선정하였다.

<Table 2> Determining the number of clusters

Number of clusters	Number of parameters	Log-likelihood	AIC	BIC
Cluster 2	61	-7238.706	14599.41	14883.86
Cluster 3	92	-6921.916	14027.83	14456.84
Cluster 4	123	-6775.474	13796.95	14370.51
Cluster 5	154	-6482.615	13273.23	13991.35
Cluster 6	185	-6029.214	12428.43	13291.11
Cluster 7	216	-6240.423	12912.85	13920.08

2. 군집 별 사고 특성 분석

<Table 3> Summary of Variables Describing Cluster Characteristics.

Variable	Probability	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Lane-changing accident in overtaking lane	Pr(0)	0.9335	1	0.8587	0.7369	0.9519	0.9144
	Pr(1)	0.0665	0	0.1413	0.2631	0.0481	0.0856
Lane-changing accident in driving lane	Pr(0)	0.7876	1	0.697	0.2631	0.9327	0.7223
	Pr(1)	0.2124	0	0.303	0.7369	0.0673	0.2777
Accident caused by stopped vehicle	Pr(0)	0.491	0	1	1	0.2779	0.538
	Pr(1)	0.509	1	0	0	0.7221	0.462
Accident caused by driving vehicle	Pr(0)	0.8161	1	0.4443	1	0.8374	0.9209
	Pr(1)	0.1839	0	0.5557	0	0.1626	0.0791
Side collision	Pr(0)	0.9428	1	1	0	0.9423	0.8584
	Pr(1)	0.0572	0	0	1	0.0577	0.1416
Mutil-vehicle accident	Pr(0)	0.6887	0.7023	0.7537	0.8949	0.5966	0.8432
	Pr(1)	0.3113	0.2977	0.2463	0.1051	0.4034	0.1568
Weather_abnormal condition	Pr(0)	0.184	1	1	0.9211	1	0.9552
	Pr(1)	0.816	0	0	0.0789	0	0.0448
Road surface_abnormal condition	Pr(0)	0.2929	1	1	1	1	1
	Pr(1)	0.7071	0	0	0	0	0
Night time	Pr(0)	0.7312	0.762	0.7438	0.719	0.8942	0.6696
	Pr(1)	0.2688	0.238	0.2562	0.281	0.1058	0.3304
Tunnel	Pr(0)	0.8866	1	1	1	0.1059	0.9789
	Pr(1)	0.1134	0	0	0	0.8941	0.0211
Right curve	Pr(0)	0.9433	1	1	0.9395	0.9904	0.6725
	Pr(1)	0.0567	0	0	0.0605	0.0096	0.3275
Left curve	Pr(0)	0.948	1	1	0.9583	0.9817	0.5472
	Pr(1)	0.052	0	0	0.0417	0.0183	0.4528
Uphill	Pr(0)	1	1	1	1	1	0.5985
	Pr(1)	0	0	0	0	0	0.4015

<Table 3>은 최적 군집 수인 여섯 개의 군집의 사고지표 특성을 나타낸 표이다. <Table 3>의 Probability는 각 군집 개별 데이터가 군집에 속할 확률을 의미한다. 군집에 속할 확률인 $Pr(1)$ 이 높을수록 군집에서 해당 변수의 특성이 높음을 나타낸다. 군집 별 사고 특성 분석에서 군집에 상관없이 주간 시간대와 남성 가해 운전자 변수는 공통변수로 도출되었다. 군집(Cluster) 1은 정차 중 추돌사고가 상대적으로 많이 발생하고, 강우 등 이상기후이면서 노면 상태가 건조하지 않은 상황에서 교통사고가 자주 발생하였다. 군집 2는 군집 1과 달리 2차 사고와 같은 정차 중 추돌사고로만 구성되었다. 군집 3은 주행 중 추돌사고가 상대적으로 많이 발생하고 정차 중 추돌사고는 없는 것으로 나타났다. 군집 4는 주행차로 내 차로변경 사고가 대부분을 차지하고 차로변경이 완전히 완료되기 전에 발생하는 측면 추돌로만 구성되었다. 군집 5는 군집 1에 비해 정차 중 추돌사고가 더 자주 발생하였으나 기상이나 노면 상태는 정상인 경우에만 교통사고가 발생하였다. 특히 타 군집과 비교할 때 다중 차량사고와 터널 내 사고가 많은 것으로 나타나 2차 사고 피해가 많은 것으로 분석되었다. 군집 6은 평면선형이 좌 곡선 부와 오르막 구간이 상대적으로 많은 것으로 나타났으나, 타 군집에 비해 사고 특성이 가장 미흡한 것으로 분석되었다.

3. 사고 심각도 예측모형 평가

앞 절에서 언급했듯이 잠재 계층분석 방법에서 최적 군집 수는 6개로 도출되었고, 이때 각 군집 별 종속 변수인 사고 심각도는 부상 사고와 중상·사망사고로 구분하여 그 발생비율을 확인하였다. <Table 4>는 군집 별 총분석자료의 구성 비율, 부상 사고(Possible injury)와 중상·사망사고(Severe injury)의 지표 값이다. 전체 표본 수 대비 구성 비율은 군집 2(32.2%)로 가장 많고 그다음으로 군집 3(24.4%), 군집 4(14.5%) 순으로 나타났다. 군집 2와 군집 3이 전체의 56.6%로 반수 이상을 차지하였다. 부상 사고 경우, 군집 3(35.7%)이 가장 많고, 그다음으로 군집 4(30.8%), 군집 2(14.7%) 순으로 나타났다. 중상·사망사고 경우, 군집 2(36.1%)가 가장 많고 군집 3(21.9%), 군집 5(15.3%) 순으로 나타났다.

<Table 4> Injury severity share across Clusters

Cluster	Severity injury		
	Cluster size	Possible injury	Severe injury
Cluster 1	(71 Case, 9.03%)	(10 Case, 6.99%)	(61 Case, 9.53%)
Cluster 2	(252 Case, 32.16%)	(21 Case, 14.69%)	(231 Case, 36.09%)
Cluster 3	(191 Case, 24.36%)	(51 Case, 35.66%)	(140 Case, 21.88%)
Cluster 4	(114 Case, 14.50%)	(44 Case, 30.77%)	(70 Case, 10.94%)
Cluster 5	(104 Case, 13.27%)	(6 Case, 4.20%)	(98 Case, 15.31%)
Cluster 6	(51 Case, 6.68%)	(11 Case, 7.69%)	(40 Case, 6.25%)

*군집 별 표본 수와 비율을 나타냄

<Table 5>는 여섯 개의 군집에 대한 사고위험도 및 인명피해 관측값이다. 군집 1, 군집 2, 군집 5는 부상 사고 대비 중상·사망사고 비율이 상대적으로 높게 나타남에 따라 사고 심각도가 높은 그룹이다. 반대로 군집 3, 군집 4, 군집 6은 부상 사고 대비 중상·사망사고 비율이 상대적으로 낮아 사고 심각도가 낮은 그룹으로 분류되었다. 사고 심각도가 높은 그룹(군집 1, 2, 5)은 정차 중 추돌사고가 대다수를 차지하고 있어 2차 사고나 급정지가 필요한 상황에서 발생하는 사고는 인명피해가 큰 것으로 분석되었다. 강우 등의 이상기후와 노면이 건조하지 않은 상태에서도 사고 심각도가 높은 것으로 나타났다. 또한 정차 중 추돌사고는 차로변경 사고

에 비해 다중 차량사고가 더 많이 발생하는 것으로 나타났으며 군집 1, 2, 5에서 이러한 경향이 확인되었다. 이는 차량 추돌에서 다중사고가 더 심각한 운전자 부상과 사망에 영향을 미치는 Wu et al.(2014)와 유사한 결과가 도출되었다. Ma et al.(2009) 연구에서 고속도로 터널에서 사고가 발생하면 부상의 심각성이 높은 것으로 나타났는데 본 연구 결과에서도 사고 심각도가 높은 군집에서 터널 내 교통사고가 다수 발생하였다. 사고 심각도가 낮은 그룹(군집 3, 4, 6)은 주행 중 차로변경 사고, 주행 중 추돌사고가 자주 발생하고, 차로변경 사고는 측면 추돌이 대다수를 차지하였다. 또한 좌·우 곡선 부와 오르막 구간 내 교통사고는 상대적으로 사고 심각도가 낮은 것으로 나타났다. 연구 결과는 기존 문헌의 추돌사고 및 차로변경 사고에 대해 주행 중 추돌 사고, 정차 중 추돌사고, 추월차로 진입을 위한 차로변경 사고, 그리고 주행차로 진입을 위한 차로변경 사고로 세분화하여 특성이 나타났다.

<Table 5> Characteristics of accidents by Clusters

Cluster	Characterization	Possible injury	Severe injury
1	High risk accident	10	61
2	High risk accident	21	231
3	Low risk accident	51	140
4	Low risk accident	44	70
5	High risk accident	6	98
6	Low risk accident	11	40

본 연구는 잠재 계층분석을 통해 도출된 여섯 개의 군집에 대해 로짓 모형(Logit model)을 통해 군집 별 사고 심각도를 예측하였다. 기존 데이터를 학습하여 새로운 데이터를 검증하는 방법으로 이항 로짓 모형을 사용하였다. <Table 6>은 데이터 세트 구축과정에서 이 상처와 결측치를 제외한 고속도로 사고 783건을 분석에 사용하였고, 최종 데이터를 학습 데이터(70%, 548개)와 평가 데이터(30%, 235개)로 구분하였다. 사고 심각도가 높은 그룹(군집 1, 2, 5)은 427개로써, 학습 데이터(70%, 298개)와 평가 데이터(30%, 129개)로 구분하였고, 사고 심각도가 낮은 그룹(군집 3, 4, 6)은 356개로써, 학습 데이터(70%, 249개)와 평가 데이터(30%, 107개)로 나누어 최종적인 예측모형을 수행하였다.

<Table 6> Data composition

	Entire data	High risk accident	Low risk accident
Total data set	783	427	356
Train data set	548	298	249
Test data set	235	129	107

로짓 모형(Logit model)의 예측성능을 평가하기 위해 오 분류표(Confusion matrix) 기반의 정확도(Accuracy), 재현율(Recall), 정밀도(Precision)를 계산하였으며, 그 결과를 <Table 7>에 제시하였다. 정확도는 전체 데이터 수 중 예측 결과와 실제값이 같은 데이터의 비율이고, 재현율은 실제 Positive인 대상 중 예측과 실제값이 Positive로 일치한 데이터의 비율이며, 정밀도는 예측을 Positive로 한 대상 중 예측과 실제값이 Positive로 일치한 데이터의 비율이다. 이진 분류 모델의 신뢰도 측면에서 정확도뿐만 아니라 재현율과 정밀도를 사용하여 나타낸다. <Table 7>에 각각의 train data 및 test data 수에 대해 명시하였고, 로짓 모형(Logit model)의 예측

성능을 평가하기 위해 랜덤 샘플링 5번 수행을 통한 평균값을 나타냈다.

<Table 7> Results of confusion matrix

Accident data	Train data	Test data	Accuracy	Precision	Recall
1. Entire data (Cluster1~Cluster6)	548	235	0.782	0.813	0.937
2. High risk cluster data (Cluster1 + Cluster2 + Cluster5)	298	129	0.816	0.876	0.923
3. Low risk cluster data (Cluster3 + Cluster4 + Cluster6)	249	107	0.789	0.825	0.910

분류기의 예측성능을 나타내는 정확도는 전체 데이터에서 78.2%로 나타났으며, 재현율은 93.7%, 정밀도는 81.3%로 나타났고, 사고 심각도가 높은 군집에서 정확도는 81.6%로 나타났으며, 재현율 및 정밀도는 각각 92.3%, 87.6%로 분석되었다. 사고 심각도가 낮은 군집은 정확도는 78.9%로 나타났고 재현율 및 정밀도는 각각 91.0%, 82.5%로 나타났다. 로짓모형 분석 결과 군집 별로 위험사고 예측을 수행했을 때 정확도(Accuracy), 재현도(Recall), 정밀도(Precision) 모두 0.75 이상으로 도출되었다. 군 집별로 수행했을 때, 기존 전체 데이터보다 사고 심각도를 높은 그룹과 낮은 그룹으로 나누어 분석했을 때 정확도와 정밀도 값이 상대적으로 높게 도출되었다. 이는 잠재 계층분석을 통해 분류된 군집 데이터에서 사고 심각도가 높은 그룹과 사고 심각도가 낮은 그룹의 사고위험 예측성능이 향상된 것으로 분석되었다. 블랙박스 사고 영상에서 도출된 운전행태(2차 사고, 진로 변경, 급정지 등)를 반영한 군집분석이 사고 심각도 예측의 입력자료로 활용하는 것이 성능 평가에 효율적으로 작용하였다. 또한 인명피해 심각도에 따라 데이터를 분류하는 것이 사고위험 예측에 효과적인 것으로 판단된다.

V. 결론

본 연구는 고속도로에서 발생한 교통사고 블랙박스 영상을 기반으로 군집분석과 예측모형 비교를 수행하였다. 분석자료는 인적요인, 도로요인, 사고요인, 기타요인 등으로 구분하였으며, 사고 직전의 도로 및 교통 상황을 파악할 수 있는 차량 주행행태, 기상 및 노면 상태 등 사고 영상에서 추출이 가능한 항목을 설명 변수로 활용하였다. 분석 데이터의 이질성을 고려하기 위해 잠재 계층분석(Latent class analysis)을 이용하여 교통사고 특성을 분류하고, 도출된 군집을 토대로 이항로짓모형을 적용하여 예측성능을 분석하였다.

유형별 교통사고 특성을 분석하기 위해 잠재 계층분석을 통해 AIC와 BIC 값을 판단지표로 6개의 최적 군집 수를 도출하였다. 6개의 군집 간의 사고 특성을 살펴보면, 정차 중 추돌사고, 이상기후(강우, 적설 등), 노면 상태가 젖거나 결빙인 상태, 주행차로 내 차로변경 사고, 터널, 다중사고, 좌 곡선부 등이 군집의 분류항목으로 분석되었다. 특히 운전자의 주행행태(driving behavior) 측면에는 2차 사고와 같은 정차 중 추돌사고가 차로변경에 의한 추돌사고에 비해 상대적으로 사고 심각도가 높은 것으로 나타났다. 이러한 결과는 주로 차량 추돌사고가 사고 심각도에 영향을 미치는 기존 논문과 차이가 있다. 본 연구 결과는 추돌사고를 정차 중 추돌사고 및 주행 중 추돌사고로 세분화하여 나타났다. 또한 이상기후나 노면 상태가 건조하지 않은 도로 환경은 사고 건수는 정상 기후에 비해 적게 발생하나 사고 심각도는 상대적으로 높은 것으로 분석되었다. 고속도로 내 터널구간은 다소 어두운 곳으로 진입하는 상황에서 압순응이 발생하고 고속주행이 가능하므로 사고

심각도가 높은 것으로 나타났다. 또한 곡선 부의 경우, 차체 빔의 시야 가림에 영향을 주는 좌 곡선 부에서 상대적으로 사고가 자주 발생하였다.

잠재 계층분석으로 분류된 각 군집을 사고 심각도 수준의 비율을 기준으로 두 그룹으로 나누고, 종속변수인 인명피해 수준을 반영하여 사고 예측 평가를 수행하였다. 사고 심각도 비율은 부상 사고 대비 중상·사망 사고 비율을 기본으로 사고 심각도가 높은 그룹과 낮은 그룹으로 구분하였다. 사고 심각도 예측모형 성능 평가를 위한 분석자료는 전체 데이터, 군집분석에서 도출된 사고 심각도가 높은 그룹과 낮은 그룹 데이터를 활용하였다. 사고 심각도가 높은 그룹은 정차 중 추돌사고, 다중사고, 그리고 터널 내 사고가 다수를 차지하고 있는 것으로 확인되었다. 2차 사고, 급정지가 필요한 상황 그리고 터널 환경에서 발생하는 사고는 인명피해가 큰 것으로 분석되었다. 사고 심각도가 낮은 그룹은 주행 중 차로변경 사고, 주행 중 추돌사고 그리고 측면 사고가 다수를 차지하고 있는 것으로 나타났다. 차로를 변경 혹은 측면 사고로 발생한 교통사고는 사고 심각도가 높은 그룹에 비해 상대적으로 인명피해가 낮은 것으로 분석되었다. 사고 예측모형은 로짓 모형(Logit model)을 적용한 결과, 전체 데이터를 분석한 경우보다 잠재 계층분석으로 두 개의 사고 심각도 그룹을 분류하여 예측했을 때 정확도와 정밀도가 더 우수하게 나타났다. 특히, 사고 심각도가 높은 그룹의 정확도는 81.6%, 정밀도 87.6%로 다른 그룹과 달리 상대적으로 80% 이상의 우수한 예측 능력을 보여주었다. 이는 군집분석을 통한 그룹별 사고 특성과 사고 심각도를 반영하여 사고위험을 예측하는 것이 사고분석과정에서 더 효과적인 것으로 판단된다.

본 연구 결과는 교통사고 직전의 다양한 사고원인을 블랙박스 영상을 통해 항목을 추출하고 이를 군집기법과 사고위험도 예측성능을 비교하는 데 활용하였다. 이를 통해 2차 사고와 같은 정차 중 추돌사고, 차로변경 중 측면 추돌사고 등이 고속도로 교통사고에 중요한 주행행태변수로 작용한다는 것으로 도출하였다. 본 연구는 기존 논문과 달리, 블랙박스 사고 영상을 통해 직접적인 사고원인을 규명할 수 있는 요인을 설명변수로 적용하였다. 정면 추돌사고는 주행 중, 정차 중(2차 사고)으로 구분하였고, 측면 추돌사고는 주행차로 진입을 위한 차로변경, 추월차로 진입을 위한 차로변경으로 세분화하여 실질적인 사고 직전의 운전 행동을 변수로 사용한 것이 기존 논문과 차별성이 있다. 이는 고속도로 사고 예방을 위한 교통안전 관리전략을 수립하는 데 활용할 수 있을 것으로 판단된다. 또한 본 연구는 종속변수가 부상 사고와 중상·사망사고로 이분형으로 구성되어, 이산 선택모형(discrete choice model) 중 관측된 독립변수를 적합 시키며 주로 사용되는 방법론인 이항로짓모형을 활용하여 사고 심각도 분석을 진행하였다. 사고 심각도를 이항로짓모형을 통해 예측하였지만 향후 연구는 정확도 높은 사고위험 예측모형 개발을 위해 Nested Logit model, 신경망(NN), AI 분석 방법론을 적용할 필요성이 있다. 위 방법론을 활용해 다양한 모형을 개발하고 평가 및 비교를 통해 교통사고의 특성을 명확하게 분석할 수 있을 것으로 판단된다.

향후 연구과제로는 특정 차량 주행행태(차로변경 사고, 추돌사고 등), 사고 직전의 도로 환경(교량과 터널 진입·출입부 등), 교통류 상태(혼잡, 비 혼잡) 등을 별도로 구분하여 많은 표본 수를 확보하면 군집분석과 사고 위험도 예측모형 수행에서보다 현실에 적용할 수 있는 의미 있는 결과를 도출할 것으로 판단된다.

ACKNOWLEDGEMENTS

본 논문은 국토교통부 자율주행기술개발혁신사업 ‘주행 및 충돌상황 대응 안전성 평가기술개발(21AMDP-C160637-01)’ 과제 지원에 의해 수행되었습니다.

REFERENCES

- Abdel-Aty, M.(2003), “Analysis of driver injury severity levels at multiple locations using ordered probit models”, *Journal of Safety Research*, vol. 34, no. 5, pp.597-603.
- Aiash, A. and Robusté, F.(2021), “Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree”, *International Journal of Injury Control and Safety Promotion*, pp.1-9.
- Akaike, H.(1974), “A new look at the statistical model identification”, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp.716-723.
- Akaike, H.(1987), “Factor analysis and AIC”, *Psychometrika*, vol. 52, no. 3, pp.317-332.
- Charlton, M., Fotheringham, S. and Brunson, C.(2009), “Geographically weighted regression”, *White Paper*, National Centre for Geocomputation, National University of Ireland Maynooth 2.
- Depaire, B., Wets, G. and Vanhoof, K.(2008), “Traffic accident segmentation by means of latent class clustering”, *Accident Analysis & Prevention*, vol. 40, no. 4, pp.1257-1266.
- Eboli, L., Forciniti, C. and Mazzulla, G.(2020), “Factors influencing accident severity: An analysis by road accident type”, *Transportation Research Procedia*, vol. 47, pp.449-456.
- Garrido, R., Bastos, A., De Almeida, A. and Elvas, J. P.(2014), “Prediction of road accident severity using the ordered Probit model”, *Transportation Research Procedia*, vol. 3, pp.214-223.
- Gray, R. C., Quddus, M. A. and Evans, A.(2008), “Injury severity analysis of accidents involving young male drivers in Great Britain”, *Journal of Safety Research*, vol. 39, no. 5, pp.483-495.
- Hong, J., Tamakloe, R. and Park, D.(2019), “A comprehensive analysis of multi-vehicle crashes on expressways: A double hurdle approach”, *Sustainability*, vol. 11, no. 10, p.2782.
- Iranitalab, A. and Khattak, A.(2017), “Comparison of four statistical and machine learning methods for crash severity prediction”, *Accident Analysis & Prevention*, vol. 108, pp.27-36.
- Jeon, H. and Kim, J.(2020), “An Analysis of Multiple-Vehicle Accidents on Freeways Using Multinomial Logit Model”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 19, no. 5, pp.1-14.
- Lei, T., Peng, J., Liu, X. and Luo, Q.(2021), “Crash prediction on expressway incorporating traffic flow continuity parameters based on machine learning approach”, *Journal of Advanced Transportation*, vol. 2021, 8820402.
- Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X. and Zhang, G.(2019), “Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes”, *Accident Analysis and Prevention*, vol. 129, pp.230-240.
- Ma, Z. L., Shao, C. F. and Zhang, S. R.(2009), “Characteristics of traffic accidents in Chinese freeway tunnels”, *Tunnelling and Underground Space Technology*, vol. 24, no. 3, pp.350-355.
- Mohamed, S. A., Mohamed, K. and Al-Harthi, H. A.(2017), “Investigating factors affecting the occurrence and severity of rear-end crashes”, *Transportation Research Procedia*, vol. 25, pp.2098-2107.
- Nagler, J.(1994), “Scobit: An alternative estimator to logit and probit”, *American Journal of Political Science*, vol. 38, no. 1, pp.230-255.
- Nylund, K. L., Asparouhov, T. and Muthén, B. O.(2007), “Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study”, *Structural*

Equation Modeling: A Multidisciplinary Journal, vol. 14, no. 4, pp.535-569.

- Raftery, A. E.(1986), “A note on Bayes factors for log linear contingency table models with vague prior information”, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 48, no. 2, pp.249-250.
- Sun, Z., Xing, Y., Wang, J., Gu, X., Lu, H. and Chen, Y.(2021), “Exploring injury severity of bicycle-motor vehicle crashes: A two-stage approach integrating latent class analysis and random parameter logit model”, *Journal of Transportation Safety & Security*, pp.1-27.
- Traffic Accident Analysis System, <http://taas.koroad.or.kr>, 2022.09.30.
- Vermunt, J. K. and Magidson, J.(2002), *Latent class cluster analysis. Applied latent class analysis*, SAGE Publishing, pp.89-106.
- Wu, Q., Chen, F., Zhang, G., Liu, X. C., Wang, H. and Bogus, S. M.(2014), “Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways”, *Accident Analysis & Prevention*, vol. 72, pp.105-115.
- Yu, R. and Abdel-Aty, M.(2013), “Multi-level Bayesian analyses for single-and multi-vehicle freeway crashes”, *Accident Analysis & Prevention*, vol. 58, pp.97-105.