

자율주행차 사고심각도의 영향요인 분석에 관한 연구: 사고데이터와 교통인프라 정보를 결합하여

A Study on Factors Influencing the Severity of Autonomous Vehicle Accidents: Combining Accident Data and Transportation Infrastructure Information

김 창 훈* · 김 정 화**

* 주저자 : 경기대학교 일반대학원 도시·교통공학과 석사과정

** 교신저자 : 경기대학교 스마트시티공학부 도시·교통공학과 조교수

Changhun Kim* · Junghwa Kim**

* Dept. of Urban Planning and Transportation Engineering., Kyonggi University

** Dept. of Urban Planning and Transportation Engineering., Kyonggi University

† Corresponding author : Junghwa Kim, Junghwa.kim@kyonggi.ac.kr

Vol. 22 No.5(2023)
October, 2023
pp.200~215

pISSN 1738-0774
eISSN 2384-1729
<https://doi.org/10.12815/kits.2023.22.5.200>

Received 18 September 2023
Revised 5 October 2023
Accepted 16 October 2023

© 2023. The Korea Institute of
Intelligent Transport Systems. All
rights reserved.

요 약

자율주행 기술이 고도로 발전하고, 관련 시장이 급격하게 성장하고 있어 머지않은 시기 내에 완전 자율주행 시대가 도래할 것으로 예상된다. 한편, 자율주행 기술의 발전과 함께 기술 안전성에 대한 의문이 제기되고 있으며, 관련 사고 소식이 보도되면서 기술에 대한 우려는 증대되고 있다. 자율주행차의 안전성 향상을 위해, 사고 사례를 분석하고 사고 원인을 규명하는 행위가 선행될 필요가 있다. 이에, 본 연구는 자율주행 사고데이터를 통해 자율차 사고의 심각도에 대한 영향요인을 분석하였다. 연구 데이터는 CA DMV에서 수집·배포하고 있는 자율주행차 사고 레포트를 중심으로 사고 지점의 공간 정보, 교통 정보를 사용하였다. 중점 데이터가 사고 레포트임을 고려할 때, 사건 발생 횟수의 기댓값이 반영될 수 있도록 포아송 회귀 분석을 사용하여 모델링을 진행하였다. 모형 분석 결과, 자율주행차 사고 심각도는 조도가 낮을 때, 자전거·버스 전용 차로가 존재할 때, 보행자와 자전거 사고 이력이 많은 지역에서 증가한다는 결과가 도출되었다. 본 연구 결과는 향후, 자율주행차 안전성 개선을 위한 알고리즘 개발 및 관련 교통 인프라 설치를 위한 기초자료로 활용될 수 있을 것이다.

핵심어 : 사고 심각도, 자율주행차, CA DMV 보고서, 차원 축소, 포아송 회귀

ABSTRACT

With the rapid advance of autonomous driving technology, the related vehicle market is experiencing explosive growth, and it is anticipated that the era of fully autonomous vehicles will arrive in the near future. However, along with the development of autonomous driving technology, questions regarding its safety and reliability continue to be raised. Concerns among technology adopters are increasing due to media reports of accidents involving autonomous vehicles. To promote the improvement of the safety of autonomous vehicles, it is essential to analyze previous accident cases and identify their causes. Therefore, in this study, we aimed to analyze the factors influencing the severity of autonomous vehicle accidents using previous accident cases and related data. The data used for this research primarily comprised autonomous vehicle accident reports collected and

www.kci.go.kr

distributed by the California Department of Motor Vehicles (CA DMV). Spatial information on accident locations and additional traffic data were also collected and utilized. Given that the primary data used in this study were accident reports, a Poisson regression analysis was conducted to model the expected number of accidents. The research results indicated that the severity of autonomous vehicle accidents increases in areas with low lighting, the presence of bicycle or bus-exclusive lanes, and a history of pedestrian and bicycle accidents. These findings are expected to serve as foundational data for the development of algorithms to enhance the safety of autonomous vehicles and promote the installation of related transportation infrastructure.

Key words : Accident Severity, Autonomous vehicle, CA DMV Report, Dimension reduction, Poisson regression

I. 서 론

자율주행 기술은 기존 교통 산업의 패러다임을 전환할 수 있는 잠재력을 가졌으며 전 세계적으로 국가 및 기업 단위의 활발한 투자를 통해 급속히 발전하고 있다. KISTI¹⁾에 따르면 레벨 3 이상 자율주행 차량의 전 세계 시장 규모는 2025년 1,500억 달러에서 2035년에 1.1조 달러로 성장할 것으로 예상된다. ICT²⁾ 기업들의 자율주행 시장 진출이 활발히 이루어지고 있으며, 완성차 업체와 ICT 기업 간에 자율주행 시장 주도권을 놓고 경쟁과 수평적 협력이 확대되고 있다. 자율주행 차량의 상용화는 도로 안전성을 높이고 교통 혼잡을 저감 시키며 보다 효율적인 교통을 제공할 것이다. 이전에도 그랬듯이 새로운 기술의 정착을 위해서는 극복해야 할 위험 요인과 도전과제가 존재한다. 자율주행 기술에 대한 가장 큰 우려 중 하나는 안전성이다. 세계 경제 포럼(2015년)이 잠재적 수용자들을 대상으로 자율주행 기술에 대해 우려하는 점을 조사한 결과, 안전, 제어 문제, 차량 고장, 혼잡한 교통 상황에서 기술 신뢰성 부족 등을 걱정하였다. 이러한 결과는 사람들이 자율주행 차량의 안전성과 기술의 신뢰성에 대한 문제를 걱정하고 있음을 나타낸다.

자율주행 기술은 인간의 실수로 인해 발생하는 다수의 사고를 예방할 수 있겠지만, 모든 사고를 방지할 수는 없다. 자율주행 차량이 복잡하고 예측할 수 없는 환경에서 운영될 때, 그들의 센서와 소프트웨어가 처리하기에 적합하지 않은 예상치 못한 상황을 직면할 수 있기 때문이다. 여기서, 자율주행 차량 사고 원인 파악을 통해 기술의 취약성을 보완하는 연구의 필요성은 강조된다. 해당 연구는 자율주행 차량의 안전성을 향상하고 새로운 기술에 대한 공중의 신뢰를 쌓는 데 도움을 줄 수 있으며, 자율주행 차량을 위한 표준과 각종 법·규제를 개발하는데 기초자료로써 활용할 수 있을 것이다. 앞서 언급한 자율주행차 사고에 관한 연구의 필요성을 고려하여, 본 연구는 도로에서 발생한 실제 자율주행차 사고데이터를 통해 사고의 원인을 도출하고 각각의 요인들이 사고에 미치는 영향력 분석을 목표로 한다.

본 연구의 수행 절차는 세 단계로 기술할 수 있다. 첫째, 기존의 차량사고 모델링 연구를 기반으로, 본 연구에 적용 가능성을 검토하였다. 둘째, 데이터 수집, 처리, 분석 과정을 진행하여 분석에 적합한 데이터를 확보하였다. 셋째, 데이터를 기반으로 적절한 분석을 통해 자율주행차 사고의 심각도에 대한 모형을 구축하였다. 연구 데이터로, 자율주행차 사고데이터는 CA DMV³⁾에서 수집하였으며, 사고 위치에 대한 공간 및 교통 데이터는 Google Earth와 CAL TRANS⁴⁾에서 수집하였다. 사고데이터, 공간 데이터, 교통 데이터 간의 시공간

1) KISTI : Korea Institute of Science and Technology Information, 한국과학기술정보연구원

2) ICT : Information and Communication Technology, 정보와 통신 기술

3) CA DMV : California Department of Motor Vehicles, 캘리포니아 차량관리국

적 특성을 최대한 일치시켰으며, 데이터 일관성을 확보하고자 일정한 기준을 통해 데이터를 수집하였다.

II. 선행연구 검토

1. 교통사고 분석 연구

자율주행 자동차의 안전성 평가와 사고 분석에 관한 대다수의 선행연구는 차체 혹은 자율주행 알고리즘 성능을 분석하는 데 초점을 두고 있으며, 대부분이 테스트베드와 같은 시험환경이나 시뮬레이션 기반의 가상 환경에서 수행되었다. 반면 자율주행차의 사고 내용이나 위험성을 분석한 연구는 매우 부족한 실정이다. 이에 따라, 일반 자동차를 대상으로 진행된 사고 분석 연구를 중점적으로 확인하여 변수 설정 및 모형화 방안을 검토하였다. 교통사고에 영향을 미치는 요인은 매우 다양하며, 사고를 예측하거나 사고의 위험성을 평가하기 위한 연구는 국내외에서 활발하게 진행되고 있다.

Oh et al.(2005)는 충청남도과 충청북도의 77개 교차로에 대해 계층적 샘플링을 적용하여 교통사고와 교통량 및 주변 환경 요인 사이의 관계를 분석하고 교통사고 예측 모델을 개발하였다. 해당 연구에서는 주요 도로와 부도로의 평균 일일 교통량 (ADT), 주요 도로의 대형 차량 비율, 부도로의 출입구 수, 주요 도로의 중앙 분리대 및 주요 도로의 속도 제한 등을 변수로 설정하고, 해당 변수들이 교통사고에 미치는 영향을 파악하기 위해 포아송 회귀 모델을 적용하였다. 분석 결과, 해당 도로의 ADT가 교통사고에 가장 큰 영향을 미치는 주요 요인임을 밝혔다. Shim and Hwang(2009)는 교통사고 발생률과 지역별 교통사고를 줄이기 위한 연구로 도로 사용자, 도로 및 네트워크 구간에 따른 사고 위험 요인을 모델링하였다. 해당 연구에서는 춘천의 교차로를 대상으로 사고 발생 상황, 기하 구조, 통제 방법, 교통량 및 회전 교통량을 분석하고, 추출된 평가 요소 간의 상관 분석을 통해 변수를 도출했습니다. 도출된 변수는 교통량, 도로 폭 및 교차로 각도로, 판별 분석 모델링 기법을 적용하여 교통사고 위험을 세 가지 그룹으로 분류하였다. 이와 유사하게 Park(2014)은 대한민국 청주의 24개 도로에서 수집한 데이터를 기반으로 교통사고 모델을 구축하였다. 2단계 최소 자승 추정 방법을 사용하여 교통사고 횟수와 교통량 사이의 쌍방 관계를 분석하고, 주요 도로와 보조 도로로 구간을 나누어 사고 횟수와 교통량에 대한 외생 변수의 영향을 비교·분석하였다. 교통사고 횟수에 영향을 미치는 공통 변수는 버스 정류장 수였으며, 교통량에 영향을 미치는 공통 변수는 차로 수와 구간 길이가 도출되었다. 또한, 좌회전 및 우회전 차로의 존재가 교통사고에 미치는 영향이 큰 것으로 분석되었다.

앞선 연구들은 교통사고 위치의 공간적 특성을 주요 변수로 채택하여 모델링을 진행하였으며, 이와 달리 운전자 특성, 환경 특성 및 특성을 주요 변수로 한 모델링 연구도 존재한다. Seong et al.(2015)은 지역 내 교통사고 감소를 위한 대안 수립을 목적으로, 토지 이용 변화에 따른 교통사고 원인을 분석하였다. 해당 연구에서는 진주의 5년간 교통사고 데이터를 분석하여, 주거 및 상업 지역을 대상으로 사고 빈도수 기반의 클러스터링 지도를 구축하였다. 교통사고 빈도는 봄에 가장 높았고 겨울에 가장 낮았으며, 낮 시간대가 저녁 시간대에 비해 약간 높았다. 토지 이용에 따른 교통사고 클러스터 지도 분석에서는 겨울철 주거 지역의 사고 밀도가 특히 낮게 도출되었다. 차량 소유유형에 따른 교통사고 심각성 모델링 연구도 존재한다. Lee et al.(2022)은 제주시의 교통사고 데이터를 기반으로 렌터카 사고와 자차 사고 간 특성 차이를 조사하고 사고 심각성에 대한 관계를 검토하였다. 렌터카 사고 그룹에서는 20대와 30대 운전자 및 여성 운전자의 비율이 높

4) CAL TRANS : California Department of Transportation, 캘리포니아 교통국

았으며 차량 간 충돌, 단독 사고 및 정면충돌의 비율이 높게 도출되었다. 해당 연구에서는 포아송 회귀분석을 사용하여 모델링을 진행하였으며, 분석 결과 렌터카 사고 그룹은 부상의 심각도가 더 높게 도출되었다.

최근 이르러서야 국외에서 자율주행 차량에 대한 사고 요인과 위험에 관한 모델링 연구가 드물게 진행되었다. Xu et al.(2019)은 캘리포니아 차량관리국(CA DMV)에서 수집한 사고 데이터로 자율주행차 사고의 특성과 패턴을 조사하였다. 사고 위치, 날씨, 운전 모드, 주행 행동, 차량 속도, 충돌 유형, 충돌 심각성 및 차량 손상 부위와 같은 변수들로 사고의 특성을 설명하였다. 이후 부트스트랩을 기반으로 한 이진 로지스틱 회귀 분석을 수행하여 사고 심각성과 충돌 유형을 모델링하였다. 연구 결과로 운전 모드, 충돌 위치, 노상 주차 존재, 일방통행 도로가 사고 심각성에 미치는 영향이 크며, 자율주행차의 주행 행동 중 일시 정지, 회전 및 앞지르기가 충돌 유형에 영향을 끼치는 것으로 분석되었다. Chen et al.(2020)은 CA DMV의 사고 보고서 데이터와 Google 지도 API⁵⁾와 POI⁶⁾ 데이터를 사용하여 자율주행차 사고의 심각성을 예측하고 사고 심각성에 영향을 미치는 주요 요인을 분석하였다. 충돌 유형, 사고 심각성, 주행 행동 및 차량 손상 정도를 고려하여 사고의 특성을 도출하였다. 두 가지 다른 분류 모델인 eXtreme Gradient Boosting(XGBoost) 및 Classification and Regression Tree(CART)를 적용하였을 때, XGBoost 모델이 자율주행차와 관련된 부상 사고를 식별하는 데 더 우수한 성능을 확인하였다. 해당 모델을 적용하여 분석하였을 때 날씨, 차량 손상 정도, 사고 위치 및 충돌 유형이 사고 심각성을 결정하는 주요 요인으로 확인되었으며, 부상을 유발하는 사고는 다양한 토지 이용 패턴이 있는 지역에서 발생 빈도가 높은 것으로 관찰되었다.

2. 자율주행차의 사고 유형과 피해 심각도

Favaro et al.(2017)는 2014년부터 2017년까지 캘리포니아에서 발생한 자율주행차량 데이터를 사용하여 사고 유형, 손상 위치 및 사고 발생 빈도를 분석하였다. 이를 통해, 교차로에서 자율주행차 사고 빈도가 가장 높으며, AV와의 후방 충돌이 가장 빈번하며, 충돌 시 상대속도가 10마일 이하인 경우가 60% 정도를 차지하는 것을 밝혔다. Kim and Cho(2020)은 자율주행차의 사고 유형을 분석하였다. 자율주행차 사고데이터를 활용하여 사고 유형을 3가지(교차로에서 직진, 좌회전, 우회전)로 구분하였고, 각 유형에서 어떠한 사고가 발생하는지를 분석하였다. 직진 상황에서는 자율주행차가 정지하고 후방 차량이 추돌할 때의 사고가 가장 빈번하게 발생하였다. 좌회전 상황에서는 후방 차량이 자율주행차를 추월하려고 할 때 사고가 가장 빈번하게 발생하였다. 우회전 상황에서는 자율주행차가 양보를 위해 대기하거나, 회전 시 감속할 때 가장 많은 사고가 발생하였음을 밝혔다. 위 3가지 유형의 사고로부터 사고의 원인은 자율주행차와 일반 운전자 간의 상호 작용 부족이 원인이 되었음을 도출하였다.

3. 차원 축소 기법

차원 축소는 데이터 집합 내 변수 개수를 줄이면서 기존의 데이터 정보를 최대한 보존하기 위해 사용되는 기술이다. 차원 축소 기법은 복잡한 데이터 집합을 분석하고 시각화하기 쉽게 만들며, 계산의 복잡성을 줄이고 기계 학습 알고리즘의 성능을 강화하는 등 여러 가지 장점이 존재한다. (Ma and Zhu, 2013). 차원 축소에는 주성분 분석(PCA), 요인 분석(FA) 및 t 분포 확률적 임베딩(t-SNE)을 포함한 다양한 기법이 존재한다. 각 기법은 고유한 장단점이 존재하며, 데이터 집합의 특성과 분석 목표에 따라 어떤 기법을 사용할지 결정하게

5) API : Application Programming Interface, 응용 프로그램 인터페이스

6) POI : Point of Interest, 관심 지점

된다(Favarò et al., 2017).

FA와 PCA는 차원 축소의 대표적인 기법이다. PCA는 데이터 집합의 차원을 줄이면서 가능한 한 원래 분산을 보존하기 위해 사용된다. 가장 많은 분산을 설명하는 주성분을 유지함으로써 데이터 집합 내 변수의 수를 줄일 수 있으며, 동시에 가장 중요한 정보를 유지할 수 있는 것이다. 이는 데이터 압축, 특성 추출 및 데이터 시각화 등 여러 응용 분야에 적용할 수 있다. 전반적으로 PCA는 고차원의 데이터 집합을 분석하고 데이터 내의 기저 구조를 식별하는 강력한 기술로 평가할 수 있으며, 데이터 과학, 기계학습 및 이미지 처리를 포함한 다양한 분야에서 널리 사용된다. FA는 차원 축소와 데이터 집합 내 잠재 변수를 식별하기 위해 사용되는 통계적 기법이다. FA의 주요 아이디어는 데이터 집합 내의 관측 변수가 잠재 변수인 요인에 의해 유발된다는 것이다. FA의 주요 목표는 내재 요인을 식별하고 관찰된 변수와의 관계를 식별하는 것에 있다(Yong and Pearce, 2013). PCA와 FA 모두 차원 축소에 사용되지만, PCA는 데이터 내 분산을 가장 잘 설명하는 핵심 요인을 발견하는 데 목적을 두며, FA는 관찰된 변수 사이의 상관관계를 설명하는 잠재적인 요인을 찾는 데 목적을 둔다. 두 기법 모두 장단점이 존재하며, 적용할 기법의 선정은 분석 목표와 연구 가설에 따라 결정된다.

4. 포아송 회귀 분석

포아송 회귀는 역학, 생태학, 금융, 사회 과학 등 다양한 분야에서 사건의 발생 횟수(Count data)를 모델링하는 데 널리 사용되는 통계적 기법이다. 이 방법은 사건 발생 변수와 하나 이상의 예측 변수 간의 관계를 추정하기 위해 사용되는 일반화 선형 모델(GLM)의 일종이다(Coxe et al., 2009).

포아송 분포는 포아송 회귀 모델 종속변수의 기본 확률 분포이다(Hayat and Higgins, 2014). 포아송 분포는 일정한 시간 또는 공간 간격 내에서 발생하는 일정 수의 사건의 확률을 모델링하는 이산 확률 분포로, 평균 발생률이 알려진 경우, 해당 확률 분포를 사용하게 된다(Frome, 1983). 고객 서비스 센터가 특정 시간 동안 받는 전화 횟수, 특정 도로 구간에서 하루 동안 발생하는 교통사고 횟수, 특정 기간 동안 인구 내 특정 질병의 새로운 감염 사례 수와 같은 데이터는 포아송 분포로 모델링 될 수 있다. 포아송 분포는 평균과 분산을 나타내는 단일 매개 변수 λ 를 가지고 있으며, 주어진 시간 또는 공간 내에서 k 사건을 관측할 확률은 포아송 확률 질량 함수에 의해 제시되며 이는 Equation(1)과 같다.

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} \dots\dots\dots (1)$$

사건의 발생 횟수를 모델링할 때 포아송 회귀는 다른 회귀 방법에 비해 여러 가지 이점이 있으며, 특히 종속변수가 비음수 및 이산형인 경우, 그리고 데이터의 분산이 평균과 비례할 때 유용하다. 포아송 회귀 모형은 비교적 해석하기 쉽고 범주형 및 연속형 예측 변수 또한 처리할 수 있다(Jessalyn et al., 2021). 포아송 회귀 모형은 종속변수의 분산이 평균보다 큰 경우 발생하는 과대산포 문제를 해결하는 데 확장될 수 있다. 이는 포아송 분포의 과대산포에 대한 가정이 위배 될 때 발생할 수 있으며, 이때 포아송 회귀 대안으로 준 포아송 모델을 사용할 수 있다(Ver. Hoef and Boveng, 2007). 요약하면, 포아송 회귀는 다양한 분야에서 카운트 데이터를 모델링할 때 적합한 통계 기법이다. 기존의 선형 회귀 모델은 교통사고 발생의 불규칙성, 이산형, 비음수적 특성을 적절하게 고려하기 어려운 한계가 있으나 포아송 회귀 모형은 교통사고의 확률 분포를 적절하게 반영할 수 있다. 기존의 많은 선행연구에서 교통사고 분석에 포아송 회귀 모델을 사용하였으며, 분석 결과로 예측 변수와 계수 간의 상호 관계를 파악할 수 있음을 확인하였다.

선행연구로부터 교통사고 발생확률 및 사고 심각성을 예측하는 모델은 시간, 공간 및 환경 요인 등 복합적인 영향요인을 기반으로 설계할 수 있다는 것이 확인되었다. 특히 교통사고 데이터와 같이 사건 발생확률에 기인한 데이터의 경우 비선형 회귀 분석의 한 종류인 포아송 회귀분석이 적절한 것으로 분석되었다. 본 연구에서는 선행연구에서 밝혀낸 교통사고 영향 변수와 모델링 방법론을 참고하여 자율주행차 사고의 심각성을 예측하기 위한 모델을 개발하고자 한다.

Ⅲ. 연구 데이터

1. 캘리포니아 차량관리국(CA DMV) 자율주행차 사고 레포트

캘리포니아 차량관리국의 자율주행차 사고 레포트는 연례 보고서로 캘리포니아 주요 도로에서 발생한 자율주행차(AV)와 관련된 데이터를 제공하고 있다. 사고 레포트는 캘리포니아 법률에 따라 의무화되며 모든 자율주행차는 사고의 종류와 관계없이 10일 이내에 차량관리국에 레포트를 보고해야 한다. 해당 보고서에는 자율주행차의 사고 내용, 차량 종류, 자율주행 기술 수준 및 충돌 원인 등 사고 상황에 대한 전반적인 정보가 포함되어 있다. 본 연구에서는 CA DMV에 등록된 자율주행차의 사고 레포트로부터 분석에 필요한 데이터를 수집하였다. 수집한 보고서의 시간적 범위는 2019년부터 2021년까지이며, 공간적 범위는 샌프란시스코 카운티에서 발생한 사고만 선별하였다. 중복된 레포트와 누락된 항목이 있는 레포트는 제외하고 총 102건의 사고 레포트를 수집하였다. 전체 데이터 표본의 크기는 모델링 학습을 위해 제한적일 수 있지만, 캘리포니아 지역에서 발생한 모든 자율주행차 사고를 포함하고 있는 전수조사 데이터이므로 충분히 의미 있는 결과를 도출할 수 있을 것으로 판단하였다. 이후에는 차량 사고와 관련된 선행연구에서 공통적으로 다루고 있는 영향요인들을 참고하여 레포트 내에서 추출할 요인들을 선별하였다. 사고 레포트, 공간 정보 데이터, 교통 정보 데이터에 대한 수집 경로가 다르므로 데이터의 원활한 결합을 위해 교차로 내에서 발생한 사고만을 분석 대상으로 진행하였으며 자율주행차 사고의 대부분이 교차로 내에서 발생하기 때문에 제외되는 데이터 또한 극히 일부였다. CA DMV 사고 레포트에서 선정한 자율주행차 사고의 영향 요인은 <Table 1>과 같으며, 시간과 장소, 차량의 피해 범위 및 피해 정도, 날씨, 조도, 노면 상태 등의 내용으로 구성된다. CA DMV 레포트의 구성요소 중 분석에 사용된 변수들에 대한 기술 통계는 <Table 2>와 같다.

<Table 1> Variables and contents collected from the CA DMV accident report

Variables	Contents
Date of Accident	Day / Month / Year
Time of Accident	Hour / Minute
Location of Accident	Accident Location
Vehicle Damage	Unknown / None / Minor / Moderate / Major
Damaged Area	Display the damaged area in units of cells
Weather	Clear / Cloudy / Raining / Snowing / Fog / Other
Lightning	Daylight / Dusk-Dawn / Dark
Roadway Surface	Dry / Wet / Snow-Icy / Slippery

<Table 2> Descriptive statistics of variables extracted from CA DMV accident report

Variables	Number	Min.	Max.	Avg.	SD
Time of Accident	102	1	2	1.66	0.477
Vehicle Damage	102	1	5	2.98	0.796
Damaged Area	102	0	4	1.72	0.927
Weather	102	1	3	1.17	0.447
Lightning	102	1	3	1.74	0.954
Roadway Surface	102	1	2	1.04	0.195

2. 공간 정보 데이터

CA DMV의 자율주행차 사고 레포트에서 제공되는 공간 정보가 충분하지 않으므로, 사고 위치의 세부적인 공간 정보는 Google Earth의 응용 프로그래밍 인터페이스(API)를 사용하여 수집하였다. 공간 정보를 추출할 때, 사고가 발생했던 시공간적 시점과 최대한 일치 시켰다. 공간 정보는 시간에 따른 변동성이 상대적으로 적기 때문에 사고 시점의 정보를 일정 수준에서 보존하고 있다는 전제하에 데이터를 수집하였다. 데이터의 정확성과 일관성을 확보하기 위해 정의한 공간 정보 변수의 유형과 추출 방법은 <Table 3>에 제시하였다. 수집된 변수들은 교차로 및 도로의 기하·공간 정보와 정류장 및 전용차로의 존재 여부 등의 내용으로 구성되며, 거리 단위는 m를 기준으로 하였다. 공간 정보 데이터의 구성 요소 중 분석에 사용된 변수들에 대한 기술 통계는 <Table 4>와 같다.

<Table 3> Spatial information variables extracted from the Google Earth API

Variables	Contents and Format
Area of Intersection	Setting the end of crosswalk as the boundary of the intersection to calculate the area(square meter)
Distance between Intersections	Distance between the stop line of the current intersection and the stop line of the next intersection(meter)
Number of Lanes	The value obtained by dividing the sum of lanes in each direction of an intersection by 2(unit of number)
Road Width	From the center line of a lane to the center line of an adjacent lane(meter)
Left and Right Turn Lanes	The number of lanes exclusively designated for left or right turns(unit of number)
Exclusive Lanes	The number of lanes exclusively designated for bicycles, buses (unit of number)
Number of Crosswalks	The Number of pedestrian crossings around accident point (unit of number)
Number of Bus stops	Bus stops within a radius of 30m from the intersection(unit of number)

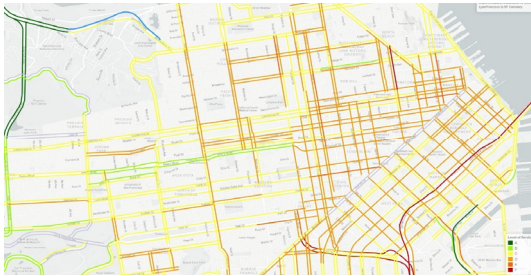
<Table 4> Descriptive statistics of variables extracted from Google Earth API

Variables	Number	Min.	Max.	Avg.	SD
Area of Intersection	102	192	2783	773.60	595.606
Distance between Intersections	102	30	250	103.75	40.825
Number of Lanes	102	2	13	5.91	2.045
Road Width	102	2.0	3.8	3.075	0.2735

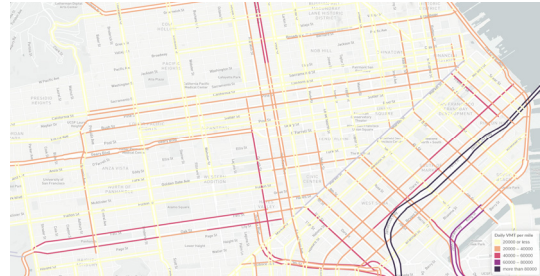
Variables	Number	Min.	Max.	Avg.	SD
Left and Right Turn Lanes	102	0	3	0.42	0.814
Exclusive Lanes	102	0	3	0.37	0.659
Number of Crosswalks	102	0	8	3.84	1.533
Number of Bus stops	102	0	9	1.04	1.469

3. 교통 정보 데이터

공간 정보와 마찬가지로, CA DMV의 사고 레포트에서 제공되는 교통 정보 데이터가 충분하지 않으므로, 샌프란시스코 교통국(SFCTA)에서 추가적인 교통 정보 데이터를 수집하였다. 수집된 교통 데이터의 예시는 <Fig. 1>, <Fig. 2>와 같으며, 해당 그림은 차량 흐름 및 각 도로 네트워크에 대한 차량 속도, 밀도, 혼잡도 및 관련 지표를 기반으로 측정된 서비스 수준(LOS) 할당 내역과 차량 주행 거리(VMT)를 나타낸다. <Table 5>은 수집된 교통 정보 데이터의 유형과 내용이며, 사고 지점을 포함하고 있는 링크(Link)의 데이터를 추출하였다. 교통 정보 데이터의 구성 요소 중 분석에 사용된 변수들에 대한 기술 통계는 <Table 6>와 같다.



<Fig. 1> LOS at accident locations using data



<Fig. 2> VMT at accident locations using data

<Table 5> Traffic information variables extracted from SF CTA data

Variables	Contents and Format
Vehicle Miles Traveled (VMT)	Normalized by the length of the segment and could be interpreted as approximate average daily traffic volume.
Auto Level-of-Service (LOS)	Grades road segments by vehicle delay with "A" describing free flow, and "F" describing bumper-to-bumper conditions. It represents the LOS during the autonomous accident time zone
Bicycle and Pedestrian Collision Data	Collision locations are snapped to the nearest intersection using the latitudes and longitudes given from SWITRS to the latitude and longitudes of the intersections in San Francisco. This data is based on accident history collected from 2006 to 2017.

<Table 6> Descriptive statistics of variables extracted from SF CTA data

Variables	Number	Min.	Max.	Avg.	SD
Vehicle Miles Traveled (VMT)	102	1	4	1.69	0.783
Auto Level-of-Service (LOS)	102	10	80	40.515	15.683

Variables	Number	Min.	Max.	Avg.	SD
Bicycle Collision Data	102	0	30	3.44	4.557
Pedestrian Collision Data	102	0	19	4.01	3.956

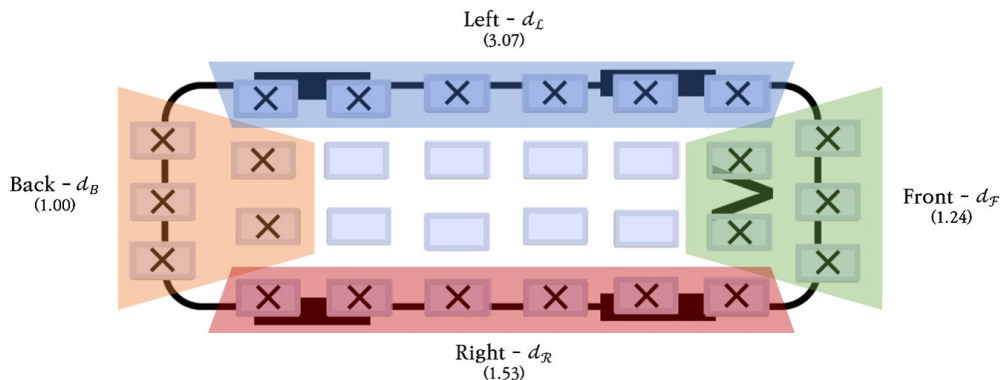
IV. 변수설정

본 연구에서는 자율주행차의 사고를 분석하고 사고와 관련된 여러 요인을 탐색한 후 사고심각도에 영향을 주는 요인을 분석하는 것을 목적으로 한다. 모델링을 진행하기에 앞서, 모델을 구성하는 변수를 선정 방법에 대해 논의하고자 한다. 사고심각도를 정량화할 수 있는 종속변수를 설정할 때, 수집된 데이터 내에서 추가적인 계층화 과정이 수행되었다. 통계적 모델링 관점에서 종속변수의 분할은 특정 기준에 따라 집합화하는 것을 의미하며, 이는 모델의 예측 능력에 유의미한 영향을 끼친다. 종속변수를 더 작고 더 동질적인 집합으로 분할 할 때, 통계 모델은 보다 미묘한 패턴을 감지하고, 예측력이 증대되며, 종속 및 독립 변수 간의 관계를 더 정확하게 이해할 수 있게 된다. 종속변수를 분할함으로써 결과의 해석 가능성을 향상할 수 있으므로, 본 연구에서 종속변수 분할 과정을 수행하였다. 종속변수는 Equation(2)로 정의된다.

$$y = y_d \times y_a \times d_x \dots\dots\dots (2)$$

여기서, y 는 사고심각도를, y_d 는 피해 정도를, y_a 는 피해 범위를, d_x 는 피해 지점별 가중치를 나타낸다.

자율주행차의 피해 정도를 나타내는 변수인 y_d 는 교통사고 조사관을 위한 차량 손상 척도(Investigators National Safety Council, 1983)에서 제시된 7단계 척도를 사용하여 충돌 보고서에서 수집된 피해 정도에 관한 내용을 양적 정보로 변환하였다. 피해 정도에 대응되는 값은 다음과 같다: 피해 없음 = 1, 경미한 피해 = 3, 중간 정도의 피해 = 5, 심각한 피해 = 7. y_a 와 d_x 는 피해 범위와 피해 위치에 따라 결정된다. 충돌 보고서에서 자율주행차 피해 범위(영역)는 30개의 단위 셀로 나누어진 차량 이미지에 표시된다. y_a 는 표시된 단위 셀의 개수를 기반으로 산출되며, d_x 는 단위 셀이 차량의 전면, 후면, 좌측 혹은 우측 영역에 속하는 위치에 따라 결정됩니다. 단위 셀이 속한 영역에 따라 서로 다른 d_x 값이 할당된다. 피해 지점에 대한 가중치 d_x 는



<Fig. 3> Collision area and Weighted variable d_x settings

Park et al.(2016)의 교통사고에서 운전자 부상 심각성 추정에 관한 연구를 기반으로 선정했다. 해당 연구에서는 2009년부터 2014년까지의 6년간 NASS-CDS⁷⁾ 데이터를 사용하여 충격 방향별 운전자 부상 심각도⁸⁾ISS 15+ 기준)의 분포 비율을 분석하였다. 변수 선정에 대한 구체적인 사항은 <Fig. 3>을 통해 나타났다.

자율주행차의 피해 정도와 피해 범위로 구성된 방정식으로부터 자율주행차의 사고심각도라는 종속변수를 정의했다. 이후에는, 앞서 수집한 데이터로부터, 독립 변수를 선정하고 모델링과 통계적 검증을 수행하고자 한다. 모델링을 위한 초기 변수와 표현 방법은 <Table 7>에 나타났다.

<Table 7> Initial input variable contents and description

Variables	Contents and Format
Severity of damages (y)	Indicates the severity of the traffic accident and is calculated through Equation2
Time of accident ($r1$)	Time of the accident, with AM represented as 1 and PM represented as 0
Lightning ($r2$)	brightness around the time of the accident, with brightness represented as 1 for daylight, 0.5 for dusk-dawn, and 0 for dark
Weather ($r3$)	Weather around the time of the accident, with clear represented as 1, cloudy as 0.5, and rain as 0
Road surface ($r4$)	Road surface condition at the time of the accident, dry represented as 1, wet represented as 0
Number of Crosswalks ($g1$)	The number of intersections within the intersection where the accident occurred
Distance between Intersections ($g2$)	The distance from the intersection where the accident occurred to the next intersection(m)
Number of Bus stops ($g3$)	The number of bus stops within a radius of 30 meters from the accident location
Number of Lanes ($g4$)	Sum of the number of lanes in all directions within the intersection where the accident occurred
Road Width ($g5$)	The width of the road at the accident location(m)
Exclusive Lanes ($g6$)	Presence of a bus or bicycle lane within the intersection at the accident location (yes=1, no=0)
Area of Intersection ($g7$)	The area of the intersection where the accident occurred(m^2)
Left and Right Turn Lanes ($g8$)	Presence of a left or right turn lane within the intersection at the accident location (yes=1, no=0)
VMТ ($t1$)	The vehicle miles traveled (VMТ) on the link where the accident occurred.
LOS ($t2$)	The level of service (LOS) on the link where the accident occurred
Pedestrian Collision ($t3$)	Count of pedestrian collisions that took place within the intersection where the accident occurred
Bike Collision ($t4$)	The count of bike collisions that took place within the intersection where the accident occurred

V. 분 석

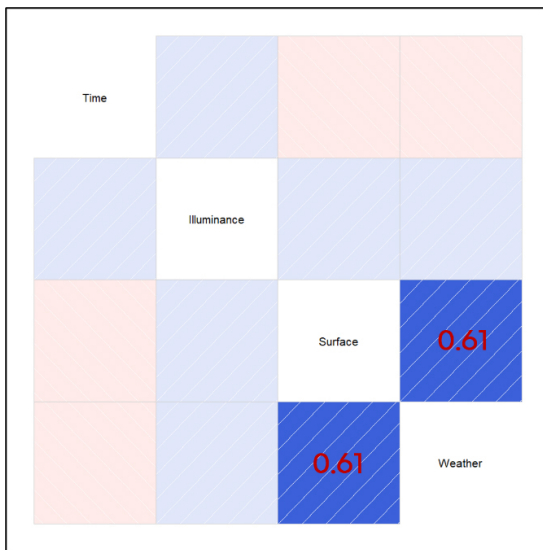
1. 차원 축소 기법의 적용

통계 모형 생성 이전에 과적합 감소와 모형 성능 향상, 모형의 해석 가능성을 높이기 위해 차원 축소 기법

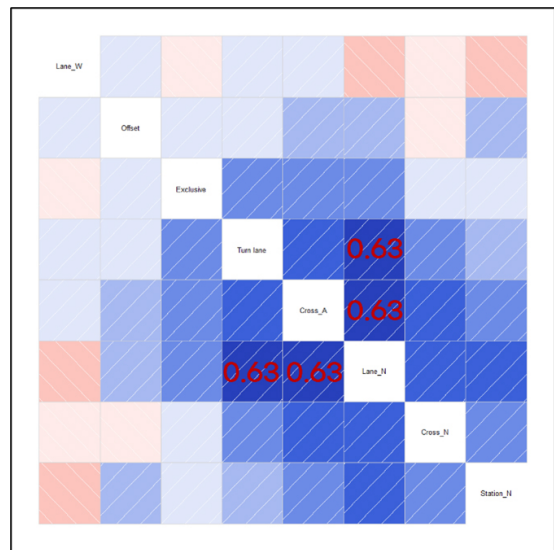
7) NASS-CDS : National Automotive Sampling System-Crash worthiness Data System

8) ISS : Injury Severity Score, 부상 심각도 점수

을 적용하였다. 첫 번째 단계로, 변수 간 상관관계 분석을 통한 요인 축소를 진행하였다. 상관 분석에는 통계 프로그램 R과 psych 패키지를 활용하였다. Psych 패키지는 탐색 및 확정적 요인 분석 수행, 신뢰성 추정, 개별 항목 분석을 제공한다. 해당 패키지에는 기술 통계량, 상관 행렬 및 주성분 분석 기능도 포함되어있으며, 산점도 및 히스토그램을 생성하는 기능도 포함되어있다. 'Psych' 라이브러리를 활용하여 다각도로 수집된 데이터에 대한 상관 분석을 수행하였다. 분석 결과 도로 표면 - 날씨, 회전차로 - 차로 개수, 교차로 면적 - 차로 개수 간 높은 상관관계가 도출되었다. 각 상관 계수의 p-값은 0.05 이하의 유의 수준에서 통계적 유의성을 만족하였다. 상관 계수 값이 0.6 이상일 때, 높은 상관관계를 갖는다고 판단했으며, 시각화 결과는 <Fig. 4>와 <Fig. 5>에 나타내었다. 해당 그림에서 빨간색과 파란색은 각각 양의 상관관계와 음의 상관 관계를 나타내며, 색상의 진하기는 상관관계의 강도를 나타낸다.



<Fig. 4> Correlation analysis (CA DMV Report)

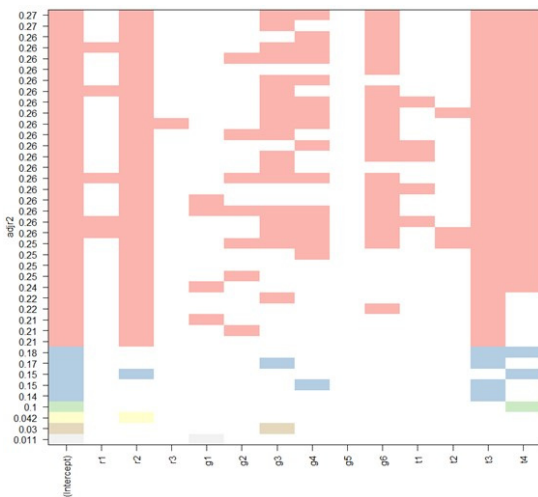


<Fig. 5> Correlation analysis (Spatial Data)

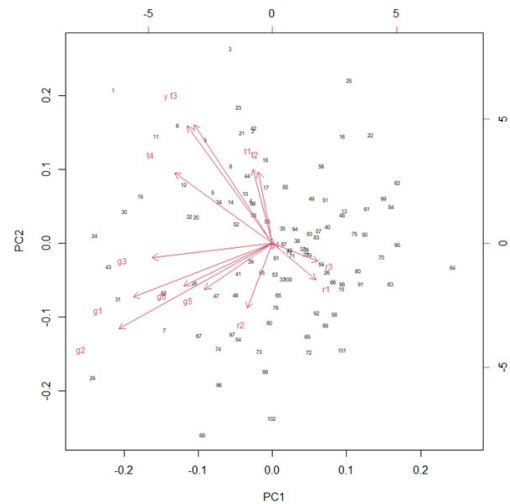
예측 변수들이 모형 내에서 높은 상관관계를 가질 때, 다중공선성 문제가 발생할 수 있다. 이는 부정확한 계수 추정치, 표준 오차의 과대 추정, 계수 부호의 불안정성 등 예측 변수를 식별하는 데 방해 요인으로 작용한다. 이러한 이유로 상관 분석 결과에서 높은 상관관계를 가진 변수들을 제거하고 통합하는 과정을 진행하였다. <Table 7>에 제시된 초기 변수에서 도로 표면(r4), 교차로 면적(g4), 회전차로(g8)를 제외하고 남은 14개의 변수를 사용하여 분석을 계속 진행하였다. 상관관계가 높은 요소들을 기반으로 차원 축소를 진행하고 나머지 요인들을 통한 최적 모델을 개발하기 위해 PCA 및 전부분 회귀분석(All Subset Regression)을 실시하였다. 전부분 회귀분석은 통계 모델링에서 결과 변수를 가장 잘 예측하는 변수 집합을 식별하기 위해 사용하는 변수 선택 기법이다. 전부분 회귀분석을 통해 변수 조합으로 생성이 가능한 모든 모델을 검토해 보았을 때, 모형 결정 계수가 높은 변수 집합에는 "조도 (r2), 버스 정류장 수(g3), 차로 수(g4), 전용차로 존재(g6), 보행자 사고 이력(t3), 자전거 사고 이력(g3) 변수가 포함되었다. 전부분 회귀 분석의 시각화 결과는 <Fig. 6>을 통해 나타내었다.

전부분 회귀분석과 함께 PCA(주성분 분석)를 실시하였다. PCA 행렬에서 화살표의 방향과 각도는 변수 간의 관계를 나타낸다. 화살표가 좁은 각도로 배열되고 서로 가까이 있을 때 화살표로 표시된 두 변수는 정의

상관관계를 가진다. 화살표가 직각으로 교차하면 두 변수는 상관관계가 없으며 화살표가 넓은 각도로 반대 방향으로 퍼져 있을 때, 두 변수는 음의 상관관계를 갖는다. 또한, 화살표가 축에 가까울수록 해당 변수와 구성 요소 사이의 관련성은 높다. 이와 같은, 이론적 사실을 기반으로 PCA 결과를 해석하였고, 제1 주성분과 높은 상관관계를 가지는 변수는 횡단보도 수(g_4), 전용차로 존재(g_6) 등의 공간 데이터였으며, 제2성분과 높은 상관관계를 가지는 변수는 사고시간(r_1), 날씨(r_3) 등의 CA DMV 및 VMT(t_1), 보행자 사고 이력(t_4) 등의 교통 데이터였다. PCA 시각화의 결과는 <Fig. 7>을 통해 제시하였다. 전부분 회귀분석과 PCA 결과를 참고하여 모델링에 사용할 변수를 선택하고 해당 변수 집합을 적절하게 조합하여, 다양한 모델에 대한 다중 회귀 분석을 수행하였다. 이를 통해 최적 모델을 형성하는 변수를 선정하였으며 최종 선정된 변수는 조도(r_2), 차로 수(g_4), 전용차로 존재(g_6), 보행자 사고 이력(t_3) 그리고 자전거 사고 이력(t_4)이며, 선정된 변수들을 사용하여 모델링을 수행하였다.



<Fig. 6> All subset regression analysis result



<Fig. 7> PCA matrix (Biplot)

2. 모델 추정

선정된 변수를 사용한 모델 추정에 앞서, 종속변수의 분포 형태를 관찰하였을 때, 종속변수로 설정한 사고 심각도(y)는 좌 편향된 분포를 보이며 정규성을 만족하지 않으므로 일반 선형 모델 범주에 속하는 포아송 회귀 모델을 채택해 모델 추정을 진행하였다. 선정된 변수들로부터 포아송 회귀분석을 수행하였고 분석 결과 모든 변수가 유의확률 0.05 이하에서 통계적 유의성을 만족하였다. 회귀 모델 자체의 유의성을 검증하기 위해, 절편만을 포함하는 NULL 모델과 생성된 모델의 편차를 카이-제곱 분포를 통해 확인하였다. p -값은 $7.578434e-28$ 로 생성된 모델의 통계적 유의성을 확인하였다. 분석을 통해, 추가 변수가 모델의 적합성을 향상하는데 유하게 기여한다는 결론을 내릴 수 있으며, 생성된 모델이 통계적으로 유의한 것으로 판단할 수 있다. 포아송 회귀분석 결과는 <Table 8>과 같다.

포아송 회귀분석 결과 모델 적합도와 각각의 독립 변수는 모두 통계적 유의성을 만족하였다. 그러나 포아송 회귀 모델에서는 종종 과대산포 문제가 발생하므로 해당 문제를 검증할 필요성이 존재한다. 포아송 분포는 평균과 분산이 동일하다고 가정하며, 분산이 관측값과 예측값의 차이로 인해 평균보다 큰 경우 이를 과대

산포라고 정의한다. 과대산포의 존재 여부 확인은 R의 qcc 라이브러리에 포함된 Overdispersion Test를 사용하였다. 분석 결과, p -값은 0으로 도출되어 과대산포가 존재하지 않는다는 귀무가설을 기각하고 모델에 과대산포 문제가 있음을 확인하였다.

<Table 8> The results of the Poisson regression analysis

Independent Variables	Coefficients	Std. Error	z value	Pr(> z)
Null	2.97011	0.43907	6.765	1.34e-11***
Lightning ($r2$)	-0.11590	0.02378	-4.875	1.09e-6***
Number of lanes ($g4$)	-0.60930	0.22299	-2.732	0.006288***
Exclusive lanes ($g6$)	0.27349	0.07241	3.777	0.000159***
Pedestrian collision ($t3$)	0.76155	0.10096	7.543	4.59e-14***
Bike collision ($t4$)	0.33913	0.09089	3.731	0.000191***
Null deviance	438.19 on 101 degrees of freedom			
Residual deviance	301.14 on 96 degrees of freedom			

- Not relevant; *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.

과대산포가 확인되면 준 포아송(Quasi-Poisson)을 사용할 수 있다. 준 포아송에서는 반응 변수의 분산이 평균에 비례하지만 비례 상수(분산 대개 변수)가 포아송 모델처럼 고정되어 있지 않고 데이터에서 추정된다. 따라서 추정 과정이 보다 유연해지며 포아송 모델의 엄격한 가정을 충족하지 않는 데이터를 활용한 모델링을 진행할 수 있다. 분석 결과, 준 포아송을 사용한 모델은 통계적 유의성을 갖는 것으로 확인되었으며 분석 결과는 <Table 9>와 같다.

<Table 9> The results of the Quasi-Poisson regression analysis

Independent Variables	Coefficients	Std. Error	z value	Pr(> z)
Null	2.97011	0.82722	3.590	0.000522***
Lightning ($r2$)	-0.11590	0.04479	-2.587	0.011168**
Number of lanes ($g4$)	-0.60930	0.42013	-1.450	0.150242
Exclusive lanes ($g6$)	0.27349	0.13642	2.005	0.047800**
Pedestrian collision ($t3$)	0.76155	0.19022	4.004	0.000123***
Bike collision ($t4$)	0.33913	0.17124	1.980	0.050515*
Null deviance	438.19 on 101 degrees of freedom			
Residual deviance	301.14 on 96 degrees of freedom			

- Not relevant; *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.

통계적 유의성 검정 결과를 기반으로, 조도($r2$), 전용차로 존재($g6$), 보행자 사고 이력($t3$), 자전거 사고 이력($t4$)이 교통사고의 심각도에 영향을 미치는 변수로 분석되었다. 정리하자면, 자율주행차 사고의 심각도는 조도가 낮을 때, 자전거나 버스 전용 차로가 존재할 때, 그리고 보행자와 자전거 사고 이력이 많은 지역에서 증가하는 경향이 있다는 것을 알 수 있다.

VI. 결론 및 향후 연구과제

본 연구는 캘리포니아 차량관리국(CA DMV)에서 수집·배포한 자율주행차 사고데이터와 사고 지점의 공간 정보 및 교통 정보를 분석하여 자율주행차 사고의 심각도에 영향을 미치는 요인을 분석하였다. 연구 데이터는 다음의 세 가지 유형으로 구분된다. 주행 환경 데이터는 날씨, 도로 상태, 조명 및 시간 등의 정보를 포함한다. 공간 데이터는 교차로 너비, 차로 수, 차로 폭, 버스 정류장 수, 전용차로 존재 등의 정보를 포함한다. 교통 데이터는 서비스 수준(LOS), 주행 마일 수(VMT), 보행자 및 자전거 사고 이력 등의 정보를 포함한다. 주행 환경 데이터는 자율주행차 사고 레포트에서, 공간 데이터는 Google Earth에서, 교통 데이터는 샌프란시스코 교통국에서 수집했다. 초기 선정된 독립 변수는 16개였으나, 과적합을 방지하고 모델 해석 가능성을 높이기 위해 차원 축소 기법을 적용하였다. 차원 축소의 첫 단계는 높은 상관성을 가진 요소를 제거하거나 병합하기 위해 상관 분석을 진행하였다. 상관 분석 결과, 주행 환경 및 공간 데이터에서 높은 상관성을 가진 변수들이 확인되었으며 적절한 조치를 통해 관련 변수를 제외하였다. 두 번째 단계에서는 전부분 회귀분석을 적용하였다. 분석 결과, 조도(r_2), 버스 정류장 수(g_3), 차로 수(g_4), 전용차로의 존재(g_6), 보행자 사고 이력(t_3) 및 자전거 사고 이력(t_4)을 포함하는 모델이 일반적으로 높은 모델 적합성을 보이는 것으로 확인하였으며, 해당 변수를 중점적으로 고려하여 다음 단계를 진행하였다. 마지막으로 차원 축소를 통한 변수 선택 과정을 마무리하기 위해 주성분 분석(PCA)을 수행하였다. PCA 결과, 공간 데이터는 주성분 1에 주로 분포하고 있었으며 주행 환경 및 교통 데이터는 주성분 2에 주로 분포되었다. 주성분 분석 행렬에서 각 구성 요소 간의 상관관계의 정도와 크기를 판단할 수 있었다. 최종적으로, 조도(r_2), 차로 수(g_4), 전용차로의 존재(g_6), 보행자 사고 이력(t_3), 자전거 사고 이력(t_4)을 최종 변수로 채택하였다.

모델 추정에 앞서 종속변수의 분포를 확인하였을 때 좌 편향된 포아송 분포를 띠고 있음을 확인하여 포아송 회귀분석을 진행하였다. 포아송 회귀분석 결과 모든 변수와 모델 자체적으로 통계적 유의성을 검증하였지만, 포아송 모델의 과대산포 문제를 확인하기 위한 추가적인 검증 절차를 진행하였다. 과대산포 문제의 검증 결과 생성된 모형에 과대산포 문제가 있음이 확인되었고 과대산포 문제를 제어할 수 있는 준 포아송(Quasi-Poisson)을 사용하여 재분석하였다. 준 포아송을 사용한 회귀 분석 결과, 차로 수(g_4)를 제외한 모든 변수와 모델 자체의 통계적으로 유의성이 검증되었다. 보행자 사고 이력(t_3)이 자율주행차 사고심각도에 가장 큰 영향을 끼치는 것으로 확인되었으며, 자율주행차 사고심각도는 조도가 낮을 때, 자전거 또는 버스 전용차로가 존재할 때, 보행자와 자전거 사고의 이력(건수)이 높은 지역일수록 증가하는 것으로 나타났다.

선행연구의 결과들로부터 본 연구의 통계 분석 결과를 해석해 보았을 때, 조도가 낮을 때의 사고심각도가 증가하는 것은 자율주행차의 센서 성능 저하와 관련이 있다고 판단된다. 자전거 또는 버스 전용차로가 존재할 때 사고심각도가 증가하는 것은, 자율주행차의 주행 프로그램 혹은 알고리즘상에 전용차로라는 개념이 미탑재되었거나 인식이 불명확함에 따라 기존의 교통환경과는 다른 이질적인 정보 수용과 관련 있다고 판단된다. 마지막으로, 보행자와 자전거 사고 이력의 높은 지역의 경우 사고심각도가 증가하는 것에 대해서는 주목해 볼 필요가 있다. 사고 이력이 높은 도로는 기하 구조 및 도로 시설물 등의 설계 및 배치 불량이 원인이 되었을 가능성이 크며, 이러한 형태의 교통환경은 자율주행 시스템의 주행에도 영향을 미칠 수 있음을 간접적으로 설명하기 때문이다.

본 연구 결과는 자율주행차 사고심각도에 영향을 미치는 요인이 차량의 성능 및 기능적 측면 이외에도 환경 요인, 공간 요인, 교통 요인을 포함할 수 있음을 시사한다. 본 연구의 한계점으로는 Google Earth API를 사용하여 수집된 공간 데이터는 이미지 왜곡 및 프로그램 오류로 인한 실제 공간 정보와 차이가 존재할 것으로 예상되는 점과 수집 가능한 자율주행차 사고데이터가 제한됨에 따라 데이터의 분량과 다양성이 제한적

이라는 점이다. 향후 연구에서는 실제 측정 데이터를 사용하는 등 데이터 완결성을 높이고, 보다 다양한 변수 구성으로 모형 추정 작업을 수행하며, 전이 학습을 통하여 국내 실정에 맞는 모형 개발을 목표로한다.

ACKNOWLEDGEMENTS

본 연구는 2023년 경기대학교 대학원 연구원장학생 장학금 지원에 의하여 수행되었음
이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-01352, 자율주행 관련 법규 및 규제 대응 서비스 시나리오 실효성 검증 기술 개발)

REFERENCES

- Chen, H., Chen, H., Liu, Z., Sun, X. and Zhou, R.(2020), “Analysis of Factors Affecting the Severity of Automated Vehicle Crashes Using XGBoost Model Combining POI Data”, *Journal of Advanced Transportation*, pp.1-12.
- Coxe, S., West, S. G. and Aiken, L. S.(2009), “The analysis of count data: A gentle introduction to Poisson regression and its alternatives”, *Journal of Personality Assessment*, vol. 91, no. 2, pp.121-136.
- Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M. and Varadaraju, N.(2017), “Examining accident reports involving autonomous vehicles in California”, *PLoS One*, vol. 12, no. 9, e0184952.
- Frome, E. L.(1983), “The Analysis of Rates Using Poisson Regression Models”, *Biometrics*, vol. 39, no. 3, pp.665-674.
- Ha, T., Kim, S., Seo, D. and Lee, S.(2020), “Effects of explanation types and perceived risk on trust in autonomous vehicles”, *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 73, pp.271-280.
- Hayat, M. J. and Higgins, M.(2014), “Understanding Poisson regression”, *The Journal of Nursing Education*, vol. 53, no. 4, pp.207-215.
- Jeon, Y., Ko, J. and Yeom, C.(2020), “Impact analysis of road safety diagnosis items on traffic accident density-Focused on the autonomous districts in Seoul”, *Journal of Seoul Studies*, vol. 21, no. 2, pp.67-84.
- Jessalyn, K. H., Amy, Y. X. Y., Moria, K. K. and Peter, C. A.(2021), “Comparing regression modeling strategies for predicting hometime”, *BMC Medical Research Methodology*, vol. 21, no. 138.
- Kim, K. and Cho, S. A.(2020), “Lessens Learned from Crash Types of Automated Vehicles:Based on Accident Data of Automated Vehicles in California, USA”, *Journal of Korean Society of Civil Engineers*, vol. 17, no.2, pp.34-42.
- Lee, H. S., Kang, J. H., Kim, W. J., Kang, Y. J., Kang, K. W., Park, H. S., Song, S. W., Park, C. B., Boo, J. H., Lee, S. K. and Ko, S. Y.(2022), “Are rental car accidents more deadly in Jeju?”, *Journal of the Korean Society of Emergency Medicine*, vol. 33, no. 4, pp.406-415.
- Ma, Y. and Zhu, L.(2013), “A Review on Dimension Reduction”, *International Statistical Review*, vol.

- 81, no. 1, pp.134-150.
- National Safety Council(1983), *Vehicle Damage Scale for Traffic Accident Investigators*, Highway Safety Research Center, Chapel Hill, North Carolina 27514.
- Oh, J., Sung, N. and Ha, O.(2005), "Development of accident prediction model for improving safety at signalized intersections on national highways", *Journal of Korean Society of Civil Engineers*, vol. 25, no. 1D, pp.9-15.
- Park, B. H., Yoo, D. S., Yang, J. M. and Lee, Y. M.(2008), "Analysis of accident characteristics and model development for rural signalized intersections (focused on Cheongju and Cheongwon)", *Journal of Korean Society of Transportation*, vol. 26, no. 2, pp.35-46.
- Park, B.(2014), "Prediction of traffic accidents and traffic volume using simultaneous equations", *Journal of Construction Engineering and Technology*, vol. 33, no. 1, pp.1-6.
- Park, S., Yoo, H., Jeong, S. and Chung, K.(2016), "Estimation of Driver Injury Severity in Traffic Collisions", *Journal of Transport Research*, vol. 24, no. 1, pp.1-15.
- Ringnér, M.(2008), "What is principal component analysis?", *Nature Biotechnology*, vol. 26, pp.303-304.
- Seong, B. J., Bae, K. H. and Yoo, H. H.(2015), "Analysis of spatiotemporal distribution characteristics of traffic accidents in Jinju", *Journal of Korean Society for Geospatial Information System*, vol. 23, no. 2, pp.3-9.
- Shim, K. and Hwang, K.(2009), "Development of Computation Model for Traffic Accidents Risk Index-Focusing on Intersection in Chuncheon City", *International Journal of Highway Engineering*, vol. 11, no. 3, pp.61-74.
- Shlens, J.(2014), *A tutorial on principal component analysis*, arXiv preprint arXiv:1404.1100.
- Ver, H. J., M. and Boveng, P. L.(2007), "Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?", *Ecology*, vol. 88, no. 11, pp.2766-2772.
- Xu, C., Ding, Z., Wang, C. and Li, Z.(2019), "Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes", *Journal of Safety Research*, vol. 71, pp.41-47.
- Yong, A. G. and Pearce, S. C.(2013), "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis", *Tutorials in Quantitative Methods for Psychology*, vol. 9, no. 2, pp.79-94.