

자율주행차 실사고 데이터를 활용한 연관규칙 분석 기반 자율주행차 도심부 사고 시나리오 방법론 개발

이호준¹ · 강민희² · 송재인³ · 황기연^{4*}

¹주식회사 스페이스인사이트 연구개발팀 선임연구원, ²홍익대학교 스마트도시전공 박사수료, ³홍익대학교 도시공학과 초빙교수, ⁴홍익대학교 도시공학과 교수

Development of Autonomous Vehicle Traffic Accident Scenarios in Urban Areas Based on Real-World Accident Data Using Association Rule Mining

LEE, Hojun¹ · KANG, Minhee² · SONG, Jaein³ · HWANG, Keeyeon^{4*}

¹Junior Researcher, R&D team, Spacelsight Co., Ltd., Seoul 07788, Korea

²Ph.D. Candidate, Industrial Convergence Interdepartmental Program, Smartcity, Hongik University, Seoul 04066, Korea

³Visiting Professor, Department of Urban Design & Planning, Hongik University, Seoul 04066, Korea

⁴Professor, Department of Urban Design & Planning, Hongik University, Seoul 04066, Korea

*Corresponding author: hwangkeeyeon@gmail.com

Abstract

Although traffic accidents are expected to decrease with the development of autonomous vehicles, safety and reliability issues are emerging as traffic accidents with autonomous vehicles are constantly occurring. To address this issue, an autonomous vehicle safety evaluation scenario is utilized that is directly linked to vehicle safety. Therefore, ensuring the reliability and specificity of the scenario is crucial. Accordingly, we proposed a methodology for creating a reliable and specific autonomous vehicle accident scenario in urban areas to evaluate the safety of autonomous vehicles. To ensure the reliability of the scenario, we utilized the DMV collision report, which contains autonomous vehicle accident data. Additionally, to improve specificity, we employed the Apriori algorithm of association rule analysis, a data mining technique. After analyzing the association rules using the DMV collision report, we identified approximately 300,000 association rules and derived a combination of various accident factors that were highly related to the type of vehicle collision. To provide specific scenarios for evaluating the safety of autonomous vehicles, we analyzed the PEGASUS project and the CETRAN project and finally selected nine scenario factors. Based on the association rules that included all of the scenario factors, we presented six autonomous vehicle traffic accident scenarios in the urban areas. These scenarios are expected to be utilized as a standardized benchmark for evaluating the safety of autonomous vehicles.

Keywords: apriori algorithm, association rule mining, autonomous vehicle, dmv collision report, traffic accident scenario

J. Korean Soc. Transp.
Vol.41, No.3, pp.375-393, June 2023
<https://doi.org/10.7470/jkst.2023.41.3.375>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 21 April 2023

Revised: 8 May 2023

Accepted: 31 May 2023

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

초록

자율주행차 개발에 따라 교통사고 감소 등이 기대되고 있으나, 자율주행차 사고가 끊임없이 발생하고 있어 안전성 및 신뢰성 문제가 대두되고 있다. 이를 해소하기 위해 자율주행차 안전성 평가 시나리오가 활용되고 있으며, 이는 차량의 안전과 직결되어 있어 시나리오의 신뢰성 및 구체성이 필수적으로 요구된다. 이에 본 연구는 자율주행차 안전성 평가를 위해 신뢰성과 구체성을 확보한 자율주행차의 도심부 사고 시나리오 생성 방법론을 제안하였다. 시나리오의 신뢰성을 확보하기 위해 자율주행차 실사고 데이터인 DMV collision report를 활용하였으며, 구체성을 확보하기 위해 데이터 마이닝 기법인 연관규칙 분석의 Apriori 알고리즘을 활용하였다. DMV collision report를 활용하여 연관규칙을 분석한 결과, 약 300,000개의 연관규칙이 확인되었으며, 차량 충돌 형태와 연관성이 높은 다양한 사고 요인의 조합을 도출하였다. 구체적인 시나리오를 제시하기 위해 PEGASUS 프로젝트와 CETRAN 프로젝트를 분석하여 최종적으로 9개의 시나리오 요인을 선별하였으며, 해당 시나리오 요인을 모두 포함한 연관규칙을 기반으로 6개의 자율주행차 도심부 사고 시나리오를 제시하였다. 본 연구에서 제시한 6개의 시나리오는 자율주행차의 안전성 평가를 위한 시나리오로 활용 가능할 것으로 기대된다.

주요어: Apriori 알고리즘, 연관규칙 분석, 자율주행차, DMV 사고 보고서, 사고 시나리오

서론

최근 자율주행자동차(Autonomous Vehicle, 이하 자율주행차)를 위한 법적, 윤리적, 사회적 측면 등 제도적인 고려와 함께 기술의 진보를 통해 자율주행차 3단계 상용화가 가시화되고 있다. SAE(Society of Automotive Engineers)는 기술 수준에 따라 자율주행차 단계를 6단계(레벨 0-레벨 5)로 구분하였으며, 단계가 올라갈수록 인간의 개입이 최소화되고 차량에 탑재된 자율주행시스템(Autonomous Driving System, ADS)이 스스로 다양한 상황을 대처해야 한다. 구글, 현대 등 여러 기업에서 자율주행차가 직면하는 다양한 상황을 스스로 대처할 수 있는 고도화된 자율주행시스템을 개발하고 있으며, 국내에서도 상암, 판교 등 시범운행지역을 선정하여 4단계 자율주행차의 시범 운행을 수행하고 있다. 그러나 자율주행차 사고는 끊임없이 발생하고 있으며, 이상적인 상황에서도 사고가 발생할 수 있다는 의견이 제시됨에 따라(Goodall, 2014; Kang et al., 2020; Kang et al., 2022a) 자율주행차 안전성에 대한 이슈가 제기되고 있다. 이처럼 단순히 기술의 진보만으로 자율주행차의 상용화를 보장할 수 없으므로 안전성을 확보할 수 있는 체계가 요구되는 실정이다.

자동차의 안전성을 확보하는 방법으로 자동차 안전도 평가(New Car Assessment System, NCAP)가 존재하며, 이는 충격 실험을 통해 차량의 종합적인 안전성을 평가하여 소비자에게 자동차의 안전도에 관한 정보를 제공하는 제도이다. 도로 위를 주행하는 모든 차량은 안전성을 보장할 수 있는 평가제도가 요구되므로 자율주행차 또한 NCAP 내 포함되어 평가가 진행되고 있다. 자율주행차 안전성 평가시스템은 자율주행시스템에 대한 평가로 전방 충돌 방지 보조 시스템(Forward Collision-Avoidance Assist, FCA), 차로 유지 보조 시스템(Lane Following Assist, LFA) 등을 활용하고 있다. 그러나, 자율주행차는 일반 차량과 달리 복잡한 환경에서 특정 상황을 대응할 수 있어야 하므로 다양한 환경에서의 평가가 요구되어 시나리오 기반 접근법이 제시되고 있다(ISO 26262, 2018; PEGASUS Project Consortium, 2019; de Gelder et al., 2020; Riedmaier et al., 2020; Steimle et al., 2021). 시나리오 접근법은 실제 도로 운행 및 시뮬레이션 등을 통해 다양한 환경을 고려하여 조절할 수 있으며, 정교하게 설계된 모든 시나리오를 다룰 수 있기 때문에 다양한 환경에서의 종합적 평가를 가능하게 한다는 장점이 존재한다. 이러한 자율주행차 평가를 위한 시나리오 접근법은 지식 기반 접근법(Knowledge-based extraction)과 데이터 기반 접근법(Data-driven extraction)으로 구분된다(Riedmaier et al., 2020). 지식 기반 접근법의 경우, 온톨로지 방식을 활용하여 전문가 지식을 통해 차선변경, 우회전, 좌회전 등 다양한 상황 대한 목록을 사전에 분류하여 시나리오를 빠르게 생성할 수 있으나 시나리오 자체에 대한 대표성의 문제가 제기될 수 있다. 반면, 데이터 기반 접근법은 실제 데이

터와 머신러닝 및 딥러닝 등을 활용하여 시나리오를 생성하는 방식으로 시나리오 대표성의 문제를 극복할 수 있으나, 데이터 정제 및 인공지능 등의 요건을 충족하지 못할 경우 시나리오를 생성할 수 없다는 전제조건이 존재한다. 이러한 조건이 충족될 경우 데이터 및 방법론에 따라 시나리오의 신뢰성 및 구체성을 확보할 수 있어 최근 데이터 기반 접근법 연구가 다양하게 수행되고 있다. 또한, 자율주행차는 고속도로 및 도심도로 환경에서 ODD(Operational Design Domain)가 상이하기 때문에 고속도로와 도심도로를 나누어 시나리오를 제시할 필요가 있다. 특히 도심 도로의 경우, 일반 차량 뿐만 아니라 복잡한 인프라 및 보행자가 혼재된 공간으로 여러 객체 간의 상호작용이 발생함에 따라 자율주행차 사고의 발생 빈도와 인명피해 확률이 높음에 따라 시나리오의 필요성이 제기되고 있다(Son et al., 2021).

이에 본 연구는 자율주행차와 일반 차량이 혼재하는 상황에서 발생하는 사고상황을 대상으로 자율주행차의 도심 부 시나리오를 도출하는 방법론을 개발하고, 이를 기반으로 신뢰성과 구체성을 확보한 자율주행차의 사고 시나리오를 제시하고자 한다.

선행연구

1. 교통사고 분석 연구

교통사고는 매년 많은 사망자와 부상자를 낳으며, 이로 인해 사회적 비용도 상당한 규모로 발생하고 있다. 이에 교통사고 예방은 매우 중요한 과제임에 따라 교통사고의 원인을 분석하는 연구가 수행되고 있다. 이는 일반 차량의 사고와 자율주행차의 사고를 대상으로 수행되고 있으며, 다양한 사고데이터를 활용하여 수행되고 있다. 일반 차량을 대상으로 교통사고의 부상 및 사고 심각도를 예측하고, 심각도와 연관된 사고 요인을 파악하여 사고의 패턴을 도출하는 연구가 수행되고 있으며(Shanthi and Ramani., 2012; Taamneh et al., 2017; Bahiru et al., 2018), 충돌 방식, 연령대, 약물 및 음주 유무, 성별 등이 심각도에 유의미한 영향을 미치는 요소임을 확인하였다. 또한, 데이터마이닝 기법을 활용하여 교통사고의 원인을 파악하는 연구가 수행되고 있으며, 사고 데이터와 방법론에 따라 날씨, 조명 상태, 신호위반 유무 등이 교통사고의 원인으로 도출되었다(Park et al., 2006; Li et al., 2017; Muhammad et al., 2017; Bhin and Son, 2018; Bhin et al., 2020; Kim et al., 2020; DrissiTouzani et al., 2020). 특히, Kumar and Toshniwal(2016)은 사고 발생 위치의 공간적 분석을 통해 사고의 특징과 원인을 분석하였다. 저자는 k-means 알고리즘을 활용하였으며, 사고 빈도를 기준으로 사고의 군집화를 진행하고 군집을 기준으로 연관규칙 분석을 수행하였다. 분석 결과, 고속도로 교차로와 보행자 충돌사고가 연관성이 높은 것을 확인하였으며, 사고가 다수 발생하는 군집은 대부분 구릉지에서 발생한 이륜차 사고로 구성되어 있는 것을 확인하였다. 자율주행차의 경우, 일반 차량에 비해 사고가 다수 발생하지 않으며, 사고데이터는 제작사의 기밀사항으로 접근이 제한되어 있다. 이에 대부분의 연구는 캘리포니아의 DMV collision report(이하 DMV 사고 보고서)를 활용하여 자율주행차의 사고를 분석하는 연구를 수행하고 있다(Leilabadi et al., 2019; Das et al., 2020; Torres et al., 2021; Ashraf et al., 2021; Kang et al., 2022c). DMV 사고 보고서는 자율주행차의 사고를 기록한 보고서로 사고 당시의 날씨, 위치, 차량의 종류 및 움직임 등 다양한 데이터를 제공한다(<https://www.dmv.ca.gov>). DMV 사고 보고서를 활용한 교통사고 분석을 통해 대부분의 사고는 자율주행차가 정지하고 있는 상황에서 후방 추돌형태인 것을 확인하였으며(Ashraf et al., 2021; Torres et al., 2021), 좌회전, 우회전 상황에서 측면 충돌사고가 빈번하게 발생하는 것을 확인하였다(Das et al., 2020; Ashraf et al., 2021). 특히, Leilabadi et al.(2019)은 DMV 사고 보고서의 심층 분석을 통해 사고의 심각도와 도로 표면 상태의 상관관계가 있는 것을 확인하였으며, 다양한 사고 요인 중 도로 표면의 변화와 기상조건을 자율주행차의 사고와 밀접한 연관이 있는 사고 요인으로 선정하였다.

2. 시나리오 도출 연구

교통사고 예방을 위해 다양한 노력이 이루어지고 있으며, 보다 적극적으로 사고를 사전에 방지할 수 있는 시나리오

오 접근법이 제시됨에 따라 시나리오를 도출하는 연구가 수행되고 있다. 사고 시나리오는 일반 차량의 사고데이터를 기반으로 인공지능(Yuan et al., 2020; Yu et al., 2020; Kang et al., 2022b)과 데이터 마이닝 기법(Lenard et al., 2014; Nitsche et al., 2017; Sui et al., 2019)을 활용하여 작성되고 있다. Kang et al.(2022b)은 일반 차량의 사고데이터와 vision transformer를 활용하여 사고 취약상황을 분류하고 사고 취약상황 시나리오를 제시하였다. 위 연구는 TTC(Time-To-Collision)를 기준으로 사고 이전의 상황을 4가지(Normal, Pre-collision, Collision, Post-collision)로 분류하였으며, vision transformer의 학습을 통해 94%의 분류정확도를 도출하였다. 또한, Pegasus project에서 제시한 시나리오 템플릿을 기반으로 구체적인 사고 취약상황의 상황 시나리오를 제시하였다. 데이터 마이닝 기법을 활용하는 연구는 대부분 클러스터링과 연관규칙 분석을 진행하였으며, 이를 통해 보행자 사고(Lenard et al., 2014), 교차로 충돌 시나리오(Nitsche et al., 2017), 자동차-이륜차 시나리오(Sui et al., 2019) 등 다양한 상황의 시나리오가 도출되었다. 이처럼 일반 차량의 사고를 예방하기 위해 단순 사고 분석이 아닌 시나리오를 도출하는 연구가 수행되고 있다. 나아가 자율주행차 사고의 발생 가능성 또한 고려되고 있어 이를 사전에 예방하기 위해 자율주행차의 시나리오를 도출하는 연구가 수행되고 있다. 자율주행차 시나리오는 DMV 사고 보고서를 기반으로 통계 분석(Liu et al., 2021; Favaro et al., 2017), 데이터 마이닝(Song et al., 2021), 토픽 모델링(Alambeigi et al., 2020) 기법 등을 활용하여 도출되고 있다. 특히, Liu et al.(2021)은 통계 분석을 통해 일반 차량과 자율주행차의 시나리오가 상이한 것을 확인하였다. 이에 저자는 일반 차량 데이터를 활용한 시나리오를 자율주행차에 그대로 적용하는 것에 한계가 존재하는 것을 지적하였으며, 자율주행차의 시나리오는 자율주행차의 데이터를 활용해야 하는 것을 강조하였다. 또한, Favaro et al.(2017)은 DMV 사고 보고서와 DMV 제어권 전환 보고서의 통계 분석을 통해 자율주행차 사고의 특징을 추출하였으며, 빈번하게 발생하는 사고 시나리오를 도출하였다. 위 연구는 동적 객체의 움직임에 따라 26개의 사고 시나리오를 제시하였으며, 통계 분석을 통해 대부분 자율주행차 사고는 교차로 사고 및 자율주행차가 전방에 위치한 후방 추돌형태임을 확인하였다. 또한, 제어권 전환 보고서 분석을 통해 후방 추돌시, 제어권 전환을 위해 인지 및 대응하는 과정에서 인간과 유사한 결함이 존재한다는 것을 주장하였다.

3. 차별성

선행연구 고찰 결과, 일반 차량과 자율주행차의 교통사고를 예방하기 위해 사고의 원인을 분석하는 연구와 사고를 선제적으로 예방하기 위해 시나리오를 도출하는 연구가 수행되고 있다. 대부분 상관분석 등의 통계적 기법을 활용하였으며, 연관규칙 분석, 클러스터링 등 데이터 마이닝 기법을 통해 사고의 원인 분석 뿐만 아니라 다양한 상황의 시나리오를 생성하는 것을 확인하였다. 특히, 선행연구를 통해 일반 차량과 자율주행차 시나리오의 특성이 상이하였으며, 자율주행차 시나리오의 신뢰성을 확보하기 위해 자율주행차 데이터를 필수적으로 활용해야 하는 것을 확인하였다. 또한, 자율주행차 데이터는 제작사의 기밀 사항으로 접근이 제한되어 있어, 대다수의 연구는 캘리포니아의 DMV 사고 보고서를 활용하였다. 그러나, 선행연구는 보고서 분석을 통해 시나리오를 제시하였으나 실질적인 평가를 위해 사용할 수 있는 시나리오의 형태를 갖추지 못해 구체성이 부족한 한계를 지닌다. 이에 본 연구는 자율주행차 데이터인 DMV 사고 보고서를 활용한 연관규칙 분석을 통해 사고 요인 간의 연관성을 분석하고, 추출된 연관규칙을 기반으로 사고 시나리오의 요인을 선정하여 자율주행차의 도심부 사고 시나리오를 도출하고자 하며, 이를 통해 시나리오의 신뢰성과 구체성을 확보한 것에 차별성이 존재한다.

자율주행차 도심부 사고 시나리오 생성 방법론

자율주행차의 사고 시나리오 생성을 위한 전체적인 프레임워크는 다음과 같다(Figure 1).

1. 자율주행차 실사고 데이터를 수집하고 전처리한다(Step 1).
2. 연관규칙 분석을 통해 사고 요인의 관계를 파악하고, 시나리오 요소를 선정한다(Step 2 & Step 3).
3. 연관규칙을 조합하여 자율주행차의 사고 시나리오를 제시한다(Step 4 & Step 5).

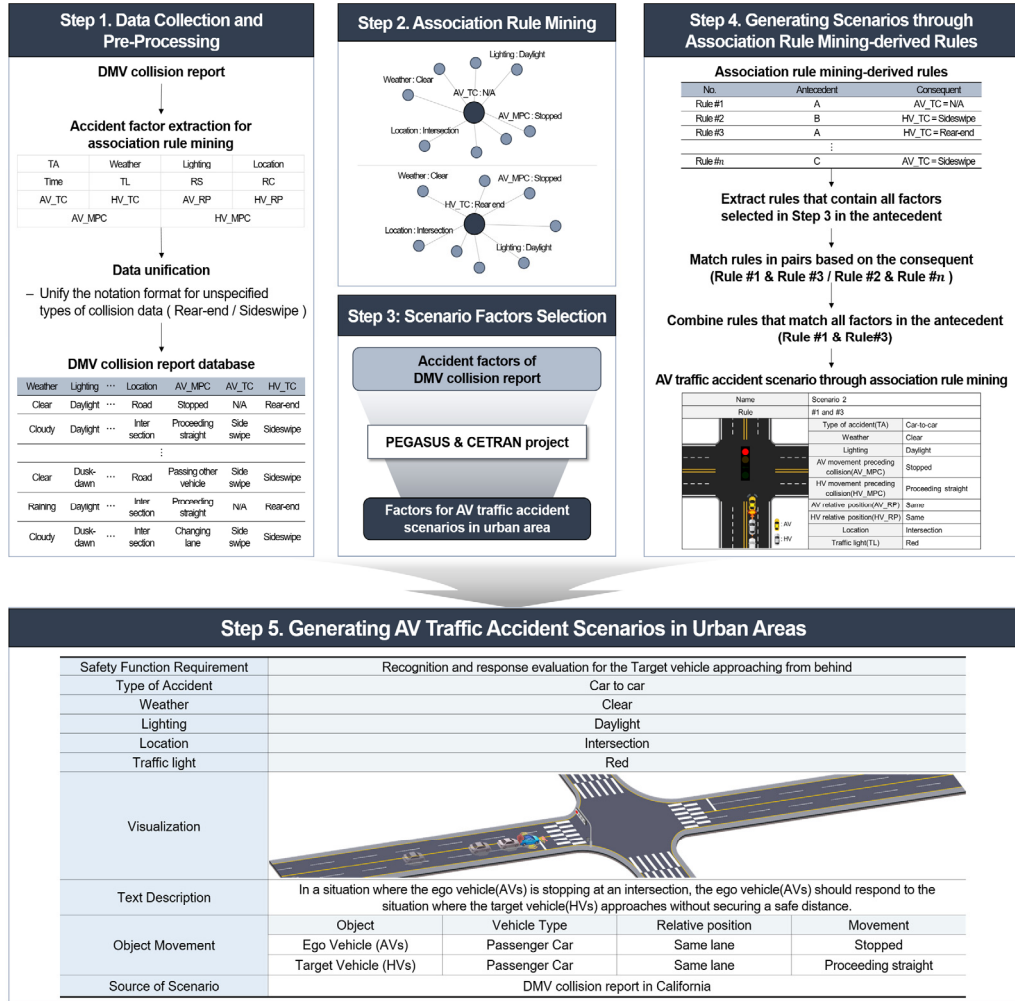


Figure 1. Process of deriving autonomous vehicles accident scenario using association rule mining

1. 자율주행차 사고 데이터 수집 및 전처리

본 연구는 미국 캘리포니아 DMV에서 제공하는 사고 보고서를 활용하여 자율주행차의 사고 데이터베이스를 구축하였다. GM, Waymo, Zoox 등의 여러 제작사는 자율주행차 학습 데이터 구축, 자율주행 기술의 고도화 등을 위해 캘리포니아에서 실도로를 주행하고 있다. 이때 사고가 발생할 경우, 정보를 기록하여 캘리포니아 DMV에 보고서를 제출해야 한다. 해당 보고서는 제작사, 주행모드와 더불어 도로 조건, 사고 발생 시간 뿐만 아니라 사고와 연관된 차량(Vehicle 1(Autonomous Vehicle, 자율주행차, 이하 AV), Vehicle 2(Human-driven Vehicle, 일반 차량, 이하 HV))의 충돌 전 움직임, 상대적 위치, 충돌 형태 등의 정보를 개별적으로 제공하고 있다. 모든 DMV 보고서를 검토한 결과, 2018년 이후에 보고서의 형태가 개정되었으며 개정 이전과 대비하여 다양한 정보를 제공하는 것을 확인하였다. 또한, 보고서에서는 수동 운전 모드(Conventional mode)와 자율주행 모드(Autonomous mode)에 대한 정보를 제공하였다. 이에 따라 자율주행차 사고 시나리오를 작성하는 본 연구의 목적에 적합하다고 판단되는 자율주행 모드에서 발생한 개정 이후의 사고 보고서 165건을 활용하였다.

165건의 데이터를 검토한 결과, 보고서를 작성하는 가이드라인이 존재하지 않아 제조사별로 충돌 형태를 작성하는 방식이 상이한 것을 확인하였다. 충돌 형태는 AV와 HV가 개별적으로 작성되어 있었으며, 후방 추돌과 측면 충돌의 경우 제조사별로 작성방식이 상이하였다. 이처럼 동일한 충돌 형태임에도 작성방식이 상이할 경우, 분석으로

도출된 결과의 해석이 불가능하다. 이에 후방 추돌은 (N/A, Rear-end)로 작성하였으며, 측면 충돌은 (Sideswipe, Sideswipe)로 작성하는 데이터 통일화를 진행하였다(Table 1).

Table 1. Result of data unification

Type of Collision	Before (AV, HV)	After (AV, HV)
Rear-end	(Rear-end, Rear-end) / (Rear-end, Head on) (Rear-end, N/A) / (N/A, Rear-end)	▶ (N/A, Rear-end)
Sideswipe	(Sideswipe, Sideswipe) / (Sideswipe, N/A) / (N/A, Sideswipe)	▶ (Sideswipe, Sideswipe)

이후, 사고 요인을 선별하였다. DMV 사고 보고서의 경우, Figure 2의 형태로 구성되어있으며, 추가적인 상세 사항의 경우 Figure 3과 같이 사고 경위가 서술되어있다. 이에 따라 본 연구에서는 고정적으로 제공되는 요소(Figure 2)를 우선적으로 활용하고 사고 경위에 표시된 요소(Figure 3)를 활용하여 총 24개의 사고 요인을 확인하였으며, 연구 목적에 맞도록 사고 요인을 선별하였다. 사고 요소 선별 과정에서 제조사 정보와 사고 후 정보(AV 파손 정도, HV 파손 정도, 인간 상해 정보 등)를 제거하였으며, 사고가 발생한 위치의 세부 주소 정보(Address/City/County/State)가 사고별로 상이하여 해당 위치의 도로 형태(Location)로 대체하였다. 또한, 사고 당시 차량의 속성을 파악할 수 있는 중요 요인인 차량 속도의 경우, 보고서에 따라 사고 경위 서술란(Figure 3)에 제공되고 있으나 80% 이상의 보고서에서 제공되고 있지 않아 이를 제거하였다.

WEATHER (MARK 1 to 2 ITEMS)	VEH 1	VEH 2	MOVEMENT PRECEDING COLLISION	VEH 1	VEH 2	OTHER ASSOCIATED FACTOR(S) (MARK ALL APPLICABLE)
A. CLEAR	X	X	A. STOPPED	X		A. CVC SECTIONS VIOLATED CITED <input type="checkbox"/> YES <input type="checkbox"/> NO
B. CLOUDY			B. PROCEEDING STRAIGHT		X	
C. RAINING			C. RAN OFF ROAD			
D. SNOWING			D. MAKING RIGHT TURN			
E. FOG/VISIBILITY			E. MAKING LEFT TURN			
F. OTHER			F. MAKING U TURN			
G. WIND			G. BACKING			B. VISION OBSCUREMENT <input type="checkbox"/>
LIGHTING			H. SLOWING/STOPPING			C. INATTENTION* <input type="checkbox"/>
A. DAYLIGHT	X	X	I. PASSING OTHER VEHICLE			D. STOP & GO TRAFFIC <input type="checkbox"/>
						E. ENTERING/LEAVING RAMP <input type="checkbox"/>

Figure 2. Default accident factors of the DMV collision report

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

Autonomous Mode Conventional Mode

A Waymo Autonomous Vehicle ("Waymo AV") was stopped in autonomous mode in the slip lane from westbound Nita Avenue to northbound San Antonio Road in Palo Alto when it was rear-ended. The Waymo AV was yielding to approaching traffic when a commercial box truck made contact with the Waymo AV's rear bumper at approximately 3 MPH. The Waymo AV sustained minor damage to its rear bumper, and the commercial box truck left without exchanging information. Waymo contacted the Palo Alto Police Department and a police report was filed. No injuries were reported at the scene.

Figure 3. Optional accident factors of the DMV collision report

이와 같은 선별 과정을 통해 연관규칙 분석에 활용할 14개의 사고 요인을 선별하였다(Table 2). 14개의 사고 요인은 사고 형태(Type of accident, TA), 날씨(Weather), 조명(Lighting), AV 충돌 전 움직임(AV movement preceding collision, AV_MPC), HV 충돌 전 움직임(HV movement preceding collision, HV_MPC), AV 상대적 위치(AV relative position, AV_RP), HV 상대적 위치(HV relative position, HV_RP), 사고 발생 위치(Location), 도로 표면 상태(Road surface, RS), 도로 상태(Road condition, RC), 사고 발생 시간(Time), 신호등 상태(Traffic light, TL), AV 충돌 형태(AV collision type, AV_TC), HV 충돌 형태(HV collision type, HV_TC)과 같다. 이는 맑음(Clear), 정지(Stopped), 차선 변경(Changing lane), 마름(Dry), 빨간불(Red light), 측면 충돌(Sideswipe) 등의 세부 속성으로 구성되어 있으며, 이를 통해 사고 당시의 상황을 세부적으로 파악할 수 있다.

Table 2. Details of accident factors used in association rule mining

Type of accident (TA)					
Description	Information of the accident type				
Detail values	Car-alone	Car-to-human	Car-to-bicycle	Car-to-car	
Weather					
Description	Weather conditions at the time of accident				
Detail values	Clear	Cloudy	Raining	Fog/Visibility	
Lighting					
Description	Lighting conditions at the time of accident				
Detail values	Daylight	Dusk-dawn	Dark-street lights		
AV movement preceding collision (AV_MPC)					
Description	The movement of AV prior to the accident occurrence				
Detail values	Stopped	Proceeding straight	Making right turn	Making left turn	Backing
	Slowing/Stopping	Passing other vehicle	Changing lane	N/A	
HV movement preceding collision (HV_MPC)					
Description	The movement of HV prior to the accident occurrence				
Detail values	Stopped	Proceeding straight	Making right turn	Making left turn	Backing
	Slowing/Stopping	Passing other vehicle	Changing lane	Entering traffic	Other unsafe turning
	Xing into opposing lane	Parked	Traveling wrong way	Other	N/A
AV relative position (AV_RP)					
Description	The relative position of the AV prior to the accident				
Detail values	Left	Right	Same	Opposite	N/A
HV relative position (HV_RP)					
Description	The relative position of the HV prior to the accident				
Detail values	Left	Right	Same	Opposite	N/A
Location					
Description	The road type of the accident				
Detail values	Intersection	Parking lot	Road		
Road surface (RS)					
Description	The road surface of the accident				
Detail values	Dry	Wet	N/A		
Road condition (RC)					
Description	The road condition of the accident				
Detail values	Construction-Repair zone	Reduced roadway width	No unusual conditions	N/A	
Time					
Description	The time of the accident occurrence				
Detail values	0-6	6-12	12-18	18-24	
Traffic light (TL)					
Description	The type of traffic light of the accident				
Detail values	Red light	Green light	Right-on-red	Stop sign	None
AV type of collision (AV_TC)					
Description	The collision type of AV				
Detail values	Head on	Sideswipe	Rear-end	Broadside	Hit object
	Vehicle/Pedestrian	Other	None		
HV type of collision (HV_TC)					
Description	The collision type of HV				
Detail values	Head on	Sideswipe	Rear-end	Broadside	Other
					N/A

2. 연관규칙 분석

연관규칙 분석은 데이터 속성 간의 관계를 파악하기 위한 데이터마이닝 기법이다. 특히, Apriori 알고리즘

(Agrawal and Srikant, 1994)은 대표적인 연관규칙 분석 방법론으로 범용성이 뛰어나며, 조건절(Antecedent)과 결과절(Consequent)을 설정하여 여러 속성 간의 연관성을 파악할 수 있다. 또한, 이는 자주 발생하는 빈발항목 집합을 활용하여 기존 방식 대비 연산 시간이 단축되는 장점이 있다.

Apriori 알고리즘은 다음과 같이 진행된다.

1. 전체 데이터를 대상으로 여러 속성의 조합을 통해 빈발항목 집합($I = I_1, I_2, I_3 \dots I_n$)을 추출하여 각 집합이 전체 데이터셋에서 포함되는 비율을 계산한다.
2. 빈발항목 집합의 최소 지지도(Support)를 설정하여 유의미한 후보 데이터군을 추출하며, 이를 활용하여 연관 규칙을 도출한다. 이때, 연관규칙은 A(조건절, Antecedent)→B(결과절, Consequent)의 형태 표현되며, A가 발생할 때 B가 발생한다는 의미이다.

이처럼 Apriori 알고리즘을 통해 도출되는 연관규칙은 지지도(Support), 신뢰도(Confidence), 향상도(Lift)라는 3가지 지표를 활용하여 규칙의 적절성을 판단한다. 지지도(Equation 1)는 전체 집합에서 A와 B를 모두 포함하는 집합의 빈도를 확률로 표현한 것으로 규칙의 중요성을 확인하는 지표이며, 1과 가까울수록 규칙이 매우 중요하다는 것을 의미한다. 신뢰도(Equation 2)는 조건부확률의 개념으로 A를 포함하는 집합에서 A와 B가 모두 포함되는 집합의 확률을 의미하며 규칙의 신뢰성을 판단하는 지표이다. 이때 신뢰도는 0과 1 사이의 값으로 신뢰도가 높은 규칙일수록 1에 가까운 신뢰도가 도출된다. 향상도(Equation 3)는 A가 집합에 없을 때 B가 포함되는 확률과 A가 집합에 존재할 때 B가 포함되는 확률의 증가 비율을 계산하여 도출되며, 규칙의 유용성을 판단하는 지표이다. 향상도가 1이 도출되는 것은 A와 B의 연관성이 존재하지 않는 것을 뜻하며, 두 집합(A, B)의 연관성이 높을수록 향상도는 크게 도출된다.

$$Support(A \rightarrow B) = \frac{P(A \cap B)}{N} \tag{1}$$

$$Confidence(A \rightarrow B) = P(B | A) = \frac{P(A \cap B)}{P(A)} \tag{2}$$

$$Lift(A \rightarrow B) = \frac{P(B | A)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \tag{3}$$

3. 자율주행차 사고 시나리오 생성을 위한 요소 선정

자율주행차의 안전성 평가를 위한 시나리오를 생성하기 위해 PEGASUS(PEGASUS Project Consortium, 2019), CETRAN(de Gelder et al., 2020) 등 다양한 프로젝트가 수행되고 있다. PEGASUS 프로젝트는 독일에서 진행된 자율주행차 안전성 평가 시나리오 접근 방식으로, 고속도로에서 발생할 수 있는 다양한 시나리오를 제시하고 자율주행 기능의 안전성 확보 방법과 이에 관련한 요구사항을 체계화하였다. 또한, 시나리오의 추상 수준에 따라 functional scenario, logical scenario, concrete scenario를 제시하였다(Figure 4). Functional scenario는 도로의 형태, 고정 및 동적 요소에 관한 정보, 환경 조건 등을 자연어로 설명한 시나리오이며, logical scenario와 concrete scenario는 functional scenario에서 정의한 시나리오 요인들의 범위와 값을 제공하는 시나리오이다. CETRAN 프로젝트는 싱가포르에서 진행되었으며, NHTSA의 pre-crash scenario를 기반으로 자율주행차의 안전성 평가를 위한 시나리오를 제시하였다. 또한, CETRAN 프로젝트는 PEGASUS 프로젝트와 달리 도심부를 대상으로 시나리오를 도출하였으며, 동적 요소(dynamic environment), 정적 요소(static environment), 환경(condition)으로 구분하여 시나리오 요인을 정의하여(Figure 4) 64개의 시나리오를 제시하였다.

본 연구의 목적은 실제 데이터를 활용하여 구체적인 시나리오를 제시하는 것에 있어, 실제 데이터 요인을

PEGASUS 프로젝트와 CETRAN 프로젝트의 시나리오 요인과 비교하여 시나리오 요인을 선정하고자 한다. 이에 따라 PEGASUS와 CETRAN 프로젝트 요인을 추출하고, 실제 데이터 요인과 비교하였을 때 공통적으로 사용되는 요인을 활용한다. 또한, PEGASUS 프로젝트와 CETRAN 프로젝트가 집중하는 공간이 상이하므로 실제 데이터 및 시나리오 제시 공간에 따라 각 프로젝트에만 해당하는 요인을 추가 활용한다. 종합적으로 시나리오 제시를 위한 시나리오 요인을 추출하고 이에 기반하여 시나리오 템플릿을 작성할 수 있다.

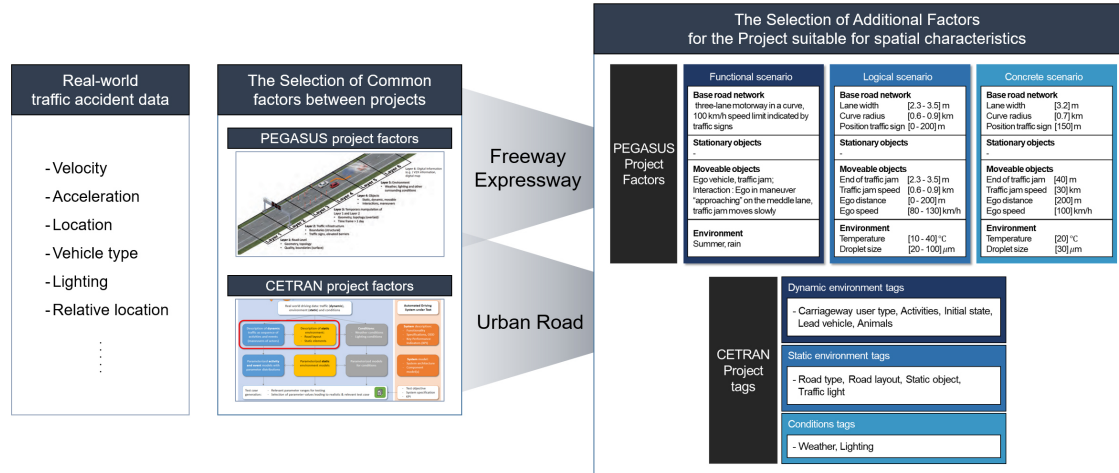


Figure 4. Scenario factor selection process based on PEGASUS and CETRAN project

위와 같은 과정을 통해 9개의 사고 시나리오 요인(Type of accident(TA), Weather, Lighting, AV movement preceding collision(AV_MPC), HV movement preceding collision(HV_MPC), AV relative position(AV_RP), HV relative position(HV_RP), Location, Traffic light(TL))을 선정하였다고 가정한다면, 이때의 시나리오는 Figure 5와 같은 형태로 제시할 수 있다. 이는 PEGASUS 프로젝트의 functional 시나리오를 기반으로 작성하였으며, 9개 요인이 모두 포함되어 있다. 사고 당시의 날씨, 조명, 위치, 신호등 상태를 Figure 5(b)–(f)에서 제공하며, 두 차량(AV, HV)의 움직임, 상대적 위치 등 차량에 대한 정보는 Figure 5(i)를 통해 파악할 수 있다. 또한, Figure 5(j)를 통해 해당 시나리오가 도출된 데이터의 출처를 제공하며, Figure 5(g), (h)를 통해 사고의 전체적인 상황을 파악할 수 있다. 마지막으로 Figure 5(a)를 통해 자율주행차가 사고를 예방하기 위한 기능의 요구사항을 정의하였다.

(a)	Safety Function Requirement				
(b)	Type of Accident				
(c)	Weather				
(d)	Lighting				
(e)	Location				
(f)	Traffic light				
(g)	Visualization				
(h)	Text Description				
(i)	Object Movement	Object	Vehicle Type	Relative position	Movement
(j)	Source of Scenario				

Figure 5. Functional scenario template of the accident presented in this study

분석 결과

1. 자율주행차 사고 데이터 기초 통계

연관규칙 분석에 앞서 165건 사고의 기초 통계 분석을 진행하였으며, 다음과 같은 특징을 확인하였다.

1. 153건(93%)의 차대차 사고가 발생함(Figure 6(a)).
2. 148건(90%)의 사고가 맑은 날씨에서 발생하였으며(Figure 6(b)), 114건(69%)의 사고가 밝은 조명 상태에서 발생함(Figure 6(c)).
3. 사고 발생 시 차량의 움직임 분포를 확인한 결과, AV 정지, HV 직진 상황은 30%(50건), AV 직진, HV 직진 상황은 10%(16건), AV 정지, HV 우회전 상황은 7%(11건)로 나타남(Figure 6(d)).
4. 사고 발생 시 차량 충돌 형태의 경우, HV가 AV 후방을 추돌한 사고((AV, HV)=(N/A, Rear-end))가 70%(115건)로 가장 많았으며 AV와 HV의 측면 충돌사고((AV, HV)=(Sideswipe, Sideswipe))가 13%(22건)로 확인됨(Figure 6(e)).
5. 두 차량이 동일 차선에 위치하였을 때((AV, HV)=(Same, Same)), 사고가 많이 발생함(83건, 50%)(Figure 6(f)).
6. 도로 표면이 마른 상태일 때와 특정 이벤트가 없는 일반적 상황(148건, 90%)일 때의 사고가 각각 151건(92%, Figure 6(g)), 148건(90%, Figure 6(h))이 발생함.
7. 오후 12시-18시에 65건(39%)의 사고 발생함(Figure 6(i)).
8. 약 77%(127건)의 사고가 교차로에서 발생하였으며(Figure 6(j)), 25건(15%)의 사고가 빨간불일 때 발생함(Figure 6(k)).

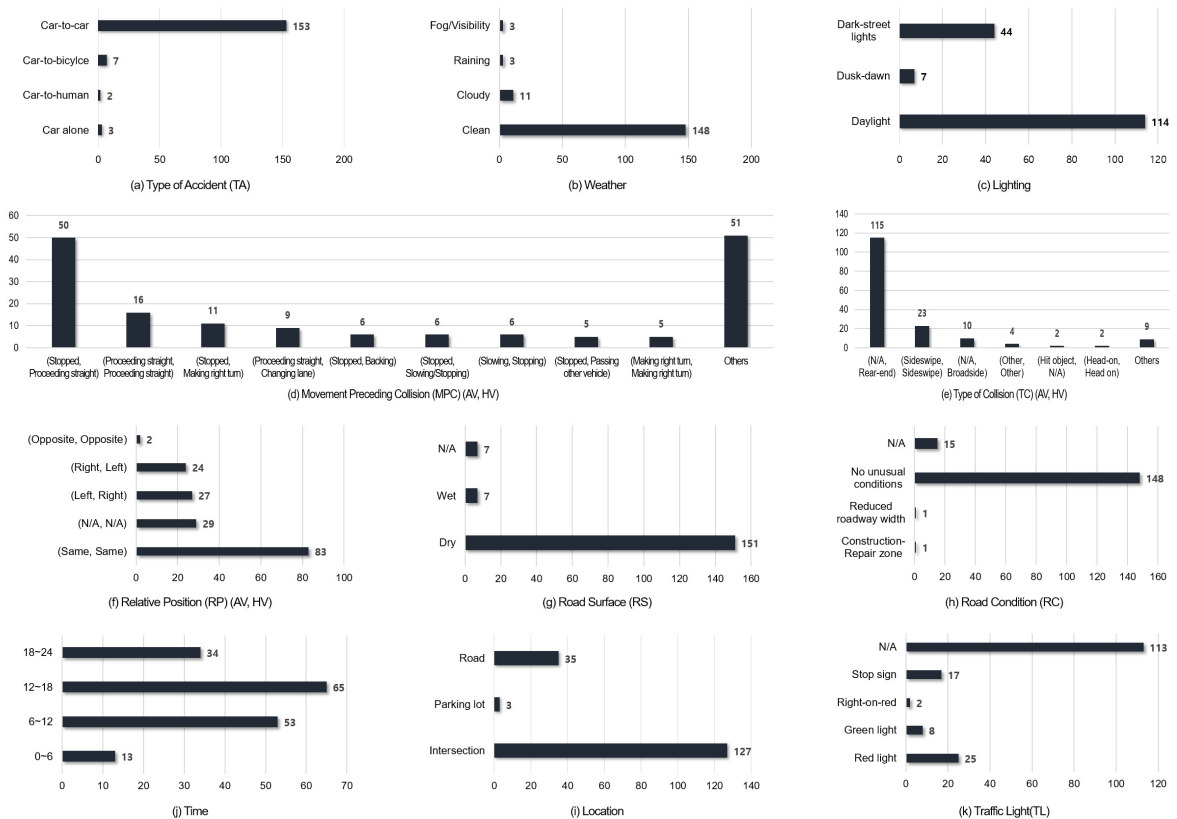


Figure 6. Result of statistical analysis of AV accident of DMV collision report

2. 연관규칙 분석 결과

연관규칙 분석을 수행하기 위해 Apriori 알고리즘의 조건을 설정하였다(Table 3). 우선, 사고 요인을 대상으로 조건절(Antecedent)과 결과절(Consequent)을 설정하였다. 통상적으로 교통사고의 결과는 사망자 수, 중상자 수, 경상자 수 등 인적 피해 또는 차량의 충돌 형태 등으로 표현된다. 또한, 결과절은 사건의 결과를 의미하므로 이에 부합하는 AV 충돌 유형(AV_TC)과 HV 충돌 유형(HV_TC)을 결과절로 설정하였다. 이후, AV 충돌 유형과 HV 충돌 유형을 제외한 12개의 사고 요인은 조건절로 설정하였다. 다음으로 빈발항목 집합을 대상으로 연관성이 높은 연관규칙을 추출하기 위해 최소 지지도(Minimum support), 최소 신뢰도(Minimum confidence), 최소 향상도(Minimum lift)를 설정하였다. 앞서 언급한 듯이 지지도, 신뢰도, 향상도는 각각 규칙의 중요성, 신뢰성, 유용성을 판단할 수 있는 지표로 이들의 최소값을 설정하여 유의미한 연관규칙을 도출하였다. 최소 지지도는 0.015로 설정하였으며, 최소 신뢰도는 0.7, 최소 향상도는 1.1로 설정하여 연관규칙 분석을 수행하였다.

Table 3. Parameters set for Apriori algorithm

Index	Value		
Antecedent	Type of accident (TA)	Weather	Lighting
	AV movement preceding collision (AV_MPC)	HV movement preceding collision (HV_MPC)	AV relative position (AV_RP)
	HV relative position (HV_RP)	Location	Road surface (RS)
	Road condition (RC)	Time	Traffic light (TL)
Consequent	AV type of collision (AV_TC)		HV type of collision (HV_TC)
Minimum support	0.015(1.5%)		
Minimum confidence	0.7(70%)		
Minimum lift	1.1		

165개의 보고서를 대상으로 연관규칙 분석을 진행한 결과, AV 충돌 형태가 결과절에 포함된 경우 164,394개의 연관규칙이 도출되었으며, HV 충돌 형태가 결과절에 포함된 경우 163,108개의 연관규칙이 도출되어 총 327,502개의 연관규칙이 도출되었다. Table 4는 약 300,000개의 연관규칙의 예시이며, 결과절로 설정한 AV 충돌 유형, HV 충돌 유형과 연관성이 높은 사고 요인은 조건절에 서술되어있다. Rule #1의 경우 도로 표면이 마르고 신호등이 없는 교차로에서 AV가 직진하고 HV가 차선을 변경할 때 발생한 차대차 사고는 AV 충돌 형태가 측면 충돌인 것과 연관성이 높다는 의미이다. Rule #3의 경우 빨간불이 점등되어있는 날씨가 맑고 밝은 상황의 교차로에서 동일차선에 위치한 AV가 정차하고 HV가 직진할 때 발생한 차대차 사고는 AV 충돌 형태가 N/A인 것과 연관성이 높다는 의미이다. 또한, Rule #12의 경우 신호등이 없는 날씨가 맑고 밝은 상황의 교차로에서 동일차선에 위치한 AV가 정차하고 HV가 직진할 때 발생한 차대차 사고는 HV 충돌 형태가 후방 추돌인 것과 연관성이 높다는 의미이다. Rule #3, #4, #5는 결과절인 AV 충돌 형태가 N/A일 때 신호등 상태를 제외하고 나머지 사고 요인이 모두 동일하였으며, Rule #12, #13, #14는 HV 충돌 형태가 후방 추돌인 연관규칙에서 신호등 상태를 제외하고 나머지 사고 요인이 모두 동일한 연관규칙임을 알 수 있다.

이처럼 조건절과 결과절을 설정하고 최소 지지도, 신뢰도, 향상도를 활용하여 연관규칙 분석을 진행한 결과, 약 300,000개의 연관규칙이 도출되었으며, AV 충돌 유형과 HV 충돌 유형에 따라 연관성이 높은 다양한 사고 요인의 조합이 존재하는 것을 확인하였다.

Table 4. Result of association rule mining (Apriori algorithm)

No.	Antecedent			Consequent
1	TA=Car-to-car Location=Intersection	AV_MPC=Proceeding straight RS=Dry	HV_MPC=Changing lane TL=None	[AV_TC= Sideswipe]
2	TA=Car-to-car HV_MPC=Passing other vehicle	Lighting=Daylight Time=6-12	AV_MPC=Stopped	[AV_TC= N/A]
3	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[AV_TC= N/A]
4	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=Red	[AV_TC= N/A]
5	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=Stop sign	[AV_TC= N/A]
6	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Slowing/Stopping Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[AV_TC= N/A]
7	TA=Car-to-car AV_MPC=Proceeding straight HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Dark-street lights AV_RP=Same TL=None	[AV_TC= N/A]
8	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Cloudy HV_MPC=Making right turn Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[AV_TC= N/A]
9	TA=Car-to-car AV_MPC=Making right turn HV_RP=Same	Weather=Clear HV_MPC=Making right turn Location=Intersection	Lighting=Dark-street lights AV_RP=Same TL=None	[AV_TC= N/A]
10	TA=Car-to-car HV_MPC=Making left turn	Weather=Clear RS=Dry	AV_MPC =Proceeding straight	[HV_TC= Sideswipe]
11	AV_MPC=Proceeding straight RS=Dry	HV_MPC=changing lane	Location=Intersection	[HV_TC= Sideswipe]
12	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[HV_TC= Rear-end]
13	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=Red	[HV_TC= Rear-end]
14	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Daylight AV_RP=Same TL=Stop sign	[HV_TC= Rear-end]
15	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Clear HV_MPC=Slowing/Stopping Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[HV_TC= Rear-end]
16	TA=Car-to-car AV_MPC=Proceeding straight HV_RP=Same	Weather=Clear HV_MPC=Proceeding straight Location=Intersection	Lighting=Dark-street lights AV_RP=Same TL=None	[HV_TC= Rear-end]
17	TA=Car-to-car AV_MPC=Stopped HV_RP=Same	Weather=Cloudy HV_MPC=Making right turn Location=Intersection	Lighting=Daylight AV_RP=Same TL=None	[HV_TC= Rear-end]

자율주행차 도심부 사고 시나리오 생성

1. 사고 시나리오 도출

PEGASUS와 CETRAN 프로젝트는 자율주행차의 안전성 평가를 위한 다양한 시나리오를 제시하였으며, 두 프로

젝트를 기반으로 시나리오를 추출하기 위한 다양한 연구가 수행되고 있다. 이에 본 장에서는 DMV 사고 보고서와 PEGASUS 및 CETRAN 프로젝트를 비교하여 최종적으로 작성할 시나리오 요인을 추출하였다. 우선, 두 프로젝트에서 모두 활용하는 8개의 요소(날씨(Weather), 조명(Lighting), AV 충돌 전 움직임(AV_MPC), HV 충돌 전 움직임(HV_MPC), AV 상대적 위치(AV_RP), HV 상대적 위치(HV_RP), 위치(Location), 신호등 상태(Traffic Light))를 최종 시나리오로 활용 요인으로 선정하였다. 이후, 본 연구는 도심부 시나리오에 집중하고 있음에 따라 도심부 시나리오를 제시한 CETRAN 프로젝트에 포함되는 요소를 선정하였으며, 이는 사고 형태(TA)가 해당되었다. 또한, AV 충돌 형태(AV_TC), HV 충돌 형태(HV_TC), 시간(Time)의 경우 PEGASUS와 CETRAN 프로젝트에서 모두 활용되지 않음에 따라 시나리오 요인에서 제외하였다. 이러한 과정을 통해 9개의 요인을 사고 시나리오 요인으로 추출하였으며, 이를 기반으로 자율주행차의 도심부 사고 시나리오를 도출하였다(Table 5).

Table 5. Accident scenario factor selection (Comparison with DMV collision report and PEGASUS, CETRAN project)

DMV collision report accident factor	PEGASUS project	CETRAN project	○/-
Type of accident (TA)	-	○	○
Weather	○	○	○
Lighting	○	○	○
AV movement preceding collision (AV_MPC)	○	○	○
HV movement preceding collision (HV_MPC)	○	○	○
AV type of collision (AV_TC)	-	-	-
HV type of collision (HV_TC)	-	-	-
AV relative position (AV_RP)	○	○	○
HV relative position (HV_RP)	○	○	○
Location	○	○	○
Road surface (RS)	○	-	-
Road condition (RC)	○	-	-
Time	-	-	-
Traffic light (TL)	○	○	○

9개의 시나리오 요인을 포함한 자율주행차의 도심부 사고 시나리오를 도출하는 과정은 다음과 같다. 우선, Table 4와 같이 도출된 327,502개의 연관규칙을 대상으로 조건절에 9개의 시나리오 요인이 모두 포함된 연관규칙을 추출하였다. 다음으로, 조건절에 포함된 9개의 시나리오 요인이 모두 동일한 연관규칙을 조합하였다. 마지막으로 결과절에 포함되어있는 AV 충돌 유형과 HV 충돌 유형을 확인하여 두 개의 연관규칙을 결합하였다. 이때, 후방 추돌은 (AV_TC, HV_TC)=(N/A, Rear-end), 측면 충돌은 (AV_TC, HV_TC)=(Sideswipe, Sideswipe)로 통일화한 것을 고려하였다. 만약, 9개의 시나리오 요인이 모두 포함된 동일한 조건절이 있는 두 개의 연관규칙의 결과절이 (AV_TC, HV_TC)=(N/A, Sideswipe)라면, 이는 DMV 사고 보고서에 기술된 충돌 유형이 아니므로 제외하였다.

우선, Table 4의 #1, #2, #10, #11은 조건절에 9개의 시나리오 요인이 모두 서술되어있지 않기 때문에 제외하였다. 또한, #12의 경우 9개의 시나리오 요인을 모두 포함하지만, AV 충돌 유형이 N/A일 때 조합이 가능한 HV 충돌 유형에 부합하는 연관규칙을 찾지 못하였다. 이에 반해, (#3, #12), (#4, #13), (#5, #14), (#6, #15), (#7, #16), (#8, #17)의 경우 조건절에 9개의 시나리오 요인이 모두 포함되어있는 것을 확인하였다. 또한, 결과절이 후방 추돌로(N/A, Rear-end) 결합이 가능하였다. 이를 통해 6개의 자율주행차 도심부 사고 시나리오가 도출되었으며, 이는 Figure 7과 같다. 사고 시나리오는 전체적인 상황을 파악할 수 있는 개략도와 함께 9개의 시나리오 요인에 대한 세부 속성으로 구성되어있다. 모든 사고 시나리오는 교차로를 대상으로 하며, AV와 HV가 동일 차선에 위치한다. Scenario 2와 3의 경우 모든 세부 속성이 Scenario 1과 동일하지만 신호등의 상태가 빨간불 또는 정지 표지판인 것을 확인할 수 있다. 또한, Scenario 5는 유일하게 AV 충돌 전 움직임이 직진이었으며, 조명이 가로등이 있는 어두운 상태임을 확인하였다. Scenario 6의 경우 유일하게 날씨가 흐림이었으며, HV 충돌 전 움직임이 우회전인 것으로 나타났다. 앞서 제시한 템플릿(Figure 5)을 활용하여 Scenario 2를 functional scenario를 작성하였으며, 이는 Figure 8과 같다.

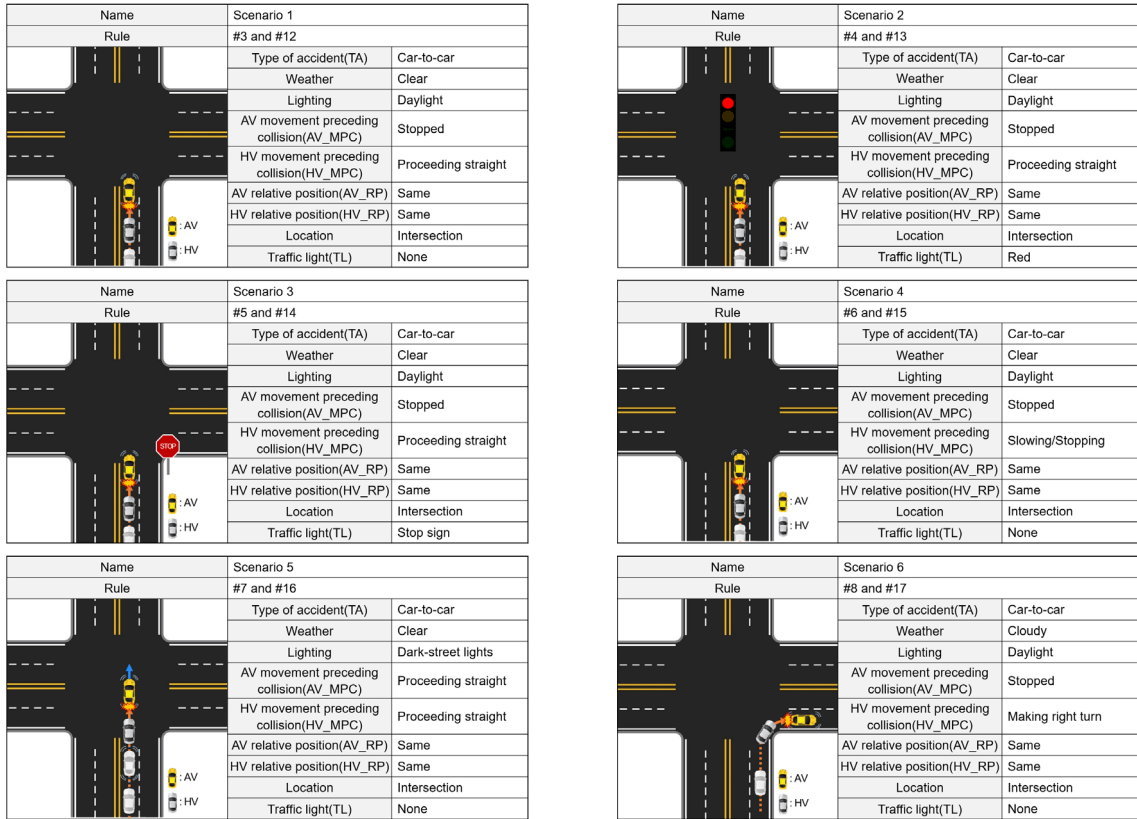


Figure 7. AV accident scenarios in the urban area generated by association rule mining

Safety Function Requirement	Recognition and response evaluation for the Target vehicle approaching from behind			
Type of Accident	Car to car			
Weather	Clear			
Lighting	Daylight			
Location	Intersection			
Traffic light	Red			
Visualization				
Text Description	In a situation where the ego vehicle(AVs) is stopping at an intersection, the ego vehicle(AVs) should respond to the situation where the target vehicle(HVs) approaches without securing a safe distance.			
Object Movement	Object	Vehicle Type	Relative position	Movement
	Ego Vehicle (AVs)	Passenger Car	Same lane	Stopped
	Target Vehicle (HV)	Passenger Car	Same lane	Proceeding straight
Source of Scenario	DMV collision report in California			

Figure 8. Scenario 2 written in a PEGASUS functional scenario format

2. 자율주행차 사고 시나리오 비교 분석

본 장에서는 연관규칙 분석을 활용하여 도출한 6개의 도심부 사고 시나리오를 세부적으로 분석하고자 한다. 우선, 165건의 DMV 사고 보고서를 대상으로 6개 사고 시나리오의 세부 속성이 포함되어있는 사고 발생 건수를 파악하였으며, 시나리오별 실제 발생한 사고 건수 및 발생 확률은 Table 6과 같다. Scenario 1은 11건이 발생하여 6.67%의 확률로 가장 높았으며, Scenario 2의 발생 확률이 4.24%로 그 뒤를 이었다. Scenario 1과 Scenario 2의 경우 날

씨, 조명, 위치, 차량의 상대적 위치 및 충돌 전 움직임 등이 통상적으로 많이 발생하는 사고의 조건과 동일함에 따라 사고 건수 및 확률이 높게 도출된 것으로 판단된다. 그러나, Scenario 3-6의 경우 사고가 5회 미만으로 발생하여 확률이 낮게 도출되었다. 만약, DMV 사고 보고서의 건수를 기준으로 시나리오를 도출한다면 4개의 시나리오는 도출되지 않을 것으로 판단된다. 이에 반해 본 연구는 연관규칙 분석을 기반으로 4개의 시나리오를 추가로 제시함에 따라 유의미한 결과를 도출하였으며, 해당 시나리오를 주목해야 할 것으로 판단된다.

Table 6. Real-world accident occurrence and probability of our AV scenarios based on DMV collision report

Scenario	Accident factors			Probability
Scenario 1	TA=Car-to-car	Weather=Clear	Lighting=Daylight	6.67% (11/165)
	AV_MPC=Stopped	HV_MPC=Proceeding straight	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=None	
Scenario 2	TA=Car-to-car	Weather=Clear	Lighting=Daylight	4.24% (7/165)
	AV_MPC=Stopped	HV_MPC=Proceeding straight	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=Red	
Scenario 3	TA=Car-to-car	Weather=Clear	Lighting=Daylight	2.42% (4/165)
	AV_MPC=Stopped	HV_MPC=Proceeding straight	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=Stop sign	
Scenario 4	TA=Car-to-car	Weather=Clear	Lighting=Daylight	1.82% (3/165)
	AV_MPC=Stopped	HV_MPC=Slowing/Stopping	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=None	
Scenario 5	TA=Car-to-car	Weather=Clear	Lighting=Dark-street lights	1.82% (3/165)
	AV_MPC=Proceeding straight	HV_MPC=Proceeding straight	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=None	
Scenario 6	TA=Car-to-car	Weather=Cloudy	Lighting=Daylight	1.82% (3/165)
	AV_MPC=Stopped	HV_MPC=Making right turn	AV_RP=Same	
	HV_RP=Same	Location=Intersection	TL=None	

다음으로, 6개의 도심부 사고 시나리오와 NHTSA의 pre-crash scenario(Najm et al., 2007)를 비교분석하였다. 이는 일반 차량의 2004 GES(General Estimates System) crash database와 CDS(Crashworthiness Data System)를 기반으로 37개의 대표 시나리오를 제시하였다. 또한, 시나리오별 부상 확률, 사고 처리 비용과 함께 날씨, 조명, 차량의 충돌 전 움직임 등에 대한 세부 속성과 이에 따른 통계적 확률을 제시하였다(Figure 9). 이에 더해 pre-crash scenario는 본 연구에서 활용한 9개의 시나리오 요인 중 사고 형태, AV 상대적 위치, HV 상대적 위치를 제외한 6개 요인에 대한 확률을 제공하는 것을 확인하였다. 이에 따라, 6개 요인에 대한 통계적 확률을 통해 시나리오별 발생 확률을 계산하였으며, 이는 Table 7과 같다. Scenario 1과 Scenario 2의 경우 일반 차량 데이터 기반의 시나리오를 기

Following Vehicle Approaching a Stopped Lead Vehicle

Typical Scenario: Vehicle is going straight in an urban area, in daylight, under clear weather conditions, at an intersection-related location with a posted speed limit of 35 mph, and closes in on a stopped lead vehicle.

Factor Over-Representation: Rural area, intersection-related, inattention, speeding and younger driver are over-represented (based on a simple comparison of percentages).

Dynamic Variations: Vehicle is decelerating in traffic lane and closes in on a stopped lead vehicle (12% of crashes). Vehicle may also be starting in traffic lane and closes in on a stopped lead vehicle (8% of crashes). In about 50 percent of the lead-vehicle-stopped crashes, the lead vehicle first decelerates to a stop and is struck afterwards by the following vehicle. This typically happens in the presence of a traffic control device or the lead vehicle is slowing down to make a turn. Thus, this particular scenario overlaps with the lead vehicle-decelerating scenario.

Scenario Severity: Table below quantifies the annual severity of this crash scenario in terms of five different metrics based on 2004 GES statistics. This table also provides the ratios of people involved by maximum injury severity using the KABCO and AIS injury scales. About 0.50 percent of all people involved in this crash scenario suffered high-level MAIS 3+ injuries (serious, severe, critical, or fatal).



Following Vehicle Approaching a Stopped Lead Vehicle

Driving Environment			Driver		Vehicle			
Lighting	Daylight	81%	Vision	Yes	4%	Contributing Factors	Yes	1%
	Dark Lighted	12%		No	96%		No	90%
	Dark	4%		No Obstruction	71%		Unknown	9%
	Dawn/Dusk	3%		Obscured	2%		Yes	0.1%
Weather	Clear	85%	Driver	Unknown	2%	Pre-Event Movement	No	100%
	Adverse	15%		Inattention	13%		No Driver Present	-
Road Surface	Dry	79%	Distractions	Sleepy	1%		No Driver Present	77%
	Wet/Slippery	21%		Not Distracted	21%		Decelerating in Traffic Lane	12%
Road	Straight	91%	Speeding	Unknown	41%		Accelerating in Traffic Lane	1%
	Curve	9%		Yes	35%		Starting in Traffic Lane	8%
Alignment	Level	80%	Other	No	61%		Stopped in Traffic Lane	0.04%
	Other	20%		Unknown	4%		Passing Another Vehicle	-
Land Use	Rural	49%	Violation	Speeding	0.1%		Parked in Travel Lane	-
	Urban	51%		Reckless	1%		Leaving a Parked Position	-
Day	Weekday	82%	Impairment	None	43%	Entering a Parked Position	-	
	Weekend	18%		Other	51%	Turning Right	0.05%	
	On Roadway	99%		Unknown	5%	Turning Left	-	
	Shoulder/Parking Lane	0.1%		Bl/Blackout	0.1%	Making U-turn	-	
Relative to Roadway	Off Roadway	0.5%	Other	Drowsy	1%	Backing Up	-	
	Left Turn Lane	0.3%		None	93%	Negotiating a Curve	2%	
	Unknown	0.03%		Other	3%	Changing Lanes	-	
	Non-Junction	34%		Unknown	4%	Merging	-	

Figure 9. An example of the scenario presented by NHTSA, along with statistical probabilities for each detailed factors as part of the pre-crash scenarios

준으로 발생 확률이 10%가 초과하는 것을 확인하였다. 그러나, Scenario 3-5의 경우 발생 확률이 매우 낮았으며, Scenario 6의 경우 0.0015%의 발생 확률이 도출되어 일반 차량을 대상으로 발생 확률이 매우 낮은 시나리오인 것을 알 수 있다. 이에 본 연구와 같이 자율주행차 시나리오는 일반 차량 데이터가 아닌 자율주행차 데이터를 기반으로 작성되어야 할 것으로 판단된다.

Table 7. Probability of our AV scenarios based on NHTSA pre-crash scenario

Index	Accident factors (Included in NHTSA pre-crash scenario)			Probability
Scenario 1	Weather=Clear	Lighting=Daylight	AV_MPC=Stopped	12.88%
	HV_MPC=Proceeding straight	Location=Intersection	TL=None	
Scenario 2	Weather=Clear	Lighting=Daylight	AV_MPC=Stopped	11.16%
	HV_MPC=Proceeding straight	Location=Intersection	TL=Red	
Scenario 3	Weather=Clear	Lighting=Daylight	AV_MPC=Stopped	2.58%
	HV_MPC=Proceeding straight	Location=Intersection	TL=Stop sign	
Scenario 4	Weather=Clear	Lighting=Daylight	AV_MPC=Stopped	2.01%
	HV_MPC=Slowing/Stopping	Location=Intersection	TL=None	
Scenario 5	Weather=Clear	Lighting=Dark-street lights	AV_MPC=Proceeding straight	2.24%
	HV_MPC=Proceeding straight	Location=Intersection	TL=None	
Scenario 6	Weather=Cloudy	Lighting=Daylight	AV_MPC=Stopped	0.0015%
	HV_MPC=Making right turn	Location=Intersection	TL=None	

결론 및 향후 연구

인적 요인에 의한 교통사고를 예방하기 위해 자율주행차가 개발되고 있으며, 자율주행차의 안전성과 신뢰성을 확보하기 위해 자율주행자동차의 안전성 평가에 활용 가능한 다양한 상황의 시나리오가 제시되고 있다. 이처럼 자율주행차의 안전성 평가를 위한 시나리오는 신뢰성과 구체성을 확보해야하며, 사고의 발생 빈도가 높은 도심부에 주목해야 한다. 이에 본 연구는 자율주행차의 사고데이터인 DMV 사고 보고서를 활용한 연관규칙 분석을 통해 자율주행차의 도심부 사고 시나리오를 제시하는 연구를 수행하였다. 연관규칙 분석을 위해 DMV 사고 보고서의 수집 및 전처리를 진행하였으며, 조건절과 결과절을 설정하는 등 연관규칙 분석의 조건을 설정하였다. Apriori 알고리즘을 활용한 연관규칙 분석 결과, 결과절로 설정한 일반 차량과 자율주행차의 충돌 형태와 연관성이 높은 사고 요인이 도출되었으며, 차선변경, 직진, 정지 등 차량의 충돌 전 움직임과 날씨, 조명, 위치 등 다양한 사고 요인의 조합이 차량의 충돌 형태와 연관성이 높은 것을 확인하였다. 이후, PEGASUS 및 CETRAN 프로젝트를 검토하여 9개의 시나리오 요인을 선정하였으며, 추출된 연관규칙을 기반으로 자율주행차의 안전성 평가를 위한 구체적인 시나리오를 제시하였다. 연관규칙을 조합한 결과, 신호등의 상태, 차량의 충돌 전 움직임 등이 다양한 6개의 사고 시나리오가 도출되었으며, 해당 시나리오는 9개의 시나리오 요인을 구체적으로 제시하였다. 또한, 6개의 시나리오를 대상으로 DMV 사고 보고서의 기초통계분석과 일반 차량 데이터 기반의 NHTSA pre-crash scenario를 활용하여 시나리오의 발생 확률을 도출한 결과, 확률이 낮은 시나리오가 도출됨에 따라 해당 시나리오를 주목해야 할 것으로 판단된다.

본 연구는 자율주행차 사고데이터의 절대적인 개수가 부족하여 교차로에서 발생한 후방 추돌 상황의 시나리오만 도출되었다는 한계가 존재하였다. 또한, DMV 사고 보고서가 제공하는 데이터 범위의 한계로 인해 평가를 위한 세부적인 실험값이 제공되지 않은 한계가 존재하였다. 이에 향후 지속적으로 축적되는 DMV 사고 보고서를 활용하여 데이터 기반의 신뢰성 있는 시나리오와 보다 다양한 상황의 시나리오를 도출하는 연구를 수행하고자 한다. 또한, DMV 사고 보고서 이외의 차량의 속도, 가속도 등이 기술된 사고 보고서 및 데이터를 활용하여 구체적인 실험값을 제공하는 logical scenario와 concrete scenario를 제시하는 연구를 수행하고자 한다. 본 연구는 자율주행차의 실사고 데이터를 활용하여 신뢰성과 구체성을 확보한 시나리오를 제시한 것에 의의가 있으며, 이는 자율주행차의 안전성 평가를 위한 기초적인 자료로 활용 가능할 것으로 판단된다.


Funding


This research was supported by a grant(RS-2023-00236825) from the Transportation and Logistics R&D Program funded by the Ministry of Land, Infrastructure, and Transportation of the Korean government and 2022 Hongik University Research Fund.


알림


본 논문은 한국ITS학회 2022년도 국제학술대회(2022.06.17)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

ORCID

LEE, Hojun  <http://orcid.org/0000-0002-0552-7712>

KANG, Minhee  <http://orcid.org/0000-0003-2548-6011>

SONG, Jaein  <http://orcid.org/0000-0002-1759-9617>

HWANG, Keeyeon  <http://orcid.org/0000-0001-9472-7996>

REFERENCES

- Agrawal R., Srikant R. (1994), Fast Algorithms for Mining Association Rules, In Proc. 20th int. conf. Very Large Data Bases, VLDB, 1215, 487-499.
- Alambeigi H., McDonald A. D., Tankasala S. R. (2020), Crash Themes in Automated Vehicles: A Topic Modeling Analysis of the California Department of Motor Vehicles Automated Vehicle Crash Database, arXiv preprint arXiv:2001.11087.
- Ashraf M. T., Dey K., Mishra S., Rahman M. T. (2021), Extracting Rules from Autonomous-vehicle-involved Crashes by Applying Decision Tree and Association Rule Methods, Transportation Research Record, 2675(11), 522-533.
- Bahiru T. K., Singh D. K., Tessfaw E. A. (2018), Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity, In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 1655-1660, IEEE.
- Bhin M. Y., Son S. K. (2018), Analysis of Factors Influencing Traffic Accident Severity According to Gender of Bus Drivers, J. Korean Soc. Transp., 36(6), Korean Society of Transportation, 440-451.
- Bhin M. Y., Son S. K., Kim J. Y., Chae W. R. (2020), A Study on the Analysis of Traffic Accidents by Road Type considering Traffic Safety Signs on the Urban Road, J. Korean Soc. Transp., 38(5), Korean Society of Transportation, 390-403.
- Das S., Dutta A., Tzapakis I. (2020), Automated Vehicle Collisions in California: Applying Bayesian Latent Class Model, IATSS Research, 44(4), 300-308.
- de Gelder E., den Camp O. O., de Boer N. (2020), Scenario Categories for the Assessment of Automated Vehicles, CETRAN, Singapore, Version, 1.
- DrissiTouzani H., Faquir S., Yahyaouy A. (2020), Data Mining Techniques to Analyze Traffic Accidents Data: Case

- Application in Morocco, In 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), 1-4, IEEE.
- Favarò F. M., Nader N., Eurich S. O., Tripp M., Varadaraju N. (2017), Examining Accident Reports Involving Autonomous Vehicles in California, PLoS one, 12(9), e0184952.
- Goodall N. J. (2014), Ethical Decision Making During Automated Vehicle Crashes, Transportation Research Record, 2424(1), 58-65.
<https://www.dmv.ca.gov>
- ISO (2018), 26262 – Road vehicles – Functional Safety.
- Kang M., Im I. J., Song J., Hwang K. (2022a), Is Only the Dedicated Lane for Automated Vehicles Essential in the Future? The Dedicated Lanes Optimal Operating System Evaluation, Sustainability, 14(18), 11490.
- Kang M., Lee W., Hwang K., Yoon Y. (2022b), Vision Transformer for Detecting Critical Situations and Extracting Functional Scenario for Automated Vehicle Safety Assessment, Sustainability, 14(15), 9680.
- Kang M., Song J., Hwang K. (2022c), The Extraction of Automated Vehicles Traffic Accident Factors and Scenarios Using Real-World Data, In Congress on Intelligent Systems: Proceedings of CIS 2021, 1, 1-15, Singapore: Springer Nature Singapore.
- Kang M., Song J., Hwang K. (2020), For Preventative Automated Driving System (PADS): Traffic Accident Context Analysis Based on Deep Neural Networks, Electronics, 9(11), 1829.
- Kim H. J., Jeon G. S., Jang J. A., Yoon I. S. (2020), Analysis of Factors Affecting Buses and Trucks Crash Severity Using Meta Analysis, J. Korean Soc. Transp., 38(6), Korean Society of Transportation, 520-535.
- Kumar S., Toshniwal D. (2016), A Data Mining Approach to Characterize road Accident Locations, Journal of Modern Transportation, 24, 62-72.
- Leilabadi S. H., Schmidt S. (2019), In-depth Analysis of Autonomous Vehicle Collisions in California, In 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 889-893, IEEE.
- Lenard J., Badea-Romero A., Danton R. (2014), Typical Pedestrian Accident Scenarios for the Development of Autonomous Emergency Braking Test Protocols, Accident Analysis & Prevention, 73, 73-80.
- Li L., Shrestha S., Hu G. (2017), Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques, In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 363-370, IEEE.
- Liu Q., Wang X., Wu X., Glaser Y., He L. (2021), Crash Comparison of Autonomous and Conventional Vehicles Using Pre-crash Scenario Typology, Accident Analysis & Prevention, 159, 106281.
- Muhammad L. J., Sani S., Yakubu A., Yusuf M. M., Elrufai T. A., Mohammed I. A., Nuhu A. M. (2017), Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, Its Prone Locations and Yime Along Kano-Wudil Highway, Int J Database Theory Appl, 10(11), 197-208.
- Najm W. G., Smith J. D., Yanagisawa M. (2007), Pre-crash Scenario Typology for Crash Avoidance Research (No.DOT-VNTSC-NHTSA-06-02), United States. National Highway Traffic Safety Administration.
- Nitsche P., Thomas P., Stuetz R., Welsh R. (2017), Pre-crash Scenarios at Road Junctions: A Clustering Method for Car Crash Data, Accident Analysis & Prevention, 107, 137-151.

- Park J. S., Kim T. Y., Yoo D. S. (2006), Development of Traffic Accident Prediction Models by Correlation between Environment factor and Traffic Accident, The 53rd Conference of KST, Korean Society of Transportation, 350-359.
- PEGASUS Project Consortium (2019), PEGASUS Method: An Overview, Available: <https://www.pegasusprojekt.de/?les/tmpl/Pegasus-Abschlussveranstaltung/PEGASUSGesamtmethode.Pdf>
- Riedmaier S., Ponn T., Ludwig D., Schick B., Diermeyer F. (2020), Survey on Scenario-based Safety Assessment of Automated Vehicles, *IEEE*, 8, 87456-87477.
- Shanthi S., Ramani R. G. (2012), Feature Relevance Analysis and Classification of Road Traffic Accident Data Through Data Mining Techniques, In Proceedings of the world congress on engineering and computer science, 1, 24-26, sn.
- Son S. N., Lee J. Y., Jang K. H. (2021), Development of Demonstration Scenarios for Autonomous Cooperative Driving on Urban Roads, *Transportation Technology and Policy*, 18(5), Korean Society of Transportation, 61-68.
- Song Y., Chitturi M. V., Noyce D. A. (2021), Automated Vehicle Crash Sequences: Patterns and Potential Uses in Safety Testing, *Accident Analysis & Prevention*, 153, 106017.
- Steimle M., Menzel T., Maurer M. (2021), `Toward a Consistent Taxonomy for Scenario-Based Development and Test Approaches for Automated Vehicles: A Proposal for a Structuring Framework, a Basic Vocabulary, and Its Application, *IEEE*, 9, 147828-147854.
- Sui B., Lubbe N., Bärghman J. (2019), A cClustering Approach to Developing Car-to-two-wheeler Test Scenarios for the Assessment of Automated Emergency Braking in China Using in-depth Chinese crash data, *Accident Analysis & Prevention*, 132, 105242.
- Taamneh M., Alkheder S., Taamneh S. (2017), Data-mining Techniques for Traffic Accident Modeling and Prediction in the United Arab Emirates, *Journal of Transportation Safety & Security*, 9(2), 146-166.
- Torres J., Li Y., Zhang J. (2021), Investigating Traffic Crashes Involving Autonomous Vehicles, In IIE Annual Conference, Proceedings (1046-1051), Institute of Industrial and Systems Engineers (IISE).
- Yu R., Ai H., Gao Z. (2020), Identifying High Risk Driving Scenarios Utilizing a CNN-LSTM Analysis Approach, In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 1-6, IEEE.
- Yuan Q., Xu X., Zhou J. (2020), Paving the Way for Autonomous Vehicle Testing in Accident Scenario Analysis of Yizhuang Development Zone in Beijing, In CICTP 2020, 62-72.