

블랙박스 영상 분석을 통한 토픽모델링 기반의 자율주행차 Edge Case 시나리오 구축 : 보행자, 이륜차, 자전거, 전동킥보드를 중심으로

김태준¹ · 주혜연¹ · 장정아² · 윤현정³ · 김정화^{4*}

¹경기대학교 도시·교통공학과 석사과정, ²아주대학교 TOD기반 지속가능 도시교통연구센터 연구교수,

³한국전자통신연구원 자율주행지능연구실 책임연구원, ⁴경기대학교 도시·교통공학과 조교수

Development of Autonomous Vehicle Edge Case Scenarios Based on Topic Modelling Through Dashcam Video Analysis : Focusing on Pedestrian, Motorcycle, Bicycle, E-Scooters

KIM, Taejun¹ · CHU, Hyeyeon¹ · Jang, Jeong Ah² · Yun, Hyunjeong³ · KIM, Junghwa^{4*}

¹Master Course, Department of Urban & Transportation Engineering, Kyonggi University, Gyeonggi 16227, Korea

²Research Professor, TOD-based Sustainable City/Transportation Research Center, Ajou University, Gyeonggi 16499, Korea

³Principal Researcher, Autonomous Driving Intelligence Research Laboratory, Electronics and Telecommunications Research Institute(ETRI), Daejeon 34129, Korea

⁴Assistant Professor, Department of Urban & Transportation Engineering, Kyonggi University, Gyeonggi 16227, Korea

*Corresponding author: junghwa.kim@kyonggi.ac.kr

Abstract

Ensuring that autonomous vehicles can safely operate in complex road environments where they interact with various road users-including pedestrians, motorcycles, bicycles, and e-scooters-remains a critical challenge for the commercialization of autonomous driving technology. In particular, Edge Case scenarios, which are characterized by their low occurrence probability but high risk, are essential for validating the safety and reliability of autonomous vehicles. However, existing research has primarily focused on safety evaluation scenarios involving interactions between autonomous and conventional vehicles, with limited consideration of diverse road users such as pedestrians, motorcycles, bicycles, and scooters in real-world traffic environments. This study aims to develop Edge Case scenarios that address potential accident risks faced by autonomous vehicles. To achieve this, 2,410 dashcam dataset were collected from the youtube channel ‘HanMoonchul TV’. The data were processed through text preprocessing, topic modelling using the Latent Dirichlet Allocation (LDA) technique to extract key accident risk factors. The identified risk factors were classified according to the Pegasus 5-layer framework. Based on this classification, various Edge Case scenarios involving interactions with pedestrians, motorcycles, bicycles, and e-scooter were constructed. The results demonstrate that the designed scenarios enable autonomous vehicles to operate reliably even in unpredictable accident situations. These scenarios serve as tools for validating the safety and performance of autonomous vehicles. By employing a data-driven approach, this

J. Korean Soc. Transp.
Vol.43, No.5, pp.554-575, October 2025
<https://doi.org/10.7470/jkst.2025.43.5.554>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 20 January 2025

Revised: 26 February 2025

Accepted: 21 June 2025

Copyright © 2025
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/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

study addresses proposes a methodology for developing scenarios to evaluate safety in complex traffic environments.

Keywords: autonomous vehicle, edge case, pegasus 5-layer, safety evaluation scenarios, topic modelling

초록

자율주행차가 다양한 도로 사용자와 혼재된 복잡한 도로 환경에서 안전하게 운행할 수 있도록 하는 것은 자율주행 기술 상용화를 위한 핵심 과제로 대두되고 있다. 특히, 발생 가능성은 낮지만 위험성이 높은 Edge Case 상황은 자율주행차의 안전성과 신뢰성을 검증하는 것이 매우 중요하다. 그러나 현재까지의 연구는 자율주행차와 차량 간 사고 중심의 안전성 평가 시나리오 구축에 집중되어 있으며, 실제 도로에서 자율주행차가 상호 작용할 보행자, 이륜차, 자전거, 킥보드 등의 다양한 객체를 고려한 연구는 부족한 실정이다. 이에 본 연구는 자율주행차와 해당 객체 간의 사고 위험상황을 다룰 수 있는 Edge Case 시나리오를 구축하는 것을 목표로 한다. 이를 위해, 유튜브 채널 '한문철 TV'에서 수집한 2,410건의 블랙박스 교통사고 데이터셋을 기반으로 LDA 토픽모델링 기법을 사용하여 주요 사고 위험 요인을 도출하였다. 도출된 위험 요인은 Pegasus 5-layer 포맷에 따라 체계적으로 분류하고, 이를 바탕으로 자율주행차와 보행자, 이륜차, 자전거, 전동킥보드 간 다양한 사고 시나리오를 구축하였다. 분석 결과, 자율주행차가 예측하기 어려운 사고 상황에서도 대응성을 향상할 수 있는 시나리오를 구축했으며, 이러한 시나리오는 실도로에서 자율주행차의 안전성과 성능 검증을 위한 도구로 활용될 수 있을 것으로 기대된다. 또한, 본 연구는 데이터 기반 접근 방식을 통해 복잡한 교통 환경에서 자율주행차의 안전성 평가를 위한 체계적인 시나리오 구축 방법을 제시하였다.

주요어: 자율주행차, 엣지 케이스, 페가수스 5-레이어, 안전성 평가 시나리오, 토픽모델링

연구배경 및 목적

세계보건기구(WHO)에 따르면, 자동차 교통사고로 인해 전 세계적으로 연간 약 135만 명이 사망하고 있다. 교통사고는 90% 이상이 운전자의 부주의로 발생하기 때문에 이러한 암울한 통계 속에서 자율주행차가 안전한 대체 교통수단으로 주목받고 있는 상황이다(Masmoudi et al., 2019; Luettel et al., 2012). 특히, 국제자동차기술협회(SAE) 기준 Level 3 이상의 자율주행차는 교통사고 사망자 수를 크게 줄이는 데 기여할 것으로 기대된다(Riedmaier et al., 2020). 이에 따라 Tesla, BMW, Mercedes Benz, Toyota, 현대 등 자동차 제조사와 Google, Uber, Waymo 등과 같은 기업들도 자율주행 기술 개발에 적극적으로 참여하고 있다. 또한, 우리나라 정부도 자율주행차의 연구·시범운행을 촉진하기 위해 2024년 6월 기준 서울, 판교, 제주 등을 포함한 17개 시·도 36곳을 자율주행 시범운행지구로 지정하여 안전 기준 일부 면제, 유상 여객·화물운송 특례 등 부여를 통해 미래 모빌리티 전환의 핵심 기술로 자율주행차 기술 발전과 면밀한 실증을 위해 계속해서 노력하고 있다(MOLIT, 2024).

그러나 이러한 자율주행차 시범운행과 기술 발전 노력에도 불구하고 자율주행 안전성 확보에 대한 우려가 지속되고 있으며 자율주행차 관련 교통사고 또한 여전히 발생하고 있는 상황이다. 미국 도로교통안전국(NHTSA)는 2021년 7월부터 2022년 5월까지 10개월간 사고를 조사한 결과 총 392건의 교통사고가 자율주행으로 인해 발생했다고 밝히면서 사고 발생 시 제조사 보고를 의무화하여 자율주행 안전성에 대한 중요성을 강조했다(NHTSA, 2022). 또한, 2016년 Google의 자율주행차 사고, Tesla Auto Pilot의 인지 실패, 2018년 Uber의 사망 사고 등과 같은 교통사고가 발생함에 따라 자율주행차 안전성 강화의 필요성이 부각되고 있다(Park et al., 2018). 이를 미루어 볼 때 자율주행차 안전성에 대한 지속적인 평가와 기술 개발이 필요하다는 것을 나타낸다.

한편, 이러한 자율주행차 안전성 평가를 위해 자율주행차 및 일반차량 교통사고 데이터를 기반으로 시나리오 구축 연구가 활발히 진행되고 있다. 이렇게 구축된 자율주행차 사고 시나리오는 실제 도로 운행 및 시뮬레이션 등을 통해 다양한 환경을 고려하여 조절할 수 있고 정교하게 설계된 모든 시나리오를 다룰 수 있기 때문에 다양한 환경에서

의 자율주행차의 안전성을 종합적 평가를 가능하게 한다는 장점이 있다(Lee et al., 2023). 아울러, 발생 가능성과 시스템 반응의 난이도에 따라 Typical Case, Critical Case, Edge Case로 분류하고 있으며 자율주행 안전성 확보를 위해 설계 단계에서 반드시 고려되어야 하는 시나리오로 강조되고 있다(Najm et al., 2007; Kalisvaart, 2021b).

그러나 기존 시나리오 연구는 사고 발생 가능성이 높은 차량 간 시나리오 구축에 집중되어 있으며, 자율주행차와 보행자, 이륜차, 자전거, 전동킥보드 등 차량 외 객체의 갑작스러운 움직임으로 인해 예측하지 못해 발생할 수 있는 Edge Case에 대한 분석과 시나리오 구축은 차량 간 사고에 비해 부족한 실정이다. 보행자, 이륜차, 자전거, 전동킥보드와 같은 객체는 이동 경로나 속도가 일정하지 않고, 갑작스럽게 방향을 바꾸거나 멈추는 경우가 많아 예측하기 힘든 움직임을 보여 자율주행차의 센서가 즉각적으로 인식하고 대응하기 어려우며 특히 도심의 복잡한 교통 환경에서 안전성에 영향을 미칠 수 있다(Kang et al., 2024). 따라서 도로 상에서 자율주행차가 보행자, 이륜차, 자전거, 전동킥보드와 같은 객체의 갑작스러운 움직임에 대응하기 위한 시나리오를 구축하여 자율주행차의 안전성 평가에 활용할 필요가 있는 시점이다.

이에, 본 연구는 차대 보행자, 차대 이륜차, 차대 자전거, 차대 전동킥보드 교통사고가 발생한 블랙박스 데이터셋을 기반으로 LDA 토픽모델링 기법과 Pegasus 5-layer 포맷을 활용하여 사고 발생 전후 도로 상태, 운행상황, 환경요인 등 복잡하고 다양한 환경을 반영하고 예측하기 어려운 Edge Case 시나리오를 구축하여 실도로에서 자율주행차의 안전성 평가 방법으로 제시하는 것을 목적으로 한다. 이를 통해 구축된 시나리오는 자율주행차 기술 개발 과정에서 안전성 향상을 위한 기초적인 자료로 활용 가능하며 실도로 상에서 발생할 수 있는 다양한 교통사고 유형을 대비할 수 있을 것으로 기대된다.

선행연구 검토

1. Edge Case

자율주행차 사고 시나리오는 Figure 1과 같이 발생 가능성과 시스템 반응의 난이도에 따라 Typical Case(일반), Critical Case(위험), Edge Case(특이) 시나리오로 구분된다. Typical Case는 자율주행 시스템이 정상적으로 작동해야 하는 일상적이고 반복적인 주행 상황으로 일반적인 도로 환경에서 자주 발생하는 시나리오를 의미한다(Najm et al., 2007). Critical Case는 시스템의 안전성과 기능 수행에 직접적인 위험을 초래할 수 있는 치명적이고 고위험한 상황으로 테스트 및 검증 과정에서 우선 고려되어야 하는 시나리오를 말한다(Durak et al., 2023; Zhang et al., 2021). 마지막으로 Edge Case는 자율주행 시스템(ADS)의 성능을 도전하게 만드는 드물고 복잡한 상황으로, 발생 가능성은 낮지만 일반적인 학습 데이터나 설계된 운행 조건을 벗어난 예외적 시나리오를 의미한다(Moradloo et al., 2024; Kalisvaart, 2021a). Typical Case는 자율주행 시스템이 잘 작동해야 하는 평범한 상황이므로 안전성 평가를 위한 검증 대상이 아니라 기본적인 조건에 해당되며, Critical Case는 자율주행 시스템의 물리적 성능 한계를 도전하게 하지만 예측 가능한 구조를 가지고 있다. 이와 반면에 Edge Case는 반드시 Critical Case처럼 치명적이고 고위험 하지는 않을 수 있으나 자율주행차의 안전성 확보를 위해 설계 단계에서 반드시 고려되어야 하는 시나리오로 강조되고 있다(Kalisvaart, 2021b). 또한, 실제 사고의 주된 원인이며 자율주행 시스템의 안전에 대한 신뢰성과 책임성 검증에서 중요한 기준으로 간주되는 중요한 시나리오로 주목받고 있다.

Moradloo et al.(2024)는 자율주행 시스템(ADS)의 안전성을 확보하기 위해 예외적 교통상황에 해당하는 Edge Case를 식별하고자 하였다. 이를 위해 미국 NHTSA에서 수집한 ADS 관련 사고 보고서 392건 중 189건을 분석 대상으로 선정하고, 이 중 실제 시스템 작동 중이었던 사고 사례를 정량적·질적 방법으로 평가하였다. 해당 연구는 센서 실패(Sensor), 비정형 객체(Content), 시간적 복합성(Temporal) 등 세 가지 유형의 Edge Case를 정의하고 사례를 분류하였으며, 그 결과 전체 사고 중 약 8%(15건)가 Edge Case로 식별되었다. 해당 사고들은 주로 보행자 돌발진입, 인식 불가능한 물체, 비·야간·혼잡 등 복합적 환경 요인과 관련되어 있었음을 확인하였다. 또한, Kalisvaart

(2021b)는 시야가 차단된 횡단보도에서 보행자가 건너는 동시에 자전거가 추월하는 상황과 원형 교차로의 반지름이 작고 해가 낮게 떠 있어 눈부심이 발생하는 환경에서 고속주행 중인 자전거와 노면의 미끄러움이 동시에 존재하는 조건과 같은 다양한 환경 요소와 객체 간의 상호작용이 모두 포함된 복합적인 상황이 Edge Case라고 설명하였다.

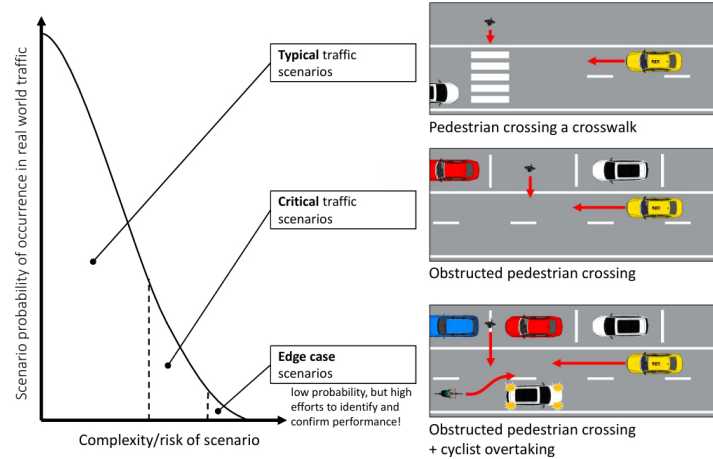


Figure 1. Scenario case definition (Kalisvaart, 2021a)

2. Pegasus 5-layer 포맷

Pegasus 5-layer 포맷은 독일 Pegasus Project에서 자율주행차 평가 프레임워크로 자율주행차 기준, 평가 방법 수립을 목표로 활용되었다. 자율주행차 실험 시나리오를 구현하기 위해 도로의 배치, 도로 형상과 같은 기하구조부터 시나리오에 참여하는 객체, 날씨, 통신 조건 등을 5단계 layer로 구성하여 자율주행 안정성 실험 수행을 돕는 시나리오 개발 포맷이다(Pegasus, 2019; MIRA, 2018). 본 연구에서는 사람이 운전하는 차량의 블랙박스 교통사고 영상에서 도출한 사고 발생 전후의 도로 상태, 운행상황, 환경요인 등 다양한 환경을 자율주행차 안전성 평가 시나리오에 적용하기 위해 사고 위험요인 키워드를 Pegasus 5-layer 항목에 따라 분류하여 자율주행차 안전성 평가를 위한 충돌 시나리오를 구축하고자 한다. 본 연구에서 활용할 Pegasus 5-layer 포맷의 layer 항목별 구성요소는 Table 1과 같다.

Table 1. Pegasus 5-layer format

Layer	Components
Layer 1 : Road-Level	Road geometry and topology, road surface quality and boundaries
Layer 2 : Traffic infrastructure	Traffic signs, barriers, road markings, lane
Layer 3 : Temporary modifications	Road construction work, traffic obstructions
Layer 4 : Objects	Dynamic of ego vehicle, traffic-relevant object, pedestrians
Layer 5 : Environment	Lightning conditions, weather (rain, snow, fog, etc.)

3. 자율주행차 시나리오 구축 연구

최근 자율주행차 안전성 평가를 위해 교통사고 데이터 기반의 시나리오 구축 연구가 다수 진행되고 있다. 이러한 시나리오 접근법은 실제 도로 주행 기반 접근법의 경제적·시간적 제약을 극복하며, 제한된 평가 환경에서도 교통사고 요인과 외부 환경을 체계적인 분석을 통해 구체적이고 신뢰성이 높은 시나리오를 도출할 수 있는 장점이 있어 자율주행차의 안전성 평가를 위한 시나리오 구축 방법으로 주목받고 있다. 이에 관련 연구에 대한 검토를 진행하였다. 먼저 Liu et al.(2024)은 캘리포니아 DMV 사고 데이터를 바탕으로 교차로에서 발생한 197건의 자율주행차 사고에

대해 연관규칙 분석을 통해 사고 특성과 위험요인을 분석하여 30가지 유형의 사고 발생 전 시나리오를 도출하였으며, Lee et al.(2023)도 연관규칙 분석을 통해 약 30만개의 연관규칙을 확인하고 차량 충돌 형태와 연관성이 높은 다양한 사고요인 조합을 도출하여 6개의 자율주행차 도심부 사고 시나리오를 제시하였다. Park et al.(2021a)은 한국 경찰청 교통사고 데이터를 활용하여 도시 주요도로와 교차로에서 발생한 사고를 선별하고 주요 특징을 추출하여 자연어 처리와 TF-IDF 기법을 통해 자율주행차를 위한 평가 시나리오 추출하는 방법론을 개발하였으며, Kang et al.(2022)은 Vision Transformer 기반의 알고리즘을 제안하여 Pegasus layer 포맷을 기반으로 자율주행차 안전 평가를 위한 기능적 시나리오를 추출하였다. So et al.(2019)은 한국 경찰청 교통사고 데이터를 수집하여 빅데이터 분석 기법 중 하나인 텍스트 가중치 분석을 활용하여 자율주행차 안전성 평가 시나리오를 도출하였다. Guo et al.(2024)는 NHTSA 사고 데이터를 기반으로 날씨, 도로 유형, 조명, 차량 행동, 충돌 각도 등 사고정보를 추출하고 유튜브에서 통계적으로 드물지만 극단적이고 예외적인 사고 상황을 수집하여 시나리오를 구축하였고, Li et al.(2024)은 중국 NAIS 사고 데이터를 현실에서 실제로 발생한 위험 조건을 담고 있음을 전제로 해당 위험 조건을 중요도 샘플링과 조합 테스트를 통해 시나리오를 구축하였다. 그 외에도 Kibalama and Tulpule(2022)는 NHTSA 사고 데이터를 통해 정면충돌, 후방충돌, 앞지르기 사고, 차선유지 실패 등 12가지 ADS 제어 목적 범주를 분류하고 이에 대한 시나리오를 구축하였으며, Chung(2024)는 도로교통공단 사고통계와 경찰청 사고 현장보고서를 수집하여 가해·피해 차량 유형, 충돌방향, 위치, 사고 유형, 운전 행태 요소 등을 검토하여 시나리오 도출했다.

4. 보행자, 이륜차, 자전거, 전동킥보드 시나리오 구축 연구

다음으로 보행자, 이륜차, 자전거, 전동킥보드와 같은 객체 기반 시나리오를 구축한 연구에 대해 검토하였다. Rampf et al.(2023)은 자율주행차와 자전거 이용자 간의 상호작용을 현실적으로 시나리오화하기 위해 자전거 시뮬레이터 데이터를 활용한 자율주행차 시뮬레이션 모델을 제안하였다. SUMO 시뮬레이터를 기반으로 차량의 인식, 경로 계획, 제어 모듈을 구성하고, 교차로 회피, 추월 등 네 가지 주행 시나리오를 구현하였다. Wang et al.(2022)는 자율주행차의 위험 대응 능력을 평가하기 위해 중국 내 일반차량-이륜차 충돌 사고 239건의 심층 데이터를 바탕으로 시나리오를 구축하였다. Cao et al.(2019)은 중국 자동차사고 심층 연구(NAIS) 데이터베이스를 활용하여 승용차와 이륜차 간 충돌을 대상으로 클러스터 분석과 통계 검증을 통해 반복적인 충돌 시나리오를 도출하였다. Sui et al.(2019)은 중국에서 2013년 기준 이륜차와 자전거가 도로교통 사망자 중 가장 높은 비율을 차지한다는 점을 언급하며, CIDS(China In-Depth Accident Study) 데이터를 활용하여 승용차와 이륜차, 자전거 간 충돌 사고 672건을 기반으로 주요한 다섯가지 시나리오를 도출했다. 아울러, 국내에서는 Baek et al.(2024)은 iGLAD(the initiative for the Global harmonization of Accident Data)가 구축한 사고 데이터 중 보행자 사고를 분류하여 특성을 확인하여 시나리오를 구축하고 시뮬레이션 수행하여 사고 예방 시스템의 성능 평가를 진행했다.

5. 선행연구 검토 결과 및 본 연구의 차별성

다음으로 본 연구 관련 Edge Case 및 Pegasus 5-layer의 개념과 시나리오 구축 관련 선행연구 검토 결과를 정리하였다. 먼저, Edge Case는 일반적인 도로 조건이나 학습 데이터에서 벗어난 복잡하고 드문 상황으로, 자율주행 시스템의 설계 운행 조건(ODD)을 초과하여 발생하는 예외적 시나리오를 의미한다. 이러한 상황은 발생 가능성은 낮지만 사고로 이어질 경우 높은 위험을 초래할 수 있기 때문에, 자율주행차의 안전성 확보를 위해 반드시 고려되어야 하는 시나리오로 강조되고 있다. Edge Case는 단일 요인에 국한되지 않고, 날씨, 조도, 노면 상태, 객체의 돌발적인 움직임, 운전자 주의력 저하 등 다양한 요소의 복합적 상호작용으로 구성되며, 실제 주행환경에서 예측이 어렵다는 특징을 가진다.

다음으로, Pegasus 5-layer 포맷은 독일 Pegasus 프로젝트에서 제안된 자율주행차 시나리오 평가 프레임워크로, 도로 기하구조, 교통 인프라, 일시적 환경, 동적 객체, 외부환경의 다섯 계층으로 구성된다. 이 구조는 실제 도로에서

발생 가능한 다양한 변수들을 계층적으로 정리할 수 있어, 자율주행차의 안전성 평가 시나리오를 정형화하는 데 효과적인 도구로 활용되어 왔다.

시나리오 구축과 관련된 선행연구는 경찰청 교통사고 데이터, 캘리포니아 DMV, GIDAS, NAIS 등 자율주행차 및 일반 차량 사고 데이터를 활용하여 전형적 사고 유형에 대한 시나리오를 구축한 사례가 다수 존재한다. 그러나 이들 연구는 사고 발생 전후의 도로 상태, 운행 환경, 교통 혼잡 등 복잡한 조건을 충분히 반영하지 못하였고, 자율주행차 사고 데이터의 표본 수가 제한적이어서 Edge Case 수준의 정밀한 시나리오 구성에는 한계가 있었다.

한편, 보행자, 이륜차, 자전거, 전동킥보드와 같은 비정형 도로 객체와의 상호작용을 다룬 일부 연구도 존재하지만, 이들 대부분은 단일 객체와의 단순 충돌 또는 교차로 등 제한된 상황에 집중되어 있었다. 복잡한 도시 환경에서 이들 객체가 보이는 불규칙적인 움직임이나 돌발적 진입과 같은 특성을 다계층적으로 재현하고 분석한 연구나 자율주행차와 해당 객체가 혼재된 상황의 다양한 조건을 구체적으로 구현한 시나리오 구축 연구가 다소 부족했던 것으로 나타났다.

이에 따라 본 연구에서는 Edge Case의 개념적 중요성을 바탕으로, 사고 발생상황을 효과적으로 구조화하기 위한 프레임워크로 Pegasus 5-layer를 사용하고자 한다. 또한, 시나리오 구축 과정에서는 한문철TV에서 수집한 블랙박스 사고 영상 데이터를 활용하여 자율주행차와 보행자, 이륜차, 자전거, 전동킥보드가 혼재된 복잡한 교통상황에서 사고 전후의 도로 기하구조, 주행환경, 주·야간, 기상 조건, 외부환경 등 다양한 평가 요소를 Pegasus 5-layer에 따라 계층적으로 분류하고, 이를 기반으로 LDA 토픽모델링 기법을 적용하여 자율주행차 안전성 평가를 위한 Edge Case 시나리오 그룹을 구축하였다.

연구방법론

1. 적용 방법론 검토

최근 방대한 텍스트 데이터를 분석하는 기법으로 일일이 직접 분석을 진행하던 전통적인 분석방법의 대안으로 떠오르고 있는 것이 텍스트 마이닝이다. 텍스트마이닝은 자연어 처리 기술을 통해 다양하고 방대한 텍스트 데이터에서 특정 주제와 관련된 패턴과 주제를 추출하여 의미와 분포를 도출해주는 텍스트 분석기법이다. 텍스트 마이닝 기법 중 토픽 모델링(Topic Modelling) 기법은 구조화되지 않은 방대한 문서 집합을 대상으로 토픽을 찾아내기 위하여 유사한 의미가 있는 단어들을 군집하는 방식을 사용하여 주제를 추론하는 방법으로 많이 활용되고 있다. 토픽 모델링은 Figure 2과 같이 비구조화된 문서 집합에서 확률적 알고리즘을 활용하여 문서 내 단어의 사용 패턴을 분석하고, 이를 기반으로 해당 문서가 포함하고 있는 주제를 추출하는 기법으로 문서 집합 내에서 주제를 추출할 뿐만 아니라 개별 주제들이 어떻게 연결되어 있는지 분석할 수 있는 기법이다(Blei, 2012).

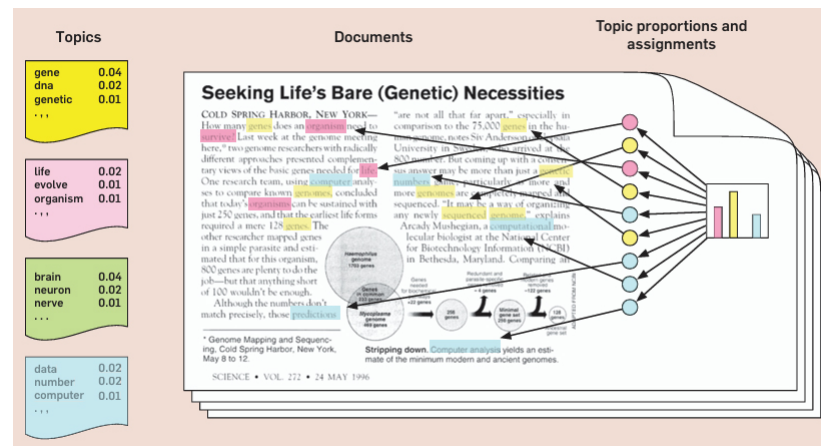


Figure 2. The intuitions behind LDA (Blei, 2012)

본 연구에서는 수집된 데이터들의 주요 사고요인 토픽들을 추출하여 시나리오를 구축하고자 Latent Dirichlet Allocation(LDA) 기법을 사용했다. LDA는 확률적 토픽모델링 기법 중 가장 보편적으로 사용하는 기법 중 하나로 문서 집합의 단어 뭉치에서 숨겨진 토픽 구조를 찾아내기 위한 확률적 기계 학습모형이다(Blei, 2012). LDA 토픽모델링은 데이터 내의 관측 변수를 통해 잠재 변수를 추론하여 전체 데이터에 토픽과 토픽의 할당 비율, 토픽 내 키워드 분포를 확인할 수 있어 기존 시각화 도구와 달리 토픽 내 키워드 분포를 확인할 수 있으며, 그 키워드를 기반으로 연구자가 해당 토픽을 해석하고 정의할 수 있게 함에 따라 비구조화된 텍스트 데이터를 분석하는 데에 많이 사용된다. Figure 3는 LDA 토픽모델링 알고리즘 과정을 도식화한 것이다. 본 연구에서 LDA 토픽모델링이 적용되는 과정은 다음과 같다. 먼저 블랙박스 사고 영상과 사고 정보 데이터를 텍스트화하여 각각의 사고를 하나의 문서(D)로 구성하였다. 다음으로 사고 관련 키워드를 전처리한 후 Pegasus 5-layer에 따라 단어 시퀀스($w_{d,n}$)로 구성한다. 이후 LDA 토픽모델링을 통해 각 문서가 어떤 시나리오 토픽에 속하는지를 나타내는 문서-토픽 분포(θ_d)와 각 단어가 어떤 토픽에서 생성되었는지를 나타내는 토픽 할당 변수($z_{d,n}$)를 학습하여 최종적으로 각 시나리오 주제에 내재된 사고 키워드 집합인 토픽-단어 분포(β_k)를 도출하는 과정으로 이루어진다. 마지막으로 도출된 토픽과 단어를 기반으로 사고의 주요 패턴과 유형을 식별하고 Pegasus Layer를 기준으로 재분류하여 시나리오를 구축하였다.

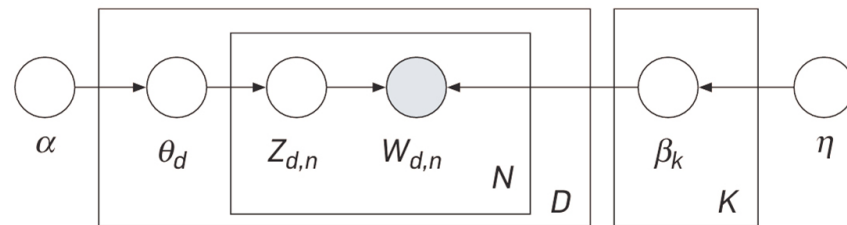


Figure 3. Graphical model of LDA (Blei, 2012)

이러한 방대하고 비정형적으로 구성된 텍스트 데이터를 다루고 분석하는 방법으로 LDA 토픽모델링 기법을 활용한 연구가 다수 진행되었다. Park et al.(2021b)은 미국 캘리포니아 차량관리국에서 제공하는 자율주행차 사고 리포트 데이터를 수집하여 텍스트 임베딩, LDA 토픽모델링 기법을 활용하여 자율주행차 교통사고의 주요 키워드를 도출하여 사고 분석에 관한 연구를 진행하여 사고요인 분석에 토픽모델링 기법을 활용할 수 있음을 시사하였다. Jo et al.(2022)는 서울시 한양도성을 중심으로 소셜 미디어 플랫폼 정보를 수집하여 여가통행 행태를 추정하였으며, 분석 결과 사고, 운동, 관광, 휴양 등 4개의 여가통행 유형을 도출하였다. Oh(2015)는 교통 및 ICT 분야의 교통 관련 논문의 텍스트 데이터를 기반으로 동시출현단어 분석과 LDA 토픽모델링 기법을 활용하여 연구 동향을 분석하였고 활성화화를 위한 필요 연구영역을 검토하였다. 이와 같이 LDA 토픽모델링 기법은 자율주행차 교통사고 키워드 도출, 여가통행 목적 추정, 교통 연구동향 파악 등 교통 분야 연구에서도 다양하게 활용되고 있다.

2. 자율주행차 안전성 평가 시나리오 구축 방법

자율주행차 사고 데이터를 활용하면 자율주행차의 기능 및 특성에 기반한 시나리오를 구축할 수 있으나, 현재 해당 데이터는 그 양과 접근이 제한적이라는 한계가 있다. 이에 따라, 자율주행차가 인간 운전 차량뿐만 아니라 보행자, 이륜차, 자전거, 전동킥보드 등과 함께 혼재된 교통 환경에서 운행한다는 점을 고려할 때, 기존 일반 차량이 경험한 위험 상황이 자율주행차에도 유사하게 발생할 가능성이 높은 것으로 판단된다. 이러한 배경 하에, 다수의 선 행연구에서는 자율주행차 안전성 평가 시나리오 구축을 위해 일반 차량의 사고 데이터를 활용하였다. Park et al.(2021a)와 So et al.(2019)은 한국 경찰청 교통사고 데이터를 활용하여 자율주행차 안전성 평가 시나리오를 구축하였으며, 특히, So et al.(2019)은 일반차량 사고 데이터를 기반으로 텍스트 가중치 분석을 통해 도출한 자율주행차 시나리오

가 실제 사고를 얼마나 설명하는지 검증하여 해당 시나리오의 유효성을 확인하였으며, GIDAS, iGLAD와 같은 국제적인 사고 데이터베이스에도 적용 가능하다고 제안하였다. 또한, Guo et al.(2024), Li et al.(2024), Kibalama and Tulpule(2022), Chung(2024) 등 국내·외 많은 연구에서 미국 NHTSA, 중국 NAIS, 한국 경찰청 사고 데이터와 같은 일반차량 사고를 기반으로 사고 당시의 도로 환경과 객체의 움직임 등을 조합하여 자율주행차 사고 시나리오를 구축한 연구가 다수 있다.

이에, 본 연구에서도 자율주행차 시나리오 구축을 위해 일반차량 교통사고 블랙박스 영상을 분석에 활용하였다. 또한, 드물지만 극단적이고 예외적인 사고 상황을 고려한 Edge Case 시나리오 상황을 중점적으로 검토하기 위해 Guo et al.(2024)의 연구와 같이 유튜브에서 특이사고 자료를 수집하였으며, 본 연구에서는 일반적으로 발생하기 어려운 교통사고 사례와 그 과실 비율을 설명하는 한문철 TV의 사고 블랙박스 영상을 분석대상으로 선정하였다. 한문철 TV는 블랙박스 영상으로 사고 전후 상황을 구체적으로 확인할 수 있었으며, 사고일시, 도로선형, 도로경사, 차로수, 발생지점, 교통 방해요소, 도로환경, 사고 직전 차량조작, 사고유형, 주간/야간, 날씨 등 다양한 환경 정보를 구체적으로 포함하고 있어 사고를 유발한 요인을 상세히 확인할 수 있어 복잡하고 다양한 환경을 반영한 시나리오를 구축할 수 있다. 그리하여 먼저 차대 보행자, 차대 이륜차, 차대 자전거, 차대 전동킥보드 사고 블랙박스 영상을 수집하고 도로 상태, 운행상황, 환경요인 등과 같은 사고 정보 및 설명자료를 텍스트로 추출하였다. 다음으로 시나리오 구성 요소를 선정하기 위해 수집된 데이터를 기반으로 자율주행차 안전성 평가에 적용하기 위해 Pegasus 5-layer 포맷에 근거하여 데이터 항목을 정의하고, 추출한 사고 정보 텍스트의 불용어와 잉여 키워드를 제거하는 텍스트 전처리 과정을 진행하였다. 다음으로 텍스트 마이닝을 실시하여 차량 운행 중 외부환경으로 인해 발생한 사고요인 키워드 추출하였고, 도출된 키워드를 LDA 토픽모델링 기법을 활용하여 Pegasus 5-layer 포맷의 layer를 기준으로 재분류하여 자율주행차와 보행자, 이륜차, 자전거, 전동킥보드 간 Edge Case 시나리오를 구축하였다. 전체적인 시나리오 구축 흐름은 Figure 4과 같다.

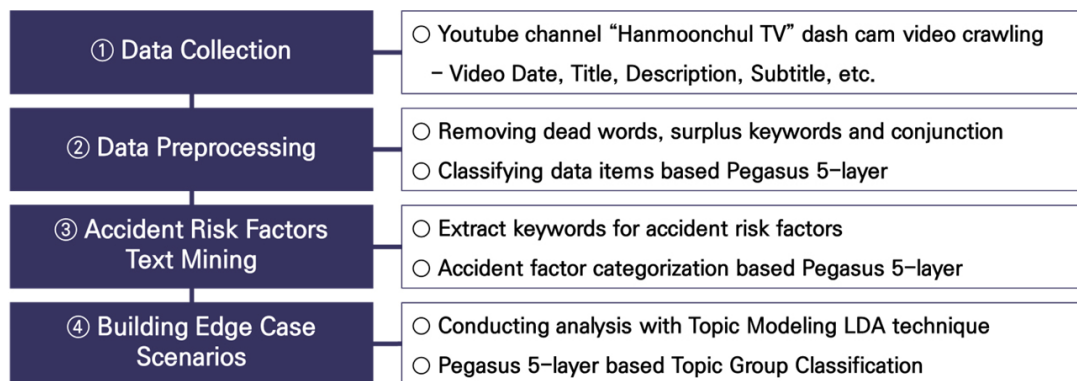


Figure 4. Research process

1) 데이터 수집 및 전처리

본 연구는 Level 3 이상의 자율주행차가 일반차량, 보행자, 이륜차 등 다양한 객체가 혼재된 교통 환경에서 Edge Case를 예측하지 못해 발생하는 위험 상황에 대처하기 위해 복잡성과 사고의 위험성이 높으나 발생 빈도가 낮아 예측이 어려운 사고 사례를 중심으로 대표적인 시나리오를 설계하고자 한다. 이를 위해 일반적으로 발생하기 어려운 교통사고 사례와 그 과실 비율을 설명하는 채널인 유튜브 채널 '한문철 TV' 블랙박스 영상을 분석대상으로 선정하였다. 이 채널에서 다루는 교통사고 상황은 자체적으로 Edge Case 성격을 지니고 있으며, 그 중에서도 이동 경로나 속도가 일정하지 않고 갑작스럽게 방향을 바꾸거나 멈추는 경우가 많아 예측하기 어려운 움직임을 보이는 보행자, 이륜차, 자전거, 전동킥보드 사고 영상을 활용하였다. 이러한 블랙박스 영상은 사고 전후 상황과 사고 당시의 도로 조

건, 환경 요인, 날씨, 조도 등 다양한 정보를 제공해 교통사고 상황 분석에 유용하다. 추가적인 진행자의 사고 설명을 통해 정형화된 사고 데이터만으로 파악하기 어려운 교통사고 맥락을 심층적으로 분석할 수 있는 장점이 있다.

분석대상을 선정한 후, 시나리오 구축에 앞서 시나리오 구성요소 선정을 위해 교통사고 위험요인 키워드에 대한 항목 정의를 진행하였다. 각각의 데이터 항목은 Pegasus 5-layer 포맷에 근거하여 Table 2와 같이 선정하였으며 블랙박스 영상과 사고 설명에서 확인할 수 있고 사고 전후 및 당시의 상황을 적절히 설명하고 추출할 수 있는 항목을 선정하였다. 실제 자율주행차가 대응 가능한 구체적인 시나리오 도출을 위해 사고일시, 도로선형, 도로경사, 차로수, 발생지점, 교통 방해요소, 도로환경, 사고 직전 차량조작, 사고유형, 주간/야간, 날씨 등을 포함하여 항목을 설계하였다.

Table 2. Dash cam data collection components

Layer	Data	Components
-	Accident situation	Records of the accident situation (ex. A vehicle traveling in two lanes changes into one lane without a turn signal, causing an accident)
Layer 1	Road alignment	Flat alignment (ex. Straight, left curve, right curve)
	Road slope	Lateral slope, longitudinal slope (ex. Flat, uphill, downhill)
Layer 2	Lane	Number of road lanes (ex. 4 lanes round trip, 3 lane one way)
	Accident point	Point type (ex. General road, side street, crosswalk)
Layer 3	Traffic flow	Reasons for congestion (ex. Illegal parked)
	Road environment	Road surface condition, work status, etc. (ex. Moisture road, no work)
	Main cause of accident	Cause category of accident (ex. Sudden lane change, reverse driven, neglect of attention, etc.)
Layer 4	Accident type	Accident type (ex. Front-end collision, side-end collision, rear-end collision)
	Maneuvering the vehicle before an accident	Vehicle driver behavior (ex. Steering wheel excessively, sudden stops, etc.)
	Accident subject type	Accident target, collision site, etc. (ex. Car to pedestrian, car to bicycle, etc.)
Layer 5	Day/night	Time information (ex. Daytime, night)
	Weather	Weather information (ex. Sunny, rainy, snowy)

다음 단계로 Youtube API를 이용하여 ‘한문철 TV’의 차 대 보행자, 이륜차, 자전거, 전동킥보드 사고 블랙박스 영상을 필터링하였다. 각 영상의 게시일자, 제목, URL 등 영상 기본정보를 수집하고 사고 정보와 진행자의 사고 설명이 포함된 영상 설명과 자막을 웹 크롤링하여 텍스트 데이터를 수집하였다.

수집된 데이터는 비정형적인 자연어 형태로 구성되어 있으므로 전처리 과정을 수행하였다. 교통사고 발생상황 및 원인, 시나리오 구축과 관련이 없어 분석 과정에서 크게 의미가 없는 조사, 어간, 어미 등 불용어, 잉여 키워드, 접속사를 제거하였다. 또한, 동사와 서술어를 명사화하고 동일한 의미를 가지고 있는 문구를 공통 용어로 변환하였다. 이후 Pegasus 5-layer 포맷을 기준으로 정의한 사고 위험요인 키워드 항목에 따라 분류하여 구조화된 데이터셋을 구축하였다.

데이터셋 구축 결과, 본 연구는 2021년 9월부터 2023년 3월까지 약 2년간 ‘한문철 TV’에 게시된 승용차 기준의 보행자, 이륜차, 자전거, 전동킥보드 관련 사고 정보 및 진행자 설명자료 데이터를 2,410건을 구축하였다. 구축된 데이터의 주요 기초통계는 Table 3과 같이 정리되어 있으며, 사고 유형 별로는 차대 이륜차(1,020건), 보행자(732건), 자전거(592건), 전동킥보드(56건) 순으로 많이 발생한 것으로 나타났다. 주요 사고 발생지점은 일반도로(1,614건), 횡단보도(464건), 골목길(198건) 등에서 주로 발생하였으며, 주요 사고원인으로는 급진입(701건), 시거불량(536건), 무단횡단(315건) 등으로 확인되었다. 영상에서 사고 원인이 명확히 확인되지 않은 사고에 대해서는 주요 사고 원인을 별도로 집계하지 않았다.

Table 3. Basic statistics of collecting data

Components	Case	Number of case	Ratio
Accident type	Car to motorcycle	1,020	42%
	Car to pedestrian	732	30%
	Car to bicycle	592	25%
	Car to e-scooter	56	2%
	Car to etc.	10	1%
Accident point	General road	1,614	67%
	Crosswalk	464	19%
	Side street	198	8%
	Unsignalized crosswalk	80	3%
	Parking lots	39	2%
	Other	15	1%
Main cause of accident	Sudden enter	701	31%
	Poor sight distance	536	24%
	Jaywalking	315	14%
	Signal violation	216	9%
	Reversing	179	8%
	Sudden lane change	176	8%
	Centerline encroachment	69	3%
	Sudden stop	43	2%
	Other	23	1%

2) 사고요인 키워드 텍스트마이닝

전처리가 완료된 블랙박스 텍스트 데이터셋을 기반으로 시나리오의 주요 키워드를 도출하기 위해 의미 있는 단어 단위로 분절하고 토큰화(Tokenization)를 수행하였다. 이를 통해 교통사고의 주요 원인을 설명할 수 있는 사고요인 키워드를 추출하였다.

이를 위해 텍스트 마이닝은 Python 기반의 한국어 자연어처리 라이브러리인 KoNLPy를 활용하여 수행하였다. KoNLPy는 Java 기반 형태소 분석 엔진을 Python 환경에서 활용할 수 있도록 지원하는 인터페이스로, Hannanum, Kkma, Komoran, Mecab, Okt 등의 형태소 분석 클래스를 포함한다. 본 연구에서는 형태소 사전 편집이 용이하다는 장점을 가진 KAIST SWRC 개발의 Hannanum 분석기를 활용하였다. 해당 분석기를 통해 추출된 주요 사고요인 키워드는 앞서 정의한 데이터 항목에 따라 Pegasus 5-layer 포맷의 layer 기준에 따라 분류하였다.

3) LDA 토픽모델링을 통한 시나리오 구축

Pegasus 5-layer 포맷을 기준으로 분류된 사고요인 키워드를 기반으로 Edge Case 시나리오를 구축하기 위해 LDA 토픽모델링 기법을 적용하였다. LDA 토픽모델링 분석을 통해 데이터 내에 사고 그룹을 생성하기에 앞서 분석을 통해 도출된 각각의 토픽들이 유의미한 결과값을 가지고 독립성을 확보할 수 있도록 적정한 토픽 그룹 수를 설정하였다.

토픽 그룹 수 설정은 각 토픽 내 단어들이 얼마나 일관되게 나타나는지를 평가하는 Coherence score를 기준으로 결정하였다. Coherence score는 그룹 내 키워드 간의 연관성을 수치화한 지표로 그 값이 높을수록 주제 내 일관성이 우수함을 의미하며, 최고점을 보이는 값이 최적의 토픽 수로 해석된다.

이렇게 설정된 최적의 토픽 그룹 수를 기반으로 실도로에서 자율주행차가 보행자, 이륜차, 자전거, 전동킥보드 등 다양한 객체와 혼재한 상황에서 마주할 수 있는 외부 사고를 반영하여, 주사고 유형(Accident type)을 기준으로 사고를 구분하여 LDA 토픽모델링 분석을 수행하였다.

분석 결과, 각 사고 유형별로 도출된 토픽 그룹의 상위 30개 키워드의 토픽과 키워드의 상호 관계를 파악하여 키워드 분포, 유사성 등을 분석하였다. 이를 통해 사고 상황을 대표하고 논리적 오류가 없는 키워드를 추출하고

Pegasus 5-layer 항목에 따라 다시 재배치하여 시나리오 그룹을 구축하였다. 이렇게 정리된 시나리오 그룹별 키워드를 기반으로 실제 사고 발생 흐름에 맞추어 도로구조, 환경요인, 사고원인, 차량조작, 사고유형 등의 논리적인 순서로 키워드를 조합하고 서술형으로 연결하여 시나리오를 구성하였다. 이러한 방식으로 계속해서 수행하여 각 토픽 그룹에서 추출된 키워드를 기반으로 Pegasus 5-layer 항목에 따라 재구성한 뒤 해당 사고의 특성과 대표성, 실현 가능성 등을 고려하여 최종 Edge Case 시나리오로 도출하였다.

시나리오 구축 결과

1. LDA 토픽모델링 분석

1) 보행자 사고 시나리오

자율주행차와 보행자 간 발생할 수 있는 사고 시나리오 구축을 위해 수집된 블랙박스 데이터셋 중 보행자와 발생한 사고 데이터를 기반으로 LDA 토픽모델링 분석을 진행하였다.

먼저, 적정 토픽 그룹 수를 설정하기 위해 Coherence Score를 확인하였다. Figure 5와 같이 도출된 그래프에서 토픽 그룹 수 4개가 Coherence Score가 가장 높게 나타남에 따라 분석을 위한 토픽 그룹 수를 4개로 설정하였다. 다음으로 LDA 토픽모델링 기법을 활용하여 보행자 사고에 대해 분석을 진행하였고, 시나리오 그룹과 layer 항목별 키워드를 Table 4과 같이 도출했다.

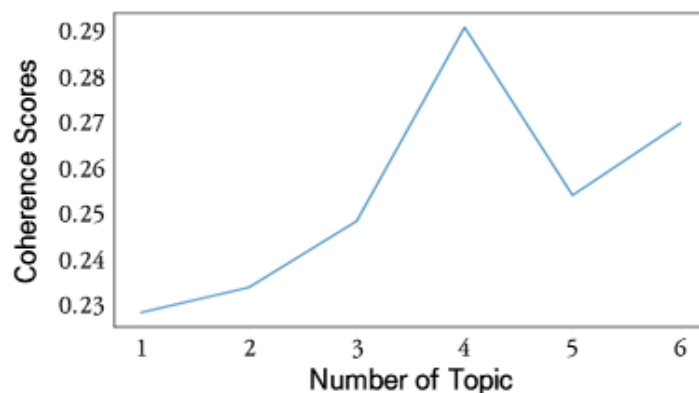


Figure 5. Pedestrian accident topic coherence score

Table 4. Pedestrian accident edge case scenario topic group

Layer	Components	Topic group 1	Topic group 2	Topic group 3	Topic group 4
Layer 1	Road alignment	1. <u>Straight</u> , 2. Right curve	1. Straight, 2. <u>Left curve</u>	1. <u>Right curve</u> , 2. Straight, 3. Left curve	<u>Straight</u>
	Road slope	<u>Downhill</u>	–	<u>Uphill</u>	<u>Uphill</u>
Layer 2	Lane	<u>10 lanes round trip</u>	<u>2 lanes round trip</u>	–	<u>4 lanes round trip</u>
	Accident point	1. General road, 2. Parking lots, 3. <u>Crosswalk</u>	1. General road, 2. <u>Crosswalk</u> , 3. Unsignalized crosswalk	1. <u>Crosswalk</u> , 2. Side street, 3. Unsignalized crosswalk, 4. Road within Apartment, 5. Parking lots, 6. General road	1. General road, 2. Crosswalk, 3. <u>Unsignalized crosswalk</u>

Table 4. Pedestrian accident edge case scenario topic group j(continued)

Layer	Components	Topic group 1	Topic group 2	Topic group 3	Topic group 4
Layer 3	Traffic flow	1. No obstacles, 2. Parked vehicles, 3. <u>Traffic congestion</u> , 4. Illegal parked	<u>No obstacles</u>	<u>No obstacles</u>	1. <u>Parked vehicles</u> , 2. Traffic congestion
	Road environment	1. Normal, 2. <u>Moisture road</u>	<u>Normal</u>	Normal	Normal
	Main cause of accident	1. <u>Poor sight distance</u> , 2. Negligence to look ahead, 3. Centerline encroachment	1. <u>Jaywalking</u> , 2. <u>Sudden enter</u>	<u>Poor sight distance</u>	1. Poor sight distance, 2. <u>Sudden enter</u> , 3. Jaywalking, 4. Signal Violation
Layer 4	Accident type	1. <u>Collision</u> , 2. Open door accidents	1. Collision, 2. <u>Side-end Collision</u> , 3. Rear-end Collision	<u>Collision</u>	<u>Collision</u>
	Maneuvering the vehicle before an accident	1. <u>Driving</u> , 2. Sudden stop, 3. Sudden braking	1. <u>Driving</u> , 2. Sudden stop	<u>Driving</u>	<u>Driving</u>
	Accident subject type	<u>Car to Pedestrian</u>	<u>Car to Pedestrian</u>	<u>Car to Pedestrian</u>	<u>Car to Pedestrian</u>
Layer 5	Day/night	<u>Night</u>	1. <u>Night</u> , 2. Daytime	<u>Night</u>	–
	Weather	<u>Rainy</u>	–	–	–

토픽 그룹 1은 ‘내리막’, ‘왕복 10차로’, ‘교통 혼잡’, ‘시거불량’, ‘급정거’ 등, 토픽 그룹 2은 ‘횡단보도’, ‘무단횡단’, ‘측방 추돌’, ‘야간’ 등, 토픽 그룹 3은 ‘커브’, ‘오르막’, ‘횡단보도’, ‘골목길’, ‘시거불량’ 등, 토픽 그룹 4은 ‘직진구간’, ‘비신호 횡단보도’, ‘주차된 차량’, ‘급진입’ 등과 같은 키워드가 도출되었다. 도출된 키워드를 기반으로 사고 발생 흐름에 따라 논리적인 시나리오 조합을 구성하였고, 이를 바탕으로 최종 자율주행차-보행자 Edge Case 시나리오를 설계하였다. 시나리오 구축에 사용된 layer 항목별 키워드는 Table 4과 같이 굵은 글씨로 강조하였다.

2) 이륜차 사고 시나리오

다음으로 자율주행차 대 이륜차 사고 시나리오 구축을 위해 수집된 블랙박스 데이터셋 중 이륜차와 발생한 사고 데이터를 기반으로 LDA 토픽모델링 분석을 수행하였다. 보행자 사고와 같이 적정 토픽 그룹 수를 설정하기 위해 Coherence Score를 확인하였다. Figure 6와 같이 도출된 그래프에서 토픽 그룹 수 4개가 Coherence Score가 가장 높게 나타남에 따라 분석을 위한 토픽 그룹 수를 4개로 설정하였다. 다음으로 LDA 토픽모델링을 활용하여 보행자 사고에 대해 분석을 진행하였고, Table 5과 같이 시나리오 그룹과 layer 항목별 키워드를 도출했다. 토픽 그룹 1은 ‘편도 1차로 도로’, ‘횡단보도’, ‘주차된 차량’ 등, 토픽 그룹 2은 ‘골목길’, ‘불법주차’, ‘시거불량’, ‘역주행’ 등, 토픽 그룹 3은 ‘편도 3차로 도로’, ‘일반도로’, ‘급차선변경’, ‘신호위반’ 등, 토픽 그룹 4은 ‘전방주시 태만’, ‘안전거리 미확보’, ‘급제동’ 등과 같은 키워드가 도출되었다. 도출된 키워드를 바탕으로 시나리오 구축을 위한 키워드를 조합하여 시나리오를 최종 구축하고자 하며, 시나리오 구축에 사용된 layer 항목별 키워드는 Table 4과 같이 굵은 글씨로 강조하였다.

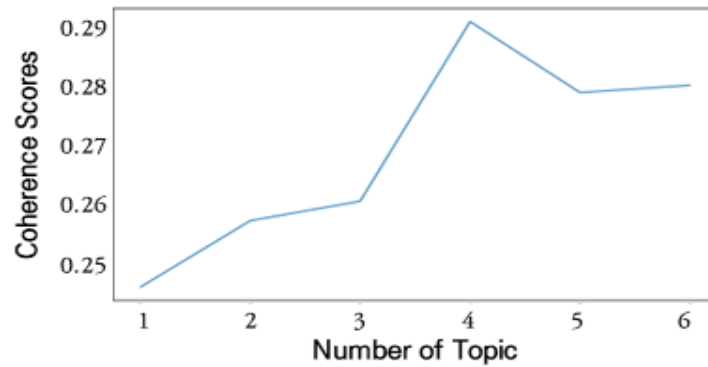


Figure 6. Motorcycle accident topic coherence score

Table 5. Motorcycle accident edge case scenario topic group

Layer	Components	Topic group 1	Topic group 2	Topic group 3	Topic group 4
Layer 1	Road alignment	1. <u>Straight</u> , 2. Left curve	1. Right curve, 2. <u>Straight</u> , 3. Left curve	1. Straight, 2. <u>Left curve</u>	1. <u>Straight</u> , 2. Left curve
	Road slope	–	–	–	–
Layer 2	Lane	<u>1 lane one way</u>	1. <u>2 lanes round trip</u> , 2. 4 lanes round trip	<u>3 lane one way</u>	<u>1 lane one way</u>
	Accident point	1. <u>Crosswalk</u> , 2. Unsignalized Crosswalk, 3. General road, 4. Parking lots, 5. Roundabout	1. General road, 2. <u>Side street</u> , 3. Rodeside	1. <u>General road</u> , 2. Parking lots, 3. Road within Apartment	1. <u>General road</u> , 2. Intersection, 3. Side street
Layer 3	Traffic flow	1. No obstacles, 2. <u>Parked vehicles</u>	1. No obstacles, 2. <u>Illegal parked</u>	<u>No obstacles</u>	<u>No obstacles</u>
	Road environment	<u>Normal</u>	<u>Normal</u>	<u>Normal</u>	<u>Normal</u>
	Main cause of accident	1. <u>Sudden enter</u> , 2. Poor sight distance, 3. Signal violation, 4. Directive violations	1. Poor sight distance, 2. Sudden enter, 3. <u>Reversing</u>	1. Sudden enter, 2. Sudden lane change, 3. <u>Signal Violation</u> , 4. Centerline encroachment, 5. Overspeeding, 6. Jaywalking, 7. Dilemma Zone	1. Negligence to look ahead, 2. <u>Unsafe distance</u> , 3. Signal Violation, 4. Negligence to look Rear-end
Layer 4	Accident type	1. <u>Collision</u> , 2. Open door accidents	1. Collision, 2. <u>Side-end Collision</u>	1. <u>Collision</u> , 2. Rear-end Collision, 3. Side-end Collision	1. Collision, 2. <u>Rear-end Collision</u>
	Maneuvering the vehicle before an accident	1. <u>Driving</u> , 2. Sudden stop	1. <u>Driving</u>	1. <u>Driving</u> , 2. Sudden stop	1. Driving, 2. Sudden stop, 3. <u>Sudden braking</u>
	Accident subject type	<u>Car to motorcycle</u>	<u>Car to motorcycle</u>	<u>Car to motorcycle</u>	<u>Car to motorcycle</u>
Layer 5	Day/night	–	1. Daytime, 2. <u>Night</u>	<u>Night</u>	–
	Weather	–	–	–	–

3) 자전거 사고 시나리오

자율주행차 대 자전거 사고 또한 Figure 7와 같이 Cohernece Score를 통해 토픽 그룹 수 4개로 설정하여 LDA 토픽모델링을 진행했다. 분석 결과 Table 5과 같이 시나리오 그룹과 layer 항목별 키워드를 도출했으며, 토픽 그룹 1은

‘내리막’, ‘골목길’ 등, 토픽 그룹 2은 ‘아파트 단지 내 도로’, ‘주차된 차량’ 등, 토픽 그룹 3은 ‘불법주차’, ‘무단횡단’ 등, 토픽 그룹 4은 ‘비신호 횡단보도’, ‘공사중’ 등과 같은 키워드가 도출되었다. 시나리오 구축에 사용된 layer 항목 별 키워드는 Table 6과 같이 굵은 글씨로 강조하였다.

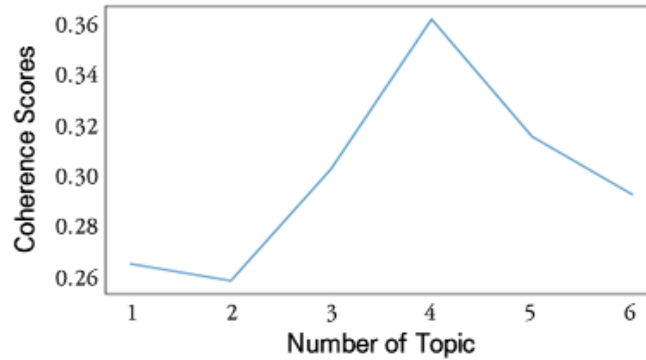


Figure 7. Bicycle accident topic coherence score

Table 6. Bicycle accident edge case scenario topic group

Layer	Components	Topic group 1	Topic group 2	Topic group 3	Topic group 4
Layer 1	Road alignment	1. Straight , 2. Right curve, 3. Left curve	1. Right curve , 2. Straight, 3. Left curve	Straight	1. Straight, 2. Left curve , 3. Right curve
	Road slope	Downhill	–	–	–
Layer 2	Lane	1. 1 lane one way , 2. 2 lanes round trip	–	1. 6 lanes round trip , 2. 2 lanes round trip	1. 4 lanes round trip , 2. 6 lanes round trip
	Accident point	1. General road, 2. Side street , 3. Parking lots, 4. Road within Apartment, 5. Crosswalk	1. General road, 2. Crosswalk, 3. Side street, 4. Road within Apartment , 5. Parking lots	1. General road, 2. Bicycle lane, 3. Crosswalk	1. Crosswalk, 2. Unsignalized crosswalk , 3. General road
Layer 3	Traffic flow	No obstacles	1. No obstacles, 2. Parked vehicles	1. No obstacles, 2. Illegal parked	1. No obstacles 2. Working on Construction
	Road environment	Normal	Normal	Normal	Normal
	Main cause of accident	1. Sudden enter , 2. Negligence to look ahead, 3. Reversing, 4. Jaywalking, 5. Poor sight distance, 6. Signal Violation, 7. Sudden lane change	1. Sudden enter , 2. Reversing, 3. Poor sight distance, 4. Sudden lane change, 5. Signal Violation, 6. Directive Violations	1. Jaywalking , 2. Poor sight distance, 3. Centerline encroachment, 4. Sudden enter, 5. Negligence to look ahead, 6. Reversing	1. Signal Violation, 2. Jaywalking, 3. Poor sight distance , 4. Sudden enter
Layer 4	Accident type	Collision	Collision	1. Collision , 2. Open door accidents, 3. Rear-end Collision	Collision
	Maneuvering the vehicle before an accident	1. Sudden stop , 2. Driving	1. Driving , 2. Sudden braking	1. Driving	1. Driving, 2. Sudden stop
	Accident subject type	Car to bicycle	Car to bicycle	Car to bicycle	Car to bicycle
Layer 5	Day/night	–	Night	Night	1. Night 2. Daytime
	Weather	–	–	–	–

4) 전동킥보드 사고 시나리오

마지막으로 전동킥보드 분석도 Figure 8와 같이 도출된 Cohernece Score 그래프에서 토픽 그룹 수 4개로 설정하여 LDA 토픽모델링을 진행했다. 분석 결과 Table 7과 같이 시나리오 그룹과 layer 항목별 키워드를 도출했으며, 시나리오 구축에 사용된 layer 항목별 키워드는 Table 4과 같이 굵은 글씨로 강조하였다.

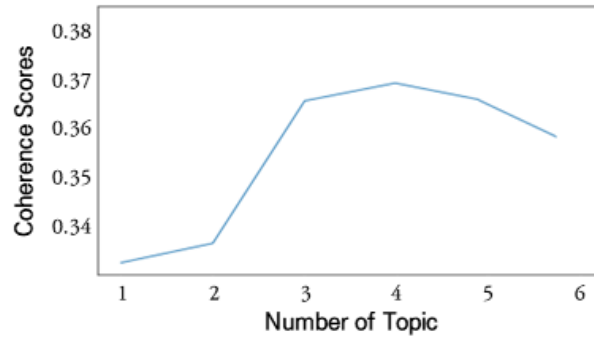


Figure 8. e-scooter accident topic coherence score

Table 7. e-scooter accident edge case scenario topic group

Layer	Components	Topic group 1	Topic group 2	Topic group 3	Topic group 4
Layer 1	Road alignment	1. Right curve, 2. Left curve, 3. Straight	1. Straight, 2. Right curve , 3. Left curve	–	1. Left curve 2. Right curve
	Road slope	–	Uphill	–	Uphill
Layer 2	Lane	1 lane one way	1. 3 lanes one way , 2. 6 lanes round trip	1. 2 lanes round trip , 2. 4 lanes round trip, 3. 10 lanes round trip, 4. 6 lanes round trip	1. 1 lane one way , 2. 4 lanes round trip
	Accident point	1. Side street , 2. General road, 3. Unsignalized crosswalk, 4. Crosswalk	1. General road, 2. Crosswalk , 3. Side street, 4. Road within apartment, 5. Parking lots	1. General road , 2. Crosswalk 3. Road within apartment	1. General road, 2. Side street
Layer 3	Traffic flow	1. No obstacles, 2. Parked vehicles	1. No obstacles, 2. Parked vehicles	No obstacles	1. No obstacles 2. Parked vehicles
	Road environment Main cause of accident	Normal 1. Sudden enter, 2. Reversing, 3. Poor sight distance , 4. Signal Violation, 5. Jaywalking, 6. Negligence to look ahead	Normal 1. Poor sight distance, 2. Reversing, 3. Sudden enter , 4. Sudden lane change, 5. Negligence to look ahead, 6. Jaywalking	Normal 1. Sudden enter, 2. Poor sight distance, 3. Centerline encroachment ,	Normal 1. Centerline encroachment, 2. Reversing, 3. Sudden enter
Layer 4	Accident type	1. Collision, 2. Open door accidents	1. Collision 2. Rear-end Collision	Collision	Collision
	Maneuvering the vehicle before an accident	1. Driving, 2. Stop	Driving	Driving	1. Driving 2. Parking
	Accident subject type	Car to e-scooter	Car to e-scooter	Car to e-scooter	Car to e-scooter
Layer 5	Day/night	1. Daytime , 2. Night	Night	1. Daytime, 2. Night	Night
	Weather	–	–	–	–

2. 최종 Edge Case 시나리오 구축 결과

최종적으로 주사고 유형(Accident Type)에 따라 승용차와 보행자, 이륜차, 자전거, 전동킥보드 간에 발생한 교통 사고 블랙박스 데이터셋을 각각 LDA 토픽모델링 기법으로 분석하여 주요 사고요인을 도출하였다. 도출된 사고요인은 Pegasus 5-layer 포맷의 항목에 따라 분류하였으며, 이를 기반으로 각 토픽 그룹별 시나리오 키워드를 구성하였다. 구성된 결과를 바탕으로 객관성과 논리성을 확보하기 위해 각 항목에서 최적의 키워드를 선별하여 최종 Edge Case 시나리오를 구축하였다.

첫 번째로 자율주행차 대 보행자 사고 Edge Case 시나리오는 다음과 같이 도출됐다.

- (1) 야간에 비로 인해 도로가 젖어 있는 왕복 10차로 도로에서 직진 주행 차량이 시거불량과 교통 혼잡으로 인해 횡단보도에 진입한 보행자와 충돌 사고 발생
- (2) 왕복 2차로 도로에서 좌회전 주행 차량이 횡단보도에서 무단횡단하는 보행자와 측면 추돌사고 발생
- (3) 오르막 도로에서 우회전 주행 차량이 시거불량으로 인해 횡단보도로 진입한 보행자와 충돌 사고 발생
- (4) 신호가 없는 횡단보도가 있는 왕복 4차로 도로 오르막 직선 구간에서 주행중인 차량이 주차된 차량 사이에서 갑자기 진입한 보행자와 충돌 사고 발생

두 번째로 자율주행차-이륜차 Edge Case 시나리오는 다음과 같이 도출됐다.

- (1) 편도 1차로 도로에서 직진 주행 차량이 주차된 차량 사이에서 갑자기 진입한 이륜차와 충돌 사고 발생
- (2) 야간에 왕복 2차로 이면도로를 직진 중이던 차량이 불법주차로 인해 중앙선을 침범하여 주행한 오토바이와 측면 추돌사고 발생
- (3) 야간에 편도 3차로 도로에서 좌회전 주행하던 차량이 신호위반 오토바이와 충돌 발생
- (4) 편도 1차로 직선 구간 도로에서 공사작업으로 인해 선행 차량이 갑자기 브레이크를 밟아, 이를 뒤따르던 오토바이의 안전거리 미확보로 인해 후미 추돌사고 발생

세 번째로 자율주행차-자전거 Edge Case 시나리오는 다음과 같이 도출됐다.

- (1) 일방통행 1차로 내리막 이면도로에서 직진 중이던 차량이 도로로 갑자기 진입한 자전거와 충돌 사고가 발생. 차량이 긴급 제동을 시도했으나 사고를 피하지 못함
- (2) 야간에 아파트 단지 내 도로에서 우회전 커브를 돌던 차량이 주차된 차량 사이에서 갑자기 진입한 자전거와 충돌 사고 발생
- (3) 야간에 불법 주차된 차량이 있는 왕복 6차로 도로를 직진 중이던 차량이 횡단보도를 무단 횡단하던 자전거와 충돌 사고 발생
- (4) 야간에 왕복 4차로 도로에서 좌회전 커브를 돌던 차량이 신호 없는 횡단보도를 건너는 자전거를 발견하고 긴급 제동을 시도했으나 충돌 사고 발생

마지막으로 자율주행차-전동킥보드 Edge Case 시나리오는 다음과 같이 도출됐다.

- (1) 주간에 가로변 주차된 차량이 있는 왕복 1차로 이면도로에서 전동킥보드가 주차된 차량의 개문을 미쳐 확인하지 못하고 시거불량으로 인해 개문사고 발생
- (2) 야간에 가로변 주차된 차량이 있는 편도 3차로 오르막 도로에서 직진 주행 중인 차량이 횡단보도에 갑자기 진입한 전동킥보드 충돌 사고 발생
- (3) 야간에 왕복 2차로 도로에서 중앙선을 침범하며 주행 중이던 차량이 전동킥보드 충돌 사고 발생
- (4) 야간에 편도 1차로 오르막 이면도로에서 좌회전 주행 하던 차량이 도로로 갑자기 진입한 전동킥보드와 충돌 사고 발생

자율주행차와 보행자, 이륜차, 자전거, 전동킥보드의 최종 Edge Case 시나리오 구축 결과는 Table 8와 같이 정리하였으며, LDA 토픽모델링 분석에 따라 도출된 키워드는 굵은 글씨로 강조하였다.

Table 8. Edge Case scenario development results

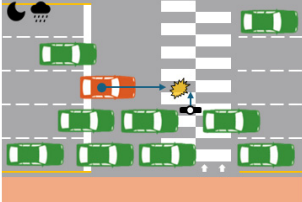
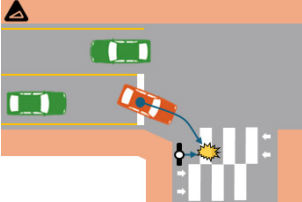
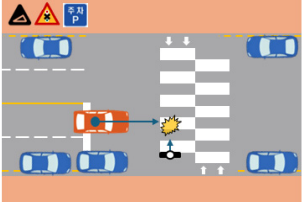
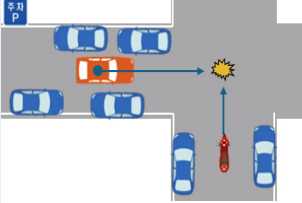
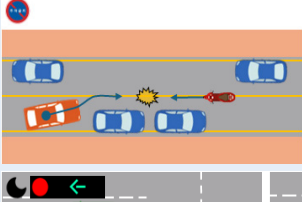
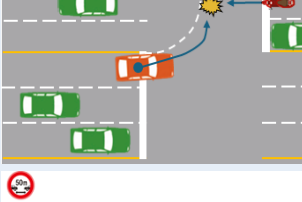
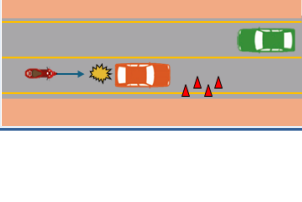
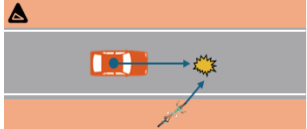
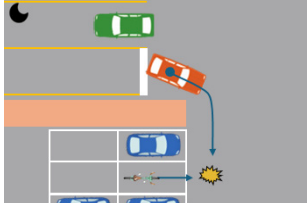
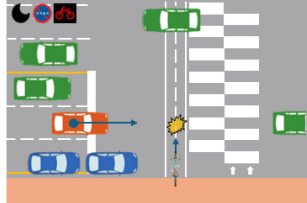
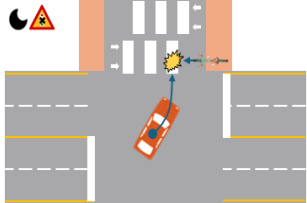
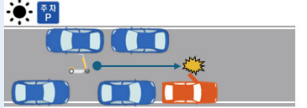
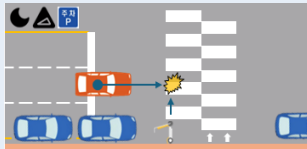
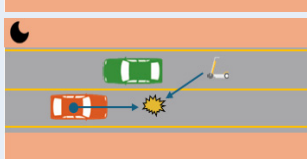
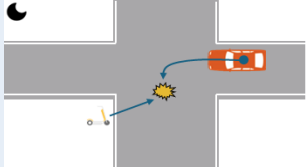
Class	No.	Contents	Scenario Situation
Pedestrian	1	A car <u>driving straight</u> on a <u>10 lanes road</u> occurs in a <u>collision</u> with a <u>pedestrian</u> entering the <u>crosswalk</u> due to <u>poor sight distance</u> because of <u>congested traffic</u> and <u>Moisture road</u> surface due to <u>rain</u> at <u>night</u> .	
	2	A car <u>driving left curve</u> turn on a 2 lanes road occurs in a <u>side-end collision</u> with a <u>pedestrian jaywalking</u> an <u>crosswalk</u> .	
	3	A car <u>driving right curve</u> turn on an <u>uphill</u> road with <u>poor sight distance</u> occurred a <u>collision</u> with a <u>pedestrian</u> entering a <u>crosswalk</u> .	
	4	A car <u>driving</u> on a <u>straight</u> uphill section of <u>4 lanes</u> at an <u>unsignalized crosswalk</u> collided with a <u>pedestrian</u> who <u>suddenly entered</u> from between <u>parked cars</u> , resulting in a <u>collision</u> .	
Motorcycle	1	A car <u>driving straight</u> on a <u>1 lane one way</u> road collided with a <u>motorcycle</u> that <u>suddenly entered</u> from between <u>parked vehicles</u> , resulting in a <u>collision</u> .	
	2	At <u>night</u> , a car <u>driving straight</u> on <u>2 lanes round trip</u> road in a <u>side street</u> was involved in a <u>side-end collision</u> with a <u>motorcycle encroaching the centerline</u> due to <u>illegal parking</u> .	
	3	At <u>night</u> , a car driving on <u>3 lanes one way</u> road <u>collided</u> with a <u>motorcycle</u> during a <u>left turn curve</u> due to the motorcycle's <u>signal violation</u> .	
	4	A motorcycle that was following a car that <u>suddenly braked</u> due to <u>construction work</u> on a <u>straight</u> section of <u>1 lane one way road</u> failed to maintain a <u>safe distance</u> , resulting in a <u>rear-end collision</u> .	

Table 8. Edge Case scenario development results (continued)

Class	No.	Contents	Scenario Situation
Bicycle	1	On <u>1 lane one way downhill side street</u> , a car driving <u>straight</u> <u>collided</u> with a <u>bicycle</u> that <u>suddenly entered</u> the road, despite the car making an emergency <u>sudden stop</u> .	
	2	At night, a car turning <u>right curve</u> on a <u>road within an apartment complex</u> <u>collided</u> with a <u>bicycle</u> that <u>suddenly entered</u> from between <u>parked cars</u> .	
	3	At night, a car <u>driving straight</u> on <u>6 lane round trip road</u> with <u>illegally parked cars</u> <u>collided</u> with a <u>bicycle jaywalking</u> at a <u>crosswalk</u> .	
	4	At night, a car making a <u>left curve</u> turn on <u>4 lane round trip road</u> attempted an emergency <u>sudden stop</u> upon spotting a <u>bicycle</u> crossing at an <u>unsignalized crosswalk</u> but <u>collided</u> due to <u>poor sight distance</u> .	
E-Scooter	1	At daytime, a <u>straight</u> section of <u>1 lane round trip side street</u> with <u>parked vehicles</u> , a <u>dooring accident</u> occurred as a <u>parked</u> car opened its door with <u>poor sight distance</u> of an <u>e-scooter</u> .	
	2	At <u>night</u> , On <u>uphill 3 lanes one way road</u> with <u>cars parked</u> on the side, a car driving <u>straight</u> <u>collided</u> with an <u>e-scooter</u> that <u>suddenly entered</u> at a <u>crosswalk</u> .	
	3	At <u>night</u> , a car driving on <u>2 lanes round trip road</u> <u>collided</u> with an <u>e-scooter</u> due to traveling in <u>centerline encroachment</u> .	
	4	At <u>night</u> , On an <u>uphill, 1 lane one way side street</u> , a car making a <u>left curve</u> turn <u>collided</u> with an <u>e-scooter</u> that <u>suddenly entered</u> the road.	

결론

본 연구는 자율주행차와 보행자, 이륜차, 자전거, 전동킥보드가 혼재되어 있는 복잡한 교통 환경에서 자율주행차의 안전성과 신뢰성을 도모하고, 비정형적이고 예측하기 어려운 사고에 대응하기 위한 Edge Case 시나리오를 구축하기 위해 진행되었다. 이러한 연구는 자율주행차 기술의 상용화와 안정성 확보를 위한 필수적인 단계로, 실제 도로

환경에서의 다양한 사고 요인을 사전 분석 및 도출을 통해 시나리오를 구축하여 자율주행차 안전성 평가에 활용하고 자율주행차가 예측하지 못한 상황에서 갑작스럽게 발생하는 위험상황을 대비할 수 있는 체계적인 시나리오 구축 방법을 제시했다.

자율주행차와 보행자, 이륜차, 자전거, 전동킥보드와 같이 예측이 어려운 객체와 복잡한 교통환경에서 발생하는 상호작용을 Edge Case 사고 유형을 중심으로 분석하기 위해 유튜브 채널 '한문철 TV'에서 수집한 차대 보행자, 차대 이륜차, 차대 자전거, 차대 전동킥보드 사고에 대한 2,410건의 블랙박스 데이터셋을 활용하였다. 해당 데이터는 사고 발생 전후 상황, 도로 환경, 교통 요소 등을 상세히 기록하고 있어 시나리오 구축을 위한 사고요인 분석과 도출에 적합하였다. 블랙박스 영상의 사고정보를 추출하여 텍스트 데이터로 구성된 사고 데이터셋을 정제하기 위해 불용어 및 잉여 키워드를 제거하고, KoNLPy 기반의 텍스트 마이닝과 LDA 토픽모델링 기법을 활용하여 주요 사고 위험요인을 도출하였다. 이를 바탕으로 자율주행차 안전성 평가에 접목시키기 위해 독일의 자율주행차 평가 테스트 프레임워크인 Pegasus 5-layer 포맷에 기반해 도로 기하구조, 교통 인프라, 환경조건 등으로 체계적으로 분류하여 Edge Case 시나리오를 설계하였다.

분석 결과, 블랙박스 데이터셋을 기반으로 자율주행차의 단순한 사고 유형을 넘어 복잡하고 예측이 어려운 상황에서도 안정적으로 작동할 수 있도록 지원하는 시나리오를 구축할 수 있었다. 보행자, 이륜차, 자전거, 전동킥보드 등 다양한 도로 사용자와의 상호작용을 포함한 시나리오는 실제 도로 환경에서의 자율주행차 운행을 보다 정밀하게 반영할 수 있는 기반을 제공하였다. Pegasus 5-layer 포맷을 활용하여 도로 기하구조, 도로조건, 교통 방해요소, 주요 사고원인 등을 체계적으로 분석함으로써 자율주행차 안전성 평가에 활용할 수 있는 방법론을 제시하였다. 특히, Pegasus 5-layer 포맷을 기반으로 시나리오를 구축함에 따라 도로 기하구조, 객체, 환경조건 등이 명확히 정의되어 있어 데이터 변환을 통해 CARLA, PreScan 등 자율주행 시뮬레이터 플랫폼에 적용 가능하며 이를 바탕으로 실도로 테스트베드 기반 평가 절차에도 활용될 수 있을 것으로 보인다. 또한, 실질적으로 실도로에서 각각의 다양한 요소의 재현 가능성을 고려하여 구축된 시나리오임에 따라 실도로 테스트베드에서도 안정성 평가가 진행될 수 있을 것으로 기대된다. 이를 통해 자율주행차 안전성 검증 및 위험요인 예측에 있어 중요한 도구로 활용될 수 있을 것으로 기대된다. 시나리오를 통해 자율주행차의 실제 운행 환경에서 발생할 수 있는 잠재적 문제를 사전에 예측함으로써 자율주행 기술의 신뢰성을 향상시킬 수 있다. 더 나아가, 본 연구에서 제안한 접근법은 자율주행차와 인간 운전 차량이 혼재된 교통 환경에서도 기술의 유효성을 평가하는 데 기여할 수 있을 것으로 기대된다.

그러나 본 연구는 블랙박스 영상에서 확인이 어렵거나 진행자의 설명으로 언급되지 않은 도로환경(Road Environment), 날씨(Weather) 등 일부 항목은 텍스트 데이터로 추출하기에 어려운 부분이 있었으며 이로 인해 LDA 토픽모델링 수행 시 해당 항목이 시나리오 구축에 과소 평가되었을 수 있다. 또한, 실제 발생한 자율주행차 사고 데이터를 활용하지 못해 모든 자율주행차의 시나리오를 설명하는 데에 부분적으로 한계가 있을 것으로 보인다. 또한, 구축된 Edge Case 시나리오가 실제 자율주행차의 안전성 평가 과정에 적용 가능 여부에 대한 시뮬레이션이나 실도로 환경에서 실증을 통해 검증되지 않았다는 한계가 있다. 따라서 후속 연구에서는 본 연구에서 도출한 시나리오를 Waymo, Tesla, Uber ATG 등의 실제 자율주행차 사고 데이터를 기반으로 비교 검토하여 시나리오의 타당성을 검증하여 시나리오를 검토하고 타당성을 제고할 필요가 있다. 아울러, 도출된 최종 Edge Case 시나리오가 실제 자율주행차 안전성 평가에 활용 가능한지를 검토하기 위해 CARLA, PreScan 등 시뮬레이션 플랫폼을 활용한 가상주행 분석과 실도로 테스트베드 기반의 실증 실험을 병행하여 시나리오의 타당성과 실도로 적용 가능성을 검증할 필요가 있다. 궁극적으로 데이터의 다양성과 양을 확대하고 보행자, 이륜차, 자전거, 전동킥보드와 같은 객체와의 상호작용뿐만 아니라 차량을 포함하여 자율주행차 상용화 시대에 대비할 수 있는 새롭고 종합적인 교통 조건을 포함하여 시나리오의 포괄성과 정밀도를 강화하는 연구를 향후 진행하고자 한다. 이를 통해 자율주행차가 모든 운행 조건에서 신뢰성 있고 안전하게 작동할 수 있는 기반을 마련할 수 있을 것이다.

Funding

This work was conducted as part of the Autonomous Driving Technology Development Innovation Project of the Ministry of Science & ICT(MSIT) and the Institute for Information & communication Technology Planning & evaluation(IITP)(2021-0-00697 Development of Technology for Verification of Operational SW Safety and Counter-measures under the Fall Back MRC for Edge Class Autonomous Driving Function.

This work was also supported by the IITP(Institute of Information & Coummunications Technology Planning & Evaluation)-ICAN(ICT Challenge and Advanced Network of HRD)(IITP-RS-2024-00436954) grant funded by the Korea government(Ministry of Science and ICT), and was supported by Kyonggi University's Graduate Research Assistantship 2025.

알림

본 논문은 한국ITS학회 2024년도 춘계학술대회(2024.04.26)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

ORCID

KIM, Taejun  <http://orcid.org/0009-0002-0605-6657>

CHU, Hyeyeon  <http://orcid.org/0009-0001-5516-6855>

JANG, Jeong Ah  <http://orcid.org/0000-0001-5629-8342>

YUN, Hyunjeong  <http://orcid.org/0009-0006-4421-7764>

KIM, Junghwa  <http://orcid.org/0000-0001-9771-5755>

REFERENCES

- Blei D. M. (2012), Probabilistic Topic Models, *Communications of the ACM*, 55(4), Association for Computing Machinery, 77-84.
- Baek S. R., Kim Y. W., Yoo T. H., Kim C. H. (2024), Development of Database-Based Pedestrian Accident Scenarios and Analysis of Accident Prevention Effects from V2X Application, *Journal of Korean Society of Intelligent Transport Systems*, 23(5), The Korean Society of Intelligent Transport Systems, 274-292.
- Cao Y., Xiao L., Dong H., Wang Y., Wu X., Li P., Qiu Y. (2019), Typical Pre-Crash Scenarios Reconstruction for Two-Wheelers and Passenger Vehicles and its Application in Parameter Optimization of AEB System based on NAIS Database, In 26th International Technical Conference on the Enhanced Safety of Vehicles(ESV), 19-0272.
- Chung S. (2024), A Study on the Methodology for Developing High-Risk Scenarios for the Validation of Autonomous Vehicles, *Transactions of the Korean Society of Automotive Engineers*, 32(3), The Korean Society of Automotive Engineers, 301-308.
- Durak U., Gupta S., Akgün K. (2023), Critical Scenario Techniques for Automated Vehicles: Literature Review, *ASIM Workshop STS/GMMS/EDU Proceedings*, 87-94.
- Guo A., Zhou Y., Tian H., Fang C., Sun Y., Sun W., Gao X., Luu A. T., Liu Y., Chen Z. (2024), SoVAR: Building Generalizable Scenarios from Accident Reports for Autonomous Driving Testing, In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering (ASE '24)*, 1-13.

- Jo J. S., Kwak J. H., Ku D. G., Yoo Y. T., Lee S. J. (2022), Estimation of Leisure Behavior Using Unstructured Data: Focusing on Seoul City Wall, J. Korean Soc. Transp., 41(1), Korean Society of Transportation, 135-147.
- Kalisvaart S. (2021a), Introducing Edge Cases, Presented at the EU-CSA ARCADE Workshop on Edge Cases, European Commission.
- Kalisvaart S., Svanberg E., Barnard Y., Innamaa S., Sanz L., Bou J., Koskinen S. (2021b), Proceedings of Workshop on Edge Cases: ARCADE D2.4 Joint CCAM Networking Report, European Commission, Horizon 2020 Programme, 2-27.
- Kang J. H., Ko W. R., Kim Y. J., Yoon J. G., Yun I. S. (2024), Development of Autonomous Vehicle Evaluation Scenarios Based on Car-to-Bicycle, Car-to-Pedestrian, and Car-to-Animal Traffic Accidents, Journal of Korean Society of Intelligent Transport Systems, 23(5), The Korean Society of Intelligent Transport Systems, 322-337.
- Kang M. H., Lee W. S., Hwang K. Y., Yoon Y. (2022), Vision Transformer for Detecting Critical Situations and Extracting Functional Scenario for Automated Vehicle Safety Assessment, Sustainability, 14(15), MDPI, 9680.
- Kibalama D., Tulpule P. J. (2022), AV/ADAS Safety-Critical Testing Scenario Generation from Vehicle Crash Data, SAE Technical Papers, 1-9.
- Lee H. J., Kang M. H., Song, J. I., Hwang K. Y. (2023), Development of Autonomous Vehicle Traffic Accident Scenarios in Urban Areas Based on Real-World Accident Data Using Association Rule Mining, J. Korean Soc. Transp., 41(3), Korean Society of Transportation, 375-393.
- Li P., Zhu X., Ren Y., Tan Z., Hu W., Zhang Y., Xu C. (2024), Generalization of Cut-in Pre-Crash Scenarios for Autonomous Vehicles based on Accident Data, Scientific Reports, 14, 17664.
- Liu Q., Wang X., Liu S., Yu C., Glaser Y. (2024), Analysis of Pre-Crash Scenarios and Contributing Factors for Autonomous Vehicle Crashes at Intersections, Accident Analysis and Prevention, 195, Elsevier, 1-12.
- Luettel T., Himmelsbach M., Wuensche H. J. (2012), Autonomous Ground Vehicles: Concepts and a Path to the Future, Proceedings of the IEEE, 100(1), IEEE, 1831-1840.
- Masmoudi M., Ghazzai H., Frikha M., Massoud Y. (2019), Object Detection Learning Techniques for Autonomous Vehicle Applications, IEEE International Conference on Vehicular Electronics and Safety (ICVES), 1-6.
- MIRA (2018), K-City Scenario Definition Consultancy.
- MOLIT (2024), https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmepage=1&id=95089899, 2024.12.15.
- Moradloo N., Mahdinia I., Khattak A. J. (2024), Safety in Higher Level Automated Vehicles: Investigating Edge Cases in Crashes of Vehicles Equipped with Automated Driving Systems, Accident Analysis & Prevention, 203, Elsevier, 107607.
- 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.
- NHTSA (2022), Summary Report : Standing General Order on Crash Reporting for Level 2 Advanced Driver Assistance Systems, United State, National Highway Traffic Safety Administration.
- Oh J. S. (2015), Identifying Research Opportunities in the Convergence of Transportation and ICT Using Text Mining Techniques, J. Korean Soc. Transp., 22(4), Korean Society of Transportation, 93-110.
- Park S. M., Park S. H., Jeong H. R., Yun I. S., So J. H. (2021a), Scenario-Mining for Level 4 Automated Vehicle Safety

- Assessment from Real Accident Situations in Urban Areas Using a Natural Language Process, *Sensors*, 21(20), MDPI, 6929.
- Park S. M., Lee H. P., So J. H., Yun I. S. (2021b), Analysis for Autonomous Vehicle Collision Using Text Embedding, *Journal of Korean Society of Intelligent Transport Systems*, 20(1), The Korean Society of Intelligent Transport Systems, 160-173.
- Park S. M., So J. H., Ko H. G., Jeong H. R., Yun I. S. (2018), Development of Safety Evaluation for Automated Vehicle Tests Using 5-Layer Format (Case of the Community Road), *Journal of Korean Society of Intelligent Transport Systems*, 18(2), The Korean Society of Intelligent Transport Systems, 114-128.
- PEGASUS project, <https://www.pegasusprojekt.de/en/home>, 2024.12.17.
- Rampf F., Grigoropoulos G., Malcolm P., Keler A., Bogenberger K. (2023), Modelling Autonomous Vehicle Interactions with Bicycles in Traffic Simulation, *Frontiers in Future Transportation*, 3, 894148.
- Riedmaier S., Ponn T., Ludwig D., Schick B., Diermeyer F. (2020), Survey on Scenario-Based Safety Assessment of Automated Vehicles, *IEEE Access*, 8, IEEE, 87456-87458.
- So J. H., Park I. S., Wee J. R., Park S. M., Yun I. S. (2019), Generating Traffic Safety Test Scenarios for Automated Vehicles using a Big Data Technique, *KSCE Journal of Civil Engineering*, 23(6), Korean Society of Civil Engineers, 2702-2712.
- Sui B., Lubbe N., Bärghman J. (2019), A Clustering 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 and Prevention*, 132, Elsevier, 105242.
- Wang X., Peng Y., Xu T., Xu Q., Wu X., Xiang G., Yi S., Wang H. (2022), Autonomous Driving Testing Scenario Generation based on in-Depth Vehicle-to-Powered Two-Wheeler Crash Data in China, *Accident Analysis & Prevention*, 176, Elsevier, 106812.
- Zhang Y., Ko S., Lee K. (2021), Autonomous Vehicle Scenario-based Safety Evaluation Using Improved Criticality Metrics, *Accident Analysis & Prevention*, 160, Elsevier, 106312.