

# 고속도로 톨게이트 교통사고 유형화 및 네트워크 클러스터링 기반 톨게이트 개선방안 수립 연구

## Study on the Establishment of Tollgate Improvement Measures through Categorization of Expressway Tollgate Accidents and Network Clustering

김 인 영\* · 정 한 솔\*\* · 박 상 민\*\*\* · 이 광 섭\*\*\*\* · 윤 일 수\*\*\*\*\*

\* 주저자 : 아주대학교 D.N.A. 플러스융합학과 석박사 통합과정  
 \*\* 공저자 : 아주대학교 D.N.A. 플러스융합학과 석사과정  
 \*\*\* 공저자 : 한국교통연구원 도로교통연구본부 부연구위원  
 \*\*\*\* 교신저자 : SK 텔레콤 Teleco data 사업팀 부장  
 \*\*\*\*\* 공저자 : 아주대학교 교통시스템공학과 교수

Inyoung Kim\* · Hansol Jeong\* · Sangmin Park\*\* · Kwangseob Lee\*\*\* · Ilsoo Yun\*

\* Dept. of Transportation Eng., Ajou Univ.  
 \*\* Dept. of Road Transport Research, Korea Transport Institute  
 \*\*\* Teleco data business team, SK telecom

† Corresponding author : Kwangseob Lee, leekase@sk.com

Vol. 23 No.5(2024)  
 October, 2024  
 pp.01~17

pISSN 1738-0774  
 eISSN 2384-1729  
<https://doi.org/10.12815/kits.2024.23.5.1>

Received 19 July 2024  
 Revised 16 August 2024  
 Accepted 4 October 2024

© 2024. The Korean Society of Intelligent Transport Systems. All rights reserved.

### 요 약

국내 톨게이트는 하이패스 차로와 Toll Collection System(TCS) 차로가 공존하는 복잡한 형태로 설계되어 있어 해당 구간에서 교통사고가 빈번히 발생하고 있다. 정부에서는 교통사고 예방을 위해 교통사고 요인 분석을 기반으로 톨게이트 개선방안을 도출하고 개선해오고 있으나 여전히 교통사고가 끊이지 않고 있다. 다만, 톨게이트는 단거리이지만 주행 시 인지해야 할 상황과 이벤트들이 다수 존재해 교통사고 분석 시 상황과 요인 등을 복합적으로 고려할 필요가 있다. 따라서 본 연구에서는 교통사고 내용 데이터에 역문서 빈도 가중치를 적용하여 교통사고 요인과 상황을 도출하였으며, 그 이후 톨게이트 교통사고를 유형화하였다. 유형화 결과를 바탕으로 네트워크 클러스터링을 수행하여 실효성 높은 톨게이트 개선방안을 제안하였다.

핵심어 : 톨게이트 교통사고, 역문서 빈도 가중치, 교통사고 유형화, 네트워크 클러스터링, 톨게이트 개선방안

### ABSTRACT

In Korea, tollgates are designed in a complex manner with the coexistence of Hi-Pass and Toll Collection System lanes, frequently leading to traffic accidents. Despite the continuous efforts of the government to improve tollgates based on an analysis of accident factors, incidents still persist. Tollgates require drivers to be aware of numerous circumstances and events within a short distance, necessitating careful consideration of several factors and circumstances when analyzing traffic accidents. Therefore, this study applied the Term Frequency-Inverse Document Frequency method to traffic accident data to identify the factors and circumstances. Subsequently, the tollgate traffic accidents were categorized. Finally, effective tollgate improvement measures were proposed based on the categorization result.

Key words : Tollgate traffic accident, Term Frequency-Inverse Document Frequency, Traffic Accident Categorization, Network Clustering, Improvement strategies of tollgate

## I. 서론

### 1. 개요

국내 톨게이트는 하이패스 차로와 Toll Collection System(TCS) 차로가 공존하는 복잡한 형태로 설계되어 있어 차량들의 급차선변경, 급가속 등 급격한 주행행태를 야기하고 있다. 이로 인해 톨게이트 구간에서의 교통사고가 빈번히 발생하고 있으며, 특히 2022년에는 톨게이트 부근이 거리가 짧음에도 불구하고 위치별 고속도로 교통사고 중 세 번째(약 8.2%)로 크게 나타났다(Korea Road Traffic Authority, 2023). 고속도로 톨게이트 내에서의 교통사고가 지속적으로 발생하면서 국내 정부에서는 교통사고 예방을 위한 다양한 대책안을 제시해오고 있다. 그 중 교통 시설물(시선유도봉, 표지판) 추가 설치는 2014년 대비 2018년에 교통사고를 32.6% 감소시킨 바 있다(Ministry of Land, Infrastructure and Transport(a), 2024). 이외에도 다차로 하이패스 구축을 통해 하이패스의 차로폭이 넓어지고 시설물이 줄어들면서 편안하고 안전한 주행이 가능해졌으며, 최근에는 지정체 해소 및 사고예방 효과를 목표로 번호판 인식방식의 무정차 통행료 결제 시스템(즉, 스마트톨링) 기술을 고속도로 일부 구간에 시범적으로 도입하였다(Korea Expressway Corporation, 2021; Ministry of Land, Infrastructure and Transport(b), 2024).

다만, 교통 시설물 및 다차로 하이패스 구간 확대에도 불구하고 여전히 톨게이트 교통사고가 끊이지 않고 있다. 이는 교통사고 분석 시 단순히 운전자, 차량, 기타 요인 등에 대해 기초 통계분석을 수행하거나 기하구조 요인 분석을 통해 개선방안을 도출하여 구체적인 교통사고 상황을 고려하지 않았기 때문인 것으로 판단된다. 톨게이트의 경우 특히, 단거리 내 분·합류, 차선수 감소 및 증가, 폐쇄된 톨부스 등 주행 시 인지해야 할 상황과 이벤트들이 다수 존재하기 때문에 복합적으로 상황과 요인 등 모두를 고려할 필요가 있다. 또한, 먼 미래에 마주하게 될 톨게이트 내 고도의 자율주행차(automated vehicle, AV)의 주행 안전성 확보를 위해서는 보다 다양한 상황과 요인들이 고려될 필요가 있으며 이를 기반으로 실효성 높은 톨게이트 개선방안을 수립할 필요가 있다.

따라서 본 연구에서는 지속 가능하고 체계적인 톨게이트 교통사고 분석 방법과 이에 따른 개선방안을 도출하고자 하였다. 이를 위해 우선, 기존 고속도로 교통사고 요인 분석 시 주로 활용되고 있는 한국도로공사의 운전자, 차량, 기타 등 세 가지 교통사고 요인들을 재구성하고자 하였다. 교통사고 요인 재구성을 위해 교통사고 상황과 이벤트가 문자열로 자세히 묘사된 한국도로공사의 교통사고 상황 데이터를 활용하여 교통사고를 유형화하고자 하였다. 해당 데이터에 대해 텍스트 마이닝 기법 중 하나인 역문서 빈도 가중치(term frequency inverse document frequency, TF-IDF) 분석을 수행하여 주요 단어들을 도출하고 단어들을 분류하는 기준을 정립하여 기준에 따라 주요 교통사고 요인들을 매칭하였다. 그다음으로는 기준별 교통사고 요인들 간의 연관성을 살펴보기 위해 louvain 알고리즘을 통해 단어 간 네트워크 클러스터링을 수행하였으며, 군집화 결과를 해석하여 향후 톨게이트의 운영 및 설계방식의 변화상을 모색하였다.

### 2. 연구의 범위 및 절차

본 연구는 고속도로 시설 중 연장당 교통사고 빈도가 높으며, 단거리 내 복잡한 도로형상과 주행규칙이 다수 존재하는 톨게이트 구간을 공간적 범위로 설정하였다. 해당 구간에서 발생한 교통사고 상황 데이터가 수집된 2018년부터 2022년까지를 시간적 범위로 설정하였다. 연구의 내용적 범위로는 텍스트 마이닝 기반 고속도로 톨게이트 교통사고 요인 유형화, 유형화된 사고요인 louvain 네트워크 클러스터링 수행, 클러스터링

결과 기반 톨게이트 개선방안 도출 등으로 설정하였다. 마지막으로 연구의 수행 절차는 연구 계획 수립, 관련 이론 및 선행 연구고찰, 교통사고 요인 유형화 및 네트워크 클러스터링 수행, 톨게이트 개선방안 제안 순으로 구성하였다.

## II. 관련 이론 및 선행 연구 고찰

### 1. 관련 이론

#### 1) 국내외 톨게이트 유형

국내외 많은 톨게이트는 manual toll collection(MTC)과 electronic toll collection(ETC) 차선이 공존하는 전통적인 톨게이트(traditional mainline toll plaza, TMTP) 방식으로 운영이 되고 있다. TMTP는 MTC와 ETC 차선이 공존하는 톨부스로 감속 혹은 정지가 요구된다. 미국의 일부 고속도로에서는 도로용량을 늘리고 교통사고 건수 감소 효과가 있는 MTC와 ETC 차선이 완전히 분리된 형태인 하이브리드 톨게이트(hybrid mainline toll plaza, HMTP) 방식으로 운영되고 있다(Xing et al., 2019; Abuzwidah and Abdel-Aty, 2018). HMTP의 경우 ETC 차량은 개방형 톨링 차로를 통해 감속 혹은 정지 없이 바로 통과가 가능하다. 국내의 경우 TMTP와 HMTP가 공존하는 방식으로 톨게이트를 운영하고 있다. 최근 다차로 하이패스와 스마트톨링 사업 추진으로 인해 ETC와 TCS 차선을 구분하고 있으나, 화물차의 오른쪽 통행으로 인해 TCS 차선에 화물차 하이패스가 공존하는 형태로 운영하고 있다. 다음 <Fig. 1>은 미국의 TMTP와 HMTP 운영사례를 보여주고 있다.



Source: Safety Evaluation of Hybrid Main-Line Toll Plazas(Abuzwidah et al., 2014)

<Fig. 1> TMTP and HMTP Examples

#### 2) 스마트톨링

스마트톨링 시스템은 기존 하이패스 시스템이 가지는 단점(TCS 차로 이용을 위한 가감속, 차선 변경 등)을 보완하고 운영 효율성을 향상시키기 위해 도입된 개념으로 하이패스 혹은 번호판 촬영 등 영상인식 기술을 활용하여 통행권을 받거나 통행요금을 납부하는 것이다(The Korea Transport Institute, 2018). 국내에서는 2016년 8월 「제1차 국가도로종합계획」에서 스마트톨링 도입계획을 밝힌 바 있다(Ministry of Land, Infrastructure and Transport, 2021). 2024년 5월부터는 고속도로 경부선(대왕관교), 남해선(서영암, 강진무위사, 장흥, 보성, 벌교, 고흥, 남순천, 순천만) 등 일부 구간에서 스마트톨링 시범사업을 수행하고 있다(Ministry of Land, Infrastructure and Transport(c), 2024).

국외에서는 번호판 인식 센서를 통해 자동으로 통행료를 수납하는 형태로 현장수납 없이 전 차선의 수납을 자동화한 시스템인 all electronic tolling(AET) 시스템을 사용하고 있다. 특히, 캐나다는 최초로 AET를 도입한 나라로 시간대별, 차종별 거리비례요금제를 적용하여 폐쇄식으로 운영하고 있다. 이에 반해 미국 등의 대부분의 AET 도입국가에서는 개방식으로 운영하고 있다(Korea Expressway Corporation, 2021).

### 3) 다차로 하이패스

다차로 하이패스는 2~4개의 하이패스 차로를 병합하여 하나의 차로처럼 이용할 수 있도록 차로 구분 시설물을 철거하고 차로폭을 확보하는 것이다. 다차로 하이패스는 기존 단차로 하이패스보다 빠른 속도로 통과가 가능하며, 차로폭이 넓어 운전자가 안전하게 통과할 수 있다. 국내 다차로 하이패스 구축으로 인한 효과로는 2021년 기준 차로당 통과대수가 최대 64%(1,100대/h → 1,800대/h) 증가했으며, 두 번째로는 교통사고가 18%(2018년 38건 → 2020년 31건) 감소한 것으로 나타났다. 이로 인해 다차로 하이패스의 효과가 입증되면서 고속도로 톨게이트가 점차적으로 다차로 하이패스로 전환되고 있으며, 2021년 기준 82개의 영업소에 구축이 완료된 상태이다(Ministry of Land, Infrastructure and Transport(d), 2024).

국외에서는 open road tolling(ORT) 시스템으로 명명하고 있으며, 이는 AET와 동일하게 번호판 인식 센서를 통해 통행료를 자동으로 수납하는 형태이다. 다만, 현장수납 차로를 운영하기 때문에 해당 차로에서는 번호판 영상인식에 의한 자동 수납 서비스는 제공하지 않는다. 캐나다 국경에 인접한 미국의 일부 주에서는 타 국가 차량의 요금 미납 사례를 고려하여 ORT로 전환하여 톨게이트를 운영하고 있다(Korea Expressway Corporation 2021). 다음 <Fig. 2>는 국내 스마트톨링과 다차로 하이패스이다.



Source: (a) Press Release(Ministry of Land, Infrastructure and Transport(c), 2024), (b) Press Release(Ministry of Land, Infrastructure and Transport, 2021)

<Fig. 2> Smart Tolling and Multi-Lane Hi-Pass System Examples

## 2. 선행 연구 고찰

### 1) 톨게이트 교통사고 분석

Xing et al.(2019)는 중국 난징 톨게이트 상류부에서의 ETC와 MTC의 안전성을 평가하였으며, 기본 구간, 분류 구간, 요금 수집 구간으로 나누어서 연구를 진행하였다. 무인항공기(unmanned aerial vehicle, UAV)를 통해 수집된 비디오 영상 중 차량 궤적 데이터를 추출하여 충돌소요시간(time to collision, TTC) 지표를 통해 차

량 충돌 위험을 평가하였다. 특히, 랜덤 파라미터 로지스틱 모형을 통해 통행료 징수 유형, 차량 유형, 목표 주행 차로 및 위치가 안전에 미치는 영향을 비교하였다. 그 결과, ETC의 TTC 값은 2.18초 MTC는 2.33초로 특히 분류 구간에서 MTC보다 ETC가 더 높은 제한속도로 인해 충돌 위험이 더 크게 나타났다. 또한, ETC와 MTC 차량이 혼합된 차선의 경우 상층에 영향을 주는 경향이 큰 것으로 나타났으며, MTC 차량이 ETC 차량보다 충돌 위험이 큰 것으로 나타났다.

Song et al.(2022)는 톨게이트 내 교통 상충 위험을 평가하기 위해 혼합로지토형(mixed logit model)을 활용하여 MTC와 ETC 차량별 상충 위험 연관성을 살펴보았다. 분석 변수로는 차량 유형, 속도, 가속도, 요금 수집 유형, 공간적 특성 등으로 선정하였다. 그 결과, ETC 차량이 요금소 부근에서 정지는 하지 않으나 공격적으로 주행하는 경향이 있으며 차선변경이 빈번하게 이루어지는 것으로 나타났다. MTC 차량은 요금소의 대기행렬을 확인하고 가장 짧은 대기행렬을 찾아 경로를 설정하는 것으로 드러났다. 또한, 평균속도가 증가하는 경우 상충 심각도가 높아지며, 후행 차량이 화물 운송 차량일 경우 상충 빈도가 증가하는 것으로 나타났다. 이는 주변 관측이 어려운 것이 주요 원인으로 판단하였다. 공간적 특징으로는 차선변경, 엇갈림, 합류 및 분류 구간, 도로 작업 구간에서 상충 위험과 연관이 있는 것으로 나타났다.

## 2) 텍스트 기반 교통사고 유형 분석 및 키워드 클러스터링

Giummarra et al.(2021)는 호주 빅토리아 주에서 발생하는 도로 충돌사고의 특성을 규명하고자 텍스트 마이닝 기법을 사용하여 교통사고 충돌 특성을 분류하고 예방방안을 도출하였다. 이를 위해 2010년부터 2016년 사이에 발생한 빅토리아 주 교통사고 데이터를 수집하였으며 qualitative data analysis(QDA) 마이너 툴을 사용하여 텍스트 데이터로부터 교통사고 상황을 분류하고 wordstat를 통해 중요 키워드를 추출하여 패턴을 확인하였다. 그다음으로는 분석된 데이터의 통계적 검증을 수행하여 교통사고 유형과 과실 귀속 간 상관관계를 조사하였다. 그 결과, 타 차량의 과실, 기계적 결함과 같은 고유 과실, 보행자 횡단 및 후방 충돌 등 과실 불분명 등이 주요 원인으로 분석되었다.

Kwayu et al.(2021)는 미시간에서 발생한 도로 차량 충돌 텍스트로부터 주요 토픽을 이해하고 토픽과 충돌로 인한 차량 파손 유형(angle, head on, rear end, sidewipe, other) 간 연관성 등에 대해 평가하였다. 이를 위해 structural topic modelling(STM)과 network topology 분석 기법을 사용하였다. 총 60개의 토픽을 추출 후 사전에 정의된 차량 파손 유형을 각 토픽에 라벨링 후 유형별 주요 교통사고 요인들을 도출하였다. 그 예시로, angle과 관련된 주요 토픽은 정지표시, 빨간불, 좌회전, 차선 변경 등이 있었다.

Esenturk et al.(2023)는 교통사고 데이터를 분석하여 AV가 마주할 수 있는 다양한 위험 상황을 식별하고 이를 기반으로 테스트 시나리오를 생성하였다. 이를 위해 2016년부터 2018년 사이의 영국에서 발생한 교통사고 데이터를 수집하였으며 robust clustering using links(ROCK) 알고리즘을 활용하여 텍스트 데이터를 군집화하였다. 그 결과, 총 26개의 클러스터가 도출되었으며 사고 패턴을 잘 나타내는 클러스터를 확인한 결과, ‘야간 시간대 고속도로와 떨어진 부근에서 미숙련 운전자가 보행자 및 도로 위 물체와 충돌’, ‘교차로에서 왼손 운전자가 차량을 운전하고 있으며 차량이 미끄러지거나 전복되는 사고’, 등이 도출되었다. 클러스터 내 키워드를 통해 위험상황을 도출하고 이를 기반으로 AV 안전성 테스트를 진행하도록 권고하였다.

Lee et al.(2020)은 키워드 네트워크 분석 중 주요 연구의 주제를 식별할 수 있는 네트워크 클러스터링 알고리즘을 비교 및 분석하였다. edge betweenness(EB), fast greedy(FG), walktrap, leading eigenvector(LE), louvain algorithm 등을 비교 및 분석하였으며, 그 결과 네트워크 클러스터링 중 louvain 알고리즘이 모듈성, 처리시간 측면에서 가장 우수한 성능을 보이는 것으로 나타났으며 그다음으로는 FG 알고리즘이 우수한 성능을 보이는 것으로 나타났다.

### 3. 시사점

톨게이트 교통사고 분석 관련 선행 연구를 고찰한 결과, 톨게이트 구간의 경우 단거리 내 잦은 변이구간으로 인해 차량 간 잦은 가감속 및 차선변경이 요구되어 충돌사고가 발생하고 있었다. 주요 충돌사고에 영향을 미치는 변수로는 공통적으로 차량 유형, 통행료 징수 유형, 복잡한 공간적 특성 등이 있었다. 또한, 텍스트 기반 교통사고 유형 분석 관련 연구를 고찰한 결과 텍스트를 군집화 및 유형화하여 도로에서 발생하는 다양한 충돌 사고요인 및 원인들을 도출하였다. 즉, 최근 연구동향을 살펴보면 구체적인 교통사고 요인 및 원인에 근거하여 교통사고 분석을 수행하고 요인과 원인 간 연관성을 분석하여 안전대책을 강구하고 있다.

다만, 국내의 경우 톨게이트 개선방안을 수립할 때 단순히 운전자, 차량, 기타 등 세 가지 요인을 통해 개선방안을 도출해오고 있다. 이는 국내의 복잡한 톨게이트 운영방식 및 기하구조 특성과 이에 따른 차량행태 변화 등을 고려하지 않았다는 한계점이 있다. 실효성 높은 톨게이트 개선방안을 도출하기 위해서는 실제 차량이 마주할 수 있는 종합적인 상황까지 고려하여 개선방안을 도출하고 정책을 수립할 필요가 있다. 이를 바탕으로 개선방안을 수립하게 된다면 장기적으로 톨게이트 운영효율 및 교통사고 감소 효과를 높일 수 있을 것으로 기대된다.

## Ⅲ. 교통사고 요인 유형화 및 네트워크 클러스터링 기반 교통사고 상황 도출

### 1. 분석데이터 수집

본 연구에서는 한국도로공사 교통사고 데이터를 톨게이트 구간의 구체적인 사고유형을 분석하기 위한 기초자료로 활용하고자 하였다. 이를 위해 2018~2022년 고속도로 교통사고 데이터 중 톨게이트 구간 내 발생한 사고 상황이 텍스트 형식으로 작성되어 있는 4,819건의 교통사고 전수 데이터를 수집하였다. 교통사고 전수 데이터는 특정 기간 동안 발생한 교통사고를 모두 포함하는 것으로 해당 기간 동안의 교통사고 현황을 대표할 수 있는 표본에 해당된다. 교통사고 피해규모가 큰 사고유형을 우선순위로 하여 톨게이트 개선방안을 도출하기 위해 4,819건의 데이터 중 교통사고 피해정도가 경미한 사고등급 D 데이터는 제외하여 최종적으로 433건의 데이터를 분석 데이터로 활용하였다. 이때, 433건의 데이터의 경우 분석 데이터로 적을 수 있으나 기존 연구를 검토한 결과 433건 이하의 텍스트 데이터를 활용하여 유의미한 연구 결과를 도출한 사례가 다수 존재하였다(Addiga and Bagui, 2022; Go and Hong, 2022; Lee et al., 2024). 또한, 텍스트 데이터의 경우 데이터 해석 가능성이나 패턴 발견을 위해 주로 사용되고 있어 433건의 데이터는 교통사고 원인과 패턴을 분석하는 데 적절한 크기라고 판단하였다.

### 2. 교통사고 요인 기초 통계분석

한국도로공사에서는 주요 교통사고 원인을 운전자, 차량, 기타 등 세 가지 요인으로 구분하고 있다. 요인별 주사고 원인으로는 운전자의 경우 주시태만, 기타, 졸음, 과속, 안전거리 미확보, 추월불량 순으로 운전자 부주의로 인한 사고가 다수 발생한 것으로 나타났다. 다음으로 차량의 경우 제동장치 결함, 기타, 단독차량 화재, 차량부품 이탈 순으로 차량 자체 오작동으로 인한 사고가 다수 발생했다. 마지막으로 기타요인의 경우 기타, 적재불량, 무단보행, 노면잡물, 포트홀, 도로사정 등의 순으로 도로 기능 불량으로 인한 사고가 발생하였다. 교통사고 요인별 주요 교통사고 원인이 차지하는 건수 및 비율은 다음 <Table 1>과 같다.

<Table 1> Result of statistical analysis of traffic accident factors

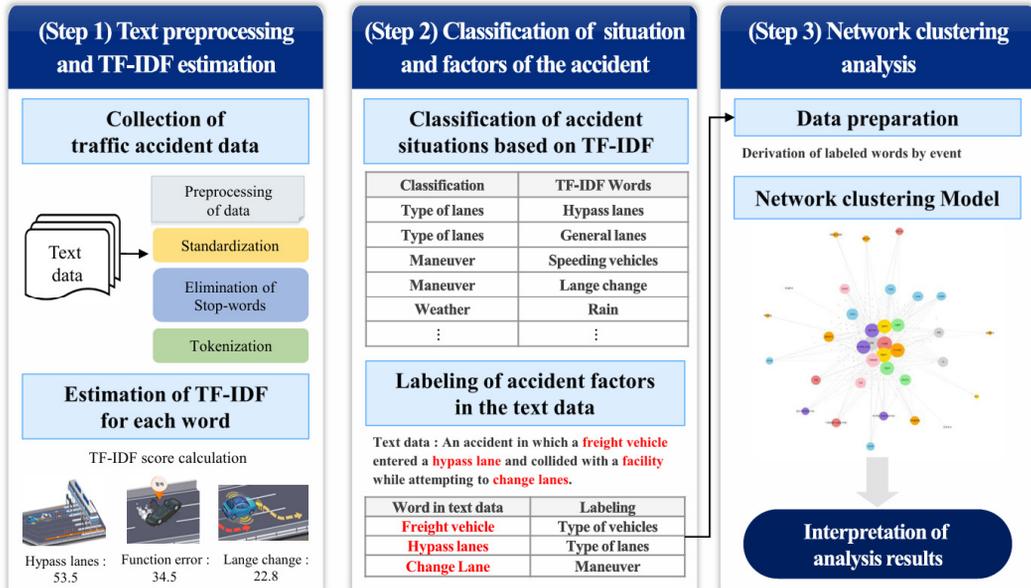
Classification	Traffic accident factors	Frequency	Ratio (%)
Driver	Speeding	23	5.3
	Failure to secure safe distance	2	0.5
	Drowsiness	52	12.0
	Attention-free	251	58.0
	Improper overtaking	1	0.2
	Others	53	12.2
Vehicles	Single-vehicle fire	6	1.4
	Brake system failure	10	2.3
	Vehicle part detachment	1	0.2
	Others	9	2.1
Others	Road debris	3	0.7
	Road conditions	1	0.2
	Jaywalking	3	0.7
	Overloaded vehicle	6	1.4
	Pothole	2	0.5
	Others	10	2.3
Total		433	100

### 3. 교통사고 요인 유형화 및 교통상황 도출

#### 1) 개요

<Table 1>에서 제시된 교통사고 요인 및 주 교통사고 원인만을 가지고는 구체적인 교통사고 상황을 확인할 수가 없다. 따라서 본 연구에서는 텍스트로 교통사고 경위를 자세하게 작성한 교통사고 내용 데이터를 분석하여 차량이 실제 마주할 수 있는 다양한 교통사고 상황들을 도출하고 이러한 교통사고를 방지하기 위한 톨게이트 개선방안을 제시하고자 한다.

이를 위해 교통사고 내용에 대해 텍스트 전처리를 우선적으로 수행하고자 한다. 텍스트 전처리는 모델의 효율성을 높이고 용도에 맞는 데이터 분석을 위해 필요한 작업이다. 그다음으로 텍스트마이닝 기법 중 TF-IDF를 사용하여 단어별 중요도를 계산 후, 중요도가 높은 순서대로 100개의 단어를 도출하여 교통사고 상황을 유형화하는 기준을 도출하고자 한다. TF-IDF란 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지 나타내는 데 유용한 기법이다(Choi et al., 2019). 100개 단어 선정 기준은 기존 연구사례를 검토한 결과이며, 기존 연구에서는 보통 50개에서 200개 사이의 단어들을 연구자 주관에 의해 활용하고 있어 본 연구에서는 100개의 단어를 도출하고자 한다(Liang and Niu, 2022; Zhou et al., 2024, Kang, 2024, Go and Hong, 2022). 여기서, 100개 단어의 신뢰성을 확보하기 위해 상위 100개의 단어가 전체 433건의 텍스트 데이터를 얼마나 커버할 수 있는지 확인하는 절차를 수행하고자 한다. 마지막으로 교통사고 상황 유형화 기준을 토대로 사고내용 주요 단어에 유형화 기준을 라벨링 후 네트워크 클러스터링을 수행하여 연관성 높은 단어 간의 조합을 통해 교통사고상황을 도출하고자 한다. 텍스트 분석 및 교통사고 상황 도출 수행 절차는 다음 <Fig. 3>과 같다.



<Fig. 3> Categorization of traffic accident factors and situations, and network clustering procedure

## 2) TF-IDF 기반 주요 단어 도출

본 연구에서는 <Table 2>에 해당하는 한국어로 작성된 원시 텍스트 데이터 전처리를 수행하기 위해 <Fig. 3> 절차에 따라 용어 통일, 오타 수정, 특수문자 및 알파벳 제거, 불용어 처리, Mecab-ko 기반 한국어 형태소 분석 순으로 전처리를 수행하였다. 용어통일 내용으로는 착오/착각/오인/잘못진입 등은 착오로, 갓길과 길어깨는 갓길로, 화물하이패스는 화물차 하이패스 등과 같이 의미가 같으나 다르게 표현된 단어와 띄어쓰기(급진입과 급진입 등) 위주로 용어통일을 수행하였다. 불용어로는 차종, 노선명, 한국어 조사 등을 불용어로 처리하였다. 이때, 불용어 처리란 자연어 분석 시 의미가 없는 단어를 제거하는 과정이다. Mecab-ko의 경우 사용자 사전을 통해 교통분야에서 사용되는 전문용어(하이패스차로, 가드레일, 축중차로 등)를 정의하였다.

<Table 2> Exemples of traffic accident raw data

Classification	Raw data
1	The accident occurred as the vehicle was entering tollbooth 3 of the Guri towards Toegyewon direction, where it collided with the vehicle classification device installed on the right side. The vehicle came to a stop facing 12 o'clock direction.
2	The accident involved the patrol car (#Vehicle 1) entering the Namyangsan tollgate entrance via the Hi-Pass lane 61. #Vehicle 3 (sedan) abruptly changed lanes from lane 2 to Hi-Pass lane 61, causing #Vehicle 2 (sedan) to brake suddenly. #Vehicle 1 then rear-ended the rear of Vehicle 2 on the right front side, and all vehicles came to a stop facing the 12 o'clock direction.
3	The accident involved a tanker truck (accident vehicle) entering the Hi-Pass lane at Seoul San toll plaza. While attempting to re-enter the weigh-in-motion lane after mistakenly entering the turnaround lane, the vehicle collided with the barrier on the turnaround lane due to driver inattention.
4	The accident involved a passenger vehicle (sedan) traveling from Naengjeong to W.Gimhae IC. near the W.Gimhae IC tollgate exit, the vehicle abruptly changed lanes from lane 3 to lane 1, striking a traffic sign and coming to a stop facing 6 o'clock direction in lane 1.

전처리를 수행 후 TF-IDF 기반 100개의 중요 단어를 도출하였다. 그다음으로는 TF-IDF 기반으로 도출된 100개의 중요 단어들에 대해 해당 단어들에 실제 교통사고 내용을 얼마나 커버하고 있는지 확인하였다. 그 결과 교통사고 내용 별 평균 75%의 TF-IDF의 중요 단어를 포함하고 있음을 확인하였다. 따라서, 해당 단어들은 각 교통사고 내용의 특성을 포괄적으로 설명할 수 있는 중요한 정보를 포함하고 있으며, 교통사고를 묘사할 수 있는 충분한 근거임을 확인하였다. 다음 <Table 3>는 100개 중요 단어 중 20개의 단어와 TF-IDF 결과 예시를 보여준다.

<Table 3> Examples of TF-IDF estimation result

Ranking	Word	TF-IDF	Ranking	Word	TF-IDF
1	Hi-Pass lane	51.607	11	Equipment	23.004
2	Entrance	48.117	12	Facade	20.746
3	Facility	47.241	13	Barrier	14.350
4	Exiting	34.015	14	Collision	12.897
5	Exit	31.916	15	Drowsiness	12.805
6	Ahead	30.592	16	Entrance	12.246
7	Freight	29.845	17	Side	11.126
8	Shoulder lane	27.092	18	Guard rail	10.491
9	Diverging area	23.275	19	Weigh-in-motion lane	10.400
10	Toll booth	23.034	20	Trailer	10.167

### 3) 텍스트마이닝 기반 교통사고 유형화 및 네트워크 클러스터링을 통한 교통사고 상황 도출

TF-IDF 기반으로 도출된 단어를 활용하여 교통사고 상황을 분류하는 여섯 가지의 기준(사고원인, 사고위치, 사고주체 차량, 충격대상, 정차구간, 날씨)과 기준별 구체적인 교통사고 요인을 도출하였다. 그다음으로 433건의 교통사고 내용에 교통사고 유형화 기준을 라벨링하여 네트워크 클러스터링을 위한 데이터를 구축하였다. 키워드 간 연관성 높은 교통사고 상황을 도출하기 위해 키워드 간 연관성을 파악하여 클러스터링을 수행할 수 있는 커뮤니티 탐지 알고리즘(louvain)을 통해 단어들을 분석하였다. 해당 알고리즘은 대규모 네트워크에서 커뮤니티를 추출하기 위한 알고리즘으로, 밀집하게 묶여있는 커뮤니티를 추정하는 것을 의미한다. 구체적으로는 계층적 커뮤니티 구조를 확장하여 다양한 커뮤니티를 군집성(modularity) 최적화를 통해 감지하는 알고리즘으로, 커뮤니티 추정 방법은 다음과 같다(Blondel et al., 2008).

우선, 한 노드를 인접한 다른 커뮤니티에 할당하여 군집성을 측정 후 군집성이 커지는 커뮤니티가 있으면 해당 노드는 해당 커뮤니티에 속하게 되며, 모든 노드에 대해 순차적으로 수행하여 군집성이 가장 높아질 때까지 반복하게 된다. 두 번째로는 첫 단계에서 생성된 커뮤니티를 가지고 새로운 커뮤니티를 생성한 후에 기존 커뮤니티 간 연결되어있던 링크 가중치를 합쳐 하나의 링크로 생성한다. 그다음으로 동일한 커뮤니티 노드 간 링크는 새로운 네트워크에서 해당 커뮤니티에 대한 self-loops로 대체하여 군집성이 더 이상 증가하지 않을때까지 두 단계를 반복하여 최종 커뮤니티를 생성하게 된다(Blondel et al., 2008).

Louvain 알고리즘을 적용하여 클러스터링을 수행한 결과, 총 세 개의 그룹으로 클러스터링 되었다. 그룹별 라벨링 및 키워드별 링크 가중치와 유형화 기준을 매칭한 결과는 다음 <Table 4>과 같다. 그룹 1의 경우 “혼합차로에 의한 화물차의 목표차로 주행 제한”으로 라벨링을 수행하였다. 해당 그룹의 주요 단어들을 살펴보면 맑은 날씨, 화물차로 인한 교통사고 발생, TCS 차로, 차대 시설물 충돌, 진입부, 운전자의 주시태만, 길어

개 내 정차, 3차로, 하이패스와 축중차로가 혼용된 차로 등의 순으로 링크 가중치가 높게 나타났다. 이때, 링크 가중치가 높을수록 군집 내에서 키워드 간 연관성이 높다는 것을 의미한다. 그 다음으로 그룹 2는 “기상 악화 및 운전자 부주의에 의한 2차 교통사고 유발”로 라벨링을 수행하였다. 해당 그룹에서는 교통사고 발생 후 진행방향과 동일한 지점에서 정차, 승용차로 인한 교통사고 발생, 진출부, 하이패스차로, 비, 시설물 및 차량과의 다중충돌, 1차로 및 2차로, 줄임운전, 6차로, 일반과속 등의 순으로 링크 가중치가 높게 나타난 것으로 보아 진행방향과 동일한 정차지점, 승용차, 진출부 등이 군집 내 연관성이 높은 것으로 판단된다. 마지막으로 그룹 3의 경우 “차량 및 도로 결함에 의한 단독사고”로 라벨링을 진행하였다. 주요 단어들을 살펴본 결과 본선TG, 호림, 광장부, 차대차 충돌, 4차로, 단독사고, 제동장치결함, 컷인, 시설물 불량, 핸들과대/급조작 등의 순으로 링크 가중치가 높게 나타났다. 즉, 본선 TG와 호림 날씨, 광장부, 차대 차 충돌 등이 그룹 내 연관성이 높은 것으로 판단된다.

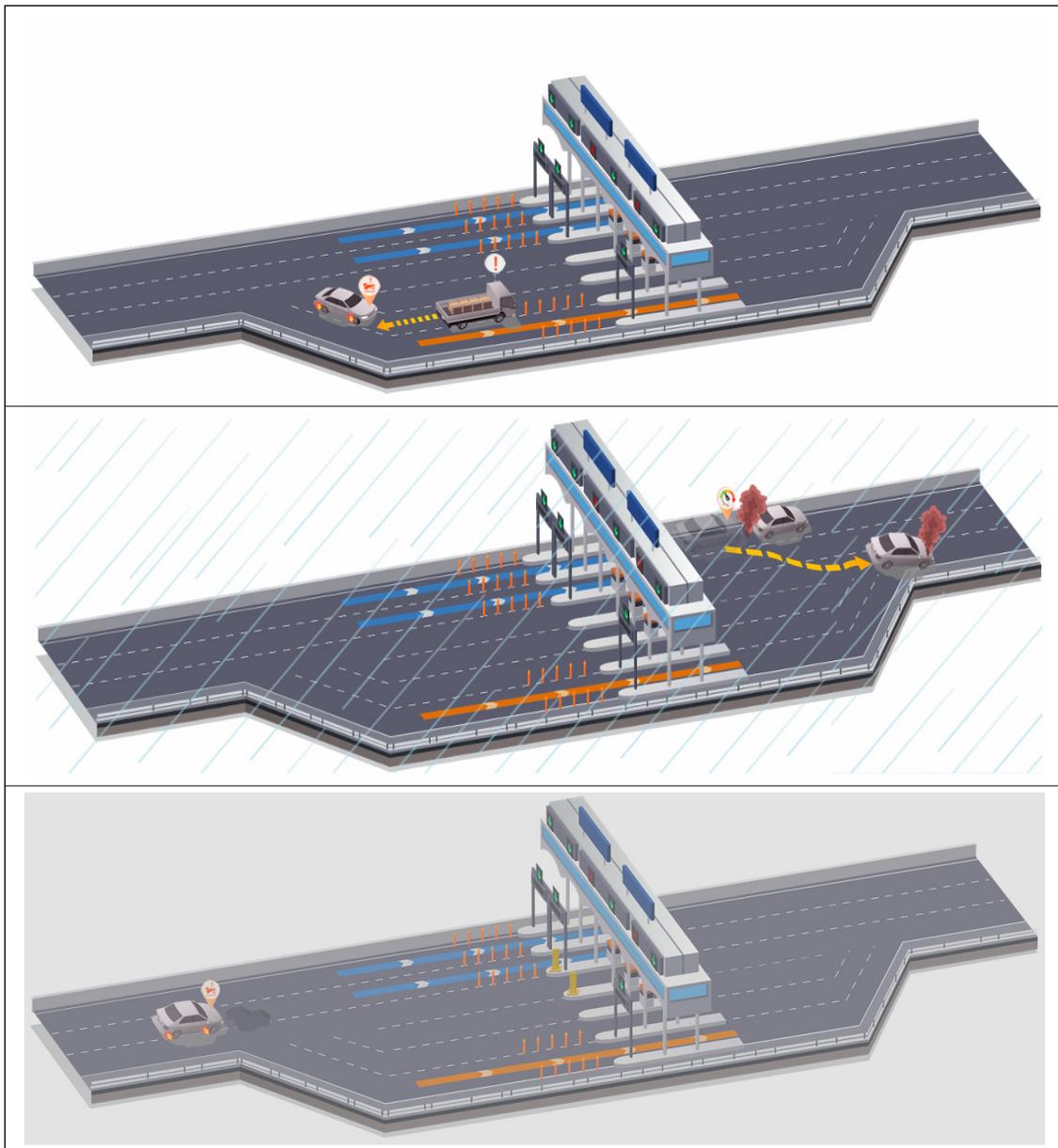
<Table 4> Keyword clustering results based on the louvain algorithm

Group	Keywords	Link weight	Accident types
Restriction of freight truck's target lane driving due to mixed lanes	Sunny	0.2129	Weather
	Freight	0.2081	Accident subject vehicle
	TCS	0.2022	Lane type
	Facility	0.2016	Collision target
	Entrance	0.2016	Accident section
	Attention-free	0.1607	Driver carelessness
	Shoulder lane	0.1602	Non-drive section
	Lane 3	0.1554	Driving lane
	Hi-pass lane and Weigh-in-motion lane	0.1281	Lane type
	Lane 5	0.1104	Driving lane
	TCS and Axle load lane	0.1052	Lane type
	Novice driver	0.1042	Driver carelessness
	Bus	0.1022	Accident subject vehicle
	Poorly loaded	0.1006	Driver carelessness
	Entering wrong-way	0.0096	Driver carelessness
Secondary accident occurrence due to worsening weather and driver carelessness	Proceeding Direction	0.2024	Non-drive section
	Passenger car	0.2021	Accident subject vehicle
	Diverging section	0.1973	Accident section
	Hi-pass lane	0.1853	Lane type
	Rainy	0.1596	Weather
	Multiple collision(facility/vehicle)	0.1396	Collision target
	Lane 1	0.1360	Driving lane
	Lane 2	0.1331	Driving lane
	Driver sleepiness	0.1295	Driver carelessness
	Lane 6	0.1080	Driving lane
	Speeding under normal conditions	0.1003	Driver carelessness
	Drunk driving	0.0999	Driver carelessness
	Closed TCS lane	0.0990	Lane type

Group	Keywords	Link weight	Accident types
	Speeding under rainy conditions	0.0903	Driver carelessness
	Pedestrian	0.0889	Driver carelessness
Single-vehicle accident due to vehicle and road defects	Mainline TG	0.1799	Accident section
	Cloudy	0.1670	Weather
	Toll plaza section	0.1566	Non-drive section
	Vehicle	0.13890	Collision target
	Lane 4	0.1316	Driving lane
	Single vehicle accident	0.1300	Collision target
	Brake system failure	0.1190	Driver carelessness
	Cut-in	0.1168	Driver carelessness
	Facility defect	0.1004	Driver carelessness
	Excessive and abrupt steering maneuvers	0.0996	Driver carelessness
	Lane 10	0.0949	Driving lane
	Safety zone	0.0916	Non-drive section
	Lane 7	0.0895	Driving lane
	Maintenance deficiency	0.0864	Driver carelessness
	Lane 8	0.0842	Driving lane

#### 4. 소결

“혼합차로에 의한 화물차의 목표차로 주행 제한” 그룹의 단어들을 살펴보면 화물차 및 버스와 같은 대형 차가 하이패스 혹은 일반차로 및 축중차로가 혼용된 차로로 진입 중 적재물 불량 혹은 착오진입 등의 운전자 부주의로 인해 시설물을 충격 후 갓길로 정차하는 사고가 발생할 수 있음을 보여준다. 실제로 착오진입 사례를 살펴보면 화물차의 경우 사각지대로 인한 주시태만과 착오진입으로 인해 시설물을 충격하는 경우가 가장 많았으며, 충돌로 인해 파손된 시설물이 타 차량의 운전을 방해하여 2차 교통사고로 이어지고 있었다. 이는 화물차의 오른쪽 차로 주행으로 인해 과도한 차선변경을 수행하면서 교통사고가 발생한 것으로 판단된다. 또한, 교통사고 위치를 살펴보면 주행차로 중 3, 5차로가 연관성이 있는 것으로 보아 TCS 차로와 하이패스차로가 분·합류되는 변이구간에서 운전자 부주의 혹은 ETC와 TCS 차량이 목표 차로를 탐색하면서 사고가 발생한 것으로 판단된다. “기상악화 및 운전자 부주의에 의한 2차 교통사고 유발” 그룹의 단어들을 종합해보면 비가 오는 날에 승용차가 하이패스차로를 통과하면서 과속, 음주운전, 졸음운전 등 운전자 부주의로 인해 시설물 및 차량 등을 다중충돌하여 진행방향에서 정차하는 사고가 발생할 수 있음을 보여준다. 특히, 빗길과 속으로 인한 충돌사고가 발생하는 것으로 보아 악의환경 또한 교통사고에 영향을 미치는 것으로 판단된다. 또한, 교통사고 위치 중 1, 2차로가 나타난 것으로 보아 교통량과 진출 후의 차로 합류가 연관성이 큰 것으로 판단된다. 마지막으로 “차량 및 도로 결함에 의한 단독사고” 그룹의 경우에는 흐린 날씨에 본선 광장부에서 TG로 주행 중 제동장치결함, 정비불량 등 차량 자체결함으로 인한 교통사고와 도로 시설물 불량으로 인해 핸들과대/급조작을 유발하는 등의 단독사고가 발생할 수 있음을 보여준다. 다음 <Fig. 4>는 그룹별 분석 결과에 대한 도식화 예시이다.



<Fig. 4> Examples of tollgate accident diagram by groups

## IV. 톨게이트 개선방안 제안

### 1. 톨게이트 개선 기본방향

교통사고 상황 및 요인을 분석한 결과 향후 톨게이트의 교통 안전성을 높이기 위해서는 설계 및 운영관점

으로 구분하여 개선방안을 도출할 필요가 있다. 특히, 단계별(단·중기 및 장기)로 적합한 톨게이트 설계 및 운영 개선방안을 도출하여 실효성을 높일 필요가 있다. 따라서 본 연구에서는 국내 도로 관련 정책 및 시장 동향을 고려하여 단·중기 및 장기로 구분하여 개선방안을 제안하고자 한다.

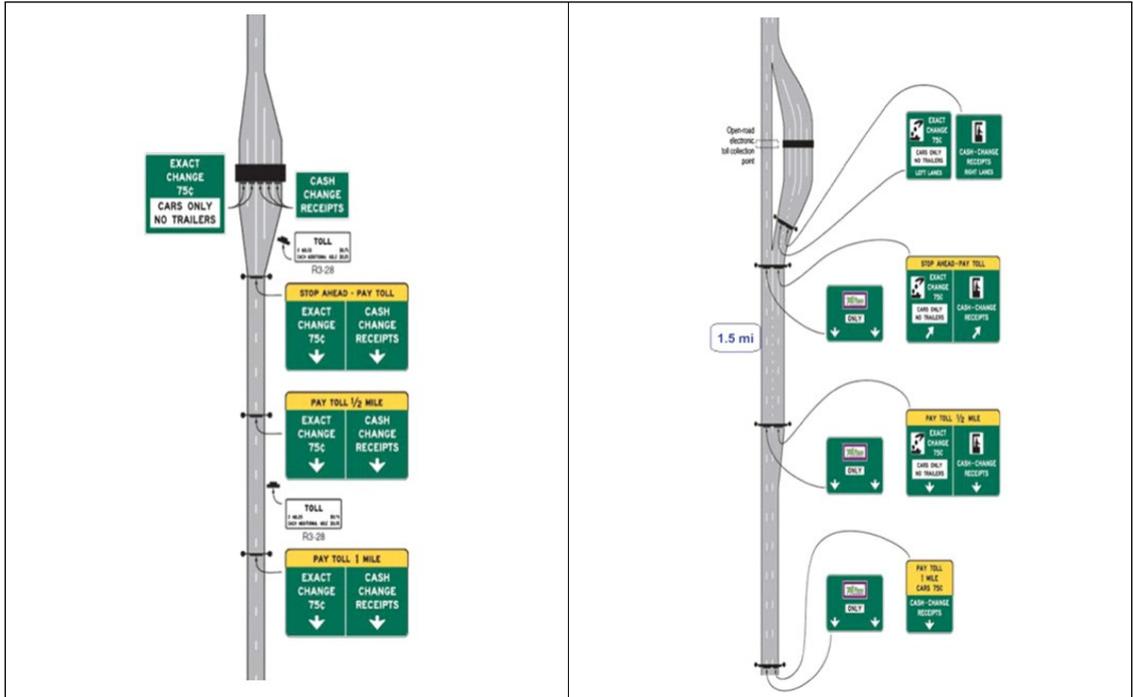
국내 도로개편 관련 주요 정책을 살펴보면 다음과 같다. 「자동차·도로교통분야 지능형교통체계 기본계획」에서는 2025년까지 도로 위험 정보를 지원하는 ITS 서비스를 고속도로, 국도, 지방도 등에 확대하고, 교통사고 위험 경고를 지원하는 C-ITS 서비스를 주요도로에 구축할 계획이다. 또한, 2030년까지 C-ITS 서비스를 전국적으로 확대해 나갈 계획이다(Ministry of Land, Infrastructure and Transport, 2022).

따라서 앞서 언급한 정책 일정을 기반으로 단·중기는 2025년부터 2029년까지, 장기는 2030년부터 2035년까지로 설정하여 구체적인 체계적인 톨게이트 개선방안을 제시하고자 한다. 특히, 단·중기적으로는 ITS 및 C-ITS 서비스 발굴 등의 도로 운영관점에서의 개선방안 그리고 장기적으로는 도로 기하구조 등 도로 설계관점에서의 개선방안을 도출하여 시의적절한 변화상을 수립하고자 한다.

## 2. 단계별 톨게이트 개선방안

단·중기 개선방안으로는 C-ITS 구축 정책 일정에 맞춰 서비스 추가 발굴을 제안한다. 현재 국토교통부에서 발표한 국가 ITS 아키텍처 3.0에 따르면 도로 위험구간, 사고 및 재난 대응·관리, 교통류 최적화, 특수 목적형 차량 이동지원 등 다수 서비스를 언급하고 있다(Ministry of Land, Infrastructure and Transport, 2023). 다만, 네트워크 클러스터링 결과를 살펴보면 목표하는 차로로 주행하고자 하는 차량들이 진입 및 진출부에서 착오진입 혹은 진출 후 합류 시 급격한 주행행태를 보이면서 교통사고가 다수 발생하고 있다. 특히, 화물차의 경우 오른쪽 차로 통행으로 인해 화물차가 과도한 주행행태를 수행하면서 교통사고가 발생하고 있다. 따라서 차종별 안전주행을 지원할 수 있는 C-ITS 서비스가 필요할 것으로 판단된다. 본 연구에서는 이러한 상황에 대응할 수 있는 톨부스 개폐 여부 및 차종별 주행 가능 차선 알림을 지원하는 C-ITS 서비스를 제안한다. 또한, 하이패스차로 통과 후 타 차로를 주행하던 차량들이 합류를 시도하면서 발생하는 교통사고를 방지하기 위해 다차로 하이패스 확대 추진을 제안한다.

장기적 개선방안으로는 도로의 기하구조 개선을 제안한다. 현재 국내 톨게이트의 경우 TMTP와 HMTP가 혼용된 방식으로 복잡한 기하구조로 설계되어 있다. 특히, TCS 차로에서는 TCS 차로와 화물차 차로(하이패스 및 축중차로), 혼용차로(TCS/하이패스, 승용차/화물 등)가 함께 운영되면서 이로 인한 주행 혼란이 지속적으로 발생하고 있다. 이로 인해 향후 TCS 차로를 통행하는 화물차의 경우 주행경로 혼란으로 인해 가감속 및 급차선변경이 지속적으로 발생할 것으로 예상된다. 따라서 전면적으로 톨게이트를 개선할 필요가 있으며, 최근 확대되고 있는 다차로 하이패스 운영 형태를 준용하되 TCS 차로에서는 TCS와 화물차 차로를 분리하는 완전 분리형 톨게이트로 개선하는 것을 제안한다. 이는 향후 교통사고 감소 효과와 더불어 차로 분리를 통해 ETC 차량의 신속한 통과가 가능해져 ETC 차량의 선호도가 올라갈 것으로 판단된다. 이를 통해 톨게이트 운영효율을 높일 것으로 사료된다. 다음 <Fig. 5>는 Abuzwidah et al.(2014)가 제시한 완전 분리형 톨게이트 예시이다.



Source: Safety Evaluation of Hybrid Main-Line Toll Plazas(Abuzwidah et al., 2014)

<Fig. 5> Example of tollgate improvement before and after

## V. 결론 및 향후 계획

국내에서는 톨게이트 교통사고가 지속적으로 발생함에 따라 해당 구간의 안전한 주행을 지원하기 위한 도로체계 개편이 검토되어오고 있다. 특히, 고속도로 톨게이트의 경우 다차로 하이패스 운영을 통해 TCS와 ETC 차선을 구분하여 ETC 차량이 정지하지 않고 곧바로 통과할 수 있도록 지원하고 있다. 이외에도 최근 일부 구간에서 C-ITS 중 스마트 톨링 서비스를 통해 번호판 인식방식의 무정차 통행료 결제 시스템을 적용하여 안전하고 신속한 주행을 도모하고 있다. 다만, 국내 정책을 살펴보면 고속도로 중 교통사고 발생 비율이 높은 톨게이트 구간의 교통사고 위험을 줄이기 위한 대응책을 발표하고 새로운 서비스를 도입하고 있으나 획기적인 개선은 지지부진한 상태이다. 이는 단순히 운전자, 차량, 기타요인 등 세 가지 요인별 주요 교통사고 원인만을 고려하고 사고위치, 사고주체 차량의 행태, 날씨 등 구체적인 상황을 고려하지 않았기 때문이라고 판단된다. 톨게이트 구간은 단거리 내 분·합류, 차선수 감소 및 증가, 개폐된 톨버스 등 주행 시 인지해야 할 상황과 이벤트들이 다수 존재해 복합적으로 상황과 요인 등 다양한 요소들을 고려하여 대응책을 마련할 필요가 있다.

본 연구에서는 한국도로공사의 고속도로 교통사고 요인 분류 기준인 운전자, 차량, 기타 등 세 가지 요인들을 교통사고 상황과 상황별 교통사고 요인들을 고려하여 재구성하고자 하였다. 이를 위해 교통사고 상황이 자세히 작성되어 있는 한국도로공사 교통사고 데이터 중 사고내용을 분석데이터로 활용하였다. 교통사고 내용을 분석하기 전 분석 정확도를 높이기 위해 전처리를 수행하였으며, 그 이후 TF-IDF 기법을 통해 교통

사고 내용에 작성되어 있는 주요 단어별 상대적 중요도 점수를 추정하였다. 중요도 점수를 기준으로 상위 100개의 단어를 추출 후 이를 토대로 교통사고 상황을 유형화하는 여섯 가지 기준을 도출하였다. 그다음으로 교통사고 내용 데이터에 여섯 가지 사고상황 기준을 라벨링하여 상황별 주요 교통사고 요인들을 도출하고 요인 간 연관성을 살펴보기 위해 네트워크 클러스터링을 수행하였다. 그 결과, 총 세 개의 그룹으로 군집화되었으며 그룹별로 “혼합차로에 의한 화물차의 목표차로 주행 제한”, “기상악화 및 운전자 부주의에 의한 2차 교통사고 유발”, “차량 및 도로 결함에 의한 단독사고” 등으로 라벨링을 수행하였다.

그룹별 연관성이 높은 단어들을 살펴본 결과 첫 번째 그룹의 경우 화물차 및 버스와 같은 대형차가 하이패스 혹은 일반차로 및 축중차로가 혼용된 차로로 진입 중 적재물 불량 혹은 착오진입 등의 운전자 부주의로 인해 시설물을 충격 후 갓길로 정차하는 사고가 발생할 수 있을 것으로 판단되었다. 두 번째 그룹의 경우 비가 오는 날에 승용차가 하이패스차로를 통과하면서 과속, 음주운전, 졸음운전 등 운전자 부주의로 인해 시설물 및 차량 등을 다중충돌하여 진행방향에서 정차하는 사고가 발생할 수 있을 것으로 판단되었다. 마지막 그룹의 단어들을 살펴본 결과 흐린 날씨에 본선 광장부에서 TG로 주행 중 제동장치결함, 정비불량 등 차량 자체결함으로 인한 교통사고와 도로 시설물 불량으로 인해 핸들과대/급조작을 유발하는 등의 단독사고가 발생할 수 있을 것으로 판단되었다.

위에서 언급한 결과를 기반으로 톨게이트 개선방안을 단·중기와 장기로 구분하여 제안하였다. 단·중기에는 C-ITS 서비스 발굴이 필요하며 특히, 주행혼란을 방지할 수 있는 톨부스 개폐여부 및 차종별 주행가능 차선 알람을 지원하는 C-ITS 서비스 도출이 필요할 것으로 판단된다. 이외에도 하이패스 통과 후 안전한 차량들의 합류를 지원할 수 있는 다차로 하이패스 확대 추진을 제안한다. 장기적으로는 도로의 기하구조 개선을 제안하였다. 화물차로 인한 교통사고 위험을 줄이기 위해 기존 국내 톨게이트 설계방식을 탈피하여 TCS 차로에서도 TCS와 화물차 차로를 분리하는 방향으로 기하구조 설계가 변화되어야 할 것으로 판단된다. 본 연구에서 제안한 개선방안은 향후 톨게이트 개선방안 수립의 기초자료로서 활용될 수 있을 것으로 기대된다.

본 연구의 한계점 및 향후 연구과제는 다음과 같다. 본 연구에서는 교통사고 유형화를 연구자의 주관으로 수행하여 유형화된 결과에 대한 객관성을 확인하지 못하였다는 한계점이 존재한다. 향후에는 대규모 언어 모델(large language models, LLM)을 활용하여 객관성을 확보할 수 있는 연구를 추가적으로 진행할 필요가 있다.

## ACKNOWLEDGEMENTS

본 연구는 국토교통부 및 국토교통과학기술진흥원의 연구비 지원(RS-2021-KA160637)에 의해 수행하였습니다.

## REFERENCES

- Abuzwidah, M. and Abdel-Aty, M.(2018), “Crash risk analysis of different designs of toll plazas”, *Safety Science*, vol. 107, pp.77-78.
- Abuzwidah, M., Abdel-Aty, M. and Ahmed, M.(2014), “Safety evaluation of hybrid main-line toll plazas”, *Transportation Research Record*, vol. 2435, no. 1, pp.53-60.
- Addiga, A. and Bagui, S.(2022), “Sentiment analysis on Twitter data using term frequency-inverse

- document frequency”, *Journal of Computer and Communications*, vol. 10, no. 8, pp.117-128.
- Blondel, V., Guillaume, J., Lambiotte, R. and Lefebvre, E.(2008), “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, pp.3-6.
- Choi, W., Yoo, K. and Choi, S.(2019), “Create list of stopwords and typing error by TF-IDF weight value”, *EasyChair*, pp.1-5.
- Esenturk, E., Turley, D., Wallace, A., Khastgir, S. and Jennings, P.(2023), “A data mining approach for traffic accidents, pattern extraction and test scenario generation for autonomous vehicles”, *International Journal of Transportation Science and Technology*, vol. 12, no. 4, pp.955-972.
- Giummarra, J., Beck, B. and Gabbe, J.(2021), “Classification of road traffic injury collision characteristics using text mining analysis: Implications for road injury prevention”, *PloS One*, vol. 16, no. 1, pp.1-19.
- Go, W. and Hong, J.(2022), “Framework for key performance index of use cases in smart transportation using text mining”, *Korean Society of Transportation*, vol. 40, no. 4, p.452.
- Kang, J.(2024), “Analysis of the maintenance factors of the apartment housing exclusive facilities using text mining”, *Journal of the Korean Housing Association*, vol. 35, no. 1, pp.79-88.
- Korea Expressway Corporation(2021), *Impact analysis and policy direction study for next-generation expressway tolling system*, pp.15-29.
- Korea Road Traffic Authority(2023), *Statistical analysis of traffic accidents*, p.200.
- Kwayu, M., Kwigizile, V., Lee, K. and Oh, S.(2021), “Discovering latent themes in traffic fatal crash narratives using text mining analytics and network topology”, *Accident Analysis & Prevention*, vol. 150, pp.1-17.
- Lee, S., Jeong, Y., Kim, J. and Kim, J.(2024), “Urban railway accident severity analysis using ordered probit model and XGBoost”, *Journal of the Korean Society for Railway*, vol. 27, no. 21, pp.162-173.
- Lee, Y., Lee, Y., Seong, J., Stanescu, A. and Hwang, C.(2020), “A comparison of network clustering algorithms in keyword network analysis: A case study with geography conference presentations”, *International Journal of Geospatial and Environmental Research*, vol. 7, no. 3, pp.1-16.
- Liang, M. and Niu, T.(2022), “Research on text classification techniques based on improved TF-IDF algorithm and LSTM inputs”, *Procedia Computer Science*, vol. 208, pp.460-470.
- Ministry of Land, Infrastructure and Transport(2021), *1st National road comprehensive Plan*, p.77.
- Ministry of Land, Infrastructure and Transport(2022), *Intelligent Transportation System(ITS) implementation plan*, pp.71-415.
- Ministry of Land, Infrastructure and Transport(2023), *National ITS architecture 3.0*, pp.4-13.
- Ministry of Land, Infrastructure and Transport(a), [https://www.molit.go.kr/USR/NEWS/m\\_72/dtl.jsp?lcmspage=2&id=95083240](https://www.molit.go.kr/USR/NEWS/m_72/dtl.jsp?lcmspage=2&id=95083240), 2024.07.15.
- Ministry of Land, Infrastructure and Transport(b), [https://www.molit.go.kr/USR/NEWS/m\\_71/dtl.jsp?id=95089803](https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?id=95089803), 2024.07.15.
- Ministry of Land, Infrastructure and Transport(c), [https://www.molit.go.kr/USR/NEWS/m\\_71/dtl.jsp?id=95089803](https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?id=95089803), 2024.07.15.
- Ministry of Land, Infrastructure and Transport(d), [https://www.molit.go.kr/USR/NEWS/m\\_71/dtl.jsp](https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp)

?id=95085605, 2024.07.15.

Song, P., Sze, N., Zheng, O. and Abdel-Aty, M.(2022), “Addressing unobserved heterogeneity at road user level for the analysis of conflict risk at tunnel toll plaza: A correlated grouped random parameters logit approach with heterogeneity in means”, *Analytic Methods in Accident Research*, vol. 36, pp.1-19.

The Korea Transport Institute(2018), *Efficient utilization of expressway tollgate space by introduction of smart tolling system*, p.1.

Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y. and Zheng, O.(2019), “Examining traffic conflicts of up stream toll plaza area using vehicles trajectory data”, *Accident Analysis & Prevention*, vol. 125, pp.174-176.

Zhou, J., Ye, Z., Zhang, S., Geng, Z., Han, N. and Yang, T.(2024), “Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data”, *Heliyon*, vol. 10, no. 16, pp.1-22.