Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6880310, 19 pages
https://doi.org/10.1155/2022/6880310

Research Article

Development of Risk-Situation Scenario for Autonomous Vehicles on Expressway Using Topic Modeling

Osung Chae,[1] Junghwa Kim,[1] Jeongah Jang,[2] Hyunjeong Yun,[3] and Shinkyung Lee[3]

[1]*Kyonggi University, Department of Urban & Transportation Engineering, Suwon, Kyonggi, Republic of Korea*
[2]*Ajou University, TOD-based Sustainable City/Transportation Research Center, Suwon, Republic of Korea*
[3]*ETRI Autonomous Driving Intelligence Research Section, Daejeon, Republic of Korea*

Correspondence should be addressed to Junghwa Kim; junghwa.kim@kyonggi.ac.kr

Received 20 April 2022; Accepted 18 July 2022; Published 29 August 2022

Academic Editor: Yuchuan Du

Growing interest has recently been paid to the development of autonomous vehicle scenarios, and corresponding research is being conducted on various methodologies and on the generation of scenarios including technological elements. However, most studies have focused on frequently-occurring accident types or representative accident situations; thus, there is a lack of studies on scenarios considering unpredictable accidents. Proper preparation is required for accident situations because even a small traffic accident that is less likely to occur can lead to fatalities if it is difficult to predict. Accordingly, this study established accident situations based on the Pegasus layer model by using unstructured text data to explain traffic accidents on expressways in Korea. The established accident situations were classified into three types (Typical Traffic/Critical Traffic/Edge Case) according to frequency. Topic modeling was applied to the Edge Case class, i.e., the least likely to occur and thus difficult to predict, to analyze the characteristics of groups and develop risk-situation scenarios for autonomous vehicles based on actual accident data.

1. Introduction

The quality of life has been consistently improving owing to advancements in Internet of Things technologies. In this context, integrating information technology with automobiles has led to the development of autonomous vehicles (AVs) which do not require human operation. The market size for AVs is growing every year owing to reductions in traffic accidents, improved accessibility to driving, and smooth travel through autonomous identification of traffic flows (Kwon and Lee) [1]. As the technology of level 2 AVs has become commercialized, the technological development of AVs at higher levels has also been accelerating (McKinsey and Company, 2021). The six levels of AVs range from 0 to 5 according to the "controlling entity" and "responsible entity"; most currently available AVs belong to level 2 (SAE, 2016). When AVs (AVs) become commercialized, the accident rate owing to human factors decreases, as the artificial intelligence (AI) plays the role of the driver. However, the possibility of accidents caused by system errors or part failures owing to AI systems being imperfect must also be considered, along with the traffic accidents occurring on roads where AVs and regular vehicles are mixed. Various studies have been carried out for this purpose. Patel et al. [2] proposed a scenario for embodying different brake strategies of AVs in a mixed traffic flow. Virdi et al. [3] calculated the potential for collision in different scenarios according to the mixed rate of AVs by intersection, and proved that potential for collisions decreased as the ratio of AVs increased. Yu and Li [4] discovered a contemporaneous correlation between the probability of a risk situation and scenario variability by developing a path analysis model, and extracted various elements of the risk scenarios induced by human drivers and AVs by adding a logistic regression model. Song et al. [5] established a systematic approach for identifying important test scenarios based on integrated tools and a workflow. The researchers suggested a method for determining parameters for explaining a scenario rather than depending on assumptions when generating scenarios. Erwin et al. [6] used a scenario representativeness metric based on the Wasserstein

distance to quantify the extent of representing an actual scenario. They generated actual parameters through the estimation of a probability density function (PDF) of the parameters, and then examined whether a scenario including the generated parameters was found in the actual scenarios. The proposed method was proven to be capable of automatically determining the PDF estimation and optimal scenario parameterization relative to other methods involving scenario parameterization and PDF estimation assumptions.

In view of the above, research is continuously being conducted on scenarios for responding to risk situations on roads. However, in general, only operation problems of AVs are considered; complicated road situations and driving environments are not considered when analyzing the risk situations of AVs according to the mixed ratio of AVs. In addition, as most scenarios for accident situations deal with frequently-occurring major accidents, there is relative lack of research on scenarios for unusual situations. Nevertheless, scenarios to prepare for unusual situations are also required, as traffic accidents can be fatal if such situations are not properly prepared for.

Among the traffic accidents that can occur on roads where autonomous vehicles and general vehicles coexist, there are accidents with a low probability of occurrence. Although such accidents occur infrequently, prepare is not possible and is likely to lead to fatal accidents. Since such an accident is difficult to predict, it is difficult to prepare in advance. Therefore, in this study, since it is difficult to predict, we try to classify main lane accidents that are difficult to prepare in advance into scenarios based on the cause of the accident and analyze them. This study used a data range corresponding to the accident data of expressways in Korea to reflect actual traffic situations and driving environments. The Pegasus five-layer format was used to distinguish between the characteristics of accidents in text, for ultimately performing an analysis using topic modeling. To consider unusual cases (i.e., those less likely to occur), cases were classified into three groups (typical traffic, critical traffic, and edge case) depending on the frequency. Representative scenarios were generated for each point of occurrence by recombining through topic modeling of the edge case, as this was the least likely to occur among the classified groups.

1.1. Related Work. Park [7] proposed five types of take-over scenarios for responding to fallback situations. Conditional AV (SAE level 3) used a safety mechanism for generating a take-over request for a driver, so as to take responsive measures when facing emergency situations where normal driving is not possible. Drivers were considered as the entity of the emergency response (fallback-ready user); the safety during the take-over process delivered from the AV to the driver was systematically evaluated to properly respond to emergency situations occurring during automated driving. The experiment results showed that the highest subjective evaluation was demonstrated in a straight-road accident vehicle scenario, proving that this particular scenario is the most urgent, dangerous, and physically/emotionally demanding, and allows for the least amount of time.

Emzivat [8] defined a dynamic driving fallback strategy executable by an automated driving system (ADS) for considering dynamic driving task (DDT) fallback, e.g. for an ADS for which the driving environment monitoring has deteriorated. An ADS significantly reduces the workload of a human driver, but excessively advanced automation may induce sleepiness and negligence. When the function for monitoring the driving environment of an AV is degraded owing to device malfunctioning or inclement weather, proper responsive measures must be taken for malfunctions or unexpected situations arising from road conditions. The time for converting the attention of a user to the driving task must be determined to immediately respond to such unexpected situations. Furthermore, certain drivers do not perceive the limitations and functions of an ADS; therefore, the DDT fallback strategy can be used to lower the accident severity and increase driver safety, as all responsibilities can be given to drivers when converted to the driving task. Most fallback strategies automatically stop vehicles within the currently driving path. However, there are certain situations where vehicles should not be stopped. For example, stopping vehicles in tunnels without a shoulder or on expressways can be dangerous; hence, stopping vehicles should be avoided as much as possible. A fallback strategy was established for various road conditions where an emergency stops can be dangerous, with three methods (ghost vehicle, considering time-to-collision (TTC) standards, or using TTC + ghost vehicle) being considered as fallback strategies. The experimental results showed that an ADS combined with TTC standards and a ghost vehicle was more effective, as the position of the vehicle was more stably determined (e.g., from not immediately deciding, based on malfunctioning for a longer period of time).

Seo and Kim [9] summarized previous studies on the security and main functions of AVs and deduced various attack scenarios from the perspective of an in-vehicle system, aiming to analyze situations requiring security. Security and safety are two factors that must be considered as AV technology rapidly advances. The communication channel for an AV largely consists of external and internal networks. The internal network, or in-vehicle system, is used for controlling vehicle functions. An attack on an in-vehicle system, even from a small threat, may have fatal consequences, as it is directly associated with the safety of the driver. Three scenarios have been suggested for detecting in-vehicle attacks, based on a control system analysis related to road conditions potentially occurring in automated driving environment.

Park et al. [10] developed a methodology for deducing AV test scenarios by utilizing traffic accident data and natural language processing (NLP). Their NLP-based test scenario mining methodology generated scenarios for urban arterial roads and regional intersections; specifically, test scenarios for 16 urban arterial roads and 38 regional intersections were generated. Generally, test scenarios are the core means for evaluating and guaranteeing the driving capabilities of AVs. These test scenarios are used to analyze

validity and efficiency by reflecting road geometries, traffic conditions, and fine vehicle maneuvers. Arterial roads have various dangerous traffic conditions which may degrade the performance of AVs. Hence, AV driving on arterial roads needs to be tested based on the test scenarios for arterial roads as generated in studies.

Nico et al. [11] suggested a scenario evaluation algorithm for analyzing and applying autonomous emergency braking, based on considering the possibility of collision avoidance.

Emre et al. [12] analyzed STATS19 accident data from Great Britain to identify traffic accident patterns, and then used the analyzed data to systematically generate scenarios for improving safety through a connected and automated vehicle test. They used clustering results containing traffic accident characteristics and association rule mining based on a market basket analysis when generating traffic scenarios.

Li. [13] defined an overall framework for scenario generation including definitions, components, data sources, and a data processing method, and further proposed a scenario-based V model. In general, creating test cases in a virtual environment is much more efficient in terms of time and resource consumption relative to building an actual scenario test model. Nevertheless, a problem can arise owing to the very low possibility of accidents occurring in a virtual environment when randomly generating scenarios. Such problem can be solved if scenarios are devised by focusing on risk situations.

Research is continuously conducted on scenarios for responding to risk situations involving AVs (Chae et al., Jeong, Choi and Lim, Kim et al.); [14–17] hence, there is growing interest in and need for research on these scenarios, owing to social changes involving rapid advancements in AV technology. However, most scenario studies have failed to consider complicated road conditions and driving environments, as they have focused only on AV malfunctions or technological issues. Furthermore, the majority of these studies have focused on analysis or standardization of scenarios for frequently-occurring representative cases. Moreover, previous studies focused on the behaviors of AVs and nearby vehicles and traffic conditions at the point of occurrence and surrounding areas when generating scenarios. This is because it is difficult to generalize the nature of scenarios, as subjective judgments must be included while considering other factors as quantitative variables. Therefore, this study constructed an accident dataset based on the Pegasus five-layer format while using the traffic accident data of actual expressways, so as to indirectly reflect complicated road conditions and driving environments. In addition, accident causes, accident types, and environmental conditions which were difficult to generalize in scenarios in the past were considered by characterizing the groups through topic modeling. The difficult-to-predict accident cases with less than a 20% frequency of occurrence were applied using topic modeling, i.e., by grouping the accident dataset according to frequency. During the topic modeling, important keywords were extracted for each layer, and then were recombined to ensure that the risk-situation scenarios reflected actual road conditions and driving environments. The scenarios proposed in this study have a significance in that they not only consider road conditions and driving environments in which AVs and regular vehicles are mixed, but also consider unusual risk situations that are difficult to predict (as the scenarios reflect the actual traffic accident data).

2. Methodology

2.1. Data Collection. This study designed representative scenarios for considering actual road conditions and responding to unpredictable risk situations for automated driving of level 4 or higher. The study aimed to implement unpredictable risk situations in the scenarios by selecting accidents with a low frequency of occurrence among traffic accident cases. Accident investigation data explaining 10,135 cases of traffic accidents occurred on expressways in Korea over a five-year period from 2016 to 2020 were used to reflect actual road conditions. The investigation data were unstructured data including the accident year, road alignment, road inclination, cut and fill, accident location, traffic obstacle, road environment, major cause of accident, accident type, vehicle operation immediately before accident, major accident type, whether the accident happened during the day or night, and weather.

(i) Accident date: The time of the accident (ex. 11 : 00 on January 17, 2020)

(ii) Road alignment: flat alignment (ex. Straight line)

(iii) Road slope: superelevation slope, longitudinal slope (ex. 7°, flat)

(iv) Section and fill division: (ex. Flat land)

(v) Occurrence point: Branch name, branch type (ex. Busan, Main Line)

(vi) Traffic situation obstacles: reasons for congestion (ex. Contingency congestion)

(vii) Road environment: road surface condition, work status, lighting facilities, etc. (ex. Dry, no work, not applicable)

(viii) Main cause of accident: A large category of cause of accident (ex. Neglect of attention)

(ix) Accident type: Accident target, collision site, etc. (ex. Car-to-facility accident, head-on collision)

(x) Manipulating the vehicle just before the accident: driver behavior (ex. Steering wheel excessively)

(xi) Day/Night: Time information (ex. Daytime: 08 : 00~20 : 00, Night: 20 : 00~08 : 00)

(xii) Weather: Weather information (ex. Sunny)

(xiii) Accident situation: Records of the situation at the time of the accident (ex. While entering the #1 Seoul tollgate 14 lane entrance, the #2 rear end, which was being issued a pass by neglecting to look forward, collided with the #1 front and proceeded after each stop at 12 o'clock It is an accident that moved to the right edge)

2.2. Layer-Based Accident Data Construction and Accident Situation Classification. The Pegasus method [18] identified and conceptualized linguistic explanatory knowledge for traffic patterns on expressways in Germany to generate knowledge-based scenarios. The Web Ontology Language was used to formulate the respective knowledge. In this ontology, knowledge was expressed through hierarchical grades, as well as through the semantic relationships and restrictions between grades. In this study, six independent layers are proposed and defined for structuralizing scenarios based on this ontology. The six layers are used to interpret and identify the characteristics of scenarios otherwise difficult to grasp. Layer 1 represents the road level (including road geometry), whereas Layer 2 represents permanent road facilities such as traffic lights and rules. Layer 3 represents temporary rules or situations that cannot be defined in Layer 2. Layer 4 represents the behaviors of movable objects such as vehicles or pedestrians, and Layer 5 represents environmental conditions such as weather on the roads and day/night. Lastly, Layer 6 represents the digital information of the AV. The accident data were reconstructed as shown in Figure 1, based on Pegasus Layers 1–5. The data corresponding to Layer 1 (road condition) include the alignment and inclination of the roads, the data corresponding to Layer 2 (point of occurrence) include the data of the main lane. The data corresponding to Layer 3 (temporary modifications) include data on accident causes, and the data corresponding to Layer 4 (road operation situation) include traffic congestion and construction work data. The data corresponding to Layer 5 (environmental conditions) include weather and day/night information. The reconfigured data are shown in Figure 1. For traffic situation scenarios, the general situations most likely to occur are referred to as Typical Traffic, those less likely to occur than Typical Traffic are referred to as Critical Traffic, and those that are less likely to occur than Critical Traffic and are difficult to predict are referred to as Edge Case(s), with respect to the inflection point on the frequency graph.

2.3. Generation of Typical Traffic Accident Situation. For suggesting risk scenarios for AVs based on vehicle-to-vehicle traffic situations, the interactions between vehicles were considered as the triggering conditions related to traffic situations (Table 1). Examples of triggering conditions include the road shape, behavior of the AV, and behaviors and locations of nearby vehicles. The risk-situation scenarios based on the triggering conditions related to traffic situations were deduced by connecting accident causes and accident situations in the actual accident data. As "Typical Traffic" in this study refers to accident situations represented by unstructured data, such data must be primary when generating the representative scenarios. Selecting test representative scenarios from unstructured data scenarios before creating detailed logical scenarios will contribute to the generation of more effective logical scenarios. The following triggering conditions, as unstructured data elements, were applied to generate the Typical Traffic accident situations.

2.4. Development of Representative Scenarios through Topic Modeling. Topic modeling is a text mining technique for discovering latent meanings by identifying abstract topics present within a vast amount of data. Using topic modeling enables the extraction of meaningful information from a large amount of unstructured data without a classification system (Blei, Ng and Jordan).

Park and Song [19] collected thesis abstracts published from 1970 to 2012, and identified research trends of library and information science in Korea since 1970 through latent Dirichlet allocation-based topic modeling. The analysis results showed that the research trends could be classified into 20 topics. Accordingly, the major research topics of interest to library information scientists were identified, and hot topics (those mostly actively researched) and cold topics (those less actively researched) were extracted through a yearly trend analysis.

Park et al. [20] demonstrated that text-based big data on disasters could be used for creating future disaster scenarios with unpredictable characteristics by using qualitative methods instead of using statistical methods.

Kayser and Shala [21] examined whether a combination of text mining and scenario planning was beneficial, and proved that such approach can efficiently select a greater amount of text as the amount of data increases.

In general, topic modeling is applied across various fields, as it allows for handling a vast amount of data and of intuitively and elaborately demonstrating the relationships between topics (Kim et al.; Nam; Yoo et al.; Lee et al.; David et al.; Jason; Miller; Fabienne; Park and Son) [22–30].

As described above, Edge Cases are defined as accident situations representing infrequently-occurring and unpredictable risk situations that are difficult to respond to. The characteristics of traffic accidents involving AVs were analyzed by applying topic modeling to Edge Case accident data based on the accident cause of each point of.

Occurrence. The keywords related to vehicle behaviors were extracted from the Edge Cases and then were used to generate representative scenarios reflecting accident situations potentially occurring on expressways on which AVs and regular vehicles both operate.

This study aimed to propose representative risk-situation scenarios that are less likely to occur and difficult to predict by combining the characteristics of each group and vehicle behavior keywords extracted through topic modeling using expressway accident data. Keywords with meanings higher than the average ratio are extracted from each of the vehicle behavior and the cause group through topic modeling. When combining the derived keywords, the sentences were rearranged according to the 5W1H principle to compose the scenario. The layer items that correspond to the 5W1H principle rules are as follows (Table 2).

3. Analysis

3.1. Classification of Traffic Accident Risk Situations. In this study, the traffic situation-triggering conditions were classified based on the International Organization for Standardization definitions, using 1,182 cases of expressway

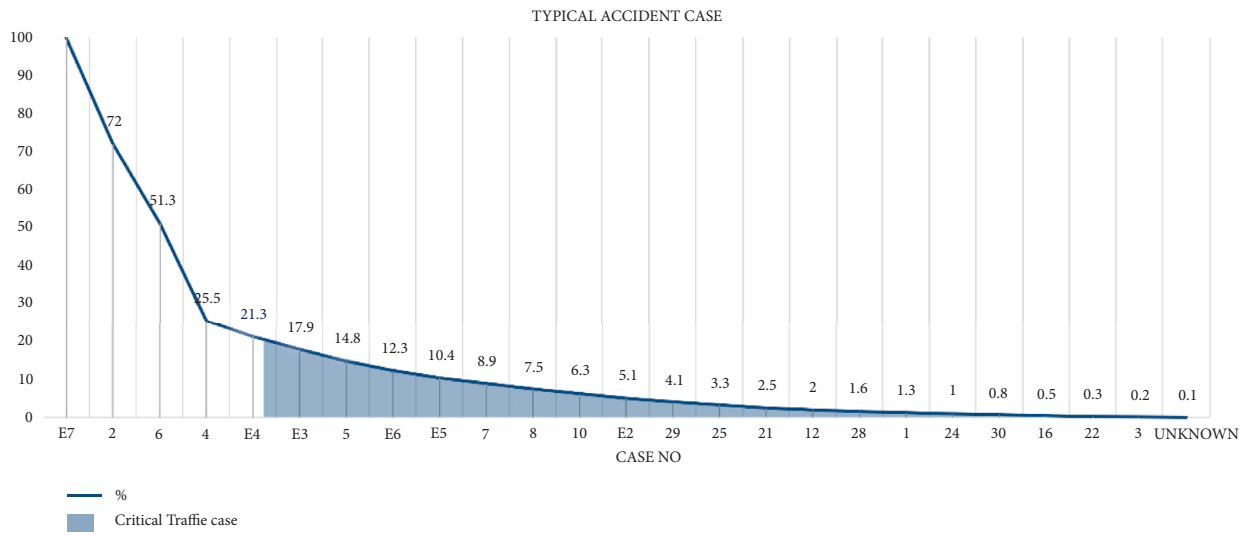


Figure 1: Typical traffic accident situation.

Table 1: Triggering conditions related to-traffic situations.

Road geometry	Behavior of automated vehicle (AV)	Location of nearby vehicles	Behavior of nearby vehicles
Main lane	Lane keeping	Trailing	Cut-in
Merging lane	Lane change	Leading	Cut-out
Split		Side-by-side	Acceleration
Ramp			Deceleration
			Synchronization

Table 2: Triggering conditions related to-traffic situations.

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Alignment	Point of occurrence	Traffic obstacle factors	Accident type	Day and night
Inclination		Road environment	Traffic obstacle factors	
Cut and fill		Main accident cause	Vehicle operation just before the accident	Weather
		Accident cause	Main accident type	

accident data that were re-established based on the Pegasus layer. As a result, 40 cases of traffic accidents (Table 3) and seven extra cases not corresponding to general traffic accidents (Table 4) were extracted. In the triggering condition-based accident situation data, when the AV drove while keeping the lane, the accident situations that were possible but did not happen corresponded to all situations except no. 9 for the non-working section and nos. 12 and 16 for the working section. When the AV drove while changing the lane, the accident situations that were possible but did not happen corresponded to all situations excluding nos. 23, 26, and 27 for the non-working section and no. 39 for the working section.

The Critical Traffic accident situations were deduced based on the accident situations (Figure 1) within the inflection point on an accident frequency graph, as generated from Typical Traffic accident situations re-established based on the triggering conditions. The factors were distinguished into vehicle factors and other factors to examine the differences according to accident cause (Figure 2). Based on the deduced Critical Traffic accident situations, unusual cases within 20% of occurrence frequency are classified into Layer 3 (vehicle factor, other factor), resulting in a total of 35 Edge Cases corresponding to 219 actual accidents (18.5%). For the vehicle factors, the cases with a lower frequency than the no. 2 case (21.4%) correspond to Edge Cases, whereas for the other factors, the cases with a frequency lower than 20% including no. 5 case (20.2%) correspond to Edge Cases (See Figures 3 and 4).

The Critical Traffic accident situations include 16 vehicle factors and 19 other factors (including those cannot be classified), thus reflecting a total of 660 cases (55.8%) of 35 accident types. The Edge Cases concerning vehicle factors include 15 types: E6 (slippery road), 5 (rear-end collision + lane keeping), E1 (other: vehicle defect), 7 (head-on collision + lane keeping), 10 (structure collision + lane keeping), 2 (leading side collision + lane change), E5 (poor loading), 21 (head-on collision + lane change), 8 (object on road + rear-end collision + lane keeping), 30 (median separation + head-on collision), 29 (trailing side collision + lane change), 24 (structure collision + lane change), E4 (pedestrian collision), 3 (rear-end collision + lane keeping + lane change of nearby vehicle), and 1 (head-on collision + lane keeping + lane change of nearby vehicle). The Edge Cases

Table 3: Typical traffic accident classification.

Ego Behavior	Traffic obstacle factors	Cut-in		Cut-out		Acceleration		Deceleration		Sync	
		v-v	v-s	v-v	v-s	v-v	v-s	v-v	v-s	v-v	v-s
Lane keep	Non-working section	1	2	3	4	5	6	7	8	9	10
	Working section	11	12	13	14	15	16	17	18	19	20
Lane change	Non-working section	21	22	23	24	25	26	27	28	29	30
	Working section	31	32	33	34	35	36	37	38	39	40

Table 4: Extra accident case.

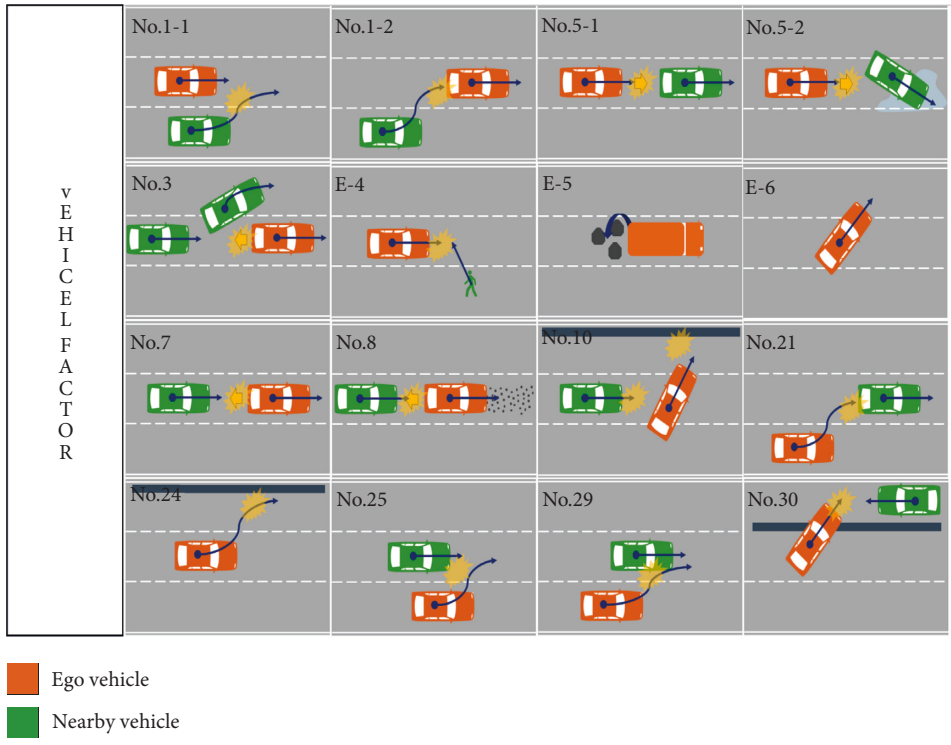
Extra no.	Accident type
E1	Others
E2	Slippery road
E3	Animal intrusion
E4	Pedestrian collision
E5	Poor loading
E6	Loss of balance
E7	Fire

concerning other factors (Figure 5) include 17 types: 5 (rear-end collision + lane keeping), E5 (poor loading), 8 (object on road + rear-end collision + lane keeping), E1 (other: object on road), E2 (slippery road), E7 (fire), 29 (trailing side collision + lane change), 7 (head-on collision + lane keeping), 10 (structure collision + lane keeping), 28 (miscellaneous thing on road + lane change + side collision), 12 (lane keeping + head-on collision + construction site), 1 (head-on collision + lane keeping + lane change of nearby vehicle), 21 (head-on collision + lane change), 22 (structure collision + lane change + lane entering of nearby vehicle), 25 (non-working section + lane change + deceleration of nearby vehicle), E6 (loss of balance), and 24 (structure collision + lane change). For Critical Cases (i.e., not corresponding to Edge Case), the vehicle factors include E7 (fire), 6 (structure collision + lane keeping + acceleration of nearby vehicle), and 2 (structure collision + lane keeping + lane entering of nearby vehicle), whereas the other factors include 2 (structure collision + lane keeping + lane entering of nearby vehicle), 4 (structure collision + lane keeping + lane change of nearby vehicle), E4 (pedestrian collision), E3 (animal intrusion), and 6 (structure collision + lane keeping + acceleration of nearby vehicle). Cases nos. 2 (structure collision + lane keeping + lane entering of nearby vehicle) and 6 (structure collision + lane keeping + acceleration of nearby vehicle) did not correspond to Edge Cases, and occurred frequently.

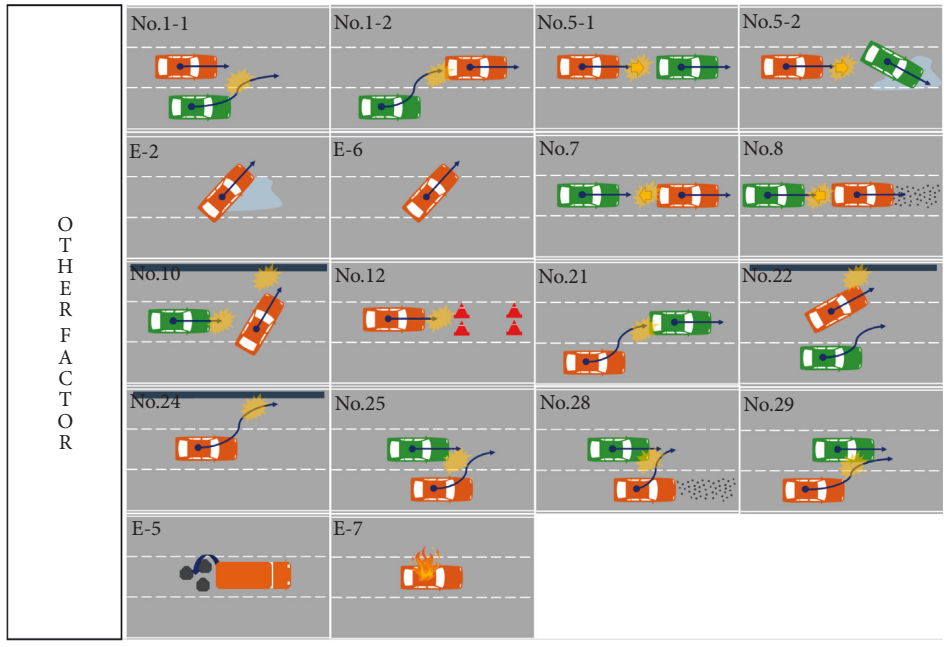
4. Topic Modeling Result of Edge Case Subjects

4.1. Edge Case Analysis Result for Vehicle Factors. Keywords were extracted to analyze the Edge Case accident situations through topic modeling. First, the coherence score was evaluated to determine the number of groups into which the data per point of occurrence were to be divided. The coherence score of the vehicle factors was the highest when the number of topics was four; therefore, the topic modeling proceeded by setting the number of groups to four (Figure 6).

For topic elements of group 1, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, left-curve), superelevation (flatness, fill), and inclination (downhill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (stopped vehicle, normal, non-working section), road environment during the accident (dry), and main accident cause (tire damage, vehicle part problem). The Layer 4-Road Operation Situation elements include accident type (side collision, single accident, read-end collision, head-on collision), vehicle operation immediately before the accident (staying in the driving lane, other), and main accident type (vehicle-vehicle, other). The Layer 5-Environmental Conditions elements include day/night (daytime, nighttime) and weather (sunny, cloud) (Table 5). For the topic elements of group 2, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, left-curve, right curve), superelevation (flatness, fill, cut), and inclination (uphill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (stopped vehicle, normal, non-working section), road environment during the accident (dry), and main accident cause (tire damage, brake device defects, vehicle part problem). The Layer 4-Road Operation Situation elements include accident type (rear-end collision, head-on collision, single accident), vehicle operation immediately before the accident (staying in the driving lane, other), and main accident type (vehicle-vehicle, other). The Layer 5-Environmental Conditions include day/night (daytime, nighttime) and weather (sunny) (Table 6). For the topic elements of group 3, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, left-curve, right curve), superelevation (flatness, fill, cut), and inclination (downhill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (stopped vehicle, normal, non-working section), road environment the during accident (dry), and main accident cause (tire damage, vehicle part problem, brake device defects). The Layer 4-Road Operation Situation elements include accident type (side collision, single accident, read-end collision, head-on collision, shoulder, other), vehicle



(a)



(b)

Figure 2: Critical traffic accident situation.

operation immediately before the accident (staying in the driving lane), and main accident type (vehicle-vehicle, vehicle-structure, other). The Layer 5-Environmental Conditions elements include day/night (daytime, nighttime) and weather (sunny) (Table 7). For the topic elements of group 4, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, right curve), superelevation (flatness, fill, cut), and inclination (uphill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (stopped vehicle, normal, non-working section), road environment the during accident (dry, normal, wet), and main accident cause (tire damage, brake device defects, vehicle part problem). The

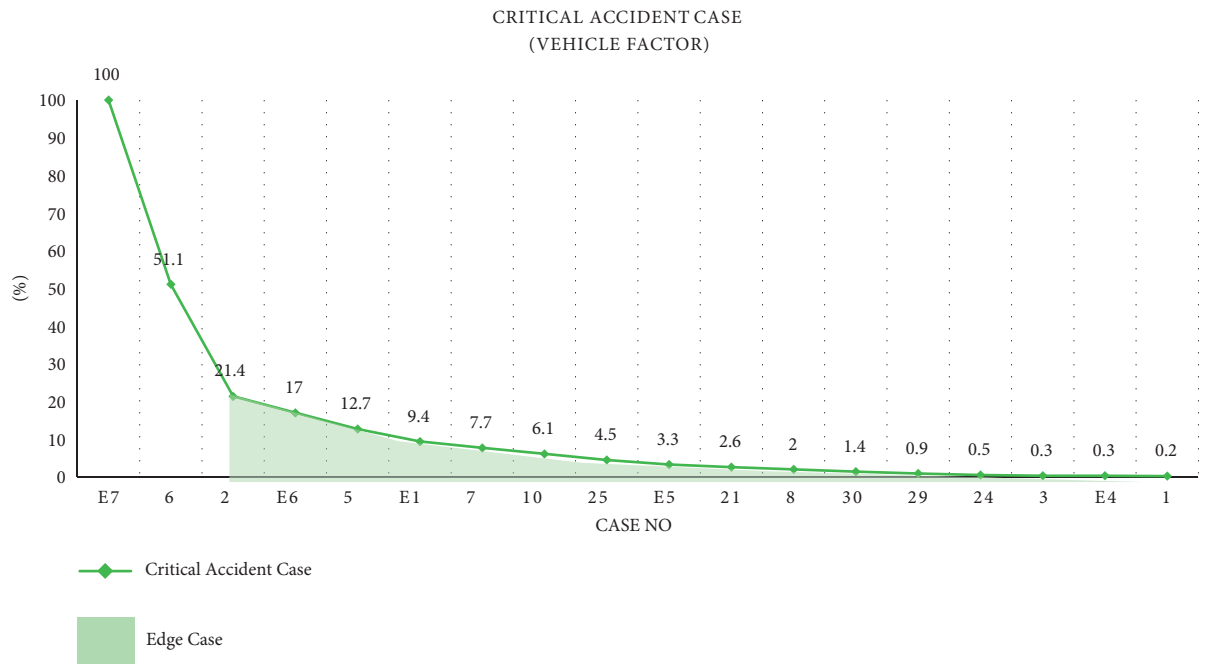


Figure 3: Edge case type of vehicle factors.

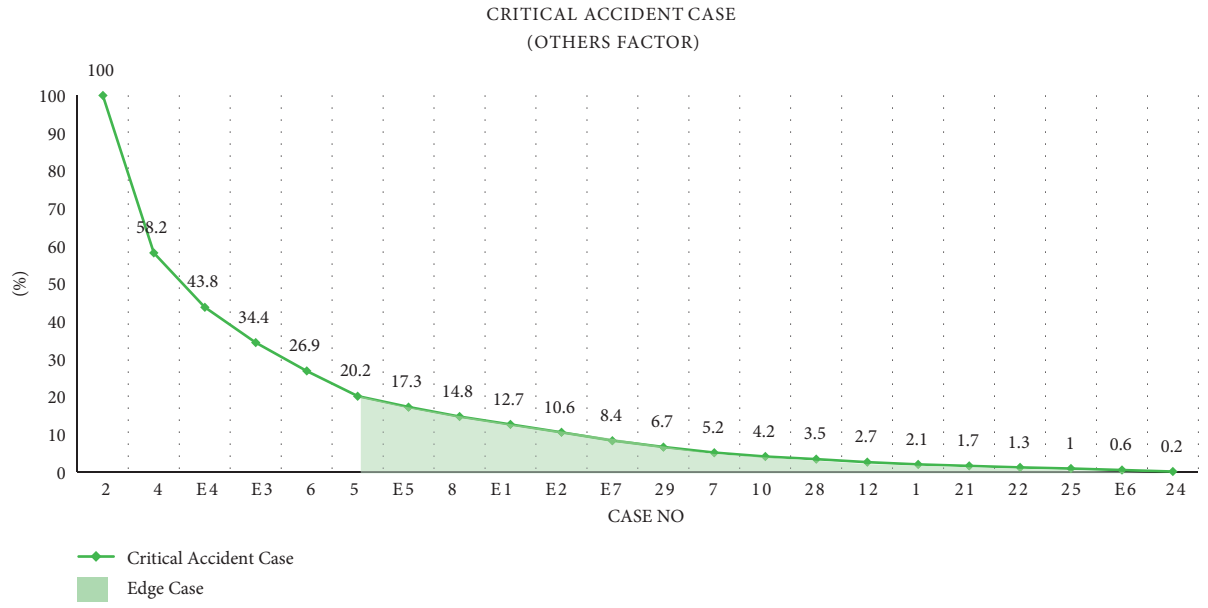


Figure 4: Edge case type of other factors.

elements corresponding to Layer 4-Road Operation Situation include the accident type (single accident, shoulder, guardrail), vehicle operation immediately before the accident (staying in the driving lane, overactive handle, other), and main accident type (vehicle-structure, other). The Layer 5-Environmental Conditions elements include day/night (daytime, nighttime) and weather (sunny, rainy) (Table 8). The table below shows the frequency from arranging the topics of each group by layer. The emphasized top three words represent the layers with the largest portions among topic groups, and the topic groups within the top three frequency ranking (See Figures 5 and 7–9).

4.2. Edge Case Analysis Result for Other Factors. The coherence score of the other factors was the highest when the number of topics was four, and therefore, the topic modeling proceeded by setting the number of groups to four (Figure 10).

For the topic elements of group 1, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, left-curve, right curve), superelevation (flatness, fill, cut), and inclination (uphill, downhill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road

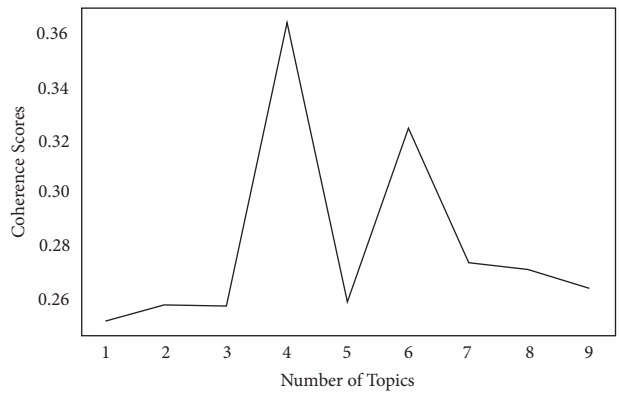


Figure 5: Coherence score of vehicle factors.

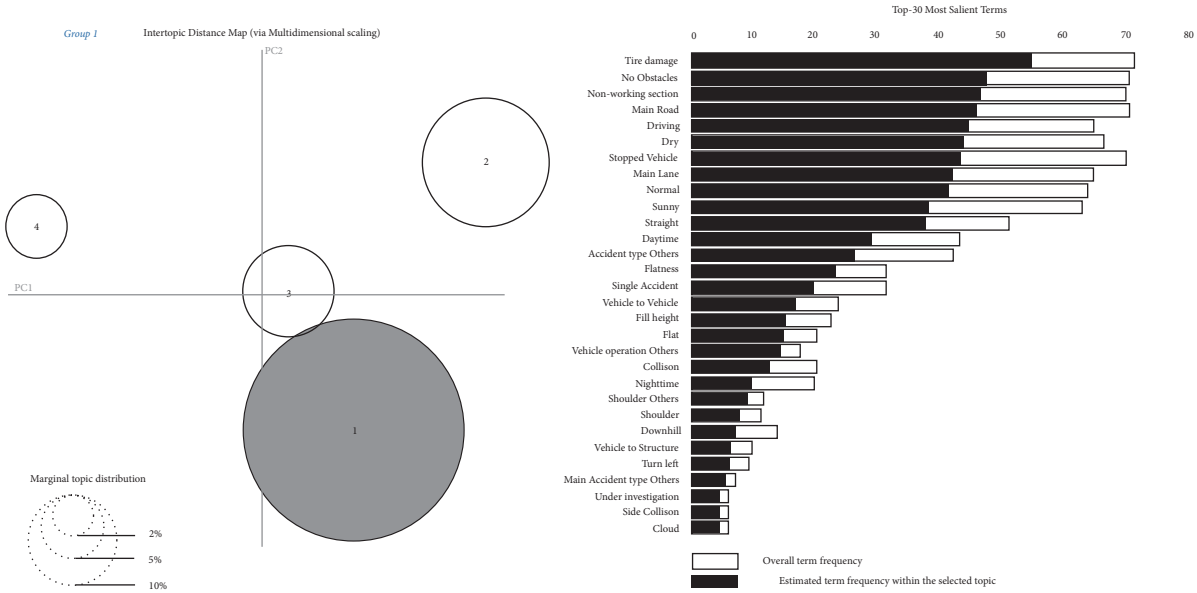


Figure 6: Vehicle factor group 1.

Table 5: Vehicle factor topic modeling group 1.

Vehicle factor group 1				
Layer		Topics		Rate
Layer 1	Alignment	Left-curve, left-curve 1000 m more, straight		5%
	Inclination	Downhill, downhill less than 1%, downhill 1% to 3%, downhill 3% more, downhill less than 500 m, flat		3%
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m		6%
Layer 2	**Point of occurrence**	**Main lane**	**1/2, 1/3, 1/4, 2/2, 2/3, 2/4, 3/3, 3/4, 3/5, 4/4, 4/5, 5/5**	**15%**
		Shoulder	Shoulder, 2/2	1%
Layer 3	**Traffic obstacle factors**	**Non-working section, stopped vehicle, normal**		**15%**
	Road environment	Dry		8%
	Main accident cause	**Tire damage, problem of device**		**10%**
Layer 4	Accident type	Side collision, single accident, shoulder others, shoulder guardrail, rear-end, head-on, angle		7%
	Vehicle operation Just before the accident	**Driving, others**		**10%**
	Main accident type	Vehicle-vehicle, others		5%
Layer 5	Day and night	Daytime, nighttime		8%
	Weather	Sunny, cloud		7%

Bold and italic: top frequency topic layer and top three frequency topic group.

Table 6: Vehicle factor topic modeling group 2.

Vehicle factor group 2			
Layer		Topics	Rate
Layer 1	Alignment	Left-curve, left-curve 1000 m more, right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, straight	5%
	Inclination	Uphill, uphill less than 1%, uphill 1% to 3%, uphill 3% more, flat	4%
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m, cut part less than 10 m, cut part 10 m more	5%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 2/3, 2/4, 3/3, 3/4, 3/5, 4/4, 4/5, 5/5**	**15%**
Layer 3	**Traffic obstacle factors**	**Non-working section, normal**	**12%**
	Road environment	Dry	7%
	Main accident cause	**Tire damage, Brake device defects, problem of device parts**	**13%**
Layer 4	Accident type	Single accident, rear-end, head-on	3%
	Vehicle operation just before the accident	Driving, others	9%
	Main accident type	Vehicle-vehicle, others	7%
Layer 5	Day and night	Daytime, nighttime	6%
	Weather	Sunny	8%

Bold and italic: top frequency topic layer and top three frequency topic group.

Table 7: Vehicle factor topic modeling group 3.

Vehicle factor group 3			
Layer		Topics	Rate
Layer 1	**Alignment**	**Right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, left-curve, left-curve 1000 m more, straight**	**10%**
	Inclination	Downhill, downhill less than 1%, downhill 1% to 3%, downhill 3% more, downhill less than 500 m, flat	9%
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m, cut part less than 10 m, cut part 10 m more	4%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 2/3, 2/4, 3/3, 3/4, 3/5, 4/4, 4/5, 5/5**	**14%**
Layer 3	**Traffic obstacle factors**	**Non-working section, normal, stopped vehicle**	**10%**
	Road environment	Dry	6%
	Main accident cause	**Tire damage, Brake device defects, problem of device parts**	**10%**
Layer 4	Accident type	Side collision, single accident, rear-end, head-on, shoulder others	4%
	Vehicle operation just before the accident	Driving, overactive handle, others	9%
	Main accident type	Vehicle-vehicle, vehicle-structure, others	6%
Layer 5	Day and night	Daytime, nighttime	8%
	Weather	Sunny	8%

Bold and italic: top frequency topic layer and top three frequency topic group.

Table 8: Vehicle factor topic modeling group 4.

Vehicle factor group 4			
Layer		Topics	Rate
Layer 1	Alignment	Right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, straight	3%
	Inclination	Uphill, uphill less than 1%, uphill 1% to 3%, uphill 3% more, uphill less than 500 m, flat	**13%**
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m, cut part less than 10 m, cut part 10 m more	2%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 2/3, 2/4, 3/3, 3/4, 3/5, 4/4, 4/5, 5/5**	**11%**
		Shoulder Shoulder, 2/2	5%
Layer 3	**Traffic obstacle factors**	**Non-working section, normal, stopped vehicle**	**15%**
	Road environment	Dry, wet	10%
	Main accident cause	Tire damage, brake device defects, problem of device parts	10%
Layer 4	Accident type	Single accident, shoulder, shoulder guardrail	6%
	Vehicle operation just before the accident	Driving, overactive handle, others	5%
	Main accident type	Vehicle-structure, others	6%
Layer 5	Day and night	Daytime, nighttime	7%
	Weather	Sunny, rainy	7%

Bold and italic: top frequency topic layer and top three frequency topic group.

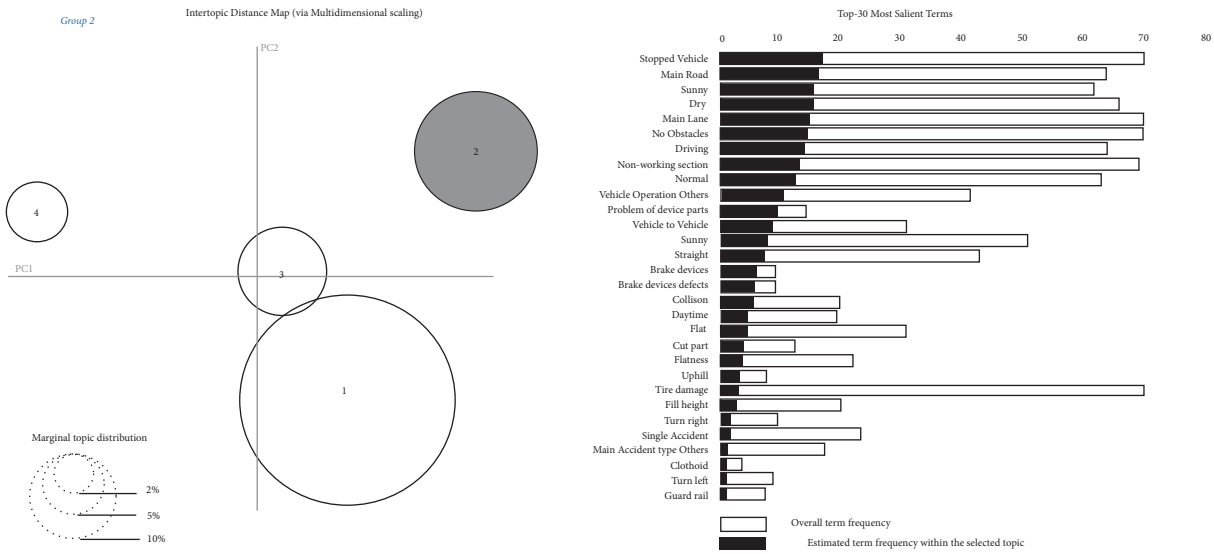


Figure 7: Vehicle factor group 2.

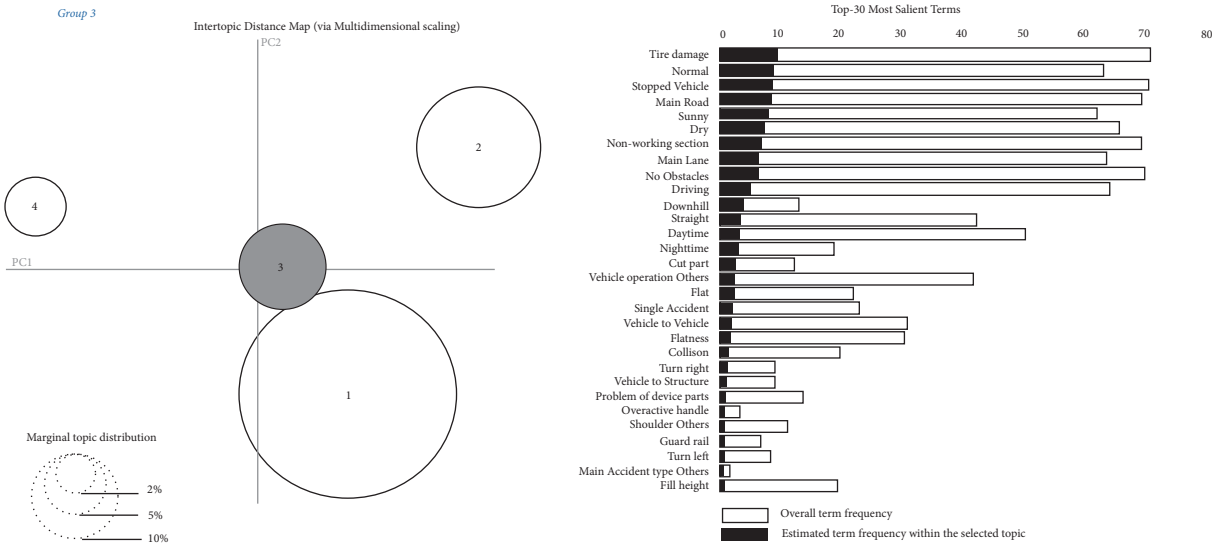


Figure 8: Vehicle factor group 3.

Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (normal, non-working section), road environment during the accident (dry, wet), and main accident cause (poor loading, object on road). The elements corresponding to Layer 4-Road Operation Situation include accident type (single accident, read-end collision, head-on collision, guardrail), vehicle operation immediately before the accident (staying in the driving lane, overactive handle, other), and main accident type (vehicle-vehicle, vehicle-structure, other). The Layer 5-Environmental Conditions elements include day/night (daytime, nighttime) and weather (sunny, cloud) (Table 9). For the topic elements of group 2, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, left-curve, right curve), superelevation (flatness, fill), and inclination (flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (normal, non-working section), road environment during the accident (dry, wet), and main accident cause (poor loading, slippery road, visual disturbance, object on roads, falling object). The elements corresponding to Layer 4-Road Operation Situation include accident type (side collision, guardrail), vehicle operation immediately before the accident (staying in the driving lane), and main accident type (vehicle-vehicle, other). The elements corresponding to Layer 5-Environmental Conditions include day/night (daytime, nighttime) and weather (rainy) (Table 10). For the topic elements of group 3, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight, right curve), superelevation (flatness, cut), and inclination (uphill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the

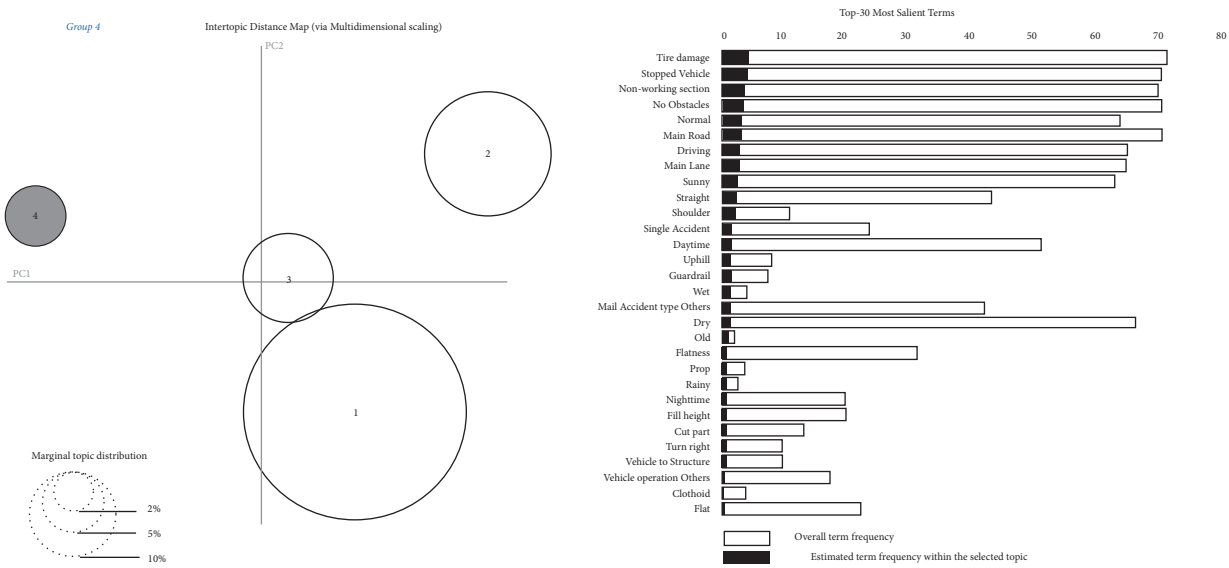


Figure 9: Vehicle factor group 4.

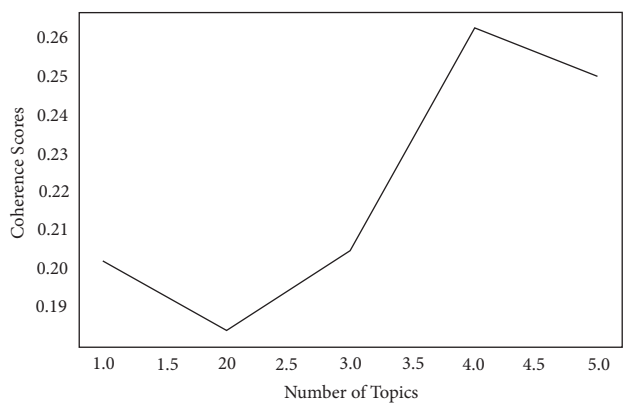


Figure 10: Coherence score of other factor.

Table 9: Other factor topic modeling group 1.

Other factor group 1			
Layer		Topics	Rate
Layer 1	Alignment	Right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, left-curve, left-curve 1000 m more, straight	7%
	Inclination	Uphill, uphill less than 1%, uphill 1% to 3%, uphill 3% more, uphill less than 500 m, downhill, downhill less than 1%, downhill 1% to 3%, downhill 3% more, downhill less than 500 m, flat	7%
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m, cut part less than 10 m, cut part 10 m more	7%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 3/3, 2/4, 3/3, 4/4**	**16%**
		Shoulder, accelerate lane	1%
Layer 3	**Traffic obstacle factors**	**Non-working section, normal**	**15%**
	Road environment	Dry, wet	8%
	Main accident cause	Poor loading, object on roads	4%
Layer 4	Accident type	Single accident, shoulder guardrail, rear-end, head-on	5%
	Vehicle operation just before the accident	Driving, others, overactive handle	7%
	Main accident type	Vehicle-vehicle, vehicle-structure, others	7%
Layer 5	**Day and night**	**Daytime, nighttime**	**9%**
	Weather	Sunny, cloud	7%

Bold and italic: top frequency topic layer and top three frequency topic group.

Table 10: Other factor topic modeling group 2.

Other factor group 2			
Layer		Topics	Rate
Layer 1	Alignment	Right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, left-curve, left-curve 1000 m more, straight	7%
	Inclination	Flat	5%
	Cut and fill	Flatness, fill height less than 2 m, fill height 2 m–5 m, fill height 5 m–15 m	3%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 3/3, 2/4, 3/3, 4/4**	**17%**
		Shoulder, accelerate lane	1%
Layer 3	Traffic obstacle factors	Non-working section, normal	9%
	Road environment	Dry, wet	10%
	Main accident cause	**Poor loading, Slippery road surface, Visual disturbance, object on roads, Falling object**	**17%**
Layer 4	Accident type	Side collision, shoulder guardrail	1%
	Vehicle operation just before the accident	Driving, others	8%
	Main accident type	Vehicle-vehicle, others	2%
Layer 5	Day and night	Daytime, nighttime	9%
	Weather	**Rainy**	**11%**

Bold and italic: top frequency topic layer and top three frequency topic group.

point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (normal, non-working section), road environment during the accident (dry, wet), and main accident cause (poor loading, object on roads). The elements corresponding to Layer 4-Road Operation Situation include accident type (collision), vehicle operation immediately before the accident (staying in the driving lane, overactive handle), and main accident type (vehicle-vehicle, vehicle-structure, other). The elements corresponding to Layer 5-Environmental Conditions include day/night (daytime, nighttime) and weather (sunny, rainy) (Table 11). For the topic elements of group 4, the elements corresponding to Layer 1-Road Condition (permanent) include alignment (straight), superelevation (flatness, cut), and inclination (downhill, flatland). The element corresponding to Layer 2-Traffic Infrastructure (permanent) is the point of occurrence (main lane). The elements corresponding to Layer 3-Road Condition and Traffic Infrastructure (temporary) include traffic obstacle factors (congestion, normal, non-working section), road environment during the accident (dry, wet), and main accident cause (poor loading). The elements corresponding to Layer 4-Road Operation Situation include accident type (single accident, rear-end collision, head-on collision, guardrail), vehicle operation immediately before the accident (staying in the driving lane, overactive handle), and main accident type (vehicle-vehicle, vehicle-structure). The elements corresponding to Layer 5-Environmental Conditions include day/night (daytime, nighttime) and weather (sunny, rainy) (Table 12). The table below shows the frequency from arranging the topics of each group by layer. The emphasized words represent the layers constituting the largest portions among topic groups, and the topic groups within the top three frequency ranking (See Figures 11–14).

5. Discussion

Topic modeling was performed for the Edge Case accident situations classified according to point of occurrence and accident case to analyze how the topics including each layer element were distributed, as well as the group characteristics (Table 13). The top three weighted topics demonstrating the group characteristics were selected. The vehicle factors were largely classified into groups having four different characteristics. As all groups of 1, 2, 3, and 4 were based on the main lane, the topic of the point of occurrence was included in all groups. The traffic obstacle factor and day/night topics had the largest distribution in group 1, whereas the main accident cause had the largest distribution in groups 2, 3, 4; the main accident cause, road shape, and road inclination were the most influential in groups 2, 3, and 4, respectively. Main lane (other factor) also had the largest distribution in all groups. The traffic obstacle factor was most influential in groups 1, 3, and 4, whereas the main accident was the most influential in group 2. This signifies that the main lane (other factor) had a similar distribution to that of the main lane (vehicle factor) insofar as the topics. The emphasized topics in Table 12 clearly demonstrate the differences in topic characteristics between vehicle factors and other factors.

When analyzing the topic modeling results overall, it can be seen that the main lane topic is distributed in similar patterns, but main lane (vehicle factor) has groups with a larger distribution for the main accident cause, whereas main lane (other factor) has groups with a larger distribution for the traffic obstacle factor. When accidents occur owing to vehicle factors, the main accident cause must be more critical than traffic obstacle factor; when accidents occur owing to other factors, the traffic obstacle factor must be more influential than main accident cause. Such results signify that the topic modeling was appropriately applied for the main lane.

Table 11: Other factor topic modeling group 3.

Other factor group 3			
Layer		Topics	Rate
Layer 1	**Alignment**	**Right-curve, right-curve 500 m–1000 m, right-curve 1000 m more, straight**	**9%**
	Inclination	Uphill, uphill less than 1%, uphill 1% to 3%, uphill 3% more, uphill less than 500 m, flat	8%
	Cut and fill	Flatness, cut part less than 10 m, cut part 10 m more	5%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 3/3, 2/4, 3/3, 4/4**	**11%**
		Shoulder	1%
Layer 3	**Traffic obstacle factors**	**Non-working section, normal**	**9%**
	Road environment	**Dry, wet**	**9%**
	Main accident cause	Poor loading, object on roads	7%
Layer 4	Accident type	Single accident, shoulder guardrail, rear-end, head-on	8%
	Vehicle operation just before the accident	Driving, overactive handle	8%
	Main accident type	**Vehicle-vehicle, vehicle-structure, others**	**10%**
Layer 5	Day and night	Daytime, nighttime	7%
	Weather	Sunny, rainy	8%

Bold and italic: top frequency topic layer and top three frequency topic group.

Table 12: Other factor topic modeling group 4.

Other factor group 4			
Layer		Topics	Rate
Layer 1	Alignment	Straight	3%
	Inclination	Downhill, downhill less than 1%, downhill 1% to 3%, downhill 3% more, downhill less than 500 m, flat	5%
	Cut and fill	Flatness, cut part less than 10 m, cut part 10 m more	5%
Layer 2	**Point of occurrence**	**Main lane** **1/2, 1/3, 1/4, 2/2, 3/3, 2/4, 3/3, 4/4**	**15%**
		Shoulder, accelerate lane	3%
Layer 3	**Traffic obstacle factors**	**Non-working section, normal, congestion**	**18%**
	Road environment	Dry, wet	9%
	Main accident cause	Poor loading, object on roads	3%
Layer 4	Accident type	Single accident, shoulder guardrail, rear-end, head-on	3%
	Vehicle operation Just before the accident	**Driving, others, Overactive handle**	**11%**
	Main accident type	Vehicle-vehicle, vehicle-structure, others	9%
Layer 5	Day and night	Daytime, nighttime	5%
	Weather	**Sunny, rainy**	**11%**

Bold and italic: top frequency topic layer and top three frequency topic group.

In emergency situations, an AV allows a driver to take-over certain rights to control. As a driver's vehicle operation can be interpreted as a reaction to accident situations in urgent situations, it can be considered as an important factor for constituting scenarios. For generating scenarios reflecting a driver's reaction to accident situations, 17 representative unpredictable risk-situation scenarios were generated by combining the top three topics and keywords related to vehicle behaviors immediately before accidents from the actual accident data corresponding to Edge Cases (Table 14). The words from Layers 1–5 were combined by having top three topic groups as main topics when generating scenarios; the remaining layer elements of the actual accident data from Edge Case accident cause were extracted as additional topics to be combined. In other words, we have configured the skeletons that make up the scenario with the top three topics. To flesh the scenario out, the top three topics found

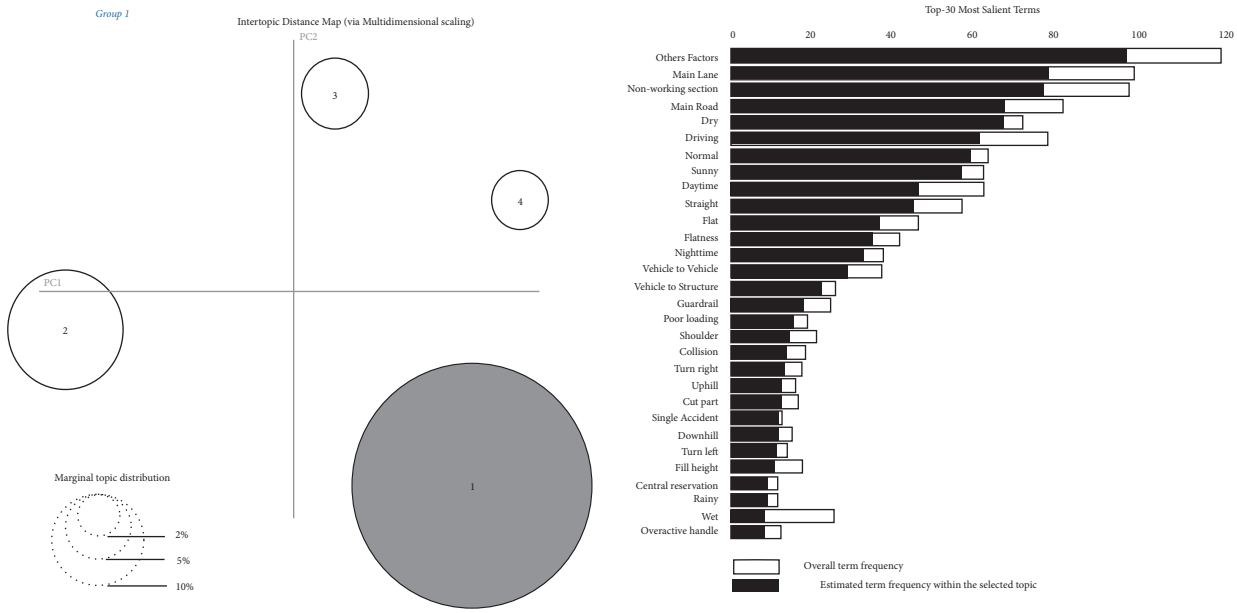


Figure 11: Other factor group 1.

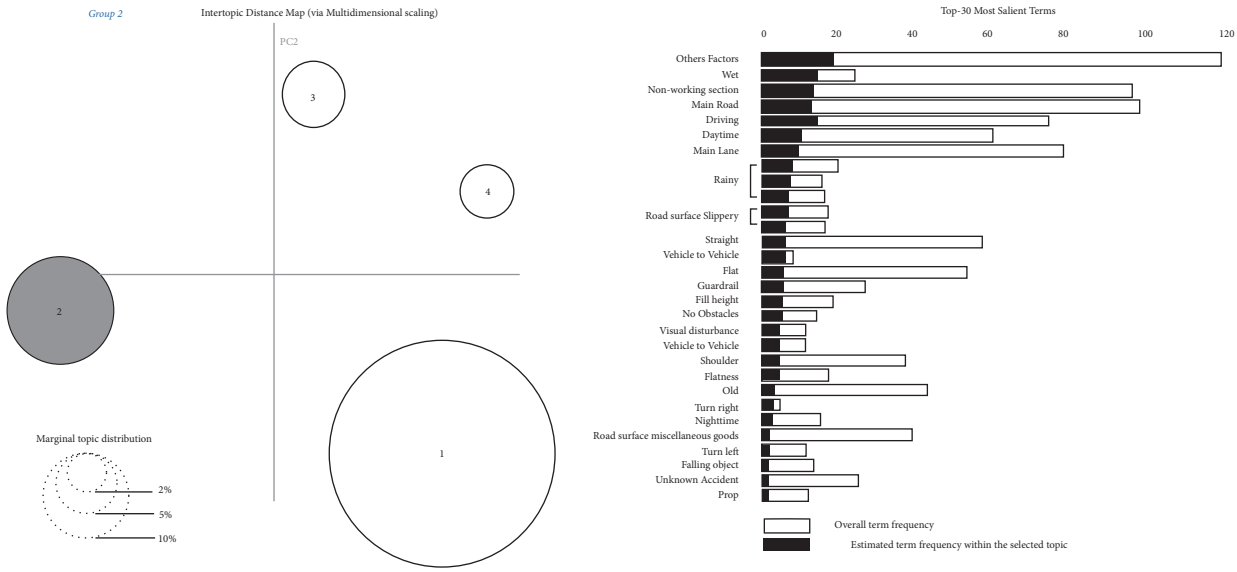


Figure 12: Other factor group 2.

the main real accidents. This created the final scenario by extracting the remaining layer topics except the layer containing the top three topics. When rearranging the text to embody the scenario, we placed the keywords corresponding to the layer according to the 5W1H principle. The emphasized main topics in the scenarios below are related to direct accident causes and driver behaviors representing accident response behaviors of AVs; other topics constitute accident situations extracted from actual accident data and classified by accident case.

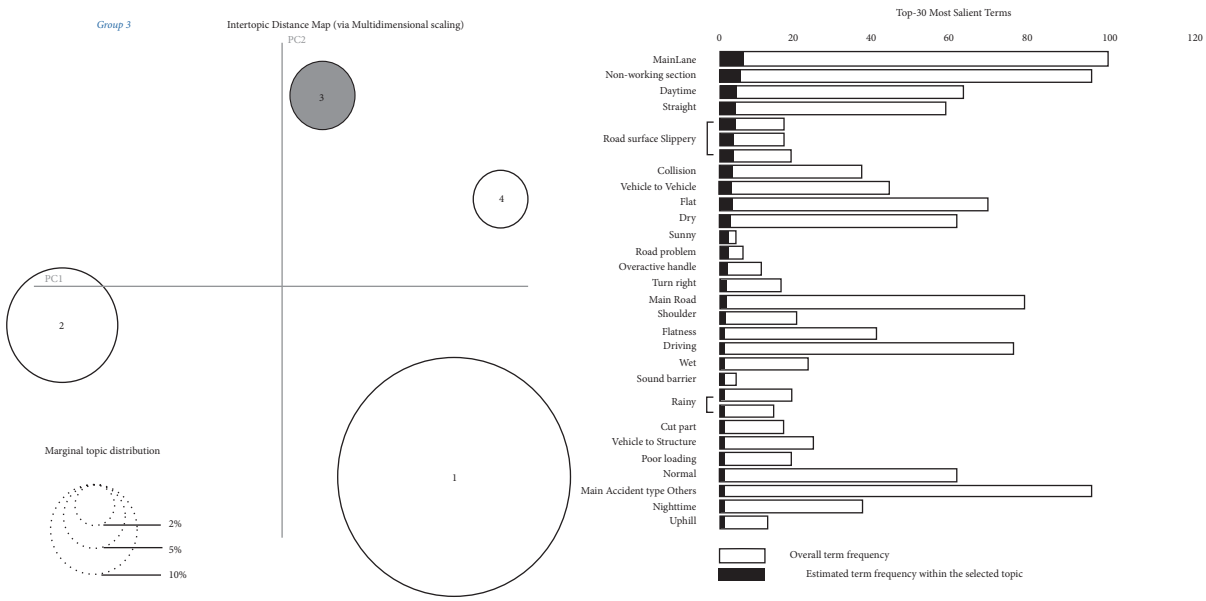


Figure 13: Other factor group 3.

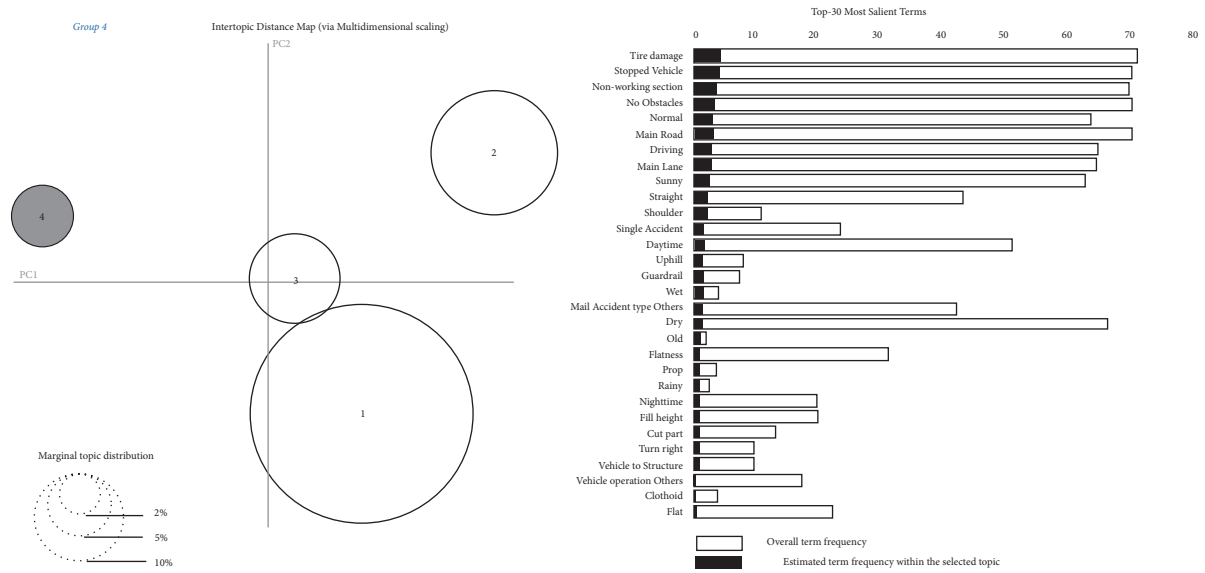
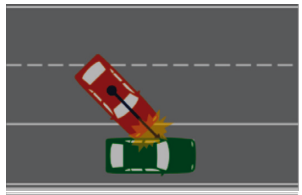
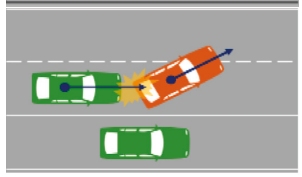
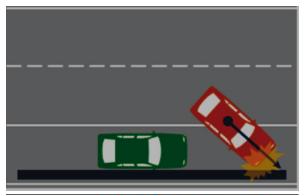
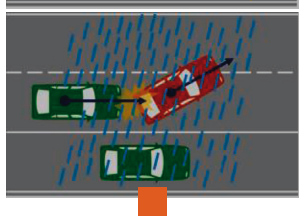
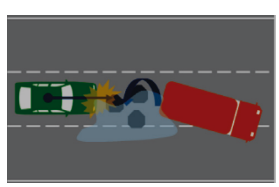
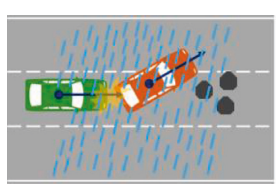


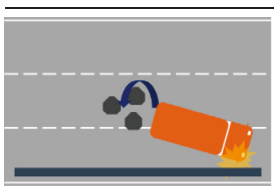
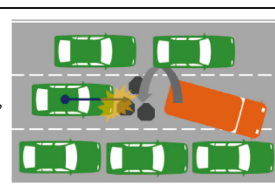
Figure 14: Vehicle factor group 4.

Table 13: Top three topic distribution by group.

Cause	Group	Layer 1		Layer 2		Layer 3		Layer 4		Layer 5	
		Alignment	Inclination	Point of occurrence	Traffic obstacle factors	Road environment	Main accident cause	Vehicle operation	Main accident type	Day and night	Weather
Vehicle	1			o	o					o	
	2			o			o		o		
	3	o		o			o				
	4		o	o			o				
Other	1			o	o					o	
	2			o			o				o
	3	o		o	o	o			o		
	4			o	o			o			o

Table 14: Representative risk-situation scenario.

Factor	Group	Illustration	Accident
Vehicle factor	1		A stopped vehicle (L3) in the shoulder when driving on the main lane (L2) which is flat (L1), straight (L1), and dry (L3) in non-working section (L3) at night (L5) on a cloudy day (L5). Vehicle-vehicle accident (L4) involving side collision (L4) into a stopped vehicle (L3) owing to tire damage
	2		When driving through normal traffic (L3) on the main lane (L2) which is dry (L3), flat (L1), straight (L1), and has a certain fill height (L1) during daytime (L5) on a sunny day (L5). Vehicle-vehicle accident (L4) involving rear-end collision (L3) of a secondary vehicle after a single accident (L4) caused by vehicle part problem (L3)
	3		A stopped vehicle (L3) in the shoulder when driving on the main lane (L2) which has right-curve (L1), downhill (L1), has a certain cut part (L1), and dry (L3) at night (L5) on a sunny day (L5). Vehicle-structure collision (L4) of a single accident (L4) owing to overactive handling (L4) caused by vehicle part problem (L3)
	4		A stopped vehicle (L3) in the shoulder when driving on the main lane (L2) which is straight (L1), uphill (L1), has a certain fill height (L1), and dry (L3) at night (L5) on a rainy day (L5). Collision by a secondary vehicle after a single accident (L4) caused by tire damage (L3)
OthersFactor	1		When driving through normal traffic (L3) on the main lane (L2) which is wet (L3), uphill (L1), straight (L1), and has a certain fill height (L1), and has a certain fill height (L1) at night (L5) on a cloudy day (L5) Vehicle-vehicle accident (L4) of a head-on collision (L4) with a nearby vehicle owing to poor loading (L3)
	2		When driving (L4) on the main lane (L2) which has left-curve (L1) and is uphill (L1) and flat (L1) in non-working section (L3) during daytime (L5) on a rainy day (L5) Collision by a secondary vehicle after a side collision (L4) caused by object on roads (L3)
	3		When driving (L4) through normal traffic (L3) on the main lane (L3) which is dry (L3), uphill (L1), straight (L1), and has a cut part (L1) during daytime (L5) on a sunny day (L5) Vehicle-structure accident (L4) caused by poor loading (L3)
	4		Congestion (L3) when driving on the main lane (L2) which is straight (L1), flat (L1), downhill (L1), and dry (L3) during daytime (L5) on a sunny day (L5) Head-on collision (L4) into another vehicle owing to overactive handling (L4) caused by poor loading (L3)

 Ego vehicle,  nearby vehicle.

6. Conclusion

Growing interest has recently been paid to the development of scenarios for evaluating the safety of AVs, and research is being conducted on various methodologies and on the generation of scenarios including technological elements. However, most studies have focused on frequently-occurring accident types or representative accident situations; thus, there is a lack of studies on scenarios considering unpredictable accidents. Proper preparation is required for such accident situations, because even a traffic accident that is less likely to occur can lead to fatal accidents if it is difficult to predict. This study used Korean expressway traffic accident data and topic modeling to develop risk-situation scenarios involving AVs based on actual accident data. The collected expressway accident data were preprocessed based on the Pegasus layer to create unstructured accident situation data; the generated accident situation included 1,182 cases. From these, the data within 20% of occurrence frequency were extracted to generate Critical Case accident situations. Furthermore, the data within 20% of occurrence frequency were extracted from the selected Critical Case accident situations to classify Edge Cases that were less likely to occur and difficult to predict. The characteristics between groups were comparatively analyzed based on the point of occurrence through topic modeling of the Edge Cases. According to the topic modeling results, most accident situations in the main lane are largely affected by the point of occurrence, in which the main accident cause and traffic obstacle factors are major influences in-vehicle factors and other factors, respectively. Most accident situations occurring on ramps are significantly affected by traffic obstacle factors, and accident situations are generally affected by weather if there is a minor impact from traffic obstacle factors. For the topic of tunnels, the traffic obstacle factor had a large influence in all groups, and there was a difference between day and night. Ultimately, 17 representative risk-situation scenarios otherwise difficult to predict owing to a low occurrence frequency were generated by recombining additionally extracted vehicle behavior keywords with topics by group. Various elements needed to be reflected to the maximum possible extent, as a variety of aspects must be considered when generating scenarios dependent on the point of occurrence and other factors. Furthermore, the method for generating scenarios may vary depending on the purpose for generating the scenarios, particularly when devising scenarios by considering actual accident situations.

This study developed scenarios using traffic accident data and topic modeling, but there are several limitations. First, the developed scenarios were not evaluated or verified; hence, actual vehicle or simulation experiments must be conducted to evaluate and verify the developed scenarios. Second, the expressway accident data used in this study are unstructured data, and thus fail to include numerical data for assuming specific situations. By performing simulations for the generated risk-situation scenarios, specific risk accident situations reflecting numerical data and scenarios including sensor data can also be generated.

Data Availability

The topic modeling data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author(s) declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by Ministry of Science and ICT and Institute for Information and Communication Technology Planning and Evaluation (Development of Validation Technology for Operation Rights SW Safety and Response according to Fallback MRC of Edge-Based Autonomous Driving Function, no. 2021-0-00697, and Development of technology for validating the autonomous driving services in perspective of laws and regulations no. 2021-0-01352).

References

[1] S. Kwon and J. Lee, "Autonomous vehicle security threats and technology trends," *Korea Institute of Information Security and Cryptology*, vol. 30, no. 2, pp. 31–39, 2020.

[2] R. H. Patel, J. Härri, and C. Bonnet, "Braking strategy for an autonomous vehicle in a mixed traffic scenario," in *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2017)*, pp. 268–275, Porto, Portugal, April 2017.

[3] N. Virdi, H. Grzybowska, S. Travis Waller, and V. Dixit, "A safety assessment of mixed fleets with Connected and Autonomous Vehicles using the Surrogate Safety Assessment Module," *Accident Analysis & Prevention*, vol. 131, 2019.

[4] R. Yu and S. Li, "Exploring the associations between driving volatility and autonomous vehicle hazardous scenarios: Insights from field operational test data," *Accident Analysis & Prevention*, vol. 166, 2022.

[5] Q. Song, K. Tan, P. Runeson, and S. Persson, "Critical Scenario Identification for Realistic Testing of Autonomous Driving Systems," *Research Article*, 2022.

[6] E. De Gelder, J. Hof, E. Cator et al., "Scenario parameter generation method and scenario representativeness metric for scenario-based assessment of automated vehicles," 2022, https://arxiv.org/abs/2106.06215.

[7] P. Jongdo, "A study on issue tracking on multi-cultural studies using topic Modeling," *Journal of the Korean Literature and Information Society*, vol. 53, no. 3, pp. 273–289, 2019.

[8] Y. Emzivat, J. Ibanez-Guzman, P. Martinet, and O. H. Roux, "Dynamic driving task fallback for an automated driving system whose ability to monitor the driving environment has been compromised," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, Redondo Beach, CA, USA, June 2017.

[9] E. Seo and H. Kim, "Security of self-driving car from the point of view of in-vehicle system," *Transaction of The Korean Society of Automotive Engineers*, vol. 26, no. 2, pp. 240–253, 2018.

[10] S. Park, S. Park, H. Jeong, I. YunYun, and J. J. SoSo, "Scenario-mining for level 4 automated vehicle safety assessment from

real accident situations in urban areas using a natural language process," *Sensors*, vol. 21, no. 20, p. 6929, 2021.

[11] N. Kaempchen, B. Schiele, and K. Dietmayer, "Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-vehicle collision scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 678–687, 2009.

[12] O. E. Efe, T. R. Aydos, and S. Emre Aydingoz, "Mechanism of acitretin-induced relaxations in isolated rat thoracic aorta preparations," *Canadian Journal of Physiology and Pharmacology*, vol. 100, no. 1, pp. 35–42, 2022.

[13] X. Li, "A scenario-based development framework for autonomous driving," 2020, https://arxiv.org/abs/2011.01439.

[14] H. Chae, Y. Jeong, M. Lee, K. Min, and K. Yi, "Development of lane change assessment scenarios for autonomous driving vehicles," in *Proceedings of the Korean Society of Mechanical Engineers Spring/Autumn Conference*, pp. 1778–1783, Seoul, South Korea, April 2015.

[15] J. Jeong, "Emergency safety for automated vehicles," *Autojournal*, vol. 42, no. 5, pp. 25–28, 2020.

[16] Y. Choi and J. Lim, "Design for AEBS test scenario applying domestic traffic accidents," *International Journal of Advanced Smart Convergence*, vol. 9, no. 4, pp. 2288–2847, 2020.

[17] D. Kim, O. Kwon, S. Park, and S. Tak, "A study on road driving stability analysis and major factors of autonomous driving based on cluster analysis," *Journal of the Korean Transportation Association*, vol. 33, 2020.

[18] Pegasus, *Pegasus Method*, 2019.

[19] J. Park and M. Song, "A Study on the Research Trends in Library & information science in Korea using topic modeling," *Journal of the Korean Society for information Management*, vol. 30, no. 1, pp. 7–32, 2013.

[20] S. Park, D. Kim, J. Kim, J. Chung, and J. Lee, "Future disaster scenario using big data: a case study of extreme cold wave," *International Journal of Design & Nature and Ecodynamics*, vol. 11, no. 3, pp. 362–369, 2016.

[21] K. Victoria and S. Erduana, "Generating futures from text—scenario development using text mining," *Anticipating Future Innovation Pathways Through Large Data Analysis*, vol. 29–245, 2016.

[22] T. Kim, H. Choi, and H. Lee, "A study on the research trends in fintech using topic modeling," *Journal of the Korea Academia-Industrial cooperation Society*, vol. 17, no. 11, pp. 670–681, 2016.

[23] C. Nam, "An illustrative application of topic modeling method to a farmers diary," *Comparative Culture Study*, vol. 22, 2016.

[24] S. Yoo, K. Park, Y. Park, S. J. LeeHwang, K. Hwang, and K.-S. Kim, "Analysis of domestic industrial security trends using LDA topic modeling," *Korean Journal of Industry Security*, vol. 10, no. 2, pp. 79–103, 2020.

[25] J. Lee, J. Park, J. Yoon et al., "Case analysis of sports match-fixing applying topic modeling," *Journal of the Korean Association for Physical Education Measurement and Evaluation*, vol. 23, no. 2, pp. 51–65, 2021.

[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[27] J. Chuang, C. D. Manning, and J. Heer, "Termite: visualization techniques for assessing textual topic models," *Advanced Visual Interfaces*, vol. 12, pp. 21–25, 2012.

[28] I. M. Miller, "Rebellion, crime and violence in Qing China, 1722-1911: a topic modeling approach," *Poetics*, vol. 41, no. 6, pp. 626–649, 2013.

[29] F. Lind, J. M. Eberl, O. Eisele, T. Heidenreich, S. Galyga, and H. G. Boomgaarden, "Building the bridge: topic modeling for comparative research," *Communication Methods and Measures*, vol. 16, no. 2, pp. 96–114, 2021.

[30] M. Park and J. Son, "Reference test scenarios for assessing the safety of take-over in a conditionally autonomous vehicle," *Transaction of the Korean Society of Automotive Engineers*, vol. 27, no. 4, pp. 309–317, 2019.