

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives



Understanding the road rage behavior and implications: A textual approach using legal cases in Korea

Hyeyeon Ryu^a, Changhun Kim^b, Junghwa Kim^{b,*}, Suejung Lee^a, Jaeyoung Lee^{c,d,f}

^a Department of Forensic Psychology, Kyonggi University, Suwon, Korea

^b Department of Urban and Transportation Engineering, Kyonggi University, Suwon, Korea

^c School of Traffic & Transportation Engineering, Central South University, Changsha, Hunan, China

^d Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL, USA

^f Faculty of Transportation, Kunming University of Science & Technology, Kunming, Yunnan, China

ARTICLE INFO

Keywords:

Road rage
Driving behavior
Traffic safety
Psychological condition
Text mining
Precedent

ABSTRACT

This study aims at exploring the road rage behavior using extreme legal cases of Korea. This study uses a topic modeling algorithm with a text embedding technique to classify the road rage by characteristics (i.e., type, behavior, damage, punishment, and context). We collected 39 representative cases of the road rage from precedent data in Korea. We used nominalizing keywords and a latent Dirichlet allocation (LDA) algorithm of topic modeling among text mining techniques to analyze each case's criminal facts and punishment. After pre-processing by indexing with a coherence score, we set the appropriate number of topics. Through the study, it was confirmed that the model has the highest explanatory power with five topics. We further analyze the five topics for occurrence and subsequent road rage, method of realizing road rage, psychological state of the offenders during road rage, method of collecting evidence of road rage, and type and extent of the damage. By categorizing the keywords, we derive the road rage cases by type. Through this study, precedents dealing with criminal cases of road rage were analyzed to derive the types and causes of retaliatory driving. This result is expected to be used as basic data for the enactment of road rage policies and laws in the future.

1. Introduction

Mobility and accessibility are becoming more important than ever in a rapidly changing modern society. Aftermath, individual demand for vehicles is increasing day by day. According to statistics between 2011 and 2020 in Korea (Korea Ministry of Land, 2021), the number of car registrations increased by 5.93 million and the number of license holders increased by 5.94 million. However, besides the increasing number of vehicles and drivers' licenses each year, traffic crimes are also steadily increasing. Among them, the incidence of violent protests from other drivers and criminal acts that cause mental and material damage using automobiles is increasing (Lee, 2021). Such acts became popular in the 1990s when the United States and Canada media began to report them as road rage (Asbridge, 2006). Road rage has been defined in a variety of ways. Shinar (1998) defined it as hostile behavior toward other road users, such as tailgating or intentional collisions. Smart et al. (2003) defined it as a malicious attempt to feel pleasure from the other driver while driving and threaten or injure the other person. Also, road rage

was defined as follows by NHTSA (National Highway Traffic Safety Administration) "an assault with a motor vehicle or other dangerous weapon by the operator or passenger of one motor vehicle or precipitated by an incident that occurred on a roadway" (Shinar, 1998). In summary, in the Western world, the term road rage is defined as a traffic offense that drives to the extent that it harms a person's body or assets (Kim, 2018). Similarly, according to the Korea Highway Traffic Authority, road rage is defined as any act of threatening or fearing the other party by intentionally using a car in the wake of a dispute on the road (Korea Road Traffic Authority, 2021). In this regard, we tried to understand the fundamental characteristics of road-raged, such as how road-raged crimes occurred in Korea and what factors occurred. According to the recent road rage arrests by the Korean National Police Agency (Agency, 2021), the number of road rage cases increased from 4,400 in 2017 to 5,500 in 2019, an increase of more than 1,000 cases. However, despite increasing annual road rage cases, the prosecution rate against road rage is gradually decreasing from 55% in 2017, 43% in 2018, and 41% in 2019 (Yoo, 2020).

* Corresponding author.

E-mail address: junghwa.kim@kyonggi.ac.kr (J. Kim).

<https://doi.org/10.1016/j.trip.2022.100712>

Received 25 March 2022; Received in revised form 7 October 2022; Accepted 26 October 2022

Available online 9 November 2022

2590-1982/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The American Automobile Association Foundation for Traffic Safety (American Automobile Association Foundation for Traffic Safety, 2013) examined road rage in the United States. It confirmed at least 218 homicides were associated with road rage, and 12,610 cases resulted in injuries. In the United States, guns are often involved in road rage. The trend of increasing retaliatory driving using firearms is accelerating as shown in Table 1. In the United States, there were 616 cases of road rage involving firearms until 2016, and after that, there were 353 additional cases by May 2021. Furthermore, people’s stress factors are worsening due to the prolonged COVID-19 (Lena et al., 2021). This factor increases the violence of road rage.

As such, automobiles are ‘dangerous objects’ that one can mobilize for felony crimes, and the risk of criminal activity using them is very substantial. However, in Korea, the number of indictments in cases is less than half of actual road rage cases (Choi et al., 2020). They are dealt with as summary trials in most cases, and complaints end with a misdemeanor punishment. The Criminal Act judges road rage which has a new punishment clause in the Road Traffic Act since 2016. The Korean National Police Agency officially managed the application of the law and the results of trials in 2017 as shown in Table 2. This study is a starting point for case studies related to road rage. It explores how the Korean courts have been making judgments on road rage cases up to the present time in 2021. Cong et al (2021) demonstrated that attributes related to drivers such as gender, age group, education, driving years, professional or not, daily driving hours, and vehicle type are correlated to road rage behavior. In order to investigate more factors related to road rage in this study, including factors that indicate the attributes of drivers, it was intended to target on real cases rather than survey-based ones. Therefore, we analyze the contents specified in the precedents and check the context of the road rage cases leading to a formal trial, including the damage or punishment of road rage using the topic modeling algorithm during text embedding (Table 3).

2. Literature review

As retaliatory driving emerged as a social problem, various relevant studies have been conducted. Psychological studies have been conducted on the causes of retaliatory driving, negative emotions such as shame, distress, nervousness, and broken heart aligned with driving behavior (Kwon & Oh, 2013). Additionally, attention was paid to the relationship between aggressive driving behavior, aggression, and impulsivity (Berdoulat et al., 2013; Krahé and Fenske, 2001). Also, a comparative study was carried out on aggressive driving behavior, emotions, and anger toward aggressive driving. It confirmed that socio-cultural differences are important in order to understand driving behavior as well (Feng, Lei, Liu, Kumfer, Zhang, Wang & Lu, 2016; McLinton & Dollard, 2010). Drivers on the road react to situational and environmental factors too. It can be imagined the situation in which stress increases when driving in heavy traffic (Deffenbacher et al., 2003). Aggressive driving behavior appears in poor road condition. It was found that congested road was a factor increasing reckless driving

Table 1
Status of Road rage incidents involving a gun and % of Incidents Resulting in Injury or Death in the United States.

Year	Road Rage Incident with a Gun	% of Incidents Resulting in Injury or Death
2016	616	34 %
2017	661	34 %
2018	645	33 %
2019	694	39 %
2020	701	48 %
2021	353	51 %
(Through May)		

Source: Burd-Sharps and Bistline(2021).

Table 2
Ratio of the Perpetrator/Victim Driver’s Gender and Passenger Transport Vehicle.

Division		Incidents (N = 39)	Percentage
Gender of Offender Driver	Male	39	100 %
	Female	0	0 %
Victim’s gender	Male	29	74.36 %
	Female	10	25.64 %
Whether it is a passenger transport vehicle	Offender’s vehicle	5	12.82 %
	Victim’s vehicle	6	15.38 %
	General passenger vehicle	28	71.79 %

Table 3
Example of Pre-processing Process.

Pretreatment Method	Contents
First Index Selection	Among the contents of the precedent, only the contents corresponding to the application of the law are extracted, including ‘name of the court,’ ‘contents of the order,’ ‘criminal facts,’ ‘gist of evidence,’ and ‘applicability of laws and regulations.’
Second Elimination of Stopwords	The defendant ‘changed’ the lane ‘to the first lane’ without turning on the winker—eliminating stopwords such as postpositions, stems, endings, and connecting words.
Third Nominalization and Common Word Conversion	Defendant’s winker was “unused” when changing lane to 1st lane: Change “not turned on” to unused and convert the verb into a noun and a common word to give consistency.
Fourth Surplus Elimination of Keywords	Change to the first lane from the non-use state of the winker: Removal of redundant keywords that appear repeatedly and do not directly affect the result analysis.

behavior (Krahé and Fenske, 2001; DePasquale et al., 2001). As the number of vehicles on the road increases, the road must also be expanded. However, if road expansion is not made, traffic congestion increases, and the resulting stress and tension are expressed as aggressive driving behavior towards other drivers. (Hennessy & Wiesenthal, 2001). Ellison et al.’s (1995) experiment conducted the relationship between driving anger and driver anonymity. The participants who were impossible to recognize other driver’s faces showed more aggressive driving behavior. Moreover, driver anonymity was found to cause drivers to have more signal violations, collisions, speeding, and pedestrian accidents (Ellison et al., 2001). Socio-demographic characteristics are one of the dominant factors influencing road rage behavior. Female drivers tend to drive less aggressively than male (Pfeiffer, Pueschel & Seifert, 2016; Berdoulat et al., 2013; Asbridge & Butters, 2013). Also, age is a key variable related to road rage behavior. It was found that the younger the driver, the more reckless driving such as signal violation or speeding or road rage driving that threatens other drivers (Asbridge & Butters, 2013; Wickens et al., 2011). Also, it was found that drivers in their 40 s were more likely to lane-hogging (lane-hogging: obstructing the passing of other drivers). Other drivers by occupying the passing lane than drivers in their 20 s or 60 s due to their confidence in their driving experience and high risk-taking tendency (Reimer et al., 2013). On the other hand, there is also a study result (Holland and Rabbitt, 1992; Holland and Rabbitt, 1994) that elderly drivers overestimate their driving skills and cause more lane-hogging of other drivers by changing lanes or occupying a passing lane (Holland and Rabbitt, 1994). According to a study examining the difference in driving behavior according to income and educational level aggressive driving behavior occurred more in the driver group with high income and college degrees

(Asbridge & Butters, 2013; Wickens et al., 2011). While extreme violence on the road, such as direct physical violence against other drivers, was found to be associated with a lower income level (Junger, West & Timman; 2001).

We try to use the textual data from precedents for analyzing road rage behavior. Judgments of the courts interpret social conflict based on the law and it serves to elicit social consensus. Studies that analyze precedents to examine how courts interpret and apply laws. Park (2020) studied the process of judicial application for hazardous driving crime by analyzing precedents. Park pointed out that it is difficult to understand the details of the judicial process for a specific crime if you are not a party or a related person to the crime. On the other hand, Park explained that systematically recorded documents such as precedents have the advantage of being able to understand the case and the judicial application process in detail. His study tried to understand the individual specificity and causal properties of each case by comparing and analyzing the case sentences. Additionally, criticisms and implications for the sentencing standards were derived by comparatively analyzing the sentencing conditions for each country. Choi, Gahng & Cho (2014) studied the role of the Psychopathy Checklist-revised (PCL-R), a psychological tool in the US courts, by analyzing precedents. They tried to determine the circumstances and purpose of the PCL-R test being used in the US courts. In addition, it was attempted to identify the legal issues and to draw implications on how these psychological test tools should be used in the courtroom. This study quantitatively and qualitatively analyzed precedents citing the PCL-R test between 2004 and 2014 and confirmed that this tool is recognized as scientific evidence with predictive power on the potential risk of criminals in most courts in the US. Jeon and Park (2021) conducted a study on analyzing judicial precedents for the protection of rights for the demented elderly on abuse caused by family. They collected the related precedents from 2004 to 2020 and they tried to find the shortcomings of the current legal system and to figure out how to protect legal rights. In particular, the main issues were the proof of crimes against elder abuse, determination of mental and physical disabilities, and applicable laws. Jeon and Park confirmed that it was difficult for the elderly with dementia to recognize and report abuse, therefore it was difficult to prove the crime, and the defendants showed an intention to reduce the sentence by claiming mental and physical disorders caused by alcohol. Additionally, they confirmed that the punishment provisions of the Criminal Act were applied in many cases due to the insufficient provision of punishment for elder abuse.

To analyze the textual data, we adopt to use the of topic modeling which is a type of text mining. It is a probabilistic model algorithm that extracts potential topics from a set of atypical documents (Blei, 2012). Topic modeling is like document clustering in that it provides many documents clustered by topic. Topic modeling has the characteristics of soft clustering. Such as one document can simultaneously respond to various topics, it is suitable for modeling real situations. Topic modeling is meaningful in that it can use insights and results in various analyses of vast amounts of documents (Kim et al., 2017). Although, it can provide unobserved latent factors in a document set including hidden variables such as topics and structures of unobserved documents through observed variables such as words included in documents, topics in the entire document set, percentage of topics assigned to each document, and topics (Kim et al., 2016). For this reason, researchers use topic modeling provides easy analysis of a large amount of data (Yoo et al., 2020) and Studies applying topic modeling in various fields have been attempted. Through topic modeling, Ian and Miller (2013) classified violence and crimes in the Qing Dynasty of China in the 18th and 19th centuries into five groups, providing insight into the nature of the crime at the time. Lee (2014) conducted a study by analyzing keyword frequency and a network of court precedents to restore the reliability of the defect judgment system in apartment housing. Through this, they proposed the TO-BE model and a new judgment process for defects by identifying the causes of the problems of the current defect judgment

system through keywords that appear mainly in the defect judgment process. Fabienne et al. (2021) used topic modeling to observe systematic patterns in text and suggested an automated analysis methodology between different languages. In transportation field, research using topic modeling is also being actively conducted. Lijun (2017) conducted an analysis of transportation research using topic modeling. Through topic modeling, an analysis of papers interposed in 22 transportation journals was conducted from 1990 to 2015, and social issues and traffic research trends for each period were presented through the analysis results. Carlos et al. (2019) analyzed the reports of 54 “Irish Road Safety Inspections (RSI)” collected over 6 years (2012–2017) through topic modeling. Through this, the types of safety accidents that occur on roads and key factors were defined, and the significance as basic data for establishing road standards and guidelines was presented. Kenneth (2021) conducted an analysis of traffic accidents through topic modeling. Traffic accident data for 10 years (2009–2018) in Michigan were used to analyze the types and causes of collision accidents. Through this, major topics related to five accidents (crash at stop-sign, crossing the centerline, unable to stop, lane change maneuver, and run-off-road crash) were derived. Park et al. (2021a), Park et al. (2021b) analyzed CADMV autonomous vehicle accident data through topic modeling based on the latent Dirichlet allocation (LDA) algorithm. As a result of the analysis, researchers drew topics and keywords for specific accident situations. They confirmed that they could draw major accident situations using topic modeling as the main contents of the relevant issues: rear-end collision while driving, rear-end collision while stopping, broadside collision, and collision with a freight vehicle. Like the basic data of this study, various legal document analysis studies using topic modeling have been attempted. Ravi and Raghuvveer (2012) conducted case analysis through topic modeling. Legal judgment criteria were grouped through case analysis in India, and it is significant that case analysis through topic modeling was conducted for the first time in India. Sim and Kim (2017) presented a case search and classification system using topic modeling. The system is designed to classify precedents based on the mutual similarity of major keywords within the precedent. Lee et al. (2021) identified significant issues in precedent through topic modeling analysis based on the LDA algorithm. Comparing the topic modeling results with the original text confirmed that manipulating the game results using the discernment method and the basis for judging the sentencing are issues of manipulating game results that appear in precedents. In this study, researchers conducted a precedent analysis to observe the latent patterns and features of road rage behavior in precedents by adopting the topic modeling technique based on the LDA algorithm. Many studies have also been conducted to visualize the analyzed topic modeling results. Since Blei developed the LDA topic modeling algorithm in 2003, research analysis and visualization tools developed actively in various fields (Sievert & Shirely, 2014). Most of the visualization tools provided bar charts, word clouds, and pie charts using a list of highly weighted keywords. This method allows the user to search for appropriate topics and keywords directly but does not examine the overall topic modeling results. Chuang et al. (2012) visualized topic modeling results in a matrix format. Although researchers could identify important keywords within all topics and the relationship between topics and keywords, there is a limitation in showing only keywords with high frequency.

3. Data

For this study, in collecting precedents, we used a specialized platform that provided a legal database. Conventional studies that analyzed precedents often extracted data through the Internet access system of the Supreme Court of Korea search engine. However, in the case of this study, when we set the period to all and searched the keyword ‘road rage,’ we only had three results available for viewing in the Supreme Court’s reading system. In contrast, we extracted 54 case sentences. We limited the cases in the study to criminal cases. We excluded judgments

Table 4

Top 30 Keywords from the Road Rage Precedents.

Rank	Keyword	Frequency	Rank	Keyword	Frequency
1	Intimidation	58	16	Overtake	21
2	Drive	50	17	Traffic	20
3	Court	49	18	Cut in	20
4	District	48	19	Aggravation	19
5	Special	47	20	Fine	19
6	Change	34	21	Workhouse	17
7	Road	30	22	Range	17
8	Imprisonment	29	23	Sudden braking	17
9	Revenge	25	24	Assault	17
10	Hazard	25	25	Bumper	17
11	Lane	24	26	Course	16
12	Destruction	24	27	Stuff	16
13	Horn	23	28	Vehicle	16
14	Probation	23	29	Interruption	15
15	Threat	21	30	Direction	14

with a change in sentencing and sentencing on two instances where the Supreme Court dismissed the appeal. Among the searched data, if the lower court and the appeals court overlapped, we counted the number of road rage cases as one, and the merged case with a filed indictment, including the road rage case, as one case of road rage.

A total of 39 independent cases were collected by integrating the lower and upper courts. Next, we classified the judgment court, case number, sentencing date, order, and reason-criminal facts. The criminal points included specific facts about the place, time, method of the crime, the perpetrator, the victim, and the crime outcome. In this study, to focus on Korean road rage and interpret it, we removed the contents corresponding to the reasons for sentencing specified in the precedent, application of laws and regulations, and the point of evidence before proceeding. Furthermore, the frequency of variables such as the gender of the perpetrator and victim and whether they used their vehicles for passenger transport (Table 4).

4. Topic modeling approaches

4.1. Latent Dirichlet allocation (LDA) algorithm

Latent Dirichlet allocation is a probability-based generative model algorithm for Topic Modeling analysis which is one of the most used methodologies in academic research. LDA provides topic modeling results by inferring latent topics and keywords that are not observed in documents by separating data structures into documents, topics, and keywords. Fig. 1 is a graphical model that explains the variables and operating mechanisms that make up the LDA algorithm.

In the graphical model, D denotes the total number of corpus documents K denotes the total number of topics (hyper parameters), and N denotes the number of words in the d -th document. A square means to repeat the number of times, and a circle refers to a variable. The variable where the arrow starts is a condition, and the variable facing the arrow is a variable corresponding to the result. The only observable variable is $w_{d,n}$ and all latent variables can be estimated through this information.

In the LDA algorithm, the combination probability of the word distribution in the topic and the topic distribution in the document is expressed by the following equation:

$$p(\phi_{i:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\phi_i | \beta) \prod_{d=1}^D p(\theta_d | \alpha) \left\{ \prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \phi_{1:K}, z_{d,n}) \right\} \quad (1)$$

The inference method of the LDA algorithm is to find the variable that maximizes the posterior. Since the value of the corresponding variable cannot be directly observed, collapsed gibbs sampling technique is used.

4.2. LDAvis: Visualize topic modeling results

Since the development of topic modeling, various visualization tools have been developed to derive and interpret clear results. Early visualization tools displayed keywords with high weights in the topic modeling results but could not interpret the overall topic modeling results. Later, in 2014, a visualization tool called LDAvis was developed by Sievert & Shirley. The visualization tool used PCA (Principal Component Analysis), a dimension reduction method, and keyword extraction method to easily grasp the relationship between the topic and the keyword. In addition, it is possible to analyze the distribution of keywords in topics and the similarity between topics through interface manipulation of the result screen. Fig. 2. is the result of topic modeling analysis derived through LDAvis. LDAvis visualizes and provides topic modeling results to analysts through Intertopic Distance Map on the left and Top-30 most Relevant Terms for Topic on the right. The left area provides a two-dimensional topic modeling result applying the dimension reduction method. Each circle represents a topic. The closer the distance between the circles, the higher the similarity between topics. In the area on the right, the relationship between topics and keywords is derived. The information expressed in the corresponding area is presented based on salience and discriminative power.

5. Analysis

5.1. Topic modeling analysis procedure

The procedure for case analysis using topic modeling is shown in Fig. 3. A detailed description of the data purification and topic modeling analysis process is as follows.

First, we carried out preprocessing to analyze the collected precedents. Thus, we extracted keywords that have meaning only after classifying case data to obtain meaningful results. Second, we investigated the keywords constituting each topic drawn through topic modeling and their weight. Third, we calculated the optimal number of topics through the coherence score to implement the LDA algorithm's topic modeling. Finally, we analyzed the contents explained through the derived topics and keywords.

In analyzing unstructured case data, we performed preprocessing to improve the accuracy and clarity of the results. We conducted preprocessing through a series of processes. First, we selected an index suitable for analysis within the precedent. We extracted only the contents corresponding to 'name of the court,' 'contents of order,' 'criminal facts,' 'substance of the evidence,' and 'application of laws' from among the indexes in precedent to observe meaningful analysis results. Second, we eliminated meaningless words such as suffixes and stems. Third, we nominalized all the words and converted words expressed in other forms into common words for analysis. Finally, we removed words with a high frequency of occurrence without affecting the interpretation of the results. Table 5 provides an example of the preprocessing method.

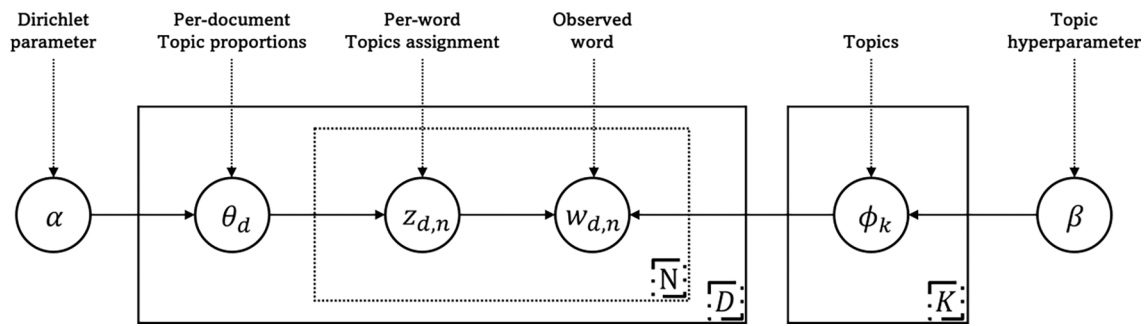


Fig. 1. Graphical model of LDA algorithm operating mechanism.

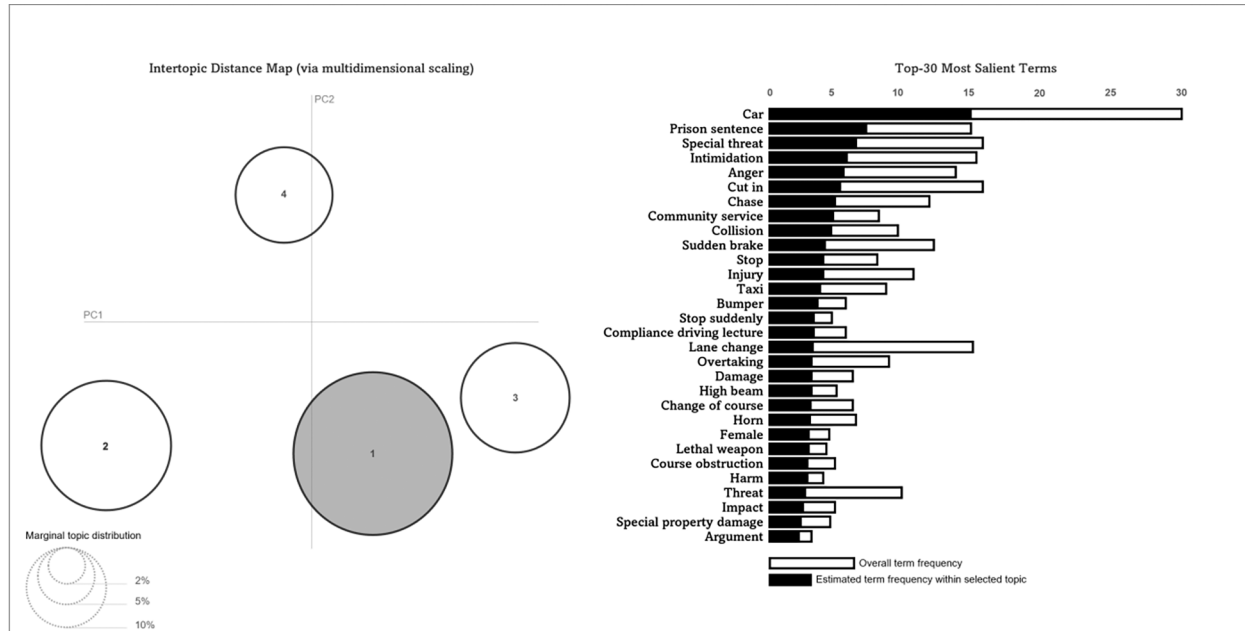


Fig. 2. Topic modeling analysis result derived from LDAvis.

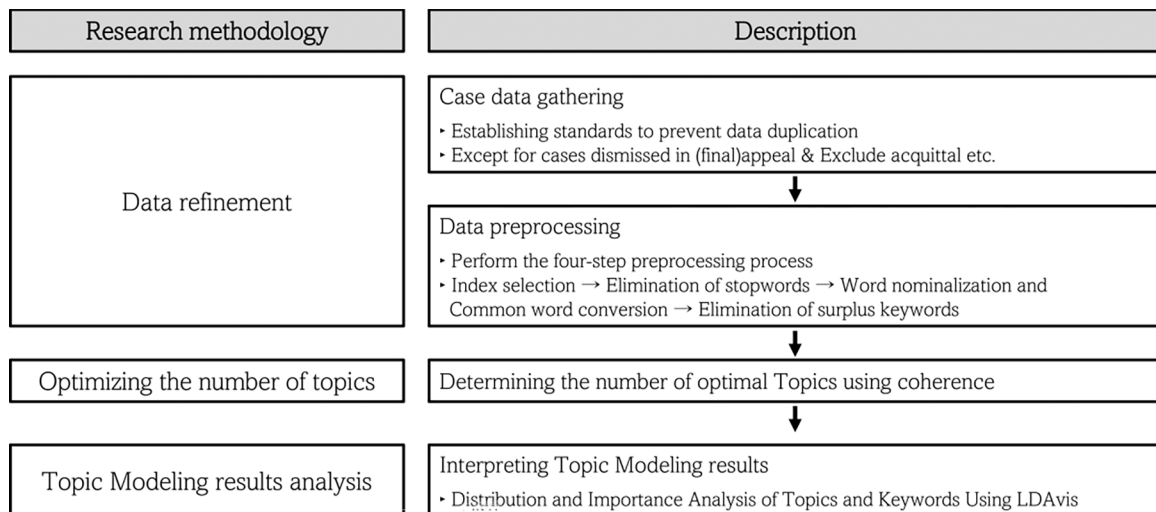


Fig. 3. Process for analysis of road rage cases using Topic Modeling.

Table 5
Topics and Top 10 Keywords from Road Rage Precedents.

Keywords Probability Distribution	Topic1 proportion	Topic2 proportion	Topic3 proportion	Topic4 proportion	Topic5 proportion
1	Threat 0.016	Intimidation 0.021	Imprisonment 0.014	Special intimidation 0.022	Wound 0.019
2	Horn 0.015	Imprisonment 0.018	Aggravation 0.013	Intimidation 0.021	Distortion 0.016
3	Lane change 0.015	Special intimidation 0.017	Probation 0.012	Road rage 0.016	Imprisonment 0.015
4	Imprisonment 0.014	Horn 0.016	Rage 0.010	Overtake 0.015	Traffic law violation 0.014
5	Fine 0.014	Aggravation 0.013	Road rage 0.010	Workhouse 0.015	Community service 0.012
%	25.1	23.4	22.4	16.9	12.1

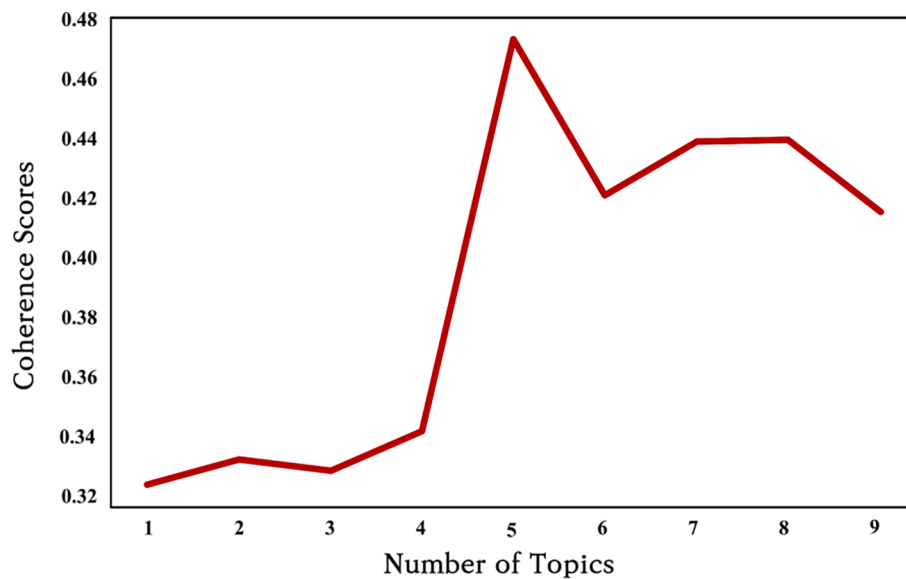


Fig. 4. Coherence Scores by Number of Topics.

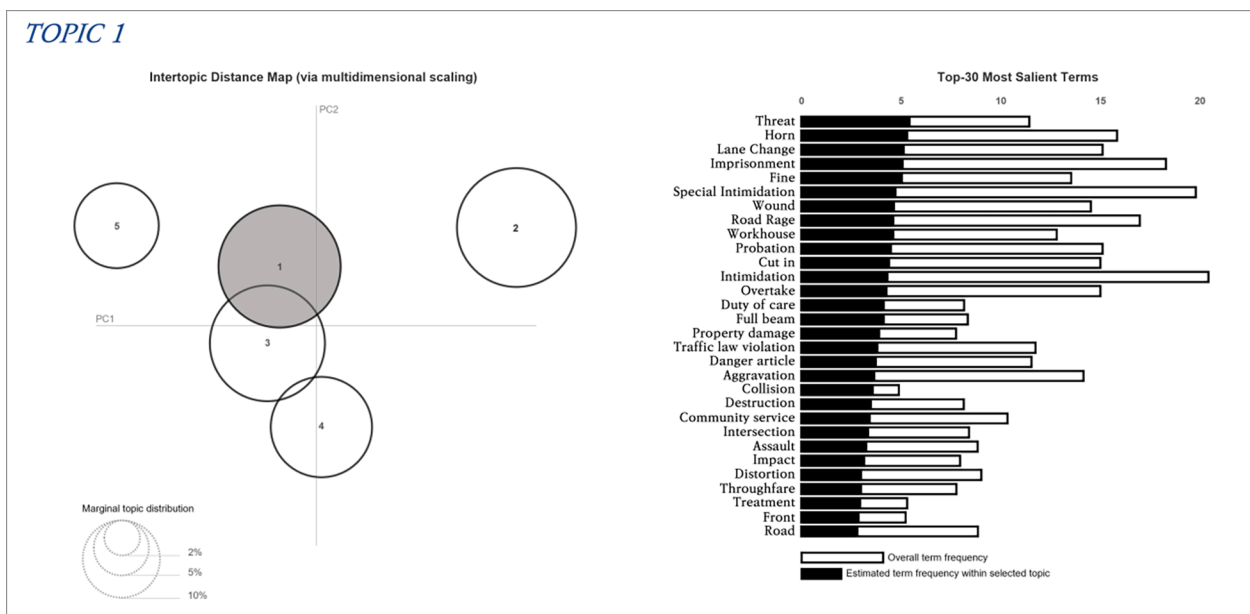


Fig. 5. Visualization of Topic1 analysis.

4.2. Keyword analysis

Table 6 provides an overview of the analysis of road rage precedents. We derived the keyword ‘intimidation’ 58 times, followed by keywords related to judgment such as ‘local,’ ‘court,’ ‘special,’ and ‘imprisonment.’ Then, we drew keywords related to road rage behavior such as ‘road,’ ‘change,’ ‘lane,’ ‘horn,’ and ‘threat’ were drawn. The derived top 30 keywords show that keywords related to ‘judgment,’ ‘driving behavior,’ and ‘method of road rage’ for road rage drivers frequently appear in precedents.

We set an appropriate number of topics so that each topic has a meaningful result to secure independence. To judge this, we used the coherence score as an indicator. The coherence score is a device that measures the consistency of keywords within a topic and measures the mutual similarity of the top N items within a topic. Thus, the higher the coherence score value, the higher the consistency of the topic. Fig. 4 shows the coherence score according to the number of topics. When the number of topics was 5, the coherence score was the highest at 0.473. This result means that when the number of topics is 5, the relevance of keywords constituting each topic is the highest. When the number of topics was 10 or more, we excluded topics with similar contents due to duplicated deduction.

4.3. Modeling result

As depicted in Fig. 5 and Fig. 6 keywords constituting Topic 1 included ‘threat,’ ‘horn,’ ‘lane change,’ ‘imprisonment,’ ‘fine,’ ‘special intimidation,’ and ‘road rage.’ The act of honking became the reason for road rage, and it caused road rage behavior that threatened the other drivers with lane changes or cut-in. Afterwards, they threatened and injured the victim secondarily. Finally, the perpetrator received a fine or imprisonment for the crime of special intimidation. Keywords constituting Topic 2 included ‘special intimidation,’ ‘intimidation,’ ‘road rage,’ ‘overtaking,’ ‘prison workshop (workhouse),’ ‘fine,’ ‘sudden braking,’ ‘cut-in,’ ‘horn/honking,’ etc. The act of honking became the reason for road rage and caused road rage behavior that threatened the other drivers with acts of overtaking, sudden braking, and cut-in. We could not draw keywords related to secondary physical victimization, and the perpetrator received a fine for the crime of special intimidation.

Fig. 7 illustrates the keywords constituting Topic 3, including ‘imprisonment,’ ‘aggravation,’ ‘suspended prison sentence,’ ‘anger,’

‘road rage,’ ‘threat,’ ‘lane change,’ and ‘winker.’ Through the keywords of ‘angry’ and ‘aggravation,’ it is possible to understand the psychological state of the offender at the time of road rage. In addition, they showed road rage behavior by lane changes without using their winker. As a result, the perpetrator received a suspended prison or imprisonment sentence for special intimidation and traffic law violation.

Fig. 8 shows the keywords constituting Topic 4, including ‘special intimidation,’ ‘road rage,’ ‘overtaking,’ ‘fine,’ ‘sudden braking,’ and ‘dashboard camera’ road rage showing behavior such as overtaking or sudden braking. In these instances, the perpetrator received a fine for the crime of special intimidation. The situation analysis for the case was through the dashboard camera.

Fig. 9 depicts keywords constituting Topic 5, including ‘injury,’ ‘distortion,’ ‘imprisonment,’ ‘traffic law violation,’ ‘community service,’ ‘suspended prison sentence,’ ‘destruction,’ ‘diagnosis/medical certificate,’ ‘tension,’ etc. We can identify the degree of damage to the victim through keywords such as ‘injury,’ ‘distortion,’ ‘tension,’ ‘diagnosis,’ and ‘destruction.’ In this instance, the perpetrator received community service, including community service for traffic law violations.

Table 7 shows the proportion of the top 10 keywords included due to topic modeling based on the LDA algorithm. Based on this, we define the shows the names and main contents of the topics collected through the analysis of road rage precedents. Firstly Topic 1 covers the overall content of road rage, from the occurrence of road rage to the behavior of the road rage driver and the resulting judgment. Topic 2 contains information on road rage methods for road rage drivers. Topic 3 was a topic that showed the psychological state of the perpetrator and victim during road rage. Topic 4 is about how to collect evidence to determine the illegality of road rage drivers. Finally, topic 5 identifies the types of injuries and vehicle damage suffered by victims of road rage. The types and characteristics of road rage can be identified based on 5 topics and the main contents derived from the topics are summarized as follows.

[Road rage occurrence and subsequent situation] The horn and full beam of the other vehicle caused road rage. It performs through sudden lane changes and cut-in. Since then, the perpetrator has made threats along with intimidation and assault. The perpetrator’s sentence included a fine and imprisonment for traffic law violation, a special threat.

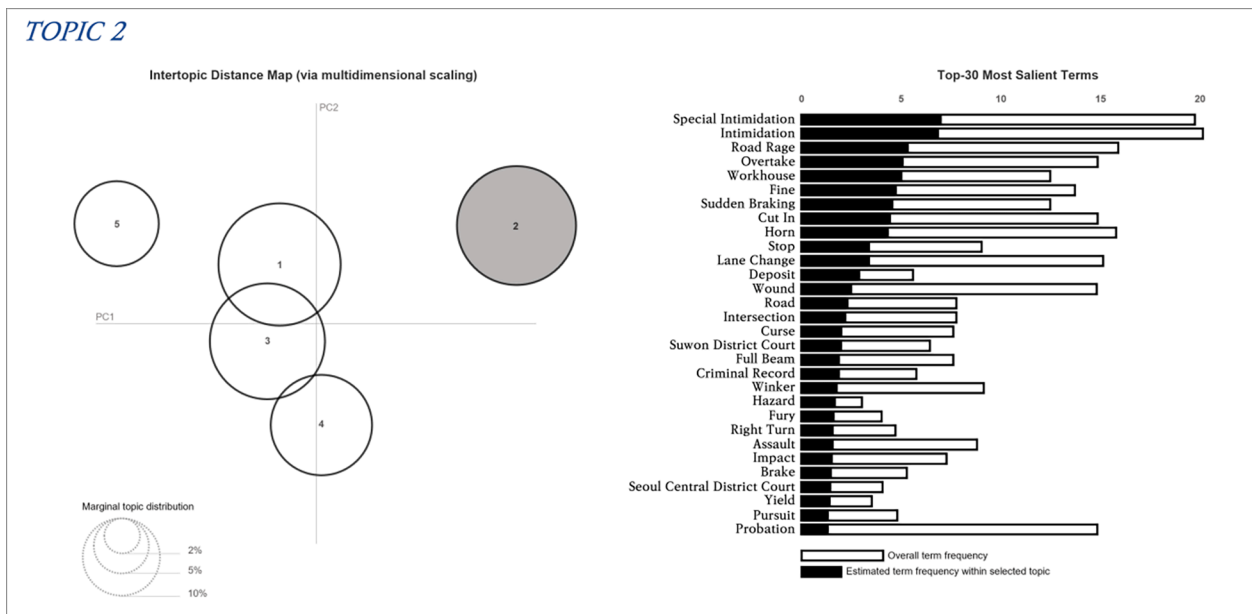


Fig. 6. Visualization of Topic2 analysis.

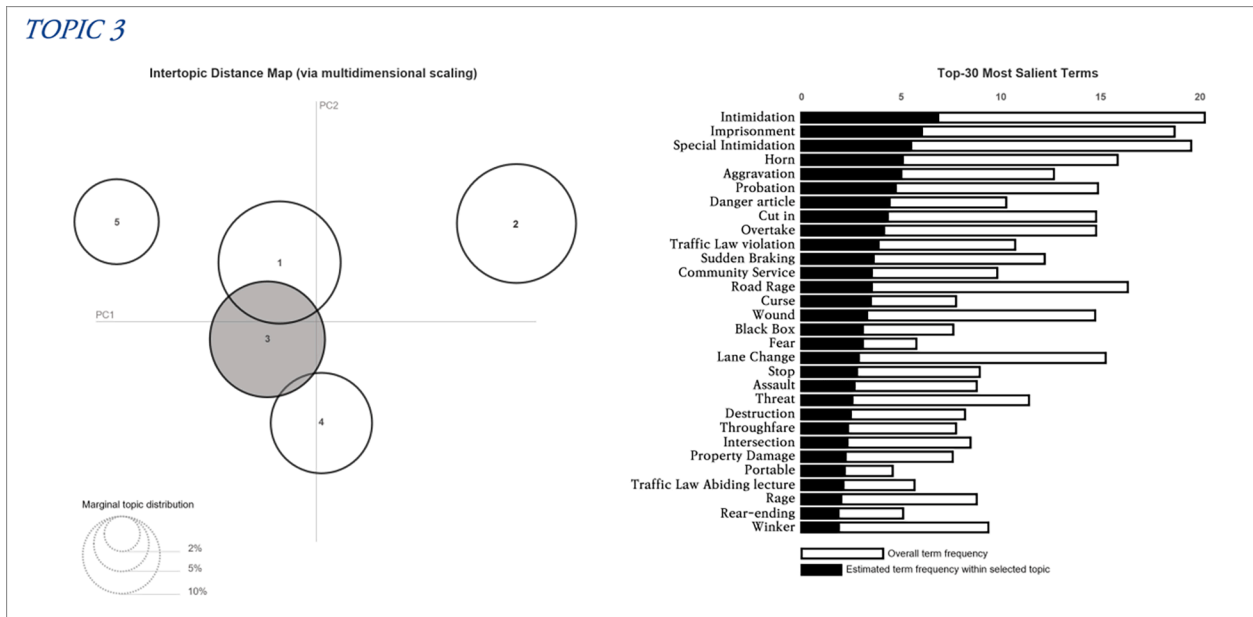


Fig. 7. Visualization of Topic3 analysis.

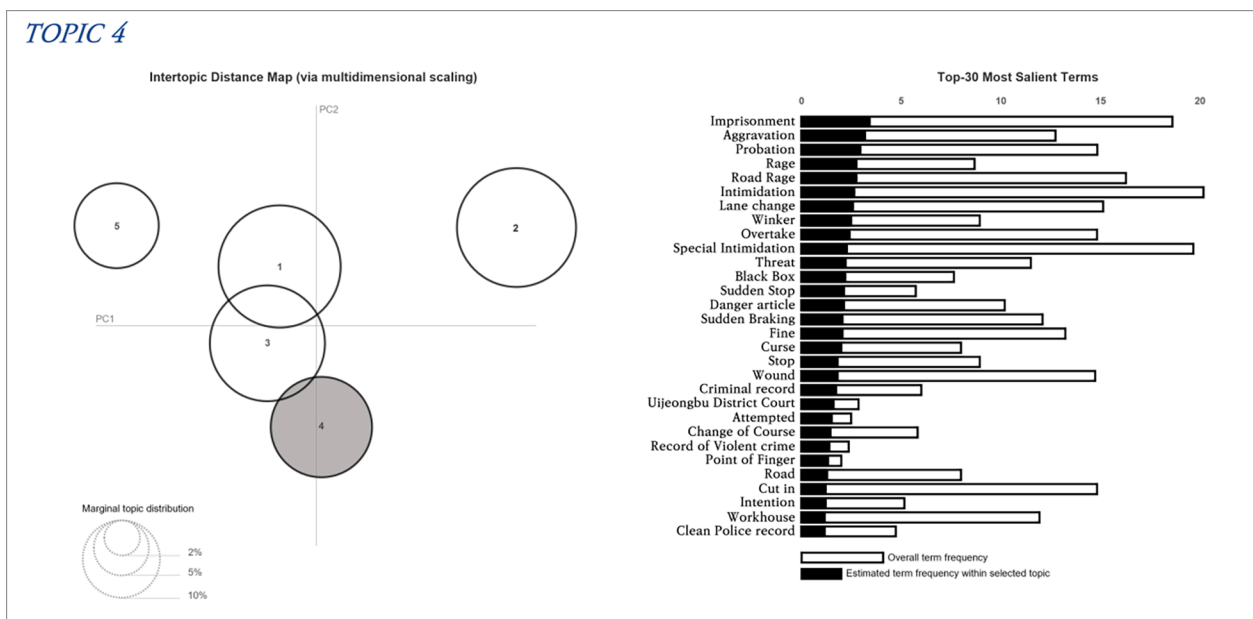


Fig. 8. Visualization of Topic4 analysis.

[Road rage behavior] The perpetrators act out road rage through overtaking, sudden braking, interruption, lane change, swearing, and no blinker.

[Psychological conditions in road rage] The perpetrator momentarily loses control of his emotions and vents his anger in extreme ways, such as swearing, assault, and collision. As a result, the victim feels scared and threatened.

[Gathering evidence] The driving behavior at the time of road rage and the criminal circumstances such as assault, cursing, and use of dangerous items are mainly collected and analyzed through dashboard camera in the vehicle.

[Type of damage] Collisions caused by road rage damage to the vehicle and injuries to the victim. There are types of injuries such as muscle tension and sprains, and the situation may worsen due to hit and run.

5. Conclusion

This study proposed a topic modeling approach based on the LDA algorithm using Korean road rage precedents. Since the proportion of unstructured data such as text is showing an overwhelming increase (Dang et al., 2014; Talib et al., 2016), this unstructured data is being used in various fields through the development of big data technology. Among them, a new methodological attempt to interpret meaningful information using widely used text mining and topic modeling techniques, this study attempted by introducing it into the criminal case judgment data for road rage crimes. From the modeling results, it is possible to analyze the characteristics of road rage, such as the main types and causes of road rage. Through the text embedding technique, it was possible to obtain topics appearing in the case sentence as five distinct categories. This method is rather than deducing the process of

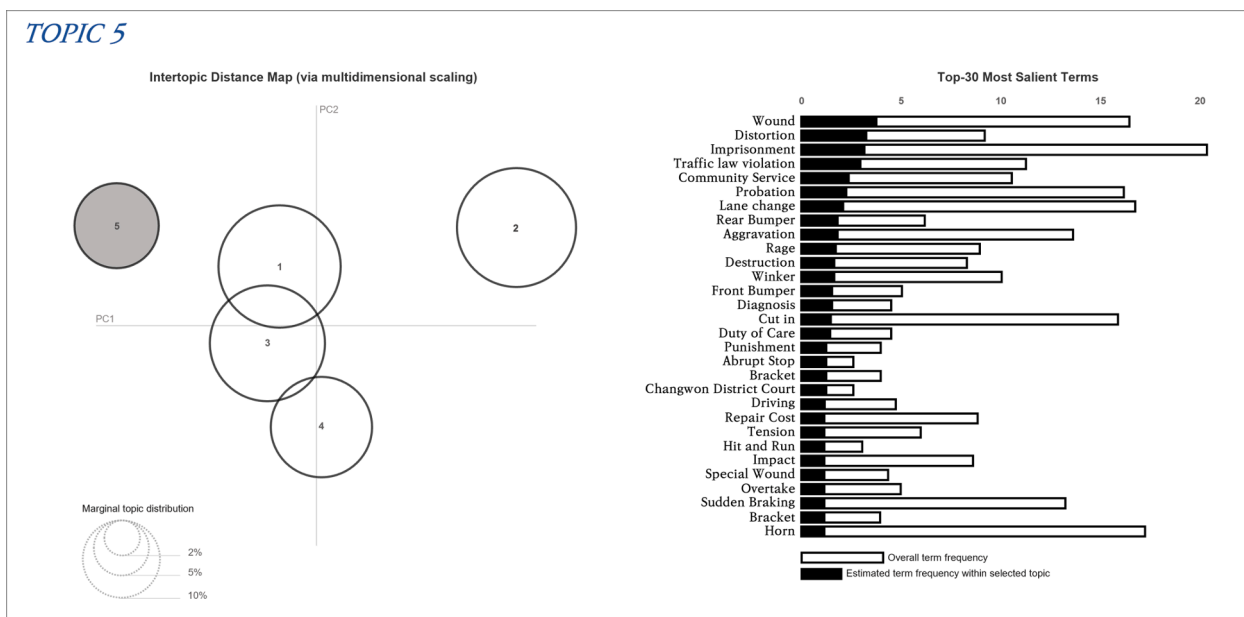


Fig. 9. Visualization of Topic 5 analysis.

road rage similar to each topic, including deducing the process before and after the occurrence of road rage, realizing road rage, the psychological state of the road rage driver, the method of collecting evidence of road rage, and the type and extent of road rage damage. You can now check the details for each category. For example, looking at the topic related to road rage occurrence and subsequent situation, we see that before the act of road rage, minor actions such as non-use of winker or use of horns acted as triggers, which we expect emphasized in traffic safety education or a driver's license test. In addition, in the topic of the psychological state, road rage is derived from the anger mechanism of the driver who drives in retaliation and the behavioral results. This type of behavior can include specialized training for road rage that the Road Traffic Authority could further consider and deal with when implementing the driver's safety driving education. As another example, the topic of damage type and severity raises awareness of the seriousness of road rage to members of society. Courts could use this as a reference point for victim-centered judicial judgment that can help to understand the psychological damage of victims of road rage crimes.

In this study, we identified the characteristics of road rage in 5 topics. We were able to confirm the results consistent with the characteristics of road rage confirmed in previous studies by other researchers. For example, Lee (2021) analyzed the characteristics and tendencies of Korean road rage driving criminal cases based on the trial judge in his study. As a result, contact road rage through sudden breaking in was the most frequent. This is consistent with the features according to the frequency of road rage behavior, such as sudden cut-in that this study identified in the topic of road rage behavior.

We also found that honking and full beam are one of the causes of conflicts on the road. According to Wu et al. (2018), honking and full beam are the most common behaviors expressing anger and disagreement among Chinese drivers as well. Furthermore, in a survey conducted by Demirkol, Tosun, & Yuksel (2015) on full-time drivers, the highest factor loading value (0.85) was given to the response 'When other drivers honk at me while I'm driving makes me angry'. The use of horns is not only a cause of anger while driving but also a way of expressing anger. For example, Sullman et al. (2013) conducted a questionnaire on the expression of anger by taxi drivers in Turkey. As a result, using honk and beam was confirmed as the most common way to express their anger towards other drivers following verbal expressions.

On the other hand, in this study, muscle tension or sprain was derived as a topic as a type of damage caused by road rage to the victim.

However, many previous studies, such as Shin (2020) and Choi, Kim, & Kang (2020), conducted a questionnaire on the psychological damage caused by road rage to victim drivers and, as a result, psychological atrophy during driving, which led to in-depth considerations on the psychological damage factors such as the sense of threat and fear felt at the time of the incident. We expect that our study obtained the victim's physical damage as a representative topic result because, given the data's nature, the description of the victim's physical damage is much larger than the description of the victim's psychological damage on legal precedents.

In this paper, we present a new method for analyzing road rage crimes. Furthermore, a series of processes in the analysis confirmed the possibility of case analysis using topic modeling. Through case analysis, it was possible to derive complex characteristics such as trigger, behavior pattern, and level of punishment of road rage in Korea. If data are managed and collected based on the characteristics and themes of retaliatory driving derived from our analysis, it is expected to be of appropriate help in deriving countermeasures against ever-increasing road rage.

Road violence is a clear criminal act committed on the road targeting certain victims. However, efforts to manage and data collection activities related to road rage are still very scarce in Korea. In addition, there is no direct bill to punish road rage, so it is being punished by the criminal law. Therefore, information related to the actual situation of retaliatory driving only exists as inside data of the National Police Agency. For this reason, accessing the data was also difficult for researchers and the general public. Empirical studies on the continuously increasing driving population or road rage cases have not been conducted sufficiently and actively in Korea and abroad. This paper serves as a starting point for analyzing road rage cases in Korea that the courts of justice have even judged. In particular, it is also academically meaningful to derive the aspects of road rage cases not treated as misdemeanors by applying text embedding technology based on the crime facts appearing in case law. Moreover, this study interpreted road rage precedents through topic modeling and contributed to identifying the characteristics and types of retaliatory driving. Through this study, we analyzed road rage in the trial stage by maximizing the resources of a limited database. As a result of this study, we expect that if the data built every year starts to be managed based on the topics divided through this study, it will be appropriate assistance for continuously studying countermeasures against the increasing numbers of road rage in the future.

Starting with this study, researchers could actively conduct follow-up studies that resolve various interdisciplinary questions based on more data through the continuous management of statistics on road rage by the Korean NPA. Although we conducted this study with the judgments of the Republic of Korea, it would be meaningful to analyze the judgments of road rage in the United States, Canada, the UK, and Australia and classify the types of road rage in each country.

We expect our study to serve as a valuable reference for future researchers as a basic outline of road rage data. Although this study has its contributions, it is not without a limitation. The number of precedents used in the study is relatively small. This is because the collectable precedents are limited. In Korea, the Supreme Court sentence data is the easiest to collect. But it is only 3 % of the overall sentences. In the case of lower courts, only 0.003 % of precedents are disclosed (Lee and Park, 2017). In the case of LDA research, hundreds of thousands of data are usually used. Therefore, further studies would be required with larger-scale data are needed for a more effective technical approach of topic modeling.

CRedit authorship contribution statement

Hyeyeon Ryu: Data curation, Writing – original draft. **Changhun Kim:** Visualization, Investigation, Software. **Junghwa Kim:** Conceptualization, Methodology, Software, Validation. **Suejung Lee:** Supervision. **Jaeyoung Lee:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgement

This research was supported by a grant(2022-MOIS41-002) of Citizen-customized Life Safety Technology Development Program funded by Ministry of Interior and Safety(MOIS, Korea).

References

- Agency, K.N.P., 2021. Driver's License Holders Status. Statistics Korea. https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1617#quick_02.
- American Automobile Association Foundation for Traffic Safety. 2013. *Road rage: How to avoid aggressive driving*. <https://exchange.aaa.com/wp-content/uploads/2013/06/Road-Rage-Brochure.pdf>.
- Asbridge, M., Butters, J., 2013. Driving Frequency and Its Impact on Road Rage Offending and Victimization: A View from Opportunity Theory. *Violence Vict.* 28 (4), 602–618.
- Berdoulat, E., Vavassori, D., Sastre, M.T.m., 2013. Driving Anger, Emotional and Instrumental Aggressiveness, and Impulsiveness in the Prediction of Aggressive and Transgressive Driving. *Accid. Anal. Prev.* 50, 758–767.
- Blei, D.M., 2012. Proximalistic topic models. *Commun. ACM* 55 (4), 77–84.
- Burd-Sharps, S., Bistline, K., 2021. Reports of Road Rage Shootings are on the Rise. *Everytown Res. Policy*. <https://everytownresearch.org/reports-of-road-rage-shootings-are-on-the-rise/>.
- Carlos, R., Joao, C., Thomas, C., Govert, S., 2019. Topic analysis of Road Safety Inspections using latent Dirichlet allocation: A case study of roadside safety in Irish main roads. *Accid. Anal. Prev.* 131C (1), 336–349.
- Choi, I., Gahng, T., Cho, E., 2014. The Role and Relevance of the Psychopathy Checklist-Revised in the U.S. courts. *Korean Institute of Criminology* 25 (4), 375–414.
- Choi, S., Kim, K., Kang, J., 2020. An Exploratory study on Aggressive Driving and Road Rage in Korea. *Korean Institute of Criminology* 1–460.
- Chuang, J., Manning, C.D., Heer, J., 2012. Termite: Visualization techniques for assessing textual topic models. *Advanced Visual Interfaces* 12, 21–25.
- Cong, H., Shi, X., Cooper, J., Suo, Z., Zhao, X., 2021. Road rage in China: An exploratory study. *J. Transport. Saf. Secur.* 13 (5), 503–524.
- Dang, S., Ahmad, P.H., 2014. Text Mining: Techniques and its Application. *Int. J. Eng. Technol. Innovat.* 1 (4).

- Deffenbacher, J.L., Deffenbacher, D.M., Lynch, R.S., Richards, T.L., 2003. Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behav. Res. Therapy* 41 (6), 701–718.
- Demirkol, I.C., Tosun, H., Yuksel, Y., 2015. What Causes Drivers to Involve in a Traffic Altercation? *J. Instit. Soc. Sci.* 33, 59–67.
- DePasquale, J.P., Geller, E.S., Clarke, S.W., Littleton, L.C., 2001. Measuring road rage: development of the Propensity for Angry Driving Scale. *J. Saf. Res.* 32, 1–16.
- Ellison, P., Bell, P., Deffenbacher, J., 2001. The Effects of Trait Driving Anger, Anonymity, and Aggressive Stimuli on Aggressive Driving Behavior. *J. Appl. Soc. Psychol.* 31 (2), 431–443.
- Ellison, P.A., Govern, J.M., Petri, H.L., Figler, M.H., 1995. Anonymity and Aggressive Driving Behavior: A field study. *J. Soc. Behav. Personal.* 10, 265–272.
- Fabienne, L., Jakob-Moritz, E., Olga, E., Tobias, H., Sebastian, G., Hajo, G., Boomgaarden. 2021. Building the Bridge: Topic Modeling for Comparative Research. *Commun. Methods Measur.*
- Feng, Z., Lei, Y., Liu, H., Kumfer, W., Zhang, W., Wang, K., Lu, S., 2016. Driving Anger in China: A Case Study on Professional Drivers. *Transport. Res. Part F: Traffic Psychol. Behav.* 42, 255–266.
- Hennessey, D.A., Wiesenthal, D.L., 2001. Gender, Driver Aggression, and Driver Violence: An Applied Evaluation. *Sex Roles* 44 (11/12), 661–676.
- Holland, C.A., Rabbitt, P.M., 1994. The Problems of Being an Older Driver: Comparing the Perceptions of an Expert Group and Older Drivers. *Appl. Ergon.* 25 (1), 17–27.
- Ian, M., Miller, 2013. Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach. *Poetics* 41 (6), 626–649.
- Jeon, H.W., Park, K.I., 2021. Study on Analyzing Judicial Precedents for Protection of Rights for the Demented Elderly on Abuse caused by Family: Focusing on the Proof of Crime, Judgment of Insanity, Application of Laws. *Korean Soc. Secur. Stud.* 37 (1), 93.
- Junger, M., West, R., Timman, R., 2001. Crime and Risky Behavior in Traffic: An Example of Cross-situational Consistency. *J. Res. Crime Delinqu.* 38 (4), 439–459.
- Kim, T.G., Choi, H.R., Lee, H.C., 2016. A Study of the Research trends in fintech using topic modeling. *J. Korea Academia-Indust. Cooperat. Soc.* 17 (11), 670–681.
- Kim, N.G., Lee, D.H., Choi, H.C., Wong, W.X.S., 2017. Investigations on techniques and application of text analytics. *J. Korean Instit. Commun. Informat. Sci.* 42 (2), 471–492.
- Kim, S., H., 2018. A Study on the Aggressive Driving in Korea. *J. Korean Public Police Secur. Stud.* 15 (1), 29–52.
- Korea Ministry of Land, Infrastructure and Transport. 2021. *Vehicle Registration Status*. Statistics Korea. https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1257.
- Korea Road Traffic Authority. 2021. *Traffic safety information*. Road rage definition https://www.koroad.or.kr/kp_web/safeDriveObligation7.do.
- Krahe, B., Fenske, I., 2001. Predicting Aggressive Driving Behavior: The Role of Machoperpersonality, Age, and Power of Car. *Aggressive Behavior.* 28 (1), 21–29.
- Kwon, M.J., Oh, Y.T., 2013. A Study on Effects of the Driver's Emotion on the Driving Behavior. *Kore. Soc. Transport. Vol* 31. Issue 6.
- Lee, K. 2014. Study on improvement of apartment housing defect judgment system based in judicial precedent cases (Doctoral dissertation). *Hanyang University*.
- Lee, J.W., 2021. An Analysis on the Tendency of Victimization of Road Rage Cases in South Korea. *Korean J. Public Saf. Criminal Justice* 30 (4), 89–116.
- Lee, J.Y., Park, J.H., Yoon, J.W., Lee, S.B., Youm, S.K., Lee, Y.S., Yoon, H.J., 2021. Analyzing Precedents for Sports Match Fixing by Applying Topic Modeling. *Kor. Soc. Measure. Evaluat. Phys. Educat. Sports Sci.* 23 (2), 51–65.
- Lee, Y.G., Park, J.H., 2017. An analysis of the research trend in The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science using Topic Models. *Kor. J. Measure. Evaluat. Phys. Educat. Sports Sci.* 19 (2), 11–22.
- Lena, S., Pfeifer, Kartin, H., Sebastian, O., Oliver, T., Wolf. 2021. Stress research during the COVID-19 pandemic and beyond. *Neurosci. Biobehav. Rev.*, 131, 581–596.
- McLinton, S.S., Dollard, M.F., 2010. Work Stress and Driving Anger in Japan. *Accid. Anal. Prev.* 42, 174–181.
- Park, J.H., 2020. A Comprehensive Study on the Process of Judicial Application of Hazardous Driving Crime – Focused on the Case Analysis between Korea and Japan. Konkuk University.
- Park, S.M., Lee, H.P., So, J.H., Yun, I.S., 2021a. Study of analysis for autonomous vehicle collision using text embedding. *J. Korea Instit. Intellig. Transp. Syst.* 20 (1), 160–173.
- Park, S.M., Lee, Y.N., Choi, A.R., Ahn, J.J., 2021b. The 'Online Access to Judgment' Service in Korea: A Study on Improving Judgment Data for the Development of Legal AI (Artificial Intelligence). *J. Police Law.* 19 (3), 3–36.
- Pfeiffer, J.L., Puschel, K., Seifert, D., 2016. Interpersonal Violence in Road Rage. Cases from the Medico-Legal Center for Victims of Violence in Hamburg. *J. Forensic Leg. Med.* 39, 42–45.
- Ravi, K.V., Raghuvver, K., 2012. Legal Documents Clustering using Latent Dirichlet Allocation. *Int. J. Appl. Informat. Syst.* 2 (6), 27–33.
- Reimer, B., Donmez, B., Lavalliere, M., Mehler, B., Coughlin, J., Teasdale, N., 2013. Impact of Age and Cognitive Demand on Lane Choice and Changing under Actual Highway Conditions. *Accid. Anal. Prev.* 52, 125–132.
- Shinar, D., 1998. Aggressive driving: the contribution of the drivers and the situation. *Transport. Res. Part F: traffic psychol. Behav.* 1 (2), 137–160.
- Sievert, C., Shirely, K. E. 2014. LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70.
- Sim, J., Kim, H., 2017. A Searching Method for Legal Case Using LDA Topic Modeling. *J. Instit. Electron. Informat. Eng.* 54 (9), 67–75.
- Smart, R., Mann, G., Stoduto, R.E., 2003. The prevalence of road rage. *Can. J. Public Health* 94 (4), 247–250.

- Sullman, M.J.M., Stephens, A.N., Kuzu, D., 2013. The Expression of Anger amongst Turkish Taxi Drivers. *Accid. Anal. Prev.* 56, 42–50.
- Talib, R., Hanif, M.K., Ayesha, S., FaTIMA, F., 2016. Text Mining: Techniques, Applications and Issues. *Int. J. Adv. Comput. Sci. Appl.* 7 (11), 414–418.
- Wickens, C.M., Mann, R.E., Stoduto, G., Ialomiteanu, A., Smart, R.G., 2011. Age Group Differences in Self-reported Aggressive Driving Perpetration and Victimization. *Transport. Res. Part F: Traffic Psychol. Behav.* 14 (5), 400–412.
- Wu, X., Wang, Y., Peng, Z., Chen, Q., 2018. A Questionnaire Survey on Road Rage and Anger-provoking Situations in China. *Accid. Anal. Prev.* 111, 210–221.
- Yoo, S.T., Park, K.S., Lee, Y.S., Hwang, S.J., Kim, K.S., 2020. Analysis of domestic industrial security trends using LDA topic modeling. *Korean Associat. Indust. Secur.* 10 (2), 79–103.
- Yoo, M. (2020, October 8). Last year, 5,536 cases of road rage incidents increased by 25% compared to the previous year. *Dailynews*. <https://www.idailynews.co.kr/m/view.php?idx=76456>.