

ARTICLE

토픽모델링 기반의 판례분석을 통한 난폭운전과 보복운전 특성 비교 연구

김창훈¹ · 이수정² · 김정화^{3*}¹경기대학교 스마트시티공학부 도시·교통공학전공 학부과정, ²경기대학교 범죄심리학과 교수,³경기대학교 스마트시티공학부 도시·교통공학전공 조교수

Exploring the Features of Aggressive Driving and Road Rage by Analyzing Judicial Precedents Based on Topic Modeling

KIM, Changhun¹  · LEE, Suejung^{2*}  · KIM, Junghwa^{3*} ¹Student, Department of Urban & Transportation Engineering, College of Creative Engineering, Kyonggi University, Gyeonggi 16227, Korea²Professor, Department of Forensic Psychology, Kyonggi University, Gyeonggi 16227, Korea³Assistant Professor, Department of Urban & Transportation Engineering, College of Creative Engineering, Kyonggi University, Gyeonggi 16227, Korea

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

Abstract

Case of damage caused by aggressive driving and road rage have been reported frequently through various media. Diversifying and serious crimes cause a lot of damage, creating anxiety for other drivers on the road. Compared with violence and social seriousness of the crime, research activities have not been conducted due to the lack of academic interest. The purpose of this study is to present a reference point of legal punishment for road rage crimes and policies to protect victims. we tried to understand the crime characteristics by case analysis using topic Modeling which infers potential variables within a set of documents. The analysis of case data was conducted through the process of collecting case data, data preprocessing, determining the optimal number of topics, topic modeling, and comparative analysis of characteristics. 23 Aggressive driving, 31 Road rage, a total of 54 case datas were collected by setting Classification criteria for avoid data duplication and quality improvement. Then data preprocessing was performed through four steps of index selection, term removal, nounization, with common word conversion and surplus keyword removal in case data. As a results of the topic Modeling analysis, four topics and sub keywords were derived to understand the characteristics of aggressive driving and road rage. Classified topic and keywords for Comparative analysis were analyzed in four types (behavior pattern, damage type, crime fact and judgment, crime target and place), then visualization work has been conducted using Radar Chart to compare the characteristics of aggressive driving and road rage include distinguish the similarities and differences between the two crimes. The methodology applied in this study confirmed specific crime forms and characteristics can be derived through case study, and it is expected to be used to develop police guidelines, crime prevention systems, and infrastructure to crack down on aggressive driving.

Keywords: aggressive driving, case study, characteristic study, road rage, topic modeling

J. Korean Soc. Transp.
Vol.40, No.5, pp.683-699, October 2022
<https://doi.org/10.7470/jkst.2022.40.5.683>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 5 August 2022

Revised: 10 August 2022

Accepted: 10 August 2022

Copyright ©
Korean Society of Transportation

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

초록

여러 매체를 통해 난폭운전과 보복운전으로 인한 피해사례가 빈번하게 보도되고 있다. 다양해지고 심각해지는 가해 수법으로 피해사례가 늘어나고 있으며, 도로 위 운전자들의 불안감이 조성되고 있다. 해당 범죄의 폭력성, 사회적 심각성과 비교했을 때 이에 대한 정책적, 학술적 관심은 현저히 낮으며, 관련된 실증연구 또한 미흡한 수준이다. 본 연구에서는 가해자 처벌 및 피해자 구호를 위한 법률적 근거를 마련하는 데 있어 기준점을 제시하고자 문서 집합 내에 잠재 변수를 추론하는 기법인 토픽모델링을 활용하여 판례분석을 진행함으로써, 범죄 특성 및 특성 간 연관성을 파악하고자 하였다. 판례 데이터의 분석은 판례 데이터 수집, 데이터 전처리, 최적 토픽 수 결정, 토픽모델링, 특성 간 비교분석에 이르는 과정을 통해 진행되었다. 데이터의 중복 방지 및 품질 향상을 위한 기준을 마련하여 난폭운전 23건, 보복운전 31건 총 54건의 판례 데이터를 수집하였으며, 이후 비정형의 판례 데이터의 분석을 위해 판례 내 인덱스 선별, 불용어 제거, 명사화 및 공통단어 변환, 잉여 키워드 제거의 4단계 절차를 통해 데이터 전처리 작업을 수행하였다. 토픽모델링 분석 결과를 통해 난폭운전과 보복운전의 특성을 파악할 수 있는 각각 4가지의 잠재 토픽과 하위 키워드를 도출하였다. 도출된 토픽과 토픽 내 키워드들에서 비교 분석이 가능한 특성을 추출하여 4가지 유형(행동양상, 피해유형, 범죄사실과 판결, 범행대상과 장소)에 대한 분석을 진행하였으며, Radar Chart를 이용한 시각화를 통해 난폭운전과 보복운전의 특성 비교 및 두 범죄의 유사점과 차이점을 구분하였다. 본 연구 수행 과정에서 적용된 방법론은 판례를 통해 특정 범죄의 형태와 특성을 도출할 수 있음을 확인하였으며, 향후 난폭·보복운전 관련 법률 제정 및 법률적 판단 기준을 마련하는 기초자료 제작과 함께 이를 활용한 난폭·보복운전 단속을 위한 경찰 가이드라인, 범죄 예방 시스템 및 인프라 개발 등에 활용할 수 있을 것으로 기대된다.

주요어: 난폭운전, 판례분석, 특성 분석, 보복운전, 토픽모델링

서론

대한민국의 자동차 등록 대수는 2011년 1,843만 대에서 2020년 2,436만대로, 9년 동안 약 593만여 대가 증가하였다(Ministry of Land, Infrastructure and Transport, 2022). 이와 더불어, 운전면허 소지자 현황 또한 2011년 2,725만 명에서 2020년 3,319만 명으로 증가하였다(National Police Agency, 2022). 대한민국 통계청의 인구통계와 비교하였을 때, 운전면허 소지 비중은 54.6%에서 64.1%로 증가하였으며, 이러한 증가 추이는 일상생활에서 자동차의 역할이 증대되고 있음을 짐작하게 해준다. 자동차의 대중화가 가져오는 편리성의 이면에 여러 문제점이 발생하고 있으며, 그중에서도 최근 난폭·보복 운전과 관련된 범죄의 피해사례가 다양한 매체를 통해 보고되면서 사회적 문제로 제기되고 있다. 경찰청이 발표한 최근 난폭·보복운전 조사 현황에 따르면 2020년의 난폭, 보복운전 접수건은 각각 9,375건, 5,537건으로 2017년에 비해 1천여 건 이상씩 증가하였다(Yoo, 2022). 난폭·보복운전 발생 건수는 매년 증가하고 있지만, 해당 사건에 대한 기소율은 약 15%정도 감소한 것으로 나타났다. 이러한 현상은 범죄 신고자와 법적 처벌 규정 사이의 이해 차이에서 기인된 것으로 분석된다. 미국의 AAA Foundation에서 조사된 난폭·보복운전 피해 현황을 살펴보면, 최소 218건의 살인 사건이 난폭·보복운전이 발단이 되어 발생하였고, 부상자가 발생한 사건은 12,610건에 달하는 것을 확인하였다(AAA Foundation for Traffic Safety, 2022). 미국 또한 난폭·보복운전 발생 및 피해사례가 증가하고 있으며, 특히 충기가 동원된 보복운전은 2016년 616건에서 2021년 800여건 이상 발생하여, 범죄의 심각성이 증대되고 있다(Deliso, 2022). 이처럼 최근 난폭·보복운전 발생 건수와 피해사례 증가하고 있어, 이에 대한 대응 전략 마련이 요구된다. 난폭·보복운전을 판단하고 처벌할 수 있는 명확한 제도적 기준이 정의되어야 하며, 다각화된 통계 자료 수집을 통해 해당 범죄에 대한 보편적 분류 체계 마련과 잠재 특성 분석이 이루어져야 할 것이다. 나아가 난폭·보복운전 예방 및 계도를 위한 정책 마련을 통해 도로 위의 잠재 위험을 예방하는 사회적 가치 실현을 도모해야 할 것이다. 국내에서는 2016년에 도로교통법에 난폭운전 처벌 조항이 신설되었으며, 보복운전의 경우 형법으로 처벌되며 해당 재판 결과는 2017년부터 경찰청에서 관리되기 시작했지만, 범죄의

원인 및 실태 파악을 위한 학술연구, 자료 수집은 미흡한 실정이다. 본 연구에서는 LDA 알고리즘 기반의 토픽모델링을 통해 대한민국 재판부가 난폭·보복운전 사건에 대해 어떤 판결을 내려오고 있었는지를 분석하여 난폭·보복운전 특성 및 관련 판례연구의 시발점이 되고자 한다.

선행연구 검토

1. 난폭·보복운전에 관한 연구

난폭운전과 보복운전에 대한 국내 연구는 아쉽게도 활발하게 진행되고 있지 않은 편이다. 난폭·보복운전과 관련된 연구 활동이 미흡한 원인으로 난폭·보복운전과 관련된 경찰청이나 관계부처의 공식적 통계집계 활동이 2017년 도부터 이루어진 점, 수집된 일부 자료들이 경찰청 혹은 관계부처의 내부자료로만 운용되어 외부 연구자들이 쉽게 접근하지 못했던 점 등이 영향을 끼쳤을 것이다(Ryu, 2021). Lee(2017)는 경기도에서 발생한 262건의 보복운전 데이터를 경찰로부터 받아 그 특성을 분석하였다. 해당 연구에서 보복운전이 발생하는 장소, 시간, 피해자 특성, 범행행동의 패턴 등을 분석한 결과, 대부분 가해자는 3-40대 남성이었으며, 특히 전과기록이 존재하는 범행자들의 보복운전 실행 빈도수가 높았음을 분석함과 동시에 도로교통 범죄에 형사법이 확장되는 경향성에 관한 연구도 이루어졌다. Joo(2015)에 따르면, 보복운전은 도로교통법상의 안전운전 의무 위반죄가 적용되는 행위였었지만, 협박·상해·손괴 등의 행위로 보복운전 가해자들이 처벌되는 사례를 통해 도로교통 범죄에 형사법이 확장되어 적용되고 있다고 분석하였다. 이를 통해, 저자는 시대의 변화에 따라 형사법이 확장 적용되는 경향이 정당한지에 대한 검토의 필요성을 시사하였다. Choi et al.(2020)은 난폭·보복운전 예방을 위한 실태조사를 진행하였다. 경찰청의 협조를 통해 난폭·보복운전에 대한 내부 관리 자료를 수집하여, 한국에서 발생하는 난폭·보복운전의 실태를 파악하고 관련 법제도와 예방 정책에 대해 검토하였다. 난폭·보복운전의 발생 양상과 특징 및 관련 처벌 조항의 한계점에 대한 조사를 통해 향후 정책 수립에 기여할 수 있는 기초자료를 제시하였다. 저자는 블랙박스와 휴대폰, 통신 장비의 성능향상과 더불어 성숙해진 시민의식의 영향으로 난폭·보복운전의 위험성과 피해사태가 미디어를 통해 알려지면서 더 이상 외면할 수 없는 커다란 사회적 문제로 자리 잡게 되었다고 분석하였다.

2. 토픽모델링 적용 연구

Park et al.(2021)은 캘리포니아 차량 관리국(California Department of Motor Vehicles, CADMV)에서 수집된 자율주행 자동차 교통사고 분석 보고서를 토픽모델링 기법을 활용하여 분석하였다. 보고서의 사고 상세 설명(Accident Details Description) 항목을 분석하여 자율주행 자동차의 사고상황을 설명하는 4개의 토픽과 키워드들을 분석하였다. 분석 결과로 주행 중 후방 충돌, 정차 중 후방 충돌, 측면 충돌, 트럭 및 기타 충돌이 주요 사고내용으로 도출되었다. 해당 연구는 비정형 데이터에 대한 토픽모델링 기법의 적용 가능성을 확인하였으며, 그 방법을 시사하고 있다. Lee(2014)는 공동주택 하자판결 시스템에 대한 신뢰도 회복을 위해 법원 판례에 대한 키워드 빈도 분석, 네트워크 분석 등의 텍스트임베딩 기법을 적용하였다. 키워드분석 결과 균열 판정 기준인 0.3mm가 주요 키워드로 도출되었으며, 이를 통해 분석된 주요 내용으로는 0.3mm 미만의 균열에 대해 피고 측에서는 하자가 아니라고 주장하나 대부분의 법원에서는 하자를 인정해주었다는 점과 이러한 판결은 집합건물법, 주택법, 민법에 의해 판결됨을 확인하였다. Lee et al.(2021)은 44건의 승부조작 판례를 LDA알고리즘 기반의 토픽모델링 기법을 적용하여 분석하였다. 분석 결과로 도출된 토픽과 키워드들의 조합을 통해 ‘승부조작 경로’, ‘E-Sports와 관련된 승부조작’, ‘정통적 승부조작’, ‘승부조작 식별법’, ‘양형 판단 근거’가 판례에서 분석되는 주요 이슈임을 확인하였다.

방법론 검토

1. 토픽모델링

토픽모델링은 텍스트 마이닝의 한 종류로 비정형 문서 집합 내에서 잠재 토픽을 추출하는 데이터 분석 기법이다. 토픽모델링은 다량의 비정형 데이터를 주제에 따라 군집화하여 제공하며 이러한 측면은 문서 군집화(Document Clustering)와 유사하게 보이나, 한 개의 문서에 하나의 주제만 대응되는 경성 군집화(Hard Clustering)와 달리 한 개의 문서가 다양한 주제에 대응될 수 있는 연성 군집화(Soft Clustering)의 특성을 가진다. 이러한 성질은 방대한 양의 데이터에 대한 통찰(insight)과 결과를 활용한 다양한 분석을 수행할 수 있게 해줌과 동시에 현실 상황의 모델링에 적합하다는 평가를 받고 있다(Kim et al., 2017).

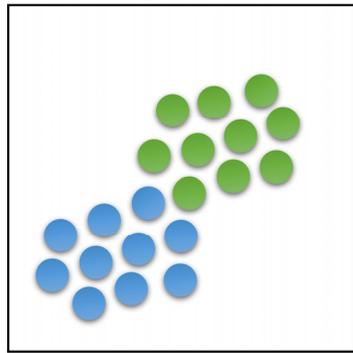


Figure 1. Illustrate of hard clustering

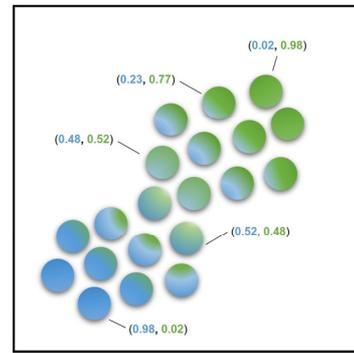


Figure 2. Illustrate of soft clustering

2. LDA 알고리즘

LDA 알고리즘(Latent Dirichlet Allocation algorithm)은 토픽모델링 분석을 진행할 때 가장 보편적으로 사용하는 확률 기반 생성 모델 알고리즘이다(Blei, 2012). LDA 알고리즘은 문헌 내 포함된 텍스트 등의 관측된 변수(Observed variable)를 통해 관측되지 않는 문헌의 주제, 구조 등의 잠재 변수(Hidden Variable)를 추론하여, 전체 문서 집합의 토픽, 문서별로 할당된 토픽 비율, 토픽 내 포함된 키워드 분포 결과를 제공한다(Kim et al., 2016).

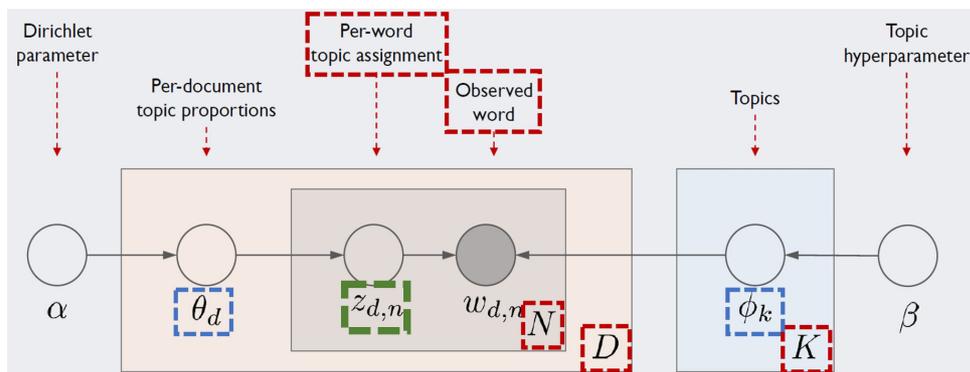


Figure 3. Graphical model of LDA algorithm

Figure 3은 LDA 알고리즘을 구성하는 변수와 작동 메커니즘을 설명하는 그래프 모형이며, 문서 내 토픽 할당 확률 계산식은 Equation 1과 같다.

$$P(z_{d,i} = j \mid z_{-i}, w) = \frac{n_{d,k} + \alpha_j}{\sum_{i=1}^K} \times \frac{v_{k,w_{d,n}} + \beta_{w_{d,n}}}{\sum_{j=1}^V (v_{k,j} + \beta_j)} = AB \quad (1)$$

Equation 1에서 사용된 각 변수값은 아래와 같다.

- $n_{d,k}$: k 번째 토픽에 할당된 d 번째 문서의 단어 빈도
- $v_{k,w_{d,n}}$: 전체 말뭉치에서 k 번째 토픽에 할당된 단어($w_{d,n}$)의 빈도
- $w_{d,n}$: d 번째 문서에 n 번째로 등장한 단어
- α : 문서의 토픽 분포 생성을 위한 디리클레 분포 파라미터
- β : 토픽의 단어 분포 생성을 위한 디리클레 분포 파라미터
- K : 분석자가 지정하는 토픽 수
- V : 문서 집합 내에 포함된 전체 단어 수
- A : d 번째 문서가 k 번째 토픽과 맺고 있는 연관성 정도
- B : d 번째 문서의 n 번째 단어($w_{d,n}$)가 k 번째 토픽과 맺고 있는 연관성 정도

3. LDAvis

LDA 알고리즘은 차원 축소 용이성과 주제 생성의 일관성이 뛰어나며(Mimno et al., 2008), 이러한 특성 덕분에 학계에서 표준으로 사용되고 있다(Kim and Jang, 2016). 2003년 Blei에 의해 LDA 알고리즘 개발된 이후부터 다양한 분야에 걸쳐 이를 활용한 연구 및 분석이 진행되었으며, 결과 도출의 명확성과 해석의 용이성을 위한 시각화 도구의 개발도 활발하게 진행되었다. 초기에는 가중치가 높은 키워드 리스트를 나타내는 막대 차트, Word Clouds, 파이 차트 등이 등장했으나 이러한 방식은 개별 토픽과 키워드를 분석할 수 있지만, 전체적인 토픽모델링 결과를 해석할 수 없었다. 2012년 Chuang et al.(2012)이 토픽모델링의 결과를 Matrix 형식으로 시각화하는 도구를 개발하였고, 전체 토픽과 키워드 사이의 관계를 파악할 수 있지만, 등장 빈도수가 높은 키워드만을 나타내는 한계점이 존재하였다(Chuang et al., 2012). 이후 2014년 Sievert, Shirley에 의해 개발된 토픽모델링 시각화 도구인 LDAvis는 차원 축소 방법인 PAC(Principal Component Analysis)와 키워드 추출법을 활용하여 토픽과 키워드의 상호 관계를 손쉽게 파악할 수 있도록 하였으며, 인터페이스 조작을 통해 토픽 내 키워드 분포 및 토픽 간 유사성 등을 분석할 수 있다.

Figure 4는 LDAvis를 활용해 표출한 토픽모델링 분석 결과 화면이다. LDAvis는 좌측의 Intertopic Distance Map(via multidimensional scaling)과 우측의 Top-30 Most Relevant Terms for Topic을 통해 분석자에게 토픽모델링 결과를 시각화하여 제공한다. 기존의 토픽은 키워드 개수만큼의 차원을 가지고 있으며, Intertopic Distance Map에서는 Principal Component Analysis(PCA)를 통해 n 차원의 벡터를 2차원으로 압축하는 차원 축소법을 이용하고 있다. Intertopic Distance Map 내에서 원의 크기는 토픽에 단어가 얼마나 속해있고 어떻게 분포되어있는지를 나타내며, 원 사이의 거리가 가까울수록 토픽 간 유사성이 높음을 의미한다. 우측의 Top-30 Most Relevant Terms for Topic 영역에서는 토픽을 형성하는 주요 키워드와 토픽과 키워드의 관계를 도출하고 있다. 해당 영역에서 표출되는 정보는 현저성(salience)과 분별력(discriminative power)을 기준으로 제시되며, 문서 집합 내 등장 비율, 명확한 토픽 지시(차별성)를 모두 고려하여 키워드를 선택하게 된다. 이때 현저성과 분별력 사이에는 부적 상관(Negative correlation)이 존재하는데, 분석자는 λ 값의 조절을 통해 두 기준의 중요성을 조절할 수 있으며, λ 값이 0에 가까울수록 분별력을, 1에 가까울수록 현저성의 중요도를 우선시하게 된다.

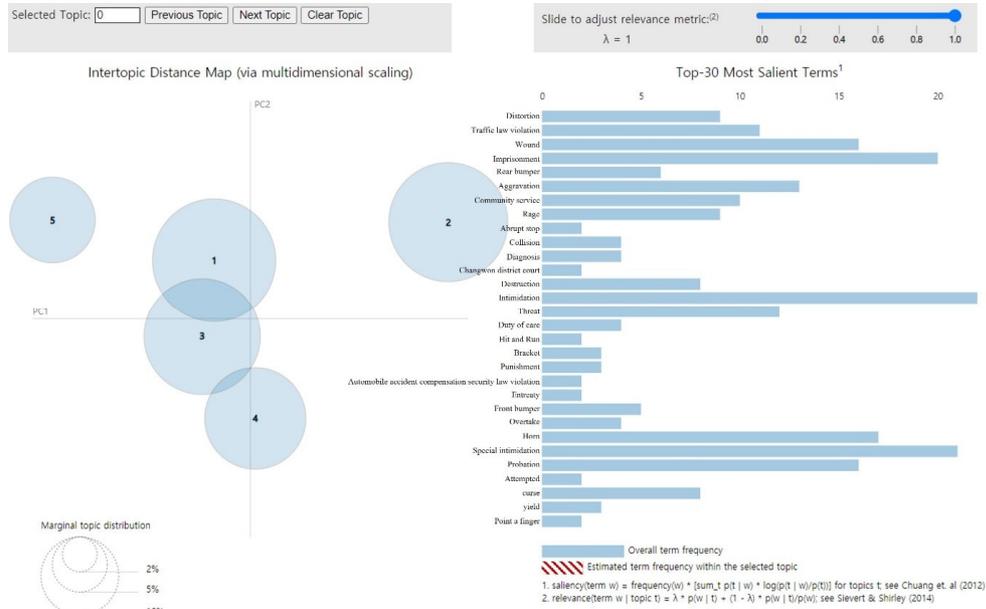


Figure 4. Interface of LDAvis

판례분석 절차

본 연구에서는 비정형의 난폭·보복운전 관련 판례를 토픽모델링을 활용하여 분석하기 위한 방법론을 개발하였다. Figure 5는 분석 절차와 방법 및 해당 내용을 나타낸다.

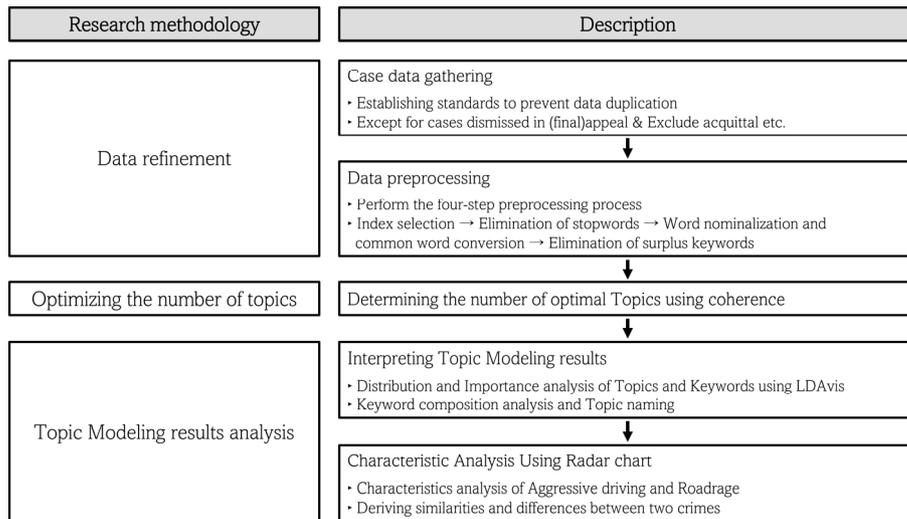


Figure 5. Case analysis process

1. 데이터 정제

본 연구에서 분석한 데이터는 범죄 사실 내에 난폭운전 혹은 보복 운전을 포함하는 판례로부터 수집되었다. 형사 사건에 대한 판례만을 다루었으며, 데이터의 중복을 방지하고, 정확한 분석 결과를 도출하기 위해 무죄 판결 제외, 고등 법원과 대법원에서 기각된 판례 제외, 하급심과 항소심이 중복되거나 병합사건의 경우 1건의 사건으로 취급

등의 분류 기준을 마련하여 데이터를 수집하였다. 이 과정에서 난폭운전에 관한 판례는 23건, 보복운전에 관한 판례는 31건이 수집되었다. Table 1은 수집된 난폭·보복운전 판례 데이터 개요를, Table 2는 판례 데이터를 구성하는 주요 항목이다.

Table 1. Case data information

	Contents
Target	Cases involving road rage in criminal facts
Trial	First trial (district court)
Period	From 2014 to 2020
Cases	23 aggressive cases · 31 road rage cases

Table 2. Case components

	Contents
Outline	Court name & date · Case number & type
Reason	Criminal fact · Evidence · Legislation
Summary	The gist of the defendant's and defense's claims
Judgment	Judges order & judgment · Reason of punishment

수집된 비정형 데이터를 분석하는 데 있어 결과의 정확성과 명확성을 향상하기 위해 전처리 작업을 수행하였으며, 이는 4단계의 과정을 통해 진행되었다. 1단계는 판례 내의 항목 중 의미 있는 분석 결과를 도출할 수 있는 구성 요소들에 대해 분류했으며, ‘법원명’, ‘주문내용’, ‘범죄 사실’, ‘증거의 요지’, ‘법령의 적용’에 해당하는 내용에 대해 중점적으로 분류하는 작업을 진행하였다. 2단계는 수집된 텍스트 내 문법적 기능만을 담당하며 단어적 의미가 없는 조사, 어간, 어미 등의 불용어를 제거하였다. 3단계는 1, 2단계에서 정제된 키워드들의 중복을 방지하고 분석 결과에서 일관성을 나타내기 위해 단어의 명사화와 유사 단어를 공통단어로 변환하는 작업을 진행하였다. 마지막으로 4단계에서는 선행 단계에서 도출된 단어 중 결과 해석에 영향을 끼치지 않으나, 등장 빈도수가 많아 결과 해석에 혼란을 줄 수 있는 단어를 제거하였다. 4단계에 걸친 데이터 전처리 방법과 단계별 예시는 Table 3과 같다.

Table 3. Preprocessing sequence and example

	Procedure	Description
First	Index selection	Among the precedent, only the contents corresponding to the application of the law are extracted.
Second	Elimination of stopwords	Eliminating stopwords such as postpositions, stems, endings, and connecting words.
Third	Word nominalize and common word conversion	For consistency of results, nounization of verbs and conversion similar words into common words.
Fourth	Elimination of surplus keywords	Remove frequently appearing words that do not affect interpretation of results.

2. 최적 토픽 수 결정

토픽모델링 분석을 통해 도출된 각각의 토픽들이 유의미한 결과값을 가져 독립성을 확보할 수 있도록 분석자는 적정 토픽 개수를 설정해야 한다. 적정 토픽 개수 산정은 Perplexity 또는 Coherence를 통해 평가하게 되며, Perplexity가 갖는 결과 해석에서의 한계점을 고려하여 본 연구에서는 Coherence를 통해 최적 토픽 수를 선정하였다. 토픽모델링 분석 결과값이 우수할수록 한 토픽 내 의미론적으로 유사한 키워드가 밀집되게 된다. 즉 토픽 내 상위 키워드 간의 유사도 계산을 통해 해당 토픽모델링 분석 결과값을 평가할 수 있으며, 이때 Coherence를 평가 지표로 사용하게 된다.

3. 토픽모델링 결과 해석

전처리 과정을 거쳐 수집된 난폭·보복운전 관련 판례를 통해 난폭·보복운전의 유형과 주요 특성을 도출하기 위해 토픽모델링을 수행하였다. 모델링 알고리즘으로는 LDA를 채택하였으며, 해당 알고리즘은 주어진 문서 집합에 대해 어떤 토픽들이 존재하는지에 대한 확률 모형을 키워드 기반으로 제시함으로써, 분류 성공률이 높고 해석이 용이하다는 장점을 갖는다. LDA 기반의 토픽모델링 분석을 구현하기 위해 머신러닝 및 빅데이터 분석용 프로그램 언어인 Python과 텍스트 임베딩을 위한 라이브러리 Soynlp, 최적 키워드 분석 및 토픽모델링 처리를 위한 라이브러리 Gensim, 결과값의 시각화를 위한 라이브러리 LDavis를 이용하였다. 토픽모델링 분석 결과를 바탕으로, 난폭운전과 보복운전의 주요 특성을 도출하였다. 난폭·보복운전에서 도출된 토픽과 키워드의 조합을 통해 각각의 토픽을 명명하고 주요 상황을 기술하였으며, 이후 난폭·보복운전에서 발견되는 공통특성에 대해 Radar Chart를 이용한 비교분석을 진행함으로써 난폭운전과 보복운전 범죄의 개별성과 주안점을 도출하였다.

판례분석 결과

1. 난폭운전 판례분석 결과

난폭운전 판례에 대한 토픽모델링 분석을 진행하기 위해 Coherence를 지표로 최적 토픽 수를 산출하였을 때 4개의 토픽이 최적 토픽 개수로 분석되었다. Topic 1을 구성하는 키워드에는 교통상 위험, 후진, 반복, 방해, 중앙선, 안전거리, 신호 위반, 앞지르기 등이 포함되었다. 해당 토픽에서 난폭운전 가해자의 주요 범죄 행동은 후진, 안전거리 미확보, 중앙선 침범, 신호 위반으로 도출되었으며, 해당 범죄 행위를 반복하면서 상대 차량에 위협을 가하며 교통상 위험을 발생시키는 상황을 유발하는 것으로 분석되었다. Topic 2를 구성하는 키워드에는 승용차, 손괴, 상해, 염좌, 치료, 충돌, 위협, 수리 등이 포함되었다. 해당 토픽에서 승용차와 택시가 주요 피해 대상으로 도출되었으며, 피해자들은 충돌에 의한 범퍼 파손 등의 차량 손상과 염좌 등의 상해를 입는 것으로 분석되었다. Topic 3을 구성하는 키워드에는 도주, 위반, 교통상 위험, 무면허 운전, 경찰, 회피, 유턴, 침범 등이 포함되었다. 해당 토픽에서 가해자의 범죄 양상은 무면허로 주행하다 경찰의 단속을 회피하기 위해 도주하는 과정에서 난폭운전을 행하였으며, 불법 유턴, 중앙선 침범 등의 범법 행위로 교통상 위험을 발생시키는 것으로 분석되었다. Topic 4를 구성하는 키워드에는 도주, 음주운전, 중앙선 침범, 유턴, 단속, 도로교통법 위반, 후진, 신호위반 등이 포함되었다. 해당 토픽에서 가해자의 범죄 양상은 음주운전 단속을 회피하기 위해 도주하는 과정에서 난폭운전을 행하였으며, 불법 유턴, 중앙선 침범, 후진, 신호 위반 등의 범법 행위로 교통상 위험을 발생시키는 것으로 분석되었다. 토픽모델링을 통해 분석된 4개의 토픽과 하위 키워드의 조합으로 토픽을 명명하였으며, 토픽별 난폭운전 특성을 나타내는 주요 상황을 Table 4에 기술하였다.

Table 4. Topic analysis results of aggressive driving cases

Topic	Topic naming	Major information
Topic 1	Behavior pattern	<ul style="list-style-type: none"> Repeat activities such as reversing, not securing a safe distance, violating the center line and signal Threatening the other vehicle and creating traffic hazards
Topic 2	Damage type	<ul style="list-style-type: none"> Cars and taxis on the road are affected The perpetrator causes damage to the vehicle or inflict an injury
Topic 3	Unlicensed driving	<ul style="list-style-type: none"> Runaway to evade police crackdown on unlicensed driving Traffic hazards are caused by reckless driving in runaway course
Topic 4	Drunk driving	<ul style="list-style-type: none"> Runaway to evade police crackdown on drunk driving Traffic hazards are caused by reckless driving in runaway course

2. 보복운전 판례분석 결과

보복운전 판례에 대한 토픽모델링 분석을 진행하기 위해 Coherence를 지표로 최적 토픽 수를 산출하였을 때 4개의 토픽이 최적 토픽 개수로 분석되었다. Topic 1을 구성하는 키워드에는 화남, 끼어들기, 택시, 진로 변경, 상향등, 경적, 여성 말다툼 등이 포함되었다. 해당 토픽에서 보복운전의 범행 결정요인은 여성 운전자의 서행, 택시의 끼어들기, 다른 차량의 진로 변경 및 경적과 상향등 사용으로 도출되었으며, 이러한 요인들에 의해 가해자 화를 내며 보복운전을 결심한 것으로 분석되었다. Topic 2를 구성하는 키워드에는 차선변경, 끼어들기, 급제동, 보복운전, 추월, 침범, 앞지르기 등이 포함되었다. 해당 토픽에서 보복운전 가해자의 주요 범죄 행동은 방향지시등 미사용, 차선변경, 갑작스러운 끼어들기와 앞지르기, 차선 침범, 급제동, 추격 등으로 도출되었으며, 해당 범죄 행위를 지속·반복적으로 행하면서 보복운전을 실현하는 것으로 분석되었다. Topic 3을 구성하는 키워드에는 상해, 염좌, 치료, 충돌, 긴장, 욕설, 손괴, 경추 등이 포함되었다. 해당 토픽에서 피해자의 피해 내용으로는 자동차의 파손, 염좌, 긴장 등의 상해가 존재했으며, 보복운전 이후 정차 상황에서 욕설과 폭행으로 피해자들을 위협한 추가적인 범행 사실이 분석되었다. Topic 4를 구성하는 키워드에는 특수협박, 징역형, 상해, 도로교통법 위반, 벌금형, 사회봉사, 일반 교통방해 치상, 폭력행위 등 처벌에 관한 법률 위반 등이 포함되었다. 해당 토픽에서 판결에 의한 가해자의 죄목은 특수협박, 도로교통법 위반, 일반 교통방해 치상, 폭력행위 등 처벌에 관한 법률 위반 등이며, 처벌은 사회봉사와 징역형을 선고받은 것으로 분석되었다. 토픽모델링을 통해 분석된 4개의 토픽과 하위 키워드의 조합으로 토픽을 명명하였으며, 토픽별 보복운전 특성을 나타내는 주요 상황을 Table 5에 기술하였다.

Table 5. Topic analysis results of road rage cases

Topic naming		Key implications
Topic 1	Determinants of crime	<ul style="list-style-type: none"> The cause of the crime is the slow movement of a female driver, the intervention of a taxi, the change of course of another vehicle, and the use of horns and high beam. The perpetrator decides to road rage based on the cause of the crime.
Topic 2	Behavior pattern	<ul style="list-style-type: none"> Roadrage methods are not used, interruption and overtaking, lane violation, sudden braking. The perpetrator continues and repeats the act to realize road rage.
Topic 3	Damage type	<ul style="list-style-type: none"> The type of damage to the victims is injury such as sprains, tension and car damage. The perpetrator threatens the victims through additional criminal acts, such as abusive language and assault, while stopping after road rage.
Topic 4	Crime facts and contents of punishment	<ul style="list-style-type: none"> The crime of the perpetrator by judgment is special intimidation, violation of the road traffic act, general traffic obstruction. The level of punishment for the perpetrator by judgment is from social service to imprisonment.

3. 난폭·보복운전 특성별 비교분석 결과

토픽모델링을 활용하여 난폭운전과 보복운전 판례의 잠재 토픽 및 하위 키워드들을 도출하였으며, 이들을 조합하여 토픽의 명명과 주요 상황을 제시하였다. 선행 분석 결과를 통해 각각의 범죄 행위의 형태와 특성을 관측하였으며, 이를 바탕으로 추가적인 분석을 위해 난폭운전과 보복운전의 비교 가능한 특성을 추출하였다. 비교 가능한 특성의 영역은 행동 양상, 피해 유형, 범죄 사실과 판결, 범행 대상과 장소이며, 두 범죄 행위의 특징 및 유사점과 차이점을 명확하게 파악하기 위해 Radar Chart를 이용한 시각화를 진행하였다. Figure 6에서는 난폭운전과 보복운전의 행동 양상에 대한 비교분석을 수행하였다. 난폭운전에서는 신호 위반, 후진, 유턴, 과속, 차선침범 등 경찰의 단속을 회피하여 도주하기 위한 주행 행동이 약 66%의 비율로 관측되었다. 보복운전에서는 급제동, 추격, 차선 침범 등 특정 운전자를 위협하기 위한 주행 행동이 약 93%의 비율로 관측되었다. Figure 7에서는 난폭운전과 보복운전으로 인해 발생한 피해 유형에 대한 비교분석을 수행하였다. 난폭운전의 경우, 가해자의 도주 과정에서 불특정 다수의 피해

자가 발생하는 특성에 의해 염좌, 긴장, 타박상 등의 경상 발생이 약 89%의 비율로 관측되었다. 보복운전의 경우, 가해 대상이 명확한 특성에 의해 경추, 요추 손상과 골절 등의 중상 발생이 약 59%의 비율로 관측되었다. 범죄 특성의 차이에 의해 보복운전 피해자의 피해 정도가 난폭운전 피해자에 비해 치명적인 것으로 나타났다.

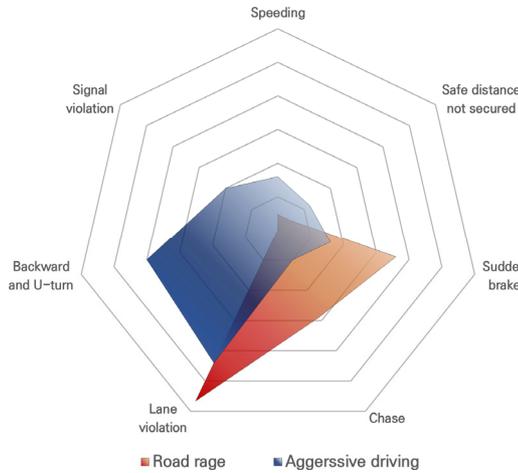


Figure 6. Behavior pattern comparison

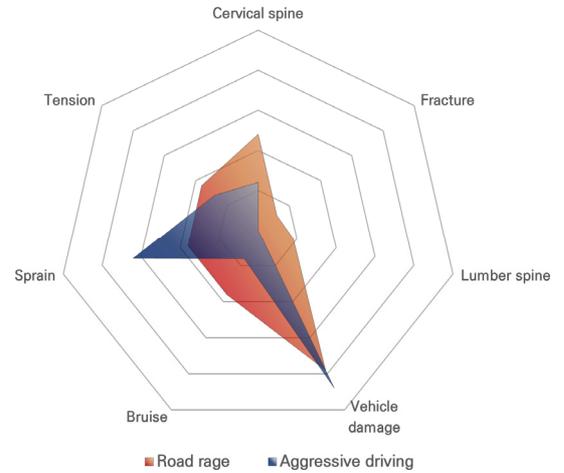


Figure 7. Damage type comparison

Figure 8에서는 난폭운전과 보복운전의 범죄 사실과 판결에 대한 비교분석을 수행하였다. 난폭운전에서는 무면허 운전, 음주운전 등 다른 범죄 행위가 수반된 결합범의 형태가 약 46%의 비율로 관측되었다. 보복운전에서는 협박, 폭행, 손괴 등의 범죄 행위가 약 87%의 비율로 관측되었다. 난폭운전에 비해 보복운전의 구형 정도가 더 중하였으며, 난폭운전은 도로교통법에 따라 처벌되지만 보복운전은 형법에 따라 처벌되기 때문에 이러한 양상을 나타내는 것으로 분석된다. Figure 9에서는 난폭운전과 보복운전의 범행 대상과 장소에 대한 비교분석을 수행하였다. 난폭운전의 피해 대상으로는 순찰차와 택시가 약 86%의 비율로 관측되었으며, 이는 가해 차량의 도주를 저지하는 과정에서 운전 숙련도가 높은 택시가 경찰차를 지원하면서 나타난 결과로 분석된다. 보복운전의 피해 대상으로는 택시, 버스, 화물차가 약 81%의 비율로 관측되었으며, 고속도로에서의 범행 비율이 높은 것으로 관측되었다. 일반 차량에 비해 다소 거친 택시와 화물차의 주행 행동이 가해자의 보복운전 유발요인으로 작용했을 것이라 분석된다.

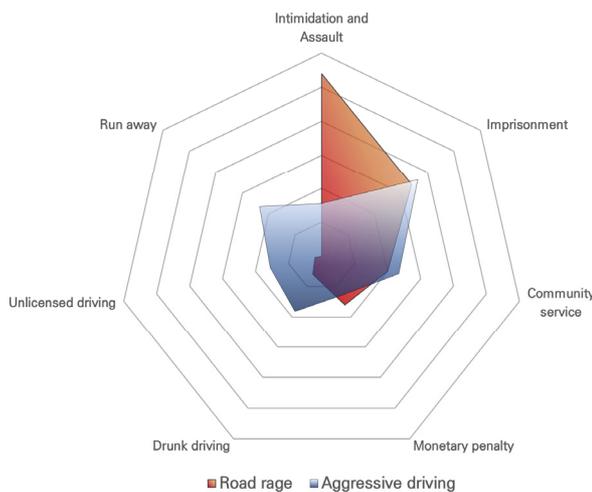


Figure 8. Criminal facts and rulings comparison

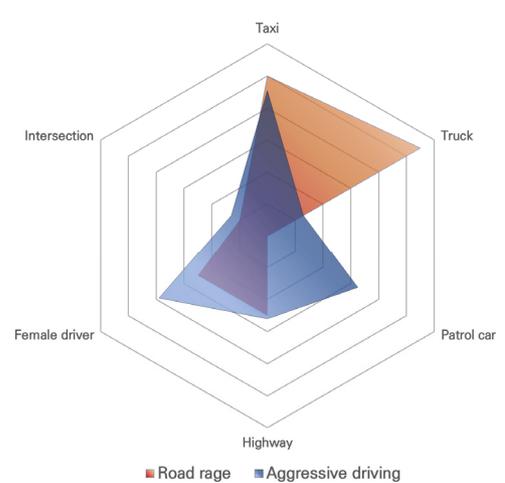


Figure 9. Crime subject and place comparison

Table 6에서는 Figure 6에서 Figure 9까지 사용된 난폭운전·보복운전 특성의 구성 비율을 나타낸다.

Table 6. The composition ratio of aggressive driving and roadrage characteristics

Figure	Figure naming	Composition of aggressive driving characteristics (%)				Composition of roadrage characteristics (%)		
Figure 6	Behavior pattern comparison	Speeding	Safety distance not secured	Sudden brake	Chase	Lane violation	Backward and U-turn	Signal violation
		3.7 / 10.1	3.7 / 7.6	27.5 / 10.1	22.0 / 6.3	43.2 / 27.8	0 / 25.3	0 / 12.7
Figure 7	Damage type comparison	Cervical spine	Fracture	Luber spine	Vehicle damage	Bruise	Sprain	Tension
		18.2 / 10.9	4.5 / 0	6.8 / 0	29.5 / 40	13.6 / 7.3	13.6 / 29.1	13.6 / 12.7
Figure 8	Criminal facts and rulings comparison	Intimidation and assault	Imprisonment	Community service	Monetary penalty	Drunk driving	Unlicensed driving	Run away
		40.9 / 10.7	25.8 / 25	15.2 / 16.1	12.1 / 8.9	4.5 / 12.5	1.5 / 10.7	0 / 16.1
Figure 9	Crime subject and place comparison	Taxi	Truck	Patrol car	Highway	Female driver	Intersection	
		30.3 / 26.9	33.3 / 7.7	0 / 19.2	15.2 / 15.4	15.2 / 23.1	6.1 / 7.7	

결론

본 연구에서는 형사사건으로 기소된 난폭·보복운전 판례를 토픽모델링을 활용해 분석하였다. 우선으로 비정형의 판례 데이터를 분석 가능한 형태로 변환하는 전처리 작업을 수행하였으며, 정제된 판례 데이터를 토픽모델링을 활용하여 잠재 토픽과 키워드를 도출하였다. 이후 도출된 토픽과 키워드의 조합으로 토픽의 명명, 주요 상황의 도출, 난폭운전과 보복운전의 범죄 형태와 특성을 도출하였으며, 난폭운전과 보복운전의 특성을 비교분석하고 Radar Chart를 이용해 시각화하였다. 일련의 분석 과정을 통해 난폭운전과 보복운전은 범행 요인, 행동 양상, 범행 대상과 장소, 피해 유형, 범죄 사실, 판결의 영역에서 구별되는 특성을 가진 것으로 분석되었다. 난폭운전 판례분석에서 도출된 주요 내용은 가해자가 후진, 불법 유턴, 신호 위반, 과속 등 도주 상황에서의 행동 특성을 보였으며, 이러한 과정에서 불특정 다수에게 경미한 피해를 입히는 비율이 높은 것으로 분석되었다. 보복운전 판례분석에서 도출된 주요 내용은 가해자가 갑작스런 끼어들기와 앞지르기, 차선 침범, 급제동, 지속적인 추격 등 피해자를 위협하려는 행동 특성을 보였으며, 이러한 과정에서 특정 피해자에게 중한 피해를 가하는 비율이 높은 것으로 분석되었다. 또한 범죄 사실과 판결에서 난폭운전은 음주운전, 무면허 운전의 범법 행위가 동반된 결합범의 형태가, 보복운전은 협박, 폭행 등 형법에 따라 처벌되는 범행의 비율이 높은 것으로 확인되었다. 일련의 분석 결과를 통해 판례 내에 잠재되어 있던 난폭운전과 보복운전의 범죄 특성을 도출하였으며, 난폭운전과 보복운전은 개별적 특성이 뚜렷한 서로 다른 범죄임을 확인하였다.

본 연구에서는 형사사건으로 기소된 난폭·보복운전 판례를 토픽모델링을 활용해 분석하였으나, 몇 가지 한계점이 존재한다. 우선, 형사사건으로 기소된 난폭·보복운전 판례 수가 부족하여 토픽모델링 분석에서 추가적인 정보도 도출되지 못한 점이다. 이는 난폭·보복운전 관련 범죄 기소율이 다른 범죄에 비해 약 20% 정도 낮으며, 사건 대부분이 약식으로 처리되어 수집 가능한 판례에 한계가 존재하였기 때문이다. 추후, 난폭·보복운전 판례 관련 데이터가 충분히 확보된다면, 추가적인 정보 도출이 가능할 것으로 판단된다. 다음으로 판례 구성요소인 주문내용, 판단, 양형의 이유에서의 분석이 용이하지 않은 한계가 있다. 보복운전의 경우 난폭운전과 다르게 관련 법이 제정되지 않아 형법에 따라 처벌되고 있으며, 판사의 판단 혹은 구형되는 양형의 수준 또한 난폭·보복운전 범죄 행위에 대해서만 이뤄지는 것이 아니라, 가해자의 전과 여부, 추가 범행 여부, 피해자의 피해정도 등을 복합적으로 고려하기 때문에 유사한 난폭·보복운전 범죄에 대해서도 처벌의 내용이나 수준에서 차이가 발생하였기 때문이다. 아직까지 국내에서는 난폭·보복운전 현황에 대한 통계가 집계되고 관리되지 않고 있으며, 이 연구자들이 관련 정보를 수집하기 쉽지

않아 난폭·보복운전 관련 실증연구 또한 활발하게 진행되지 않는 실정이다. 본 연구에서는 한정적인 데이터 자원을 최대한 활용하여 토픽모델링을 이용한 난폭·보복운전 판례 분석을 통해 난폭·보복운전 형태 및 특성을 도출하였다. 연구 결과를 바탕으로 향후 난폭·보복운전 관련 법률 제정, 법률적 판단 기준을 마련하는 기초자료 제작과 함께 이를 활용한 난폭·보복운전 단속을 위한 경찰 가이드라인, 범죄 예방 시스템 및 인프라 개발 등에 활용할 수 있을 것으로 기대한다.

Funding

This work was supported by Kyonggi University Research Grant 2021.

알림

본 논문은 대한교통학회 제86회 학술발표회(2022.04.22)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

ORCID

KIM, Changhun  <http://orcid.org/0000-0003-0676-1292>

LEE, Suejung  <http://orcid.org/0000-0002-4717-4871>

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

References

- AAA Foundation for Traffic Safety, Road Rage: How to Avoid Aggressive Driving, <https://exchange.aaa.com/wp-content/uploads/2013/06/Road-Rage-Brochure.pdf>, 2022.03.18.
- Blei D. M. (2012), Provarilistic topic models, *Communications of the ACM*, 55(4), 77-84.
- 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, 74-77.
- Deliso M., “Disconcerting” Rise in Road Rage Shootings Resulting in Death or Injury, Data Shows, ABC News, <https://abcnews.go.com/US/disconcerting-rise-road-rage-shootings-resulting-death-injury/story?id=78181165>, 2022.03.18.
- Joo H. K. (2015), The Expansion of Special Criminal Laws in the Field of Road Traffic, *Journal of Criminal Law*, 27(4): 65, 173-202.
- Kim N. G., Lee D. H., Choi H. C., Wong W. X. S. (2017), Investigations on Techniques and Applications of Text Analytics, *The Journal of Korean Institute of Communications and Information Sciences*, 42(2), 471-492.
- Kim S. K., Jang S. Y. (2016), A Study on the Research Trends in Domestic Industrial and Management Engineering using Topic Modeling, *Journal of the Korea Management Engineers Society*, 21(3), 71-95.
- Kim T. G., Choi H. R., Lee H. C. (2016), A Study of the Research trends in fintech using topic modeling, *Journal of the Korea Academia-Industrial Cooperation Society*, 17(11), 670-681.
- Lee J. (2017), A Study on Road Rage: Focusing on Kyonggi Province, Master’s Thesis, Kyonggi University.

- 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, *Korean Society for Measurement and Evaluation in Physical Education and Sports Science*, 23(2), 51-65.
- Lee K. (2014), Study on Improvement of Apartment Housing Defect Judgment System Based in Judicial Precedent Cases, Doctoral dissertation, Hanyang University.
- Mimno D., Wallach H., McCallum A. (2008), Gibbs Sampling for Logistic Normal Topic Models with Graph-based Priors, In *NIPS Workshop on Analyzing Graphs*, 1-8.
- Ministry of Land, Infrastructure and Transport, Vehicle Registration Status, Statistics Korea, https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1257, 2022.03.04.
- National Police Agency, Driver's License Holders Status, Statistics Korea, [https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1617#quick_02](https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1617#quick_02;);, 2022.03.04.
- Park S. M., Lee H. P., So J. H., Yun I. S. (2021), Study of Analysis for Autonomous Vehicle Collision Using Text Embedding, *The Journal of The Korea Institute of Intelligent Transport Systems*, 20(1), 160-173.
- Ryu H. (2021), A Study on Road Rage: Focusing on Kyonggi Province, Master's Thesis, Kyonggi University.
- Yoo M., (Last year 2020), 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>, 2022.03.15.

Appendix A

Topic modeling result of aggressive driving

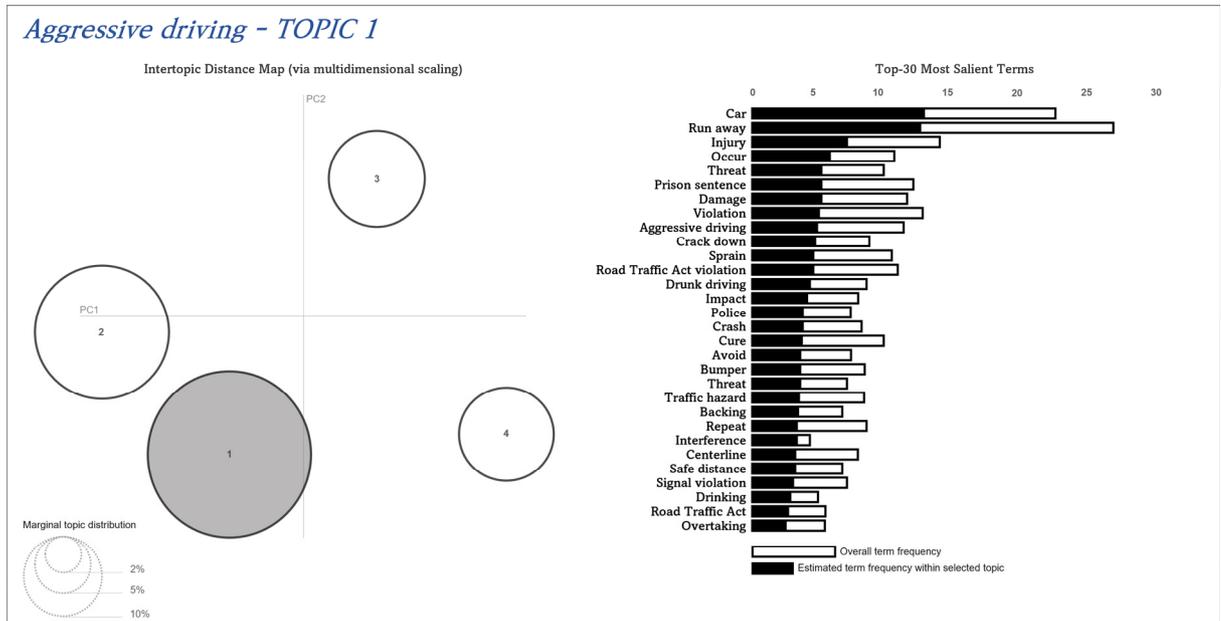


Figure 10. Topic modeling result of aggressive driving - Topic 1

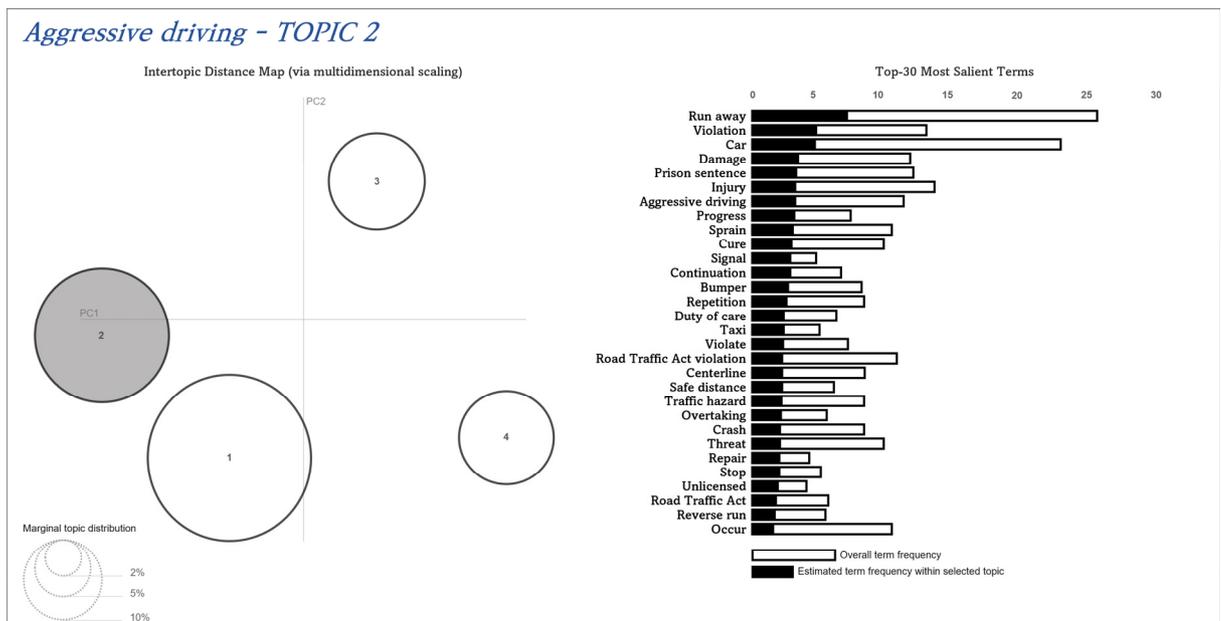


Figure 11. Topic modeling result of aggressive driving - Topic 2

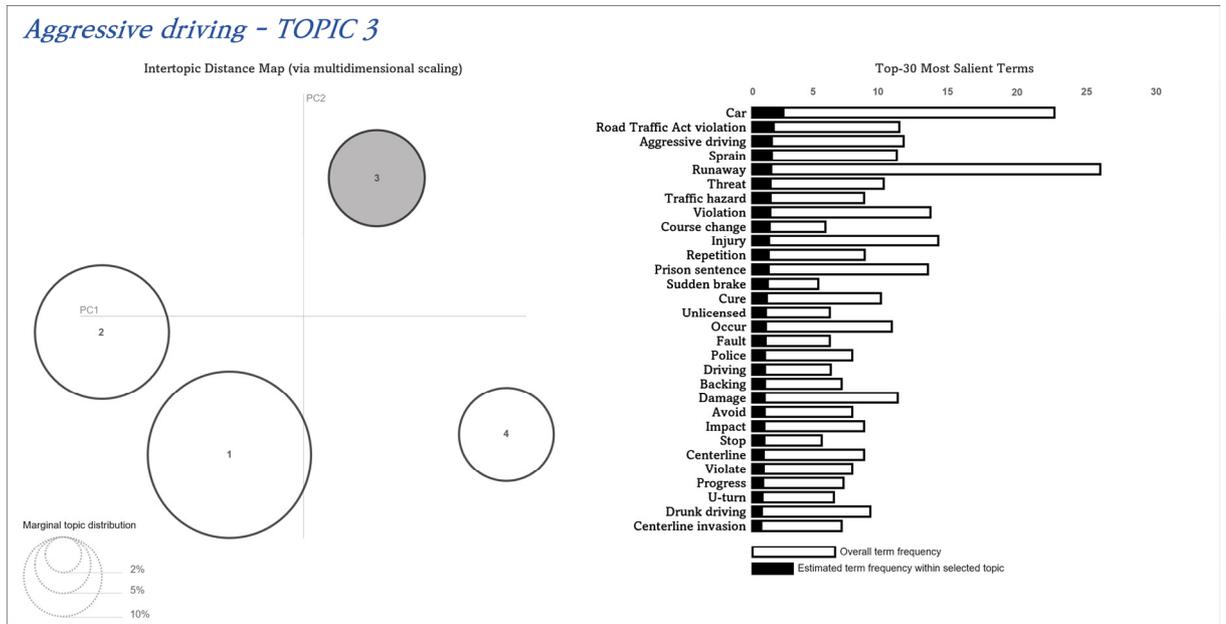


Figure 12. Topic modeling result of aggressive driving - Topic 3

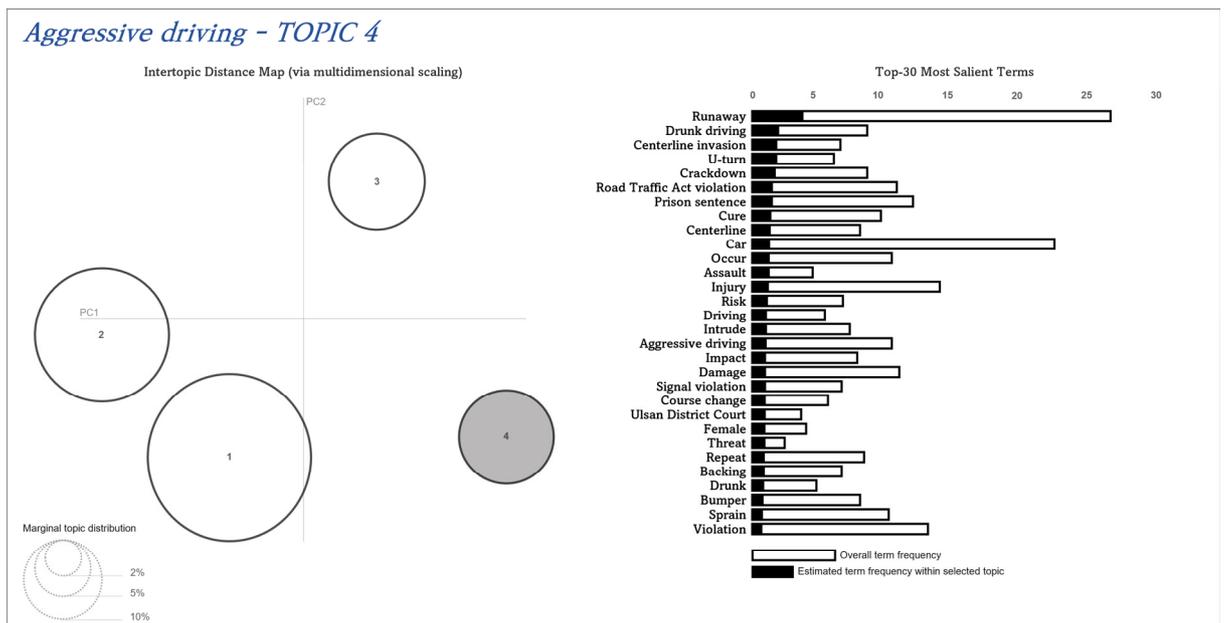


Figure 13. Topic modeling result of aggressive driving - Topic 4

Appendix B

Topic modeling result of road rage

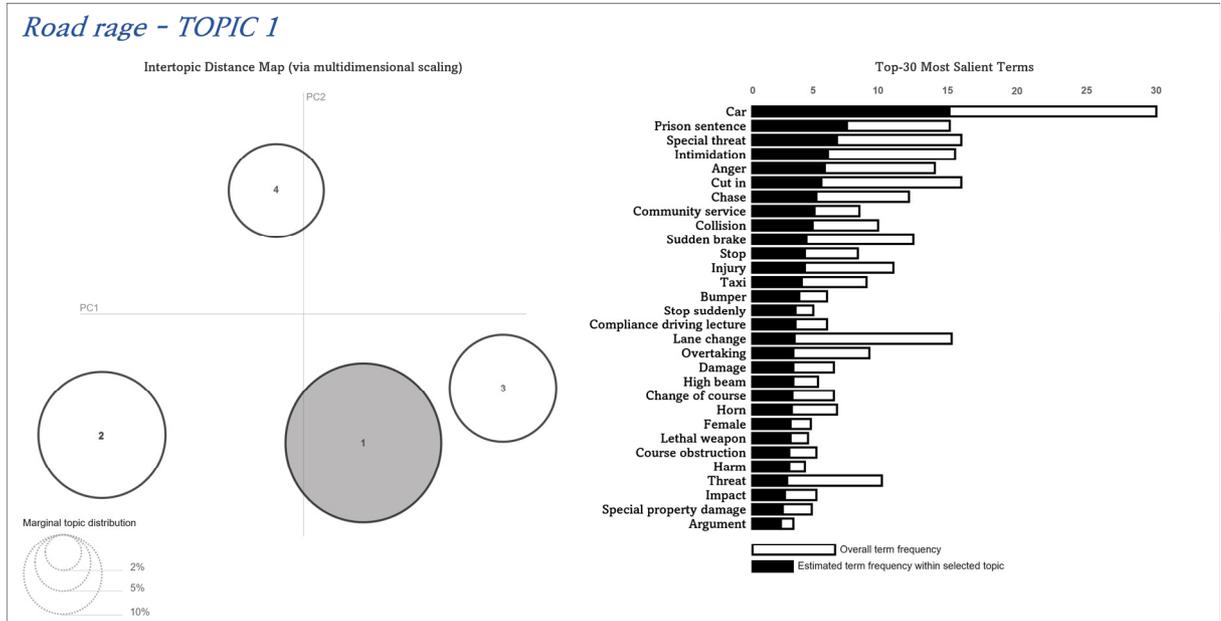


Figure 14. Topic modeling result of road rage - Topic 1

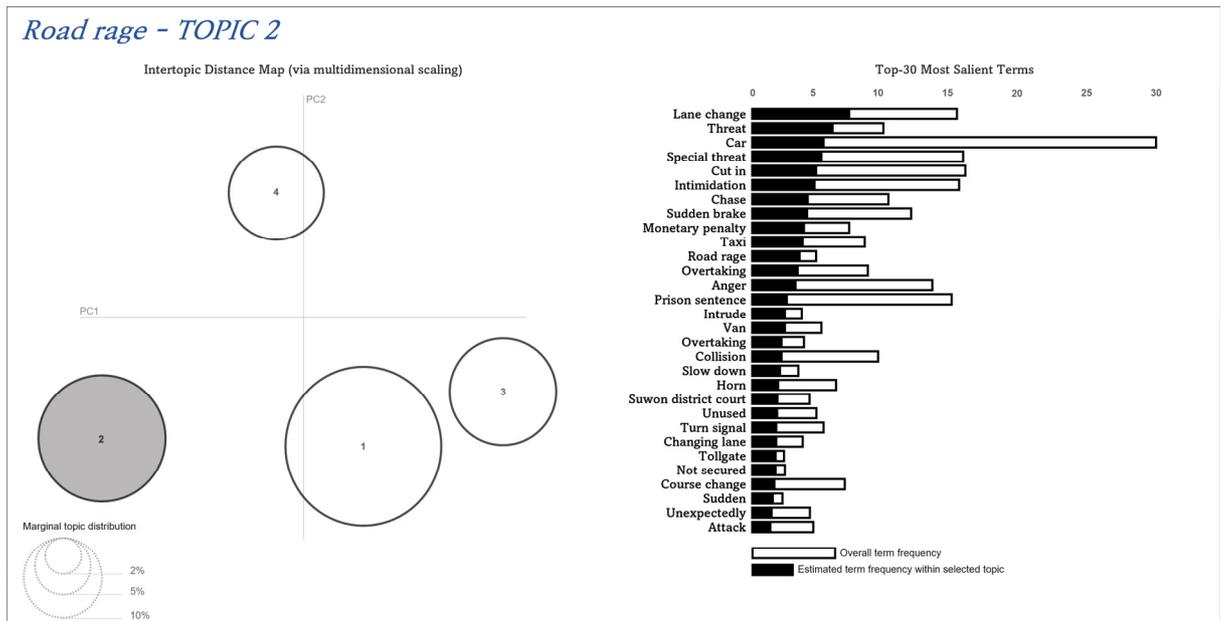


Figure 15. Topic modeling result of road rage - Topic 2

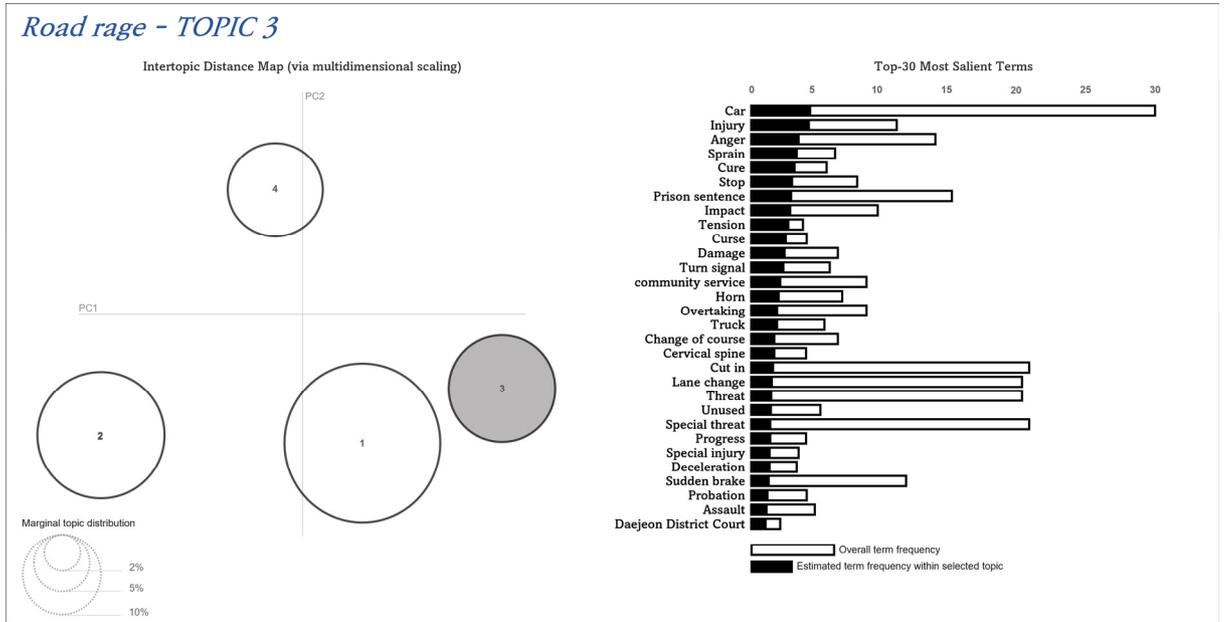


Figure 16. Topic modeling result of road rage - Topic 3

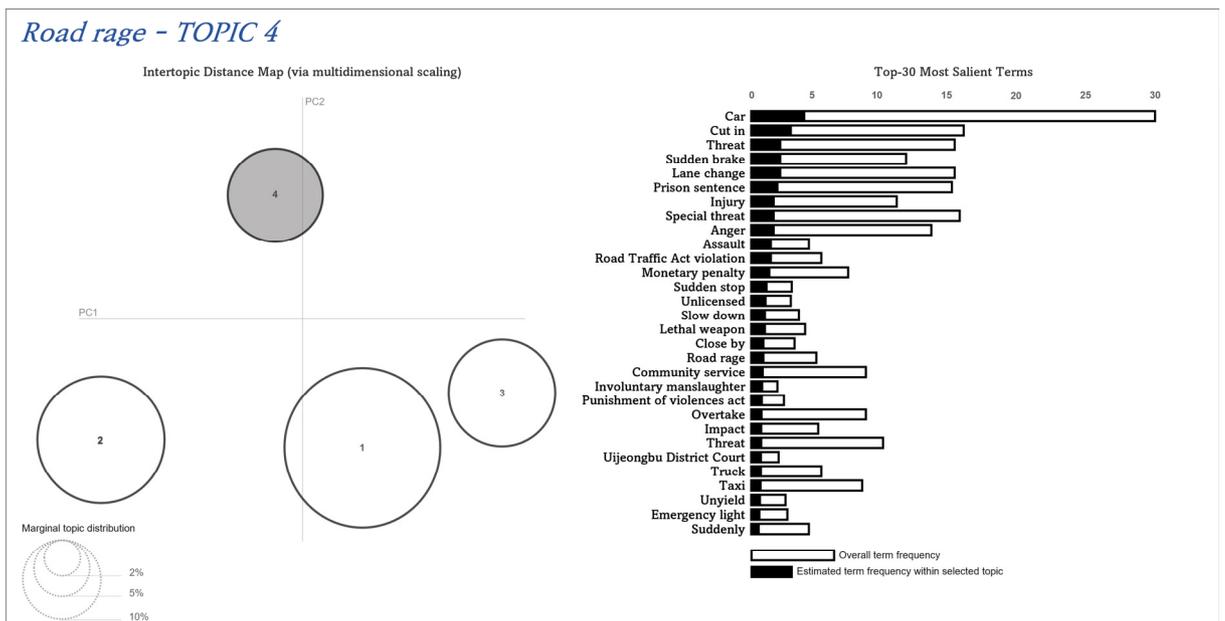


Figure 17. Topic modeling result of road rage - Topic 4