

토픽모델링을 이용한 지역별 일반국도 도로불편민원 신고 패턴 분석

김진국, 전우훈

한국건설기술연구원

jingukkim@kict.re.kr, cwhoon@kict.re.kr

Analysis of Regional Reported Road Inconvenience Complaints Pattern of National Highway using Topic Modeling

Kim Jin Guk, Jeon Woo Hoon

Korea Institute of Civil Engineering and Building Technology

요약

최근 이상기후로 인한 폭염, 한파, 그리고 태풍으로 인한 돌풍, 집중호우 등이 자주 발생함에 따라 도로 노면 및 안전시설 파손 등 운전자 안전을 위협하는 민원신고가 지속적으로 증가하고 있다. 본 논문에서는 지역별로 자주 신고되는 일반국도 도로불편신고 민원에 대한 유형을 분석하였다. 이를 위해 약 8년간 일반국도에서 발생한 도로불편민원 신고데이터를 기반으로 비정형 자료 분석에 주로 활용되는 토픽모델링(Topic modeling) 기법 중 LDA(Latent Dirichlet Allocation)를 이용하여 지역별 일반국도에서 자주 신고 되는 민원들에 대한 유형을 도출하였다. 본 연구결과는 지역별 일반국도의 효율적이고 선제적인 도로 유지관리 정책 수립을 위한 기초자료로 활용될 수 있을 것이다.

I. 서론

매년 도로연장의 지속적인 증가와 국민 삶의 질 향상으로 안전하고 깨끗한 도로환경 조성을 요구하는 도로 이용자들이 늘고 있다. 이로 인해 최근 인공지능 및 빅데이터 기반의 스마트 첨단 기술을 적용한 다양한 도로유지관리 기술들이 개발되고 있다.

도로 상에서 발생하는 각종 도로 파손(포트홀), 시설물 파손, 낙하물, 낙석 등은 도로 이용자의 안전을 위협할 뿐만 아니라 2차 사고의 위험으로 이어질 수 있다. 이로 인해, 국토교통부에서는 2014년부터 일반국도에서 발생하는 도로불편민원을 스마트폰 앱으로 신고 받아 운영·관리하고 있다.

따라서, 본 연구에서는 일반국도 상에서 자주 발생하여 신고 되는 도로불편민원에 대한 유형을 지역별(서울권, 강원권, 충청권, 경상권, 전라권)로 분석하였다.

II. 연구방법론

본 연구에서는 일반국도에서 발생한 도로불편민원 신고 데이터를 기반으로 비정형 자료 분석에 주로 활용되는 텍스트마이닝(Text mining) 중 토픽모델링 기법을 이용하여 지역별 신고민원에 대한 유형을 분석하였다. 토픽모델링은 주로 비구조화된 텍스트 문서에서 새로운 정보를 추출하기 위한 연구에 많이 활용되고 있으며, 특정 분야의 주제 및 동향을 파악하는데 효과적인 수단으로 제안되고 있다. 대량의 텍스트 자료를 이용하여 핵심 주제나 유형에 대한 키워드 도출이 가능하다.

지역별로 자주 발생한 도로불편민원 신고 유형을 분석하기 위해 2014년부터 2021년까지 “적척해결서비스 App(국토교통부)”을 통해 수집된 민원 Database를 수집하였다. 구축된 데이터를 기반으로 문서집합을 생성한 후 토픽모델링 분석기법 중 LDA를 활용하여 지역별로 일반국도에서 자주 신고되는 민원들에 대한 유형을 도출하였다. 본 연구의 수집 데이터에 대한 설명은 Table 1과 같다.

Table 1. Data collection

Contents	Complaints database
Period	2014.04 ~2021.12
Area	National Highways
Region	Metropolitan, Gangwon, Chungcheong, Gyeongsang, Jeolla
# of complaints	16,956

2014년부터 2021년까지 전국 일반국도에서 발생한 도로불편민원 신고건수는 약 1만 7천 건에 달하며, 지역별로는 전라권이 약 5천 5백 건으로 제일 많은 민원이 신고되었으며, 강원권이 약 2천여 건으로 제일 적은 민원이 접수된 것으로 나타났다(Fig.1).

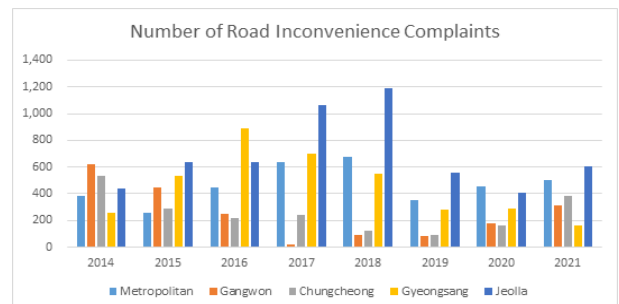


Fig.1 Number of road inconvenience complaints on national highways

III. 분석결과

본 연구에서는 LDA 토픽모델링 기법을 통해 핵심 주요 토픽과 키워드를 분석하였다. 토픽 수 결정을 위해 LDA 실루엣(Silhouette) 계수를 이용하여 최적 토픽 수를 도출하고, 알파 값을 0.1, 베타 값을 0.01로 설정한 후 샘플링 분석을 1,000회 반복하여 수행하였다. Zhao(2015)는 선행연구를

Table 2. LDA topic modeling results

Region	Topic results	1 st		2 st keyword		3 st keyword		4 st keyword		5 st keyword	
		keyword	Prob	keyword	Prob	keyword	Prob	keyword	Prob	keyword	Prob
Metropolitan	Topic 1	차선	0.061	안전시설	0.039	배수	0.038	시설물	0.031	도색	0.027
	Topic 2	불법	0.053	현수막	0.042	노면	0.030	차량	0.029	사고	0.020
	Topic 3	포트홀	0.054	동물사체	0.038	차선	0.034	포장	0.021	시설물	0.018
	Topic 4	쓰레기	0.120	노면	0.100	포트홀	0.079	길어깨	0.072	토사	0.048
Gangwon	Topic 1	표지판	0.058	시선유도봉	0.046	로드킬	0.029	사고	0.027	차선	0.025
	Topic 2	포트홀	0.080	포장	0.063	교량	0.059	중앙분리대	0.055	가드레일	0.028
	Topic 3	횡단보도	0.019	사고	0.019	도색	0.019	시선유도봉	0.019	교차로	0.018
	Topic 4	포트홀	0.076	배수	0.049	낙하물	0.026	낙석	0.022	노면	0.021
Chungcheong	Topic 1	포트홀	0.030	노면	0.027	표지판	0.025	교차로	0.018	신호	0.018
	Topic 2	로드킬	0.059	중앙분리대	0.032	포트홀	0.021	차선	0.021	낙하물	0.017
	Topic 3	시선유도봉	0.051	충격흡수시설	0.034	가드레일	0.033	시설물	0.025	교차로	0.020
	Topic 4	불법	0.105	배수	0.025	쓰레기	0.024	광고	0.019	표지판	0.018
Gyeongsang	Topic 1	사고	0.029	노면	0.022	차선	0.019	통행	0.018	신호	0.016
	Topic 2	쓰레기	0.176	로드킬	0.062	교차로	0.033	표지판	0.029	불법	0.022
	Topic 3	차선	0.042	낙하물	0.033	포트홀	0.028	노면	0.024	중앙분리대	0.016
	Topic 4	시설물	0.096	표지판	0.050	포트홀	0.040	가드레일	0.021	노면	0.021
Jeolla	Topic 1	노면	0.106	포트홀	0.092	쓰레기	0.083	낙하물	0.043	청소	0.012
	Topic 2	로드킬	0.060	차선	0.041	포트홀	0.036	중앙분리대	0.026	포장	0.020
	Topic 3	사고	0.023	표지판	0.020	노면	0.019	차선	0.014	노면표시	0.013
	Topic 4	표지판	0.080	불법	0.071	교차로	0.037	중앙분리대	0.028	시설물	0.021

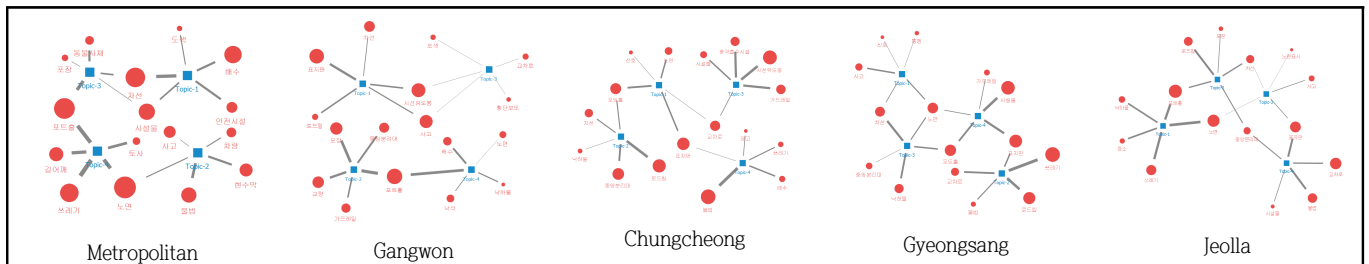


Fig.2 Topic & keyword network by LDA topic modeling

통해 알파 값은 0.1~0.01, 베타 값은 0.01이 최적 값인 것으로 제시하였다.

지역별로 도출된 토픽모델링의 결과를 살펴보면, 수도권의 경우 Topic별 키워드 중 차선 시인성 불량, 불법 시설물 설치, 노면상태 불량, 포트홀 등과 관련된 민원이 가장 많이 신고된 것으로 나타났다. 강원권의 경우 포트홀, 포장상태 불량, 교량난간 파손 중앙분리대 파손, 배수시설 불량 등과 관련된 민원이 가장 많이 신고된 것으로 분석되었고, 충청권은 불법 시설물 설치, 로드킬, 시선유도봉 등 안전시설물 파손 등과 관련된 민원이 가장 많이 신고된 것으로 나타났다. 경상권의 경우에는 도로 쓰레기, 도로안전 시설물, 로드킬, 표지판 등과 관련된 민원이 가장 많이 신고된 것으로 나타났으며, 전라권은 포트홀, 도로 표지판, 불법 시설물 설치, 로드킬과 관련된 민원이 가장 많이 신고된 것으로 분석되었다.

IV. 결론

최근 이상기후로 인한 폭염, 한파, 그리고 태풍으로 인한 돌풍, 집중호우 등이 자주 발생함에 따라 도로 노면 및 안전시설 파손 등 운전자 안전을 위협하는 민원신고가 지속적으로 증가하고 있다.

본 연구에서는 약 8년 동안 수집된 일반국도의 도로불편신고 민원을 대상으로 지역별로 자주 발생하는 민원 유형을 분석하였다. 지역별 도로관리기관의 도로 보수 및 유지관리 인력은 제한적이므로 이를 체계적으로 관리하는 데에는 현실적으로 어려움이 있다. 따라서 본 연구결과는 지역별 일반국도의 효율적이고 선제적인 도로 유지관리 정책 수립을 위한 기초자료로 활용될 수 있을 것이다.

참 고 문 헌

- [1] Zhao, W., Chen, J. and Zen, W. "Best Practices in Building Topic Models with LDA for Mining Regulatory Textual Documents", CDER 9th November, 2015.
- [2] S. M. Kim and Y. J. Kim. "Research Trend Analysis on Living Lab Using Text Mining", Journal of Digital Convergence, Vol. 18, No. 8, pp. 37-48, 2020.
- [3] Y. J. Kim, D. S. Yoo, et al. "Latent Crime Reporting Types Analysis Based on LDA Topic Modelg for Unclassified Types of 112 Crime Reporting Data in Police", The Journal of Korean Institute of Communications and Information Sciences(J-KICS) Fall Conference, pp. 0891-0892, 2020.
- [4] T. H. Won, H. H. Yoo "Pattern Analysis for Civil Complaints of Local Governments Using a Text Mining", Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and
- [5] K. C. Park and C. H. Lee "A Study on the Research Trends for Smart City using Topic Modeling", J. Internet Comput. Serv., Vol. 20, No. 3, pp. 119-128, 2019.
- [6] W. H. Jeon and I. C. Yang "Development of a Location-based Smartphone App Reporting Road Problems With User Participation", Journal of Digital Contents Society, Vol. 19, No. 10, pp. 1825-1832, 2018.