

비정형 텍스트 기반의 토픽 모델링을 이용한 건설 안전사고 동향 분석

이상규
한국건설기술연구원

A Study on the Trends of Construction Safety Accident in Unstructured Text Using Topic Modeling

Sang-Gyu Lee

Korea Institute of Civil Engineering and Building Technology

요약 본 연구는 건설 안전사고에 대한 트렌드 분석을 위해 LDA(Latent Dirichlet Allocation) 기반의 토픽모델링(Topic Modeling)을 제시하여 분석하고자 한다. 특히, 건설산업의 안전사고를 예방하기 위해 제시되고 있는 기존의 다양한 정형데이터 분석에서 벗어난 비정형 데이터 분석 기반의 토픽 모델링을 통해 건설 안전사고 주요 핵심 키워드의 흐름에 대해 파악이 가능하다.

본 방법론을 적용하기 위해 540개의 건설 안전사고 관련 뉴스데이터를 수집하였다. 이를 기반으로, 10가지 토픽과 각 토픽 내의 10가지 키워드를 통해 주요 이슈를 도출하였고 각 토픽에 대한 2017년 1월부터 2018년 2월까지의 뉴스 데이터를 월별 시계열 분석을 통해 향후 토픽에 관한 이슈를 예측한다.

본 연구를 바탕으로 향후 건설 안전사고의 다양한 이슈를 선제적으로 예측하고 이를 기반으로 건설 안전사고 정책과 연구에 좋은 방향을 제시할 것으로 판단한다.

Abstract In order to understand and track the trends of construction safety accident, this study shows the topic trends in the construction safety accident with LDA(Latent Dirichlet Allocation)-based topic modeling method for data analytics. Especially, it performs to figure out the main issue of construction safety accident with unstructured data analysis based on the topic modeling rather than a variety of structured data analysis for preventing to safety accident in construction industry.

To apply this methodology, I randomly collected to 540 news article data about construction accident from January 2017 to February 2018. Based on the unstructured data with the LDA-based topic modeling, I found the 10 topics and identified key issues through 10 keyword in each 10 topics. I forecasted the topic issue related to construction safety accident based on analysis of time-series trends about the news data from January 2017 to February 2018. With this method, this research gives a hint about ways of using unstructured news article data to anticipate safety policy and research field and to respond to construction accident safety issues in the future.

Keywords : Unstructured data, Construction Safety Accident, Topic Modeling, Latent Dirichlet Allocation (LDA), News Article, Time-Series

본 논문은 국가과학기술연구회 창의 융합클러스터 지원사업(20180221-001)의 일환으로 수행되었음.

*Corresponding Author : Sang-Gyu Lee(Korea Institute of Civil Engineering and Building Technology.)

Tel: +82-31-910-0526 email: sanggyulee@kict.re.kr

Received June 27, 2018

Revised (1st July 18, 2018, 2nd August 18, 2018, 3rd August 20, 2018, 4th August 20, 2018, 5th September 11, 2018, 6th September 13, 2018)

Accepted October 5, 2018

Published October 31, 2018

1. 서론

본 연구는 2017년부터 2018년 초까지의 건설 안전사고와 관련한 비정형 텍스트 뉴스 기사를 중심으로 핵심 키워드를 분류하여 현 트렌드를 파악한다.

이를 위해서 건설 안전사고 트렌드 분석을 위해서 해당 기간의 웹 뉴스 기사를 수집하고 비정형화된 데이터를 필터링한다. 더불어, 해당 데이터를 토픽 모델링 (Topic Modeling)에 적용하여 토픽별 소주제에 따른 주요 키워드 추출을 바탕으로 주요 이슈를 확인하고, 월별 시계열과 토픽별 의미 연관성을 그래프를 통해 확인하여 예측 분석한다.

본 연구의 결과로서 제시되는 다양한 건설 안전 키워드 및 트렌드 파악을 바탕으로 향후 건설 안전 정책 및 연구 방향에 도움을 줄 것으로 예상된다.

2. 본론

2.1 연구 배경

한국건설산업연구원의 2017년도 연구 보고서 <건설 안전사고 근절을 위한 혁신적인 안전관리 개선 방안 연구>에 따르면, 지난 5년간의 (2010~2015년) 사망만인율을 비교한 결과 한국의 건설산업 사망만인율(근로자 10,000명당 발생하는 사망자 수의 비율)이 4개 국가(한국, 미국, 영국, 싱가포르)중에서 한국이 가장 높은 것으로 나타났다[1]. 특히, 한국의 건설산업 사망만인율은 영국의 8.5배, 싱가포르의 2.9배, 미국의 2배가 높은 것으로 나타났다.

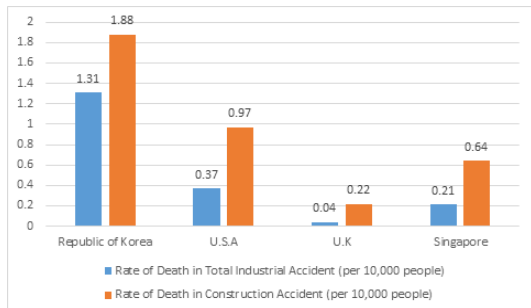


Fig. 1. Mortality Rate Comparison between Total Industry and Construction in 4 Nations (During 2010~2015)

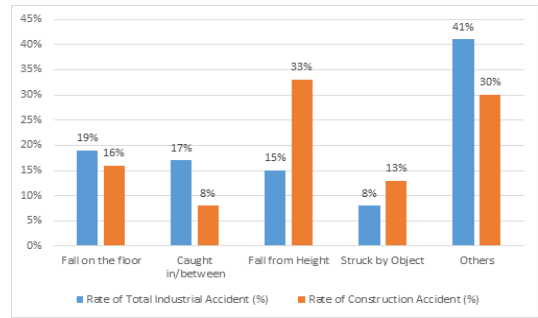


Fig. 2. Main Causes of Accident Comparison between Total Industry and Construction (%) (During 2010~2015)

더불어, 건설산업에서 사고 유형별로 일어나는 사고로는 떨어지는 사고와 물체에 부딪히는 사고 비율이 전체 산업보다 각 2.2배, 1.6배 더 높게 나타났다. 이에 비해 넘어짐과 끼임에 의한 사고는 건설산업이 전체산업에 비해 낮은 것으로 확인되었다.

또한, 평균 사고 유형별 분포도 상 떨어지는 사고 (33%)가 가장 높은 사고 유형별로 확인되었고 다음으로는 넘어지는 사고(16%), 물체에 부딪힘 (13%), 끼임 (8%) 순으로 나타났다.

제시된 건설 사고에 관한 정형데이터는 설문 작성자의 설계를 기초로 작성자의 주관적 이슈로 한정되어 확인되고 있다.

특히, 기존 정형 데이터와는 다르게 현재 SNS, 뉴스, 문서 등과 같이 방대한 양의 비정형 데이터가 쏟아지고 있는 가운데 보다 의미있는 데이터를 수집하고 분석하여 유의미한 결론을 도출하는 것이 필요하다[2]. 따라서, 비정형 데이터의 분석 방법으로서 제시되고 있는 텍스트/오픈이언 마이닝, 소셜네트워크 분석, 군집 분석 가운데 본 연구는 텍스트 마이닝의 분석기법으로서 자연어를 기반으로 정보를 가공하고 의미있는 정보를 추출하여 건설 안전사고에 대한 주요 토픽별 연계성을 도출하게 된다.

최근 비정형 뉴스 데이터를 기반으로 트렌드 분석에 대한 연구는 활발하게 수행되어 왔다. 뉴스데이터의 조회기록을 통해 인터넷 사용자의 호감도 흐름을 분석[3], 뉴스데이터를 통해 토픽간의 시간 격차를 확인하는 연구 [4], 뉴스를 활용한 이슈 유사도를 기간별 분석[5] 등이 있다. 이를 기반으로 본 연구에서 비정형 뉴스 데이터를 바탕으로 월별로 바뀌는 주요 키워드를 확인하여 이슈 흐름을 파악하는 것이 필요하다.

기계학습 기반의 토픽모델링을 바탕으로 비정형데이터를 분석하여 현 건설산업의 안전사고에 대한 이슈를 정리하고 이를 통해 건설 안전 강화 정책에 밑바탕이 될 수 있는 키워드 제안을 하고자 한다.

2.2 연구 방법

본 연구 방법은 비정형 뉴스 데이터 수집 및 구축 과정, 형태소 분석, 토픽 모델링 기법을 통해 건설 안전사고에 관한 현 동향을 분석한다. 특히, 기계학습 기반의 토픽모델링은 비정형화되어 있는 텍스트 자료에 속해있는 의미 있는 토픽을 추출해주는 확률모델로서 주로 사용되고 있는 기법이 LDA (Latent Dirichlet Allocation)로 본 연구에서 이를 적용하여 최근 건설 안전사고의 동향을 파악하고자 한다.

2.2.1 데이터 수집 및 구축과정

건설 안전사고의 키워드 분석의 첫 단계로서 데이터 수집을 통한 구축과정이 필요하다. 따라서, 텍스트 기반의 비정형 데이터 수집을 위해서는 보고서, 잡지, 뉴스기사, 논문 등을 확인할 수 있다. 특히, 뉴스기사와 같은 대중미디어는 과학기술에 대한 대중의 이해와 태도에 많은 영향을 미치고 있으며 국가 정책 파악을 쉽게 할 수 있다.

더불어, 비정형 데이터임에도 불구하고 뉴스기사의 형식을 갖추는 내용이 많기 때문에 NLP (자연어 처리: Natural Language Processing) 과정에 매우 용이하다[6]. 특히, 육하원칙을 기반으로 논문이나 잡지에 비해 뉴스기사는 내용은 많지 않지만 축약된 형식으로 NLP 처리 속도 및 정확성에 높은 장점을 가지고 있다. 이러한 이유를 바탕으로 데이터 수집 대상은 뉴스기사 데이터로 정하였다.

본 연구에서는 최근에 발생한 건설 안전사고 관련 키워드 중심으로 2017년 1월부터 2018년 2월까지의 약 540여개의 뉴스 기사를 랜덤하게 추출하였다. 뉴스 제목, 내용의 중복성을 띄는 기사는 제거하였고, 조사 및 전치사와 같은 키워드 외의 단어는 연구 데이터에 적용하지 않았다.

2.2.2 형태소 분석

뉴스와 같은 문서 또는 텍스트 형태의 파일을 데이터 전처리 과정을 통해 기계적으로 이해하고 분석할 수 있는 형태로 바꾸어야 한다.

프로그래밍 R 툴의 KoNLP (한글 자연어 처리) 패키지를 통해 비정형화된 문서 또는 텍스트 파일의 데이터를 단어 형태로 바꾸고 해당 단어 형태만 추출하도록 한다.

특히, 의미 있는 토픽 도출의 목적으로 사용될 토픽 모델링을 구현하기 전에 선행되어야 하는 것이 바로 형태소 분석이다. 따라서, 보다 의미 있는 토픽을 도출하기 위해서는 텍스트 파일 내의 글자의 수가 2 글자 이상의 단어 형태의 추출과 띄어쓰기 오류 및 미등록어와 같은 무의미한 단어를 제거함으로써 자동색인 방지 목적으로 사용하지 않는 용어를 파악해야 한다[7].

2.2.3 토픽 모델링 적용

전처리된 데이터 안에서 의미 있는 토픽 도출을 위해서 사용되는 가장 대표적인 기법은 LDA이다.

- * K : 토픽의 수
- * α : 문서별 토픽 K 의 Dirichlet prior weight, 값을 결정하는 파라미터
- * η : 토픽별 단어 W 의 Dirichlet prior weight, 값을 결정하는 파라미터
- * θ_d : 문서별 토픽의 비율
- * β_k : 토픽별 단어 W 의 생성확률
- * $Z_{d,n}$: 문서 d 의 n 번째 단어의 토픽
- * $W_{d,n}$: 문서 d 의 n 번째 단어

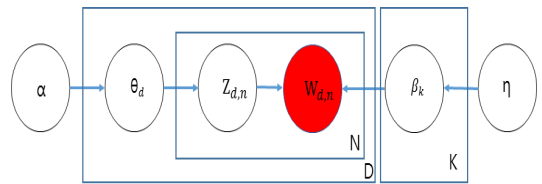


Fig. 3. Latent Dirichlet Allocation (LDA) Model

LDA 기법은 문서별 토픽과 토픽별 특정 단어의 생성 확률을 계산한 확률 모형으로 문서 내에 분포된 주제를 추정하고 해당 문서가 특정 토픽들과 같이 분포되어 있는지를 찾아내고 있다[8].

LDA 기반의 토픽모델링을 통해 이슈를 분석하는 연구가 현재 활발하게 진행되고 있다.

LDA 분석방법론을 기반으로 사회 문제의 이슈를 분석하였고[9], 국내 문헌정보학에 관련한 학술지의 분야별 트렌드 분석을 LDA 기반의 토픽모델링으로 분석 수

행하였고[10], LDA 기반의 토픽모델링을 통해 4700건의 특허 DB를 기반으로 핀테크 기술에 대한 주요 토픽별 이슈를 제시하였다[11].

본 연구는 프로그램 R를 통해 깃스 샘플링 3000번을 반복하였고, 총 10개의 토픽 주제를 선정하였다. 각 10개의 토픽에 대한 각 주제를 선정하였다[12].

2.3 연구 결과

아래 Table 1과 같이 토픽 모델링의 LDA 기법을 통해 선정된 10개의 토픽을 바탕으로 제시된 토픽 단어 10개씩을 나누었다.

Table 1. Results of Topic Modeling

	Topic Word
Topic 1	safety management, construction sites, safety accident, thaw, Ministry of Public Safety and Security, fraudulent construction, persons concerned, local government, manual
Topic 2	engineering, thaw, industrial revolution, infrastructure, Hyun-mi Kim, construction technology, technical skills, research and development, smart, competitiveness
Topic 3	Shin Gori, Jae-in Moon, public opinion, energy, Korea Hydro & Nuclear Power Company, social, debate, de-nuclear, cooperative society, activation
Topic 4	construction company, construction engineer, subcontract, building, construction sites, amendment, main contractor, temporary equipment and material, Seoul, obligation
Topic 5	construction sites, system, worker, safety management, construction company, Seoul, safety accident, real time, smart, safety education
Topic 6	facility, road facility, maintenance, Vietnam, continuous, construction technology, real time, Ministry of Land, Infrastructure, and Transport, undersea tunnel, fine dust
Topic 7	safety accident, construction sites, worker, Jeju province, tower crane, apartment, mortality, Gyeonggi province, persons concerned, mold
Topic 8	construction company, construction sites, worker, safety accident, industrial accident, Ministry of Employment and Labor, a place of business, safety management, mortality, subcontractor
Topic 9	Ministry of Land, Infrastructure, and Transport, construction machinery, infrastructure, monitoring, service, industrial revolution, disaster safety accident, safety policy, automation, prompt
Topic10	tower crane, Gyeonggi province, secret fund, distribution center, occurrence, health and safety, inspection agency, government employee, fall, Ministry of Land, Infrastructure, and Transport

2.3.1 토픽별 시계열 분석

분석된 토픽별 주요 키워드를 토대로 문서 내의 특정 단어의 생성확률을 월별 가중치의 값으로 계산하여 토픽별 시계열분석을 통해 월별 추이를 확인한다. 이러한 분석은 토픽에 대한 시간적 분포가 다르게 나타날 것으로 예상되기 때문에 시계열 분석을 통해 토픽 분포의 차이를 확인한다[13].

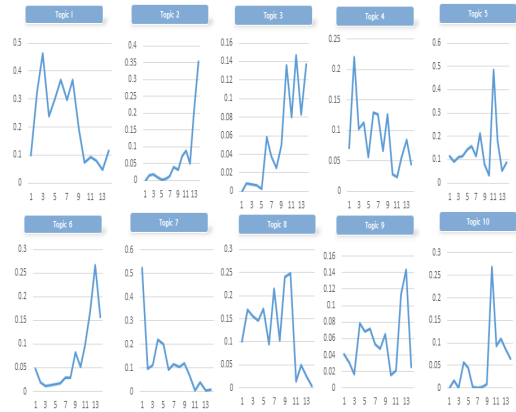


Fig. 4. Results of Time-Series by Topic (Graph)

제시된 Fig. 4는 2017년 1월부터 2018년 2월까지의 뉴스 비정형 데이터를 토대로 토픽별 추이도를 제시한다.

Topic 1의 경우, 건설업계에서 건설 관계자의 수준 높은 건설현장 관리능력 확보와 부실시공을 사전에 예방하기 위해 안전관리 및 메뉴얼 제공을 통해 건설 안전에 대비하고 있다. 더불어, 국민안전처는 건설현장 안전사고를 예방하기 위해 현장 관계자의 사고 주의를 지속적으로 요구하고 있다.

Topic 2의 경우, 제4차 산업혁명에 대응하는 건설 기술 개발을 위해서 스마트 건설 자동화와 같은 건설 현장의 노동생산성 향상과 안전사고 감축을 위한 정책이 발표되고 있다. 특히, 현장에서 어려운 시공과 안전사고 발생을 예방하기 위해서는 스마트 건설 자동화 기술을 본격적으로 연구개발과 건설 엔지니어링의 국제 경쟁력 강화를 위한 다양한 제도가 필요하다.

Topic 3의 경우, 문재인 정부 이후 탈원전 에너지 정책을 통해 2017년 초부터 신고리 5호기와 6호기의 원전 공사를 중단하고 국민적 공론화를 통해 원전 건설 재개 결정되었다. 하지만, 아직까지 건설 안전의 이슈로 탈원전에 대한 문제는 지속적으로 논의되고 있다.

Topic 4의 경우, 가설공사 안전사고의 주요 원인이 기자재 노후화와 자재결함 보다는 가설구조물의 설치기준을 준수하지 않아 발생되고 있다.

더불어, 건설 재하청을 거칠수록 단가가 줄어들고 고용책임도 불분명해지기에 원청의 무리한 요구를 따를 수밖에 없는 것도 안전사고의 주요 요인이라고 볼 수 있다. 즉, 건설 안전사고를 유발하는 하도급 업체에 대한 공사 참여 배제하고 부실공사와 임금체불과 같은 안전사고 유발 요인을 사전에 차단하는 노력이 필요하다.

Topic 5의 경우, 건설 안전관리를 위해 스마트 건설 기술개발 R&D 추진과 ICT와 IoT의 첨단 기술을 현장에 적극 적용하고 체계적인 안전관리를 위한 실시간 모니터링 시스템 구축을 계획하고 있다. 더불어, 안전 역량 강화를 위한 건설 안전교육 수업을 통해 안전사고 예방에 다양한 노력을 기울이고 있다.

Topic 6의 경우, 국토교통부는 건설 안전을 위한 시설물 유지관리에 대한 R&D 투자가 점차적으로 증가하고 있다. 이는 지진과 같은 재난 재해 그리고 미세먼지와 같은 사회 이슈 문제 해결을 대응 지원하기 위해 관련 연구개발에 집중적으로 투자하고 있다. 더불어, 국내 건설엔지니어링 업계의 지속적인 해외진출을 지원하기 위해 해외 전문 건설인력을 양성을 추진하고 있다.

Topic 7의 경우, 제주지역에서 대형 숙박 및 주거 단지의 건설공사가 진행되고 있는 가운데 건설 사고가 빈번히 일어나고 있다. 더불어, 경기지역에서는 아파트 건설 공사장에서 거푸집 해체 과정에서 사용되는 타워크레인 사고로 인해 다수의 사망자가 발생하고 있다.

Topic 8의 경우, 건설현장의 안전관리 부실 문제를 해결하기 위해 고용노동부에서는 산업재해 발생 시 발주자와 원청의 책임 강화 및 건설 하청업체의 위험 작업의 외주화 방지 제도를 추진하고 있다.

Topic 9의 경우, 건설산업에서 사용하는 타워 크레인 및 건설장비와 같은 다양한 건설기계의 선제적 사고 방지를 위해서는 안전기준 강화가 필요하다. 더불어, 신축 건축물의 부실공사 예방과 재난안전을 위해서 건축설계 및 시공 단계의 전문가 모니터링을 통해 신속한 건설 안전 정책을 추진한다.

Topic 10의 경우, 경기도지역에서 발생한 타워크레인 해체작업 중 크레인이 추락하여 많은 사상자가 발생한 난 뒤 사건 무마를 위해 비자금을 조성하였다. 더불어, 건설업체가 산업재해 예방을 위해 사용될 안전보건관리

비를 공무원 비자금에 사용되고 있다.

본 연구의 데이터 분석 방법론으로 제한한 LDA기법의 토픽모델링을 통해 제시된 토픽별 10가지의 키워드를 확인하였다. 더불어, 문서 내의 주요 단어의 생성확률을 월별 가중치 값으로 계산하였고 토픽별 시계열 분석을 활용하여 2017년 1월부터 2018년 2월까지의 각 토픽의 월별 추이도를 Fig. 4를 통해 확인하였다.

앞서 제시된 토픽별 시계열 분석 결과를 토대로 선형 회귀분석을 통해 향후 이슈 예측을 확인한다.

선형회귀분석은 사회현상에 대한 원인 및 결과의 관계를 함수를 통해 선형으로 근사시키는 기법으로서, 단순하게 선형을 통해 원인과 결과를 밝히기는 어렵고 선형에 가깝게 해석하는 것은 매우 높은 오차를 발생할 수도 있다고 한다[14]. 하지만, 단순 월별에 따라 토픽 가중치값을 통해 토픽의 추세를 제시하는 대다수의 토픽모델링 연구에서 선형 회귀 분석의 회귀계수 값을 사용하여 확인하고 있었다.[11],[15]

앞서 기존 문헌을 바탕으로 본 연구에서 독립변수는 연도별로 적용하며 종속변수는 건설 안전사고 토픽별 평균값을 이용하여 선형 회귀분석의 회귀계수로 판별한다.

이를 바탕으로 아래 Table 2는 시계열적 동향 결과를 적용하여 유의미한지를 확인한다. 유의수준은 0.1을 기준으로 통계적으로 유의미하다고 판단한다. 더불어, 회귀계수가 0보다 크면 토픽별 이슈가 상승한다고 판단하여 Hot, 0보다 작으면 하락한다고 판단하여 Cold로 이슈를 구별하였다[16].

Table 2. Analysis of Time-Series (Hot/Cold)

	Coefficient	P-Value	Hot/Cold
Topic 1	-0.2141	0.0957	Cold
Topic 2	0.8761	8.637E-07	Hot
Topic 3	0.2515	0.06769	Hot
Topic 4	-0.064	0.3829	X
Topic 5	-0.05541	0.4179	X
Topic 6	0.6538	0.0004634	Hot
Topic 7	-0.1857	0.124	X
Topic 8	-0.338	0.02921	Cold
Topic 9	0.1072	0.2532	X
Topic 10	0.0319	0.5413	X

더불어, Table 2는 Fig. 4의 시계열 분석을 통해 유의미한 결과를 의미적 연관성으로 판단하여 주요 건설 안

전사고 이슈 예측을 제시한다. 예측 결과, 전체 10개의 Topic 중에서 5개의 Topic (Topic 1, Topic 2, Topic 3, Topic 6, Topic 8)이 유의한 것으로 확인된다. Topic 2 (안전사고 감축을 위한 스마트 건설기술 개발과 국제 경쟁력 강화 제도)와 Topic 3 (새 정부 출범 이후 탈원전 에너지 정책 논의), Topic 6 (사회 이슈 해결 및 시설물 유지관리 R&D 투자 확보와 해외 건설엔지니어링 전문 인력 양성) 이슈는 상승할 것으로 예측된다. 반대로, Topic 1(건설관계자의 현장 관리 및 매뉴얼을 통한 안전 사고 방지 요구)과 Topic 8 (고용노동부의 산재발생 시 원청 책임 강화 및 위험 작업의 외주화 방지 제도)의 이슈는 점차 하락 할 것으로 예측된다.

위 Table 2의 예측 결과를 기반으로 향후 건설 안전 사고 이슈에 대한 정책 또는 연구 방향을 적극적으로 활용할 필요가 있다.

3. 결론 및 고찰

비정형 텍스트 데이터 중에서 뉴스 기사 데이터를 중심으로 토픽 모델링 LDA 기법을 통해 선정된 10개의 토픽을 확인하였다. 더불어, 선정된 10개 토픽별 키워드 중심으로 시계열 그래프를 통해 향후 예상되는 결과를 확인하였다.

하지만, 건설 안전사고에 대한 키워드를 중심으로 뉴스 데이터를 한정하여 트렌드 분석을 하는 것은 몇 가지 한계점을 제시하기도 한다.

첫째, 뉴스 데이터 수집을 진행함에 있어서 중립적인 관점의 데이터를 수집하기 위해서는 Sentiment Analysis (감성분석)를 통해 긍정, 부정, 중립에 대한 의견을 구분하여 연구에 적용할 필요가 있다[17]. 만약, 데이터 수집의 오류로 인해 중립적 관점보다 편향적 뉴스 데이터로 인해 비현실적 정책 이슈로 바뀔 가능성이 있다.

둘째, 각 토픽 키워드들에 대한 뉴스 데이터의 추이를 장기적인 추세도로 확인할 필요가 있다. 이를 위해서는 연도별 데이터 수집이 필요하며 향후 지속적인 데이터 수집이 필요할 것으로 보인다. 이러한 시계열적 데이터가 축적이 된다면, 지속적인 관심도 상승 및 하락에 대한 추세를 파악하여 보다 효율적으로 정책 반영 및 흐름 정보를 구체적으로 추론할 수 있다는 점이다.

마지막으로, 비정형 뉴스 데이터로 국한된 데이터 수

집 및 분석으로 인한 적정 모수 데이터가 부족한 것을 확인하였다. 이에 따라 소셜 네트워킹 서비스(SNS: Social Networking Service)와 같은 다양한 비정형 데이터 수집을 통해 적정 모수 데이터를 해결할 필요가 있다.

본 연구는 토픽 모델링 기법을 통한 기존 트렌드 분석에서 사용되는 워드 클라우드 시각화와 같은 키워드 분석과는 다르게 기계학습 방법을 적용하여 토픽별 키워드를 분류하여 기존 텍스트 마이닝 분석과는 차별성을 두었다. 또한 토픽 모델링상 토픽별로 제시된 키워드의 생성 확률을 가중치값으로 계산되어 유의수준을 적용하여 통계 검증을 확인하였다.

그러나 현 건설 안전사고 정책 제시에서 참고되는 자료로 사용되는 분석으로는 비정형 데이터 기반의 분석 방법론을 적용하는 것보다는 설문 조사 기반의 정형 데이터 분석을 활용한 연구가 활발한 편이다. 따라서, 실용성 높은 비정형 데이터 분석 방법론을 기반으로 앞서 제시한 한계점을 보완한다면 향후 건설 안전사고 이슈를 활용한 다양한 예방 정책과 관련 연구에 대해서 좋은 방향을 제시할 것으로 고려된다.

감사의 글: 이 논문은 국가과학기술연구회 창의 융합클러스터 지원사업(20180221-001)의 일환으로 수행되었습니다.

References

- [1] S. Y. Choi, S. I. Choi, W. S. Yu, "A Study in Improvement of Current Problems of Safety Management System in Korean Construction Industry", Construction Economy Research Institute of Korea Report, 2017.
- [2] J. W. Lee, S. K. Kim, "Complementary Research and Analysis for Hadoop", *The Korea Society of Computer and Information Winter Conference 2012*, Vol.20, No.2, pp.3-6, July, 2012.
- [3] C. Liu, N. Kim, "Individual Interests Tracking : Beyond Macro-level Issue Tracking", *Journal of Information Technology Services*, Vol.13, No.4, pp.275-287, 2014. DOI: <https://dx.doi.org/10.9716/kits.2014.13.4.275>
- [4] D. H. Jeong, M. Song, "Time gap analysis by the topic model-based temporal technique", *Journal of Informetrics*, Vol.8, No.3, pp.776-790, 2014. DOI: <https://dx.doi.org/10.1016/j.joi.2014.07.005>
- [5] S. H. Min, "Investigating Dynamic Mutation Process of Issues Using Unstructured Text Analysis", *Journal of Intelligence and Information Systems*, Vol.22, No.1, pp.139-157, 2016.

DOI: <https://dx.doi.org/10.13088/jiis.2016.22.1.139>

- [6] D. M. Park, "News source network analysis as big data analytics of news articles", *Korean Society for Journalism and Communication Studies*, Vol.57, No.6, pp.234-262, 2012.
- [7] J. H. Lee, H. K. Lee, "A study on unstructured text mining algorithm through R programming based on data dictionary", *Journal of the Korea Industrial Information Systems Research*, Vol.20, No.2, pp.113-124, 2015
DOI: <https://dx.doi.org/10.9723/jksis.2015.20.2.113>
- [8] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol.3, pp.993-1022, 2003.
- [9] D. Jeong, J. Kim, G. N. Kim, J. U. Heo, B. W. On, M. Kang, "A Proposal of a Keyword Extraction System for Detecting Social Issues", *Journal of Intelligence and Information Systems*, Vol.19, No.3, pp.1-23, 2013.
DOI: <https://dx.doi.org/10.13088/jiis.2013.19.3.001>
- [10] J. H. Park, M. Song, "A Study on the Research Trends in Library and Information Science in Korea using Topic Modeling", *Journal of the Korean Society for Information Management*, Vol.30, No.1, pp.7-32, 2013.
DOI: <https://dx.doi.org/10.3743/kosim.2013.30.1.007>
- [11] T. K. Kim, H. R. Choi, H. C. 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, 2017.
DOI: <https://dx.doi.org/10.5762/kais.2016.17.11.670>
- [12] J. H. Park, H. J. Oh, "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea : focused on LDA and HDP", *Journal of Korean Library and Information Science Society*, Vol.48, No.4, pp.235-258, 2017.
- [13] J. H. Bae, N. G. Han, M. Song, "Twitter Issue Tracking System by Topic Modeling Techniques", *Journal of Intelligence and Information Systems*, Vol.20, No.2, pp.109-122, June, 2014.
DOI: <https://dx.doi.org/10.13088/jiis.2014.20.2.109>
- [14] B. Kang, M. Song, W. Jho, "A Study on Opinion Mining of Newspaper Texts based on Topic Modeling", *Journal of the Korean Library and Information Science Society*, Vol.47, No.4, pp.315-334, 2013.
DOI: <https://dx.doi.org/10.4275/kslis.2013.47.4.315>
- [15] J. S. Park, S. G. Hong, J. W. Kim, "A Study on Science Technology Trend and Prediction Using Topic Modeling", *Journal of the Korea Industrial Information Systems Research*, Vol.22, No.4, pp.19-28, 2016.
- [16] K. S. Shin, H. R. Choi, H. C. Lee, "Topic Model Analysis of Research Trend on Renewable Energy", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.16, No.9, pp.6411-6418, 2015.
DOI: <https://dx.doi.org/10.5762/KAIS.2015.16.9.6411>
- [17] B. Kang, M. Song, W. Jho, "A Study on Opinion Mining of Newspaper Texts based on Topic Modeling", *Journal of the Korean Society for Library and Information Science*, Vol.47, No.4, pp.315-334, 2013.
DOI: <https://dx.doi.org/10.4275/kslis.2013.47.4.315>

이 상 규(Sang-Gyu Lee)

[정회원]



- 2015년 2월 : 서울대학교 경영대학원 경영정보학 (경영학석사)
- 2016년 12월 ~ 현재 : 한국건설기술연구원 전임연구원

<관심분야>

건설정책, 데이터마이닝, 기술수용모델(TAM)