

건설현장 안전 지적 사항 분석

Vocabulary Analysis of Safety Warnings in Construction Site

강 경 수*

Kang, Kyung-Su

류 한 국**

Ryu, Han-Guk

Abstract

The purpose of this study is to analyze the vocabulary related to safety accidents based on the reports recorded on the violation of safety rules at the construction sites. We used Word2Vec and Topic Model as natural language processing techniques to analyze the safety accidents presented in the reports of the large enterprise. The words that appeared based on the occupational accident types such as the fall, falling objects, and others were derived and visualized. We derive the frequency and similarity of the words and topics of the accident that occur at the construction site. In future studies, we will be able to proceed with the generation of texts from pictures based on images and this reports.

키 워 드 : 건설안전, 자연어 처리, 워드 임베딩, 토픽 모형

Keywords : construction safety, natural language processing, word embedding, topic model

1. 서 론

2018년 전 산업 분야에서 건설업의 재해 발생빈도는 두 번째로 높으나, 중대 재해인 사망사고는 건설업이 가장 높은 것으로 나타났다. 건설현장은 잦은 안전재해가 발생하지만, 한번 발생하면 중대 재해로 연결되는 매우 위험성이 높은 산업이다. 이러한 특징으로 인해 건설현장에서는 재해를 예방하기 위해서 안전 관리자를 고용하여 현장을 관리하고 감시하고 있다. 특히, 안전 관리자는 안전수칙을 지키지 않거나, 불안정한 현장을 지적하고 이를 기록하고 있다. 하지만 안전지적을 기록한 문서는 관계형 데이터처럼 분석이 쉽지 않기 때문에 자세하게 기록된 이런 데이터를 현장에서 활용하지 못하고 있는 실정이다. 따라서 본 연구는 안전지적 보고서와 같이 텍스트로 이뤄진 비정형 데이터를 분석하여 데이터 내에서 의미 있는 패턴과 공통된 주제어를 도출하고자 하였다.

2. 데이터 및 방법론

총 250,437건의 안전지적 보고서를 통합하여 불필요한 문자열을 삭제 및 대치하고 띄어쓰기나 오타를 처리하였다. 그리고 각 보고서에서 명사만 추출하여 수치로 재표현하였다. 데이터 전처리 전 점검내용의 최대 문자열 길이는 288개이며 평균 18개를 사용하는 것으로 나타났다. 본 연구에서 활용한 Word2Vec과 토픽 모델링은 다음과 같다. Word2Vec은 텍스트를 수치화하는 단어 임베딩(word embedding) 기법 중 하나로 딥러닝을 활용하여 기존의 빈도기반 기법보다 문서에서 나타나는 단어에 대한 더 많은 정보를 저차원으로 표현할 수 있다. 토픽 모델링은 문서에서 토픽을 추출하는 모형으로 주어진 문서에 어떤 주제들이 존재하는지에 대한 확률모형이다.

3. 분석 결과

안전지적 보고서에서 나타나는 단어의 유사도와 토픽 모델링으로 도출된 주제어들은 다음 그림 1과 같다. 추락, 감전, 붕괴, 화재, 기타에서 나타나는 단어들의 밀집성은 매우 좋으나, 전도, 충돌, 낙하는 밀집력이 떨어진다. 특히 낙하는 추락, 붕괴, 전도 재해에 넓게 포진되어있어 낙하가 다른 사고와 연관성이 높은 것으로 보인다. 감전의 경우 다른 사고들과 확연히 차이 나는 발생형태로 보이며, 화재는 기타로 분류한 것들과 연관되어있다.

토픽 모델링에 안전지적 보고서 데이터를 적용한 결과 다음 표 1과 같다. 토픽의 개수는 보고서에서 나타나는 8개의 업무상 재해발생형태로 설정했으며, 8개와 같이 분류되는지 확인하였다. 그리고 추출한 주제어의 기여도를 활용하여 단어를 재표현하였다.

* 창원대학교 경영학과 박사수로

** 삼육대학교 건축학과 교수, 공학박사, 교신저자(ryuhanguk333@gmail.com)

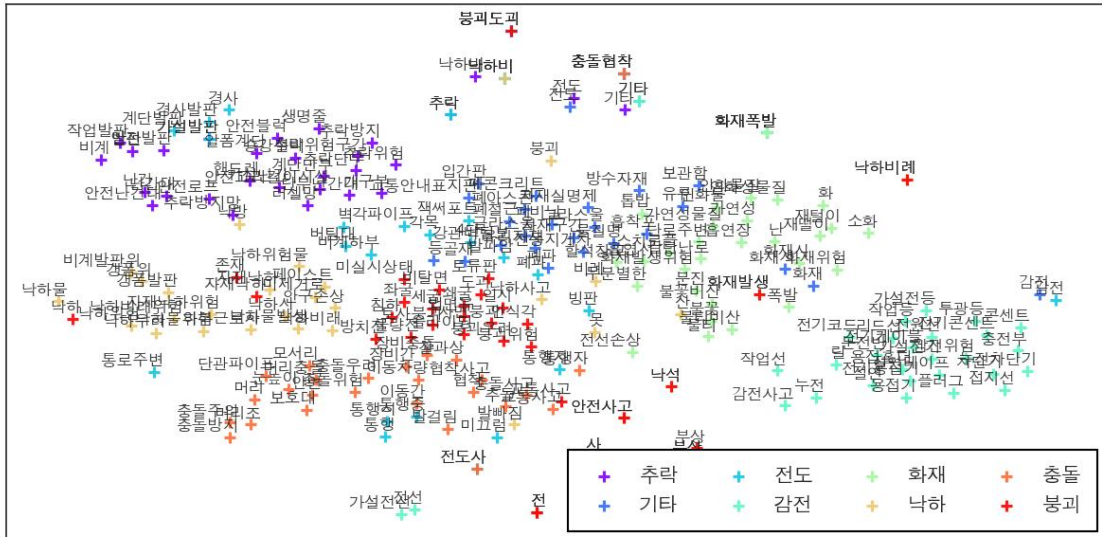


그림 1. Word2Vec 결과를 TSNE 2차원 시각화

표 1. 업무상 재해유형별 주제어

토픽	관련 단어	단어 재표현
1	추락, 설치, 발판, 미설치, 불량, 요망, 난간대, 단부,...	해체안한 시스템동바리상 발판단부 난간대 2줄 설치 승강사다리 설치
2	화재폭발, 방치, 기타, 소화, 조치요망, 미배치, 바닥, 점건, 사용, 보관,...	아르곤 고압용기 사용후 현장내 방치 위험물 저장이동보관 미설치 화재폭발
3	기타, 전도, 붕괴도괴, 요망, 자재정리, 내, 자재, 제거, 적치, 하부,...	알폼 써포트 자재 적체 불량 전도위험 전도
4	추락, 미설치, 미착용, 개구부, 안전고리, 미체결, 난간대, 해체, 사다리, 작업자,...	단부 난간대 미설치 작업자 안전벨트 미착용 안전고리 미체결, 추락
5	전도, 충돌협착, 기타, 위험, 미확보, 통, 이동통, 필요, 통행,...	지게차 자재운반시 운전자 시야 미확보 및 신호수 미배치 충돌위험 충돌협착
6	감전, 사용, 전도, 불량, 미설치, 작업선, 거푸집, 미조치, 상태불량,...	발전기 공급되는전기선 바닥방치 한 피복손상 예상 감전
7	낙하비, 기타, 미사용, 미비, 슬링벨트, 미부착, 보안경, 안전난간, 손상, 낙하물,...	슬링벨트 승 스티커 사용 중 기타
8	기타, 불량, 미흡, 정리정돈, 전도, 정리, 자재, 상태, 조치, 요망,...	작업 종료 후 정리 정돈 미흡 작업 환경 저하 기타

4. 결 론

본 연구는 안전지적 보고서에서 잠재된 의미 있는 패턴과 공통적인 주제어를 추출하여 분석하였다. 23만 건에 대한 문서에서 Word2Vec을 적용하여 재해발생유형과 밀접한 관계가 있는 단어들을 군집하여 시각화하였고, 토픽 모델링을 이용하여 각 주제어 및 결과를 이용하여 단어를 재표현하였다. 쉽게 분석할 수 없는 비정형 데이터인 안전지적 보고서에서 잠재된 의미를 끌어내었다. 향후 이에 관한 지속적인 연구를 통해 보고서에 첨부된 사진을 활용하여 텍스트를 자동으로 생성하고 안전 관리자의 업무를 줄일 수 있는 프로그램을 개발하도록 하겠다.

Acknowledgement

이 성과는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2017R1A2B4A003722).

참 고 문 헌

1. Mikolov, T., et al, Efficient Estimation of Word Representations in Vector Space, arXiv:1301.3781, 2013.
2. Bengfort, B., et al, Applied Text Analysis with Python: Enabling Language-aware Data Products with Machine Learning, O'Reilly Media, Inc., 2018.