TF-IDF와 K-Means 군집분석을 활용한 중대재해 사고 유형 분류

이영휘*, †

1. 서론

산업현장에서 반복적으로 발생하는 근로자 사망 및 상해 사고를 예방하고 근로자의 안전권을 보장하기 위하여, 정부는 2021년 1월 중대재해 처벌 등에 관한 법률을 제정하여 2022년 1월부터 시행하고 있다. 그러나 이러한 법적 장치가 마련된 이후에도 산업현장의 재해는 여전히 발생하고 있으며, 최근에는 국내 대형 베이커리 제조업체 SPC에서 작업자가 컨베이어 벨트에 끼여 사망하거나 손가락이 절단되는 등의 중대재해가 잇달아 발생하여 사회적 우려를 낳고 있다.

본 연구는 과거 중대재해 사고 데이터를 정량적으로 분석하여 반복적으로 발생하는 대표적 사고 유형을 규명하고, 이를 토대로 실효성 있는 예방 대책을 제안하고자 한다. 이를 위해 미국 노동부 산하 산업안전보건청(OSHA: Occupational Safety and Health Administration)이 공개하는 중대재해 보고서를 분석 대상으로 하였으며, 2015년부터 2025년까지 약 10년간 축적된 사고 데이터 중 사고 경위(narrative) 텍스트를 대상으로 텍스트 마이닝을 수행하였다. 구체적으로는 TF-IDF(Term Frequency-Inverse Document Frequency)로 텍스트를 벡터화한 후 K-Means 군집분석을 적용하여 유사한 사고 유형 그룹화를 통해 빈번하게 발생하는 사고 패턴을 도출하였다.

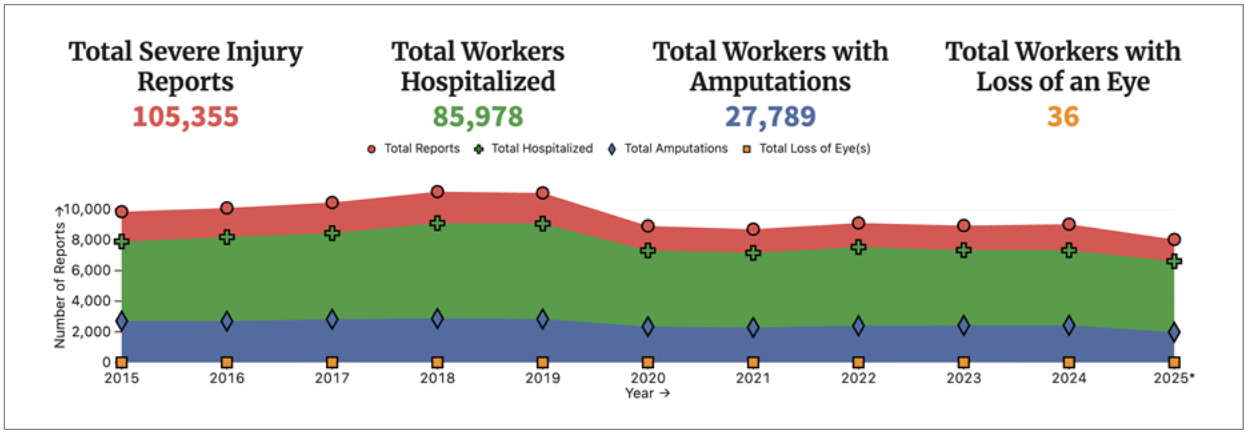


Fig 1. 미국 산업안전보건청의 중대재해 데이터 Dashboard

* 한국생산기술연구원 지속가능기술연구소 지역산업혁신부문(배출저감) / 기술원 / hwee@kitech.re.kr / 교신저자

2. 본론

가. 데이터 전처리 및 분석 방법

텍스트 마이닝(Text Mining)은 비정형 텍스트 데이터로부터 유의미한 인사이트를 도출하는 기법이다. TF-IDF(Term Frequency-Inverse Document Frequency)는 문서 내 단어의 출현 빈도와 희귀성을 함께 고려하여 각 문서를 특징 벡터화 하고, 단어별 중요도를 산출하는 알고리즘이다. 이러한 특성으로 인해 TF-IDF는 서술형으로 작성된 사고 경위 데이터에서 유사한 사고 유형을 군집화(Clustering)하거나, 각 유형을 특징짓는 핵심 단어를 도출하는 데 적합하다. 본 연구는 TF-IDF 벡터화와 K-Means 군집분석을 결합하여 중대재해 사고 유형을 분석하였으며, 전체적인 연구 절차는 다음과 같다.

Table 1. Research Framework

Input	->	전처리	->	분석	->	Out put	->	전략도출
중대재해 105,318건		사고 경위 추출, 불용어 제거		TF-IDF 벡터화, 클러스터링		키워드 도출, 유형 라벨링		예방 전략 도출

분석 데이터는 총 105,318건(업데이트 전 37건 제외)이며, 사고 ID, 발생일자, 사업장명, 사업장 주소, 발생 위치, 입원자 수, 절단 사고 여부, 상해·질병 분류 코드 등 총 28개 컬럼으로 구성되어 있다. 전처리 과정에서는 관사, 전치사, hospitalized, injury와 같이 사고 유형 판별과 상관없는 불용어 73개를 제거하여 전체 단어 출현 빈도를 약 345만 회에서 171만 회로 49.7% 감소시켰다. 이후 1-gram 및 2-gram을 포함한 TF-IDF 벡터화를 수행하고, 지나치게 희소하거나 과다 출현하는 단어를 제외하여 최종적으로 13,240개의 단어(특성)로 구성된 105,318×13,240 크기의 TF-IDF 행렬을 구축하였다. 이 행렬을 입력으로 K-Means 군집분석을 적용하여 사고 유형을 8개 군집(클러스터)으로 분류 후 상위 키워드를 도출하고 사고 유형으로 라벨링했다.

Table 2. 군집(클러스터)별 상위 키워드 및 사고 유형

군집	상위 키워드	사고 유형
1	지게차, 팰릿, 발, 충돌, 운전/작동, 다리, 사이, 발목, 랙(선반)	차량·지게차 충돌(지게차·운반장비)
2	사다리, 떨어짐, 피트(높이), 지면, 지붕, 골절, 아래, 바닥, 오르기	사다리·고소추락
3	톱, 톱날, 절단, 베임, 테이블톱, 절단됨, 손가락, 밴드톱, 목재	톱 절단 및 베임
4	화상, 온도, 고온, 뜨거운, 얼굴, 플래시, 물, 전기, 불, 아크, 손	화상·화학물질·전기 화상
5	떨어짐, 걸려 넘어짐, 걷기, 미끄러짐, 고관절, 바닥, 골절, 얼음	미끄러짐·넘어짐
6	기계, 손가락, 손, 절단, 끼임, 손가락끝, 검지, 중지, 가동/작동	기계 협착 및 손가락 절단(회전기계)
7	손가락, 절단, 손가락끝, 끼임, 손, 사이, 검지, 중지, 컨베이어, 벨트	기계 협착 및 손가락 절단(컨베이어)
8	트럭, 충돌, 떨어짐, 다리, 머리, 등/허리, 발, 골절, 트레일러	차량·지게차 충돌(화물차·트레일러)

나. 분석 결과

텍스트 마이닝과 군집분석을 실시한 후 중대재해와 관련성이 높은 클러스터별 산업재해 비율, 산업별 사고유형 비율, 사고 유형별 신체 부위 비율을 분석했다.

첫째, 클러스터별 산업재해 비율은 차량·지게차 충돌과 기계 협착 38.8%, 손가락 절단(컨베이어) 19.3%로 높게 나타났으며, 두 유형이 전체 사고의 58.1%를 차지한다.

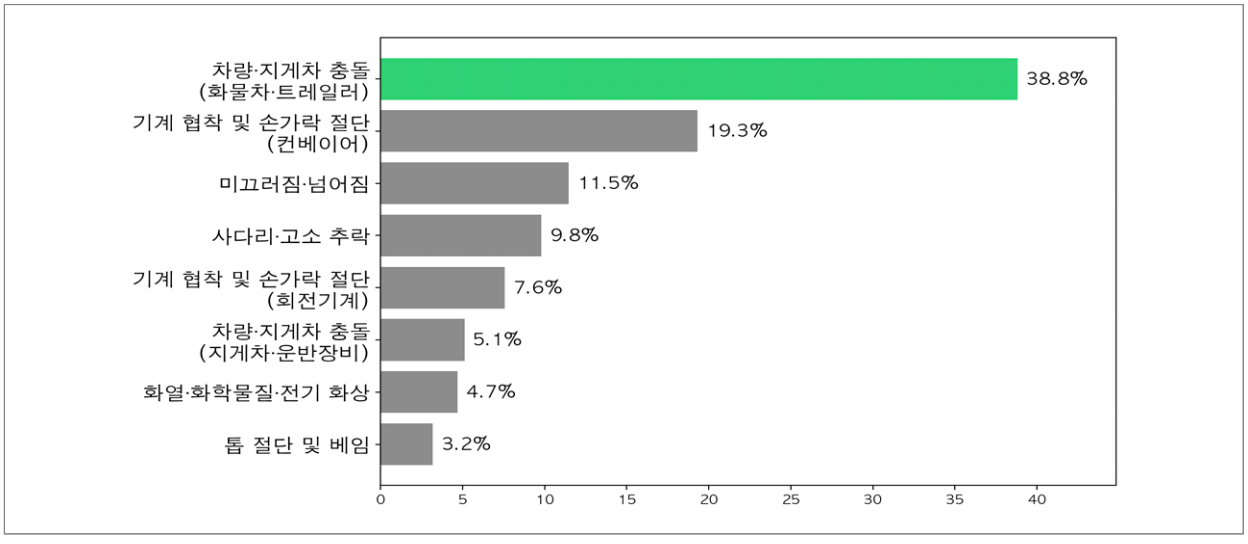


Fig. 2 클러스터별 산업재해 비율

차량·지게차 충돌 사고는 자재의 운반·적재·출하 등을 위해 사업장 내부에서 상시 운행되는 지게차의 작업 특성에 기인하는 것으로 판단된다. 지게차 운전자는 적재물로 인하여 전방 시야 확보 제약으로 보행 중인 작업자를 인지하지 못하는 경우가 발생하며, 이와 반대로 작업자 또한 시야 밖(측면 또는 후방)에서 접근하는 지게차를 인지하지 못하여 충돌로 이어지는 경우도 확인되었다. 「산업안전보건기준에 관한 규칙」은 이러한 사고를 예방하기 위하여 지게차에 전조등, 후미등, 헤드가드, 백레스트, 좌석 안전띠, 후진 경보기 등 5종의 안전장치 설치를 의무화하고 있다(제2관 지게차). 그러나 이들 안전장치는 보행자와 운전자 양측이 상호 시야를 확보하고 있음을 전제로 할 때에만 실효성을 가지므로, 사각지대에서 발생하는 충돌 사고를 근본적으로 예방하는 데에는 한계가 있다.

기계 협착 및 손가락 절단(컨베이어) 사고는 컨베이어 벨트가 저속 또는 고속으로 회전운동을 하는 과정에서 작업자의 손, 신체, 의류, 장갑 등이 말려 들어가면서 발생하는 것으로 확인되었다. 이러한 위험성은 실제 산업재해 사례를 통해서도 뒷받침 되는데, 최근 국내 SPC의 베이커리 공장에서 발생한 사망사고 역시 작업자가 설비 점검을 위해 윤활유를 주입하는 과정에서 컨베이어 벨트에 끼여 사망한 사고로 보고된 바 있다. 「산업안전보건기준에 관한 규칙」은 컨베이어에 역주행 방지장치, 비상 정지장치, 덮개, 트롤리 연결, 건널다리 설치 등을 의무화하고 있다(제11절 컨베이어). 하지만 오래된 설비는 안전장치가 없고, 현장에서 기본 안전 수칙과 방호조치가 충실히 이행되지 않는 데에서 기인하는 것으로 판단된다. 컨베이어 협착 사고는 기존 안전기준 준수만으로도 상당 부분 예방이 가능하다는 점에서, 예방을 통해 감소시킬 수 있다.

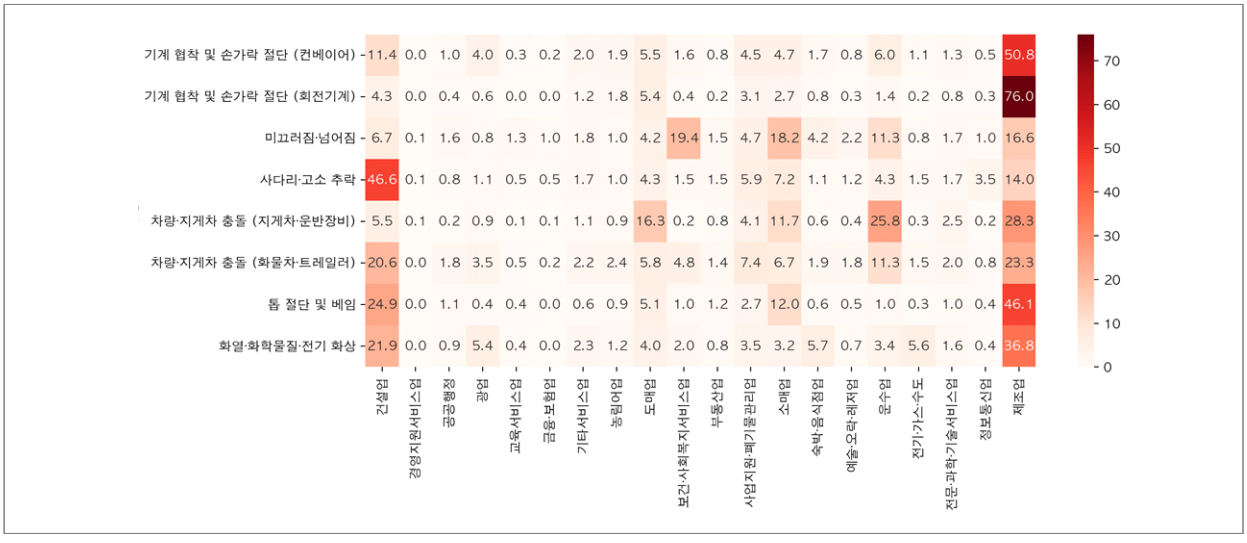


Fig 3. 산업별 사고유형 비율

둘째, 산업별 사고 유형을 분석한 결과, 기계 협착 및 손가락 절단 사고는 제조업에 집중되는 것으로 나타났다. 회전기계에 의한 협착 사고의 76.0%, 컨베이어에 의한 협착 사고의 50.8%가 제조업에서 발생하여, 제조업에서 기계 설비와 관련된 재해 발생 위험이 상대적으로 높은 것으로 보인다. 반면 사다리·고소 추락(46.6%), 톱 절단 및 베임(24.9%), 화물차·트레일러 충돌(20.6%)은 건설업에서 높은 비중을 차지하였다. 이는 고소작업과 중량물 운반이 빈번하게 이루어지는 건설 현장의 작업 특성이 반영된 결과로 보인다. 특히 사다리·고소 추락 사고는 건설업에서 가장 높은 발생 비율을 보였다. 이러한 사고는 근로자의 추락, 고소 작업대 전도, 붐(Boom) 파손 등 다양한 형태로 발생한다. 과거에는 비계나 사다리를 이용한 작업 중 발생한 사고가 주를 이루었으나, 최근에는 수직형 고소 작업대와 같은 장비의 결함이나 지면의 기울어짐으로 인해 전도되는 사고 빈도가 높은 것으로 보인다.

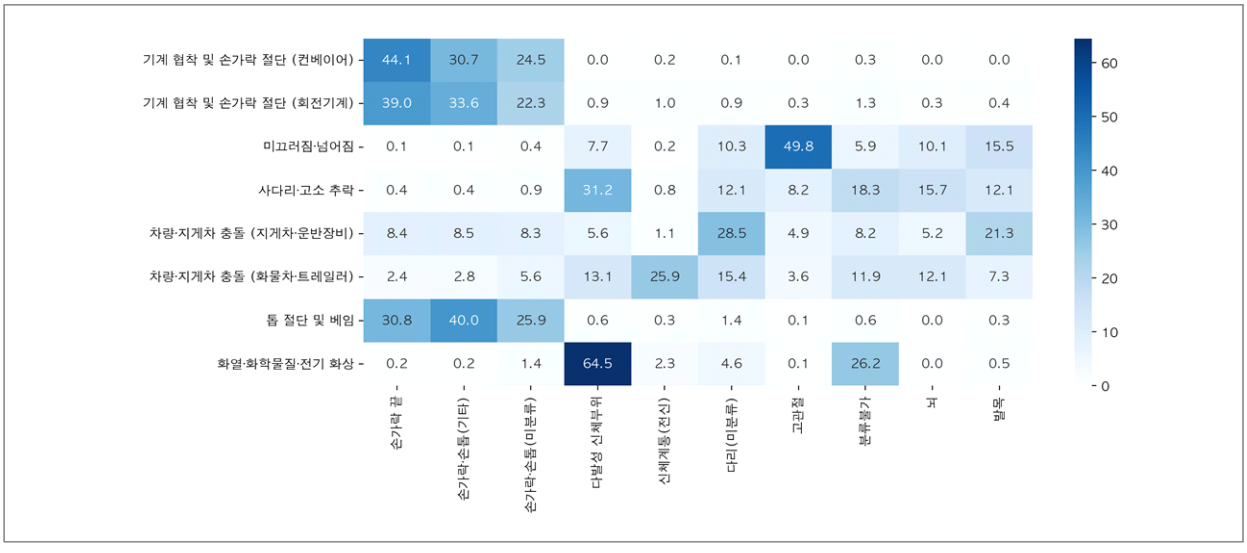


Fig 4. 사고 유형별 신체 부위 분포 비율

셋째, 사고 유형별 신체 부위 분포 비율을 분석한 결과, 신체 손상 부위에도 뚜렷한 차이가 나타났다. 컨베이어 및 회전기계에 의한 협착 사고는 손가락 끝(각각 44.1%, 39.0%)과 손가락·손톱 부위(각각 30.7%, 33.6%)에 손상이 집중되어 있었다. 특히 화학물질, 전기 화상의 경우 다발성 신체부위가 64.5%로 나타나 신체 전반에 피해가 발생하며, 미끄러짐의 경우 주로 고관절 손상에 집중된 다는 것을 알 수 있다.

3. 결론

본 연구는 OSHA 중대재해 데이터 105,318건을 대상으로 TF-IDF와 K-Means 군집분석을 수행하여 산업현장에서 반복적으로 발생하는 중대재해의 주요 사고 유형을 분류하였다. 분석 결과, 차량·지게차 충돌과 컨베이어 벨트 협착 사고가 전체 사고의 과반을 차지하는 대표적인 사고 유형으로 확인되었으며, 제조업에서는 기계 협착 사고가, 건설업에서는 추락 및 차량 충돌 사고가 상대적으로 높은 비중을 차지하는 것으로 나타났다. 특히 사고 발생 원인은 작업 환경, 장비 특성, 작업 방식 등 복합적인 요인에 의해 결정되므로, 사고 유형별 위험 특성을 반영한 예방 대책이 필요하다. 따라서 스마트 안전기술 도입 확대를 통해 작업장 내에서 근접 경보 센서, AI 영상인식 시스템 등을 도입하고, 컨베이어 등 위험 설비에는 자동 비상정지 및 작업자 감지 기능을 적용하여 사고 발생 가능성을 최소화해야 한다.

참고문헌

[1] Occupational Safety and Health Administration (OSHA), Severe Injury Reports, U.S. Department of Labor. https://www.osha.gov/severeinjury

[2] 국가법령정보센터, 「중대재해 처벌 등에 관한 법률」, https://www.law.go.kr