

# 텍스트 마이닝을 활용한 계절별 건설현장 추락사고 특징 분석

김준수<sup>1</sup> · 김병수\*  
<sup>1</sup>경북대학교 토목공학과

## Characteristics Analysis of Seasonal Construction Site Fall Accident using Text Mining

Kim, Joon-Soo<sup>1</sup>, Kim, Byung-Soo\*

<sup>1</sup>Department of Civil Engineering, Kyungpook National University

**Abstract :** The death rate of industrial accidents per 10,000 people in Korea is two to three times higher than that of major countries. Falling accidents at the construction site happened to have caused the most deaths. Analysis of existing research and measures by national institutions showed that the industrial accident management concentrated on falling accidents was insufficient and the seasonal safety management measures were not enough. There is thus the need for research that provides detailed and enough information on falling accidents. This study, therefore, aims to overcome the limitations of existing research and safety management accident response using a methodology that provides the necessary information for the prevention of fall accidents by deriving seasonal crash characteristics of the construction site. In order to provide enough information, 387 cases of seasonal construction site falling were collected, which describes the causal relationship of accidents. Text mining using principal component analysis and cluster analysis was carried out. The analysis showed that: In the spring, snowfall and unreasonable operation of equipment including lifts were the major cause. In summer, most accidents were caused by form, insufficient safety inspection, and installation work. In autumn, weather factors such as wind and typhoon were the cause. In winter, material transportation, exterior wall work, and ignore safety precautions were the cause of the crash.

**Keywords :** Industrial Accidents per 100,000 People, Cluster Analysis, Principal Component Analysis, Falling Accidents, Text Mining

### 1. 서론

#### 1.1 연구의 배경 및 목적

우리나라는 산업 전반적으로 안전에 대한 관심과 조치 및 대응은 활발하지만 2016년 기준으로 세계 주요 국가와의 사망사고만인율을 비교해보면 2~3배 높은 수준이다. 2016년 기준 산업재해 사망자 969명 중 건설현장에서 사망한 근로자수는 499명이며 산업재해 사망자수의 51.49%에 해당하여 건설 산업은 ‘고위험 분야 집중관리’로 구분이 되었다 (The Office for Government Policy Coordination, 2018).

그 중 고소작업으로 인한 ‘추락사고’는 근로자 사망사고 유형 중 가장 많은 비율(499명 중 281명, 56.3%)을 나타내고 있다.

이러한 문제를 해결하기 위하여 정부는 지난 2012년부터 건설공사 규모별로 일용근로자를 대상으로 기초안전보건교육을 이수한 근로자에 한해 건설현장에 투입하고 있지만 건설현장 재해자 수는 크게 감소되지 않았다(Park, 2018).

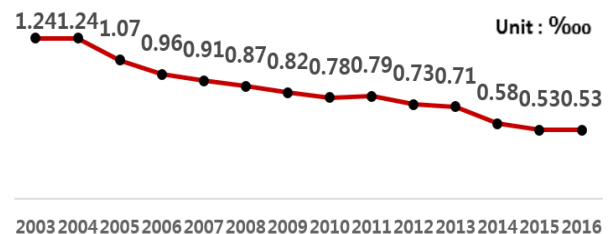


Fig. 1. Industrial accidents per 100,000 people at Construction Site in Korea

또한, 추락재해를 예방, 방지하기 위하여 안전캠페인 및 결의대회 개최, 방송 및 매체 홍보 등 다양한 예방사업을 전개 (Ministry of Employment and Labor, 2018)하고 있지만 다양하고 복잡한 공종과 유해하며 위험사항이 수시로 발생하는

\* Corresponding author: Kim, Byung-Soo, Department of Civil Engineering, Kyungpook National University, Deagu 41566, Korea  
E-mail: bskim65@knu.ac.kr  
Received January 12, 2019; revised -  
accepted February 12, 2019

건설업의 특성상 근로자들은 항상 산업재해에 노출되어 있다.

이러한 노력에도 불구하고 최저가낙찰제로 인한 저가수주, 효율성만 강조한 공사진행, 무리한 공사기간, 사업주의 근로자 안전의식 개선노력 미흡과 같은 제도적 문제점이 건설현장 사망사고율이 높은 원인으로 판단된다.

2011~2015년간 철도시설공단의 안전사고 분석결과를 보면, 저가낙찰현상이 전체의 77.8% 발생하였으며, 한국건설산업연구원 발표에서도 최저가 낙찰제 현장에서의 재해율이 16배나 높은 것으로 나타났다(Shin, 2015).

결국 정부는 이러한 행태를 바로잡고자 안전관리 의무 법제화 공표와 발주자 중심의 책임 선도모델을 정립하여 건설산업의 전반적인 안전기강 확립을 구축하고 있다(Ministry of Employment and Labor, 2018).

하지만 이러한 제도적 조치는 효과가 발현되기까지 상당 기간 소요되며 공사규모가 크지 않은 현장에는 효력이 닿지 않을 가능성이 높다.

결국 건설현장에서 공사규모와 관계없이 활용가능성이 용이하며 즉각적인 효과가 높은 추락사고 방지책이 필요하다고 판단된다.

본 연구는 공사기간이 긴 건설업의 특성을 고려하여 계절별로 추락사고 특징을 도출하여 추락사고를 예방하기 위한 정보를 제공하는데 목적이 있다. 이를 위해 2011년부터 2017년 까지 7년 동안 계절별로 건설현장에서 발생한 추락사고 인터넷 기사 387건을 웹크롤링하였다. 수집된 기사들은 형태소 분석을 한 후 추락사고의 요인으로 추정되는 변수(단어)들을 정량화한 텍스트 마이닝을 실시하였다. 그리고 텍스트 마이닝은 고차원이며 다변량 특징이 있기 때문에 유의미한 정보추출을 위해 주성분분석과 군집분석을 활용 후 계절별 건설현장 추락사고 특징을 도출하였으며 이 과정은 오픈소스 통계분석 프로그램인 R로 진행하였다.

### 1.2 연구의 범위 및 방법

빅데이터 분석의 특징은 개별적인 데이터는 독립적으로 존재하면 가치가 없지만 많은 데이터가 누적되었을 경우 의미 있는 소수의 데이터로부터 새로운 통찰과 패턴 등 유의미한 결과를 얻을 수 있다. 이와 같은 맥락에서 다소 적은 양의 데이터라 할지라도 유용한 정보가 도출된다면 빅데이터의 의미를 가질 수 있다고 판단된다. 또한 하루에 작성되는 양이 27,000여 건이며 사회 전반적인 이슈에 대한 사실과 의견을 담고 있는 뉴스의 경우는 대표적인 비정형 빅데이터다(Park & Oh, 2013). 그렇기 때문에 건설현장에서 발생한 계절별 추락사고 387건 뉴스를 텍스트 마이닝하여 도출된 결과는 계절별 추락사고에 대하여 설명할 수 있으며 특징이 나타날 수 있을 것으로 판단된다.

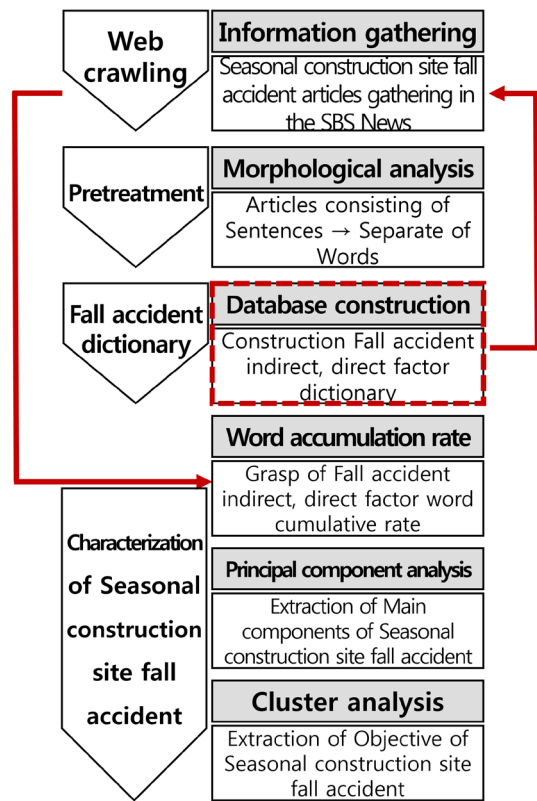


Fig. 2. Study Process

본 연구의 범위는 웹크롤링 구현이 용이한 HTML (Hypertext Markup Language)구조로 구성된 SBS (Seoul Broadcasting System)뉴스 웹페이지에서 2011년부터 2017년까지 생성된 건설현장 추락사고 뉴스 399건 중 웹크롤링으로부터 수집된 387건을 연구의 대상으로 하였다.

건설현장 추락사고 뉴스를 계절별로 수집하기 위하여 웹크롤링을 코딩하였다. 계절의 구분은 3월~5월까지지는 봄, 6월~8월까지지는 여름, 9월~11월까지지는 가을, 12월과 1월~2월은 겨울로 설정하였다.

수집된 계절별 건설현장 추락사고 기사는 문장으로 이루어져 있어 분석에 알맞지 않아 별도의 전처리가 필요하다. 형태소 분석을 실시하여 불필요한 문장구성을 제거한 후 단어들을 추출 하였다. 추출된 단어들 중 건설현장 안전사고에 직간접적인 영향을 준다고 판단되는 134개 요인을 선정 후 '추락사고 요인 단어장'을 구축 하였다.

구축된 '추락사고 요인 단어장'은 다시 형태소 분석된 387건의 계절별 건설현장 추락사고 뉴스로부터 직간접적인 추락사고 요인 단어들의 누적치를 산정하였다.

누적된 단어들 중 최빈값의 단어들만 추출 후 계절별 건설현장 추락사고의 특징을 도출하기에는 설명력이 부족하다고 판단되어 다변량 분석기법인 주성분 분석을 활용하여 추락사고 안전요인을 설명할 수 있는 성분들을 선정하였다.

하지만 주성분 분석결과를 해석할 경우 연구자의 주관적인 판단과 경험이 요구되어 왜곡된 결과를 얻을 수 있는 단점이 있다. 이에 주성분 분석을 통해 추출된 요인들을 대상으로 계층별 군집분석을 실시하여 객관성을 확보한 계절별 건설현장 추락사고 특징을 도출하였다.

### 1.3 기존 연구 고찰

국내외 건설 산업재해의 연구는 크게 재해원인 규명 후 개선대책 제시와 제도개선의 연구로 구분할 수 있다.

개선대책과 관련된 연구로서 Kwon (2009)은 건설현장에서의 안전 확보를 위한 안전관리의 실태를 체계적으로 분석한 후 개선방안으로 위험성관리 및 안전관리 계획서, 산업안전보건위원회 활성화, 맞춤형 안전교육, 자율 안전점검, 작업장 환경개선, 안전 활동 평가를 제시하고 있다.

Lee et al. (2011)은 법적용 미흡, 안전교육의 미흡, 노후 장비 등과 같은 건설현장의 재해 집중 원인분석을 통해 안전교육, 현장 관리 등의 개선방안을 제시하였다.

Oh (2016)은 건설현장의 안전사고 유발 요인 중 안전수칙 이행 책임전가의 의식을 예방하기 위하여, 현장관리자와 근로자의 책임전가의 의식을 분석하고 영향요인과의 상관성분석을 활용하여 책임전가의 의식을 감소시키는 위한 방안을 제시하였다.

Heinrich (1980)는 재해의 발생은 수많은 사고 요인들의 연쇄반응에 의한 결과로 나타나는 것이라고 분석하고 직접적 원인을 제거한다면 재해예방을 기대할 수 있을 거라 주장하였다.

Hinze (1996, 1998)는 건설현장의 안전사고에 의한 부상을 예방하기 위하여 부상유형별 안전사고 리스크를 분석 후 치명적인 건설재해 예방에 기여를 하였다. 그리고 사전사고 예방 안전프로그램 개발을 위하여 세부적인 사고요인 인과관계를 분석하였다.

Jimmie H. (2003)는 건설현장의 안전 활동 측정 및 분석을 통하여 재해율과 연관성이 높은 주요 항목을 제시하여 안전관리를 유도하는 연구를 진행하였다.

제도개선과 관련된 연구로서 Yun (2015)은 우리나라 건설현장의 안전관리시스템의 운용수준의 미흡한 점을 고려하고자 건설현장 관리감독자와 근로자의 안전의식을 제고시켜 건설재해 감소를 도모하기 위하여 건설안전보건교육의 개선방안을 모색하였다.

Kim (2015)은 공사금액 120억원 미만의 중소기업 건설현장의 재해현황 파악, 기초안전교육제도 운영현황과 법정안전교육의 실태 및 문제점을 분석하여 건설현장 근로자 안전교육의 실효성 제고방안을 제시하였다.

Koehn et al. (1995)은 개발도상국인 인도와 이미 선진화된 미국, 새롭게 선진국 대열에 들어선 대만의 안전관리 제

도를 비교 분석하였다. Charles J. and Richard J. (1995)는 많은 건설현장에서 환경적 문제가 발생하므로 국가적 차원에서 안전과 환경규제를 통합하여 하나의 지침을 제공함으로써 비용절감과 분쟁의 감소를 가져올 수 있도록 하는 방안을 제시하였다.

근로자의 고용정책과 더불어 사무를 총괄 및 산업재해 예방을 하는 고용노동부와 안전보건공단에서는 안전관리비 계상검토, 안전보건조치가 불량한 사업의 책임자 사법처리, 상하반기 건설현장 기획감독, 안전점검 기준표 개발 등 사업장을 대상으로 전반적인 산업재해를 관리하기 위한 제도적인 조치와 가이드라인을 제시하였다(Korea Occupational Safety & Health Agency, 2018; Ministry of Employment and Labor, 2017).

본 연구가 기존 연구와의 차별성을 가지는 것은 건설현장의 추락 사고를 집중적으로 다루며 체계적인 정보관리가 어려운 건설현장의 특성상 추락사고 데이터 수집이 어려운 문제점을 해결하기 위하여 뉴스 텍스트 데이터를 활용하여 추락사고의 특징을 도출하고 건설관리자에게 추락사고를 예방할 수 있는 정보를 제공한다는 점이다.

## 2. 이론적 고찰

### 2.1 R programing과 웹 크롤링

R 프로그램은 오픈소스이며 무료 소프트웨어로 통계 데이터 분석과 수정에 사용되는 프로그램 중 하나이다.

R 프로그램은 데이터 분석을 위한 다양한 패키지들을 제공한다. R은 통계와 그래픽스관련 기능을 확장할 수 있고, 고유 언어 내장 프로그램 기능과 더불어 수백 가지의 통계적 함수를 제공한다. 또한 데이터를 수집하거나 가공 할 수 있는 다양한 라이브러리를 제공하고 있다.

웹크롤러는 인터넷상에 존재하는 웹문서들을 추적하여 필요한 정보를 수집하는 기술을 말하며(Seo et al., 2013), 인터넷 검색시스템, 전자상거래 상품검색등 대부분의 인터넷 산업의 근간이 되는 핵심기술이다(Know et al., 2003). 웹크롤러는 웹서버를 순환하면서 각 홈페이지에 있는 수많은 정보를 수집하는 프로그램으로 사람이 홈페이지의 각 링크를 따라가서 정보를 얻는 반복적인 작업을 대신하여 프로그램이 자동으로 웹페이지의 내용을 분석한 후 URL을 하나씩 접속하여 정보를 수집한다.

### 2.2 텍스트 마이닝

데이터마이닝의 한 분야인 텍스트 마이닝은 대량의 텍스트를 효율적으로 활용하는 연구이다(Beak, 2017). 또한 정형 또는 비정형의 텍스트 데이터나 업무 중 발생된 문서,

E-mail, PDF (Portable Document Format)등에서 의미 있는 패턴을 도출한다. 도출된 패턴은 문서요약, 특성추출 등의 연구에 활용된다(Fabrizio Sebastiai, 2002; Koller & Sahami, 1997).

수집된 건설현장 추락사고 뉴스는 문장으로 이루어진 비정형 데이터이기 때문에 별도의 전처리 작업을 실시하여 추락사고에 직·간접적으로 영향을 미치는 단어를 추출 후 누적된 결과를 제시한다.

텍스트데이터의 가장 큰 특징은 고차원(High-dimensionality)이다. 이는 텍스트의 단락, 문장이 단어로 구성되어 있기 때문이다. 단어는 하나의 차원을 가진다. 그렇기 때문에 텍스트 데이터는 고차원으로 취급한다. 결국 텍스트 마이닝은 이런 변수들을 분석함으로써 문장의 성격을 파악하고 단락의 특징을 추정함으로써 통계분석을 통한 내용을 전반적으로 이해할 수 있다(Beak, 2017).

### 2.4 주성분 분석과 군집분석

주성분 분석은 상관성이 있는 다변량 데이터를 내포된 정보의 손실을 최소화하여 저차원의 데이터로 축약하는 방법이다. 즉, 복잡한 데이터 프레임을 단순화하는 다변량 통계기법이다.

주성분은 각 변수에 가중치를 연산 후 합성하여 선정된다. 이때 최초 변수가 가지고 있는 정보를 최대한 활용하여 변수간의 분산이 최대가 되도록 가중치를 연산한다. 원래의 변수가 가지고 있는 다수의 주성분이 존재하는 경우에는 각각의 주성분은 서로 상관이 없도록 하기 위하여 공분산이 0이 되도록 결정한다(Lee et al., 2013).

주성분 분석의 수학적 원리는 다음과 같다.

$$\begin{aligned}
 z_1 &= a_{11}x_a + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n \\
 z_2 &= a_{21}x_a + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n \\
 z_3 &= a_{31}x_a + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n \\
 &\dots \\
 z_m &= a_{m1}x_a + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n
 \end{aligned}
 \tag{1}$$

여기서,  $z_1 \sim z_m$ 은 주성분,  $x_1 \sim x_n$ 은 원래의 변수,  $a_1 \sim a_{mn}$ 은 가중치를 뜻한다(Lee et al., 2013).

군집분석은 각 객체의 유사성을 측정하여 유사성이 높은 대상 집단을 분류하고, 같은 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체간의 상이성을 규명하는 통계분석방법이다. 이때에 유사성은 측정한 변수들을 이용하여 모든 개체들 간의 거리(Distance) 또는 비유사성(Dissimilarity)을 통하여 계산된다. 군집분석에서 이러한 유사성은 유클리드(Euclid) 또는 마할로비스(Mahalanobis) 거리를 이용하여 측정되며 거리가 가까울수록 유사성이 높다(Michael Crawley, 2014).

## 3. 건설현장 추락사고 특징분석

### 3.1 데이터 전처리

SBS뉴스에서 웹 크롤링한 387건의 뉴스는 분석에 용이하지 않기 때문에 별도의 전처리 과정이 필요하다.

텍스트 마이닝은 통계를 기반으로 한 텍스트의 맥락과 주요 내용을 파악하여 패턴 및 의미를 파악하는 기법이다(Beak, 2017). 그렇기 때문에 문장으로 구성된 계절별 건설현장 추락사고는 분석에 불필요한 단어, 조사, 접속사 등을 제거해야 한다. 이를 제거하기 위하여 형태소 분석을 실시하여 추락사고와 직간접적으로 영향을 미치는 단어를 도출 후 이들의 누적치를 구하였다.

〈Table 1〉은 각 계절별 건설현장 추락사고의 직간접인 요인들을 누적치가 높은 순으로 나타낸 것이다. 〈Table 1〉은 텍스트 데이터의 특징인 고차원, 다변량의 형태를 나타내고 있으며 단순 빈도만으로는 주요 맥락과 내용을 파악하여 계절별 건설현장 추락사고 특징분석의 통계적 근거를 제시하기에는 어려움이 있다.

Table 1. Cumulative results of seasonal fall accident factor

Spring - 99 Words			Summer - 114 Words		
Words	Cumulative		Words	Cumulative	
	Value	Accumulation rate		Value	Accumulation rate
Installation work	49	5.36	Crane	58	5.33
Apartment Construction site	43	10.07	Vehicle	40	9.01
Lift operation	43	14.77	Inexperience driving	39	12.59
Vehicle	39	19.04	Installation work	38	16.08
Lack of inspection	36	22.98	Scaffolding work	37	19.49
...			...		
Protective sphere	1	100.00	Dropping	1	100.00

Autumn - 101 Words			Winter - 103 Words		
Words	Cumulative		Words	Cumulative	
	Value	Accumulation rate		Value	Accumulation rate
Crane	160	14.18	Crane	80	8.54
Structure construction site	58	19.33	Apartment Construction site	47	13.55
Vehicle	46	23.40	Vehicle	43	18.14
Building construction site	42	27.13	Building construction site	38	22.20
Factory construction site	35	30.32	Installation work	28	25.61
...			...		
Fence	1	100.00	Dropping	1	100.00

### 3.2 주성분 분석

고차원 다변량의 특성을 가진 텍스트데이터의 특성상 차원을 축소하여 데이터를 분석하는 방법이 필요하다.

그 중 차원을 축소하기 위하여 저차원에서 중요한 정보를 도출하는 다변량 통계분석 방법 중 하나인 주성분 분석 (Michael Crawley, 2014)을 활용하였다.

주성분 분석을 진행하기 위해서는 변수간의 독립성유지, 분석에 필요한 성분 수 결정, 분석의 유의미함의 증명이 중요하다. 이 연구에서는 각 변수들을 직교회전하여 변수의 독립성을 유지해주는 배리맥스 회전을 실시하여 변수간 독립성을 확보하였으며 전처리된 데이터를 상관성 분석을 실시 후 고윳값이 1 이상인 상관행렬의 차원수 성분 수가 각 계절별로 4로 산출되어 모든 계절의 주성분의 수는 4로 선정하였다. 그리고 제3의 변수를 제외하고 두 변수간의 순수한 상관관계(편상관계, Partial Correlations)를 나타내는 KMO (Kaiser - Meyer - Olkin, 0.5이상일 경우 양호함) 값으로 주성분의 유의미함 증명(Herry F., 1974) 하였다.

Table 2. Fall accident principal component analysis in spring

Division	PCA1	PCA2	PCA3	PCA4
Installation work	-0.03	0.78	0.16	0.17
Crane	0.85	-0.06	0.03	-0.03
Ignore of safety precautions	-0.10	0.86	0.01	-0.06
Lack of inspection	0.52	0.44	-0.32	-0.07
Apartment Construction site	0.36	0.04	0.62	0.11
Building construction site	-0.16	0.16	0.51	0.49
Concrete work	0.59	-0.36	0.13	-0.08
Collapse	-0.19	0.02	0.86	-0.06
Lift operation	-0.20	0.13	-0.20	0.84
Equipment Operation	0.93	0.09	-0.16	0.04
Snowfall	0.92	-0.08	0.09	0.00
Material transportation	-0.22	0.09	-0.25	-0.61
KMO	0.52			
Eigen value	4			
Proportion (%)	28	14	14	11
Cumulative proportion (%)	28	42	56	68

주성분 분석에 사용한 주성분 점수는 1에 근접 할수록 양의 관계이며 -1에 근접 할수록 음의 관계를 가진다. 즉, 양의 관계는 추락사고에 밀접한 영향을 미치기 때문에 0.80이상의 주성분 점수의 요인만 분석에 고려하였다.

봄의 건설현장 추락사고 제 1주성분은 <Table 2>에서와 같이 장비운용, 강설, 크레인운으로 도출되어 강설 시 무리한 장비와 크레인 운용이 추락사고를 유발함을 알 수 있다. 제 2, 3, 4주성분은 안전수칙, 붕괴발생, 리프트로 인하여 추락

사고가 유발함을 알 수 있었다. 그리고 위 주성분들은 봄 건설현장 추락사고의 68%를 설명하고 있다.

여름의 건설현장 추락사고 제 1주성분은 <Table 3>에서와 같이 장비와 기계 운전미숙, 건물 현장, 차량이 추락사고를 유발함을 알 수 있었다. 제 2주성분은 교량현장, 거푸집작업, 점검미비로 도출되었다. 제 3주성분은 크레인운용과 설치작업으로 도출되었다. 제 4주성분은 리프트, 비계작업이 추락사고를 유발함을 알 수 있다. 그리고 위 주성분들은 봄 건설현장 추락사고의 81%를 설명하고 있다.

Table 3. Fall accident principal component analysis in summer

Division	PCA1	PCA2	PCA3	PCA4
Crane	0.25	0.13	0.92	0.08
Inexperience driving	0.95	-0.05	-0.04	0.00
Scaffolding work	-0.07	-0.11	0.04	0.97
Concrete work	0.09	-0.14	0.79	0.40
Installation work	0.05	0.27	0.90	-0.16
Material transportation	0.53	-0.10	0.29	0.04
Lift operation	-0.05	-0.07	0.15	0.94
Bridge construction site	-0.09	0.82	-0.14	0.01
Vehicle	0.92	-0.17	0.23	-0.07
Building construction site	0.93	0.16	0.06	-0.04
Form work	0.22	0.86	0.20	-0.06
Lack of inspection	-0.03	0.80	0.42	0.05
Foot scaffolding	0.56	0.21	-0.10	0.58
Illegal and Violation	-0.13	0.76	-0.01	-0.11
KMO	0.51			
Eigen value	4			
Proportion (%)	24	20	19	17
Cumulative proportion (%)	24	45	64	81

Table 4. Fall accident principal component analysis in autumn

Division	PCA1	PCA2	PCA3	PCA4
Crane	-0.09	0.94	0.04	-0.15
Wind	0.99	0.03	0.01	-0.06
Construction site	-0.06	-0.06	0.78	0.11
Explosion	0.99	0.06	-0.01	0.03
Bridge construction site	-0.11	-0.04	0.45	0.82
Factory construction site	0.89	-0.02	-0.04	-0.14
Tower	0.20	0.93	-0.15	-0.05
Fire	0.57	-0.21	-0.57	0.10
Housing construction site	0.99	0.07	0.02	-0.04
Typhon	0.99	0.07	0.03	-0.03
Illegal and violation	0.26	-0.15	0.59	-0.20
Vehicle	0.84	0.01	0.08	-0.17
Concrete work	-0.09	-0.11	-0.28	0.86
Scaffolding work	-0.03	-0.26	-0.35	0.39
KMO	0.63			
Eigen value	4			
Proportion (%)	42	14	12	12
Cumulative Proportion (%)	42	56	68	80

가을의 건설현장 추락사고 제 1주성분은 <Table 4>에서와 같이 바람, 폭발발생, 주택현장, 공장현장, 태풍, 차량이 추락사고를 유발함을 알 수 있다. 제 2주성분은 크레인 운용과 타워, 제 3주성분은 주성분 점수가 0.80이상인 요인이 도출되지 않았으며 제 4주성분은 콘크리트 작업, 교량 현장이 추락사고를 유발함을 알 수 있다. 그리고 제 3주성분을 제외한 주성분들은 가을 건설현장 추락사고의 68%를 설명하고 있다.

겨울의 건설현장 추락사고 제 1주성분은 <Table 5>에서와 같이 구조물 현장, 안전수칙 미 준수, 자재운반이 추락사고를 유발함을 알 수 있다. 제 2주성분은 건물현장, 화재발생, 제 3주성분은 크레인 운용과 외벽이 추락사고를 유발함을 알 수 있다. 제 4주성분은 주성분 점수가 0.80이상인 요인이 도출되지 않았다. 그리고 제 4주성분을 제외한 주성분들은 겨울 건설현장 추락사고의 72%를 설명하고 있다.

본 주성분결과들의 결과는 계절별 추락사고 맥락과 특징을 도출하기 위해서는 분석자의 경험과 주관적 판단을 요구하게 된다. 이때 분석자의 따라 도출되는 결과 값이 상이할 수 있으며 왜곡될 가능성이 있다라고 판단되어 명확한 계절별 추락사고 특징도출에는 어려움이 있다.

Table 5. Winter Construction Site Fall Accident Principal Component Analysis

Division	PCA1	PCA2	PCA3	PCA4
Crane	0.33	0.10	0.85	0.02
Structure construction site	0.93	0.14	0.25	0.09
Vehicle	0.07	-0.11	-0.45	-0.68
Factory construction site	0.44	0.64	0.42	0.15
Building construction site	0.31	0.81	0.04	0.19
Material transportation	0.92	0.05	0.12	0.33
Concrete work	0.30	0.13	-0.30	0.77
Fire	-0.18	0.92	-0.06	0.07
Machine	0.70	0.19	0.09	-0.50
Lack of inspection	0.51	0.75	0.17	-0.27
Ignore of safety precautions	0.87	0.24	0.38	-0.01
Exterior wall work	0.18	0.02	0.86	-0.04
KMO	0.51			
Eigen value	4			
Proportion (%)	32	22	18	13
Cumulative proportion (%)	32	54	72	85

### 3.3 군집분석

표본의 집합 중 통계적으로 비슷한 표본을 동일한 트리에 가치를 묶어주며 다른 표본은 다른 가치에 묶어주는 군집분석은 직관적인 해석이 가능하여 분석자의 주관적인 판단이 요구되는 주성분분석 결과를 보완할 수 있다.

군집화하는 방법에는 공간상의 거리에 기초한 여러 가지 방법이 있는데 본 연구에서는 와드법(Ward's method)을 사용하였다. 와드법은 오차의 제곱합을 최소화시켜 정보의 손실량이 최소화가 되도록 하는 방법이다(Michael Crawley, 2014). 군집을 이루고 있는 요인들은 유사성을 지니고 있기 때문에 계절별 추락사고의 특징을 나타낼 수 있으며, 군집을 결정하는 기준은 Height(군집간의 거리)값의 범위가 작은 것으로 한다.

봄 추락사고의 군집은 <Fig. 3>에서와 같이 군집간의 거리를 감안하면 크레인, 안전수칙 미 준수, 붕괴·리프트, 장비·눈의 4개 그룹으로 나누어진다. 하지만 군집간의 거리가 짧은 붕괴·리프트, 장비·눈은 하나의 그룹으로 구분할 수 있으며 가장 많은 요인이 분포되어 봄 추락사고의 특징을 나타낼 수 있다. 즉, 봄 추락사고는 다른 계절과 비교하여 강설과 무리한 리프트와 장비운용이 추락사고를 유발하며 또한 붕괴사고로 인한 추락사고가 발생됨을 알 수 있다.

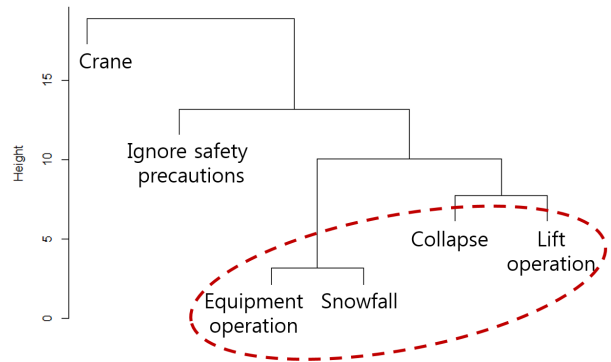


Fig. 3. Cluster analysis of fall accident in spring

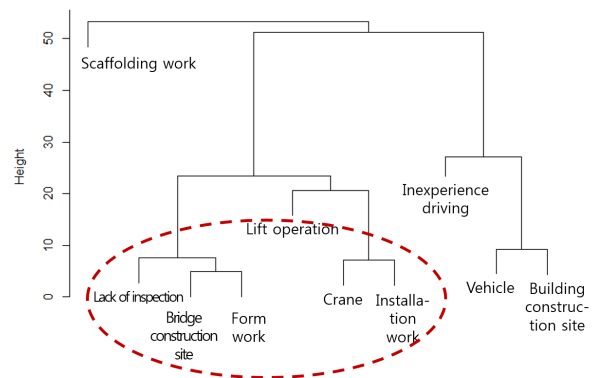


Fig. 4. Cluster analysis of fall accident in summer

여름 추락사고의 군집은 <Fig. 4>에서와 같이 ‘비계작업’, ‘리프트’, ‘점검미비’, ‘교량시공과 거푸집 작업’, ‘크레인인 설치작업’, ‘운전미숙’, ‘차량과 건물시공’ 7개의 그룹으로 나누어진다. 하지만 군집간의 거리가 짧은 ‘리프트, 점검미비, 교량시공, 거푸집 작업, 크레인, 설치작업’은 하나의 그룹으로

구분 할 수 있으며 가장 많은 요인이 분포되어 여름 추락사고의 특징을 나타낼 수 있다. 즉, 여름 추락사고는 다른 계절과 비교하여 거푸집 및 설치 작업으로 인한 추락사고와 점검미비로 인한 추락사고의 특징을 나타내고 있으며, 교량 시공의 장소적 특징과 무리한 크레인 운용의 특징을 나타내고 있다. 그리고 리프트의 경우는 봄과 공통적인 추락사고 요인으로 도출되었다.

가을 추락사고의 군집은 <Fig. 5>에서와 같이 ‘크레인’, ‘타워’, ‘교량과 콘크리트 작업’, ‘바람과 폭발’, ‘공장현장과 차량’, ‘주택과 태풍’ 6개의 그룹으로 나누어진다. 하지만 군집간의 거리가 짧은 ‘바람, 폭발, 주택, 태풍, 공장현장, 차량’은 하나의 그룹으로 구분할 수 있으며 가장 많은 요인이 분포되어 가을 추락사고의 특징을 나타낼 수 있다. 즉, 가을 추락사고의 특징은 바람과 태풍의 날씨 요인 특징이 나타나며 폭발사고가 추락사고에 영향을 주며 현장의 차량 사고가 추락사고와 관련되어 있는 것으로 나타났다. 또한, 장소적 특징으로는 주택과 공장 현장이 나타났다.

겨울 추락사고의 군집은 <Fig. 6>에서와 같이 ‘크레인’, ‘구조물 시공’, ‘자재운반 작업’, ‘안전수칙 미 준수와 외벽’, ‘건물시공과 화재발생’ 5개의 그룹으로 구분할 수 있다. 하지만 군집간의 거리가 짧은 ‘자재운반 작업, 안전수칙 미 준수, 외벽, 건물 시공, 화재 발생’은 하나의 그룹으로 구분할 수 있으며 가장 많은 요인이 분포되어 겨울 추락사고의 특징을 나타낼 수 있다.

즉, 겨울 추락사고의 특징은 자재운반 및 외벽작업에 의

한 추락사고와 안전수칙 미 준수와 현장의 화재 발생은 추락사고로 이어짐을 알 수 있다. 또한 장소적 특징은 건물시공의 특징이 나타났다.

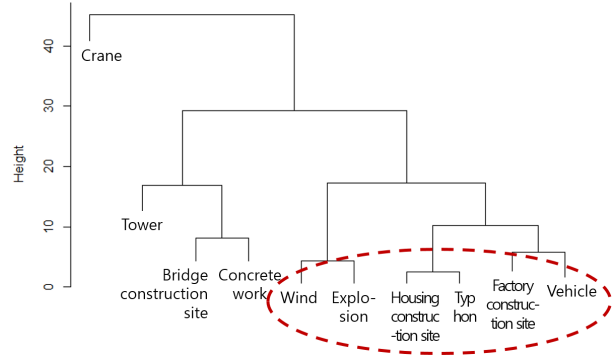


Fig. 5. Cluster analysis of fall accident in autumn

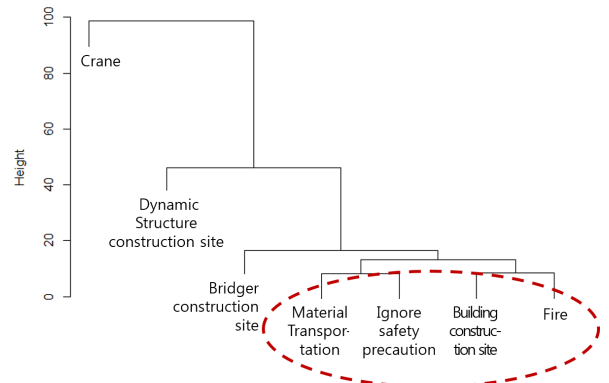


Fig. 6. Cluster analysis of fall accident in winter

Table 6. Characteristics of seasonal construction fall accident and status of data reduction

Classification	Pre-treatment	Principal component analysis	Cluster analysis
Spring	Installation work, Apartment construction site, Lift operation, Vehicle, Lack of inspection etc, 99 words	Crane, Ignore of safety precautions, Collapse, Lift operation, Equipment operation, Snowfall (6 Factors)	Equipment operation, Snowfall, Collapse, Lift operation (4 Factors)
Amount of data	100.00%	6.06%	4.04%
Characteristics	1. Snowfall, 2. Excessive equipment and Lift operation, 3. Fall accident due to collapsing		
Summer	Crane, Vehicle, Inexperience driving, Installation work, Scaffolding work etc, 114 words	Crane, Inexperience driving, Scaffolding work, Installation work, Lift operation, Bridge construction site, Vehicle, Building construction site, Scaffolding work, Lack of Inspection (9 Factors)	Lack of inspection, Bridge construction site, Scaffolding work, Lift operation, Crane, Installation work (6 Factors)
Amount of data	100.00%	7.89%	5.26%
Characteristics	1. Scaffolding and Installation work 2. Excessive Crane and Lift operation (Lift operation is a common cause of fall accident in spring), 3. Lack of inspection, 4. Bridge construction site		
Autumn	Crane, Dynamic structure construction site, Vehicle, Building construction site, Factory construction site etc, 101 words	Crane, Wind, Explosion, Bridge construction site, Factory construction site, Tower, Housing construction site, Typhoon, Vehicle, Concrete work (10 Factors)	Wind, Explosion, Housing construction site, Typhoon, Factory construction site, Vehicle (6 Factors)
Amount of data	100.00%	9.90%	5.94%
Characteristics	1. Wind and typhoon 2. Explosion and vehicle accident in construction site 3. Housing and factory construction site		
Winter	Crane, Apartment construction site, Vehicle, Building construction site, Installation work etc, 103 words	Crane, Dynamic structure construction site, Building construction site, Material transportation, Fire, Ignore safety precautions, Exterior wall work (7 Factors)	Material transportation, Ignore safety precautions, Exterior wall work, Building construction site, Fire (5 Factors)
Amount of data	100.00%	6.80%	4.85%
Characteristics	1. Material transportation and exterior wall work 2. Ignore safety precautions 3. Fire 4. Building construction site		

## 4. 결론

오늘날 건설산업의 특징 중 하나인 고소작업은 근로자에게 가장 위험한 산업재해와 가장 많은 사망자를 발생시키는 원인이다. 이러한 문제를 해결하기 위하여 정부와 고용노동부는 많은 대책을 내놓았지만 그 효과는 미미한 실정이다. 이에 본 연구는 공사기간이 긴 건설업의 특징을 고려하여 계절별 추락사고 특징을 주성분분석과 군집분석을 활용하여 텍스트 마이닝 하였다.

주성분 분석에서는 68%~85%의 설명력을 갖춘 결과 값을 얻었지만 분석자의 주관적인 판단으로 왜곡된 결과 값이 나타날 가능성을 배제하기 위하여 군집분석을 추가 실시한 결과 계절별 추락사고 특징을 나타내는 요인들이 명료하게 도출되어 <Table 6>에 나타내었다.

또한 주성분 분석과 군집 분석으로 도출된 결과는 원데이터량의 9.90%~6.06%, 5.94%~4.04%임을 알 수 있다. 이는 의미 있는 소수의 데이터로부터 유용한 정보를 얻는 빅데이터의 특징이 나타나 빅데이터 연구로서 의미가 있다고 판단된다.

텍스트 마이닝으로 도출된 계절별 건설현장 추락사고 특징은 추락사고 방지책 마련에 도움이 될 것으로 기대하며 본 연구방법은 추후 추락사고 뿐만 아니라 건설현장에서 발생하는 다양한 사고에도 적용될 수 있을 것으로 판단된다.

## 감사의 글

이 논문은 2017년도 경북대학교 연구년 교수 연구비에 의하여 연구되었음.

## References

Kwon, B.S. (2010). "A Study on the Improvement of Current Construction Safety Management System." Master Thesis, Incheon University, Korea.

Beak, Y.M. (2017). "Text Mining Using R." Hanul Academy.

Charles J. Kiber, and Richard J. Coble (1995). "Integrating Safety and Environmental Regulation of Construction Industry." *Journal of Construction Engineering and Management*, 121(1), pp. 95-99.

Enno Koekn, Rupesh K, Jothari, and Ching-Shing Pan (1995). "Safety in Developing Countries: Professional and Bureaucratic Problems." *Journal of Construction Engineering and Management*, 121(3), pp.261-265

Fabrizio Sebstiai (2002). "Machine Learning in Automated Text Categorization." *ACM Computing Surverys*, 34(1), pp. 1-47.

Heinrich, H.W. (1980). "Industrial Accident Prevention." McGraw-Hill.

Herry F. Kaiser (1976). "An index of Factorial Simplicity." *Journal of Psychometrika*, 39(1), pp. 31-36.

Hinze, J., Fredley, and J., Pedersen, C. (2000). "Identifying Root Causes of Construction Injuries." *USA Journal of Construction Engineering and Management*, ASCE, 126(1), pp. 52-56.

Hinze, Jimmie, and Gambatese, John. (1996). *Addressing Construction Worker Safety in Project Design*, Research Report Team of Washington; Texas at Austin University, 1996-06.

Jimmie Hinze (2003). "Fator That Influence Safety performance of Specialty Contractors." *USA Journal of Construction Engineering and Management*, ACSE, 129(2), pp. 159-167.

Kim, C.H. (2015). "A study on analysing basic safety-health training and improving effectiveness of safety training in small and medium-sized construction site." Master Thesis, Kyunghee University, Korea.

Koller D., and Sahami M. (1997). "Hierarchically classifying documments using very few words." *Proceeding of 1997 the Fourteenth International Conference on Machine Learning*.

Korea Occupational Safety & Health Agency (2018). "Press Release." 2018,02.

Lee, Y.S., Kim, D.Y., and Oh, H.U. (2013). "Assessment of CO<sub>2</sub> Emissions of Vehicles in Highway Sections Using Principal Component Analysis." *Korean journal of the korean society of civil engineers*, KSCE, 33(5), pp. 1981-1987.

Lee, Y.S., Kim, J.S., Choi, S.C., and Kim, C.E. (2011). "A study on the plan of improvement to reduce the occurrence of accidents of small construction site." *Korean Journal of Korea Management Engineering Society*, KMES, 16(3), pp. 99-113.

Michael Crawley (2014). "The R Book." Acorn publisher.

Ministry of Employment and Labor (2017). "Press Release." 2017,07.

Ministry of Employment and Labor (2018). "Press



- Release.” 2018.01.
- Oh, G.S. (2010). "Analysis of Relationship between Fatal Accidents Causes and Consciousness of Scapegoatism in Construction Site," *Proceeding of 2010 KIBC Conference*, Korea.
- Park, D.M., and Oh, S.W. (2016). "Big Data Analytics of Broadcasting News." *Broadcast Research Institute*, 28(1), pp. 7–51.
- Park, H.G. (2018). "An Empirical Study on the Efficient Operation of the Safety Monitoring Team in Construction Site." Doctor Thesis, Myungji university, Korea.
- Quan, C.H., Lee, Y.T., Kim, Y.J., and Lee, Y.D. (2003). "Design and Implementation of a High Performance Web Crawler." *Journal of Korea Society of Industrial Information System*, KSIE, 13(4), pp. 64–72.
- Seo, D.M., and Jung, H.M. (2013). "Intelligent Web Crawler for Supporting Big Data Analysis Services." *Journal of The Korea Contents Association*, JKCA, 13(12), pp. 575–584.
- Seo, J. (2015). *Data R Love: Analysis of Data for Introductory Using R*, Ar-Eum.
- Shin, J.Y. (2015). "Analysis and Prevention Measures for Construction Accidents." *SSANGYONG ENGINEERING & CONSTRUCTION Technology Research Institute*, 71, pp. 10–16.
- The Office for Government Policy Coordination (2018). "Countermeasures to Reduce Industrial Accident Deaths." 2018.01.
- Yun, J.M. (2015). "Improvement Measures of Safety & Health Education for Construction Disaster Reduction." Master Thesis, Hanbat National University, Korea.

---

**요약 :** 우리나라의 사망사고만인율은 주요 국가와 비교하면 2~3배 높으며 그 중 건설현장의 추락사고는 가장 많은 사망사고를 유발하고 있다. 기존연구와 국가기관의 조치를 살펴보면 추락사고에 집중된 산업재해 관리의 미흡과 계절별 안전관리 대책의 미비가 그 원인인 것으로 판단된다. 본 연구에서는 기존 연구 및 조치의 한계점을 해결하기 위하여 계절별 건설현장 추락사고 특징을 도출하고, 추락사고 방지책을 마련하는데 필요한 적정 정보를 제공하고자 한다. 정보의 제공을 위하여 사고발생 인과관계에 대하여 기술한 계절별 건설현장 추락사고 기사데이터 387건을 수집 후, 주성분분석과 군집분석을 활용한 텍스트 마이닝을 실시하였다. 분석결과 계절별 특징으로서 봄에는 강설, 리프트를 포함한 장비의 무리한 운용 등이 원인이며, 여름에는 거푸집, 점검미비, 설치작업 등이 원인으로 나타났다. 그리고 가을에는 바람과 태풍의 날씨요인 등이 원인이며 겨울에는 자재운반 및 외벽작업, 안전수칙 미 준수 등이 추락사고의 원인으로 나타났다.

**키워드 :** 사망사고만인율, 군집분석, 주성분 분석, 추락사고, 텍스트 마이닝

---