

데이터마이닝 기법을 활용한 건설 중대 재해요인 간 연관성 분석

임지선¹ · 한상욱² · 강영철³ · 강상혁^{4*}

¹인천대학교 건설환경공학과 석사과정 · ²한양대학교 건설환경공학과 부교수 · ³연세대학교 건축공학과 부교수 · ⁴인천대학교 도시환경공학부 부교수

Affinity Analysis Between Factors of Fatal Occupational Accidents in Construction Using Data Mining Techniques

Lim, Jiseon¹, Han, Sanguk², Kang, Youngcheol³, Kang, Sanghyeok^{4*}

¹Graduate Student, Department of Civil and Environmental Engineering, Incheon National University

²Associate Professor, Department of Civil and Environmental Engineering, Hanyang University

³Assistant Professor, Department of Architecture and Architectural Engineering, Yonsei University

⁴Associate Professor, Department of Civil and Environmental Engineering, Incheon National University

Abstract : Governments and companies are trying to reduce occupational accidents in the construction industry; however, the number of disasters are not decreasing significantly. This study aims to identify the correlation between factors affecting construction disasters quantitatively. To this end, 1,197 cases of serious disasters provided by Korea Occupational Safety and Health Administration (KOSHA) were analyzed using affinity analysis, one of the data mining techniques. The data from KOSHA were preprocessed and analyzed with variables of accident type, project type, activity type, original cause materials, sensory temperature, time of the accident, and fall height, and the association rules were derived for fall accidents and the others. For fall accidents, 64 association rules with lift ratios of 1.38 or greater were derived, and for the other accidents, 59 association rules with lift ratios of 1.54 or greater were derived. After analyzing the derived association rules focusing on the relationship among accident factors, this study presented the significance of applying the affinity analysis to address the study's limitations. The significance of this study can be found in that the correlation among factors affecting construction accidents is presented quantitatively.

Keywords : Construction Occupational Accident, Affinity Analysis, Association Rules, Data Mining, Construction Safety, Safety Management

1. 서론

1.1 연구의 배경 및 목적

건설업은 산업재해의 빈도와 강도가 높은 산업으로 알려져 있다. 안전보건공단에서 공개한 2019년도 산업재해 현황분석에 따르면, 건설업 관련 재해는 전체 108,434건의 산업재해 중 27,024건(24.92%)으로 제조업에 이어 두 번째로 높았다. 건설업 관련 사망 재해자는 2,020명 중 517명(25.59%)으로 전 산업 가운데 가장 큰 비중을 차지하였다(KOSHA, 2020).

정부는 100대 건설사들의 사망사고를 매년 20% 줄여 2017년도 대비 2022년까지 절반으로 감소시키는 것을 목표로 삼고 있다. 또한, 산업재해를 예방하고 노무자의 안전 및 보건을 유지·증진함을 목적으로 개정된 산업안전보건법이 2020년 시행되었고, 산업현장에서 사망사고 등 중대재해가 발생하였을 때, 기업 CEO와 임원, 대주주까지 최소 1년 이상의 징역을 집행할 수 있도록 하는 법인 중대재해처벌법이 2021년 1월에 제정되어 시행을 1년 정도 앞두고 있다.

많은 근로자가 건설업에 종사하고 있으며, 고용 지표 또한 큰 비중을 차지하고 있다. 건설업의 특성상 대형화, 고층화, 복잡화 등으로 인해 건설 재해는 단일의 원인으로 발생하기보다 다양한 원인이 복합적으로 작용하여 발생하므로 위험성이 높고 안전관리에 어려움이 많다. 따라서 건설 재해 예방을 위해서는 건설 재해에 영향을 주는 다양한 요인과 그 요인들 간의 규칙을 도출하는 것이 필요하다.

본 연구의 목적은 중대재해에 영향을 미치는 요인 간의

* **Corresponding author:** Kang, Sanghyeok, Department of Civil & Environmental Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, Korea

E-mail: lifesine@inu.ac.kr

Received April 12, 2021; **revised** July 17, 2021

accepted August 9, 2021

연관성을 정량적으로 규명하는 것이다. 이를 통해 현장 안전 관리에 도움이 될 수 있는 기초자료를 제시할 수 있다. 본 연구에서 다룬 재해요인에는 사고 발생 형태(accident type), 건설업종(project type), 작업내용(activity type), 기인물(original cause material), 체감온도(sensory temperature), 시간대(time), 그리고 추락 높이(fall height)이다.

1.2 연구의 범위 및 방법

본 연구는 안전보건공단에서 홈페이지를 통해 제공하는 ‘건설 중대 재해사례와 대책’에 수록된 2009년부터 2018년까지 1,197건의 재해사례를 분석하였다. 여기서 중대재해란 1) 1인 이상의 사망자가 발생한 재해, 2) 3개월 이상의 요양을 요하는 부상자가 동시에 2인 이상 발생한 재해, 또는 3) 부상자 또는 직업성 질병자가 동시에 10인 이상 발생한 재해를 말하며, 본 연구에서 활용한 데이터는 질병에 의한 사망사고를 제외한 사고에 의한 사망사고만을 포함한다.

본 연구는 다음과 같은 절차로 이루어졌다. 첫째, 건설 재해요인 분석 및 건설 안전 관리에 관한 선행연구를 고찰하였다. 둘째, 안전보건공단의 중대재해 사례를 분석이 가능한 형태의 데이터로 전처리하였다. 셋째, 안전보건공단에서 제공하는 데이터는 한계가 있으므로 재해요인이 될 만한 외부 변수를 추가하여 데이터 세트를 구성하였다. 넷째, 연관성 분석 기법(affinity analysis)을 적용해서 연관규칙(association rules)을 도출하고, 그 결과를 해석하였다.

2. 이론적 고찰

2.1 연관성 분석

데이터마이닝 기법 중 연관성 분석은 ‘무엇이 무엇과 관련이 있는지’를 밝혀내는 비교사학습 방법으로 데이터의 집합에서 발생하는 숨겨진 패턴을 찾는 분석기법이다. 유통업체가 고객의 장바구니에 담긴 아이템 간의 관계를 분석하면서 만들어진 분석기법이라 ‘장바구니 분석(market basket analysis)’이라고도 불린다(Shmueli et al., 2018). 이 기법은 시간의 순서에 따른 원인과 목적의 관계를 밝히는 것이 아닌 비목적성 기법이며, 두 개 또는 그 이상의 변수들 사이의 상호 연관성을 파악하는 것으로, 교사학습의 예측이나 분류의 방법과는 다르다.

연관규칙 분석의 결과는 수십 또는 수천 개가 발생하기 때문에 도출된 수많은 연관성 규칙 중 다양한 평가척도를 사용하여 규칙성이 높은 항목의 집합을 도출하는 것이 관건이다. 연관규칙 분석에서 사용되는 평가척도는 지지도(support), 신뢰도(confidence), 향상도(lift ratio) 등이 있으며 그에 대한 설명은 다음과 같다.

지지도는 데이터베이스에서 임의로 선택된 항목 집합(item set)이 조건부(A)와 결론부(B)의 모든 항목 집합을 포함할 확률이다(식 1). 즉, 지지도는 두 개의 항목 집합이 동시에 일어날 확률과 같다. 지지도는 상호 대칭적이며 규칙 ‘A⇒B’의 지지도와 규칙 ‘B⇒A’의 지지도는 같다.

$$Support = P(A \text{ and } B) \tag{1}$$

신뢰도는 A와 B의 모든 항목 집합을 포함하는 레코드의 수(즉, 지지도)와 B의 모든 항목 집합을 포함하는 거래의 수의 비율이다. 신뢰도는 임의로 선택된 거래가 A의 모든 항목을 포함한다고 할 때 B의 모든 항목도 포함할 조건부 확률이다(식 2). 신뢰도는 0에서 1 사이의 값을 갖는데, 1에 가까울수록 의미 있는 연관성을 갖는다고 볼 수 있다.

$$Confidence = \frac{P(A \text{ and } B)}{P(B)} = P(B|A) \tag{2}$$

향상도는 신뢰도를 두 개의 사건이 독립적이라는 가정하에 신뢰도(기준신뢰도, benchmark confidence)로 나눈 값이다. 이 지표는 어떤 규칙에 대하여 조건이 없을 때 기대되는 결과보다 조건이 추가될 때의 결과가 얼마나 향상되는지를 알려주는 척도이다. 따라서 향상도 값이 1보다 크면 A와 B 사이에 의미 있는 연관 관계가 있다고 볼 수 있다.

$$Lift \ Ratio = \frac{Confidence}{Benchmark \ Confidence} \tag{3}$$

결국, 지지도, 신뢰도, 향상도가 일정 기준을 충족시키는 연관규칙은 항목 집합 간에 밀접한 연관성을 가진 것으로 판단한다. 결과적으로 연관성 분석이란 여러 항목 집합 사이의 수많은 연관성 규칙 중에서 지지도, 신뢰도, 그리고 향상도에 근거해 일반화할 수 있는 의미 있는 규칙을 도출하는 방법이다.

2.2 선행연구 고찰

여러 연구자가 다양한 분석기법을 활용하여 건설 재해요인에 관한 연구를 수행하였다. <Table 1>은 데이터마이닝 기법 중 교사학습을 이용한 연구를 요약한 것이다.

데이터마이닝 기법은 여러 가지가 있으며, 각 기법은 저마다의 장·단점을 가지고 있다. 따라서 데이터의 유형, 특성, 그리고 분석의 목적에 따라 적용될 데이터마이닝 기법이 달라진다. 의사결정나무모델, 인공신경망, 판별분석 등의 데이터마이닝 기법은 지도학습의 일종으로 결과변수에 미치는 요인들의 기여 정도를 파악하는 데 활용된다. 한편, 본 연구는 요인들 사이의 연관성을 파악하고자 비지도 학습인 연관

Table 1. Studies using supervised learning techniques

Authors	Summary
Leem et al. (2005)	The performance of each algorithm was evaluated in order to select an analysis model optimized for the construction industry using industrial accident data. Performance evaluation of CHAID, CART, C4.5, and QUEST, which are representative algorithms of decision trees, was conducted to select an algorithm optimized for the construction industry.
Cho et al. (2017)	In this study, the accident type was set as a target variable based on the case data of construction accidents, and the relationship with the factors influencing it was established as a tree-model.
Kim et al. (2017)	Using construction safety accident cases, an artificial neural network model with the accident type as a target variable was constructed, and the prediction accuracy of the model was compared by comparing it with discriminant analysis.
Min et al. (2018)	Cases of occupational diseases that occurred at construction sites were used. This study found a model that can predict possible occupational diseases in the construction industry using case-based reasoning(CBR) and discriminant analysis.
Choi and Ryu (2019)	A tree model was built by applying Ada Boost, one of the CART models, to predict deaths and injuries by using data from fall accidents among the data of injured persons approved for industrial accident compensation insurance.

성 분석기법을 적용하였다. 연관성 분석은 강력한 비목적성 분석기법으로 이미 발생한 재해의 요인들 간의 연관성을 분석하기에 용이하다. 건설 재해는 다양한 요인이 복합적으로 연계되어 발생하므로 연관성 분석은 사고의 발생 요인 간의 관계를 이해하는 데 많은 도움이 된다.

〈Table 2〉는 연관성 분석을 이용하여 건설 재해의 요인 간의 관계를 규명하고자 한 연구들을 보여 준다. Shin et al. (2012)는 사망자와 중환자를 공사금액, 공사규모, 근속기간, 연령, 기인물, 발생형태, 공정률, 직종, 재해요일, 재해시간을 요인으로 하는 연관규칙을 도출하였고, Lee (2016)는 발생 형태, 기인물, 직종, 공정률, 공사규모의 변수를 활용하여 기후 조건에 따라 건설현장에서 사망자가 발생하고 있음을 발견하여 기후요소별 안전관리 방안 수립의 기초를 제안하였다. Lee (2020)는 발생형태별, 공사규모별, 공정률별로 나누어 발생형태, 공정률, 공사규모, 기인물, 직종, 연령, 요양기간, 근속기간, 재해발생시간, 시설물에 대한 변수를 활용하여 연관규칙을 도출하였다. Son and Ryu (2019)는 연관규칙을 통해 낙하물에 기인한 건설 사망, 부상 재해를 분석하였고, Lee (2020)는 소규모 건설현장에서의 사망재해를 연구하였다. 그러나 기존의 연구에서는 건설업종이나 근로자의 작업내용에 대한 정보가 변수에 포함되지 않았기 때문에 실질적으로 현장에서 안전관리 대책을 마련하는 데 한계가 있다.

본 연구에서는 발생형태, 건설업종, 작업내용, 기인물, 체감온도, 재해시간, 추락높이를 건설 재해요인으로 구성하여

연관규칙을 도출하였다. 건설업종, 작업내용, 체감온도, 추락높이와 같은 변수는 지금까지 연관성 분석을 활용해 분석된 사례가 없는 것으로 조사되었다. 이를 통하여 작업현장에서 근로자와 안전관리자가 주의해야 할 요인들에 대한 적절한 방안을 마련할 수 있는 정보를 제공할 수 있다.

Table 2. Studies using affinity analysis technique

Authors	Contents
Shin et al. (2012)	Using KOSHA data, the association was analyzed by dividing the fatality into the dead and the critically ill by setting the process rate, construction size, information on the injured, the type of occurrence, the cause of the accident, the size of the personnel, the type of job, age, service period, the day of injuries, and the time of disaster. As a result, the result was that the tendency of the fatal and critically ill accidents was similar.
Lee (2016)	Using data on deaths from construction in seven major cities in Korea, a rule of association for each climate factor (temperature, humidity, wind speed, precipitation) was derived. They conducted research on the type of occurrence, cause, job type, process rate, and construction scale (manpower) in fatal accidents.
Son and Ryu (2019)	Among KOSHA data, the accident caused by falling objects in the construction site was used to divide into two groups: death and injury, and the correlation between the working period, personal safety equipment, disease name, work contents, causes, and unstable behavior was analyzed to find the disaster rules and to find the key disaster factors through hierarchical clustering.
Lee et al. (2020)	The data of the accident death list of KOSHA were used to analyze the rules among the factors that occurred at the small construction site. They utilized variables such as scale, process rate, duration of service, occupation, age, gender, occurrence type, and person's life.
Lee (2020)	They analyzed the construction industry disaster data by type of occurrence, construction scale, and process rate. They conducted a correlation rule analysis with age, disaster time, causes, occupation, duration of medical treatment, duration of service, and facilities.

3. 건설 중대재해 사례 데이터

3.1 데이터 개요

본 연구에서는 안전보건공단에서 2009년부터 2018년까지 발간한 ‘건설 중대재해 사례와 대책’에 수록된 사례를 활용하였다. 약 10년간의 건설 중대재해 사례는 1,197건으로, 공사명, 발생일시, 재해 형태, 재해 정도, 소재지, 공사 규모, 재해 개요, 재해 상황도, 안전대책으로 구성되어 있다.

사례의 재해 개요, 안전대책은 주로 서술형으로 기술되어 있는데, 이를 통해 사고 발생 원인과 결과에 대한 정보를 파악할 수 있다. 다시 말해 해당 재해가 무슨 공사에서 어떤 작업을 하다가, 어떤 원인(기인물)에 의해 발생하였는지를 파악할 수 있으며 사고의 발생형태(재해유형)도 알 수 있다. 이

Table 3. Variable Description

Variable	Category
Accident type (12)	Falling, Electric Shock, Pressed/overturned, Caught in between, Slip down, Hit, Collapse, Struck by, Exposure to hazardous material, Amputation/cutting/ stabbing, Fire/explosion/burst, Others *Others' include oxygen deficiency, exposure to abnormal atmospheric pressure, exposure to abnormal temperatures, and contact. *Variable names: acci01~acci12
Project type (13)	Metal joiner's roof assembling construction, Mechanical construction, Painting masonry waterproof stone construction, Scaffold demolition construction, Water and sewage facility construction, Elevator ropeway construction, Maintenance management, Interior construction, Electric construction, Foundation pavement construction, Iron and steel structure construction, Steel reinforcement concrete construction, Others construction *Others' include gas facility construction, heating construction, industrial environment equipment construction, landscape planting facility installation construction, railroad track construction, and dredging construction. *Variable names: proj01~proj13
Activity type (25)	Welding flame cutting, Structure connection/ assembling/installation, Excavating, Installation/dismantling/maintenance of equipment, machinery, other construction work, Tamping, Painting, Laying, Plastering Rub finishing, Waterproof caulking, Plumbing wiring, Repair, Installing stone, Signal, Transportation, loading: lifting-unloading, Driving/ operation, moving, arrangement cleaning, Masonry, Piercing, Placing of concrete, Leveling, Pavement, Dismantling· demolishing, Checking-examination *Variable names: act01~act25
Original cause materials (18)	Hypothetical structure, Form · floor post, construction and mining machinery, main structural parts and components, stairs and ladder, Transportation, Other structures, Other equipment and machinery, End and opening, Scaffold walk plate, Slope and bedrock, Equipment and machinery parts and accessories, Transportation/ lifting of equipment and machinery, Hazardous and dangerous material, Working environment and natural environment, Electrical equipment and component, Portable machinery *Variable names: fact01~fact18
Sensory temperature (5)	Less than 5°C, 5°C - 15°C, 15°C - 25°C, Over 25°C, Non classified *Variable names: temp1~temp5
Time (9)	00~03, 03~06, 06~09, 09~12, 12~15, 15~18, 18~21, 21~24, Non classified *Variable names: time1~time9
Fall height (7)	Less than 2m, 2-3m, 3-5m, 5-10m, 10-20m, Over 20m, Non classified *Variable names: height1~height7

러한 재해 내용은 서술형으로 기술되어 분석이 가능한 형태로 전처리하는 과정이 필요하다.

3.2 데이터 전처리

본 연구에서는 건설 재해의 여러 요인 간의 연관성을 분석하기 위한 카테고리의 재분류를 위해 산업안전보건공단의 '산업재해 기록·분류에 관한 지침', 건설산업기본법시행령의 '건설업의 업종과 업종별 업무 내용'과 기존 선행연구를 바탕으로 분류체계를 수정·보완하는 형식으로 진행하였다.

기존 건설 중대재해 사례 데이터의 재분류를 통해 변수는 '재해 발생 시점, 재해 발생지역, 체감온도, 기인물, 재해 형태, 부상 정도, 추락 높이, 건설업종, 작업내용'으로 구분하였다. 도출된 변수 중 발생형태를 떨어짐과 떨어짐을 제외한 재해로 구분하였고, 떨어짐을 제외한 재해는 '건설업종, 작업유형, 기인물, 발생시간, 체감온도', 떨어짐 재해는 추락 높이를 추가하여 연관성 분석을 진행하였다. 그 세부 분류는 <Table 3>과 같다.

3.2.1 발생형태

건설 중대재해 사례의 발생형태는 과거 사용하던 용어가

혼재되어 있었기 때문에 '산업재해 기록·분류에 관한 지침'과 기존의 산업재해 현황분석 통계 결과를 참고하여 재분류하였다. 전체 중 떨어짐이 약 55% 정도 차지하고 있으며, 무너짐, 깔림·뒤집힘, 맞음 순으로 빈도가 높았다.

Table 4. Accident type

Accident type	Frequency (%)	Accident type	Frequency (%)
Falling	661 (55.2%)	Electric Shock	37 (3.1%)
Collapse	112 (9.4%)	Fire, explosion, burst	25 (2.1%)
Pressed, overturned	100 (8.4%)	Exposure to hazardous material	18 (1.5%)
Hit	100 (8.4%)	Slip down	12 (1.0%)
Caught in between	58 (4.8%)	Amputation, cutting, stabbing	11 (0.9%)
Struck by	56 (4.7%)	Others	11 (0.9%)

3.2.2 건설업종

건설산업기본법시행령에 따르면 현재 전문건설업종 체계를 28개로 나누고 있으나, 2022년부터 14개로 통합된다. 본 연구에서는 기존의 전문건설업종체계에 따라 분류하고 업종 분류가 불가능했던 전기공사는 따로 추가하고, 정보가 부

족한 사례는 기타로 분류하였다. 이후 개정된 시행령에 따라 통합하고 건수가 20건보다 적은 공사 또한 기타로 분류하였다. 기타를 포함한 13개의 업종 분류 중 철근·콘크리트 공사, 도장·습식·방수·석 공사, 금속·창호·지붕·건축물 조립 공사, 철·강·구조물 공사가 전체의 54.3%를 차지하였다.

Table 5. Project type

Project type	Frequency (%)	Project type	Frequency (%)
Steel reinforcement · concrete construction	247 (20.6%)	Mechanical construction	73 (6.1%)
Painting masonry · waterproof · stone construction	151 (12.6%)	Interior construction	61 (5.1%)
Metal · joiner's roof · assembling construction	145 (12.1%)	Maintenance management	51 (4.3%)
Iron and steel structure construction	107 (8.9%)	Water and sewage facility construction	35 (2.9%)
Others construction	100 (8.4%)	Electric construction	32 (2.7%)
Foundation · pavement construction	98 (8.2%)	Elevator · ropeway construction	20 (1.7%)
Scaffold · demolition construction	77 (6.4%)		

3.2.3 작업내용

건설 중대재해 사례집의 재해 개요, 안전대책 등의 내용을 통해 재해가 발생했을 때 작업자의 행동을 분류하였다. 그 결과 구조체 연결·조립·설치작업, 운반·상·하역 작업, 해체·철거작업, 이동 중에 발생한 재해가 전체의 60.6%를 차지하였다(Table 6).

3.2.4 기인물

안전보건공단에서 발표한 '산업재해 기록·분류에 관한 지침'에서 직접적으로 재해를 유발하거나 영향을 끼친 에너지원(운동, 위치, 열, 전기 등)을 지닌 기계·장치, 구조물, 물체·물질, 사람 또는 환경을 기인물로 정의하였다. 기인물 분류 기준과 분류코드의 대부분을 바탕으로 분류를 진행하였다. 분류 결과 건설업의 특성상 건축물·구조물 및 표면에 의한 재해가 약 60%를 차지하였다. 이는 '산업재해 기록·분류에 관한 지침'이 건설 산업에 최적화된 분류체계가 아닌 전체 산업에 적용 가능한 분류체계를 가지고 있기 때문이다. 따라서 해당 분류체계를 기본으로 중분류 기준에 의한 추가 분류를 진행하였다.

그 결과 비계 및 작업발판, 운반·인양설비 및 기계, 기타 건물·구조물, 건설·광산용 기계에 의한 재해가 전체의 50.7%를 차지하였다. 이때 기타 건물·구조물에는 철골빔, 트러스, 지붕, 탑, 전주 등을 포함하고 있다(Table 7).

Table 6. Activity type

Activity type	Frequency (%)	Activity type	Frequency (%)
Structure connection · assembling · installation	234 (19.5%)	Plumbing wiring	23 (1.9%)
Transportation, loading, lifting, unloading	201 (16.8%)	Driving · operation	18 (1.5%)
Dismantling, demolishing	174 (14.5%)	Repair	15 (1.3%)
Moving	123 (10.3%)	Laying	13 (1.1%)
Welding · flame cutting	65 (5.4%)	Installing stone	11 (0.9%)
Plastering · Rub finishing	40 (3.3%)	Masonry	10 (0.8%)
Painting	39 (3.3%)	Signal	10 (0.8%)
Arrangement · cleaning	39 (3.3%)	Excavating	10 (0.8%)
Other construction work	36 (3.0%)	Tamping	8 (0.7%)
Checking, examination	29 (2.4%)	Leveling	8 (0.7%)
Placing of concrete	29 (2.4%)	Piercing	7 (0.6%)
Installation · dismantling · maintenance of equipment, machinery	26 (2.2%)	Pavement	5 (0.4%)
Waterproof · caulking	24 (2.0%)		

Table 7. Original cause materials

Original cause materials	Frequency (%)	Original cause materials	Frequency (%)
Scaffold · walk plate	207 (17.3%)	Stairs and ladder	49 (4.1%)
Transportation/ lifting of equipment and machinery	141 (11.8%)	Electrical equipment and component	39 (3.3%)
Other structures	137 (11.4%)	Hypothetical structure	35 (2.9%)
Construction and mining machinery	122 (10.2%)	Portable machinery	20 (1.7%)
Main structural parts and components	119 (9.9%)	Hazardous and dangerous material	18 (1.5%)
Form · floor post	91 (7.6%)	Working environment and natural environment	10 (0.8%)
End and opening	69 (5.8%)	Other equipment and machinery	9 (0.8%)
Slope and bedrock	67 (5.6%)	Transportation	8 (0.7%)
Equipment and machinery parts and accessories	54 (4.5%)	General manufacturing & processing equipment / machinery	2 (0.2%)

3.2.5 체감온도

건설업은 주로 옥외작업으로 인해 여름철 높은 온도, 겨울철 낮은 온도, 습도, 강한 바람, 강우 등 기상적인 영향을 많

이 받는다. 이런 기상조건은 작업자의 집중력 및 생산성에도 영향을 미친다. 따라서 여름철에는 기온과 습도, 겨울철에는 기온과 풍속을 찾아 기상청에서 사용하고 있는 체감온도 산출 지표에 따라 체감온도를 구하여 변수로 추가하였다.

Table 8. Sensory temperature

Temperature category	Frequency (%)	Temperature category	Frequency (%)
Less than 5°C	342 (28.6%)	Over 25°C	158 (13.2%)
5°C - 15°C	317 (26.5%)	Non classified	6 (0.5%)
15°C - 25°C	374 (31.2%)		

3.2.6 발생시간

사례집에 작성된 사고 시점을 분류하였다. 근무 시간인 오전 9시부터 점심시간 전, 점심시간 이후 이른 오후, 늦은 오후, 저녁 시간대로 3시간씩 나누어 분류하였다.

Table 9. Time by 3 hours

Time category	Frequency (%)	Time category	Frequency (%)
00-03	7 (0.6%)	15-18	297 (24.8%)
03-06	2 (0.2%)	18-21	23 (1.9%)
06-09	148 (12.4%)	21-24	9 (0.8%)
09-12	421 (35.2%)	Non classified	12 (1.0%)
12-15	278 (24.8%)		

3.2.7 추락 높이

안전보건공단의 '산업재해 기록·분류에 관한 지침'에서 추락 높이는 '떨어짐' 재해가 발생한 당해 시설물의 높이를 정의하며, 2m 미만, 2m 이상 3m 미만, 3m 이상 5m 미만, 5m 이상 10m 미만, 10m 이상으로 분류코드를 구분하고 있다. 본 연구에서는 위 지침에 따라 분류하였을 때, 10m 이상에서의 재해 빈도가 전체 떨어짐 재해의 40%가 넘게 발생하여 10m 이상~20m 미만을 추가로 구분하여 최종 추락 높이에 대한 범주를 설정하였다.

Table 10. Fall height

Height category	Frequency (%)	Height category	Frequency (%)
Less than 2m	32 (4.8%)	10m - 20m	167 (25%)
2m - 3m	30 (4.5%)	Over 20m	112 (16.8%)
3m - 5m	98 (14.7%)	Non classified	44 (6.6%)
5m - 10m	185 (27.7%)		

4. 분석 결과

건설 사고는 하나의 원인으로 발생하지 않는다. 공사가 이루어지는 현장은 다수의 다양한 요인이 복합적으로 작용하

기 때문이다. 따라서 본 연구에서는 연관성 분석을 수행하여 건설 재해요인 변수 간의 연관규칙을 도출하였다. 이때 R Studio 4.0.2의 'arules' 패키지를 이용하여 분석을 수행하였다. 규칙의 연관성 정도를 평가하기 위하여 지지도, 신뢰도, 향상도 측도를 사용하였다. 연관성 분석은 최소 지지도와 최소 신뢰도의 평가척도 기준이 정해져 있지 않기 때문에 통계적 유의수준의 기준을 정하기 어렵다. 따라서 지지도, 신뢰도, 그리고 향상도의 수치를 조정해 가며 유용한 규칙을 도출해 낸다. 알려진 바와 같이 향상도가 1보다 큰 경우 연관성이 높은 의미 있는 규칙이라는 것을 나타낸다. 따라서 본 연구에서는 시행착오를 고려하여 연관성 분석을 하였고, 지지도 0.02 이상, 신뢰도를 0.50 이상으로 하는 규칙을 생성하여 향상도의 순으로 정리하였다(Table 11, 12).

전체 데이터를 대상으로 분석을 수행한 결과 지지도가 너무 낮아지고, 통찰을 도출할 수 없다는 것을 확인하였다. 이것은 떨어짐 재해가 전체 중 약 55%를 차지하여 대부분 연관규칙이 떨어짐 재해를 포함하고 있고, 과도한 아이템 수로 지지도가 상당히 낮아졌기 때문이다. 따라서 본 연구에서는 떨어짐 재해와 떨어짐을 제외한 재해로 구분하여 분석을 수행하였다.

4.1 떨어짐 재해의 연관규칙

떨어짐 재해 661건을 대상으로 연관성 분석을 진행하여 64개의 규칙을 발견하였다. 전반적으로 철근·콘크리트 공사에서는 기인물이 거푸집 및 동바리인 것과 추락높이가 20m 이상, 해체·철거작업에 관한 규칙이 많이 도출되었다. 또한, 철·강·구조물 공사에서는 기인물이 기타 건물, 구조물인 것과 추락높이가 10m 이상 20m 미만, 구조체 연결·조립·설치 작업에 관한 규칙이 다수 도출되었다. 금속·창호·지붕·건축물 조립공사에서는 기인물이 기타 건물·구조물이고 추락높이가 5m 이상 10m 미만에 관한 규칙이 도출되었고, 비계·구조물·해체공사에서는 해체·철거작업, 기인물이 비계 및 작업발판과 관련된 규칙들이 도출되었다.

도출된 연관규칙을 조건부와 결론부의 항목 집합 사이의 발생 가능성을 나타내는 향상도를 기준으로 나열하였을 때, 향상도의 값이 모두 1.38 이상의 값을 가지고 있어 항목 집합의 요인들 간의 연관성이 높다는 것을 알 수 있다. 향상도가 가장 높은 규칙은 "건설업종이 철근·콘크리트 공사이고 추락높이가 20m 이상이면 기인물이 거푸집 및 동바리이다"로 나타났다. 즉, 떨어짐 재해에서 철근·콘크리트 공사와 추락높이 20m 이상, 그리고 기인물이 거푸집 및 동바리라는 세 아이템 간에는 매우 높은 연관성이 존재한다는 것을 알 수 있다. 이는 철근·콘크리트 공사이고 작업이 20m 이상에서 이루어질 때는 거푸집 및 동바리에 의한 떨어짐 사고를

Table 11. Association rules of fall accident

No	Lift	Support	Confidence	Rules
1	8.62	0.0227	0.652	proj12, height6 → fact2
2	7.34	0.0227	0.556	proj12, act24 → fact2
3	6.61	0.0318	0.500	proj12, temp3 → fact2
4	6.48	0.0303	0.588	act24, fact10 → proj4
5	5.70	0.0227	1.000	act24, fact2 → proj12
6	5.70	0.0227	1.000	fact2, height6 → proj12
7	5.70	0.0318	1.000	fact2, temp3 → proj12
8	5.36	0.0242	0.941	fact2, time5 → proj12
9	5.31	0.0212	0.667	act2, fact7, time4 → proj11
10	5.24	0.0696	0.720	fact2 → proj12
11	4.99	0.0212	0.875	fact2, temp1 → proj12
12	4.82	0.0393	0.605	act2, fact7 → proj11
13	4.79	0.0227	0.833	act7, fact10 → proj3
14	4.67	0.0393	0.813	act9 → proj3
15	4.67	0.0393	0.813	act7 → proj3
16	4.62	0.0272	0.581	fact7, height5 → proj11
17	4.55	0.0242	0.571	fact7, time6 → proj11
18	4.32	0.0212	0.778	proj1, temp3, height4 → fact7
19	4.09	0.0212	0.737	proj11, act2, time4 → fact7
20	3.70	0.0393	0.667	proj11, act2 → fact7
21	3.64	0.0212	0.667	fact7, temp3, height4 → proj1
22	3.55	0.0242	0.640	proj11, time6 → fact7
23	3.45	0.0272	0.621	proj11, height5 → fact7
24	3.37	0.0303	0.606	proj11, time4 → fact7
25	3.28	0.0348	0.590	proj1, height4 → fact7
26	3.19	0.0212	0.583	fact7, time5 → proj1
27	3.08	0.0696	0.554	proj11 → fact7
28	3.06	0.0242	0.552	proj11, height4 → fact7
29	2.73	0.0303	0.500	fact7, temp3 → proj1
30	2.71	0.0287	0.792	proj4, act2 → fact10
31	2.65	0.0257	0.773	proj4, height5 → fact10
32	2.61	0.0212	0.700	proj11, fact7, time4 → act2
33	2.57	0.0272	0.750	proj4, time4 → fact10
34	2.51	0.0212	0.700	proj1, fact7, temp3 → height4
35	2.45	0.0303	0.714	proj4, act24 → fact10
36	2.40	0.0635	0.700	proj4 → fact10
37	2.18	0.0212	0.583	temp4, height5 → act2
38	2.18	0.0212	0.583	time4, temp1, height4 → act2
39	2.16	0.0227	0.600	fact1 → height4
40	2.15	0.0408	0.628	proj3, time4 → fact10
41	2.15	0.0287	0.576	proj11, time4 → act2
42	2.11	0.0393	0.565	proj11, fact7 → act2
43	2.09	0.0212	0.560	proj1, time6 → act2
44	1.98	0.0227	0.577	proj3, act7 → fact10
45	1.93	0.0227	0.517	proj11, height5 → act2
46	1.93	0.0272	0.563	act7 → fact10
47	1.92	0.0212	0.560	proj3, temp1 → fact10
48	1.92	0.0212	0.609	proj1, fact7, height4 → temp3
49	1.89	0.0318	0.525	fact7, temp3 → height4
50	1.83	0.0227	0.536	proj3, height5 → fact10
51	1.83	0.0393	0.510	fact7, time4 → height4
52	1.78	0.0272	0.500	time5, height4 → temp1
53	1.76	0.0257	0.515	proj3, temp2 → fact10
54	1.76	0.0893	0.513	proj3 → fact10
55	1.61	0.0212	0.583	act2, temp1, height4 → time4
56	1.57	0.0257	0.567	act24, height4 → time4
57	1.49	0.0212	0.538	proj11, act2, fact7 → time4
58	1.49	0.0439	0.537	act2, height4 → time4
59	1.43	0.0439	0.518	temp3, height4 → time4
60	1.43	0.0227	0.517	proj11, height4 → time4
61	1.41	0.0333	0.512	proj2 → time4
62	1.41	0.0393	0.510	fact7, height4 → time4
63	1.38	0.0212	0.500	act24, temp3 → time4
64	1.38	0.0439	0.500	act2, temp1 → time4

더욱 주의해야 한다는 것을 시사한다. 두 번째로 향상도가 높은 규칙은 “건설업종이 철근·콘크리트 공사이고 철거작업이면 기인물이 거푸집 및 동바리이다”이며, 세 번째는 “건설업종이 철근·콘크리트 공사이고 체감온도 15도 이상 25도 미만이면 기인물이 거푸집 및 동바리이다”이다. 이 규칙들을 통해 철근·콘크리트 공사에서 거푸집 및 동바리와 관련된 떨어짐 재해의 연관성이 높다는 것을 알 수 있다. 그다음으로 향상도가 높은 규칙에는 “해체·철거작업이고 기인물이 비계 및 작업발판이면 건설업종이 비계·구조물·해체 공사이다”와 “해체·철거작업이고 기인물이 거푸집 및 동바리이면 건설업종이 철근·콘크리트 공사이다” 등이 있었다. 연관성 분석은 변수들 사이에 완전히 독립적이어야 하는데 도출된 일부 규칙은 변수들 사이에 완전히 독립적이지 않은 결과로 본 연구의 한계점으로 보인다.

4.2 떨어짐을 제외한 재해의 연관규칙

떨어짐을 제외한 재해 536건을 대상으로 연관성 분석을 수행하여 59개의 연관규칙을 도출하였다. 떨어짐을 제외한 재해에서는 감전에 관한 규칙이 높은 향상도를 보였다. 철근·콘크리트 공사는 콘크리트 타설작업, 거푸집 및 동바리에 의한 무너짐과 관한 규칙이 많이 도출되었고, 지반·조성·포장 공사는 건설·광산용 기계, 사면 및 암반과 무너짐, 부딪힘과 관한 규칙이 많은 것으로 나타났다.

도출된 연관규칙들의 향상도는 1.54 이상으로 떨어짐 재해 분석에서 도출된 연관규칙과 마찬가지로 재해요인 간의 연관성이 높다는 것을 알 수 있다. 향상도가 가장 높은 규칙은 “기인물이 유해 및 위험물이면 발생형태가 화재·폭발·파열이다”로 향상도가 13.87로 나타났다. 두 번째로 향상도가 높은 규칙은 “발생형태가 감전이고 체감온도가 15도 이상 25도 미만이면 기인물이 전기설비, 부품이다”이고, 다음은 “발생형태가 감전이고 발생시간이 9-12시이면 기인물이 전기설비, 부품이다”로 나타났다. 향상도 순으로 나열하였을 때 2번째부터 7번째까지 감전과 전기설비, 부품에 관련된 규칙들이다. 이것들은 연관성이 높다고 예상 가능한 결과인데, ‘감전’과 ‘전기설비, 부품’이라는 변수 사이에 완전히 독립적이지 않아 향상도가 높은 것으로 판단된다. 그다음으로 향상도가 높은 규칙에는 “기인물이 휴대용 인력용 기계·기구이면 발생형태가 화재·폭발·파열이다”, “발생형태가 무너짐이고 건설업종이 상하수도설비공사이면 기인물은 사면 및 암반이다”, “기인물이 휴대용 인력용 기계·기구이면 작업내용 가공·취급작업이다”, 그리고 “건설업종이 철근·콘크리트 공사이고 작업내용이 구조물 연결·조립·설치작업이면 기인물이 거푸집 및 동바리이다” 등이 있었다.

Table 12. Association rules of non-fall accident

No	Lift	Support	Confidence	Rules
1	13.87	0.0205	0.647	fact14 → acci11
2	13.74	0.0317	1.000	acc2, temp3 → fact17
3	13.74	0.0224	1.000	acc2, time4 → fact17
4	13.68	0.0317	0.944	fact17, temp3 → acci2
5	13.00	0.0653	0.946	acc2 → fact17
6	13.00	0.0653	0.897	fact17 → acci2
7	12.42	0.0224	0.857	fact17, time4 → acci2
8	11.79	0.0205	0.550	fact18 → acci11
9	8.61	0.0205	1.000	acc7, proj5 → fact11
10	8.42	0.0205	0.550	fact18 → act1
11	7.19	0.0205	0.550	proj12, act2 → fact2
12	7.09	0.0373	0.833	acc7, proj10 → fact11
13	6.54	0.0354	0.500	acc7, proj12 → fact2
14	5.32	0.0243	0.765	proj4 → act24
15	4.82	0.0205	0.647	act4 → proj13
16	4.79	0.0205	1.000	proj5, fact11 → acci7
17	4.40	0.0224	0.632	acc7, fact4 → act24
18	4.36	0.0243	0.813	fact12, temp2 → acci6
19	4.22	0.0280	0.882	fact11, time4 → acci7
20	4.09	0.0485	1.000	act21 → proj12
21	4.09	0.0765	1.000	fact2 → proj12
22	4.09	0.0280	1.000	acc7, act21 → proj12
23	4.09	0.0205	1.000	act21, time5 → proj12
24	4.09	0.0224	1.000	act21, temp1 → proj12
25	4.09	0.0205	1.000	act2, fact2 → proj12
26	4.09	0.0205	1.000	acc3, fact2 → proj12
27	4.09	0.0354	1.000	acc7, fact2 → proj12
28	4.09	0.0261	1.000	fact2, time5 → proj12
29	4.09	0.0242	1.000	fact2, time6 → proj12
30	4.09	0.0280	1.000	fact2, temp1 → proj12
31	4.09	0.0205	1.000	fact2, temp3 → proj12
32	3.92	0.0336	0.818	fact11, temp3 → acci7
33	3.83	0.0373	0.800	proj10, fact11 → acci7
34	3.83	0.0224	0.800	fact11, time6 → acci7
35	3.55	0.0243	0.722	acc8, proj10 → fact3
36	3.49	0.0858	0.730	fact11 → acci7
37	3.42	0.0230	0.696	acc3, proj10 → fact3
38	3.32	0.0205	0.786	proj12, fact13 → act15
39	3.19	0.0224	0.667	fact11, temp1 → acci7
40	3.10	0.0205	0.647	fact11, temp2 → acci7
41	2.91	0.0354	0.543	fact12 → acci6
42	2.91	0.0243	0.591	acc8, temp1 → fact3
43	2.88	0.0243	0.500	acc8, fact3 → proj10
44	2.88	0.0205	0.500	fact11, temp3 → proj10
45	2.81	0.0261	0.667	acc6, temp3 → act15
46	2.76	0.0280	0.577	act21 → acci7
47	2.76	0.0280	0.577	proj12, act21 → acci7
48	2.66	0.0243	0.542	proj10, time6 → fact3
49	2.61	0.0224	0.545	act24, fact4 → acci7
50	2.58	0.0243	0.684	acc6, fact12 → temp2
51	2.56	0.0243	0.520	proj10, temp2 → fact3
52	2.37	0.0336	0.563	acc6, proj12 → act15
53	2.19	0.0280	0.536	acc7, time5 → proj12
54	2.14	0.0205	0.534	act24, time5 → proj12
55	1.89	0.0205	0.550	fact3, time5 → temp1
56	1.72	0.0243	0.500	acc7, fact3 → temp1
57	1.68	0.0373	0.571	act1 → time4
58	1.65	0.0261	0.560	proj10, temp2 → time4
59	1.54	0.0224	0.522	act24, temp2 → time4

5. 고찰

본 연구에서는 1,197건의 재해사례 데이터를 활용하여 떨어짐 재해와 떨어짐을 제외한 재해에 대하여 지지도는 0.02 이상, 신뢰도를 0.50 이상으로 하는 연관규칙을 도출하고 그 의미를 해석하였다. 분석 결과 충분히 의미 있는 규칙도 있으나 그렇지 않은 규칙들도 다수 도출되었다.

애초에 연관성 분석 기법은 구매자들의 장바구니에 담긴 아이템 간의 연관성을 유추하고자 개발된 기법으로 실물 아이템이 아닌 가상 아이템(virtual item)을 대상으로 분석이 수행될 때는 다양한 주의가 따라야 한다. 첫째, 연관성 분석을 적용하기 위해서는 아이템 간 독립성이 보장되어야 한다. 그러나 본 논문에서 분류한 변수들은 온전히 독립적이지 않을 수 있다. 예를 들어 철근콘크리트 공사라는 변수와 콘크리트 타설작업은 별개의 아이템으로 취급되었지만, 콘크리트 타설작업은 철근·콘크리트 공사에서 행해지는 작업으로서로 완전히 독립적이라고 보기 어렵다. 따라서 분석의 결과가 효과적으로 활용되기 위해서는 아이템의 독립성이 보장될 수 있도록 사전 검토가 철저히 이루어져야 한다. 이를테면 발생형태별, 건설업종별, 작업내용과 같이 건설현장에서 관리에 용이하게 요인을 나누고 발생 빈도가 유사한 요인들을 분석하면 분석의 정확성과 신뢰성을 향상시킬 수 있을 것이다.

둘째, 데이터 사례 수에 비해 요인 아이템 수가 너무 많지 않아야 한다. 요인별 아이템 수가 많아지면 높은 수준의 지지도를 만족하는 연관규칙을 도출하기 어렵고 빈도가 극단적으로 높은 아이템이 속해 있는 경우 신뢰성이 낮은 결과가 도출되기 때문이다. 예를 들어 <Table 11>의 규칙 9와 규칙 12를 살펴보면, 항목 집합이 거의 유사한데 규칙 9에는 재해 발생시간 변수가 추가되었음에도 향상도가 오히려 더 높은 것을 알 수 있다. 이는 떨어짐 재해는 구조체 연결·조립·설치 작업이고 기인물이 기타 건물·구조물일 때 건설업종이 철·강·구조물 공사인 사고가 발생시간과 연관되어 있다고 해석할 수 있다. 그러나 이것은 아이템의 특성이 반영되어 재해석되어야 한다. 왜냐하면, 9시부터 12시 사이에 발생한 재해가 다른 시간대에 발생한 재해에 비해 많기 때문에 나타난 결과로 보는 것이 더욱 합리적이기 때문이다.

건설 사고가 발생하였을 때의 사고 상황이 더욱 상세하게 기록될 필요가 있다. 소위 “Garbage in, garbage out”이라고 하듯이 신뢰성 없는 데이터를 아무리 근사한 기법으로 분석한다 해도 신뢰성 있는 결과를 기대할 수 없다. 근로자가 어느 작업 시 추락사고가 발생하여 사망에 이르렀을 경우, 어떤 공사에 어떤 작업을 하고 있었으며, 떨어진 높이는 지면으로부터 몇 m인지, 어디 부위를 다쳤는지, 근로자가 안전장

치를 제대로 착용하고 있었는지, 작업을 하고 있을 때 불안정한 요인은 없었는지, 그 외 부가적인 날씨, 풍속, 재해자의 상황 등의 자세한 상황을 서술할 필요가 있다. 아울러 이러한 내용을 작성할 때 통일된 언어가 사용되어야 하고, 데이터를 공개하여 안전사고 예방을 위한 연구가 촉진되어야 할 것이다.

6. 결론

정부와 기업이 재해를 줄이기 위해 지속적으로 노력하고 있음에도 불구하고 건설재해는 크게 줄어들지 않고 있다. 본 연구는 건설업에서 발생한 재해에서 요인 간의 연관성을 정량적으로 제시하기 위하여 연관성 분석을 수행하여 추후 근로자들과 현장관리자가 건설현장에서 적절한 안전대책을 마련할 수 있는 기초자료를 제시하였다.

‘건설 중대재해 사례와 대책’에서 재해사례를 활용하여 1,197건의 중대재해 데이터를 구축하고 데이터 전처리과정을 진행하였다. 해당 데이터를 떨어짐 재해와 떨어짐을 제외한 재해로 구분하고, 이를 바탕으로 연관성 분석의 평가척도인 지지도와 신뢰도가 일정 기준을 만족하는 연관규칙을 도출하였다. 본 연구에서 도출된 결과는 다음과 같다.

1) 떨어짐 재해를 분석하여 향상도가 1.38 이상의 연관규칙 64개를 도출하였다. 분석 결과 철근·콘크리트 공사이고 20m 이상에서 작업할 때 거푸집 및 동바리가 기인물인 재해가 서로 독립적이라고 가정했을 때에 비해 8.62배 연관성이 있다는 결과를 도출하였다.

2) 떨어짐을 제외한 재해를 분석하여 향상도 1.54 이상의 연관규칙 59개를 도출하였다. 분석 결과 상하수도설비공사이고 기인물이 사면 및 암반이면 발생형태는 무너짐인 재해는 지반·조성·포장 공사이고 기인물이 사면 및 암반이면 발생형태는 무너짐인 재해보다 지지도는 낮았지만, 연관성이 높았으며 세 아이템 사이에 서로 독립적이라고 가정했을 때보다 4.79배 연관성이 있다는 결과를 도출하였다.

3) 떨어짐 사고에서는 건설업종이 철근·콘크리트 공사, 기인물이 거푸집 및 동바리, 추락높이가 20m 이상, 해체·철거 작업에 관한 규칙이 많이 도출되었고, 떨어짐을 제외한 재해에서는 감전에 관한 규칙이 향상도가 가장 높았으며, 건설업종이 지반·조성·포장 공사, 기인물이 건설·광산용 기계, 사면 및 암반과 무너짐, 부딪힘과 관한 규칙이 많이 도출되었다.

본 연구 결과 충분히 의미있는 규칙도 있으나 그렇지 않은 규칙들도 다수 도출되었다. 연관성 분석 기법을 활용할 때는 분석에 활용한 아이템 간의 독립성이 온전히 독립적이어야 하는데 본 연구에서 분류된 변수가 온전히 독립적이지

않을 수 있다. 또한, 연관규칙을 도출할 때 높은 수준의 지지도를 만족하기 위해서는 데이터 사례 수에 비해 아이템 수가 많지 않아야 한다. 그러나 본 연구에서 활용한 데이터는 1,000여 건 정도로 데이터가 충분하지 않을 수 있다. 따라서 분석의 결과가 효과적으로 활용되기 위해서는 추후 이러한 한계점을 보완하기 위해 아이템의 독립성을 보장될 수 있도록 발생형태별, 건설업종별, 작업내용과 같이 건설현장에서 관리에 용이하게 요인을 나누고 발생 빈도가 유사한 요인들을 분석하면 분석의 정확성과 신뢰성을 향상시킬 수 있을 것이다. 그리고 건설사고 발생 시 사고 상황에 대한 기록을 자세하게 작성되어 데이터가 공개된다면 데이터 분석을 통한 건설 재해 예방을 위한 연구가 촉진될 것이다.

감사의 글

본 연구는 2019년도 과학기술정보통신부 한국연구재단 신진연구지원사업의 연구비 지원 (과제번호: NRF-2019R1C1C1009979)을 받아 수행되었습니다.

References

- Cho, Y.R., Kim, Y.C., and Shin, Y.S. (2017). "Prediction Model of Construction Safety Accidents using Decision Tree Technique." *Journal of the Korea Institute of Building Construction*, 17(3), pp. 295-303.
- Choi, J.H., and Ryu, H.G. (2019). "Analysis of Occupational Injury and Feature Importance of Fall Accidents on the Construction Sites using Adaboost." *Journal of the Arcotectiral Institute of Korea Structure & Construction*, 35(11), pp. 155-162.
- Kim, Y.C., Yoo, W.S., and Shin, Y.S. (2017). "Application of Artificial Neural Networks to Prediction of Construction Safety Accidents." *The Journal of the Korean Society of Hazard Mitigation*, 17(1), pp. 7-14.
- Korea Occupational Safety and Health Administration (KOSHA). (2016). "Guidelines on occupational accidents recorded and classified." KOSHA GUIDE G - 83 - 2016.
- Korea Occupational Safety and Health Administration (KOSHA). (2020). "2019 industrial accident status analysis."
- Lee, G.H., Lee, C.S., Koo, C.W., and Kim, T.W. (2020). "Identification of Combinatorial Factors Affecting Fatal Accidents in Small Construction Sites: Association Rule Analysis." *Korean Journal of Construction Engineering and Management*, KICEM, 21(4), pp. 90-99.
- Lee, G.Y. (2020). "Association Rule Mining Approach

- to Extracting Relationships of Accident Factors in Construction Sites.” MS Thesis, Pukyong University.
- Lee, K.H. (2016). “An Analysis on the Correlation Between Climate Factors and Fatal Accident in Construction Site.” MS Thesis, Semyung University.
- Leem, Y.M., and Choi, Y.H. (2005). “Selection of an Optimal Algorithm among Decision Tree Techniques for Feature Analysis of Industrial Accident in Construction Industries.” *Journal of the Korea Safety Management & Science*, 7(5), pp. 1-8.
- Min, Y.E., Cho, S.J., Ji, H.J., Yoo, W.S., and Shin, Y.S. (2018). “Predicting Model for Occupational Disease using Complex Analysis in Construction Site.” *The Korean Society of Science & Art*, 36, pp. 109-120.
- Ministry of Land, Infrastructure and Transport (MoLIT). (2020). “Enforcement Decree of the Framework Act on Construction Industry.” Law No. 31516.
- Shin, D.P., Son, C.B., and Lee, D.E. (2012). “Association Analysis of Construction Accident Attributes Causing Fatalities.” *Journal of the Architectural Institute Of Korea Structure & Construction*, 28(2), pp. 87-94.
- Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R., and Lichtendahl Jr. K.C. (2018). “Data Mining for Business Analytics”, Wiley.
- Son, K.Y., and Ryu, H.G. (2019). “Association Rules Analysis of Safe Accidents Caused by Falling Objects.” *Journal of the Korea Institute of Building Construction*, 19(4), pp. 341-350.

요약 : 정부와 기업이 건설업의 산업재해를 줄이기 위해 지속적으로 노력하고 있지만, 재해는 크게 줄어들지 않고 있다. 본 연구는 건설 재해에 영향을 미치는 요인들 간의 연관성을 정량적으로 규명하고자 하였다. 산업안전공단에서 공개한 중대재해 사례 1,197건을 대상으로, 데이터마이닝 기법 중 하나인 연관성 분석을 이용하여 연구를 수행하였다. 산업안전공단에서 제공하는 데이터와 외부 변수를 포함하여 재해 발생 형태, 건설업종, 작업내용, 기인물, 체감온도, 사고 시간대, 추락높이의 변수로 아이টে를 구성하여 분석하였으며, 떨어짐 재해와 그 외의 재해로 구분하여 연관규칙을 도출하였다. 떨어짐 재해의 경우 향상도가 1.38 이상인 64개의 연관규칙을 도출하였으며, 떨어짐을 제외한 재해의 경우 향상도가 1.54 이상인 59개의 연관규칙을 도출하였다. 도출된 연관규칙을 재해요인 간의 연관성에 초점을 두고 해석한 후, 고찰에서 연구의 한계와 건설재해 요인 간의 관련성을 파악할 때 연관성 분석 기법을 적용함에 있어 유의사항을 제시하였다. 본 연구는 건설 재해에 영향을 미치는 요인들 간의 연관성을 정량적인 수치로 제시하여 추후 근로자들과 현장관리자가 건설현장에서 적절한 안전대책을 마련하는 기초자료를 제공하였다는 점에서 의미를 찾을 수 있다.

키워드 : 건설 재해, 연관성 분석, 연관규칙, 데이터마이닝, 건설안전, 안전관리
