



Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
<http://scientiairanica.sharif.edu>



Research Note

Using association rules to investigate causality patterns of safety-related incidents in the construction industry

M. Mohajeri, A. Ardeshir*, and M.T. Banki

Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran.

Received 14 March 2019; received in revised form 26 April 2020; accepted 1 July 2020

KEYWORDS

Construction;
 Association rule;
 Occupational accidents;
 Safety management;
 Data mining.

Abstract. The objective of this study is to investigate causality patterns of safety-related incidents in the construction industry. Although there are many studies on finding cause-and-effect relationships in the accident database, retrieving useful knowledge from the last database and taking additional variables into account are needed. Therefore, the present study utilized the association rule method to investigate a significant scope of historical accident data in Iran's construction industry in the years 2014-2017. Based on results of association rules, the combination of worker's individual and behavioral factors and supervisory conditions are related to serious accidents. These results can provide practical insights for construction managers who need to be more concerned about the negative impact of the combination of some factors on serious construction accidents.

© 2022 Sharif University of Technology. All rights reserved.

1. Introduction

The construction industry is one of the most important causes of injury and fatality throughout the world because it is of dynamic and unpredictable nature [1,2]. Approximately 20-40% of all occupational fatal accidents occur in the construction industry, while construction employees comprise only 10% of the workforce [3]. Despite recent efforts and improvements in safety in this industry, injury and fatality accidents have not been significantly reduced [4]. Then, safety in construction management remains an open issue.

To reduce fatal and injury accidents and improve safety performance, researchers must be more concerned with identifying and analyzing factors influencing accident [5,6]. Heinrich (1959) stated that

if contributory factors in accidents were recognized, most of the accidents would be controllable [7]. Many researchers investigated factors influencing accidents [8] in order to find causality patterns of construction accidents [9]. These factors include safety management [10], environmental conditions [11], worker demographic characteristics [12], workers' behavioral characteristics [13], and type of accident, time of day, and month of the year [14]. However, it is crucial to recognize what kinds of combination of these variables can result in accidents. Investigating and analyzing historical accident data have always been important in recognizing the combination of contributory factors to accidents [7]. Therefore, gathering historical accident data may ensure an opportunity to not only find the combination of factors that result in construction accidents but also predict similar accidents in the future. The data mining methods are used to identify cause-effect relationships in the database [14–16]. These methods are suitable and applicable to analyze data related to safety occupational accidents to discover useful knowledge [17] and then, predict future events [18].

*. Corresponding author.

E-mail addresses: Mehdi.mohajeri@aut.ac.ir (M. Mohajeri); ardashir@aut.ac.ir (A. Ardeshir); bankimt@yahoo.com (M.T. Banki)

There are several data mining algorithms to apply to construction accident data, namely Decision Trees (DT) [19], Classification and Regression Tree (CART) [20], association rules [21], and Bayesian network [22–24]. The association rules method is widely applied to analyze occupational accidents to obtain cause-and-effect patterns from the accident database [25]. This method does not need the assumption that variables must be independent [14]. In the construction safety research done in the past decade, Cheng et al. (2010) used the association rules to identify cause-and-effect relationships in the accident database in Taiwan's construction industry [7]. Verma et al. (2014) applied the association rules mining approach to identify patterns of safety-related incidents at a steel plant [25]. Similarly, Amiri et al. (2016) used the association rules method for extracting patterns of 'falls' and 'falling objects' accidents in the Iranian construction industry [26]. Li et al. (2017) employed association rules to investigate the causality patterns of the non-helmet-use behavior of construction workers [9]. Shin et al. (2017) determined the incident patterns among serious injury and fatal accident data on Korean construction sites based on association rules [14]. However, because construction accident data are collected and recorded every year in databases, retrieving useful knowledge from the last database and taking additional variables into account are needed. Therefore, the present study utilizes the method of association rules to investigate a large scope of accident data of Iran's construction sites. This study can contribute to safety studies by identifying the combination of various contributory factors, which are related to construction accidents.

Another considerable contribution in this paper is that rules are ranked based on the risk assessment matrix. Previous studies have mostly employed risk assessment to assess hazard and rank activities [21]; however, in this study, a risk assessment method is applied to rank the rules. Valuable results are extracted from ranked rules. These results provide practical insights for the management to determine important rules and then, predict the reoccurrence of similar accidents.

2. Materials and methods

2.1. Data collection

According to Iran's Labor Ministry, each accident on construction sites should be reported. The accident report form involves worker information (age, work experiences, marital status, job, and education), type of accident, and other contributory variables associated with the accident. In this study, historical accident data in Iran's construction industry from 2014 to 2017 were obtained from Iran's Labor Ministry. Unfortunately, data have missing values for some variables.

Each raw dataset that had more than 30% missing values was removed. In this respect, the statistical method solved the problem of having less than 30% missing values. Although 18615 data sets of construction accidents were obtained from Iran's Labor Ministry between 2014 and 2017, 17846 of them were accepted in terms of missing values. They contained 95% of all accidents.

2.2. Variables

A total of 14 variables were recorded for each construction accident by an inspector. These variables are used in this research. After the pre-processing stage, the transformation step is required to prepare data for data mining. Therefore, variables are converted into categorical variables, as shown in the following:

- *Type of Accident (TA)*: This variable indicates the mechanism of the accident. It has been considered into five categories. These categories are "fall from height (TA1)", "fall of objects (TA2)", "struck by objects (TA3)", "caught in-between (TA4)", and "the other (TA5)".
- *Experience (EXP)*: This variable suggests the injured workers' previous experiences in this type of work. Experiences are categorized into five divisions namely $EXP1 \geq 1$, $EXP2: 1-4$, $EXP3: 5-10$, $EXP4: 11-20$, and $EXP5 > 20$ years of experience.
- *Time of Day (TIM)*: It indicates the time of day for each fault sample. Time of day is classified as 9–12 (TIM1), 12–15 (TIM 2), 15–18 (TIM 3), and 18–9 (TIM 4).
- *Age (AGE)*: It indicates the age of the workers when injured. Four classes are considered: 15–24 (AGE1), 25–34 (AGE2), 35–44 (AGE3), and 45–80 (AGE4).
- *Marital Status (MAR)*: Two classes are classified: single (MAR1) and married (MAR2).
- *Education (EDU)*: The following three groups are classified as elementary (EDU1), diploma (EDU1), and bachelor (EDU1).
- *Month of the Incident (MI)*: It indicates the month in which the accident occurs.
- *Supervision (SUP)*: This attribute indicates whether there is any supervisor (SUP1) or 'no supervisor present at the site (SUP2)' when the accident occurs on the construction site.
- *Accident Cause (CAU)*: The causes of the accident are classified into four groups based on the accident's reports: workers' errors (carelessness or negligence) (CAU1), non-use of PPE (CAU2), unsafe conditions and environment (CAU3), and combination of causes (CAU4).
- *Type of Injury (INJ)*: This variable points to the type of injury. The consequences of the accidents

are classified into three classes: “death or disability (INJ1)”, “serious injury (INJ2)”, and “minor injury (INJ3)”.

2.3. Data mining

Association rules method is a popular method for discovering the relation between variables in large data sets. In comparison with another data mining approach, the method is more suitable due to no need for introducing the dependent variable [27]. Also, the obtained rules can demonstrate casualty patterns for accident data. The apriori algorithm is the most popular algorithm for mining the association rules, as introduced by Agrawal et al. 1993 [28]. The apriori algorithm finds frequent item sets and generates the association rule from frequent items [25]. Therefore, this study applies this algorithm to explore the association rules among the construction accident data.

The apriori algorithm comprises two steps. First, an iterative search is carried out by scanning the database for frequent item sets. Second, strong association rules are produced from frequent item sets [28].

Let $I = I_1, I_2, I_3, \dots, I_m$ denote item sets and m be the total number of item sets. Then, in the association rules, a rule is defined as an implication of the form $X \Rightarrow Y$, with two conditions of $X \subseteq I$ and $Y \subseteq I$ where X and Y are two distinct subsets of I and $X \cap Y = \emptyset$. The variables X and Y are defined as the antecedent and consequent of the rule, respectively [29].

In the association rules, support, confidence, and lift are used as the three main parameters to discover association rules. Support defines the frequency of applying a certain rule to a given data set. Confidence, on the other hand, is characterized as the conditional probability of the occurrence of the consequent, given that the antecedent is true. Meanwhile, lift is an indicator of the strength of a rule over the probability of the co-occurrence of the antecedent and the consequent [16]. Support, confidence, and lift can be mathematically expressed as Eqs. (1), (2), and (3), respectively [16]:

$$\text{Support}(A \rightarrow B) = \frac{\#(A \cap B)}{N}, \quad (1)$$

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)}, \quad (2)$$

$$\text{Lift}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A) * \text{Support}(B)}, \quad (3)$$

where N is the number of transactions in the samples. If the lift value is larger than 1.0, the interdependence and correlation between the antecedent and the consequent are more significant. The higher the lift ratio, the more significant the rule [16]. The threshold

values for three indicators in this research were set as $S \geq 15\%$, $C \geq 65\%$, and $L \geq 1$.

2.4. Risk assessment

Risk assessment is traditionally defined by two variables:

1. The probability or frequency of risk occurrence;
2. The consequence or severity of risk occurrence [30].

Multiplier of the two parameters defines the level of risk, which is shown by risk matrix [31]. In the past research, risk assessment was often focused on the risk of activities [21][9], while this study targeted the rules. The aim of using the risk assessment matrix in this study is to rank the rules of predicting occupational accidents.

In this study, the probability of occurrence was calculated based on the combination of confidence and support indicators for each rule. In this work, according to the suggestion made by Munier (2013), the probability of occurrence was divided into four classes: frequent, likely, occasional, and unlikely [32]. Also, the severity of a risk occurring based on the reported accident data was categorized into 3 categories: death or disability, major injury, and minor injury. Then, by using the risk assessment matrix proposed by Li et al. (2017) [9], the rules were classified into four categories: extreme, high, medium, and low risk levels.

3. Results and findings

3.1. Overview of occupational accidents distribution

The results of the statistical analysis of data are shown in Table 1. The key results are shown as follows.

3.1.1. Cause of accident

The results of the statistical analysis demonstrated that most injuries were caused by unsafe behavior (71% of accidents), out of which 52.1% and 19.1% accounted for carelessness or negligence and non-use of PPE, respectively. The second most frequent cause of accidents resulted from the combination of causes (unsafe condition and unsafe behavior) that was calculated as 20.7%. Eventually, 8.2% of accidents were caused by unsafe conditions on sites including unprotected edge or openings.

3.1.2. Work experience

The work experience range of the injured group was from 0 to 30 years. The highest percentage (34%) belonged to the group under one year experience. It was found that work experience ranging from 1 to 5 years included 20% of accidents. These results demonstrated that new workers and less experienced workers were more risk-takers than others.

Table 1. Frequency distribution of variables.

| Factors | Level of factor | Description | Freq. | % |
|-------------------|--------------------|-------------|-------|-------|
| Age | Under 24 | AGE1 | 3913 | 21.92 |
| | 25–34 | AGE2 | 6820 | 38.21 |
| | 35–44 | AGE3 | 3819 | 21.40 |
| | Over 45 | AGE4 | 3294 | 18.45 |
| Marital status | Single | MAR1 | 5259 | 29.47 |
| | Married | MAR2 | 12587 | 70.53 |
| Experience | Under 1 year | EX1 | 6152 | 34.47 |
| | 1–5 | EX2 | 4474 | 25.07 |
| | 5–10 | EX3 | 3267 | 18.30 |
| | 10–20 | EX4 | 2690 | 15.07 |
| | Over 20 years | EX5 | 1263 | 7.08 |
| Education | Elementary | EDU1 | 14060 | 78.78 |
| | Diploma | EDU2 | 3377 | 18.92 |
| | Bachelor | EDU3 | 409 | 2.29 |
| Time of day | Time1 | TIM1 | 8087 | 45.31 |
| | Time 2 | TIM2 | 3801 | 21.29 |
| | Time 3 | TIM3 | 3349 | 18.76 |
| | Time 4 | TIM4 | 2609 | 14.61 |
| Type of accident | Fall from height | TA1 | 9581 | 53.7 |
| | Falling objects | TA2 | 2760 | 15.5 |
| | Struck by objects | TA3 | 2325 | 13.0 |
| | Caught in-between | TA4 | 1908 | 10.7 |
| | Other | TA5 | 1272 | 7.1 |
| Type of injury | Dead or disability | INJ1 | 3068 | 17.19 |
| | Major injury | INJ2 | 9439 | 52.89 |
| | Minor injury | INJ3 | 5339 | 29.91 |
| Cause of accident | Human errors | CAU1 | 9302 | 52.12 |
| | Non-use PPE | CAU2 | 3400 | 19.05 |
| | Unsafe conditions | CAU3 | 1480 | 8.29 |
| | Combine Causes | CAU4 | 3664 | 20.53 |
| Supervisor | Yes | SUP1 | 3812 | 21.36 |
| | No | SUP2 | 14034 | 78.64 |
| Total | – | – | 17846 | 100 |

3.1.3. Time of accident

A large proportion of accidents occurred between 9 and 12 (45.3%), followed by 12 to 15 (21.29%). According to the results, the frequency of accidents around meal break time is high.

Occupational accidents in the construction industry had the highest frequency from May to September

(51.4%) (Figure 1). These months take the whole summer in Iran and most construction activities are performed in these months.

The most common type of accidents was ‘falling from height’ (54%, 9581/17846), followed by ‘falling objects’ (15.0%, 2760/ 17846). ‘Struck by objects’ occurred (13%, 2325/17846) among all accident

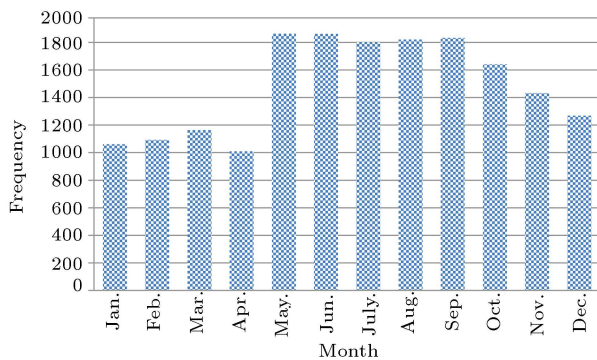


Figure 1. The month of occurrence of accidents.

types, closely followed by ‘caught in-between’ (11%, 1908/17846) (see Figure 2). The sources of injury in falling accidents involve the structure and construction facilities including roofs, openings, and scaffolding. Similarly, Ardeshtir and coauthors (2016) found that the risk of falling from height was the most probable risk incident and the most harmful accident in construction in Iran [33]. Also, falls, struck by objects, and caught in-between were the leading causes of workers’ fatalities on construction sites [34].

3.2. Association rules

Tables 2, 3, and 4 show association rules for death or disability as well as major and minor accident

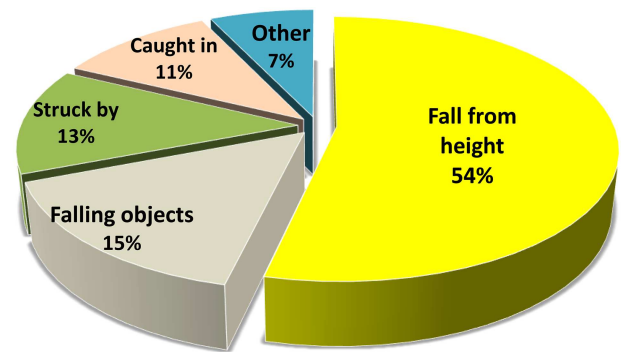


Figure 2. Type of accident in Iran in the period of 2014-2017.

categories, respectively, in which rules are sorted based on their confidence level. The best 15 rules for each class of the event are shown. These rules were chosen according to support, confidence, and lift. The lift values of all rules are larger than 1. In this study, “*type of accident*” was considered as the target variable because of its importance. Eight rules related to the type of accident were also obtained and shown. Based on the results, cause of accident, education, supervision, work experience, and marital status are the most frequent variables in rules.

Two of the best association rules for “*death and disability*” from “*type of injury*” show that workers aged

Table 2. The best association rules for death or disability (instances: 3068).

| Rule | Predictor 1 | Predictor 2 | Predictor 3 | Target Variable | Confidence | Support |
|------|-------------|-------------|-------------|-----------------|------------|---------|
| 1 | EDU1 | AGE4 | | MAR2 | 0.973 | 0.208 |
| 2 | EDU1 | SUP2 | AGE4 | MAR2 | 0.973 | 0.153 |
| 3 | AGE4 | – | – | MAR2 | 0.972 | 0.221 |
| 4 | TA1 | CAU1 | – | SUP2 | 0.943 | 0.207 |
| 5 | AGE4 | – | – | EDU1 | 0.941 | 0.214 |
| 6 | AGE4 | – | – | MAR2 & EDU1 | 0.916 | 0.208 |
| 7 | EDU1 | CAU1 | – | SUP2 | 0.906 | 0.302 |
| 8 | CAU1 | – | – | SUP2 | 0.904 | 0.363 |
| 9 | EDU1 | MAR2 | CAU1 | SUP2 | 0.893 | 0.209 |
| 10 | TA1 | EXP1 | – | EDU1 | 0.889 | 0.219 |
| 11 | CAU4 | – | – | SUP2 | 0.267 | 0.874 |
| 12 | MAR2 | TIM1 | – | EDU1 | 0.253 | 0.866 |
| 13 | SUP2 | MAR2 | CAU1 | EDU1 | 0.862 | 0.209 |
| 14 | EXP1 | SUP2 | – | EDU1 | 0.860 | 0.270 |
| 15 | EXP1 | – | – | EDU1 | 0.859 | 0.342 |
| 1* | SUP2 | – | – | TA1 | 0.818 | 0.473 |
| 2* | SUP2 | PPE | – | TA1 | 0.800 | 0.150 |
| 3* | SUP2 | EDU1 | MAR1 | TA1 | 0.690 | 0.150 |
| 4* | EDU1 | MAR1 | – | TA1 | 0.680 | 0.170 |
| 5* | SUP2 | AGE1 | – | TA1 | 0.664 | 0.150 |
| 6* | SUP2 | EXP1 | EDU1 | TA1 | 0.663 | 0.180 |
| 7* | SUP2 | EXP1 | – | TA1 | 0.643 | 0.202 |
| 8* | SUP2 | TIM1 | – | TA1 | 0.608 | 0.209 |

Table 3. The best association rules for major accidents (instances: 9439).

| Rule | Predictor 1 | Predictor 2 | Predictor 3 | Target variable | Confidence | Support |
|------|-------------|-------------|-------------|-----------------|------------|---------|
| 1 | EDU1 | AGE4 | – | MAR2 | 0.974 | 0.180 |
| 2 | AGE3 | EDU1 | – | MAR2 | 0.941 | 0.181 |
| 3 | SUP2 | AGE3 | – | MAR2 | 0.940 | 0.170 |
| 4 | EDU1 | CAU1 | TA1 | SUP2 | 0.923 | 0.251 |
| 5 | TA1 | CAU1 | | SUP2 | 0.923 | 0.311 |
| 6 | MAR2 | CAU1 | TA1 | SUP2 | 0.919 | 0.230 |
| 7 | TIM1 | CAU1 | – | SUP2 | 0.914 | 0.226 |
| 8 | EDU1 | CAU1 | – | SUP2 | 0.914 | 0.386 |
| 9 | CAU1 | – | – | SUP2 | 0.912 | 0.480 |
| 10 | EDU1 | CAU1 | MAR2 | SUP2 | 0.911 | 0.300 |
| 11 | EDU1 | CAU1 | TIM1 | SUP2 | 0.918 | 0.181 |
| 12 | CAU1 | EXP1 | | SUP2 | 0.915 | 0.160 |
| 13 | SUP2 | CAU1 | MAR2 | EDU1 | 0.852 | 0.300 |
| 14 | MAR2 | CAU1 | | EDU1 | 0.852 | 0.329 |
| 15 | MAR2 | CAU1 | TA1 | EDU1 | 0.852 | 0.213 |
| 1* | CAU2 | – | – | TA1 | 0.73 | 0.147 |
| 2* | EDU1 | – | – | TA1 | 0.676 | 0.537 |
| 3* | EDU1 | AGE2 | MAR2 | TA1 | 0.678 | 0.147 |
| 4* | EXP1 | SUP2 | – | TA1 | 0.673 | 0.173 |
| 5* | EXP1 | EDU1 | – | TA1 | 0.673 | 0.175 |
| 6* | SUP2 | MAR2 | AGE2 | TA1 | 0.668 | 0.150 |
| 7* | EDU1 | MAR2 | Tim1 | TA1 | 0.660 | 0.185 |
| 8* | EDU1 | SUP2 | MAR2 | TA1 | 0.700 | 0.325 |

Table 4. The best association rules for minor (instances: 5339).

| Rule | Predictor 1 | Predictor 2 | Predictor 3 | Target variable | Confidence | Support |
|------|-------------|-------------|-------------|-----------------|------------|---------|
| 1 | AGE3 | – | – | MAR2 | 0.957 | 0.205 |
| 2 | AGE3 | EDU1 | SUP2 | MAR2 | 0.955 | 0.153 |
| 3 | CAU1 | EXP1 | – | SUP2 | 0.876 | 0.163 |
| 4 | CAU1 | TA1 | – | SUP2 | 0.875 | 0.200 |
| 5 | CAU1 | AGE2 | – | SUP2 | 0.870 | 0.200 |
| 6 | CAU1 | | – | SUP2 | 0.873 | 0.479 |
| 7 | MAR2 | CAU1 | – | SUP2 | 0.866 | 0.331 |
| 8 | EDU1 | CAU1 | – | SUP2 | 0.866 | 0.355 |
| 9 | TIM1 | CAU1 | – | SUP2 | 0.863 | 0.218 |
| 10 | EDU1 | MAR2 | CAU1 | SUP2 | 0.855 | 0.266 |
| 11 | CAU1 | EDU1 | TIM1 | SUP2 | 0.850 | 0.160 |
| 12 | MAR2 | EXP1 | – | EDU1 | 0.822 | 0.171 |
| 13 | Time1 | MAR2 | – | EDU1 | 0.819 | 0.262 |
| 14 | MAR2 | CAU1 | – | EDU1 | 0.813 | 0.311 |
| 15 | MAR2 | TIM1 | – | EDU1 | 0.819 | 0.262 |

above 45 years AGE4 are more likely to be exposed to death and disability in accidents. Also, education, workers' behavior, and work experience are important factors contributing to deadly events (see Table 2). Rule 6 in Table 2 shows that according to the death or disability accident to which workers aged more than 45 years AGE4 are exposed, the probability that an individual who has an elementary level of education ELEM and is married MAR2 is 91.6%.

Rules 1* to 8* explain why death and disability accidents of falling from height occur on construction sites (see Table 2). These rules show that the site with no supervisor present SUP2 is the most frequent predictor variable for the target variable. Death or disability accidents due to falling from height TA1 occur on sites with no supervisor present in the location SUP2 and these events target workers with less than one year of previous work experience EXP1 or those with an elementary level of education ELEM.

According to Table 3, two main variables of rules involved in a major accident include a site with no supervisor SUP2 and workers' error CAU1, because these are observed 9 out of 15 times in rules.

The first best rule of falling from height shows that the non-use of PPE CAU2 is one of the most significant variables related to falling from height (Rule 1* in Table 3).

As shown in Table 4, the best association rules for “minor accident” from “type of injury” show that a site with no supervisor SUP2 and workers' error CAU1 are quite frequent in a minor accident. Rules 1, 2, and 12 indicate that a combination of individual workers including those aged 35–44 years AGE3 and those with less than one year of work experience, low level of education, and married status is more related to minor accidents. Also, Rules 9, 11, 13, and 15 indicate that the above combination has most

probably led to minor accidents from 9 to 12 TIM1. Given the minimum support, confidence, and lift thresholds, there is no rule related to the type of accident as a consequent or target variable.

3.3. Risk level assessment

This study provides a risk assessment matrix. First, confidence and support indices are classified into four intervals. Then, a matrix is created for them, as shown in Table 5. All rules are classified based on the occurrence probability of accidents, as shown in Table 5. The occurrence probability of the accident is divided into four categories: frequent, likely, occasional, and unlikely levels.

Two indices namely the probability and the severity of the accident are combined to obtain the risk level of each rule. The severity index is obtained from *Type of injury*, namely death or disability, major accident, and minor accident. Then, the risk matrix is drawn, as shown in Table 6. It represents the impact of cognitive factors on construction accidents. The risk is classified into four levels: extreme, high, moderate, and low.

According to Table 6, the rules are classified by risk level and represented in Figures 3 and 4. The

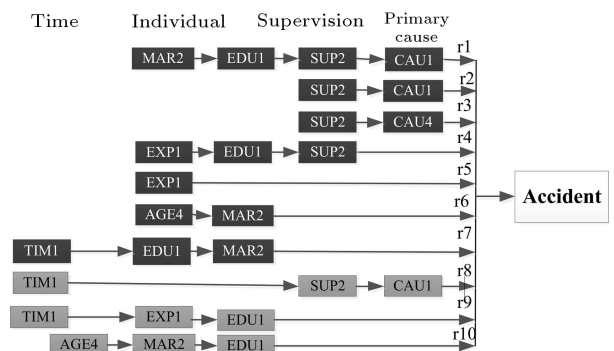


Figure 3. Association rules for construction accidents.

Table 5. Probability of accident.

| | | Support | | | |
|------------|----------|------------|------------|------------|------------|
| | | [20,22] | [22,26] | [26,30] | [30,100] |
| Confidence | [75,100] | Occasional | Likely | Frequent | Frequent |
| | [70,75] | Occasional | Occasional | Likely | Frequent |
| | [65,70] | Unlikely | Occasional | Occasional | Likely |
| | [60,65] | Unlikely | Unlikely | Occasional | Occasional |

Table 6. Risk assessment matrix.

| | | Probability | | | |
|----------|-------|-------------|----------|------------|----------|
| | | Frequent | Likely | Occasional | Unlikely |
| Severity | Death | Extreme | Extreme | High | Moderate |
| | Major | Extreme | High | Moderate | Low |
| | Minor | High | Moderate | Low | Low |

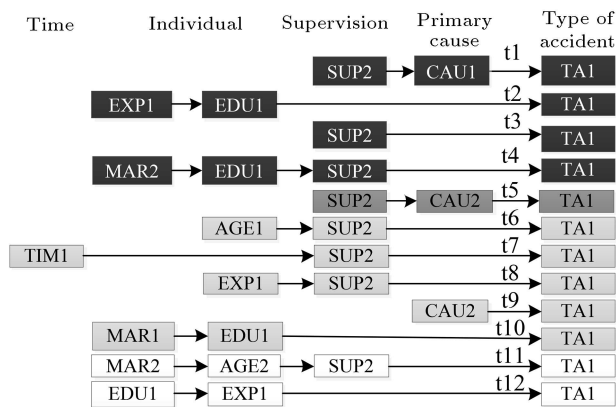


Figure 4. Association rules based on the type of accident.

first most useful rule for the extreme risk level in construction accidents includes the married workers, elementary level of education, the site with no supervisor, and workers' errors (Rule r1). It is implied that accidents are most likely to occur due to a combination of several factors, namely carelessness or negligence of workers, lack of supervision, and individual characteristics. Rules r2 and r3 state that an extreme risk level occurs on the site with no supervisor SUP2 and workers' error CAU1, or combinations of causes CAU4. Rule r5 implies that serious incidents are related to workers with less than one year of work experience EXP1. Rule r7 shows that the time of day from 9 to 12 TIM1, elementary level of education, and married workers are more likely to lead to accidents.

The first most useful rule for the high-risk level in the case of construction accident shows that the time of day from 9 to 12 TIM1, the site with no supervisor SUP2, and workers' errors CAU1 are more likely related to accidents.

It is also observed that association rules based on the *type of accident* are classified into risk levels (see Figure 4).

Figure 4 shows the following patterns: one of the most useful rules for an extreme risk level implies that on a site with workers' error CAU1 and no supervisor SUP2, it is more likely that falling from height happens (Rule t1). Rule t2 implies that falling from height is more likely related to the workers with less than one year of work experience and elementary level of education. Rule t3 shows that falling from height is the consequence of a site with no supervisor SUP2.

There is only one rule for high risk. This rule suggests that falling from height is also likely to occur on a site with no supervisor SUP2 and non-use of PPE.

Rules t6 and t8 for the medium-risk level imply that the workers less than 30 years old AGE1 or those with less than one year of work experience on sites with no supervisors most probably suffer the fall from height.

4. Discussion

In this investigation, contributory factors were identified and association rules were generated for accident databases in the construction industry. For this purpose, four groups of contributory factors were investigated, namely primary cause, supervision and inspection system, individual characteristics, and the time of day. The individual characteristics of workers refer to age, work experiences, marital status, job, and education. The results illustrate that the combination of factors results in the occurrence of incidents.

Visualization and classification of rules based on risk level bring about more interpretable and accurate results of association rules. The most important rules with extreme- and high-risk levels that led to the accident are shown in Figure 3. It can be seen that the combination of individual factors and supervision condition is more related to accidents. For instance, the combination of factors such as workers' errors who are married and have a low level of education as well as 'a site with no supervisor' could lead to accidents with the extreme risk level. There is no clear evidence to confirm exactly this pattern; however, Khosravi et al. (2015) stated that a combination of factors (e.g., individual characteristics and supervision) resulted in the occurrence of accidents on construction sites [35]. In addition, it was proved that supervisors had a direct influence on accident prevention [36]. Another most important rule demonstrated that new workers with no work experience had an extreme risk level that might cause accidents. This rule was supported by previous findings that demonstrated the positive link between inexperienced workers and unsafe behavior [37].

As shown in Figure 3, *time of day* and *age* of workers appeared in some important rules. The predominant time of day during which accidents happen is between 9 and 12. This result is in line with those of the previous research [38]. Also, most of the injured workers are above 45 years old because the possibility of serious injury increases with age [15].

The results obtained by *Type of accident* suggest that many fatal accidents related to fall from height can be prevented by simply the presence of safety supervisors and training the workers. The results also demonstrate that workers with less than one year of work experience and elementary level of education are more likely to fall from height. These results are similar to those obtained by Wang et al. who concluded that the construction workers with high knowledge and experience were more rational and less likely to take the risk [39]. The findings also show that workers' carelessness or negligence and non-use of PPE are considerable reasons for the occurrence of fall accidents on a site with no supervisor. Therefore, it can be inferred that a safety supervisor has a direct influence

on workers' safe behavior. This result coincides with past findings [40].

The main contribution of this research is the identification and ranking of causality patterns of safety-related incidents in the construction industry. To our knowledge, this research is the first study that used the risk assessment matrix to extract valuable patterns leading to accidents based on four groups of contributory factors of the accident, namely primary cause, supervision and inspection system, individual characteristics, and the time of day. This study found that negative effects of the combination of some factors result in the occurrence of serious accidents in the construction industry.

In addition, there are two main practical implications of the findings in this study. First, it is shown that inexperienced and uneducated workers as well as young workers are more likely to be the cause of and exposed to fall accidents. Therefore, construction managers need to consider these individual variables and their combinations based on obtained rules to prevent such accidents. For this purpose, project managers need to assign workers with a lower risk level in dangerous areas. For instance, managers need to keep an eye on new young workers, allowing them to work only in low-risk areas. Second, among the discovered rules, the supervisory condition (the site with no supervisor present SUP2) was the most frequent predictor variable for construction accidents. Occupational fatal accidents are reduced by assigning a permanent supervisor to construction sites.

In summary, the results of this research advise project managers to:

1. Assign workers with a lower risk level to dangerous areas;
2. Assign a safety supervisor to construction sites, especially in high-risk areas;
3. Train new employees and those who are inexperienced and uneducated;
4. Pay more attention to workers' behavior between 9 am and 12 pm.

5. Conclusions

The objectives of this study were to identify the factors contributing to the occurrence of occupational accidents by using the association rules method. For this reason, 17846 instances of historical accident data in Iran's construction industry between the years 2014 and 2017 were analyzed. Descriptive statistics were applied to the dataset to analyze occupational accident distribution. The results demonstrated that most injuries were caused by unsafe behavior (71% of all accidents) and the most common accident type was falling from height (54%, 9581/17846).

Association rules analysis was employed to identify cause-and-effect patterns in accident data. Then, 15 rules were identified by applying the limitation on the minimum support 15%, confidence 60%, and lift 1. Finally, by using a risk assessment matrix, the rules were classified into four risk levels. Based on results, the most useful rule showed that the combination of unsafe behavior of workers who are married and have a low level of education and a site with no supervisor could lead to accidents with the extreme risk level. In addition, the results of association rules related to *Type of accident* suggest that it can prevent many fatal accidents related to 'fall from height' through the presence of safety supervisor and training for workers. This paper provides practical insights for construction safety supervisors. They need to be more concerned about the negative effects of the combination of some factors in the occurrence of serious accidents. For future work, other data mining methods can be used for construction accident database to validate the results of cause-and-effect patterns among accident data.

Acknowledgments

The authors gratefully acknowledge the support provided by the Ministry of Labor and Social Affairs in Iran.

References

1. Teizer, J. and Cheng, T. "Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas", *Autom. Constr.*, **60**, pp. 58–73 (2015). DOI:10.1016/j.autcon.2015.09.003
2. Rashid, K.M., *BearWorks Coupling Mobile Technology, Position Data Mining, and Attitude toward Risk to Improve Construction Site Safety*, Missouri State University (2017).
3. Raheem, A.A. and Hinze, J.W. "Disparity between construction safety standards: A global analysis", *Saf. Sci.*, **70**(1), pp. 276–287 (2014). DOI: 10.1016/j.ssci.2014.06.012
4. Bureau of Labor Statistics US Department of Labor, *National Census of Fatal Occupational Injuries in 2014*, Bur. Labor Stat. US Dep. Labor (2015).
5. Yang, H., Chew, D.A.S., Wu, W., et al. "Design and implementation of an identification system in construction site safety for proactive accident prevention", *Accid. Anal. Prev.*, **48**, pp. 193–203 (2012). DOI: 10.1016/j.aap.2011.06.017
6. Sheykhsfard, A. and Haghighi, F. "Behavioral analysis of vehicle-pedestrian interactions in Iran", **25**, pp. 1968–1976 (2018). DOI: 10.24200/sci.2017.4201
7. Cheng, C.W., Lin, C.C., and Leu, S.S. "Use of association rules to explore cause-effect relationships

- in occupational accidents in the Taiwan construction industry”, *Saf. Sci.*, **48**(4), pp. 436–444 (2010). DOI: 10.1016/j.ssci.2009.12.005
8. Haslam, R.A., Hide, S.A., Gibb, A.G., et al. “Contributing factors in construction accidents”, In *Applied Ergonomics* (2005). DOI: 10.1016/j.apergo.2004.12.002
 9. Li, H., Li, X., Luo, X., et al. “Investigation of the causality patterns of non-helmet use behavior of construction workers”, *Autom. Constr.*, **80**, pp. 95–103 (2017). DOI: 10.1016/j.autcon.2017.02.006
 10. Tam, C.M., Zeng, S.X., and Deng, Z.M. “Identifying elements of poor construction safety management in China”, *Safety Science*, **42**(7), pp. 569–586 (2004). DOI: 10.1016/j.ssci.2003.09.001
 11. Fabiano, B., Curró, F., Reverberi, A.P., et al. “A statistical study on temporary work and occupational accidents: Specific risk factors and risk management strategies”, *Saf. Sci.*, **46**(3), pp. 535–544 (2008). DOI: 10.1016/j.ssci.2007.05.004
 12. Barlas, B. and Izci, F.B. “Individual and workplace factors related to fatal occupational accidents among shipyard workers in Turkey”, *Saf. Sci.*, **101**, pp. 173–179 (2018). DOI: 10.1016/j.ssci.2017.09.012
 13. Goh, Y.M., Ubeynarayana, C.U., Wong, K.L.X., et al. “Factors influencing unsafe behaviors: A supervised learning approach”, *Accid. Anal. Prev.*, **118**, June, pp. 77–85 (2018). DOI: 10.1016/j.aap.2018.06.002
 14. Shin, D.-P., Park, Y.-J., Seo, J., et al. “Association rules mined from construction accident data”, *KSCE J. Civ. Eng.*, **0**(0), pp. 1–13 (2017). DOI: 10.1007/s12205-017-0537-6
 15. Sanmiquel, L., Rossell, J.M., and Vintó, C. “Study of Spanish mining accidents using data mining techniques”, *Saf. Sci.*, **75**, pp. 49–55 (2015). DOI: 10.1016/j.ssci.2015.01.016
 16. Doostan, M. and Chowdhury, B.H. “Power distribution system fault cause analysis by using association rule mining”, *Electr. Power Syst. Res.*, **152**, pp. 140–147 (2017). DOI: 10.1016/j.epsr.2017.07.005
 17. Liao, C.W. and Perng, Y.H. “Data mining for occupational injuries in the Taiwan construction industry”, *Saf. Sci.* (2008). DOI: 10.1016/j.ssci.2007.04.007
 18. Khaksar, H. and Sheikholeslami, A. “Airline delay prediction by machine learning algorithms”, *Sci. Iran.*, **26**(5), pp. 2689–2702 (2018).
 19. Mistikoglu, G., Gerek, I.H., Erdis, E., et al. “Decision tree analysis of construction fall accidents involving roofers”, *Expert Syst. Appl.*, **42**(4), pp. 2256–2263 (2015). DOI: 10.1016/j.eswa.2014.10.009
 20. Cheng, C.W., Yao, H.Q., and Wu, T.C. “Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry”, *J. Loss Prev. Process Ind.*, **26**(6), pp. 1269–1278 (2013). DOI: 10.1016/j.jlp.2013.07.002
 21. Wang, X., Huang, X., Luo, Y., Pei, J., and Xu, M. “Improving Workplace Hazard Identification Performance Using Data Mining”, *J. Constr. Eng. Manag.*, **144**(8), p. 4018068 (2018). DOI: 10.1061/(ASCE)CO.1943-7862.0001505
 22. Rivas, T., Paz, M., Martín, J.E., Matías, J.M., et al. “Explaining and predicting workplace accidents using data-mining techniques”, *Reliab. Eng. Syst. Saf.*, **96**(7), pp. 739–747, (2011). DOI: 10.1016/j.res.2011.03.006
 23. Martín, J.E., Rivas, T., Matías, J.M., Taboada, et al. “A Bayesian network analysis of workplace accidents caused by falls from a height”, *Saf. Sci.*, **96**(7), pp. 739–747 (2009). DOI: 10.1016/j.ssci.2008.03.004
 24. Gerassis, S., Martín, J.E., García, J.T., et al. “Bayesian decision tool for the analysis of occupational accidents in the construction of embankments”, *ASCE J. Constr. Eng. Manag.*, **143**(2), pp. 1–8 (2015). DOI: 10.1061/(ASCE)CO.1943-7862.0001225
 25. Verma, A., Khan, S.D., Maiti, J., et al. “Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports”, *Saf. Sci.*, **70**, pp. 89–98 (2014). DOI: 10.1016/j.ssci.2014.05.007
 26. Amiri, M., Ardeshir, A., Fazel Zarandi, M.H., et al. “Pattern extraction for high-risk accidents in the construction industry: a data-mining approach”, *Int. J. Inj. Contr. Saf. Promot.*, **23**(3), pp. 264–276 (2016). DOI: 10.1080/17457300.2015.1032979
 27. Weng, J., Zhu, J.Z., Yan, X., et al. “Investigation of work zone crash casualty patterns using association rules”, *Accid. Anal. Prev.*, **92**, pp. 43–52 (2016). DOI: 10.1016/j.aap.2016.03.017
 28. Agrawal, R., Imielinski, T., and Swami, A. “Mining Association in Large Databases”, *ACM SIGMOD Rec.* (1993). DOI: 10.1145/170036.170072
 29. Kim, J.W. “Construction and evaluation of structured association map for visual exploration of association rules”, *Expert Syst. Appl.*, **74**, pp. 70–81 (2017). DOI: 10.1016/j.eswa.2017.01.007
 30. Hallowell, M.R. and Gambatese, J.A. “Activity-based safety risk quantification for concrete formwork construction”, *Journal of Construction Engineering and Management*, **135**(10), pp. 990–998 (2009).
 31. Malekitabar, H., Ardeshir, A., Sebt, M.H., et al. “On the calculus of risk in construction projects: Contradictory theories and a rationalized approach”, *Saf. Sci.*, **101**, no. September 2016, pp. 72–85 (2018). DOI: 10.1016/j.ssci.2017.08.014
 32. Munier, N., *Project Management for Environmental, Construction and Manufacturing Engineers*, Springer Science & Business Media (2013).
 33. Ardeshir, A., Mohajeri, M., and Amiri, M. “Evaluation of safety risks in construction using Fuzzy Failure Mode and Effect Analysis (FFMEA)”, **23**, pp. 2546–2556 (2016).
 34. Ghodrati, N., Yiu, T.W., Wilkinson, S., et al. “A new approach to predict safety outcomes in the construction industry”, *Saf. Sci.*, **109**, no. September 2016, pp. 86–94 (2018). DOI: 10.1016/j.ssci.2018.05.016

35. Khosravi, Y., Asilian-Mahabadi, H., Hajizadeh, E., et al. “Factors influencing unsafe behaviors and accidents on construction sites: A review”, *Int. J. Occup. Saf. Ergon.*, **20**(1), pp. 111–125 (2014). DOI: 10.1080/10803548.2014.11077023
36. Winge, S., Albrechtsen, E., and Mostue, B.A. “Causal factors and connections in construction accidents”, *Saf. Sci.*, **112**, no. October 2018, pp. 130–141, (2019). DOI: 10.1016/j.ssci.2018.10.015
37. Wang, D., Wang, X., and Xia, N. “How safety-related stress affects workers’ safety behavior: The moderating role of psychological capital”, *Saf. Sci.*, **103**, no. November 2017, pp. 247–259 (2018). DOI: 10.1016/j.ssci.2017.11.020
38. Suárez-cebador, M., Rubio-romero, J.C., and López-arquillos, A. “Severity of electrical accidents in the construction industry in Spain”, *J. Safety Res.*, **48**, pp. 63–70 (2014). DOI: 10.1016/j.jsr.2013.12.002
39. Wang, J., Zou, P.X.W., and Li, P.P. “Critical factors and paths influencing construction workers’ safety risk tolerances”, *Accid. Anal. Prev.*, **93**, pp. 267–279 (2015). DOI: 10.1016/j.aap.2015.11.027
40. Fang, D., Wu, C., and Wu, H. “Impact of the supervisor on worker safety behavior in construction projects”, *J. Manag. Eng.*, **31**(6), p. 4015001 (2015). DOI: 10.1061/(ASCE)ME.1943-5479.00003355

Biographies

Mehdi Mohajeri is pursuing his PhD in Construction Engineering and Management (Civil Engineering) at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. His main research interests are construction safety management, data-mining, behavior analysis and risk management.

Abdollah Ardeshir received the PhD degree in Hydraulics Engineering (Civil Engineering) in 1993 from the University of Kentucky, Lexington, USA. He is currently Head of Environmental Research Center and Associate Professor of Civil and Environmental Engineering Faculty at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. His main research interests are hydraulics, construction management, construction safety, health and environmental management, risk and crisis management.

Mohammad Taghi Banki received the PhD degree in Construction Management (Civil Engineering) in 1979 from the University of Missouri, USA. He is currently Assistant Professor of Civil and Environmental Engineering Faculty at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. His main research interests are construction management, construction and risk management.