

# Accident Prediction Model Using Environmental Sensors for Industrial Internet of Things

Jung-Hyok Kwon and Eui-Jik Kim\*

School of Software, Hallym University,  
1 Hallymdaehak-gil, Chuncheon-si, Gangwon-do 24252, South Korea

(Received October 15, 2018; accepted December 3, 2018)

**Keywords:** accident prediction model, association rule, big data, industrial Internet of Things, safety management

We present an accident prediction model using environmental sensors for industrial Internet of Things (IIoT), with the aim of preventing various accidents that occur at construction sites. The model is expressed as association rules generated by analyzing data collected from environmental sensors that periodically measure the changes in their surrounding environment. To develop the prediction model, we conduct the following three steps: preprocessing, association rule generation, and visualization. In the preprocessing step, the continuous value within the dataset is converted into the categorical value. In the association rule generation step, the association rules used for the prediction model are generated to represent the relationship between the accident types and causes. Finally, in the visualization step, the generated association rules are visualized in the form of a matrix plot and network graph. To demonstrate the accident prediction model, we performed an experimental implementation using open-source R. The results show that the generated association rules enable the prediction of various accidents including heatstroke, asphyxiation, collapse, and fire on the basis of the environmental factors of the construction site.

## 1. Introduction

Recently, the industrial Internet of Things (IIoT) has been widely used in a variety of industries, such as manufacturing, logistics, and construction industries.<sup>(1–3)</sup> In the IIoT, environmental sensors generate a huge amount of data every hour, and this generated data is stored on a server or the cloud in real time.<sup>(4)</sup> By analyzing such big data, we can derive new insights and meaningful results.<sup>(5)</sup> Furthermore, the results of the big data analysis can help us make better decisions and present new directions for solving existing problems. Accordingly, the use of big data analysis is spreading across a variety of industries.

In particular, many construction industries consider big data analysis to be the most important technology for safety management on construction sites.<sup>(6–8)</sup> This is because the accident rate has been increasing over time at construction sites owing to various dangerous

---

\*Corresponding author: e-mail: ejkim32@hallym.ac.kr  
<https://doi.org/10.18494/SAM.2019.2162>

factors such as heavy equipment, toxic materials, and ignition sources. Big data analysis can predict the occurrence of accidents at construction sites by analyzing the data collected from the deployed environmental sensors, thereby reducing the accident rate. In the construction site, various types of accident may occur for different reasons. Therefore, the relationships between the accident types and causes should be analyzed for accurate prediction. However, most existing work focuses only on the occurrence of a specific accident, and thus, the prediction models therein may not help to prevent accidents at construction sites.

A number of studies have been conducted to prevent accidents at construction sites. Teizer *et al.* proposed a real-time proactive radio frequency (RF) warning system to prevent collisions between workers and equipment.<sup>(9)</sup> In that system, the device sends an alert signal in real time to devices in the vicinity when it receives a signal that exceeds the user-specified signal strength. However, the proposed system considered only a specific accident type (i.e., collision). Chen and Luo investigated various data analysis methods to predict workers' fall accidents at construction sites.<sup>(10)</sup> For this, the authors compared the prediction results of the decision-tree learning algorithm, artificial neural network, and clustering algorithm. The decision-tree learning algorithm had the most accurate prediction result. However, the authors only focused on the workers' fall accidents, and thus failed to provide an appropriate data analysis method to predict different types of accidents at construction sites. Shin *et al.* generated the association rules for safety management at construction sites by analyzing the relationships between accident types and causes.<sup>(11)</sup> Various accidents could be predicted through the generated association rules. However, since the authors considered only the behavior and personal information of workers as the causes of accidents, it was difficult to predict the accidents caused by the various environmental factors of the construction sites.

In this paper, we propose an accident prediction model that is expressed as association rules generated by analyzing data collected from environmental sensors. In our work, the purpose of the proposed model is to reduce the accident occurrence rate at construction sites. For this, the proposed model accurately identifies the relationships between accident types and causes at the construction site using association rules. The development of the accident prediction model consists of three steps: preprocessing, association rule generation, and visualization. First, in the preprocessing step, all the continuous values within the dataset are converted into a categorical value. Second, in the association rule generation step, the association rules are generated by analyzing the relationships between the accident types and causes. Finally, in the visualization step, the association rules are visualized in a matrix plot and network graph. To verify the feasibility of the proposed prediction model, we conducted an experimental implementation using open-source R. In the implementation, we used a dataset consisting of six accident causes and four accident types. The implementation results showed that fifteen association rules with the minimum support of 0.2 and the minimum confidence of 0.8 are extracted via the accident prediction model.

The remainder of this paper is organized as follows. In Sect. 2, the system architecture is presented. In Sect. 3, the accident prediction model is described in detail. The implementation results are shown in Sect. 4. Finally, this paper is concluded in Sect. 5.

## 2. System Architecture

Figure 1 shows the system architecture consisting of environmental sensors, a gateway, and a big data analysis server. All the components interact with each other through wireless communication technologies such as Wi-Fi and Bluetooth. The environmental sensors periodically measure the changes in their surrounding environment and transmit the measured data to the big data analysis server via the gateway. The gateway forwards the data received from environmental sensors to the big data analysis server through the Internet. The big data analysis server stores the data measured by the environmental sensors (i.e., accident causes) and the accident history data recorded by a director of the construction site (i.e., accident types), and combines both types of data to create the dataset for the accident prediction model. The dataset includes the value of accident causes that are measured whenever the accident occurs. Figure 2 shows an example of the dataset, which consists of  $n$  accident causes and  $m$  accident types; the  $n$ -th accident cause and  $m$ -th accident type are denoted as  $C_n$  and  $T_m$ , respectively. The value of the accident cause is determined by the sensor device and is expressed by a numerical value. The value of accident type is determined by the occurrence of the accident, and is expressed by 1 if the accident occurs, and 0 otherwise.

To generate the accident prediction models, the big data analysis server has three core functions: preprocessing, association rule generation, and visualization. In preprocessing, the training dataset is transformed into a form suitable for data analysis. In association rule

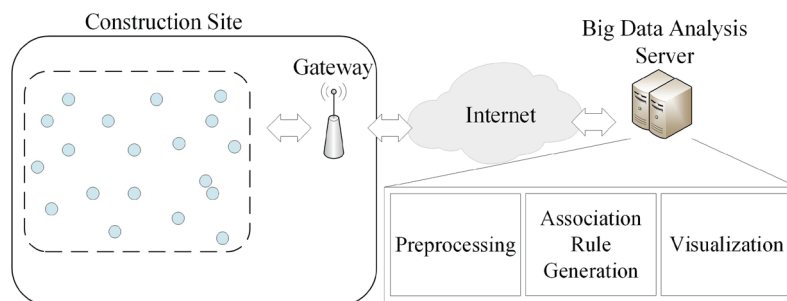


Fig. 1. (Color online) System architecture.

No	Accident Causes							Accident Types				
	$C_1$	$C_2$	$C_3$	...	$C_{n-2}$	$C_{n-1}$	$C_n$	$T_1$	$T_2$	...	$T_{m-1}$	$T_m$
1	12.5	9.1	31.2	...	459	0.43	13.0	1	0	...	0	0
2	10.8	5.3	27.0	...	398	0.13	17.4	0	0	...	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...
k	17.3	6.0	20.1	...	412	0.98	9.2	0	1	...	0	0

Fig. 2. Example of dataset.

generation, the relationships between accident types and causes are analyzed and the association rules are generated. Finally, in visualization, the generated association rules are transformed into plots and graphs for easy understanding.

### 3. Accident Prediction Model

The development of the accident prediction model consists of preprocessing, association rule generation, and visualization. Preprocessing is an important step for improving the accuracy of the model. The cause of accident can be represented as a continuous value that can be any numerical value. On the other hand, the accident type is represented only as a categorical value (i.e., 1 or 0). In preprocessing, the continuous value is categorized into three risk levels to convert them to a categorical value (i.e., discretization). The value of a cause of accident is represented as 1 when it belongs in the range of a specific risk level, and 0 otherwise. Figure 3 shows an example of discretization in which the oxygen concentration is discretized into three categories [i.e., 0–12, 12–18, and over 18 milligrams per liter (mg/L)].

After preprocessing, the association rules are generated. The association rules denote the relationships between the causes and types (i.e., items) of accidents, and are expressed in the form  $A \rightarrow B$ . A and B are referred to as the antecedent and consequent of the association rule, respectively. In our study, the causes of accident correspond to A, and the accident types correspond to B. Association rules expressed in an IF-THEN structure are evaluated in terms of support, confidence, and lift. Support indicates how frequently the association rule appears in the dataset. Confidence indicates how frequently the association rule is found to be true in the dataset. Lift indicates how accurately the association rule predicts the accident compared with random chance.

All the possible combinations of items (i.e., itemsets) are considered when association rules are generated. Thus, even a small number of items can cause a large amount of computation. To reduce the computation burden, we use the Apriori algorithm, which follows the principle of eliminating infrequent itemsets. In the Apriori algorithm, if a particular itemset is infrequent, the subsets of this itemset are also infrequent itemsets. Therefore, the use of the Apriori algorithm allows the big data server to generate association rules with a reduced number of itemsets. Figure 4 shows an example of the Apriori algorithm. In the example, the items are a, b, c, and d, and the itemsets (i.e., combinations of a, b, c, and d) are the circles. Moreover,

No	Oxygen Concentration	discretization			
No	Oxygen Concentration	Oxygen Concentration 0–12	Oxygen Concentration 12–18	Oxygen Concentration over 18	Risk Level
1	21.4	0	0	1	Low
2	15.9	0	1	0	Intermediate
...	...	...	...	...	...
k	6.1	1	0	0	High

Fig. 3. Example of discretization.

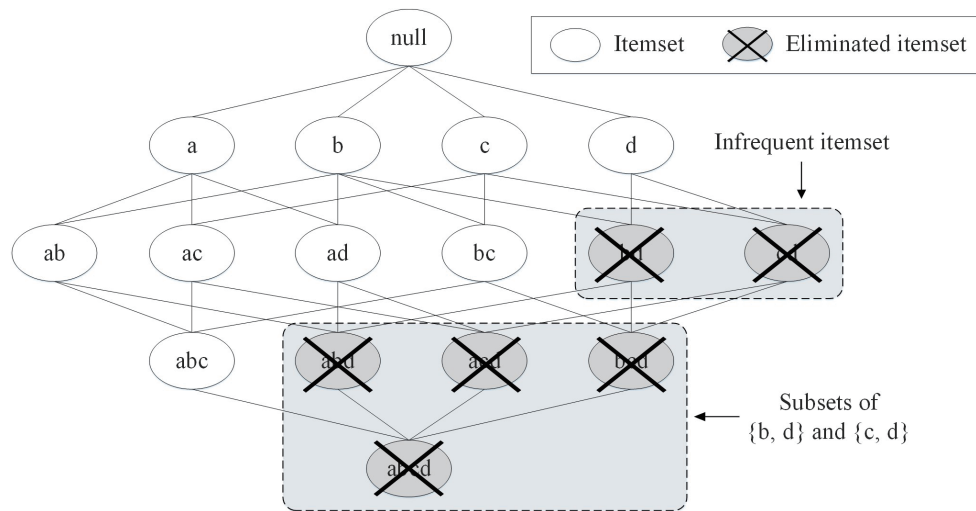


Fig. 4. (Color online) Example of the Apriori algorithm.

the infrequent itemsets are  $\{b, d\}$  and  $\{c, d\}$ . The infrequent itemsets and the subset thereof (i.e.,  $\{a, b, d\}$ ,  $\{a, c, d\}$ ,  $\{b, c, d\}$ , and  $\{a, b, c, d\}$ ) are eliminated through the Apriori algorithm. Therefore, in this case, only ten out of a total of sixteen itemsets are used for association rule generation.

Finally, the generated association rules are visualized as a matrix plot and network graph. The matrix plot represents each association rule as a rectangle. The  $x$ -axis and  $y$ -axis of the matrix plot represent the antecedent and consequent of the association rule, respectively. The color of the rectangle representing each association rule indicates support. The higher the support, the darker the color of the rectangle. The network graph indicates the association rules by “{accident cause}  $\rightarrow$  circle  $\rightarrow$  {accident type}”. The size and color of the circle respectively indicate support and lift of the association rule. A larger circle means higher support, and a darker circle indicates higher lift. The matrix plot and network graph can provide users with the intuitive insight to readily understand the association rules.

#### 4. Implementation

In this section, we demonstrate the accident prediction model for safety management at construction sites by conducting experimental implementation using open-source R. The experimental dataset used in the implementation consists of the accident types and causes collected from a construction site. The accident causes consist of oxygen concentration, hydrogen sulfide ( $H_2S$ ) concentration, humidity, flame size, construction vibration, and temperature. The accident types include fire, asphyxiation, collapse, and heat stroke. Therefore, the dataset contains a total of ten items. The accident causes with continuous values are preprocessed to categorize the continuous values as low, intermediate, or high depending on the risk level. The value of the accident type is set to 1 or 0, depending on whether the accident occurs or not, respectively.

Table 1  
Fifteen association rules.

Left-hand side	Right-hand side	Support	Confidence	Lift
{Temperature_high}	{Heatstroke}	0.23	1.00	4.41
{Oxygen_concentration_intermediate}	{Asphyxiation}	0.30	0.91	1.33
{Construction_vibration_intermediate}	{Collapse}	0.39	1.00	2.55
{Oxygen_concentration_high}	{Asphyxiation}	0.38	0.90	1.32
{Flame_size_high}	{Fire}	0.43	1.00	1.97
{Oxygen_concentration_intermediate, H <sub>2</sub> S_concentration_high}	{Asphyxiation}	0.27	1.00	1.46
{Humidity_high, Construction_vibration_intermediate}	{Collapse}	0.21	1.00	2.55
{H <sub>2</sub> S_concentration_high, Construction_vibration_intermediate}	{Collapse}	0.32	1.00	2.55
{Oxygen_concentration_high, Construction_vibration_low}	{Asphyxiation}	0.24	0.92	1.34
{Oxygen_concentration_high, H <sub>2</sub> S_concentration_high}	{Asphyxiation}	0.34	1.00	1.46
{Humidity_high, Flame_size_high}	{Fire}	0.21	1.00	1.97
{Flame_size_high, Construction_vibration_low}	{Fire}	0.26	1.00	1.97
{H <sub>2</sub> S_concentration_high, Flame_size_high}	{Fire}	0.34	1.00	1.97
{Oxygen_concentration_high, H <sub>2</sub> S_concentration_high, Construction_vibration_low}	{Asphyxiation}	0.21	1.00	1.46
{H <sub>2</sub> S_concentration_high, Flame_size_high, Construction_vibration_low}	{Fire}	0.20	1.00	1.97

To find the meaningful relationships between the accident types and causes, the association rules are generated on the basis of the experimental dataset. Table 1 shows the association rules generated when the minimum support is set to 0.2 and the minimum confidence level is set to 0.8. The table lists fifteen association rules. The left-hand side (LHS) of Table 1 represents the antecedent of the association rule and the right-hand-side (RHS) represents the consequent of the association rule. Therefore, the LHS includes the accident causes, and the RHS includes the accident types. Each rule has support, confidence, and lift. The generated association rules show the relationship between the accident types and causes as follows. If the risk levels of oxygen concentration and H<sub>2</sub>S concentration are intermediate or high, asphyxiation occurs. Moreover, if the risk level of construction vibration is more than intermediate, collapse accidents tend to occur. Heatstroke occurs when the risk level of temperature is high. According to the association rules, accident occurrence is affected not by only one item but by multiple items.

To readily understand the relationships between the accident types and causes, we visualize the association rule results. We use the *arulesViz* package supported by open-source R for visualization. Figures 5 and 6 show the visualization results of the generated association rules. Figure 5 shows fifteen generated association rules as a matrix plot. In the figure, the highest support value of the association rules is exhibited in the darkest color. The highest support values of association rules for asphyxiation, collapse, fire, and heatstroke are 0.38, 0.39, 0.43, and 0.23, respectively. From the support (i.e., color) of the square, we conclude that asphyxiation is more likely to occur when the risk level of oxygen concentration is high. We further find that collapse, fire, and heatstroke may occur when the risk levels of construction\_vibration, flame\_size, and temperature are intermediate, high, and high, respectively.

Figure 6 shows a network graph for the generated association rules. In the figure, the direction of arrows is concentrated on asphyxiation and fire. This means that there are a

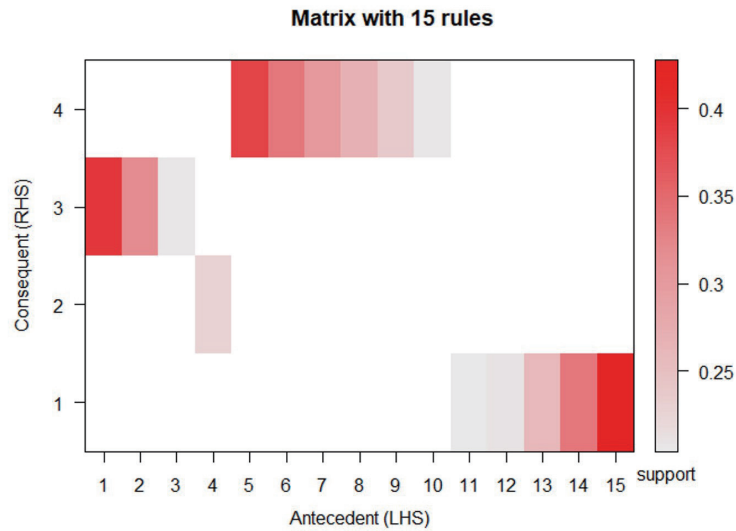


Fig. 5. (Color online) Matrix plot with fifteen association rules.

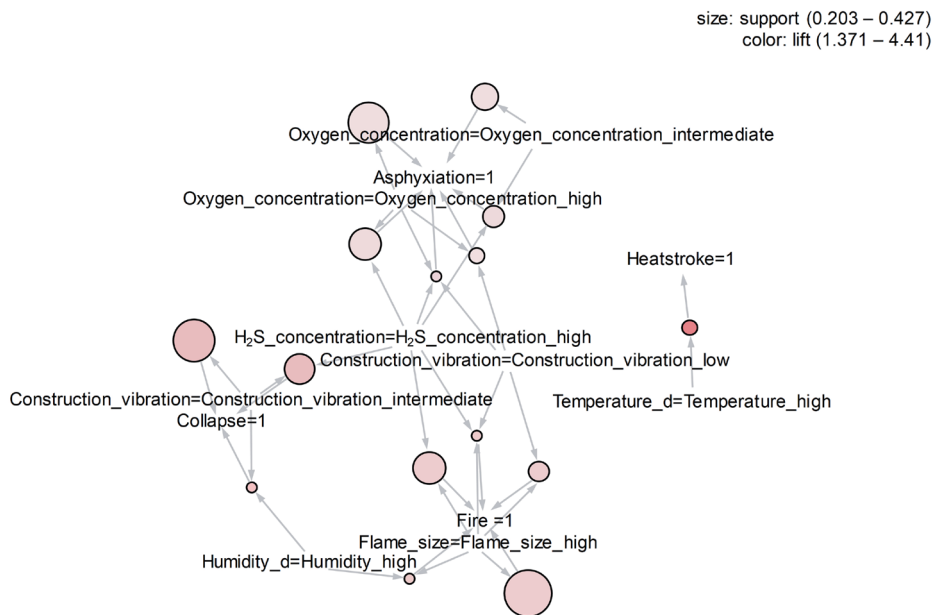


Fig. 6. (Color online) Network graph with fifteen association rules.

number of association rules related to asphyxiation and fire. Among the fifteen association rules, six are related to asphyxiation and five are related to fire. A large number of association rules implies that this type of accident occurs owing to various accident causes. The association rules “{flame\_size\_high} → circle → {fire}” and “{construction\_vibration\_intermediate} → circle → {collapse}” have large circles. The size of the circle indicates the level of support. Therefore, fire and collapse frequently occur when the risk levels of flame size and construction vibration are high and intermediate, respectively. A circle with a darker color indicates an association rule having a higher lift. Thus, there is a high probability that heatstroke will occur when the risk level of temperature is high.

## 5. Conclusions

In this paper, we proposed an accident prediction model to reduce the occurrence of accidents at construction sites. To predict the various accidents that might occur at construction sites, we analyzed data collected from environmental sensors and generated association rules that represent the relationship between the accident types and causes. For this, preprocessing, association rule generation, and visualization are conducted step by step. To demonstrate the accident prediction model, we conducted an experimental implementation using open-source R. The results showed that the proposed accident prediction model can accurately predict the occurrence of various accidents. We expect the proposed prediction model to be applicable in a worker alert service to prevent accidents at construction sites. In future work, we will focus on the improvement of prediction accuracy. Specifically, we plan to develop a knowledge-based accident prediction model that predicts the occurrence of accident considering both datasets collected from environmental sensors and those extracted from accident statistics posted on the web (i.e., knowledge).

## Acknowledgments

This research was supported by Hallym University Research Fund, 2018 (HRF-201806-011).

## References

- 1 T.-Y. Kim and E.-J. Kim: *J. Supercomput.* **74** (2018) 3549.
- 2 H.-H. Lee, J.-H. Kwon, and E.-J. Kim: *J. Supercomput.* **74** (2018) 4385.
- 3 J.-H. Kwon, H.-H. Lee, Y. Lim, and E.-J. Kim: *Appl. Sci.* **6** (2016) 427.
- 4 A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash: *IEEE Commun. Surv. Tutorials* **17** (2015) 2347.
- 5 R. Kune, P. K. Konugurthi, A. Agarwal, R. R. Chillarige, and R. Buyya: *Software Pract. Exper.* **46** (2016) 79.
- 6 M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha: *Adv. Eng. Inform.* **30** (2016) 500.
- 7 C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos: *J. Big Data* **2** (2015) 21.
- 8 T. Liu, G. Xiong, J. Hou, T. R. Nyberg, and X. Li: *Proc. 2016 IEEE Int. Conf. Service Operations and Logistics, and Informatics (SOLI)* (IEEE, 2016) 168.
- 9 J. Teizer, B. S. Allrerad, C. E. Fullerton, and J. Hinze: *Autom. Constr.* **19** (2010) 630.
- 10 H. Chen and X. Luo: *Procedia Eng.* **164** (2016) 439.
- 11 D.-P. Shin, Y.-J. Park, J. Seo, and D.-E. Lee: *KSCE J. Civ. Eng.* **22** (2018) 1027.