BIM-based Personnel Safety Level Evaluation Approach Using Machine Learning Methods for Public Buildings

Yang Peng[1], Zhenzhong Hu[2]

1) Master Degree Candidate, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: pengy15@mails.tsinghua.edu.cn
2) Ph.D., Assoc. Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: huzhenzhong@tsinghua.edu.cn

Abstract:

Ensuring personnel safety is mandatory in the operation and maintenance (O&M) management of large public buildings. Personnel safety hazards should be controlled in the O&M phase. As a crucial step in controlling hazards, the evaluation of the safety level should be based on the most accurate information. However, difficulties arise in such evaluation because of the complexity of variables that influence personnel safety. For this reason, this paper proposes a novel approach for evaluating personnel safety in public buildings. The proposed technique is based on two machine learning methods that apply building information model/modeling (BIM) technology. The proposed framework implements a supervised machine learning process through the following steps: (1) a training set, which is an integration of already known personnel safety levels derived from different sources, such as expert opinion, historical accident records, and attributes of the region, and the necessary data are extracted from BIM; (2) both decision tree and support vector machine are employed to summarize the evaluation rules; (3) evaluation of the safety levels of all regions of the building is performed automatically in accordance with these rules. The outputs would help O&M managers in dealing with each safety hazard. The proposed approach was implemented through a prototype and applied to a real public building, thereby proving that the machine learning methods were accurate and feasible in terms of safety evaluation. The proposed approach provided valuable information regarding safety management during the O&M phase.

Keywords: building information model; personnel safety; machine learning; decision tree; support vector machine

1. INTRODUCTION

Ensuring personnel safety is a vital aspect of the operation and maintenance (O&M) management of large public buildings. Safety issues have several types. Although fire safety appears as one of the mostly discussed dangers in public buildings, the design codes for such has been implemented for decades. Other safety issues include electrical accidents and mechanical failures, both of which can seriously harm residents. In extreme cases, hazards in public buildings can cause life losses. These hazards should be controlled in the O&M phase. Evaluating the personnel safety level is a crucial step in controlling hazards. The key to making such evaluation is to define the evaluating rules, which refer to how variables would influence the safety level. After the current status of a certain region of the building is examined, the obtained information will be input into evaluating rules, and the safety level will be received as output. These rules are generally based on historical records, expert interview, direct analysis of design drawings, etc. These methods can expose several problems in real practice (further discussed in Section 2.2).

In contrast to traditional works, the evaluating rules in this study are automatically summarized by machine learning methods. Machine learning is widely used to discover deeper rules behind original datasets, and these rules are in turn used to predict trends. A learning rule is a classification process. That is, each region of the building to be analyzed corresponds to a record storing the values of variables and classification results (which reflect the relevant safety levels). Machine learning works through three steps. First, a number of records are marked with safety levels by other methods. Then, machine learning algorithms attempt to learn hidden rules from these already known records. This step works like a training process for the computer; thus, this group of known records is called the "training set". Finally the remaining records are automatically classified in accordance with the rules. In general, accuracy evaluation is also required. This kind of machine learning is called "supervised", as opposed to "unsupervised" learning, which does not involve a training set.

Building information model/modeling (BIM) is a suitable data source that integrates necessary information to support machine learning. BIM serves as a complete data model that provides many views for certain uses. Basic information has been built during the design and construction process. Then, during the O&M phase, numerous attributes and files will be gradually added upon building elements. With the help of BIM, original data can be extracted conveniently, exactly, and automatically.

This paper proposes a BIM-based approach for evaluating personnel safety in public buildings, and this proposed approach is based on two machine learning methods. In Section 2, related studies are reviewed, and background knowledge is introduced. In Section 3, the two machine learning methods are presented in detail, and an application framework is formed together with automated safety level evaluation. In Section 4, BIM is integrated into the approach. Then, a case study of the data from a public building is illustrated in Section 5. Finally the drawn conclusion is provided in Section 6.

2. LITERATURE REVIEW

Many studies have tackled safety management via BIM. For example, a BIM-based framework of safety monitoring (Hu et al. 2012) has been proposed and verified in large-scale projects. BIM has also been integrated into safety planning on construction sites (Ganah and John 2015). However, the majority of these applications have focused on the construction phase; only a few have works explored the O&M phase, including a study that presented a BIM-based framework to support safe maintenance and repair practices during the facility management phase (Wetzel and Thabet 2015), which involved classification and rule checking skills. Another study developed a new game involving fire evacuation in a BIM environment (Rüppel and Schatz 2011). In other words, BIM has become a useful data source and analysis tool for safety management.

Safety evaluation is typically conducted on the basis of three aspects: historical records (Yi and Langford 2006), expert interview (Rozenfeld et al. 2010), and direct analysis of design drawings (Zhang et al. 2013). Such an approach renders a number of problems in real practice. For example, the formulated rules are often complex, fuzzy, and domain knowledge-based, thereby hindering automatic evaluation. Moreover, most of these methods consider static variables only. These methods have been implemented for many years, but they lack the ability to consider unique variables of a certain building. Other methods can directly address on-site conditions, but their rules greatly depend on experience. In addition, the relationships among variables from different methods are difficult to describe, preventing them from being integrated to create a better evaluation model.

Machine learning was developed to automatically generate design rules (Arciszewski et al. 1994). Machine learning has been proven useful in several instances in the building industry. Several support vector machine models have conducted structural health monitoring by applying signal processing (Ying et al. 2013). The performance of these models has been identified and evaluated. In these studies, "if–then" decision sentences are typically output as final results. Another study combined machine learning with data processing approach with data mining, where two time series classification methods were employed to detect the behavior of construction workers (Ying et al. 2013). Data transfer from 3D capture systems requires tremendous effort. Nonetheless, the quality of the original data is important for facilitating fast and accurate learning. Therefore, preprocessing is always necessary. In practice, two or more integrated methods are preferred rather than a single one. Kargah-Ostadi (2014) provided a comparison of nine machine learning algorithms in terms of performance and complexity.

3. METHODOLOGY AND FRAMEWORK

3.1 Evaluating Rules Learning by Decision Tree

Decision tree algorithms are a deeply developed classification method that first developed in the last century (Breiman 1984; Quinlan 1993). After the algorithm learns from the training set, the classification result is presented as a tree structure. As shown in Figure 1, each node in the tree represents a group of records that will be split in the next step. Every split is based on the value of a certain variable. Finally, the leaves of the tree represent the result of the classification. The earliest decision tree algorithm only generates binary trees; i.e., one node will only be split exactly into two. Later on, more complex algorithms were proposed (Breiman 1984; Quinlan 1993). For conciseness of use, Figure 1 does not depict a binary tree.

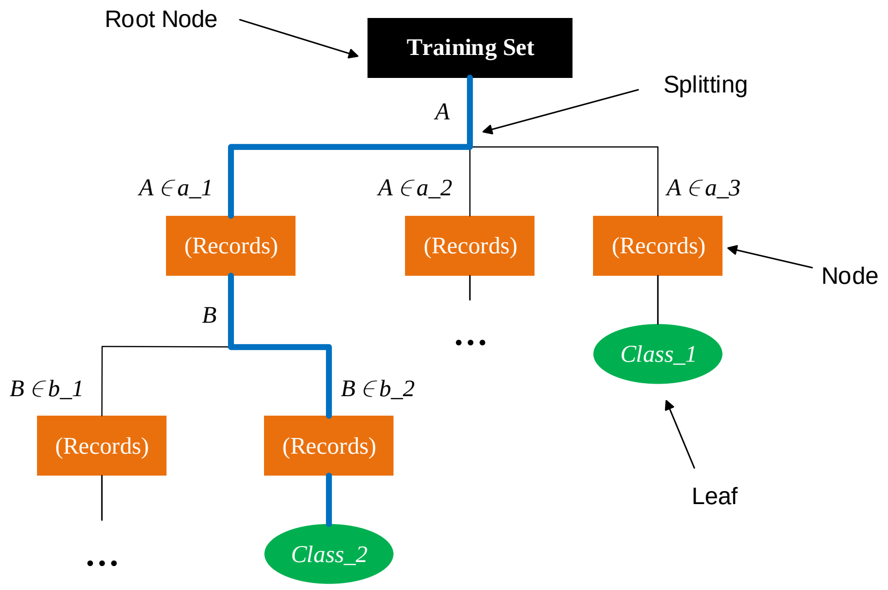


Figure 1. Part of a typical decision tree

When a record is evaluated by a certain decision tree, the variables of the record are checked along the tree, starting from the root node. The tree will lead this record to a leaf, where it can find a class to which it should belong. Moreover, a decision tree can easily be translated into rules by tracing every path of the tree and writing it as an "if–then" rule. For example, the path marked by a thick line in Figure 1 can be written as

If ($A \in a\_1$) and {If ($B \in b\_2$) and …}…} Then *class_2*

These classification rules are as many as tree leaves. Several strategies may be used to split them. In this study, classic CART (Breiman 1984) is chosen because of its good speed performance.

3.2 Evaluating Rules Learning by SVM

The support vector machine (SVM) is a kind of mathematical model that involves linear optimization. Proposed two decades ago, this algorithm id originally based on statistical learning (Vapnik and Chervonenkis 1971) (Boser et al. 1996). In this study, only a number of basic principles were considered. A record with n variables is represented by a vector in n-dimension linear space. SVM classifies these vectors by separating their endpoints with hyper planes, called separating hyper planes, in the space. Each hyper plane works as a classification rule. By applying linear optimization theory, the algorithm can determine the most effective separating hyper planes.

If the variable vector is $\boldsymbol{X} = (x_1, \ldots, x_n)^{\mathrm{T}}$, where x_i is the value of variable i, the formula of a hyper plane has the form of

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} + b = 0$$

where b is constant and weight vector $\boldsymbol{W} = (w_1, \ldots, w_n)^{\mathrm{T}}$; i.e., w_i is the weight of variable x_i. The importance of the different variables is scaled by $\boldsymbol{W}$. Classic SVM is employed in this study.

3.3 Automated Safety Level Evaluation

Applications of machine learning methods usually follow a similar workflow and can be highly automatic. This framework starts with data extraction and preprocessing. In fact, evaluating rules are actually classification models. As soon as these models are generated by the algorithm, they can address any input data without the need for rebuilding, and the records are classified as output. With machine learning, such evaluation becomes automated and stable Moreover, users need not know details about the inner algorithm. Therefore, the output result can be well understood by non-professionals.

4. BIM-BASED DATA EXTRACTION AND PREPROCESSING

Each record should possess several attributes or variables that represent the factors that influence the safety level. In considering what variables should be included in personnel safety evaluation, three aspects were examined in this study, i.e., escape issues, region attribute issues, and management issues. All information was extracted from the relational database of a BIM management system. “Safety level” had six categories, namely, “5. Very Safe”, “4. Safe”, “3. General”, “2. Dangerous”, “1. Very Dangerous”, and “0. Undefined”.

In escape issues, “escape distance” and “story height” are both important, as they are related to the evacuation behaviors of residents. These distance-related variables are extracted in BIM on the basis of bounding boxes. The escape distance of REGION_23 is shown as an example in Figure 2. First, the bounding boxes of this region and the nearby exit doors in the same story are generated (marked by thick dashed lines). Then, their center coordinates (x_{region}, y_{region}, z_{region}) and (x_{door}, y_{door}, z_{door}) are calculated. Finally, the distance between them is calculated as the Manhattan distance scale

$$dist=\min(|x_{region}-x_{door}|+|y_{region}-y_{door}|)$$

Here, the z coordinate is ignored. The straight distance is never used because people will never directly pass through obstacles such as walls. Story height is calculated similarly from the model.

Figure 2. Extraction of distance-related variables

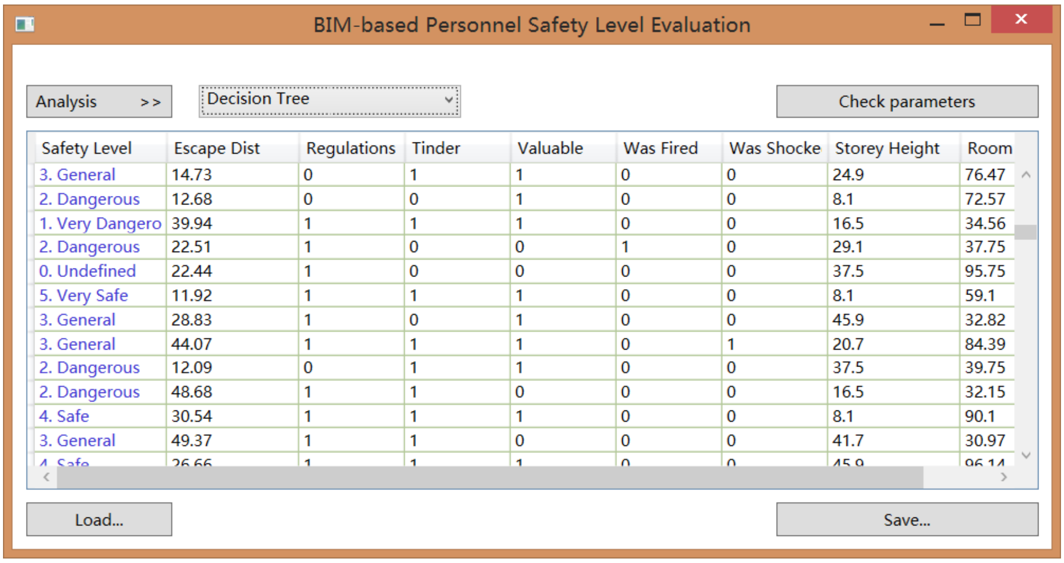
In region issues, “tinder” and “valuable” are markers of tinder or valuable things. They were input when the BIM was being established. “Room areas” were automatically indicated in BIM after the rooms were defined. In management issues, “was fired” and “was shocked” represent historical fire and electrical accidents; this information can be retrieved from repair records in the O&M phase in the facility management oriented BIM. The occurrence of “regulations” is always essential in safety management, and they can be looked up in attached files of components.

As discussed in Section 2.3, data preprocessing is necessary prior to machine learning. Incomplete records existed in original data set, including missing “safety level” values or failure to read other variables. These missing values may greatly decrease the accuracy of result, so they were detected and eliminated before analysis. Moreover, variables should be transformed somehow for better comprehension. In this study, some variables were reduced as 0–1 values. For example, “tinder” contains the kind and amount of tinder in detail, but the algorithm cannot afford to deal with every detailed information. What really matters is “yes” or “no”. After preprocessing, “tinder” had only two values: “0”, which means there was no tinder, and “1”, which means there was some tinder regardless of kind and amount.

5. CASE STUDY

5.1 Overview of the Training Set

All data in the case were obtained from the BIM database of a public building. A prototype was developed for BIM-based personnel safety level evaluation. In this study, 219 records were extracted by following the procedure described Section 4. Figure 3 illustrates the main interface. After data preprocessing, the original data set can be viewed in the list box.



Safety Level	Escape Dist	Regulations	Tinder	Valuable	Was Fired	Was Shocke	Storey Height	Room
3. General	14.73	0	1	1	0	0	24.9	76.47
2. Dangerous	12.68	0	0	1	0	0	8.1	72.57
1. Very Dangero	39.94	1	1	1	0	0	16.5	34.56
2. Dangerous	22.51	1	0	0	1	0	29.1	37.75
0. Undefined	22.44	1	0	0	0	0	37.5	95.75
5. Very Safe	11.92	1	1	1	0	0	8.1	59.1
3. General	28.83	1	0	1	0	0	45.9	32.82
3. General	44.07	1	1	1	0	1	20.7	84.39
2. Dangerous	12.09	0	1	1	0	0	37.5	39.75
2. Dangerous	48.68	1	1	0	0	0	16.5	32.15
4. Safe	30.54	1	1	1	0	0	8.1	90.1
3. General	49.37	1	1	0	0	0	41.7	30.97

Figure 3. Variables and records of the data set

Records that had undefined safety levels were meaningless for further analysis and were therefore eliminated. Finally, 208 valid records remained, of which 166 and 42 records served as the training set and the test set, respectively. Among these records, 72 were marked "general" and 68 were marked "safe". The levels of "very safe" and "very dangerous" were relatively few.

5.2 Output of Evaluating Rules

As proposed in Section 3.1, CART method was chosen to generate the tree structure. The decision tree algorithm worked for less than 1 second to output the final result (illustrated in Figure 4). Inner nodes are denoted by rectangles, and leaves are denoted by hexagons. Not all classes were pure; therefore, only the main classes were marked in leaves. At the bottom of each leaf, the amount of records in the main class is indicated in front of the slash, and the total amount of this leaf is reflected after the slash.

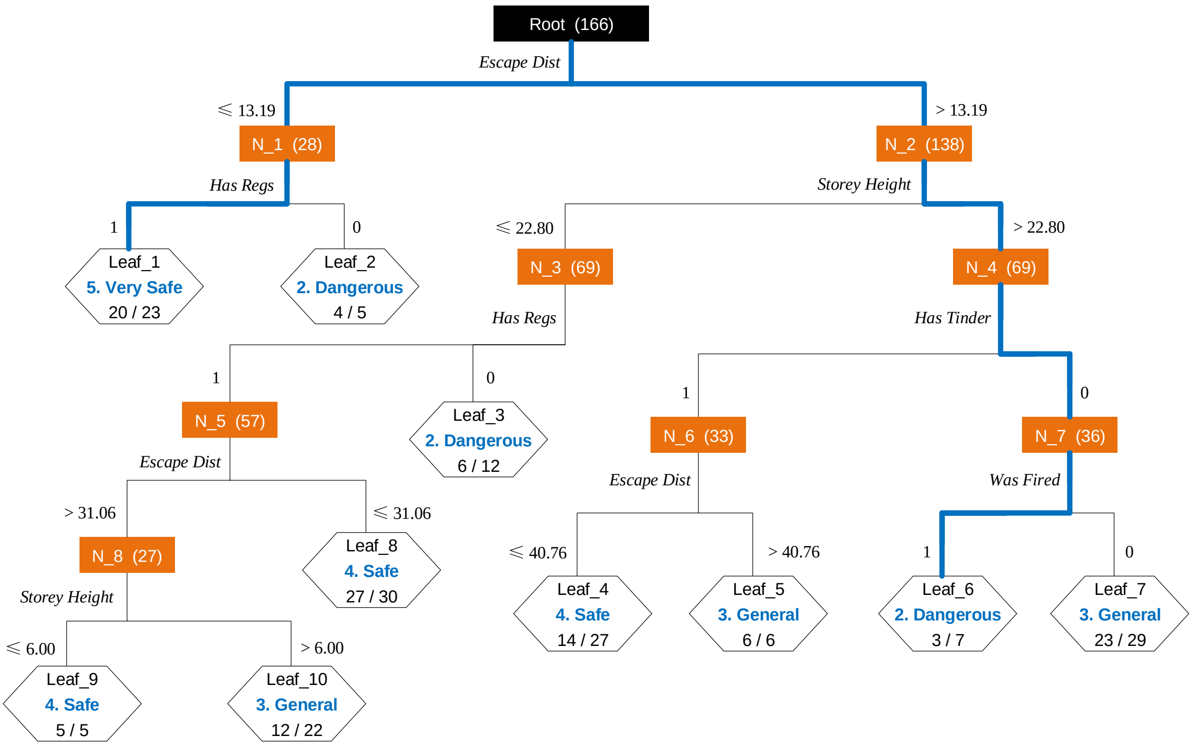


Figure 4. Output of the decision tree

Section 3.1 describes the method for generating evaluation rules. Two of them were marked by thick lines. The thick line path on the right of Figure 4 to Leaf_1 states a rule that when the region has a short escape distance and owns some safety regulations, then that region is very safe. If the area on the left thick line path to Leaf_6 experienced a fire, this area would be dangerous, especially for higher storeys. Other reasonable rules can also prove the validity of the decision tree.

The output of SVM included too many mathematical issues, which were was redundant to list here. Given that the weight vector $\boldsymbol{W}$ defined the separating hyper plane, and thus further defined the major part of the evaluation rules, it can represent the final result of SVM. From $\boldsymbol{W}$, we can calculate the importance of all variables (shown in Figure 5). In addition, the decision tree algorithm can describe another kind of importance (with a different inner process), and is thus also illustrated for comparison. The two methods agreed that escape distance, story height, and the existence of regulations were major factors in safety level evaluation, although the two methods exhibited differences in other variables of less importance.

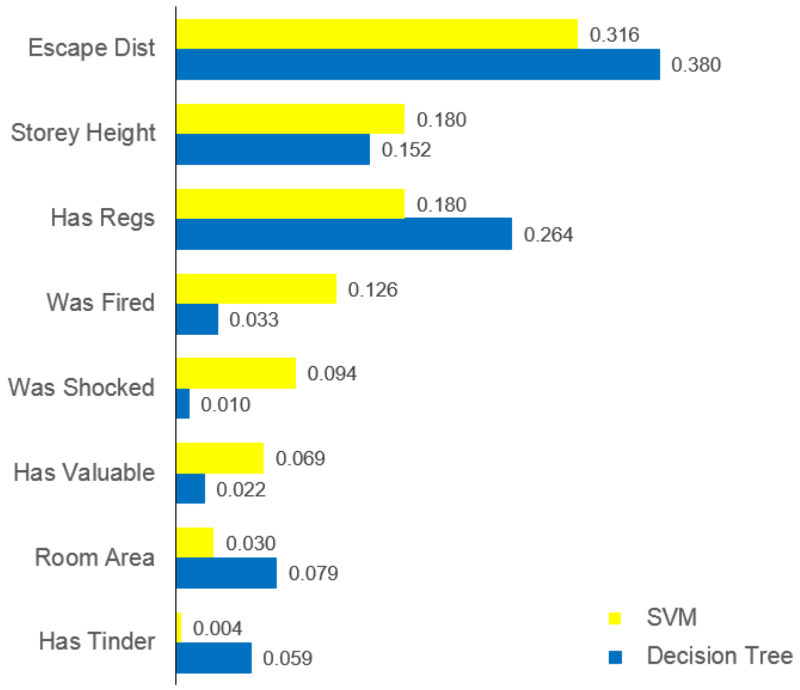


Figure 5. Importance of variables by two methods

5.3 Result of Evaluation

The test set of 42 records was used to verify the two kinds of methods. The decision tree and the SVM evaluated these records individually without any given safety levels. The actual levels and evaluation result are illustrated in Figure 6.

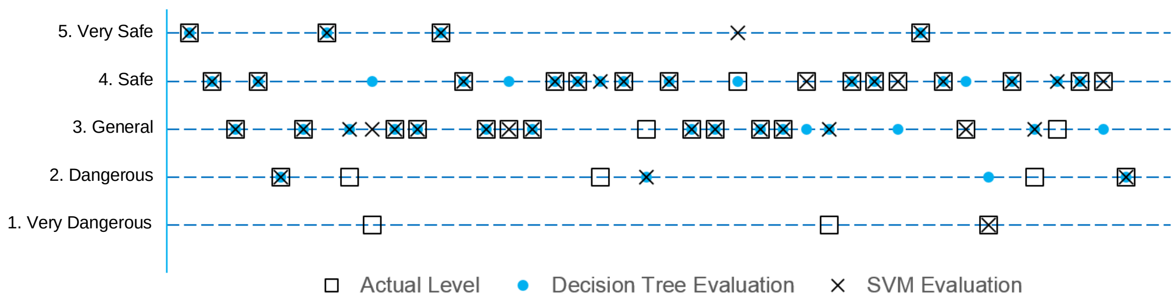


Figure 6. Actual levels and evaluation result

An evaluation result that coincides with a circle dot is a correct result; otherwise; it is an error, and the relevant error distances are recorded in Table 1. Both methods performed satisfactorily. SVM was more accurate than the decision tree with fewer large errors. In general, SVM requires more time to make calculations but performs better than the decision tree. In addition, these two methods show considerable consistency with each other (31 identical results). After evaluation, O&M managers can deal with each safety hazard on the basis of the output.

Table 1. Accuracy of two methods (total =42)

	Decision tree	SVM
Correct	29 (69%)	34 (81%)
Error (distance =1)	10	5
Error (distance =2)	2	3
Error (distance =3)	1	0

6. CONCLUSION

This paper proposes a BIM-based approach for evaluating personnel safety in public buildings. The proposed method is based on two machine learning methods. The proposed framework implements a supervised machine learning process and automated safety level evaluation approach. First, a training set and necessary data were extracted from the BIM. The training set is an integration of already known personnel safety levels derived from different sources, such as expert interview, historical accident records, and attributes of the region. Decision tree and SVM are both employed to summarize the evaluation rules. After the evaluation rules are generated, safety levels of all regions of the building can be evaluated automatically. Then, O&M managers can address each safety hazard on the basis of the output. The proposed approach was implemented through a prototype and applied to a real public building. The output result of the evaluating rules was illustrated, and the accuracy of methods was shown as well. The case study indicates that machine learning methods are accurate and feasible in terms of safety evaluation, providing valuable information for safety management during the O&M phase.

ACKNOWLEDGMENTS
This research was supported by the National Natural Science Foundation of China (No. 51478249), the National High-tech Research and Development Program of China (No.2013AA041307), and the Tsinghua University-Glodon Joint Research Centre for Building Information Model (RCBIM).

REFERENCES

Arciszewski, T., Bloedorn, E., Michalski, R., Mustafa, M., and Wnek, J. (1994). "MACHINE LEARNING OF DESIGN RULES: METHODOLOGY AND CASE STUDY." *Journal of Computing in Civil Engineering*, 3(8), 286-308.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1996). "A Training Algorithm for Optimal Margin Classifiers." *Proceedings of Annual Acm Workshop on Computational Learning Theory*, 5, 144-152.

Breiman, L. (1984). *Classification and Regression Trees*, Wadsworth International Group, Boca Raton, Florida.

Ganah, A., and John, G. A. (2015). "Integrating Building Information Modeling and Health and Safety for Onsite Construction." *Safety and Health at Work*, 6(1), 39-45.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining : concepts and techniques*, Morgan Kaufmann, Waltham.

Han, S., Lee, S., and Pena-Mora, F. (2012). "A Machine-Learning Classification Approach to Automatic Detection of Workers' Actions for Behavior-based Safety Analysis." International Conference on Computing in Civil Engineering, American Society of Civil Engineers, Florida, United States.

Hu, Z., Chen, X., Zhang, J., and Zhang, X. (2012). "A BIM-based research framework for monitoring and management during operation and maintenance period." 14th International Conference on Computing in Civil and Building Engineering, The International Society for Computing in Civil and Building Engineering, Moscow, Russia.

Kargah-Ostadi, N. (2014). "Comparison of Machine Learning Techniques for Developing Performance Prediction Models." 2014 International Conference on Computing in Civil and Building Engineering, Florida, United States.

Quinlan, J. R. (1993). *C4.5 Programs for Machine Learning*, Morgan Kaufmann, San Mateo, California.

Rozenfeld, O., Sacks, R., Rosenfeld, Y., and Baum, H. (2010). "Construction Job Safety Analysis." *Safety Science*, 48(4), 491-498.

Rüppel, U., and Schatz, K. (2011). "Designing a BIM-based serious game for fire safety evacuation

simulations." *Advanced Engineering Informatics*, 25(4), 600-611.
Vapnik, V. N., and Chervonenkis, A. Y. (1971). "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities." *Theory of Probability And Its Applications*(16), 264-280.
Wetzel, E. M., and Thabet, W. Y. (2015). "The use of a BIM-based framework to support safe facility management processes." *Automation in Construction*, 60, 12-24.
Yi, K., and Langford, D. (2006). "Scheduling-Based Risk Estimation and Safety Planning for Construction Projects." *Journal of Construction Engineering and Management*, 132(6), 626-635.
Ying, Y., Garrett, J., Jr., O. I., Soibelman, L., Harley, J., Shi, J., and Jin, Y. (2013). "Toward Data-Driven Structural Health Monitoring: Application of Machine Learning and Signal Processing to Damage Detection." *Journal of Computing in Civil Engineering*, 6(27), 667-680.
Zhang, S., Teizer, J., Lee, J., Eastman, C. M., and Venugopal, M. (2013). "Building Information Modeling (BIM) and Safety: Automatic Safety Checking of Construction Models and Schedules." *Automation in Construction*, 29, 183-195.