Building fire risk assessment based on machine learning

Aiming Xu[1], *Beibei Sun*[1]

[1] Southeast University
Nanjing, Jiangsu Province, China

ABSTRACT

In recent years, with the rapid development of economy, the continuous expansion of trading areas and buildings has caused more serious fire risks. In order to reduce the incidence of fire accidents and effectively improve building fire safety management, it is necessary to explore the application of the machine learning (ML) algorithms in fire risk assessment. This study aims to propose a ML framework for building quantitative fire risk assessment and use four regression algorithms with the data set which is collected by the Fire Safety Management System of Social Units in Jiangsu Province to get fire risk score of each company and the Mean Square Error (MSE) is used to evaluate the models. The final result shows DNN has the best performance in the experiment, which is of great significance to promote the intelligence and accuracy of fire prevention and control in smart city construction.

Keywords: Building fire risk, Risk assessment, Regression algorithm, Deep learning

INTRODUCTION

According to the fire data of China Fire and Rescue Bureau, there are 252 thousand fires which are received and reported in all in 2020. Compared with 2019, the number of fires, casualties and property losses are declining, however, there are still many large-scale fires, and the direct loss of property is still up to 4.409 billion yuan. How to reduce fire damage has become a research focus, and fire risk assessment is an effective way to solve the problem.

At present, scholars have extensively explored fire risk assessment methods which can be roughly divided into three categories: Qualitative, quantitative and semi quantitative assessment methods. In general, the results of the qualitative method are not intuitive enough, the quantitative method requires lots of statistics data, semi-quantitative method incorporates the characteristics of qualitative method and quantitative method. And with the advantages of specific results and moderate cost, semi-quantitative method has been used widely. For example, Wang proposed an integrated method which includes evidence theory, fuzzy theory and sensitivity analysis technique and this method effectively deal with some uncertainty problems (Wang et al. 2020). Zou also combines quantified safety checklist and structure entropy weight to calculate fire risk score (Zou et al. 2021). Most of the semi-quantitative methods need to combine the expert opinions to construct the index system and determine the weight, so the results are inevitably subjective. It is necessary to find more scientific and accurate methods to evaluate fire risk based on local fire basic statistics and architectural characteristics, so that more valuable and detailed assessment report can be provided to companies and the fire occurrence probability can be further reduced.

Machine learning (ML) is an important technology of artificial intelligence. It makes computers learn and think like human beings in a data-driven way and deal with the complex problems that human beings cannot solve (Alzubi et al. 2018). It can conduct independent evaluation by itself and get more accurate results. In addition, the evaluation model can be adjusted according to the occurrence of fire to make the evaluation results more accurate. Dang applied AdaBoost, XGBoost, MLP and RF algorithms on fire risk prediction in Humberside area and the result showed AdaBoost has the best performance (Dang et al. 2019). However, the actual situations of fire risk are different, the methods developed in the literature may not suit the conditions in different region.

METHODOLOGY

The process of machine learning framework construction is shown in Figure 1. First, the index system is established according to the existing nation standards and relevant data is extracted from the Fire Safety Management System. Then, the training data is put into ML model and hyper parameters is optimized. Finally, the test data is used to validate model and the result will be output.

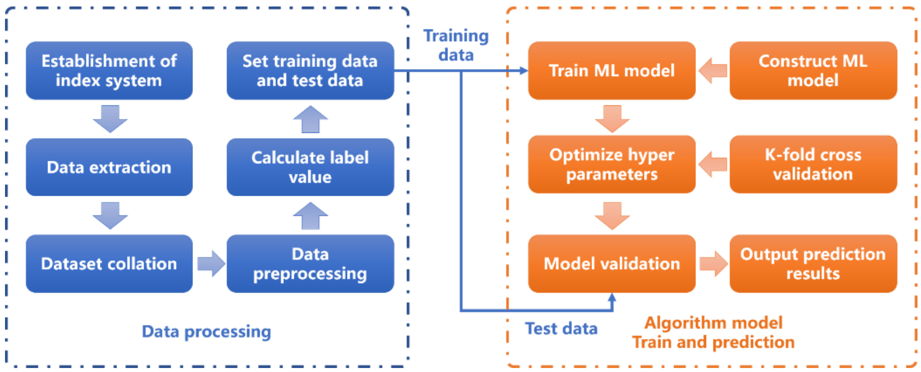

Figure 1. The process of algorithm model construction

Establishment of An Evaluation Index System

Fire is caused by many factors, so it is important to establish an index system to understand the related data, evaluation method and significant factors of fires. After reading related literatures, the final evaluation index system which references the index system in nation standard XF/T 3005-2020 and T/JFPA 000*-2021 is established. However, because the data that can be collected is different from the index factors in the index system, it is necessary to match up them before data processing. The adjustment mainly takes the data that can be collected in the Fire Safety Management System in consider. There are 48 indicators in the updated index system in total.

Data Extraction and Collation

The Fire Safety Management System can collect lots of fire data of Social Units in Jiangsu Province, but most of the data is irrelevant, so selecting the data corresponding to the index system is an important work. According to the influencing factors in the index system, a lot of unimportant data has been deleted and the dataset is sorted again. The values of each factor are shown in the column, and the data of a sample is displayed in the row. In order to facilitate the calculation of the data and make the display of the dataset that only has two values more concise, the numbers 0 and 1 are used to represent unqualified and qualified respectively, the partial data list as shown in Table 1. Because the online time of the Fire Safety Management System is short and it can not provide a large amount of data, some data are expanded by random filling.

Table 1: Data list

Sample number	**Fire separation**	**Fire compartment**	**Refuge layer**	**…**
1	1	1	0	…
2	0	1	0	…
3	1	1	0	…
…	…	…	…	…

Data Pre-processing

In order to improve the computing performance of machine learning, it is necessary to find the redundant values, missing values and outliers.

There are many methods can be used to fill the missing value. The most common methods are univariate and multivariate methods. Univariate method contains mean imputation, forward or backward imputation, and moving average methods. These methods fill the missing value according to the data characteristics of the variable and Yu proved these methods are effective in processing the dataset with a few missing values (Yu et al. 2019). Multivariable methods apply some ML algorithms on missing value imputation, such as k-nearest neighbour (KNN), multiple linear regression (MLR). Compared with univariate method, this method can provide more accuracy value and when the missing data ratio is large, it can still achieve satisfactory performance.

If there are duplicate sample data, only one sample is retained and the rest samples are deleted. Outliers in the sample are filled with 1 or 0 according to the rest of the data of the sample.

Calculation of Target Value

The calculation method of the target value references the method provided in T/JFPA 000*—2021, and it is modified to match up the data that can be collected by the system. In order to simplify calculation processing and improve efficiency, the whole calculation process is implemented by Python language.

Model Construction

Gradient boosting decision tree(GBDT), Extreme Gradient Boosting (XGBoost）, Support Vector Regression（SVR）and Deep neural network（DNN）are used to predict the fire risk score of each company in this study. These four models are all supervised learning model. The input and output of models are the training data and the target value respectively.

GBDT. GBDT is an ensemble learning models, it is also called strong evaluator. GBDT takes the value of the negative gradient of the loss function in the model as the approximate value of the residual to fit the regression tree (Zhang et al. 2021). In addition, the misclassified samples attained by previous leaners will influence the later learning, the model of GBDT is shown in Figure 2.

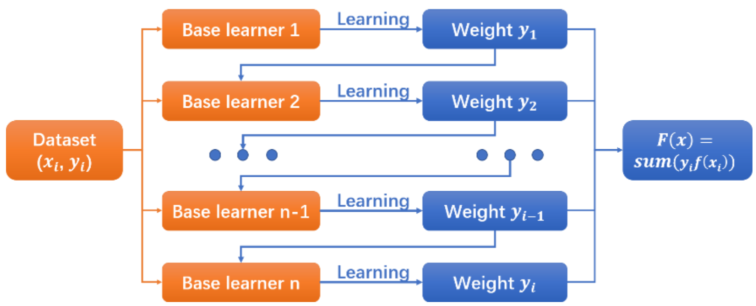

Figure 2. The model of GBDT

XGBoost. XGBoost algorithm was proposed by Chen and Guestrin, its excellent performance in Kaggle's ML competitions make it popular (Chen et al. 2016). Compare to GBDT, it introduces more methods to avoid overfitting, such as the regularization term.

SVR. Using Support Vector Machines (SVM) to solve regression problems model is called SVR. It maps data samples to higher dimensional space by kernel function and find an optimal hyperplane to minimize the variance between the support vector and the hyperplane (Abbasi et al. 2019).

DNN. DNN is one of the types of artificial neural network. If the number of the hidden layer in a neural network is more than three, this neural network can be called DNN. By combining multiple types of network layers, the input data is calculated. In order to make the result more accurate, the back propagation algorithm is applied to error reduction. Information in the back propagation algorithm is transmitted in two directions and the neurons can receive their own feedback information. It has memory function and the ability of optimization is better than the feedforward neural network.

Model Training

To measure the final performance of each model, some evaluation metrics should be used. R Squared (R^2) is used to express the relationship between the sum of squares of residuals and the total sum of squares. It usually represents the fitting degree of the model. If the value of R^2 is closer to 1, it means the fire risk value calculated by the model is valuable. its calculation method is shown in Eq. (5)

$$R^2 = 1 - \frac{u}{v} \tag{5}$$

Mean Square Error (MSE) compares the difference between the predicted fire risk value and the actual fire risk value.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2 \tag{6}$$

Where f_i, y_i are the predicted fire risk value of the model and the average of the actual values, N represents the number of samples.

Hyper Parameters Optimization

In order to make the model more accurate, the hyper parameters need to be adjusted while training the model. However, The smallest calculation error does not represent the best performance, so the complexity of the algorithm and the prediction error should be considered together in the progress of optimization. Compare to other algorithm, although grid search takes a long time to calculate, it can consider all possibilities of the set of hyper parameters. Therefore, grid search is used to find the best set of hyper parameters in this study. Table 2 shows some critical hyperparameters of GBDT, XGBoost, SVR. Besides, DNN model has the best performance when its combination of neuron nodes is 30,30,30,30,30 and the MSE becomes larger when it exceeds 30.

Table 2: Hyperparameters optimization results

Algorithm	Hyperparameters	Search Ranges	Optimal Values
GBDT	n_ estimators	(50, 800)	800
	learning rate	(0.01, 0.2)	0.2
	max_ depth	(2,10)	3
	min_samples_leaf	(2, 5)	2
XGBoost	n_ estimators	(50, 800)	800
	learning rate	(0.01, 0.2)	0.2
	max_ depth	(2,10)	3
	subsample	(0.1, 1)	0.5
SVR	C	(1,20)	5
	ε	(0.1,1)	0.1
	gamma	(0.01, 0.1)	0.05

RESULTS AND DISCUSSION

5-fold cross validation is used in this experiment. Table 3 shows the performance of four algorithms with optimal value.

Table 3: Performance of the Algorithms

Algorithm	**Evaluation metrics**	
	R^2	MSE
GBDT	0.99	0.53
XGBoost	0.99	0.64
SVR	0.99	0.10
DNN	0.99	0.09

According to the result shown in the table, the MSE of DNN is the smallest. This result indicates that the performance of DNN model is the best. In addition, the R^2 of these four algorithms are the same and they are very close to 1. This phenomenon shows that the fire risk values that calculated by these four algorithms are all acceptable in the application of fire risk assessment.

ACKNOWLEDGMENTS

I would like to acknowledge my tutor for taking the time to carefully read my paper and help me to improve it. At the same time, I am also very grateful to my classmates for giving some useful suggestions when I look for relevant information. In addition, I also want to thank the company that provided the dataset for me.

CONCLUSIONS

In this study, a machine learning framework of fire risk assessment model is put forward and the model is trained by using the data of Fire Safety Management System of Social Units in Jiangsu Province. The dataset is organized according to the data that can be collected in the system and the expected fire risk values are calculated by the method provided in the regulation. Then, four models include GBDT, XGBoost, SVR and DNN are constructed to evaluate the value of fire risk after optimizing hyper parameters. The result shows that DNN is the best algorithm to evaluate the fire risk to assess specific dataset in this study.

Except the basic structure of DNN, there are lots of deep learning networks with different structure, such as CNN. It contains convolution layer, pooling layer and fully connected layer except input layer, hidden layer and output layer. Further-more, the Resnet and RNN are also popular in recent years. For example, even though these structures are mostly used in classification, the input layer and the output layer can be adjusted to be adapt to the dataset and it can be applied to evaluate the value of fire risk. The next step is to try these structures on the assessment of fire risk and hope it can bring a better result.

REFERENCES

China Fire and Rescue Bureau (in Chinese), https://www.119.gov.cn/article/41kpo4CAQyv

Wang, W., Xiao, G., Liu, Y., and Mi, H. (2020), “An integrated method for fire risk assessment in residential buildings.” *Mathematical Problems in Engineering*, Volume 2020.

Zou, Q., Zhang, T., and Liu, W. (2021), “A fire risk assessment method based on the combination of quantified safety checklist and structure entropy weight for shopping malls.” *Proceedings of the Institution of Mechanical Engineers*, 235, 610-626.

Alzubi, J., Nayyar, A., and Kumar, A. (2018), Machine learning from theory to algorithms: an overview. *2nd National Conference on Computational Intelligence*. vol. 1142, 012012.

Dang. T. T., Cheng. Y., Mann. J., Hawick. K. and Li. Q. (2019), Fire Risk Prediction Using Multi-Source Data: A case study in Humberside area. *25th International Conference on Automation and Computing*, 1-6.

Implementation Regulations for Assessment of Fire Prevention and Control Ability in Enterprises and Institutions (in Chinese), T/JFPA 000*—2021 (2021)

Fire safety assessment for places of employment (in Chinese), XF/T 3005-2020 (2020)

Yu, X., Ergan, S., and Dedemen, G. (2019), “A data-driven approach to extract operational signatures of HVAC systems and analyze impact on electricity consumption.” *Applied Energy*. 253, 113497.

Zhang. W.H., Yu. J.Q., Zhao A.Z. and Zhou, X.W. (2021), “Predictive model of cooling load for ice storage air-conditioning system by using GBDT.” *Energy Reports*. 7, 1588-1597.

Chen, TQ., and Guestrin, C. (2016): Xgboost: A scalable tree boosting system. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Abbasi, M., Farokhnia, A., Bahreinimotlagh M., and Roozbahani. R. (2021), “A hybrid of Random Forest and Deep Auto-Encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction.” *Journal of Hydrology*. 597, 125717