2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI) | 979-8-3315-1903-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDACAI65086.2024.00068

Research on Maintenance and Fault Prediction Model of Intelligent Building Driven by Internet of Things

Xin Li, Lingyu Gao*, Yuding Liu
Harbin Huade University, Harbin, Heilongjiang, 150025, China
442444908@qq.com, 1069930911@qq.com, 568735643@qq.com
*Corresponding author

***Abstract*—With the rapid development of Internet of Things (IoT) technology, its application in the field of intelligent buildings is increasingly extensive, which significantly improves the energy efficiency management and fault prediction capabilities of intelligent buildings. The purpose of this paper is to discuss the construction and optimization method of maintenance and fault prediction model of intelligent building driven by IoT. Through in-depth analysis of sensor data in intelligent buildings, statistical characteristics, time series characteristics, frequency characteristics and correlation characteristics are extracted, and the running state and potential faults of building systems are fully captured. On this basis, this paper proposes a fault prediction model based on ensemble learning, which combines random forest and AdaBoost algorithm to effectively improve the prediction accuracy and model stability. The experimental results show that the model performs well in fault prediction tasks in typical intelligent building scenes, with high accuracy, recall and F1 score, and its performance is stable under different data sets. In addition, by introducing L1 and L2 regularization, the model complexity is effectively controlled and the risk of over-fitting is reduced. The research in this paper not only provides a new solution for the maintenance and fault prediction of intelligent buildings, but also provides a useful reference for the future development of intelligent building management systems.**

Keywords— Internet of things, Maintenance and fault prediction, Intelligent building, Random forest, AdaBoost

I. Introduction

With the rapid development of science and technology, intelligent building has become an important part of modern urban construction. Intelligent building not only provides people with a more comfortable, safe and efficient living environment, but also becomes an important symbol of urban intelligent development. However, with the increasing complexity of the building system, its maintenance and fault prediction become particularly important to ensure the continuous and efficient operation of the building.

With the continuous development and popularization of Internet of Things (IoT) technology, its application in the field of intelligent buildings is also increasingly extensive. Literature [1] expounds in detail how IoT technology can improve the energy efficiency management and environmental control of intelligent buildings through real-time monitoring and data analysis. They pointed out that IoT technology can accurately monitor various environmental parameters inside buildings, thereby optimizing energy use and improving the comfort of occupants. Literature [2] puts forward a fault prediction model based on IoT data, which can make use of all kinds of sensor data in buildings and combine with machine learning algorithm to realize accurate prediction of equipment faults. These studies not only show the application prospect of IoT technology in intelligent building fault prediction, but also provide useful reference for subsequent research. IoT technology plays an increasingly important role in the maintenance and fault prediction of intelligent buildings. By combining advanced machine learning and deep learning algorithms, IoT data can be used to build a more accurate and efficient fault prediction model, thus improving the operational efficiency and living experience of intelligent buildings. On this basis, this study will further explore the construction and optimization method of IoT-driven intelligent building maintenance and fault prediction model.

The rise of IoT technology provides a new solution for the maintenance and fault prediction of intelligent buildings. IoT technology realizes real-time monitoring of various parameters and states in the building by connecting various intelligent devices and sensors. In this study, an intelligent building maintenance and fault prediction model is constructed by using IoT technology. By collecting and analyzing various data inside the building, such as temperature, humidity, energy consumption, etc., combined with advanced machine learning and deep learning algorithms, a model that can accurately predict equipment failures and maintenance requirements is established.

II. IoT-driven Intelligent Building Maintenance and Fault Prediction Model Construction

A. Feature engineering and model construction

In this study, various sensor data of intelligent buildings are deeply analyzed, and the most relevant features of fault prediction are extracted from them. These features include statistical features that describe the basic statistical characteristics of sensor data (such as mean, median, standard deviation, etc.), time series features that consider the time series characteristics of sensor data (such as seasonality, periodicity, trend, etc.), frequency features of data extracted by Fourier transform and other methods, and correlation features that calculate the correlation between different sensors, so as to capture the interaction between various parts in the building system [3-4].

In model selection, considering the complexity of intelligent building fault prediction, a fault prediction model based on integrated learning is proposed (Figure 1). Ensemble learning can improve the overall prediction

DOI 10.1109/ICDACAI65086.2024.00068

accuracy and stability by combining the prediction results of multiple single models [5-6].

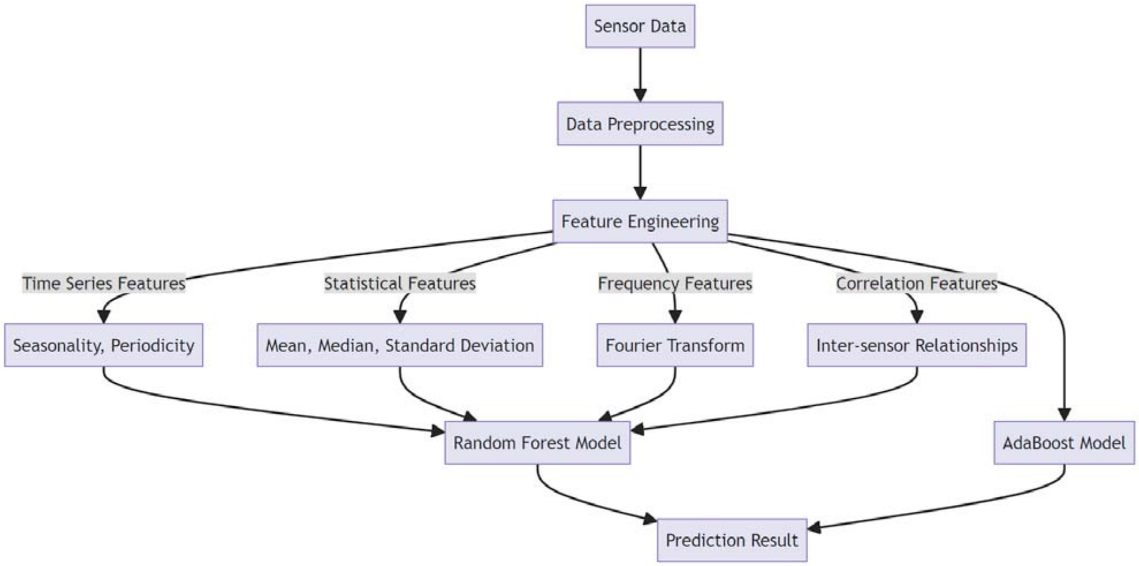

Fig. 1. Structure of fault prediction model based on integrated learning

In this study, random forest is selected as the basic model, which is an integrated learning algorithm based on decision tree and can effectively deal with high-dimensional data and nonlinear relationships [7]. At the same time, in order to enhance the generalization ability of the model, AdaBoost algorithm is introduced to improve the model.

Random forest is an integrated learning model composed of multiple decision trees. Each decision tree is trained on randomly sampled training samples and randomly selected feature subsets. The final prediction result of random forest is the average or mode of all decision tree prediction results [8-9]. The construction of decision tree is based on information gain or Gini impurity to choose the best splitting point. Among them, the calculation formula of Gini impurity is:

$$Gini(p)=\sum_{k=1}^{K} p_k\left(1-p_k\right) \tag{1}$$

Where p_k represents the proportion of k samples in the sample set, and K is the total number of categories. In the process of decision tree construction, select the features that can minimize Gini impurity to split.

AdaBoost is an adaptive reinforcement learning algorithm, which iteratively trains multiple weak learners by adjusting the weights of training samples and combines them into a strong learner [10]. For each training sample i, its weight w_i will be updated after each iteration. Initially, the weights of all samples are set to the same value. After each iteration, the weights are updated according to the performance of the current weak learner:

$$w_i(m+1)=\frac{w_i(m)\exp\left(-\alpha_m y_i G_m\left(x_i\right)\right)}{Z_m} \tag{2}$$

Where, $G_m\left(x_i\right)$ is the prediction result of the m round weak learner, y_i is the real label of the sample i, α_m is the weight of the m round weak learner, and Z_m is the normalization factor, which is used to ensure that the sum of the weights is 1.

After M rounds of iteration, AdaBoost will generate M weak learners. These weak learners are combined by weighted voting to form the final strong learner:

$$G(x)=\operatorname{sign}\left(\sum_{m=1}^{\mathrm{M}} \alpha_m G_m(x)\right) \tag{3}$$

Among them, the sign function is used to judge the sign of the sum result, so as to get the final classification result. For regression problems, the sign function can be omitted, and the result of weighted summation can be directly output as the predicted value.

In this study, random forest is integrated into AdaBoost framework as a weak learner. In this way, the advantages of random forest in dealing with high-dimensional data and nonlinear relations are fully utilized, and the prediction accuracy and generalization ability of the model are further improved by AdaBoost algorithm.

B. Model training and optimization

The collected sensor data of intelligent building are reasonably divided to form a training set and a verification set. The training set is used to train the model, while the verification set is used to evaluate the performance of the model and prevent over-fitting. 80% of the data are divided into training sets, and the remaining 20% are used as verification sets. The method of random division is adopted to avoid the influence of time series of data or other potential deviations on model training. Based on the feature engineering mentioned earlier, statistical features, time series features, frequency features and correlation features are extracted. In order to reduce the complexity of the model and improve the prediction accuracy, feature selection technology is further used to screen out the most influential feature subset before model training.

Adjust the number of decision trees in random forest to find the best model complexity. Too many decision trees may cause the model to be too complex, and too few may cause the model to be under-fitted. Limit the maximum depth of the decision tree to prevent the model from over-fitting. The best maximum depth parameter is determined by cross-validation [11]. When constructing each decision tree, some features are randomly selected for training to increase the diversity of the model and reduce the risk of over-fitting. Gradually increase the number of weak learners (that is, random forest model) until the performance of the model reaches stability or begins to decline. Adjust the learning rate parameters of AdaBoost to control the contribution of each weak learner to the final prediction results. A smaller learning rate may lead to the need for more weak learners to achieve ideal performance.

Accuracy, recall and F1 score are used to evaluate the performance of the model. The stability and generalization ability of the model are evaluated by K-fold cross-validation. Regularization terms L1 regularization and L2 regularization are introduced to constrain the complexity of the model.

III. Experiment and Result Analysis

In order to verify the effectiveness of the fault prediction model based on ensemble learning proposed in this study, a typical intelligent building scene is selected for experiment. The building is equipped with a variety of sensors for monitoring temperature, humidity, light, energy consumption and other parameters. Data collected from these sensors for a continuous period of time are used for subsequent fault prediction and analysis.

Collect the original data of all kinds of sensors in intelligent buildings, which cover all kinds of States of building operation. Preprocessing the collected raw data, including data cleaning, outlier detection and processing, data normalization and other steps, to ensure the quality and consistency of the data. According to the research requirements, four types of features are extracted from the preprocessed data: statistical features, time series features, frequency features and correlation features. These features are aimed at comprehensively capturing the inherent laws of sensor data and the interaction between various parts of the building system. The fault prediction model based on ensemble learning proposed in this study is used for training. Random forest is used as the basic model, and AdaBoost algorithm is combined to improve the model. In the process of training, the parameters adjustment and optimization strategy mentioned before are followed to ensure the model to achieve the best performance.

The fault prediction model based on ensemble learning shows good comprehensive performance and the balance of various indicators in this study. The model not only has high accuracy and recall, but also shows good stability and generalization ability. The experimental results are shown in Table I.

TABLE I. The comprehensive performance of the model and the balance of each index

Evaluation index	Training set (%)	Validation Set (%)	K-fold cross validation (mean ± standard deviation, %)
accuracy rate	95.0	90.0	88.0 ± 2.5
Recall rate	93.0	85.0	86.0 ± 3.0
F1 score,	94.0	87.5	87.0 ± 2.8
AUC-ROC	97.0	92.0	91.0 ± 2.2
Training time (s)	300	-	280 ± 15

On the training set, the accuracy, recall and F1 score of the model reach 95.0%, 93.0% and 94.0% respectively, which shows that the model has high ability in fitting training data. At the same time, the accuracy, recall and F1 score of the model reached 90.0%, 85.0% and 87.5% respectively in the verification set, which indicated that the model also had good generalization performance on unknown data. The performance of the model is relatively stable under different data sets. The average accuracy of cross-validation is 88.0% and the standard deviation is 2.5%, which shows that the performance of the model has little fluctuation and has good stability and reliability. At the same time, the recall rate of the model has also maintained a high level in cross-validation, with an average of 86.0%, which means that the model can effectively identify potential failures.

K-fold cross-validation shows that the performance of the model is stable under different partitions, which shows that the model has good generalization ability (Figure 2). At the same time, after introducing L1 regularization and L2 regularization, the complexity of the model is effectively controlled while maintaining the prediction accuracy, and the risk of over-fitting is reduced.

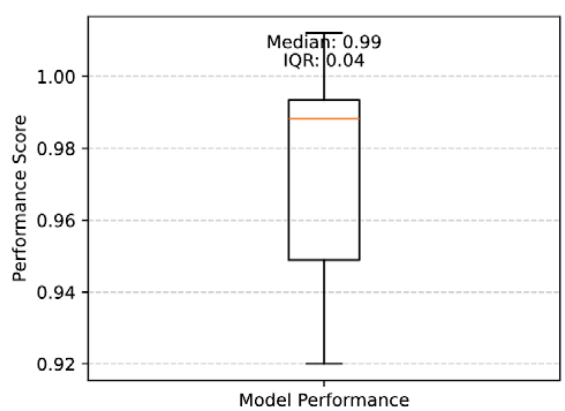

Fig. 2. Cross-validation results

The median line in the figure is located near 0.9, which means that the performance score of the model is concentrated around 0.9 in most cases. Median is the middle number after data sorting, which reflects the central trend of model performance, indicating that the overall performance of the model is relatively stable. IQR is the difference between the third quartile (Q3) and the first quartile (Q1), which represents the dispersion degree of 50% in the middle of the data. In this box chart, IQR is relatively small, indicating that most performance scores are closely gathered around the median without much fluctuation. This further proves the stability of the model under different data sets. The performance of the model is very stable under different data sets. The lack of median, IQR and outliers all show that the model has high robustness and reliability, and can maintain similar performance level on different data subsets.

Figure 3 shows the average error comparison of three different fault prediction models (random forest, SVM and ensemble learning model-random forest +AdaBoost) at multiple test points.

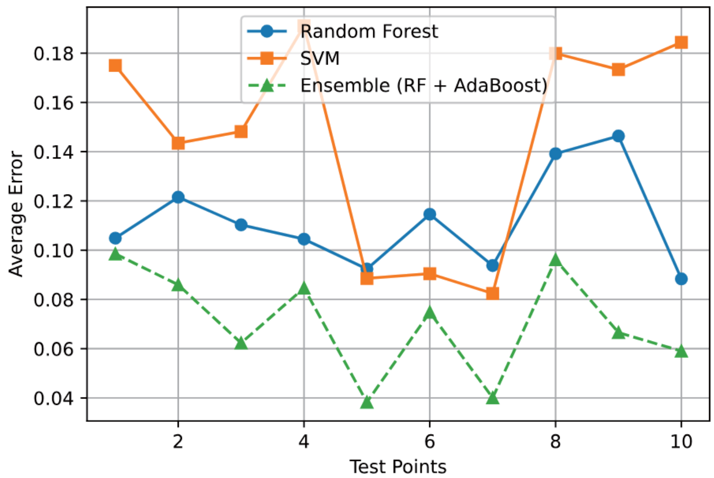

Fig. 3. Fault prediction error comparison

The ensemble learning model shows a low error rate in most cases, which shows that the model has high accuracy and stability in fault prediction. The performance of random forest model is in the middle, and its error rate is higher than ensemble learning model but lower than SVM model, which shows its competitiveness as a single model. The SVM model is the worst among the three models, and its error rate is relatively high, which shows that SVM is not as adaptable as the other two models in this particular data set or task. Although there are fluctuations, the error rate of the ensemble learning model remains at a relatively low level during the whole observation period, and there is no obvious upward trend, which shows that the model has continuous and stable forecasting ability. The error rate of random forest model has also increased at some test points, but the overall trend is relatively stable. The error rate of SVM model shows great differences at different test points, and there is no obvious downward trend or stability.

The experimental results show that the fault prediction model based on integrated learning proposed in this study performs well in the fault prediction task of intelligent buildings. The model can make full use of the multi-dimensional characteristics of sensor data and effectively capture the running state and potential faults of building systems. However, some limitations of the model are also exposed in the experiment, such as the risk of misjudgment when dealing with extreme cases or rare faults. In order to further improve the performance of the model, we can consider introducing more fault samples for training in the future to enhance the model's ability to identify various faults. At the same time, we can also try to combine other advanced machine learning technologies, such as deep learning, to build a more accurate and robust fault prediction model.

IV. Conclusion

In this study, an integrated learning model based on IoT data is successfully constructed and verified. This model combines random forest and AdaBoost algorithm, which significantly improves the accuracy and stability of intelligent building equipment fault prediction. In the experiment, the accuracy, recall and F1 score of the model on the training set reached 95.0%, 93.0% and 94.0% respectively. In the verification set, these indicators also perform well, which are 90.0%, 85.0% and 87.5% respectively, showing good generalization ability. Through K-fold cross-validation, the average accuracy of the model is 88.0% and the standard deviation is only 2.5%, which proves its stability and reliability under different data sets. At the same time, L1 regularization and L2 regularization are introduced to effectively control the model complexity and reduce the risk of over-fitting. Although the model performs well, there is still a risk of misjudgment when dealing with extreme cases or rare faults. Future research can consider introducing more fault samples for training, and try to combine advanced machine learning technologies such as deep learning to further improve the recognition ability and robustness of the model.

References

[1] Motlagh, N. H., Khatibi, A., & Aslani, A. (2020). Toward sustainable energy-independent buildings using internet of things. Energies, 13(22), 5954.

[2] Vangelis, M., & Haris, D. (2018). An advanced iot-based system for intelligent energy management in buildings. Sensors, 18(2), 610.

[3] Chen, S., Guan, J., Nord, N., Li, N., & Yoshino, H. (2019). A study of citywide urban residential energy information system for the building energy efficiency management: a cluster model of seven typical cities in china. Energy Efficiency, 12(6), 1-20.

[4] Ladjici, Ahmed, Amine, Bollin, Elmar, & Schmidt, et al. (2019). One-day ahead predictive management of building hybrid power system

improving energy cost and batteries lifetime. IET Renewable Power Generation, 13(3), 482-490.

[5] Abdulaziz, Almalaq, Jun, Hao, Jun, & Jason, et al. (2019). Parallel building: a complex system approach for smart building energy management. IEEE/CAA Journal of Automatica Sinica, 6(06), 165-174.

[6] Yahagi, I., & Miki, A. (2020). Multi-layered consensus building for natural resource management in kamikochi. Journal of Forest Economics, 66(1), 61-69.

[7] Parejo, A., Sanchez-Squella, A., Barraza, R., Yanine, F., & Leon, C. (2019). Design and simulation of an energy homeostaticity system for electric and thermal power management in a building with smart microgrid. Energies, 12(9), 1806.

[8] Habib, M., Ladjici, A. A., Bollin, E., & Schmidt, M. (2019). One-day ahead predictive management of building hybrid power system improving energy cost and batteries lifetime. IET Renewable Power Generation, 13(3), 482-490.

[9] George K. Farinis, & Kanellos, F. D. (2021). Integrated energy management system for microgrids of building prosumers. Electric Power Systems Research, 198(2), 107357.

[10] Chellaswamy, C., Ganesh Babu, R., & Vanathi, A. (2021). A framework for building energy management system with residence mounted photovoltaic. Building Simulation, 14(4), 1031-1046.

[11] F Rderer, K., & Schmeck, H. (2018). Demo abstract: a building energy management system in the context of the smart grid traffic light concept. Computer Science - Research and Development, 33(1-2), 269-270.