Generalized Prediction of Commercial Buildings Cooling and Heating Load Based on Machine Learning Technology

Jianing Liu¹, Kaiweng Zeng¹, Haizhu Wang¹, Bin Du¹ and Yufeng Tang²

¹ Guangdong Power Grid Power Dispatch Control Center, Guangzhou, 75 Meihua Road, Guangdong, 510000, China
² GDEPRI Power Control Systems & Equipment Ltd, Guangzhou, China

Email: *zengkaiwen@gddd.csg.cn

Abstract: Factors such as building area, materials, light transmittance, and housing orientation have significant effects on cold and heat loads of commercial buildings. Accurate prediction of cold and heat loads based on the characteristics of the building can provide a reliable data basis for power grid dispatch. In order to quantitatively study the influence of the commercial building factors on the cooling and heating load, this paper collected the load of 768 buildings in a certain area on a typical day in spring and the relative density, surface area, room orientation, and light transmission area of the building group. This paper uses typical machine learning methods such as random forest, extra random tree, bagging, and deep neural network to perform regression modeling on the data. Finally, three models with the best effects are selected for blending to obtain the final building cooling and heating load prediction model. In preset 156 test cases, the mean square error of the model for heat load prediction is 0.201 and the mean square error of cooling load forecast is 2.56. The blending model reduces the error to a smaller range and reduces the phenomenon of data overfitting.

1. Introduction
In recent years, the continuous acceleration of China's urbanization process has driven the rapid growth of building energy consumption. China's building energy consumption accounted for 20.62% of the country's total energy consumption in 2016, of which public building energy consumption accounted for 38.53% of the total building energy consumption [1]. Commercial buildings are typical units in public buildings, with the characteristics of high power consumption per unit area. With the rapid economic and social development, the area of commercial buildings is also rapidly expanding, which makes the load level of commercial buildings continue to increase. Accurate prediction of the cooling and heating loads of commercial buildings provides reliable data for power grid dispatching, and plays a prominent role in guiding power equipment operation, power grid economic dispatch and power marketing.

Previous studies carried out on building cooling and heating load prediction methods are mainly divided into energy consumption software simulation prediction, traditional prediction methods based on statistical regression analysis, and artificial intelligence prediction algorithms based on machine learning. The method of using energy consumption simulation software is generally based on software platforms such as EnergyPlus, TRANSYS, DOE-2, etc., to build a detailed building model and give relevant design parameters, and obtain relevant load forecast values through computer simulation. T.T. Chow [2] realizes the cooling load prediction of commercial buildings, office buildings and residential buildings in a certain area of Kowloon, Hong Kong through DOE-2 software, and achieved good results. However, the method of energy consumption simulation software for load
forecasting requires relatively high requirements for operators and basic data, and is often only used as an auxiliary method rather than the mainstream method of load forecasting. The traditional forecasting method based on statistical regression analysis, which mainly uses a large amount of basic data to establish the functional relationship between load and influencing factors, has a simple principle and a wide range of applications. Linda Pedersen [3] uses regression analysis to predict the daily time-of-use load of buildings. However, the model established by this algorithm has low prediction accuracy and has difficulty in describing the complex nonlinear model. Therefore, experts and scholars tend to adopt artificial intelligence prediction algorithms with self-learning capabilities.

Artificial intelligence prediction algorithms have a high degree of nonlinear fitting ability and high prediction accuracy. Commonly used artificial intelligence prediction algorithms include artificial neural networks [4], support vector machines, random forest regression, bagging and extra random trees. The Bagging algorithm proposed in [5] is a group learning algorithm in machine learning, which can not only further improve the accuracy and stability, but also reduce the variance of the results to avoid overfitting. The random forest algorithm is an ensemble learning model that uses decision trees as the basic classifier. It not only overcomes the problem of overfitting in decision trees that may reduce accuracy, but also has good tolerance to noise and outliers. The random forest algorithm also has excellent expansion capabilities for high-dimensional data classification problems [6-7]. The extra random tree [8] is a variant of the random forest algorithm. Each decision tree in the extra random tree uses all training samples, which makes the calculation amount of the extra random tree increase relative to the random forest algorithm. Random forest uses optimal bifurcation features in the feature subset, while extra random tree uses random selection bifurcation features, which reduces the calculation process of information gain and slightly reduces the amount of calculation.

Artificial neural network was used to predict the load type of each prediction point [9], and then creates a support vector machine model based on the data points of the same type as the prediction point. This method [9] indicates that the prediction accuracy and calculation speed of the model are better than other traditional methods. Another research [10] proves the effectiveness of bagging algorithm in short-term load forecasting of smart grid. It firstly trains the large-load data set through the convolutional neural network [10], and then divide the actual load data set into multiple subsets, and autonomously sample from the subsets for weak learning to execute the bagging algorithm. Jetcheva [11] combines neural networks and cluster analysis algorithms in data mining to predict the load of commercial buildings. The obtained prediction accuracy is about 5% higher than the traditional time series analysis method. Huo J [12] applies the support vector machine algorithm and the random forest algorithm to different data sets to compare the differences between the two machine learning algorithms. The results show that both algorithms perform well in load forecasting. The choice of algorithms should be determined by specific parameters and data, and support vector machines have higher requirements for parameter settings.

Previous studies carried out on building cooling and heating load forecasting did not consider building feature factors comprehensively, and the modeling method was relatively simple. In this paper, the area, material, light transmittance, housing orientation, and relative density of the commercial building itself are considered as factors, and typical machine learning methods such as random forest, extra random tree, bagging, and deep neural network are used for regression modeling. Three of the best models were selected for blending to generate the final building cooling and heating load model. The method proposed in this paper has the characteristics of high accuracy and low error.

2. Feature Engineering
This paper collects the cold loads, heat loads and attributes parameter of 756 buildings in a certain area on a typical day in spring. The building attributes including relative density, surface area, wall area, roof area, overall height, house orientation, insulation material area, and insulation area distribution. Figure 1 shows the data distribution of each attribute.
In order to further analyze the data characteristics, we first calculate the correlation coefficient of each attribute according to formula (1).

$$r(x,y) = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x)\text{var}(y)}}$$

(1)

Where $\text{cov}$ denotes covariance, $\text{var}$ denotes variance.

Figure 2 shows the results of the correlation analysis.

According to the correlation distribution in the figure 2, it can be seen that the load is highly correlated with building height, wall area and relative density. However, this distribution only measures the linear relationship between the two variables, and cannot fully demonstrate the relationship between the load and the building attributes.

In order to further filter the important features in the data set, we use the Recursive Feature Elimination method. We choose Extra Random Tree as the base model, and train this model on the dataset for multiple rounds. After each round of training, one of the features is removed according to the training error rate. Then, the model is trained again. The influence of this removed feature on the load can be judged by the effect difference.
before and after removal.

Table 1 shows the results of the feature ranking. From the results, it can be seen that the main features that affect the cooling load and the heating load are different. Heat load is more sensitive to building surface area, roof area, height and light transmission area; while cooling load is more sensitive to surface area, wall area, height and light transmission area. The result provides a basic direction for feature selection for subsequent model construction.

| Heat Load | Cold Load |
|-----------|-----------|
| Relative compactness | Surface area | Wall area | Roof area | height | orientation | Glazing area | Glazing distribution |
| 4         | 1         | 3         | 1         | 1       | 6         | 2           | 5              |
| 3         | 1         | 1         | 6         | 1       | 5         | 2           | 4              |

### 3. Regression Models

This paper firstly applies linear regression, Lasso regression, Ridge regression, Bagging, random forest, extra random tree, KNN, deep neural network and other typical machine learning models to building load forecasting. Then choose the best three models for blending to further improve the accuracy of prediction.

#### 3.1. Linear Regression

The linear regression model uses the least square function of the linear regression equation to model the relationship between multiple independent variables and dependent variables. The linear regression equation is shown as function (2).

\[
y = \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_n x_n \quad (2)
\]

Where \( y \) is the dependent variable, \( x \) is the independent variable, \( \beta \) is the impact factor of each variable.

#### 3.2. Lasso Regression

The Lasso regression model improves the linear regression model to prevent the regression model from overfitting the basic model. For linear regression models, the optimized objective function is to minimize the error between the true value and the predicted value. The Lasso regression model adds an L1 norm penalty factor to the optimization target, which limits the complexity of the model and prevents overfitting. The specific optimization function is as follows:

\[
J(\omega) = \frac{1}{m} \sum_{i=1}^{m} (y_i - \omega^T X_i)^2 + \lambda ||\omega||_1
\]

where \( \omega \) denotes matrix of all impact factors \( \beta \), \( \lambda \) denotes the penalty factor. The penalty term for Ridge regression is the L2 norm.

#### 3.3. Bagging

Bagging is an aggregation learning technique that repeatedly sampled and trained from data according to a uniform probability distribution. Firstly, the algorithm samples the data set, generates a self-service sample set of the same size as the original data set, and generates a weak regression model based on the self-service sample set. Then, the algorithm repeats the above process and average all regression models to form a strong regression model.

#### 3.4. Decision Tree

Decision tree is a supervised learning algorithm. It applies to categorical and continuous input (features) and output (predictor) variables. The tree-based method divides the feature space into a series of rectangles, and
then places a simple model for each rectangle. In the construction of the regression tree, the space is divided by hyperplanes, and the current space is divided into two during each division, so that each leaf node is a disjoint area in the space. When making a decision, the decision tree will search step by step according to the value of each dimension of the input sample. Finally, the sample falls into one of the N regions (assuming there are N leaf nodes), and what is obtained at this time is the predicted value of the sample.

3.5. Random Forest
Random forest also obtains multiple sample data sets through multiple replacement sampling. Unlike Bagging, Random Forest uses decision trees as the base model. The random forest algorithm first randomly selects k features from the data set, and then builds n decision trees through different random combinations based on these k features. Random variables are passed to each decision tree to predict the outcome. The forest stores all the predicted results and can get n kinds of results from n decision trees. The forest calculates the number of votes for each predicted target, and uses the predicted target with the highest number of votes as the final prediction of the random forest algorithm.

3.6. Extra Random Tree
The extra random tree model is very similar to random forest, except that it has a more random feature selection process. Compared with the random forest method: the random forest obtains the best bifurcation attribute of the tree in a random subset, while the extra random tree, realizes the bifurcation of the decision tree by obtaining the bifurcation value completely randomly for all data.

3.7. Blending
After training the first-layer model on the data set, this paper uses model fusion to fuse the prediction results of these models to further improve the prediction accuracy. The overall flow chart of the algorithm is shown in Figure 3.

![Figure 3. Curve: system result of standard experiment](image)

4. Case Studies

4.1. Cross-validation
Based on the data of 768 buildings, this paper uses the leave-one-out method to cross-validate each model to verify the stability of the model. First, we divide the data set into 10 parts, each time 9 parts are taken to train the model, and the other part is used to test the reliability of the model, and so on for 10 times, recording its tie value and standard deviation.

Table 2 shows the cross-validation results of different algorithms. The results show that linear regression (Lir), Ridge regression, Lasso regression and ElasticNet regression have deviations in the performance of cold load and heat load prediction. The above algorithms are more sensitive to data changes, cannot guarantee stable performance, and the prediction accuracy is not as good as other algorithms. The results of KNN are better than various linear regression models. The CART classification tree model divides the feature space into several units, and each divided unit gets a specific output value.

This paper uses all architectural features to perform regression operations. After the decision tree is
obtained, Reduced-Error Pruning (REP) is adopted for pruning. If the effect of the current decision subtree is lower than the threshold, stop the branch of the current decision tree. The results of the pruning algorithm have been significantly higher than the linear regression and KNN model, which approximates the best model performance. The results also show that the fusion models Bagging, Random Forest and Extra Random Forest have high prediction accuracy and stability. The fusion model controls the standard deviation at a low level, which significantly improves the data fitting performance of the model compared with other models.

|                | Lir | Ridge | Lasso | ElasticNet | Bagging | RF    | ETR   | KNN   | CART  | SVM  |
|----------------|-----|-------|-------|------------|---------|-------|-------|-------|-------|------|
| Heat Load      | 0.894| 0.888 | 0.739 | 0.751      | 0.969   | 0.968 | 0.969 | 0.885 | 0.968 | 0.837 |
| Heat Load      | 0.065| 0.073 | 0.185 | 0.184      | 0.083   | 0.0827| 0.0823| 0.184 | 0.082 | 0.194 |
| Cold Load      | 0.786| 0.870 | 0.751 | 0.762      | 0.961   | 0.961 | 0.9512| 0.924 | 0.954 | 0.859 |
| Cold Load      | 0.032| 0.036 | 0.101 | 0.098      | 0.023   | 0.024 | 0.020 | 0.08  | 0.02  | 0.076 |

4.2. Commercial Building Load Forecast and Error Analysis.

In order to further improve the accuracy of the prediction model, according to the results of cross-validation, we first choose Bagging, Random Forest and Extra Random Tree as the basic model, and then merge the three to improve the performance of the model. In this paper, mean square error (MSE) is selected as the evaluation index of load forecasting.

Figure 4 and Table 3 show the prediction results of various models and the effect of model fusion. It can be seen from the results that these three models have high performance in the task of building load forecasting, especially the heat load forecasting, but there are still obvious errors in the cooling load forecasting. Therefore, in order to improve the accuracy of load forecasting and the generalization performance of the model, we adopt the method of model fusion to weight the prediction results of the three models. It can be seen from the results that, compared with Bagging and Random Forest, Extra Random Forest has better performance and higher randomness in the generation of tree bifurcation attributes.

Based on model fusion, the prediction accuracy of heating load and cooling load has been improved. Model fusion gives a higher weight to the objects with more accurate predictions of each model, which is equivalent to retaining the advantages of each model and further improving the performance of the model.

**Figure 4. Curve: system result of standard experiment**
### Table 3. System resulting data of standard experiment

|                  | RF    | Bagging | ET    | Blending |
|------------------|-------|---------|-------|----------|
| Heat Load (MSE)  | 0.255 | 0.249   | 0.22  | 0.201(+0.019) |
| Cold Load (MSE)  | 2.887 | 2.89    | 2.76  | 2.56(+0.2) |

### 5. Conclusions

This article applies linear regression, Lasso regression, Ridge regression, Bagging, random forest, extra random tree, KNN and other typical machine learning models to building load forecasting. According to the results of cross-validation, three models with better performance are selected for model fusion. The following conclusions can be drawn from error analysis:

1) The linear regression method has the lowest prediction accuracy among various machine regression models. The prediction accuracy of linear regression can be improved by adding feature selection and regularization algorithms.

2) The model stability of KNN and decision tree algorithm is not high enough, which causes overfitting of predictions that originally meet the extreme value.

3) Bagging, random forest and extreme tree algorithms have great performance on load prediction. The idea of ensemble forecasting effectively describes the nonlinear relationship between building load and various factors. Moreover, ensemble prediction avoids over-fitting and improves the stability and generalization performance of the model. By integrating the advantages of various models, model fusion further improves the prediction performance on the basis of existing models.

### Acknowledgement

The work is sponsored by the scientific and innovation project of Guangdong Power Grid Corporation under the grand number GDKJXM20180577.

### References

[1] Building. 2018 China Building Energy Research Report[J]. 2019(02):26-31.
[2] T.T. Chow, K.F. Fong, A.L.S. Chan. Energy modelling of district cooling system for new urban development[J]. Energy & Buildings, 2004, 36(11): p.1153-1162.
[3] Linda Pedersen, Jacob Stang, Rolf Ulseth. Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems[J]. Energy & Buildings, 2008, 40(7):1124-1134.
[4] Dengfu Z, Tao Z, Zenghui Y, et al. Short-term Load Forecasting Using Radial Basis Function (RBF) Neural Networks Baswd on GN BFGS Algorithm[J]. Automation of Electric Power Systems, 2003(04):33-36.
[5] Breiman L. Bagging Predictors[J]. Machine Learning, 1996, 24(2):123-140.
[6] Leo Breiman. Random Forests[J]. Machine Learning, 2001, 45(1):5-32.
[7] Shishi D, Zhexue H. A Brief Theoretical Overview of Random Forests[J]. Integrated Technology, 2013, 002(001):1-7.
[8] Pierre Geurts, Damien Ernst, Louis Wehenkel. Extremely randomized trees[J]. Machine Learning, 2006, 63(1):3-42.
[9] Niu D X, Wanq Q, Li J C. Short term load forecasting model using support vector machine based on artificial neural network[C]// Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on. IEEE, 2005.
[10] Dong X, Qian L, Huang L, et al. A CNN based bagging learning approach to short-term load forecasting in smart grid[C]. Ubiquitous intelligence and computing, 2017: 1-6.
[11] Jetcheva, Jorjeta G, Majidpour, Mostafa, Chen, Wei-Peng. Neural network model ensembles for building-level electricity load forecasts[J]. Energy & Buildings, 2014, 84:214-223.
[12] Huo J, Shi T, Chang J. Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources[C]// IEEE International Conference on Software Engineering & Service Science. IEEE, 2017.