A combined forecasting method for short term load forecasting based on random forest and artificial neural network

Chunming Yuan\(^1,\ast\), Yuanying Chi\(^1,a\), Xiaojing Li\(^2,b\)

\(^1\)Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing, China
\(^2\)State Grid Control Center of Gansu Electric Power Company, Lanzhou, China

\ast\)Corresponding author e-mail: 857431571@qq.com, \(a\)goodcyy@126.com, \(b\)luckey369@qq.com.cn

Abstract. Electric energy is closely related to people's life, in recent years, the construction of smart grid has already been proposed. Short-term load forecasting is a research hotspot in the process of smart grid. In this paper, we proposed a combined forecasting method based on random forest and artificial neural network, the final result is the weighted sum of the two single models, and the weight of each single model is obtained by the least square method. The data of experiment is the load data of a power plant in Hunan province from 2012 to 2017, and the corresponding weather information, the sampling granularity of the data is 15 minutes. The combined model we proposed can combine the advantages of random forest and artificial neural network, and the result of experiment shows that the combined model improves the accuracy of short term load forecasting.

1. Introduction
Accurate load forecasting can bring huge economic benefits to power companies, according to the study by British researchers, if the error of load forecasting increases by 1%, the operating cost of power companies will increase by 17.7 million. Therefore, power companies pay more attention to the load forecasting methods. Load forecasting methods are divided into three types according to the time intervals, the time interval for short-term load forecasting is the next few hours to weeks, medium-term load forecasting is the next weeks to months, usually within one year, and we call it long-term load forecasting when the time interval exceeds one year. Different types of load forecasting have different meanings for power companies. Short-term load forecasting is of great significance, the power company can perform machine scheduling and system maintenance based on the result of short-term load forecasting. Short term load forecasting often relies on the weather information of forecast day as auxiliary information, this can further improve the accuracy of prediction.

In recent years, many scholars have conducted in-depth research in this field, and put forward a lot of effective methods. These methods are mainly divided into two categories. One is statistical model, for example, regression analysis and time series method etc, this traditional method does not consider the effect of temperature and holidays on the load value, so the effect of prediction is usually not very good. For example, S. N. Dodamani who is a professor from Gogte Institute of Technology, proposed a short-term load forecasting method based on time series analysis [1]. Another is the method of artificial intelligence, including support vector machine, artificial neural network and expert system etc. This
method takes full account of the features that affect load value, so the accuracy of this method is relatively high. Saurabh Singh applied artificial neural networks to short-term load forecasting, meteorological information is also taken into account, [2]. Weicong Kong proposed a short term residential load forecasting method based on residents behavior learning, it tells us that there is a great relationship between electricity consumption and residents’ living habits [3]. The scholars from Madan Mohan Malaviya University of Technology compared the performance of different neural network models, which including Levenberg Marquardt back propagation, Bayesian Regularization and Scaled Conjugate Gradient algorithms etc [4]. Researchers in Turkey proposed a novel short-term load forecasting approach using adaptive neuro-fuzzy inference system [5]. Glen Mitchell from the university of the West Indies proved that artificial neural network is better than SVM[6]. The above all are the methods of the single model, Some scholars have put forward the combined forecasting model methods. For example, Zhang Yi proposed a method based on load trend and fuzzy C-means clustering algorithm [7]. The scholars from Quang Ninh University of Industry proposed an effective approach to ANN-based short-term load forecasting model using Hybrid Algorithm GA-PSO [8]. In [9], the author proposed a short-term load forecasting model based ridge let neural network optimized by particle swarm optimization algorithm, which further improved the precision and accuracy of prediction.

In this paper, we proposed a combined model based on random forest and artificial neural network. The random forest is a widely used integrated algorithm, strong generalization, fast learning. Artificial neural network has strong self-learning ability, but it is easy to get into the local optimal solution. So the better result can be obtained by combining the two algorithms. The result of experiment also proves that the combined model is more effective. The paper is organized as follows, Section II introduces the methods used in the experiment, which include random forest and artificial neural network. Section III is the whole experiment process, including feature engineering, the structure of the combined model and the evaluating indicator. Section IV is the analysis of the experimental results. The last part is the summary of the whole paper and the plan of the future work.

2. Method

2.1. Random forest
Random forest is a popular integrated algorithm, which consists of two parts, classification and regression tree (CART) and bagging algorithm. Random forest can process high-dimensional feature data, and it does not need to do feature deletion, it can give the importance of each feature. But when dealing with regression problems, it cannot give continuous output values.

2.1.1. CART. CART is a decision tree algorithm proposed by Breiman. L in 1984. Like other decision trees, for example, ID3, C4.5 etc, it consists of three steps, selection of the features, generation of the trees and pruning of the trees. Figure 1 is a classical decision tree model, which consists of Root Node, Sub Node and Leave Node etc. The decision tree divides sub-nodes according to the selected features, and the order of feature selection should be based on specific indicators. CART can deal with both classification and regression problems, when dealing with classification problems, it regards Gini index as the criterion of splitting nodes. As shown in formula (1), the sample has a total of N types, the probability of belonging to the kth type is \( p_k \). When dealing with regression problems, it takes minimum variance as the criterion of splitting nodes. The formula for calculating variance is shown in below (2).

\[
\text{Gini}(p) = \sum_{k=1}^{N} p_k (1 - p_k)
\]

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sum_{i=1}^{n} x_i^2 - n \mu^2}}
\]
2.1.2. Bagging. Bagging is an integrated idea which is introduced by Breiman. It selects samples from the original samples by bootstrap sampling, this method can improve the generalization of decision tree. The bootstrap sampling makes 36.8% samples will not be chosen, so this method can effectively solve the problem of over-fitting. Random forest adopts this integration idea, the generation process of random forest is shown as below.

1) Selecting n samples from the original sample by bootstrap sampling;
2) Selecting K features from all the features randomly;
3) Using the samples and features selected above to generate decision tree, repeat M times, and will get M decision trees;
4) For the new input data, using the results of M decision trees as the final prediction.

For classification problems, the most common category in the M decision trees is the final classification results, for regression problems, the mean of M models is used as the final result, as shown in formula (3). \( h_k (x) \) is the result of the k-th decision tree., and M is the number of the decision trees.

\[
Y = \frac{1}{M} \sum_{k=1}^{M} h_k (x)
\]  

(3)

2.2. Artificial neural network

Artificial Neural Network (ANN) has a strong ability of non-linear fitting, it also called multilayer perceptron. The structure of the artificial neural network is shown in figure 2, it has one hidden layer, layer X is the input layer, layer H is the hidden layer, and layer Y is the output layer. Forward propagation of the network is shown in formula (4)-(6), W is the weight value between neurons, and b is the bias term, among f is the activation function, which is sigmoid, tanh etc.
3. The combined model

The data of this experiment are the load value and weather data of a power plant in Hunan province from 2012 to 2017, the data of the first five years are used as the training set, and the data of 2017 is used as the test set, which is used to evaluate the effect of the models. The sampling granularity of the data is 15 minutes, such fine-grained data can improve the accuracy of prediction, and the weather data includes temperature and humidity.

3.1. Feature Engineering

There are many unexpected situations in the process of power production, so there are missing values and outliers in raw data, these abnormal values will reduce the accuracy of the model, therefore, we have to do the data preprocessing firstly. If the difference between the load at t and the load at t-1, or the difference between the load at t and the load at t+1, exceeds the fixed threshold $\beta$, the value of the load at t is considered to be an outlier. Because the time interval of the raw data is very short, the difference between adjacent moments should not be too large. For these abnormal values, we use the mean of the before and after values to represent, as shown in Formula (7).

$$x(t) = \frac{1}{2} x(t - 1) + \frac{1}{2} x(t + 1)$$  \hspace{1cm} (7)

The input features of the model include the moment of a day, workday or not, load value at the same moment last week, temperature and humidity. Besides, the holiday is considered as non-working day. For the moment of a day, [0,1,2,3,...,95] represents the 96 moments of a day, and 0 represents the working day, 1 represents the non-working day. The input features of the model are shown in Figure 3. Because the range of intervals for each input feature is very different, we need to normalize all features to the range of [0,1] by formula (8). this can speed up the convergent rate of the model.
\[
\hat{y} = \frac{y - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}} \tag{8}
\]

3.2. Principle of the combine model

The combine model consists of two stages, firstly, obtains the prediction of the two single models, then, the prediction of the combine model is the weighted sum of two single models. As shown in formula (9), Load combine is the final result, the sum of weights must be equal to 1, as shown in formula (10). The optimal weight is obtained by least square method, formula (11) is the objective function which we want to minimize. The cost is the sum of the differences between the predicted and the actual values of all samples, M is the number of samples, and the second stage of the training is to minimize the cost. Figure 4 is the flow chart of the combine model.

\[
\text{Load}_{\text{combine}} = W_1 \ast \text{Load}_{RF} + W_2 \ast \text{Load}_{ANN} \tag{9}
\]

\[
W_1 + W_2 = 1 \tag{10}
\]

\[
\text{cost} = \frac{1}{2} \sum_{i=0}^{M} (\text{Load}_{\text{combine}}^i - \text{Load}_{\text{actual}}^i)^2 \tag{11}
\]
3.3. Experimental parameters and evaluation indexes

In order to evaluate the predictive effect of the different models, we use mean absolute percent error as the criterion for judging. The calculation formula is shown in (12). N is the number of the samples, Load\textsubscript{pre} is the predicted value, and the Load\textsubscript{act} is the actual value.

\[
\text{MAPE} = 100 \times \frac{1}{N} \sum_{i=0}^{n} \left| \frac{\text{Load}_{\text{pre}} - \text{Load}_{\text{act}}}{\text{Load}_{\text{act}}} \right|
\]  \hspace{1cm} (12)

The experimental environment is python (3.5.2), the framework of deep learning is Keras (2.1.4), the library of machine learning which is sklearn (0.19.1). The parameters of random forest and neural network are shown in table 1, including the number of trees, the number of candidate variables, minimum node size, and the ANN has a hidden layer, the number of neurons in the hidden layer is 20. Activation function is relu, the learning rate is 0.001.

| Model            | Parameters                                      |
|------------------|-------------------------------------------------|
| Random Forest    | $N_{\text{tree}}=500$ (number of trees)        |
|                  | $M_{\text{try}}=3$ (number of candidate variables in each split) |
|                  | $Z_{\text{min}}=5$ (minimum node size)         |
| ANN              | ANN has a hidden layer. The number of neurons in the hidden layer is 20. Activation function is relu, learning rate is 0.001. |

Figure 4. The flow chart of the combine model.
4. Analysis of experimental result

In order to clearly observe the predictive effect of three models, we only show the forecast curve of May 1, 2017, it contains 96 moments. Figure 5 is the fitting curve of artificial neural network, figure 6 is the fitting curve of random forest, and figure 7 is the fitting curve of the combine model. The curve of the original load is plotted with red, the curves of predicted values are plotted with other colors. From the graph, we can see that the curve of our combined forecasting model and the original curve are more fitting, this shows that the combined model we proposed is better than the single models.

Figure 5. Fitting Curve of ANN

Figure 6. Fitting Curve of Random Forest
Figure 7. Fitting Curve of the combine model

Random forest is widely used as a popular integration method, it learns very fast, and the importance of each feature can be given. But when there is a specific noise in the data, random forest will be overfitting. The learning speed of artificial neural network is very slow, and it is easy to get into the local optimal solution, but it has strong non-linear fitting ability. The combined model proposed in this paper combines the advantages of the two single models. The mean absolute percent error of the three models for the whole year of 2017 is shown in table 2. The MAPE of the combine model is the smallest, which is 2.385%, it is a good result for the short-term load forecasting.

Table 2. MAPE for the whole year 2017.

| Model          | Random forest | ANN  | Combine model |
|----------------|---------------|------|---------------|
| MAPE           | 2.793         | 3.730| 2.385         |

5. Conclusion
In this paper, we conducted a short-term load forecasting experiment. We introduced two single models to short-term load forecasting, which are random forest and artificial neural network. The proposed combine model is based on the two single models, the final prediction result is the weighted sum of the predicted values of the two single models. The model takes into account the meteorological information and the past load values. The result of experiment shows that the proposed combine model is better than the two single models. So this method has great significance for short-term load forecasting, it will bring more economic benefits to power companies. In the future work, we can try to introduce the combine method to medium and long term load forecasting.

Acknowledgments
This work was financially supported by State Grid Gansu Electric Power Company (52272216002B) and National Natural Science Foundation of China (91646201).

References
[1] S.N.Dodamani, V.J.Shetty ,R.B.Magadum, Short Term Load Forecast Based on Time Series Analysis: A Case Study, (2015) 299 - 303.
[2] Saurabh S, Shoeb H, Mohammad A B, Short Term Load Forecasting Using Artificial Neural Network, (2017) 159 - 163.

[3] Weicong K, Zhao Y, Short-Term Residential Load Forecasting Based on Resident Behaviour Learning, (2018) 1087 - 1088.

[4] Kishan B S, Suneet S, Pragya S, Short-Term Load Forecasting of Toronto Canada by Using Different ANN Algorithms, (2016).

[5] Emre A, Fatih Onur H, A Novel Short-term Load Forecasting Approach Using Adaptive Neuro-Fuzzy Inference System, (2018) 160 - 163.

[6] Glen M, Sanjay B, A Comparison of Artificial Neural Networks and Support Vector Machines for Short-term Load Forecasting using Various Load Types.

[7] Zhang Y, Zhang F, Zhu B Q, A novel ultra-short term load forecasting method based on load trend and fuzzy c-means clustering algorithm, (2014).

[8] Manh-H Pham, T-A Tho Vu, Thu-H Y Dang, An Effective Approach to ANN-Based Short-Term Load Forecasting Model Using Hybrid Algorithm GA-PSO, (2018).

[9] Wang Q, Zhang Y B, Zhu X Y, Short-term Load Forecasting Model Based on Ridgelet Neural Network Optimized by Particle Swarm Optimization Algorithm, (2017) 727 - 731.