The Load Forecasting Method based on Characteristic Analysis and Combination Learning

Zhijun Zhang*, Ang Li, Xin Qi and Qiang Li
State Grid Tianjin Chengnan Power Supply Branch Company, Tianjin, China

*Corresponding author e-mail: patrol20@126.com

Abstract. Load forecasting is an important basis for power system planning and safe and economic operation. This paper presents a load forecasting model based on feature ranking and combination learning. In view of the great difference of regional load, firstly, the random forest algorithm is used to sort the factors which have great influence on the prediction target. Then, the model selects the characteristic attributes with high characteristic contribution, dynamically combines the prediction results of the extreme learning machine, AdaBoost and neural network model, and updates the weights in a certain period through Lasso algorithm to obtain the final prediction results. Finally, the actual load data of Tianjin Power Grid is used for example verification. The results show that the prediction model established in this paper has good prediction accuracy and stability.

1. Introduction
Load forecasting of power system refers to forecasting the demand of power load in a certain period of time in the future by some mathematical method [1-2]. For a long time, scholars at home and abroad have done a lot of research on the theory and method of short-term load forecasting. The traditional method is based on the prediction principle of time series, including autoregressive method, autoregressive moving average method and cumulative autoregressive moving average method. The method is widely used because it requires less data and a simple model. But it requires high stability of the original time series. What is more, the prediction error is large [3-4].

In addition, the other is based on machine learning, which has many mature models, such as Support Vector Machine [5], Grey Model [6], Markov Chain [7], Combined Prediction Model [8], etc. The method based on historical data has higher prediction accuracy. Through in-depth analysis of the correlation between load and weather, a load forecasting model based on the combination of weather forecast and load classification is proposed [9]. Considering the correlation factors of daily features, and selecting the most similar date load of the predicted target, some scholars introduce the concepts of model prediction accuracy and prediction validity, study the transfer law of model validity, and propose a quantitative estimation method of prediction accuracy based on Markov Chain and cloud model [10]. Considering the uncertain factors of wind power and hydropower, some scholars simplify the attributes of historical samples based on differential evolution rough sets and propose a net load prediction method that considers the access of distributed power [11]. However, the above-mentioned traditional model does not analyze the difference between load forecasting in multiple scenarios and conventional load forecasting, and does not propose targeted forecasting methods. The prediction
accuracy of the existing methods has reached a certain bottleneck. How to effectively use the artificial intelligence and machine learning algorithm with strong learning ability is the key to solve the problem of load prediction according to the application scenarios and characteristics of various loads.

In this paper, a load prediction model combining feature sorting algorithm with machine learning model is proposed by referring to the latest research results of artificial intelligence algorithm. The following section introduces the mechanism of extreme learning machine with kernel and Adaboost. In section 3, the updated Lasso prediction model is established, in which the prediction results of each sub-model are weighted. The data of Tianjin Power Grid is used for example verification in section 4. The results show a reasonable configuration of the combined learning model can help to obtain the best prediction results. Especially, feature sequencing can effectively select the features with the greatest impact, and combination learning has good accuracy and stability for solving the load prediction problem. Conclusions are outlined in the last section.

2. Prediction Model and Algorithm Mechanism

2.1. Extreme Learning Machine with Kernel

Extreme learning machine with kernel belongs to single-layer feedforward neural network algorithm. The basic extreme learning machine is shown as follows.

\[ f(x) = h(x)\beta \] (1)

Where \( h(x) \) represents the calculated output of hidden layer, \( \beta = [\beta_1, \cdots, \beta_L]^T \) is the connection weight between the hidden layer and the output layer.

The error expression of extreme learning machine is as follows.

\[ \lim_{L \to \infty} \|f(x) - f_o(x)\| = \lim_{L \to \infty} \|\sum \beta h_i(x) - f_o(x)\| = 0 \] (2)

Where \( L \) is the number of neurons, \( f_o(x) \) is the true mark.

The basic architecture of the extreme learning machine is shown in Figure 1. The output function can be expressed as follows.

\[ f_L(x) = h(x)\beta = \sum_{i=1}^L g_i(x)\beta_i = \sum_{i=1}^L G(a_i, b_i, x)\beta_i \] (3)

Where \( g_i(x) \) and \( G(a_i, b_i, x) \) are the output function of the \( i \)-th hidden node, \( a_i \) and \( b_i \) are the parameters of the hidden layer, \( \beta_i \) is the weight vector.

The optimal binary solution for training the feedforward neural network is as follows.

\[ H\beta = T \] (4)

\[ \|H\hat{\beta} - T\| = \min_{\beta} \|H\beta - T\| \] (5)

Where \( T \) is the predicted target.

The minimum standard square solution of the output weight of the system is as follows.

\[ \beta = H^T = H^T (HH^T)^{-1}T = H^T (\frac{1}{C} + HH^T)^{-1}T \] (6)
Where $H$ is the matrix of the hidden layer of the neural network, $H^\dagger$ is the generalized inverse matrix of $H$. By increasing $1/C$, the solution becomes more general.

$$f(x) = h(x)H^T \left( \frac{I}{\lambda} + HH^T \right)^{-1} T = \begin{bmatrix} K(x,x_1) \\ \vdots \\ K(x,x_N) \end{bmatrix} \left( \frac{I}{\lambda} + \Omega_{ELM} \right)^{-1} T \tag{7}$$

$$\Omega_{ELM} = K(u,v) = \exp(-\gamma \|u - v\|) \tag{8}$$

Where $\Omega_{ELM}$ is the gaussian kernel function, $N$ is the dimensions of the input layer.

2.2. Adaboost

Adaboost (Adaptive Boosting) is a superimposed integration model, which trains several weakly fitting models and then combines them to form a strong prediction model. The correct samples are given a lower weight, while the wrong samples are given a higher weight. The performance of the prediction model is improved by continuous weighted operation and combination learning.

First, $n$ groups of training data are selected from the sample, and the distribution weight of the data is initialized.

$$D_1(i) = \frac{1}{n} \tag{9}$$

Second, when training the $t$-th weak learner, the training data is used to train the decision tree to obtain the sum of errors.

$$e_t = D_t(i) \tag{10}$$

Then, the weight of the weak learning model is calculated based on the sum of prediction errors. The formula for calculating the weight is as follows.
The basic architecture of the extreme learning machine is shown in Figure 1. The output function can be expressed as follows.

$$\alpha_i = \frac{1}{2} \ln \left( \frac{1 - e_i}{e_i} \right)$$  \hspace{1cm} (11)

Where \( i = 1, 2, \ldots, n \), \( B \) is the normalized factor.

After training \( T \) round, \( T \) group weak learners are obtained, and then a strong learner \( h(x) \) is obtained. The formula is as follows.

$$h(x) = \text{sign} \left( \sum_{i=1}^{T} \alpha_i \times f(g_i, \alpha_i) \right)$$  \hspace{1cm} (13)

### 2.3. Lasso

Lasso is a linear regression analysis method for feature selection and regularization at the same time. Its basic idea is to minimize the sum of squares of residuals when the sum of absolute values of regression coefficients is less than a threshold.

For the ordinary linear model, it can be expressed as follows.

$$Y = X\beta + \epsilon$$  \hspace{1cm} (14)

Where \( Y \) is the predicted value of the final load, \( X = \left( X^{(1)}, X^{(2)} \right) \) correspond to the predicted values of the first and second layer models respectively, \( \epsilon \) is random error term, \( \epsilon_i \sim N(0, \sigma^2) \), \( \epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^T \), \( \beta = (\beta_1, \beta_2, \cdots, \beta_d)^T \) is the regression coefficient, \( n \) and \( d \) are the number of regression coefficients. Penalty items are added to Lasso regression, resulting in Lasso estimates as follows.

$$\hat{\beta}_{\text{lasso}} = \arg \min \left( \| Y - X\beta \|^2 + \lambda \sum |\beta_j| \right)$$  \hspace{1cm} (15)

Where \( \lambda \) is the adjustment coefficient, \( \hat{\beta}_{\text{lasso}} \) is the dynamic weight corresponding to the predicted values of the final sub-models. In the process of model training, weights need to be solved, and the completion of weight calculation means the completion of load prediction model.

### 3. Load Forecasting Model based on Data Cleaning and Combination Learning

Firstly, the random forest algorithm is used to analyze the characteristics of the input data, which is helpful to find the different factors that affect the load in different regions. Secondly, considering that a single model tends to fall into a local minimum and the statistical hypothesis space is limited, this paper adopts a combined model to improve the reliability of the prediction algorithm. The prediction results of ELM, Adaboost and ANN (Artificial neural network) are further studied by Lasso to learn the advantages of each model.

At the same time, in order to ensure the real-time performance of the parameters in Lasso, the parameters in Lasso are changed through dynamic update, so as to obtain the time-series rolling load
prediction model and ensure the mutual matching between the current model and data. Figure 2 shows the time-series rolling mode of load forecasting.

![Figure 2. Load forecasting time horizon-rolling.](image)

The overall training process of the model is as follows. Firstly, random forest algorithm is used to sort the input data. Complete data are input into Adaboost model, ANN model and ELM model respectively for training. Finally, the prediction results and the original data are input into the Lasso linear model to obtain the final load power prediction results. Meanwhile, the relevant parameters in the Lasso model are updated in real time within a certain time threshold or when the error reaches a certain threshold. The model integration makes the prediction more stable. The block diagram of the model in this paper is shown in Figure 3.

![Figure 3. PV output forecasting based on data cleansing and model aggregation.](image)

4. Case Study
The actual operation data of Tianjin Power Grid is selected for example verification. In addition, in order to better analyze the weather information of the algorithm in this paper, the weather-related information comes from Numerical Weather Prediction (NWP). The forecast target is the load for the next hour. It should be noted that the data for the whole year of 2017 is training data, while the data for January 2018 is test data. The test machine uses Win 10 system and Core i5 processor, and compiles related programs in Matlab 12a.

The error indexes used in the example include Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f_i - t_i}{f_i} \right| \times 100\% \tag{16}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i - t_i)^2} \times 100\% \tag{17}
\]
Where $n$ is the number of samples, $f_i$ and $t_i$ are the actual and predicted load of the bus at time $i$.

4.1. Analysis of Feature Ranking
In this paper, stochastic forest algorithm is used to rank and analyze the importance of load prediction features. The features initially selected include three categories: meteorology, historical load, and time. Meteorological characteristics are five candidate variables obtained from Numerical Weather Prediction, including temperature, humidity, wind speed, rainfall and atmospheric pressure. The historical data is the load historical data of the first 6 moments before the predicted target time. Holidays and working days are taken into account in the time information. After the feature importance is calculated by the random forest algorithm, each feature score is shown in Figure 4.

![Input feature ranking](image)

**Figure 4. Input feature ranking**

Holidays have the highest score for input attributes, indicating that holidays have the greatest impact on load forecasting, which is consistent with common sense. Residents' habits of working and rest during holidays and working days are quite different, which makes their energy usage patterns quite different. The score of temperature characteristics ranked second, because residential load is also sensitive to temperature changes. The change of temperature difference may cause the switching of large-scale air-conditioning load, and the electricity consumption curve also changes accordingly. In addition, the load history information of several adjacent moments also provides good information for the prediction, and each moment also corresponds to a certain degree of characteristic importance.

4.2. Comparative Analysis of Prediction Effect
The prediction model in this paper includes a variety of prediction algorithms. Lasso algorithm dynamically learns the weight of each algorithm's prediction results and analyzes the dynamic trend of the corresponding weight predicted by each sub-model. The Lasso model weights for the temporal changes are shown in Figure 5, including the corresponding weights for ELM, Adaboost and ANN.
In order to further verify the accuracy and superiority of the algorithm in this paper, the combined learning prediction model in this paper was compared with the prediction results of neural network and time series algorithm under the prediction scenarios in the first week of June. The prediction results of each algorithm are shown in Table 1. The comparison with the prediction results of neural network is shown in Figure 6.

Table 1. PV output forecasting error evaluation considering multiple models.

| Scenario                          | Algorithm         | MAPE   | RMSE  |
|-----------------------------------|-------------------|--------|-------|
| forecasting results in the first week of June | Model in this paper | 3.71%  | 2.91% |
|                                   | Neural network    | 5.45%  | 4.43% |
|                                   | Time series       | 6.59%  | 6.02% |
| forecasting results in the third week of June | Model in this paper | 4.22%  | 3.13% |
|                                   | Neural network    | 5.97%  | 5.25% |
|                                   | Time series       | 8.55%  | 9.55% |

According to Table 1, compared with other algorithms, the combined learning model can better track the variation trend of load output. In terms of load prediction, the algorithm in this paper has obtained the highest accuracy, because the dynamic combination of various learning algorithms can
improve the generalization and robustness of the model. Compared with other algorithms, neural network, as a single learning method, tends to fall into local minima, which makes the overall generalization not high. Although this algorithm has achieved good prediction accuracy, it is still lower than the method used in this paper. The time series algorithm can only account for the historical information of the load. The prediction accuracy is low and the learning ability is the weakest. It can be seen that the combined forecasting model in this paper has high reference value for the power system to formulate the day-to-day operation strategy.

5. Summary
This paper presents a load prediction model based on stochastic forest feature analysis and combination learning algorithm. First of all, the random forest algorithm is used to rank the factors that have a greater impact on the predicted target and select the feature attributes with a higher contribution to the feature. Then, ANN, ELM and Adaboost models are selected in the model training phase. In addition, Lasso algorithm is used to learn and track the load variation trend. Finally, an example is given to verify that the feature ranking and combination learning algorithm has a good effect in load forecasting.

In the future, in view of load forecasting, the following aspects can still be further studied to improve the accuracy of load forecasting.

1) The scale access of distributed power supply brings new challenges to load forecasting. Constructing the characteristics of distributed power in substation and considering the influence of distributed power supply on local net load are problems worthy of further study.

2) The quality of the load data at the level of distribution network is generally poor. There are often many outliers and missing values in the load data, which require data supplement, cleaning and other processes to improve the quality of data. The optimization of data quality can effectively improve the prediction effect.

Acknowledgments
This work was financially supported by Kj-19-1-06 research and application of load structure analysis and accurate prediction based on big data under power market reform fund.

References
[1] R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, Solar forecasting methods for renewable energy integration, J. Progress in energy and combustion science. 6. 39 (2013) 535-576.
[2] H.T. Yang, C.M. Huang, Y.C. Huang, Y.S. Pai, A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output, J. IEEE Transactions on Sustainable Energy. 5. 3 (2014) 917-926.
[3] M.A. Eltawil, Z. Zhao, Grid-connected photovoltaic power systems: Technical and potential problems—A review, J. Renewable and Sustainable Energy Reviews. 14. 1 (2010) 112-129.
[4] Y. Ren, P.N. Suganthan, N. Srikanth. Ensemble methods for wind and solar power forecasting—A state-of-the-art review, J. Renewable and Sustainable Energy Reviews. 50 (2015) 82-91.
[5] P. Bacher, H. Madsen, H.A. Nielsen. Online short-term solar power forecasting, J. Solar Energy. 10. 83 (2009) 1772-1783.
[6] M. Rana, I. Koprinska, V.G. Agelidis, 2D-interval forecasts for solar power production, J. Solar Energy. 122 (2015) 191-203.
[7] F. Almonacid, P.J. Pérez-Higuera, E.F. Fernández, et al. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator, J. Energy conversion and Management, 85 (2014) 389-398.
[8] M. Bouzerdoum, A. Mellit, A.M. Pavan, A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant, J. Solar Energy, 98(2013) 226-235.
[9] F. Zhu, Study on PV generation power forecasting method based on KPCA and shuffled frog leaping algorithm, J. Renewable Energy Resources. 2. 36 (2018) 236-240.

[10] K. Cheng, L.M. Guo, Y.K. Wang, Application of cluster analysis in forecasting photovoltaic power generation. J. Renewable Energy Resources. 5. 35 (2017) 696-701.

[11] K. Ding, H.X. Ding, Y. Wang, L. Wang, L. Gao, Z.F. Liu, Short-term power prediction of photovoltaic array based on lifting wavelet transform- BP neural network, J. Renewable Energy Resources. 4. 35 (2017) 566-571.