

Short-term Load Forecasting Based on Wavelet Decomposition and XGBoost

Ningning Zheng¹, Yunfeng Shao², Suli Zou¹ and Zhongjing Ma¹⁺

¹ School of Automation, Beijing Institute of Technology, Beijing, P.R.China

² Lvliang Power Supply Company, Shanxi, P.R.China

Abstract. The development of intelligent power systems and the large-scale access of distributed power sources have continuously deepened the impact on the power distribution side, and have placed higher requirements on the accuracy of load forecasting. In this paper, a short-term load forecasting method based on wavelet analysis and XGBoost is proposed. First, use wavelet analysis to classify power loads in different frequency bands, and then use the XGBoost model for training and prediction of the classified loads. Finally, using real power data in a certain area as a sample, the average absolute percentage error (MAPE) and the mean squared error(MSE) are used to compare and analyze the 24th hour data predicted by XGBoost and SVM and LSTM, respectively. The results show that XGBoost has a better fit and higher accuracy for short-term power loads.

Keywords: power system; load forecasting; wavelet analysis; XGBoost

1. Introduction

With the rapid development of social data collection and storage technology, the amount of power load data and its complexity are increasing exponentially. Discovering the value of mass data on the user side and efficiently and rationally applying a large amount of data has become an urgent problem in the era of the smart grid[1].

Load forecasting is the basis of distribution network planning. Its accuracy directly affects whether the power system operates economically and safely, and affects the utilization of output power and economic dispatch. Therefore, accurate load forecasting is of great significance to the optimal combination of units, economic dispatch, optimal power flow and power market transactions[2].

At present, methods for short-term load forecasting are mainly divided into two categories. One category is the traditional statistical method. The classic models are multiple linear regression (MLR) and autoregressive moving average (ARMA)[3], etc. . This type of method has a simple model, but it can only process a small amount of sample data, and it cannot add too many influencing factors. The second category is artificial intelligence methods, including BP neural network[4], grey system prediction[5], support vector machine (SVM)[6], deep belief network (DBN)[7], etc., among which BP and SVM methods are the most widely used. Long short time memory (LSTM)[8] is an improved recurrent neural network model that has been popular in recent years. It can take into account both the time series and non-linearity of load data, and has high prediction accuracy. However, it is complex and the training speed is slow.

With the development of machine learning and computer technology, XGBoost came into being[9]. It has fast training speed and requires a small number of samples. Besides, XGBoost only needs time series data as character, and does not need temperature, humidity, rainfall, and date type to achieve the accuracy effect comparable to LSTM.

+ Corresponding author. Tel.: +86-13146916588.
E-mail address: zyy_bit@163.com.

The power load itself fluctuates in cycles of days, weeks, and years. In this paper, wavelet decomposition technology is used to smooth the original time series^[10], and then the XGBoost model is used to predict each subsequence. Finally, the real data from a regional in China was used for verification. The prediction results showed that after adding wavelet analysis, the accuracy of prediction used XGBoost was improved apparently.

2. Introduction and Application of Wavelet Analysis

The so-called wavelet transform is selecting a suitable mother wavelet $\psi(t)$ to generate derived wavelets based on displacement and scaling and amplitude changes on the time axis, and use these wavelets to compare the translation of the signal to be analyzed on the time axis. Because the derived wavelet can reach arbitrarily small prescribed accuracy and can accurately measure the finite-length signal, it can obtain the information of the local time interval. This method will have broad applications in many fields such as pattern recognition, state monitoring, fault diagnosis, harmonic analysis, and transient stability of power systems.

In continuous wavelet transform, we consider $\phi(t) \in L^2(\mathbf{R})$ and $\phi^*(w)$ is the fourier transform of $\phi(t)$. $\phi^*(w)$ satisfied:

$$\int_{-\infty}^{+\infty} \frac{|\phi^*(w)|^2}{|w|} dw < \infty \quad (1)$$

$\phi(t)$ is a mother wavelet function. After stretching and translation, $\phi(t)$ can be expressed as:

$$\phi_{a,\tau}(t) = \frac{1}{\sqrt{|a|}} \phi\left(\frac{t-\tau}{a}\right), a, \tau \in \mathbf{R}; a \neq 0 \quad (2)$$

Where a is a scaling variable and τ is a translation variable.

The number of wavelet decomposition layers is determined by the characteristics of the data and experience. Generally, three layers are used for testing. By performing the first-scale wavelet decomposition of the original load sequence to obtain the high-frequency signal d_1 , you can see the time at which the load suddenly changes, that is, the obvious sudden clusters in the figure. It can be seen that the places with large changes in load are roughly equal in interval, so there may be law. In addition, it can be seen that the load sequence changes very drastically in this frequency domain. At the same time, the low-frequency signal d_2 is obtained. Then d_1 is further decomposed into scale 2 and scale 3 to obtain signals: high-frequency signals d_3 , and low-frequency signal a_3 , as shown in Figure 1.

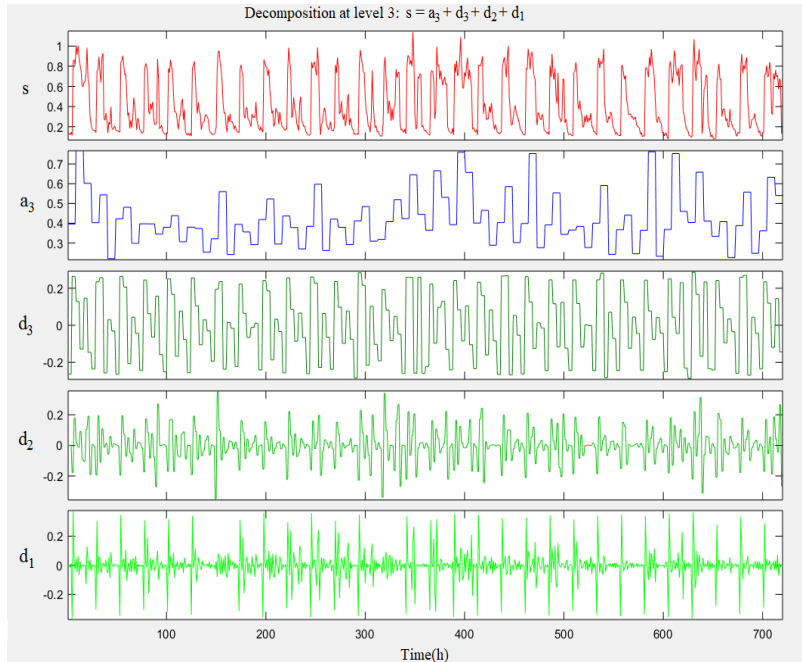


Fig. 1: Original load sequence and each decomposition signal.

After obtaining the decomposed signals, the XGBoost model training is performed on each wavelet sequence to obtain the prediction model of the load wavelet sample, and the error of the prediction model is analyzed to obtain the model prediction accuracy. Then the fourth order wavelet decomposition is performed on the original load sequence, and the model prediction is used again. It is found that the accuracy is almost not improved. Therefore, this paper uses the third order wavelet decomposition. The structure of the load prediction model under wavelet transform is shown in Figure 2.

In order to improve the accuracy of the reconstructed signal after wavelet transform, the selection of scale variables and translation variables should be small. The time series of the actual load data is a non-stationary signal. In this paper, the collected samples are analyzed and processed to obtain the input and output models. The data of each column is decomposed by wavelet transform. Considering the accuracy of signal reconstruction after wavelet transform, wavelet decomposition is performed on the samples according to the stationarity of the signal. For this load data sample, three-level wavelet decomposition is used, and the obtained wavelet signal is predicted and researched to judge the signal stability.

This paper uses the June load data from sampling sites in Shanxi Province, China. A total of $24 \times 30 = 720$ data were sampled, $24h$ per day, for a total of 30 days, recorded as the original sequence S . The load sequence S is decomposed using the Mallat algorithm in MATLAB, and Daubechies wavelet 1 (db1) was selected as the wavelet function.

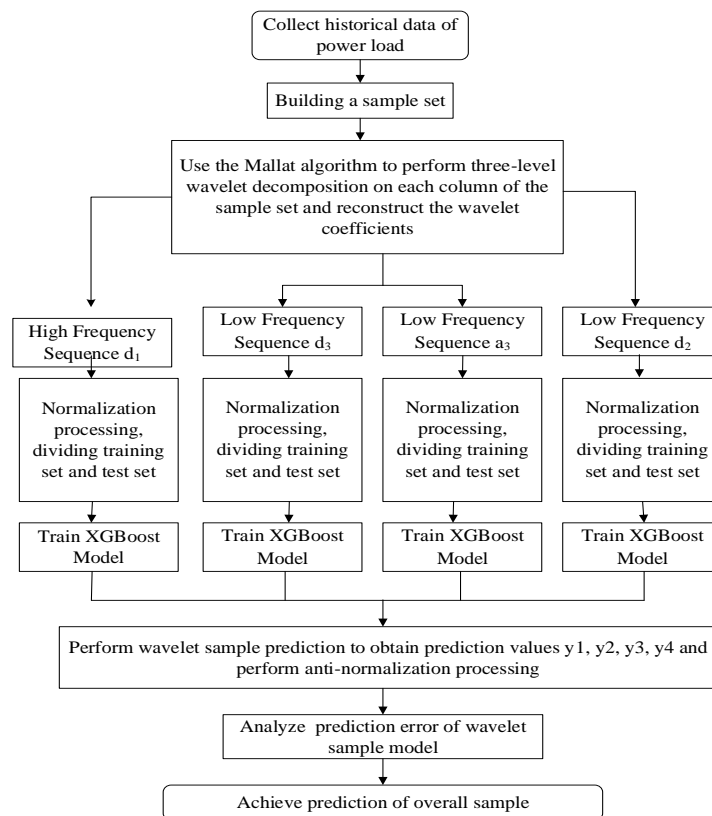


Fig. 2: Structure of XGBoost prediction model based on wavelet transform.

3. Introduction and Application of XGBoost

3.1. Design of XGBoost Model

Extreme Gradient Boosting (XGBoost) is used for supervised learning problems, by training sample data to predict target values.

In this paper, data preprocessing is performed before data prediction. In machine learning, the data preprocessing step is crucial because of the amount of grid data that can be collected is small, the number of data feature dimensions is small, and the relationship between the feature quantities is unknown. Therefore, this paper uses the small amount of data required by the XGBoost model and does not care about the feature relationship. The load data is used as the model input quantity. Only the vacant value filling and outlier removal

are performed on the sample load data, and then the data set is divided into a training set, a validation set, and a test set by using a “cross-validation method”.

Then the preprocessed training data is input into the XGBoost prediction model after the preliminary adjustment of the hyperparameters. The model uses the greedy algorithm to calculate the parameter values of each decision tree layer by layer to minimize the objective function. The validation set is used to judge the model. If the performance on the validation set is not good, continue to optimize the parameters until a better result is achieved and then the prediction model is saved.

Finally, the test data set is input to the saved prediction model to predict the load, and the evaluation indicators are calculated based on the true and predicted load values: mean absolute percentage error (MAPE) and mean square error (MSE), and the performance of the adopted model is evaluated by comparing these two values.

3.2. Construction of XGBoost Model

In this paper, the XGBoost model uses a CART tree as the classifier. The leaf nodes of the CART tree correspond to an actual score rather than a specific category, which facilitates the addition of multiple CART tree scores, which makes XGBoost algorithm more efficient when optimizing. The model building steps are as follows:

1. Initialize the common parameters of the algorithm and generate a new CART tree each iteration;
2. Before the start of each iteration, calculate the first derivative g_i and the second derivative h_i of the loss function at each training sample point;
3. Generate a new CART tree through the greedy strategy, and calculate the predicted value corresponding to each leaf node using (3):

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (3)$$

4. Add the newly generated decision tree $f_i(x)$ to the model: $\hat{y}_i^t = \hat{y}_i^{t-1} + f_i(x_i)$;
5. After the model is generated, use the function `sklearn.Model_selection.GridSearchCV` in the Scikit-learn library to adjust the parameters, and use the scoring function to judge the model.
6. Put the best combination of parameters into the model to train and get the prediction result.

In step 4, in order to avoid overfitting the model, change the model update formula to $\hat{y}_i^t = \hat{y}_i^{t-1} + \alpha f_i(x_i)$, where α is called the step size or learning rate.

The XGBoost model flowchart built in this paper is shown in Figure 3.

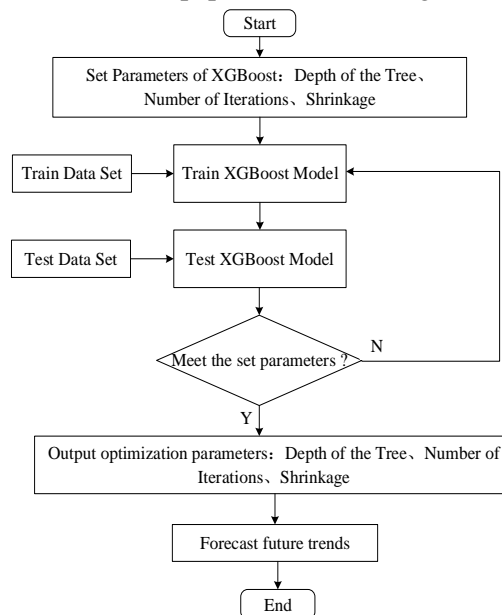


Fig. 3: XGBoost-based ultra-short-term load forecasting process for power grids.

4. Load Forecasting based on Wavelet Analysis and XGBoost

In order to reduce the complexity of the model and improve the generalization ability of the XGBoost model, the data input in this paper is 23 dimensions, the output is 1 dimension, the input vector is the load of the 23 historical moments before the current time to be predicted, and the output is the current load.

The sample of this time series is subjected to column-wise three-level wavelet decomposition using wavelet decomposition, and four wavelet subsequences are obtained as a_3, d_3, d_2, d_1 .

The steps of short-term load forecasting algorithm based on wavelet and XGBoost model are as follows:

1. Collect and analyze historical load data, use the 23 load values x_1, x_2, \dots, x_{23} before the current moment as input vectors, and use the load values at the 24th time in the future as output vectors to construct a training sample set.

2. The training samples are decomposed by the third-order wavelet using Mallat's algorithm to obtain four wavelet training sample sets.

3. Normalize the wavelet training samples.

4. Establish a XGBoost model to train each wavelet sample to obtain the prediction value of each wavelet sample and the prediction error of the model.

5. Add the prediction results of each wavelet sample to get the total predicted value.

6. Calculate the MAPE and MSE values predicted by the model after adding wavelet analysis, and compare with the corresponding values when only using the XGBoost model to evaluate its performance.

5. Simulation Results and Performance Comparison

This paper uses the load and temperature data from the 1997 and 1998 EUNITE load data competition as samples to establish the SVM model, LSTM network model, and XGBoost model, and select the most suitable model-XGBoost after comparing the accuracy and training time. The accuracy of each model is shown in Figures 4, 5, and 6, and the MAPE and MSE created by each model is shown in Table 1. For the total data, the first 90% of the sample data is selected as the training set, and the last 10% of the data set is used as the test set. The simulation uses Spyder to run the program with Python3. The processor is Intel (R) Core (TM) i5-8265U dual-core processor (1.60GHz, 1.80GHz), and the operating system is windows 10 operating system.

Forecasting uses a rolling forecasting method, that is, every 23 hours of data is used to predict the 24th hour of data. All data samples contain three characteristics: load, date type, and temperature of the day.

The average absolute percentage error (MAPE) and mean square error (MSE) are used to judge the prediction effects of various methods.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\% \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

Among them, y_i represents the true value of the test set, \hat{y}_i represents the predicted value of the test set, and n represents the number of samples in the test set.

(1) SVM prediction model

During SVM model training, historical load and date type are used as input features.

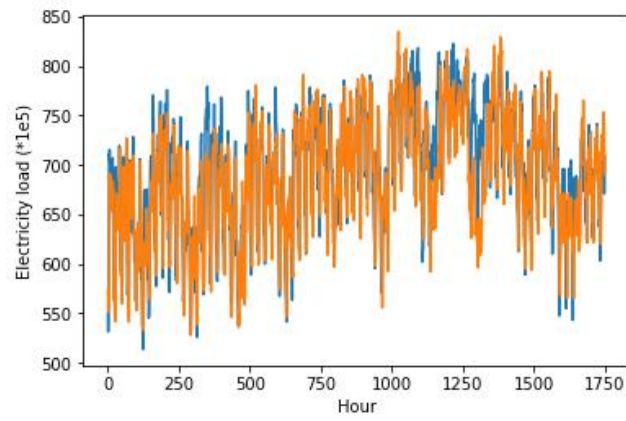


Fig. 4: Prediction results on the SVM model.

(2) LSTM prediction model

When training LSTM models, historical load, date type, and temperature are used as input features.

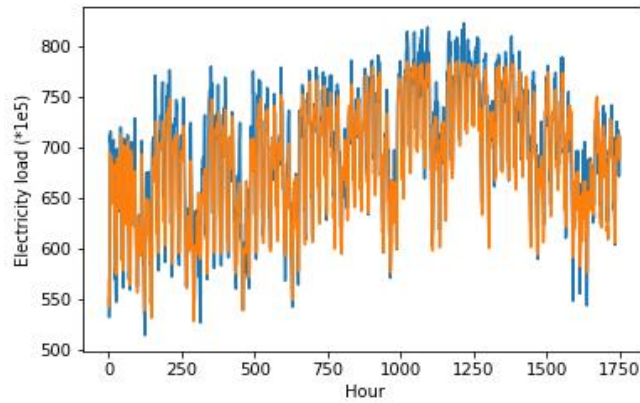


Fig. 5: Prediction results on the LSTM model.

(3) XGBoost prediction model

The XGBoost model uses only historical load as the input feature when training.

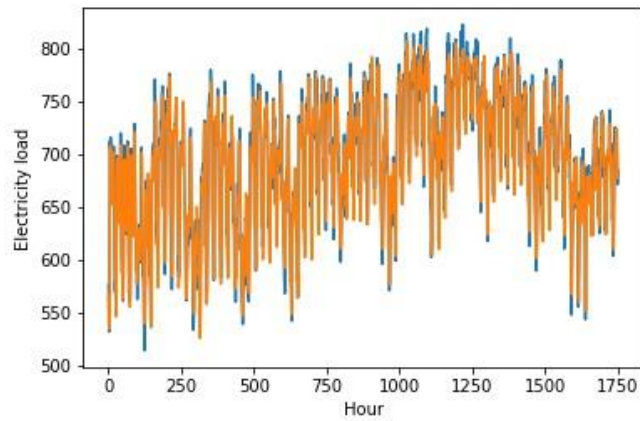


Fig. 6: Prediction results on the XGBoost model.

The prediction errors and training time of the three models are shown in Table 1. As can be seen from Table 1, the XGBoost machine learning model has a short training time and high prediction accuracy.

Table 1: Prediction errors and training time of different models

Model	MAPE(%)	MSE(%)	T(s)
SVM	2.80	40.4	420
LSTM	2.30	25.6	1200
XGBoost	2.01	14.1	180

According to the results shown in Table 1, the XGBoost model was finally applied to the forecast of real power load in a region of Shanxi Province, China. The results are shown in Figure 7.

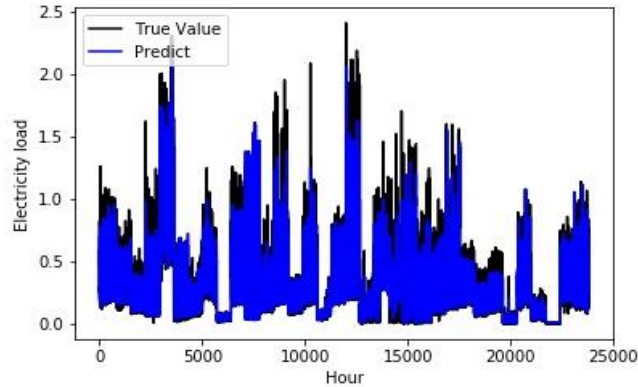


Fig. 7: The result of actual data used XGBoost model.

The average absolute percentage error $MAPE = 1.108$ and the mean square error $MSE = 20.2$. The error is relatively large, because the quality of the real data is poor. There are outliers, that is, noise interference, and the error of some points is too large, so this paper adds wavelet analysis to denoise before prediction.

Wavelet analysis was used to decompose the original load into 4 frequency bands, and the XGBoost model was used to predict the 24th hour load of each frequency band signal, and then the predicted values of the 4 frequency bands were directly added. The final result is shown in Figure 8.

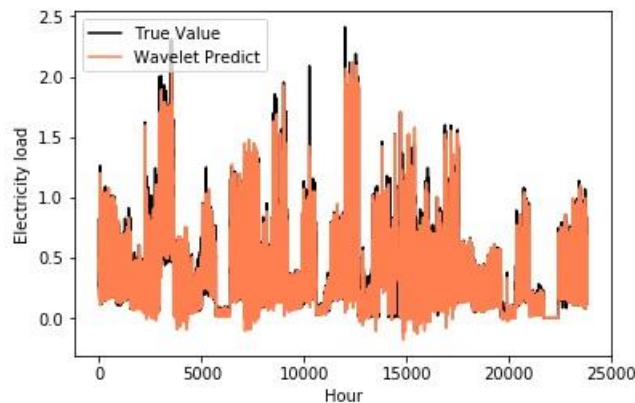


Fig. 8: The result of actual data used XGBoost model after Wavelet analysis

Calculate that $MAPE = 0.561$ and $MSE = 3.7$. The prediction accuracy is greatly improved than when only the XGBoost model is used, which shows that the method in this paper is effective.

6. Acknowledgements

This work is supported by National Science Foundation(NNSF) of China under Grant 61873303.

7. References

- [1] Dongxia Z , Xin M , Liping L , et al. Research on Development Strategy for Smart Grid Big Data. Proceedings of the Csee, 2015.
- [2] Kang Chongqing, Xia Qing, and Liu Mei. Power System Load Forecasting. China Electric Power Press, 2007.
- [3] Hong, Tao , P. Wang , and H. L. Willis. A Naïve multiple linear regression benchmark for short term load forecasting. 2011.
- [4] Dong Xiucheng, Li Qin, and Xu Qiang. The Application of BP Algorithm in Short-term Power Load Forecast. Journal of Instrumentation, 2003, 024(0z2): 379–381.
- [5] Shi Deming, Li Linchuan, and Song Jianwen. Power System Load Forecasting Based on Grey Forecast and Neural Network. Power System Technology, 2001, 25 (12): 14-17.

- [6] Chen B J , Chang M W , and Lin C J . Load forecasting using support vector Machines: a study on EUNITE competition 2001. IEEE Transactions on Power Systems, 2004, 19(4): 1821-1830.
- [7] Kong Xiangyu, Zheng Feng, and E Zhijun, et al. Short-term Load Forecasting Based on Deep Belief Network. Automation of Electric Power Systems, 2018, 042 (005): 133-139 .
- [8] Chen Zhuo and Sun Longxiang. Short-term power load forecasting method based on deep learning LSTM network. Electronic Technology, 2018, 047 (001): 39-41.
- [9] Yang Xiude, Wang Jinmei and Zhang Lina. Application of XGBoost in ultra-short load forecasting. Automation of Electric Drives, 2017, 039 (004): 21-25.
- [10] Ma Xiuhong, Cao Jiping and Dong Shengfei. Wavelet analysis and its application. Computer Technology and Development, 2003, 013 (008): 93-94,100.