Short-term load forecasting based on LSTM-RF-SVM combined model

Short-term load forecasting based on LSTM-RF-SVM combined model

Wei Wei[1*a], Jing Huang[2]

[1]Information Institute, Zhejiang Sci-Tech University, Hangzhou, Zhejiang Province, 310018, China

[2]Information Institute, Zhejiang Sci-Tech University, Hangzhou, Zhejiang Province, 310018, China

[a]201820602014@mails.zstu.edu.cn [*]Corresponding author's e-mail: ww_hunan@163.com

Abstract. Aiming at the problem of low accuracy of single model forecasting power load, a combined LSTM-RF-SVM prediction model based on Long Short-Term Memory network (LSTM), Random Forest (RF) and Support Vector Machine (SVM) is designed. Firstly, the preprocessed load data, meteorological data and date data are input into the three models according to the sliding time window. The prediction results of each model are obtained through training. Then the optimal weighted combination method is uesd to calculate the weight coefficients of the three models in the prediction results. Finally, the final prediction results are obtained by integrating the weight coefficients. Experimental results show that compared with other methods, the established forecasting model improves the accuracy of short-term load forecasting.

1. Introduction

Due to the stable operation of power system, there are strong load balancing constraints, it is important for the reliable operation of power system to realize dynamic balance between power generation and load changes[1]. At present, short-term load forecasting is mainly divided into traditional methods and machine learning based forecasting methods[2]. For example, in the traditional method, literature[3] uses partial least squares regression analysis, literature[4] uses the optimization method of the nearest neighborhood points to establish multivariate time series model, and reference[5] uses the improved Kalman filter algorithm. Due to the simple principle and single mode of the traditional method, when the amount of data is large, the modeling effect is not obvious, and the prediction accuracy is difficult to meet the requirements of modern power. The other one is based on machine learning. For example, the paper[6] uses random forest, and reference[7] uses wavelet mutation Drosophila optimized support vector machine, but the above methods lack the consideration of time series correlation. LSTM is one of the most widely used recurrent neural networks. It has a strong ability to model time series[8], but it has a poor effect when the feature is discontinuous.

This paper designs a short-term load forecast based on the LSTM-RF-SVM model, based on the problem of a single algorithm. First, it processes a large amount of weather information, date information, and load data as features. Then the processed features are input into LSTM model, RF model and SVM model respectively. Finally, the best weighting method determines the weighting factors to get the final forecasted load result. The validity of the model is verified by comparing the predicted load results with the actual load results.

2. Principle of load forecasting

2.1. Long Short-Term Memory

LSTM network is an improved RNN, which uses memory cells to overcome the problem of gradient disappearance in the process of parameter backpropagation[9]. Each neuron of the LSTM network includes a forgetting gate, an input gate and an output gate[10]. The basic unit of the network is shown in Figure 1.

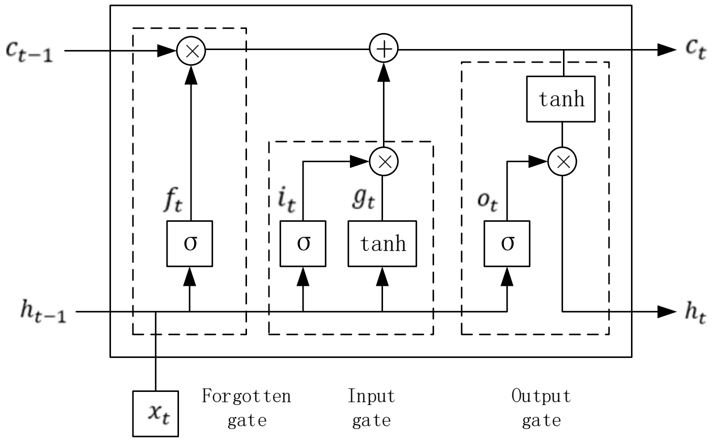

Figure 1. Basic unit of LSTM network.

The following expressions are the mathematical expressions of LSTM cell model:

$$f_t = \sigma\left(W_f[h_{t-1} \ \ x_t]\right) + b_f \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1} \ \ x_t]) + b_i \tag{2}$$

$$g_t = tanh\left(W_g[h_{t-1} \ \ x_t]\right) + b_g \tag{3}$$

$$o_t = \sigma(W_o[h_{t-1} \ \ x_t]) + b_{fo} \tag{4}$$

$$c_t = g_t \odot i_t + c_{t-1} \odot f_t \tag{5}$$

$$h_t = \phi(c_t) \odot o_t \tag{6}$$

Among them, f_t is the output of forgetting gate; i_t is the input gate output; o_t is the output of the output gate; W_f, W_i, W_g, W_o is the weight matrix of the input variable in the corresponding gate; x_t and h_{t-1} is the unit input; h_t is the unit output; c_t is the state of state unit; σ is the sigmoid function change; ϕ is the tanh function change.

2.2. Random Forest

The core idea of supervised learning stochastic forest algorithm is: bootstrap resamples the original samples to obtain multiple sample sets, and establishes the classification decision tree model for the extracted sample sets, and obtains the final prediction result as the mean value of cart decision tree[11]. Let a random forest consist of a series of cart decision trees$i = 1,2, \dots, n$. When constructing the i decision tree, a fixed number of variables are randomly selected from the input variables as the feature space of the decision tree. After the b decision tree is constructed, a complete RF regression model can be constructed[12]. For test sample X, each decision tree is used to predict, and the corresponding b decision tree prediction results are obtained. The prediction result of test sample x is the average value of b decision tree results.

2.3. Support Vector Machine

SVM is a binary classification model that finds a hyperplane to classify samples. The principle of segmentation is to divide the data correctly while ensuring that the classification interval is maximized. In the case that the sample is non-linear and separable, the non-linear problem becomes a linear problem by introducing a non-linear inner product kernel function. Specifically, the linearly inseparable sample data of the original space is mapped to a higher-dimensional space to make them linearly separable, and then linear division is used in the high-dimensional space[13].

3. LSTM-RF-SVM combination model

3.1. Combination forecasting process

The flow chart of prediction using LSTM-RF-SVM combination model is shown in Figure 2. Firstly, the original data is pre-processed, and then the processed data is divided into training set and test set. Next, set the hyperparameters and feature sequences to train each model separately. According to the prediction effect, the hyperparameter optimization model is adjusted to improve the prediction effect. Finally, the final prediction results are obtained by the optimal weighted combination method. The prediction results based on the combination model are compared with the actual sampling load.

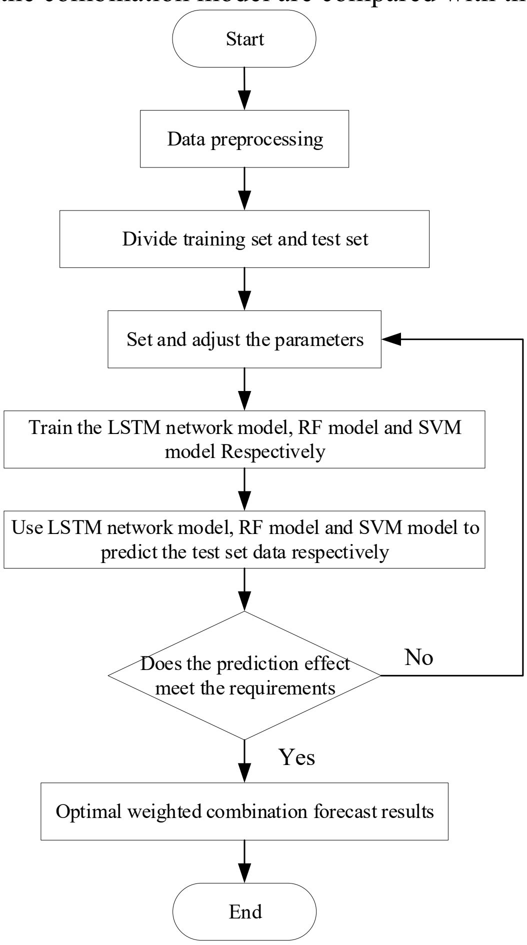

Figure 2. Combination forecast flow chart.

3.2. Combination forecasting process.

The common methods to determine the weight include arithmetic average method, simple weighting method and optimal weighting method, among which the optimal weighting method has the best effect in the combination model[14]. The application steps of the optimal group method in this paper are as follows[15].

Firstly, the deviation matrix e is obtained, as shown in equation (7).

$$E = \begin{bmatrix} \sum_1^n \varepsilon_{1t}^2 & \sum_1^n \varepsilon_{1t}\,\varepsilon_{2t} & \sum_1^n \varepsilon_{1t}\,\varepsilon_{3t} \\ \sum_1^n \varepsilon_{2t}\,\varepsilon_{1t} & \sum_1^n \varepsilon_{2t}^2 & \sum_1^n \varepsilon_{2t}\,\varepsilon_{3t} \\ \sum_1^n \varepsilon_{3t}\,\varepsilon_{1t} & \sum_1^n \varepsilon_{3t}\,\varepsilon_{2t} & \sum_1^n \varepsilon_{3t}^2 \end{bmatrix} \tag{7}$$

Among them, n is the total load sampling; ε_{1t}, ε_{2t} and ε_{3t} is the error between predicted value and true value of KNN model, LSTM network model and SVM model at t time.

Then, the optimal weight K of the combination is obtained by Lagrange method, as shown in equation (8).

$$K = \frac{E^{-1}R}{R^T E^{-1} R} \tag{8}$$

Among them, $K = (w_1, w_2, w_3)^T$, w_1, w_2 and w_3 is the weight coefficient of KNN model, LSTM model and SVM model.

Finally, according to the load prediction results of the three models and the weight coefficients in the combined model, the final prediction results are obtained, as shown in equation (9).

$$y_t = w_1 y_{1t} + w_2 y_{2t} + w_3 y_{3t} \tag{9}$$

Among them, y_{1t}, y_{2t}, and y_{3t} is the load forecasting of KNN model, LSTM network model and KNN model at time t; $y_t$is the prediction result of the combined model at time t.

4. Experimental results and analysis

4.1. Data set.

In this paper, the data of the Ninth "China Electrical Engineering Association Cup" national undergraduate electrical mathematical modeling competition are used as experimental data sets. This data is 1106 days of power load data and corresponding meteorological data from January 1, 2012 to January 10, 2015. Load sampling is carried out every 15 minutes in a day, and a total of 96 sampling points are taken as load data. The highest, lowest and average daily temperature, the relative humidity of the air and the day's rainfall are used as meteorological data. The model is trained from January 1, 2012 to December 11, 2015 as the training set, and the last 30 days (from December 12, 2015 to January 10, 2015) is used as the test set for load forecasting.

4.2. Data processing and input data selection.

This experiment uses interpolation method to fill in the missing values. Then, the influence of different dates on load is found, and the date characteristics are extracted. Combined with historical load data and meteorological characteristics information, the input series is constructed by sliding time window method as the input of each forecast. The extracted features are shown in table 1. The sliding time window is 4. A total of 4 * 12 features are input to predict the load of the next sampling point.

Table 1. Input features.

Features	Description
LOAD	Load at a sampling point.
MAX_TEMP	The highest temperature of the day.
MIN_TEMP	The lowest temperature of the day.
AVE_TEMP	Average temperature of the day.

RH	Relative humidity.
PRCP	Rainfall in a day.
WEEK	1-7: Monday to Sunday.
WEEKEND	1: Monday to Friday; 2: Weekend.
MONTH	1-12: January to December
DAY	Day of month.
SEASON	1-4: Spring, Summer, Autumn, Winter
HOUR	1-96: Sampling point

As shown in Figure 3, the correlation between each sampling point is shown by thermal diagram.

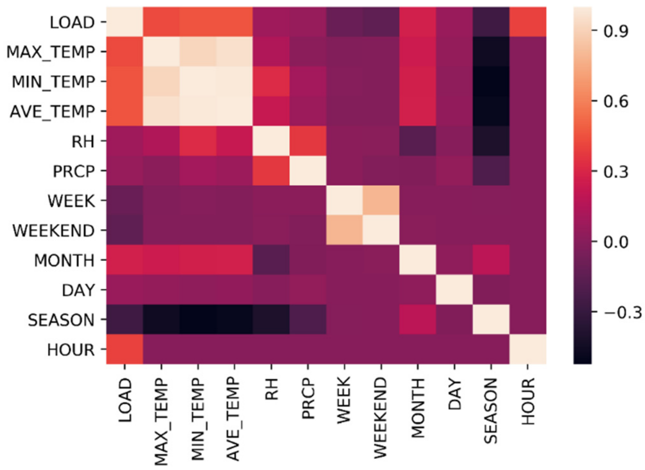

Figure 3. Feature heat map.

The dimension of different features may vary greatly, which affects the accuracy of the experiment. Therefore, the input and output data must be normalized. Each feature sequence is normalized in the range of [0,1]. The formula of normalization is shown in equation (10).

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

Among them, x^* is the normalized value; x_{min} is the minimum value in the feature sequence; x_{max} is the maximum value in the feature sequence.

4.3. Experimental evaluation criteria.

This paper selects RMSE and MAPE as indicators for evaluating the combination model and other methods mentioned in the article. The calculation formulas are shown in equations (11) and (12).

$$E_{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2} \tag{11}$$

$$E_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i^*|}{y_i} \times 100\% \tag{12}$$

Among them, n is the number of samples; $y_i$and y_i^* is the actual load value and forecast load value of the i sampling point on the forecast day.

4.4. Analysis of experimental results.

The setting of super parameters greatly affects the accuracy of model prediction. The main parameters involved in this model are: LSTM network model has three layers, the number of neurons is 64, 32 and 16, the loss function is MSE, the activation function is relu, and the optimizer is Adam; the RF model uses 50 decision trees, and the maximum depth of the tree is 20.

In this paper, the short-term load forecasting of each sampling point is carried out continuously from December 12, 2015 to January 10, 2015. The RMSE and MSPE of the proposed model are compared with those of KNN, GUR and BP neural networks. The results show that the accuracy of the combined model is higher than other methods in random day and 30 day continuous load forecasting, as shown in table 2.

Table 2. Comparison of prediction results

model	one day		30 days	
	RMSE(kW)	MAPE(%)	RMSE(kW)	MAPE(%)
KNN	456.888	5.739	610.913	7.093
GUR	291.691	2.601	270.622	2.584
BP	346.467	2.869	329.823	2.860
LSTM-RF-SVM	245.552	2.077	240.248	2.201

Figure 4 shows the line chart of predicted value and actual value of 96 sampling points of different models in a certain day. It can be seen from the figure that the forecast of the four models on this day is basically in line with the change trend of the actual load. The load curve predicted by the combination model mentioned in this paper is more consistent with the actual load change curve, and the load forecasting curve obtained by the combination model has the highest accuracy.

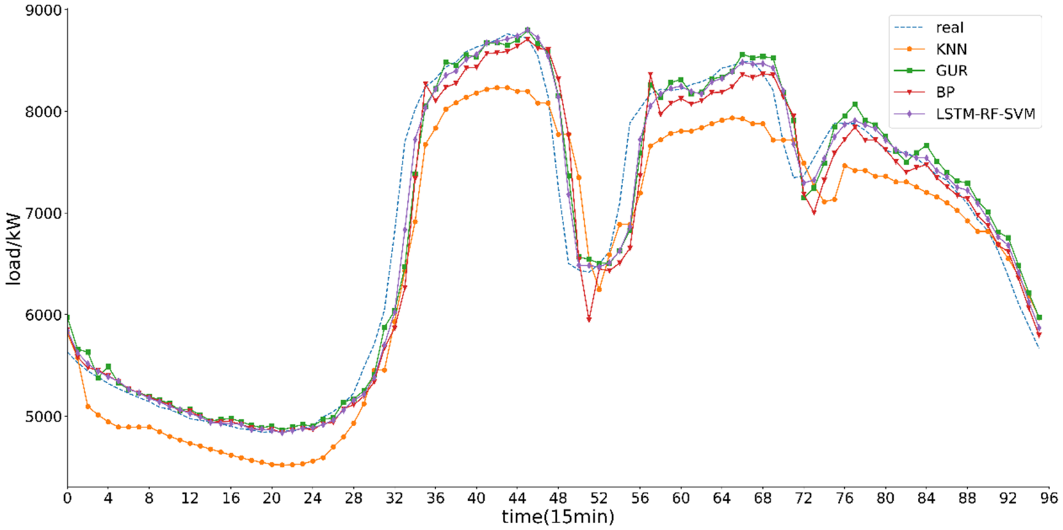

Figure 4. Feature heat map.

5. Summary

This paper proposes a short-term load forecasting method based on LSTM-RF-SVM for the accuracy of load forecasting in power systems. This method combines the characteristics of the three models, not only takes into account the time sequence between the load data, but also can mine the hidden information of the load data. Compared with a single model, the combined model of text can predict the load change of the next sampling point and improve the accuracy of short-term load forecasting. In future work, the date information and weather information related to the load data can be combined to classify the load data and then forecast.

References

[1] Zhu Juncheng, Yang Zhile, Guo Yuanjun, Yu kunjie, Zhang Jiankang, Mu Xiaomin. Application of deep learning in power load forecasting [J]. Journal of Zhengzhou University (Engineering Edition), 2019,40 (05): 13-22.

[2] Wang zengping, Zhao Bing, Ji Weijia, Gao Xin, Li Xiaobing. Short term load forecasting method based on gru-nn model [J]. Power system automation, 2019,43 (05): 53-62.
[3] Zhang Fusheng, Wang Hong, Han Ti, sun Xiaoqiang, Zhang Zhenyu, Cao Jin. Short term load forecasting based on partial least squares regression analysis [J]. Power grid technology, 2003 (03): 36-40.
[4] Lei Shaolan, sun Caixin, Zhou Fu, Zhang Xiaoxing. Study on multivariable time series linear regression prediction method for short-term power load [J]. Acta electrical engineering Sinica, 2006 (02): 25-29.
[5] Chen Jun, Tao Wei, LV Yingfei, he Jianping. Short term load forecasting based on Kalman filter [J]. Electrical switch, 2014,52 (02): 43-45.
[6] Li Wanhua, Chen Hong, Guo Kun, Guo Songrong, Han Jiamin, Chen Yuzhong. Study on power load forecasting based on Stochastic Forest algorithm [J]. Computer engineering and application, 2016,52 (23): 236-243.
[7] Liang Ni Xiao. Short term load forecasting based on improved fruit fly algorithm optimized support vector machine [D]. Guangxi University, 2014. Zhang Fusheng, Wang Hong, Han Ti, sun Xiaoqiang, Zhang Zhenyu, Cao Jin. Short term load forecasting based on partial least squares regression analysis [J]. Power grid technology, 2003 (03): 36-40.
[8] Zhang Yuhang, Qiu Caiming, he Xing, Ling Zenan, Shi Xin. A short-term load forecasting method based on LSTM neural network [J]. Power information and communication technology, 2017,15 (09): 19-25.
[9] Guo Fu Ao, Liu Daming, Zhang Zhenzhong, Tang Fei. Gpso-lstm short-term load forecasting based on characteristic correlation analysis [J/OL]. Electrical measurement and table: 1-10 [2020-08-14].http://kns.cnki.net/kcms/detail/23.1202.TH.20200730.1417.012.html.
[10] Zhang Fusheng, Wang Hong, Han Ti, sun Xiaoqiang, Zhang Zhenyu, Cao Jin. Short term load forecasting based on partial least squares regression analysis [J]. Power grid technology, 2003 (03): 36-40.
[11] Huang Qingping, Zou Xiaoming, Liu chuqun, ye Mingwu, Huang Qijun. Short term load forecasting based on wavelet decomposition and random forest [J]. Power information and communication technology, 2019,17 (09): 24-29.
[12] Liu Qiang, Hu Zhiqiang, Zhou Yu, LV pengpeng, Wang Jinxin. Short term wind power prediction based on ceemd and stochastic forest algorithm [J]. Smart power, 2019, 47 (06): 71-76 + 94.
[13] Lu Ning, Wu benling, Liu Ying. Application of SVM model based on Adaptive Particle Swarm Optimization in load forecasting [J]. Power system protection and control, 2011,39 (15): 43-46 + 51.
[14] Quan Yi, Zhang Yongchuan. Weight algorithm and its application in combination forecasting method [J]. Science and technology entrepreneurship monthly, 2006 (05): 171-172.
[15] Chen Weinan, Hu Zhijian, Yue Jingpeng, Du Yixing, Qi Qi. Short term load forecasting based on the combination model of long-term memory network and lightgbm [J/OL]. Power system automation: 1-9 [2020-08-15]. http://kns.cnki.net/kcms/detail/32.1180.TP.20200610.1819.012.html.