

Day ahead load forecasting of integrated energy system based on multi-model combination

Jianhua Ye^{a,*}, Fengzhang Luo^b, Jing Cao^c, Li Yang^d

^aTianjin Key Laboratory of Information Sensing & Intelligent Control, Tianjin University of Technology and Education, Tianjin 300222, China; ^bKey Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China; ^cState Grid Tianjin Electric Power Company, Tianjin 300010, China; ^dState Grid Chongqing Yongchuan Power Supply Company, Chongqing 402160, China

ABSTRACT

The load of user level integrated energy system changes rapidly and it is difficult to predict accurately. Therefore, a day ahead load forecasting method of integrated energy system based on multi-model combination was proposed. Firstly, the long short-term memory (LSTM) network model, convolutional neural network (CNN) model and harmony search (HS) optimized light gradient boosting machine (LightGBM) model were established. Then, the inverse root mean square error method (IRMSE) was used to combine the forecasting results of the three models to obtain the final forecasting value. The effectiveness of the proposed method was verified by the actual data of an integrated energy system. The results show that the proposed method is superior to the single prediction model and the simple average combination model, and has the best prediction accuracy for electric, cooling and heat loads.

Keywords: Load forecasting, integrated energy system, LSTM, CNN, LightGBM.

1. INTRODUCTION

Integrated energy system (IES) can diversify energy supply, effectively improve the comprehensive utilization efficiency of all kinds of energy, and play a positive role in environmental protection. Load forecasting is the premise to ensure the reliable and economic operation of energy system. There is a certain coupling relationship between multi-energy loads in IES, and the user level loads of IES change frequently. Consequently higher requirements are put forward for IES multi-energy load forecasting (MELF).

With the acceleration of energy transformation process, the research on MELF is gradually increasing. At present, the research of MELF mainly belongs to deterministic forecasting, which can be divided into two categories according to the number of forecasting models used. The first category is single model prediction. This kind of method usually uses Copula theory¹, Pearson coefficient², grey correlation analysis³ and other methods for correlation analysis, extracts the influencing factors with great correlation with MELF. These factors construct the original feature set of forecasting models. On this basis, convolutional neural network (CNN)², encoder-decoder model based on long short-term memory (LSTM) network⁴, gated recurrent unit (GRU)⁵ are utilized to further extract features. Traditional machine learning algorithms widely used in power load forecasting are also used in MELF, such as generalized regression neural network², support vector regression (SVR)², gradient boosting decision tree (GBDT)⁴, extreme learning machine (ELM)⁵, least square support vector machine (LSSVM)⁶. In recent years, deep learning has been more and more applied in MELF with its good learning performance and generalization ability, including deep belief network (DBN)⁷, LSTM network⁸. The second category is the combined forecasting method. These methods mostly use decomposition algorithms, such as wavelet packet⁹ and quadratic mode decomposition¹⁰, to decompose multi-energy loads of IES into different components. Then different methods, such as recurrent neural network (RNN)⁹, deep bidirectional LSTM (DBiLSTM) network and multiple linear regression (MLR)¹⁰, are adopted to predict components respectively.

*yjh7985@126.com

To sum up, considerable achievements have been made in the research of MELF, in which the prediction accuracy of combined forecasting model is generally higher than that of single forecasting model. Light gradient boosting machine (LightGBM) has good performance in power load forecasting¹¹. Therefore, this paper constructs a weighted combination forecasting method with LSTM network, CNN model and LightGBM model optimized by harmony search (HS)¹². Different features composed of load data, meteorological information and calendar information are input into these models. The weight coefficients are determined by inverse root mean square error (IRMSE), and the weighted combination of the predicted loads of the three models is the final MELF values. The experimental results show that the presented model is better than the single models and surpasses the simple average combination model.

2. MELF BASED ON LSTM NETWORK

2.1. LSTM network

LSTM network is an improved RNN¹³, which can learn the long-term dependence information of sequences and overcome the gradient disappearance and explosion problems of traditional RNN. The structure of LSTM cell is shown in Figure 1. t is current time. f_t , i_t , o_t , g_t and c_t represent forgetting gate, input gate, output gate, the candidate value of cell state and cell state at time t respectively. x_t is current input. h_{t-1} represents the output at time $t-1$. σ is Sigmoid function. \otimes represents the product of the corresponding elements of two vectors.

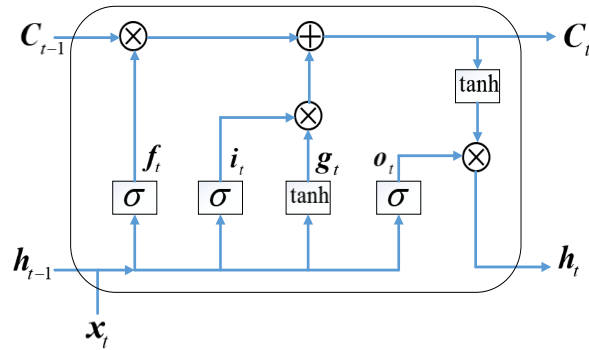


Figure 1. The structure of LSTM cell.

2.2. Construction of forecasting model based on LSTM

For LSTM forecasting model, its input and output data should be set first. Assuming that the IES load (electric, cooling and heat load) at time t on day d is predicted, the input data are composed of the 16-dimensional features of time t on the previous 7 days ($d-7, \dots, d-1$). The features include IES load, meteorological information and calendar information, as shown in Table 1. DNI, DHI and GHI are abbreviation for direct normal irradiance, diffuse horizontal irradiance and global horizontal irradiance. Therefore, the input sequence length of LSTM network is 7 and the feature dimension of each time is 16. For the working day type in Table 1, the working day is 1, otherwise it is 0. For the holiday type, the holiday is taken as 1, otherwise it is taken as 0.

The feature values of different dimensions vary dramatically. Therefore, the normalization operation is carried out as follows:

$$x_i^{\text{norm}} = \frac{x_i - x_i^{\min}}{x_i^{\max} - x_i^{\min}} \quad (1)$$

where x_i and x_i^{norm} represent the values before and after normalization. x_i^{\max} and x_i^{\min} are the maximum and minimum values of the i th feature respectively.

LSTM network structure can be described as follows: the normalized input sequence data are sent to the input layer, then cascade multiple hidden layers, and finally the electric, cooling and heat load are predicted through the fully connected layer.

Table 1. Feature set of LSTM.

Feature number	Feature name
1-3	Electric/cooling/heat load
4-6	DNI/DHI/GHI
7-10	Dew point/wind speed/humidity/temperature
11	Time t (1-24)
12	Working day type
13	Holiday type
14	Day (1-31)
15	Day in a week (1-7)
16	Month (1-12)

3. MELF BASED ON CNN MODEL

3.1. CNN model

CNN is a special feedforward neural network¹⁴. Its structure is similar to that of multi-layer perceptron (MLP). It has the advantages of equivalent representation, sparse interaction and parameter sharing. CNN usually consists of three parts: convolution layer, pooling layer and full connection layer. Among them, the convolution layer generally convolves with input data with multiple kernels, and then adopts nonlinear activation function to extract features. The dimension of the pooling layer is reduced by down sampling. Then the features are flattened into one-dimensional vector and sent to the fully connected layer for prediction.

3.2. Construction of forecasting model based on CNN

The input structure of CNN model is slightly different from that of LSTM network. The 16-dimensional features of LSTM network input data are arranged into 4×4 matrix, taking the input sequence length as the number of channels. The output is the same as that of LSTM network.

4. MELF BASED ON LIGHTGBM

4.1. LightGBM model

LightGBM is an efficient GBDT model based on histogram¹¹. It adopts gradient based single-sided sampling and leaf growth strategy with depth limit, makes it have better training speed, less memory consumption and higher accuracy. LightGBM constructs one regression tree at a time by fitting the residual of the previous regression tree. LightGBM combines weak learners into a single strong learner in the iterative process by minimizing the loss function, as shown in equation (2).

$$f(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x}) \quad (2)$$

where \mathbf{x} is the input feature vector, $f(\mathbf{x})$ denotes the final predicted value and $f_i(\mathbf{x})$ accounts for the output of the i th regression tree.

4.2. Construction of forecasting model based on LightGBM

4.2.1. Determination of input and output data. The input and output of LightGBM model are different from LSTM network. The IES load (electric, cooling or heat load) at time t on day d is predicted, and the input data is 25-dimensional features. The input features include the electric, cooling and heat load at time t and $t-1$ on day $d-1$ and $d-7$, weather forecast information and calendar information at time t on day d as shown in Table 1. The output is the load at time t on

day d , and the dimension is 1. That is, the prediction model based on LightGBM needs to be established for electric, cooling and heat loads respectively.

4.2.2. *Hyper Parameter Optimization Based on Harmony Search.* When building LightGBM model, it is necessary to determine its hyper parameters, among which the most important parameters are the number of regression trees N_{tree} and the maximum number of leaves per tree N_{leaf} . Therefore, harmony search (HS)¹² algorithm is used to optimize these two hyper parameters. HS algorithm is inspired by the improvisation process of musicians. In the process of music creation, musicians will repeatedly adjust the tones of various instruments to find the best harmony (hyper parameters).

5. COMBINED FORECASTING MODEL

After the predicted values of electric, cooling and heat loads are obtained by LSTM network, CNN model and LightGBM model respectively. IRMSE method¹⁵ is adopted to determine the weights of the three models. The calculation process is as follows:

$$f_{cb} = \sum_{i=1}^{N_{mod}} w_i f_i \quad (3)$$

$$w_i = \frac{1/I_{RMSE}^i}{\sum_{j=1}^{N_{mod}} 1/I_{RMSE}^j} \quad (4)$$

$$I_{RMSE}^i = \sqrt{\frac{1}{N_v} \sum_{k=1}^{N_v} (y_k - \hat{y}_k)^2} \quad (5)$$

where f_{cb} is the final predicted value of electric, cooling or heat load. f_i is the predicted value of the i th model. I_{RMSE}^i represents the RMSE of the i th model. w_i represents the weight of the i th model and N_{mod} is the number of models, which is 3 in this paper. y_k and \hat{y}_k are the actual and predicted electric, cooling or heat load of the k th sample point. N_v is the number of samples in the validation set. It can be seen from equation (4) that the model with smaller error has larger weight coefficient, which can reduce the error of the combined model and improve the prediction accuracy.

The framework of the proposed method is shown in Figure 2.

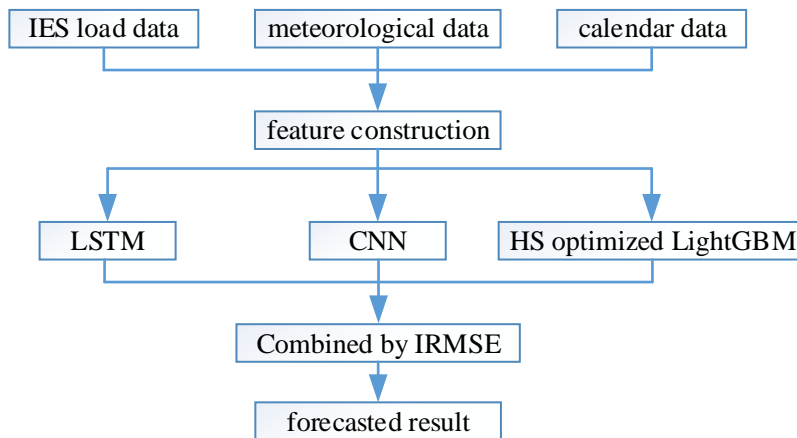


Figure 2. The framework of the proposed method.

6. CASE STUDY

6.1. Experimental data and platform

The user level IES multi-energy load data is from Tempe campus of Arizona State University¹⁶. Meteorological data are from the National Renewable Energy Laboratory of the United States¹⁷.

The data from May 25, 2017 to August 31, 2019 are selected as the experimental data, and the sampling interval is 1 hour. The data from August 25 to August 31, 2019 constitute the test set, and the other data are randomly divided into training set and validation set according to 4:1.

The experimental hardware is configured with Intel Core i5-4200 CPU and 8G memory. Python language is used for programming, and each model is implemented by calling PyTorch and scikit-learn libraries respectively.

6.2. Evaluation indices

In order to comprehensively evaluate the performance of MELF methods, mean absolute percentage error (MAPE) and RMSE indicators are selected to evaluate the effect of single load and integrated load forecasting. MAPE is defined as:

$$I_{MAPE}^j = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (6)$$

$$W_{MAPE} = \sum_{j=1}^3 \alpha_j I_{MAPE}^j \quad (7)$$

where m is the number of samples in the test set. I_{MAPE}^j denote MAPE for class j load. W_{MAPE} is the integrated MAPE of IES loads, which represents the overall performance of MELF. α_j is the importance ratio of class j load (electric, cooling or heat load).

6.3. Model parameter setting

The parameter setting has a great impact on MELF performance of each model. In this paper, the grid search method is used to obtain the optimal hyper parameters of LSTM network and CNN model.

The LSTM network is provided with a hidden layer with 16 neurons. Two convolution layers are set in CNN model, the number of convolution kernel is 8 and 16 respectively, and the size of convolution kernel is 2×2 . There is no pooling layer, two fully connected layers are set, the number of neurons is 10 and 3 respectively, and the activation function is ReLU. Both LSTM network and CNN model are optimized by Adam algorithm, the learning rate is 0.01, the training batch size is set to 24, and epochs are 100.

In order to evaluate the performance of the combined method (denoted as IRMSE-LCL) in MELF, it is compared with SVR and MLP. The hyper parameters of LightGBM, SVR and MLP are optimized by HS algorithm. The relevant parameter settings of HS are shown in Table 2.

Table 2. Parameter setting of HS.

Parameter name	Parameter value
HMS	30
HMCR	0.8
PAR	0.3
T_{\max}	100

The inputs of SVR and MLP are the same as that of LightGBM model. Similar to the LightGBM model, the output of the SVR model is one-dimensional, so it is necessary to train the SVR model for electric, cooling and heat loads respectively. The output of MLP model is 3-dimensional, so only one MLP model needs to be trained to predict the electric, cooling and heat loads at the same time. The SVR model adopts Gaussian kernel function, and the hyper

parameters optimized by HS include insensitive loss parameter, penalty coefficient and kernel function parameter. For the MLP model, the hyper parameters optimized by HS include the number of hidden layer neurons and L₂ norm penalty parameter. The activation function also selects ReLU. The optimization algorithm is Adam, with adaptive learning rate and 500 iterations.

6.4. Result analysis

IRMSE-LCL model is compared with single LSTM network, CNN, LightGBM, SVR and MLP, and with the simple average combination model of LSTM, CNN and LightGBM (represented by Avg-LCL). The MAPE, RMSE and integrated MAPE of day ahead MELF on the test set are shown in Tables 3 and 4. When calculating the integrated MAPE, the importance ratios of electric, cooling and heat load are set to 0.4, 0.4 and 0.2 respectively³.

From Tables 3 and 4, we can conclude that:

- (1) The forecasting accuracy of any single model is no better than other models in all types of load.
- (2) The single models have different prediction accuracy for different types of load. MLP and LightGBM have better prediction accuracy on electric load. LSTM network and CNN have the best prediction performance on cooling load. LSTM network, CNN and LightGBM have less prediction error on heat load.

Table 3. MAPE comparison of different models.

Model	MAPE/% (electric/cooling/heat)	Integrated MAPE/%
LSTM	5.25/5.90/3.54	5.17
CNN	5.61/5.79/3.87	5.33
LightGBM	4.28/7.30/3.87	5.41
SVR	4.41/7.84/5.66	6.03
MLP	4.24/8.38/4.25	5.90
Avg-LCL	4.49/5.77/2.66	4.64
IRMSE-LCL	4.08/5.60/2.55	4.38

Table 4. RMSE comparison of different models.

Model	Electric load (kW)	Cooling load (tons)	Heat load (mmBTU)
LSTM	1836.85	1289.82	0.23
CNN	1921.97	1251.45	0.27
LightGBM	1581.61	1555.98	0.25
SVR	1628.05	1540.96	0.37
MLP	1631.00	1716.54	0.28
Avg-LCL	1584.66	1244.11	0.18
IRMSE-LCL	1495.18	1180.99	0.17

(3) LSTM network, CNN and LightGBM have smaller integrated MAPE, that is, their overall prediction performance is the best.

(4) The two combination models outperform all the single models. The IRMSE-LCL method not only has the best overall prediction performance, but also has the highest prediction accuracy for all types of load.

The proposed method combines LSTM network, CNN and LightGBM, which can effectively combine the respective advantages of each model and learn from each other. It can not only enhance the model's perception of timing

information, but also fully mine the characteristic information of discontinuous data. Giving more weight to the model with higher prediction accuracy can effectively improve the overall prediction performance.

7. CONCLUSION

In this paper, a combined MELF method based on LSTM network, CNN and model is proposed. The inverse root mean square error method is used to weighted combine the prediction results of the three models to obtain the final prediction values. Compared with the single models and the simple average combination model, the proposed model has better prediction accuracy. The next step is to study the load decomposition algorithm to further improve the prediction performance of IES loads.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (51977140).

REFERENCES

- [1] Ma, J., Gong, W. and Zhang Z., *Adv. Technol. Electr. Eng. Energy*, 39, 24-31 (2020).
- [2] Luo, F, Zhang, X., Yang, X., Yao, L., Zhu, L. and Qian, M, *High Volt. Eng.*, 47, 23-32 (2021).
- [3] Tian, H., Zhang, Z. and Yu, D., *Proc. CSU EPSA*, 33, 130-7 (2021).
- [4] Wang, S., Wang, S., Chen, H. and Gu, Q., *Energy*, 195, 116964 (2020).
- [5] Sun, X., Li, J., Zeng, B., Gong, D. and Lian, Z., *Control Theory A.*, 38, 63-72 (2021).
- [6] Tan, Z., De, G., Li, M., Lin, H., Yang, S., Huang, L. and Tan, Q., *J. Clean. Prod.*, 248, 119252 (2020).
- [7] Shi, J., Tan, T., Guo, J., Liu, Y. and Zhang, J., *Power Syst. Technol.*, 42, 698-707 (2018).
- [8] Sun, Q., Wang, X., Zhang, Y., Zhang, F., Zhang, P. and Gao, W., *Autom. Electr. Power Syst.*, 45, 63-70 (2021).
- [9] Zhu, L., Wang, X., Ma, J., Chen, Q. and Qi, X., *Electr. Power Constr.*, 41, 131-8 (2020).
- [10] Chen, J., Hu, Z., Chen, W., Gao, M., Du, Y. and Lin, M., *Autom. Electr. Power Syst.*, 45, 85-94 (2021).
- [11] Chen, W., Hu, Z., Yue, J., Du, Y. and Qi, Q., *Autom. Electr. Power Syst.*, 45, 91-7 (2021).
- [12] Geem, Z. W., Kim, J. H. and Loganathan, G. V., *Simulation*, 2, 60-8 (2001).
- [13] Hochreiter, S. and Schmidhuber, J., *Neural Comput.*, 9, 1735-80 (1997).
- [14] Khan, A., Sohail, A., Zahoor, U. and Qureshi, A. S., *Artif. Intell. Rev.*, 53, 5455-516 (2020).
- [15] Moreira, J., Soares, C., Jorge, A. M. and Sousa, J., *ACM Comput. Surv.*, 45, 1001-40 (2012).
- [16] Campus metabolism, <http://cm.asu.edu/Online>, accessed 4-July-2021.
- [17] NSRDB data viewer, <https://maps.nrel.gov/nsrdb-viewer/Online>, accessed 14-July-2021.