Multivariate load prediction method for integrated energy system based on CEEMD-LSTM

Wanjia Zhang1,* and Wei Huang1
1School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China

*Corresponding author e-mail: zhangwanjia@ncepu.edu.cn

Abstract. Accurate load prediction is increasingly important for the economic dispatch of the integrated energy system. A multivariate load prediction method based on Complete Ensemble Empirical Mode Decomposition and Long Short Term Memory network (CEEMD-LSTM) is raised. Pearson correlation coefficient was used to analyze the influencing factors, and the influencing factors with high correlation degree were screened out. CEEMD was used to decompose cold, heat and electric load sequences into the mean components of Intrinsic Mode Function (IMF), and the ones with highly related to load forecasting are screened and retained. The selected IMF mean components and influencing factors were input into the LSTM model for training and learning to obtain the final CEEMD-LSTM load forecasting model. The simulation results are verified by an integrated energy system in a province. Compared with other load forecasting methods, the proposed load forecasting method has higher prediction accuracy, and considers the difference between cold, heat, electric load and the correlation of influencing factors.

1. Introduction

The access of large-scale renewable energy system enhances the coupling between various energy systems, which has requirements on the accuracy and reliability of load forecasting. Therefore, the prediction of cool, heat and electric load has great engineering influence on the research of the integrated energy system [1].

In this paper, a multivariate load prediction method based on CEEMD-LSTM model is raised. After CEEMD decomposition of load data, the prediction regression model of long and short term memory network was constructed together with the influencing factors to predict the load. Using the data of a province's integrated energy system for training and verification, and comparing with other load forecasting algorithms, the feasibility of the algorithm is verified.

2. The principles of the CEEMD-LSTM model

2.1. The principle of CEEMD

CEEMD is to add n groups of white noise with opposite symbols on the basis of Ensemble Empirical Mode Decomposition (EEMD). The two complementary signals are then decomposed by EEMD respectively to generate the components of the IMF. The 2n groups IMF mean values were used as the final results [2]. The calculation formula is shown below:
\[ C_i^+ (t) = x(t) + n_i (t) \]  \hspace{1cm} (1)
\[ C_i^- (t) = x(t) - n_i (t) \]  \hspace{1cm} (2)
\[ IMF_j = \frac{1}{2n} \sum_{i=1}^{n} IMF_{ij} \]  \hspace{1cm} (3)

In the formula, \( C_i^+ (t) \) and \( C_i^- (t) \) are the data to be decomposed after adding positive and negative white noise, \( x(t) \) is the original signal, \( n_i (t) \) is the added white noise. \( IMF_{ij} \) is the jth IMF component of the ith signal, \( IMF_j \) is the jth mean IMF component of the original signal.

2.2. The principle of LSTM

LSTM can not only learn the data rules in the long time scale, but also reflect the data characteristics in the short time scale, and realize the combination of long-term and short-term memory [3, 4]. The LSTM’s structure is shown below in figure 1.

![Figure 1. LSTM network structure.](image)

3. The model of CEEMD-LSTM

3.1. Data preprocessing

In this paper, standard data with zero mean value and unit variance are obtained by using standard fraction normalization method [5]. The standardized formula is as follows:

\[ x_m = \frac{x - \mu}{\sigma} \]  \hspace{1cm} (4)

In the formula, \( x \) is the sample value to be normalized, \( \mu \) and \( \sigma \) are the average value and standard deviation of the sample data, \( x_m \) is the normalized sample data.

3.2. Division of training set and test set

The initiate 85% of the sample data is used to be the training data set, and the last 15% is used to be the test data set.

3.3. Pearson correlation coefficient

Pearson correlation coefficient is a indicator to measure the variable’s correlation. The calculation formula is as follows:

\[ \rho_{XY} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2 \cdot \sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \]  \hspace{1cm} (5)

In the formula, \( \overline{X} \) and \( \overline{Y} \) are the mean value of time series \( X_i \) and \( Y_i \); \( N \) is the data number.
3.4. **CEEMD-LSTM model steps**

In this paper, after taking into account the influence of historical load data, the weather conditions are added to be the influencing factors. First, the sample data is preprocessed by the standard fraction normalization method, and the training set and test set are divided. Secondly, the paper uses Pearson correlation coefficient to calculate the correlation between load and weather influencing factors. After that, CEEMD was performed on the load sequences in all samples, and the IMF mean components of each group were calculated, analyzed and screened by Pearson correlation coefficient. N groups of IMF mean components of cool, heat and electric load were obtained respectively. Then LSTM neural network training was carried out. The CEEMD-LSTM load prediction model was obtained by inputting IMF mean component data of 3N groups of historical cold, heat and electric load decomposed by CEEMD as well as weather influencing factors. The final result is obtained by adding the predicted N groups of IMF mean components and performing inverse transformation. The CEEMD-LSTM model is shown in figure 2.

![CEEMD-LSTM model](image)

**Figure 2.** CEEMD-LSTM model.

4. **Case study**

The experimental data came from an industrial park in a province. The sampling interval of cool, heat and electric load was 15 min. From September 25 to November 12, 4704 groups of data were used as training data, and 576 groups of data from November 13 to November 18 were used as test data. The plot matrix diagram of total sample load, temperature and humidity is shown in figure 3, and the plot matrix diagram of sample load is shown in figure 4. In this paper, plot matrix is used to show the correlation between variables, which provides a reference for Pearson correlation calculation below.
4.1. Selection of influencing factors

Based on 5280 groups load data of an integrated energy system from September to November and weather data. The results are shown below in Table 1.

|            | electric load | heat load | cold load |
|------------|---------------|-----------|-----------|
| electric load | 1             | 0.26      | 0.36      |
| heat load   | 0.26          | 1         | 0.32      |
| cold load   | 0.36          | 0.32      | 1         |
| temperature | 0.71          | 0.63      | 0.64      |
| humidity    | 0.23          | 0.28      | 0.25      |
| wind speed  | 0.04          | 0.03      | 0.08      |

According to Table 1, the absolute value of correlation coefficient is greater than 0.1 and less than 1, showing a correlation, while the absolute value of correlation coefficient between them and wind speed is less than 0.1, showing no correlation. Therefore, this paper only takes the historical data of cool, heat and electric load, temperature and humidity as the factors to be considered.

4.2. Model parameter

4.2.1. CEEMD parameter selection. During CEEMD decomposition, 200 groups of positive and negative white noise sequences were added to the original load sequence, and the standard deviation of each group of Gaussian white noise sequences was 0.2.

4.2.2. LSTM parameter selection. In this paper, through continuous simulation and calculation, a LSTM neural network is created. The network has 50 hidden unit layers, 50 fully connected layers and
a discarded layer with probability of 0.5. The solver "Adam" was used to carry out 200 rounds of training in small batch with size of 20. In addition, to prevent the gradient explosion, the gradient threshold is set to 1.

4.3. The result of CEEMD

According to formula (6), the number of components of cool, heat and electric load data in this paper is 11.

$$m = \text{fix}(\log_2(\text{xsize})) - 1$$  \hspace{1cm} (6)

In the formula, \(m\) is the number of IMF components. \(\text{fix}\) is rounding function. \(\text{xsize}\) is the sampling points.

In this paper, Pearson correlation calculation is made between 11 groups IMF mean components and the load value at the next moment, IMF mean components \(\text{IMF}_i \sim \text{IMF}_j\) with correlation coefficient greater than 0.2 are selected, as shown in figure 5, which effectively reduces mode mixing.

4.4. CEEMD-LSTM load prediction result

The forecast results of cool, heat and electric load on November 18 are shown in figure 6.

To ensure the justification, the same test set is used to evaluate the load forecasting errors of different algorithms [6, 7]. The results are shown below in figure 7.
Figure 7. Algorithm error comparison radar chart.

By comparing the prediction results of different algorithms, it can be seen that ARIMA model, LSTM and the improved LSTM all have the ability to approximate the load curve. The accuracy of ARIMA model and BP is poor, followed by LSTM, EMD-LSTM and EEMD-LSTM prediction error is significantly reduced, and the calculation accuracy of CEEMD-LSTM model is the highest.

5. Conclusion

In this paper, the coupling between the cool, heat and electric load and the nonlinear randomness of each load is considered. The historical load, temperature and humidity and other weather factors are comprehensively considered to forecast the cool, heat and electric load.

1) The historical load data, temperature and humidity are selected as inputs by Pearson correlation coefficient. The CEEMD is used to screen the IMF mean component. The coupling between cold, heat, electric load and weather factors is considered, and the prediction accuracy is improved. The weather factors and IMF mean components decomposed by CEEMD are selected appropriately to reduce workload and improve the operation speed.

2) Compared with other traditional models, the raised CEEMD-LSTM model can better fit the load value, and the prediction accuracy has been significantly improved. It has a strong adaptability in solving the prediction of cool, heat and electric load in the integrated energy system.

References

[1] LIANG Rong, WANG Hongtao, WU Kuihua, SUN Wei, FU Chunmei, ZHANG Xiaolei. Short-term Forecasting of Cooling, Heating and Electrical Load based on Neural Network and ARIMA Model[J/OL]. Proceedings of the CSU-EPSA: 1-8.
[2] ZHAO Min. Microgrid Short-Term Load Forecasting Model Based on CEEMD and VHBFO_SVM[J]. Colliery Mechanical & Electrical Technology, 2019, 40(05): 38-43.
[3] LU Jixiang, ZHANG Qipei, YANG Zhihong, TU Mengfu, LU Jinjun, PENG Hui. Short-term Load Forecasting Method Based on CNN-LSTM Hybrid Neural Network Model[J]. Automation of Electric Power Systems, 2019, 43(08): 131-137.
[4] Kody M. Powell, Akshay Sriprasad, Wesley J. Cole, Thomas F. Edgar. Heating, cooling, and electrical load forecasting for a large-scale district energy system, Energy, 2014(74): 877-885.
[5] LIU Zhiqing, Xu Huafeng, WANG Chaoyong, LIU Yamei. Research on Normalization Method of Scores of Scientific Research Based on Sigmoid Function[J]. Journal of Xinxiang University, 2019, 36(12): 19-22.
[6] Chen Feixiang, Xu Jianqun, Wang Chenyang, et al. Research on building cooling and heating load building cooling and heating load prediction model on user’s side in energy internet system[J]. Proceedings of the CSEE, 2015, 35(14): 3678-3684.
[7] ZHAO Jun, SU Pengwei, AN Qingsong, et al. Progress in Researches on Load Forecasting Technology for Optimization of Distributed Energy System[J]. Automation of Electric Power Systems, 2018, 30(08): 12-18.