An Efficient Short-Term Electricity Forecasting Approach Based on EEMD-LSTM Model with Feature Factors

Y J Tian¹, J R Luo², M Wen² and J G Li²
¹State Grid Shanghai Municipal Electric Power Company, Shanghai, China
²Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China

E-mail: 13901712348@163.com

Abstract. The forecasting of electricity consumption data plays an important role in the operation, planning, and security of the power grid. However, electricity data is affected by multiple factors and large fluctuations, which makes it difficult to accurately forecast. Traditionally, ARIMA and SVM are widely used for electricity forecasting based on historical consumption data. However, for non-stationary multi-feature data, traditional schemes cannot achieve deep feature mining of them, and the forecast results are inaccurate. To address this problem, this paper proposes an efficient short-term electricity forecasting approach based on EEMD-LSTM model. Firstly, we perform Savitzky-golay (SG) smoothing on the original data, and then introduce feature factors to the feature analysis. In particular, the proposed approach can reduce the random noise in data, as well as reduce the impact of data fluctuations, and effectively learn the long-term characteristics of the data. The simulation results show that, compared with ARIMA, LSTM, EMD-SVM, EMD-LSTM, the proposed approach can achieve better accuracy in the electricity forecasting.

1. Introduction

With the usage of smart meters, electricity consumption data grows every day. In the big data era, analysis of the inherent characteristics among the electricity consumption data can make more accurate electricity forecasting, which is vital to the electric power dispatching, planning and other power operation. However, the data is affected by different features, traditional manners cannot be applied directly to analyse them. Correspondingly, machine learning and deep learning have begun to be applied to do so. Generally, electricity forecasting is classified into four categories: ultra-short-term, short-term, medium-term and long-term. Among them, the ultra-short term is a forecast for the next few minutes to several hours, the short-term is a forecast for the next few days, the medium-term is a forecast for the next few weeks or months, and the long-term is for the next few years. In particular, short-term electricity forecasting is the most challenging one because it is closely related to the implementation of grid monthly dispatching.

A lot of researchers conducted predictive assessments. Legendre proposed the least squares (LS) [1], which is suitable for the case where the sequence shows a linear change trend, fitting the past sequence, and forecasting the subsequent sequence. In order to better fit the curve, YuLe proposed an autoregressive model (AR) [2]. Based on this model, Walker proposed a moving average model (MA) [3] and an autoregressive moving average model (ARMA) [4], while ARMA cannot fit non-stationary data. In response, Box and Jenkins proposed the autoregressive integrated moving average model
(ARIMA) [5]. Since ARIMA cannot capture non-linear relationships, Rumelhart and McClelland proposed the back propagation neural network (BP) [6], which is a multilayer feedforward neural network trained according to the error back propagation algorithm and it is the most widely used neural network at present. For large amount of data, the above schemes have limitations in use and low execution efficiency. Then, more and more current models are proposed by combinations, making full use of the advantages of multiple models. In [7], empirical mode decomposition (EMD) [8] was used to process the data and the long short-term memory (LSTM) model was used for forecasting. In [9], a combination model of EMD and SVM was used for forecast. For both models, the influence of the feature factors on the forecasting and the mode aliasing of EMD when decomposing the data are not considered. EEMD is an improvement of EMD. Huang can effectively address the mode aliasing problem in EMD through the method of noise addition [10]. LSTM is a special kind of recurrent neural network that can address long-term dependence problems. It was proposed by Hochreiter and Schmidhuber in 1997 [11]. Compared with recurrent neural network (RNN) [12], it can remember long-term Characteristic relationship and address the problem of gradient disappearance and gradient explosion. However, all the existing literatures cannot get more detailed feature correlation, and they did not consider the impact of noise in the data on the prediction. To overcome the above drawbacks of the existing literatures, this paper proposes an efficient short-term electricity forecasting approach. Specifically, it uses ensemble empirical mode decomposition (EEMD) to further decompose the original data, and fully consider the feature factors, and combine it with the LSTM model for forecasting.

To achieve better forecasting accuracy, the main contributions are as follows:

- First, this paper uses savitzky-golay (SG) [13] and EEMD to perform noise reduction and smooth processing on the original electricity consumption data.
- Secondly, the maximum information coefficient is used to perform correlation analysis on intrinsic mode functions (IMFs) to find the features that are highly relevant to these IMFs for data combination.
- Third, compared with traditional recurrent neural networks, the proposed approach can solve the problem of long-term dependence, avoid gradient disappearance and gradient explosion. Also, it can get high-precision forecast results.

The remainder of this paper is organized as follows. In Section II, conduct a preliminary knowledge introduction. In Section III, introduce this paper proposed approach. Then in Section IV, perform model performance evaluation and experimental analysis. Finally, conclude this paper in Section V.

2. Preliminaries

2.1. Maximal information coefficient

Maximum information coefficient (MIC) was proposed in 2011, and it is the latest method for detecting non-linear correlations between variables. MIC is universal and fair. Compared with other correlation analysis methods, its advantages are shown in Table 1.

| Method | Requirements | Standardization | Complexity | Robustness |
|--------|--------------|-----------------|-------------|------------|
| Pearson | Linear       | Yes             | Low         | Low        |
|        | Linear       | Non-linear      |             |            |
| Spearman| Linear       | Yes             | Low         | Medium     |
|        | Non-linear   |                 |             |            |
| KNN    | Linear       | No              | High        | High       |
|        | Non-linear   |                 |             |            |
| MIC    | Linear       | Yes             | Low         | High       |
|        | Non-linear   |                 |             |            |
$$I(x; y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dxdy$$  \hspace{1cm} (1)

$$mic(x; y) = \max_{a+b \leq B} I(x; y) \log_2 \min(a, b)$$  \hspace{1cm} (2)

where, $x, y$ is a sequence of two features. $a, b$ is the number of divisions on $x, y$. $B$ as a variable, usually 0.6 power of the total data.

2.2. Ensemble empirical mode decomposition

EEMD is an improvement of EMD. EMD is a method proposed by Huang et al. That can decompose signals into modal modes without using any well-defined functions. As the basis, only the IMF need to be generated adaptively based on the signal, which is suitable for analyzing non-linear and non-stationary signal sequences. EMD needs to satisfy when decomposing: the signal has at least two extreme points, and the time scale feature is determined by the time scale between the extreme points. Each IMF must satisfy: within the data range, the number of local extreme points and zero crossings must be equal or at most 1; at any time, the average value of the envelope of the local maximum and local minimum must be 0. However, mode aliasing occurs when EMD is decomposed, and the number of iterations in the process of decomposing IMF lacks a stopping criterion. In order to solve the above-mentioned problems of EMD, Huang uses a noise-assisted signal processing (NADA) to add noise to the signal for auxiliary analysis. White noise is added to the original signal, and the uniform distribution of the white noise spectrum is used. When the signal is added to a white noise background that is uniformly distributed throughout the time-frequency space, signals at different time scales are automatically distributed to the appropriate reference scale. The characteristics of zero mean, after several averages, the noise finally cancels each other.

$$IMFs = \sum_{m=1}^{MaxIter} \frac{C_{j,m}}{MaxIter}$$  \hspace{1cm} (3)

where, $MaxIter$ is the total number of times EMD was performed. $m$ represents the current number of times EMD has been performed. $C_{j,m}$ represents the $j$-th IMF decomposed from the $m$-th time.

2.3. Long short-term memory

LSTM is an efficient neural network that can rely on input gate, forget gate, and output gate to address long-term dependence problems as well as gradient explosions and gradient disappearances. The state of the input gate is obtained by merging the output of the hidden layer at the previous moment and the input state of the current moment and obtaining it through the sigmoid function. The state of the hidden layer at the previous moment and the state of the current moment are merged, and multiplied by sigmoid and tanh to obtain the state of the forget gate. The output gate state is a combination of the hidden state at the previous moment and the input state at the current moment, and is multiplied by the cell state processed by tanh through the sigmoid function. Updating the state is to multiply the state of the cell at the previous moment by the state of the input gate and add the state of the forget gate. Its internal structure is shown in Figure 1.
The LSTM formulas are as follows:

\[ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \]  (4)
\[ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \]  (5)
\[ C'_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \]  (6)
\[ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \]  (7)
\[ C_t = f_t \cdot C_{t-1} + i_t \cdot C'_t \]  (8)
\[ h_t = o_t \cdot \tanh(C_t) \]  (9)

where, \( i_t, f_t, o_t \) represent input gate, forget gate and output gate respectively. \( h_t \) is the layer output at time \( t \), \( h_{t-1} \) is the layer output at time \( t-1 \). \( C'_t \) is the cell status update, \( C_t \) is the cell state output. \( x_t \) is the current time input. \( W_i, W_f, W_o, W_c \) are the update weights of input gate, forget gate, output gate, and cell state. \( b_i, b_f, b_o, b_c \) are input gate, forget gate, output gate, and cell state bias.

3. The proposed approach
This paper proposes an efficient short-term electricity forecasting approach by combining SG and EEMD with the LSTM model, and performs feature analysis to date highly correlated data combination features. The research object is the forecasting of electricity consumption in a certain area of Shanghai and the data situation is given in detail in Section IV. The proposed approach mainly consists of the following three phases: system model, data processing, and the proposed forecasting model.

3.1. System model
The forecasting approach proposed in this paper will forecast the electricity data of the next 5 days in a rolling form based on the electricity data of the past 25 days, and finally forecast the electricity situation of the next 5 days, as shown in Figure 2.
3.2. Data preprocessing
Find the missing part of the data value is present, with a mean value of data before and after filling, and the data analyzed using box plots, which identify outlier data, the same data before and after smoothed average. Because the LSTM model is sensitive to data, the data is pre-processed using normalization.

\[ y = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  

where, \( x, y \) are sequences. \( x_{\text{min}} \) is the minimum value of the \( x \) and \( x_{\text{max}} \) is the maximum value of the \( x \).

3.3. Forecasting model
The preliminary knowledge involved in this model has been described in the section II, and present the proposed model in detail here. Divide the data into consumption and other features, which include holidays, rainfall, wind speed, pressure, humidity, and temperature. For the electricity consumption data, construct supervised learning, using the past 25 days as the training sample \( x \), and the next 5 days as the label \( y \), and constructing the training samples in a sliding window manner, as shown in Figure 2. The structure of this model is shown in Figure 3.

**Figure 3. LSTM model structure based on EEMD combined with feature processing**

Next, use the SG to smooth the training sample \( x \), Let the filter window width be \( n=2m+1 \) for each measurement point \( x=(-m, -m+1, \ldots, 0, 1, \ldots, m-1, m) \), Use k-1 degree polynomial to fit the data points in the window, which is \( y=a_0+a_1x+a_2x^2+\ldots+a_kx^k \). After constructing into matrix form, get \( Y(2m+1) \times 1 = X(2m+1) \times k \times A(k+1)+E(2m+1)+1 \). The least square solution of \( A \) is \( A' = (X'X)^{-1}X'Y \) and the filter value \( Y' = XA = X(X'X)^{-1}X'Y \). After the data is processed by SG, random noise is reduced. The processed data can be used for EEMD to obtain multiple IMFs. Stop conditions are controlled by standard deviation. The specific process is shown in Figure 4.

**Figure 4. Decomposition flowchart of EEMD**

\[ x(t) = \sum_{i=1}^{n} imfi(t) + r_n(t) \]  

(11)
\[ S_d = \frac{1}{T} \sum_{t=0}^{T} \left( \frac{|h_{j-i}(t) - h_j(t)|^2}{h_j^2(t)} \right) \]  

where, \( imf_i(t) \) is the \( i \)-th IMF of the EMD decomposition, \( r_n(t) \) is the residual after decomposition and screening of \( n \) IMFs, \( S_d \) is the standard deviation, \( h_j(t) \) is the residual, \( h_{j-i}(t) \) is the residual from the \( i \)-th component.

As a result, the data pre-processing part is completed, and the EEMD processing result is shown in Figure 5. The IMFs and the separated features are subjected to MIC calculation, and the IMFs are classified into two parts with and without feature correlation. The specific formulas are shown in (1) and (2).

![IMFs derived from electricity data decomposition](image)

**Figure 5.** IMFs derived from electricity data decomposition

Then, the obtained IMFs and features are solved for the maximum information coefficient. As shown in Table 2, the highly relevant data are selected as the features of the corresponding IMFs and combined to obtain new data.

**Table 2.** MIC analysis between IMFs and features

|          | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 | IMF7 | IMF8 | IMF9 | IMF10 |
|----------|------|------|------|------|------|------|------|------|------|-------|
| Holiday  | 0.17 | 0.20 | 0.14 | 0.13 | 0.11 | 0.12 | 0.13 | 0.15 | 0.13 | 0.03  |
| Rainfall | 0.11 | 0.11 | 0.10 | 0.12 | 0.13 | 0.14 | 0.15 | 0.17 | 0.20 | 0.10  |
| Wind speed | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.15 | 0.16 | 0.16 | 0.20 | 0.11  |
| Pressure | 0.13 | 0.13 | 0.13 | 0.14 | 0.24 | 0.37 | 0.37 | 0.41 | 0.74 | 0.12  |
| Humidity | 0.11 | 0.12 | 0.13 | 0.13 | 0.15 | 0.16 | 0.17 | 0.24 | 0.27 | 0.12  |
| Temperature | 0.13 | 0.13 | 0.14 | 0.14 | 0.40 | 0.50 | 0.37 | 0.53 | 0.85 | 0.11  |

Finally, the constructed new data set is input into the LSTM model. Since the data is processed multiple times, the LSTM can learn more effectively. Through multiple iterative training of the model, relying on the characteristics of the neural network, it is possible to learn the deep correlation of the data. The unique input gate, forget gate, and output gate structure can record useful historical information and update the weights, so you can learn more feature correlations and achieve the purpose of final forecasting.

### 4. Performance evaluation

#### 4.1. Experimental data
In this paper, the used data set are daily electricity consumption in a certain area of Shanghai. There are more than 2.2 million electricity users in this area, including low-voltage residents, low-voltage industrial and commercial, and high-voltage industrial and commercial. The area has high temperatures, heavy precipitation, and low sunshine hours; disaster weather is frequent, and short-term strong convective weather is more, so weather factors have a greater impact on electricity. The time range is from January 9, 2014 to June 19, 2018. The electricity amount from January 9, 2014 to July 6, 2017 was used as the training set, and the electricity amount from July 7, 2017 to June 19, 2018 was used as the test set. Use the electricity data of the past 25 days as a training sample to forecast the electricity of the fifth day in the future, and use the sliding window to forecast the electricity of the next 5 days.

4.2. Feature selection
The multiple IMFs obtained by EEMD are shown in Figure 5. The data of this experiment are electricity, rainfall, wind speed, pressure, humidity, temperature, and holidays as features. Correlation analysis was performed for each IMFs component and feature using the maximum information coefficient.

As shown in Table 2, the maximum information coefficient of average pressure and IMFs9 reached 0.74, the maximum information coefficient of average temperature and IMFs6 reached 0.50, the maximum information coefficient of IMFs8 reached 0.53, and the maximum information coefficient of IMFs9 reached 0.85. So for IMFs6, this article will introduce temperature characteristics, for IMFs8, introduce temperature characteristics, and for IMFs9, introduce pressure and temperature characteristics. The evaluation indicators are MAPE and RMSE:

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{o_t - y_t}{o_t} \right| * \frac{100}{n}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (o_t - y_t)^2}
\]

where, \(o_t\) is true value at time \(t\), \(y_t\) is forecast value at time \(t\).\(n\) is the total amount of data.

4.3. Performance and analysis
Compared with the traditional scheme or the existing combination model, from Figure 6, that the global analysis, the proposed approach in this paper can better fit the curve. In particular, from Figure 7’s local analysis, the proposed approach can achieve closer results to peaks and valleys than other approaches.
Figure 7. Comparison of local details

As shown in Table 3, ARIMA has achieved better results. However, from both Figures 6 and 7, it can be found that the effect of ARIMA is good, because the results cannot be forecasted well, and the obtained results are only lagging fluctuations of the input values. By observing the graphs and tables, it can be found that the proposed approach in this paper has achieved the best results regardless of the curve fitting situation or RMSE and MAPE.

Table 3. Comparison of experimental results

| Model       | RMSE (10^6) | MAPE (%) |
|-------------|-------------|----------|
| ARIMA [4]   | 5.2         | 4.9      |
| LSTM [11]   | 9.2         | 8.79     |
| EMD-LSTM [7]| 4.2         | 4.12     |
| EMD-SVM [9] | 6.1         | 5.71     |
| Proposed    | 3.7         | 3.29     |

For model prediction, in addition to performance, model processing time is also important. As shown in Table 4, comparison of model processing time was performed. The running time of ARIMA, LSTM, EMD-SVM is short, while EMD-LSTM and the method proposed in this paper take about 10 minutes. But for power forecasting, 10 minutes can reduce MAPE to 3.29, which is a great improvement, and the purpose of this time is to achieve a short-term forecast of 5 days, so it will not affect the ultimate purpose of the experiment.

Table 4. Comparison of experimental time

| Model       | Time (s) |
|-------------|----------|
| ARIMA [4]   | 61       |
| LSTM [11]   | 77       |
| EMD-LSTM [7]| 531      |
| EMD-SVM [9] | 87       |
| Proposed    | 652      |

5. Conclusion

In this paper, an efficient short-term electricity forecasting approach is proposed based on SG filtering, EEMD and LSTM. It performs data analysis on the data set, and obtains features that are highly related to IMFs. Comparative experiments are performed. The experimental results show that the proposed approach can get better electricity forecasting accuracy that others. For our future work, we
will introduce heuristic algorithms to optimize model parameters and improve the LSTM model so that it can provide more efficient performance for specific scenarios.

Acknowledgements
This work is supported by the State Grid Shanghai Municipal Electric Power Company Project of China (No.5209280002B).

References
[1] Kim J M and Jung H J 2018 Time series forecasting using functional partial least square regression with stochastic volatility, garch, and exponential smoothing J. Forecast. 37 269-80
[2] Osborn D R and Smith J P 1989 The performance of periodic autoregressive models in forecasting seasonal UK consumption J. Bus. Econ. Stat. 7 117–27
[3] Robinson P M 1977 The estimation of a nonlinear moving average model Stoch. Process. Their Appl. 5 81–90
[4] Perez E G and Ceballos R F 2019 Malaria incidence in the Philippines: prediction using the autoregressive moving average models Int. J. Eng. Futur. Technol. (2019) 16 1-10
[5] Zhang L Y, Lin J, Qiu R Z, Hu X S, Zhang H H, Chen Q Y, Tan H M, Lin D and Wang J K 2018 Trend analysis and forecast of PM2.5 in Fuzhou, China using the arima model Ecol. Indic. 95 702–10
[6] Chi W L 2018 Forecasting stock index based on bp neural network algorithm 2018 Int. Conf. Mech., Electron., Control Autom. Eng. (MECAE 2018)
[7] Zhang B N 2018 Foreign exchange rates forecasting with an emd-lstm neural networks model J. Phys. Conf. Ser. 1053
[8] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q A, Yen N C, Tung C C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci. 454 903–95
[9] Yaslan Y and Bican B 2017 Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting Measurement 103 52–61
[10] Tan Q F, Lei X H, Wang X, Wang H, Wen X, Ji Y and Kang A Q 2018 An adaptive middle and long-term runoff forecast model using eemd-ann hybrid approach J. Hydrol. 567 767–80
[11] Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory Neural Comput. 9 1735–80
[12] Hajibotorabi Z, Kazemi A, Samavati F F and Ghaini F M M 2019 Improving dwt-rnn model via b-spline wavelet multiresolution to forecast a high-frequency time series Expert Syst. Appl. 138 112842
[13] Cao R, Chen Y, Shen M, Chen J, Zhou J, Wang C and Yang W 2018 A simple method to improve the quality of ndvi time-series data by integrating spatiotemporal information with the savitzky-golay filter Remote Sens. Environ. 217 244–57