Application of Variational Mode Decomposition and Deep Learning in Short-Term Power Load Forecasting

Ping Yu¹, Jie Fang¹, Yubao Xu¹, Qiang Shi²

¹College of Electronical and Photoelectronic Engineering West Anhui University, Lu'an, Anhui 237012, China
²Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd, Guangzhou 510620, China

*Corresponding author’s e-mail: yuping@wxc.edu.cn

Abstract. Accurate load forecasting of power system operation and development is of great significance. Because of the power load time series has strong nonlinear, the traditional forecasting model does not apply. Therefore, a short-term load forecasting model based on variational model (VMD) and length (LSTM) is proposed. Firstly, the VMD decomposes the original load sequence to get the modal of different size frequency component. Then, phase space reconstruction (PSR) organizes the modal components into deep learning inputs. Then, the LSTM network is employed to predict each group of modal components. Finally, all the modal component predictive value addition to the power load to predict the future. The experimental results show that compared with the BP, LSTM and EEMD-LSTM model, the model completely weakens the non-stationary load sequence, minimize the prediction error, reached the highest prediction accuracy.

1. INTRODUCTION

Traditional power load prediction models mainly include: meteorological physics [¹], mathematical statistics [²] and artificial intelligence [³]. However, in nonlinear power load sequence prediction, it is difficult for this model to obtain good prediction effect [⁴]. Therefore, at present, the research focuses more on weakening the non-stationarity of load time series and reducing the difficulty of forecasting. Then, a hybrid forecasting model combining multiple methods is adopted to achieve accurate forecasting of load series. In research [⁵], this paper proposes a BP neural network based on particle swarm optimization for short-term load forecasting method, vector of the improved BP network, and have achieved good prediction effect. In [⁶], combined with the empirical mode decomposition (EMD) and LSTM neural network to predict the impact of load similar days (SD), the experimental results verify the effectiveness of the model. In [³], combined with the influence of empirical mode decomposition (EMD) and LSTM neural network on predicting load similar days (SD), the experimental results verify the validity of the model. In [⁷], considering the poor effect of the empirical mode decomposition, using the variational mode decomposition (VMD) and extreme learning machine (ELM) for load forecasting, which can effectively weaken the nonlinear load sequence and improve the prediction accuracy. Therefore, this paper proposes a combined VMD, PSR and LSTM network hybrid model. The model aims to fully decompose the load sequences, weaken their non-stationarity, and analyze and learn the temporal relationship of load sequences to improve the accuracy of the short-term
load forecasting model. Through example calculation and contrast analysis, we verify the accuracy of the model.

2. BASIC THEORY

2.1 Variational mode decomposition

VMD is a kind of adaptive, completely non-recursive modal change and signal processing aliasing method. It overcomes the endpoint of the EMD method facing effect and modal component aliasing problems, has a more solid mathematics theory foundation, can reduce the time series non-stationarity, high complexity, and strongly nonlinear. This method can get different frequency scale subsequence, has a relatively stable subsequence, suitable for non-stationary short-term power load sequences.

VMD core idea is to construct and solve the variational problem. Assuming that the original signal \( f \) is decomposed into \( k \) components, to ensure that the decomposition sequence is a finite bandwidth mode with a center frequency component. Besides, In addition, the modal estimate, minimizing the sum of bandwidth constraint is the sum of the modal estimate bandwidth is equal to the original signal. The specific variational constraint expression is:

\[
\min_{\{u_i, \omega_i\}} \sum_k \left\| \frac{\partial}{\partial t} \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_i(t) \right] e^{-j\omega_i t} \right\|^2_2
\]

\[
\text{s.t.} \sum_{k=1}^{K} u_k = f
\]

Where \( K \) is the number of modal that need to be decomposed (a positive integer), \( \{u_i\}, \{\omega_i\} \) correspond to the \( k \)-th modal component, and the center frequency, \( \delta(t) \) is the Dirac function, and * is the convolution operator.

To solve the equation (1), introduce Lagrange multiplier, then convert constrained variational problem into unconstrained variational problem, so, augmented Lagrange expression is:

\[
L(\{u_i\}, \{\omega_i\}, \lambda) = \alpha \sum_k \left\| \frac{\partial}{\partial t} \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_i(t) \right] e^{-j\omega_i t} \right\|^2_2 + \left\| f(t) - \sum_k u_k(t) \right\|^2_2 + \left( \lambda(t), f(t) - \sum_k u_k(t) \right)
\]

Where \( \alpha \) is a quadratic penalty factor, whose function is to reduce the interference of Gaussian noise. The expressions of \( u_i, \omega_i \) and \( \lambda \) after the iterative alternating optimization are given by equations (3)-(5). The detailed process can be studied in [7,8].

\[
u_k^{n+1}(\omega) \leftarrow \frac{f(\omega) - \sum_{j=1}^{n} \hat{u}_j(\omega) + \hat{\lambda}(\omega)}{1 + 2\alpha(\omega - \omega_j)^2}
\]

\[
\omega_k^{n+1} \leftarrow \frac{\int_0^\omega \left\| u_k^{n+1}(\omega) \right\|^2 d\omega}{\int_0^\omega \left\| u_k^{n+1}(\omega) \right\|^2 d\omega}
\]

\[
\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \gamma \left( f(\omega) - \sum_{k=1}^{n+1} \hat{u}_k(\omega) \right)
\]

Where: \( \gamma \) is the noise tolerance. It meets the requirement of the fidelity of signal decomposition. \( u_k^{n+1}(\omega), \hat{u}(\omega), f(\omega) \) and \( \hat{\lambda}(\omega) \) correspond to the \( u_k(t), u(t), f(t) \) and \( \hat{\lambda}(t) \) form of Fourier transform.
VMD main iterative solving process is:
Step1: initialize \( \hat{u}_k \), \( \hat{\omega}_k \), \( \lambda^1 \) and maximum the number of iterations \( N \), \( n \leftarrow 0 \);
Step2: use formulas (3) and (4) to update \( \hat{u}_k \) and \( \hat{\omega}_k \);
Step3: update \( \lambda \) using formula (5);
Step4: accuracy convergence criterion \( \varepsilon > 0 \). If \( \sum_{k}||u_{n+k} - \hat{u}_k||^2_2 / ||u_k||^2_2 < \varepsilon \) and \( n < N \) are not satisfied, return the second step. Otherwise, the iteration is completed, and the last \( \hat{u}_k \) and \( \hat{\omega}_k \) are the outputs.

2.2 Phase space reconstruction
Phase space reconstruction (PSR) technology was put forward by Packard in 1980 with the initial purpose of recovering chaotic attractors from high-dimensional phase space\(^9\). See formulas (6) - (7) for the specific reconstruction process. This technique has two key parameters: delay time \( \tau \) and embedded dimension \( d \), which need to be selected according to the actual situation.

In this paper, coordinate delay reconstruction method, construction delay time \( \tau \) and embedding dimension \( d \) modal components in phase space vector sequence. Therefore, for the modal component sequence \( h \) decomposed by VMD, the PSR expression of \( x = \{x_i | i = 1, 2, ..., N\} \) at different prediction layers can be expressed as:

\[
X = \begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_L \\
\end{bmatrix} = \begin{bmatrix}
x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\
\vdots & \vdots & \ddots & \vdots \\
x_i & x_{i+\tau} & \cdots & x_{i+(d-1)\tau} \\
\vdots & \vdots & \ddots & \vdots \\
x_L & x_{L+\tau} & \cdots & x_{L+(d-1)\tau} \\
\end{bmatrix}
\]

(6)

Where: \( L = N - (d-1) \cdot \tau - h \), \( N \) are the total number of load samples, \( \tau \) and \( d \) are delay time and embedding dimension respectively, and \( X_i \) \( (i = 1, 2, ..., L) \) is the the phase space of \( i \)-th space vector. The output matrix corresponding to the prediction model can be derived from the following formula:

\[
O = \begin{bmatrix}
O_1 & O_2 & \cdots & O_i \\
\end{bmatrix}^T = \begin{bmatrix}
x_{1+h+(d-1)\tau} & x_{2+h+(d-3)\tau} & \cdots & x_N \\
\end{bmatrix}^T
\]

(7)

Where \( O_i \) represents the predicted value of the \( i \)-th vector corresponding to the phase space matrix.

2.3 Long-short time memory network
LSTM as RNN in deep learning is one of the best variant, avoids the RNN long-term dependency problem, solve the problem of the gradient of the gradient transfer process disappear. LSTM is very suitable for processing and predicting highly correlated problems of the time series\[^{10}\]. The important structures to achieve its flexible operation are the forget gate, the input gate and the output gate in turn, as shown in Fig.1. The cell state is controlled by the above three gates, and state information of LSTM networks is selectively retained and forgotten.
The first step of LSTM is to check the information of the current $h_{i-1}$ and $x_t$ through the sigmoid unit in the forget gate, and determine which information should be discarded or retained by the network status $C_{i-1}$ at the last moment. Where $C_{i-1} \in [0,1]$ , 0 represents total forgetting, and 1 represents retention.

$$f_t = \sigma(W_f \cdot [h_{i-1}, x_t] + b_f) \tag{8}$$

The $h_{i-1}$, $x_t$ and tanh network layers in the input gate are then used to determine which information to update to obtain the new candidate network status $C_t$.

$$i_t = \sigma(W_i \cdot [h_{i-1}, x_t] + b_i)$$

$$C_t = \tanh(W_c \cdot [h_{i-1}, x_t] + b_c) \tag{9}$$

Using the forgotten information of cell status $C_{i-1}$ at the last moment and the input information of candidate cell status $C_t$, the new cell information $C_t$ was obtained. Where $\odot$ represents the Hadamard product, following the multiplication of corresponding elements.

$$C_t = f_t \odot C_{i-1} + i_t \odot C_t \tag{10}$$

Finally, the cell state information was determined. The sigmoid unit of the output gate needs to combine the input $h_{i-1}$ and $x_t$ information to obtain the output judgment condition $o_t$, and the cell information $C_t$ gets the output vector through the tanh layer. The judgment condition and the output vector are multiplied by the Hadamard product, to obtain the output of the LSTM unit.

$$o_t = \sigma(W_o \cdot [h_{i-1}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(C_t) \tag{11}$$

**3. A HYBRID MODEL FOR SHORT-TERM LOAD FORECASTING**

3.1 VMD-PSR-LSTM short-term load forecasting model

To reduce the nonlinearity and volatility of short-term power load series, reduce the difficulty of prediction and improve the prediction accuracy, a hybrid short-term power load-forecasting model VMD-PSR-LSTM is proposed. The specific operating steps of the model are shown in Figure 2.
As can be seen from Fig. 2, this model firstly decomposes the signals of the original load sequence with VMD, weakens the noise information of the sequence, and then uses the adaptive decomposition algorithm to obtain the different frequency-scale components (IMF) of each mode. Then, by using the PSR technique of chaotic timing sequence analysis, appropriate delay time and embedded dimension are selected to optimize and recombine the modal components obtained by the decomposition to form the input of deep learning. After the reconstructed modal component is normalized, the LSTM neural network with strong nonlinear fitting ability in deep learning is used to train and predict each modal component. In the end, all the predicted values were normalized and summed up to complete the solution of the prediction model, and finally, the power load prediction was realized.

3.2 Evaluation indicators
The following evaluation indexes were used to test the prediction effect of the model in a mathematical statistic. Where $y_i$ represents the actual value of the power load, $\bar{y}$ represents the average value, and $\hat{y}_i$ represents the predicted value.

i. **Mean square error (MAE)**

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$  \hspace{1cm} (12)

ii. **Root mean square error (RMSE)**

$$RMSE = \left( \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right)^{\frac{1}{2}}$$  \hspace{1cm} (13)

iii. **Mean absolute percentage error (MAPE)**

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$  \hspace{1cm} (14)

iv. **Determination coefficient ($R^2$)**
\[ R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \]  

(15)

3.3 Set the contrast

To verify the effectiveness of the proposed model, three comparative models (BP, LSTM, and EEMD-LSTM) were developed and compared with the proposed hybrid model. The predictive performance of each model was demonstrated through the size of the error value and the fitting effect.

4. EXAMPLE CALCULATION AND MODEL SOLUTION

4.1 Data collection

This study selected data are from Jiangyin city, Jiangsu province, China, August 1 and August 7, 2018. The sampling frequency is 15 minutes per time. The data volume of a day is 96 samples, a total of 672 samples, as shown in Figure 3. Where, 576 samples from the first 6 days were selected as the training set, and 96 samples from the last 1 day were selected as the test set.

Fig. 3. Measured load sequence

You can see from Figure 3, the original power load sequence has stronger nonlinearity and volatility, which makes it difficult to predict the future load sequence.

4.2 Model solution

To reduce the case of mode aliasing and fully separate the modal components, three important parameters in the VMD are set as \( K = 5 \), \( \gamma = 0 \) and \( \alpha = 2000 \). Five different frequency scale modal components of the original power load sequence are obtained, as shown in Figure 4.

Fig. 4. The modal component IMF1-IMF5
After several attempts, two parameters of phase space reconstruction were set as $\tau=1$ and $d=6$. This is beneficial to simulate the randomness of load sequence and to mine its chaotic temporal relation$^{[9]}$. The optimized and recombinated five modal components of PSR are normalized and input into the LSTM network, which is conducive to the rapid learning and solution of the LSTM network. Using root-mean-square error MSE as LSTM error transfer function. To improve learning efficiency, the activation function was selected as relu, and the parameters of LSTM network were optimized by the Adam algorithm. The number of network iterations is set to 200. After the inverse normalization, the predicted values of all the solved modes were summarized to realize the final short-term load prediction. The prediction effect of the proposed hybrid model is shown in the following figure 5.

Fig. 5. The prediction effect diagram of the proposed model

5. ANALYSIS OF EXAMPLES

5.1 Model contrast

BP, LSTM, and EEMD-LSTM models were used for comparison models to compare the accuracy of different models in predicting the actual power load sequence. The above three models adopt the same iteration times, and the relevant parameters take the default values. The above three models were solved, and the prediction effects of all models were obtained as shown in Figure 6.

Fig. 6. The prediction effect of each model
In picture 6, the model proposed in this paper is significantly better than BP, LSTM and EEMD-LSTM in the fitting effect of test set. Table 1 shows the indices prediction error of the model on the test set and the fitting effect. Tab.1 shows the prediction error index and fitting effect of all models on the test set.

Tab. 1 Prediction results of models in test sets

| Models      | MAE/MW | RMSE/MW | MAPE/% | R²/% |
|-------------|--------|---------|--------|------|
| BP          | 119.49 | 143.06  | 3.03   | 53.09|
| LSTM        | 104.97 | 128.86  | 2.68   | 61.94|
| EEMD-LSTM   | 62.32  | 73.41   | 1.59   | 87.65|
| The proposed| 41.03  | 50.11   | 1.06   | 94.24|

5.2 Result analysis

Through the above research and comparative analysis, the VMD-PSR-LSTM hybrid model obtained the best prediction effect, effectively predicted the short-term power load sequence, the evaluation index MAE was 41.03 MW, RMSE was 50.11 MW, MAPE was 1.06%, and the correlation coefficient R was 94.24%. The validity of the proposed model is verified. Figure 7 shows the comparison of evaluation indicators for all models. Among them, the BP model shows the worst prediction results, and the LSTM network is slightly better than the BP model, indicating that the direct use of the neural network to predict nonlinear load sequence cannot achieve better prediction results. Compared with the LSTM model, the EEMD-LSTM hybrid model shows significantly improved results. Since the original sequence is preprocessed by EEMD, the nonlinearity of the original sequence is reduced, and consequently, the LSTM model is improved to some extent. However, the "endpoint effect" and "mode aliasing" still exist in the mode decomposition using EEMD. It limits the prediction performance of the LSTM model.

The proposed model adopts VMD decomposition, avoids the weakness of EEMD. The relatively stable load sequence is obtained and the difficulty of forecasting is reduced. Meanwhile, the modal component sequence is optimized by using PSR, and the LSTM network with strong nonlinear fitting and learning ability in deep learning is adopted for combined prediction, and the future load sequence is predicted successfully. MAPE model was taken as an example to compare with EEMD-LSTM model. The error of the proposed model was reduced by 33.33%.

6. CONCLUSION

This paper proposes a hybrid model based on VMD, PSR, and LSTM network to fully weaken the nonlinear power load sequence, reduce the difficulty of prediction, and improve the prediction accuracy. The stable modal component is obtained through the example application, and the optimized load component sequence is input into the LSTM network with a strong learning ability for training and prediction. By comparing BP, LSTM, EEMD-LSTM, and other prediction models, The model has the
smallest prediction error, the highest fitting accuracy and the best prediction effect. It verifies the effectiveness and accuracy of the proposed hybrid model in predicting non-stationary short-term power loads.

About the author:
Yu ping (1980 --), female, Anhui province, master's degree, lecturer, research direction: electronic circuit.

ACKNOWLEDGEMENTS
This work was supported by the National Natural Science Foundation (NO. 61302179).

REFERENCES
[1] Li WQ, Chang L. (2018) A combination model with variable weight optimization for short-term electrical load forecasting[J]. Energy, 164:575-593.
[2] Amjady N. (2001) Short-term hourly load forecasting using time-series modeling with peak load estimation capability[J]. IEEE Transactions on Power Systems, 16(4):798-805.
[3] Zhao SH, Li CH, Shen YL, et al. (2017) A Hybrid Model Equipped with the Minimum Cycle Decomposition Concept for Short-Term Forecasting of Electrical Load Time Series[J]. Neural Processing Letters, 46(2):1059-1081.
[4] Eugene A. Feinberg, Dora Genethliou. (2006) Load Forecasting[M]. Applied Mathematics for Restructured Electric Power Systems, pp.269-285.
[5] Sun W, Zou Y. (2007) Short Term Load Forecasting Based on BP Neural Network Trained by PSO, International Conference on Machine Learning & Cybernetics. IEEE.
[6] Zheng HT, Yuan JB, Chen Long. (2017) Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation[J]. Energies, 10(8):1168.
[7] Li W, Quan CX, et al. (2018) Short-Term Power Load Forecasting Based on a Combination of VMD and ELM[J]. Polish Journal of Environmental Studies,27(5):2143-2154
[8] Dragomiretskiy, K; Zosso, D. (2014) Variational mode decomposition [J]. IEEE Trans. Signal Process. 62(3),531–544.
[9] Packard, N H, Crutchfield, J P, et al. (2008) Geometry from a time series [J]. Phys Rev Lett, 45:712.
[10] S. Hochreiter and J. Schmidhuber. (1997) Long short-term memory. Neural Computation, 9(8):1735–1780.
[11] Li W, Shi Q, et al. (2020) A Hybrid Forecasting Model for Short-Term Power Load Based on Sample Entropy, Two-Phase Decomposition and Whale Algorithm Optimized Support Vector Regression[J]. IEEE Access, 8: 166907-166921.