Ultra-Short-term Electric Load Forecasting Based on VMD-BiLSTM Model

Jiakun Chen [1, a], Wanxing Ma [2, b], Zhimin Chen [1, c]

[1]School of Electronic and Information, Shanghai Dianji University, Shanghai 201306, China;

[2] Ningguo Power Supply Company, State Grid Anhui Electric Power Company, Anhui 242300, China.

[a] 1204110173@qq.com, [b] mawx2020@163.com, [c] chenzm@sdju.edu.cn

Abstract. Fast and accurate ultra-short-term load forecasting is beneficial for building an efficient and modern smart grid. This paper proposes an ultra-short-term load forecasting model based on Variational Mode Decomposition (VMD) and improved Bi-directional Long Short-Term Memory (BiLSTM) network. The model first decomposes the complex load historical sequence using VMD, reducing the non-stationarity and complexity of the load sequence to facilitate the prediction task using neural networks. Then, the improved BiLSTM network is used to allow input data to skip some transformations in the network, thereby deeply and bidirectionally mining various hidden information from historical sequences. Finally, this paper selects temperature as the factor most correlated with power load forecasting and incorporates it into the prediction process to achieve effective ultra-short-term load forecasting. Compared with existing models such as LSTM, SVM, LeNet, DRNet4-7, and unimproved VMD-BiLSTM based on example data, the simulation results show that the proposed model can predict the load changes within one hour in the future, achieving ultra-short-term load forecasting, and has good prediction accuracy and algorithm robustness.

Keywords: smart grid, ultra-short-term electric load forecasting, variational mode decomposition, improved bidirectional long short-term memory network.

1. Introduction

Intelligent scheduling technology and advanced demand-side management are the main components of smart grid construction [1]. Accurate load forecasting, as a key technology, is an important prerequisite for building smart grids and virtual power plants. However, with the integration of different types of clean energy sources into the grid, the non-stationarity and complexity of power load sequences have increased significantly, further increasing the difficulty of forecasting. Traditional forecasting methods such as regression analysis and exponential smoothing are no longer able to meet the accuracy requirements of modern smart grids.

In order to further reduce the impact of non-stationarity and complexity in ultra-short-term power load sequences, some scholars have adopted adaptive empirical mode decomposition [2] to decompose the signal without the need for complete prior knowledge, but there is a phenomenon of mode mixing. On the other hand, variational mode decomposition can set the number of sub-sequences generated by a sequence to a predetermined value during sequence decomposition, overcoming the phenomenon of mode mixing [3], and is often used to deal with non-stationary and nonlinear long sequences.

The mainstream load forecasting model, DRNet4-7, was proposed by Xiaokang [4] He et al. This model adopts a residual learning approach to address the gradient vanishing and exploding problems in deep learning models. At the same time, the design of DRNet4-7 using multi-scale inputs and a multi-branch network effectively enhances the model's expressive power. However, the ability of DRNet4-7 [5] to handle outliers and noise still needs further improvement. P. Miao [6] formulates channel equalization as a memory prediction problem with time series characteristics using a clever input representation. LSTM inherits the chain-like structure of RNN that consists of repeated neural network modules. It can handle long-term dependencies and store memory parameters, making it capable of accurately recovering the original transmission signal with complex channel characteristics and fast convergence speed. However, LSTM only considers the information contained in the historical load sequence and cannot capture hidden information at future time steps. Therefore, by adding a bidirectional LSTM with a backward propagation input information module on the basis of the LSTM model [7], it is possible to more comprehensively and efficiently mine the context hidden information in the time series.

This article proposes a model based on VMD-BiLSTM. Firstly, decompose the sequence using VMD, and by limiting the bandwidth of the sub-sequences, load sub-sequences with more pronounced periodicity and concentrated frequencies are obtained. Inspired by the ResNet network, an improved two-layer bidirectional LSTM network is employed to allow input data to skip certain transformations in the network, enabling deep bidirectional exploration of various hidden information in historical sequences, thereby improving the stability and accuracy of network convergence. Next, based on the decomposed load sequences and temperature information, the model predicts the load for the next hour.

2. Basic Principles of Variational Modal Decomposition

Variational mode decomposition is a new signal decomposition method proposed by Dragomiretskiy [8] in 2014. It decomposed the time-series signal into a series of intrinsic mode functions (IMF) with different central frequencies, effectively overcoming the end-point effect and mode mixing phenomenon.

The variational problem involves minimizing the sum of estimated bandwidth for all IMFs, with the constraint that the sum of all IMFs is the signal before decomposition.

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \sum\limits_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ \text{s.}t. \sum\limits_{k=1}^{K} u_k(t) = f(t) \end{cases} \tag{1}$$

Among them, u_k is the sum of all IMFs, ω_k is the set of center frequencies for each mode, which is the Dirac distribution, K is the decomposition number, and $f(t)$ is the original sequence.

Obtaining unconstrained variational problems by adding penalty factor α and Lagrange multiplication operator λ. The expression for this problem is as follows:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = & \\ & \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 \\ & + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \end{aligned} \tag{2}$$

Use the alternating direction multiplier method to solve the unconstrained variational problem mentioned above, in order to obtain the optimal solution of the problem [9]. The formula for iteratively updating the intrinsic mode components and center frequency is as follows:

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}(\omega)-\sum_{i<k}\hat{u}_i^{n+1}(\omega)+\frac{\hat{\lambda}(\omega)}{2}}{1+2\alpha\left(\omega-\omega_k^n\right)^2} \tag{3}$$

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega\left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega} \tag{4}$$

Among them, $\hat{u}_k^{n+1}(\omega)$ is the Wiener filtering of the current remaining signal, and $\hat{u}_k(\omega)$ is subjected to inverse Fourier transform. The real part is $\{u_k(t)\}$, and ω_k^{n+1} is the center of gravity of the current IMF power spectrum.

3. BiLSTM Network and Its Improvement

3.1 BiLSTM networks

Traditional single-layer unidirectional LSTM networks train parameters in a forward direction according to the time sequence, which cannot capture the implicit information of future time steps and explore the internal features within the sequence. To address this issue, this paper introduces a bidirectional temporal prediction network that deeply explores various hidden information in the historical sequence. The structure of the BiLSTM model [11] is shown in Figure 2. Compared to the traditional LSTM network structure, the BiLSTM model can simultaneously perform forward and backward training. All A are trained in the forward direction, and all A' are trained in the backward direction, sharing their respective directional network parameters. Additionally, forward training and backward training are independent of each other, allowing for better exploration of the potential features in the data.

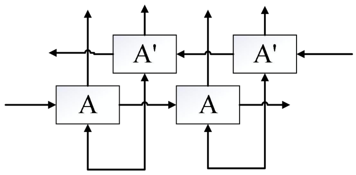

Fig.2 The structure of one-layer BiLSTM model

3.2 Improvement of the network

In our proposed method, we enhance the two-layer BiLSTM by concatenating the original input data with the outputs of both the first-layer and second-layer networks. This concatenated data is then passed through a fully connected linear layer to obtain the final prediction result [12]. Compared to the original approach, the improved network structure presented in this article demonstrates improvements in both training speed and accuracy without introducing additional computational burden. The improved network structure proposed in this article is shown in Figure 3.

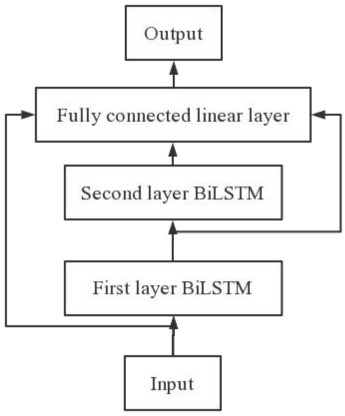

Fig.3 Improved network structure

The formula used for the fully connected linear layer in the improved two-layer BiLSTM model is as follows:

$$y = \boldsymbol{W}\boldsymbol{x}^{\mathrm{T}} + b \tag{5}$$

Where $y \in \mathbb{R}$ represents the output of the fully connected linear layer, with bias b and weights $\boldsymbol{W} \in \mathbb{R}^N$. $\boldsymbol{x} \in \mathbb{R}^N$ represents the N dimensional input obtained by concatenating the original input data with the outputs of each layer of the network.

4. Ultra-Short-term Electric Load Forecasting Based on VMD-BiLSTM Model

4.1 The model structure

The power load sequence has non-stationary and strong randomness, and exhibits periodicity with people's daily rhythms. The non-stationary load sequence is decomposed using VMD to obtain load sub-sequences with more obvious periodicity and more concentrated frequencies. Each load sub-sequence is then processed by an improved two-layer BiLSTM to achieve effective ultra-short-term load forecasting. The model structure is illustrated in Figure 4.

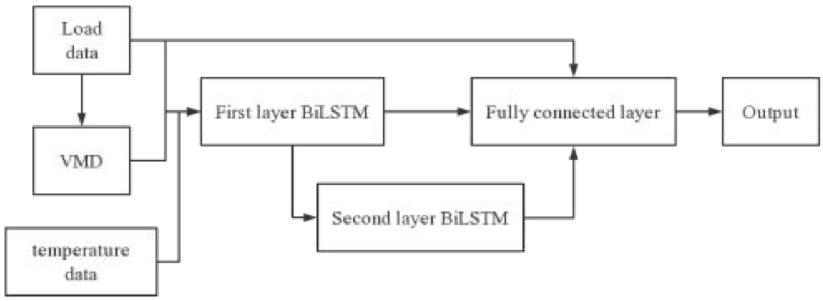

Fig.4 The structure of forecasting model based on VMD-BiLSTM network

The model utilizes VMD decomposition to decompose the non-stationary load data into multiple IMFs. Considering the impact of temperature on people’s daily electricity consumption, both the temperature data and the decomposed load data are inputted into the network. To maintain consistent data structure, the original load data is directly inputted into the fully connected linear layer of the improved double-layer BiLSTM network. The model evaluation indicators selected in this article are MAE, MSE and MAPE.

4.2 Method for optimizing network model parameters

The parameter optimization framework for the improved two-layer BiLSTM model in this paper is shown in Figure 5. Firstly, VMD decomposition is performed on the load data to obtain multiple IMF sequences, followed by data normalization with temperature data. Divide the training and testing sets by 8:1. The network-related hyperparameters and learning rate are initialized. The improved double-layer bidirectional LSTM network is iteratively trained using the pre-processed training set data. The MAPE of the network on the validation set is compared with the historical best MAPE. If the error is lower, the current network parameters are saved. This process continues until the maximum number of iterations is reached or the optimal error threshold is met. Finally, the network parameters are tested using the testing set to evaluate their performance.

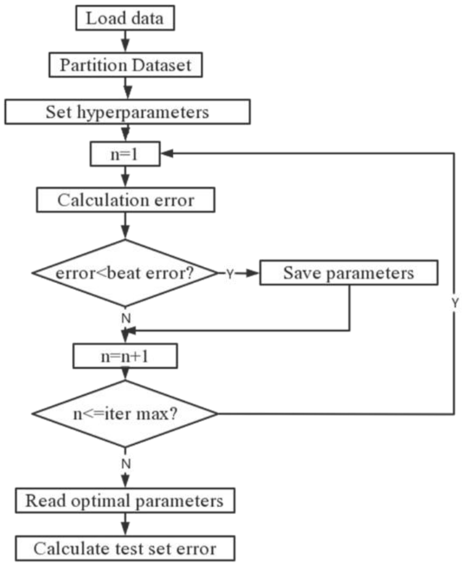

Fig.5 The flowchart of model parameters optimization

During the iteration process, the test set partitioned from the original data is not involved in the training of network parameters. Therefore, the various errors obtained on the test set by the final network can truly reflect the generalization performance of the network.

5. Example Analysis

5.1 Data analysis and processing

The experimental data in this paper is obtained from temperature data and electricity load data collected in a residential area in Jiangsu, China. The time period selected for analysis is from January 2016 to December 2017. The dataset is divided into three sets in an 8:1:1 ratio, corresponding to the training, validation, and testing sets. The training set consists of the original load data, the decomposed sub-sequences obtained from VMD decomposition, and the temperature data. The divided dataset is then normalized using the maximum absolute value normalization method, scaling it to the range of [-1, 1].

$$y_i = \frac{x_i}{\left\| x \right|_{\max} \right|}, \quad i = 1, 2, \ldots, N \tag{6}$$

In the formula, x_i is the i -th element in the dataset, $\left\| x \right|_{\max} \right|$ is the absolute value with the highest absolute value of the element in the dataset, y_i and x_i The corresponding element value obtained after normalization.

This paper proposes a load signal decomposition method based on VMD, which uses the number of sub-sequences automatically obtained from the EMD decomposition of the original sequence as the basis for selecting the value of K. After decomposing the original sequence into 14 sub-sequences using EMD, the value of K is set to 14. Experimental comparisons and verifications are conducted with K selected from 11 to 16, and the comprehensive experimental results show that selecting K as 14 can achieve the best decomposition effect.

5.2 Hyperparameter setting

Our simulation is performing on the Win 10 system, using the PyTorch 15.1+cu92, the graphics card is GTX 1660, the CUP is Intel Core i5 9400, and the RAM is 8G. We set the time window length T is 48, chose L1loss as the loss function, the bath size is set to 128, and using the Adam optimizer with the learning rate is lr=lr*0.7, the intial value of learning rate is 0.01. According to the model structure and optimization method proposed in Section 4 of this article, the preprocessed data was modeled. From Figure 9, it can be concluded that the MAPE errors of both the training and validation sets have steadily decreased, indicating that the proposed model can converge stably.

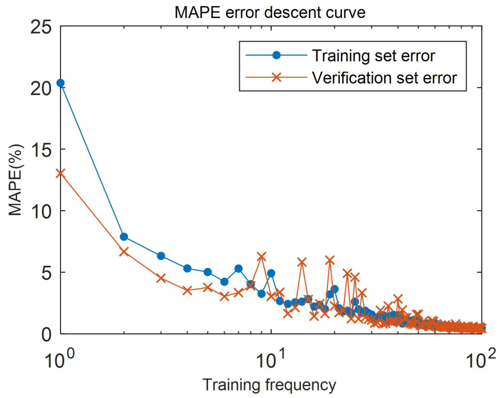

Fig.9 Error drop curve during model training

5.3 Experiment Results

The comparison of various errors for the VMD-BiLSTM model proposed in this paper is shown in Table 3. Since the test set data is an unknown dataset of actual electricity consumption in a specific location in Jiangsu, which has not been involved in the model training, Table 3 shows that the various errors of the test set are close to the minimum error of the model. It can be verified that the VMD-BiLSTM model proposed in this article has high prediction accuracy and good generalization ability.

Table.3 Error comparison of proposed VMD-BiLSTM

Data set	Error		
	MAE（kW）	MSE（kW^2 ）	MAPE（%）
Training set	1.8091	5.9944	0.4265
Validation set	2.1460	9.8606	0.4435
Test set	1.9666	7.2962	0.4491

The comparison of various errors obtained from testing different models using the same dataset is shown in Table 4. Among them, this work proposes that the VMD-BiLSTM combined prediction model has better prediction errors than other constructed models after training with test sets.

Table.4 The performance comparation of different models

Model structure	Error		
	MAE（kW）	MSE（kW^2 ）	MAPE（%）
SVR	48.2979	3939.9516	11.8607
LeNet	24.3795	1085.4309	5.2800
ANN	23.4826	1048.9150	5.0521
LSTM	24.9184	1127.1912	5.3971

DRNet4-7	38.335	2854.7436	8.2857
DRNet3-11	38.2588	2832.9189	8.2326
Improved two-layer BiLSTM	24.5527	1136.8838	5.3322
VMD- two-layer BiLSTM	2.8044	14.5408	0.6483
VMD - Improved two-layer BiLSTM	1.9666	7.2962	0.4491

6. Summary

This work proposes a combined forecasting model based on VMD-BiLSTM for short-term load forecasting in smart grid systems. The model utilizes VMD signal processing method to decompose non-stationary and nonlinear load sequences, and combines it with an improved two-layer BiLSTM network to skip certain transformations in some input data and deeply mine various hidden information in historical sequences in both directions The results showed that it can not only track the trend of load changes, but also ensure high prediction accuracy during periods of significant load changes.

7. Literature References

[1] I. COLAK, R. BAYINDIR and S. SAGIROGLU, "The Effects of the Smart Grid System on the National Grids," 2020 8th International Conference on Smart Grid (icSmartGrid), Paris, France, 2020, pp. 122-126.

[2] R. Ho and K. Hung, "A Comparative Investigation of Mode Mixing in EEG Decomposition Using EMD, EEMD and M-EMD," 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Malaysia, 2020, pp. 203-210.

[3] Q. Cheng, T. Chen and Y. Lei, "Research on the Separation Method of LFM Signal Based on VMD-FastICA," 2021 World Conference on Computing and Communication Technologies (WCCCT), Dalian, China, 2021, pp. 75-79.

[4] S. G. N, G. S. Sheshadri and N. C, "ANN Based Short Term Load Forecasting for Karnataka State Demand," 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), Vijaypur, India, 2022, pp. 1-5.

[5] KO M，LEE K，KIM J K，et al．Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting[J]．IEEE Transactions on Sustainable Energy，2021，12(2)：1321-1335．

[6] P. Miao, G. Chen, K. Cumanan, Y. Yao and J. A. Chambers, "Deep Hybrid Neural Network-Based Channel Equalization in Visible Light Communication," in IEEE Communications Letters, vol. 26, no. 7, pp. 1593-1597, July 2022．

[7] G．Dudek，P．Pełka and S．Smyl. A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(7): 2879-2891．

[8] Yao, Jiachi , et al. "Noise source identification of diesel engine based on variational mode decomposition and robust independent component analysis." Applied Acoustics 116(2017):184-194.

[9] Tang X , Zhong P , Li Z ,et al.Evaluation of biological speckle activity: Using variational mode decomposition[J].Optik - International Journal for Light and Electron Optics, 2021, 243(1):167475.

[10] Xu and Z. Wang, "Fault Detection for Satellite Gyroscope Using LSTM Networks," 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China, 2023, pp. 633-638.

[11] Xu X , Liu C , Zhao Y ,et al.Short-term traffic flow prediction based on whale optimization algorithm optimized BiLSTM_Attention[J].Concurrency and computation: practice and experience, 2022(10):34.

[12] He K , Zhang X , Ren S ,et al.Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition.IEEE, 2016.