Ultra-short-term Wind Power Prediction Model Based on VMD Decomposition and LSTM

Xingdou Liu¹, Li Zhang¹*, Zhirui Zhang¹, Tong Zhao¹ and Liang Zou¹

¹School of Electrical Engineering, Shandong University, Jinan250061, Shandong, China

*zhleee@sdu.edu.cn

Abstract. With the rising demand for environmental protection, wind energy has made considerable progress as a clean energy source. Due to the uncertainty of wind power generation, accurate forecasting of wind power is conducive to achieving stable grid-connected control of wind turbines. This paper proposes an ultra-short-term wind power forecasting model composed of variational modal decomposition (VMD) and long short-term memory (LSTM) networks. Obtain the recorded historical wind speed and historical power from the wind field data acquisition and monitoring control system (SCADA), and use VMD to decompose and reconstruct the historical wind speed of the wind field to form 5 training data sets. Using historical power, historical wind speed, and predicted wind speed as input, wind power forecasts for the next 4 hours (15 minutes apart) are made. In order to evaluate the prediction ability of the proposed model, two algorithms, LSTM and XGBoost, are used for prediction. Taking root mean square error (RMSE) and mean absolute error (MAE) as evaluation indicators, the results show that the prediction accuracy of wind power is significantly improved after VMD processing, and the performance of LSTM is better than XGBoost in the two evaluation indicators.

1. Introduction

In order to ensure the stability of wind turbine grid connected, it is necessary to predict the wind power. At present, wind power prediction is divided into point prediction and probability distribution prediction [1]. Point prediction gives the prediction value of power point at different time intervals [2]. Probability prediction is the range prediction for the uncertainty of wind power prediction [3-4]. Forecasting methods mainly include autoregressive integrated moving average method (ARIMA) [5], autoregressive and moving average method (ARMA) [6], continuous method (PM) [7], Kalman filter method (KF) [8], Gaussian Process (GP) [9], Support Vector Machine (SVM) [10], Evolutionary Algorithm (EA) [11], Neural Network (NN) [12], Wavelet Analysis (WA) [13].

There are some deficiencies in wind speed collection of wind field, which can not represent the wind speed of the whole wind field. Figure 1 is a schematic diagram of the distribution of wind turbines in an actual wind farm. A wind farm can reach a scale of 10×10 km. Due to cost issues, the number of wind measurement towers is small, and there is even only one wind measurement tower. Moreover, the collected wind speed fluctuates greatly (Figure 4), and even extremely high points with extreme fluctuations are recorded, or the record is missing due to the failure of the collection system, which is difficult to represent the wind speed change of the entire wind field.
To solve this problem, reference [14] proposed a secondary decomposition technique composed of wavelet algorithm and singular spectrum analysis, which decomposed the wind speed sequence into several subsequences. In reference [15], adaptive noise complementary integrated empirical mode decomposition (ceemdan) and variational mode decomposition (VMD) are combined to establish a two-stage decomposition scheme. Even if some decomposition techniques are used to mine the features of different subsequences, the randomness of high-frequency wind speed components is very significant, and the high-frequency band is over subdivided, which has little physical significance. For the prediction of the output power of the whole wind field, it is more practical to remove the high-frequency random components of the wind speed series than to over excavate the fluctuation characteristics of the wind speed series.

This paper uses variational mode decomposition (VMD) to decompose it into five intrinsic mode functions (IMF) according to different center frequencies, and then superposes the IMF with different frequencies to reconstruct historical wind speed. According to the different reconstruction components, five training data sets are formed. LSTM and xgboost training models were used to compare the prediction effect, and root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the error. The results show that using the superposition of IMF1 and IMF2 as the historical wind speed data set, the best results are achieved in both algorithm models, and the error indexes of LSTM in different training sets are better than XGBoost.

2. Symbol definition

In order to clearly express the specific work of this paper, the definitions of parameters and variable symbols are given in this section. The defined symbols include the predicted power, the predicted wind speed of the wind farm and the operation data and meteorological data recorded by the wind farm SCADA system, the naming method is shown in in Table 1.

After VMD processes the recorded $W_{\text{history}}$ of wind field, IMF with different frequencies will be superimposed as a new $W_{\text{history}}$ to form five data sets. The naming method is shown in Table 2, and $W_1$ is the $W_{\text{history}}$.

| Table 1. Symbol definition | Table 2. Divide the data set |
|----------------------------|------------------------------|
| Characteristic quantity    | Symbol                       | Component superposition | $W_{\text{history}}$ | Dataset   |
| Predicted power            | $P_{\text{predict}}$         | IMF (1+ 2+ 3+ 4+ 5)     | $W_1$                  | Dataset1  |
| Predicted wind speed       | $W_{\text{predict}}$         | IMF (1+ 2+ 3+ 4)        | $W_2$                  | Dataset2  |
| Historical wind speed      | $W_{\text{history}}$         | IMF (1+ 2+ 3)           | $W_3$                  | Dataset3  |
| Historical power           | $P_{\text{history}}$         | IMF (1+ 2)              | $W_4$                  | Dataset4  |
| Future power               | $P_{\text{future}}$          | IMF (1)                 | $W_5$                  | Dataset5  |
3. Method introduction

3.1. Variational mode decomposition

VMD is an adaptive, completely non recursive method of modal variation and signal processing [16]. VMD defines the intrinsic mode function (IMF) as an FM-AM signal. The steps to construct the variational problem are as follows:

\[
\begin{align*}
\min_{\{u_k\}, \{w_k\}} & \left\{ \sum_k \left[ \int \left| \frac{\partial}{\partial t} \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right| e^{-j w_k t} \right| dt \right\}^2 \right. \\
\text{s.t.} & \sum_k u_k(t) = f(t)
\end{align*}
\]

(1)

Where \( u_k(t) \) is the mode function of the input signal; \( \{u_k\} \) represents the mode set; \( w_k \) is the center frequency of the k-th mode corresponding to the input signal; \( \{w_k\} \) represents a set of center frequencies corresponding to the decomposed mode; \( \delta(t) \) is the unit impulse function. Introducing the Lagrangian multiplier \( \lambda \) and the quadratic penalty factor \( \alpha \), the formula (2) can be rewritten as:

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

(2)

Using the alternating direction method of the multiplication algorithm to solve (3), a set of modal components and their respective center frequencies are obtained. Each mode can be estimated from the solution in the frequency domain, expressed as:

\[
\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i=1}^{n+1} \hat{u}_i^{n+1}(w) - \sum_{i=1}^{n} \hat{u}_i^{n}(w) + \hat{\lambda}^n(w)}{1 + 2\alpha(w - w_k^2)}
\]

(3)

Where \( n \) is the number of iterations; \( \hat{f}(w) \), \( \hat{u}_i^{n+1}(w) \), \( \hat{u}_i^{n}(w) \) and \( \hat{\lambda}^n(w) \) is the form after Fourier transform.

In formula (4), it has the characteristics of Wiener filter structure, which directly updates the mode in the Fourier domain. In addition, these modes can be obtained in the time domain by extracting the real part of the inverse Fourier transform of the filter analysis signal.

\[
W_{k+1}^{n+1} = \frac{\int_0^\infty w \left| \hat{u}_k^{n+1}(w) \right|^2 dw}{\int_0^\infty \left| \hat{u}_k^{n+1}(w) \right|^2 dw}
\]

(4)

3.2. Long short-term memory network

LSTM network is a special recurrent neural network (RNN) [17], which effectively handles practical training problems such as information disappearance and gradient internal explosion brought by RNN network. Therefore, LSTM provides more accurate training effects for complex time series data. The unit structure of the LSTM network is shown in Figure 2.
The LSTM unit can be defined by the following set of equations:

The input gate:
\[
i_t = g(w_{ix}x_t + w_{ih}h_{t-1} + b_i)
\]  \hspace{1cm} (5)

The forget gate:
\[
f_t = g(w_{fx}x_t + w_{fh}h_{t-1} + b_f)
\]  \hspace{1cm} (6)

The output gate:
\[
o_t = g(w_{ox}x_t + w_{oh}h_{t-1} + b_o)
\]  \hspace{1cm} (7)

Input conversion:
\[
c_{in} = \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_{c_{in}})
\]  \hspace{1cm} (8)

Status update:
\[
c_t = f_t h_{t-1} + i_t c_{in}
\]  \hspace{1cm} (9)
\[
s_t = o_t \tanh(c_t)
\]  \hspace{1cm} (10)

Where \(x_t\) is the input sequence at time \(t\); \(s_t\) is the output of the hidden layer; \(c_{in}\) is the initial information transmitted to the cell layer; \(f_t, i_t\) and \(o_t\) represent the output of forgetting gate, input gate and output gate respectively; \(g\) and \(\tanh\) are the activation functions; \(w\) and \(b\) are the weights and deviations of neurons in each cell respectively.

4. Case study

The case of this paper is from a wind farm in Jining City, Shandong Province, China. LSTM and XGBoost are used to verify the proposed method. The installed capacity of the wind farm is 93MW, and the time length of data collection is from January 1, 2019 to September 28, 2020. These data are used to form a data set for training and testing of the model, and to predict the wind power in the next four hours (with an interval of 15 minutes).

4.1. Evaluative index

In this paper, absolute value error (MAE) and root mean square error (RMSE) are selected as error evaluation indexes, which are \(P_{\text{future}}\) and \(P_{\text{predict}}\).

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| 
\]  \hspace{1cm} (11)
\[
\text{RMSE} = \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right)^{1/2} 
\]  \hspace{1cm} (12)

4.2. VMD processing \(W_{\text{history}}\)

VMD is used to decompose the whitory. The five IMF cases after decomposition are shown in Figure 3. It can be seen that there are some isolated extreme points in the original whitory. The reconstructed \(W_{\text{history}}\) sample is shown in Figure 4. It can be seen that after removing a certain high frequency component, the abnormal extremely high wind speed points are removed, and the missing data of wind speed record caused by system fault is also corrected to a certain extent.
4.3. Model training and prediction effect

The LSTM and XGBoost algorithms are used to predict the wind power in the next 4 hours with the interval of 15 minutes, using the five data sets (1-5) training model formed by the $W_{\text{history}}$ processed by VMD.

MAE and RMSE are taken as error evaluation indexes. The results of each algorithm in different data sets are shown in table 3, and the error change curve is drawn as shown in Figure 5.

Table 3. Divide the data set

|         | LSTM  |         | XGBoost |
|---------|-------|---------|---------|
| Dataset1| MAE   | RMSE    | MAE     | RMSE    |
|         | 7.299 | 9.603   | 7.618   | 9.910   |
| Dataset2| 7.221 | 9.528   | 7.587   | 9.878   |
| Dataset3| 7.018 | 9.279   | 7.578   | 9.886   |
| Dataset4| 5.914 | 7.893   | 6.766   | 8.952   |
| Dataset5| 6.555 | 8.592   | 6.975   | 9.065   |
5. Conclusion

(1) The recorded wind speed in the first four hours can be used to predict the power in the next four hours. The recorded wind speed history is decomposed into five IMFs by VMD, and the high frequency IMF is removed in turn for reconstruction. The extremely high frequency wind speed component can be removed gradually, and the missing wind speed points can be corrected. After processing, the wind speed curve is smoother and more representative of the wind energy change of the whole wind field, which improves the generalization ability and power prediction accuracy of the model.

(2) For time continuous power prediction task, LSTM has unique memory and forgetting structure design, which can deeply learn and selectively extract the continuous variation of wind speed and power, and has better prediction effect than XGBoost.

6. References

[1] Du P, Wang J Z, Yang W D and Niu T 2019 Appl. Soft. Comput. 80 93-06
[2] Ernst B, Oakleaf B, Ahlstrom M L, Lange M, Moehren C, Lange B, Focken U and Rohig K 2007 IEEE Power and Energy Magazine 5 78-89
[3] Yang M, Lin Y, Zhu S, Han X and Wang H 2015 J. Modern Power Syst. Clean Energy 3 361-70
[4] Wan C, Xu Z, Pinson P, Dong Z Y and Wong K P 2014 IEEE Trans. Power Syst. 29 1033-44
[5] Chen P, Pedersen T, Bak-Jensen B and Chen Z. 2010 IEEE Trans. Power Syst. 25 667-76
[6] Rajagopalan S and Santoso S 2009 Canada IEEE Power Energy Soc. Gen. Meeting 1-6
[7] Bludszuweit H, Dominguez-Navarro J and Llombart A 2008 IEEE Trans. Power Syst. 23 983-91
[8] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I and Kallos G 2008 J. Wind Eng. Ind. Aerodyn. 96 2348-62
[9] Chen N, Qian Z, Nabney I and Meng X 2014 IEEE Trans.Power Syst. 29 656-65
[10] Zeng J and Qiao W 2012 IEEE Trans. Sustain. Energy 3 255-64
[11] Elattar E 2011 Proc. 2nd IEEE PES Int.Conf. Exhib. Innovative Smart Grid Technol. (ISGT Europe) 1–7
[12] Quan H, Srinivasan D and Khosravi A 2014 IEEE Trans. Neural Netw. Learn. Syst. 25 303-15
[13] Shi J, Liu Y, Yang Y and Lee WJ 2011 IEEE Ind. Commer. Power Syst. Tech. Conf. (ICPS) 1-7
[14] Yu C, Li Y and Zhang M 2017 Energy Convers. Manage. 148 895-04
[15] Peng T, Zhou J, Zhang C and Zheng Y 2017 Energy Convers. Manage. 153 589-02
[16] Dragomiretskiy K and Zosso D 2014 IEEE Trans.Signal Process 62 531-44
[17] Hochreiter S and Schmidhuber J 1997 Neural Computation 9 1735-80