Research on Wind Power Ultra-short-term Forecasting Method Based on PCA-LSTM

Siying Wu
School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055, China;
Corresponding author’s e-mail: wusiying028@163.com

Abstract. Wind power ultra-short-term forecasting can provide the support for adjusting the intraday power generation plan, carrying out the incremental spot trading of wind power, and improving the utilization of wind power. In order to improve the forecast accuracy of wind power, a wind power ultra-short-term power forecast method based on long-term-term memory (LSTM) network is proposed. First, the principal component analysis method is used to reduce the multivariate meteorological time series dimension. Then by using the cyclic memory characteristics of LSTM network to model multi-dimensional time series, the nonlinear mapping relationship between meteorological data and power data is established, and the wind power forecast is finally realized. The actual data of the eastern China wind farm is used to verify the results. It shows the method established in this paper can effectively use the meteorological and power data to forecast the wind power, and compared with the traditional time series, BP neural network method, the method in this paper has higher forecast accuracy and has broad application potential.

1. Introduction
Wind power has maintained a sustained and rapid development trend in recent years. Accurate wind power forecast technology can allow dispatch centre to incorporate wind power into the startup plan, reducing the thermal power startup capacity under the premise of ensuring system security, and maximize the wind power accommodation level. Wind power ultra-short-term forecast is mainly to forecast the theoretical wind power in the next 15 minutes to 4 hours. The current wind power ultra-short-term forecast methods mainly include physical methods and statistical methods [1]. The physical method [2] mainly includes three key processes: downscaling model, wind power conversion model, and model output statistics. The modelling process is complicated and the calculation cost and difficulty are relatively high, and it is often used for power forecast of new wind farms. The statistical method mainly analyses the statistical distribution characteristics of historical wind power data, a non-linear relationship between actual wind power and historical wind power, meteorological forecast data is established for forecast, the main methods include kalman filter model, ARMA model, and artificial neural networks model [3-6], but because no future meteorological change information is introduced, the forecast results are mainly the copies of historical neighbour fluctuations, and the forecast results have a significant time delay characteristics. Wind power forecasting based on deep learning has also become a research hotspot. LSTM neural networks for wind power forecasting is used in Ref[7-8], but only historical data is used, and meteorological forecast data is not considered. In order to improve the shortcomings of the above methods, in this article, more meteorological factors which affect wind power output are considered in the input model, and PCA method is used to do the multi-dimensional meteorological factors dimension.
reduction[9]. A new PCA-LSTM wind power ultra-short-term forecast model is built, and the practicality and advancement of the method are verified based on actual cases.

2. Wind power affecting factors analysis and feature selection

2.1. Wind power affecting factors selection

Wind power is related to wind turbine parameters and wind speed characteristics. The wind turbine output can be calculated as equation (1):

\[ P_w = \frac{C_p S \rho V^3}{2} \]  

Where \( P_w \) is power of wind wheel, \( C_p \) is wind turbine power coefficient, \( S \) is the area of wind wheel, \( \rho \) is air density, \( V \) is wind speed.

The air density \( \rho \) is a function of air pressure, temperature, humidity. The empirical calculation formula is shown as equation (2).

\[ \rho = 3.48 \frac{P}{T} \left[ 1 - 0.378 \frac{\phi P_b}{P} \right] \]  

Where \( T \) is thermodynamic temperature, \( P \) is normal pressure, \( P_b \) is horizontal pressure, \( \phi \) is relative air humidity.

It can be known from equation (1) and equation (2) that the power of a wind turbine is affected by several factors such as air pressure, temperature, wind speed, wind direction and humidity. Wind speed, wind direction at different altitudes, humidity, temperature and air pressure need to be obtained from numerical meteorological forecast need to be used as necessary inputs for forecast models. Some of the affecting factors have correlation between each other, so fewer factors can be used to comprehensively express all the information in each influencing factor. The PCA method can be used to do feature dimension reduction and determine the minimum input features number. While reducing the number of input variables and try to minimize the reducing dimension errors.

2.2. Principal component analysis

The main idea of the principal component analysis method [10] is to reduce the dimension through the spatial feature transformation of the vector. The original \( n \)-dimensional variable is spatially transformed and projected onto the new \( m \)-dimension orthogonal feature \( (m < n) \). The newly formed \( m \)-dimensional variable reflects most information of the history variable, which is a linear shown of the original \( n \)-dimensional variable. It is called the principal component variable. The original data information contained in it is reflected on the variance of each variable.

There are data sample sets \( X = \{x_1, x_2, \ldots, x_m\} \). It is assumed that the sample set has been centralized, namely, \( \sum x_i = 0 \), the new coordinate system after the projection is set \( \{w_1, w_2, \ldots, w_d\} \), in which \( w_i \) is the standard orthogonal basis vector, \( \|w_i\|_2 = 1 \). The projection of a certain data sample point \( x \) on the newly created spatial plane is \( W^T x \), to ensure that the projections of all sample points are as independent as possible and there is no interaction of information, it is necessary to maximize the variance of all data sample points after projection. The variance of data sample points after projection can be shown by \( \sum W^T x_i x_i^\top W \), therefore, the optimization goal of the projection transformation should be:

\[ \max_w \text{tr} \left( W^T X X^T W \right) \quad \text{s.t.} \quad W^T W = 1 \]  

(3)
Transform (3) using Lagrange multiplier method, we can obtain:

\[ X^T W = \lambda W \]  
(4)

By decomposing and sorting the eigenvalues of the covariance matrix \( X^T \) of the data samples, we can obtain \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \); Finally, when the cumulative contribution rate of variance exceeds 90\%, the first \( k \) principal components corresponding to the variance are shown in equations (5) and (6).

\[ \eta_i = \frac{\lambda_i}{\sum_{m} \lambda_m} \times 100\% \]  
(5)

\[ \eta_k(p) = \sum_{i=1}^{k} \eta_i \]  
(6)

After the first \( k \) eigenvalues are selected, the \( k \) eigenvectors \( W = (w_1, w_2, \ldots, w_k) \) that correspond one to one are the new vectors obtained after the feature dimension reduction.

3. Wind power ultra-short-term forecast model

3.1. The structure of LSTM network

LSTM is a RNN network with a special structure. It has special hidden units that can save the characteristics of the input for a long time, can make full use of the characteristics of long-term historical information, Fig.1 shows the typical LSTM network.

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

Compared with the traditional recurrent neural network, in addition to the hidden state \( h_t \), the cell state \( C_t \) is also added in the forward recurrent neural network of the LSTM network. Each network unit of LSTM is composed of an input gate (circle dotted line) and a forget gate (square dotted line), the output gate (dot dotted line part) and the cell state (solid line). The forward calculation formula in each part is as follows, \( x_t \) is the input sequence, \( W_f, W_i, W_a, W_o, U_f, U_i, U_a, U_o \) are the weight matrices connecting the hidden state \( h_t \) and the input sequence \( x_t \), and \( b_f, b_i, b_a, b_o \) are bias vectors.

1) The hidden state \( h_{t-1} \) of the previous sequence of the forget gate and the input sequence \( x_t \) are converted into the output of the forget gate, as shown in equation (7):

\[ f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \]  
(7)

2) The input gate consists of two parts, the previous hidden state \( h_{t-1} \) and this input \( x_t \) are multiplied using tanh and sigmoid functions respectively, as shown in equations (8) and (9):

\[ i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \]  
(8)

\[ a_t = \tanh(W_a h_{t-1} + U_a x_t + b_a) \]  
(9)
3) Update the cell state, calculate the product of the previous cell state $C_{t-1}$ and the forget gate output $f_t$, and the multiply and sum of the output gate $i_t$ and $a_t$ to calculate the new cell state to the next neuron:

$$C_t = C_{t-1} \times f_t + i_t \times a_t$$  \hspace{1cm} (10)

4) Update the hidden state $h_t$ using the output gate. In the first step, the hidden state $h_{t-1}$ of the previous sequence and the input sequence $x_t$ are activated by the sigmoid function. In the second step, the hidden state $h_{t-1}$ is multiplied by the above output $o_t$ to obtain a new hidden state $o_t$ by tanh activation function:

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o)$$  \hspace{1cm} (11)

$$h_t = o_t \times \tanh(C_t)$$  \hspace{1cm} (12)

LSTM is similar to the RNN in structure. It is a recurrent network that can persistent keep and transfer information. It can perform sequential information transmission through multiple clones of the same neuron. Fig.2 shows the timing loop of the LSTM network model. Each dashed box contains a complete LSTM structure.

![Figure 2. Expansion of LSTM network model in time](image)

3.2. Data normalization
In the wind power forecast, to eliminate the each input variable value difference, a normalized method is used to process the input data. Project various types of data between the determined maximum and minimum values and normalize them from zero to one in (13).

$$x_{\text{norm}} = \frac{x_t - x_{\text{min}}}{x_{\text{maximum}} - x_{\text{min}}}$$  \hspace{1cm} (13)

Where $x_t$ is actual output, $x_{\text{max}}$ is the maximum output, $x_{\text{min}}$ is the minimum output, $x_{\text{norm}}$ is the normalization output.

For wind direction data in the numerical weather forecast, since its value range is 0-360 degrees, in order to maintain the smoothness of the data, this paper uses the sin and cos values of the wind direction angle for normalization processing, as shown in equation (14).

$$\begin{align*}
\alpha_{\sin}^* &= \sin(\alpha) \\
\alpha_{\cos}^* &= \cos(\alpha)
\end{align*}$$  \hspace{1cm} (14)

For power data, use the wind power $P_{\text{out}}$ divide its installed capacity $S_{\text{op}}$ for normalization, as shown in equation (15):

$$P_{\text{out}}^* = \frac{P_{\text{out}}}{S_{\text{op}}}$$  \hspace{1cm} (15)
3.3. Model parameter selection
The designed LSTM model includes input layer, network layer, and output layer. Sigmoid function is used as an activation function, the loss function is mean square error, and the dropout technology is used to prevent over-fitting. The input vector includes \( n \)-dimensional meteorological data, for example, wind speed of different layers, air pressure, wind direction of different layers, temperature, and humidity. The \( k \)-dimensional meteorological data and wind power are obtained after the principal component analysis. The time step is 16, that is, the data of the past 16 time steps is used as input vector, including the historical power data of the past 16 points and the \( k \)-dimensional meteorological data corresponding to the forecasted time point \( t + 1 \) to \( t + 16 \).

The network layer includes two hidden layers. There are separately 30 and 60 neurons in first and second layer, the output layer nodes is set 16, it means the output is the forecasted wind power from time \( t + 1 \) to \( t + 16 \). In training process, some parameters are set as follows: the maximum training number is 5000, the study rate is 0.01, the dropout rate is 0.5 and the mini-batch size is 60. The time period to be forecasted is 15 minutes to 4 hours in the future, and the time resolution is 15 minutes.

3.4. Forecast evaluation index
RMSE, MAE and forecast pass rate are used to evaluate the forecast result, the equations are from (17) to (18). According to the requirements of the industry standard " Wind Power Forecasting System Functional Specification " (NB/T 31046-2013), the RMSE of the 4h of the ultra-short-term forecast cannot exceed 15%. Therefore, if the error is less than 15% of the installed capacity, the result is qualified.

\[
E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{P_{\text{mi}} - P_{\text{pi}}}{S_{\text{op}}} \right)^2}, \quad i = 1, \ldots, n \tag{16}
\]

\[
E_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|P_{\text{mi}} - P_{\text{pi}}|}{S_{\text{op}}} \right), \quad i = 1, \ldots, n \tag{17}
\]

\[
Q_{S} = \frac{1}{n} \sum_{i=1}^{n} B_{i}, \quad i = 1, \ldots, n
\]

\[
\begin{align*}
B_{i} = 1 & \quad \text{if } \frac{|P_{\text{mi}} - P_{\text{pi}}|}{S_{\text{op}}} \geq 0.85 \\
B_{i} = 0 & \quad \text{if } \frac{|P_{\text{mi}} - P_{\text{pi}}|}{S_{\text{op}}} < 0.85
\end{align*}
\tag{18}
\]

Where \( P_{\text{mi}} \) is the actual wind power, \( P_{\text{pi}} \) is the forecast wind power, \( S_{\text{op}} \) is the installed capacity.

4. Case study
4.1. Data Source and Pre-processing
The data used in this article is the actual numerical weather forecast data and power data of a wind farm in the eastern province of China from January 2018 to December 2018. The installed capacity is 100MW and the time resolution is 15min. There are a total of 35040 sets of data. 32064 sets of data of the first 11 months are selected as the training data, and 2976 sets of data for the last month are used as the test data. The numerical weather forecast data after normalization includes 10m high wind speed, 70m high wind speed, 10m high wind direction sin value, 10m high wind direction cos value, 70m high wind direction sin value, 70m high wind direction cos value, temperature, humidity, sea level pressure and surface pressure, total 10 dimensional data. The above variables are represented by \( X_{1} \sim X_{10} \). In order to reduce input variables dimension and minimize the loss of information contained in the variables, the PCA method is used to perform the dimension reduction process to obtain the meteorological factors.
for wind power ultra-short term power forecast. The calculated results of each principal component are shown in table 1.

| Component variable name | Eigenvalue | Variance contribution rate | Cumulative contribution rate |
|-------------------------|------------|----------------------------|------------------------------|
| Y1                      | 16.3365    | 0.6446                     | 0.6446                       |
| Y2                      | 7.9128     | 0.2597                     | 0.9043                       |
| Y3                      | 1.1435     | 0.0303                     | 0.9346                       |
| Y4                      | 0.6043     | 0.0223                     | 0.9569                       |
| Y5                      | 0.5382     | 0.0146                     | 0.9715                       |
| Y6                      | 0.3692     | 0.0102                     | 0.9817                       |
| Y7                      | 0.3161     | 0.0085                     | 0.9902                       |
| Y8                      | 0.2631     | 0.0069                     | 0.9971                       |
| Y9                      | 0.1054     | 0.0022                     | 0.9993                       |
| Y10                     | 0.0721     | 0.0007                     | 1                            |

From table 1, we can see that the cumulative variance contribution rate of the first two principal components Y1 and Y2 is 90.43%, exceeding 90%, it means first two principal components Y1 and Y2 can be considered to include nearly all original meteorological data information, so the principal components are selected as Y1 and Y2 to replace the original variables. It reduces the dimension and prevents the loss of information to the greatest extent. It can improve the computing efficiency while ensuring the accuracy of the data.

4.2. Result analysis

After obtaining the two input meteorological principal component variables, use them and wind power data as the input variables. Forecast model is built in the keras deep learning framework. Fig. 3 shows the iteration loss, with the training iterations number increase, the loss function quickly converges.

![Figure 3. Loss function variation](image_url)

At the same time, traditional BP neural network, classical time series ARMA method and LSTM prediction method without PCA method are used to forecast the wind power. The RMSE and the MAE are compared to evaluate the advantages and disadvantages of several algorithms and verify the feasibleness and accuracy of the model proposed in this article. After simulation training and verification, a 16-point multi-step forecast result is obtained, the RMSE and MAE calculation result of each step is shown in Fig.4 and Fig.5. At the same time, the statistics of the various error evaluation indicators of the 16th point (4th hour) forecast are emphasized, which is shown in table 2, the forecast result in the 16th point is compared in Fig.6, which can focus on the performance of ultra-short-term wind power forecast.
It can be seen from the error comparison curve that the longer the forecast time scale, the larger the forecast error. When time scale is small, the forecast effect obtained by the ARMA model is comparable to or better than that of the traditional neural network and deep neural network algorithms, especially within 30min. This is because since the forecast step increasing, the dependence between the forecast wind power and the historical wind power is weakened. With the increase of the time scale, the advantages of the PCA-LSTM method gradually emerged. The growth of RMSE and MAE was significantly lower than other forecast methods. The 16 points average value of RMSE is 9.46%, and compared with a single LSTM model based on multi-dimensional meteorological data input without dimensionality reduction, it has improved a lot. This shows that the role of the PCA method in wind power ultra-short-term forecast is very important, it can help improve forecast accuracy. In addition, the PCA-LSTM model has fewer input variables. When large-scale wind farm cluster power need to be forecasted, the calculation efficiency can be improved.

From table 2, it can be seen that the method proposed in the article has achieved good results on the error parameters at the 16 point, which proves that PCA-LSTM method proposed in the article can reduce the forecast error. It should be noted that the model and model parameters in this paper are based
on specific wind farms. When making forecast for other wind farms, the hyper-parameters of the model need to be adjusted.

5. Conclusion

In this paper, a wind power ultra-short-term forecast method based on PCA-LSTM model is proposed. On the basis of traditional LSTM forecast model, large number of meteorological factors such as wind speed, wind direction, temperature, humidity and pressure are fully considered. The PCA method is used to reduce the input data dimension and choose the input variables. The simulation results show that with the time step increases, the forecast increasing effect is more and more obvious, which verifies its practicability and superiority in wind power forecast. Next step, the application of other deep learning methods such as convolutional neural networks in wind power forecast need to be further researched to reduce the forecast error.

References

[1] Xue Y, Yu C, Zhao J, et al. (2015) A review on short-term and ultra-short-term wind power prediction. Automation of Electric Power Systems, 39:141-151.
[2] Feng S, Wang W, Liu C, et al. (2010) Study on the physical approach to wind power prediction. Proceedings of the CSEE, 30:1-6.
[3] Yang M, Sun Y, Wang Dong, et al. (2014) Multi-Step Prediction Study on the Multi-Sampling Scale Wind Power Based on the Time Series Method. Electrical Measurement & Instrumentation, 51: 55-60.
[4] Peng X, Xiong L, Wen J, et al. (2016) A summary of the state of the art for short-term and ultra-short-term wind power prediction of regions. Proceedings of the CSEE, 36:6315-6326.
[5] Yang J, Huo Z, He Y, et al. (2018) Ultra-short-term wind power prediction based on wavelet and minimum resource allocation network. Power System Protection and Control, 46: 55-61.
[6] Wang H, Hu Z, Chen Z, et al. (2013) A hybrid model for wind power forecasting based on ensemble empirical mode decomposition networks. Transactions of China Electro technical Society, 28:137-144.
[7] Barbounis T. G, Theocharis J. B, Alexiadis M. C, Dokopoulos P. S. (2006) Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21:273-284.
[8] Zhu Q, Li H, Wang Z, et al. (2017) Short-Term Wind Power Forecasting Based on LSTM. Power System Technology, 41:3797-3802.
[9] He D, Liu R. (2013) Ultra-short-term wind power prediction using ANN ensemble based on the principal components analysis. Power System Protection and Control, 41:50-54.
[10] Filippo M.B, Enrico D.S, Antonello R, Alireza S. (2015) Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. IEEE Access, 3:1931-1943.