Short-Term Prediction of Wind Power Based on Deep Long Short-Term Memory

Qu Xiaoyun, Kang Xiaoning, Zhang Chao, Jiang Shuai, Ma Xiuda
Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University
Xi'an 710049, China
Email: xiaoyun_Qu@126.com

***Abstract*—This paper proposes a wind power prediction model based on the Long Short-Term Memory model, one of the deep learning method. Deep learning conforms to the trend of big data and has powerful capability of learning and generalization for mass data. Principal component analysis (PCA) is used to choose input samples and reduce the dimensions of the input variables of the LSTM prediction model based on numerical weather prediction(NWP) data. Simulation results show that, compared with BP neural network and support vector machine(SVM) model, the LSTM prediction model has higher prediction accuracy and greater potential for engineering applications. It is effective to apply the Long Short-Term Memory model to the field of wind power prediction.**

***Index Terms*--deep learning; Long Short-Term Memory; numerical weather prediction; principal component analysis; wind power prediction**

I. INTRODUCTION

As the proportion of wind power in the power network continues to increase, it will bring a challenge to the safety and stabilization of the power network and then restrict the scale of wind power development [1]. The accurate prediction of the wind power can effectively reduce or avoid the adverse effect of wind farm on power network. And it is very important for the sustainable development of the wind power.

There are physical methods, statistical methods, learning methods and combinations of these three methods for wind power prediction. Different methods are used for different time scales and different data sources [2]. Statistical methods and learning methods are used based on history data for short-term prediction of wind power within six hours. Any of these three methods can be used based on numerical weather prediction(NWP) data for the prediction within forty-eight hours. Physical methods aim to describe the physical process of converting wind to power and models all of the steps involved. Statistical methods aim at describing the connection between predicted wind and power output directly by statistical analysis of time series from data in the past [3]. Common statistical methods are time series method [4], regression analysis method and kalman filter method [5] etc. Learning methods use artificial intelligence algorithms that are able to implicitly describe nonlinear and highly complex relations between input data and output data. In [6], the wind speed and power were forecasted by neural network based on time series, but the effect was relatively poor for long time scales. As to [7], an artificial neural network (ANN) model for wind power prediction was constructed based on NWP data to forecast the error band.

Deep learning is the development of artificial neural network. Especially when AlphaGo defeated the human, the field of artificial intelligence has set off a wave of deep learning. Deep learning conforms to the trend of big data and has the strong learning and generalization ability for massive data [8]. This paper studies the wind power prediction model based on the Long-Short Term Memory, one of deep learning methods. The effective information which reflects the characteristics of the wind farm is extracted by PCA. The principal components are chosen as inputs of the LSTM prediction model. Training and learning the LSTM model based on a large number of numerical weather prediction data.

II. SELECTION OF FORECAST FACTOR

The output power of a wind turbine is given by:

$$P_w = C_P A \rho v^3 / 2 \tag{1}$$

where P_w is the output power (kW), C_p is the rotor coefficient of performance, ρ is the air mass density (kg/m^3), A is the swept blade area (m^2), v is the speed of the wind (m/s). Air density is closely related to temperature, humidity and pressure:

$$\rho = 3.48 \frac{P}{T} (1 - 0.378 \frac{\varphi P_b}{P}) \tag{2}$$

where P is normal atmosphere, T is thermodynamic temperature, P_b is saturation vapor pressure, φ is relative air humidity.

As shown in (1) and (2), wind turbine output power is related to wind speed, temperature, humidity, pressure. Among them, the wind speed is the most important factor. In addition, considering the wake effect, the wind direction has a great influence on the wind power. Among these factors, some are related but some are not related. Few predictors will result

978-1-5090-5417-6/16/ $ 31.00 ©2016 IEEE

in lack of information, which can't fully reflect the variation of wind power. Too many predictors will lead to redundant information and the decrease of generalization performance. In order to solve this problem, important predictors are selected by PCA.

The main principle of PCA is finding an appropriate linear transformation to transform variables related to each other into new variables that are independent of each other. Each new variable has its own unique meaning, among them, the variables with larger variance can reflect the main information contained in the original multiple variables.

The NWP data include air density, pressure, temperature, the wind speed and wind direction in 100m (denoted by Y_1 to Y_5). Analyzing the NWP data by PCA. The values of eigenvalues and contribution rates of principal components are shown in TABLE I and Fig. 1.

TABLE I. EIGENVALUES AND CONTRIBUTION RATE

Principal Component	**Eigenvalue**	**Contribution Rate （%）**	**Cumulative contribution rate （%）**
Z_1	0.2531	92.8984	92.8984
Z_2	0.0189	6.9387	99.8371
Z_3	0.0004	0.1402	99.9773
Z_4	4.87e-05	0.0179	99.9952
Z_5	1.34e-05	0.0049	100

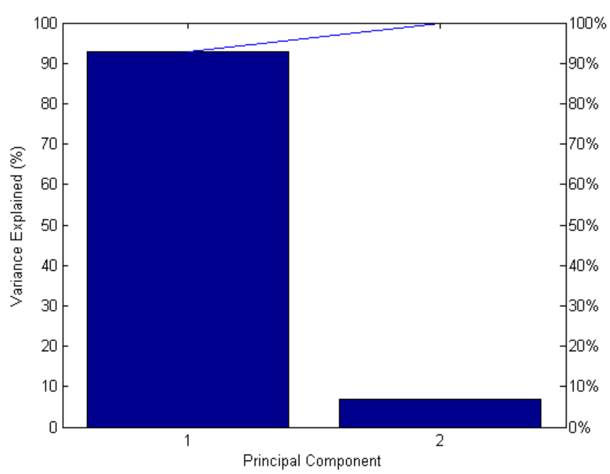

Figure 1. The contribution rate of principal component

From TABLE I and Fig. 1, the cumulative contribution rate is up to 99.8% when the first two principal components are selected. And other principal components can be discarded. There are five meteorological inputs at each time before processing, now there are only two principal components as inputs. Besides, the two principal components can reflect the main information contained in the original multiple variables.

TABLE II shows the characteristic vector of eigenvalue of covariance matrix. The first principal component Z_1 is mainly related to the wind speed(Y_4), and the second principal component Z_2 is mainly related to the wind direction(Y_5).

TABLE II. THE CHARACTERISTIC VECTOR OF EIGENVALUE OF COVARIANCE MATRIX

NWP data	Z_1	Z_2	Z_3	Z_4	Z_5
Y_1	-0.0007	0.0220	0.7318	0.5302	-0.4276
Y_2	0.0011	-0.0083	0.0966	0.5409	0.8355
Y_3	-0.0031	-0.0115	-0.6743	0.6529	-0.3448
Y_4	0.9998	-0.0155	-0.0013	0.0018	-0.0024
Y_5	0.0155	0.9995	-0.0231	0.0003	0.0123

III. WIND POWER PREDICTION MODEL

Long Short-Term Memory (LSTM) is a deep learning method proposed in 1997 by Sepp Hochreiter and Jürgen Schmidhuber [9]. Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations. Compared with the traditional shallow model, it has a number of layers of nonlinear transformation, which create the conditions for more complex task modeling. With sufficient training samples, deep learning model can achieve full potential and dig out the abundant information contained in the massive data.

A. *Long Short-Term Memory*

LSTM can be used as a complex nonlinear unit to construct a larger deep neural network, which can reflect the effect of long-term memory and has the ability of deep learning. LSTM network consists of an input layer, an output layer, and a plurality of hidden layers. The hidden layer is composed of the memory cell, the basic structure is shown in Fig. 2. One cell consists of three gates (input, forget, output), and a recurrent connection unit. The input to this unit is x_t, the current input at step t, and s_t, the current hidden state. The current output is o_t, c_t is the internal memory of the unit.

Gates use a sigmoid activation (denoted by g), while input and cell state are often transformed with tanh. LSTM cell can be defined with a following set of equations:

Input gate:

$$i_t = g(W_{xi}x_t + W_{hi}s_{t-1} + b_i) \tag{3}$$

Forget gate:

$$f_t = g(W_{xf}x_t + W_{hf}s_{t-1} + b_f) \tag{4}$$

Output gate:

$$o_t = g(W_{xo}x_t + W_{ho}s_{t-1} + b_o) \tag{5}$$

Input transform:

$$c\_in_t = \tanh(W_{xc}x_t + W_{hc}s_{t-1} + b_{c\_in}) \tag{6}$$

State update:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c\_in_t \tag{7}$$

$$s_t = o_t \cdot \tanh(c_t) \tag{8}$$

where W_{ij} is the connection weights of neuron i to j, b is deflection.

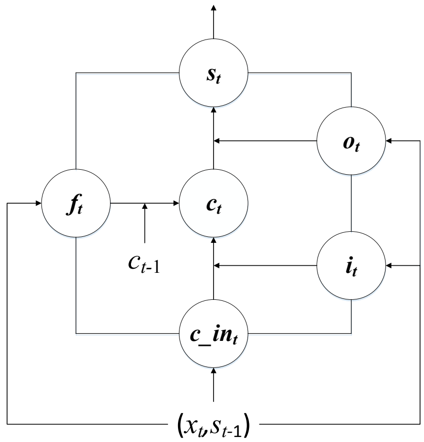

Figure 2. The basic structure of LSTM

The number of hidden layers and the number of neurons in each layer play a key role in the neural network training and impact the prediction accuracy. The more the number of hidden layers and neurons in each layer, the more complex the model. If there are few neurons in each layer, the network may not be trained or the performance is poor. If there are many neurons, although the network system error can be reduced, but on the one hand, the network training time is prolonged and the training may also appear over-fitting phenomenon.

LSTM is a recurrent neural network (RNN) architecture. Fig. 3 shows the network structure unfolded in time. *U, V, W* are the network parameters.

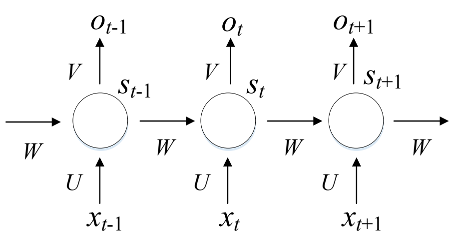

Figure 3. Network structure unfolded in time

B. *LSTM prediction model*

The historical wind power data and NWP data are selected as samples of the prediction model. According to the network structure shown in Figure 3, the model can be expressed as follows:

$$\begin{aligned} o(t+1) = F(&o(t), o(t-1), \ldots, o(t-n), \\ &x(t+1), x(t), x(t-1), \ldots x(t-n)) \end{aligned} \tag{9}$$

where $o(t+1)$ is the predicted wind power, $o(t)$, $o(t-1)$, …, $o(t-n)$ are the current and past observed wind power. $x(t+1)$, $x(t)$, …, $x(t-n)$ are the new variables selected by PCA.

The network topology of LSTM prediction model is shown in Fig. 4. In this paper, the LSTM network consisting of three hidden layers is built. After the principal component analysis of the original NWP data, the two principal components which have great influence on the wind power are obtained as inputs of the prediction model. And the power of wind farm is the output of the model.

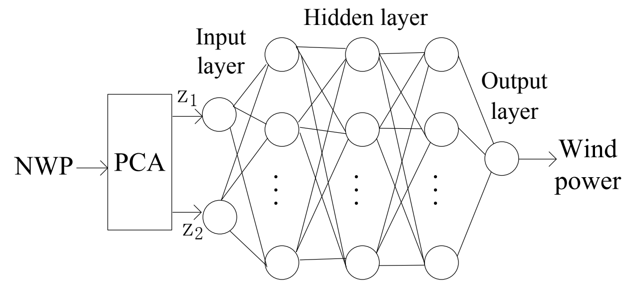

Figure 4. LSTM prediction model based on NWP

IV. EXPERIMENTAL VALIDATION

In this section, the case study is carried out using the data of Manchester wind farm to validate the proposed method. The installed capacity of the wind farm is 120MW. The data from 2010 to 2011 contains the wind power measurement data and the NWP data of Manchester wind farm including air density, pressure, temperature, the wind speed and wind direction in 100m.The time interval is 5 min. Processing and normalizing the original data.

A. *Forecasting accuracy evaluation*

This paper will use Normalized Mean Absolute Error (NMAE) and Normalized Root Mean Square Error (NRMSE) and to study the prediction accuracy [10]. The error measures are defined as follows:

$$e_{NMAE} = \frac{1}{N}\sum_{i=1}^{N}\left|x'(i) - x(i)\right| \tag{10}$$

$$e_{NRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x'(i) - x(i))^2} \tag{11}$$

where $x(i)$ represents the normalized actual observation value at time t, $x'(i)$ represents the forecast normalized value for the same period, N is the number of forecasts.

B. *Results of the prediction model*

Building the LSTM prediction model based on Keras deep learning framework of Python platform. In order to avoid the over-fitting problem when the network is trained and to ensure that the generalization ability of the network is good enough, it is needed to find the best number of hidden layers and the number of neurons in each layer. After training the network for multiple times, the best parameters are set as follows: the number of hidden layers is three, the number of neurons in three layers are 300,500,200 and iteration times are 100.

The original NWP data and principal components are respectively used as inputs of the prediction model for comparison. The data from May 01,2010 to May 31,2011 are taken as training samples. The target is to forecast the future 24 h power of wind farm on June 01, 2011.The wind power forecast results are shown in Fig.5. Fig.6 is the extracted segment from Fig.5. TABLE III shows the results of NMAE and NRMSE of LSTM and PCA-LSTM methods.

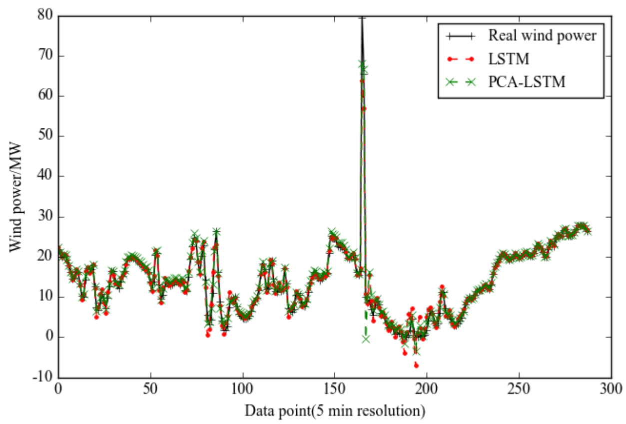

Figure 5. Forecast results of LSTM and PCA-LSTM

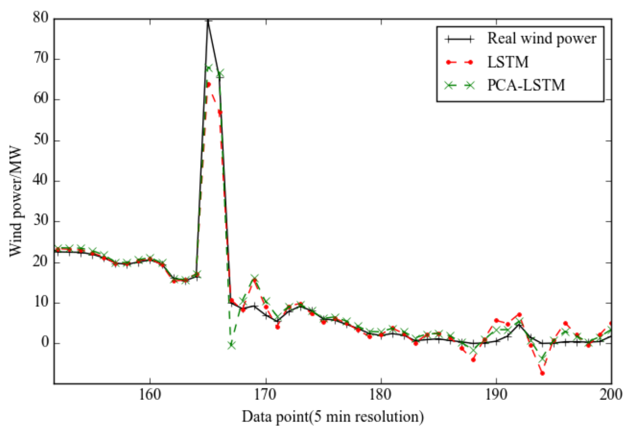

Figure 6. Partial results of LSTM and PCA-LSTM

TABLE III. PREDICTION ERRORS OF LSTM AND PCA-LSTM

Prediction model	**NMAE(%)**	**NRMSE(%)**
LSTM	0.6856	1.3614
PCA-LSTM	0.5672	1.0732

It can be seen that prediction results of both LSTM and PCA-LSTM methods are close to the actual wind power curve, besides the prediction accuracy of PCA-LSTM is higher than LSTM model which bases on original NWP data. The PCA-LSTM model can also reduce the complexity of the network and enhance the generalization ability of the model.

C. Comparison with different models

In this paper, PCA-LSTM model is compared with BP neural network and support vector machine(SVM) model. The prediction results of different models are shown in Fig.7. Fig.8 is the extracted segment from Fig.7. TABLE IV shows the errors of different prediction models.

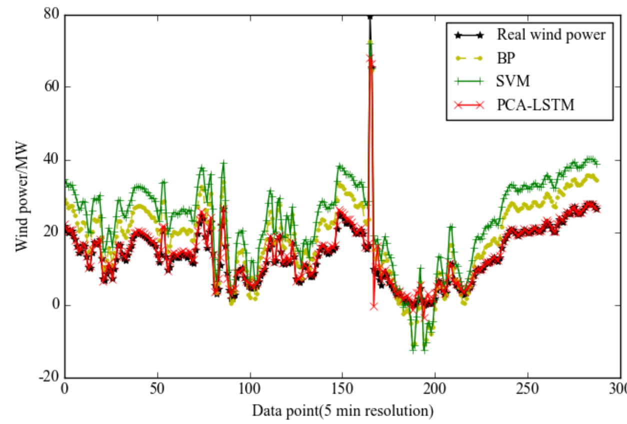

Figure 7. Prediction results of different models

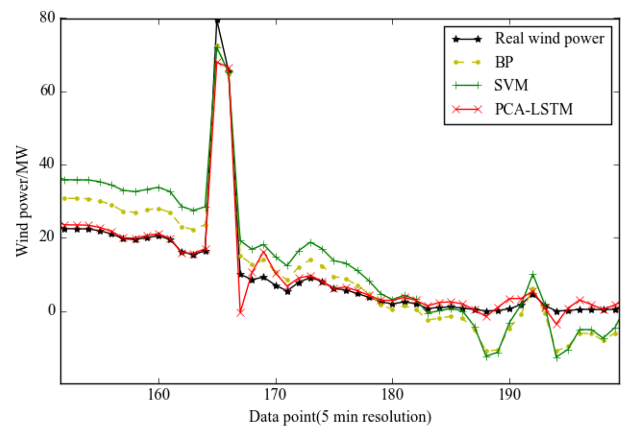

Figure 8. Partial results of different models

TABLE IV. ERRORS OF DIFFERENT PREDICTION MODELS

Prediction model	**NMAE(%)**	**NRMSE(%)**
BP	4.7505	5.1090
SVM	8.2126	8.3681
PCA-LSTM	0.5672	1.0732

As we can see from Figure 7-8 and TABLE IV, for the same test samples, the PCA-LSTM model is close to the actual wind power curve. Compared with BP neural network and SVM model, the errors of PCA-LSTM model are reduced by about 4% and 7%. LSTM method combined with principal component analysis can obtain more effective information from massive data.

V. CONCLUSION

Compared to traditional learning methods, deep learning method has the advantage of data learning and generalization ability. This paper proposes the deep Long-Short Term Memory method applied to wind power prediction. In order to reduce the dimension of the input variables and to reduce the complexity of the network, suitable input samples are selected by principal component analysis. By training and learning the LSTM model, the future 24 h wind power is forecasted. Compared with BP neural network and SVM model, the prediction accuracy of PCA-LSTM is greatly improved. LSTM method can effectively analyze massive data. It has the advantage of generalization and high dimensional function approximation ability. Accordingly, it is shown that the Long-Short Term Memory model is advanced and practical in the field of wind power prediction.

REFERENCES

[1] Tande J,"Grid integration of wind farms," Wind Energy, 2003, 6: 281-295.

[2] XUE Yusheng, YU Chen, ZHAO Junhua, et al, "A Review on Short-term and Ultra-short-term Wind Power Prediction," Automation of Electric Power Systems, 2015, 39(6):141-151.

[3] Bernhard Ernst, Brett Oakleaf and Mark L. Ahlstrom, "Predicting the Wind," IEEE Power and Energy Magazine, vol. 5 no. 6, pp. 78–89, Nov. 2007.

[4] Billinton R, HUA Chen, Ghajar R, "A sequential simulation technique for adequacy evaluation of generating system including wind energy," IEEE Trans on Energy Conversion, 1996, 11(4): 728-734.

[5] BOSSANYI E A, "Short-term wind prediction using Kalman filters," Wind Engineering, 1985, 9(1): 1-8.

[6] XIAO Yongshan, WANG Weiqing, HUO Xiaoping, "Study on the Time-series Wind Speed Forecasting of the Wind Farm Based on Neural Networks," Energy Conservation Technology, 2007, 25(2): 106-108.

[7] FAN Gaofeng, WANG Weisheng, LIU Chun, DAI Huizhu, "Wind Power Prediction Based on Artificial Neural Network," Proceeding of the CSEE, 2008, 28(34): 118-123.

[8] YIN Baocai, WANG Wentong, WANG Lichun, "Review of Deep Learning," Journal of Beijing University of Technology,2015(01).

[9] S. Hochreiter, J.Schmidhuber, "Long short-term memory," Neural Computation, 1997, 9(8): 1735-1780.

[10] SIDERATOS G, HATZIARGYRIOU N D, "An advance statistical method for wind power forecasting," IEEE Trans on Power Systems, 2007, 22(1): 258-265.