Multi-Step wind power forecasting model Using LSTM networks, Similar Time Series and LightGBM

Yukun Cao
School of Computer Science and Technology Shanghai University of Electric Power, Shanghai, China
Email: marilyn_cao@163.com

Liai Gui
School of Computer Science and Technology Shanghai University of Electric Power, Shanghai, China
Email: 1018615946@qq.com

Abstract—**Intermittent and fluctuating wind forces are detrimental to the grid. A multivariate model was proposed to improve the accuracy of wind power generation prediction in order to induce system operators to reduce risks. The model consists of three steps. First, the meteorological data such as wind speed are predicted by LSTM networks on the basis of traditional time series approaches. Then a method of similar time series matching with hierarchical search is proposed to highlight the main factors and save computing time. We use similar disparity as a criterion to select similar meteorological series and power data as training sets. Finally, similar data are inputted into LightGBM for modeling, training, and prediction. Industrial data of the wind power plant is examined case. The results are clearly display that the proposed method can effectively predict wind power in the next 6 hours and achieve high precision, which has certain engineering practical value.**

Keywords-Wind power; Multi-step prediction； similar time series; LSTM network； LightGBM

I. INTRODUCTION

Prediction of wind speed and wind power plays an important role in wind power generation management of government and energy companies. Reliable and accurate prediction offers excellent control and power planning, e.g. daily and hourly dispatch, transmission, power storage and management of other wind power related operations [1]. In the past decade, many approaches to predicting wind power generation in the short term, medium term or long term have been addressed and categorized mainly into the following three types: physical models, time series models [2] and artificial intelligence models [3-4].

The physical method uses physical and weather information such as wind direction, roughness, obstruction, pressure, temperature, etc. to model wind power and predict multi-step values. These methods have advantages for long-term forecasting of wind power [5]. However, due to its high dependence on numerical weather forecast data, accuracy is greatly affected by the precision of numerical weather prediction. The time series models use historical power time series to predict future wind power generation. The modeling is simple and does not depend on numerical weather forecast data, but the prediction accuracy is low, and the accuracy decreases rapidly with the increase of prediction time. The time series approach to building wind forecasting models requires some key weather variables including wind speed and wind direction [6]. Historical dataset of wind power generation can also be directly used in time series models to predict wind power. The artificial intelligence methods use different artificial intelligence models, which establish a nonlinear mapping relationship between input variables and output variables through abundant historical samples' training, and predict future wind power based on the trained model. Artificial intelligence includes BP neural net-work, support vector machine [7] and ANN [8], etc.

Review of existing research, there are still three problems to be solved: First, meteorological data is directly related to wind power output. When it is impossible to directly measure and predict the wind power output, accurate prediction of meteorological data becomes particularly important. Second, model complexity control. In other words, the question is how to overcome the over-fitting of the training stage to improve the generalization ability of the prediction stage. Thirdly, filtering of input variables, i.e., preprocessing inputted varia- bles, controlling redundant information and improving com-putational efficiency of model training.

For the first problem, traditional statistical models, e.g. autoregressive (AR) models and autoregressive and moving average (ARMA) models [9-10] are often used for wind speed prediction. However, the accuracy is generally not high, only better than the persistence method. As a special RNN model, the long-short-term-memory (LSTM) [11] network can consider the temporal correlation and effectively avoid the gradient disappearance and gradient explosion during the training stage through its special gate structure design. LSTM can be trained more effectively to make effective use of historical sequence information [12].

For the second problem, there have been some methods to prevent over-fitting, including Bayesian methods, bagging methods, and boosting methods [13]. LightGBM [14] uses the boosting method to integrate multiple weak learners to form a strong learner. The algorithm has strong expressive ability and can be used for most regression problems. Compared with the general regression tree algorithm, it has the merits of preventing over-fitting, excellent generalization ability, faster speed and lower memory consumption [15].

For the third problem, there are some similarity methods to filter inputted data. In [16-17], the maximum wind speed, minimum speed values and average ones are taken as feature vectors, and the historical wind speed series with the best similarity are selected as the model input. In the litera- ture above, Euclidean distance is often used as the similarity basis, which can only reflect the magnitude of wind speed or the numerical similarity of other characteristics, but cannot take into account the similarity of wind speed variation trend. Similar disparity [18] is a new statistical measure that can objectively compare the similarity between samples. It can reflect not only the similar numerical values, but also the similar trends. Considering the diversity of meteorological features, a hierarchical searching method for similar time series is adopted to highlight the influence of wind speed and wind direction as the dominant factors.

This paper proposed an LSTM meteorological numerical prediction network and a multi-step wind power forecasting method based on similar time series and LightGBM. First, the forecasting day's meteorological series are predicted by LSTM networks. Then we use similar disparity as a criterion to select similar meteorological series and power data from original data. Finally, similar data are inputted into LightGB-M for modeling, training, and prediction.

II. THE PROPOSED FORECASTING MODEL

A. Feature Selection

The dataset in this study is from a wind farm in Shang-hai, which provides wind power, wind direction, wind speed, air temperature, atmospheric density, humidity and other meteorological information. Based on the data set, we exp-lore the influence of various meteorological factors on wind power generation, so as to select appropriate feature vectors. The Pearson similarity in the distance analysis method is used to approximate the correlation between the power generation data and the other data of the wind farm. The formula for calculating the Pearson correlation coefficient of two n-dimensional vectors x and y is shown in (1):

$$\mathrm{r}_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \tag{1}$$

where $\overline{x}$ and $\overline{\mathrm{y}}$ are the average of all values in x and y, respectively. The Pearson-correlation coefficient r_{xy} is a real number in [-1,1]. When r_{xy}>0, the two variables are positiv-ely correlated, otherwise, they are negatively correlated. The larger $|\mathrm{r}_{xy}|$, the higher the correlation between x and y. Fina-lly, according to the calculation results, wind speed, wind direction, atmospheric pressure and temperature with high correlation coefficient are selected as feature vectors. The correlation coefficient between them is shown in Table I. It is worth pointing out that in different wind farms, there is often a difference in the degree of correlation between wind power and various meteorological factors. When the subject chang-es, the correlation conclusion may no longer apply.

TABLE I. CORRELATION COEFFICIENT BETWEEN WIND POWER GENERATION AND METEOROLOGICAL FACTORS

Variable	WD	WS	AT	AP	WP
WD	1.0000	–0.3736	–0.1754	–0.0179	0.3254
WS	–0.3736	1.0000	0.4473	–0.3789	0.9065
AT	–0.1754	0.4473	1.0000	0.6329	–0.0521
AP	–0.0179	–0.3789	–0.6329	1.0000	–0.2736
WP	0.3254	0.9065	–0.0521	–0.2736	1.0000

B. LSTM Networks

LSTM is a special type of recurrent neural network (RNN) that performs well in long-term and short-term dependencies due to its special structure. Especially, it's good at solving time series problems. As shown in Fig.1, the memory cells that replace the hidden layers of traditional neurons are the core of the entire LSTM networks [19-20].

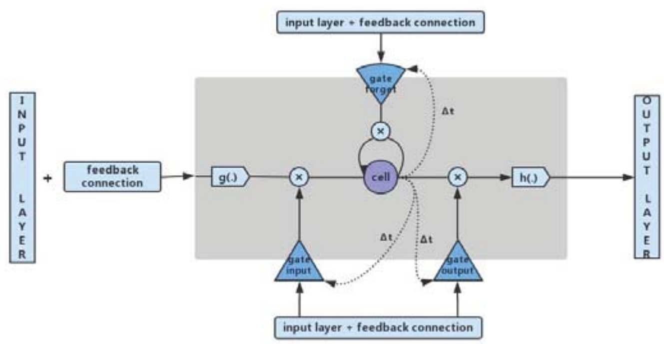

Figure 1. LSTM unit

With the input gate, output gate and forget gate, the LSTM network can affect the cell status by adding or removing information. Updating unit status and computing LSTM network output can be calculated as follows:

$$i_t = \sigma(w_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_c) \tag{2}$$

$$f_t = \sigma\left(w_{fx}x_t + w_{fm}m_{t-1} + w_{f_c c_{t-1}} + b_f\right) \tag{3}$$

$$\mathrm{c}_t = f \odot \mathrm{c}_{\mathrm{t-1}} + i_t \odot \mathrm{g}(\mathrm{W}_{\mathrm{cx}}\mathrm{x}_{\mathrm{t}} + \mathrm{W}_{\mathrm{cm}}\mathrm{m}_{\mathrm{t-1}} + \mathrm{b}_{\mathrm{c}}) \tag{4}$$

$$\sigma_{\mathrm{t}} = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \tag{5}$$

$$\mathrm{m}_t = o_t \odot \mathrm{h}(\mathrm{c}_{\mathrm{t}}) \tag{6}$$

$$\mathrm{y}_t = W_{ym}m_t + b_y \tag{7}$$

where x_t and y_t are input data and output data, respectively; i_t, f_t and o_t represent input gate, forgetting gate and output gate respectively ; And c_t is an activation state of each unit. m_t is the same one of memory block. In addition, W_{ix}, W_{im}, W_{ox}, W_{om}, W_{oc}, W_{fm}, W_{fc}, W_{cx}, W_{ic} W_{fx}, W_{cm}, and W_{ym} are the corresponding coefficients of weight; σ, h and g represent the gate, the output and input activation functions. The activation function often selects the Sigmoid function or the tanh function. ⊙ represents multiplication of elements (i.e., Hadamard product) between two vectors. b_c, b_o, b_i, b_f, and b_y are dissimilar bias vectors[21].

At present, for the recurrent neural network model such as LSTM, there are two mainstream training methods: the Back-Propagation Trough Time algorithm (BPTT) [22] and the Real-Time Recurrent Learning (RTRL) [23]. Since the BPTT algorithm is clear in concept and efficient in calculation, it is adopted in this paper to train LSTM network. The process of BPTT training LSTM is shown in Fig.2. Finally, four LSTM networks are generated to predict the four feature vectors for the next six hours.

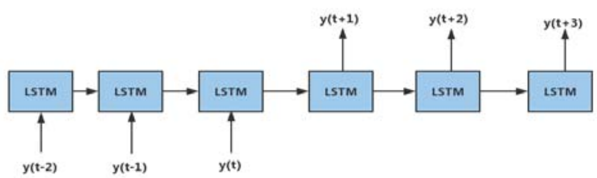

Figure 2. LSTM training by BPTT

C. Similar Time Series Selection

The traditional similar method is to select the historical wind power time series based on the similarity of weather information, and assume that it is close to the power of the prediction day [24]. In this paper, a 6×4 dimensional weather series and corresponding six historical power values are selected. It can be found out from Table I that wind speed and wind direction are the dominant factors. Therefore, the dominant factors should be treated more carefully in the similar time series matching.

This paper adopts a layered approach to the search of similar time series. Firstly, we match similar wind speed and wind direction series. Then we search for temperature and air pressure series. It's worth mentioning that the similarity of wind speed and wind direction must be greater than a certain threshold. The temperature and pressure's similarity should be no less than a certain threshold.

The first layer searches for wind speed and wind direction series, and the time series of test set is:

$$x_1 = [\vec{s}, \vec{d}] \tag{8}$$

$$\vec{s} = [s_1, s_2, ..., s_6] \tag{9}$$

$$\vec{d} = [d_1, d_2, ..., d_6] \tag{10}$$

The second layer searches for temperature and air pressure, and the time series of test set is:

$$x_2 = [\vec{t}, \vec{p}] \tag{11}$$

$$\vec{t} = [t_1, t_2, ..., t_6] \tag{12}$$

$$\vec{p} = [p_1, p_2, ... p_6] \tag{13}$$

The similar disparity is used as the criterion to measure the similarity between two time series, which is represented by the symbol $C(X,Y)$:

$$C(X,Y) = (\alpha R_{xy} + \beta D_{xy}) / (\alpha + \beta) \tag{14}$$

$$D_{xy} = \frac{M(X,Y)}{n} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{15}$$

$$R_{xy} = \frac{\sum_{i=1}^{n} |x_i - y_i - E_{xy}|}{n} \tag{16}$$

$$E_{xy} = \frac{\sum_{i=1}^{n} (x_i - y_i)}{n} \tag{17}$$

where $C(X,Y)$ represents the similarity between two samples, and the smaller the value, the more similar the two samples are. D_{xy} is actually the hamming distance, which can accurately reflect the difference between the two samples in the total average value. R_{xy} can reflect the difference between two samples (x_i, y_i) and their dispersion of E_{xy}. If the degree of dispersion is smaller, the shape of the two sample curves is more similar. Similarity is determined by R_{xy} and D_{xy} .While α and β represent their contribution to the total similarity coefficient. This paper they are both set to 0.5.

D. LightGBM Forecasting

LightGBM [14] is Microsoft's open source GBDT algorithm in 2016. Histogram-based algorithm [25] is used to accelerate the training procedure and reduce memory consumption; In addition, It uses parallel voting decision tree algorithms combined with advanced network communication techniques to qualifying parallel learning. The training data is divided into multiple machines, and a local voting decision to select the top-k attribute and a global voting decision to receive the top-2k attribute are performed in each iteration. LightGBM applies the leaf-wise method to search for leaves with maximum gain [26]. We use matched similar weather time series as training sets for LightGBM. The weather time series predicted by the LSTM networks is a test set for LightGBM.

A schematic diagram is shown in Fig.3 in order to facilitate the detailed description of the proposed multi-step wind power generation prediction method. Firstly, the meteorological time series of the initial sets and the weather series predicted by LSTMs are analyzed for similarity matching. The meteorological values and corresponding power of the similar time series are extracted to form the training sets. Finally, the LightGBM model is used as the prediction model.

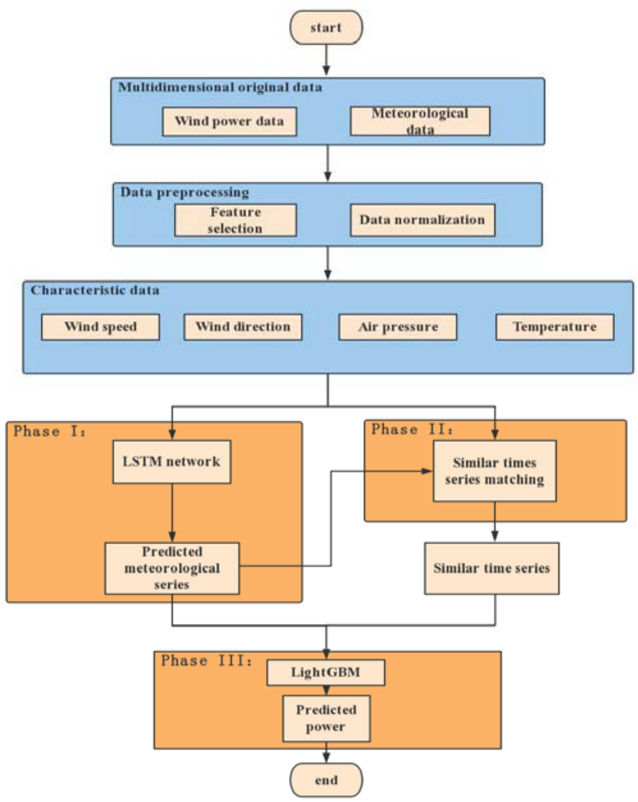

Figure 3. The schematic diagram describing the proposed approach

III. APPLICATION

The data used in this paper was from a wind farm in Shanghai. The observation interval of the data is 1 hour, including multiple related variables such as wind speed, wind direction, air temperature, air pressure and power generation. It starts from 01.01.2007 00:00:00 A.M to 12.31.2012 24:00:00 P.M, providing 52,560 samples in total. The wind farm's total data is divided into two parts, Data Set 1 and Data Set 2, which include 47,760 and 4800 samples, respectively. Data set 1 is the training set, and data set 2 is for testing.

A. *Data Normalization*

Using the normalization method, the wind speed, temperature, atmospheric pressure and wind power are normalized by statistical limit values, and the values are reduced to the interval [-1, 1]:

$$x' = \frac{x-(x_{\max}-x_{\min})/2}{(x_{\max}-x_{\min})/2} \tag{18}$$

where $x_{\max}$ and $x_{\min}$ are the largest and smallest values of the variables, respectively

To make the predicted values have physical significance, the calculation formula of the inverse normalization is as follows:

$$x = 0.5[x'(x_{\max}-x_{\min})+(x_{\max}+x_{\min})] \tag{19}$$

B. *LSTM Weather Series Forecasting*

The LSTM networks for weather series prediction mainly needs to determine five hyperparameters of the model. They are the input layer time step, the input layer dimension, the number of hidden layers, each hidden layer dimension, and the output variable dimension.

The input layer time step is equal to the length of the variable time series used for prediction. In this paper, the parameter is set to 9 through trial, i.e., the historical data of the first 9 moments is inputted for prediction. Input layer dimension is the number of variables, the value of the four LSTM networks in this article is set to 1. The number of hidden layers is the number of LSTM layers. This forecasting task is based on historical time series to predict the meteorological time series in next six hours. Therefore the output variable dimension is set to 6. The final prediction results are shown in Fig. 4.

C. *Similar Time Series Matching*

The specific steps for matching similar time series are as follows:

1) Extract the test set of wind speed and direction series as:

$$X = [s_1, s_2, \ldots, s_6, d_1, d_2, \ldots, d_6]$$

2) Extract historical wind speed and wind direction series as:

$$Y = [Y_1, Y_2, \ldots, Y_m]$$

$$\begin{cases} Y_1 = [s_1, s_2, \ldots s_6, d_1, d_2, \ldots, d_6] \\ Y_2 = [s_2, s_3, \ldots s_7, d_2, d_3, \ldots, d_7] \\ Y_3 = [s_3, s_4, \ldots s_8, d_3, d_4, \ldots, d_8] \\ \ldots \\ Y_i = [s_{m-6+1}, s_{m-6+2}, \ldots s_m, d_{m-6+1}, d_{m-6+2}, \ldots, d_m] \end{cases}$$

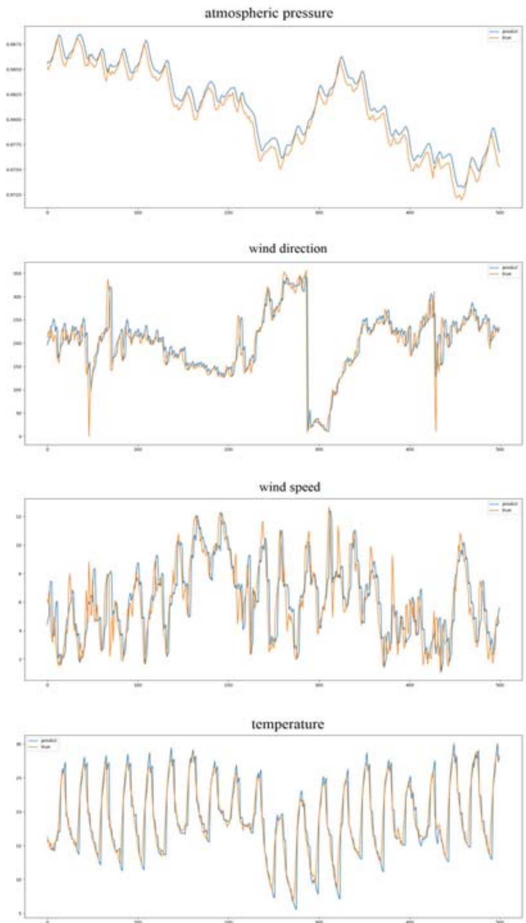

Figure 4. The LSTM weather series forecasting curve

3) Calculate the similarity between Y_i and X by calculating the similar disparity.

Since the similar disparity is as small as possible, the match threshold of the first layer is set to 0.4. The second layer searches for temperature and air pressure series, and the steps are similar to the above. The second threshold is set to 0.7.

Similar disparity and Euclidean distance are used as similarity criteria, respectively. The best historical time series similar to the test set series are searched for training, and the results are compared as shown in Fig. 5.

D. Wind Power Forecasting

1) Baselines

After getting similar power sequences and weather time series, we obtained training set and test samples. For a clearer comparison, we apply the ANN model, the SVM model, and the three-stage multivariate model to the same data set.

2) Evaluation

To fully prove the proposed model's superiority, we chose three common prediction accuracy metrics to compare

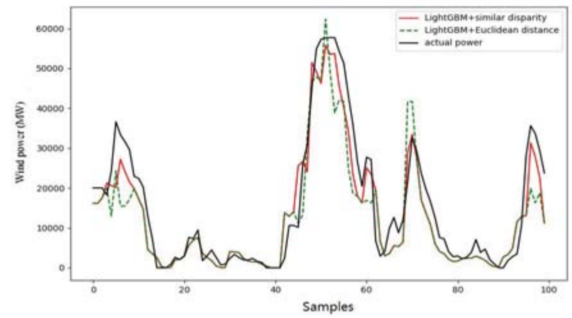

Figure 5. The result of similarity criterion comparison

the ANN, the SVM, and the proposed model. These criteria are the root mean square error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). They are shown in formula (21), (22) and (23):

$$e_i = x_i - \hat{x}_i \tag{20}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{21}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2} \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}| \frac{x_i - \hat{x}_i}{x_i} \times 100 | \tag{23}$$

where x_i is the actual wind power at the ith time point, $\hat{X}_i$ represents the model prediction value at the same point, and n is the total number of predictions.

3) Experimental results

Fig. 6 shows the forecasting results about the 3955th-4057th data points under 6 steps by three different models. The forecasting errors are presented in Table II. From the chart we can see that the model is more accurate than the other two models for multi-step ahead prediction, especially when predicting non-stationary wind power.

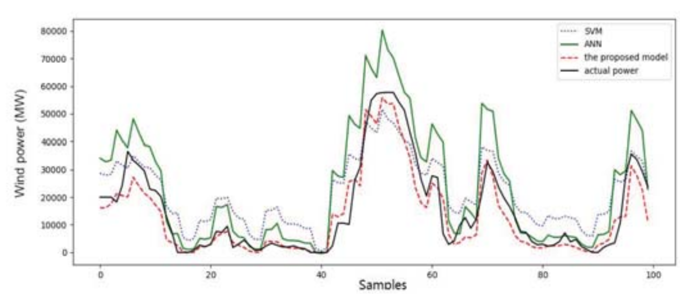

Figure 6. The 6-step ahead forecasting curve

TABLE II. ERROR METRICS OF THREE MODELS

	RMSE(MW)	MAE(MW)	MAPE(%)
ANN	3549.5679	3040.5854	33.145
SVM	2412.3102	1994.3915	26.39
The Proposed Model	852.618	835.8336	16.397

IV. CONCLUSION

In order to improve prediction accuracy of wind power generation prediction, this paper proposes a three-stage complex multivariate model. Industrial data from the Shanghai wind farm were used for the study and four typical meteorological variables were used for prediction. The conclusions are summarized as follows:

(1) From results of Table I, we can see that weather fluctuations have different effects on the wind process. Wind changes are complex reactions of multiple variables, and predictions through multiple variables can make the results more accurate.

(2) Selecting similar time series can reduce the number of input samples and mine more useful historical information. The result of Fig.5 illustrates that similar weather series are more effective in wind power prediction than original data in modeling. The similarity based on similar disparity is better than Euclidean distance because it also considers the trend.

(3) The three-stage prediction model was applied to the actual wind farm, and the performance was better verified under three error metrics. Summarizing all the research results in this paper, it is proved that the proposed method can improve the accuracy of wind power prediction. It is also feasible in multi-step wind power prediction, and keeps the better performance.

REFERENCES

[1] Maatallah, O., Achuthan, A., Janoyan, K., and Marzocca, P., ‘‘Recursive wind speed forecasting based on Hammerstein auto-regressive model,’’Appl. Energy,2015, 145, pp. 191–197.

[2] Kavasseri RG, Seetharaman K.,‘‘Day-ahead wind speed forecasting using f-ARIMA models,” Renew Energy, 2009, 34, 5, pp.1388–1393.

[3] Yuan X, Chen C, Yuan Y, Huang Y, Tan Q. ‘‘Short-term wind power prediction based on LSSVM–GSA model,’’ Energy Conversion and Managemen, 2015,101,pp.393–401.

[4] Ramasamy, P., S. S. Chandel, and A. K. Yadav,‘‘Wind speed prediction in the mountainous region of India using an artificial neural network model.’’ Renewable Energy,2015,80,pp.338–347.

[5] Schicker I, Papazek P, Kann A, and Wang, Y., ‘‘Short-range wind speed predictions for complex terrain using an interval-artificial neural network,’’ Energy Procedia, 2017, 125, pp.199–206.

[6] Amjady N., Keynia F., Zareipour H., ‘‘Wind Power Prediction by a New Forecast Engine Composed of Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization,’’ IEEE Transactions on Sustainable Energy, 2011,2(3), pp.265–76.

[7] Liu Y, Shi J, Yang Y, and Lee,W.J., ‘‘Short-Term Wind-Power Prediction Based on Wavelet Transform–Support Vector Machine and Statistic-Characteristics Analysis,” IEEE Transactions on Industry Applications,2012,48(4),pp.1136–1141

[8] Quan H, Srinivasan D, Khosravi A., ‘‘Short-term load and wind power forecasting using neural network-based prediction intervals,” IEEE Transactions on Neural Networks and Learning Systems,2014,25(2) pp.303–315.

[9] Erdem E, Shi J., ‘‘ARMA based approaches for forecasting the tuple of wind speed and direction,” Appl Energy, 2011,88,pp.1405–1414.

[10] Kavasseri R G, Seetharasman K., ‘‘Day-ahead wind speed forecasting using ARIMA models,” Renewable Energy, 2009, 34(5),pp.1388-1393.

[11] Hochreiter S , Schmudhuber J., ‘‘Long short-term memory,” Neural computation,1997,9(8),pp.1735–1780.

[12] Tian Y, Pan L.,‘‘Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network,” IEEE International Conference on Smart City, 2015,pp.153–158.

[13] Varnek A., ‘‘Bagging and Boosting of Regression Models,” Tutorials in Chemoinformatics. John Wiley & Sons, Ltd, 2017, pp.249–255.

[14] Meng, Q., Ke, G., Wang, T., Chen, W., Ye, Q., Ma, Z. M., and Liu, T., ‘‘A communication-efficient parallel algorithm for decision tree,” In Advances in Neural Information Processing Systems,2016,pp. 1271–1279.

[15] Wang D, Zhang Y, Zhao Y.,‘‘LightGBM: An Effective miRNA Classification Method in Breast Cancer Patients,” International Conference,2017,pp.7–11.

[16] Zhang X, Wang R, Liao T,and Zhang T,‘‘Short-Term Forecasting of Wind Power Generation Based on the Similar Day and Elman Neural Network,”Computational Intelligence,2015 IEEE Symposium.IEEE, 2015, pp.647-650

[17] Wang, S. Z., Gao, S., Zhao, X., and Zhang, N., ‘‘A multi-timescale wind power forecasting method based on selection of similar days,” China International Conference on Electricity Distribution. IEEE,2016.

[18] Zhang A, Ding D, Li X,Yin Z, and Han C, ‘‘Application of similarity degree in pavement erosion prediction in Beijing,”. Meteorological Science and Technology Progress, 2012, 02(1),pp.38–42.

[19] Sainath, T. N., Vinyals, O., Senior, A., and Sak, H., “Convolutional, long short-term memory, fully connected deep neural networks,” IEEE Int Conf Acoust Speech Signal Process 2015,2015,4580–4584.

[20] López, E., López, E., Valle, C., Valle, C., Allende, H., and Allende, H., et al. “Wind power forecasting based on echo state networks and long short-term memory,” Energies,2018, 11(3), 526.

[21] Hu,Y.L., Chen,L., ‘‘A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm,”Energy Conversion and Management, Vol.173,2018,pp. 123-142.

[22] Y. Bengio, P. Simard and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult," in IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157-166, March 1994.

[23] Chang F, Chang C, Huang H,“Real-time recurrent learning network for stream-flow forecasting,” Hydrological Processes, 2002,16(13),pp. 2577–2588.

[24] Renani E T, Elias M F M, Rahim N A., “Using data-driven approach for wind power prediction: A comparative study," Energy Conversion & Management, 2016, 118, pp.193-203.

[25] Ranka, S. and Singh, V., “Clouds: A decision tree classifier for large datasets,” In Proceedings of the 4th Knowledge Discovery and Data Mining Conference, 1998, 2–8.

[26] Wang, D., Zhang, Y., and Zhao, Y., “LightGBM: An Effective miRNA Classification Method in Breast Cancer Patients,” International Conference,2017, pp.7-11.