

A Data Mining Approach Combining K-Means Clustering with Bagging Neural Network for Short-term Wind Power Forecasting

Wenbin Wu and Mugen Peng

Abstract—Wind power forecasting (WPF) is significant to guide the dispatching of grid and the production planning of wind farm effectively. The intermittency and volatility of wind leading to the diversity of the training samples have a major impact on the forecasting accuracy. In this paper, to deal with the training samples dynamics and improve the forecasting accuracy, a data mining approach consisting of K-means clustering and bagging neural network is proposed for short-term WPF. Based on the similarity among historical days, K-means clustering is used to classify the samples into several categories, which contain the information of meteorological conditions and historical power data. In order to overcome the over fitting and instability problems of conventional networks, a bagging-based ensemble approach is integrated into the back propagation neural network. To confirm the effectiveness, the proposed data mining approach is examined on real wind generation data traces. The simulation results show that it can obtain better forecasting accuracy than other baseline and existed short-term WPF approaches.

Index Terms—Wind power forecasting, k-means clustering, ensemble learning, bagging neural network

I. INTRODUCTION

The development and utilization of renewable energy has been one of the hottest spots around the world. Wind power generation is rapidly expanding into a large-scale industry due to the cleanness and wide availability, and has been characterized as an fluctuating and intermittent power. Unfortunately, it is difficult to ensure the security and stability while accessing to the electricity grid, especially for large-scale application. An accurate and reliable wind power forecasting approach is essential for power quality, reliability management while reducing the cost of supplying spinning reserve.

A large number of wind power forecasting approaches have been proposed in many literatures, which can be roughly classified into three categories: 1) physical forecasting approach; 2) statistical forecasting approach; 3) combination approach [1]. The physical approach uses detailed topological and meteorological descriptions to model the conditions at the location of the wind farm [2]. Then the wind speed is predicted by numerical weather forecasting approach and converted into wind power via the power curves generated by wind turbines. The accuracy of this approach largely depends on the amount of physical information. The statistical approach aims at establishing the relationship between wind power

and a set of variables including historical data and online measured data such as wind speed and wind direction [3]. This approach is applicable for most scenarios without considering geographical conditions. The core idea of the combination approach is to take advantage of the physical and statistical methods and improve the forecast accuracy [4].

The physical approach can establish a specialized wind power forecasting model for a wind farm without large amount of historical data. In particular, this approach needs detailed physical characteristics of wind turbines and wind farm to achieve an accurate model [5]. However, it is difficult to collect these physical characteristics in a short period of time. Besides, the model has poor versatility due to the factors of specific geographical conditions and operation state of wind turbines.

Compared with the physical approach, the statistical approach needs to collect a lot of historical data to build the forecasting model. In conventional statistical approaches, the time series model is usually applied to predict the wind power depending on the principle of continuity. There are several kinds of time series models, including autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA), and autoregressive integrated moving average model (ARIMA) [9]- [12]. Reference [13] concentrates on multi-time series and multi-time scale modeling in wind speed and wind power forecasting. Then AR, MA, ARMA and ARIMA approaches are utilized to model multi-time series. The Kalman filter approach [14] is applied to optimize the time series model, which can modify the weights dynamically on the basis of ARMA model. Much work have shown that the time series models can obtain good forecasting accuracy only in the very short-term scale scenarios. In addition, this approach is applicable for the continuous sequence without considering the similarity among historical data.

Data mining approaches are catching researcher's attention [6]- [8]. It has been reported that neural network (NN) outperformed other methods in short-term forecasting problems [19]. NN is able to model the complex nonlinear relationship between the historical data and forecasting power [15]. Besides, NN needs to use the historical data which have a major impact on the wind power as the input variables, such as wind speed and temperature. Back propagation (BP) NN is easy to formulate and often used as a reference model. But there are some potential improvements for the BPNN. Firstly, the result of the BPNN is sensitive to the initialization of weights and biases, and the network is easy to fall into local optimum. Secondly, the overtraining problem may arise if the BPNN has

Wenbin Wu (e-mail: bingeb@bupt.edu.cn) and Mugen Peng (e-mail: pmg@bupt.edu.cn) are with the Key Laboratory of Universal Wireless Communications (Ministry of Education), Beijing University of Posts and Telecommunications, Beijing, China.

too many parameters to be estimated in the training samples [16]. Finally, several improvements for the BPNN are achieved by 1): optimizing parameters 2): tuning the structure of NN 3): changing the kernel of NN. Particle swarm optimization (PSO) [20], the Elman NN [17] and Radial Kernel Function (RBF) as examples respectively correspond to the above approaches. However, these improvements don't consider that the selection of training samples has an important impact on the forecasting accuracy.

In order to improve the forecasting accuracy, clustering approaches are also applied to select proper training set. Reference [17] proposes an approach which selects the similar days as the training samples. The K-means clustering is utilized to classify wind speed into several categories. In reference [21], the ARMA model is presented and the K-means clustering is used to classify wind direction. However, only the weather conditions are clustered and the actual operation state of wind turbines is not considered.

In this paper, a data mining approach consisting of the K-means clustering and bagging neural network is proposed to predict the short-term wind power of individual wind turbine. The main contributions of this paper are summarized as follows.

- 1) To mitigate the impact of the diversity of training samples, the K-means clustering is utilized to classify the original data sets into several categories according to meteorological conditions and historical power. Person correlation coefficient is used to calculate the similarity between each category and the forecasting day. The most similar category is selected as the training sample.
- 2) To overcome the instability and over fitting problems of the BPNN, a bagging-based approach is integrated into the BPNN as the forecasting engine. The random sampling is conducted on the training sample by the Bootstrap sampling technique to form N subsets. The BP algorithm is applied to each subset and trained on the individual network. The final result is determined by calculating the average value of N networks.
- 3) The performance of the proposed approach is evaluated. Simulation results show that the root mean squared error (RMSE) and mean absolute error (MAE) of the proposed approach are less than other baseline approaches, which demonstrates the proposed approach can improve the forecasting accuracy.

Besides, many researchers have focused on the platform of collecting and processing data in wind turbines because they are important for managing and forecasting the wind energy. In order to make use of the manifold potential, the available data of wind turbines needs to be analyzed and the advanced measuring technologies and communication networks should be implemented. The Cyber Physical System (CPS) is a powerful approach to develop the wind energy sector, which has been widely used in CMS or SCADA systems [24].

The rest of the paper is organized as follow. Section II presents the process of establishing the prediction model. Section III carries out the results of the simulation and Section IV concludes the paper and summarizes the future work.

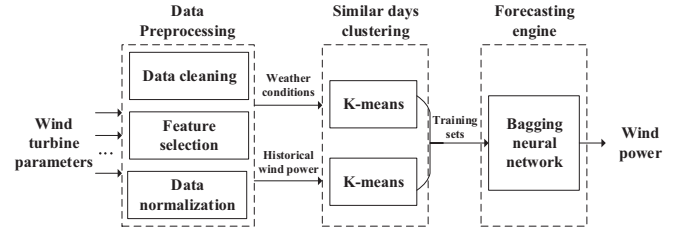


Fig. 1. The system architecture for the proposed WPF approach

II. THE ESTABLISHMENT OF THE WIND POWER FORECASTING MODEL

Data mining approaches have been widely used for classification and prediction problems. The proposed approach is based on data mining, which consists of the K-means clustering and bagging neural network. Fig. 1 shows the wind power forecasting model. Firstly, data preprocessing is conducted on the vector space to clean unreasonable data, normalize the training samples and select the most related variables as the inputs of the neural network. Secondly, data after preprocessing are clustered by the K-means clustering to select the training set which is most similar to the forecasting day. Finally, the wind power is forecasted by the bagging neural network, which is able to alleviate the instability and over fitting problems of the BPNN.

A. Data preprocessing

A number of wind turbine parameters are collected as the training samples via the sensor unit. However, these samples may contain unreasonable data. Besides, using too many parameters as the training features would increase the computing complexity and obtain undesired results for the reason that some variables are irrelevant or redundant in this model. Selecting features which are most related to the wind power is able to improve the accuracy. Finally, data normalization has an effect on the convergence rate and accuracy of the training algorithm. Thus, in order to obtain accurate forecasting results, data preprocessing is necessary.

- *Data cleaning*: The original samples may contain data whose values of some characteristics are unreasonable. For example, the values of wind speed and wind power are less than zero. It is obvious that these data are not available and need to be removed. Then, it is necessary to fill the vacancy of the deleted data. The mean value method is applied, which makes use of the mean value ahead and back of the deleted data.
- *Feature selection*: In theory, more input variables can carry more discriminating power. But in practice, excessive variables are prone to cause many problems. Therefore, selecting a suitable set of input variables from the raw data has a great impact on the forecasting performance. Relief algorithm is a kind of feature weighting algorithms. The core idea is that the different weights are assigned to the corresponding features according to the correlation, and the feature whose weight is less than the threshold would be removed. The formulation of Relief

algorithm is given by [18]. The running time of Relief algorithm increases linearly with the sampling times and original features so that this method has high operating efficiency. In addition, Relief algorithm can achieve the purpose of physical dimensions reduction compared with the principal component analysis (PCA).

- *Data normalization*: The aim of data normalization is to transform the raw data to the same orders of magnitude so that the convergence rate and forecasting accuracy can be improved. The min-max method is applied for normalization, which can be expressed as $\bar{x} = (x - \min)/(max - \min)$. In this equation, x is the original data, and max and min represent the maximum and minimum value of the training set. The result \bar{x} is mapped to [0,1].

The three steps above are significant to more accurate forecasting results. After data preprocessing, the proposed approach can be implemented and compared with other WPF approaches.

B. Similar day clustering based on meteorological conditions and historical power

Large amount of data are generated by wind turbines when operating normally. The training samples can be regarded as the time series and auto regressive method is often utilized to build the forecasting model. However, the random behavior of wind can lead to the inconsistencies of training samples. Then, it is difficult to obtain desire results when using this kind of training samples. Therefore, selecting similar days and classifying them into the same category as the training samples can be considered as a good method to improve the forecasting accuracy.

1) *The clustering sample*

Wind speed and temperature have a major impact on wind power generation. Selecting the samples which have the most similar wind speed and temperature with the forecasting day from the historical data is significant. A sample consisting of wind speed and temperature is used as the clustering benchmark, which is constructed as

$$S_1 = [WS_{max}, WS_{min}, WS_{mean}, T_{max}, T_{min}, T_{mean}], \quad (1)$$

where WS_{max} , WS_{min} , WS_{mean} , T_{max} , T_{min} and T_{mean} represents the maximum, minimum and average values of wind speed and temperature respectively.

According to the investigation results, wind power generation is largely determined by the meteorological conditions. However, the similar meteorological conditions can not represent that the wind power of the training samples is similar to the same of the forecasting day because of the different geographical conditions and the operation state of wind turbines. Therefore, the historical data which have the similar wind power with the forecasting day is also used for clustering. Then the sample S_2 could be expressed as

$$S_2 = [WP_{max}, WP_{min}, WP_{mean}], \quad (2)$$

where WP_{max} , WP_{min} , WP_{mean} represents the maximum, minimum and average values of daily wind power generation. Although the S_2 of the forecasting day is unknown, the day which is before the forecasting day can be assumed to have the same S_2 . According to the principle of wind power generation, the daily wind power generation curve has the same trend for next few days in the same season. Then the samples S_1 and S_2 are clustered respectively. It is assumed that cluster 1 from S_1 and cluster 2 from S_2 are most similar to the forecasting day. Finally, the same days from cluster 1 and cluster 2 are collected as the training samples of the forecasting model.

2) *The clustering algorithm*

Clustering belongs to unsupervised learning, which has no label on the training data. In this paper, it is needed to use the clustering algorithm to mine the similar samples and category them into one class. Typical clustering algorithms contain the K-means clustering and expectation maximization (EM).

In this paper, the K-means clustering is used as the clustering method for the reason that K-means algorithm can handle large amount of data sets effectively. K-means algorithm is a kind of clustering method based on partitioning, which usually judges the similarity by computing the distance. The core of K-means algorithm is to choose k center points randomly and partition the data according to the distance between the data and k center points. By means to the Euclidean distance, the algorithm assigns each data to its closest center point P_k , which is calculated by

$$P_k = \frac{1}{N_k} \cdot \sum_{i=1}^{N_k} x_i^k, \quad (3)$$

where x_i^k is the i-th data in the cluster k, and N_k is the number of data points in the respective cluster. The clustering center needs to be updated by computing the average value of every cluster until the value doesn't change any more. Then the initial data set is divided into k clusters in which the data have high similarity.

The process of clustering is as follows:

- *Step 1*: K objects are selected as the initial clustering centers from the data set which obtains N objects.
- *Step 2*: The distance between the data objects and clustering center is calculated and classifying them to the nearest cluster.
- *Step 3*: The mean value of the cluster is computed and used to update the clustering center.
- *Step 4*: The iterative method is applied in Step 2 and 3 until the value of the clustering center changes no more. If not, the process continues.

The K-means clustering is conducted on the data sets of S_1 and S_2 . The data series are cut up into M components, and the training set can be expressed as $D = d_1, d_2, \dots, d_M$, where D represents the training days. The K-means algorithm can select similar days from D as one class. Then several classes are produced.

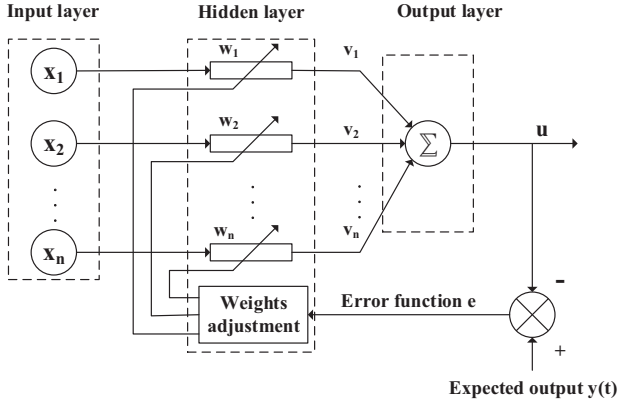


Fig. 2. The model of three-layers BPNN

According to Pearson correlation coefficient, the most related class is chose as the final input of the neural network.

C. Bagging neural network

Neural network (NN) has been one of the most effective data mining approaches for prediction. NN can deal with nonlinear problems well without establishing complex mathematical model. Back propagation neural network (BPNN) as one of the most common NNs is usually used as the forecasting algorithm. The basic BPNN consists of three layers: an input layer, a hidden layer and an output layer.

Fig. 2 shows the principle of the BPNN. The BPNN consists of two processes: forward propagation of data stream and back propagation of the error signal. In the process of forward propagation, the state of neurons in each layer only affect ones in the next layer. If the expected output could't be obtained in the output layer, the algorithm turns to the process of back propagation of the error signal. The gradient descent method is conducted on the weights vector space. It is needed to dynamically search for a set of weights vector and minimize the error function. As for the hidden layer, the neural numbers of this layer is usually $2M + 1$ according to the experience, where M represents the neural numbers of the input layer. However, different neural numbers have an effect on the results of the output layer so that the network is tested when $M = [2, 3, \dots, 10]$ to find the best forecasting result.

In Fig. 2, $[x_1, x_2, \dots, x_n]$ are the input variables, $[w_1, w_2, \dots, w_n]$ are weights between the input layer and the hidden layer and $[v_1, v_2, \dots, v_n]$ are weights between the hidden layer and the output layer. $Y(t)$ is the expected output. In addition, the general Sigmoid function is used as the transfer function to handle the nonlinear problem, which can be expressed as

$$y = f(x) = \frac{1}{1 + e^{-x}}. \quad (4)$$

The output of the hidden layer z can be formulated as

$$z = f_1 \left(\sum w_i x_i \right). \quad (5)$$

Similarly, the output of the out layer can be formulated as

$$u = f_2 \left(\sum v_k z_k \right). \quad (6)$$

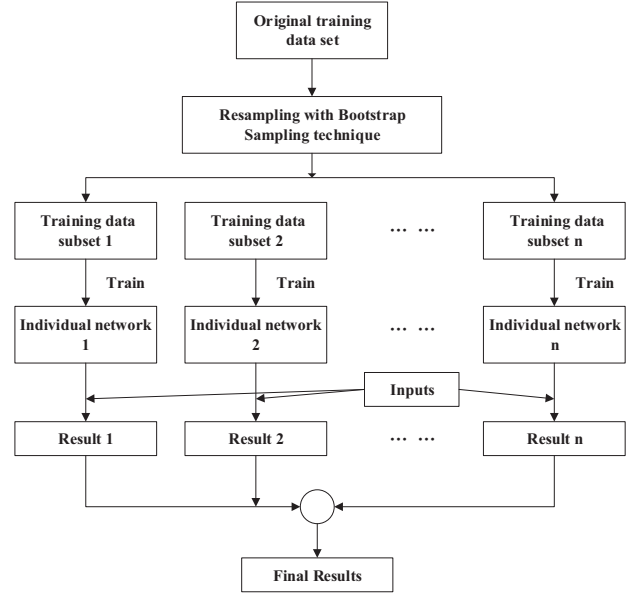


Fig. 3. The process of sampling and training of the bagging algorithm

Thus, the process of forward propagation is completed. Then, the error signal e is produced by the function of u and $Y(t)$, which is calculated as

$$e = \frac{1}{2} \sum (y_i(t) - u_i)^2. \quad (7)$$

The gradient descent method is utilized to modify w_i and minimize the error function e . This method needs to calculate the partial derivatives to update the weights of w_i and v_i . It is incessantly required to be repeated until the value of e equals to zero, which means that the values of u and $Y(t)$ are identical. In this process, several weights need to be modified, which consumes a lot of time.

The BPNN can obtain good results compared with other algorithms such as linear aggression and support vector machine (SVM) when dealing with large amount of nonlinear samples. However, it is easy to trap into the local optimization and the forecasting result would be instable when the training samples change little. Besides, the over fitting problem would arise for the BPNN.

In order to improve the problems above, ensemble learning is applied to optimize the BPNN. Ensemble learning can combine several learning modules to enhance the stability and forecasting accuracy of the model. Bootstrap aggregating (bagging) algorithm as a kind of ensemble learning methods is used to improve the performance of the BPNN. It is verified that the bagging-based approach is effective for instable learning algorithms.

The bagging algorithm can be effective on the premise that a weak learner is applied to the training samples. Because of the poor accuracy of the weak learner, it is needed to use the learner several times on the training samples. Then a sequence of forecasting functions are produced and the most accurate function is determined by voting.

Fig. 3 shows the process of the bagging algorithm. It can be seen that the original training data set is sampled again by Bootstrap sampling technique and the learning algorithm is

applied for the subsets. Suppose $T(x)$ is a classifier, such as a tree, producing a predicted class label at input point x . To bag T , we draw bootstrap samples $\{(x_i^*, y_i^*)\}_1^B, \dots, \{(x_i^*, y_i^*)\}_B^B$ each of size n with replacement from the training data. Then the results can be expressed as follows:

$$C_{bag}(x) = \text{MajorityVote} \{T^{*b}(x)\}_{b=1}^B \quad (8)$$

The output $C_{bag}(x)$ is the optimal classifier, which is determined by voting. In this paper, the bagging-based method is used for prediction. The key factor that bagging algorithm can improve the forecasting accuracy is the stability of the learning algorithm. Firstly, bagging algorithm can select K sets with n samples randomly from the initial training set. Secondly, BPNN is utilized to train the subsets for several times. Then a sequence consisting of the forecasting models h_1, h_2, \dots, h_K is produced. Finally, the chosen forecasting model H is determined by calculating the average value of K models. In particular, all the forecasting models have the same importance.

Applying the BPNN based on the bagging algorithm can not only improve the forecasting accuracy but also reduce the time costs.

III. SIMULATION RESULTS

In this section, the proposed forecasting approach is simulated and compared to other approaches. The 10-min data obtained from the Supervisory Control And Data Acquisition system (SCADA) of a wind farm is adopted as research data. The six numbers for an hour are sum to obtain hourly data, which could avoid that 10-min data fluctuated greatly so that the forecasting accuracy could be improved.

In order to evaluate the performance, two kinds of error calculation methods: 1) root mean squared error (RMSE) and 2) mean absolute error (MAE) are proposed:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2} \quad (9)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i| \quad (10)$$

In the above equations, y_i is the actual power, y'_i is the predictive power, N is the number of the test data.

A. Data preparing

In this paper, the measured data from September 29th to September 30th for a total of 24 hours are selected as test data. Data in three months before this period of time are used as training samples for clustering. Fig. 4 shows the wind power generation of all the year. It can be observed that the power generation is much different before and after July. Even in the same season, the power curve varies dramatically, which means that it is necessary to apply the clustering method. There are 39 variables collected by the SCADA system. Relief algorithm is conducted on the vector space to select proper variables as the inputs of neural networks. The results of Relief

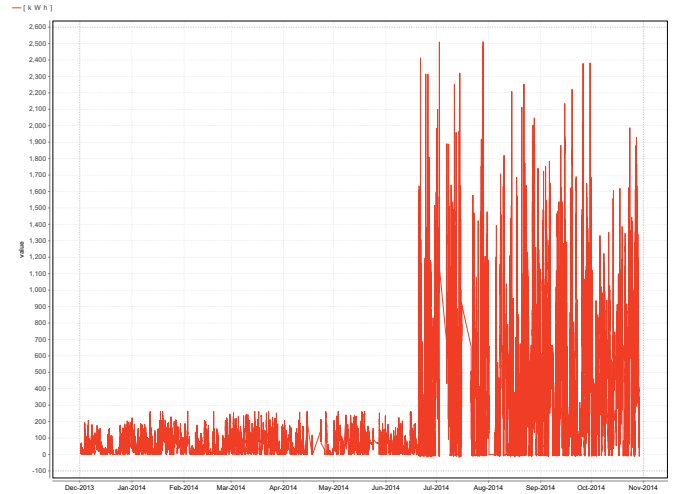


Fig. 4. The wind power generation curve from Dec. 2013 to Nov. 2014

TABLE I
THE RESULTS OF RELIEF ALGORITHM

Variable	Weight
Average wind speed per second	0.052
Machine side semiconductor temperature	0.022
Network side semiconductor temperature	0.016
The maximum angle of blade 1	0.013
The angle of blade 1	0.003
Generator bearing temperature	0.003
The ambient temperature	0.002
Rotor winding temperature of generator	0.002

algorithm have been shown in Table I, in which the variables whose weight is large than 0.01 are listed. According to the results, the first four variables have better weights. Besides, the ambient temperature is also significant. Then, the whole input variables are selected.

Table I illustrates that wind speed is the most important variable in wind power forecasting. Although temperature from the wind turbine units plays an important role according to Relief algorithm, it can not be utilized as the forecasting variable for the reason that it is unable to obtain the temperature of machines in the forecasting day. The generator bearing temperature can not be used for the same reason. Therefore, we select average wind speed per second, the angle of blade 1 and the ambient temperature as the input variables, where the angle of blade represents the wind direction. Besides, the value of the angle of blade is decomposed into sine and cosine values.

Fig. 5 shows the relationship between wind speed and power. It can be seen that when actual wind speed is less than cut-in wind speed, and the wind turbines do not operate and the wind power is zero. When the actual wind speed is larger than cut-in wind speed, the value of wind power is increasing as the wind speed growing until the wind speed is less than cut-out wind speed. When the actual wind speed is larger than cut-out wind speed, wind turbines would operate in the state of rated power. It is obvious that the relationship between them is nonlinear so that applying the neural network is necessary.

According to the described clustering approach above, it is

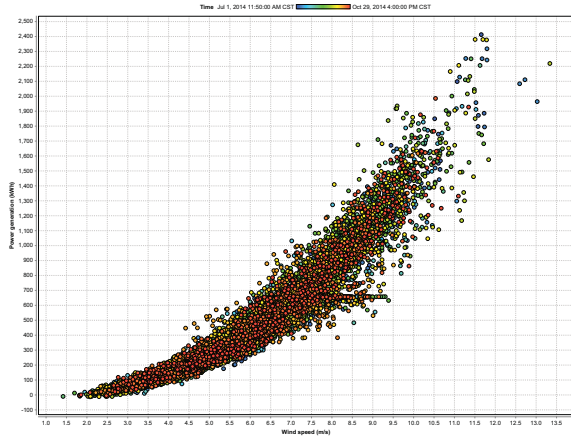


Fig. 5. Strong correlation between wind power generation and wind speed

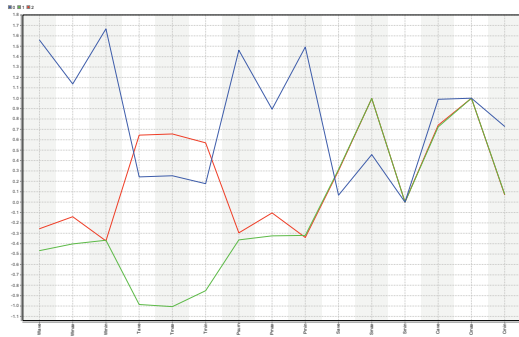


Fig. 6. The K-means clustering results with meteorological conditions and historical power

needed to divide all the training samples into pieces and regard one day as the training unit, which consists of 144 items. The average, maximum and minimum values of wind speed and temperature are calculated as the clustering features. Then the sum, maximum and minimum of the power generation are also collected for clustering. For the forecast period, the power generation before this period is regarded as having the same features of power generation with the forecasting period. Then, the data are clustered by the K-means clustering.

Fig. 6 shows the clustering result, and three lines mean that the initial samples are classified into 3 categories. It can be seen that the three categories are much different in wind speed, temperature and wind power, which are the main factors of prediction. In this process, the blue line is represented by $k = 0$, the green line is for the case of $k = 1$, and the red line is for $k = 2$. The initial samples have different features, and the K-means method can classify the samples which have similar features into one category.

The distance between the forecasting day and the 3 categories is calculated according to the Pearson correlation coefficient:

$$P = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (11)$$

where X-axis and Y-axis represent the forecasting sample and the 3 categories, respectively. The most related category is chosen and there are 29 days totally adopted as the training

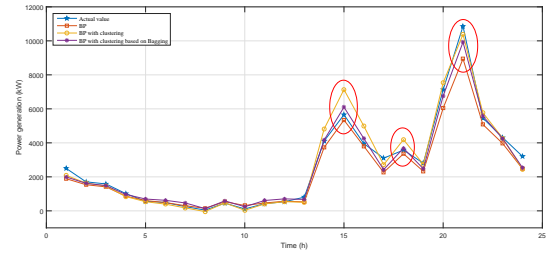


Fig. 7. The 24-hours forecasting results of BPNN, BPNN with clustering and the proposed approach

TABLE II
THE PERFORMANCE OF BASELINE APPROACHES AND THE PROPOSED APPROACH

Forecasting approach	RMSE(kW)	MAE(kW)
BP neural network	558.098	365.091
BPNN with clustering	487.718	323.596
Bagging-BPNN with clustering	342.548	255.156

data. In order to compare the performance of clustering with non-clustering, a data set containing 29 days from September 1th to September 29th is used as the training set.

After the feature selection based on Relief algorithm, the training data consisting of the wind speed, the blade angle, and the temperature, are used as the input variables to construct a BPNN model based on the bagging algorithm.

B. Forecasting results

To better verify the effectiveness of the proposed approach, this paper compares three methods of predicting power generation: the BPNN without clustering pretreatment, the BPNN with clustering pretreatment and the proposed approach.

Fig. 7 shows the comparison of the three approaches. It is obvious that the purple line is most similar to the actual wind power value. In particular, the peak values marked on Fig. 7 shows that the proposed approach has better performance compared with other approaches.

From the prediction results in Table II, it can be seen that the BPNN with clustering can reduce the forecasting error compared with the BPNN, in which RMSE and MAE are decreased by 12.7% and 11.5%, respectively. Furthermore, the proposal can obtain the best performance among these three approaches, which can decrease 38.7% and 29.8% for RMSE compared with the BPNN, and 30.1% and 21.1% for MAE compared with the BPNN with clustering. The simulation results verify that the bagging neural network with the similar meteorological conditions and historical power clustering approach can obtain better forecasting accuracy.

Besides, the Approach 1 in [17], the Approach 2 in [22], and the Approach 3 in [23] have been simulated and the results are listed in Table III. It can be seen that the proposed approach has the best forecasting accuracy among these approaches. The Approach 1 has the worst results because the different data has different characteristics, and it may be not applicable to this case. The Approach 2 uses the ANN ensemble based on PCA to predict the wind power. The proposal decreases 19.9% and 17.1% for RMSE and MAE compared with the Approach 2,

TABLE III
THE PERFORMANCE OF THREE METHODS AND THE PROPOSED METHOD

Forecasting approach	RMSE (kW)	MAE (kW)
Approach 1	1053.736	956.211
Approach 2	427.912	307.379
Approach 3	389.472	279.528
Bagging-BPNN with clustering	342.548	255.156

respectively. The Approach 3 uses the artificial neural network as the forecasting approach and it has better performances than the Approach 1 and 2, but is worse than the proposal. As a result, the proposal can obtain best forecasting accuracy than the other baselines.

IV. CONCLUSION

In this paper, a data mining approach for wind power forecasting has been proposed, which consists of the K-means clustering method and bagging neural network. The historical data are clustered according to the meteorological conditions and historical power. Pearson correlation coefficient is used to calculate the distance between the forecasting day and the clusters. The input variables of the neural network are selected by Relief algorithm to reduce the complexity and the Bagging algorithm is applied to optimize the stability and accuracy of the BPNN. To demonstrate the effectiveness, the proposed approach has been tested according to the actual data in the practical wind farm. The RMSE and MAE results show that the proposals have significant gains.

Although the proposals are not specially designed for the individual wind turbine, the idea of clustering is still important and effective when a large-scale wind farm is built. Particularly, in a wind farm, the location of wind turbines may lie in one direction, then the wind speed of these wind turbines can be classified into one category. With this way, the proposal can be extended and widely used in all real wind farm, which not only increases the forecasting accuracy, but also reduces the computational complexity.

To improve the forecasting accuracy, the effective meteorological forecasting should be researched, and the corresponding optimal method for the BPNN should be designed. Besides, the multi-dimensional clustering problem should be formed and the wind power forecasting model for wind farms should be researched.

REFERENCES

- [1] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renew. Sust. Energ. Rev.*, vol. 31, no. 2, pp. 762-777, Jan. 2014.
- [2] D. Carvalho, A. Rocha, M. Gomez-Gesteira, and C. Santos, "A sensitivity study of the WRF model in wind simulation for an area of high wind energy," *Environ. Model. Softw.*, vol. 33, no. 7, pp. 23-34, Mar. 2012.
- [3] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1-8, Jan. 2012.
- [4] Z. Yin, Z. Chen, D. Yin, and Q. Li, "A new combination model for short-term wind power prediction," in *Proc. IEEE DRPT*, Changsha, China, Nov. 2015, pp. 1869-1873.
- [5] Q. Xu, D. He, N. Zhang, C. K. Q. Xia, J. Bai, and J. Huang, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1283-1291, Oct. 2015.
- [6] M. Khalid and A. Savkin, "Closure to discussion on 'a method for short-term wind power prediction with multiple observation points'," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1898-1899, May. 2013.
- [7] J. Tastu, P. Pinson, P. J. Trombe, and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 480-489, Jan. 2014.
- [8] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033-1044, May. 2014.
- [9] X. Zhang, J. Zhang, Y. Li, and R. Zhang, "Short-term forecasting of wind speed based on recursive least squares," in *Proc. IEEE ICECE*, Wuhan, China, Oct. 2011, pp. 367-370.
- [10] M. S. Miranda and R. W. Dunn, "One-hour-ahead wind speed prediction using a bayesian methodology," in *Proc. IEEE Power Engineering Society General Meeting*, Montreal, Canada, Oct. 2006, pp. 1-6.
- [11] J. L. Torres, A. Garcia, M. D. Blas, and A. D. Francisco, "Forecast of hourly average wind speed with ARMA models in Spain," *Solar Energy*, vol. 79, no. 1, pp. 65-77, Jul. 2005.
- [12] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Trans. Energ. Convers.*, vol. 24, no. 1, pp. 125-136, Mar. 2009.
- [13] I. Colak, S. Sagiroglu, M. Yesibudak, E. Kabalci and H. I. Bulbul, "Multi-time series and -time scale modeling for wind speed and wind power forecasting part I: Statistical methods, very short-term and short-term applications," in *Proc. IEEE ICRERA*, Paris, France, Nov. 2015, pp. 209-214.
- [14] Y. Tian, Q. Liu, Z. Hu, and Y. Liao, "Wind speed forecasting based on time series - adaptive kalman filtering algorithm," in *Proc. IEEE FENDT*, Chengdu, China, Jun. 2014, pp. 315-319.
- [15] K. Bhaskar and S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Trans. Sustain. Energy*, vol. 3, no. 2, pp. 306-315, Apr. 2012.
- [16] S. Li, P. Wang, and L. Goel, "Wind power forecasting using neural network ensembles with feature selection," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1447-1456, Jul. 2015.
- [17] X. Zhang, R. Wang, T. Liao, T. Zhang, and Y. Zha, "Short-term forecasting of wind power generation based on the similar day and elman neural network," in *Proc. IEEE SSCI*, Cape Town, South Africa, Dec. 2015, pp. 647-650.
- [18] C. Hyuk-Gyu, H. Park, and H. Kwon, "Similarity measurement among sectors using extended Relief-F algorithm for disk recovery," in *Proc. IEEE ICCIT*, Busan, South Korea, Nov. 2008, pp. 790-795.
- [19] L. Ma, S. Y. Luan, C. W. Jiang, H. L. Liu, and Y. Zhang, "A review on the forecasting of wind speed and generated power," *Renew. Sust. Energy Rev.*, vol. 13, no. 4, pp. 915-920, May. 2009.
- [20] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal," *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 50-59, Jan. 2011.
- [21] E. Erdem, J. Shi, and Y. Peng, "Short-term forecasting of wind speed and power - a clustering approach," in *Proc. ISERC*, Toronto, Canada, May. 2014, pp. 1-8.
- [22] D. He and R. Liu, "Ultra-short-term wind power prediction using ANN ensemble based on PCA," in *Proc. IEEE IPEMC*, Harbin, China, Aug. 2012, pp. 2108-2112.
- [23] A. S. Kumar, T. Cermak, and S. Misak, "Short -term wind power plant predicting with Artificial Neural Network," in *Proc. IEEE EPE*, Kouty nad Desnou, Czech Republic, Jul. 2015, pp. 584-588.
- [24] P. Kunzemann, G. Jacobs, and R. Schelenz, "Application of CPS Within Wind Energy-Current Implementation and Future Potential," in *Industrial Internet of Things*, Switzerland: Springer, 2017, pp.647-670.



Wenbin Wu received his bachelor degree in Telecommunications Engineering with Management from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. He is currently a Ph.D. student in the Key Laboratory of Universal Wireless Communications (Ministry of Education) at BUPT. His main research interests include wind power forecasting and non-orthogonal multiple access for cloud radio access network.



Mugen Peng (M'05-SM'11) received the B.E. degree in electronics engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000, and the Ph.D. degree in communication and information systems from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2005. Afterward, he joined BUPT, where he has been a Full Professor with the School of Information and Communication Engineering since 2012. In 2014, he was an Academic Visiting Fellow with Princeton University,

Princeton, NJ, USA. He leads a Research Group focusing on wireless transmission and networking technologies with the Key Laboratory of Universal Wireless Communications (Ministry of Education), BUPT. His main research areas include wireless communication theory, radio signal processing, and convex optimizations, with a particular interests in cooperative communication, selforganization networking, heterogeneous networking, cloud communication, and internet of things. He has authored/coauthored over 60 refereed IEEE journal papers and over 200 conference proceeding papers.

Dr. Peng was a recipient of the 2014 IEEE ComSoc AP Outstanding Young Researcher Award, and the best paper award in IEEE WCNC 2015, WASA 2015, GameNets 2014, IEEE CIT 2014, ICCTA 2011, IC-BNMT 2010, and IET CCWMC 2009. He received the First Grade Award of the Technological Invention Award in the Ministry of Education of China, and the First Grade Award of Technological Invention Award from the China Institute of Communications. He is on the Editorial/Associate Editorial Board of the *IEEE Communications Magazine*, *IEEE Access*, *IEEE Internet of Things Journal*, *IET Communications*, and *China Communications*. He has been the guest leading editor for the special issues in the *IEEE Wireless Communications*. He is the Fellow of IET.