Short Term Forecasting for Wind Power Based on Cluster Analysis

Yang Gao[1], Jing Xing[1*], Aoran Xu[1], Liu Zhang[1], Gang Wang[1], Quanping Zou[1]

[1]*Electric power college, Shenyang Institute of Engineering, Pu Chang Str.18 of Shenbei New Area. Shenyang City, Liaoning Province, China*

Received 26 November 2014, www.cmnt.lv

Abstract

In order to make full use of historical wind speed information behind the data, according to daily similarity of the wind speed and wind power, short term power forecasting method based on cluster analysis is presented in this paper. Through the original sample data is preprocessed, election history daily data that is similar with characteristic parameters of NWP of forecast day, so as to establish training samples of model. NWP information of forecast day provided by Meteorological Department will be as the characteristic parameters of forecast day, and calculating Euclidean distance between characteristic parameters will be regarded as a basis of similarity measure. Finally forecasting model is founded by adopting similar samples based on cluster. Using NWP data as input parameters, the actual wind power as a target value, many kinds of short-term wind power forecasting model is gained by training. Through the actual wind farm test, forecasting accuracy is improved obviously.

Keywords: Short-term wind power forecasting; cluster analysis; K means cluster algorithm; daily similarity; Numerical Weather Prediction (NWP).

1 Introduction

Having the obvious characteristic difference, wind speed also has certain rule with periodic change at the same time from sunrise to sunset every day, it is found that the wind speed has existed certain daily similarity characteristics [1]. In order to make full use of historical wind speed information behind the data, according to daily similarity of the wind speed and wind power, short term power forecasting method based on cluster analysis is presented, point of departure lies in the important role of sample pretreatment to improve the forecasting results, the ultimate goal is to improve the forecasting accuracy.

Looking for the basis of cluster is the key of wind power forecasting on adopting cluster method, i.e. obtain basis of different types of samples divided. At present, both at home and abroad, research of short-term wind power forecasting and forecasting accuracy improvement method is also more, but cluster method with a similar day is treated as before forecasting, to improve the forecasting accuracy of research method is not arisen horizon. A similar day on power change trend of wind power, exist some common features, and is reflected on basic consistency with through the NWP change trend of information.

The method proposed in this paper is applying the cluster analysis method to the wind power forecasting, based on the original sample data pretreatment, extracting history daily data that is similar with characteristic parameters of NWP of forecast day as the modeling of training samples. Regarding NWP information of forecast day as characteristic parameters of forecasting day, determine various data of similar day by computing the Euclidean distance of characteristic parameters that are obtained, finally using similar sample of cluster establish power forecasting model. The NWP data as the input parameters of the model, the data of actual wind power as the target value of the model, through the model training can get short term wind power multi-step forecasting model.

2 Daily similarly of wind power

The size of the actual output power of wind turbine depends mainly on the local wind energy resources, and the feature of wind resources mainly refers to the change of wind speed. Wind speed that is distance in unit time air moving in the horizontal direction, is mainly influenced by weather factors and terrain, surface obstacle factors etc. The alternative rotation of day and night for the earth showed some degree of similarity on the weather of some day, so wind power changing trend of the weather similar day is very similar.

Figure 1 is curves of the Yilan wind farm in 2013 January to June, the wind changes of some day can be seen, the trend has some similarity, wind power for corresponding also showed a similar tendency, as shown in Figure 2. Because of the existence of a certain change trend of wind power with the weather that day, for a wind farm, wind power can be used on the similarity relation of weather information similar to judge. Meteorological information can be obtained by forecasting mode, according to the similarity search similar sample of forecast time interval, and regarded as training samples of a short-term forecasting model. Cluster analysis is an effective way to solve searching the similar samples.

***Corresponding author's** e-mail:309904066@qq.com

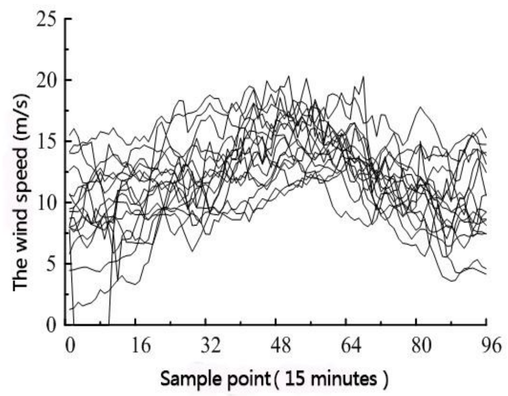

FIGURE 1 Wind speed change curve.

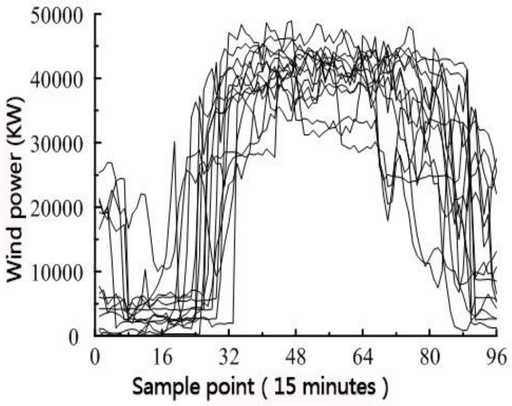

FIGURE 2 Wind power part, power trend curves.

3 Cluster analysis on similar day

Cluster analysis is the statistical analysis technique which will make the research object dividing into different categories according to certain metric standard[2]. Cluster belongs to unsupervised classification method, and is a kind of exploratory analysis, there is no pre-specified category. In the field of machine learning, cluster is unsupervised learning. Cluster analysis has been widely applied in many fields, including pattern recognition, data analysis, image processing, and market research. Both in terms of business, networking, and biological applications have played an important role[3].

3.1 THE BASIC STEPS OF CLUSTER ANALYSIS

Using cluster method generally includes the following four steps[4][5][6]:

(1) Find out the characteristics of classification object;

In general, object of the research data contains multiple characteristics, among them the representative characteristics can help to identify data object itself, this kind of characteristic attribute is the standard of distinguish between different data object, so when performing the cluster analysis, the first thing is extracting the features which make the object different from other data.

(2) To determine the similarity of classification object;

Cluster is that data object of the similar characteristic attribute is divided into a category, and different characteristic attribute is divided into another category. Therefore, the similarity of measuring data need have a metric standard. Cluster results obtained in different standards may be different. About the similarity measure is described in Section 3.3.

(3) To propose the steps or cluster algorithm that can achieve classification;

One of the key steps of the cluster process is to use different methods for dividing the data object. The main cluster analysis method includes the partitioning cluster analysis method and the hierarchical cluster analysis method. Then the hard cluster and fuzzy cluster are two more commonly techniques in dividing methods.

(4) To evaluate the cluster results;

That to evaluate the quality of the cluster result is good or bad is no objective standard. You can choose more realistic cluster results according to the actual situation.

3.2 THE METHODS OF CLUSTER ANALYSIS

Traditional cluster algorithms include five categories: partitioning method, hierarchical method, density-based method, grid-based method and model-based method [7][8][9]. Partitioning and hierarchical methods are based on the distance to judge the similarity, and the density-based approach is based on the concept of density to judge the similarity. Modern cluster algorithms include high-dimensional cluster analysis method and dynamic cluster algorithms, two types of algorithms.

K-means cluster method is the most classic one among the dynamic cluster algorithm, the basic idea is to divide each sample to the nearest category from the mean, which is based on distance as the standard for cluster.

K-means cluster algorithm generally includes the following process steps:

(1) All data will be divided into K initial classes, then selecting K sample points as the initial cluster centers, which are denoted by $z_1(l), z_2(l), ..., z_k(l)$,and the initial value is $l=1$;

(2) According to the nearest neighbor rule all samples will be assigned to $\omega_j(K)$ in K class which is represented by each cluster center. The number of samples included in various types is $N_j(l)$;

(3) Calculate all kinds of mean vector, and the vector as a new cluster center:

$$z_j(l+1) = \frac{1}{N_j(l)} \sum x(i) \ . \quad (1)$$

Among $j = 1, 2, ..., k$, $i = 1, 2, ..., N_j(l)$.

(4) If $z_j(l+1) \neq z_j(l)$, it is said that the cluster results are not the best. Then returns to step (2) to continue the iterative calculation.

(5) If $z_j(l+1) = z_j(l)$, iterative process ends, at this time of the cluster result is that the optimal cluster results.

3.3 MEASURE OF SIMILARITY

Before the starting to group the samples, the measure of similarity must be defined. Each sample must be compared with the other samples, then the very "similar" samples are placed in the same class, the samples which are not similar are placed in a different class. Similarity between the two elements can be used a variety of different methods to measure, there are mainly distance measure, correlation measure and a measure of the amount of information or the like.

The common distance include Minkowski Distance, Mahalanobis Distance and Canberra Distance.

This paper uses the Euclidean distance measuring the similarity degree between the different trend of wind speed in different days. Every day will be as a data object, by a 7 d vector said, called the day NWP vector, expressed as:

X = [Pav, Vmin, Vmax, Tmin, Tmax, Dsin, Dcos],

which variables in turn is the daily pressure average, daily minimum wind speed, daily wind speed maximum, minimum daily temperatures, maximum daily temperature, the daily wind sine average, the daily wind cosine average. Due to differrent dimensions of the each component in the day NWP vector, it needs for normalization processing, atmospheric pressure, wind speed and temperature were divided by their respective historical maximum, the sine and cosine values of the wind direction are normalized values without further processing.

Distance is defined as:

$$d_i = (\sum_{k=1}^{7}(x_m(k)\text{-}x_i(k))^2)^{1/2} . \tag{2}$$

Among them, d_i is Euclidean distance between forecast day and historical samples i. x_m is the day NWP vector of forecast day, x_i is the day NWP vector of historical data, i = 1,2, ..., n, n is the number of samples.

4 Short-term wind power forecasting model based on cluster analysis

4.1 MODEL STRUCTURE

Figure 3 shows a schematic diagram of the structure of short-term wind power forecasting model based on cluster analysis. First, according to the K-means cluster algorithm, historical samples are divided into K classes by the NWP vector closest principle, using the computer programming to search automatically classification which forecast day belongs to, and the sample data for each category as training samples which using neural network build forecasting model.

When the model is trained, we use NWP pressure, NWP wind speed, NWP temperature, NWP wind direction sine and NWP wind direction cosine, etc of the classification samples which forecast day belongs to as input, and target value of model is regarded as the output power of the wind farm. When using this model to predict, we can obtain predictive value of wind power by simply inputting the model forecasting NWP date information.

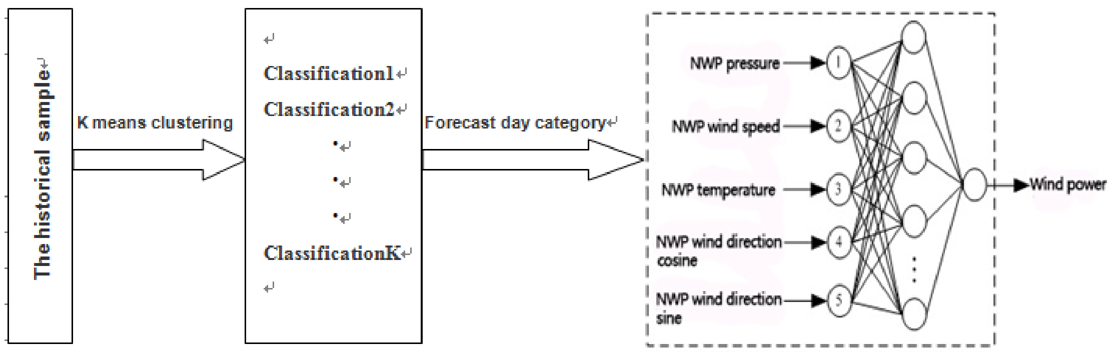

FIGURE 3 Short-term wind power forecasting model based on cluster analysis.

4.2 CLUSTER CALCULATION AND ANALYSIS OF RESULTS

Using the NWP data in Yilan wind farms of Heilongjiang in January to February, 2013 and measured wind power data get on analysis, modeling and forecast. Data resolution is 15 minutes. Selecting the February 4, 2013 as forecast day, forecasting steps is 96 points.

Selecting 20 days of before historical data February 4, 2013 to do cluster analysis, using K-means cluster algorithm, then criterion function and classification K's curves is shown in Figure 4, and according to the method how to determine the optimal classification number as 2.2 described, take the point K of Criterion function curve inflection as the best classification number, and get K = 3.

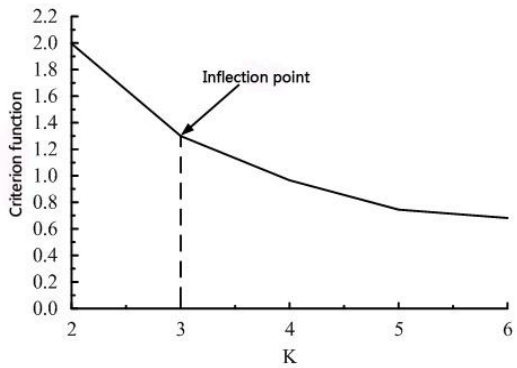

FIGURE 4 Criterion function and the classification number K diagram.

In the case of classification number K = 3, 20 historical samples which belong to categories is shown in Table 1. Among them, four days belong to category 3, one day belong to category 2, the others all belongs to category 1. By the formula (1) we can calculate these three cluster centers (normalized),which are follows:
Category 1:[0.988 0.183 0.438 -1.130 -0.804 0.042 0.051]
Category 2: [0.988 0.555 0.863 -1.151 -0.853 0.119 0.189]
Category 3: [0.9930.047 0.268-0.856-0.551 -0.020-0.125]

Predicting the day and normalizing NWP date vector on February 4, 2013 is

[0.981 0.3400.801 -0.932-0.5790.113 -0.052],

by the formula (2) the Euclidean distance of the three types of cluster centers are 0.51, 0.48 and 0.63 respectively, from Category 2 is the nearest cluster center, so classification which forecast day belongs to is Category 2, the first 19 samples day belong to Category 2 samples, February 2, 2013.

TABLE 1 The table of sample cluster

Sample date	1	2	3	4	5	6	7	8	9	10
Category	1	3	1	3	3	3	1	1	1	1
Sample date	11	12	13	14	15	16	17	18	19	20
Category	1	1	1	1	1	1	1	1	2	1

4.3 CLUSTER RESULTS FOR SHORT-TERM FORECASTING

The NWP air pressure, wind speed, air temperature, wind direction sine and wind direction cosine on February 2, 2013 as input and measured power data as output set up model, and data resolution is 15 minutes, and the training sample data number is 96, and the neural network structure uses GRNN, and the window width parameter value is 0.15 s, the model structure shown in Figure 3.

After the model training is completed, inputting NWP Pressure, NWP wind speed, NWP temperature, NWP wind direction sine and NWP wind direction cosine of the forecast day on February 4, 2013 can get wind power forecasting value. Forecasting error NMAE and NRMSE were 10.67% and 14.01% respectively.

In order to verify the nature of this method, the neural network model based on cluster analysis will be compared with the continuous model, then we can obtain contrast curve of predicted results which shown in Figure 5. Power forecasting error situation is shown in Table 2.

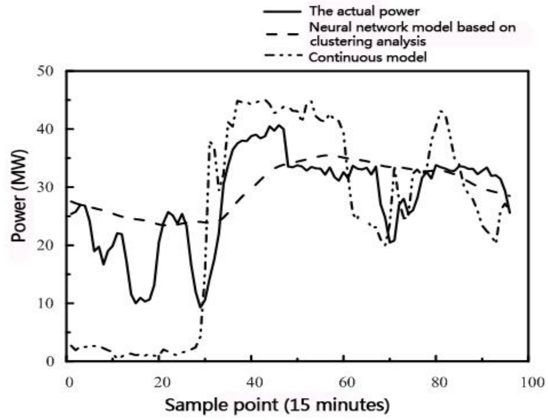

FIGURE 5 The model of power forecasting curve comparison.

TABLE 2 Power forecasting error.

Wind power forecasting model	*NMAE*/%	*NRMSE*/%
Based on Cluster analysis Neural network model	10.67	14.01
Continuous model	22.12	26.81

5 Conclusion

Because of the different wind speed characteristics of the four seasons in a year, and daily wind speed also has certain rules with the periodic change of sunrise and sunset, in order to make full use of historical wind speed information behind the data, this paper analyzes daily similarity of wind power, and the corresponding meteorological data is also similar, meteorological data is available to obtain the forecast value in advance, which provides a way for short-term forecasting of wind power for using the similarity. After the actual wind farm test, verifying the validity and progressiveness of the proposed method in this paper, the specific conclusions are as follows:

(1) Processing properly before wind power forecasting model can improve the forecasting accuracy of short-term wind power effectively;

(2) By cluster analysis method, analyzing the NWP data can classify the data of different weather types effectively;

(3) Connecting cluster analysis method and neural network can establish effective short-term wind power forecasting model and improve the forecasting accuracy.

References

[1] Zhu F, Sun H, and Zhou W. Predicting Wind Speed Based on Neural Network of Clustering Similar Days. New energy power generation ,Twenty-fourth Annual Conference Proceedings of the Chinese Society of Universities for Electric Power System and its Automation. pp 2627-2630.

[2] Cluster Analysis-Data Mining-Computer Encyclopedia com. Computer science, computer, software, hardware, networking, digital, IT Encyclopedia [DB/OL].http://www.jsjbk.cn/index.php?doc. -view-8293.

[3] Li Y . 2010 Hebei Science and Technology Resource Allocation Situation Assessmentand Integration Strategies. Yanshan University: Yanshan.

[4] Li Y F. 2009 Application and Research of Image processing techniques in hardness measuring system. Nanjing University of Technology: Nanjing.

[5] Huang S S. 2011 Design implemented and application of Nanjing telecommunication network user behavior analysis system. Nanjing University of Technology: Nanjing.

[6] Gu J S. 2013 The application of data mining technology in the life insurance customer. South China University of Technology: Guangzhou.
[7] Liu Y. 2009 The integrated production of heuristic-based methods - research of distribution scheduling. Tianjin University: Tianjin.
[8] Luo Y J. 2010 The study of anomaly detection algorithm based on anomaly detection data stream. Harbin Engineering University: Harbin.
[9] Wang Y. 2009 Application of data mining in the gas system. Nanjing University: Nanjing.

Authors

Yang Gao, 1974.07. 3, Shenyang city of Liaoning province

Current position, grades: Associate professor, subdecanal of the institute of electric power.
University studies: PhD degree in Power System and Automation from Shenyang Agricultural University, Shenyang, China, in 2012.
Scientific interest: Microgrid power system analysis; interests include fault diagnosis and uncertain theory.

Jing Xing, 1989.05.02, Yanji city of Jilin province

Current position, grades: Bachelor's degree from changchun institute of technology, Changchun, China, in 2013.
University studies: power system analysis
Scientific interest: power system analysis; New energy direction

Aoran Xu, 1983.07. 05, Tonghua city of Jilin province

Current position, grades: Instructor
University studies: M.S degree in Power System and Automation from Beijing Jiaotong University, Beijing, China, in 2008. Study on Micro-grid Power Storage Converter.
Scientific interest: Power system analysis; Rural power; New energy direction; fault diagnosis and uncertain theory

Zhang Liu, 1981.02. 20, Anshan city of Liaoning province

Current position, grades: Instructor
University studies: M.S degree in Power System and Automation from Northeast Dianli University, Jili, China, in 2008.
Scientific interest: Power System and Automation; Power system analysis; Data set for power performance measurement.

Gang Wang, 1980.03. 15, Benxi city of Liaoning province

Current position, grades: Instructor
University studies: M.S degree in Power System and Automation from Northeast Electric Power University, Changchun, China, in 2008.
Scientific interest: Power System and Automation; power system analysis;

Quanping Zou, 1979.08. 26, Shenyang city of Liaoning province

Current position, grades: Instructor
University studies: M.S degree in Power System and Automation from Northeast Dianli University, Jilin, China, in 2004.
Scientific interest: Power System and Automation; power system analysis; Power flow calculation and control.