IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 8 (August. 2018), ||V (IV) || 32-39

Improving Ensemble Learning in Ultra Medium-Term Wind Power Prediction Based On Machine Learning Techniques

Dr.G.Anandharaj[1], S.Meena [2]
[1](*Assistant Professor, Department of Computer Science and Application, Adhiparasakthi college of arts and science (Autonomous), India) (younganand@gmail.com)*
[2](*M.Phil (CS) Research Scholar, Department of Computer Science and Application, Adhiparasakthi college of arts and science (Autonomous), India) (s.meenachitra@gmail.com)*
Corresponding Author: Dr.G.Anandharaj1

Abstract : This paper Discuss Medium term wind power prediction based wind speed, wind direction, air temperature and power utilizing wind turbine information collected at 72's intervals. A time-series model display way to deal wind behavior is studied. Both exponential smoothing and data-driven models are created for wind prediction. An overview of the ongoing improvements in machine learning techniques focused on prediction using Ensemble Learning. The machine learning techniques rapidly developed and Ensemble Learning techniques have been used for prediction. In this paper, to deal with the training samples dynamics and improve the forecasting accuracy, K-means clustering is utilized to classify the samples into several categories, which contain the data of meteorological conditions and historical power data. A data mining approach consisting of K-means clustering Uses to choose similar days Training sets. And compared various neural networks in WEKA tool is proposed for medium-term WPF. The technique is validated by wind power system data, and the forecast error is calculated and analyzed. Experimental results show that our proposed method has high accuracy, which provides reference to medium- term forecasting of wind power generation.

Keywords - wind power forecast, k-means clustering, ensemble learning, bagging neural network, WEKA Tool

---

Date of Submission: 06-08-2018 Date of acceptance: 23-08-2018

I. INTRODUCTION

Wind power generation has the advantages of safety, reliability, noiseless and low pollution, in recent years which has been rapid developed. Easily affected by the wind speed, weather, temperature and some random factors, the power of wind power generation has the characteristics of intermittence and volatility, uncontrolled, and so on. The large-scale distributed wind power generation integrated on grid has brought great challenges to security and stability of power grid operation, effective scheduling and other work. Therefore, accurately forecast the output of wind power generation system has become particularly important, and it is also a prerequisite for electric power dispatching, power grid planning, which has practical significance.

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 a 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. The physical approach uses detailed topological and meteorological descriptions to model the conditions at the location of the wind farm. 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. 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.

In this paper, the selection method of the similar day is proposed. The similar day of output power curve of wind power generation system has a good similarity. According to the similarity of meteorological information and history output power data, the output power of the solar medium term direct prediction model is established using Ensemble learning method and the prediction model can reflect the change trend of wind power.

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

1) For forecasting wind speed, time series method is a practical method. It has no restrictions on history data. Only single wind speed sequence is required. Considering the samples used in training would affect the precision of forecasting model, the statistic clustering technique is proposed to filter history data prior to modeling.
2) With the same data, setting average of wind speed and its maximum on the anticipated day as indicators for clustering analysis, and using Euclidean distance as the criterion of similarity, the dates which metrological characteristics is similar to the anticipated day are filtered.
3) With the filtered samples, another ARIMA model is constructed. Its daily average of relative errors has dropped to 22.16%. Compared with the result without clustering prior to modeling, the predicted precision has been improved.

II. WIND POWER FORECASTING

With the emerging significance of wind power, wind power forecasting (WPF) is an important tool to assist and efficiently address wind integration challenge, and considerable efforts have been made to develop more accurate wind power forecasts.

2.1 WPF Applications

–Allocation of reserves based on the expected wind power feed.
– Optimization of the scheduling of conventional power plants by functions such as economic dispatch etc.
– Optimization of the value of the produced electricity in the market. Such predictions are required by different types of end-users (utilities, TSOs, etc.) and for different functions such as unit commitment, economic dispatch, dynamic security assessment, participation in the electricity market, etc.
– In addition, even longer time scales (7 days) would be interesting for the maintenance planning of large power plant components, wind turbines or transmission lines.

2.2 Challenges: Forecasting & Scheduling

- The important challenges include scheduling, system control and dispatch; Reactive power supply and voltage control; Regulation and frequency response reserve; Energy imbalance service; operating synchronized reserve and supplemental reserve.
- Incorporation of Wind Power Forecasting (WPF) in real time power system day-to-day operational planning.
- A survey of grid operators worldwide** found near unanimous agreement that integrating a significant amount of wind will largely depend on the accuracy of wind power forecast.
- The world wide figures of day-ahead hourly load forecast errors are typically in the range of 1% to 3%.
- Errors in wind forecasts have the range of 15% to 20% mean absolute error (MAE) per wind plant.
- With available forecasts and through large-scale wind integration studies, system operation can be improved using day ahead wind power forecasts for unit commitment, thus reducing overall operating costs, unserved energy, and wind curtailment, while maintaining required levels of system reliability.

2.3 Forecast Terminology

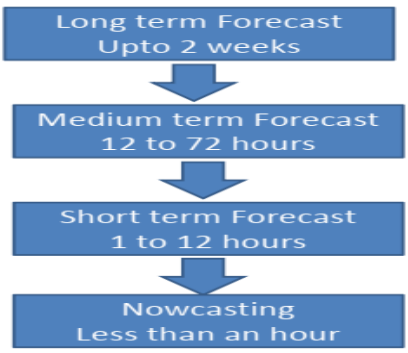

Fig 1 Forecasting Terminology

III. 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. 2 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.

3.1 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**.**

3.1.1 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.

3.1.2 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. 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, BESTFIRST Search algorithm can achieve the purpose of physical dimensions reduction compared with the principal component analysis (PCA).

3.1.3 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 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 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.

3.2 Similar Day Clustering

Due to the random changing of wind as well as the inconsistencies of training data, prediction of the dates often difficult to achieve the desired results on this training set to establish model which the variation of the wind characteristics is inconsistent. Proper selection and classification of modeling data can improve the similarity and consistency of the data, which is beneficial to improve the accuracy of the model. On the basis of the above, a new method of power generation based on similar day clustering is proposed.

At present, the main short-term wind power forecast demands a day for the unit, so the sample data clustered by day can match the training data and forecast results. In order to reflect the diurnal variation of light and temperature, the relevant physical quantities are selected, and a sample of similar days is constructed as follows:

$$x = [S_{max}, S_{avg}, S_{min}]$$

Where S_{max}, S_{avg}, S_{min} represents the maximum, mean and minimum values of the wind speed, respectively.

3.3 K-means algorithm

The clustering method used in this paper is the traditional K-means algorithm. The basic idea is to use K points in the space as the center of clustering then classifying the objects closest to them. By iteration, the value of each clustering center is updated until the best clustering results are got. Prior input of N data objects are partitioned into K clusters so that they obtained clustering is consistent with that. The objects in same cluster have high similarity, and objects in different clusters have smaller similarity. Similarity clustering is calculated by a center of each cluster obtained by the mean value of the objects.

Suppose the sample set is divided into C classes, the clustering steps are described as follows:

- C cluster center is initialized by properly selecting from the N sample.
- In the K iteration, for any sample, the distance from the sample to the C center is calculated, and then this sample is classified to the nearest center.
- Update the central values of the class using the mean value.
- For all C clustering centers, if use the iterative method of second and three, the value remains unchanged, then end the iteration, otherwise continue.

3.4 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. 3 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 affects ones in the next layer. If the expected output couldn’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 $M = [2.3...10]$ to find the best forecasting result.

IV. EXPERIMENT RESULT

4.1 The Explorer Interface of WEKA

In WEKA application issue, this is probably the most confusing part of becoming familiar with WEKA because presented screen is quite complex. Initially “preprocess” will have been selected when commanding WEKA where to find the data set to be used. WEKA processes data sets that are in its own ARFF format. Conveniently, the download will have set up a folder within the WEKA-3.8.1 folder called “data”. This contains a selection of data files in ARFF format

4.2 ARFF format files

There is no need to know about ARFF format to convert data from other formats. However, it is useful to see the information that such files provide to WEKA.

```
@attribute 'Date/Time (UTC)'
@attribute Temperature numeric
@attribute 'Wind Speed ' numeric
@attribute 'Wind Direction ' numeric
@attribute Hr01 numeric
@attribute Hr24 numeric
@attribute Hr48 numeric
@attribute Hr72 numeric

@data
'12/16/2017 13:00',31.23,17.48,237.62,868,440,1570,749
'12/16/2017 14:00',30.49,21.09,242.56,691,434,1529,766
'12/16/2017 15:00',30.03,22.55,245.48,592,474,1522,983
'12/16/2017 16:00',29.52,23.22,251,486,561,1561,1218
'12/16/2017 17:00',29.26,22.45,254.18,424,622,1566,1436
'12/16/2017 18:00',28.9,22.18,256.87,265,704,1507,1655
```

Fig 2 ARFF file format for Final dataset in this paper

4.3 Opening a data set

In the Explorer window, click on "Open file" and then use the browser to navigate to the 'data' folder within the WEKA-3.8.1 folder.

Select the file called historical_weather_conditions.arff. (This is in fact the file listed above). This is a 'WPF evaluation' data set, like the ones used in class for demonstration purposes. In this case, the normal usage is to learn to predict the 'Acceptation' attribute from four others providing information about the WPF evaluation.

The Explorer window should now look like this:

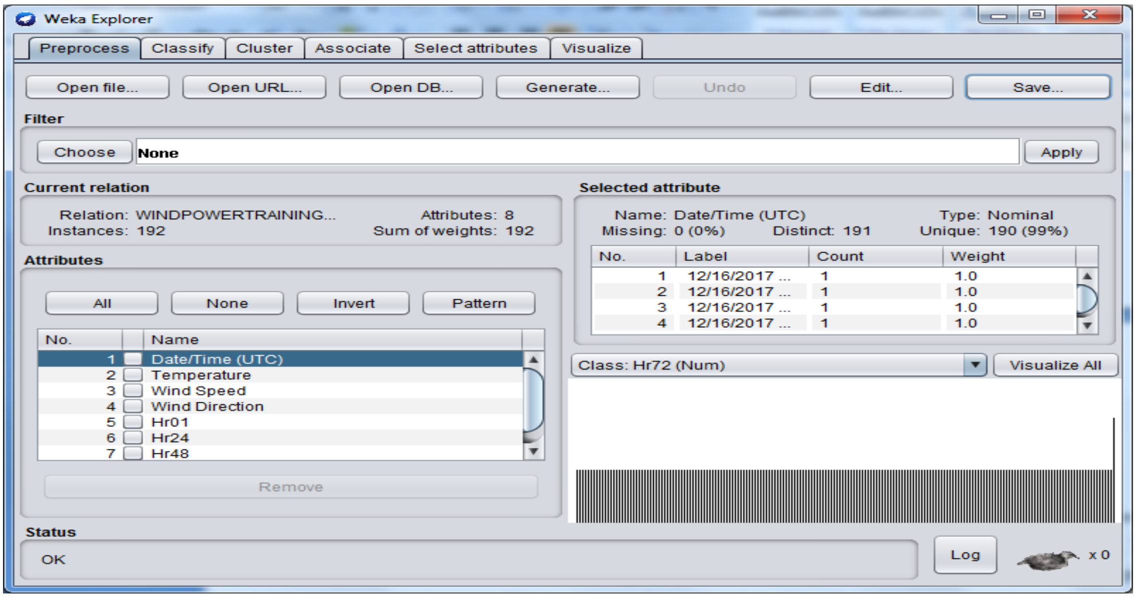

Fig 3 Dataset

By using BestFirst Algorithm Select attributes in historical Weather dataset and in historical WPD Dataset. In historical Weather dataset selected attributes are DateTime, Temperature and Wind Directions. Like Wise in historical WPD Dataset selected attributes is Hr04.Hr37, Hr71.

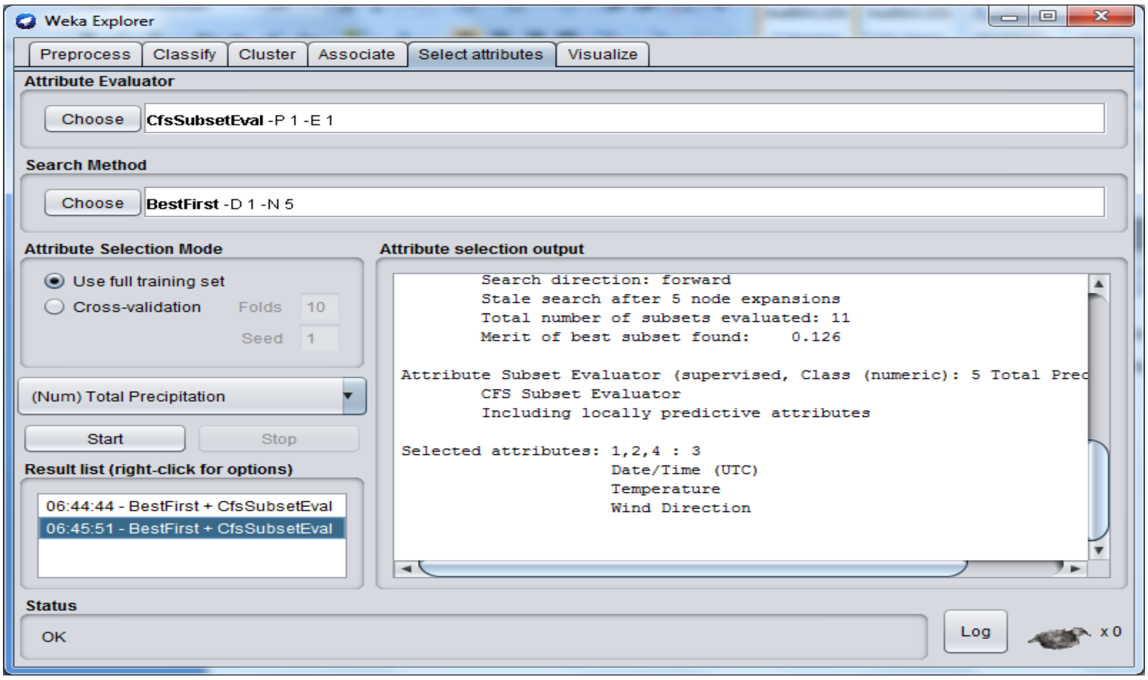

Fig 4 Dataset

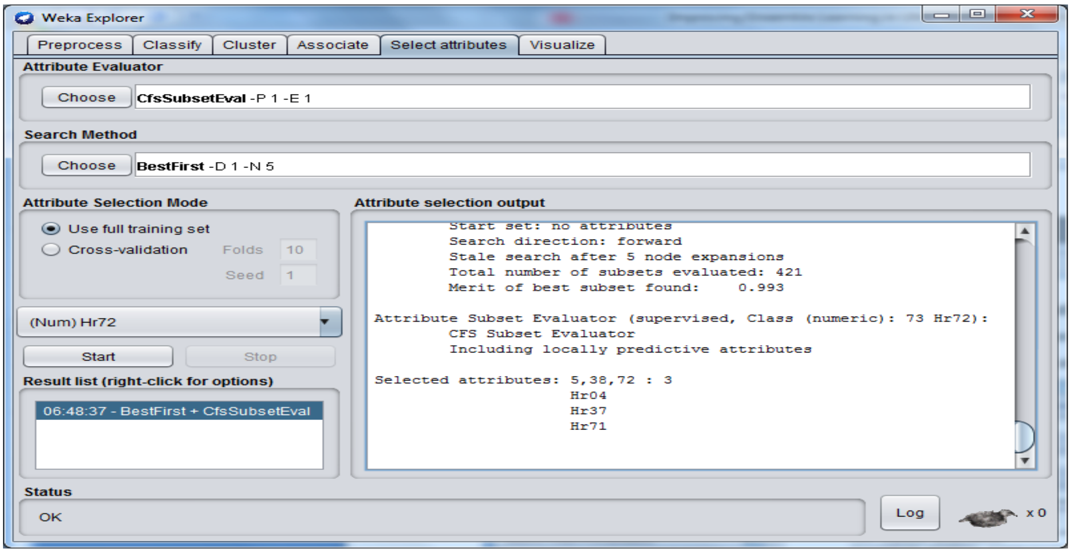

Fig 5 Attribute selection

By Applying K-means Clustering select similar day's data from in historical Weather dataset and in historical WPD Dataset.

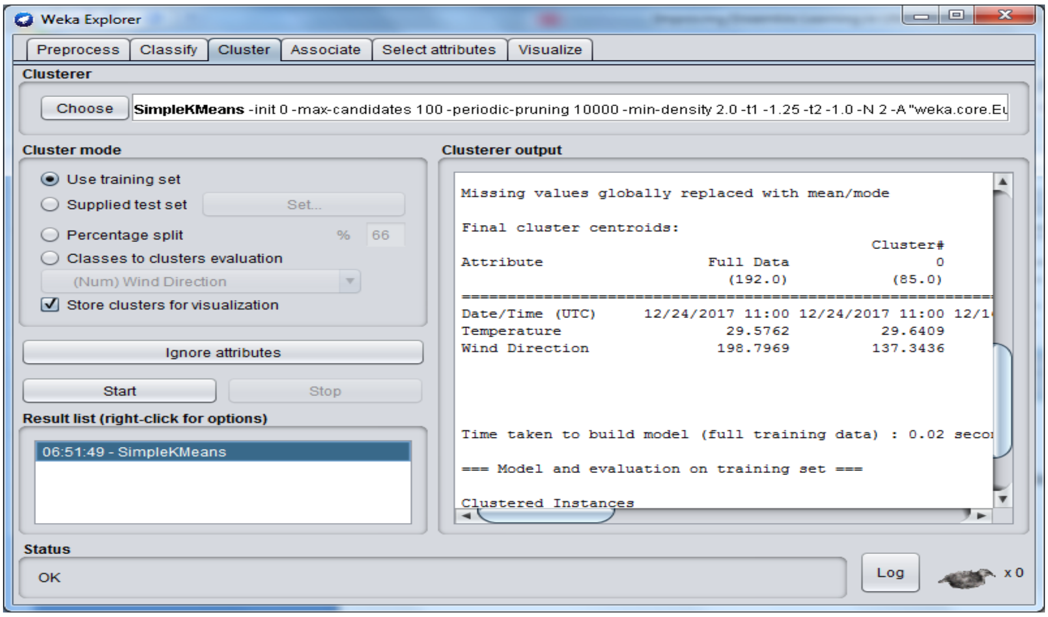

Fig 6 Attribute selection

After Clustering Join Similar days data to generate Training dataset. By applying Bagging Neural Network by using training dataset generate model and applying testing data reevaluate the result.

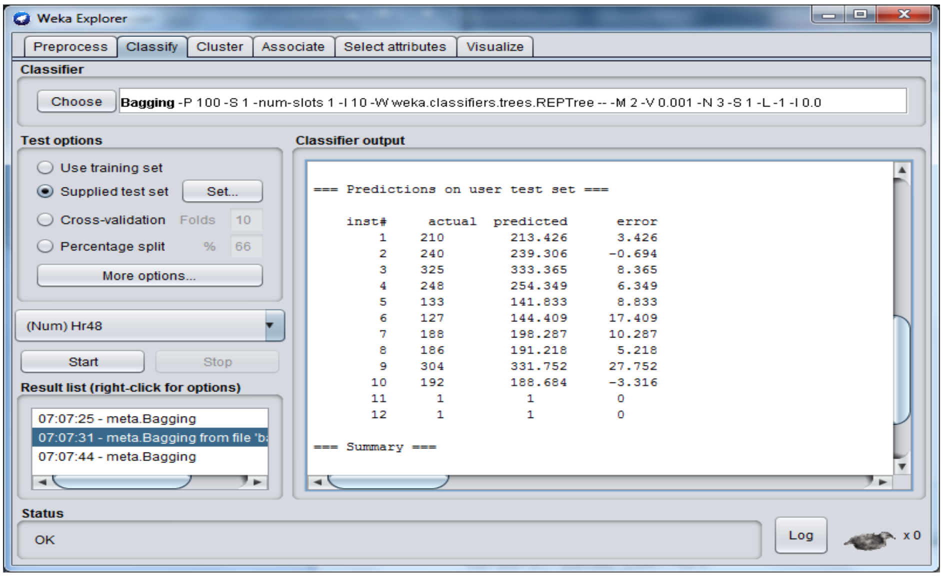

Fig 6 Classify

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

Table 1 The Performance of Baseline Approaches and the Proposed Approach

V. 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 BESTFIRST Search 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, and 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 farms, which not only increases the forecasting accuracy, but also reduces the computational complexity.

FUTURE ENHANCEMENT

To improve the forecasting accuracy, the effective mete orological 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 shortterm 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 shortterm wind power prediction with multiple observation points'," *IEEE Trans. Power Syst.*, 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]. M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Network*, vol. 29, no. 2, pp. 6–14, Mar. 2015.

[10]. M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, Third quarter, 2016.

[11]. M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog computing based radio access networks: Issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46–53, Aug. 2016.

[12]. A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Trans. Energ. Conver.*, 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 shortterm applications," *in Proc. IEEE ICRERA*, Nov. 2015, pp. 209-214.

[14]. M. Peng, D. Liang, Y. Wei, J. Li, and H. Chen, "*Self-configuration and self-optimization in LTE-Advanced heterogeneous networks*", *IEEE Communications Magazine*, May 2013.

[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*, 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.*, 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*, Koutynad 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

Dr.G.Anandharaj." Improving Ensemble Learning in Ultra Medium-Term Wind Power Prediction Based On Machine Learning Techniques ." IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 8, 2018, pp. 32-39.