

Wind Power Prediction based on Random Forests

Zehong Zhou^{1, a}, Xiaohui Li^{1, a} and Huaren Wu^{1, a}

¹School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing 210042, China.

^azoe_elite@163.com.

Keywords: Wind Power Generation; Wind Power Prediction; Random Forest

Abstract. With a massive increase of wind power, the prediction of wind power is becoming increasingly important. The algorithm of Random forests has many advantages such as less adjustable parameters, higher precision of prediction and better generalization ability. This algorithm has been widely applied in numerous fields such as medical science, management and economics. However there is no application in short-term wind power prediction yet. In this paper, the random forest algorithm will be applied to the short-term wind power prediction. The random forest regression model is established. The powers of a wind farm are predicted. The effectiveness of random forest regression algorithm adopted is verified in wind power prediction.

1. Introduction

With the exhaustion of fossil energy, to exploit and utilize renewable energy and to realize the sustainable development of energy have become significant measures of energy development strategy across the globe. Nevertheless, wind power generation has gradually become a new power generation technology which is comparatively mature in recent years due to the fact that wind resources are widespread, renewable, and pollution-free as well. By the end of 2015, the cumulative installed capacity of wind power in the world has reached 432, 419MW. It will expect to meet 12% of electric demand in the world by 2020.

Wind power prediction refers to analyzing and forecasting the active power generation of wind farms in the next period of time in advance by employing certain mathematical method. According to its purpose of using and the forecasted time scale, wind power prediction is divided into the super short-term 0-4h prediction, the short-term 0-72h prediction as well as mid-and-long term prediction with a longer time scale. The valid wind power prediction can be dealt with unstable output of wind power triggered by the wind itself with stochastic volatility and intermittent features, and correctly specified wind power plan to ensure safe and stable operation of electrical power system. Therefore, enhancing the accuracy of wind power prediction is one of the important measures to improve the safe and economic operation of power system including the large-scale wind power system. [1]

At present, the technology of wind power prediction develops relatively fast. MABEL establishes a power generation prediction model (including three input variables, i.e. wind speed, relative air humidity and electricity generation time) by using artificial neural network [2]. Reference [3] adopts the artificial neural network to predict wind power. Compared with statistical method, this kind of prediction method has a higher prediction accuracy, but it needs a mass of raw data and exists problems, such as the slow training speed, poor generalization ability and etc. Reference [4] describes a prediction model to consider wind power probability with climbing characteristics. It has acquired very high short-term wind power prediction accuracy.

Random Forest (RF) is an algorithm based on Bagging ensemble learning theory and stochastic subspace method. This algorithm can make full use of artificial intelligence model with excellent fitting ability. It has been widely used in medical science, biology, economics so on and so forth. However, it is applied little in wind power generation prediction. Citing the wind power of a wind farm a month as an example, this paper has constructed a wind power prediction model based on

random forest and tested the effect of prediction, aiming at providing a new idea for the short-term wind power prediction.

2. Random Forest Algorithm

Random Forest (RF) was proposed by Leo Breiman [5] in 2001. The random forest algorithm is based on statistical learning theory, by using bootstrap as sampling method extracting multiple samples from the original sample, building decision tree modeling according to each bootstrap sample, then integrating the prediction of multiple decision trees, and coming up with the final results by voting ultimately. The essence of RF is a classifier containing a number of decision trees, which are formed by adopting random method. Random Forest Regression (RFR) can be regarded as a strong predictor integrating a lot of weak predictors (decision trees). The realization process of random forest is as follows.

1. The original training set is N , the application of bootstrap method has been put back to the random extraction of K as a new self-help samples, and the resulting the classification trees, each time has not been drawn out of the sample composed of the out-of-bag data;

2. There is m variables, M try variables are randomly selected at each node of each tree, then choose the most ability of classification variables in M try variables and categorical variables threshold by examining each point determined;

3. Each tree growth the maximum and do not do any pruning;

4. The tree is composed of random forests, and the new data are identified and classified according to the random forest classifier.

Supposing the training set is extracted independently from the distribution of random vector X , Y ; as a result, any numerical prediction value $H(X)$ of the mean square generalization error is [6,7]:

$$E_{X,Y} [Y - h(X)]^2 \quad (1)$$

The predicted value of the random forest regression is the average of the K decision trees $\{h(\theta, X_k)\}$. It is similar to random forest classification. The theorem can be seen here [6,7]:

Theorem 1: $k \rightarrow \infty$

$$E_{X,Y} [Y - \text{av}h_k(X, \theta_k)]^2 \rightarrow E_{X,Y} [Y - E_\theta(X, \theta_k)]^2 \quad (2)$$

Record type (1.2) is the PE^{**} on the right side. It means the generalization error of random forest. The average generalization error of each decision tree PE^* can be defined as:

$$PE^* = E_\theta E_{X,Y} [Y - h(X, \theta)]^2 \quad (3)$$

Theorem 2: For all of the θ , $EY = E_X h(X, \theta)$

$$PE^{**} \leq \bar{\rho} PE^* \quad (4)$$

In this type, $\bar{\rho}$ is the residual error $Y - h(X, \theta)$ and $Y - h(X, \theta')$'s weighted correlation coefficient and also θ and θ' are mutually independent.

The theorem 2 gives the exact regression forest condition: low correlation between residual errors and decision tree with low error. Random forest reduces the average error of decision tree by weighted correlation coefficient $\bar{\rho}$.

Random forest prediction can be viewed as an adaptive neighborhood classification and regression process. For each one $X = x$, both of them can get the original n observed value of the weight set $\omega_i(x)$, $i=1, 2, \dots, n$. The estimation of random forest prediction or conditional mean is equivalent to the weighted mean of dependent variables.

3. Wind Power Prediction using Random Forests

In accordance with the random forest algorithm mentioned, the algorithm has fewer parameters to adjust and need not worrying about the features such as over fitting, speedy classification and high-efficient processing large sample data, estimable characteristic factor importance, strong ability to resist noise and so on. Hence, random forest can fully reflect the advantages of data mining and does not need to assume the implementation of function form in order to avoid the hypothesis error. In wind power prediction, using random forest regression method can effectively analyze the nonlinear and interaction data. It does not need to assume the provided model of mathematical form in advance. It has good regression analysis results.

The study of wind farm power prediction by using random forest method is based on the learning rule of high precision fitting in samples and the ability of the high confidence to promote knowledge out of samples. Random forest has two sorts of techniques, namely, classification and regression. The prediction of wind farm power in this paper belongs to the regression prediction. The following part demonstrates the calculation procedure of the wind farm power prediction model[8] based on random forest:

Wind power prediction of random forests is a collection of B trees $\{T_1(X), \dots, T_B(X)\}$. Among them, $X = \{x_1, \dots, x_p\}$ is the dimension p characteristic vectors of the wind power. The collection will produce B results $\hat{Y}_1 = T_1(X), \dots, \hat{Y}_B = T_B(X)$. The \hat{Y}_b ($b = 1, \dots, B$) is the predicted value of wind power about the tree b. For regression problem, \hat{Y} is the mean of all trees prediction.

A complete random forest training set is established in line with the process of stochastic algorithm, and then put the independent variables into the test set, and the result of wind power prediction comes out. The basic idea of establishing the random forest model is shown in Figure 1.

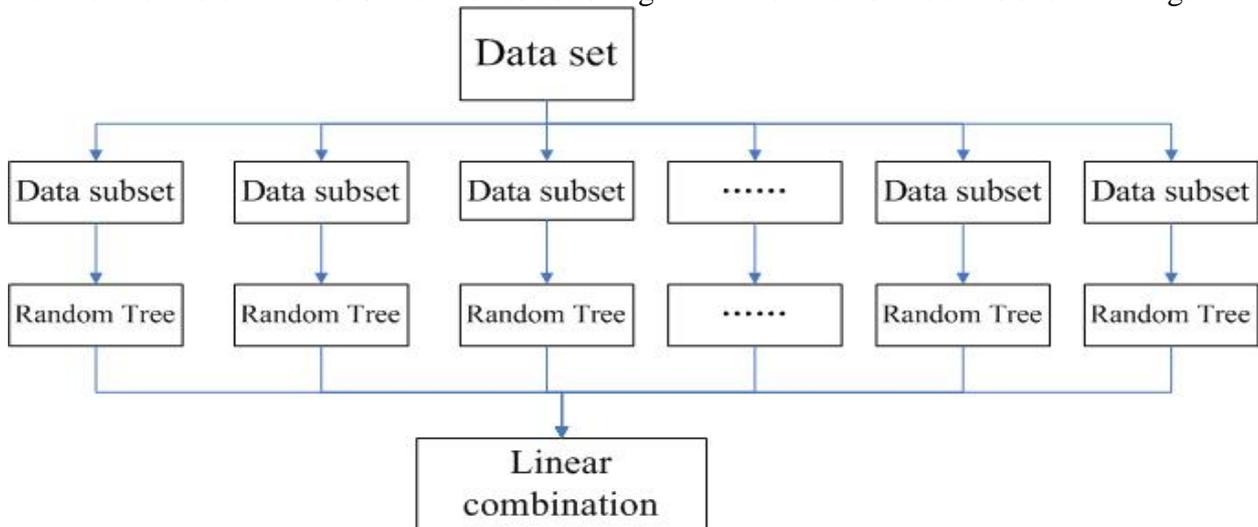


Fig. 1 The establishing of a random forest model

4. Case analysis

The historical data of wind farm in a certain area in March 2016 is analyzed in this paper. The total of installed capacity wind power is 49.5MW. The sample time is 15 minute. The 1,600 data in a month are selected. For the purpose of processing conveniently, the wind power value will be conducted a normalization processing. 1,000 samples are randomly selected as training set. 500 test samples, use TreeBagger function in Matlab software to verify the effectiveness of this method in this paper.

The Treebagger function of MATLAB software is used to establish the model of random forest regression algorithm as follows:

$B = \text{TreeBagger}(nTree, \text{train_data}, \text{train_label}, \text{'Method'}, \text{'regression'})$. Among them, nTree represents the number of decision trees, train_data and train_label represent the training data and the corresponding labels respectively. 1,500 data randomly selected from 1,600 data are analyzed.

$[\text{Predict_label}, \text{Scores}] = \text{predict}(B, \text{test_data})$. Among them, B represents the model obtained from the above training, test_data represents the test data.

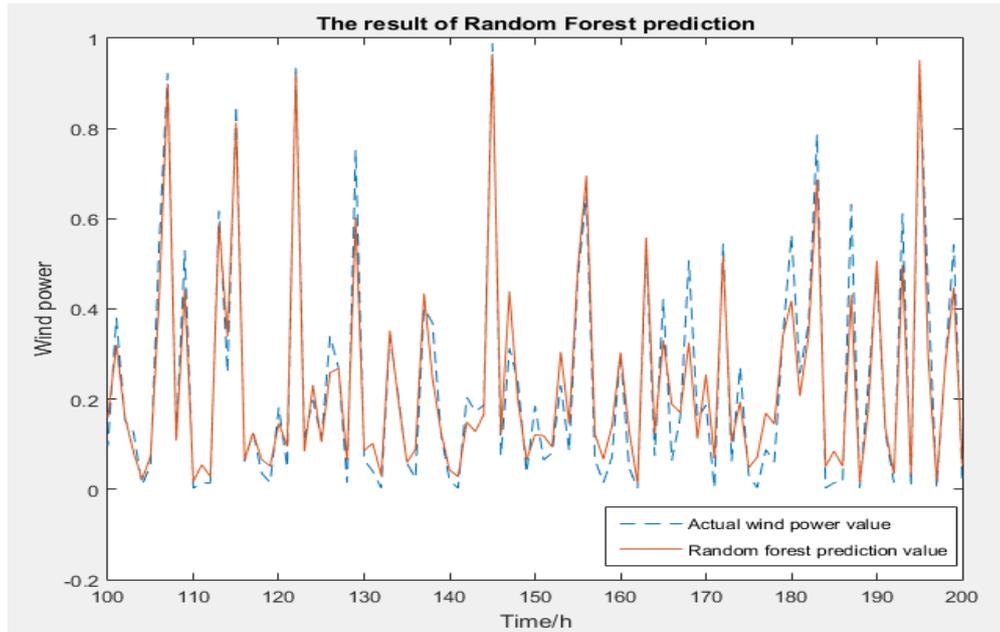


Figure 2 Random forest prediction

It can be seen from the figure that the predicted power is close to the actual power, so this method can be better applied to predict the power. As a whole, the random forest algorithm plays a very important role in application.

Table 1 prediction results under different decision trees

decision tree	MAE/%	MSE/%
50	0.048	2.476
100	0.050	2.379
200	0.049	2.129

In this table, MAE is short for Mean Absolute Error and MSE is the short for Mean Square Error.

Figure 4 shows the random forest algorithm for wind power prediction is feasible, and the table 1 also lists the mean absolute error and mean square error under different decision trees. This method has its certain superiority.

5. Summary

A random forest model of wind farm power prediction is constructed in this paper, and an empirical study is carried out on the basis of the actual data from a wind farm. The results suggest that the prediction accuracy of random forest model is obviously better than the traditional method in both training samples and test samples. The research of this paper provides a scientific and effective solution to the practical problem of wind power prediction.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (51177074).

References

- [1] WANG Jian, YAN Gangui, SONG Wei, MU Gang. The review of Wind power prediction technology [J]. Journal of Northeast Dianli University, 2011,03:20-24. In Chinese.
- [2] MABELC, FERNANDEZ E. Analysis of wind power generation and prediction using ANN: A case study [J]. Renewable Energy, 2008 (33) : 986 — 992.
- [3] HUANG Jin-hua , PENG Hui. Study of wind power short-term prediction of wind farm based

on neural network [J]. *Electro-technics Electric*, 2009, 9.

[4] GAN Di, KE Deping, SUN Yuanzhang, CUI Mingjian. Short-term probabilistic wind power forecast considering ramp characteristics[J]. *Electric Power Automation Equipment* , 2016,04:145-150. In Chinese.

[5] Breiman L. Random Forests[J]. *Machine Learning*,2001, 45(1):5-32.

[6] FANG Kuangnan. Random forest combination forecast theory and its application in financial[M]. Xiamen university press, 2012,5. In Chinese.

[7] Koczyk G, Chelkowszki J. An assessment of the resistance gene analogues of *Oryza sativa ssp japonica*: their presence and structure[J]. *CellularMolecular Biology Letters*, 2003, 8(4): 963–972.

[8] XU Ge, ZHANG Ke. Based on random forest model of real estate price evaluation[J]. *Statistics and Decision*, 2014,17:22-25.In Chinese.