IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 15, 2020, accepted December 28, 2020, date of publication December 31, 2020,
date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048382

Wind Power Short-Term Forecasting Model Based on the Hierarchical Output Power and Poisson Re-Sampling Random Forest Algorithm

JIE HAO[1,2], (Member, IEEE), CHANGSHENG ZHU[1], AND XIUTING GUO[1], (Member, IEEE)
[1]School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
[2]School of Electrical Engineering, Northwest Minzu University, Lanzhou 730030, China

Corresponding author: Changsheng Zhu (viewjie1@126.com)

This work was supported in part by the Central University Fund Project of China under Grant 31920190066.

ABSTRACT Under the background of big data, the use of massive online data to improve the real-time characteristics and reliability of wind power prediction and to reduce the impact of wind farms on the power grid makes the power supply and demand balance important problems to solve. This paper provides a new solution for short-term wind power forecasting to address these problems. In this paper, an improved random forest short-term prediction model based on the hierarchical output power is proposed, and it is used to forecast the power output of a real wind farm located in Northwest China. First, a chi-square test is adopted to discretize the power data to divide the large-scale training data and remove abnormal data. The novelty of this study is the establishment of a classification model with the output wind power as the classification target and the use of Poisson re-sampling to replace the bootstrap method of the random forest, that is, to improve the training speed of the random forest algorithm. The results indicate that the proposed technique can estimate the output wind power with an MSE of 0.0232, and the comparison illustrates the effectiveness and superiority of the proposed method.

INDEX TERMS Chi-square test, data discretization, Poisson re-sampling, random forests, wind power prediction, weighted k-nearest neighbors algorithm.

I. INTRODUCTION

Wind energy is a renewable and clean energy source with large storage capacity and wide distribution. Wind power has become the fastest growing renewable energy power generation technology in the world [1], [2]. However, wind power is volatile and intermittent, and large-scale wind power grid connections represent severe challenges to the safe and stable operation of power systems.

Wind power prediction technology is an effective method to mitigate the negative effects of wind power grid connections [3]. Accurate and reliable wind power prediction has greatly contributed to dynamic economic dispatch in power systems. In addition, the wind power permeability is increased, the rotating spare capacity is reduced, and the wind farm capacity coefficient is stabilized [4].

The associate editor coordinating the review of this manuscript and approving it for publication was B.Chitti Babu.

Wind power prediction methods can be divided into two categories according to the physical objects of prediction. One way is to predict the wind speed first and then obtain the wind power output according to the wind turbine or wind farm power curve. The other is to directly predict the output power of wind farms. At present, the research methods for wind power prediction around the world have shifted from physical methods to statistical and machine learning methods [5], including support vector machines (SVMs) [6], time series analysis [7], [8], neural networks [9], [10], fuzzy control [11], [12], and intelligent optimization algorithms [13]–[15]. These methods usually train a prediction model according to historical sampling data and then find the optimal model parameters. Furthermore, these models achieve high precision in short-term wind power prediction. However, when the amount of historical data is large and the confidence intervals are too large, the above models suffer problems such as slow convergence speed, serious overfitting and easily falling into

local minimum. Thus, the accuracy of wind power prediction is greatly reduced.

Wind speed is a key factor affecting wind power. In fact, even at different times of the same day, the frequency and amplitude of wind speed fluctuations can vary greatly, which increases the difficulty of wind power prediction. In large-scale wind farms, the spatial and temporal distributions of wind conditions are more diverse. If all the data are used to predict wind power directly through a single prediction model, it is difficult to describe the complexity and diversity of wind conditions, which affects the prediction accuracy of the model.

To address these problems, a wind power prediction algorithm based on weather classification was proposed in [16]. Numerical weather forecast (NWP) data are classified according to daily changes, and prediction models for different types of NWPs are established. In [17], a prediction model for wind power probability was proposed based on wind scenario recognition. The wind speed and direction were selected as the reference variables to divide the wind scenario and a cluster analysis model of the wind scenario was established. In [18], a combination of the wind speed variation period and wind power prediction was proposed, and a combination forecast model was developed. These methods have overcome the limitation of single prediction models to some extent. However, it is not sufficient to reflect the impact of wind changes on the output power of wind turbines by simply classifying the wind speed according to the season or time period.

This paper proposes a hierarchical wind power prediction model, and the larger data set is divided into smaller training sets for modelling. By discretizing historical data, the regression problem is transformed into a classification problem, and wind speed and direction data are grouped into narrow confidence intervals according to different output wind power levels. Then, Poisson re-sampling is used, instead of the bootstrap method in random forests, to simplify the pruning process when establishing the wind power prediction models. Finally, a dataset from the Supervisory Control and Data Acquisition (SCADA) system of a wind farm in Northwest China is used to assess the validity of the proposed model.

II. DISCRETE METHOD OF OUTPUT WIND POWER BASED ON CHI-SQUARE TEST

A. CHI-SQUARE TEST

The chi-square test, also known as the χ^2 test, is a non-parametric hypothesis testing method that is mainly used for statistical inference of unordered categorical variables [19]. The basic idea of the chi-square test is to compare the consistency between theoretically inferred values and actual observed values. Assuming that the composition ratios of the categorical variables are the same, the chi-square statistic χ^2 is constructed as:

$$\chi^2 = \sum \frac{(A-T)^2}{T} \tag{1}$$

where A is the actual observed value and T is the theoretical inferred value. According to (1), the smaller the difference between A and T is, the closer χ^2 is to 0. Therefore, χ^2 can be used to reflect the similarity between the actual observed value and the theoretical inferred value. The chi-square test is often used in feature selection, anomaly detection and correlation analysis.

B. CHI-SQUARE SORTING AND CHI2 ALGORITHM

In data mining modelling, continuous variable sometimes need to be transformed into discrete variables in a process called data discretization. Furthermore, feature discretization can be used to simplify a regression model. This process is robust to abnormal data because nonlinearity is introduced in data discretization, which makes the model more stable and reduces the risk of overfitting.

Binning is a common data discretization method that includes unsupervised binning and supervised binning. Chi-square binning is a supervised method that relies on the chi-square test. The chi-square statistic χ^2 is selected as the statistical index for discrimination. The basic idea of chi-square binning is to judge whether there is a distribution difference between two adjacent intervals [20]. If the two adjacent intervals have very similar distributions, they can be merged; otherwise, they should be kept apart. When the data are discretized on the basis of the chi-square statistic, the bottom-up merge method is adopted until reaching the restricted condition.

Data discretization algorithms based on chi-square binning include the ChiMerge algorithm, Chi2 algorithm, and mod-Chi2 algorithm [21]. However, an appropriate threshold is not easy to obtain in the original ChiMerge algorithm, whereas the improved Chi2 algorithm can automatically determine an appropriate chi-square threshold and maintain the restoration degree of the original continuous data set.

In this paper, the Chi2 algorithm is used to discretize wind power into different levels. The input data of the Chi2 algorithm include vector data for discretization and evaluation standards for classification.

The aims of this paper is to discretize wind power, and based on this discretization, train different wind power prediction models. First, according to the Chi2 algorithm, wind speed and wind direction are used as the classification evaluation criteria. The data set in this paper is from the SCADA system of a wind field in Northwest China, and the wind speed and direction data are the average values measured by two wind sensors on the wind turbine at a given time. Taking into account the computing power of the computer we used, the operation data of a wind turbine for one month, that is, from 2016-6-1 to 2016-6-30, are selected, and the data are collected with one-minute intervals.

Fig. 1 and Fig. 2 show the wind speed distribution and wind roses of the wind field, respectively. The wind speed is distributed mainly in the range of $5\ m/s \sim 15\ m/s$, and the prevailing wind direction in this month is south. In this paper, the considered data attributes include wind speed, wind

Chi2 Algorithm

Input: Continuous data set, significance level α and inconsistency rate δ
Output: Discrete data set
Stage 1: Start with the initial significance level α, discretize all the numerical attribute values. Each attribute is sorted by its value
Stage 2: Perform the following steps:
1. Calculate the Chi-square statistic χ^2 of adjacent intervals (at the beginning, each interval contains only one attribute value);
2. Combine the two intervals with the lowest χ^2 in the adjacent interval;
3. Continue the merge process until the χ^2 values of all adjacent intervals exceed the parameter α;
4. Repeat the process and gradually reduce the parameter α until the inconsistency of the data is less than the parameter δ.

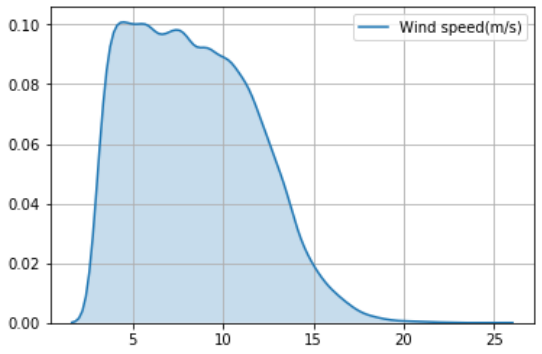

FIGURE 1. The wind speed probability distribution of the wind field.

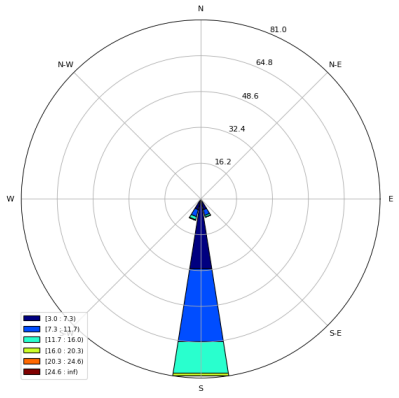

FIGURE 2. The wind direction rose of the wind field.

direction and the output power of the wind turbine. Data that do not meet the cut-in speed ($3\ m/s$) and cut-out speed ($20\ m/s$) are removed, and there are a total of 26796 pieces of data.

The sample data are shown in Table 1 (only the first 10 pieces of data are displayed in the table).

TABLE 1. The data set from the SCADA system.

Time	Wind speed (m/s)	Wind direction (°)	Wind power (kw)
2016-06-01 09:01:00	9.9274	181.0262	1678
2016-06-01 09:02:00	11.667	175.8787	1885
2016-06-01 09:03:00	11.2031	177.9269	1867
2016-06-01 09:04:00	10.6232	174.8918	1856
2016-06-01 09:05:00	11.435	176.9168	1865
2016-06-01 09:06:00	9.5215	178.757	1677
2016-06-01 09:07:00	11.3191	174.0502	1872
2016-06-01 09:08:00	10.1594	176.1071	1764
2016-06-01 09:09:00	10.6232	175.0758	1826
2016-06-01 09:10:00	12.1309	178.7631	1853

TABLE 2. Contrast of discretized results.

Sequence Number	Discretization result	Actual wind power (kw)
17199	18	2120
9891	1	37
2843	7	486
17896	18	2081
9290	15	1613
14758	11	1034
22560	4	197
5138	6	329
6775	12	1108
11451	9	637

According to the wind energy density formula, wind speed is the decisive factor of wind power. The dominant wind direction remained essentially unchanged for a month; therefore, we focus on the influence of wind speed on wind power discretization. The relationship and distribution of the original wind speed and wind power data are represented as a scatter diagram in Fig. 3. Several distinct outliers can be observed in Fig. 3. These outliers must be removed; otherwise, the accuracy of the prediction model will be affected.

In the Chi2 algorithm, the initial value of parameter α is 0.5, and that of δ is 0.05. Finally, wind power was discretized into 18 levels, and a comparison between the discretization result and the actual wind power is shown in Table 2 for 10 randomly selected pieces of data (where the sequence number represents the ordinal relation of the data in the original data set).

To ensure that the discrete data set accurately represents the original data set, a consistency check is used as the stop criterion for the Chi2 algorithm. This approach automates the discretization process by introducing an inconsistency rate as the stopping criterion, and the significance value is automatically selected. Therefore, the algorithm has good robustness in data discretization.

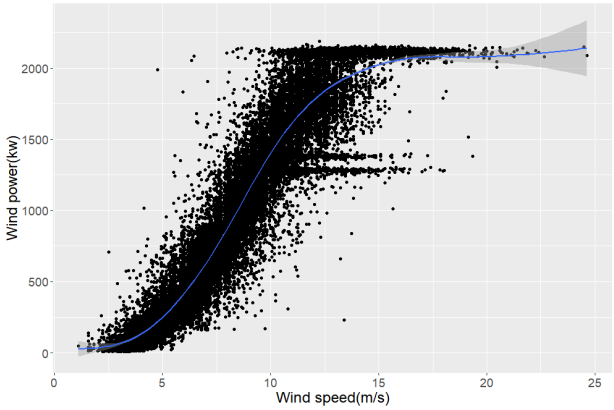

FIGURE 3. Scatter plot of wind power.

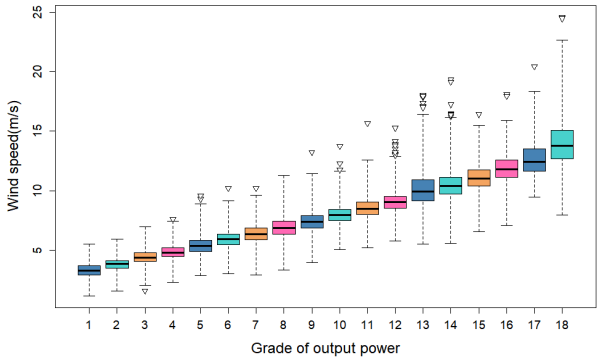

FIGURE 4. Box plot of wind power.

In addition, we can convert the wind power level into a label variable and draw a boxplot of wind power and wind speed, as shown in Fig. 4. The relationship of wind speed and the discretized wind power is reflected in the boxplot, and the outliers are also shown in Fig. 4. The outliers were identified and deleted based on the boxplot. The triangular-shaped points in the figure are extreme outliers: to verify the anti-noise ability of the classification and prediction model, we kept some mild outliers.

The scatter diagram used to delete extreme abnormal data is shown in Fig. 5.

III. WIND STATE CLASSIFICATION MODEL BASED ON THE WEIGHTED K-NEAREST NEIGHBOR ALGORITHM

Commonly used machine learning classification algorithms include SVM, logistic regression, K-nearest neighbors (KNN) [22], and decision trees. In view of the wind power prediction model proposed in this paper and according to the discretization results of wind power, each measurement value has a different classification possibility, which makes the problem a multi-category classification problem with a single label. The KNN algorithm can predict the classification of new sample points by dividing data points into several classes; moreover, the algorithm is suitable for multi-category classification problems. Furthermore, compared to other algorithms, the KNN algorithm is simple to implement and has a good classification effect when the data size is large.

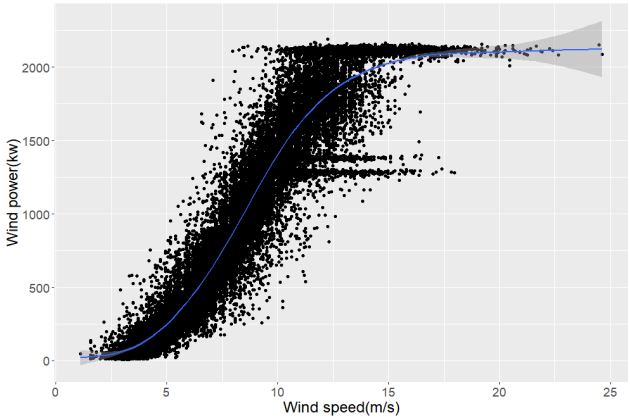

FIGURE 5. Scatter plot of wind power after deleting outliers.

A. KNN ALGORITHM

On the basis of a given training data set, the KNN algorithm can find the k instances closest to the new instance in the training data set according to the given training data set. If most of the k instances belong to the same class, the input instance is categorized into this class:

KNN Algorithm

Input: The training data set $T = \{(x_1, y_1), (x_2, y_2),..., (x_N, y_N)\}$

where $x_i \subseteq R^n$ represents the eigenvector ($i = 1, 2,..., N$) and $y_i \in \Upsilon = \{c_j\}$, which is the category ($j = 1, 2,..., K$).

Output: The classification to which the input variable belongs.

Stage 1: Based on the given distance measurement, find the k points closest to x in the training set T; the neighborhood of x that covers these k points is denoted as $N_k(x)$;

Stage 2: In $N_k(x)$, determine the category of x according to the classification decision rule:

$$y = \arg\max_{c_j} \sum_{x \in N_k(x)} I(y_i = c_j), \quad i = 1, 2, \ldots, N;; j = 1, 2, \ldots, K$$

where I is an indicator function. When y_i is equal to c_j, I is 1; otherwise, I is 0.

B. WEIGHTED KNN ALGORITHM

The KNN algorithm is sensitive to the uneven distribution of samples, which can easily cause classification errors. The error is more obvious when the distribution of samples has a large skew towards a particular classification. Because the density of the data sample is large, more samples may belong to a certain category, and the test data will be more likely to be assigned to this category, which could result in misjudgement. Therefore, we add a weight to the distance between the test

data and the known value; that is, the shorter the distance is, the greater the weight.

The weighted KNN must first sort the distance values. The nearest k elements [23] are selected, and the weighted average is calculated; that is, each distance value is multiplied by the corresponding weight, and the results are summed. The final distance can be expressed as:

$$f(x) = \frac{\sum_{i=1}^{k} d_i w_i}{\sum_{i=1}^{k} w_i} \tag{2}$$

where d_i represents the distance between the nearest neighbor i and the value to be predicted x, w_i is the weight, and $f(x)$ is the numerical result of the distance.

This paper uses a Gaussian function [24] to select the weights. When the distance is 0, the weight reaches its maximum value of 1. As the distance increases, the weight decreases continuously; however, it does decay rapidly and never reaches 0. The Gaussian weight is expressed as:

$$w_i = e^{-\frac{d_i^2}{2c^2}} \tag{3}$$

where c is the half-peak height of the Gaussian function, which is generally 0.3∼0.5.

In the traditional KNN algorithm, all training samples are considered equally important, that is, all the attributes are treated equally. This approach does not reflect the influence of different sample characteristic variables on the classification results. Compared to wind speed, wind direction has a weaker influence on wind power. In this paper, we assign different penalty factors to wind speed and wind direction when using the weighted KNN algorithm to reflect the influence of wind speed on the output power and to improve the classification accuracy.

C. WIND STATE CLASSIFICATION MODEL

The wind power prediction model proposed in this paper establishes different prediction models for different wind power grades. Wind speed and wind direction can be classified according to the results in Section II. However, 18 different prediction models must be trained, and according to the classification results, different grades of data sets can use different models to predict the output power, which greatly increases the workload of the wind power prediction task. Therefore, we must fully recognize the similarity principle between samples, and multiple classification labels should be merged to reduce the number of prediction models. To maintain the balance between the classified samples and reduce the influence of an uneven sample distribution on the weighted KNN classifier, according to the discretization results in Section II, the label merging results are shown in Table 3.

To verify the effectiveness of the algorithm, 70% of the data set is randomly selected as the training set and 30% is selected as the test set. The distance function used in this paper is Euclidean distance.

TABLE 3. The label merging results.

Class label	Discretization result	Data sample size
Grade Ⅰ	1~5	6346
Grade Ⅱ	6~10	7010
Grade Ⅲ	11~14	6627
Grade Ⅳ	15~18	6812

$$L\left(x_i, x_j\right) = \left(\sum_{l=1}^{n} \left|x_i - x_j\right|^2\right)^{\frac{1}{2}} \tag{4}$$

where x_i and x_j are two different instance points, each of which includes wind speed and wind direction characteristics.

In general, the model with the least empirical risk is the best model. When the sample size is sufficiently large, empirical risk minimization (ERM) can ensure a good learning effect. For a training set with N samples, the ERM function f_N is:

$$f_N = \arg\min \frac{1}{N} \sum_{i=1}^{N} L(y_i f(x_i)) \tag{5}$$

Second, an appropriate k must be chosen. The value of k should not be excessively large or small. A small k corresponds to a small training error, but for the overall model, overfitting is likely to occur. A larger k increases the training error and can lead to inaccurate prediction. Therefore, cross validation can be used to select the optimal k. This paper adopts 10-fold cross validation; that is, the data set is randomly divided into 10 disjoint subsets of the same size. Verification and comparison indicate that the optimal k is 11.

The trained model is then used to classify the test set. The wind power is classified according to the given grade, as shown in Fig. 6. Different colours in the figure represent different wind power grades, and the classification accuracy reaches 98%.

IV. A RANDOM FOREST PREDICTION MODEL BASED ON POISSON APPROXIMATE RE-SAMPLING

Random forest is one of the most popular machine learning methods [25] that has good advantages in different data sets. In the process of creating a random forest, the unbiased estimation method is used to estimate the generalization error, that is, the mathematical expectation of an estimator is equal to the estimated parameter. The difference between random forest algorithms is how randomness is introduced into the process of tree generation. Therefore, a random forest is a collection of trees with sample sampling and feature sampling; that is, the randomness of a random forest is reflected in two aspects: sample sampling (row sampling) and attribute sampling (column sampling). The basic random forest generation process is shown in Fig. 7.

x is the input characteristic variable, and y is the weighted output of the random forest. Given a data set containing m

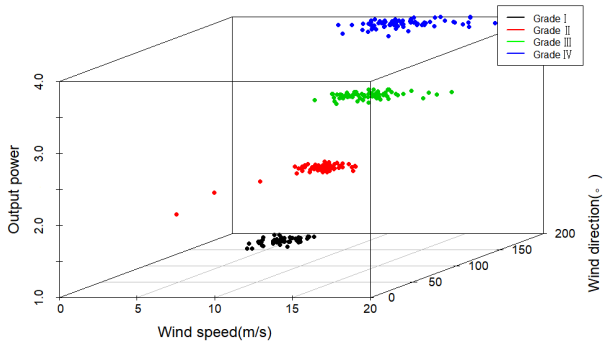

FIGURE 6. Scatter plot of wind power after deleting outliers.

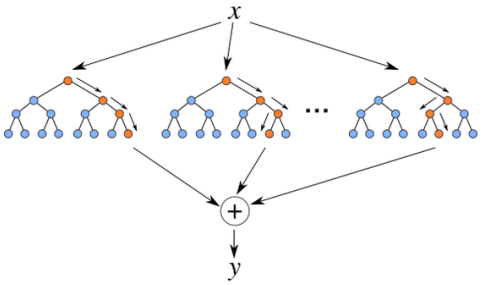

FIGURE 7. Random forest algorithm model.

samples, we sample the data at random with replacement, so the same sample can be selected more than once.

After random sampling n ($n \leq m$) times, we obtain a sample set with n samples. Some samples in the initial training set appear many times in the sampling set, while others never appear. Suppose that each sample has k characteristics and randomly select j attributes from k. Attribute sampling usually uses sampling without replacement. The split point of the decision tree is obtained according to the selected j attributes.

A. POISSON APPROXIMATE RE-SAMPLING ALGORITHM

In view of the problem proposed in this paper, only wind speed and wind direction are used for wind power prediction; thus, we have to consider only row sampling. To reflect the randomness of row sampling, bootstrap or bagging methods are generally used in the formation of classified regression trees in random forests. The training set used by each tree is sampled from the total training set; therefore, some samples from the total training set may appear multiple times in a tree's training set or may never appear. Sampling with replacement is the most effective method to generate a sample set for model fitting. This sampling approach is equivalent to sampling with a uniform binomial distribution; that is, the probability of each data point being selected follows a binomial distribution.

If the above process is repeated T times, we obtain T decision trees. For the regression problem, the output dependent variable is the average of each tree and can be expressed as:

$$y = \frac{1}{T}\sum_{i=1}^{T} h_i(x) \tag{6}$$

where $h_1, h_2, \ldots, h_T$ denote different decision trees and $h_i(x)$ is the output of h_i on sample x.

Sampling with replacement is the most effective way to generate a sample set from the initial data for model fitting. The probability of each sample data point being selected follows a binomial distribution and can be calculated as:

$$P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n \tag{7}$$

The Bernoulli trial is repeated n times under the same conditions. Only two opposing outcomes per experiment are possible, that is, A and A'. The probability of A occurring is p, where p is between 0 and 1. Therefore, the probability of A' occurring is 1-p (represented as q). Suppose that k is the total number of times the result is A in n trials and that k is a random variable. That is, the binomial distribution, also called the Bernoulli distribution, is denoted as $X \sim b(n, q)$.

However, in practical applications, the Bernoulli distribution usually has a large n and relatively small p, and the product of n and p is moderate. In this case, we can use the Poisson distribution to approximate the binomial distribution because the Poisson distribution is relatively simple to calculate [26].

The Poisson distribution is defined as:

$$\lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2..., n \tag{8}$$

where p_n is the probability of A occurring in n Bernoulli trials, which depends on the number of trials.

According to (7) and (8), the binomial distribution can be approximated by the Poisson distribution with parameter $\lambda = np$:

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \tag{9}$$

Generally, each tree is generated with the same amount of data as the total number of training samples. This approach is equivalent to using the bootstrap algorithm to sample all the initial training sets such that each model contains the same number of data samples as the initial data [27]. However, the bootstrap algorithm is not suitable for very large data sets. In addition to improving the prediction accuracy, improving the fitting speed is also very important.

B. IMPROVED RANDOM FOREST

Wind power forecasting involves a large quantity of data, such as wind speed and wind direction, which can be collected at one-minute intervals. When sampling is conducted in a large amount of data, the probability of any sample being selected is very small. This scenario conforms to the conditions under which the binomial distribution approximates the Poisson distribution. In this paper, the bootstrap algorithm is replaced by Poisson re-sampling to improve the training speed of the random forest model. We call the improved algorithm Poisson re-sampling random forest, and the specific process is as follows.

(1) Conduct separate Poisson re-sampling for each input sample of the training data set $\{(x_1, y_1),(x_2, y_2),\ldots, (x_i, y_i),\ldots,$

$(x_n, y_n)\}$, where x_i is a multidimensional vector. The Poisson sampling parameter is:

$$\lambda = \frac{t}{n} \tag{10}$$

where n is the total number of samples in the initial training set and t is a data sample contained in any model.

(2) In the process of regression tree generation, ramification is needed. For all input variables, we must determine the optimal ramification variable and the optimal ramification point. That is, a characteristic variable is selected, the input space is divided into two parts according to its value, and the operation is repeated. In this paper, wind speed and direction are predicted, and the optimal split variable is selected according to the predicted target. Taking wind speed as an example, wind speed is the optimal split variable j. The objective function of searching for the optimal split point s is:

$$F_{\min} = \min_{j,s}\left[\min_{c_1}\sum_{x_i \in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2(j,s)}(y_i - c_2)^2\right] \tag{11}$$

where R_1 and R_2 represent the two subspaces after ramification. c_1 and c_2 are the predicted values of the two subspaces, which are equal to the output mean value of each sample in different subspaces, as in (12):

$$c_m = ave(y_i | x_i \in R_m) \tag{12}$$

where m represents the different subspaces that the input space has divided.

(3) Step (2) is repeated in the two subspaces until the objective function F_{min} satisfies the stop condition. Finally, the regression tree is generated as:

$$h(x) = \sum_{m=1}^{M} c_m I(x \in R_m) \tag{13}$$

where I represents different input spaces. A regression tree corresponds to an input space and the output values on this input space.

(4) Sampling T times generates T regression tree prediction models $\{h_k(x)\}$ $(k = 1, 2, ..., T)$. That is, the sampling frequency of any data sample is subject to a Poisson distribution, and the parameter λ is Tt/n. Then, the results of all the learners are averaged mathematically to obtain the output of the model as:

$$H(x) = \frac{1}{T}\sum_{k=1}^{T} h_k(x) \tag{14}$$

The large quantity of wind power data entails some challenges in wind power prediction. Compared to the bootstrap algorithm, the Poisson re-sampling algorithm simplifies the sampling process and improves the modelling speed.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, first, the feasibility of the improved random forest algorithm prediction model is verified and compared to other algorithms in terms of the prediction accuracy and real-time performance. Second, we analyse the prediction effect before and after wind power grade classification.

A. POISSON RE-SAMPLING ALGORITHM

To address the power prediction problem in actual wind farms, the effectiveness and superiority of the Poisson approximate re-sampling method in the improved random forest algorithm are verified. The experimental data are the full data without wind power classification. For the random forest algorithm, by means of cross validation, the mean square error (MSE) is minimized when the number of trees in the random forest is approximately 150. In consideration of the real-time performance of the prediction algorithm, the number of regression trees was set to 100. First, to compare the performance of different prediction models on this problem, we chose gradient boosting regression tree (GBRT) and multilayer perceptron (MLP) to compare with the random forest algorithm for the same training data.

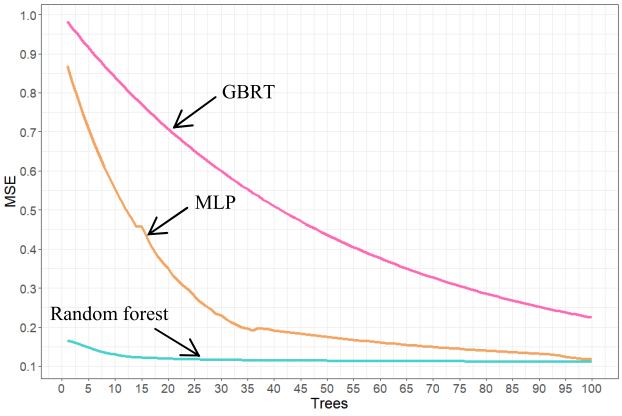

FIGURE 8. Comparison of iteration curves.

The iterative curves of these three models are shown in Fig. 8. Considering the training time and memory consumption of the model, the learning rate of GBRT is 0.01. GBRT is mainly used for regression prediction, but more regression trees are needed to achieve better prediction performance. Therefore, when the number of regression trees was limited to 100, as shown in Fig. 8, the prediction results were worse than those of MLP and random forest. For this problem, the numbers of nodes in the input layer, hidden layer and output layer of the multilayer perceptron were 2, 8, and 1, respectively. The MSE of the network after iteration is similar to that of the random forest algorithm. However, the network training speed is slow, which affects the real-time performance of prediction, and it is easy to fall into local extrema. Therefore, the random forest algorithm is selected and optimized in this paper.

TABLE 4. Performance comparison of four algorithms.

Method	MSE	Training time(s)
GBRT	0.224	15.98
MLP	0.117	26.23
Random forest with Bootstrap sampling	0.111	16.71
Random forest with Poisson re-sampling	0.096	12.35

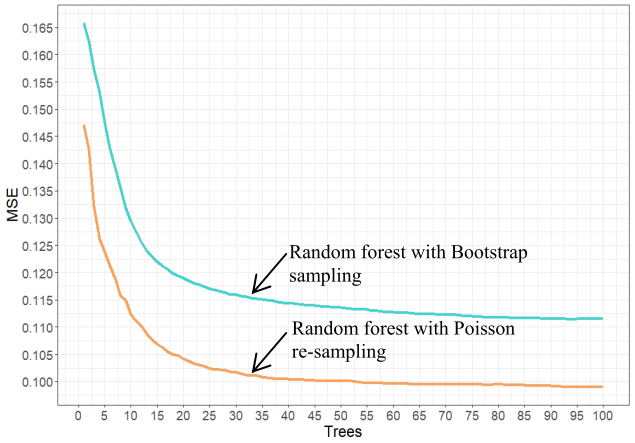

FIGURE 9. Comparison of iteration curves with different sampling algorithms.

TABLE 5. Performance of the wind power prediction algorithm for different grades.

Class label	MSE	Training time (s)
Grade Ⅰ	0.0278	2.96
Grade Ⅱ	0.0205	3.05
Grade Ⅲ	0.0238	3.20
Grade Ⅳ	0.0207	3.11

The iterative curves of the random forest algorithm before and after improvement are shown in Fig. 9. The brown curve in the figure is the iterative curve of the improved random forest algorithm. When the size of the regression tree is 100, the MSE is 0.096. In addition, to achieve the same error, compared with the unimproved random forest algorithm, the method proposed in this paper requires fewer regression trees; that is, the real-time performance is better.

TABLE 4 compares the performance of the four algorithms, including the MSE and training time. The algorithms proposed in this paper perform better in terms of training accuracy and training time.

The predictive model after training is used to test the wind motor output power per minute in the next hour to verify the fitting of the predictive model.

Fig. 10 shows the error curve and fitting curve of ungraded wind power prediction. The fitting curve shows the wind power output at 60 time points to be predicted (in minutes). As the wind power fluctuates greatly, the prediction error is large and the fitting effect is not ideal.

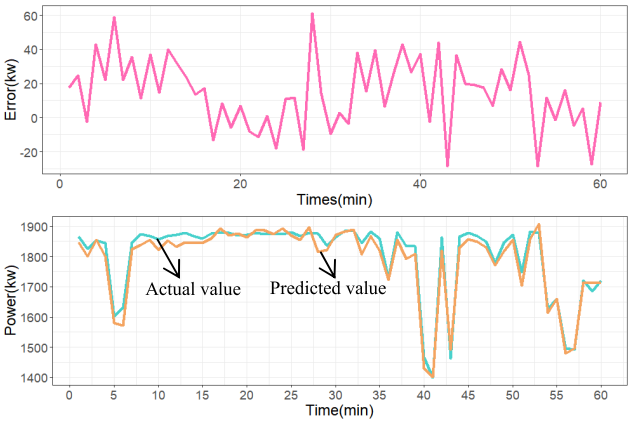

FIGURE 10. Prediction curve of the improved random forest.

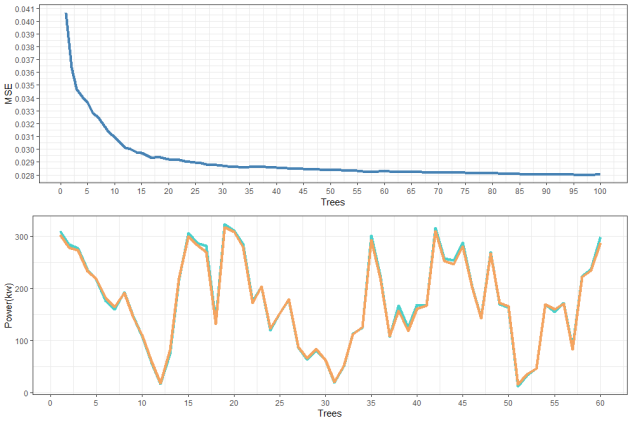

FIGURE 11. Training and prediction curve for grade I.

B. THE HIERARCHICAL WIND POWER PREDICTION ALGORITHM

According to the algorithm proposed in this paper and the above classification model, the sample data are first classified; then, the classified data are used for wind power prediction. Additionally, the improved random forest algorithm proposed in this paper is adopted. In the experiment, the number of regression trees in the random forest was also set as 100, and the four grades of data were trained and predicted to verify the effectiveness and superiority of the hierarchical wind power prediction algorithm.

Figs. 11 to 14 show the iteration curves and fitting curves of the four grade prediction models. TABLE 5 shows the performance of the four prediction models.

In the above four figures, the blue curve is the iterative curve. According to the iteration curve, when the number of regression trees reaches 50, a small MSE can be obtained. The green curve is the measured value of wind power, and the brown curve is the predicted value. The real value and the predicted value of each grade basically coincide, and only a few large errors occur at the turn of the curve.

As shown in TABLE 5, compared with the prediction results using the full data, the hierarchical wind power

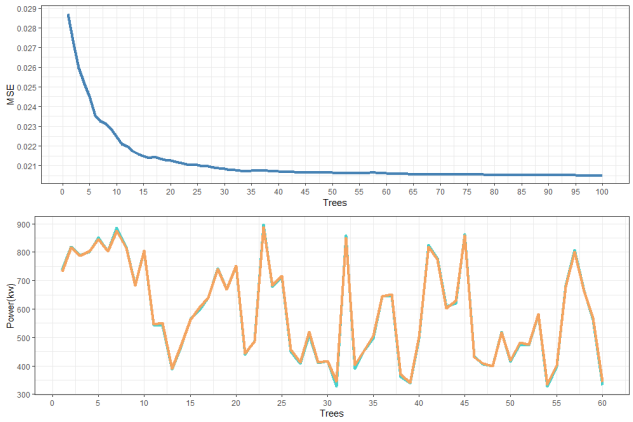

FIGURE 12. Training and prediction curve for grade II.

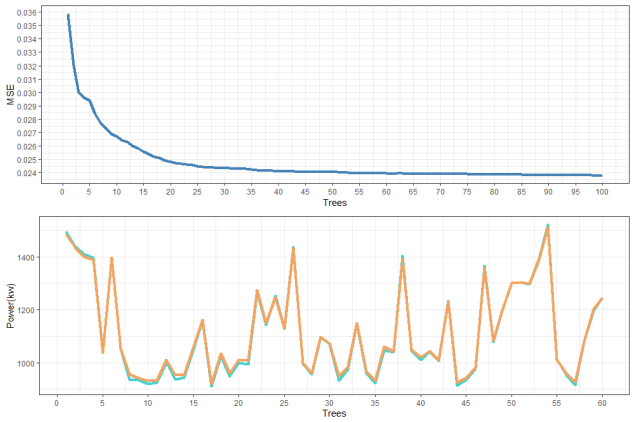

FIGURE 13. Training and prediction curve for grade III.

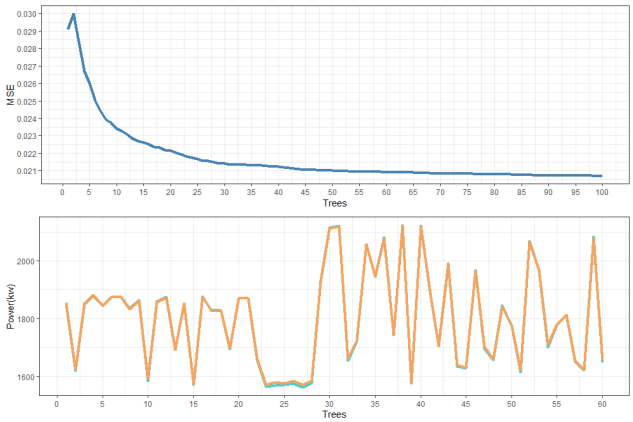

FIGURE 14. Training and prediction curve for grade IV.

prediction model has a shorter iteration time and smaller prediction error. This is because all the data are first classified, and the prediction model is established in a data set with small differences, which reduces the prediction error to a great extent and improves the real-time performance of the prediction algorithm.

To compare the results obtained with the improved random forest algorithm applied to all the data, the prediction accuracy and real-time performance after classification are improved to a great extent. However, the classification time is not considered. Classification is equivalent to the segmentation of a large data set before the prediction. In addition, the hierarchical wind power is predicted separately, and the total forecast time of this model is not counted. Therefore, the real-time performance of the model is still lacking, and parallel processing would improve the method.

VI. CONCLUSION

The large quantity of wind power data entails challenges in the prediction of wind power. The wind power hierarchical algorithm proposed in this paper as a data segmentation method can reduce the data scale without changing the original characteristics of the data to transform prediction problems into classification problems, which is conducive to the establishment of refined prediction models. Second, the weighted KNN algorithm shows good classification performance when the number of classification variables is relatively small. In addition, compared with the bootstrap algorithm, the random forest algorithm based on Poisson resampling is more suitable for big data modelling. The experimental results show that the algorithm proposed in this paper shows superior classification and prediction performance. The new solution can estimate the output wind power with an MSE of 0.0232, better than that of the algorithms that are not hierarchical.

In the next step, to account for the characteristics of big data in wind power, parallel modelling of the prediction algorithm will be realized to further improve the accuracy and speed of prediction.

REFERENCES

[1] M. R. Islam, S. Mekhilef, and R. Saidur, "Progress and recent trends of wind energy technology," *Renew. Sustain. Energy Rev.*, vol. 21, pp. 456–468, May 2013.

[2] J. Zhao, J. Wang, and Z. Su, "Power generation and renewable potential in China," *Renew. Sustain. Energy Rev.*, vol. 40, pp. 727–740, Dec. 2014, doi: 10.1016/j.rser.2014.07.211.

[3] S. C. Liu, D. X. Zhang, C. Y. Zhu, "Thanking of energy Internet big data technology," *Autom. Electr. Power Syst.*, vol. 40, no. 8, pp. 14–21, Jun. 2016, doi: 10.7500/AEPS20151106007.

[4] Y. N. Zhao, "Research on short-term prediction method of large-scale wind power based on temporal and spatial correlation," Ph.D. dissertation, Dept. Electron. Eng., China Agricult. Univ., Beijing, China, 2019.

[5] J. M. Lima, A. K. Guetter, S. R. Freitas, J. Panetta, and J. G. Z. de Mattos, "A meteorologica–statistic model for short-term wind power forecasting," *J. Control, Autom. Electr. Syst.*, vol. 28, no. 7, pp. 679–691, Jul. 2017, doi: 10.1007/s40313- 017-0329-8.

[6] Z. Y. Ding, P. Yang, X. Yang, "Wind power prediction method of support vector machine based on clustering of continuous time period," *Autom. Electr. Power Syst.*, vol. 36, no. 14, pp. 131–135, Sep. 2012, doi: 10.3969/j.issn.1000-1026.2012.14.025.

[7] J. Yang, L. Zhang, M. Q. Wang, and X. Han, "Wind-power forecasting error simulation considering output level and self-correlation," *Electr. Power Autom. Equip.*, vol. 37, no. 9, pp. 96–102, 2017, doi: 10.16081/j.issn.1006-6047.2017.09.013.

[8] K. P. Shi, Y. Qiao, and W. Zhao, "Short-term wind power prediction based on entropy association information mining of historical data," *Autom. Electr. Power Syst.*, vol. 41, no. 3, pp. 13–18, Feb. 2017. [Online]. Available: http://aeps.alljournals.ac.cn/aeps/article/pdf/20160504020

[9] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, Apr. 2014, doi: 10.1109/TSTE.2014.2300150.

[10] Y. C. Zhang, Y. C. Wang, H. Deng, X. Xiong, and H. Chen, "Short-term wind power prediction based on IAFSA-BPNN," *Power Syst. Protection Control*, vol. 45, no. 7, pp. 58–63, Jun. 2017, doi: 10.7667/PSPC160483.
[11] J. Zhao, Z.-H. Guo, Z.-Y. Su, Z.-Y. Zhao, X. Xiao, and F. Liu, "An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed," *Appl. Energy*, vol. 162, pp. 808–826, Jan. 2016, doi: 10.1016/j.apenergy.2015.10.145.
[12] S. R. Moreno and L. S. Coelho, "Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system," *Renew. Energy*, vol. 126, pp. 736–754, Oct. 2018, doi: 10.1016/j.renene.2017.11.089.
[13] R. Rahmani, R. Yusof, M. Seyedmahmoudian, and S. Mekhilef, "Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting," *J. Wind Eng. Ind. Aerodynamics*, vol. 123, pp. 163–170, Dec. 2013, doi: 10.1016/j.jweia.2013.10.004.
[14] T. Ouyang, X. Zha, L. Qin, Y. Xiong, T. Xia, and H. Huang, "Short-term wind power prediction based on kernel function switching," *Electr. Power Autom. Equip.*, vol. 36, no. 9, pp. 480–486, 2016, doi: 10.16081/j.issn.1006-6047.2016.09.012.
[15] D. Gan, D. Ke, Y. Sun, and M. Cui, "Short-term probabilistic wind power forecast considering ramp characteristics," *Electr. Power Autom. Equip.*, vol. 36, no. 4, pp. 145–150, 2016, doi: 10.16081/j.issn.1006-6047.2016.04.022.
[16] B. Wang, S. L. Feng, and C. Liu, "Wind power forecasting method based on weather classification," *Power Syst. Technol.*, vol. 38, no. 1, pp. 93–98, Jul. 2014, doi: 10.13335/j.1000-3673.pst.2014.01.015.
[17] J. Yan, Y. Q. Liu, and H. Zhang, "Dynamic wind power probabilistic forecasting based on wind scenario recognition," *Modern Electr. Power*, vol. 33, no. 2, pp. 51–58, Apr. 2016, doi: 10.3969/j.issn.1007-2322.2016.02.009.
[18] C. Zhang, Q. L. Tan, and S. Y. Tang, "Classification and combination prediction model for wind power based on change period of wind speed," *Acta Energiae Solaris Sinica*, vol. 38, no. 4, pp. 999–1006, Jun. 2017. [Online]. Available: http://en.cnki.com.cn/Article_en/CJFDTOTAL-TYLX201704019.htm
[19] X. Xiong, G. Y. Huang, and X. D. Wang, "Adaptive robust cubature Kalman filter integrated navigation algorithm based on chi-square test," *J. Detection Control*, vol. 41, no. 5, pp. 125–134, Oct. 2019. [Online]. Available: http://www.cnki.com.cn/Article/CJFDTotal-XDYX201905022.htm
[20] C. Chen, N. Wang, H. Y. Yu, "Evaluation model of converter operation process based on chi-square boxing method and logistic regression algorithm," *J. Mater. Metall.*, vol. 18, no. 2, pp. 87–91, Apr. 2019, doi: 10.14186/j.cnki.1671-6620.2019.02.002.
[21] Y. M. Wei, X. S. Gan, and H. P. Li, "Chi-square discretization algorithm of continuous attribute based on information entropy," in *Proc. 5th Int. Conf. Comput., Automat. Power Electron.*, Shenyang, Liaoning, 2018, pp. 126–131. [Online]. Available: https://www.webofproceedings.org/proceedings_series/ECS/CAPE%202017/CAPE_1113024.pdf
[22] A. F. Sumantri, I. Wijayanto, R. Patmasari, and N. Ibrahim, "Motion artifact contaminated functional near-infrared spectroscopy signals classification using K-nearest neighbor (KNN)," in *Proc. Int. Conf. Electron. Represent. Algorithm*, Jakarta, Indonesia, 2019, pp. 1–8. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1201/1/012062/pdf
[23] G. F. Fan, Y. H. Guo, J. M. Zheng, and W. C. Hong, "Application of the weighted K-nearest neighbor algorithm for short-term load forecasting," *Energies*, vol. 12, no. 5, pp. 1–19, Mar. 2019, doi: 10.3390/en12050916.
[24] S. Qiao, B. Yan, and J. Li, "Ensemble learning for protein multiplex subcellular localization prediction based on weighted KNN with different features," *Int. J. Speech Technol.*, vol. 48, no. 7, pp. 1813–1824, Jul. 2018, doi: 10.1007/s10489-017-1029-6.
[25] G. Li and X. D. Jiang, "Wind power generation ensemble forecasting method based on boosted regression tree and random forest," in *Proc. CSU-EPSA*, Nov. 2018, vol. 30, no. 11, pp. 70–74. [Online]. Available: http://en.cnki.com.cn/Article_en/CJFDTotal-DLZD201811013.htm
[26] M. F. Ren and L. Q. Liu, "Property and parameter estimation of discretization of compound possion distribution and exponential distribution," *Statist. Decis.*, vol. 19, no. 11, pp. 25–29, Nov. 2019, doi: 10.13546/j.cnki.tjyjc.2019.19.005.
[27] D. S. Terzi and S. Sagiroglu, "A new big data model using distributed cluster-based resampling for class-imbalance problem," *Appl. Comput. Syst.*, vol. 24, no. 2, pp. 104–110, Feb. 2019, doi: 10.2478/acss-2019-0013.

JIE HAO (Member, IEEE) received the master's degree from the Institute of Communication and Information System, Lanzhou University. She is currently pursuing the Ph.D. degree with the School of Computer and Communication, Lanzhou University of Technology. She is a Lecturer with Northwest Minzu University. Her research interests include machine learning, intelligent system modeling, and data mining. She has presided over two central university projects, published five articles, and applied for three national patents.

CHANGSHENG ZHU is currently a Professor and a Doctoral Supervisor with the Lanzhou University of Technology. He is also the Director of the Traditional Chinese Medicine Big Data Engineering Research Center, Gansu. His research interests include high performance computing, big data mining, and engineering manufacturing information systems. In 2010, he was selected as a Leading Talent in Gansu Province.

XIUTING GUO (Member, IEEE) received the master's degree in engineering. She is currently pursuing the Ph.D. degree with the School of Computer and Communication, Lanzhou University of Technology. Since December 2012, she has been teaching with the School of Science, Lanzhou University of Technology.