Received January 20, 2019, accepted February 3, 2019, date of publication February 27, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901920

A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting

YUN JU, GUANGYU SUN, QUANHE CHEN, MIN ZHANG, HUIXIAN ZHU, AND MUJEEB UR REHMAN
School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Corresponding author: Yun Ju (50501482@ncepu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2018MS030 and Grant 2016MS35

ABSTRACT The volatility and uncertainty of wind power often affect the quality of electric energy, the security of the power grid, the stability of the power system, and the fluctuation of the power market. In this case, the research on wind power forecasting is of great significance for ensuring the better development of wind power grids and the higher quality of electric energy. Therefore, a lot of new forecasting methods have been put forward. In this paper, a new forecasting model based on a convolution neural network and LightGBM is constructed. The procedure is shown as follows. First, we construct new feature sets by analyzing the characteristics of the raw data on the time series from the wind field and adjacent wind field. Second, the convolutional neural network (CNN) is proposed to extract information from input data, and the network parameters are adjusted by comparing the actual results. Third, in consideration of the limitations of the single-convolution model in predicting wind power, we innovatively integrated the LightGBM classification algorithm at the model to improve the forecasting accuracy and robustness. Finally, compared with the existing support vector machines, LightGBM, and CNN, the fusion model has better performance in accuracy and efficiency.

INDEX TERMS Convolutional neural network, fusion model, LightGBM, ultra-short-term wind power forecasting, wind energy.

I. INTRODUCTION

Continuous consumption of fossil energy has caused severe energy shortages and environmental pollution problems. To alleviate these situations, vigorous development of clean, sustainable and renewable energy has become the major topic recently. Wind energy, as the most efficient energy source, has the advantages of zero emissions, free pollution, and no fuel cost, and becomes the third largest energy source after thermal power and hydropower. However, wind power has its disadvantages, like uncertainties, which poses great challenges to power quality, grid security, and power system stability. Accurate and timely wind power forecasting has great significance for the safe operation of the grid [1]–[3].

According to the length of forecasting time, wind power forecasting can be divided into the medium-long term, short-term and ultra-short-term forecasting. The medium-long term forecasting is mostly used for making quarterly power generation plans of grid and wind farm construction. The short-term forecasting can help with the rationality of the economic system and maintenance of the wind turbine. The ultra-short-term forecasting is theoretically the most demanding forecasting method, mainly used control the daily operation of the wind farm unit [4], [5]. According to the research object, wind power forecasting can be divided into physical methods and statistical methods. The main difference between them is that the former uses real-time data such as weather, while the latter pays more attention to the mining of historical data [6].

For statistical methods, commonly used single model prediction schemes include persistence methods [7], support vector machines [8]–[10], the linear model including

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

ARMA, ARIMA and autoregressive model [11], [12], and nonlinear models such as Fuzzy model [13], wavelet-based model [14] and ANN [15]–[19]. Since the ultra-short-term forecast just takes a few minutes as the forecasting unit, which causes the data to fluctuate drastically. Meanwhile, the single-model is sensitive to the data, none of the methods mentioned above can well solve the problem. For the poor performance of forecasting, researchers begin to pay more attention to hybrid models, including a weather ensemble prediction framework for wind power forecast was constructed [20], an ensemble learning method for sample entropy technique, empirical mode decomposition (EMD) and extreme learning machine was proposed in [21]. In recent years, good results have been achieved. The major difference between the hybrid models and the single-model is that the former uses multi-algorithm, which can make remedies to the disadvantages of each method, and acquire the complementary strengths, in consequence, much more stable and effective performance will be achieved.

One similarity of some of the statistical learning methods mentioned above is that they are shallow deep learning models, while considering the wind power data is complicated, these shallow models may not be sufficient to extract corresponding nonlinear characteristics [22]. In addition, with the rapid development of smart grids, people pay more and more attention to the importance of data, forcing us to enter the era of big-data. It is difficult to extract the deep features of wind power data by traditional methods [23]. So, in order to make full use of these data, deep learning method may be the most effective way, which can effectively learn the abstract features in the complicated data.

In summary, the instability of the single model prediction and the shallow model's unsatisfactory data feature mining have inspired us to move toward deep learning and ensemble learning. There have been many papers in this area in recent years, including model combining stacked autoencoder (SAE) and backpropagation algorithm [24], model based on wavelets theory, wavelet packets, time series analysis and ANNs [25]. The disadvantage of these deep learning methods is that the ability to search for a global optimum is limited, and the number of parameters to be substantially increases as the neural network goes deeper. Because CNN adopts the weight sharing technique, so it has fewer parameters than the previous method, and it also can effectively extract the intrinsic features in wind power data [26]. However, the author of this article only converts wind power data from 1d data to 2d and uses CNN for feature extraction. This data construction method is unreasonable and does not optimize CNN according to data. Meanwhile, the spatiotemporal correlation of wind power data is not well utilized. Paper [27] proposed a wind speed prediction method based on correlation analysis of adjacent wind turbine data, and proved the value of the correlations in this data.

The main contributions in the paper are listed as follows:

(1) By analyzing the correlation between adjacent wind turbines, a way of combining adjacent wind turbine data is proposed, it further reduces the forecasting errors from original data.
(2) Construct time-order character and introduce convolutional neural networks to explore the correlation and nonlinear characteristics of data at adjacent time points.
(3) An ensemble learning method based on LightGBM and CNN is proposed, which acquires better performance in accuracy and reliability than usual wind power forecasting model.

II. THE CONVOLUTIONAL NEURAL NETWORK MODEL

A. CNN OVERVIEW

In the deep learning theory, the most developed part is the convolutional neural network [28], which has many achievements in the field of image recognition and pattern recognition. The core of CNN is a convolutional layer, which contains a lot of different convolution kernels to extract various features. In the meanwhile, the convolutional layer cooperates with the pooling layer to reduce the number of parameters and speed up the calculation. Finally, a large number of features extracted by the convolution kernel are all passed to the fully connected layer, which is used to combine the previously extracted features to achieve the final forecasting. In this way, we can fetch the implicit information of the data to achieve faster and more stable predictions.

B. DATA NORMALIZATION

Almost all of the wind power forecasting work requires dimensionless processing of data to avoid errors caused by different units. The types of normalized methods mainly include Min-Max normalization and Zero-mean normalization.

1) MIN-MAX NORMALIZATION

It's a linear transformation method, also known as deviation standardization, that makes the result fall to the interval [0,1]. The transformation formula is as follows:

$$\mathrm{x}^{*} = \frac{x - min}{max - min} \tag{1}$$

where max is the maximum value of the sample data and min is the minimum value of the sample data.

2) ZERO-MEAN NORMALIZATION

This method needs to calculate the mean and standard deviation before transforming the original data. The processed data conforms to the standard normal distribution, that is, the mean value is zero, the standard deviation is one. The transformation function is as follows:

$$x^{*} = \frac{x - \mu}{\sigma} \tag{2}$$

where μ is the mean of all sample data and σ is the standard deviation of all sample data.

In this paper, we use Min-Max normalization to transform power data of the wind farm. Because temperature,

TABLE 1. Comparison of the model error with or without pooling layer.

Pooling layer	MSE 10e-3	MAE 10e-2
with	**1.8028**	**2.473**
without	**1.7692**	**2.395**

wind directions and other features of the data set have positive and negative values and the zero-mean normalization can retain its direction information, so we adopt zero-mean normalization.

C. CONVOLUTIONAL NEURAL NETWORK CONSTRUCTION

The construction parameters of the CNN include the size of the convolution kernel, selection of pooling layer, and the number of convolution layer. This section will focus on these topics, analyze the reasons for selection and compare with actual results.

1) ACTIVATION LAYER

The purpose of the activation layer is to change linearly inseparable problem into a separable one. This change can make the model more adaptable. Since the ReLU (Rectified Linear Unit) function always has a derivative value of 1 when the input value is greater than 0, the gradient diffusion phenomenon is easy to overcome. and the function is as follows:

$$a^{l(i,j)} = f\left(y^{l(i,j)}\right) = \max\{0, y^{l(i,j)}\} \tag{3}$$

where $\mathbf{a}^{\mathbf{l(i,j)}}$ is the activation value of the convolutional layer output $\mathbf{y}^{\mathbf{l(i,j)}}$.

2) POOLING LAYER

The purpose of the pooling layer at the beginning of the design is to enlarge the local features of the image, speed up the calculation and reduce the possibility of overfitting. In the meanwhile, it can reduce the impact of data fluctuations on predictions to some extent. The commonly used pooling layer is the max-pooling layer, and the function is as follows:

$$P^{l(i,j)} = \max_{(j-i)W+1\leq t\leq jW} \left\{a^{l(i,t)}\right\} \tag{4}$$

where $a^{l(i,t)}$ represents the activation value of the t neuron in the I layer of the l layer, and $\mathbf{P}^{\mathbf{l(i,j)}}$ represents the width of the pooled area.

The wind power data used in this paper is numerical data and theoretically does not have the above characteristics of picture data. For the difference between with and without pooling layer in the model, the following table lists the results of different structures. It can be seen that the model without the pooling layer has a better forecasting effect than the other.

3) FULLY CONNECTED LAYER

The fully connected layer is generally located in the last of the convolutional neural network. Its function is to summarize the features extracted by the previous convolution layers, and then make a forecasting. The function is as follows:

$$z^{l+1(j)} = \sum_{i=1}^{n} w_{ij}^{l} a^{l(i)} + b_j^l \tag{5}$$

where w_{ij}^l is the weight of the i-th neuron ($a^{l(i)}$), and b_j^l is the bias value.

4) THE SIZE OF THE CONVOLUTION KERNEL

The convolution layer is to extract features by sliding the convolution kernel on data and convolving with the covered data. Because this method has the property of shared weights, it reduces the risk of overfitting and increases the speed of calculation. The calculation process is as follows:

$$y^{l(i,j)} = k_i^l * x^{l(x^j)} = \sum_{j'=0}^{\omega-1} k_i^{l(j')} x^{l(j+j')} \tag{6}$$

where $\mathbf{k}_\mathbf{i}^{\mathbf{l}(j')}$ represents the j' weight of the I convolution kernel of layer l, $x^{l(x^j)}$ represents the j convolution local area in layer l, and ω is the width of the convolution kernel. There are not many parameters that can be adjusted in the convolutional layer. If the size of the convolution kernel is too large, the parameters of the network will increase, this will result in training speed decrease. But if it is too small, the accuracy of feature extraction will be affected, this will result in a worse predictive effect. Therefore, the selection of the convolution kernel size is particularly important.

Network depth has the similar considerations, too many layers of the convolutional layer will result in too many network parameters, slow training speed, and risk of overfitting. However, an overly simple network cannot guarantee sufficient learning of the power of wind power changes. At the same time, in order to reduce the noise caused by padding, the convolutional layer in the structure does not adopt padding. As a result, after the feature data entered CNN, the length of time-order characters declined rapidly, so the number of hidden layers could not be set too much. Therefore, according to this rule, we set three models with convolutional layers of one, two, and three, and compare the effects of different convolution kernel sizes. Finally, according to Table 2, for any number of CNN networks, the size of the convolution kernel is set to 3. And two-layer convolution is most suitable for the wind power data used in our experiments, effectively extracting information without over-fitting. Therefore, the CNN structure adopted in this paper is shown in Fig 1. It consists of two convolutional layers, two Dropout layers, a Flatten layer, a fully connected layer, and an output layer.

III. MODEL IMPROVEMENT

In general, the convolutional neural network has three types of hidden layer—the convolutional layer, the pooling layer, and the fully connected layer. The role of the fully connected layer is to integrate features that have been highly abstracted

TABLE 2. The errors of CNN with different structures.

The number of the convolution layer	The size of the convolution kernel	1	2	3	4	5	6	7	8
1	MSE 10e-3	1.874	1.811	**1.776**	1.780	1.788	1.797	1.827	1.838
	MAE 10e-2	2.402	2.398	**2.394**	2.390	2.404	2.412	2.420	2.440
2	MSE 10e-3	1.869	1.787	**1.766**	1.770				
	MAE 10e-2	2.395	2.393	**2.385**	2.387				
3	MSE 10e-3	1.918	1.867	**1.835**					
	MAE 10e-2	2.447	2.433	**2.415**					

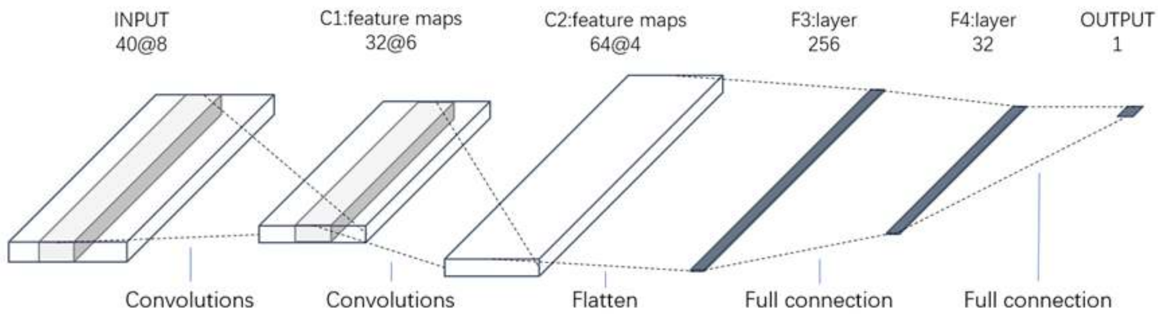

FIGURE 1. The structure of the convolutional neural network.

after multiple convolutions and normalization. These non-linear combination features are learned in a simple manner by a fully connected layer to forecast output. Essentially, the convolutional layer provides a meaningful, low-dimensional, and almost constant feature space, and then fully connected layer learns a non-linear equation in this space. However, when the amount of data is not enough or the quality of data is worse, the learning of this nonlinear equation will fall into a local optimum situation, which can be solved by replacing a fully connected layer with a stronger classifier.

A. THE INTRODUCTION OF LightGBM

LightGBM is a data model based on GBDT proposed by Microsoft in 2017. Like other boosting algorithms, GBDT combines weak learners to form a strong one. However, the decision tree used in the GBDT algorithm can only be a regression tree which is because each tree of the algorithm learns the conclusions and residuals of all previous trees. By using the residual of each predicted result and target value as the target of next learning, a current residual regression tree is obtained. And the results of multiple decision trees are added together as the final predicted output [29]. Although GBDT has achieved good learning effects on many machine learning tasks, GBDT is faced with adjustment of accuracy and efficiency with the geometric growth of data volume in recent years. LightGBM [30] algorithm has been put forward at this time. While not reducing the accuracy of the prediction, it greatly speeds up the forecasting speed and reduces the memory utilization.

The computational time of traditional GBDT algorithm is often consumed in the construction of a decision tree. The construction of a decision tree needs to find the optimal segmentation point. The general method is to sort feature values and then enumerate all possible feature points. This method wastes time and needs lots of memory. LightGBM algorithm uses an improved histogram algorithm. It divides the continuous eigenvalues into k intervals, and the division points are selected among the k values. So, it is better in training speed and space efficiency than GBDT algorithm. At the same time, the decision tree is a weak classifier. The use of the histogram algorithm will have a regularization effect and can effectively prevent overfitting.

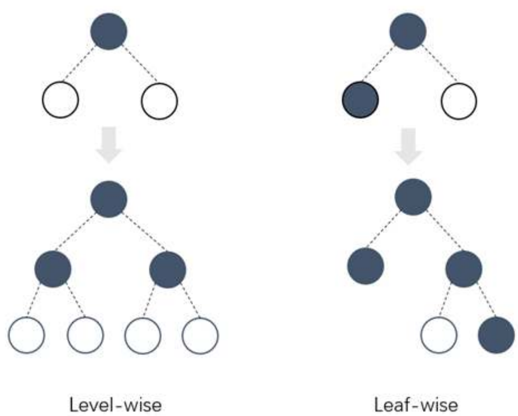

FIGURE 2. The generation strategy of tree in LightGBM.

In terms of reducing training data, LightGBM algorithm uses a leaf-wise generation strategy. Compared with the traditional method like level(depth)-wise, the leaf-wise can reduce more losses when growing the same leaf. Furthermore, the extra parameter is also used to limit the depth of the decision tree and can avoid overfitting.

In terms of reducing the number of features, the traditional and mainly used method is PCA. This method is established in the case where the features are redundant, so this method has some limitations. The EFB algorithm used by LightGBM puts many features of high-dimensional data together in a sparse feature space to avoid calculation of redundant features.

And the histogram was constructed according to the algorithm, which can accelerate calculation speed.

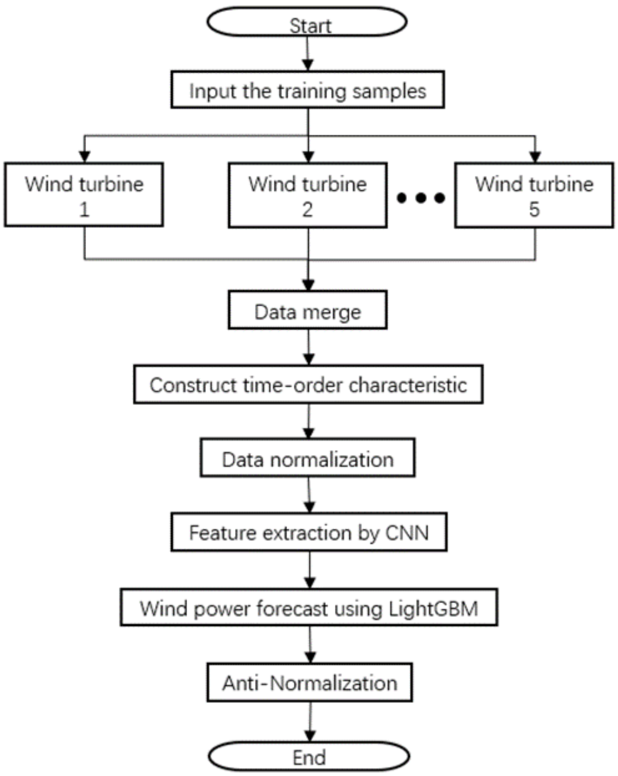

FIGURE 3. The forecasting process of ensemble learning model.

B. THE ENSEMBLE LEARNING MODEL

In this paper, we construct a model combining convolution neural network and LightGBM algorithm. The ensemble process is shown in Fig 3. The features of wind power data are extracted and filtered through the convolution layer of the convolutional neural network. Then use the model of LightGBM to fetch information and classification again on the output of flatten layer. Improve forecasting effect of the model.

The final prediction process of the ensemble learning model is as follows:

(1) According to the conclusions of the data analysis afterwards, the data of multiple wind turbines are spliced together, constructed the time series and normalized.
(2) Input the preprocessed data into the constructed CNN model for pre-training, and obtain the parameters of the convolutional layer and the fully connected layer.
(3) Freeze the parameters of the convolutional layer in the CNN model, and input the data from the Flatten layer into the LightGBM model, and retrain the model.
(4) Predict the test set data, anti-normalize the results, and output the results.

IV. DATA ANALYSIS AND PREPROCESSING

The good predictive effect of a model is often not only related to the model itself, but also to good data or good data analysis and data processing methods. They can always directly or indirectly affect the final prediction effect. The content of this chapter is to analyze the wind power data in general, so as to construct features on the basis of data. Meanwhile, these analyses will help us to better understand data and fetch information.

A. SAMPLE OVERALL EXPLORATION

The data set used in this paper is real data collected from a wind farm in the northwest of China. These data included 5 groups of wind turbine for the whole 2013 year, with a sampling period of 5min. Each sampling is the instantaneous data of the current state. The data stored in the initial data set includes time, ambient temperature, fan status, generated power, wind speed, motor speed, ID, wind direction, daily power generation, pitch angle. The features that need to be deleted included date, time and ID. Because the daily power generation is total power generation of the wind turbine between early morning and the current time point, this feature will induce the model to divide the data of each day into a certain degree. This condition for the continuous prediction of timing is unfavorable, so we convert this feature into total power generation of 5 minutes before the current time point.

Because ultra-short-term wind power forecasting is generally predicting wind power within three hours. We try to draw a numerical graph of power on time series and explore its numerical characteristics. Fig 4 (a) depicts the variation of wind power over the one day in 2013. Because the power of wind field is mainly affected by surrounding environment, like wind speed, atmospheric density, temperature and so on, thus leading the power of wind field to occur an irregular change in a very short time. The red curve in the figure is the power curve after smoothing. It can be seen from the figure that even ultra-short-term wind power has the great numerical correlation in timing, so we construct the time-order character as the model inputs.

There are often several wind turbines on a wind farm. These turbines are highly consistent due to their close proximity to weather, climate, batch size, and even operator impact. This will cause them to have a certain connection in the power generation. Fig 4(b) shows the power curve of the five adjacent wind turbines of the same wind farm on the same day. It can be seen that there is a strong correlation between them.

For the power grid, they are more concerned with the prediction of the total wind power compared to the accuracy of the single wind turbine power prediction. Therefore, as with ensemble learning, we can use this correlation to reduce the interference that may exist in the single wind turbine power data. So, we combine the characteristics of adjacent

TABLE 3. Description and unit of data features used in the experiment.

Feature	Description	Unit
Temperature	The temperature near the wind turbine	℃
Fan state	1——Normal operating state 2——Sleeping state of wind speed too low 3——Protection state of wind speed too fast	—
Wind speed	Measured by a digital anemometer outside the wind turbine cabin	m/s
Motor speed	The speed of the fan generator shaft	r/min
Wind direction	The deviation angle between the wind direction and the center line of the engine room is mainly measured by the wind vane installed on both sides of the top of the engine room.	°
5 minutes of power generation	5 minutes of power generation before the current time point	KW·h
Pitch angle	The angle between the chord line of the airfoil and the rotating plane	°
power	Instantaneous power generation of wind turbines	KW

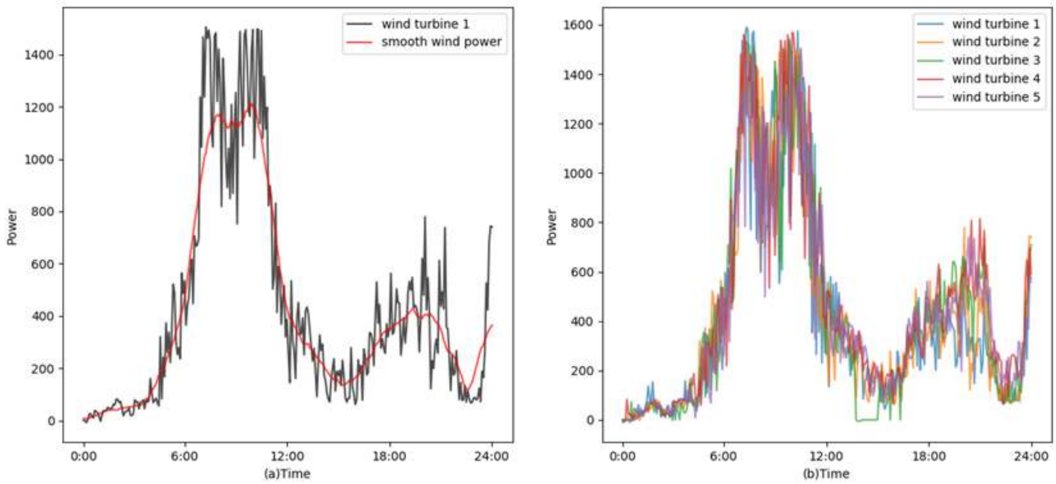

FIGURE 4. The daily variation graph of wind power.

TABLE 4. Description and unit of data features used in the experiment.

	wind turbine 1	wind turbine 2	wind turbine 3	wind turbine 4	wind turbine 5	Error between actual value and mean predicting value in 5 single models	Error between actual value and predicting value by 5 group data sets
MSE 10e-3	4.345	4.765	4.268	4.558	4.703	**2.175**	**1.766**
MAE 10e-2	3.703	3.936	3.691	3.748	3.754	**2.766**	**2.385**

wind fields at the same time point as the characteristics of a time point. And, their power is averaged as output power. To demonstrate the effectiveness of the method in a subsequent experiment, we compared the error between the prediction of the five wind turbines, the error between actual value and mean predicting value in 5 single models and the error between actual value and predicting value by 5 group data sets. We put the results in Table 4. By comparing the error between the prediction error of the first five sets of single wind turbine power and the average error of the group 6, we can see that the addition of multiple wind turbines to predict wind field power can largely compensate for the impact of single wind turbine data fluctuations on forecasting results, and the relationship between single wind turbine can also be learned. These can be seen through the comparison of the data in groups 6 and 7. To some extent, this method can reduce the disturbance of single wind turbine power data and improve the model robustness.

B. TIME-ORDER CHARACTER

From the first subsection, we can see that there is a timing correlation between the values of wind power. In this

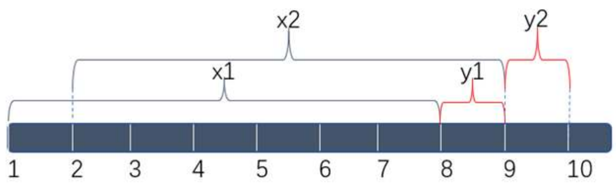

FIGURE 5. The structure of a time-order character.

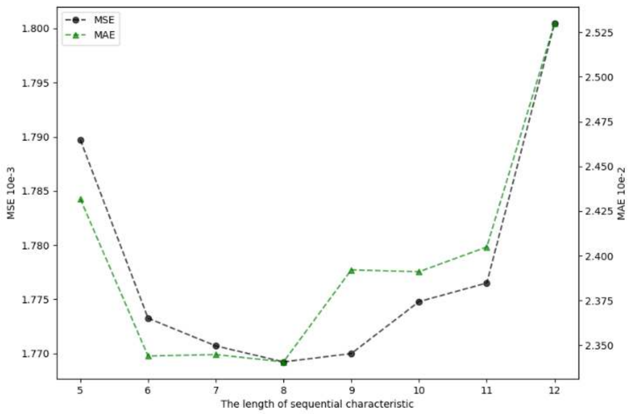

FIGURE 6. The error curve of different lengths time-order characters.

subsection, we will build time-order character according to the correlation, and the time-order character will be used to predict the power of the next time point. At the same time, the power of all time points is predicted by sliding-window prediction and compared with the actual value. For example, as shown in figure 5, all the features of the first eight data are selected as the first input x1 of a model to predict the power y1 of the next time point. In the same way, we select the data from the second to the ninth time point as the next input x2 to predict y2.

Although the CNN can better learn the temporal correlation in wind power data by building time-order characters, if the time-order characters are too long, there will be too much redundant data in the features. These redundant data will reduce the accuracy of the prediction.

Fig.6 shows the error magnitude of MSE and MAE corresponding to the prediction of time-order characters of different lengths by CNN. The horizontal axis is the length of time-order characters and the vertical axis is the error magnitude. It can be seen that the first half of the error curve shows a downward trend, and the second half of the error curve starts to rise with the increase in length. According to the curve, both MSE and MAE errors are small at the same time, so we set the length of time-order character to eight. As thus, we can make use of the original data.

V. RESULTS ANALYSIS

In the experiment, to fully verify the proposed model's performance, two widely used forecast accuracy evaluation criteria are chosen to compare with SVM, LightGBM, DNN, CNN and CNN_LGBM. The smaller the value, the better forecasting effect. These criteria are the Mean Absolute Error (MAE), and the Mean Square Error (MSE). Each criterion is defined using the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

The data split ratio adopted by the test was 8:1:1. That is, eighty percent of original data are used as the training set to train the model parameters, ten percent of the original data is used as the validation set, which is used for model optimization during model training, and another ten percent of original data is used as the test set to test the forecasting effect of model.

Since we added the Dropout layer and the early stop mechanism in CNN to prevent the model from overfitting, the results of the experiment have some randomness. So, we trained each model separately for 10 times to prevent the contingency of a single forecasting and to guarantee the fairness of the result. Meanwhile, in order to ensure the validity of the evaluation criteria, the input of all models in the experiment is the same. The parameters used in all the models are tuned by grid search or experiment, and the parameters of the ensemble learning model are the same as the parameters of the single model, which ensures the validity of ensemble learning. The key parameters of all the models used in the experiment are summarized in table 5.

Table 6 shows the error for different model predictions.

(1) It can be seen from the Mean Square Error (MSE) that the three deep learning models have different degrees of advantages over LGBM and SVM, indicating that in the case of more complex data, machine learning methods are not sufficient to learn non-structural features and it is difficult to predict accurately. Although the ensemble learning model is inferior to CNN in two out of ten results, the overall trend of the ensemble learning model is better. Facing data extracted through CNN, it is often more effective to use simple machine learning methods to predict than the fully connected layer.

(2) It can be seen from the Mean Absolute Error (MAE) that the performance of the CNN and the DNN is slightly worse than the LGBM. However, our ensemble learning model effectively overcomes this defect and has lower errors. The reason for this phenomenon is that the CNN and DNN prediction errors of very small samples are much larger than the traditional machine learning methods. This also proves that the deep learning method is easy to over- when the training data is not good enough. The effect of other models is obviously better than the SVM which indicates that SVM is not adaptable enough to deal with wind power, and does not dominant in accuracy and speed;

(3) At the end of the table is the average results of the model running 10 times. On the MSE error, the ensemble learning method has a 4%2% 2.5% improvement

TABLE 5. The parameter values for different models.

Model	Parameter	Description	Value
SVM	c	Punishment coefficient of the SVM	1.0
	kernel	Specifies the kernel type to be used in the algorithm.	rbf
LightGBM	num_leaves	The maximum number of leaves on a tree	25
	learning_rate	Improve learning rate	0.1
	early_stopping_rounds	If the model loss does not drop in the specified rounds, the training will be stopped.	10
CNN	n1 n2	The convolution kernel number of convolution layers	32
	n3 n4	The neuron number of FC layers	32/1
	convolution kernel size	The length of the 1D convolution window	3
	padding	The padding type	'valid'
	activation	Activation function to use	'ReLU'
	strides	The stride length	1
DNN	n1 n2 n3 n4 n5	The neuron number of different layers	128/128/128/10/1
	activation	Activation function to use	'ReLU'
Ensemble		The parameters are consistent with CNN and LightGBM	

TABLE 6. The results of different models.

	MSE (10e-3)					MAE (10e-2)				
Model	SVM	DNN	LightGBM	CNN	Ensemble	SVM	DNN	LightGBM	CNN	Ensemble
1	2.446	1.806	1.848	1.752	1.759	3.535	2.393	2.315	2.344	2.321
2	2.446	1.788	1.843	1.770	1.727	3.534	2.394	2.333	2.432	2.273
3	2.445	1.792	1.844	1.763	1.776	3.532	2.387	2.305	2.345	2.246
4	2.448	1.775	1.807	1.765	1.774	3.537	2.373	2.295	2.341	2.291
5	2.446	1.791	1.841	1.770	1.724	3.535	2.406	2.303	2.392	2.283
6	2.444	1.806	1.813	1.764	1.730	3.532	2.404	2.296	2.391	2.267
7	2.446	1.798	1.814	1.738	1.732	3.535	2.435	2.323	2.335	2.297
8	2.446	1.807	1.836	1.823	1.788	3.535	2.396	2.305	2.530	2.298
9	2.445	1.791	1.833	1.756	1.728	3.532	2.397	2.308	2.331	2.236
10	2.443	1.770	1.796	1.759	1.741	3.530	2.398	2.272	2.408	2.292
Average	2.445	1.792	1.827	1.766	1.748	3.534	2.398	2.306	2.385	2.280

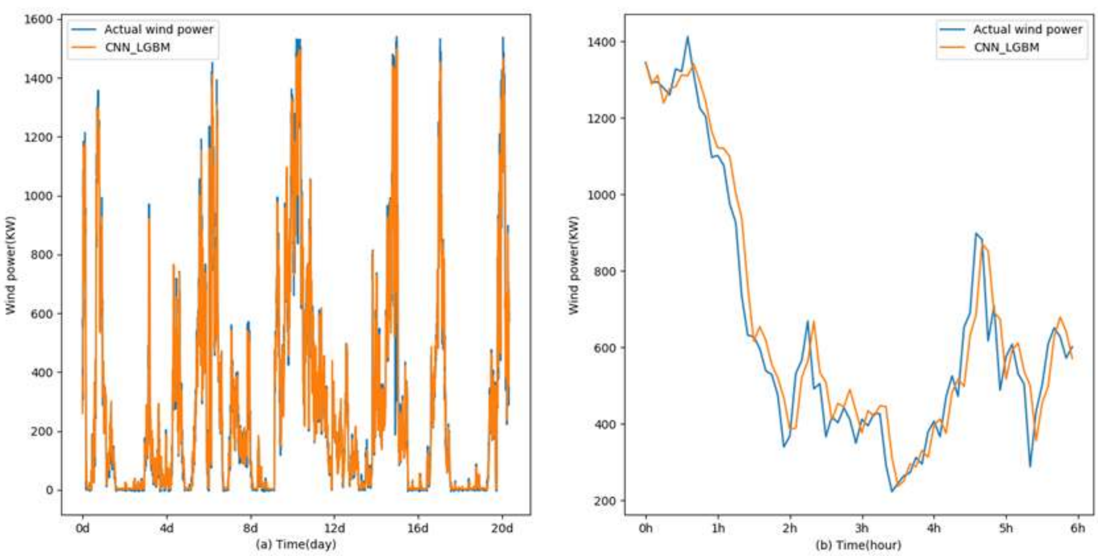

FIGURE 7. The whole test set and last six hours' forecasting curve.

compared to the LGBM CNN DNN. There are also 1%4.4% and 5% improvements on MAE. At the same time, MSE and MAE increased by 28.5% and 35% respectively compared with SVM;

Fig 7 shows the forecasting results of the test set by the fusion model. Fig 7(a) is the forecasting curve for the whole test set, and the Fig 7(b) is the forecasting curve for the last 6 hours. It can be seen from the figure that even if the wind power changes in a short period of time, the ensemble learning model can extract the conventional information and have a good prediction effect.

VI. CONCLUSION

This paper proposed a wind power forecasting model based on CNN and LightGBM algorithm for high accuracy of wind power forecasting. By analyzing the diurnal variation numerical curves of wind power, it is determined that wind power has a great correlation in time series. Then combined with forecasting results of the CNN, the optimal length of the time-order character is determined. By calculating the Pilton correlation coefficient of adjacent wind turbines power, the characteristics of adjacent wind fields are constructed to improve the stability of the forecasting model.

An ensemble learning model of convolutional neural network and LGBM is constructed to achieve higher accuracy based on above. Finally, comparing the errors between the actual and forecast values of different models, we can see that CNN can effectively fetch information of wind power data, LightGBM can summarize this information and improve model robustness. It also shows that the ensemble learning model of this paper has better prediction accuracy than the traditional prediction model when dealing with the volatility of ultra-short-term wind power.

One of the limitations of this paper is that whether it is CNN or LGBM, whether it is a single model or an ensemble learning model, their robustness is difficult to guarantee for abnormal data and false data, but compared to two single models, the model generated by ensemble learning is still better, but it will consume more computing resources. Fortunately, the computational complexity of these two algorithms is actually relatively small, and the main purpose of this paper is to explore the performance of ensemble learning methods. However, if we can solve the above problems, it will be more conducive to the improvement of the algorithm, this part of the task will be placed in our follow-up work.

REFERENCES

[1] M. Mar iukaitis, I. utautaite, L. Martišauskas, B. Jokšas, G. Gecevi ius, and A. Sfetsos, "Non-linear regression model for wind turbine power curve," *Renew. Energy*, vol. 113, pp. 732–741, Dec. 2017.

[2] B. Ernst *et al.*, "Predicting the wind," *IEEE Power Energy Mag.*, vol. 5, no. 6, pp. 78–89, Nov./Dec. 2007.

[3] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "Wind power forecasting: State-of-the-art 2009," Argonne Nat. Lab., Lemont, IL, USA, Tech. Rep. ANL/DIS-10-1, 2009.

[4] J. M. Lujano-Rojas, G. J. Osório, J. C. O. Matias, and J. P. S. Catalão, "A heuristic methodology to economic dispatch problem incorporating renewable power forecasting error and system reliability," *Renew. Energy*, vol. 87, no. 3, pp. 731–743, 2016.

[5] P. Kou, F. Gao, and X. Guan, "Stochastic predictive control of battery energy storage for wind farm dispatching: Using probabilistic wind power forecasts," *Renew. Energy*, vol. 80, no. 8, pp. 286–300, Aug. 2015.

[6] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009.

[7] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, "A fuzzy model for wind speed forecasting and power generation in wind parks using spatial correlation," *IEEE Trans. Energy Convers.*, vol. 19, no. 2, pp. 352–361, Jun. 2004.

[8] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed forecasting," *Renew. Energy*, vol. 29, no. 6, pp. 939–947, 2004.

[9] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 52, no. 4, pp. 1990–1998, 2011.

[10] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach," *Appl. Energy*, vol. 113, pp. 690–705, Jan. 2014.

[11] J. L. Torres, A. García, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Sol. Energy*, vol. 79, no. 1, pp. 65–77, 2005.

[12] L. Jian, Y. Zhao, Y.-P. Zhu, M.-B. Zhang, and D. Bertolatti, "An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China," *Sci. Total Environ.*, vol. 426, pp. 336–345, Jun. 2012.

[13] R. Iqdour and A. Zeroual, "A rule based fuzzy model for the prediction of daily solar radiation," in *Proc. Int. Conf. Ind. Technol.*, Hammamet, Tunisia, Dec. 2004, pp. 1482–1487.

[14] G. Capizzi, F. Bonanno, and C. Napoli, "A wavelet based prediction of wind and solar energy for Long-Term simulation of integrated generation systems," in *Proc. Int. Symp. Power Electron., Elect. Drives, Autom. Motion*, Pisa, Italy, Jun. 2010, pp. 586–592.

[15] M. C. Mabel and E. Fernandez, "Analysis of wind power generation and forecasting using ANN: A case study," *Renew. Energy*, vol. 33, no. 5, pp. 986–992, 2008.

[16] S. Han, J. Li, and Y. Liu, "Tabu search algorithm optimized ANN model for wind power prediction with NWP," *Energy Procedia*, vol. 12, no. 39, pp. 733–740, 2011.

[17] P. Ramasamy, S. S. Chandel, and A. K. Yadav, "Wind speed prediction in the mountainous region of India using an artificial neural network model," *Renew. Energy*, vol. 80, pp. 338–347, Aug. 2015.

[18] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "An artificial neural network approach for short-term wind power forecasting in Portugal," in *Proc. 15th Int. Conf. Intell. Syst. Appl. Power Syst.*, Curitiba, Brazil, 2009, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5352853&isnumber=5352811. doi: 10.1109/ISAP.2009.5352853.

[19] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[20] S. Alessandrini, S. Sperati, and P. Pinson, "A comparison between the ECMWF and COSMO ensemble prediction systems applied to short-term wind power forecasting on real data," *Appl. Energy*, vol. 107, pp. 271–280, Jul. 2013.

[21] G. Zhang, Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H.-C. Iu, "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2706–2715, Sep. 2015.

[22] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renew. Energy*, vol. 85, pp. 83–95, Jan. 2016.

[23] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[24] R. Jiao, X. Huang, X. Ma, L. Han, and W. Tian, "A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting," *IEEE Access*, vol. 6, pp. 17851–17858, 2018.

[25] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks," *Appl. Energy*, vol. 107, pp. 191–208, Jul. 2013.

[26] H.-Z. Wang, G.-Q. Li, G.-B. Wang, J.-C. Peng, H. Jiang, and Y.-T. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Appl. Energy*, vol. 188, pp. 56–70, Feb. 2017.

[27] Y. S. Wang and Z. Q. Su, "A prediction method for wind speed based on the adjacent wind power generator measured data correlation analysis," (in Chinese), *J. North China Electr. Power Univ., Natural Sci. Ed.*, vol. 42, no. 2, pp. 91–96, 2015.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[30] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.

YUN JU received the master's and Ph.D. degrees from the School of Control and Computer Engineering, North China Electric Power University, Beijing, China, in 2006 and 2014, respectively, where he is currently a Lecturer. He has authored or co-authored more than 50 publications. His current research interests include control theory and control engineering. He received the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award, in 2011.

MIN ZHANG is currently pursuing the master's degree with North China Electric Power University. Her research interests include smart grid and data mining based on deep learning.

GUANGYU SUN is currently pursuing the master's degree in computer with North China Electric Power University. His research interests include machine learning and load forecasting.

HUIXIAN ZHU is currently pursuing the master's degree with North China Electric Power University. Her research interests include data analysis and signal processing based on deep learning.

QUANHE CHEN is currently pursuing the master's degree in computer with North China Electric Power University. His research interest includes load forecasting based on deep learning.

MUJEEB UR REHMAN was born in Multan, Punjab, Pakistan, in 1994. He received the B.Sc. degree in information technology from the University of Education, Lahore, in 2016. He is currently pursuing the master's degree in computer application technology with North China Electric Power University, Beijing. His research interests include tomography and signal processing.