Wind Power Prediction Based on a Convolutional Neural Network

Anwen Zhu, Xiaohui Li, Zhiyong Mo, Huaren Wu
School of Electrical Engineering and Automation
Nanjing Normal University
Nanjing, China
e-mail: 568933009@qq.com

***Abstract*—Wind power has recently become one of the most important renewable energy sources due to its advantages including less pollution, flexible investment, short construction period and less land occupation. The uncertainty of the speed and direction of wind causes wind power prediction to be extremely difficult to wind power generation. The Convolutional Neural Network (CNN) has the advantage of big data processing. CNN addresses data in the form of a two-dimensional matrix and is widely applied in the field of image processing. This paper applies CNN to wind power prediction. With historical data of wind power from a wind farm as input, this paper sets and trains the CNN model in MATLAB. The results of the prediction prove the feasibility of CNN applied to regression prediction.**

Keywords-wind power; convolutional neural network; regression prediction

I. INTRODUCTION

With population growth and the development of science and technology, energy shortages due to increased power consumption and environmental pollution have become primary issues in modern society. Using renewable energy, such as wind power, solar power, tidal power, etc., is one of the effective solutions to energy and environmental problems; these solutions are also in accordance with the principles of sustainable development strategy. Due to its advantages of less pollution, flexible investment, shorter construction period and less land occupation, wind energy has become one of the most important modern day renewable energy sources. However, wind energy has its limitations, mainly affected by meteorological conditions, which result in the uncertainty of power output; therefore, the power system will have operational difficulty and the quality of the power supply will be affected. This has become a major problem of wind power generation. Furthermore, when large-scale wind farms are connected to the power grid, wind power, with strong volatility and intermittent operations, will have great effects on the distribution, dispatching mode, stability, reactive compensation, peak load and frequency modulation of the power grid. To solve this problem, it is necessary to achieve accurate prediction of wind power. The prediction accuracy determines the probability of lowering the spinning reserve capacity of the grid, reducing the cost of the wind power generation system, and it provides the basis for power grid dispatching, so that the power grid can operate reliably, safely and economically.

As the power supply of wind turbines, wind speed plays an important role in the wind power fluctuation. The main methods of wind speed and wind power prediction at present are the continuous prediction method, the stochastic time series method, the spatial correlation method, the Kalman filter method, the wavelet analysis, the support vector machine (SVM), the artificial neural network (ANN), etc. [1] discusses short-term wind speed and power prediction, comprehensively using the wavelet analysis, the stochastic time series method, the chaos theory and the artificial neural network. The accuracy of the predicted values form single models that still need to be improved. So [1] combines the single models and calculates the combined values of wind speed and wind power. Finally, the combined results are used to predict the output power of wind turbines, so that the accuracy can be further improved. [2] builds a model based on the wavelet analysis and the least squares support vector machine (LS-SVM), and carries out a confidence test according to wind speed distribution characteristics and the confidence theory. With the test results, appropriate decisions will be made in order to avoid the errors caused by a single prediction. However, the methods above have their own limits. Results of the continuous prediction method are not stable. The Kalman filter method has a difficulty to estimate the statistics of noise, and LS-SVM is short of sparsity. Relatively, ANN has a ability of self-learning, self-organization and self-adaptation, realizing prediction by adjusting the threshold of the internal nodes. It has a broad application prospect in the field of wind power prediction.

As one of the ANN algorithms, deep learning is developing rapidly and has garnered much interest in recent years. Compared with other methods, the prediction errors of the deep neural network are at a lower level in most cases, and the prediction effect is better [3], [4] introduces a fast learning algorithm for Deep Belief Network (DBN) and [5] applies DBN to wind speed prediction for the first time, and good results are achieved, which lays the foundation for further study in the field of wind speed prediction.

A convolution neural network (CNN) is one of the most important methods of deep learning and is a popular research direction in the field of image recognition and voice analysis. Its weight sharing network structure, which is similar to a biological neural network, can reduce the complexity of the network model and the number of weights. If the network input is a multidimensional image, the advantage will be more obvious that it can put the image directly into the network, so that the complex process of feature extraction and data reconstruction in the traditional recognition algorithm can be avoided. As a multi-layer perceptron, which is specially designed for the recognition of two-dimensional shapes, CNN has a high degree of invariance for translation, scaling, inclination, or deformation in other forms [6]. The stuctures of CNN implementatioins usually include 5 hidden layers between the input layer and outout layer, so does this paper.

Ref. [7] introduces the concept, classification and structure of CNN, improves the algorithm and applies it to optical character recognition and traffic sign recognition. [8] gives a gradual network expansion algorithm and applies CNN to the recognition of handwritten numerals on gray images and the wood defects on color images. It is proved that the accuracy and efficiency of the CNN algorithm are better than those of other mainstream algorithms such as SVM.

According to the needs of wind power generation and the development of the power system, this paper studies the wind speed and power generation of wind farms and puts forward the short-term prediction of wind power with CNN. Taking the Elia [9] wind farm as an example, of which the installed wind power generation capacity is 1960.91 MW in 2016, this paper establishes the CNN model, pre-processes the historical data, inputs and trains the network, sets the parameters, and then predicts wind power for the proceeding 4 hours.

II. CNN Applied to Regression

CNN, proposed by New York University professor Yann LeCun in 1988 [10], is widely used in the field of image recognition, such as handwriting recognition, traffic sign recognition, etc. CNN is a multi-layer perceptron with good fault tolerance and self-learning ability. It has a fast running speed, good adaptive performance and high resolution [11]. This chapter introduces the structure and training methods of CNN.

A. *Structure of CNN*

CNN is a feed-forward neural network. It extracts features from a two-dimensional image and uses the back-propagation algorithm to optimize the network structure, to solve the unknown parameters in the network. The samples are put into the network, and the required features are extracted through pre-processing. Then, the classification or regression will be done in order to obtain the output. The progress is shown in Figure 1.

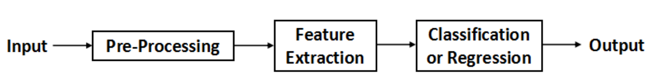

Figure 1. Structural model of CNN.

1) *Convolution*

The neural nodes of each layer in the BP neural network are arranged in one dimension and are fully connected between each layer [12]. In CNN, the neurons between each layer are no longer fully but locally connected. That is, the neurons of each hidden layer are only connected to a part of the input layer neurons.

In biology, when receptors are simulated, they become excited and transmit the sensory information to the superior center through the neurons in the sensory organ. The stimulation area of neurons is called the receptive field, the concept of which is also applied to CNN [13]. In CNN, all the convolution filters act repeatedly on the receptive field and extract the local features of the images with the convolution kernel of the image. The convolution outcome forms a feature map of the input image. Each convolution filter shares the same parameters, including the same weight matrix and bias, which is called weight sharing. The advantage of weight sharing is that it does not take the location of the local features into account in the image feature extraction and greatly reduces the number of parameters [14].

Suppose the size of a high-resolution image X_{large} is r×c. First, obtain a low-resolution image X_{small} from X_{large} and calculate the following:

$$f = \sigma(W^{(1)} X_{small} + b^{(1)}) \tag{1}$$

and find k features. σ is a activation function, such as sigmoid, $W^{(1)}$ and $b^{(1)}$ represent the weight and bias between the visual layer and the hidden layer, respectively.

For each low-resolution image X_S, the size of which is m×n, calculate the following:

$$f_s = \sigma(W^{(1)} X_s + b^{(1)}) \tag{2}$$

and find the k×(r-m+1)×(c-n+1) convolved feature matrix.

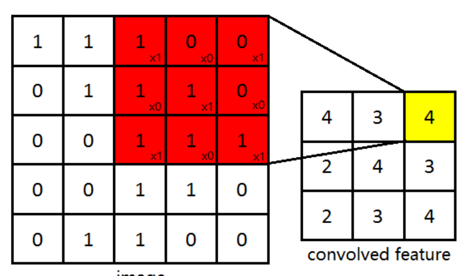

Figure 2. Convolution process.

Figure 2 shows the convolution process. The size of the input image is 5×5, and the kernel size is 3. That is, each low-resolution image with a size of 3×3 gets 1 feature. After the convolution, the size of the feature matrix is 3×3.

2) *Pooling*

Pooling is a proper process of CNN. According to different calculation methods, there are two types of pooling: average pooling and maximum pooling.

Images are static, and every image part has the same characteristics. Thus, the pooling method does the same calculation, such as average or maximum calculation, on each section of the high-resolution images. After pooling, the dimension of the characteristic statistics is greatly decreased and the generalization ability of the model is increased. The result is optimized and has a lower probability of over fitting.

Figure 3 shows the pooling process. There are 8×8 convolved features. The scale is chosen to be 2. The number of features decreased to 4×4 after pooling.

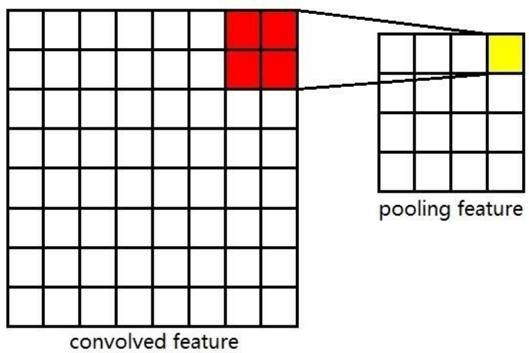

Figure 3. Pooling process.

Pooling has translating invariance. When sections of the image are linked together and the pooling rules are the same, the features will not change after the image moves a certain distance. In many practical applications, such as voice recognition and object detection, obtaining features with translating invariance is necessary.

3) CNN model

The basic structure of CNN consists of two kinds of neurons: the convolution layer and the pooling layer. The neurons in the convolution layer are locally connected with the previous layer, and their local features are extracted. The pooling layer is used to find the local sensitivity and extract the features again. The alternate occurrence of the convolution layers and the pooling layers reduces the feature resolution and the number of network parameters that need to be optimized.

Figure 4 shows the structure of CNN. It can be used for classification or regression. The map size of an input sample is 20×20. The kernel size is 5. The scale is 2. The first convolution layer C1 consists of 6 feature maps, the map sizes of which are 20 – 5 + 1 = 16. The first pooling layer S1 also consists of 6 feature maps, the map sizes of which are 16 / 2 = 8. The second convolution layer C2 consists of 12 feature maps, the map sizes of which are 8 – 5 + 1 = 4. The second pooling layer S2 also consists of 12 feature maps, the map sizes of which are 4 / 2 = 2. The 12 × 2 × 2 = 48 features are put in order, and 10 output data points are obtained through a single-layer perceptron applied to classification or regression [15].

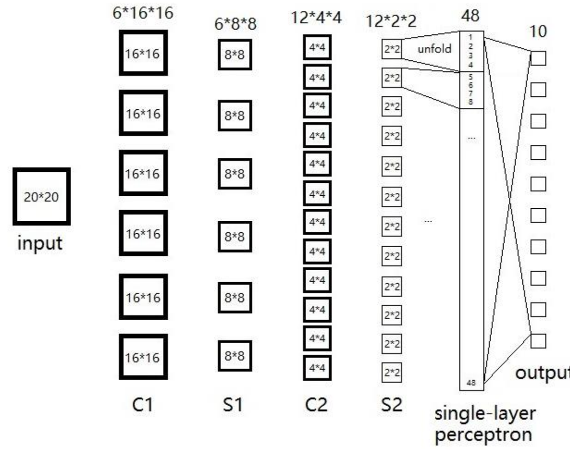

Figure 4. CNN structure.

III. WIND POWER PREDICTION

A. Regression Method

CNN typically addresses data in the form of a two-dimensional matrix and is widely used in the field of image processing. This paper applies CNN to wind power prediction, processing a one-dimensional array of data. In the pre-processing, the one-dimensional data are rearranged into a two-dimensional matrix, which conforms to the CNN input form. The property file and the response file are then created. The property file is the input of CNN, and the data of the response file are the expected output value. Each line of the property file and the response file make up a sample. With a sufficient number of samples to train the CNN, weights and biases can be obtained. The regression results are compared with the response values. The training will not end until the error reaches the expected value. When the training is done, the trained CNN model can be used to obtain the desired prediction results.

B. Wind Power Prediction

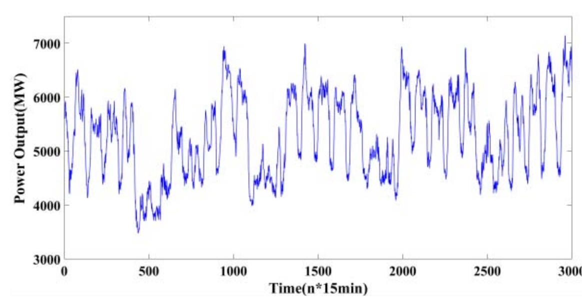

Figure 5. Power output of Elia wind farm.

Elia is a transmission system operator in Belgium. The load, interconnection, generation and balancing data can be downloaded on the elia.be website [9]. Based on the historical data of wind power in 2015 and 2016 from elia.be, the model is established with the CNN from the deep learning toolbox. After training and testing, the model gives a multi-step prediction for wind power in the proceeding 4 hours. The power output of the wind farm is shown in Figure 5.

First, pre-process the original data. The wind farm records the wind power every 15 minutes. Take the 35000 data points and normalize them with equation (3)

$$p_i = \frac{p_i^{'} - p_{\min}}{p_{\max} - p_{\min}} \tag{3}$$

where $p_i^{'}$ is the measured value of wind power.

Then, the property file and the response file will be created. The property file has 20 × 20 = 400 data in each line, which means there are 400 inputs that are a continuous sequence of sampling every 15 minutes. The response file has 16 data in each line, which stand for the wind power data for the proceeding 4 hours. Each row of the property file and the response file make up a sample. 30000 samples are used for training, and the latter 1000 samples are used to test the CNN model.

Then, the model is ready to be trained. The CNN model structure is shown in Figure 4. The size of the input matrix is 20×20, so the 30000×400 training sample matrix needs to be reshaped to a 20×20×30000 matrix. The sigmoid function is also used on the neurons of the last layer. First, initialize the parameters of the CNN model randomly. Then, the training samples are used to train the model and optimize the parameters. During the training, 400 properties of each sample will undergo the convolution and pooling process twice, then the perceptron in the last layer gets 16 outputs. The difference value between the 16 outputs of the CNN and the corresponding 16 sample response values is calculated. The BP algorithm is used to calculate and transfer the error of the neural network, and then the gradient is calculated to obtain the weight modifier. Finally, the gradient is added to the original model to update the parameters of the model. After many generations of training, the parameters are determined.

When the training is done, the test samples are used to test the prediction effect of the CNN model. With 400 wind power historical data as inputs, 16 regression prediction outputs are obtained, which stand for the output power in the proceeding 4 hours. Figure 6 and Figure 7 show two results of the multi-step wind power prediction using CNN. The red curve is the actual wind power, and the blue curve is the predictive value.

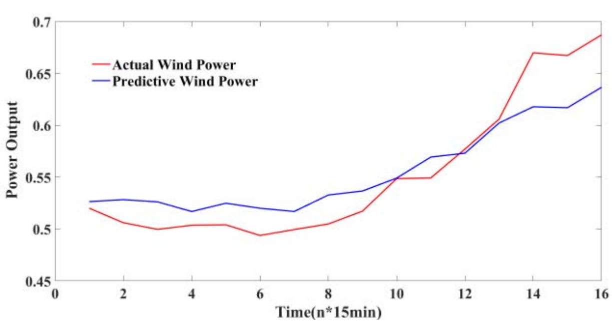

Figure 6. First result of the wind power prediction.

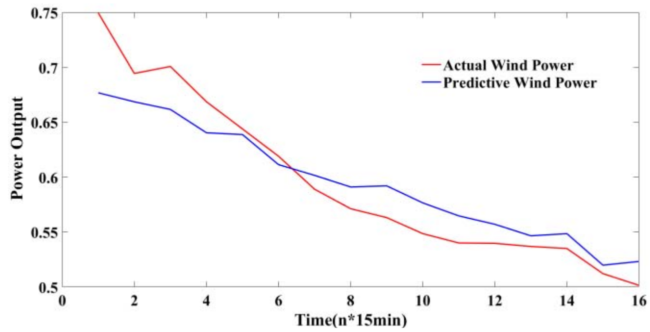

Figure 7. Second result of the wind power prediction.

The computational formula of MSE is:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(p_{fi} - p_i\right)^2 \tag{4}$$

where N = 16, p_i stands for the normalized wind power in practice， p_{fi} stands for the outcome of prediction.

MSE = 0.00076 in Figure 6. MSE = 0.00078 in Figure 7. The results show that the predicted value and the actual value coincide, so CNN may be applied to regression prediction.

IV. CONCLUSION

Wind energy has become one of the most important renewable energy sources in the world because of its advantages with less pollution, flexible investment, short construction period and less land occupation. The uncertainty of wind speed and wind direction makes it difficult to predict wind power, however. This paper presents a method that applies CNN to regression prediction. First, the wind power data arranged in the time dimension are reshaped into a two-dimensional matrix in accordance with the CNN input form, when the property file and the response file are generated. The property file is taken as the CNN input, and the response file is used to compare with the predicted data. Each line of the property file and the response file makes up a sample. The weights and bias of the network are set, and enough samples are used to train the model. The training results should be compared with the response values, and training will not end until the error reaches the expected value. Finally, the latest historical data is input and predicted results from the model are obtained.

This paper applies CNN to wind power prediction. The wind power of Elia wind farm is taken as the original data. After training and testing the network, the CNN model outputs the predicted wind power value for the next 4 hours. The result shows that CNN can be applied to regression prediction.

In this paper, the convolution neural network is used to predict wind power for the first time, and some experimental results are achieved, which applies deep learning theory to the power system.

REFERENCES

[1] LI Junhao. “Research on Short-term Prediction Methods of Wind Speed and Wind Power at Wind Farm,” Thesis, Hebei University of Technology, China, 2012.

[2] LU Kun. “Short-term Wind Speed Forecasting and Confidence Interval Test of Wind Speed Prediction,” Thesis, Lanzhou University of Technology, 2011.

[3] Bengio Y. “Deep Learning of Representations: Looking Forward,” arXiv preprint arXiv: 1305.0445, 2013.

[4] Hinton G E, Osindero S, and Teh Y-W. “A Fast Learning Algorithm for Deep Belief Nets,” Neural Computation, vol. 18, 2006, pp. 1527–1554.

[5] WAN Jie, CHEN Ning, QIAN Minhui, GUO Yufeng, Hu Qinghua, and YU Daren. “Day-ahead Wind Speed Prediction based on Hybrid Deep Belief Network,” Energy Conservation Technology, 2016.

[6] Y. LeCun, K. Kavukcuoglu, and C. Farabet. “Convolutional Networks and Applications in vision,” In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium, pp. 253–256. IEEE, 2010.

[7] CHEN Xianchang. “Research on Algorithm and Application of Deep Learning Based on Convolutional Neural Network,” Thesis, Zhejiang Gongshang University, 2013.

[8] XU Shanshan. “Research and Application of the convolution neural network,” Thesis, Nanjing Forestry University, 2013.

[9] Elia, "Wind-Power Generation Data" [Online], Available: http://www.elia.be /en/grid-data/data-download

[10] Y. Lecun, B. Boster, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel. “Backpropagation Applied to Handwritten Zip Code Recognition, ”Neural Computation, 1(4): 541-551, Winter 1989.

[11] LI Feiteng. “Convolutional Neural Network and its applications,” Thesis, Dalian University of Technology, 2014.

[12] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. “Handwritten Digit Recognition with a Back-propagation Network,” In Advances in neural information processing systems, 1990.

[13] Y lecun and Y benquo. “Handbook of Brain Theory and Neural Networks,” MIT Press, 1995, pp. 255–258.

[14] K. Fukushima. A Neural Network Model for Selective Attention in Visual Pattern Recognition[J]. Biological Cybernetics, vol. 55(1), Oct. 1986, pp. 5–15.

[15] D. Cire¸san, U. Meier, and J. Schmidhuber. “Multi-column Deep Neural Networks for Image Classification,” arXiv preprint arXiv: 1202.2745, 2012.