

Wind Power Output Forecasting Based on AC-BiLSTM

Fuqiang Li^{1,*}, Yuexuan Zhao², Yongyuan Ji¹, Tianyuan Chen²

¹ School of Electrical & Electronic Engineering, North China Electric Power University, Baoding 071003, China

² School of Energy Power & Mechanical Engineering, North China Electric Power University, Baoding 071003, China

* Corresponding author: Fuqiang Li (Email: 15132621395@163.com)

Abstract: Wind power output prediction is crucial to the balanced dispatch and market operation of the power system, and accurate prediction of wind power output can guarantee the security and stability of power grid operation. In order to realize highly accurate wind power output prediction to support the low-carbon transformation of the power system, a wind power output prediction method integrating convolutional neural network (CNN), bi-directional long and short-term memory network (BiLSTM) and Attention mechanism is proposed. To address the nonlinear and temporal characteristics of wind power output data, the CNN layer extracts local nonlinear features, the BiLSTM layer captures bidirectional temporal dependencies, and the Attention mechanism reduces redundant interference and improves the model interpretability by dynamically weighting the key temporal information. Finally, the final output of the output power prediction result is output through the fully connected layer. The example analysis is carried out with the actual wind power data of China region in 2012, and the model generalization ability is optimized by using normalization processing and Dropout technique, and the parameters are trained by using the mean square error (MSE) as the loss function in combination with Adam's algorithm, which verifies the model's validity and robustness in complex fluctuation scenarios, and provides technical support for the planning and stable operation of the power system. The model can provide technical support for power system planning and stable operation.

Keywords: Wind power output forecasting; CNN; BiLSTM; Attention mechanism; Electric power system

1. Introduction

As an energy hub, the power system is crucial to achieving Chinese carbon peak carbon neutrality goal. In order to support the realization of this goal, China's wind power installed capacity is increasing year by year, and wind power prediction has also become a research hot spot for many scholars. Improving the accuracy of wind power prediction, optimizing the structure of power industry and promoting energy transformation have positive practical significance for implementing the action plan of carbon emission peaking before 2030.

At present, time series prediction is widely used in finance, meteorology, energy and other fields. Although traditional methods such as autoregressive and moving average models perform well in some cases, they are limited in dealing with complex nonlinear and long-term dependence problems. Deep learning models, especially CNN and LSTM networks, have shown effectiveness in time series prediction. In Reference [1], LSTM neural network is used in the field of wind power output prediction. Experiments show that compared with feed-forward neural network, LSTM model has higher prediction ability and applicability. Reference [2] used CNN neural network to predict wind power output, and proved the superiority of CNN in feature extraction by comparing with RNN neural network.

However, these single models are difficult to balance the time series data. Local feature extraction and long-term dependence capture, taking into account the volatility and intermittency of wind power generation, scholars have begun to integrate a single model. Reference [3] proved that the prediction ability of BiLSTM model in time series analysis is better than that of one-way LSTM model, which can be used for power load forecasting. Reference [4] proposed an ultra-short-term power prediction considering the health status of

wind turbines and the dual attention mechanism CNN-BiLSTM. The hybrid method of convolutional neural network and bidirectional long-term and short-term memory neural network is used for ultra-short-term power prediction. This model can adapt to complex data and enhance robustness. Compared with a single model, the prediction accuracy is improved.

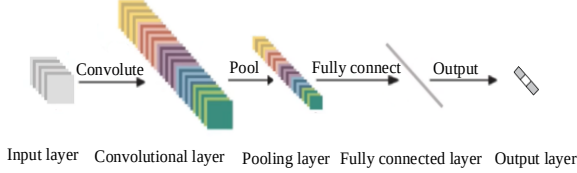
However, in the prediction of wind power output, local fluctuations may lead to over-fitting of the model, which needs to rely on strong regularization or data enhancement. At the same time, because CNN-BiLSTM is a black box model [5], it is difficult to intuitively explain the contribution of features (such as the causal relationship between meteorological factors and power fluctuations), which is not conducive to fault diagnosis in practical engineering. Therefore, attention mechanism is introduced to analyze the importance of features, so as to enhance the performance of the model, reduce the risk of overfitting, enhance the interpretability of the model and further improve the prediction accuracy.

Based on the advantages of CNN-BiLSTM and Attention mechanism, this paper proposes a new wind power output prediction method based on AC-BiLSTM for the nonlinear and time-series characteristics of wind power output [6]. Among them, the convolutional neural network CNN can effectively extract the nonlinear characteristics of wind power output data; the BiLSTM layer is used to extract the temporal features of sequence data. By inputting BiLSTM layer features into the Attention mechanism, the importance of time information is distinguished and the influence of redundant information is reduced. Finally, through experimental prediction, it is verified that the AC-BiLSTM model has good performance in wind power forecasting.

2. Model Structure

2.1. Convolution Neural Network

Convolutional neural network is a kind of feed forward neural network with convolution calculation and deep structure. It is one of the representative algorithms of deep learning. Because of its efficient feature extraction ability, it has been applied in the field of power system research. CNN was first proposed by Yann Lecun of New York University in 1998 for feature extraction of images [7]. In 2012, the emergence of AlexNet marked a major breakthrough in the field of image recognition by convolutional neural networks. Its structure often includes five parts which are input layer,



convolution layer, pooling layer, fully connected layer and output layer, as shown in Figure 1.

Figure 1. CNN structure diagram

Among them, the convolution layer uses the convolution kernel to extract the effective nonlinear local features of wind power output data. The pooling layer is used to compress the extracted features and generate more important feature information, which not only accelerates the operation speed, but also avoids the occurrence of over-fitting to a large extent and improves the generalization ability of the neural network.

2.2. LSTM Neural Network

LSTM is a special RNN neural network, which was proposed by Hochreiter & Schmidhuber in 1997 [8] and improved and promoted by Alex Graves [9]. LSTM introduces three gates, namely input gate, forgetting gate and output gate, and memory cells with the same shape as the hidden state to record additional information. LSTM has a “gate” structure to remove or increase information to the cell state, so it can solve the long-term dependence problem and the problem of RNN gradient disappearance and gradient explosion. In addition, LSTM can also make the network converge more efficiently and quickly by learning the long-term dependencies in time series data, which helps to improve the accuracy of wind power output prediction. The structural unit of LSTM is shown in Figure 2.

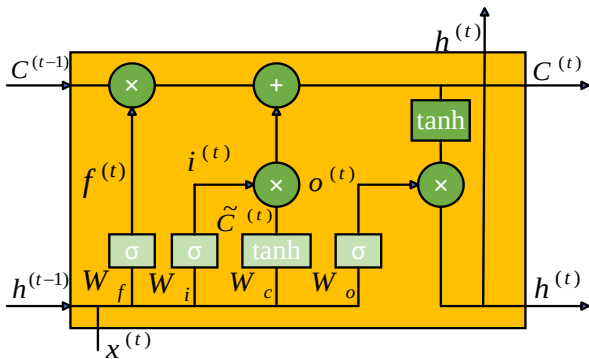


Figure 2. LSTM unit structure

(1) The first step is to decide what information to discard from the cellular state, which is accomplished by a structure called the forgetting gate. The forgetting gate determines what information is discarded from the cell state $C^{(t-1)}$ at $t-1$ time.

The forgetting gate reads the $t-1$ time hidden layer state $h^{(t-1)}$ and the time input sequence $x^{(t)}$, makes a Sigmoid nonlinear mapping and outputs the values between 0~1, and then outputs a vector $f^{(t)}$.

$$f^{(t)} = \sigma(W_f \bullet [h^{(t-1)}, x^{(t)}] + b_f) \quad (1)$$

Where $f^{(t)}$ is the state of forgetting the door at time t ; W_f and b_f are the weight and bias of the forgetting gate, respectively; σ is the bipolar Sigmoid activation function.

(2) The next step is to determine what new information is stored in the cell state. It contains two parts as followings.

First, the Sigmoid layer, also known as the “input gate layer”, reads the input $x^{(t)}$ at time t and determines the information stored in the neuron;

second, the tanh layer creates a new candidate value vector $\tilde{C}^{(t)}$ and adds it to the cell state $C^{(t)}$ to update the cell state.

The update process of the input gate is shown in the following formulas (2)-(4).

$$i^{(t)} = \sigma(W_i \bullet [h^{(t-1)}, x^{(t)}] + b_i) \quad (2)$$

$$\tilde{C}^{(t)} = \tanh(W_c \bullet [h^{(t-1)}, x^{(t)}] + b_c) \quad (3)$$

$$C^{(t)} = f^{(t)} \otimes C^{(t-1)} + i^{(t)} \otimes \tilde{C}^{(t)} \quad (4)$$

In the formula, $i^{(t)}$ is the state of the input gate at time t , and the amount of information transmitted from $x^{(t)}$ to $C^{(t)}$ is selected; W_i and b_i are the weight and bias of the input gate, respectively; W_c and b_c are the weight matrix and bias term of the cell state, respectively; $\tanh$ is a hyperbolic tangent activation function; $\otimes$ is Hadamard product.

(3) Finally, determine the output value. This output will be based on the cell state, but is also a selected version. First, run a Sigmoid layer to determine which parts of the cell state are exported. Then, the cell state is processed by the tanh function (to obtain a value between -1 and 1) and multiplied by the output of the Sigmoid gate to obtain the output value $h^{(t)}$, which is also the input value of the next hidden layer. The calculation formula of the output gate is shown in formula (5) and formula (6).

$$o^{(t)} = \sigma(W_o \bullet [h^{(t-1)}, x^{(t)}] + b_o) \quad (5)$$

$$h^{(t)} = o^{(t)} \otimes \tanh C^{(t)} \quad (6)$$

Here $o^{(t)}$ is the output gate state at time t ; W_o and b_o are the weight matrix and bias term of the output gate, respectively.

2.3. BiLSTM Neural Network

Bidirectional long-term and short-term memory network BiLSTM is an improved recurrent neural network RNN, which is specially designed to process sequence data [10]. BiLSTM can capture the bidirectional dependencies in the sequence by combining the output of the forward and backward LSTM networks. Compared with one-way LSTM, BiLSTM can take into account both the front and back information of the sequence, which is very beneficial to tasks that need to understand the global context (such as named entity recognition, machine translation, etc.). BiLSTM fully considers past and future information, which is conducive to improving model performance and further improving the accuracy of model prediction. The BiLSTM structure is shown in Figure 3. Among them, $x_1, x_2, x_3, \dots, x_t$ represents the

corresponding input data of $t_1 \sim t_i (i \in [1, t])$ at each moment, $A_1, A_2, A_3, \dots, A_t, B_1, B_2, B_3, \dots, B_t$ represent the corresponding forward and backward iterative LSTM hidden state respectively, $Y_1, Y_2, Y_3, \dots, Y_t$ represents the corresponding output data, and $\omega_1, \omega_2, \omega_3, \dots, \omega_6$ represents the corresponding weight of each layer.

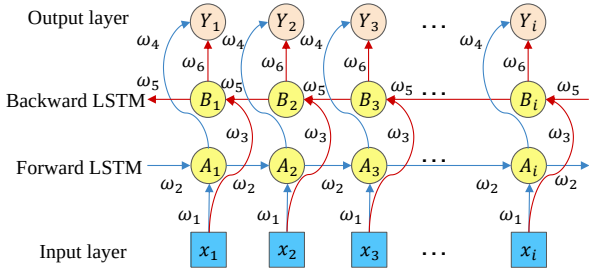


Figure 3. BiLSTM neural network structure

The hidden layer update status of forward LSTM, backward LSTM and the final output process of BiLSTM are shown in formula (7)-(9).

$$A_i = f_1(\omega_1 x_i + \omega_2 A_{i-1}) \quad (7)$$

$$B_j = f_2(\omega_3 x_i + \omega_5 B_{i+1}) \quad (8)$$

$$Y_i = f_3(\omega_4 A_i + \omega_6 B_i) \quad (9)$$

In the formula, f_1, f_2, f_3 is the activation function between different layers.

2.4. Attention Mechanism

Attention originates from the simulation of human brain attention features, which can help us filter out irrelevant or useless information and focus on important stimuli, tasks or goals. In 2014, the Google mind team published "Recurrent Models of Visual Attention." In this paper, the Attention mechanism is used for image classification on the RNN model for the first time [11]. The Attention mechanism assigns a larger weight to the key content and a smaller weight to other content through carefully designed weight distribution, reducing unnecessary calculations, thereby improving the computational efficiency of the model. The Attention mechanism also provides an intuitive way to explain the decision-making process of the model, and improves the generalization ability of the model on unseen data by more intelligently selecting features with strong correlation. The Attention mechanism unit structure is shown in Figure 4.

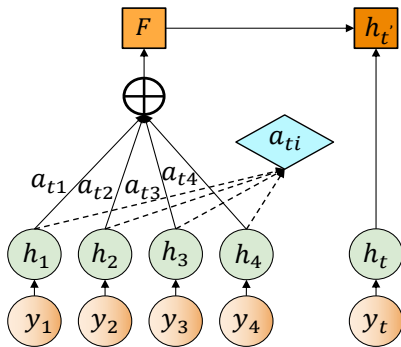


Figure 4. Attention unit structure

The attention state transition process is shown in formula (10)-(13).

$$S_{ti} = V \tanh(W h_t + U h_i + b), i = 1, 2, 3, \dots, t-1 \quad (10)$$

$$a_{ti} = \frac{\exp(S_{ti})}{\sum_{i=1}^t \exp(S_{tk})}, i = 1, 2, 3, \dots, t-1 \quad (11)$$

$$F = \sum_{i=1}^t a_{ti} \times h_i, i = 1, 2, 3, \dots, t-1 \quad (12)$$

$$h'_t = f(F, h_t, y_t) \quad (13)$$

Here a_{ti} is the attention weight value of the BiLSTM hidden layer output value h_i to the current input; $y_1, y_2, y_3, \dots, y_t$ is the input sequence; $h_1, h_2, h_3, \dots, h_t$ is the hidden layer state value corresponding to the input sequence, $y_1, y_2, y_3, \dots, y_t$ that is, h_t is the hidden layer state value corresponding to the input y_t ; h'_t is the final feature vector; V, W, U, b are the learning parameters of the model, and these parameters are constantly updated with the model training process.

2.5. AC-BiLSTM Prediction Model Structure

In this paper, a wind power output prediction method based on AC-BiLSTM model is proposed. The specific implementation process is shown in Figure 5.

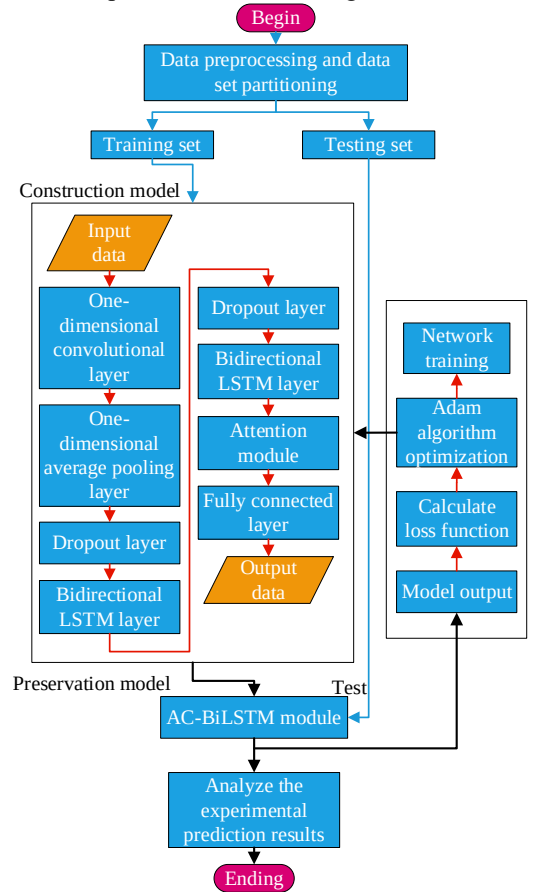


Figure 5. AC-BiLSTM flow chart

Firstly, the collected output data is preprocessed and divided into training set and test set.

Secondly, the AC-BiLSTM model is constructed. In this paper, the CNN framework composed of a one-dimensional convolution layer and a pooling layer is used to automatically extract the internal features of the output data. The convolution layer is used to extract the effective nonlinear local features of the wind power output data. The pooling layer selects the average pooling method to compress the extracted features and generate more critical feature information.

Thirdly, BiLSTM hidden layer modeling learns the internal

dynamic changes of local features extracted by CNN, and iterates to extract more complex global features from local features.

Fourthly, taking the features generated by the BiLSTM hidden layer as the input of the Attention mechanism, the attention mechanism is used to automatically distinguish the importance of the time information extracted by the BiLSTM hidden layer by weighting, which can more effectively use the time series attributes of the output data itself and mine the deep time correlation. The attention mechanism can effectively reduce the loss of historical information and highlight the information of key historical time points to weaken the influence of redundant information on the output prediction results.

Finally, the output of the Attention layer is used as the input of the fully connected layer, and the final output prediction result is output through the fully connected layer.

In addition, this paper introduces random deactivation (Dropout) technology to prevent over-fitting [12], that is, a Dropout layer is attached to each BiLSTM hidden layer. While preventing over-fitting and premature convergence, it can also improve the generalization of the model and reduce the training time of the model. In the part of network parameter optimization, Adam (Adaptive Moment Estimation) optimization algorithm is used to update the network parameters of each layer, and the loss function is Mean Squared Error (MSE).

After that, the trained AC-BiLSTM model is saved, and the test set is used to verify the effectiveness of the model. By analyzing the output prediction results, the deficiencies are determined, and the prediction model is continuously optimized and improved.

3. Data Processing

In order to verify the performance of the AC-BiLSTM model proposed in this paper, the real wind power output data (sampling period is 1h, unit MW) from January 1, 2012 to May 9, 2012 in a certain area of China are selected for wind power output prediction. The historical output data from January 1, 2012 to April 1, 2012 (a total of 2160 samples) are taken as the training set, and the output data from April 2, 2012 to May 9, 2012 (a total of 888 samples) are taken as the prediction data set.

3.1. Data Normalization

In order to improve the convergence speed of the loss function, prevent the gradient explosion and improve the calculation accuracy, the data are normalized. In the actual output prediction process, the input of the model usually contains a variety of different dimensional data. In order to eliminate the influence of different dimensions on the prediction results and improve the accuracy and efficiency of the model, it is necessary to normalize the data. In this paper, min-max standardization is used to normalize the data to [0, 1], and the expression is as shown in formula (14).

$$X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (14)$$

Where X_i is the original measured data of the i sampling point; X_i^* is the value after X_i normalization; $X_{\max}$ and $X_{\min}$ are the maximum and minimum values in the measured data, respectively. When the normalized prediction result is obtained by using the normalized data input model, the

formula (15) is used to denormalize it.

$$x^* = y^* (X_{\max} - X_{\min}) + X_{\min} \quad (15)$$

Here y^* is the normalized output prediction value; x^* is the predicted value of actual wind power output obtained after inverse normalization.

3.2. Model Evaluation

In this paper, mean absolute percentage error (MAPE), root mean square error (RMSE) and determination coefficient (R^2) are used to evaluate the performance of wind power output prediction methods. The smaller the value of MAPE and RMSE in wind power output prediction, the closer the value of R^2 is to 1, the more accurate the output prediction is. The calculation expressions are shown in formulas (16)-(18).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y(i) - y'(i)}{y(i)} \right| \times 100\% \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y(i) - y'(i))^2} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y(i) - y'(i))^2}{\sum_{i=1}^n (y(i) - \bar{y})^2} \quad (18)$$

In the formula, $y'(i)$ and $y(i)$ are the predicted output value and the real output value at i time respectively; n represents the total number of test samples; $\bar{y}$ represents the average output value of the total sample, and its formula is as follows.

$$\bar{y} = \frac{\sum_{i=1}^n y(i)}{n} \quad (19)$$

4. Example Analysis

In this paper, the AC-BiLSTM model is established by using the method of artificial search hyperparameters and the prediction comparison is carried out on the test set, which proves the advantages of the method proposed in this paper in wind power output prediction. At the same time, the reliability of the model is proved.

4.1. An example of Wind Power Output Prediction in Training Set

In order to verify the learning performance of the proposed model, this paper uses the artificial search method to select the hyperparameters for testing. The model runs 2160 times continuously, and the wind power output from January 1, 2012 to April 1, 2012 is predicted, and the MAPE, RMSE and R^2 of each method are calculated.

Figure 6 is the output comparison line chart of the model prediction training set in the example. It can be seen from the figure that the proposed AC-BiLSTM model fitting curve is similar to the actual output curve fluctuation. The lines of the two curves are basically coincident. When the actual output changes greatly (i.e., the volatility of wind power output is strong), the AC-BiLSTM model can still better fit the change

trend of the actual output, especially at the valley peak of the curve.

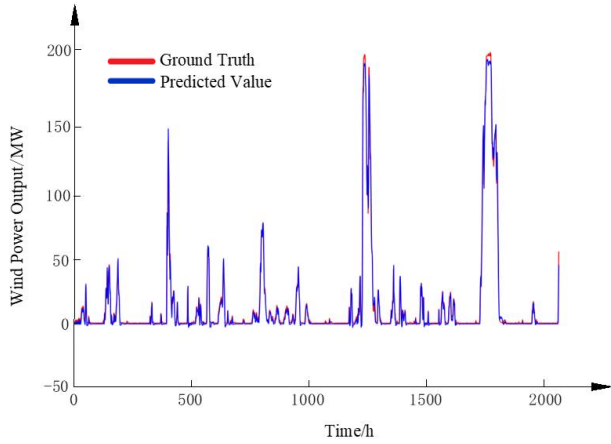


Figure 6. Comparison of training set prediction results

4.2. Test Set Wind Power Output Prediction Example

The test set model runs 888 times continuously, and the wind power output from April 2 to May 9, 2012 is predicted, and the MAPE, RMSE and R^2 of each method are calculated respectively by it.

Figure 7 is the output comparison line chart of the model prediction training set in the example. From the diagram, it can be seen that the fluctuation of the model fitting curve and the actual output curve is still relatively close, but the prediction coincidence at several maximum values has decreased. This is due to the lack of key external features, such as real-time meteorological data (gust acceleration, turbulence intensity), fan state (pitch angle, fault log) or terrain effect, which makes the model unable to correlate peak causes.

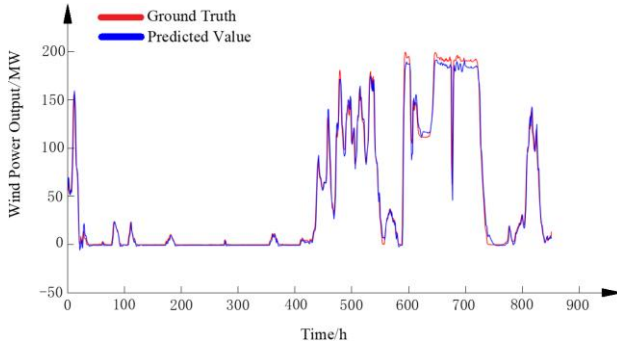


Figure 7. Comparison of test set prediction results

4.3. Error Analysis

In the example, the prediction error of the model based on AC-BiLSTM is shown in Table 1.

Table 1. Model predictive error

Data set	MAPE	RMSE	R^2
Training set	2.97	7.33	0.992
Testing set	3.58	8.67	0.984

It can be seen from Table 1 that in the output results of the prediction training set, the MAPE index of the AC-BiLSTM model is 2.97, the RMSE is 7.33, and the coefficient of determination (R^2) reaches 0.992. The model shows high accuracy on 2098 of 2160 samples, which shows that the

model is highly consistent with the actual data in the training stage. However, on the test set, the results of Table 1 show that the performance decreases slightly, MAPE rises to 3.58, RMSE rises to 8.67, and the coefficient of determination (R^2) decreases to 0.984, indicating that although the model still maintains a good predictive ability, the accuracy is reduced in the face of unseen data. This performance change may be caused by overfitting or potential distribution differences between training data and test data. Although there are the above shortcomings, the residual prediction residual (RPD) on the test set is 7.852, indicating that the model prediction maintains a relatively stable consistency with the actual value, confirming the effectiveness of the AC-BiLSTM model in wind power generation prediction.

5. Conclusion

In this paper, a wind power output prediction method based on AC-BiLSTM is proposed. By combining CNN, BiLSTM and Attention mechanism, the nonlinear characteristics and time series characteristics of wind power data are effectively extracted, and the importance of time information is distinguished to reduce the influence of redundant information. The followings are the main conclusions being summarized.

(1) The AC-BiLSTM model combines the feature extraction ability of CNN, the bidirectional dependence capture ability of BiLSTM and the weight distribution advantage of Attention mechanism, which can adapt to the nonlinear and temporal characteristics of wind power output, improve the prediction accuracy and enhance the interpretability of the model.

(2) The experimental results show that the MAPE of the model on the training set is 2.97, RMSE is 7.33, and R^2 is 0.992, which shows that the model is highly consistent with the actual data in the training stage. On the test set, MAPE increased to 3.58, RMSE increased to 8.67, and R^2 decreased to 0.984, indicating that the model still maintained good predictive ability on unseen data, but the accuracy was reduced slightly.

(3) The performance degradation of the model on the test set may be attributed to the over-fitting phenomenon in the training phase or the domain distribution difference between the training set and the test set.

In the future, more external features (such as meteorological data and fan state information) can be introduced, the model structure and parameters can be optimized, and more effective regularization methods can be explored to further improve the prediction accuracy and generalization ability.

References

- [1] Banik A, Behera C, Sarathkumar T V, et al. Uncertain wind power forecasting using LSTM-based prediction interval[J]. IET Renewable Power Generation, 2020, 14(14): 2657-2667.
- [2] Hong Y Y, Rioflorido C L P P. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting[J]. Applied Energy, 2019, 250: 530-539.
- [3] Siarni-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series[C]//2019 IEEE International conference on big data (Big Data). IEEE, 2019: 3285-3292.

- [4] Zou M, Holjevac N, Đaković J, et al. Bayesian CNN-BiLSTM and vine-GMCM based probabilistic forecasting of hour-ahead wind farm power outputs[J]. IEEE Transactions on Sustainable Energy, 2022, 13(2): 1169-1187.
- [5] EZhao C, Huang X, Li Y, et al. A double-channel hybrid deep neural network based on CNN and BiLSTM for remaining useful life prediction[J]. Sensors, 2020, 20(24): 7109.
- [6] Liu F, Liang C. Short-term power load forecasting based on AC-BiLSTM model[J]. Energy Reports, 2024, 11: 1570-1579.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [8] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [9] Graves A, Graves A. Long short-term memory[J]. Supervised sequence labelling with recurrent neural networks, 2012: 37-45.
- [10] Xu G, Meng Y, Qiu X, et al. Sentiment analysis of comment texts based on BiLSTM[J]. Ieee Access, 2019, 7: 51522-51532.
- [11] Knudsen E I. Fundamental components of attention[J]. Annu. Rev. Neurosci., 2007, 30(1): 57-78.
- [12] Baldi P, Sadowski P J. Understanding dropout[J]. Advances in neural information processing systems, 2013, 26.