

Multi-time scale wind speed prediction based on WT-bi-LSTM

Jinyong Xiang, Zhifeng Qiu*, Qihan Hao and Huhui Cao

Automation School, Central South University, Changsha 410083, Hunan Province, China

Keywords: Wind speed prediction, wavelet decomposition, bi-LSTM, deep learning.

Abstract. The accurate and reliable wind speed prediction can benefit the wind power forecasting and its consumption. As a continuous signal with the high autocorrelation, wind speed is closely related to the past and future moments. Therefore, to fully use the information of two direction, an auto-regression model based on the bi-directional long short term memory neural network model with wavelet decomposition (WT-bi-LSTM) is built to predict the wind speed at multi-time scales. The proposed model are validated by using the actual wind speed series from a wind farm in China. The validation results demonstrated that, compared with other four traditional models, the proposed strategy can effectively improve the accuracy of wind speed prediction.

1 Introduction

In the field of renewable energy power generation, wind power resource is the most popular one due to their advantages of cleanness, wide distribution and easy to be utilized on large scales, and has attracted attention of all countries[1]. Wind speed is one of the main factors affecting wind power generation capability. However, with the characteristics of intermittent and randomness, it always causes some unstable phenomena in wind power system, such as oscillation, fluctuation, slope climbing, which result in certain limits and difficulties to integrate the clear power into the grid. Accurate wind speed prediction is one of the key technologies to ensure the safe operation of wind farm and grid system [2].

Traditional wind speed prediction methods can be divided into physical model-based methods and statistical model-based methods[3]. These methods have been studied by scholars in various countries, and have developed rapidly in prediction precision and accuracy. However, there are still many problems to be solved in order to make the speed forecast more accurate and at larger time scales. With the rapidly development of computer technology and artificial intelligence, modern predicting methods are developing in the direction of intelligence, integration and depth, which makes it possible to obtain better forecasting results compared with traditional methods[4,5].

* Corresponding author : zhifeng.qiu@csu.edu.cn

Wind speed is a highly autocorrelated time-varying signal whose past information has a great effect on the forthcoming velocity[6]. Therefore, not only will the current inputs influence outputs, but the continuity and dependence between input samples need to be calculated when a speed forecasting model is established. For example, Lei Jinhao[7] et al. ever pointed out that long-term dependence was particularly important for generating separable and predictable features. Therefore, extracting the feature between speed signal using artificial intelligence technology is becoming a research hotspot. For example, long short term memory (LSTM) neural network model can make full use of the information of continuous samples, especially the information with long-term dependence. It is often used in time series problems such as speech recognition, machine translation, handwriting recognition and so on. It also achieves good results in wind speed and power prediction[8-11].

Jie Chen[8] and others thought that when it comes to using the integration method to build wind speed series models, it is insufficient to only use the linear combination method. Instead, it should take the structure of non-linear combination into consideration. Therefore, an integrated LSTM model using multiple LSTMs as sub-models and Extremal Optimization Algorithms (EO) as the top combiner to optimize the parameters of Support Vector Regression Machine (SVRM) is proposed. The numerical results demonstrate the superiority of the model. Yanfei Li et al[10]. established the main model of LSTM neural network based on the wind speed sub-sequence decomposed by Empirical Wavelet Transform (EWT); used Regularized Extreme Learning Machine (RELM) as the sub-model to model the prediction error of LSTM; and finally used Inverse Empirical Wavelet Transform (IWET) to obtain the predicted values.

The research results above confirm that LSTM model can make full use of the historical information of the sequence, and this model has been more and more widely used. At the same time, some scholars put forward the rule of learning signals from the opposite direction[12, 13], making important progress in speech recognition, Chinese event extraction and some other aspects. However, in the wind speed prediction, it has never been considered and studied from this point of view. In addition to considering the past information, whether the reversal rule of the future to the present can be used in wind speed prediction is the research focus of this paper.

In order to obtain higher accuracy of wind speed prediction, a neural network model based on bidirectional long-short memory (bi-LSTM) is proposed to extract historical and future information of wind speed after the signal is transformed by wavelet. Firstly, through the Wavelet Transform (WT), the wind speed signal is decomposed into low-frequency and high-frequency components. Therefore, the different frequencies of the signal are separated to reduce the interaction. Secondly, the bi-LSTM model is established for each frequency signal, obtaining the hidden state in two directions. By making full use of the double side information, the signal components are predicted in multi-steps. Finally, the predicted components are added together to obtain the multi-time scale prediction results. After analyzing the wind speed series in the actual wind farm, five models are established. Among them, the proposed model in this paper which considers bidirectional information has the highest prediction precision and accuracy, especially in extracting trend information of wind speed in large time scale.

The rest of this paper is organized as follows. Section 2 introduces the methods used in this paper, which is mainly about WT and bi-LSTM. Section 3 compares the precision and accuracy of different models and discusses some possible reasons for the results. Finally, Section 4 presents the conclusions and prospects of this paper.

2 WT and bi-LSTM

2.1 Wavelet Transform

Wavelet transform can analyze signals from the time domain and the frequency domain at the same time. Through stretch and translation operations, WT performs a multi-scale refinement analysis of the signals, which plays an important role in decomposing local signals[14]. For any arbitrary continuous signal $x(t)$, assume that its energy is limited, that is, $\int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty$, make $x(t) \in L^2(R)$, where $L^2(R)$ is the signal space, then the continuous WT of $x(t)$ is:

$$W_f(a, b) = \int_{-\infty}^{+\infty} x(t) \bar{\Psi}_{(a,b)}(t) dt \quad (1)$$

where, $\bar{\Psi}_{(a,b)}(t)$ is the complex conjugate of the base wavelet function $\Psi_{(a,b)}(t) = \frac{1}{\sqrt{a}} \Psi(\frac{t-b}{a})$, which is represented as Mother wavelet, with a representing scale parameter and b the translation parameter. Different mother wavelets $\Psi(t)$ acting on the signal $x(t)$, leading to different decomposition results. However, there is no specific method to guide the selection of the mother wavelet, which is often determined by comparing the advantages and disadvantages of the transformed signal modeling effect.

However, the sampling of the actual wind speed signal has a certain time interval, which belongs to a discrete signal in the time domain. When decomposing such signals, a discrete wavelet transform is required. Let the scale parameter of the continuous wavelet transform $a = a_0^j$, the translation parameter $b = ka_0^j b_0$, ($j, k \in Z$), so the parameter symbol changes from (a, b) to (j, k) , and the discrete Wavelet Transform (DWT) is obtained.

$$W_f(j, k) = \int_{-\infty}^{+\infty} x(t) \bar{\Psi}_{(j,k)}(t) dt \quad (2)$$

Mallat put forward the concept of multi-resolution analysis when constructing orthogonal wavelet bases in 1989[15], explaining the characteristics of wavelet spatially, proposing the Mallat wavelet packet algorithm for signal decomposition on this basis. The process of decomposition is shown in Figure 1.

The signal $x(t)$ is orthogonally projected into the spaces V_j and W_j to obtain a low-frequency discrete approximation signal $c_j(t)$ and a high-frequency discrete detail signal $d_j(t)$ at a resolution j . Ignoring the high-frequency $d_j(t)$, each low-frequency $c_j(t)$ signal can be decomposed further into another two parts of detail signal $d_{j+1}(t)$ and approximation signal $c_{j+1}(t)$ step by step with j increasing until the end condition. After decomposing, the signal can be expressed as equation (3):

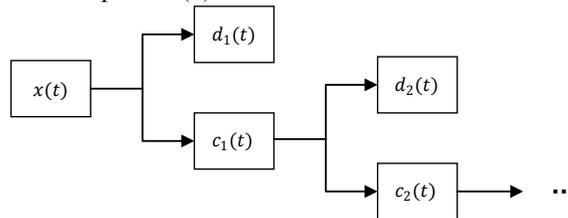


Fig. 1. Mallat signal multiresolution analysis.

$$x(t) = \sum_{j=1}^n d_j(t) + c_n(t) \quad (3)$$

The original signal decomposed by n times wavelet transform can be cumulatively reconstructed by an n -th stage approximation signal and detail signal of each stage, that is, in total, $n+1$ sub-signal sequences. For example, after wavelet transform, the wind speed signal can be decomposed into sub-sequence signals of different frequencies, effectively separating the high-frequency part of the original signal caused by atmospheric circulation, randomness, and volatility, solving the problem of irregular fluctuation of the sequence. It has been pointed out in the literature that the tightly supported bi-orthogonal Daubechies4 wavelet which was used as the mother wavelet to perform the decomposition of 2-4 layers could make a good analysis of the wind speed signal[16].

2.2 Bidirectional long short term memory recurrent neural networks(bi-lstm)

2.2.1 Long Short Term Memory (LSTM)

In the process of reverse optimization, the gradient information transported in Recurrent Neural Network (RNN) may disappear or explode as the number of layers increases, resulting in a weak reverse optimization and a low learning efficiency[17]. That is the reason why RNN is not suitable for dealing with long time series problem.

As a deep learning algorithm, the LSTM network is a special type of improved RNN. The LSTM cell unit undergoes a subtle combination of forgetting gates, input gates, and output gates, and introduces cell-state connections in the RNN network to resolve the problems of gradient disappearance or explosion during deep propagation. It is often used to deal with long-term dependent time series[18]. The structure of a basic LSTM cell unit is shown in Figure 2.

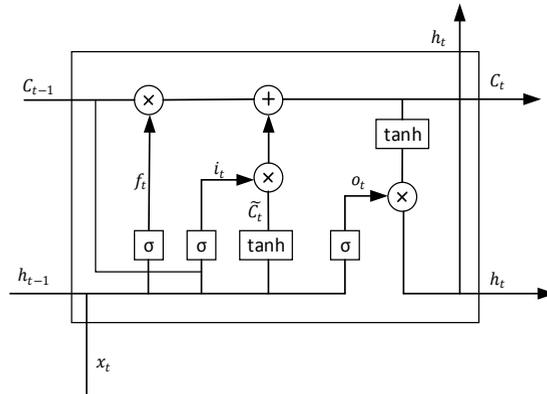


Fig. 2. Typical cell of LSTM.

Compared with simple RNN, LSTM no longer simply uses the cell state at the previous moment as the input of the current cell. Instead, it involves flexible choices after the training and learning through the forgetting gate f_t , the input gate i_t , and the output gate o_t . When processing information within the LSTM cell unit, the operation usually involves the following steps.

From the formula above, we can see that each gate in the cell receives the hidden state h_{t-1} at the previous moment, the current input x_t , and the state information C_{t-1} of the internal memory unit of the cell to perform different operations, and determine whether to

activate by a logic function. The forgetting gate outputs a number between 0 and 1 through a non-linear transformation, determining the influence that the information of the last memory cell C_{t-1} has on the current memory state. State information C_t that will be updated consists of two parts. One is determined by the forgetting gate; the other is determined by the input gate and inputs at time t . That is, a part of the input gate is used to find the state that needs to be updated, and the other part is used to make the updated information go to the cell state. Finally, the state h_t of the unit, the output at time t and the input hidden state at time $t + 1$ are determined by non-linearly activating $\tanh(\cdot)$ and the information of the output gate.

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}C_{t-1} + b_f) \tag{4}$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}C_{t-1} + b_i) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{6}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}C_t + b_o) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

Among them, $\sigma(\cdot)$ is the sigmoid activation function, $\tanh(\cdot)$ is the tanh activation function, W and b are the connection weight matrix and the bias vector between the forgetting gate, the input gate, the output gate and different input quantities, which are usually unknown parameters to be learned in the deep learning network.

2.2.2 bi-LSTM

From the previous analysis, we can infer that the LSTM network can completely transmit and utilize forward information, but this way of unidirectional propagation makes the future information in the sequence, that is, the reverse information in the signal, cannot be fully used. The Bi-LSTM network joins a backward propagation LSTM network, which not only deals with the past information in forward propagation, but also considers the future sequence information in reverse recursive process, making it possible to learn the forward and backward rules of sequence at the same time [12].

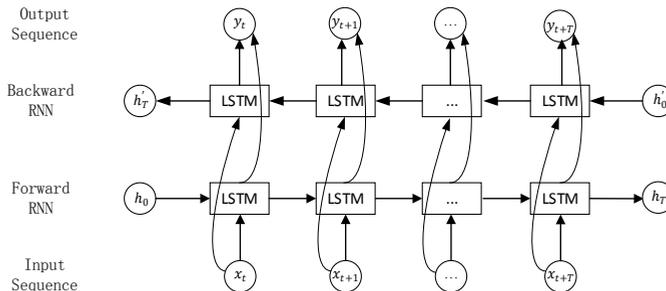


Fig. 3. Bi-LSTM neural network structure deployed in time direction.

As shown in Fig. 3, the bidirectional LSTM network includes two recurrent neural networks in positive and negative directions: the forward propagation process extracts long-term dependency information from historical signal; and the reverse process extracts future long-term dependency information in the signal. The output of the same neuron connects two LSTM meta cells, resulting in two hidden states h_t , h'_t in the forward and reverse directions.

The state of the output layer at time t is $H_t = [h_t, h'_t]$ [19]. In the model proposed in this paper, the hidden state H_t of the bi-LSTM network is passed to a fully connected layer to complete the non-linear mapping of the intermediate state; and then a regression layer is used to realize the regression prediction of the wind speed signal.

3. Case study

The data analyzed in the case study comes from a domestic wind farm, including three measured wind speed time series. The length is 1050; the sampling interval is 15min; the first 80% of the data is used as the training set, and the remaining 20% is used as the test set. The basic information of the sequence is shown in Figure 4 and Table 1. In order to construct the training and test samples of the multi-scale wind speed prediction model, we need to scroll the sequence signal back and forth, which causes a part of moved sequence to be inconsistent with the label, thus leading to some meaningless values which need to be deleted directly. A total of 18 meaningless data were deleted in experimental sequence. The error index in case study uses three general error evaluation criteria[10]: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

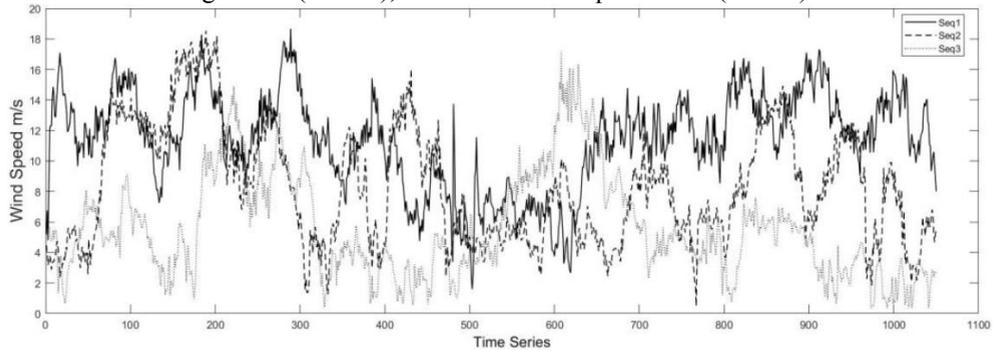


Fig. 4. 3 wind speed time series from the same wind farm.

Table 1. Basic information of three wind speed time series.

No.	Max	Min	Range	Average	variance
1	18.670	1.597	17.073	11.385	3.147
2	18.532	0.487	18.044	8.283	3.847
3	17.215	0.380	16.835	5.566	3.365

3.1 Experimental setting and steps

In order to verify the precision and accuracy of the proposed method, five prediction models, namely Persistence Model (PM), SVM, bi-LSTM, WT-SVM and the proposed WT-bi-LSTM model, are established on three wind speed sequences. Among them, PM is the simplest model and is often used as a benchmark for evaluation of other higher-complexity models.

Although the wind speed is affected by many factors, this paper only considers the influence of the wind speed signal itself as input on the prediction. The wind speed values at time $t, t-1, t-2$ and $t-3$ are used as the input to model wind speed at time $t+1, t+2, \dots, t+7$. This paper mainly describes sequence 1 and the main steps based on the WT-bi-LSTM model are as follows.

According to the above analysis, the three layers wavelet decomposition of the wind speed time series is performed by the tightly supported bi-orthogonal wavelet db4, obtaining three high-frequency detail signal components and one low-frequency approximation signal component. The decomposition results of sequence 1 is presented in Fig.5.

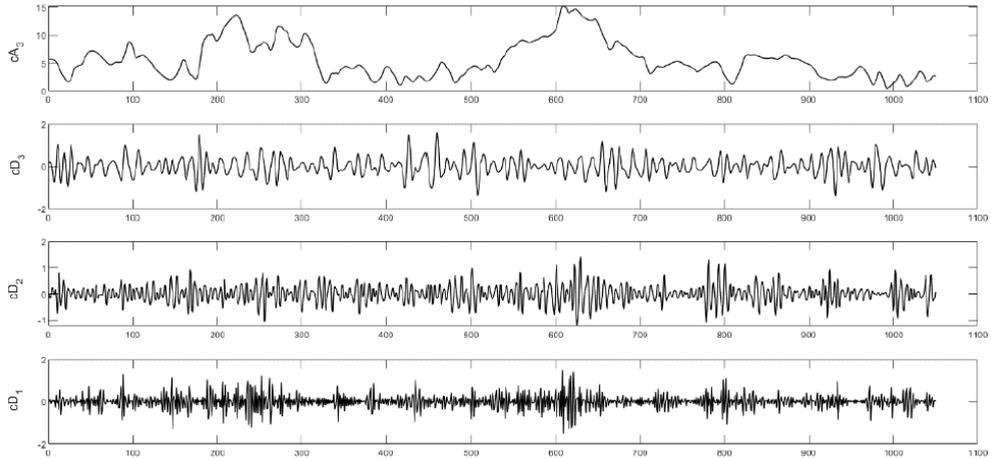


Fig. 5. Decomposition result of sequence 1.

The four signal components after the decomposition are scrolled back and forth to form an input-output sample matrix, respectively. After dividing the training set and the test set, the former is used to train the bi-LSTM network to obtain a prediction model, with which the test set is evaluated. In this paper, a deep learning network with a depth of 4 is established. The number of neurons is 4, 65, 27, and 1 respectively in the order of input layer, bi-LSTM layer, fully connected layer, and regression layer, with main hyperparameters illustrated as follows: the training cycles are 550; Adam optimization algorithm is employed to optimize the learning parameters; the initial learning rate is 0.01, decreasing with a discount factor of 0.5 in a frequency of 60 cycles.

The results of last step are superimposed with each other according to equation (3) as an inverse transform of WT to obtain the prediction result of the test set.

It can be seen from Fig. 5 that after the three-layer decomposition of the wind speed signal, the low-frequency part becomes smoother and the trend of the wind speed is highlighted explicitly; the high-frequency part is composed of three parts, vibrating up and down around the value of 0 at different frequencies. Each high-frequency component is a stationary signal, indicating that the high-frequency part can be regarded as white noise, mainly caused by factors such as topography and underlying surface other than wind speed signal.

3.2 Result and discussion

According to the experimental steps, seven models are established for 7 different time steps ahead of time t , respectively. The prediction results of one step to seven step are shown in table 2—table 4, from which we can draw the following conclusions. One step and seven step prediction results of the first 100 values in test set are presented in Fig 6.

As the time scale increases, the numerical performance of the prediction results deteriorates further. Taking the benchmark PM method as an example, the MAE value is 0.765 m/s in one step prediction, and the values of 3 steps, 5 steps, and 7 steps in advance are 1.042 m/s, 1.207 m/s, and 1.444 m/s respectively, decreasing 36.209%, 57.778%, 88.758%, respectively. MAPE and RMSE values are the same as the MAE; both deteriorate further as the time scale increases. The reasons for this phenomenon are various, mainly because of the

wind speed signal itself: as the time span increases, the abrupt change of meteorological conditions causes increasingly instable factors in the original signal and increasing uncontrollability, thus leading to worse prediction effects.

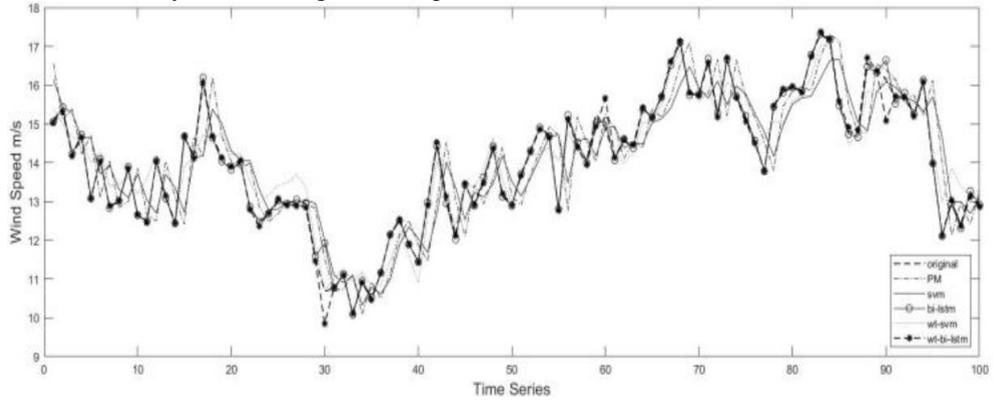


Fig. 6a. One step prediction result of sequence 1.

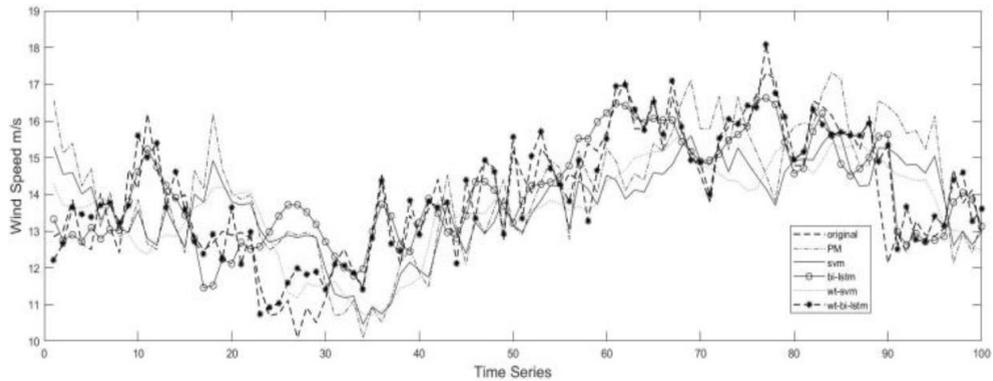


Fig. 6b. Seven step prediction result of sequence 1.

Compared with the benchmark method PM, the prediction results of other more complex models on various time scales are improved, but the improvement effects of different models are inconsistent. Taking the one-step prediction as an example. Compared with PM, the increase rates of SVM, Bi-LSTM, WT-SVM, and WT-Bi-LSTM in MAE, MAPE, and RMSE values are 4.134%, 92.497%, 51.503%, and 90.392%; 3.505%, 92.460%, 50.120%, 90.527%; 3.842%, 81.163%, 50.480%, 83.223%, respectively. Mainly because the PM model only takes the current wind speed at time t as the prediction result of time $t + N$ without any linear and non-linear calculation or mapping. Therefore, establishing a linear or non-linear prediction model for the wind speed signal can effectively improve the prediction accuracy.

Table 2. The MAE result of multi-scale wind speed prediction.

Model	MAE(m/s)						
	1	2	3	4	5	6	7
PM	0.765	0.947	1.042	1.139	1.207	1.294	1.444
SVM	0.732	0.891	0.968	1.052	1.134	1.220	1.324
Bi-LSTM	0.057	0.132	0.271	0.446	0.588	0.727	0.805
WT-SVM	0.371	0.535	0.639	0.791	0.903	1.016	1.128
WT-Bi-LSTM	0.074	0.128	0.181	0.299	0.398	0.607	0.656

Table 3. The MPE result of multi-scale wind speed prediction.

Model	MAPE(%)						
	1	2	3	4	5	6	7
PM	5.820	7.246	7.982	8.743	9.285	10.020	11.152
SVM	5.616	6.820	7.416	8.065	8.689	9.359	10.183
Bi-LSTM	0.439	1.012	2.077	3.491	4.568	5.596	6.190
WT-SVM	2.903	4.148	4.963	6.107	6.957	7.827	8.672
WT-bi-LSTM	0.551	0.970	1.360	2.277	3.044	4.638	5.024

Table 4. The RMSE result of multi-scale wind speed prediction.

Model	RMSE(m/s)						
	1	2	3	4	5	6	7
PM	0.937	1.182	1.330	1.429	1.519	1.611	1.742
SVM	0.901	1.107	1.239	1.312	1.403	1.493	1.578
Bi-LSTM	0.177	0.331	0.600	0.865	0.903	1.063	1.096
WT-SVM	0.464	0.670	0.801	0.965	1.084	1.207	1.341
WT-bi-LSTM	0.157	0.285	0.341	0.500	0.619	0.843	0.901

Compared with the SVM model, the prediction accuracy of the bi-LSTM model is greatly improved. As can be seen from the table, the improvement rates of the 1st, 3rd, 5th, and 7th steps in MAE, MAPE, and RMSE are: 92.158%, 72.045%, 48.166%, 39.237%; 82.187%, 71.990% 47.426%, 39.208%; 80.41%, 51.574%, 35.638%, 30.545%, respectively. The SVM model only considers the mapping relationship between input and output data during the mapping process of the model, ignoring the interaction between samples. However, while the bi-LSTM model considers the mapping relationship between input and output, the rule that includes positive and negative two directions between samples is also explored in depth. Therefore, modeling the time-continuous signal of wind speed and considering the bidirectional rule of history and future can significantly improve the prediction accuracy.

Compared with the SVM model, the prediction accuracy of the wind speed multi-time scale of the WT-SVM model is significantly improved. As can be seen from the table, the improvement results of the 1st, 3rd, 5th, and 7th steps in the MAE, MAPE, and RMSE values are: 49.317%, 99.988%, 20.370%, 14.804%; 48.308%, 33.077%, 19.933%, 14.838%; 48.502%, 35.351%, 22.737%, 15.019%, respectively. Mainly because WT decomposes the signal, the low frequency part of the trend signal and the high frequency part of the noise signal are stripped. For the model, the WT process obtains more feature attributes, which is more conducive to grasping the main part of the wind speed signal, thereby improving the prediction accuracy.

Compared with the bi-LSTM model, the WT-bi-LSTM model has a better numerical result in multi-step prediction of wind speed. As can be seen from the table, the prediction results of the 1st, 3rd, 5th, and 7th steps in the MAE, MAPE, and RMSE values are: -28.049%, 33.112%, 32.273%, 18.421%; -25.638%, 34.532%, 33.364%, 18.844%; 10.934%, 43.246%, 31.440%, 17.801%, respectively. In the one-step prediction results, the predictive numerical indicators MAE and MAPE of the WT-bi-LSTM model are not as good as bi-LSTM, but their absolute values are low. Probably the bi-LSTM model does not require a WT process in one-step ultra-short-term prediction, because the information required for high-precision prediction has been obtained. For longer-term predictions, the bi-LSTM model also needs to use the WT process to obtain longer trend information which can be clearly seen from the results of multi-scale predictions.

In order to verify the extension of the proposed WT-bi-LSTM model, the same modeling experiment as Sequence 1 is performed on the wind speed time series 2 and 3. The model structure and hyperparameters are not changed, the prediction results are shown in Table 7.

It can be seen from the table that for different wind speed sequences, the multi-time scale prediction numerical indicators of the proposed model perform best, and the optimal prediction results are obtained.

Table 5. The multi-scale prediction result of wind speed sequence 2, 3.

Model	Step	MAE(m/s)		MAPE(%)		RMSE(m/s)	
		Seq2	Seq3	Seq2	Seq3	Seq2	Seq3
PM	1	0.698	0.457	10.662	26.549	0.958	0.643
	3	1.030	0.782	17.116	42.695	1.464	1.025
	5	1.318	0.998	21.789	58.628	1.847	1.269
	7	1.606	1.076	27.595	70.078	2.139	1.367
SVM	1	0.686	0.474	11.003	32.668	0.938	0.642
	3	1.030	0.777	17.997	53.908	1.428	0.984
	5	1.311	0.956	23.200	69.724	1.757	1.209
	7	1.578	1.012	27.997	81.336	2.033	1.316
bi-LSTM	1	0.061	0.041	0.856	2.313	0.158	0.142
	3	0.194	0.179	3.047	9.154	0.371	0.341
	5	0.470	0.365	7.437	20.712	0.709	0.646
	7	0.746	0.637	12.124	40.237	1.036	0.986
WT-SVM	1	0.379	0.341	7.017	26.512	0.515	0.460
	3	0.589	0.523	9.928	38.435	0.779	0.695
	5	0.799	0.701	12.955	53.081	1.062	0.939
	7	1.125	0.911	18.647	73.361	1.491	1.196
WT-bi-LSTM	1	0.088	0.057	1.381	2.930	0.195	0.092
	3	0.164	0.125	2.591	6.186	0.296	0.234
	5	0.317	0.274	5.121	14.787	0.518	0.486
	7	0.622	0.483	10.209	32.856	0.912	0.820

In summary, the model WT-bi-LSTM has the best prediction effect and the highest prediction accuracy and precision. In the ultra-short-term one-step prediction, although the WT-bi-LSTM model error index is not as good as the performance of bi-LSTM, but the numerical result has been improved greatly compared to other models. Therefore, the WT-bi-LSTM model is more capable of extracting wind speed signals over long time span. It is the next step in this paper to continue to study and improve the accuracy of ultra-short-term one-step prediction on the proposed wind speed sequence.

4 Conclusion

In this paper, a neural network model of bidirectional long short term memory using wavelet decomposition is proposed to model the wind speed time series and perform multi-step

prediction. From the analysis above, we can conclude that the proposed model can make full use of history and future information of sequences. Modeling and analysis results on the measured wind speed sequence show that the proposed method is far superior to the benchmark PM method and other commonly used model. Compared with merely Bi-LSTM networks model, signal decompose strategy based on wavelet transform can make bi-LSTM have a better forecasting result on multi-scale wind speed signal prediction, especially in the long and large time scale.

References

1. Hui L, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM [J]. *Energy Conversion & Management*, 2018, 159(54-64).
2. Jiang Zhao-Yu, Jia Qing-Shan et.al. A Review of Multi-temporal-and-spatial-scale Wind Power Forecasting Method [J]. *Acta Automatica Sinica*, 2019, 45(1): 51-71.
3. Khoddayar M, Wang J. Spatio-temporal Graph Deep Neural Network for Short-term Wind Speed Forecasting [J]. *IEEE Transactions on Sustainable Energy*, 2018, PP(99): 1-.
4. Yu C, Li Y, Zhang M. Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting [J]. *Energy Conversion and Management*, 2017, 147(75-85).
5. L.Ngkvist M, Karlsson L, Loutei A. A review of unsupervised feature learning and deep learning for time-series modeling [J]. *Pattern Recognition Letters*, 2014, 42(11-24).
6. Da L, Niu D, Hui W, et al. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm [J]. *Renewable Energy*, 2014, 62(3): 592-7.
7. Lei J, Liu C, Jiang D. Fault diagnosis of wind turbine based on Long Short-term memory networks [J]. *Renewable Energy*, 2019, 133(422-32).
8. Chen J, Zeng G-Q, Zhou W, et al. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization [J]. *Energy Conversion and Management*, 2018, 165(681-95).
9. Hu Y-L, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm [J]. *Energy Conversion and Management*, 2018, 173(123-42).
10. Li Y, Wu H, Liu H. Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction [J]. *Energy Conversion and Management*, 2018, 167(203-19).
11. Zhu Q-M, Li H-Y, Wang Z-Q, et al. Ultra short term prediction of wind power generation based on short-term memory network [J]. *Power System Technology*, 2017, 41(12): 3797-802.
12. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. *Neural Netw*, 2005, 18(5): 602-10.
13. Zhen Y, Yang H, Feng Y, et al. A convolution BiLSTM neural network model for Chinese event extraction [M]. *Natural Language Understanding and Intelligent Applications*. Springer. 2016: 275-87.

14. Wang L-J, Dong Lei, et al. Short-term Power Prediction of a Wind Farm Based on Wavelet Analysis [J]. Proceeding of the CSEE, 2009, 29(28): 30-3.
15. Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1989, 7): 674-93.
16. Mi X-W, Liu H, Li Y-F. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm [J]. Energy Conversion and Management, 2017, 151(709-22.
17. Hocheriter S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(02): 10.
18. Graves A. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-80.
19. Bai Jing, Li Fei, Ji D-H. Attention based Chinese microblog position detection model of bilstm-cnn [J]. Computer application and software, 2018, 35(3): 266-74.