A Combined Model Based on Society Cognitive Optimization Algorithm for Wind Speed Forecasting

Zhaoshuang He
Xi’an University of Posts & Telecommunication

Yanhua Chen (ieyanhuachen@zzu.edu.cn)
Zhengzhou University

Min Li
Lanzhou University

Research Article

Keywords: wind speed forecasting, ELM, Elman, LSTM, SCO

DOI: https://doi.org/10.21203/rs.3.rs-871753/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
A combined model based on society cognitive optimization algorithm for wind speed forecasting

Zhaoshuang He, Yanhua Chen and Min Li

Abstract: Wind energy, as renewable energy, has drawn the attention of society. The use of wind power generation can reduce the pollution to the environment and solve the problem of power shortage in offshore islands, grassland, pastoral areas, mountain areas, and highlands. Wind speed forecasting plays a significant role in wind farms. It can improve economic and social benefits and make an operation schedule for wind turbines in large wind farms. At present, researchers have proposed a variety of methods for wind speed forecasting; artificial neural network (ANN) is one of the most commonly used methods. This paper proposes a combined model based on the existing artificial neural network algorithms for wind speed forecasting at different heights. We first use the wavelet threshold method to the original wind speed data set for noise reduction. After that, the three artificial neural networks, extreme learning machine (ELM), Elman neural network, and Long Short-Term Memory neural network (LSTM), are applied for wind speed forecasting. In addition, variance reciprocal method and society cognitive optimization algorithm (SCO) are used to optimize the weight coefficients of the combined model. In order to evaluate the forecasting performance of the combined model, we select wind speed data at three heights (20m, 50m, and 80m) in National Wind Technology Center M2 Tower. The experimental results show that the forecasting performance of the combined model is better than the single model, and it has a good forecasting performance for the wind speed at different heights.

Keywords: wind speed forecasting; ELM; Elman; LSTM; SCO

1. Introduction

Wind energy, an essential renewable green energy, has large reserves and wide distribution, and has been widely used in many fields. At present, wind power generation is the emphasis of wind energy utilization. The latest released wind power data by the World Wind Energy Association revealed that the worldwide wind capacity has approximately reached 744 Gigawatts in 2020, wherein an unprecedented 93 Gigawatts was added [1]. There are many advantages of wind power generation. It can reduce the pressure brought by the shortage of traditional energy, and make a significant contribution to local life and development. Wind speed is the direct incarnate of wind energy, and accurate wind speed forecasting has a great significance. Therefore, it is the current research priority.

However, there is some trouble obtaining a high accuracy of wind speed forecasting results due to the random fluctuation of wind speed data caused by the weather factors. Researchers have proposed a variety of methods for wind speed forecasting, including statistical methods, physical method, ANN [2, 3], and support vector machines (SVM) [4]. Statistical forecasting method is a method based on actual historical data, theoretical knowledge, and mathematical model to make quantitative forecasting about the development of things, mainly including the trend extrapolation method, regression forecasting method [5], Delphi method [6], subjective probability method, exponential smoothing (ES) method [7, 8], autoregressive integrated moving average (ARIMA) [9], fuzzy system (FS) [10] and other methods. Singh et al. [11] proposed a new Repeated wavelet transform (WT) based ARIMA (RWT-ARIMA) model, which has improved accuracy for very short-term wind speed forecasting. Liu et al. [12] proposed a hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA, in which ARIMA is employed to predict the low frequency sub-sequences and one residual.

The artificial neural network is a commonly used method due to its particular strengths in regression and classification. At the time of forecasting, BP neural network and Elman neural network are frequently used algorithms. Altan et al. [13] developed a new WSN model based on long short-term memory (LSTM) network and decomposition methods with grey wolf optimizer (GWO). Zhang et al. [14] proposed a novel model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. Catalão et al. [15] used wavelet transform and an artificial neural network to forecast the wind speed data of Portugal. Later, with maturity technology, the researchers proposed the combination forecasting model and the hybrid forecasting model for wind speed forecasting. Many experimental results showed that the combination forecasting and hybrid forecasting model could improve the forecasting accuracy and
stability. Determining the appropriate weight coefficients is the critical step to obtain better forecasting results. Compared with the method of using the algorithms to determine the weight coefficients directly, the method of using the modern intelligent optimization algorithms like genetic algorithm (GA) [16] and particle swarm optimization algorithm (PSO) [17] to further optimize the weight coefficients is more conducive to accurate results. Li et al. [18] proposed a combination model based on variable weight for wind speed forecasting, which combined ARIMA, ENN, and BPNN. Zhang et al. [19] proposed a combined model for short-term wind speed forecasting, which included flower pollination algorithm based on chaotic local search (CLSPFA), five artificial neural networks, non-negative constraint theory (NNCT), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). Liu et al. [20] combined a data pretreatment strategy, a modified multi-objective optimization algorithm, and several forecasting models to forecast the wind speed in China. By combining a convolutional neural network and a long short-term memory neural network, Chen et al. [21] developed a multifactor spatio-temporal correlation model for wind speed forecasting. Physical method [22] is based on precise mathematical and physical laws. Although the calculation is complex, some research papers are still about the physical wind speed forecasting methods [23, 24]. The precision of the physical method is high, but the application of the physical method is not very common.

Wind speed at different heights is of great importance for a wind energy assessment of wind farms and the design of wind-resistance coefficients for high-rise buildings. In this paper, a combined forecasting model ELM-Elman-LSTM based on SCO algorithm is proposed for wind speed forecasting at different heights. Firstly, the original wind speed data is de-noised by wavelet threshold algorithm, and then input data into the model for calculation. The calculation results of each model are combined with the variance reciprocal method to obtain the intermediate forecasting results. Finally, we use SCO algorithm to optimize the weight coefficient to get the final forecasting result. The social cognitive optimization algorithm is an optimization algorithm based on the core of the social cognitive theory. We will introduce it in detail in later chapters. In order to test the forecasting performance of the combined model and the forecasting accuracy at different heights, this paper selects the wind speed data at the height of 20 meters, 50 meters, and 80 meters at National Wind Technology Center M2 Tower. Experimental results show that the combined model ELM-Elman-LSTM can improve the forecasting accuracy and have good forecasting results for wind speed data at different heights.

The structure of this paper is as follows: The first chapter mainly introduces the background, significance, and several methods of wind speed forecasting, such as artificial neural network, combined forecasting method, and hybrid forecasting method. The second chapter introduces the theoretical knowledge of artificial neural networks involved in the combination model proposed in this paper. The third chapter proposes the combination model and the evaluation index of the model selected in this paper. The fourth chapter introduces the experimental data and data set division. In the fifth chapter, we experiment and analyze the results of experiments. In order to illustrate the forecasting performance of the combined model, the combined model is compared with three single models. In the last chapter, we summarize the experiment.

2. Materials and Methods

There are diverse ANNs for wind speed forecasting, such as BP neural network, Radical Basis Function (RBF) neural network, Generalized Regression Neural Network (GRNN). Every artificial neural network has certain superiority and limitation for wind speed forecasting. In this paper, we choose three different artificial neural networks that are Extreme learning machine (ELM), Elman neural network and Long Short Term Memory networks (LSTM) respectively.

2.1. Extreme learning machine

Extreme learning machine [25, 26] proposed by Huang Bin is a new type of fast learning algorithm. ELM has a faster learning speed not merely because ELM can obtain the output weights by randomly initializing the input weight and bias, but because it can determine parameters without iteration when compared with traditional artificial neural networks, especially single hidden-layer feedforward neural networks (SLFNs) [25]. The basic network architecture of ELM has the input layer, the hidden layer, and the output layer. It is assumed that there are neuron nodes in the input layer, \( L \) neuron nodes in hidden layer, and only one neuron node in the output layer of ELM, for \( N \) samples \( X_j = [x_{j1}, x_{j2}, ..., x_{jn}], j = 1, 2, ..., N \), the operating principle of ELM are as follow:

a. The weight matrix \( w_i \) from the input layer to the hidden layer and the bias \( b_i \) of the hidden layer neuron are randomly set.

b. Calculate output matrix \( H \) of the hidden layer, select a wireless differentiable function \( g(*) \) as the
hidden layer neuron activation function, the output matrix $H$ is as follow:
\[
h(i) = g\left(\sum_{j=1}^{L} x_j w_{ij} + b_i\right) \quad i = 1, 2, \cdots, L
\]
\[
H = [h_1, h_2, \cdots, h_L]
\]
\[
\bullet \quad \text{Calculate the weight } \beta \text{ of output layer. The final output of the network is as follow:}
\]
\[
o(j) = \sum_{i=1}^{L} \beta_i h(i) \quad j = 1, 2, \cdots, N
\]
The cost function of ELM is as follow:
\[
E(w_i, b_i, \beta) = \sum_{j=1}^{N} \|o_j - t_j\|
\]
The ultimate goal of ELM is to get the smallest $E$, which is as follow:
\[
\min(E(w_i, b_i, \beta)) = \min((H\beta - T)^T(H\beta - T))
\]
$T$ is the target value matrix of the sample set. According to the singular value decomposition [27] and Moore-Penrose generalized inverse [28, 29], the weight matrix of the output layer is obtained.
\[
\beta = H^+ T
\]

2.2. Elman neural network

The Elman neuron network, a recursive neural network with a local memory unit and local feedback connection, was proposed by J. L. Elman. The structure of Elman neuron network mainly includes the input layer, correlation layer, hidden layer, and output layer. The correlation layer is a special hidden layer, which receives feedback signals from the hidden layer. In the Elman neural network structure, there are full connectivity among the neuron node of the hidden layer and that of the correlation layer. The learning algorithms of the Elman neuron network are as follows:

The connection weight between the hidden layer and the output layer is $w^4$, the weight between the input layer and the correlation layer is $w^2$, and the weight between the hidden layer and the output layer is $w^3$. The input and output of each layer are as follows:

Input layer: The input layer only plays the role of signal transmission, so the input and output all are $u(t)$.

Correlation layer:
\[
\begin{align*}
\text{Input:} & \quad net_n(t) = \sum w^2(t-1) u(t-1) + \sum w^3(t-1) x_c(t-1) \\
\text{Output:} & \quad x_c(t) = h(\text{net}_n(t))
\end{align*}
\]

Hidden layer:
\[
\begin{align*}
\text{Input:} & \quad net_n(t+1) = \sum w^2(t) u(t) + \sum w^3(t) x_c(t) \\
\text{Output:} & \quad x_n(t+1) = f(\text{net}_n(t+1))
\end{align*}
\]

Output layer:
\[
\begin{align*}
\text{Input:} & \quad net_m(t+1) = \sum w^4(t+1) x_n(t+1) \\
\text{Output:} & \quad y(t+1) = g(\text{net}_m(t+1))
\end{align*}
\]

2.3. Long short-term memory neuron network

Hochreiter and Schmidhuber first proposed Long short-term memory (LSTM) neuron network [30]. In the follow-up work, it was improved and promoted by Alex Graves [31]. LSTM, a special kind of recurrent neural network (RNN), achieved good performance in many fields like speech recognition, image recognition, and data forecasting. Different from the traditional artificial neural network, RNN introduces a directed loop in the network. That is to say, and there is a full connection between the hidden layer and the hidden layer, which makes the RNN can effectively deal with the related problems. The structure of RNN contains input layer, hidden layer, and output layer, and the training algorithm of RNN is Back Propagation through Time (BPTT) algorithm. However, there is a problem that BPTT algorithm cannot solve the problem of long-term dependence. Therefore, the
emergence of LSTM is to solve the long-term dependence.

The structure of LSTM is shown in Figure 1, as a special kind of RNN. The difference between the LSTM and RNN is that LSTM introduces a processor Cell to judge whether the information is useful, and the core of LSTM is Cell. There are three gates in the cell, namely input gate, forget gate and output gate respectively. The input gate determines how much the input of the network at the current moment can reach the cell. The Output gate controls the output of the cell. The forget gate determines whether the output of the previous state is fully retained, partially retained, or completely forgotten in the current state. The training algorithm of LSTM also selects BPTT algorithm. The learning steps of LSTM mainly include forward propagation and reverse propagation.

**Figure 1.** The structure of LSTM neural network.

Forward propagation:

Input gate:
\[ \text{input}_t = \delta(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}C_{t-1} + b_i) \]  
(13)

Forget gate:
\[ \text{forget}_t = \delta(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}C_{t-1} + b_f) \]  
(14)

Cell:
\[ C_t = \text{forget}_t C_{t-1} + \text{input}_t \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \]  
(15)

Output gate:
\[ \text{output}_t = \delta(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}C_t + b_o) \]  
(16)

The final output:
\[ h_t = \text{output}_t \tanh(C_t) \]  
(17)

In the above formula, \( \delta(x) \) is the sigmoid function, and \( \tanh(x) \) is hyperbolic function, the mathematical expressions are as follows:
\[ \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \]  
(18)

\[ \delta(x) = \text{sigmoid}(x) = \frac{1}{1+e^{-x}} \]  
(19)

Back propagation is a process of error propagation that includes two directions, one is the reverse error in time, and the other is the reverse error in space. Besides, back propagation is a process of updating the weight matrix and threshold of the input gate, the weight matrix and threshold of the output gate, the weight matrix and
threshold of the forgetting gate, and the weight matrix and threshold of the Cell. Because partial differential der-
ivation is very complex in the back propagation, so it is not described in detail in this paper, and detailed content
can refer to literature [32].

3. Proposed combined model

As the core of the combination model, there are many ways to determine the weight coefficients, such as the com-
bination model with equal weights or unequal weights. This paper chooses the variance reciprocal method to calculate
the weight of the combined model ELM-Elman-LSTM, and then optimizes the weight coefficients using the society
cognitive optimization (SCO) algorithm.

3.1. SCO-VRW

In this paper, we use the SCO-VRW method to calculate the weight coefficient of the combined model. That is to
say, we first use variance reciprocal weighting (VRW) to determine the weight coefficients of the combined model, and
then we use the SCO algorithm to optimize the weight coefficients. Variance reciprocal weighting is a more commonly
used method for determining weight coefficient, a model with a smaller sum of square error was given a high weight.
On the contrary, a model with a higher sum of square error was given a smaller weight. The calculation method is as
follow:

\[
Q_i = \frac{D_i^{-1}}{\sum_{m=1}^{m} D_m^{-1}}
\]

\[
D_i^{-1} = \sum_{t=1}^{n} (d_t - Y_t)^2
\]

At present, there are many bionic intelligent optimization algorithms used to optimize the weight coefficient. The
bionic intelligent optimization algorithm is an efficient approximation algorithm, such as ant colony optimization
(ACO), particle swarm optimization (PSO), and cuckoo search algorithm (CS). Most of these algorithms are based on
insect society. Literature [33] proposed a society cognitive optimization algorithm (SCO) based on social cognitive the-
yory. SCO algorithm simulates social learning ability in social cognitive theory through competitive selection and domain
search. In the specific implementation of the algorithm, agents are used to representing people in society. The
knowledge in society is expressed by the knowledge base. Furthermore, the social learning process is simulated through
the interaction between agents and the knowledge library to achieve optimization. There are four significant concepts
SCO algorithm:

(a) Knowledge point: A knowledge point is composed of position \( x \) and its horizontal description.

(b) Library: A base is a table that contains the knowledge points.

(c) Social cognitive agent: A social cognitive agent is a behavioral individual that dominates the knowledge point in
the base, besides, it contains a memory and \( s \) set of action rules.

(d) Neighborhood search: It is assumed that there are two knowledge points that are \( x_1 \) and \( x_2 \) respectively,
the neighborhood search of \( x_1 \) use \( x_1 \) as the reference point and \( x_2 \) as the center point to calculate a new knowledge
point variable \( x_3 \):

\[
x_3 = x_1 + 2 \cdot Rand(x_2 - x_1)
\]

Assuming that the number of knowledge points in the library is \( N_{pop} \), the number of social cognitive agents is \( N_c \),
and iterative frequency is \( T \). The flow chart of SCO algorithm is shown in Figure 2, according to Figure 2, the basic steps
of SCO algorithm are as follows:

(1) Initialization. The \( N_{pop} \) knowledge points are randomly initialized, and the knowledge points in the library
are randomly assigned to each social cognitive agent. However, it is not allowed to assign a knowledge point to multiple
social cognitive agents repeatedly.

(2) Alternative learning process. Perform the following actions for each social cognitive agent.

- Imitation selection. First, two or more points (the selected knowledge points differ from the current social
cognitive agent) are randomly selected from the library, then compare the fitness values of the selected knowledge
points and choose the best one based on the principle of competitive selection.

- Observational learning. First, compare the fitness value of the knowledge point selected by imitation selec-
tion with that of the social cognitive agent. The best point is taken as the central point and another point as the reference
point. The social cognitive agent is updated based on the neighborhood searching principle, and the new knowledge
point is saved to the base.
Updating library. There are $N_c$ new knowledge points increased in the process of alternative learning. $N_c$ knowledge points with the worst fitness needs to be removed from the library in order to keep the number of knowledge points in the library remain unchanged.

Determine whether the result satisfies the end conditions. If not, repeating step 2) to step 4) until the end condition is reached.

Figure 2 shows the flowchart of weight coefficient optimization, and the following points should be paid attention to in the process of optimization.

- In the initialization process of the library, the knowledge points in the library should meet the precondition that the sum of the weight coefficients is 1. Otherwise, it must be reinitialized.

There will be a new knowledge point in the observational learning process, and new knowledge points need to be tested to meet the precondition that the sum of the weight coefficient is 1.

3.2. ELM-Elman-LSTM model

The structure diagram of the combined ELM-Elman-LSTM is shown in Figure 3. According to Figure 3, the basic forecasting steps using ELM-Elman-LSTM are as follows.

Step 1: Data de-noising. The wavelet threshold algorithm is used to de-noise the original wind speed data.

Step 2: Three single models are used individually to forecast the data after noise reduction.

- Using the ELM to train the wind speed data after the noise reduction and get the forecasting result $y_1$.
- Using the Elman to train the wind speed data after the noise reduction and get the forecasting result $y_2$.
- Using the LSTM to train the wind speed data after the noise reduction and get the forecasting result $y_3$.

Step 3: Forecasting the wind speed data with the ELM-Elman-LSTM model.

- Using variance reciprocal weighting method to calculate the weight of the combined model. Using the SCO algorithm to optimal weight, the final forecasting results are as follows.

$$\frac{1}{Q_1 + Q_2 + Q_3} Y = Q_1 y_1 + Q_2 y_2 + Q_3 y_3$$

(22)
3.3. Evaluation index

Generally, the evaluation index used commonly is MAE, MSE, RMSE, MAPE, and SSE. Each evaluation index has its advantages and limitations, so it cannot evaluate the model’s forecasting performance from one angle. This paper chooses MAE, MSE, MAPE and R-square as the evaluation indexes to evaluate the proposed model ELM-Elman-LSTM, the mathematical expressions are as follows. The smaller the value of MAE, MSE, and MAPE, the better the forecasting results, while R-square represents a fitting effect through the change of data, and the range of value is between [0,1], the closer to 1, the better fitting effect of the model.

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\text{observed}_i - \text{predicted}_i| \quad (23)
\]

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\text{observed}_i - \text{predicted}_i)^2 \quad (24)
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\text{observed}_i - \text{predicted}_i}{\text{observed}_i} \right| \times \frac{100}{N} \quad (25)
\]

\[
\text{R}^2 = \frac{\sum_{i=1}^{N} (\text{predicted}_i - \text{mean}_{\text{observed}})^2}{\sum_{i=1}^{N} (\text{observed}_i - \text{mean}_{\text{observed}})^2} \quad (26)
\]

4. Data sets information
In order to test the forecasting performance of the proposed model in this paper and the effect of wind speed forecasting at different heights, we selected wind speed data at National Wind Technology Center M2 Tower. The data are from April 10 to April 12 and April 16 in 2017 as experimental data, and they are Monday, Tuesday, Wednesday, and Sunday. The experimental data includes three heights: wind speed data at heights of 20 meters, 50 meters, and 80 meters, respectively. The data were sampled every five minutes and collected from 0:00 to 23:59, so there are 288 observations per day. The wind speed data of each height is shown in Figure 4. As can be seen from the figure, the general trend of wind speed data at each height is the same, and the higher the height, the greater the wind speed. In addition, comparing the four maps in the figure, we can see that the wind speed data selected in this paper is not periodic, and the wind speed data varies from one day to another day.

![Figure 4. Observations of wind speed at different heights.](image)

When wind speed data is measured and collected, noise is generated for various reasons. The original wind speed data will be noise-reduced to improve the forecasting accuracy. Noise reduction methods used commonly are EMD, EEMD, and PCA. In this paper, we choose the wavelet threshold method for data de-noising. We use soft threshold de-noising and hard threshold de-noising to process the same set of data. A better method with an excellent de-noising effect is selected. Generally speaking, the effect of de-noising is evaluated by SNR and mean square error (MSE). The higher the SNR, the better the de-noising effect, and the smaller the MSE, the better the effect. The results of noise reduction are shown in Table 1. It can be seen from the table that the effect of hard threshold de-noising is better than that of soft threshold de-noising.

| Data     | MSE       | SNR       |
|----------|-----------|-----------|
|          | 20m | 50m | 80m | 20m | 50m | 80m |
| Monday   | Soft threshold | 0.1518 | 0.1146 | 0.0804 | 23.93 | 25.98 | 27.72 |
|          | Hard threshold | 0.0995 | 0.0636 | 0.0333 | 25.78 | 28.55 | 31.56 |
| Tuesday  | Soft threshold | 0.0543 | 0.1312 | 0.0801 | 23.84 | 21.04 | 23.29 |
|          | Hard threshold | 0.0187 | 0.0939 | 0.0485 | 28.05 | 22.51 | 25.49 |
| Wednesday| Soft threshold | 0.0697 | 0.0823 | 0.0614 | 21.62 | 22.04 | 23.57 |
|          | Hard threshold | 0.0337 | 0.0396 | 0.0230 | 24.80 | 25.24 | 27.87 |
| Sunday   | Soft threshold | 0.0218 | 0.0214 | 0.0286 | 27.33 | 28.61 | 27.68 |
|          | Hard threshold | 0.0066 | 0.0078 | 0.0130 | 32.55 | 33.03 | 31.12 |
Figure 5 shows the input and output of the ELM-Elman-LSTM model. The data at the three heights has the same input and output modes. For example, at 20 meters high, using data from time 1 to 5 to forecast the wind speed data at time 6, using data from time 2 to 6 to forecast the wind speed data at time 7. In this way, the wind speed data of K to K+4 is used to forecast the wind speed data at time K+5. Therefore, the input of the model is the wind speed data at time K to time K+4, and the output is the wind speed data at the time of K+5. The input vector has a dimension of 5, and the output vector has a dimension of 1. In other words, for the artificial neural network selected in this paper, the number of nodes in the input layer is 5, and the number of nodes in the output layer is 1.

5. Experiment results

In order to better illustrate the forecasting performance of the combined forecasting model ELM-Elman-LSTM, we compare the forecasting results of the combined model with that of the three single models (ELM, Elman, and LSTM) from the four evaluation indexes (MSE, MAE, MAPE and R-square).

5.1. Forecasting results of individual models

In forecasting with single models, the evaluation indexes for the model we selected are MAE, RMSE, and MAPE. The number of hidden layers and neurons in each hidden layer greatly impacts the forecasting results. For the ELM network, there is only one hidden layer. For the data selected in this paper, compared with the 20 hidden layer neurons network, the forecasting effect is better than the ELM network of 10 hidden layer neurons. The forecasting results of ELM are shown in Figure 6. Table 2 shows the evaluation indexes of the ELM model.

First, by comparing the forecasting results at the same height of different datasets, we can get the following results. For wind speed data at the height of 20 meters, ELM had the best forecasting effect on Tuesday with the minimum value of three evaluation indexes, in which case the value of MSE is 0.1773, the value of MAE is 0.3475, and the value of MAPE is 9.89%. For wind speed data at the height of 50 meters and 80 meters, there were the smallest values of MSE and MAE on Tuesday, while the smallest value of MAPE on Sunday. For example, the value of MSE is 0.1914, the value of MAE is 0.3387, and the value of MAPE is 7.03%.

Second, we compare the forecasting results at different heights in the same data set. It can be seen that the value of MAPE decreases gradually as the height increases for all data sets. However, the changes are not regular. For example, in the forecasting results on Tuesday, the MAPE is 9.89%, 7.68%, and 7.51% at the height of 20, 50, 80 meters. There is no particular relationship between height and the value of MSE and MAE. For example, in the forecasting results on Tuesday, the MSE and MAE are 0.1773, 0.3475 respectively, at the height of 20 meters. The MSE and MAE are 0.1914, 0.3387 respectively, at the height of 50 meters. At the same time, the MSE and MAE are 0.1603, 0.3222 respectively, at the height of 80 meters.

In addition, the forecasting results at the height of 80 meters on Sunday are the best, in which the MAPE is 6.34%.
Table 2. The evaluation index of ELM.

| Date     | MSE   | MAE   | MAPE (%) |
|----------|-------|-------|----------|
|          | 20m   | 50m   | 80m      | 20m   | 50m   | 80m      | 20m   | 50m   | 80m      |
| Monday   | 0.2413| 0.2380| 0.2352   | 0.3685| 0.3690| 0.3591   | 14.68 | 11.46 | 10.50    |
| Tuesday  | 0.1773| 0.1914| 0.1603   | 0.3475| 0.3387| 0.3222   | 9.89  | 7.68  | 7.51     |
| Wednesday| 0.2876| 0.2587| 0.3668   | 0.3983| 0.3803| 0.4310   | 17.67 | 15.59 | 13.70    |
| Sunday   | 0.5074| 0.3192| 0.3307   | 0.4705| 0.3673| 0.3616   | 12.68 | 7.03  | 6.34     |

Figure 6. The bar chart for the evaluation index of ELM.

When using Elman neural network to forecast the wind speed, the same attention should be paid to the selection of hidden layer nodes. The result is shown in Figure 7.

First, by comparing the forecasting results of the data at the same height of different datasets, we can get the following results. For wind speed data at the height of 20 meters, the forecasting result of Elman has the smallest value of MSE and MAPE on Tuesday. The value of MSE is 0.2540 and MAPE is 11.89%, while the value of MAE is 0.4179, which is higher than that on Wednesday. For wind speed data at the height of 50 meters, there is the smallest value of three evaluation indexes on Tuesday. The value of MSE is 0.1936, MAE is 0.3315, and the MAPE is 7.46%. For wind speed data at the height of 80 meters, there is the smallest value of MSE and MAPE on Tuesday, while the value of MAE is higher than that on Monday.

Second, by comparing the forecasting results of wind speed at different heights in the same data set. We can see that there is no definite rule for the change of three evaluation indexes. For example, on Monday, the value of MSE and MAPE decreased as the height increased, while the value of the MAE did not have the same regular. On Wednesday and Sunday, the value of MSE and MAE increased as the height decreased, but the change of MAPE is not the same. In addition, on Monday and Wednesday, the value of MAPE is decreased as the height increased. In short, when using Elman neural network for wind speed forecasting at different heights, it is not sure that the value of MAPE will decrease as the height increases.

In addition, the forecasting results at the height of 50 meters on Tuesday are the best, in which of the situation the value of MAPE is 7.46%.

Table 3. The evaluation index of Elman.

| Date     | MSE   | MAE   | MAPE (%) |
|----------|-------|-------|----------|
|          | 20m   | 50m   | 80m      | 20m   | 50m   | 80m      | 20m   | 50m   | 80m      |
| Monday   | 0.3618| 0.3333| 0.2494   | 0.4538| 0.4735| 0.3727   | 17.73 | 13.97 | 10.81    |
| Tuesday  | 0.2540| 0.1936| 0.2257   | 0.4179| 0.3315| 0.3870   | 11.89 | 7.46  | 8.75     |
| Wednesday| 0.2989| 0.3316| 0.3766   | 0.3852| 0.4144| 0.4446   | 16.13 | 15.82 | 14.05    |
| Sunday   | 0.4560| 0.6091| 0.7100   | 0.4728| 0.4964| 0.6270   | 12.62 | 8.71  | 9.69     |
The wind speed data after the noise reduction are entered into the LSTM, and the forecasting results are shown in Figure 8 and Table 4. According to the figure and table, we can get the following conclusions.

First of all, we compare the forecasting results at the same height of different data sets. It can be concluded that the LSTM model has a better forecasting performance at the height of 20 meters on Tuesday, while the value of MAPE is 10.87% and MSE is 0.2336, the smallest value in all forecasting results. For a height of 50 meters and 80 meters, there is the smallest value of MSE and MAE on Wednesday, but the value of MAPE is relatively large at this time. For example, at the height of 50 meters, the value of MAPE is 16.26% on Wednesday, while the value of MAPE was 8.94% on Tuesday.

Secondly, comparing the forecasting results at different heights of the same dataset, it can be seen that the LSTM does not have specific rules for wind speed forecasting at a different height from the three evaluation indexes. However, from the perspective of MAPE alone, among the four-day forecasting results, except for Wednesday, there are three days that the value of MAPE is the smallest at the height of 50 meters. That also shows that LSTM has some advantages for wind speed forecasting at the height of 50 meters.

### Table 4. The three evaluation index of Elman.

| Date     | MSE 20m | MSE 50m | MSE 80m | MAE 20m | MAE 50m | MAE 80m | MAPE (%) 20m | MAPE (%) 50m | MAPE (%) 80m |
|----------|---------|---------|---------|---------|---------|---------|--------------|--------------|--------------|
| Monday   | 0.3543  | 0.2515  | 0.2613  | 0.4746  | 0.3775  | 0.4155  | 19.47        | 11.58        | 12.29        |
| Tuesday  | 0.2336  | 0.3285  | 0.4641  | 0.4030  | 0.4174  | 0.5223  | 10.87        | 8.94         | 10.42        |
| Wednesday| 0.2844  | 0.2979  | 0.3965  | 0.3748  | 0.4033  | 0.4748  | 15.67        | 16.26        | 15.37        |
| Sunday   | 0.7401  | 0.7100  | 1.3588  | 0.6389  | 0.5866  | 0.7314  | 16.56        | 10.54        | 11.19        |

5.2. **Forecasting results of ELM-Elman-LSTM**

The forecasting results of ELM, Elman, and LSTM are combined with a certain weight using variance reciprocal weighting method, and then the weight coefficient is optimized by the SCO algorithm. Figure 9 shows the forecasting results of ELM-Elman-LSTM model. It can be seen from the figure that the forecasting curve of the combined model ELM-Elman-LSTM is basically in line with the actual price curve, especially at the height of 50 meters and 80 meters on Sunday, the forecasting curve is closer to the actual price curve. In comparison, the fit between the forecasting curve and the actual price curve at the height of 20 meters is not as good as that of the remaining two heights.
In order to better prove the forecasting performance of the combined model, the model ELM-Elman-LSTM is compared with three single models (ELM, Elman, and LSTM), respectively. Table 5 records the three evaluation indexes (MSE, MAE, and MAPE), and Table 6 records the value of the R-square of the four models.

First, we make a simple comparison of three individual models. From the perspective of MAE and MSE, the MAE and MSE of the ELM model are higher than that of the other two models at the height of 20 meters on Wednesday, and the MAE and MSE of ELM are less than the other two models in the remaining data sets. From the perspective of MAPE, the MAPE using the ELM model is higher than the Elman at the height of 50 meters on Wednesday and 20 meters on Sunday. The MAPE of ELM is higher than that of the other two models at the height of 20 meters on Wednesday. Except for these two cases, the three evaluation indexes of ELM are smaller than those of the other two models. Based on this, it can also be said that ELM has better forecasting performance than Elman and LSTM. In addition, it can also be seen from the table that the evaluation indexes of Elman and LSTM fluctuate up and down, and the two models have different advantages for different data sets. Next, we compare the combined model with single models.

Compare the forecasting results of the ELM-Elman-LSTM model with the ELM neural network. The value of MSE using ELM-Elman-LSTM is higher than that of ELM at the height of 50 meters on Wednesday and Monday. For example,
for the wind speed at the height of 50 meters on Wednesday, the MSE is 0.2600 of ELM-Elman-LSTM, while the MSE of the ELM is 0.2587. Except for the above case, the values of the three evaluation indexes of the remaining data sets are less than that of ELM. For example, for the wind speed data at the height of 20 meters on Monday, the MAPE of ELM is 14.68%, and the MAPE of ELM-Elman-LSTM is 13.46% that decreased by 1.22%.

Compare the forecasting results of the ELM-Elman-LSTM model with the Elman neural network. The value of the three evaluation indexes forecasted by ELM-Elman-LSTM is smaller than that of the Elman model. It can also be said that the forecasting performance of the ELM-Elman-LSTM model is better than Elman neural network. For example, for the wind speed data at the height of 20 meters on Tuesday, the MSE is 0.2540, the MAE is 0.4179, and the MAPE is 11.89% forecasted by Elman neural model. While the MSE is 0.1618, the MAE is 0.3140, and the MAPE is 8.63%, forecasted by ELM-Elman-LSTM, the value of MAPE decreased by 3.26%.

Compare the forecasting results of the ELM-Elman-LSTM model with the LSTM neural network. The value of the three evaluation indexes forecasted by ELM-Elman-LSTM is smaller than that of the LSTM model for all datasets, observed from the angle of MAPE. It can be seen that the value of MAPE is greatly reduced forecasted by ELM-Elman-LSTM. For example, at the height of 80 meters on Sunday, the MAPE of LSTM is 11.69%, MAPE of ELM-Elman-LSTM is 5.96% which is reduced by 5.73%.

In conclusion, when using MSE, MAE, and MAPE to evaluate the model, the forecasting performance of the ELM-Elman-LSTM model for the wind speed at different heights is superior to the three single forecasting models. According to Table 6 and Figure 10, for the wind speed at the height of 50 meters on Sunday, the R-square value forecasted by ELM-Elman-LSTM is smaller than ELM. For the remaining data sets, the R-square value of ELM-Elman-LSTM is greater than the other three individual models. Moreover, the R-square of the combined model ELM-Elman-LSTM is close to 1. For example, for the data at the height of 80 meters on Tuesday, R-square of ELM-Elman-LSTM can reach 0.9388, but for the data at the height of 20 meters on Wednesday, the value of R-square of the four methods is not very close to 1, the highest value is 0.5447.

![Figure 10. The value of R-square for the four models.](image-url)
### Table 5. The three evaluation indexes value of the ELM-Elman-LSTM and the single models.

| Day    | height | MAE  | MSE  | MAPE (%) | ELM | Elman | LSTM | ELM-Elman-LSTM | ELM | Elman | LSTM | ELM-Elman-LSTM | ELM | Elman | LSTM | ELM-Elman-LSTM |
|--------|--------|------|------|----------|-----|-------|------|---------------|-----|-------|------|---------------|-----|-------|------|---------------|
| Monday | 20m    | 0.3685 | 0.4538 | 0.4746 | 0.3474 | 0.2413 | 0.3618 | 0.3543 | 0.2416 | 14.68 | 17.73 | 19.47 | 13.46 |
|        | 50m    | 0.3690 | 0.4735 | 0.3775 | 0.3426 | 0.2380 | 0.3333 | 0.2515 | 0.2266 | 11.46 | 13.97 | 11.58 | 10.68 |
|        | 80m    | 0.3591 | 0.3727 | 0.4155 | 0.3564 | 0.2352 | 0.2494 | 0.2613 | 0.2307 | 10.50 | 10.81 | 12.29 | 10.40 |
| Tuesday| 20m    | 0.3475 | 0.4179 | 0.4030 | 0.3140 | 0.1773 | 0.2540 | 0.2336 | 0.1618 | 9.89  | 11.89 | 10.87 | 8.63  |
|        | 50m    | 0.3387 | 0.3315 | 0.4174 | 0.3247 | 0.1914 | 0.1936 | 0.3285 | 0.1809 | 7.68  | 7.46  | 8.94  | 7.35  |
|        | 80m    | 0.3222 | 0.3870 | 0.5223 | 0.3139 | 0.1603 | 0.2257 | 0.4641 | 0.1494 | 7.51  | 8.75  | 10.42 | 7.36  |
| Wednesday| 20m | 0.3983 | 0.3852 | 0.3748 | 0.3736 | 0.2876 | 0.2989 | 0.2844 | 0.2820 | 17.67 | 16.13 | 15.67 | 15.63 |
|        | 50m    | 0.3803 | 0.4144 | 0.4033 | 0.3789 | 0.2587 | 0.3316 | 0.2979 | 0.2600 | 15.59 | 15.82 | 16.26 | 15.54 |
|        | 80m    | 0.4310 | 0.4446 | 0.4748 | 0.4300 | 0.3668 | 0.3766 | 0.3965 | 0.3599 | 13.70 | 14.05 | 15.37 | 13.48 |
| Sunday | 20m    | 0.3805 | 0.4278 | 0.6389 | 0.4240 | 0.5074 | 0.4560 | 0.7401 | 0.3987 | 12.68 | 12.62 | 16.56 | 11.82 |
|        | 50m    | 0.3673 | 0.4964 | 0.5866 | 0.3592 | 0.3192 | 0.6091 | 0.7100 | 0.3278 | 7.03  | 8.71  | 10.54 | 6.87  |
|        | 80m    | 0.3616 | 0.6270 | 0.7314 | 0.3337 | 0.3307 | 0.9796 | 1.3588 | 0.2543 | 6.34  | 9.69  | 11.19 | 5.96  |

### Table 6. The R-square of ELM-Elman-LSTM and the three single models.

| Day    | height | ELM | Elman | LSTM | ELM-Elman-LSTM | ELM | Elman | LSTM | ELM-Elman-LSTM | ELM | Elman | LSTM | ELM-Elman-LSTM | ELM | Elman | LSTM | ELM-Elman-LSTM |
|--------|--------|-----|-------|------|---------------|-----|-------|------|---------------|-----|-------|------|---------------|-----|-------|------|---------------|
| Monday | 20m    | 0.7768 | 0.8478 | 0.8653 | 0.8644 | 0.8916 | 0.9343 | 0.5356 | 0.7568 | 0.7909 | 0.8266 | 0.8397 | 0.8639 |
|        | 50m    | 0.7192 | 0.7944 | 0.8572 | 0.8162 | 0.8903 | 0.9075 | 0.5175 | 0.7739 | 0.7781 | 0.8118 | 0.6942 | 0.5968 |
|        | 80m    | 0.6673 | 0.8374 | 0.8504 | 0.8399 | 0.8139 | 0.8089 | 0.5409 | 0.7893 | 0.7825 | 0.7962 | 0.6435 | 0.4408 |
| Tuesday| 20m    | 0.7908 | 0.8526 | 0.8679 | 0.8672 | 0.8975 | 0.9388 | 0.5447 | 0.7836 | 0.7913 | 0.8350 | 0.8354 | 0.8953 |
|        | 50m    | 0.7363 | 0.7973 | 0.8410 | 0.8153 | 0.8524 | 0.8903 | 0.5175 | 0.7829 | 0.7850 | 0.8118 | 0.6942 | 0.5968 |
|        | 80m    | 0.6873 | 0.8447 | 0.8532 | 0.8399 | 0.8139 | 0.8089 | 0.5409 | 0.7893 | 0.7825 | 0.7962 | 0.6435 | 0.4408 |
| Wednesday| 20m | 0.8069 | 0.8644 | 0.8865 | 0.8864 | 0.9166 | 0.9433 | 0.5447 | 0.7836 | 0.7913 | 0.8350 | 0.8354 | 0.8953 |
|        | 50m    | 0.7500 | 0.8162 | 0.8572 | 0.8162 | 0.8903 | 0.9075 | 0.5175 | 0.7739 | 0.7781 | 0.8118 | 0.6942 | 0.5968 |
|        | 80m    | 0.7000 | 0.8447 | 0.8532 | 0.8399 | 0.8139 | 0.8089 | 0.5409 | 0.7893 | 0.7825 | 0.7962 | 0.6435 | 0.4408 |
| Sunday | 20m    | 0.8200 | 0.8844 | 0.9065 | 0.8864 | 0.9166 | 0.9433 | 0.5447 | 0.7836 | 0.7913 | 0.8350 | 0.8354 | 0.8953 |
|        | 50m    | 0.7633 | 0.8362 | 0.8872 | 0.8162 | 0.8903 | 0.9075 | 0.5175 | 0.7739 | 0.7781 | 0.8118 | 0.6942 | 0.5968 |
|        | 80m    | 0.7133 | 0.8644 | 0.8865 | 0.8864 | 0.9166 | 0.9433 | 0.5447 | 0.7836 | 0.7913 | 0.8350 | 0.8354 | 0.8953 |
Next, we compare the forecasting result of the ELM-Elman-LSTM model on wind speed data at different heights. Figure 11 shows the value of the R-square coefficient and MAPE of the combined model ELM-Elman-LSTM at different heights. According to Figure 11 and Table 6, we can see that the value of the R-square of the wind speed data at the height of 80 meters is maximum in the three heights and is also closest to 1. For example, the maximum value of R-square can reach 0.9388. The value of MAPE at the height of 80 meters is greater than that at 50 meters only on Tuesday. It also can be said that ELM-Elman-LSTM has the best forecasting performance for the wind speed at the height of 80 meters. In addition, if the model is only evaluated from the perspective of R-square, only the value of R-square for wind speed at the height of 20 meters on Wednesday is smaller, which is only 0.5447. For the rest of the data set, the forecasting performance of the wind speed at the height of 20 meters is as good as that at the height of 50 meters. However, if we evaluate the model from the perspective of MAPE, the forecasting performance of the data at the height of 50 meters is better than that at 20 meters.

In short, it can be seen from the forecasting results that the ELM-Elman-LSTM can improve the forecasting accuracy compared to a single forecasting model and obtain better forecasting results for wind speed data of different heights. Significantly, the forecasting result at the height of 80 meters is perfect.

6. Conclusion

As a kind of renewable green energy, wind energy utilization has become an increasingly important concern. Using wind power to replace traditional energy combustion power generation is an essential application method, which can reduce the pressure caused by the gradual reduction of traditional energy sources, reduce environmental pollution and contribute to the sustainable development of society. Therefore, how to effectively and reasonably apply wind energy has become the focus of research. Wind speed is a direct manifestation of wind energy, and accurate wind speed forecasting has a significant impact on the allocation of wind farms and the stable development of the electricity market. This paper proposes a combined forecasting model ELM-Elman-LSTM, and the society cognitive optimization algorithm (SCO) is used for the optimization of weight coefficients. In addition, in order to obtain better forecasting results, the original wind speed data is de-noised by the wavelet threshold method. The wind speed data at three different heights are chosen to test the model. The experimental results show that the ELM-ELman-LSTM can improve forecasting accuracy compared with a single forecasting model. Furthermore, the ELM-Elman-LSTM has better forecasting performance for wind speed at different heights.

Author Contributions: Conceptualization, S.Z.H. and M.L.; Methodology, H.Y.C.; Software, S.Z.H.; Validation, M.L., H.Y.C.; Formal Analysis, S.Z.H. and M.L.; Investigation, M.L.; Resources, H.Y.C.; Data Curation, S.Z.H.; Writing-Review & Editing, S.Z.H. and Y.H.C.; Visualization, M.L. and H.Y.C.; Supervision, H.Y.C.; Funding Acquisition, M.L.

Funding: Supported by New Star Team of Xi'an University of Posts & Telecommunications.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References
1. Global Status of Wind Power http://gwec.net/global-figures/wind-energy-global-status/
2. Cadenas E, Rivera W. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model J]. Renewable Energy, 2010, 35(12): 2732-2738.
3. Shukur O B, Lee M H. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA J]. Renewable Energy, 2015, 76: 637-647.
4. Liu M, Cao Z, Zhang J, et al. Short-term wind speed forecasting based on the Jaya-SVM model[J]. International Journal of Electrical Power & Energy Systems, 2020, 121: 106056.
5. Santamaria-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: A method based on support vector regression [J]. Renewable Energy, 2016, 85: 790-809.

6. Deveci M, Ozcak E, John R, et al. A study on offshore wind farm siting criteria using a novel interval-valued fuzzy-rough based Delphi method [J]. Journal of Environmental Management, 2020, 270: 110916.

7. Cadenas E, Jaramillo O A, Rivera W. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method[J]. Renewable Energy, 2010, 35(5): 925-930.

8. Zhang K, Qu Z, Dong Y, et al. Research on a combined model based on linear and nonlinear features-A case study of wind speed forecasting[J]. Renewable Energy, 2019, 130: 814-830.

9. Suyono H, Prabawanti D O, Shidiq M, et al. Forecasting of Wind Speed in Malang City of Indonesia using Adaptive Neuro-Fuzzy Inference System and Autoregressive Integrated Moving Average Methods[C]/2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP). IEEE, 2020: 131-136.

10. Moreno S R, dos Santos Coelho L. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system[J]. Renewable energy, 2018, 126: 736-754.

11. Singh S N, Mohapatra A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting[J]. Renewable energy, 2019, 136: 758-768.

12. Liu M D, Ding L, Bai Y L. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction[J]. Energy Conversion and Management, 2021, 233: 113917.

13. Altan A, Karasu S, Zio E. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer[J]. Applied Soft Computing, 2021, 100: 106996.

14. Zhang Y, Chen B, Pan G, et al. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting[J]. Energy Conversion and Management, 2019, 195: 180-197.

15. J.P.S. Catalão, H.M.I. Pousinho, V.M.F. Mendes. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. Renewable Energy, Volume 36, Issue 4, April 2011, Pages 1245-1251.

16. Khelli K, Berrezzek F, Bouadjila T. GA-based design of optimal discrete wavelet filters for efficient wind speed forecasting[J]. Neural Computing and Applications, 2021, 33(9): 4373-4366.

17. He Z, Chen Y, Shang Z, et al. A novel wind speed forecasting model based on moving window and multi-objective particle swarm optimization algorithm[J]. Applied Mathematical Modelling, 2019, 76: 717-740.

18. Hongmin Li, Jianzhou Wang, Haiyan Lu, Zhenhai Guo. Research and application of a combined model based on variable weight for short term wind speed forecasting. Renewable Energy, Volume 116, Part A, February 2018, Pages 669-684.

19. Wenyu Zhang, Zongxi Qu, Kequan Zhang, Wenhuan Mao, Xu Fan. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Conversion and Management. Volume 136, 15 March 2017, Pages 439-451.

20. Liu Z, Jiang P, Zhang L, et al. A combined forecasting model for time series: Application to short-term wind speed forecasting[J]. Applied Energy, 2020, 259: 114137.

21. Chen Y, Zhang S, Zhang W, et al. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting[J]. Energy Conversion and Management, 2019, 185: 783-799.

22. Wang, Han, et al. “Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system.” Applied Energy 237 (2019): 1-10.

23. Yan, Jie, et al. “Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping.” IEEE Transactions on Power Systems 33.3 (2017): 3276-3284.

24. Hoolohan, Victoria, Alison S. Tomlin, and Timothy Cockerill. “Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data.” Renewable Energy 126 (2018): 1043-1054.

25. Huang G B, Zhu Q Y, Siew C K. Extreme Learning machine: a new learning scheme of feed forward neural network[C]. In proceedings of the International Joint Conference on Neural Networks (IJCNN 2004). Budapest, Hungary, July 2004:985-990.

26. Huang G B, Zhu Q Y, Siew C K. Extreme Learning machine: Theory and applications [J]. Neurocomputing, 2006, 70:489-501.

27. Ortega JM, Matrix Theory [M]. New York and London: Plenum Press, 1987.

28. Rong H J, Ong Y S, H T A, et al. A fast pruned-extreme learning machine for classification problem [J]. Neurocomputing. 2009, 72: 359–366.

29. Lan Y, Soh Y C, Huang G B. Constructive hidden nodes selection of extreme learning machine for regression [J]. Neurocomputing, 2010, 73: 3193–3199.

30. Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems[J]. Advances in neural information processing systems, 1997: 473-479.

31. Alex Graves, Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, Volume 18, Issues 5–6, July–August 2005, Pages 602-610.

32. S. Hochreiter, J. Schmidhuber. Long short-term memory. Neural Comput., 9 (8) (1997), pp. 1735-1780.

33. Xie Xiaofeng, Zhang Wenjun. [C]//Proc of Genetic and Evolutionary Computation Conference. Washington, USA: [s. n.2004:261-262.