Long-Term Weather Prediction Based on GA-BP Neural Network

Kening Dou, Xiaoqi Sun*
School of Mathematics and Statistics, Qingdao University, Shandong, 266071, China
Email: sunxiaoqi@live.com

Abstract. A genetic algorithm is used to optimize the weights and thresholds of the BP neural network in this paper, this method improves the disadvantages of the BP neural network. For example, it is insensitive to weights and thresholds. In addition, it has a slow convergence rate and is prone to fall into local minima, etc. To verify the improvement effect, the BP neural networks and the improved BP neural networks were used to simulate the weather in Rizhao city, Shandong Province, respectively. The results of simulation show that the GA-BP neural network model is superior to BP neural network model in the accuracy of temperature prediction in long-term weather. Next, the GA-BP neural network was further optimized by selecting the optimal number of hidden layers and changing the impact factors, and the prediction results were fitted. The analysis shows that the accuracy of the improved GA-BP model in prediction is higher.

Keywords: BP neural network, Weather prediction, Genetic algorithm

1. Introduction

The human brain is a network structure composed of interwoven neurons of different organisms. It can complete complex processes such as thinking, perception, association, and memory, so it is known as an effective and perfect information processing system. Artificial Neural Networks (ANNs) [1], composed of many interconnected neurons, are complex network systems that simulate the structure and function of the human brain. The artificial neural network has the ability to self-adaptability and solving nonlinear problems. It can process information by adjusting the connection strength between neurons. The research on the neural network can be divided into research on its theory and its application. In its application research, one is the analysis and research on the software simulation and hardware implementation of ANNs, and the other is the research on its specific application in some fields, such as pattern recognition, combination optimization, signal processing, and so on. By the middle and late 1980s, different types of artificial neural network models were proposed. At the same time, scholars at home and abroad accelerated the exploration of their theories and applications, and a large number of research results emerged, which further promoted the development of an artificial neural network.

The application of artificial neural networks in meteorology started late, and now its main applications [2] include short-term climate prediction, agricultural meteorology, typhoon prediction, soil moisture prediction, visibility prediction, air pollution prediction, etc. Compared with the traditional numerical forecast model, an artificial neural network is applied to the weather forecast, which reduces the complexity of the implementation process, reduces the limitation of data, and improves the accuracy of the forecast by comprehensively considering various nonlinear factors. Therefore, it is more in line with people's needs for the accuracy of the weather forecast. Gradually,
the traditional mathematical equations and probability statistics are replaced by the ANN.

With the rapid development of artificial neural network technology, the BP algorithm is introduced into ANN, and then BP neural network is obtained. The BP neural network is widely applied in prediction, catching the eyes of many scholars, and abundant research results are obtained. For example, Cao X F [3] established a feedforward BP neural network, and used the method of a variable gradient to carry out correlation analysis on the meteorology of wheat scab in the Yangtze River Basin, and obtained effective and accurate prediction results, which provided theoretical guidance for controlling the occurrence of wheat scab in this area. Yuan M Y [4] used BP prediction model to make the precipitation and temperature forecast of Heilongjiang Province, they predicted the temperature in 1998 and 1999 and compared the results with the real temperature. The results were not far from each other, and the prediction was more accurate. Later, based on BP neural network model, some scholars further improved the BP neural network model by using different algorithms to influence the input factors at the input layer. For example, in literature [5], rainfall-runoff is divided into flood season and non-flood season through using the partition algorithm and then predicted by combining with BP neural network model. The results show that the improved model can significantly improve the prediction of rainfall-runoff. Liu D [6] took the probability of daily and lunar phases as the BP neural network’s input factors, then he established a long-term weather forecast model, and the results showed that compared with the method of direct prediction of long-term weather with the probability of daily and lunar phases, the model combined with BP’s prediction accuracy was higher. BP neural network is used for prediction greatly improves the prediction accuracy, but because the neural network training is usually a complex process, and the weights of along with the network and the change of the training sample set, will lead to the randomness of network training results and easy to fall into local extremum, so there are still many problems to be studied in the field of prediction. Therefore, some scholars aim at the shortcomings of the BP neural network and try to improve it with a genetic algorithm. For example, Ai H F [7] added a genetic algorithm to the BP neural network model to establish the GA-BP prediction model and simulated the PM2.5 content in Changchun. Zhao Z [8] predicted and simulated the PM2.5 value of a certain place by GA-BP model. Yang Y [9] used GA-BP neural network to predict and analyze the air index during the heating period. Zhang P D [10] used GA-BP and BP neural networks to predict and simulate urban air quality respectively, and the results all showed that the improved BP neural network model had higher prediction accuracy.

In this research, a long-term BP model and its optimized prediction model are used to train the temperature related data in Rizhao city, Shandong Province, and then predict the temperature in recent years. The results show that the improved GA-BP model is more accurate than the single BP model, which has practical application value. The improved GA-BP model can be used for the meteorological demand of long-term weather forecast and can provide reference for the prevention of some climatic disasters.

2. GA-BP Algorithm Implementation Process

BP (Back Propagation) neural network [11] is a common network model in the artificial neural network. Its algorithm process is as follows:

Step 1: When the sample data is transmitted forward, it first goes through the input layer, connects and weights it to the hidden layer, and then each neuron in the hidden layer summarizes all the inputs, then produces some output through the transfer function, and outputs through the next layer connect and weights it through the output layer.

Step 2: Compare the error between the expected output \( d_o(k) \) and the actual output \( y_o(k) \).

\[
\text{Error function: } e = 2 \sum_{k=1}^{n} [d_o(k) - y_o(k)]^2
\]

Step 3: In the error back transmission stage, the error between them should be sent back, and the weight between each connection layer should be adjusted in reverse order for repeated learning and training, and the cycle should continue until the output layer outputs a relatively satisfactory result.

The purelin function is used to transfer parameters in output layer of the BP model, while the tansig
function is used to transfer parameters in hidden layer. The expressions are: 
\[ \text{purelin}(x) = x, \]
\[ \text{tansig}(x) = 2 \left(1 + \exp(-2x)\right)^{-1} - 1. \]

Its mathematical model is shown in figure 1.

![Transfer function](image1.png)

**Figure 1.** Transfer function.

Genetic Algorithm (GA) is an evolutionary Algorithm developed on the basis of natural selection and the principle of gene genetics. Figure 2 is the neural network algorithm program of genetic algorithm optimization.

![GA-BP program diagram](image2.png)

**Figure 2.** GA-BP program diagram.

3. Establishment of Forecast Model

In this paper, monthly meteorological data of Rizhao, Shandong province from January 1973 to December 2018 are selected (from China Meteorological Network). Relevant contents of weather data include monthly average wind speed, mean sea level pressure, monthly precipitation, monthly precipitation days, and monthly average visibility, etc. A total of 552 sets of data were simulated by MATLAB, among which the first 456 sets of data from 1973-2010 were used for simulation training of the model, and the last 96 sets of data from 2011-2018 were used for simulation prediction.

In the construction process of the BP model, the forecast model includes five meteorological input parameters, namely, monthly average wind speed, mean sea level pressure, monthly precipitation, monthly precipitation days, and monthly average visibility, and one output parameter, namely, monthly average temperature. There are 5 nodes in the input layer and 1 node in the output layer.

The concrete steps to implement GA-BP algorithm [12] are as follows:

Step 1: Import sample data and normalize the data to build a network. In the GA-BP model, 60 iterations are performed, setting the population size as 60, crossover probability as 0.3, and mutation probability as 0.01, and the random initial population is initialized.

Step 2: The fitness value of the population is calculated to find the optimal individual. Then, through selection, crossover, and mutation operation, determine the optimal and the worst individual with the fitness value and their positions in the team and change the best individual.

Step 3: Determine if evolution is over, and if so, proceed to the next step. If not, go back to step 2 for loop iteration optimization.

Step 4: Finish the iterative optimization of genetic algorithm optimization BP network algorithm, and assign the network parameters of the optimal individual obtained by optimization to BP neural network for training.

Step 5: Set the network parameters of BP neural network, and train the network for 100 times, the learning rate was set as 0.1, the momentum coefficient as 0.8, and the network target accuracy as 0.001.
Finally, the optimal network parameters are assigned to the BP neural network for the simulation test of temperature prediction. The values of each parameter in the neural network are given (table 1).

**Table 1.** Each neural network algorithm parameter settings.

| Neural network | Data quantity | Population size | Evolution algebra | iteration times | Learning rate | Momentum coefficient | Training target error |
|----------------|---------------|-----------------|-------------------|-----------------|---------------|----------------------|----------------------|
| BP             | 552           | 60              | 60                | 100             | 0.1           | 0.8                  | 0.001                |
| GA-BP          | 552           | 60              | 60                | 100             | 0.1           | 0.8                  | 0.001                |

4. Simulation Results

4.1. Prediction of Temperature by BP Model

BP neural network takes into account many nonlinear factors that cause temperature change, and it has strong mapping ability to nonlinear fitting. When there are 5 nodes in the input layer and 1 node in the output layer, according to experience, the number of nodes in the hidden layer is set as 10. Thus, its network model structure can be obtained, as shown in figure 3 below.

**Figure 3.** Network model structure.

4.1.1. Training Neural Network. Figure 4 shows the error change curve of the network in the training process. Through analysis, when BP neural network iterations for 8 times, the output performance reaches the optimal level, and the output error is 0.012946. Moreover, when the network iterates 14 times and the iteration stops, and the BP neural network error accuracy target 0.001 during the training process, the error performance of the model network does not reach the expected error accuracy target.

**Figure 4.** BP error performance curve. **Figure 5.** BP simulation curve. **Figure 6.** BP error curve.

First of all, the simulated training experiment was conducted on the first 456 samples (air temperature of Rizhao Station, Shandong Province, 1973-2018 and the relevant data of the five influencing factors of air temperature) with the trained network. Then, this model was used to train the final 96 sample data (relevant data of air temperature influencing factors of Rizhao Station, Shandong Province, 2011-2018), and then the predicted value of the model for air temperature of Rizhao Station, Shandong Province, 2011-2018 was obtained. The predictive fitting results of the model are given in figure 5, which facilitates a more intuitive comparison between predicted and real value. Figure 6 is
error curve. According to the figure, the maximum error is 9 degrees Celsius, so the prediction error of BP neural network model is relatively large.

4.2. Prediction of Temperature by GA-BP Model

The improved BP neural network model based on a genetic algorithm overcomes the shortcomings of the traditional BP neural network that is sensitive to threshold and weight. After analyzing the error performance curve (figure 7) in the process of network training, the optimized neural network reached the best output performance with an error of 0.01215 when iterating for 3 times. When the number of iterations reaches 9, the iteration is stopped. Compared with BP, although the output error is reduced, it still fails to meet the expectation.

Based on the prediction steps of the BP model, the first 456 samples were simulated and predicted by using the trained network, and the fitting results of their real values and predicted values were obtained. The simulation prediction of the final 96 sample data was carried out with this model, and the fitting curve of the real value and the predicted value (figure 8) and the error curve of this model were obtained (figure 9). According to figure 9, the maximum error of this model is 7 degrees Celsius.

Through the above figure, the prediction results of the two network models are analyzed and compared, and the prediction accuracy of GA-BP model for temperature of Rizhao, Shandong Province is higher than that of BP model.

5. Improvement of GA-BP Model

Adding the input nodes of the network in the GA-BP neural network model is meaningful to improve the accuracy of the prediction results, so the input layer’s nodes were increased from 5 to 7, and the improved model is denoted as GA-BP1. Considering the different network output error performance corresponding to the number of nodes in different hidden layers, the error performance is measured by mean-variance MSE, and hidden layers’ nodes number are determined according to the empirical formula [13] \( n = \sqrt{n_i + n_o} + c \). Where \( n_i \) represents the number of nodes at the input layer; \( n_o \) represents the number of nodes in the output layer, and \( c \) takes any constant between 1 and 10.

MSE calculation formula is as follows:

\[
MSE = N^{-1} \sum_{i=1}^{N} (e_i)^2 = N^{-1} \sum_{i=1}^{N} (t_i - a_i)^2
\]

Where \( t_i \) is the network output value, \( a_i \) is the expected output value of the sample, and \( N \) is the number of samples.

After repeated training experiments, when the input layer and output layer’s nodes are 7 and 1, the network output error data corresponding to the number of nodes in different hidden layers is obtained (table 2). It can be known that when the hidden layer is 10, the mean-variance is the lowest, with a value of 0.0012. Therefore, the hidden layer is determined as 10 here.
Table 2. The number n of hidden layer nodes and the relationship between the MSE.

| N   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| MSE | 0.0021 | 0.0018 | 0.0016 | 0.0016 | 0.0017 | 0.0016 | 0.0012 | 0.0018 | 0.0017 |

To make the improved results more convincing, the model GA-BP2 was added, and the difference among GA-BP, GA-BP1 and GA-BP2 were given in table 3.

Table 3. Part four models of information.

| Number of nodes | Input layer | Hidden layer | Output layer | Input layer influence factor | Output layer |
|-----------------|-------------|--------------|--------------|-----------------------------|--------------|
| BP model        | 5           | 10           | 1            | Visibility, wind speed, sea level pressure, precipitation, rainfall days | Average monthly temperature |
| GA-BP model     | 5           | 10           | 1            | Visibility, wind speed, sea level pressure, precipitation, rainfall days | Average monthly temperature |
| GA-BP1 model    | 7           | 10           | 1            | Visibility, wind speed, sea level pressure, precipitation, days of rainfall, maximum sustained wind speed, minimum temperature | Average monthly temperature |
| GA-BP2 model    | 7           | 10           | 1            | Visibility, wind speed, sea level pressure, precipitation, days of rainfall, maximum sustained wind speed, dew point temperature | Average monthly temperature |

Then the network was trained, and the error performance curves of GA-BP1 and GA-BP2 models (figure 10a, figure 10b), the simulation curves of the real and predicted air temperature (figure 11a, figure 11b) and the error curves (figure 12a, figure 12b) were obtained in turn, just like the steps of BP and GA-BP models in predicting air temperature.

Figure 10a. GA-BP1 error performance curve.  
Figure 11a. GA - BP1 simulation curve.  
Figure 12a. GA - BP1 error curve.
According to the network error curve (figure 10a, figure 10b), it can be seen that the optimal output errors of GA-BP1 and GA-BP2 are 0.0013240 and 0.00054331 respectively, both of which meet the prediction error expectation, and GA-BP2 has a better error performance. According to the simulation curves of the real and predicted air temperature (figure 11a, figure 11b), it can be seen that the predicted value of the model is close to the real value. Then, according to the error curve (figure 12a, figure 12b), the difference between the predicted value of GA-BP1 and the real value is close to 4 degrees Celsius, while the maximum difference temperature between the predicted value of GA-BP2 and the real value is still less than 2.5 degrees Celsius. It is easy to know that the prediction errors are all smaller than GA-BP model. In conclusion, the GA-BP1 and GA-BP2 model are superior to GA-BP model in prediction performance.

6. Comparison and Analysis of Results
It can be seen from the simulation results of the above four models that the prediction accuracy is gradually increasing. However, in order to observe the improvement degree of each model more intuitively, according to the prediction results of each model, the fitting curves of four models are respectively made below (figure 13).

Note: The vertical axis represents the true value and the horizontal axis represents the predicted value; the corresponding model of each fitting curve is marked out in the figure.

If the training data of the model is too little or the features are too complex in the simulation training, the simulation results are prone to overfitting and underfitting phenomena. The data involved
in simulation training in this paper is sufficient, and the input layer is set with 7 dimensions, including visibility, wind speed, sea level pressure, precipitation, days of rainfall, maximum sustained wind speed, dew point temperature, etc. The model has been well simulated and implemented, so there is no overfitting or underfitting phenomenon. The following is the evaluation of the four prediction models in the paper. See table 4 for specific data.

| Model       | SSE   | RMSE  | Adj-$R^2$ | $R^2$ | MAE  |
|-------------|-------|-------|-----------|-------|------|
| BP          | 1039  | 3.3431| 0.8654    | 0.8682| 2.6901|
| GA-BP       | 1001  | 3.2815| 0.8703    | 0.8730| 2.6746|
| GA-BP1      | 184.5739| 1.4088| 0.9761    | 0.9766| 1.1396|
| GA-BP2      | 43.0338| 0.6802| 0.9944    | 0.9945| 0.6857|

By comparing the fitting curves of the four models and their relevant data, the correlation coefficients between the predicted value and the true value of BP, GA-BP, GA-BP1, and GA-BP2 models are 0.8654, 0.8703, 0.9761, and 0.9945, respectively. Therefore, in terms of temperature prediction, the prediction accuracy of GA-BP and its improved model is higher than that of the BP neural network alone. By comparing GA-BP, GA-BP1, and GA-BP2 models, the optimal number of nodes in the hidden layer can be determined, and increasing the number of nodes in the input layer can improve the error performance of network prediction, thus the prediction accuracy of the model can be improved. GA-BP1 and GA-BP2 model have the same network structure, but the GA-BP2 prediction accuracy is higher, and therefore more in accord with the requirement of people for temperature prediction accuracy, comparing the two models, that illustrate the impact of different factors on the model prediction accuracy is also different, so the need to forecast on the influence factors of the collection before the screening, thus improve the performance of BP model in prediction to the greatest extent.

7. Conclusion

Respectively in this paper using BP neural network; The GA-BP is obtained by adding genetic algorithm to BP model; The GA-BP1 is obtained by increasing the maximum sustained wind speed and minimum temperature in the GA-BP model; The GA-BP2 is obtained by increasing the maximum sustained wind speed and dew point temperature in the GA-BP model. The weather-related data of Rizhao city, Shandong Province from 1973 to 2010 were analyzed by using the above four forecasting models, and the monthly average temperature of Rizhao city from 2011 to 2018 was predicted. After comparative analysis, we can get:

1. In terms of prediction accuracy, GA-BP model is superior to single BP model.
2. By comparing the prediction models of GA-BP1 and GA-BP2, it can be seen that different influencing factors also have an impact on the accuracy of temperature prediction, and the dew point temperature has a deeper impact on the temperature. The accuracy of the model prediction will be further improved if the influence degree analysis and screening of various influencing factors of weather change can be carried out.

Acknowledgement

This project was supported by Shandong Provincial Natural Science Foundation, No. ZR2019PA007. Thanks to all those who helped me during the writing of this paper. First of all, I would like to thank my supervisor, Dr. Sun Xiaoqi, who patiently gave me professional guidance and encouragement in the process of my thesis writing. Secondly, I would like to thank my friend Liu Qian for her help and encouragement in the difficult process of my thesis. Finally, I would like to thank Dr. Roberto Xu and other editors for their direct and indirect assistance to me.
References

[1] Sun X Q 2015 Analysis on Dynamics Networks Behavior of Stochastic Neural with Time Delays [D] Ocean University of China.

[2] Sun F F, Wang M 2015 Application research of artificial neural network in weather forecast Urban Construction Theory Research 2015(22) 3633-3633

[3] Cao X F 2012 A bp neural network meteorological prediction Bulletin of Science and Technology 28 (08) 55-57.

[4] Yuan M Y, Zhou X J, Chen W 2000 Application of artificial neural network in temperature and precipitation forecast Heilongjiang Meteorology 3 17-19.

[5] Ju Q, Yu Z B, Hao Z C, Ou G X, Zhao J, Liu D D 2009 Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model Neurocomputing 72(13) 2873-2883.

[6] Liu D, Li J, Li Y H, Xiang B, Li G 2006 Application experiment of BP neural network model in long-term weather forecasts Meteorological Science and Technology 34 (03) 250-253.

[7] Ai H F 2019 Prediction of “Haze” weather based on GA-BP neural network Green Technology (22) 95-96, 99.

[8] Zhao Z 2015 Research on optimization of BP neural network haze prediction model based on genetic algorithm Technology Outlook 27 226-226, 228.

[9] Yang Y, Yang Y 2016 Prediction and analysis of air quality index in heating period based on GA-BP neural network Journal of Shaanxi University of Science and Technology 34 (04) 171-176, 186.

[10] Zhang P D 2014 Prediction model of urban ambient air quality based on BP neural network Automation Technology and Application 33 (01) 9-11, 19.

[11] Yuan B Q, Cheng G, Zheng L G 2018 Basic principle of BP neural networks Digital Communication World (08) 28-29.

[12] Zhang Z G, Yin J C, Liu C, Zhang X G 2017 SAPSO-BP network in tidal level prediction of port Water Transport Engineering (01) 34-40.

[13] Cheng H Y 2014 Application of GA-PSO-BP Neural Network in the Prediction of Atmospheric Pollutant Concentration [D] Huazhong University of Science and Technology.