Air Quality Index Prediction Based on Improved PSO-BP

Xiali Sun¹, Shixin Li¹*, Xi Chen² and Kun Wang¹

¹School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, China
²School of Electrical and Information Engineering, Tianjin University, Tianjin, China

*Corresponding author: lishixin@tute.edu.cn

Abstract. For air quality index prediction problem, this paper puts forward the optimization based on improved PSO - BP algorithm, the method using particle swarm weights and threshold of BP neural network is optimized, and the update each particle's position and speed of the weight of adaptive adjustment strategy, to balance the global optimization and local optimization ability, and it has been verified by experiment that the improved PSO - BP neural network model is compared with PSO - BP and GA - BP and BP on prediction accuracy improved.

Keywords: IPSO – BP, PSO, BP, air quality index, prediction.

1. Related theories

1.1. Particle swarm algorithm

PSO was first put forward by Kennedy and Eberhart [5] in 1995. When birds are praying, the easiest way to find food is to search the area around the bird closest to the food. The PSO initializes a group of particles in the feasible solution space, and each particle shows a potential optimal solution of the extremal optimization problem. The characteristics of the particle are represented by three indexes of position, velocity and fitness. The fitness value is derived from the fitness function, and the quality of the value represents the quality of the particle [6]. In formula [7], we can find the new update position and speed of particles:

\[
\begin{align*}
    v_{i,n}^{k+1} &= w v_{i,n}^k + c_1 r_1 (p_{i,n}^k - x_{i,n}^k) + c_2 r_2 (g_n^k - x_{i,n}^k) \\
    x_{i,n}^{k+1} &= x_{i,n}^k + v_{i,n}^k
\end{align*}
\]

Among them, \( k; i \) and \( n \) are iteration times, particle number and dimension separately; \( v_{i,n}^k \) is the velocity of the \( i \) particle on the \( n \) dimension after the \( k \) iteration; \( p_{i,n}^k \) is the optimal position of the \( i \) particle in the \( n \) dimension after the \( k \) iteration; \( g_n^k \) is the position of the global optimal in the \( n \) dimension after the \( k \) iteration; \( w \) is the inertia weight coefficient, usually between with [0.4–0.9], whose value can adjust the convergence speed of particle swarm; \( c_1, c_2 \) is the learning
factor, usually set to 2; \( r_1, r_2 \) is a random number equally distributed on \([0, 1]\). In order to prevent the blind search of particles, its position and velocity are usually limited within a certain range.

PSEO has the following characteristics: it can remember the best position in history and transfer it to other particles; it relies on the particle speed to complete the search. In the iterative process, only the optimal particle information is passed to other particles, so the search speed is fast. Few adjustable parameters and simple structure.

1.2. BP neural network

The basic principle of the BP [8] is: input a series of neurons in the input layer into the hidden layer after weighting, hidden layer neurons after summarizing all the input minus the threshold, and then produced by transfer function response output, through the next layer connection weights weighted loss to the output layer, summary output layer neurons after all inputs to produce a response output. By comparing it with the expected value, the error value is returned, and the weight and threshold of the neural network are adjusted according to the predicted error value, so the predicted output of BP approximates the expected output. We can find the BP is topological structure in Figure 1 below:

![BP neural network topology diagram](image)

Figure 1. BP neural network topology diagram.

Figure 1, \( x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_m; w_{ij}, w_{jk} \) are the input value, prediction value and the weight of the BP. The input value and predicted value of the BP are the independent variable and the dependent variable of the function respectively, which are non-linear functions. The BP should be trained before prediction. Through training, the network has associative memory and prediction functions.

2. Model building

2.1. PSO-BP

Due to the BP has the weak points of low convergence efficiency and easily trapped in local minima, compared with the genetic algorithm and particle swarm algorithm without the process of "crossing" and "variation", is simpler, and has global search optimal, bright characteristics such as high precision and fast convergence speed, to enhance the prediction accuracy, this paper will combine PSO and the BP [9] their separate advantages, establish and optimize AQI prediction model.

PSO - BP model is based on PSO to optimize the BP weights and thresholds, each particle represents the weight and threshold of BP. The optimal initial weights and thresholds of the network are found by particle optimization, and the values obtained by PSO algorithm are assigned to the neural network for training and prediction, it can make PSO - BP [10] to get the global optimal solution.
2.2. IPSO-BP
When the speed weight value \( w \) is large, the global optimization ability is strong but accuracy is low, while the local optimization ability is weak and the convergence speed is fast. When the value \( w \) is small, the local search ability is strong, but the global search ability is weak and the convergence speed is slow. Therefore, an IPSO-BP based on adaptive weight\[11\] is proposed. Adaptive weight value \( w \) is adopted for the position and velocity of each particle to balance global optimization and local optimization. The steps of this method are as follows:

1. Calculate the average fitness value of the population particles \( f_{\text{avg}} \);
2. When the fitness value \( f \) is greater than the average fitness \( f_{\text{avg}} \), the weight value \( w \) changes with \( f_{\text{avg}} \), \( f_{\text{min}} \) linearly between the maximum and minimum speed weights.

The updating formula \[12\] based on adaptive weight \( w \) is shown in Formula 3 below.

\[
w = \begin{cases} 
    w_{\text{max}} - \left( w_{\text{max}} - w_{\text{min}} \right) \frac{f - f_{\text{min}}}{f_{\text{avg}} - f_{\text{min}}} & , f \leq f_{\text{avg}} \\
    w_{\text{max}} & , f > f_{\text{avg}} 
\end{cases}
\]

(3)

Among them, \( f \), \( f_{\text{avg}} \) and \( f_{\text{min}} \) are divided into the current fitness value, average fitness value and minimum fitness value, \( w \), \( w_{\text{max}} \) and \( w_{\text{min}} \) are the current speed weight, maximum speed weight and minimum speed weight respectively.

The prediction model of AIR quality index \[13\] was built according to IPSO-BP neural network. The specific flow chart is shown in Figure 2.

![Figure 2. Prediction flow chart based on IPSO-BP neural network.](image)

In the setting of the parameters of the prediction model, for achieve rapid convergence and prevent overfitting, the learning rate of the network is set as 0.1, the error is set as 0.00001, and the training times are set as 100. The IPSO-BP is similar to PSO-BP, with three layers: input, hidden and output layer. The hidden layer adopts the method of double hidden layer and sets the number of nodes as \[31,1\]. Compared with the single hidden layer, the IPSO-BP neural network has strong generalization ability and high prediction accuracy. The average absolute error\[14\] is used to evaluate the prediction
results. MAE (Mean Absolute Error) refers to the average of absolute error, which is defined as formula 4 and can better reflect the actual situation of the predicted value error.

$$MAE = \frac{1}{N} \sum_{i=1}^{N}|f_i - y_i|$$  \hspace{1cm} (4)

The algorithms used in the full text are all implemented in the MATLABR2019b environment.

3. Experimental results and analysis

3.1. Data collection

The data used in this paper are all from the Air Quality Data Enquiry Network. The specific data of PM$_{2.5}$, PM$_{10}$, SO$_2$ and NO$_2$ (numerical unit : ) for a total of 334 days from January 1 to November 29, 2020 in tianjin are used. This data is true and reliable, and there is no invalid data. In order to obtain reasonable prediction results, the contents of PM$_{2.5}$, PM$_{10}$, SO$_2$ and NO$_2$ were studied and predicted respectively, and then the comprehensive error was considered, without considering the influence of other factors. We can find the original data in Table 1.

| data     | PM$_{2.5}$ | PM$_{10}$ | SO$_2$ | NO$_2$ |
|----------|------------|-----------|--------|--------|
| 2020/1/1 | 62         | 97        | 14     | 74     |
| 2020/1/2 | 84         | 127       | 14     | 88     |
| 2020/1/3 | 141        | 190       | 28     | 107    |
| ...      | ...        | ...       | ...    | ...    |
| 2020/11/28 | 12       | 24        | 5      | 31     |
| 2020/11/29 | 39       | 65        | 11     | 55     |

3.2. Data preprocessing

For the original monitoring data, in order to improve the processing effect of the model, a normalized processing method is adopted. The specific processing formula [15] is shown in Equation 5:

$$y' = \frac{y - y_{\text{min}}}{y - y_{\text{max}}} \times 0.75 + 0.2$$  \hspace{1cm} (5)

The normalized data interval is between [0.2,0.95]. $y'$ is the processed data, $y$ is the original data, $y_{\text{min}}$ is the minimum value in the sample data, $y_{\text{max}}$ is the maximum value in the sample. The normalized data after data processing are shown in Table 2.

| data     | PM$_{2.5}$ | PM$_{10}$ | SO$_2$ | NO$_2$ |
|----------|------------|-----------|--------|--------|
| 2020/1/1 | 0.3803     | 0.4391    | 0.5462 | 0.6875 |
| 2020/1/2 | 0.4511     | 0.5207    | 0.5462 | 0.7925 |
| 2020/1/3 | 0.6345     | 0.6918    | 0.9500 | 0.9350 |
| ...      | ...        | ...       | ...    | ...    |
| 2020/11/28 | 0.2193   | 0.2408    | 0.2865 | 0.3650 |
| 2020/11/29 | 0.3062   | 0.3522    | 0.4596 | 0.5450 |
For the 334 groups of data in Table 2, the first 317 groups and the last 17 groups are training and test samples separately.

3.3. Experimental results and analysis

In this paper, 317 groups of data collected from 334 groups were set as training samples, and 17 groups of data were set as test samples. Through the IPSO-BP prediction model, the MAE of PM2.5, PM10, SO2 and NO2 and the MAE of these four groups of overall samples were respectively obtained, and the prediction results of BP, GA-BP, PSO-BP and the IPSO-BP were compared. We can find in Table 3.

Table 3. Simulation prediction results.

| Network structure | Prediction model | PM$_{2.5}$ | PM$_{10}$ | SO$_2$ | NO$_2$ | Overall |
|-------------------|------------------|------------|----------|-------|-------|--------|
| Training sample error | BP | 0.061974 | 0.070908 | 0.071514 | 0.069453 | 0.050153 |
| Test sample error   |    | 0.075245 | 0.071102 | 0.082606 | 0.098468 | 0.32869 |
| Network structure   | GA-BP |          |          |        |       |        |
| Prediction model    | PM$_{2.5}$ | 0.061852 | 0.070898 | 0.071363 | 0.069342 | 0.042766 |
| Test sample error   |    | 0.084082 | 0.072502 | 0.083154 | 0.010071 | 0.085050 |
| Network structure   | PSO-BP |          |          |        |       |        |
| Prediction model    | PM$_{2.5}$ | 0.061415 | 0.070705 | 0.070112 | 0.06921 | 0.037693 |
| Test sample error   |    | 0.080796 | 0.072346 | 0.082807 | 0.10057 | 0.094350 |
| Network structure   | IPSO-BP |          |          |        |       |        |
| Prediction model    | PM$_{2.5}$ | 0.061016 | 0.070688 | 0.070193 | 0.067244 | 0.032661 |
| Test sample error   |    | 0.080551 | 0.071509 | 0.082672 | 0.10521 | 0.072812 |

From the above table, we can see that the MAE of training and test samples on PM2.5, PM10, SO2, and NO2 for BP, GA-BP, PSO-BP and IPSO-BP have a little different, but in the overall sample, The MAE of the test samples and training samples of the BP neural network are 0.0501 and 0.3286 separately. The MAE of the test samples and training samples of the GA-BP are 0.0428 and 0.0851 separately. The test samples and training samples of the PSO-BP are The MAE are 0.0377 and 0.0943 separately. The MAE of the test samples and training samples of the IPSO-BP are 0.0326 and 0.0728 separately.

The trend of the actual measured and predicted values of air quality index (PM$_{2.5}$, PM$_{10}$, SO$_2$, NO$_2$) in Tianjin from January 1 to November 29 on the IPSO-BP training set and test set is shown in Figure 3 and Figure 4 below.
Figure 3. Actual value and predicted value of IPSO-BP training set

Figure 4. Actual and predicted values of IPSO-BP test set

Figure 3 shows the actual and predicted values of the AQI on the IPSO-BP training set, and Figure 4 shows the actual and predicted values of the AQI on the IPSO-BP test set. It can be seen from the figure that the actual values on the test set are basically in line with the predicted values.

The experimental results show that the IPSO-BP has a smaller mean error value, which reflects the actual prediction error, and further verifies the superiority of the IPSO-BP.

4. Conclusions
So far, people pay more attention to environmental quality issues and the monitoring of important AQI. Therefore, it is important to effectively predict AQI. The prediction model based on BP[3] has achieved good results, but it has some defects and needs to be improved. In this paper, the particles search for the optimal initial value and threshold value of BP, the advantage of global searching of particle swarm to the initial value and threshold value of BP, and build an IPSO-BP forecasting model, apply it to Tianjin AQI prediction, the experimental data show that by this way the MAE is smaller, and enhances the prediction accuracy.

Acknowledgments
Tianjin enterprise science and technology special commissioner project 19JCTPJJC41500.

References
[1] Liu Hewang, Liu Chi, Zheng Shilin. Can the implementation of ambient air quality standard (2012) boost the high-quality development of Chinese enterprises? [J]. China soft science, 2020 (10): 45-55.
[2] Yang Yun, Fu Yanli. Study on prediction of PM2.5 mass concentration in air [J]. Computer simulation, 2016,33 (03): 413-418.
[3] AI Hongfu, Shi Ying. Research on haze forecast based on BP artificial neural network [J]. Computer simulation, 2015, 32 (1): 402-405.
[4] ZHOU L, SHI Y, LI Y, et al. Parameter selection, analysis and evaluation of an improved particle swarm optimizer with leadership [J]. Artificial Intelligence Review, 2010, 34(4): 343-367.
[5] Kennedy J, Eberhart R C. Particle swarm optimization [C]/ Proceedings of IEEE International Conference on Neural Networks , Perth , WA , 1995. Piscataway, NJ : IEEE Service Center , 1995 : 1942-1948.
[6] Wang Xiaochuan, Shi Feng, Yu Lei. Matlab neural network analysis of 43 cases [M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2013.
[7] KENN EDY J,EBERHART R C,SHI Y.Swarm intelligence [M].San Francisco:Morgan Kaufmann Publishers,2001:227-229.
[8] Yu X, Efe M O, Kaynak O. A general backpagation algorithm for feedforward neural network learning [J]. IEEE Transactions on Neural Networks, 2002, 13(1): 251-254.

[9] Y. Lu, L. Yuping, L. Weihong, S. Qidao, L. Yanjun and Q. Xiaoli, "Vegetable Price Prediction Based on PSO-BP Neural Network," 2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA), Nanchang, 2015, pp. 1093-1096, doi: 10.1109/ICICTA.2015.274.

[10] Shen Yigao. Application of improved PSO BP neural network algorithm in weather prediction [J]. Computer age, 2019, 000 (008): 18-20, 36

[11] Shi Y and Eberhart R C. Parameter selection in Particle Swarm Optimization[C]. Proceedings of the 7th Annual Conference on Evolutionary Programming, New York, USA, 1998, 69-73.

[12] Chen Jinhui, Chen Chen, Dong Biao. Improved particle swarm optimization algorithm based on adaptive strategy [J]. Computer simulation, 2015, 32 (03): 298-303.

[13] Wu Huijing, He Xiaohui. Prediction of air quality index based on GA-BP neural network [J]. Journal of Anhui Normal University (NATURAL SCIENCE EDITION), 2019, 42 (04): 360-365.

[14] He Minghui, Xu Yi, Wang Ran, Hu Shanzhong. Neural network optimization and application based on improved particle swarm optimization [J]. Computer engineering and applications, 2018, 54 (19): 107-113 + 128.

[15] AI Hongfu. Research on "haze" weather prediction based on GA-BP neural network [J]. Green technology, 2019 (22): 31.