Research on China's environmental governance based on GA-BP neural network and TOPSIS method

*Zherong Zhang**

Shenyang Aerospace University, Liaoning, China

ABSTRACT

Regarding the construction of a mathematical model between air quality index (AQI) and different pollutant concentrations, firstly, a genetic algorithm was used to optimize the BP neural network model with PM2.5, PM10, SO_2, CO, NO_2 and O_3 as the main air pollutants, and then the air pollution in Beijing from 2015 to 2021 was used as the validation object, and it was found that the fitted R-squared on the basis of 20% test set is greater than 0.95 or above, and finally the Spearman correlation model is used to analyze the main pollutants associated with AQI index to provide solutions for the subsequent treatment of air pollution.

Regarding the construction of a comprehensive water quality evaluation model, firstly, the weight integration method based on the moment estimation theory was used to assign subjective and objective weights to the indicators, and dissolved oxygen, temperature, turbidity, ammonia nitrogen, permanganate index and hydrogen ion concentration index were used as indicators for evaluating water quality, and then the optimal weights of each indicator were derived as 7.766%, 9.509%, 37.962%, 19.666%, 17.726% and 7.7371%, and finally the comprehensive evaluation of water quality in each city was carried out by TOPSIS method.

Regarding the construction of urban noise pollution monitoring network, firstly, a genetic algorithm based on Gaussian kernel support vector machine was used to optimize the model, and Guangzhou was used as the validation object to solve the Gaussian response surface with the building density as the noise index, and the optimal number of monitoring points of 10 was obtained by using genetic algorithm for optimization.

Keywords: Genetic algorithm; BP neural network; Spearman correlation; Theory of moment estimation; TOPSIS; Gaussian kernel support vector machine

1 INTRODUCTION

In recent years, as the economy and population continue to grow, China has faced serious environmental problems such as air and water environmental pollution and urban noise pollution. In order to better understand these problems, three questions are posed in this paper, and they are explored in depth. First, we established a mathematical model between air quality index (AQI) and different pollutant concentrations, used genetic algorithm to optimize the BP neural network model, and developed corresponding measures based on the correlation analysis of relevant pollutants to finally rank the 10 cities with the best air quality in China. Next, we established a comprehensive evaluation system of water environment quality and ranked the 10 cities with the best water environment in China by weight integration method and TOPSIS comprehensive evaluation method. Finally, we took Guangzhou as an example and used Gaussian kernel support vector machine to fit the response surface, and optimized the number and location of monitoring points by genetic algorithm to minimize the number of monitoring points to ensure the accuracy and comprehensiveness of the monitoring results. Through these studies, we hope to provide some useful information and suggestions for improving environmental problems in China.

2 RELEATED WORK

The problems that need to be addressed in this paper are as follows:

Based on the data collected by the team, establish a mathematical model between the air quality index (AIQ) and the concentration of different pollutants to better understand the air quality situation; based on the above findings, take corresponding measures to improve the air quality; list the 10 cities with the best air quality condition in the country.

Establish a comprehensive evaluation system of water environment quality to reflect the pollution level and treatment effect of water environment; list the 10 cities with the best water environment in China.

Based on the selected cities, consider the accuracy and comprehensiveness of the detection results, and optimize the detection network of urban noise pollution with the goal of minimizing the number of monitoring points.

In addition, we need to collect the data from the national statistical yearbook, such as air pollution data and water quality data of each city.

3 MODEL ESTABLISHMENT AND SOLUTION

3.1 BP modeling

Since there is a large correlation between AQI and different pollutant concentrations, and also a large nonlinearity, traditional prediction models such as multiple linear regression are not good at mining the relationship between AQI and pollutant concentrations, while machine learning BP neural network fitting models are not only good at mining the nonlinear relationship between variables but also can better handle a large number of data samples. In addition, genetic algorithm is the most effective optimization algorithm in today's intelligent optimization algorithm, and the combination of genetic algorithm and BP neural network can further optimize the BP neural network.

Therefore, in this paper, we use BP neural network model to fit AQI and pollutant concentration for prediction, and combine genetic algorithm to optimize the fitting error and construct genetic algorithm optimized BP neural network model (GA-BP neural network) Finally, for taking corresponding measures to improve air quality problem we can first calculate the correlation size of each pollutant concentration to AQI and then intervene in a targeted way. important pollutant concentration to effectively reduce the AQI value.

3.2 Genetic algorithm optimization of BP neural network model

3.2.1 Genetic Algorithm

Genetic algorithm (GA) is derived from Darwin's theory of evolution of life (reproduction, mating and mutation). In GA, the optimal solution is obtained by the reproduction and evolution of the population.

In GA, there are three types of genetic operators: selection, crossover and mutation.

Selection. Selecting certain data among a portion of regular data as the next set of data is the selection operator. Commonly used selection operators include: roulette wheel method, tournament method, etc. In this paper, we use the roulette wheel method:

$$f_i = \frac{k}{F_i} \tag{1}$$

$$p_i = \frac{f_i}{\sum_{i=1}^{N} f_i} \tag{2}$$

Where F_i denotes the fitness value of an individual i, p_i denotes the selection probability of i, k is the coefficient; N is the number of individuals in the population .

Crossover. The crossover operator simulates the genetic recombination process in order to transfer the current best genes to the next population and obtain new individuals. The specific steps of the crossover operator are as follows:

Step1: Random selection of objects;

Step2: According to the selected object length, randomly select the intersection position.

Step3: Define the crossover probability $P_c(0 < P_c \leqslant 1)$, run the crossover operator and change the genes. The intersection of the k chromosome a_k and chromosome one a_{kl} at the j position is as follows:

$$\begin{cases} a_{ki} = a_{ki}(1-b) + a_{lj}b \\ a_{li} = a_{li}(1-b) + a_{kj}b \end{cases} \tag{3}$$

Where b is a random number in the interval 0-1.

Mutation. This operator simulates the phenomenon of gene mutation in biology, and new individuals are obtained according to the probability of mutation (mutation probability). The individual that carries out the mutation is the j gene of the i individual a_{ij} and the mutation is performed as follows:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{max}) \times f(g) \\ a_{ij} + (a_{min} - a_{ij}) \times f(g) \end{cases} \tag{4}$$

where: the maximum value of gene a_{ij} is a_{max}; the minimum value of gene a_{ij} is a_{min}; $f(g) = r_2(1 - g/G_{\max})^2$; the random number is r_2; g is the current iteration number; G_{max} is the maximum evolution number; r is the random number between [0,1] [1].

BP neural networks have been quite influential; In areas such as pattern recognition and signal processing, however there is still a challenge on the way of attacking the design network, namely the determination of the structure. This paper takes the condition that the genetic algorithm can reach a specific value to find the global optimal solution, which in turn is used to optimize the connection weights and thresholds of the neural network, and then in taking the boost [2].

3.1.2 Spearman Correlation Model

The Spearman rank correlation coefficient is used to measure the correlation between two variables and is usually used to measure non-linear relationships. It is based on the ranking position of each variable in the sample rather than a specific numerical magnitude [3].

The formula for Spearman's rank correlation coefficient is expressed as follows:

$$P = 1 - \frac{6\sum {d^2}_i}{n(n^2-1)} \tag{5}$$

The Spearman's rank correlation coefficient takes values between [-1,1], where -1 indicates a perfectly negative correlation, 0 indicates no correlation, and 1 indicates a perfectly positive correlation. Unlike the Pearson correlation coefficient, the Spearman rank correlation coefficient can be used to measure the correlation between any two variables, both linear and nonlinear, and does not require the relationship between the two variables to be linear [4].

3.2 Problem 1 model solving

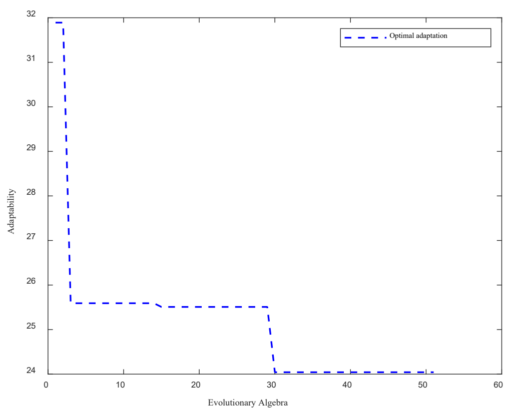
Before solving, we need to review the relevant literature to determine what are the main pollutants that affect AQI [5]. By reviewing the relevant literature, we choose PM2.5, PM10, SO2, CO, NO2 and O3 as the main air pollutants in this paper. In order to demonstrate the feasibility of our AQI model, the daily AQI values and the concentrations of each pollutant in Beijing from 2015 to 2021 were collected and obtained from the public data of the National Environment Bureau [6].

Table 1: GA-BP neural network parameter settings

Initializing the population	Number of iterations	Crossover probability	Mutation probability
10	50	0.4	0.15
Number of neural network iterations	Number of hidden layers	Learning Rate	Training target value
200	10	0.1	0.00001

Meanwhile, in order to evaluate the fitting effect of GA-BP neural network, the root mean square error is used in this paper as an evaluation of the reasonableness and accuracy of the model [7].

First, the first 80% of Beijing 2015~2021 is selected as the training set and the last 20% as the test set, and the results are solved using MATLAB software as follows:



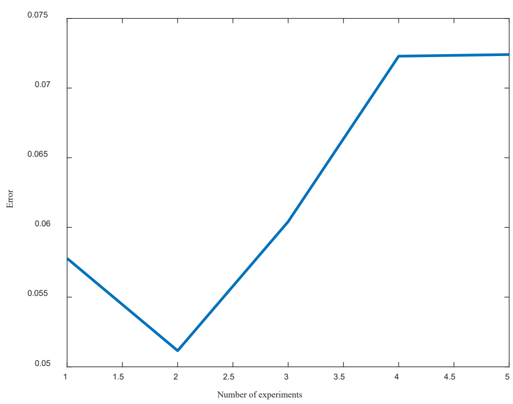


Fig. 1: GA-BP neural network error drop curve

From the left side of Figure 1, it can be seen that the fitting error of the BP neural network after combining the genetic algorithm decreases significantly and reaches the optimal value at the 30th iteration; on the right side of Figure 2, it can be seen that the relative root mean square error of the GA-BP neural network decreases significantly after 5 iterations of training [8].

Finally, we obtain the fitting results for the test set of GA-BP neural network:

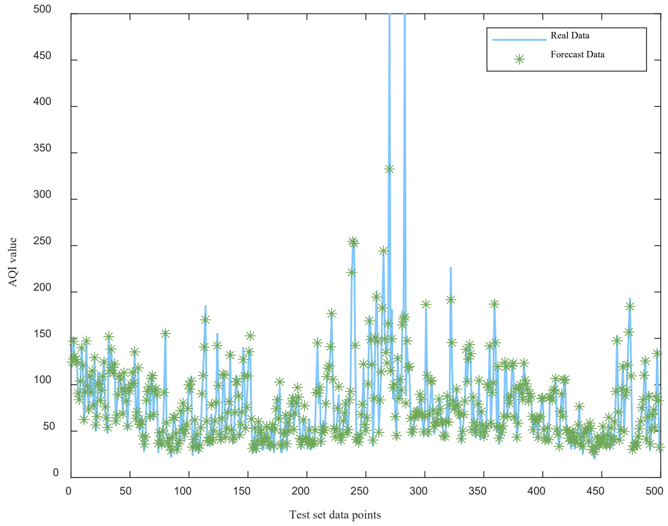


Fig. 2: GA-BP neural network test set fitting results

From Figure 2, it can be seen that the root mean square error of the GA-BP neural network test set is 13.8635, which is a small error and high model accuracy. Besides, the GA-BP neural network regression R-square is greater than 0.95 on both the training and test sets [9].

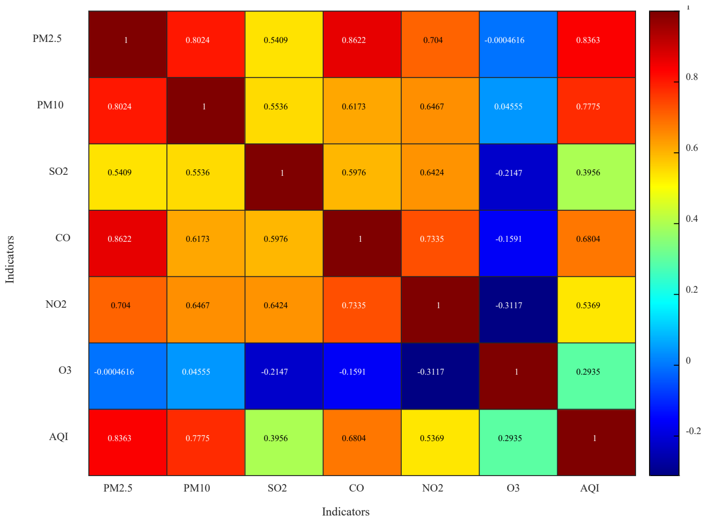


Fig. 3: Heat map of Spearman's correlation coefficient

From Figure 3, we can see that PM2.5, PM10, and CO have a strong correlation to AQI, so in the treatment of Beijing air pollution, we can first target the above pollutants with a strong correlation.

Next for air quality improvement measures, we can improve from reducing industrial production of air pollution, in order to introduce the results of the above Spielman correlation analysis, we can establish the relative degree of intervention of the relevant pollutants equation, assuming that the environmental sector requires the development of relevant industrial pollution emissions policy is expected in the future AQI need to be reduced by 10%, which corresponds down to the concentration of each pollutant to reduce emissions The equation related to the strength of the policy is then [10].

To verify the feasibility of our air quality treatment measures, we use Beijing November 2021 as the validation dataset and compare the post-intervention AQI values with the pre-intervention AQI values [11].

Table 2: Predicted AQI values for the top 10 cities with the best air quality in China

City \ day	1	2	3	4	5	6	7	8	9	10	Average value	Ranking
Yushu Prefecture	37	36	36	35	33	32	33	27	46	32	34.7	1
Chamdo	34	35	35	37	36	39	35	33	34	31	34.9	2
Linzhi City	30	31	36	37	41	38	38	36	33	30	35.0	3
Qujing City	38	28	26	38	40	38	31	34	35	43	35.1	4
Lhasa	39	42	43	39	36	39	39	40	36	39	39.2	5
Lijiang City	42	39	43	41	44	42	45	44	40	39	41.9	6
Shigatse	46	50	46	44	44	44	45	40	41	44	44.4	7
Haibei Prefecture	46	45	47	43	41	43	45	47	47	41	44.5	8
Ahri area	52	49	50	49	49	48	45	46	48	50	48.6	9
Nagchu City	47	52	47	49	46	50	59	49	64	59	52.2	10

Table 4: Genetic algorithm to solve the optimal monitoring points

	Optimal monitoring points at minimum building density		Optimal monitoring points at maximum building density	
Serial number	Dimensionality	Longitude	Dimensionality	Longitude
1	40.000	113.25	40.000	113.392
2	107.083	113.224	107.083	113.449
3	22.968	113.516	22.89	113.416
4	22.858	113.58	22.887	113.361
5	23.345	113.237	22.849	113.414
6	22.914	113.245	23.15	113.281
7	22.875	113.236	23.342	113.448
8	23.176	113.263	23.088	113.388
9	23.391	113.33	23.145	113.232
10	23.263	113.561	23.252	113.524
11	23.228	113.319	22.893	113.265
12	23.339	113.485	22.917	113.536

According to the accuracy and comprehensiveness requirements of the monitoring points established above.

4 DISCUSSION

Based on the research on environmental governance in China in this paper, the following conclusions are drawn:

The GA-BP neural network model is used to fit the air pollutant concentration to the air quality index (AQI value), and Beijing is used as an example, and it is found that the fitting effect is excellent, the fitted R-square is greater than 0.95 or more, and the root mean square error of the test set is small, which indicates the superiority of the GA-BP neural network model in studying the relationship between the air quality index and air pollutants.

5 CONCLUSION

Based on the research on environmental governance in China in this paper, the following conclusions are drawn:

The GA-BP neural network model is used to fit the air pollutant concentration to the air quality index (AQI value), and Beijing is used as an example, and it is found that the fitting effect is excellent, the fitted R-square is greater than 0.95 or more, and the root mean square error of the test set is small, which indicates the superiority of the GA-BP neural network model in studying the relationship between the air quality index and air pollutants [12].

6 DATA SOURCES

The article includes some data to support the results of this research. In order to protect the privacy and security of the Chinese NAIS database, NAIS restricted other information. This information can be obtained from NAIS for researchers who meet the criteria for accessing confidential data.

7 ACKNOWLEDGEMENTS

Thanks for the data support provided by the NAIS database and the China-PCS, as well as the project funding of Vehicle Measurement Control and Safety Key Laboratory of Sichuan Province (QCCK2021-011) and the State Administration of Market Administration Project (202248).

REFERENCES

[1] He, T. I., Tian, X. Z., Li, S. N., et al. Improved BP neural network based on swarm algorithm for wind power prediction. Journal of Electric Power Science and Technology, 2018,33(04):22-28.

[2] Wu, Kaifeng, Zhang, Lixin, Kan, Xi, Wang, Wang, Sai. A pressure sensor calibration method based on improved GA-BP neural network. Foreign Electronic Measurement Technology,2023,42(02):38-44.

[3] Yu Qun, Huo, S. D., He Jian, Li Lin, Zhang Jianxin, Feng Yuyao. Trend prediction of power grid outages in China based on Spearman correlation coefficient and system inertia. Chinese Journal of Electrical Engineering:1-12[2023-06-04].

[4] Gao Mingyu, Cai, Sun C. H., Liu C. M., Zhang Z. F., Dong Z. K., He Zhiwei, Gao Weiwei. A short-circuit fault detection algorithm based on Spearman rank correlation combined with neural networks inside a battery pack. Journal of Electronics and Information,2022,44(11):3734-3747.

[5] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11) [Online Application Software]. Retrieved from https://www.spsspro.com.

[6] Wen, Shao-Chun, Luo, Fei, Fu. A review on the application of genetic algorithms in artificial neural networks. Computational Technology and Automation, 2001(02):1-5.

[7] Gan L. X., Zhang H. Z., Lu T., et al. Water traffic safety factors based on entropy power method. China Navigation,2021,44(2):53-58.

[8] Liu Panpan, Ren G. Y., Duan Xu, L. L., Zhao L. J., Ren Xing, Miao Junwei. Evaluation of the quality and flavor of white radish by different drying methods based on the coefficient of variation method. Food and Fermentation Industry:1-11[2022-03-31]. DOI:10.13995/j.cnki.11-1802/ts.029017.

[9] Jiang S. Q., Liu S. F., Liu Z. M. A general gray number correlation decision model based on moment estimation theory with applications. Statistics and Decision Making,2019,35(24):29-33. DOI:10.13546/j.cnki.tjyjc.2019.24.006.

[10] Xue, L., Su, Chao, Cui, Chaoqun, and Sun, Yan-Ding. Application of optimal combination assignment method under moment estimation theory in mineral resource evaluation. China Mining,2017,26(04):41-46.

[11] Fu Jia-Feng, Liu Qian, Ma Zhan-Yun, Ying Na. A comprehensive evaluation of carbon peaking capacity of 30 provinces in China. Ecological Economics,2023,39(06):18-24.

[12] Yin Y, Zou permanently, Du Tali, Zhang Y. SSA-SVM based fault diagnosis of cryogenic freshwater cooling system of ships. Computer Simulation:1-7[2023-06-04].