# GA-BP Air Quality Evaluation Method Based on Fuzzy Theory

# Ma Ning<sup>1,\*</sup>, Jianhe Guan<sup>1</sup>, Pingzeng Liu<sup>2</sup>, Ziqing Zhang<sup>3</sup> and Gregory M. P. O'Hare<sup>4</sup>

Abstract: With the rapid development of China's economy, the scale of the city has been continuously expanding, industrial enterprises have been increasing, the discharge of multiple pollutants has reached the top of the world, and the environmental problems become more and more serious. The air pollution problem is particularly prominent. Air quality has become a daily concern for people. In order to control air pollution, it is necessary to grasp the air quality situation in an all-round way. It is necessary to evaluate air quality. Accurate results of air quality evaluation can help people know more about air quality. In this paper, refers to previous research results and different evaluation methods, combined with artificial neural network, fuzzy theory, genetic algorithm, GA-BP hybrid algorithm based on fuzzy theory is proposed to evaluate air quality. At the same time, for the problem that the two-grade standard of air quality annual evaluation is not suitable for practical application, the four-grade standard for annual air quality evaluation has been proposed, and its practicality has been verified through experiments. By setting contrast experiments and comparing the air quality evaluation model based on standard BP algorithm, it is proved that the fuzzy GA-BP evaluation model is better than the standard BP model, both in efficiency and accuracy.

Keywords: Air quality evaluation, fuzzy theory, genetic algorithm, BP neural network.

# **1** Introduction

Environmental quality evaluation is the basis of environmental management, while the environment is greatly influenced by external factors, and at the same time, there are many kinds of influencing factors [Onkal-Engin, Demir and Hiz (2004)]. Therefore, considering the influence of every factor on the environment and the relationship among factors, we can evaluate the environmental quality more accurately [Arhami, Kamali and Rajabi (2013)]. Air quality evaluation is an important content of environmental quality evaluation, and it is the basis of air pollution control [Yang and Wang (2017)]. Therefore, in order to manage air quality scientifically and control air pollution, we must use scientific evaluation method and reasonable evaluation standard to evaluate air quality accurately [Gurjar,

<sup>&</sup>lt;sup>1</sup> China University of Geoscience, Beijing, 100083, China.

<sup>&</sup>lt;sup>2</sup> Shandong Agricultural University, Taian, 271018, China.

<sup>&</sup>lt;sup>3</sup> Shandong Vocational Institute of Fashion Technology, Taian, 271000, China.

<sup>&</sup>lt;sup>4</sup> University College Dublin (UCD), Belfield, Dublin, 999014, Ireland.

<sup>\*</sup> Corresponding Author: Ma Ning. Email: yuetianhewu@126.com.

Butler, Lawrence et al. (2008)]. Experts from different countries use different methods and models to evaluate air quality [Chang and Hanna (2004)].

Combined with the knowledge of BP neural network, genetic algorithm and fuzzy theory, a hybrid GA-BP algorithm based on fuzzy theory is designed in this paper, which is used to establish an evaluation model for air quality evaluation. At the same time, the existing two-grade evaluation standard is not suitable for practical use. Therefore, based on the existing two-grade standard, this paper puts forward the four-grade standard of annual evaluation, and verifies its feasibility through experiments.

#### 2 Basic theory

# 2.1 Fuzzy theory

Fuzziness, refers to the objective things in real life presents a "both one and the other" characteristics. Fuzzy phenomenon is objective existence, people always understand the world from fuzzy to accurate, and then to fuzzy again, the process of constantly learning, researching and improving. Professor L. A. Zadeh, an American cybernetics expert, published a famous paper "Fuzzy Sets" in 1965 in the journal Information and Control. The publication of this paper marks the emergence of fuzzy theory. Fuzzy theory is a theory that uses the basic concept of fuzzy set or continuous membership function. Fuzzy theory is a theoretical tool to study many things with unclear boundary in life. The fuzzy theory holds that the logic of human thinking is fuzzy, and even if the conditions and data are not clear, we must make a judgment. While the modern computer is a two-level logic, either 0 or 1, which is just the opposite of the human way of thinking, too absolute, not in line with real life.

#### 2.1.1 Brief introduction of fuzzy set

The basis of fuzzy theory is fuzzy set theory. Fuzzy set is the reflection of objective existence of fuzzy concept. Fuzzy concept is the concept of unclear boundary and uncertain extension. The original classical set is replaced by fuzzy set. Fuzzy mathematics based on fuzzy set is produced by fuzziness of classical mathematics.

The fuzzy set can represent the property of "both one and the other" in real life. Let the domain be U, then the fuzzy subset A on U is a set described by its membership function  $\mu_A$ . From the above,  $\mu_A$  is actually the map of the universe U to the interval [0,1]:

$$\mu_A: U \rightarrow [0,1]$$

(1)

Among them, for any element u in the set,  $\mu(u)$  represents the degree of the element u belonging to the fuzzy subset, that is, the degree of membership of u. At this time, the degree of membership of u is expressed as A(u). It can be known from the definition that when A(u) takes the value 0 or 1, the set A is an ordinary set, representing a classical mathematical set. Therefore, the ordinary set can be regarded as a set under a special state of a fuzzy set. When solving a fuzzy problem, it is generally necessary to determine the membership function before proceeding to the next step. Using fuzzy sets can overcome the shortcomings of classical set theory.

### 2.1.2 Fuzzification and defuzzification

In order to make the computer to understand the content of fuzziness, it is necessary to solve the problem of the transformation between the fuzzy natural language of human beings and the clear value. At the same time, in order to facilitate the operation, it is also necessary to map clear values to natural languages.

The fuzzy set theory well solves the problem of conversion between clear values and fuzzy values. Therefore, it is necessary to first convert the precise parameters into fuzzy quantities, and then perform other data processing. This process is called fuzzification. Fuzzification is the method of mapping input values to fuzzy sets. It is a method of describing variables in a fuzzy language. Some variables that cannot be processed accurately will be more easily manipulated after they are fuzzied.

The process corresponding to fuzzification is defuzzification. In the system based on fuzzy theory, due to the use of fuzzy data for data analysis and statistics, the corresponding results obtained are mostly fuzzy sets or membership functions. In practical applications, people need to use precise quantities to represent data results. Therefore, it is necessary to convert the fuzzy set or membership function into a corresponding precise set or exact quantity. This process is defuzzification.

#### 2.2 BP artificial neural network theory

BP neural network, which called (Error Back Propagation Neural Network, BPNN), trained by error backpropagation, achieve the goal of minimizing the mean squared error by changing the network weights and threshold, finally can fit the data accurately. BP neural network has many advantages. The most important advantage is that it has the characteristics of self-learning and self-adaptation. Moreover, BP network is also robust and generalization, which makes it widely used in many areas.

The main working ideas of BP neural network are divided into two phases. The first is the forward propagation phase of the input sample, and the second is the backward propagation phase of the error signal.

First of all, the input sample is transmitted in the forward direction. The data enters the neural network from the network input layer. After the data is processed through the hidden layer, it is output through the output layer. Since the data flow direction is along the positive direction of the neural network, it is called positive spread.

Then, through the comparison between the actual output of the network output layer and the expected output, the mean square error of the expected output and the actual output is calculated, and the mean square error is reversed in the direction of the input layer, so that the reference errors of the neurons in each layer are obtained. This reference error is used as a credential to adjust the weight or threshold of each unit. The whole process above is a learning of a neural network. This process is continuously repeated through the neural network, and the forward propagation of the sample input and the back propagation of the error are repeated, and the error signal is constantly adjusted the weight or threshold of the hidden layer unit, until the error signal corresponding to the requirements.

Although the BP has many advantages, there are still some problems, including:

#### 218 Copyright © 2019 Tech Science Press

1) How to determine the initial value. The initial values such as neural network weights, have a direct impact on the network. However, the determination of initial values often depends on experience and there is no corresponding theoretical basis.

2) Local minimum problem. The local minimum value refers to the point where the error value reaches a local minimum when the network is training. This is mainly because the principle is the gradient descent, which leads to local minimum problems that are unavoidable for the algorithm. For some problems, it may be that the solution is a local minimum solution.

To solve these problems, we combine genetic algorithm with BP algorithm, optimize the initial weights of neural network in advance by the powerful global search ability of genetic algorithm, so as to avoid the inherent defects of BP algorithm.

## 2.3 Genetic algorithm theory

The Genetic Algorithm (GA) was proposed by Professor Holland in the United States in 1975 in the book "Adaptation in Natural and Artificial System". Genetic algorithm is a research method that artificially simulates the biological evolution process in nature. It is an artificial simulation search algorithm. Genetic algorithm is essentially a direct search method that does not rely on specific problems and gradient information. It follows the survival and elimination rule of biological evolution. It generates the following hypotheses by mutating and reconstructing the best existed hypothesis and makes it possible to solve the problem. Solution space continues to evolve so that optimal solutions can be selected in the space.

In the system of biological evolution, the chromosome is the carrier of biological genetic information, and it is composed of genes. Similarly, due to the genetic algorithm simulating the biological evolution process, in the genetic algorithm, the data form of the onedimensional data string represents the chromosome or individual gene. It represents a possible solution of a problem [He, Deng, Gao et al. (2017)]. The degree of fitness in a biological evolution system is an ability to represent an individual's ability to adapt to the environment, as well as the ability of the individual to reproduce offspring. The fitness function of a genetic algorithm, also called the evaluation function, is an index used to determine the degree of the individual's superiority and inferiority in the group. It is based on the objective function of the problem to be evaluated.

In genetic algorithms, a population is a set of possible solutions to a problem. In the genetic algorithm, generally, starting from one set of solutions of the problem, improving to another set of better solutions, and further optimizing from this set of improved solutions, iteratively proceeding until the obtained set of solutions satisfy the required conditions of the problem. In the optimization process, each set of solutions is a population, in which each solution is called the individuals, mainly by the selection operator, crossover operator and mutation operator produces the new population, the three operators are equivalent to the evolution of gene mutation and gene hybridization. The following are the key steps in the genetic algorithm:

1) Coding and decoding. Coding is one of the key steps in the design of genetic algorithms. Because genetic algorithm for parameter encoding operation, rather than the parameters themselves, and therefore need to address the relationship between the parameters and the chromosome through encoding and decoding. The mapping from the design space to the coding space is called coding, and the mapping from the coding space to the design space is called decoding. The chromosome arrangement and decoding methods of individuals are determined by the coding method.

2) Population setting. The main problem of population setting is the size of population, which, as one of the control parameters of genetic algorithm, directly affects the efficiency of the final result of genetic algorithm optimization.

3) Fitness function in genetic algorithm. In genetic algorithm, fitness function is a function to calculate individual fitness. Fitness function is a simulation of the adaptability of biological individuals to natural environment in the course of biological evolution. Individuals with higher fitness have higher survivability in nature, belong to good individuals, they are not easily eliminated. A better fitness function can guide the search of optimal solutions, and also can help to overcome the problem of premature convergence and the slowly convergence in genetic algorithms.

4) Genetic operators. From the beginning of encoding to the initial population, genetic manipulation uses random selection, crossover, and mutation operations to achieve the evolutionary process of survival of the fittest, resulting in a new generation of excellent groups that are more adaptable to the environment. There are three main genetic operators in basic genetic operations: selection, crossover, and mutation.

The genetic algorithm solution steps are as follows:

1) Randomly generate an initial population, the initial population consists of the string data of determined length;

2) Calculate the current population fitness function;

3) Through selection, mutation and crossover operations to create new populations.

4) Evaluate the individual's fitness in the new population. If the conditions are met, then the optimal individual generated during the iteration is used as the output of the genetic algorithm. The individual is an approximate solution to the problem; otherwise, the process returns to Step 2 for iteration.

#### 3 Air quality evaluation model construction

The simple BP algorithm has the disadvantages of difficult initial value setting, easy to fall into local minima and slow convergence speed. According to the characteristics of global optimization of genetic algorithm, a hybrid GA-BP algorithm is formed by combining the two algorithms to train neural network. Because air quality evaluation belongs to evaluation involving fuzzy attributes, fuzzy theory is introduced into the design process of hybrid algorithm to evaluate air quality.

The design idea of the hybrid algorithm proposed in this paper is as follows:

First of all, through the excellent global search ability of genetic algorithm, the initial value of BP neural network can be optimized, so that the initial value can be effectively selected near the global optimal value, so that the learning of the network has a better starting point. It can shorten the learning process of network and make the decision process faster and better.

Secondly, for the air quality evaluation, the fuzzy theory was introduced. After the input layer of the GA-BP algorithm, a fuzzification layer is set to fuzzify the output signal of the input neuron. Because the fuzzy signal is input into the neural network, the final output of the output layer is a fuzzy signal, and finally the defuzzification layer is added after the output layer, so that the fuzzy signal can be transferred to the accurate signal. This algorithm is called the fuzzy GA-BP algorithm. The hybrid algorithm consists of three parts: the fuzzy information processing section, the genetic algorithm processing the initial value of the network and the training part of the BP neural network.

#### 3.1 Evaluation standard setting and data preprocessing

According to the relevant laws and regulations and evaluation standards of environmental quality evaluation, combined with the historical data of air quality, urban air quality evaluation and research will be conducted [Pirovano, Colombi, Balzarini et al. (2015); Pepe, Pirovano, Lonati et al. (2016)]. At present, there are six major pollutant indicators, namely PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO; the following six indicators are selected as the evaluation factor: the annual average concentration of SO<sub>2</sub>, the annual average concentration of PM<sub>2.5</sub>, the annual average concentration of PM<sub>10</sub>, the 95th percentile of the CO 24-hour average, and the 90th percentile of the maximum 8-hour sliding average of the O<sub>3</sub>.

According to the actual situation, in order to better distinguish the degree of pollution in the target area, this paper proposes the idea of classifying air quality into four grades when conducting air quality evaluation studies. The original two grades were expanded to four grades, and grades one to four were: excellent, good, light pollution, and moderate pollution. Tab.1 is the corresponding grade of pollutant concentration limits:

Name	Time span	Concentration limit (mg/m <sup>3</sup> )					
		Grade1	Grade2	Grade3	Grade4		
$SO_2$	Annual average	0.02	0.06	0.10	0.15		
$NO_2$	Annual average	0.04	0.04	0.08	0.15		
$PM_{10}$	Annual average	0.04	0.07	0.12	0.20		
PM <sub>2.5</sub>	Annual average	0.015	0.035	0.075	0.15		
CO	24-hour average	4	4	6	10		
$O_3$	Maximum 8-hour average	0.10	0.16	0.21	0.26		

Table 1: Environmental rating grades 4

When inputting data, normalization of the data is first performed. For the data of interval [0, n], it needs to be mapped to the normal work interval [0,1] of the neural network. Data normalization can avoid the problem of large difference in data levels due to different dimensions between different attributes. In this study, the "min-max" method was used.

$$\mathbf{x}' = \frac{x - \min}{\max - \min} \tag{2}$$

In the formula, x represents the original value of the data, x' represents the value after normalization of the data, and max and min represent the maximum value and the minimum

value of the attribute. At the same time, after the data is output, the same processing is required to restore the output to the original form so that the corresponding prediction result can be obtained.

y = u(max - min) + min

Among them, *y* is the predicted value, *u* is the output of the neural network model, and max and min have the same meaning as the normalized formula.

# 3.2 Determine the network structure

From the above, we can see that the input item of the network is the concentration of six kinds of air pollutants, so there are six neurons in the input layer. In terms of network output, only one item is the air quality grade. According to the air quality four-grade standard, the output vector should correspond to the grade. Since the model is fuzzied and the output is a membership function, the output range should include four ranges: grade1, grade2, grade3 and grade4. We determined that there were four neurons in the output layer.

At the same time, due to the ability of neural network algorithms to infinitely approach nonlinear quantities, for any continuous function in closed interval can be approximated by a hidden layer of the network. So in this study, we used one hidden layer network structure. The number of neurons in the hidden layer is one of the optimization goals of the genetic algorithm, the maximum possible number of the current settings is 50. The network learning rate is set to 0.1, the momentum factor is 0.5, the maximum number of learning is 10000, target learning error is 0.001.

# 3.3 Genetic algorithm optimization neural network

After the neural network is set up, the corresponding setting of genetic algorithm should be carried out. Because the optimization of neural network by genetic algorithm is the optimization of the initial value, the number of neurons in the hidden layer, the weight and the threshold value should be set as chromosome code. The corresponding adaptive function is set up for population evolution. When the population evolution is optimal, it is the best initial value. The corresponding parameters of genetic algorithm, including population size of 40, evolution times limit 200. Fig. 1 shows the chromosome coding scheme.

the maximum number of hidden neurons N	output layer neuron thresholds	the threshold of the first hidden layer neuron and the connection weight of input and output layers	the threshold of the secon hidden layer neuron and the connection weight of input and output layers	d threshold of the N hidden layer neuron and the connection weight of input and
---	--------------------------------------	--	--	--

Figure 1: Chromosome coding scheme

The fitness function is the key of the genetic algorithm, decides whether the individual satisfies the conditions, and decides whether the evolution stops. For the BP neural network, the following fitness function is selected:

$$F(g) = \frac{1}{SSEg} = \frac{1}{\sum e^2 g}$$
(4)

(3)

Among them, SSEg is the sum of squared errors of the real output and expected output of the BP neural network when the genetic algorithm is applied to the *g*-th generation. The larger the f(g) is, the smaller the sum of squares of the representative errors is, which means that the error of the network is smaller.

The cross rate is 0.5 and the mutation rate is 0.08. These two parameters decide the probability of the cross operations and mutation operations. Tab. 2 shows the relevant parameters of the genetic algorithm and neural network:

Name	Value	Name	Value
Input neurons	6	Learning times	10000
Output neurons	4	Weight	(-1,1)
Number of hidden layers	1	Learning rate	0.1
Hidden layer neurons	5	Expectation error	0.001
Threshold	(-1,1)	Momentum factor	0.5
Population	40	Evolutional generation	200
Cross rate	0.5	Mutation rate	0.08

Table 2: Relative parameters of fuzzy GA-BP



Figure 2: Fuzzy neural network structure

The fuzzy neural network has added the fuzzification layer and the defuzzification layer. The network structure is changed from three layers of an ordinary BP neural network to five layers. Fig. 2 is the structure diagram of the fuzzy neural network. The data flow is exactly the same as the normal BP neural network, but it only adds the operation of fuzzification and defuzzification.

### 3.4 Network fuzzification and defuzzification operations

According to the four-grade air quality standard set above, data in Tab. 3 is used as an example to perform data fuzzification processing [Ishibichi, Kwon and Tanaka (2010)]. The triangular membership function is used this time. Tab. 4 shows the membership functions for the six fuzzy subsets of pollutants.

Name	$SO_2$	$NO_2$	$PM_{10}$	PM <sub>2.5</sub>	CO	O <sub>3</sub>
Concentratio	0.01	0.048	0.092	0.073	3.2	0.199
$n(mg/m^3)$						

ues

Table 4:	The men	bership	function	of six	kinds	of input	factors
$\mathbf{I}$ <b>u</b> $\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$	I HC HICH	locismo	runcuon	OI DIA	KIIGO	or mout	ractory

Name	Grade1	Grade2	Grade3	Grade4
$(A)SO_2$	(0,0.02,0.06)	(0.02,0.06,0.10)	(0.06,0.10,0.15)	(0.10,0.15,0.15)
$(B)NO_2$	(0, 0.04, 0.04)	(0.04, 0.04, 0.08)	(0.04,0.08,0.15)	(0.08,0.15,0.15)
(C)PM <sub>10</sub>	(0, 0.04, 0.07)	(0.04,0.07,0.12)	(0.07,0.12,0.20)	(0.12,0.20,0.20)
(D)PM <sub>2.5</sub>	(0,0.015,0.035)	(0.015, 0.035, 0.075)	(0.035,0.075,0.15)	(0.075, 0.15, 0.15)
(E)CO	(0,4,4)	(4,4,6)	(4,6,10)	(6,10,10)
$(F)O_3$	(0,0.10,0.16)	(0.10,0.16,0.20)	(0.16,0.20,0.25)	(0.20, 0.25, 0.25)

The input vector includes 6 attributes, and each basic attribute corresponds to 4 membership functions. Therefore, each input sample corresponds to 24 membership functions, so the fuzzification layer neurons amount to 24. In the first sample, the input vectors (A<sub>1</sub>, B<sub>1</sub>, C<sub>1</sub>, D<sub>1</sub>, E<sub>1</sub>, F<sub>1</sub>), the corresponding output is S<sub>1</sub>, the mapping relation of input and output denoted (A<sub>1</sub>, B<sub>1</sub>, C<sub>1</sub>, D<sub>1</sub>, E<sub>1</sub>, F<sub>1</sub>) $\rightarrow$ S<sub>1</sub>, the data processed by the fuzzification layer are represented by A', B', C', D', E', F', so the input membership vector and output membership vector can be expressed as follows,

 $(A_1'...A_4', B_1'...B_4', C_1'...C_4', D_1'...D_4', E_1'...E_4', F_1'...F_4') \rightarrow (S_1'...S_4')$ (5)

Through this process, we convert the ordinary data samples into usable training samples of the fuzzy neural network.

When processing network output, we need to use the defuzzification method. In this study, the membership degree limiting element mean method is used for the defuzzification. The membership value 'a' of the network output is used to segment the curve of the membership function it belongs to, and summing all the elements equal to the membership degree after the division, and then calculate the average. The resulting average value is the output value.

The fuzzy output layer is similar to the fuzzy layer. The number of neurons is equal to the sum of all fuzzy subsets of the attribute of the target vector, and the output value of the fuzzy output layer is the membership degree of the fuzzy sub sets corresponding to every attribute in the target vector. Shown as follows.

$$(A_1^{1}(x_1), A_2^{1}(x_1)...A_{k1}^{1}(x_1))...(A_1^{n}(x_n), A_2^{n}(x_n)...A_{kn}^{n}(x_n)) \rightarrow (B_1^{1}(y_1), B_2^{1}(y_1)...B_{j1}^{1}(y_1))...(B_1^{m}(y_m), B_2^{m}(y_m)...B_{jm}^{m}(y_m))$$
(6)

 $A_{k1}{}^{1}(x_1)$  represents the membership function value of the first fuzzy sets of attribute  $x_1$  in the input vector, and k1 is the number of the fuzzy sets of the attribute x1.  $B_{j1}{}^{1}(y_1)$  represents the membership value of the fuzzy set of the attribute  $y_1$  in the output vector, and j1 is also the number of the fuzzy sets of the attribute y1.

## **4** Experimental verification

#### 4.1 Fuzzy GA-BP evaluation model training results

After completing the design of the whole model structure, the training of the network can be carried out. Through MATLAB, 500 new samples were generated between each grade, a total of 2000 new samples, 80% of which were training samples, a total of 1600, 20% for test samples, a total of 400. After the fuzzy GA-BP algorithm network training process, the performance of the network is tested by the test sample. Tab. 5 shows the output statistics.

		-			
Grade	1	2	3	4	Total
Total number of samples	100	100	100	100	400
Correct sample number	96	96	98	95	385
accuracy rate%	96	96	98	95	96.25

 Table 5: Statistical results of fuzzy GA-BP model

After the statistics of the output of all the tested samples, the accuracy rate is 96.25%. It is proved that the GA-BP hybrid algorithm is feasible to evaluate the air quality and has reached a certain requirement.

#### 4.2 Standard BP evaluation model comparison experiment

After the evaluation of air quality through the fuzzy GA-BP model, a comparative test was set up to compare the results of the air quality evaluation conducted under the standard BP algorithm under the same conditions, and to compare the accuracy of the evaluation results with the evaluation efficiency. When using the BP algorithm to evaluate, the evaluation criteria adopt the same four-grade criteria, and the evaluation factor is the same as the GA-BP evaluation model. That is, 6 input vectors and 1 output vector.

Grade	1	2	3	4	Total
Total number of samples	100	100	100	100	400
Correct sample number	93	92	95	90	370
accuracy rate%	93	92	95	90	92.5

Table 6: Statistical results of standard BP model

For the hidden layer, the three-layer network structure of one single hidden layer is also selected. According to the empirical formula, the number of neurons in the hidden layer is 5. Other relevant parameters, number of learning, learning rates, expected errors, and momentum factors were consistent with the fuzzy GA-BP evaluation model, and the weight

and the threshold was taken as a random number between (-1, 1). Tab. 6 shows the actual output and expected output of sample data of the BP evaluation model.

From Tab. 6, the accuracy rate of the standard BP evaluation model is 92.5%, which also achieves the effect of air quality evaluation.

# 4.3 Analysis of results

Analyzing the evaluation results of two air quality evaluation models, we found that the accuracy of the standard BP evaluation model was slightly lower than the fuzzy GA-BP evaluation model. However, in terms of evaluating efficiency, compare the training time and the number of learning times. As shown in Fig. 3 and Fig. 4, the number of learning times and the error change curve for the two evaluations. As shown, after the GA-BP evaluation model is optimized by the GA algorithm, the initial error value of 0.006 is much lower than the initial error value of the standard BP algorithm by 0.028. Although the absolute difference in data is only 0.022, it represents a huge improvement in the network training process. This shows that the GA algorithm is very successful in optimizing the initial value of the BP network. The initial value is chosen to reduce the error.

Subsequently, comparing the training time of the two models, GA-BP took a total time of 28.4 s, while the standard BP model was 42.1 s. This shows that GA-BP also has a great advantage in training time. The optimization of the GA algorithm reduces the time for the BP neural network to adjust the weight and threshold itself and achieves the goal of improving work efficiency. At the same time, for the GA-BP algorithm, the greater the amount of data and the higher the accuracy requirement, the more obvious the advantages will be.



Figure 3: Error curve of fuzzy GA-BP model



Figure 4: Error curve of standard BP model

# **5** Conclusion

This paper combined fuzzy theory and genetic algorithm to optimize the standard BP neural network, designed a GA-BP hybrid algorithm based on fuzzy theory, and made the fuzzy hybrid algorithm for air quality evaluation GA-BP model. By evaluating the quality of air through the fuzzy GA-BP algorithm, proved the feasibility of fuzzy GA-BP algorithm. It verified that the four-grade standard for air quality evaluation proposed in the paper is feasible. By setting up comparison experiments, comparing with the standard BP algorithm evaluation model, it is proved that the fuzzy GA-BP algorithm does achieve the goal of optimization and is superior to the standard BP algorithm evaluation model in terms of efficiency and accuracy.

Because the fuzzy GA-BP hybrid algorithm is the same as the BP algorithm, it is based on a large number of training data to fit the network to judge, so the evaluation accuracy is higher than the general method. Moreover, during subsequent use, new training data can be continuously added to the network, so as to increase the generalization ability of the network and make it do not occur inaccurate evaluation as time goes by.

# References

Arhami, M.; Kamali, N.; Rajabi, M. M. (2013): Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science & Pollution Research*, vol. 20, no. 7, pp. 4777-4789.

Chang, J. C; Hanna, S. R. (2004): Air quality model performance evaluation. *Meteorology & Atmospheric Physics*, vol. 87, pp. 167-196.

**Gurjar, B. R.; Butler, T. M.; Lawrence, M. G.; Lelieveld, J.** (2008): Evaluation of emissions and air quality in megacities. *Atmospheric Environment*, vol. 42, no. 7, pp. 1593-1606.

He, P.; Deng, Z. L; Gao, C. Z; Wang, X. N; Li, J. (2017): Model approach to grammatical evolution: deep-structured analyzing of model and representation. *Soft Computing*, vol. 21, no. 18, pp. 5413-5423

Ishibichi, H.; Kwon, K.; Tanaka, H. (2010): Learning of fuzzy neural networks from fuzzy input and fuzzy targets. *Fifth International Fuzzy Systems Association World Congress*, vol. 1, pp. 218-232.

**Onkal-Engin, G.; Demir, I.; Hiz, H.** (2004): Assessment of urban air quality in Istanbul using fuzzy synthetic evaluation. *Atmospheric Environment*, vol. 38, no. 23, pp. 3809-3815.

**Pepe, N.; Pirovano, G.; Lonati, A.; Balzarini, A.; Toppetti, G. M. et al.** (2016): Development and application of a high resolution hybrid modelling system for the evaluation of urban air quality. *Atmospheric Environment*, vol. 141, pp. 297-311.

**Pirovano, G.; Colombi, C.; Balzarini, A.; Riva, G.; Gianelle, V. et al.** (2015): PM2.5 source apportionment in Lombardy (Italy): comparison of receptor and chemistry-transport modelling results. *Atmospheric Environment*, vol. 106, pp. 56-70.

**Yang, Z. S.; Wang, J.** (2017): A new air quality monitoring and early warning system: air quality assessment and air pollutant concentration prediction. *Environmental Research*, vol. 158, pp. 105-117.