$PM_{2.5}$ Forecasting Based on Artificial Neural Network and Genetic Algorithm

Hongxia Zhu, Liqiang Fan

Department of Mathematics & Information Science, Langfang Teachers University, Langfang 065000, Hebei, China

***Abstract* — In order to predict $PM_{2.5}$ concentrations value more accurately, we propose a new prediction technique based on BP artificial neural network model and the multi-population quantum genetic algorithm (this new technique is shorted as MP-QGA-BP) to improve the $PM_{2.5}$ concentrations value forecasting system. The MP-QGA-BP model is compared with the traditional BP artificial neural network model for daily maximum of $PM_{2.5}$ concentrations value forecasting. When we use the real-time monitoring time-series data sample including PM_{10}, CO, NO_2 and SO_2 pollutant data which are closely related to the value of $PM_{2.5}$ concentrations value and Mat lab language to program, the result of simulation has shown that the model established by MP-QGA-BP artificial neural network has a smaller training and predicting error and a better general ability than the model established by traditional BP artificial neural network.**

Keywords - $PM_{2.5}$, BP neural network, quantum genetic algorithm, Mat lab language, training error, predicting error

I. INTRODUCTION

The air quality has always been a major issue which relates to the future and destiny of the mankind. with the rapid development of industrialization and urbanization, in recent years, the fog and haze weather gradually draws people's attention and it has become a hotspot for research. The inhalable particle $PM_{2.5}$ has become the main cause of fog and haze in most major cities in China, and it not only damages the ecological environment and pollutes the environment. More importantly, it is a great threat to human health[1-2]. Therefore, it is very important to establish a scientific and effective $PM_{2.5}$ concentrations value forecasting model.

There are a large number of studies [2-7] on the generating, spread, forecast and other aspects of $PM_{2.5}$. At present, the artificial neural network method [4-7] is commonly used for $PM_{2.5}$ forecast. Since the artificial neural network is a human-like neural network which has the ability of self-learning and self-organizing and self-adjusting according to different internal and external drives. So compare with the classical linear regression model, it can better adapt to the $PM_{2.5}$ forecast with multiple factors, full of randomness, uncertainty and nonlinearity. The application of traditional BP neural network [7] is more typical. But due to the shortcomings that BP neural network is easy to fall into local minimum point and its network structure is not easy to determine, this model has certain limitations in the aspect of accuracy when forecasting $PM_{2.5}$.

For this reason, we propose a new BP artificial neural network forecast model based on multi-population quantum genetic algorithm. We test the performance of the two forecast methods: BP artificial neural networks optimization model and traditional BP artificial neural network. The value of $PM_{2.5}$ concentration is forecasted by the new model and the traditional BP artificial neural network model and we compare and analyze simulation results with test.

II. PRINCIPLES

A. *BP artificial neural network model*

BP artificial neural network consists of three layers: an in-put layer, a number of hidden layers and an output layer. The network structure is shown in Fig. 1. Each layer is composed of a number of independent neurons. The learning process of BP neural network consists of two processes: forward propagation of information and back propagation of error [7].

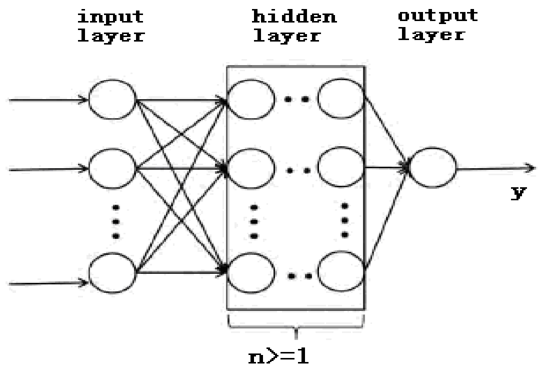

Figure 1. BP Structure Diagram

In the forward propagation process, the input information is calculated from the input layer through the hidden layers to the output layer. And the neuron state of each layer only affects that of the next layer. If the desired output does not appear in the output layer, the error change value of output layer is calculated and will turn to back propagation. Through the network, the error signals are back transmitted along the original connection path to modify the

DOI 10.5013/IJSSST.a.16.02.10 ISSN: 1473-804x online, 1473-8031 print

Weight and threshold value of neurons at each layer till the desired goal is reached [8]. Repeat training the network, until it meets the desired error.

B. MP-QGA-BP artificial neural network model

1) Quantum genetic algorithm(QGA)

Genetic algorithm (GA) is an exchange mechanism that simulates the survival of the fittest of biological evolution and the chromosomes to find the best individuals through selection, crossover and mutation. The basic process is to encode the data of questions' solution space to form a gene coding sequence, and then to do the operations of selection, crossover and mutation for the coding sequence so as to continuously generate the new individuals for optimal selection and ultimately get the most optimal solution[9].

Quantum genetic algorithm is essentially a genetic algorithm. Therefore, traditional genetic algorithm can be applied in the fields which are also applicable for the quantum genetic algorithm. Since the introduction of quantum computing, its results are obviously better than traditional evolutionary algorithm. The process of quantum genetic algorithm is as follows[10-11]:

Step 1: Initialize the population $Q(t_0)$, to randomly generate N chromosomes coded by quantum bits;
Step 2: Measure each individual of the initial population $Q(t_0)$ to obtain the corresponding certain solution $P(t_0)$;
Step 3: Evaluate the fitness of each certain solution;
Step 4: Record the best individual and its fitness;
Step 5: Judge whether the computational process can be completed, if the computational process meets termination condition then exit, or continue to calculate;
Step 6: Measure each individual of population $Q(t)$ to obtain the corresponding certain solution;
Step 7: Evaluate the fitness of each certain solution;
Step 8: Use quantum rotation gate $U(t)$ to adjust individual and obtain a new population $Q(t+1)$;
Step 9: Record the best individual and the corresponding fitness;
Step 10: Plus Step 1 with the iterations t and then return to Step 5.

2) Multi-population quantum genetic algorithm (MP-QGA)

The basic idea of MP-QGA is based on the following concepts:

- Use a number of populations for optimization calculation using QGA separately;
- The bridge between populations is immigration operator through which all the populations can evolve together. So the best individual is the coevolution result of all populations.
- Use artificial selection operator to store best genes of every evolution generation in each population.

The process of MP-QGA is shown as Fig. 2:

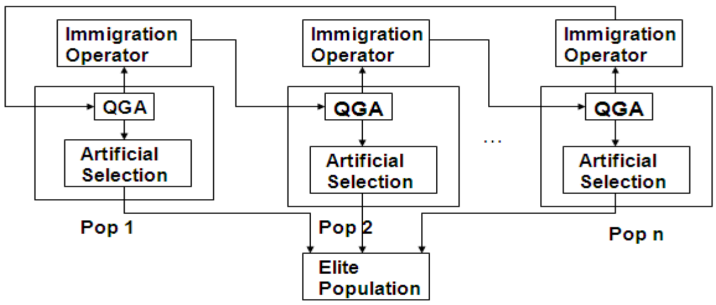

Figure 2. Schematic Diagram of MP-QGA Algorithm Structure

C. Combined MP-QGA algorithm: a new optimizing BP artificial neural network scheme

Because the choice of the initial connection weights and threshold value of the traditional BP artificial neural network has a great influence on the training of the network, but it can not be obtained accurately. For this reason, we propose a simple and effective method that uses MP-QGA algorithm to optimize its initial connection weights and threshold value that makes the BP artificial neural network forecast variance to achieve the minimum.

The optimization process is divided into two main steps: the first step is to confirm the coding scheme of BP network weights and threshold; and the second step is to use Multi-population Quantum Genetic Algorithm to complete the optimization. For certain BP network structure, the process of optimizing network initial weight and threshold has eight steps as follows:

Step 1: Confirm the encoding mode of weights and thresholds (the binary encoding mode is used in this paper) to obtain several initial populations;
Step 2: Decode the genes in all populations to obtain corresponding weights and thresholds, so as to construct the corresponding BP network with different initial weights and thresholds;
Step 3: Calculate the fitness value of each gene based on the performance evaluation criteria (such as forecast variance, etc.);
Step 4: Confirm the corresponding quantum rotation gate according to fitness values;

Step 5: Update all populations to obtain several next generation populations according to the new quantum rotation gate;
Step 6: Immigration operation is done among the new populations through which the worst genes of target population are replaced by the optimal genes of source population;
Step 7: Use artificial selection operator to obtain the best gene of each population and save it in the elite group;
Step 8: Return to the second Step 2 and finish the optimization process until it reaches the optimal gene with the least number of generations.

D. $PM_{2.5}$ forecast based on MP-QGA-BP artificial neural network

1) Data acquisition and pre-processing

The air quality data used in this study (including NO_2, SO_2,CO, $PM_{2.5}$ and PM_{10}) are from large weather forecast website " Weather Report Website". Part of the data obtained is shown in TABLE I below. Since these variables have different dimensions and magnitudes, before inputting the data into the artificial neural network, they need to be normalized to reduce the influences on model performance that are caused by large change of variables (not belong to the same magnitudes). The processing formula is as follows:

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \tag{1}$$

y' represents the normalized data, y represents the true data, $y_{\min}, y_{\max}$ respectively represent the minimum value and the maximum value. The normalized results are shown in TABLE II.

TABLE I. THE MONITORING DATA FROM NOVEMBER 1 TO 15, 2014 (UNIT: $\mu g/m^3$ (THE UNIT OF CO IS mg/m^3))

Date	$PM_{2.5}$	PM_{10}	CO	NO_2	SO_2
2014-11-1	102	151	1.1	50	12
2014-11-2	12	52	0.36	24	7
2014-11-3	43	79	0.8	47	15
2014-11-4	132	195	1.98	81	57
2014-11-5	114	169	1.37	61	23
2014-11-6	11	37	0.32	14	9
2014-11-7	45	80	1.27	44	27
2014-11-8	66	102	1.16	51	14
2014-11-9	89	135	1.22	54	14
2014-11-10	142	205	1.85	76	25
2014-11-11	59	123	0.92	41	12
2014-11-12	7	75	0.28	13	8
2014-11-13	31	65	0.79	49	17
2014-11-14	58	92	1.2	74	37
2014-11-15	170	239	2.53	105	102

TABLE II. THE NORMALIZED DATA

Date	$PM_{2.5}$	PM_{10}	CO	NO_2	SO_2
2014-11-1	0.52	0.32	0.17	0.22	0.05
2014-11-2	0.03	0.07	0.02	0.07	0.00
2014-11-3	0.20	0.14	0.10	0.21	0.08
2014-11-4	0.69	0.43	0.34	0.41	0.49
2014-11-5	0.59	0.36	0.22	0.29	0.16
2014-11-6	0.02	0.03	0.01	0.01	0.02
2014-11-7	0.21	0.14	0.20	0.19	0.20
2014-11-8	0.32	0.20	0.18	0.23	0.07
2014-11-9	0.45	0.28	0.19	0.25	0.07
2014-11-10	0.74	0.45	0.32	0.38	0.18
2014-11-11	0.29	0.25	0.13	0.17	0.05
2014-11-12	0.00	0.13	0.00	0.00	0.01
2014-11-13	0.13	0.10	0.10	0.22	0.10
2014-11-14	0.28	0.17	0.19	0.37	0.29
2014-11-15	0.90	0.54	0.45	0.56	0.93

2) The optimized BP artificial neural network topological structure and parameter settings based on MP-QGA

The three-layer ($s-t-1$) BP artificial neural network model is created with the number of neurons in the hidden layer of t . According to the characteristics of the issues involved in this research, the number of neurons in the network input layer s is equal to 4, and the number in the hidden layer t is equal to 5, the forecast model structure of MP-QGA-BP is shown in Fig. 3. The gene number of various groups N is equal to 40; the least number of generations of the optimal gene is defined as MP which is equal to 10; the binary code length is equal to 10, the search space number (the number of parameters to be optimized) $T = t \times s + t \times 1 + s + t + 1 = 35$; the generating of initial individuals in all kinds of groups is this: $N \times T$ initial values within the specified range are generated randomly, the target error of BP network $GOAL = 0.0001$; the training maximum number $EPOCHS = 1500$; the training sample number m_1 is equal to 25 , the number of test samples m_2 is equal to 5. Network training error E_1 and test error E_2 are respectively the formula (2) and the formula (3):

$$E_1 = \left(\sum_{k=1}^{m_1} \left(Y_k - T_k \right)^2 \right)^{\frac{1}{2}} \tag{2}$$

$$E_2 = \left(\sum_{k=1}^{m_2} \left(Y\_T_k - T\_T_k \right)^2 \right)^{\frac{1}{2}} \tag{3}$$

Wherein, m_1 is the training sample size, m_2 is the test sample size, Y_k and Y_T_k respectively are expected outputs of training samples and forecast samples, T_k and T_T_k respectively are network simulation outputs of training samples and forecast samples.

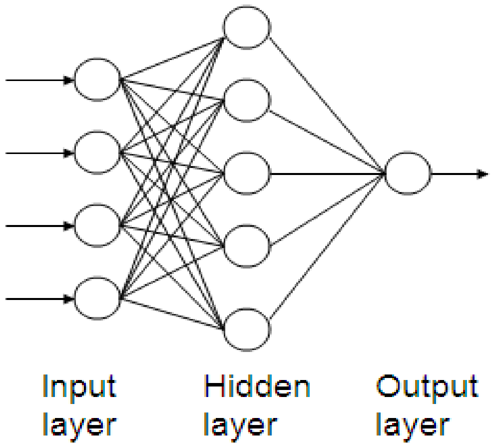

Figure 3. Structure Chart of BP neural Network

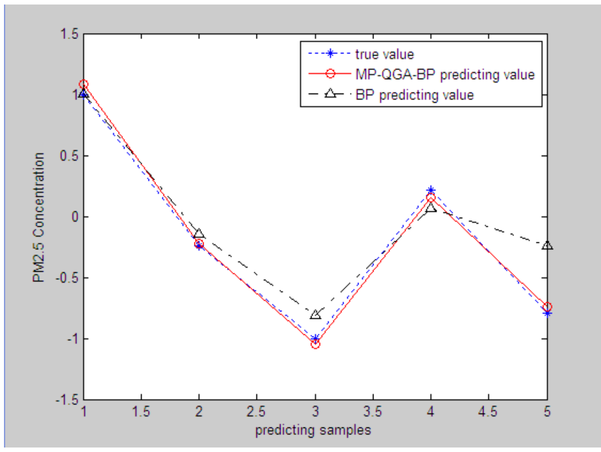

Figure 4. Contrast curve of prediction value of $PM_{2.5}$ and true value of $PM_{2.5}$

3) $PM_{2.5}$ concentration forecast process of MP-QGA-BP artificial neural network

The $PM_{2.5}$ concentration forecast model based on MP-QGA-BP artificial neural network fully combines with good nonlinear processing ability of the BP neural network and good optimization ability of MP-QGA so as to establish a multivariate nonlinear prediction model which can predict the $PM_{2.5}$ concentration more accurately. And the specific process is as follows:

Step 1: Collect the real-time data of PM2.5 concentration and other parameters;
Step 2: Use the formula (1) to normalize data, and divide the normalized data into training and testing samples of two parts;
Step 3: Use MP-QGA to optimize the BP neural network to find the best initial weights and thresholds;
Step 4: Use the training samples to train BP neural network according to the best parameters of Step 3;
Step 5: Use the trained network model to predict the concentration of PM2.5.

III. RESULTS AND DISCUSSION

1) Simulation results

The results of 30 groups of data in Fig. 1 are shown in Fig.2 after normalization. Take the first 25 groups of data in Fig. 2 as the training samples and take the remaining five groups of data as the forecast samples. Input the value to be forecast into MP-QGA-BP artificial neural network and compare the training error, forecast error and forecast results with that of the traditional BP network. The results are shown in Fig. 4 (draw traditional BP predicting value, MP-QGA-BP predicting value and actual value on the same chart) and TABLE III.

TABLE III. TRAINING ERROR AND THE PREDICTION ERROR

Type / Error	MP-QGA-BP	BP
Prediction error	0.15328	0.50871
Training error	0.26057	0.37433

2) Simulation results discussion

As can be seen from Table III, by contrast, the values of the training error or prediction error of MP-QGA-BP artificial neural network is significantly less than the value of traditional BP neural network. And the prediction curve in Fig. 4 also shows that prediction effect of MP-QGA-BP artificial neural network is better than that of the traditional BP artificial neural network. So compared with traditional BP artificial neural network, the optimized BP artificial neural network has better stability and prediction accuracy, which has some guiding significance for haze weather forecast.

IV. CONCLUSIONS

The BP artificial neural network learning algorithm based on MP-QGA has been put forward in this paper in order to solve the defects of the traditional BP artificial neural network learning algorithm which is easy to fall into the local minimum and the premature convergence of the genetic algorithm. This algorithm can give full play of the advantages of the multi-population quantum genetic algorithm, and also can effectively avoid the local minimum. The simulation studies have shown that this algorithm can effectively solve the local minimum problems of traditional BP artificial neural network training and the prediction error has been greatly improved.

ACKNOWLEDGMENT

This work was supported by Langfang Science and Technology Research and Development Self Financing Project (2015011063) and Hebei Province Department of Education Fund(QN2016114).

REFERENCES

[1] Dexiang Wang,Jianlin Hu,Yong Xu,Di Lv, Xiaoyang Xie, Michael Kleeman, Jia Xing, Hongliang Zhang,Qi Ying, "Source contributions to primary and secondary inorganic particulate matter during a severe wintertime PM2.5 pollution episode in Xi'an, China",Atmospheric Environment, vol. 97, no.11, pp.182-194, 2014.

[2] F. Yang; J. Tan, Q. Zhao,et al., "Characteristics of PM2.5 speciation in representative megacities and across China", Atmospheric Chemistry and Physics Discussions, vol.11, no. 1, pp.1025-1051, 2011.

[3] W. Wang; Y. Guo, "Air pollution PM2.5 data analysis in Los Angeles long beach with seasonal ARIMA model", Proceedings of International Conference on Energy and Environment Technology, pp.7-10,2009.

[4] Mckendry I G, "Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting", Journal of the Air & Waste Management Association vol. 52, no.9,pp.1096-1101,2002.

[5] L.Yao,N. Lu,S.Jiang, "Artificial neural network (ANN) for multi-source PM2.5 estimation using surface MODIS and meteorological data", Proceedings of International Conference on Biomedical Engineering and Biotechnology, pp. 1228-1231, 2012.

[6] Zheng H M, Shang X X, "Study on prediction of atmospheric PM2. 5 based on RBF neural network", Proceedings of Fourth International Conference on Digital Manufacturing and Automation, pp.1287-1289, 2013.

[7] Z. Jiang, X. Meng, C.Yang, et al., "A BP neural network prediction model of the urban air quality based on rough set", Proceedings of Fourth International Conference on Natural Computation, pp.362-370, 2008.

[8] J. L. Chu, H. Y. Li, X. J. Chen, "Research on Improved BP Learning Algorithm of BP Neural Network", Advanced Materials Research, vol. 765-767, pp. 1644-1647, 2013 .

[9] A.Brindle, "Genetic algorithms for function optimization", [Ph.D Dissertation], University of Alberta, 1981.

[10] Kedar Nath Das, Rajashree Mishra, "Chemo-inspired genetic algorithm for function optimization", Applied Mathematics and Computation,vol.220, pp.394-404, 2013.

[11] A.Narayanan,M. Moore, "Quantum-inspired genetic algorithms", Proceedings of IEEE InternationalConference on Evolutionary Computation, pp. 61-66,1996.