Advances in Engineering Research, volume 140
2nd International Conference on Electrical, Control and Automation Engineering (ECAE 2017)

Multi-Source Information Fusion Based on Neural Networks in Air Quality Forecasting

Xiaoqiang Zhao[1, 2], Yubing Chen[1, 2,*], Qiang Gao[1] and Dan Deng[1]
[1]Xi'an University of Posts and Telecommunications, School of Communications and Information engineering, Chang'an West Street Chang'an District, Xi'an, China, 710121
[2]Shaanxi Key Laboratory of Information Communication Network and Security, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi, China, 710121
[*]Corresponding author

***Abstract*—To forecast the air quality accurately, the model of air quality using multi-source information fusion technology based on neural network is proposed. The back propagation (BP) neural network models with time-series and no time-series training samples, the nonlinear auto-regressive (NARX) neural network with time-series training sample are respectively established on the MATLAB platform. The daily data of NO_2, O_3, PM_{10} and AQI are predicted using the models respectively. The conclusions are as follows: the three models with reliability, high prediction accuracy for air quality forecasting are successfully established. The accuracy of NARX with dynamic feedback capability is higher than BP neural network, while the BP neural network of larger non time-series training sample is of higher prediction accuracy.**

Keywords—multi-source information fusion; air quality forecasting; time series; BP neural network; NARX neural network

I. INTRODUCTION

With the rapid development of industrialization and urbanization, a large amount of industrial waste gas, automobile exhaust emissions and other environmental factors have led to the increasing decline of air quality, which seriously threaten people's health. Air quality forecast is the basis of air quality assessment, management, and decision-making, it also can provide reference and basis for air pollution prevention and control works validly. Therefore, it is very necessary to effectively forecast the air quality and improve the prediction accuracy [1].

Multi-source of meteorological information make the air quality shows complex dynamic nonlinear characteristics [2], plus with the traditional predictive models of nonlinear system such as bilinear model, the threshold autoregressive model, and the ARCH model and so on [3], above all need assumptions of sequence relation. Hence, there exist great difficulties in the aspect of theoretical research and practical application. In order to overcome the deficiency of traditional nonlinear prediction in practical application, taking the complexity and nonlinear dynamic characteristics of air quality into full consideration, the non-linear combination forecasting model of air quality by using multi-source information fusion technology upon neural network is proposed.

Based on multi-source information fusion, target recognition comprehensively utilizes the performance advantages of all kinds of sensors. It can improve the stability and reliability of automatic target recognition system and strengthen the anti-interference ability and adaptability of the system, and it has become one of the main research directions in the field of automatic target recognition; In addition, features of neural network such as distributed parallel processing, non-linear mapping, adaptive learning, robustness, fault tolerance and strong generalization ability make it to be a favorable tool to realize the multi-source information fusion technology[4,5].

In view of different training samples, three forecasting models are established by making use of BP neural network and NARX neural network, which achieve the nonlinear regression of the air quality on the MATLAB platform.

II. THEORETICAL BASIS

A. The Theory and Model of BP Neural Network

BP neural network is a kind of static neural network with non-feedback and non-memory. Three-layer BP neural network is presented, which is composed of input, hidden and output layers. The network structure diagram as shown in FIGURE I.

In FIGURE I, W_{ij} and W_{jk} is the weights between node i and j, node j and k respectively. The learning process is shown in figure 2.

In FIGURE II, the $x_1, x_2 ... x_n$ are input samples, $W_1, W_2 ... W_n$ are weights coefficients. By means of adjustment of weight coefficients, input signals generate output results in the U. The error signal e is the result of the comparison between the expected output signals and U. By changing the weights among neurons through the learning rule of Widrow–Hoff constantly to reduce the reverse error e, till the error e reach to an expected state [6].

The whole working process is divided into two processes: one is work signals transfer forward; the other is error signal back propagation. The process of work signal transfer forward can be described as the follow equations:

$$S_j = \sum_{i=0}^{m-1} w_{ij} + b_j \tag{1}$$

$$x_j = f(S_j) \tag{2}$$

In the process of back propagation, the mean square error is described as equation:

$$E(w,b) = \frac{1}{2}\sum_{j=0}^{n-1}(d_j - y_j)^2 \tag{3}$$

The output and input can be described as equation:

$$y(t) = f(x(t)) \tag{4}$$

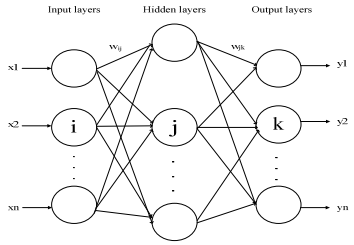

FIGURE I. THE BP NEURAL NETWORK OF THREE LAYERS STRUCTURE DIAGRAM

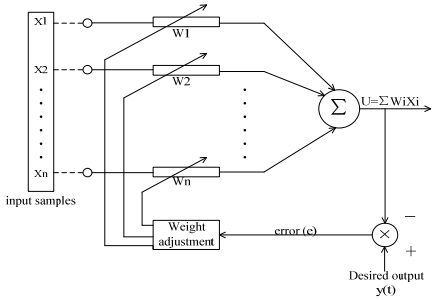

FIGURE II. THE LEARNING PROCESS OF BP NEURAL NETWORK

In the equation (4), $x(t)$, $f(t)$ and $y(t)$ are the meteorological monitoring data at a certain moment, the mapping relation is determined by the neural network and the output which is pollutant data respectively.

B. The Theory and Model of Time-Series Dynamic Neural Network

Nonlinear auto-regressive models(NARX)is a typical regression neural network, which consists of static neurons and feedback of network outputs[7], the relation between output and input can be described as equation:

$$y(t) = f(y(t-1), y(t-2) \cdots y(t-n), x(t-1), x(t-2) \cdots x(t-n)) \tag{5}$$

In this equation, x (t), y(t)and n are meteorological monitoring data at a certain moment, the mapping relation is determined by the neural network between input and output, the delay order numbers, respectively .The model structure is shown in FIGURE III.

In FIGURE III, d, W, and b are delay order numbers, the weight, and the bias respectively; f1 and f2 are activation function of hidden layer and output layer, the tansig and purelin function are used respectively.

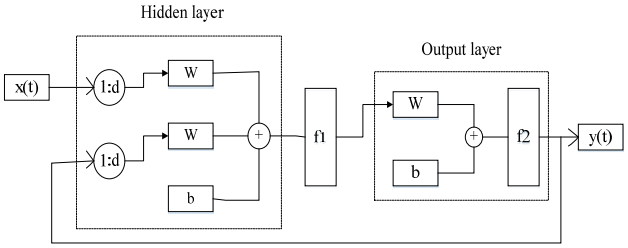

FIGURE III. NEURAL NETWORK STRUCTURE OF NARX

Due to the expected output is known in the process of NARX training, the Series-Parallel (open loop) model on the basis of the Parallel (closed-loop) model can be established. The two model structures are shown in FIGURE IV.

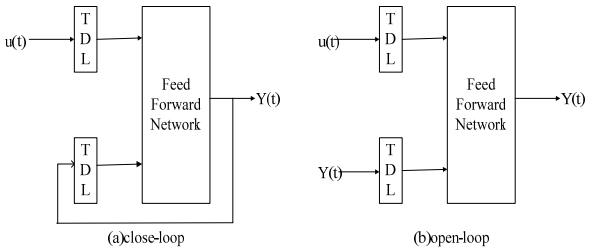

FIGURE IV. TWO KINDS OF MODEL STRUCTURES OF NARX

In this paper, the open- loop mode of recursive gradually multi-step prediction is adopted to change the NARX dynamic neural network into a simple forward neural network so that to build the model by static BP neural network directly. After output delay, the output vector is introduced through external feedback into the input vector [8].

III. THE RESEARCH METHODS

A. Data Sources and Sample Selection

The daily meteorological and pollutant data from 2013 to 2014 of the region of Wuhan city are selected as the training data set to predict the site 3 pollutants index during July 2014. The input training set is composed of nine groups of daily meteorological data, they are the temperature, humidity, air pressure, wind direction, wind speed, the maximum and minimum temperature, the maximum and minimum wind speed at a certain moment of the city, and the output of pollutants index consist of NO_2 and O_3, PM_{10} and AQI value.

In order to eliminate the difference among orders of magnitude and unify the data under the same dimension, the data should be converted into [0, 1] by using normalized processing [9] before the sample data are used. It can not only speed up the convergence of network learning and training process, but also have higher prediction accuracy. The normalization function of map minmax in MATLAB is used to cope with the training and prediction data.

B. Establish the Model of BP Neural Network

1) *Select the transfer function*

The output can be calculated via neurons by use of transfer function of the hidden and output layers. The difference performance of the prediction error of different transfer functions is shown in table I.

TABLE I. COMPARISON TABLE OF DIFFERENT TRANSFER FUNCTION PREDICTION ERROR

Hidden layer transfer function	Output layer transfer function	percentage error	MSE
logsig	tansig	40.63%	0.9025
logsig	purelin	0.08%	0.0001
logsig	logsig	352.65%	181.25
tansig	purelin	1.70%	0.0107
purelin	tansig	120.08%	113.02

From table I, logsig and purelin are selected to be transfer function of hidden layer and output layer respectively.

2) *Select the training function*

According to the error of the table I, training function modifies the weights and threshold continuously, four training functions of trainbr (standardization Bayesian back propagation algorithm), trainscg (BP training function of quantitative the connection gradient), trainrp (Bounce back propagation algorithm) and trainlm (Levenberg-Marquardt) are selected, the test results are as shown in FIGURE V.

FIGURE V illustrates that the training functions of traingdm and trainscg are not yet reached to the target error after 1000 iterations, the MSE of traingdm and trainscg are equal to 0.0675and 0.0368 respectively. The training function of trainlm and trainbr reach to the target error after 248 iterations and 172 iterations, the MSE are equal to 0.0282 and 0.0363 respectively. Thus the best training function is trainlm.

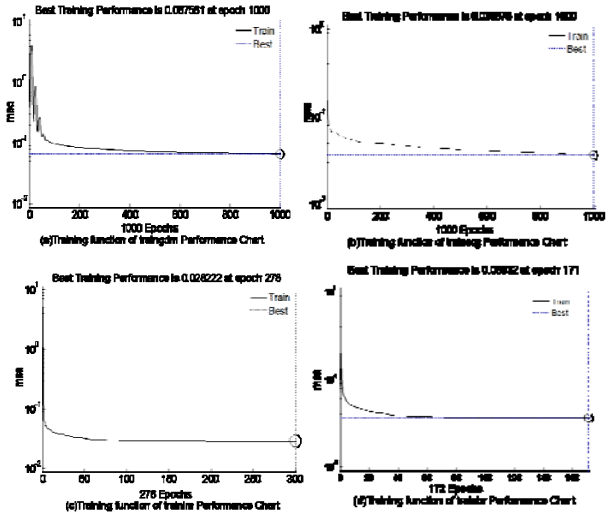

FIGURE V. PERFORMANCE THE TRAINING FUNCTIONS

3) *Select the number of hidden layer neurons*

The number of input and output layer neurons is respectively decided by the dimension of input and output of training set in BP neural network. It is an important step to select the number of hidden layer neurons. Excessive number of neurons will add unnecessary weight correction process and too less will weaken the generalization ability. Therefore, in the actual selection of hidden layer neurons, the first thing is to calculate the range of neurons according to the empirical formula and then to choose the best number of hidden layer by doing multiple experiments. Set the number of input layer, hidden layer and output layer as m, h, n respectively, and the empirical formulas are as follows:

$$h = \sqrt{m+n} + a \ , \ (1 < a < 10) \tag{6}$$

$$h = \log_2^m \tag{7}$$

The number of neurons in the hidden layer is determined by 6~14.

C. Establish the Model of Time-Series Dynamic Neural Network

NARX network is composed of static neurons and output feedback. Take NO_2 as an example. The number of hidden layer neurons of the prediction model of NO_2 is selected as 10, delay orders is selected as 3, the auto-correlation error and input cross-correlation error of NO_2 is shown in FIGURE VI.

Through the analysis of FIGURE VI, the expected error of the initial value of auto-correlation and input-error cross-correlation value is higher, the rest of values are all near or within the limit, which meet the requirement.

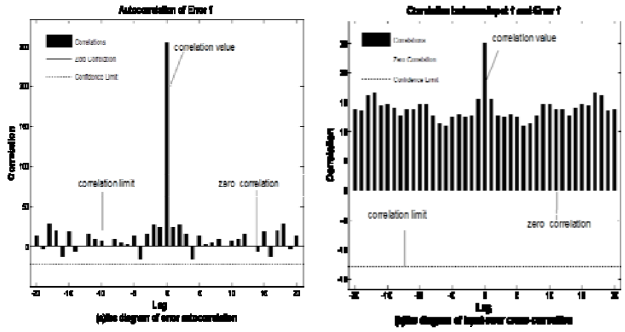

FIGURE VI. RELATED ERROR FIGURE OF NO2

D. Simulate the Network

1) Simulate the BP network and time-series dynamic neural network

The training set is composed of 1000 sets of no time- series meteorological data and its corresponding pollutants data. Take the daily meteorological data from June 30, 2014 to July 30, 2014 as the predictive input to predict the values of NO_2, O_3, PM_{10}, and AQI during the whole month. According to the experiment, the best hidden layer neuron number of each forecast model is identified as 10. This prediction model is recorded as model 1, and the predicted and actual values are shown in FIGURE VII.

The training set is composed of 540 sets of meteorological time-series data and its corresponding pollutants data. Take the

daily meteorological data from June 30, 2014 to July 30, 2014 as the predictive input to predict the daily values of NO_2, O_3, PM_{10}, and AQI during the whole month. According to the experiment, the best hidden layer neuron number of NO_2, O_3, AQI and PM_{10} forecast model is identified as 10,10,10,12 respectively. This prediction model is recorded as model 2, and the predicted and actual values are shown in FIGURE VIII.

Establish prediction model of time-series dynamic neural network using the same data of model 2. According to the experiment, the delay time d and the best hidden layer neuron number of NO_2, O_3, AQI and PM_{10} of the model is identified as 3, 10, 10, 10, 9 respectively. And then select 75%, 15%, and 15% as the training data, validation data and the test data .This prediction model is recorded as model 3, and the predicted and actual values are shown in FIGURE IX.

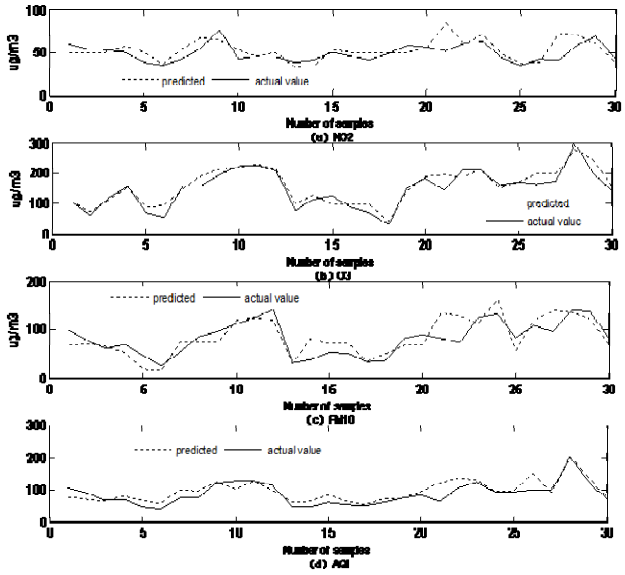

FIGURE VII. PREDICTIVE VALUE DIAGRAM OF MODEL ONE

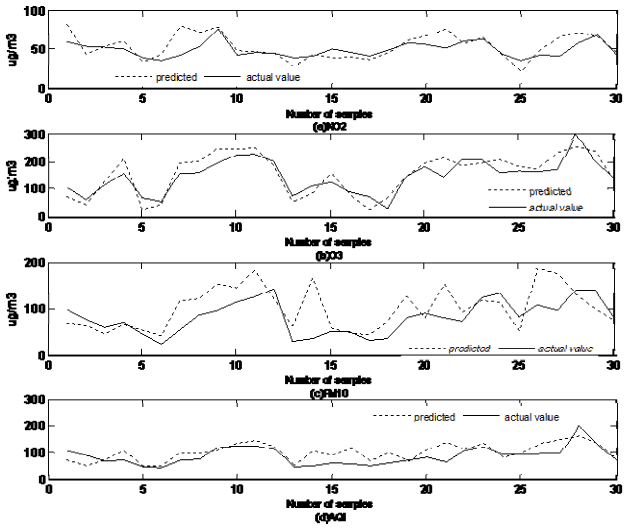

FIGURE VIII. PREDICTIVE VALUE DIAGRAM OF MODEL TWO

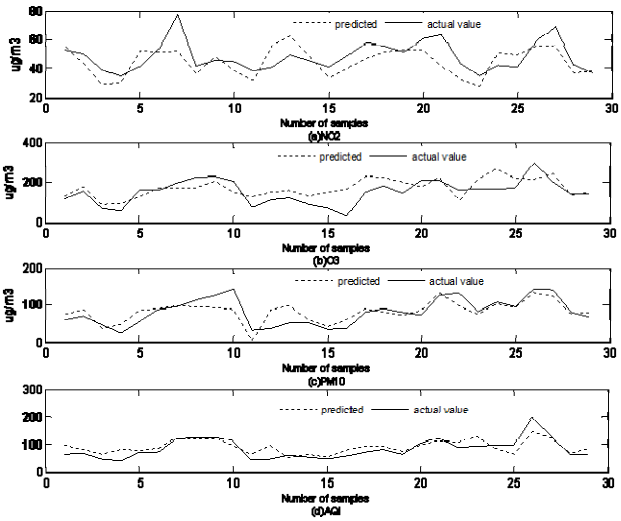

FIGURE IX. PREDICTIVE VALUE DIAGRAM OF MODEL THREE

2) Results analysis

Take the mean absolute error (MAE), mean relative error (MRE), root mean square error (RMSE), and the average accuracy E of the predicted values of four pollutants as the evaluation standard of judging prediction effect[10]. The three models according to the results of the experiments are given a contrastive analysis. The formulas for evaluation standard are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x^*_i - x_i\right| \tag{8}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|x^*_i - x_i\right|}{x_i}\times 100\% \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x^*_i - x_i\right)^2} \tag{10}$$

$$E = 1 - \frac{\left|x_i - x_i^*\right|}{x_i} \tag{11}$$

In the formulas of (8) to (11), x_i is the actual value of the sample i, x_i* is the predicted value of the sample i, n is the number of samples. The results are shown in table II.

The table II shows that all of the three neural network models can predict the air quality effectively. In contrast, the prediction accuracy of each parameter of model 1 is the highest, and the model 2 is the lowest. For model 1, the prediction accuracy of NO_2, O_3, PM_{10} and AQI are respectively 0.06% higher than model 2 and 0.01% higher than model3, 0.13% higher than model 2 and 0.07% higher than model 3, 0.44% higher than model 2 and 0.02% higher than model 3, 0.19% higher than model 2 and 0.05% higher than model 3. Comparing each error value of the three models, the error of model one is the smallest and the model two is the

largest.

TABLE II. THE EVALUATION OF THE THREE MODELS

Prediction of pollutant	Statistical parameter	model 1	model 2	model 3
NO_2	MAE	7.99	10.66	8.18
	MRE	16%	21%	16%
	RMSE	11.02	14.26	9.91
	E	0.83	0.78	0.82
O_3	MAE	16.83	29.53	28.41
	MRE	15%	25%	21%
	RMSE	21.35	34.59	37.22
	E	0.84	0.74	0.78
PM_{10}	MAE	19.58	32.87	17.60
	MRE	27%	49%	29%
	RMSE	24.50	43.94	22.69
	E	0.72	0.50	0.70
AQI	MAE	14.50	24.15	17.02
	MRE	19%	32%	23%
	RMSE	19.90	30.80	21.46
	E	0.80	0.67	0.76

IV. CONCLUSION

In this paper, three kinds of nonlinear prediction models are established by means of neural network. The information processing of BP neural network is a kind of mapping relation via the interaction among neurons. It has strong generalization ability for nonlinear network and also relatively simple and effective; The Open-loop model of NARX dynamic neural network establish the output feedback model on the basis of static BP neural network, it builds the output delay and introduces the output via the external feedback into the input. It can be said that the three models with reliability, high precision accuracy for air quality forecasting were successfully established. The experiments show that the NARX with dynamic feedback capability is better than BP neural network in higher precision accuracy; the BP neural network of bigger training sample is of higher precision accuracy.

ACKNOWLEDGMENTS

This work was supported by The Science and Technology Co-ordination Innovation Project of Shaanxi Province (2016KTCQ01-26), The Science and Technology Innovation Team of Shaanxi Province (2017KCT-30-02), Xi'an Science and Technology Project (2017084CG/RC047-XAYD004), The Science and Technology Innovation Team of Shaanxi Province for Broadband Wireless and Application (2017KCT-30-02), and Graduate innovation fund project (101-602080004).

REFERENCES

[1] M. S. Ji, K. K. Wan, S. L. Xu, "Study on urban air pollution control planning index system," Environmental Science and technology, vol.38, no. 12Q, pp. 440-444, 2015.

[2] E. Esposito, S. D. Vito, M. Salvato, V. Bright, R. L. Jones, & O. Popoola, "Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems," Sensors & Actuators B Chemical, vol. 231, pp. 701-713, Aug. 2016.

[3] A. Gautam, Y. C. Soh, "Stabilizing model predictive control using parameter-dependent dynamic policy for nonlinear systems modeled with neural networks," Journal of Process Control, vol. 36, pp. 11-21, Sept. 2015.

[4] D. J. Chen, "Application of multi information fusion in reservoir prediction of seismic attributes," Chengdu University of Technology, 2015.

[5] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, & S. Yu, "Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network," Applied Energy, vol. 114, pp. 1-9, Feb. 2014.

[6] X. D. Li, Q. Y. Gu, Z. H. Li, and J. G. Yang, "Thermal error modeling of spindle based on time series and neural network," Combined machine tools and automatic processing technology, no. 09, pp. 13-16, 2015.

[7] M. Zhang, "Research on stock forecasting model based on improved dynamic neural network," Inner Mongolia University, 2015.

[8] S. L. Badjate, V. DudulS, "Multi step ahead prediction of North and South hemisphere sun spots chaotic time series using focuced time lagged recurrent neural network model," Wseas Transactions on Information Science and Applications, vol. 6, no. 4, pp.684-693, Apr. 2009.

[9] J. Y. Niu, H. G. Wang, Z. Z. Shao, C. C. Song, "Application of BP artificial neural network algorithm in prediction of bird time series," Information technology and information technology, no. 03, pp. 93-97, 2013.

[10] M. Arhami, N. Kamali, M. M. Rajabi, "Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations," Research Article, no. 20, pp. 4777-4789, 2013.