

Atmospheric Environmental Quality Assessment RBF Model Based on the MATLAB

Zhonghua Fei¹, Dinggui Luo², Zhefei He¹, Bo Li¹

¹School of Mathematics and Physics Changzhou University, Changzhou, China; ²School of Environmental Science and Engineering, Guangzhou University, Guangzhou, China.
Email: 32321006@sina.com, fzh@cczu.edu.cn

Received February 14th, 2012; revised March 21st, 2012; accepted April 26th, 2012

ABSTRACT

A new method-RBF model is found to assess the atmospheric quality by use of the PREMNMX function in MATLAB to pretreat the original data and the RAND function to construct enough training samples, checking samples and outputs of their targets through linear interpolation between grades of the atmospheric quality evaluation standard. A favorable assessment result is achieved by applying this method to assess atmospheric environmental quality in a city, which shows this new method is meaningful in improving the precision and scientificity of atmospheric environmental quality assessment.

Keywords: Atmospheric Environmental Quality Assessment; The BP Network; The RBF Network; Artificial Neural Network

1. Introduction

Artificial neural network to simulate human brain, having the traits of self-organized, self-learning, adaptive, fault tolerance and other characteristics; they can be widely used in pattern recognition, etc. Environmental quality assessment in essence is a kind of pattern recognition on samples according to standards for environment quality, and therefore, it's of very realistic significance to research neural network approach in the application of atmospheric environmental quality assessment.

This paper introduces the principles of the radial basis function (RBF) network, training methods and the realizing functions in the Toolbox of MATLAB6.5. The realizing function has advantageous properties such as adaptation for determining the network construction and independence of the output on initial weight values. We apply this function to atmospheric quality assessment in a city, attempting to pretreat original data with help of PREMNMX functions in MATLAB, using the RAND function to construct enough training samples, checking samples and outputs of their targets through linear interpolation between grades of the atmospheric quality assessment standard, and determining the boundaries of atmospheric quality grades. The outcome we achieved is quiet satisfied. So this method is meaningful in improving the accuracy and objectivity of atmospheric environment quality assessment. In addition, compared with

RBF, the BP network showed artificiality on structure identification and random impact on results by initial weight value.

2. The Principle of Radial Basis Network

We will introduce RBF basic principle [1-2], training and its realization methods referenced by MATLAB6.5 toolbox function. The radial basis network is a three-layer feedforward network composed of input layer, hidden and output layer, see **Figure 1** (with a single output neurons as an example) where hidden neurons use radial basis function as activation function, usually with Gaussian function as radial basis function.

About neural network information transmission, the input layer is only responsible for information transmission, and the input and output are the same.

For hidden layer, each neuron inputs the product of the distance between the vectors $W1_i$ and the vector X^q multiplied by its own offset value $b1_i$. The vector $W1_i$ is the connected weight value between neuron of hidden layer and of input layer and also known as the i^{th} hidden layer neuron function (RBF) center. The vector X^q represents the q^{th} input vector denoted by

$$X^q = (x_1^q, x_2^q, \dots, x_j^q, \dots, x_m^q).$$

From the **Figure 2**, we can see that the i^{th} neuron input for the hidden layer produced by the corresponding the q^{th} input of the input layer is k_i^q :

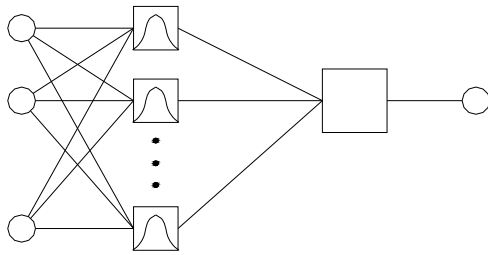


Figure 1. Construction of RBF network.

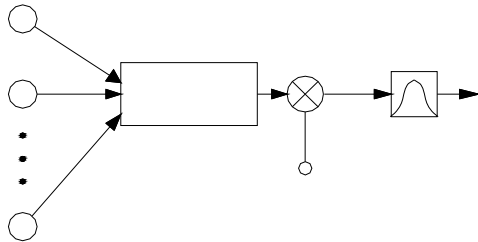


Figure 2. Sketch map for input and output about the hidden nerve unit in RBF network.

$$k_i^q = \sqrt{\sum_j (w1_{ji} - x_j^q)^2} \times b1_i.$$

By Gaussian transformation, the i^{th} output from the i^{th} neuron input of the hidden layer is

$$r_i^q = e^{-(k_i^q)^2} = e^{-\left(\sqrt{\sum_j (w1_{ji} - x_j^q)^2} \times b1_i\right)^2} = e^{-\left(\|W1_i - X^q\| \times b1_i\right)^2}.$$

In MATLAB neural network function it sets

$$b1_i = 0.8326/C_i.$$

And then the hidden layer neurons output is changed to:

$$r_i^q = e^{-\left(\frac{\|W1_i - X^q\| \times 0.8326}{C_i}\right)^2} = e^{-0.8326^2 \left(\frac{\|W1_i - X^q\|}{C_i}\right)^2}.$$

The values of C reflects response width of output for input, shown in **Figure 3**. For example: if C takes 4, then the response width will be large, often 0.5 above when the distance between the input vector and corresponding weights vector is less than 4, conversely, response width is small. The bigger C takes, the better smoothness between two neurons we will get, caused by the response range of the hidden neurons to input vector expand with it.

$$r_i^q = \begin{cases} > 0.5 & \|W1_i - X^q\| < C \\ = 0.5 & \|W1_i - X^q\| = C \\ < 0.5 & \|W1_i - X^q\| > C \end{cases}$$

The output is weighted summation of each hidden

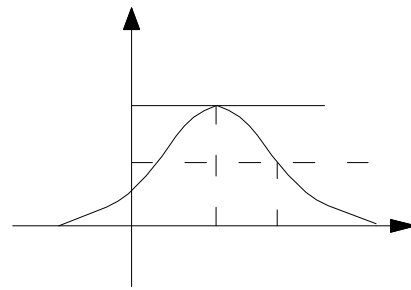


Figure 3. Relationship among $W1_i$, X^q , C and r_i^q .

layer neurons output, excitation function using pure linear function. Then the neuron output y^q produced by the corresponding the q^{th} input of the input layer is

$$y^q = \sum_{i=1}^n r_i^q \times w2_i.$$

RBF network training is divided into two steps, the first step for the supervised learning training the weights $W1$ between input layers and hidden layer, the second step for supervised learning training the weights $W2$ between hidden layer and the output layer. Network training needs to provide input vector (X), corresponding target vector (T) and radial basis function expansion constants (C). When the number of neurons in hidden layer equals the number of input vector, the value of C can take a smaller one. The purpose of the training is to get the weights $W1$, $W2$, and the offset value $b1$, $b2$. (when the number of hidden units equals the number of input vector, we will take $b2 = 0$).

In RBF networks training, one of the key problems is to decide the number of neurons in hidden layer. In the past, we often make it equal with the number of the input vector. Apparently, for many input vector, too much hidden units is difficult to acceptable. Therefore we will improve the method. The basic principle is: 0 as a neuron started training, by checking the output error to make the network automatically increase neurons, after the training sample looping once, using the training sample which make the network produce have the maximum error as the weight vector $W1_i$ to generate a new hidden neuron, then recalculating, checking the error of the new network, repeating this process until it reaches the required error or maximum number of hidden neurons. This training algorithm is realized by NEWRB functions in MATLAB6.5.

It can be seen that NEWRB function has properties such as adaptation for determining the network construction and independence of initial weight value on person, which reduce randomness of the network training.

3. Application of the Radial Basis Network

The following by introducing an atmospheric environ-

mental quality assessment is given to show the entire process of the application of radial basis network.

3.1. Raw Data

The assessment standards of atmospheric environmental quality and the atmospheric monitoring data of all the year round in one city can be seen in **Tables 1** and **2** [3]. Then we will use the radial basis network to evaluate the atmospheric quality of the city.

3.2. Preparations for Neural Network

1) Training set, test samples and the formation of the desired objective.

Training set: we use RAND function in MATLAB to generate the training sample by linear interpolation with the random uniform distribution at different evaluation levels. By 500 samples generated when less than Level 1, 500 between Level 1 and Level 2 and so on, 2000 training samples are formed. This method solves the problem that the training samples are too few to construct test samples in the past when they only used the evaluation standard as the training sample.

Test samples: we generate the test samples in the same way of generating the training sample. 100 generated when less than Level 1, 100 between Level 1 and Level 2, by analogy, 400 test samples are formed.

The desired objective (for the training set and test samples): adopting an output neurons, when both the training

Table 1. Evaluation standards for atmospheric environmental quality.

The evaluation index	Level 1	Level 2	Level 3	Level 4
SO ₂ (mg/m ³)	0.002	0.02	0.06	0.1
NO _x (mg/m ³)	0.016	0.05	0.1	0.15
TSP (mg/m ³)	0.02	0.15	0.3	0.5
Dust fall (t/km ² -month)	3.4	6.8	10.2	13.6

Table 2. Monitoring data for atmospheric environmental quality and assessment results.

The evaluation index	The sample for evaluating			
	spring	summer	autumn	winter
SO ₂ (mg/m ³)	0.11	0.093	0.057	0.085
NO _x (mg/m ³)	0.038	0.036	0.014	0.014
TSP (mg/m ³)	0.551	0.514	0.35	0.34
Dust fall (t/km ² -month)	14.3	16.2	9.6	10.2
Network output	3.0421	3.0139	2.7118	2.8419
The rating	Level 4	Level 4	Level 3	Level 3

samples and test sample are less than Level 1, according to their interpolation proportion it will generate a number between 0 and 1 as the corresponding desired objective; when both the training samples and test sample are between Level 1 and Level 2, according to their interpolation proportion it will generate a number between 1 and 2 as the corresponding desired objective; when between Level 2 and Level 3, a number between 2 and 3 as the corresponding desired objective; and so on.

2) The demarcation of atmospheric quality evaluation grades.

According to the above method it can be sure that the network output of Level 1, Level 2, Level 3 and Level 4 are respectively less than 1, between 1 and 2, between 2 and 3 and larger than 3.

3) The original data preprocessing.

There are two kinds of pretreatment plans. One is normalization, namely, by use of PRENMX function normalizing raw data to between -1 and 1; the other is no normalization, that is, the original data not being pre-processed.

3.3. Radial Basis Net Constructions, Training and Testing

1) Radial basis network construction.

The number of input layer neurons in RBF network depends on the number of atmospheric quality evaluation index which is 4, and the number of the output layer neurons is set to 1. The number of the hidden units can be adaptively determined by the use of MATLAB NEWRB function training network. The excitation function of the hidden units is RADBAS, the weighted function DIST, and the input functions NETPROD. The excitation function of the output layer neurons is pure linear function PURELIN, the weighted function DOTPROD, and the input functions NETSUM [4].

2) Network training, testing and atmospheric quality assessment.

Using the functions provided by MATLAB6.5, network training, testing and application can be realized easily by programming. With the method of normalization, network training 6 times, we can get mean square error of the training sample and test sample is 0.0014 and 4.6858×10^{-4} respectively. Randomly choosing 12 training samples and 12 test samples, the relative errors can be seen in **Table 3**, which shows network has good generalization ability. Applying the trained network for atmospheric quality assessment, we get the network output for the spring, summer, autumn and winter (see **Table 2**), which is 3.0421, 3.0139, 2.7118 and 2.8419 respectively.

3) Analysis of the results.

According to the network output the evaluating samples can be ranked in the ascending order which is autumn, winter, summer and spring. It reflects the air quality

Table 3. Network output errors of the training samples and checking samples.

The training sample				Test sample			
Serial number	Target output	The actual output	The relative error (%)	Serial number	Target output	The actual output	The relative error (%)
1	0.8952	0.8865	0.9741	1	0.3770	0.3514	6.7701
2	0.9424	0.9387	0.3884	2	0.9073	0.8999	0.8189
3	0.3351	0.3145	6.1538	3	0.6702	0.6415	4.2811
4	1.8952	1.8781	0.9023	4	1.8364	1.8284	0.4356
5	1.9424	1.9176	1.2768	5	1.1453	1.1599	1.2748
6	1.3351	1.3615	1.9774	6	1.1715	1.1883	1.4341
7	2.8952	2.9005	0.1831	7	2.8364	2.8490	0.4442
8	2.9424	2.9411	0.0442	8	2.1453	2.1199	1.1840
9	2.3351	2.3335	0.0685	9	2.1715	2.1489	1.0408
10	3.8952	3.8882	0.1797	10	3.8364	3.8477	0.2945
11	3.9424	3.9140	0.7204	11	3.1453	3.1263	0.6041
12	3.3351	3.3202	0.4468	12	3.1715	3.1517	0.6243

order of the four samples (see **Figure 4**). The graph intuitively shows the distribution of each index of the four samples. It is clear that: in autumn, only one index exceeded Level 3, in winter, there are two indexes exceeded Level 3, in summer, there are two indexes exceeded Level 4 and in spring there are three index exceeded Level 4.

According to the demarcation of atmospheric quality assessment grades, we know the level of the samples that is spring and summer as Level 4, autumn and winter as Level 3.

The above analysis shows that RBF network method not only can distinguish quality level, but also can accurately give their difference at the same quality level.

3) The original data normalized problem.

Many experiments show that if the network training uses the original data not normalized, no matter what we adjust the mean square error, the network can not output the right result.

3.4. Application Effect of the BP Network

Compared with radial basis network, we construct BP

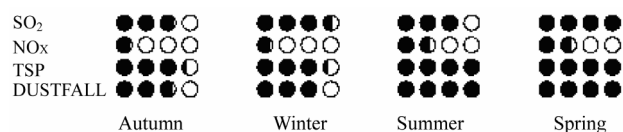


Figure 4. Order for the environment quality of dividing samples.

network to solve the problem. The process is as follows:

1) BP network construction.

Taking three-layer network, the number of the input and output layer neurons is determined as 4 and 1 respectively. There has not been a uniform method how to determine the number of the hidden units. Here we follow this common empirical formula [5]:

$$m = \sqrt{n_i + n_o} + \alpha \quad m = \log_2 n_i \quad m = \sqrt{n_i n_o}$$

where m is the number of the hidden units, n_i is of input layer, n_o is of the output layer and α is between 1 and 10.

The numbers of two kinds of hidden units are 10 and 3 respectively.

The input and output functions (excitation function) for hidden units and the output units are respectively by means of hyperbolic tangent function and linear function, namely, TANSIG and PURELIN functions in MATLAB. Network is trained by using Powell-Beale conjugate gradient back propagation algorithm, namely TRAINCGB function in MATLAB [4].

2) BP network's dependence on network structure and initialized weights.

Normalizing raw data and controlling the mean-square error as 0.0001, we make the experiments for two different structures (hidden units) and twice for the same structure (reflect the influence of different weights) to assess the atmospheric quality. The result can be seen in **Table 4**. There exists large differences among the assessment results of the four samples and the training times

Table 4. The effect for the construction of BP net and initial weight value on atmospheric quality assessment.

Number of neurons in hidden layer	Training serial number	Spring	Summer	Autumn	Winter	Train number
10	1	4.1893	4.7583	3.1918	3.7585	35
	2	1.0842	2.1688	1.0065	0.5074	20
3	1	1.1726	1.9724	0.26575	0.3653	34
	2	3.4301	3.5867	2.2894	2.8927	65

which shows great randomness.

4. Conclusions

1) NEWRB-the realizing function of RBF network provided by MATLAB6.5 toolbox has properties such as adaptation for determining the network construction and independence of initial weight value on person, which can reduce the randomness of network training, so as to improve the objective scientificity of the atmospheric quality assessment.

2) Using RAND function in MATLAB to construct enough training samples, checking samples and the corresponding target output, it can avoid the past problem only with the evaluation standard as the training samples which are too few to construct test samples. The method used in this paper has remarkable effect in improving generalization of network.

3) It is suggested that the original data can be normalized in the application of MATLAB toolbox functions to implement radial basis function; otherwise, the network is hard to convergence.

4) A favorable outcome appeared after we apply the radial basis network mentioned in this paper to evaluate atmospheric environmental quality in a City. The network output of the four evaluating samples is ranked in the ascending order which is 2.7118, 2.8419, 3.0139 and 3.0421. The corresponding samples are autumn, winter, summer, and spring respectively which perfectly reflect their atmospheric quality order. According to the demar-

cation of atmospheric quality grades, we know the level of the samples. Spring and summer are at Level 4, autumn and winter are at Level 3. In practice, we can use many network outputs from the monitoring stations to create regional air quality isoclines for the atmospheric quality assessment which can reflect variation of regional atmospheric quality more accurately.

5) Compared with RBF networks, the BP network has random of determining the network construction and initial weight value. So we suggest that RBF network is used to assess the atmospheric quality.

REFERENCES

- [1] P. D. Wasserman, "Advanced Methods in Neural Computing," Van Norstrand Reinhold, New York, 1993.
- [2] S. Cong, "The Function Analysis and Application Study of Radial Basis Function Network," *Computer Engineering and Applications*, Vol. 38, No. 3, 2002, pp. 85-87.
- [3] W. D. Yang and H. I. Cheng, "The Applied Research of Neuro-Network in Evaluation of Air Quality," *Industrial Safety and Environmental Protection*, Vol. 27, No. 9, 2001, pp. 31-33.
- [4] D. Xu and Z. Wu, "The System Analysis and Design Based on MATLAB6.X Neural Network," Xidian University Press, Xi'an, 2002.
- [5] L. Q. Han, "The Theory of Artificial Neural Network, Design and Application," Chemical Industry Press, Beijing, 2001.