Acta Technica 62 (2017), No. 6B, 1259–1268

Pollution Types and Risk Assessment based on RBF Neural Network

Hua Zhang[1]

Abstract. In order to improve effectiveness of evaluation result of pollution types, an evaluation method of pollution type based on RBF neural network of empirical function was proposed. Firstly, central value of Gaussian function of all hidden layer nodes was confirmed according to all input vectors and elementary training of network was realized; secondly, RBF neural network learning was implemented by using learnability owned by radial basis function network (RBFN) and adopting linear least square method (LLS) and gradient descent method to improve learning efficiency; finally, effectiveness of algorithm was verified by simulation experiment.

Key words. Two stages, Radial basis function network, Pollution type, Empirical function.

1. Introduction

Environmental pollution evaluation and forecast not only can realize environmental pollution change dynamics better, but also can provide real-time, systematic and accurate environmental quality information. It also strengthens control of environmental and avoids occurrence of serious pollution incidents at the same time.

At the present stage, research of environmental quality evaluation and analysis methods is no longer limited to index evaluation method, principal component analysis, grey system analysis method and fuzzy mathematical analysis method. Multiple environmental pollution evaluation methods also include artificial neural network method, matter-element extension method, projection pursuit analysis method, set pair analysis method and so on. Artificial neural networks are artificial intelligence machine to simulate structure and function of human brain. It has been widely used for image processing, financial market simulation, earthquake prediction, hydraulic power and so on at present. Compared with other methods, artificial neural networks have the ability of self-learning, associative storage function and finding optimal solution with high speed. It is a powerful tool to handle non-linear, uncertain and complex problems. At present, evaluation and forecast report of urban

[1]Shaanxi Institute Of International Trade & Commerce, Xianyang Shanxi, 712046, China

environmental pollution is mainly evaluated by converting into air pollution index (API). The result is simple and direct, and it is applicable to show urban short-term environmental quality status and change trend. Extension neural network (ENN) integrates advantages of two emerging disciplinary fields as organic combination of extenics and artificial neural networks, thus processing stronger function and advantages. It receives better application in the aspects of transformer fault diagnosis, exhaust gas detection and classification and so on, and has outstanding effect on solving classification, clustering and recognition of feature vector based on interval.

An evaluation method of pollution type based on RBF neural network of empirical function is proposed in the Thesis. RBF neural network learning is implemented by using learnability owned by radial basis function network (RBFN) and adopting linear least square method (LLS) and gradient descent method to improve learning efficiency. Effectiveness of algorithm is verified by simulation experiment.

2. Introduction to artificial neural networks

2.1. Conception of artificial neural networks

Artificial neural networks (ANNs for short) can also be called neural network (NNs) for short or called connectionist model. It is a mathematical model of algorithm to model behavior characteristics of animal neural network and implement distributed parallel information processing. Such a network relies on complexity of system and achieves the purpose of processing information by adjusting interconnection relation among a large number of internal nodes. Artificial neural networks have the ability of self-learning and self-adapting. It can analyze and master potential rules between the two by a batch of mutually corresponding input-output data provided in advance, and use new input data to calculate output result according to these rules finally. Such a learning and analyzing process is called "training".

Artificial neural networks is an artificial intelligence technology to simulate biological process of human brain, is a multilayer network architecture abstracted according to electrochemistry activity of cerebral neurons and is complex non-linear system formed by a mass of neuronal interconnection. Neuronal structure is shown in Fig 1. All input M adding threshold value d after weighted sum by a weight K, and then through function of transfer function is output a of the neuron.

$$a = f(MK + d)\,. \tag{1}$$

2.2. Radial basis function network theory

Radial basis function network theory is a three-layer feed forward neural network, including one input layer, one radial basis layer (namely hidden layer) and one output layer. The basic principle is that radial basis function is taken as basis of hidden unit to form hidden layer space, and hidden layer converts input vector and transfers modal input data of lower dimension into higher dimensional space,

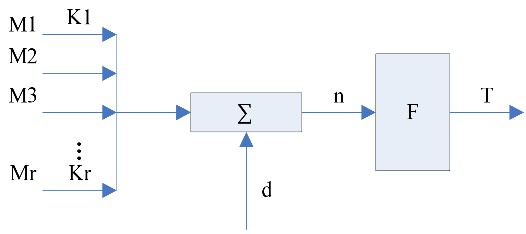

Fig. 1. Artificial neuron model

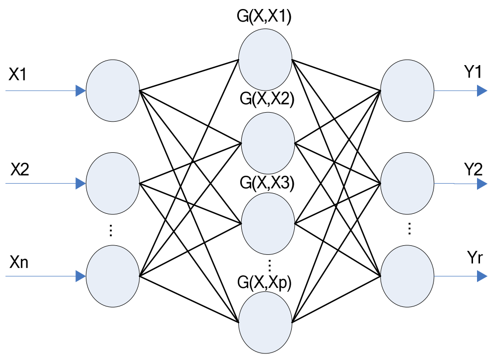

Fig. 2. RBF network structure diagrams

making linear inseparable problem in lower dimensional space separable in higher dimensional space. Radial basis function simulates local adjustment and receptive field covered mutually in human brain, so it is a local approaching network. Scientific community has proved that it can approach arbitrary function with arbitrary precision. The topological structure is shown in Fig. 2.

Input vector shall be transferred to hidden layer after input layer nodes obtain input vector. Hidden layer nodes are composed of radial basis function, and multiple forms can be adopted for radial basis function (Gaussian function is adopted usually). Hidden layer implements non-linear change and map input space into a new space. Output layer is simple linear function usually. Hidden layer nodes are connected completely with output layer nodes with different weight. Activation function of hidden layer nodes generates a local response for input excitation. Input vector is closer to the center of basis function, the response made by hidden layer nodes will be greater. Output response of the jth node of hidden layer shall be:

$$G_j(x) = \exp\left\{-\frac{||x-\mu_j||^2}{\sigma_j^2}\right\}. \tag{2}$$

Output layer is weighted sum of all units of hidden layer

$$\hat{y} = f(x) = \sum_{j=1}^{c} \varpi_j G_j(x). \tag{3}$$

Where x =[X1, X2 , $\cdots$, Xn] is input vector, μ_j and σ_j are center and size of the jth neuron, c is number of neuron, and ϖ_j is corresponding weight of the jth neuron.

3. Environmental pollution evaluation algorithm based on RBF

3.1. Sample collection of training data

Sample data required by training is directly related to the quality of empirical function accuracy of training, so a set of good training samples are guarantee of empirical function accuracy. 5000 valid data are selected as sample in the Thesis, in which 4000 are used to train empirical function and the later 1000 are used to detect training effect.

3.2. Hybrid learning algorithm of RBF neural network

Learning of RBF network is divided into two processes. The first process: confirm central value c_j of Gaussian function of all hidden layer nodes according to all input vectors. The second process: obtain weight ϖ_{jt} of output layer by using principle of least square method according to samples after parameters of hidden layer j is confirmed. The key problem of establishing RBF neural network is to confirm the center of radial basis functions according to given training samples, because G_j and anticipated output y_t are known for all training samplesG_j in case the center c_j of radial basis function is confirmed. Output weight ϖ_{jt} can be obtained by least square method.

As for construction process of gradient descent method, error function shall be defined firstly:

$$E = \frac{1}{2}\sum_{n=1}^{N} E^n \,. \tag{4}$$

Where E^n is error at the time of inputting the nth sample and N is sample number. Definition of E^n is:

$$E^n = \sum_{k=1}^{s} (t_k^n - y_k^n)^2, \quad n = 1, 2, 3 \ldots, N \,. \tag{5}$$

If error function shall be minimized, correction of parameter shall be directly proportional to its negative gradient, and then the following equation can be obtained after $\Delta C_j = -\eta_1 \frac{\partial E}{\partial C_j}$ and $\Delta \sigma_j = -\eta_2 \frac{\partial E}{\partial \sigma_j}$ are put into:

$$\Delta C_j = 2\eta_1 \sum_{n=1}^{N}\sum_{k=1}^{s} (t_k^n - y_k^n) \cdot R_j^n \cdot w^n(k,j) \cdot \frac{p^n - c_j^n}{(\sigma_j^n)^2} \,. \tag{6}$$

$$\Delta\sigma_j = 2\eta_2 \sum_{n=1}^{N} \sum_{k=1}^{s} (t_k^n - y_k^n) \cdot R_j^n \cdot w^n(k,j) \cdot \frac{||p^n - c_j^n||^2}{(\sigma_j^n)^3} \,. \tag{7}$$

When all sample input is finished, iterative method shall be adopted to adjust parameters, as follows:

$$C_j(m+1) = C_j(m) + \Delta C_j \,. \tag{8}$$

$$\sigma_j(m+1) = \sigma_j(m) + \Delta\sigma_j \,. \tag{9}$$

It is learning rate of the center, is learning rate of Gaussian width and m is iterations. In order to guarantee generalization performance of classifier, learning rate of adopted Gaussian width is usually bigger than learning rate of the center, because small learning rate makes algorithm convergence slow, but overlarge learning rate may cause unstable algorithm.

3.3. Training flow diagram of empirical function

Set input matrix to be $M \in R^{r\times N}$, input matrix of hidden layer to be $P \in R^{u\times N}$ and output layer matrix to be $K \in R^{s\times N}$ firstly, where n is training sample. If undetermined output layer weight of RBF network W $\in R^{s\times u}$, the relation of the three is:

$$K = W \times P \,. \tag{10}$$

Target output of samples is $T = (t_1, t_2, ..., t_s)^T \in R^{s\times N}$. Linear least square (LLS) method is adopted to make error between target output and actual network output reach the minimum here. R^+ of R is used to obtain R^+ of W.

Algorithm flow diagram of empirical function trained in the Thesis is divided into two stages. The first stage is sample processing. As samples have repeatability, samples stored in sample database are guaranteed to be representative after sample processing. Repeated training is eliminated and training rate is improved. The second stage is training empirical function. Training process is shown in Fig. 3.

4. Experimental analysis

At present, there are mainly 3 urban pollution monitoring items in China according to key points of environmental pollution characteristics and pollution prevention, including SO2, NO2 and inhalable particle PM10. Class of environmental pollution can be divided into: superior, good, slight pollution, moderate pollution, medium-heavy pollution and serious pollution according to different pollution concentration. More than 5000 historical data in 4 areas such as Songyuan section of Songhua River, Harbin section of Songhua River, Gansu section of Yellow River, Guanmenlazi reservoir of Huadian and so on are taken as samples for training. Unknown environmental classes are obtained by establishing good network finally to achieve the purpose of environmental quality evaluation.

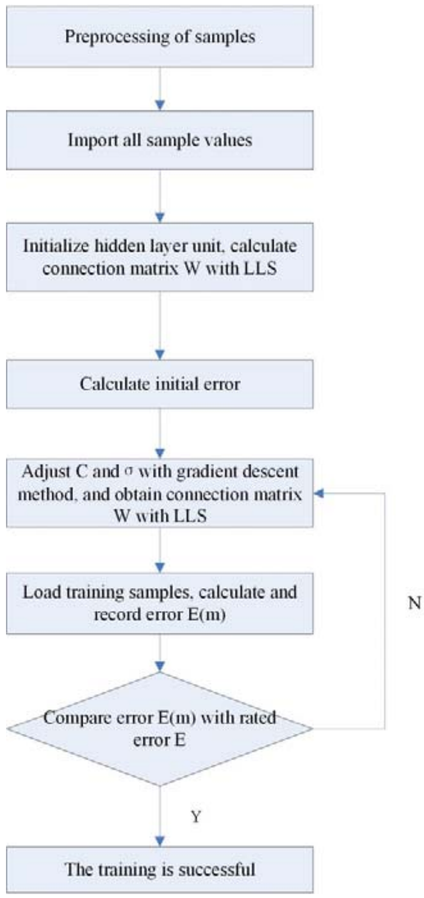

Fig. 3. Algorithm flow diagram of empirical function

4.1. Data processing

As singular sample data may exist in obtained samples, so-called singular sample data refer to especially large or small sample vectors compared with other input samples. Existence of singular samples will cause increase of network training time, and may cause that network cannot converge. Therefore, normalization processing needs to be implemented for training data and testing data before training. The maximum and minimum value method shown in equation (11) is adopted in the Thesis to normalize sample data in the range of [0, 1].

$$y = \frac{x - \min(x)}{\max(x) - \min(x)}\,. \tag{11}$$

Detailed information of experimental data is shown in Table 1.

Table 1. Detailed information of experimental data

Data source	Sample No.	Monitoring position	Monitoring items
Songyuan section of Songhua River	545	Pasture in Songyuan, Xidazuizi, slop vat	COD, BOD, ammonia nitrogen, VP and petroleum
Harbin section of Songhua River	1157	Sanjiazi, second water source, cement plant, Daliangzi, wetland of golden bay, Hulan River mouth	pH, DO, PI, COD, BOD, ammonia nitrogen, TP, TN and fecal coliform
Gansu section of Yellow River	2161	Fuhe Bridge, Baolan Bridge, Shichuan Bridge, Jingyuan Bridge, Five Buddhist Temple, Degao Bridge, Gutter Bridge, Zhe Bridge, Yehu Gorge, Yu Well, Taoyuan Bridge, Nieshui Bridge, Birch Forest, Boyang Bridge, vineyard, Pingchen Bridge, flood-control dam, bridgehead of Ning County	COD, ammonia, TP, VP and fecal coliform
Guanmenlazi reservoir of Huadian	2142	Guanmenlazi reservoir of Huadian	DO, PI, TP, ammonia, and TN

4.2. Result analysis

2/3 of data in each set of data source are selected as training samples and 1/3 of data are selected as testing data in the Thesis. 5 cross verification is adopted in training process. Monitoring index of water quality is taken as input vector and water quality grade is taken as output vector. Commonly used RBF function is adopted for kernel function. Proposed method is adopted to make an optimization for parameters of RBF. Relationship curves of evolution algebra and adaptation of four data sets are shown in Fig. 4-7 respectively. When penalty factor C, RBF kernel function parameter σ and accuracy rate of cross validation are selected from numerical values in Table 1 respectively, classification accuracy and generalization ability of algorithm is optimal.

The result is compared with result of classical SVM in order to verify effectiveness of method in the Thesis. Classical SVM is calculated by widely used free software package LIBSVM developed by professor Lin Zhiren in Taiwan University at present. Default is adopted for main parameters in the algorithm, namely penalty factor C=1. Reciprocal of attribute number of sample data is adopted for RBF kernel function parameter σ. Evaluation result of water quality is shown in Table 2.

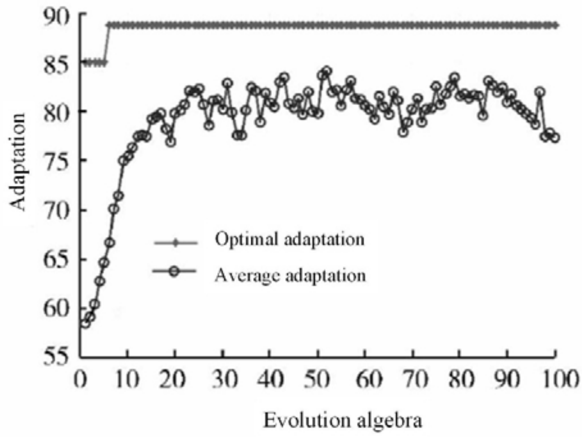

Fig. 4. Adaptation curve of songyuan section of songhua river

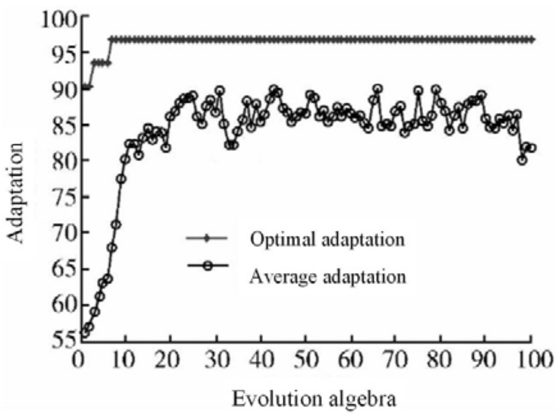

Fig. 5. GA adaptation curve of heilongjiang section of songhua river

Table 2. Evaluation result of water quality

Method	Songyuan section of Songhua River	Harbin section of Songhua River	Gansu section of Yellow River	Guanmenlazi reservoir of Huadian
BP	72.22	85.71	92.19	67.04
SVM Classical SVM	66.67	76.19	70.31	88.52
SVM Evolved SVM	77.78	85.71	93.75	94.44

It can be seen from Table 2 that recognition accuracy of water quality evaluation model of evolved SVM is improved by 16.7%, 12.5%, 33.3% and 6.7% compared with that of water quality evaluation model of classical SVM respectively, improved by 7.7%, 0%, 1.7% and 40. 9 % compared with that of BP neural network method,

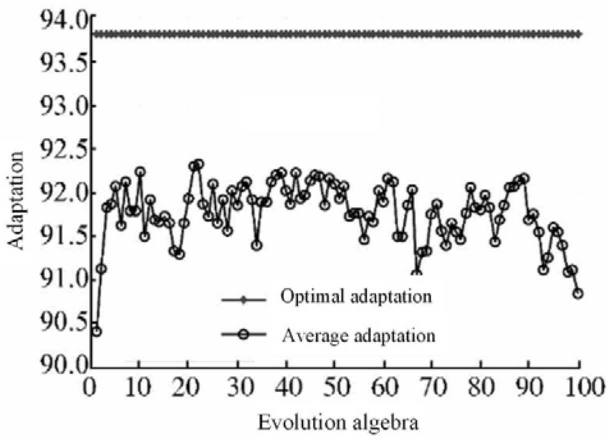

Fig. 6. GA adaptation curve of gansu section of yellow river

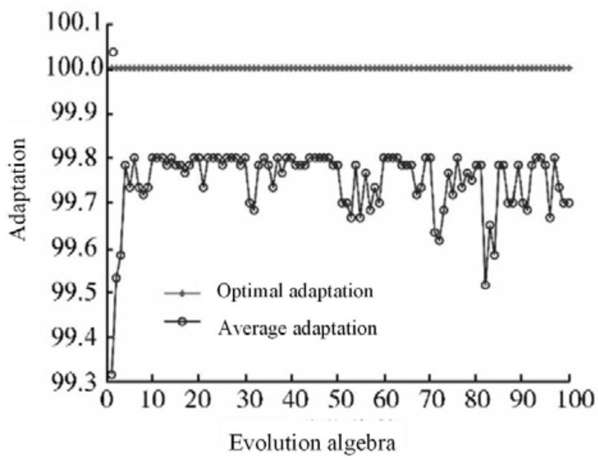

Fig. 7. GA Adaptation curve of guanmenguizi reservoir of huadian

which shows that the method in the Thesis has good classification accuracy and generalization performance.

5. Conclusions

An evaluation method of pollution type based on RBF neural network of empirical function is proposed in the Thesis. Central values of Gaussian function of all hidden layer nodes are confirmed according to all input vectors and elementary training of network is realized. RBF neural network learning is implemented by using learnability owned by radial basis function network (RBFN) and adopting linear least square method (LLS) and gradient descent method to improve learning efficiency.

References

[1] Pahlavan K, Krishnamurthy P and Geng Y: (2015) *Localization challenges for the emergence of the smart workd.* IEEE Access, 3(1), pp. 3058-3067

[2] Lv, Z., Tek, A., Da Silva, F., Empereur-Mot, C., Chavent, M., & Baaden, M.: (2013). textitGame on, science-how video game technology may help biologists tackle visualization challenges. PloS one, 8(3), e57990.

[3] Lv, Z., Halawani, A., Feng, S., Li, H., & Réhman, S. U.: (2014). textitMultimodal hand and foot gesture interaction for handheld devices. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 10.

[4] Weisen Pan, Shizhan Chen, Zhiyong Feng: (2013) textitAutomatic Clustering of Social Tag using Community Detection. Applied Mathematics & Information Sciences, 7(2): 675-681.

[5] Yingyue Zhang, Jennifer W. Chan, Alysha Moretti, and Kathryn E. Uhrich: (2015) textitDesigning Polymers with Sugar-based Advantages for Bioactive Delivery Applications, Journal of Controlled Release, 219, 355-368.

[6] Yingyue Zhang, Qi Li, William J. Welsh, Prabhas V. Moghe, and Kathryn E. Uhrich: (2016) textitMicellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity, Biomaterials, 84, 230-240.

[7] Jennifer W. Chan, Yingyue Zhang, and Kathryn E. Uhrich: (2015) textitAmphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation, Bioconjugate Chemistry, 26(7), 1359-1369.

[8] Liu D, Dong W, Wu J, et al.: (2014) textitA risk assessment method based on RBF artificial neural network-cloud model for urban water hazard[J]. Yellow River, 27(5):2409-2416.

[9] Li L, Li J, Ma J H :(2010) textitPollution assessment of heavy metals in soils based on RBF neural network.[J]. Environmental Science & Technology.

[10] Jia Z, Gong L: (2008) textitThe Project Risk Assessment Based on Rough Sets and Neural Network (RS-RBF)[C]// International Conference on Wireless Communications, NETWORKING and Mobile Computing. IEEE, 2008:1-4.

[11] Ma Z Q, Meng Q B: (2009) textitThe Research on Risk Evaluation of Real Estate Development Project Based on RBF Neural Network[C]// International Conference on Information and Computing Science. IEEE, 2009:273 - 276.

[12] Chen F, Xie H T: (2013) textitOn the urban subway construction safety risk assessment based on the rough set and RBF neural network[J]. Journal of Safety & Environment, 13(4):232-235.

[13] Mi Y, Rui S, Dong S, et al.: (2014) textitOptical fiber sensor system for monitoring leakage current of post insulators based on RBF neural network[C]// International Conference on High Voltage Engineering and Application. IEEE, 2014:1-4.

[14] Cao A Z, Tian L, Cai C F, et al.: (2006) textitThe Research and Prediction of Atmospheric Pollution Based on Wavelet and RBF Neural Network[J]. Journal of System Simulation, 18(5):1411-1413.

[15] Liu Z B, Yang X W: (2012) textitAssessment of the Underground Water Contaminated by the Leachate of Waste Dump of Open Pit Coal Mine Based on RBF Neural Network[J]. Advanced Materials Research, 599:272-277.

Received May 7, 2017