Management Science and Engineering Vol.2 No.2 June 2008

# IMPROVED ANFIS TO FORECAST ATMOSPHERIC POLLUTION[1]

**XU Hua**[2] **XUE Hengxin**[3]

**Abstract:** This paper studies the Adaptive neural fuzzy inference systems (ANFIS) based on the fuzzy model Sugeno, on the basis of the network is given the right to amend the value of the algorithm, that is the most comprehensive steep decline in law and the least-squares method to be a mixed learning algorithm. And this paper gives its improved optimization, using conjugate gradient method to improve its premise parameters of learning speed. The concentration of air pollutants has strong nonlinear characteristics, we should conduct more accurate forecasts, it must take to capture nonlinear changes of the forecasting methods. The simulation presented a choice of the most relevant input technology, and achieved very good results.

**Key words**: Sugeno fuzzy model, ANFIS, The conjugate gradient method, Data fitting

## 1. INTRODUCTION

[1] Fund projects: State 863 Project Fund (2007AA1Z158) Adaptive Multi-channel selection mechanisms and the key technical user model, Wenzhou city cleaner production and management information system design and environmental protection projects funded. China.

[2] (1978 -), female, NanJing University Of Science & Technology in Reading doctoral students, the Jiangnan Information Engineering University Lecturer, the main research direction for the environmental pollution, fuzzy systems and management information systems.
NanJing University Of Science & Technology, Nanjing, 210094, P.R. China; School of Information Engineering, Southern Yangtze University, Wuxi, 214122, P.R.China.
Email:joanxh2003@163.com
[3] NanJing University Of Science & Technology, Nanjing, 210094, P.R. China

* Received 4 February 2008; accepted 19 April 2008

Forecast methods commonly used at home and abroad are composite Index, gray clustering method, regression analysis, the number of theoretical prediction, fuzzy comprehensive evaluation method, fuzzy gray model,[4] and so on. These methods or models exist too complex, difficult to calculate, or have more subjective wishes, or the accuracy of forecasting existence poor inadequate.

ANN is a powerful tool solving nonlinear problems. In recent years, artificial neural network research develops rapidly, and has been widely used for system control and forecasting Etc.[5] BP neural network is the most important type of artificial neural network. BP is the most widely used network model of atmospheric environmental forecasting. But BP neural network also has many shortcomings. Its structure is determined by the performance of man-made. It is the slow pace of training, and the initial weight on the results is random. While BP neural network learning process into local minimum easy, easy existence of a concussion and redundant network connections or nodes defect.

Using ordinary neural network identification must also consider the choice of network, structure determination, and other issues. ANFIS based on the fuzzy model Sugen[6] can address these deficiencies, it combines fuzzy inference system and neural network structure, and it uses the most and least squares with the steep decline mixed learning algorithm[7] to study parameters, the earth reduces the complexity of the learning process. On this basis, this paper improved conjugate gradient method further, greatly improved its premise parameters of learning speed, improved its performance.

Simulation tests, we found that if you have multiple inputs, select two most relevant input than all of the technical data required for the importation, and modeling of time can be greatly reduced, and lower error can be indicators of nonlinear mappings. This show that atmospheric pollution forecast and other nonlinear prediction can use this method, and it can be popularized.

## 2. ANFIS PRINCIPLE

### 2.1 Basic structure and learning rules of ANFIS

As a rather unique neural network, with the same approximation to any arbitrary precision linear or nonlinear functions, and fast convergence, small error for samples less training, ANFIS apply nonlinear system identification possible.

Considering N input, an output, the system is divided into M fuzzy sets to each input, for a band model Sugeno, it's $M^N$ fuzzy rules:

If $x_1$ is $A_{1i_1}$ , $x_2$ is $A_{2i_2}$ , … , $x_N$ is $A_{N1}$ ,

[4] SHI Xiaoxin , Xia Jun. Integrated Environmental Impact Assessment Model of gray levels [J]. *Shanghai Environmental Science*,1996, 15 (10): 9～12

XU Zaozhong. Chen Liang. Fuzzy-Gray theory in the urban environment of Air Quality Evaluation of the City of Hubei Province to study atmospheric environmental quality assessment as an example[J]. *Journal of Environmental Science*,1997, 17 (3): 260～267

[5] SHI Chun, Guo Zhongyang, Xu Shiyuan. Artificial neural network in the coastal region of complex systems environment[J]. *Pollution and Control*, 2002, 24(5): 300-302.

Widrow B, et al.30 Years of Adaptive Neural Networks: Perception. Madaline and B-P[J].*Proc. IEEE*, 1990, 78 (9): 1415～1442

[6] HOU Zhixiang, Shen Quntai, Li Heqing.Based on ANFIS identification of nonlinear systems [J]. *System Engineering and Electronic Technology*,2005,27(1): 108-110

[7] J. S. Roger Jang .Fuzzy modeling using generalized neural networks and Kaman filter algorithm [C] . America: In Proceedings of the Ninth National Conference on Artificial Intelligence (AAA 91), 1991,762-797

J.-S.Roger Jang . ANFIS: Adaptive-Network-based Fuzzy Inference Systems [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, 23(03) : 665-685

Then $y_{i_1 i_2 ......i_N} = {}_{k=1}^{N} \; p_{i_1 i_2 ......i_N}(k)x_k + q_{i_1 i_2 ......i_N}$ ,

$i_1, i_2, ......, i_N \in \{1,2,......M\}$

Figure 1 is N = 2, M = 3 of ANFIS network structure. Other sequentially.

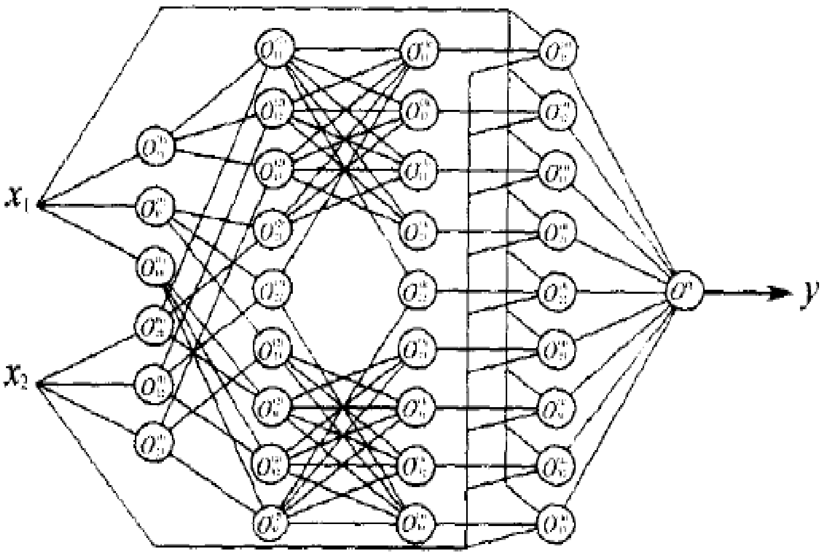


**Figure 1 N = 2, M = 3 of ANFIS network structure**

The first layer: the output function of each node:

$$O_{ki_k}^{(1)} = \mu_{ki_k}(x_k), \qquad k = 1,2,......, N, i_k = 1,2,......M \qquad (1)$$

Where: $x_k$ is paragraph ( k , $i_k$ ) nodes input, $A_{ki_k}$ is language fuzzy sets (for example: when M =3, and often is "small", "middle" or "large"), apparently $O_{ki_k}^{(1)}$ belonging to the corresponding input $x_k$ fuzzy sets membership, here membership in $A_{ki_k}$ is the parameters of membership function, is solely by the membership function determined by the limited parameters, such as bell-shaped function:

$$\mu_{A_{ki}}(x) = (1 + \left|(x - c_{i_k})/a_{ki_k}\right|^{2b_{ki_k}})^{-1} \qquad (2)$$

It is decided by three parameters $\{a_{ki_k}, b_{ki_k}, c_{ki_k}\}$.

The second layer: the layer has $M^N$ nodes, the output of each node is all imported, but the output of the layer is not necessarily in the form of the product, in fact, as long as it is representative of fuzzy "and", any T-modulus operator can be.

$$O_{i_1 i_2 ...i_N}^{(2)} = {}_{k=1}^{N} \; O_{ki_k}^{(1)} = {}_{k=1}^{N} \; \mu_{ki_k}(x_k)$$
$$i_1, i_2, ......, i_N = 1,2,......, M \qquad (3)$$

The third layer: the layer has the same number of nodes with the first layer, its output:

$$O_{i_1 i_2 ...i_N}^{(3)} = O_{i_1 i_2 ...i_N}^{(2)} / {}_{i_1 i_2 ...i_N = 1}^{M} \; O_{i_1 i_2 ...i_N}^{(2)} = ({}_{k=1}^{N} \mu_{A_{ki_k}}(x_k)) / {}_{i_1 i_2 ...i_N}^{M} \; {}_{k=1}^{N} \; \mu_{ki_k}(x_k)$$
$$i_1, i_2, ......, i_N = 1,2,......, M \qquad (4)$$

The fourth layer: The layer has the same number of nodes with the third layer, each node point is an adaptive node, the output:

$$O^{(4)}_{i_1 i_2 \ldots i_N} = O^{(3)}_{i_1 i_2 \ldots i_N} y_{i_1 i_2 \ldots i_N} = ((\,{}_{k=1}^{N} \mu_{ki_k}(x_k)) / {}_{i_1 i_2 \ldots i_N}^{M}\,{}_{k=1}^{N} \mu_{ki_k}(x_k))(\,{}_{k=1}^{N} p_{i_1 i_2 \ldots\ldots i_N}(k) x_k + q_{i_1 i_2 \ldots\ldots i_N}) \quad (5)$$

$$i_1, i_2, \ldots\ldots, i_N = 1,2,\ldots\ldots,M$$

Type: $p_{i_1 i_2 \ldots\ldots i_N}(k)$ *and* $q_{i_1 i_2 \ldots\ldots i_N}$ are adjustable parameters.

The fifth layer: Only one node, it is the output of all input Combined. That is the network output:

$$y = O^{(5)} = {}_{i_1, i_2, \ldots, i_N = 1}^{M} O^{(4)}_{i_1 i_2 \ldots i_N} \quad (6)$$

## 2.2 Mixed learning algorithm

To identify the network parameters and to improve learning speed, J. Roger S. Jang proposed mixed learning algorithm.

If we consider only an adaptive network output, said:

O=F(i, S) (7)

Type: i is input vector; S is parameters set; F is the overall function by the adaptive network achieved.

If there function H make composite function $H \circ F$ certain elements in S is linear, then these elements can be adopted by the least square method identification. If parameters set S can be divided into two sets:

$$S = S_1 \oplus S_2 \qquad (type: \oplus\, said \quad direct\, combined) \quad (8)$$

Here: S = the parameters set； $S_1$ = Premise (non-linear) parameters;

$S_2$ = Conclusion (linear) parameters

Make $H \circ F$ is linear to elements of $S_2$, Right-through formula (7) imposing H operator, is:

$$H(O) = H \circ F(B_i, S) \quad (9)$$

H(·)and F(·,·), each is unit function and fuzzy inference system function.

It is linear to elements of $S_2$. Give the elemental values of $S_1$, through formula (9), give the equation P training data input, and get a matrix equation:

$$A\theta = y \quad (10)$$

Type: $\theta$ is unknown vector, its elemental is parameters in $S_2$. Clearly this is a standard linear least squares problems, make $\| A\theta - y \|^2$ the smallest, the θ is the optimal solution for least-squares estimation $\theta^*$:

$$\theta^* = (A^T A)^{-1} A^T y \quad (11)$$

Type: $A^T$ is purchase of A, if $A^TA$ is not singular, $(A^T A)^{-1} A^T$ is pseudo –inverse of A. It can also use recursive least squares method $\theta^*$.

Mixed learning algorithm using batch mode, each cycle including a forward transfer process and a reverse transfer process. The information the forward transfer process, using input data and function signal transfer forward, calculating the output of each node layer by layer, until a calculated A and y in matrix A $\theta$ = y. Signal to move forward, the parameters in $S_2$ can by pseudo-inverse formula or by the least square method in $\theta^* = (A^T A)^{-1} A^T y$ for identification, until calculated error indicator of each training data. In the process of signal reverse transmission, error of the output from the reverse transfer to the input terminal, which the parameters in $S_1$ can be updated by the steep decline. $S_1$ parameters after

Nth learning training were adjusted to the smallest of errors.

These mixed learning rules not only reduce search space dimension of the steep decline, which can greatly reduce the time required for convergence. Mixed learning algorithm reduce the original pure back-propagation algorithm search space series, it greatly enhanced the convergence of parameters.

## 2.3 With conjugate gradient method to improve

Conjugate gradient method for unconstrained optimization problems of an important and effective method, and (to be) compared to Newton's law, it has low calculation cost. The conjugate gradient method do not must calculate or store information on the second derivative and has the Second order method function, such as wider use of the multi - variable optimization. Since 1964, Fletcher and Reeves first proposed nonlinear conjugate gradient method, the parameters $\beta_k$ of the method has drawn much attention. One, $\beta_k^{HS} = \dfrac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$ $(HS\ formula\ )$ [8] numerical good performance, particularly suitable for large-scale unconstrained optimization issue.[9] Convex quadratic function in the exact line search has its limited convergence, but if the objective function strictly non-convex quadratic function, even in the precise line search may not have a limited convergence, global convergence can not be guaranteed[10] too. In recent years, the literature[11] gives HS formula new exploration, on the right for formula amendment and global convergence, and so on. Literature[12] on the basis of the above literature, combining the characteristics and advantages of the structure of HS formula give $\beta_k$ a new method:

$$\beta_k^{\theta} = \frac{\left\| g_{k+1} - g_k \right\|^2}{d_k^T(\theta g_{k+1} - g_k)} (\theta\sigma < 1) \qquad (12)$$

In consideration of the mixed learning algorithm, the updated formula of the premise parameters and conclusions parameters is separated, so this paper adopted conjugate gradient method above to enhance the premise parameters learning speed. Experiments proved very effective indeed.

# 3. EXAMPLES OF AIR POLLUTION FORECAST

Now improved ANFIS for the District of Wuxi LAKESIDE District Cai Xing village NOx concentrations of atmospheric pollutants forecast. In January and February 2007, to choose 150 data points as a training sample, in March and April data as 150 samples tested. The original data are from

---

8 HESTEN ESM R, ST IEFEL E L. Methods of conjugate gradients for solving linear systems [J]. *J Res Nat BurStandards Sect*, 1952, 49: 4092436.

9 AL-BAAL IM. Descent property and global convergence of the Fletcher-Reeves method with inexact line search[J]. *IMA J Numer Anal*, 1985, 5: 1212124.

10 GILBERT J C, NOCEDAL J. Global convergence of conjugate gradient methods for optimization [J ]. *S IAM Journal of Optimization*, 1992, 2: 21-42.

11 DAI Zhifeng, Chen Lanping. A mixture of HS-DY conjugate gradient method [J]. *Computational mathematics*, 2005, 27(4): 429-436

SHI zhenjun. Improved HS conjugate gradient method and its global convergence[J]. *Computational mathematics*, 2001, 23(4): 393-406

DAI Hong, Yuan Yaxiang. *Nonlinear conjugate gradient method*[M]. Shanghai: Shanghai Science and Technology Press, 2001: 30-50

12 DONG Xiaoliang, Li Bingliang, Tang Qinggan. A search of the Wolfe conjugate gradient method and its global convergence [J]. *Guangxi Science*, 2007, 14 (1) : 44～ 46

Wuxi LAKESIDE District Stations. Literature[13] through the analysis obtained that the case of fixed region, we can think of hour NOx content by following factors: atmospheric stability, Reuters for observation, temperature, cloud cover, wind direction, wind speed, before a moment of pollutants content.

Representatives of various factors symbols are as follows:

*P*( *t*) —t moment of atmospheric stability grade, A, B, C, D and E all levels were recorded as 1, 2, 3, 4, 5; *D*( *t*) —t moment of pollutants content;

*D*( *t* + 1) —( *t* + 1) moment of NOx content; *H*( *t* + 1) —( *t* + 1) moment of Reuters;

*T*( *t* + 1) —( *t* + 1) moment of temperature; *C*( *t* + 1) —( *t* + 1) moment of cloud cover;

*WS* ( *t* + 1) —( *t* + 1) moment of wind speed; *WD*( *t* + 1) —( *t* + 1) moment of wind direction;

Among them, that the t said the t hours, (t + 1) said the (t + 1) hours. It is seven inputs, respectively P (t), D (t), H (t + 1), T (t + 1), C (t + 1), WS (t + 1), WD (t +1) and an output D (t + 1).

As usually for n input fitting data problems, we often need 10n data point to be a good model, if n is quite strong, we will need quite a lot of data points. For example, the importation of seven input, ideal circumstances, we should have 107 data points. In response to this situation, we consider reducing input dimension and lower error method. Choose two most relevant input, need to establish $C_2^7 = 21$ Fuzzy Logic Toolbox anfis in order to train the fuzzy model. All models in figure 2 shows. The results of each ANFIS training model are the result of Step 5. In figure 2, curve in solid lines is the root mean square error of training, dashed line is the root mean square error testing.

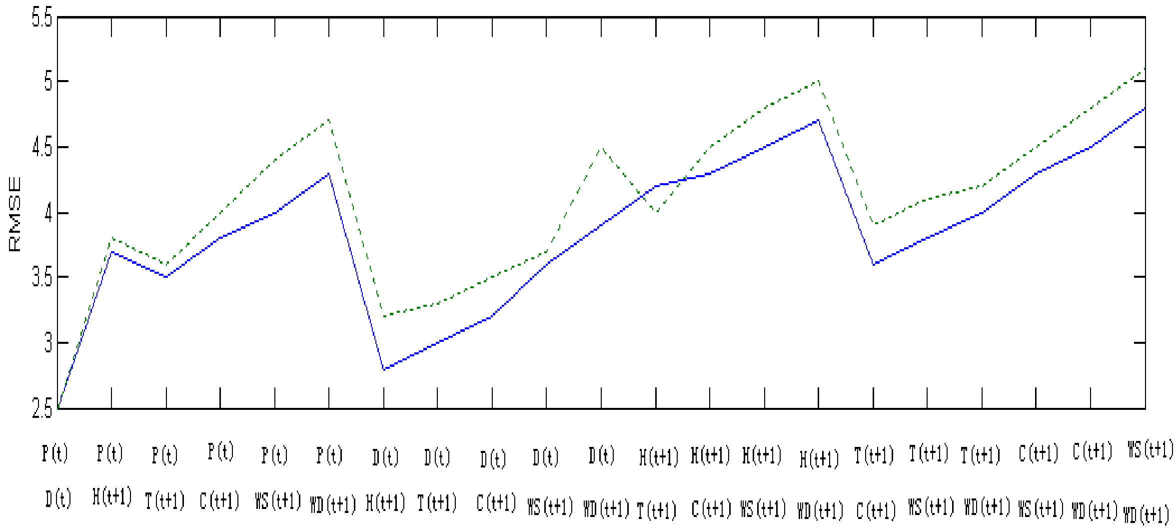


**Figure 2. The training and testing of the root-mean-square error**

Figure 2 can be seen in the two most relevant input variables should be P(t)and D(t). Now anfis order can be used to further train the model performance. Figure 3 shows the training error curve after 100 steps training, which curve in solid lines is the root-mean-square error of training, dashed line is the root mean square error testing. When the error testing Minimum, it is the best model.

[13] WANG Jian, HU Xiaomin, ZHEN Longxi, LIU Zhen. Based on the BP model of atmospheric pollution prediction method of study [J]. *Environmental science study*,2002, 15（5）: 62-64

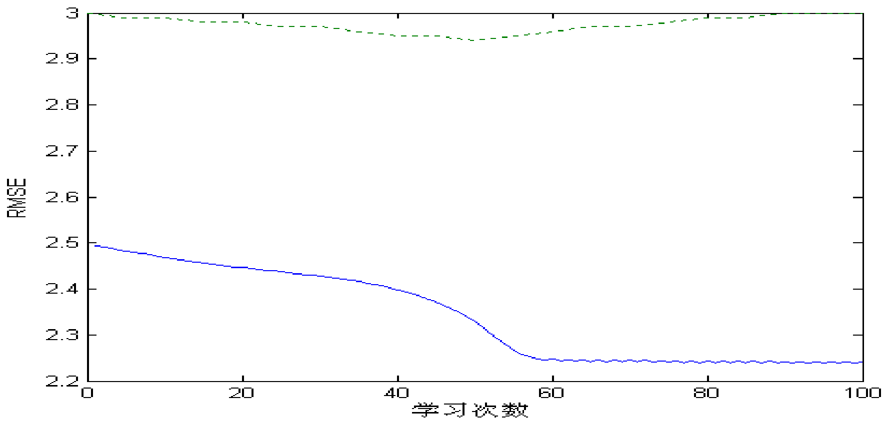


**Figure 3. Error after 100 infantry training**

Table 1 identifies the part of the forecasting data points, we can see the measured and predicted values With better.

**Table 1. Hour NOx in the forecast results**

| Date | Time | **NOx** predictive value y (mg/m$^3$) | **NOx** measured value y'(mg/m$^3$) | The absolute error | Relative error γ ( % ) |
|---|---|---|---|---|---|
| … | … | … | … | … | … |
| March 19, 2007 | 7：00-8：00 | 0.142 | 0.134 | 0.008 | 5.97 |
| | 10：00-11：00 | 0.055 | 0.067 | -0.012 | 17.91 |
| | 14：00-15：00 | 0.101 | 0.099 | 0.002 | 2.02 |
| | 17：00-18：00 | 0.098 | 0.090 | 0.008 | 8.89 |
| March 20, 2007 | 7：00-8：00 | 0.142 | 0.137 | 0.005 | 3.65 |
| | 10：00-11：00 | 0.064 | 0.054 | 0.01 | 18.52 |
| | 14：00-15：00 | 0.048 | 0.056 | -0.008 | 14.29 |
| | 17：00-18：00 | 0.110 | 0.109 | 0.001 | 0.92 |
| March 21, 2007 | 7：00-8：00 | 0.072 | 0.066 | 0.006 | 9.09 |
| | 10：00-11：00 | 0.093 | 0.100 | -0.007 | 7.00 |
| | 14：00-15：00 | 0.097 | 0.093 | 0.004 | 4.30 |
| | 17：00-18：00 | 0.039 | 0.048 | -0.009 | 18.75 |
| March 22, 2007 | 7：00-8：00 | 0.053 | 0.047 | 0.006 | 12.77 |
| | 10：00-11：00 | 0.122 | 0.142 | -0.02 | 14.08 |
| | 14：00-15：00 | 0.056 | 0.062 | -0.006 | 9.68 |
| | 17：00-18：00 | 0.054 | 0.053 | 0.001 | 1.89 |
| March 23, 2007 | 7：00-8：00 | 0.117 | 0.125 | -0.008 | 6.40 |
| | 10：00-11：00 | 0.070 | 0.063 | 0.007 | 11.11 |
| | 14：00-15：00 | 0.060 | 0.055 | 0.005 | 9.09 |
| | 17：00-18：00 | 0.092 | 0.095 | -0.003 | 3.16 |
| … | … | … | … | … | … |

## 4. CONCLUDING REMARKS

This paper studies the functions of a class and Fuzzy Inference System equivalent ANFIS adaptive network. And through mixed learning algorithm to reduce the original pure back-propagation algorithm search space series, improving the convergence rate. In this paper, based on its parameters and the parameters inadequate and slow using convergence conjugate gradient method to improve, greatly enhancing its performance. In the simulation test use improved ANFIS for NOx pollution prediction, and a bold choice of the two most relevant input technology can be found in lower error indicators, and forecasting and measured fit better. It shows that the method shown in its unique advantages, and will have pollution prediction of the potential value of good, worthy of promotion and application.