Chinese J. Chem. Eng., 12 (2) 234 – 239 (2004)

Fuzzy Neural Network Model of 4-CBA Concentration for Industrial Purified Terephthalic Acid Oxidation Process*

LIU Ruilan(刘瑞兰), SU Hongye(苏宏业), MU Shengjing(牟盛静), JIA Tao(贾涛), CHEN Weiquan(陈渭泉) and CHU Jian(褚健)**
National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, China

Abstract A fuzzy neural network (FNN) model is developed to predict the 4-CBA concentration of the oxidation unit in purified terephthalic acid process. Several technologies are used to deal with the process data before modeling. First, a set of preliminary input variables is selected according to prior knowledge and experience. Secondly, a method based on the maximum correlation coefficient is proposed to detect the dead time between the process variables and response variables. Finally, the fuzzy curve method is used to reduce the unimportant input variables. The simulation results based on industrial data show that the relative error range of the FNN model is narrower than that of the American Oil Company (AMOCO) model. Furthermore, the FNN model can predict the trend of the 4-CBA concentration more accurately.
Keywords purified terephthalic acid, 4-carboxybenzaldchydc, fuzzy neural network, soft sensor, input variables selection, fuzzy curve, dead time detection

1 INTRODUCTION

Purified terephthalic acid (PTA) is a kind of important raw material for polyester production widely used in textile and packaging industries. It is produced by catalytic oxidization of PX (paraxylene) followed by subsequent purification of the crude terephthalic acid by selective hydrogenation. 4-carboxybenzaldchydc (4-CBA) is one of the byproducts and its concentration is an important quality criteria in PTA process. Reference[1] showed that the lower the 4-CBA concentration is, the more the energy costs. Thus it is necessary to control the 4-CBA concentration of the oxidation unit on-line for saving energy and ensuring the purity of PTA.

In practice, it is rarely the case that the 4-CBA concentration is directly used as a controlled variable, because on-line accurate measurement of the 4-CBA concentration is difficult. Usually, the 4-CBA concentration is analyzed only three times each day by spectroscopic analyzer because of the cost involved. Since the spectroscopic analysis is a laboratory technique with obvious time delay, the analytical values of 4-CBA concentration are not available for real-time control adjustment if required. In order to monitor 4-CBA concentration on-line an alternative method is to build a soft sensor. By soft sensor technique the 4-CBA concentration can be inferred from other measured process variables such as temperature, pressure, flow rate, etc, and their relationship which can be constructed by regression methods and the first principle modeling methods. If the 4-CBA concentration can be estimated accurately through the soft sensor, we can apply an advanced control strategy such as dynamical matrix control (DMC) to directly control the concentration for improving the production performance.

A simple flowsheet of the industrial PTA oxidation process based on American Oil Company (AMOCO) technics is shown in Fig. 1. Paraxylene, acetic acid solvent, promoter, and catalyst are continuously metered into the feed mixing tank. The residence time is approximately 25 minutes. The mixed stream is pumped into the reactor, and the air is fed to the reactor through four inlets. The oxidation reaction is conducted in two stages. The first stage is the agitated oxidation reactor, while the second stage is the agitated first crystallizer. Exothermic heat of reac-

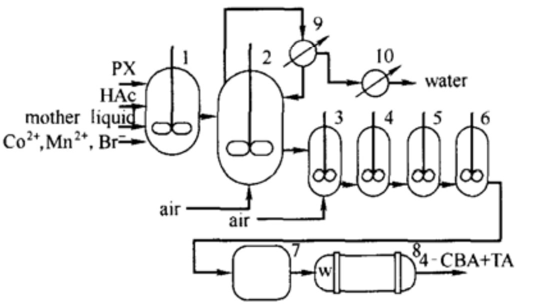

Figure 1 Simple flowsheet of the industrial PTA oxidation process
1—mixing tank; 2—reactor; 3—first crystallizer; 4—second crystallizer; 5—third crystallizer; 6—buffer tank; 7—vacuum filter; 8—crude TA dryer; 9,10—cooler

Received 2003-07-21,accepted 2003-12-30.
* Supported by the National Outstanding Youth Science Foundation of China (No. 60025308).
** To whom correspondence should be addressed. E-mail: rlliu@iipc.zju.edu.cn

tion is removed by condensing the vapor of reaction solvent. A portion of this condensate is withdrawn to control the water concentration in the reactor, and the remainder is refluxed to the reactor.

Reactor effluent is depressurized and cooled to filtering conditions in a series of three crystallizing vessels. Air is fed to the first crystallizer for additional oxidation reaction. The precipitated terephthalic acid is recovered by filtering and drying. The solids are conveyed to the purification section feed silos for additional processing. The more detailed discussion about the oxidation theory and mechanism of paraxylene can be found in the references[1-3].

There are more than 30 patents about PTA oxidation process and its oxidation reactor in the past decade[4]. Many empirical models are proposed[2,5] which are very simple, but the parameters of these empirical models are different under different operating condition and often suitable to a limited operation region. In addition, Cao *et al.* and Wang proposed the first principle models separately based on bench-scale laboratory results[1,3]. However, these models were only verified by a few industrial data, and many of industrial application problems were not settled[6].

In this work, we propose a fuzzy neural network to model the nonlinear relationship between 4-CBA concentration and measured process variables in order to predict the 4-CBA concentration in a real time manner.

2 FUZZY NEURAL NETWORK

Fuzzy neural network (FNN) can be thought of as a nonlinear mapping between input variables and output variables. Based on fuzzy neural network a complex nonlinear model can be built and is suitable to different operation region in industrial processes. The architecture of four layers' fuzzy neural network[7] with m inputs and one output is shown in Fig. 2, The four layers are input layer, fuzzification layer, inference layer and defuzzification layer respectively. There are m neurons connected with m input variables in the first layer, $m \times R$ neurons in the fuzzification layer, R neurons in the inference layer and one neuron in the output layer. Each m neurons in the fuzzification layer represents one fuzzy rule, so there are R rules in total.

The ith rule is,

R^i: If x_1 is $A_{i,1}$ and $\cdots$ and x_j is $A_{i,j}$ and $\cdots$ and x_m is $A_{i,m}$,

Then y is B_i.

where x_j is the jth input variable, y is an output variable, $A_{i,j}$ is a fuzzy set in the input space and B_i is the ith fuzzy set in the output space.

We assume that $A_{i,j}$ and B_i have Gaussian type membership function as follows,

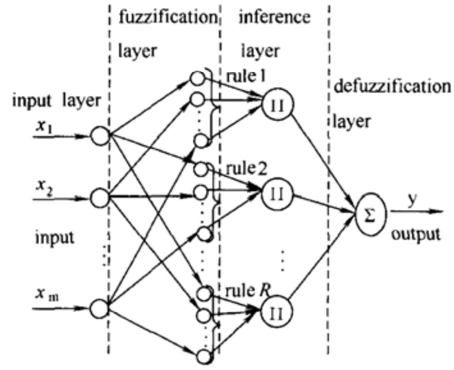

Figure 2 The architecture of fuzzy neural network

$$\mu A_{i,j}(x_j) = \exp\left[-\left(\frac{x_j - a_{i,j}}{c_{i,j}}\right)^2\right] \tag{1}$$

$$\mu B_i(y) = \exp\left[-\left(\frac{y - b_i}{d_i}\right)^2\right] \tag{2}$$

where $a_{i,j}$ and $c_{i,j}$ ($i = 1, 2, \cdots, R$; $j = 1, 2, \cdots, m$) represent the center and width of the input membership functions respectively; b_i and d_i reprensent the center and width of the output membership functions, respectively. On the basis of multiplicative inference, we get

$$w_i = \prod_{j=1}^{m} \mu A_{i,j} \tag{3}$$

The inference result coming from R rules follows a standard center of gravity formula,

$$y_{\text{out}} = \left(\sum_{i=1}^{R} d_i b_i w_i\right) \Big/ \left(\sum_{i=1}^{R} d_i w_i\right) \tag{4}$$

and the learning of FNN is accomplished by adjusting the input/output widths and the centers of membership functions and follows backpropagation (BP) algorithm[8,9]. In this study, we use an Euclidean distance, that is

$$E = \frac{1}{2}(y_{\text{out}} - y)^2 \tag{5}$$

where E is the error, y_{out} is the actual output value and y is the target output value.

We only take $a_{i,j}$ as an example for brevity. By using the BP algorithm, the following update formula can be derived

$$a_{i,j}(k+1) = a_{i,j}(k) - \eta\frac{\partial E}{\partial a_{i,j}} + \alpha[a_{i,j}(k) - a_{i,j}(k-1)] \tag{6}$$

$$\frac{\partial E}{\partial a_{i,j}} = \frac{\partial E}{\partial y_{\text{out}}}\frac{\partial y_{\text{out}}}{\partial w_i}\frac{\partial w_i}{\partial \mu A_{i,j}}\frac{\partial \mu A_{i,j}}{\partial a_{i,j}} \tag{7}$$

where η is the learning rate and α represents the momentum coefficient.

It is important to initialize the parameters of fuzzy neural network because BP algorithm is sensitive to the initial parameters. In order to select a set of suitable parameters, fuzzy C means clustering (FCM) algorithm[10-12] is applied to initialize parameters $a_{i,j}$ and b_i. The initial values $c_{i,j}$ and d_i are selected stochastically in open set (0, 1).

3 INPUT VARIABLES SELECTION AND SAMPLE SET COLLECTION

3.1 Preliminary selection of process variables

Hundreds of process variables are recorded respectively one time per 30 seconds by the distribution control system (DCS) in PTA process. When building the soft sensor model, the selection of an appropriate subset from these variables is very important. Too many unimportant variables included in the soft sensor model will lead to the difficulty of training and usage. On the other hand, the accuracy of model cannot be guaranteed if some important variables are not included.

According to the prior knowledge and experience, twelve variables are preliminarily selected, including flow rate, reaction pressure, temperature, solvent ratio in the reactor and catalyzer liquid level, *etc.* These twelve process variables are measured in the reactor and the first crystallizer. After collecting the sample data, a fuzzy curve method[7,13] is applied to reduce the input variables in section 3.4.

3.2 Dead time detection

In this section, we propose a new method based on the maximum correlation coefficient to detect dead time between the process variables and response variables. Dead time is the delay between the time when the value of a process variable changes and the time when the dependent variable begins to change in response which depends on the structure and scale of the production equipment. It must be considered to align sample data to guarantee the performance of the soft sensor to be modeled.

In practice there are two main operation state including normal operation and load down operation in the PTA process. In the case of normal operation, the change range of the 4-CBA concentration is relatively narrow. However, in the case of load down operation step disturbances are manually introduced only to one variable each time and the change range of the 4-CBA concentration is very broad. Thus we can especially increase the sample frequency of 4-CBA concentration during load down. For the reason of cost and other factors, we analyzed the 4-CBA concentration with interval of one hour at most.

In Fig. 3 a total of 11 samples of 4-CBA concentration collected are shown from 10 to 20 o'clock. On the other hand the process variables were re-sampled each 6 minutes and 141 samples were collected from 6 o'clock to 20 o'clock. Only process variable FIC114, which is the feed flow rate to reactor, corresponding to x_2 in the Table 1, is shown in Fig. 3 for simplicity. It is found that the obvious time delay exists between the change of 4-CBA concentration and the change of FIC114. For convenience of analysis, all the sample data were normalized to unit length.

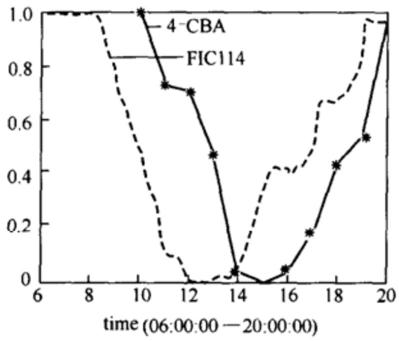

Figure 3 Obvious time delay exits between the change of FIC114 and the change of the 4-CBA concentration

Dead time τ is calculated by the following formulas

$$\tau = \arg\max_{l}[C_{x,y}(l)] \times T \tag{8}$$

$$C_{x,y}(l) = \frac{R_{x,y}(l)}{\sqrt{R_x(l) \times R_y}} \tag{9}$$

$$R_{x,y}(l) = E\{[x(t-l) - \mu_x(l)][y(t) - \mu_y]\} \tag{10}$$

$$R_x(l) = E\{[x(t-l) - \mu_x(l)]^2\} \tag{11}$$

$$R_y = E\{[y(t) - \mu_y]^2\} \tag{12}$$

where $x(t-l)$ is the value of the process variable at time $(t-l)$, $y(t)$ is the value of the response variable at time t, $\mu_x(l)$ and μ_y represent the mean value of the variables $x(t-l)$ and $y(t)$, respectively, $l = 0, 1, \cdots, 30$ denotes the number of the time delay , t denotes the sample moment of the response variable and $T = 6$ minutes, which is the sampling interval of the process variable.

According to Eq. (9), correlation coefficients between the process variable and 4-CBA concentration with different time delay are calculated. We can determine the dead time by searching the value of l corresponding to the maximum correlation coefficients shown in Eq. (8).

Dead time of different variable is shown in Table 1 according to the above formulas. It is worth to notice that the calculation of correlation coefficient between the process variable and response variable ignores the effects of the other process variables. Consequently,

Table 1 The 12 inputs and one output of the 4-CBA concentration and their dead time

No.	Variable	Dead time, min	Sample frequency
	inputs		
1	x_1: flow rate of paraxylene to feed mixing tank	144	30 s
2	x_2: feed flow rate to the reactor	144	30 s
3	x_3: catalyst concentration	150	30 s
4	x_4: level of the reactor	114	30 s
5	x_5: reactor temperature	102	30 s
6	x_6: vent O_2 concentration from the reactor	108	30 s
7	x_7: reactor condenser to water withdrawal	108	30 s
8	x_8: total water withdrawal	120	30 s
9	x_9: temperature of the first crystallizer	18	30 s
10	x_{10}: vent O_2 concentration from the first crystallizer	84	30 s
11	x_{11}: vent CO_2 concentration from the reactor	150	30 s
12	x_{12}: vent CO concentration from the reactor	168	30 s
	output		
13	y: 4-CBA concentration in the crude TA	0	8 h

all of the dead time shown in Table 1 is approximate values. The sample frequency shown in Table 1 denotes the sampling interval of the process variables in the case of normal operation. In fact, the dead time determined by the maximum correlation coefficient is consistent with the PTA process.

3.3 Data collection and preprocessing

As mentioned in section 3.1, 4-CBA concentration is sampled only three times a day while the process variables are recorded almost three thousand times a day. In other words, three samples at most are collected for training fuzzy neural network each day. The data set with 129 samples was collected according to the dead time of process variables and the sampling moment of 4-CBA concentration. That is,

$$X = [x_1(t-\tau_1)\cdots x_i(t-\tau_i)\cdots x_{12}(t-\tau_{12})] \quad (13)$$

$$Y = [y(t)] \quad (14)$$

where X and Y represent process variables values and 4-CBA concentration collected, respectively, τ_i, $i = 1, 2, \cdots, 12$, is the dead time of the ith process variable and t is the sampling moment of the 4-CBA concentration.

Before training a fuzzy neural network, it is necessary to pre-process data to identify bad outliers and filter noise. Bad outliers can result from sensor failure or misreading from lab tests. We delete samples that have outliers identified by using prior knowledge. Noise in the process variables can be filtered by average filtering method defined by

$$x_i(t-\tau_i) = \frac{1}{20}\sum_{j=1}^{20} x_i(t-\tau_i+10-j) \quad (15)$$

The right term of Eq. (15) is applied to calculate the average value of the 20 samples around τ_i.

3.4 Reduce input variables using fuzzy curve method

In section 3.1, we only select the possible input variables set based on the prior experience. Some unimportant variables may be included in the candidate set. In order to simplify the final model, we use the fuzzy curve method to reduce the unimportant variables based on the practical sample data.

Considering a system that has m possible extraneous inputs $X = [x_1 \cdots x_i \cdots x_m]$ and one output $Y = y$. The number of training data points is n. Let $x_{k,i}$ be the ith variable in the kth data point, and y_k is the output value in the kth data point. The fuzzy curve method is briefly described as follows.

The fuzzy membership function $\phi_{k,i}(x_i)$ for each input x_i is defined by,

$$\phi_{k,i}(x_i) = \exp\left[-\left(\frac{x_{k,i}-x_i}{b}\right)^2\right], k = 1, 2, \cdots, n \quad (16)$$

where b is the width of the fuzzy membership function, which is typically taken as about 20% of the length of the input interval of x_i.

For n data points, we have n fuzzy rules for each inputs and the kth rule is

R^k: If x_i is $\phi_{k,i}(x_i)$, then y is y_k.

We use the center of gravity algorithm for defuzzification to produce a fuzzy curve c_i for each input x_i by

$$c_i(x_i) = \frac{\sum_{k=1}^{n} \phi_{k,i}(x_i) y_k}{\sum_{k=1}^{n} \phi_{k,i}(x_i)} \quad (17)$$

The range of corresponding c_i can be obtain by

$$Rc_i = \max(c_i) - \min(c_i) \quad (18)$$

On the basis of the value of Rc_i, the importance of the input variables can be recognized. The larger the range of Rc_i is, the more important this input variable is.

According to Eqs. (16)—(18) the ranges of the fuzzy curves, Rc_i, $i = 1, 2, \cdots, 12$, are shown in Table 2.

As shown in Table 2, the range of the fuzzy curve for x_7 and x_{12} is smaller than that of the other variables. Thus, we delete these two variables for simplifying the model. However, this does not mean that these two variables have no effect on the 4-CBA concentration. It is just because that the range of these two variables is too small to produce any obvious error in the model prediction. In addition, if high correlation coefficient exits among several process variables, some of these variables should be deleted to reduce the prediction variance of the model. In this process, the correlation coefficient between x_1 and x_2 is 0.998, which are almost completely linear dependent. Based on the practical experience x_1 , rather than x_2, is deleted despite the larger range of the fuzzy curve between x_1 and y shown in Table 2. Finally, 9 input variables including x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10}, and x_{11} are selected to build the soft sensor based on fuzzy neural network.

Table 2 The range of the fuzzy curves

Input variables	Rc_i	Input variables	Rc_i
x_1	0.7130	x_7	0.0918
x_2	0.7109	x_8	0.7041
x_3	0.6927	x_9	0.2182
x_4	0.5046	x_{10}	0.3839
x_5	0.5966	x_{11}	0.6981
x_6	0.5353	x_{12}	0.0864

4 SIMULATION RESULTS

Using Matlab 5.3 as simulation tool, we construct a fuzzy neural network model for predicting the 4-CBA concentration. A total of 129 sample data including normal operation data and load down operation data from a practical PTA oxidation process are collected. The data are divided into two sets, one set has 100 samples used for training FNN, and the other set has 29 samples used for testing. FCM clustering algorithm and gradient descent algorithm are used to initialize and train the parameters of the fuzzy neural network respectively. After training FNN with different number of fuzzy rules and learning rate, it is found that the most suitable rule number is 3, and the learning rate is 0.16. The training and testing relative errors after 350 iterations are shown in Fig. 4. For comparison, the empirical regression model of AMOCO[2] is applied to the same data set. Under the same condition as in the FNN approach, the training and testing relative error are also given in Fig. 4.

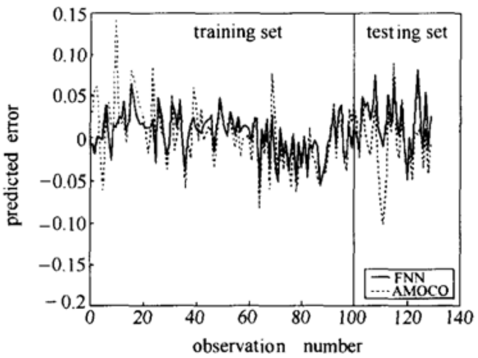

Figure 4 Predicted relative errors for 4-CBA concentration using the fuzzy neural network method and AMOCO regression model method

Table 3 lists the performance of two models including the maximum relative error, the minimum relative error and the root-mean-square-error (RMSE) in detail. In the training set the maximum relative error of FNN model is about 6.5% while that of AMOCO model is up to 14.5%; in the testing set the maximum relative error of FNN model is about 7% while that of AMOCO model is close to 10%. In addition, the RMSE of FNN model in all data set is about 0.035 while that of the AMOCO model is about 0.04. Though the RMSE's of the two models are close to each other, the relative error range of the FNN model is narrower. Therefore, the FNN model can predict the trend of 4-CBA concentration more accurately.

5 CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

This paper presents a fuzzy neural network model to predict the 4-CBA concentration of the oxidation unit in PTA process. Several technologies are used to deal with the process data before modeling. Suitable input variable subset has been selected according to the prior knowledge, experience and fuzzy curve method. The maximum correlation coefficient based method has been proposed to detect the dead time

Table 3 Training and testing performance for 4-CBA concentration using the FNN method and AMOCO model

Model	Training set			Testing set		
	max. rel. error	min. rel. error	RMSE	max. rel. error	min. rel. error	RMSE
FNN	0.0653	−0.0603	0.0256	0.0788	−0.0509	0.0353
AMOCO	0.1427	−0.083	0.0376	0.0877	−0.1035	0.0404

between the process variable and response variable. The simulation results show that the performance of the FNN model is better than that of the AMOCO model.

Further directions for research include: collecting samples as many as possible to improve the model accuracy; applying the FNN model to the PTA process to estimate the 4-CBA concentration online

NOMENCLATURE

$A_{i,j}$	the fuzzy set in the input space
$a_{i,j}$	the center of the input membership function
B_i	the fuzzy set in the output space
b	the width of the fuzzy membership function
b_i	the center of the output membership function
$C_{x,y}(l)$	the correlation coefficient between x and y when delay time is lT
$c_i(x_i)$	the fuzzy curve for each input x_i
$c_{i,j}$	the width of the input membership function
d_i	the width of the output membership function
E	the square error between the output of the fuzzy neural network and the target output value
l	the number of delay time
m	the number of the process variables
n	the number of training data
Rc_i	the range of corresponding $c_i(x_i)$
$R_x(l)$	the variance of the x when delay time is lT
$R_{x,y}(l)$	the covariance between x and y when delay time is lT
R_y	the variance of the y
T	the sampling interval of the process variable
t	the sampling moment of the process variable
X	the data set of the process variables
$x_i(t-\tau_i)$	the value of the ith process variable when delay time is τ_i
x_j	the input variable of the fuzzy neural network
$x_{k,i}$	the ith variable value in the kth data point
Y	the data set of the 4-CBA concentration
y	the output variable of the fuzzy neural network
y_k	the output value in the k th data point
y_{out}	the output of the fuzzy neural network
α	the momentum coefficient
η	the learning rate
τ	the dead time
τ_i	the dead time of the ith process variable
$\phi_{k,i}(x_i)$	the fuzzy membership function for each input x_i
$\mu A_{i,j}(x_j)$	input membership function
$\mu B_i(y)$	output membership function

REFERENCES

1 Gao, G., Servida, A., Pisu, M., Morbidelli, M., "Kinetics of *p*-xylene liqui-phase catalytic oxidation", *AIChE J*, **40**, 1156—1166 (1994).

2 Lindahl, H.A., Abrams, K.J., Stark, L.E., Zeitlin, M.A., "Method and apparatus for controlling the manufacture of terephthalic acid to control the level and variability of the contaminant content and the optical density", U.S. Pat., 4835307 (1989).

3 Wang, L.J., "Studies on the kinetics of *p*-xylene oxidation and the reactor simulation", Dissertation for Master Degree, Zhejiang Univ., China (2001). (in Chinese)

4 Li, X., Xie, G., Hua, W., "Key problems and research program for PTA process domestic development", *Polyester Industry*, **14** (1), 1—7 (2001). (in Chinese)

5 Takano, T., Suzuki, H., Taniguchi, N., "Process for producing purified terephthalic acid", U. S. Pat., 5973196 (1999).

6 Lin, Y.H., Cunningham, G.A, "A new approach to fuzzy-neural system modeling", *IEEE Trans., Fuzzy Systems*, **3** (2), 190—198 (1995).

7 Mu, S.J., Su, H.Y., Gu, Y., Chu, J., "Muti-objective optimization of industrial purified terephalic acid oxidation process", *Chinese J. Chem. Eng.*, **11** (5), 536—541 (2003).

8 Er, M.J., Liao, J., Lin, J.Y., "Fuzzy neural networks-based quality prediction system for sintering process", *IEEE Trans.*, Fuzzy Systems, **8** (3), 314—324 (2000).

9 Horikawa, S.I., Furuhashi, T., Uchikawa, Y., "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm", *IEEE Trans., Neural Networks*, **5** (3), 801—806 (1992).

10 Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Dlenum Press (1981).

11 Linkens, D.A., Chen, M.Y., "Input selection and partition validation for fuzzy modelling using neural network", *Fuzzy Sets and Systems*, **107**, 299—308 (1999).

12 Rouai, F.A., Ahmed, M.B., "A new approach for fuzzy neural network weight initialization", *Proceedings of IJCNN'01*, (2), 1322—1327 (2001).

13 Sung, A.H., "Ranking importance of input parameters of neural networks", *Expert Systems with Application.*, **15** (3), 405—411 (1998).