PERGAMON

AVAILABLE AT www.ComputerScienceWeb.com POWERED BY SCIENCE DIRECT®

Neural Networks 16 (2003) 779–784

Neural Networks

www.elsevier.com/locate/neunet

2003 Special issue

Extension neural network and its applications

M.H. Wang*, C.P. Hung

Department of Electrical Engineering, National Chin-Yi Institute of Technology, 35, 215 Lane, Sec. 1, Chung Shan Road, Taiping, Taichung, Taiwan, ROC

Abstract

In this paper, a novel extension neural network (ENN) is proposed. This new neural network is a combination of extension theory and neural network. It uses an extension distance (ED) to measure the similarity between data and cluster center. The learning speed of the proposed ENN is shown to be faster than the traditional neural networks and other fuzzy classification methods. Moreover, the new scheme has been proved to have high accuracy and less memory consumption. Experimental results from two different examples verify the effectiveness and applicability of the proposed work.

Keywords: Neural networks; Extension neural network; Extension theory

1. Introduction

Neural networks are parallel systems used for solving regression and classification problems in many fields (Carpenter, Grossberg, & Rosen, 1991; Kohonen, 1988; Rumelhart & McClelland, 1986; Specht, 1990). They can estimate a relation function between the inputs and outputs from a learning process, and also can discover the mapping form feature space into space of classes. Classification or cluster analysis is one of the most important applications of neural networks. The goal of classification is to partition a set of patterns into a group of desired subsets. There are many popular methods for applying neural networks to pattern recognition such as, multiplayer perceptrons (MLP) (Rumelhart & McClelland, 1986), Kohonen neural networks (KNN) (Kohonen, 1988), probabilistic neural network (PNN) (Specht, 1990), learning vector quantization (LVQ) (Bezdek & Pal, 1995), counter propagation networks (CPN) (Hecht-Nielaen, 1987), and adaptive resonance theory (ART) networks (Carpenter et al., 1991). There have been many successful applications in many fields.

The MLP is a continuous input and output pattern recognition, it experts in supervised learning. The most popular training method is error-back-propagation. The drawbacks are that it is not a good strategy to decide the number of neurons in hidden layers and it is time-consuming in training. The KNN is unsupervised training pattern recognition. It employs a winner-take-all learning strategy to store similar patterns in one neuron. KNN has good applications in phonetic or image pattern recognition, but it is not a good strategy to decide the learning parameters and the region of neighborhood. The ART network is an unsupervised learning and adaptive pattern recognition system. It can quickly and stably learn to categorize input patterns and permit an adaptive process for significant and new information. On the other hand, many methods have been proposed to design fuzzy classification systems for dealing with fuzzy classification problems (Hong & Chen, 1999, 2000; Wang, Liu, Hong, & Tseng, 1999; Yu and Chen, 2002). The fuzzy approaches can take human expertise, and have been successfully applied in this field. However, there are some intrinsic shortcomings, such as the difficulty of acquiring knowledge and maintaining a database.

In our world, there are some classification problems whose features are defined in a range. For example, boys can be defined as a cluster of men from age 1 to 14 and the permitted operation voltages of a specified motor may be between 100 and 120 V. For these problems, it is not easy to implement an appropriate classification method using current neural networks. Therefore, a new neural network topology, called the extension neural network (ENN) is proposed to solve these problems in this paper. In other words, the ENN permits classification of problems, which have range features, supervised learning, or continuous input and discrete output. This new neural network is

* Corresponding author. Tel.: +886-4-23924505-7233; fax: +886-4-23924419.

E-mail address: wangmh@chinyi.ncit.edu.tw (M.H. Wang).

0893-6080/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0893-6080(03)00104-7

a combination of the extension theory (Cai, 1983; Huang & Chen, 1999; Wang & Chen, 2001) and the neural network, the ENN uses a modified extension distance (ED) to measure the similarity between data and cluster center; it permits adaptive process for significant and new information, and gives shorter learning times than traditional neural networks. Moreover, this ENN has shown higher accuracy, less memory consumption in application.

2. Review of extension theory

Extension theory was originally invented by Cai to solve contradictions and incompatibility problems in 1983 (Cai, 1983). The extension set extends the fuzzy set from [0,1] to $(-\infty, \infty)$. As a result, it allows us to define a set that includes any data in the domain.

2.1. Matter-element theory

(1) Definition of matter-element

Defining the name of a matter by N, one of the characteristics of the matter by c, and the value of c by v, we can use an ordered ternary $R = (N, c, v)$ as the fundamental element to state a matter and call it a matter-element in extension theory. For example, $R =$ (Wang, Weight, 80 kg) can be used to state that Wang's weight is 80 kg. If the value of the characteristic has a classical domain or a range, we define the matter-element for the classical domain as follows:

$$R = (N, c, W) = (N, c, \langle w^{L}, w^{U} \rangle) \tag{1}$$

Where w^{L} and w^{U} are the lower bound and upper bound of classical domains, respectively.

(2) Multi-dimensional matter-element

If $R = (N, C, V)$ is a multi-dimensional matter-element, $C = [c_1, c_2, \ldots, c_n]$ a characteristic vector and $V = [v_1, v_2, \ldots v_n]$ a value vector of C, then a multi-dimensional matter-element is defined as follows:

$$R = (N, C, V) = \begin{bmatrix} N, c_1, v_1 \\ c_2, v_2 \\ \vdots \\ c_n, v_n \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} \tag{2}$$

Where $R_i = (N, c_i, v_i)(i = 1, 2 \ldots n)$ is the sub-matter-element of R.

2.2. Summary of extension set theory

(1) Definition of extension set

If U is a space of objects and x a generic element of U, then an extension set A in U is defined as a set of ordered pairs:

$$A = \{(u, y) | u \in U, y = K(x) \in (-\infty, \infty)\} \tag{3}$$

Where $y = K(x)$ is called the relational function for extension set A. The $K(x)$ maps each element of U to a membership grade between $-\infty$ and ∞. The higher the degree, the closer the element belongs to the set.

(2) Primitively extended relation function

Let $X_o = \langle a, b \rangle$, $X = \langle c, d \rangle$ and $X_o \in X$, then the extended relation function can be defined as follows:

$$K(x) = \frac{\rho(x, X_o)}{D(x, X_o, X)} \tag{4}$$

where

$$\rho(x, X_o) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2} \tag{5}$$

$$D(x, X_o, X) = \begin{cases} \rho(x, X) - \rho(x, X_o) & x \notin X_o \\ -1 & x \in X_o \end{cases} \tag{6}$$

When $K(x) \geq 0$, it can describe the degree to which x belongs to X_o and $K(x) < 0$ can describe the degree to which x does not belong to X_o. Apparently, those values that are not inside the set are not discussed in the fuzzy set. When $-1 < K(x) < 0$, this domain is called an extension domain, which means that the element x still has a chance to become an element of the set.

3. Extension neural network

The proposed ENN is a combination of the neural network and the extension theory. The extension theory proves a novel distance measurement for classification processes, and the neural network can embed the salient features of parallel computation power and learning capability. In other words, the ENN permits classification of problems, which have range features, supervised learning, or continuous input and discrete output.

3.1. Structure of ENN

The schematic structure of the ENN is depicted in Fig. 1. It comprises both the input layer and the output layer. The nodes in the input layer receive an input feature pattern and use a set of weighted parameters to generate an image of the input pattern. In this network, there are two connection values (weights) between input nodes and output nodes; one connection represents the lower bound for this classical domain of the features, and the other connection represents the upper bound. The connections between the j-th input node and the k-th output node are w_{kj}^{L} and w_{kj}^{U}.

This image is further enhanced in the process characterized by the output layer. Only one output node in the output layer remains active to indicate a classification of the input pattern. The operation mode of the proposed ENN can be

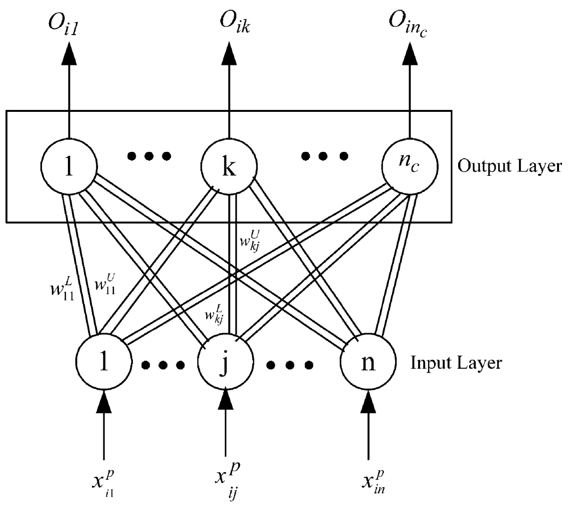

Fig. 1. The structure of extension neural network (ENN).

separated into the learning phase and the operation phase. The learning algorithm of the ENN is discussed in the next section.

3.2. Learning algorithm of the ENN

The learning of the ENN can be seen as supervised learning; the purpose of learning is to tune the weights of the ENN to achieve good clustering performance or to minimize the clustering error. Before the learning, several variables have to be defined. Let training pattern set be $X \equiv \{X_1, X_2, \ldots, X_{N_p}\}$, where N_p is the total number of training patterns. The i-th pattern is $X_i^p \equiv \{x_{i1}^p, x_{i2}^p, \ldots, x_{in}^p\}$, where n is the total number of the feature of patterns, and the category of the i-th pattern is p. To evaluate the clustering performance, the total error number is set as N_m, and the total error rate E_T is defined below:

$$E_\mathrm{T} = \frac{N_m}{N_p} \tag{7}$$

The detailed supervised learning algorithm can be described as follows:

Step 1: Set the connection weights between input nodes and output nodes by the matter-element model of extension theory:

$$R_k \begin{bmatrix} N_k, c_1, V_{k1} \\ c_2, V_{k2} \\ \vdots \\ c_n, V_{kn} \end{bmatrix} \qquad k = 1, 2, \ldots, n_c \tag{8}$$

In the extension theory, c_j is the j-th characteristic (feature) of N_k and $V_{kj} = \langle w_{kj}^\mathrm{L}, w_{kj}^\mathrm{U} \rangle$ are the classical domains of the k-th cluster (N_k) about the j-th feature c_j. The range of classical domains can be directly obtained from previous requirement, or determined from training data as follows:

$$w_{kj}^\mathrm{L} = \underset{i \in N_p}{\mathrm{Min}} \{x_{ij}^k\} \tag{9}$$

$$w_{kj}^\mathrm{U} = \underset{i \in N_p}{\mathrm{Max}} \{x_{ij}^k\} \tag{10}$$

Step 2: Calculate the initial cluster center of every cluster

$$Z_k = \{z_{k1}, z_{k2}, \ldots, z_{kn}\} \tag{11}$$

$$z_{kj} = (w_{kj}^\mathrm{L} + w_{kj}^\mathrm{U})/2 \tag{12}$$

for $k = 1, 2 \ldots n_c; j = 1, 2 \ldots n$

Step 3: Read the i-th training pattern and its cluster number p

$$X_i^p = \{x_{i1}^p, x_{i2}^p, \ldots, x_{in}^p\}, \qquad p \in n_c \tag{13}$$

Step 4: Use the proposed ED to calculate the distance between the training pattern X_i^p and the k-th cluster as follows:

$$\mathrm{ED}_{ik} = \sum_{j=1}^{n} \left[\frac{|x_{ij}^p - z_{kj}| - (w_{kj}^\mathrm{U} - w_{kj}^\mathrm{L})/2}{|(w_{kj}^\mathrm{U} - w_{kj}^\mathrm{L})/2|} + 1 \right] \tag{14}$$

for $k = 1, 2, \ldots, n_c$

The proposed distance is a modification of ED (Cai, 1983). It can be graphically presented as Fig. 2 it can describe the distance between the x and a range $\langle w^\mathrm{L}, w^\mathrm{U} \rangle$. From Fig. 2, we can see that different ranges of classical domains can arrive at different distances due to different sensitivities. This is a significant advantage in classification applications. Usually, if the feature covers a large range, the requirement of data is fuzzy or low in sensitivity to distance. On the other hand, if the feature covers a small range, the requirement of data is precision or high sensitivity to distance.

Step 5: Find the k^*, such that $\mathrm{ED}_{ik^*} = \mathrm{Min}\{\mathrm{ED}_{ik}\}$. If $k^* = p$ then go to Step 7; otherwise Step 6.

Step 6: Update the weights of the p-th and the k^*-th clusters as follows:

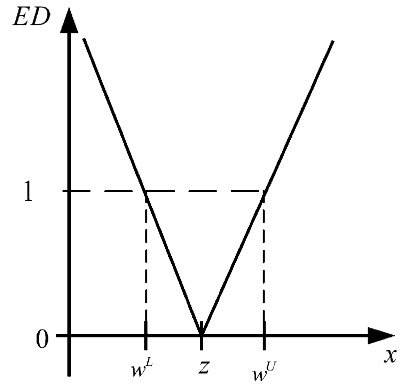

Fig. 2. The proposed extension distance (ED).

(a) Update the centers of the p-th and the k^*-th clusters:

$$z_{pj}^{\text{new}} = z_{pj}^{\text{old}} + \eta(x_{ij}^{p} - z_{pj}^{\text{old}}) \tag{15}$$

$$z_{k^*j}^{\text{new}} = z_{k^*j}^{\text{old}} - \eta(x_{ij}^{p} - z_{k^*j}^{\text{old}}) \tag{16}$$

(b) Update the weights of the p-th and the k^*-th clusters:

$$\begin{cases} w_{pj}^{\text{L(new)}} = w_{pj}^{\text{L(old)}} + \eta(x_{ij}^{p} - z_{pj}^{\text{old}}) \\ w_{pj}^{\text{u(new)}} = w_{pj}^{\text{u(old)}} + \eta(x_{ij}^{p} - z_{pj}^{\text{old}}) \end{cases} \tag{17}$$

$$\begin{cases} w_{k^*j}^{\text{L(new)}} = w_{k^*j}^{\text{L(old)}} - \eta(x_{ij}^{p} - z_{k^*j}^{\text{old}}) \\ w_{k^*j}^{\text{u(new)}} = w_{k^*j}^{\text{u(old)}} - \eta(x_{ij}^{p} - z_{k^*j}^{\text{old}}) \end{cases} \tag{18}$$

Where η is a learning rate. The result of tuning two cluster's weights is shown in Fig. 3 clearly indicating the change of ED_A and ED_B. The cluster of pattern x_{ij} is changed from cluster A to B due to $ED_A > ED_B$. From this step, we can clearly see that the learning process is only to adjust the weights of the p-th and the k^*-th clusters. Therefore, the proposed method has a speed advantage over other supervised learning algorithms, and can quickly adapt to new and important information.

Step 7: Repeat Step 3 to Step 6, if all patterns have been classified, then a learning epoch is finished.

Step 8: Stop, if the clustering process has converged, or the total error rate E_T has arrived at a preset value, otherwise, return to Step 3.

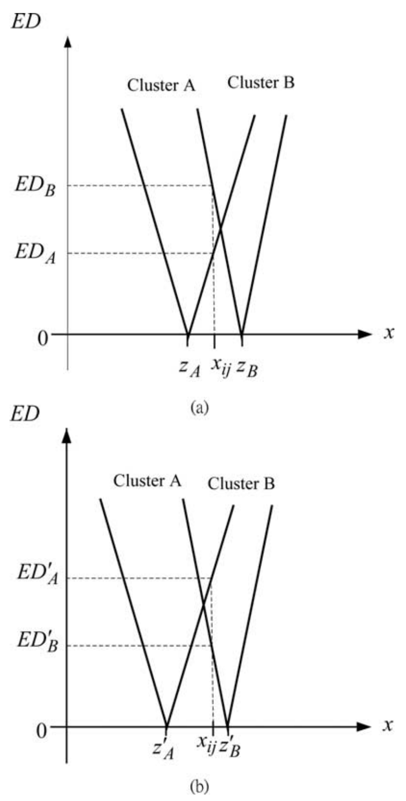

Fig. 3. The results of tuning cluster weights: (a) original condition; (b) after tuning.

3.3. Operation phase of ENN

Step 1: Read the weight matrix of ENN

Step 2: Calculate the initial cluster centers of every cluster using Eqs. (11) and (12)

Step 3: Read the tested pattern

$$X_t = \{x_{t1}, x_{t2}, \ldots, x_{tn}\} \tag{19}$$

Step 4: Use the proposed ED to calculate the distance between the tested pattern and every existed cluster by Eq. (14)

Step 5: Find the k^*, such that $\text{ED}_{ik^*} = \text{Min}\{\text{ED}_{ik}\}$, and set the $O_{ik^*} = 1$ to indicate the cluster of the tested pattern.

Step 6: Stop, if all the tested patterns have been classified, otherwise go to Step 3.

4. Experimental results

In this paper, the Iris data classification problem (Chien, 1978) and vibration diagnosis problems (Li, Sun, Liao, Chen & Hu, 1999; Li, Sun, Hu, Yue, Tang & Wang, 2000) are used to illustrate the effectiveness of the proposed ENN.

4.1. Iris data classification

There are 150 instances in the Iris data; it can be divided into three categories with the distinguishing variables being the length and width of sepal and petal. In this case, the structures of the proposed ENN are three output nodes and four input nodes. To prove the efficiency of the proposed ENN, two test cases is given in the following:

Case 1: If the system randomly chooses 75 instances from the Iris data as the training data set, and let the rest of the instances of the Iris data are the testing data set. Fig. 4 shows the learning curves of the proposed ENN with different learning rates. It is clear that the training time of the proposed ENN is quite economical with about 4 epochs for $\eta = 0.1$ and about 13 epochs for $\eta = 0.01$. Table 1 shows the comparison of the experimental results of the proposed ENN with other typical neural networks. It should

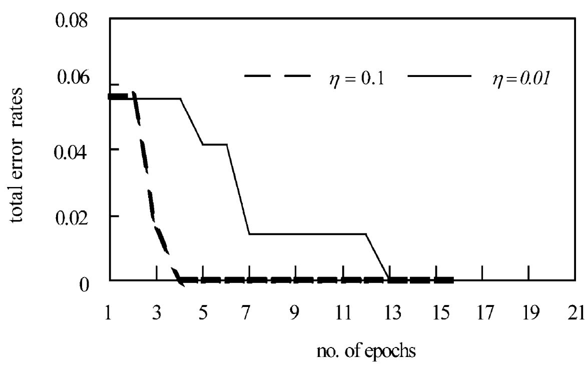

Fig. 4. The learning curves of Iris data classification problem.

Table 1
Comparison of the classification performance of various neural networks

Model	Structure	No. of connections	Learning times (epochs)	Training error	Testing error
Perceptron	4-3	12	200	0.173	0.213
MLP	4-4-3-3	37	50	0.027	0.040
PNN	4-75-3	525	1	0.0	0.053
LVQ	4-15-3	105	20	0.8	0.053
CPN	4-20-3	140	60	0.107	0.160
ENN	4-3	24	4	0.0	0.040

MLP, Multilayer perceptron; PNN, probabilistic neural network; LVQ, learning vector quantization; CPN, counter propagation network; ENN, extension neural network; training set comprised 75 data points and testing set comprised the remaining data points.

be noted that structure of the proposed ENN is the very simple, only 7 nodes and 24 connections is needed. Moreover, the proposed ENN-based also permits fast adaptive process for a large amount of training data or new information, because the learning of ENN is only to tuning low-bound and upper bound of the excited connections. It can be seen from the Table 1 that the proposed ENN has a shorter learning time than the traditional neural networks. As well, the accurate rates are quite high with about 100 and 96% for training and testing patterns, respectively. On the other hand, the proposed ENN can take expert experiences before learning, and can also produce meaningful output after learning, because the optimal classified boundary of the features are clearly determined.

Case 2: If the training data set contains 150 training instances (i.e., the full Iris data) and the testing data set is equal to the training data set containing 150 training instances. Table 2 compares the performance of the proposed ENN with several other fuzzy classification methods on the same data set. It can be seen that the proposed ENN is at least as good as the other methods, but it has a shortest learning times. Moreover, the proposed ENN permits fast adaptive process for significant and new information, and it is easy to acquire knowledge and maintain the classification database.

4.2. Vibration diagnosis of generator sets

The vibration diagnosis of generator set is based on the principle that components in engineering systems and plants produce vibration during operation. If a generator set is operating properly, vibration conditions are usually small and constant, but when faults grow or some of the dynamic processes in the machine change, the vibration signature also changes. Therefore, diagnostic information can be supplied by the spectrum of the vibration signal. In agreement with past studies (Li et al., 1999, 2000), the typical six values (amplitude of <0.4f, 0.4f−0.5f, f, 2f, 3f, and >3f) are selected for vibration fault diagnosis, and the detailed trained data are shown in Table 3. To compare diagnosis performance, the diagnosis results with different two different classification methods (Li et al., 1999, 2000),

Table 2
Comparison of the average classification accuracy rate for different method

Methods	Learning times (epochs)	Average classification accuracy rate
Hong and Chen's method	200	96.67%
Hong and Chen's method	200	97.33%
Wang-et al., method	200	97.33%
Yu and Chen's method	200	97.33%
Proposed ENN method	4	97.33%

Table 3
Tested data of generator sets

Gen. no.	Input data						Fault types
	<0.4f	0.4f−0.5f	1f	2f	3f	>3f	
1	3.35	46.6	12.15	1.94	2.3	3.67	F_1
2	4.43	51	11.02	3.02	1.3	2.43	F_1
3	3.29	50	11.61	1.24	0.9	1.3	F_1
4	5.72	46.3	12.31	3.62	1.5	0.59	F_1
5	6.32	45.8	15.23	3.56	2.3	3.19	F_1
6	1.51	3.29	52.92	6.59	2.5	2.54	F_2
7	2.43	1.19	54.49	4.64	0.8	1.78	F_2
8	0.54	2.92	48.82	6.64	3.9	1.51	F_2
9	0.81	1.73	52	6.43	3.6	1.89	F_2
10	1.24	1.35	49.79	4.64	1.0	2.27	F_2
11	1.78	1.46	22.46	23.8	19	8.59	F_3
12	0.92	1.24	30.08	22	16	5.67	F_3
13	0.65	2.11	21.98	26.2	18	11.1	F_3
14	1.13	0.92	24.46	22.3	15	15.8	F_3
15	0.92	1.40	26.08	26	20	11.4	F_3

F_1, Oil-resonance fault; F_2, imbalance fault; F_3, misalignment fault.

Table 4
Learning results using different neural networks

Classifiers	MLP	AWN	ENN
Structure	6-13-3	6-13-3	6-3
Learning speed (epochs)	2561	900	2
Accuracy	100%	100%	100%

AWN, Adaptive wavelets network.

i.e. MLP and AWN, are shown in Table 4. The two traditional neural networks were capable of pointing toward faults, but both need to learn about 2561 and 900 epochs before fault diagnosis. Contrary, the proposed ENN only need to learn 2 epochs with equivalent accuracy, and the structure of the proposed ENN is simpler than the other neural networks.

5. Conclusions

This paper presents a novel ENN based on the extension theory and neural network. Compared with traditional neural networks and other fuzzy classification methods, it permits an adaptive process for significant and new information, and gives shorter learning times. The proposed ENN can solve some special classification problems that the feature is defined in a range. Moreover, the proposed ED, the different ranges of classical domain can arrive at different distances due to different sensitivities, which is a significant advantage in classification applications. From the tested examples, the proposed ENN has been proved to have the advantage of less learning time, higher accuracy and less memory consumption. Future studies will be carried out to develop an unsupervised learning algorithm of the proposed ENN, and to extend this classification technique to control and develop fault diagnosis systems.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Council, Taiwan, ROC, under the grant no. NSC-91-2213-E-167-014.

References

Bezdek, J. C., & Pal, N. R. (1995). Two soft relative of learning vector quantization. *Neural Networks*, *8*, 729–743.

Cai, W. (1983). The extension set and incompatibility problem. *Journal of Scientific Exploration*, *1*, 81–93.

Carpenter, G., Grossberg, S., & Rosen, D. B. (1991). ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, *4*, 493–504.

Chien, Y. T. (1978). *Interactive pattern recognition*. New York, USA: Marcel Dekker, pp. 223–225.

Hecht-Nielaen, R. (1987). Counter-propagation networks. *ICNN-87*, *2*, 19–32.

Hong, T. P., & Chen, J. B. (1999). Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, *103*, 389–404.

Hong, T. P., & Chen, J. B. (2000). Processing individual fuzzy attributes for fuzzy rule induction. *Fuzzy Sets and Systems*, *112*, 127–140.

Kohonen, T. (1988). *Self-organization and associative memory* (2nd ed.). Berlin, Germany: Springer.

Li, H., Sun, C. X., Liao, R. J., Chen, W. G., & Hu, X. S. (1999). Improved BP algorithm in vibration fault diagnosis of steam turbine-generator set. *Journal of Chongqing University*, *22*, 47–52.

Li, H., Sun, C. X., Hu, X. S., Yue, G., Tang, N. F., & Wang, K. (2000). Study of method on Adaptive wavelets network for vibration fault diagnosis of steam turbine-generation set. *Transactions of China Electrotechnical Society*, *15*, 52–57.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing explorations in the microstructure of cognition*. Cambridge: Bradford Books, MIT Press.

Specht, D. F. (1990). Probabilistic neural networks and the polynomial adaline as complementary techniques for classification. *IEEE Transactions on Neural Networks*, *1*, 111–121.

Wang, C. H., Liu, J. F., Hong, T. P., & Tseng, S. S. (1999). A fuzzy inductive strategy for modular rules. *Fuzzy Sets and Systems*, *103*, 91–105.

Wang, M. H., & Chen, H. C. (2001). Application of extension theory on the fault diagnosis of power transformers. *Proceeding of the 22nd Symposium on Electrical Power Engneering, Taiwan*, 797–800.

Yu, C. H., & Chen, S. M. (2002). A new method for handling fuzzy classification problems based on clustering techniques. *Proceeding of the Seventh Conference on Artificial Intelligence and Applications*, 107–112.