Extension Neural Network-Type 2 and Its Applications

Mang-Hui Wang, *Member, IEEE*

***Abstract*—A supervised learning pattern classifier, called the extension neural network (ENN), has been described in a recent paper. In this sequel, the unsupervised learning pattern clustering sibling called the extension neural network type 2 (ENN-2) is proposed. This new neural network uses an extension distance (ED) to measure the similarity between data and the cluster center. It does not require an initial guess of the cluster center coordinates, nor of the initial number of clusters. The clustering process is controlled by a distanced parameter and by a novel extension distance. It shows the same capability as human memory systems to keep stability and plasticity characteristics at the same time, and it can produce meaningful weights after learning. Moreover, the structure of the proposed ENN-2 is simpler and the learning time is shorter than traditional neural networks. Experimental results from five different examples, including three benchmark data sets and two practical applications, verify the effectiveness and applicability of the proposed work.**

***Index Terms*—ENN-2, extension neural network (ENN), neural networks (NNs), unsupervised learning.**

I. Introduction

The learning algorithms of neural networks can be categorized into two classes—supervised learning and unsupervised learning—according to the availability of the goal output for given input data. Supervised learning is a process that incorporates an external teacher and environmental information, so it requires an external goal output to respond to input data. A supervised learning neural network (NN) can estimate a relation function between the inputs and outputs from a learning process and also can discover mapping from feature space into a space of classes. Unsupervised learning is a process that incorporates no external teacher; it results in exposition of clusters for given input patterns. The goal of cluster analysis is to partition a set of patterns into a group of desired subsets.

Classification or cluster analysis is one of the most important applications of neural networks. There are many popular neural networks using unsupervised learning, which have been used for solving classification problems in many fields [1]–[6]. Typically neural networks include Kohonen neural networks (KNNs) [1] and adaptive resonance theory (ART) networks [3], [4]; there have been many successful applications in many fields. The KNN employs a winner-take-all learning strategy to store similar patterns in one neuron. KNN has good applications in phonetic or image pattern recognition. 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 incremental learning for significant and new information. ART neural networks have the ability to learn new objects and keep old memories; the system has the two characteristics of stability and plasticity, in the same system, which is the same capability as human memory systems.

It is very important to develop one memory system to keep stability and plasticity at the same time and also keep the system low in computation cost and in memory consumption in the modern commodities [7]–[9]. On the other hand, there are some classification problems whose features are defined over an interval of values in our world. 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, several new neural networks have been proposed for where the features are defined over an interval of values, such as fuzzy lattice networks [10], [11] and radial basis function networks [12], [13]. There have been many successful applications in some fields. Nevertheless, the applicability domain of all previous neural schemes is more or less restricted. Therefore, a new neural network topology, called the extension neural network (ENN), was proposed to solve these problems in our recent paper [14]. In other words, the ENN permits classification of problems whose features are defined over an interval of values in our world, supervised learning, or continuous input and discrete output. This new neural network is a combination of the extension theory [15] and the neural network. The ENN uses a modified extension distance (ED) to measure the similarity between data and cluster center; it permits adaptive processes for significant and new information and gives shorter learning times than traditional neural networks. Moreover, this ENN has shown higher accuracy with less memory consumption in applications [14].

In this sequel, an ENN unsupervised learning pattern cluster relative called the extension neural network type 2 (ENN-2) is presented. The architecture of the ENN-2 is almost the same as the ENN; it uses the proposed extension distance and a distance parameter to control the clustering process. The learning algorithm of the ENN-2 implements a follow-the-leader approach. It does not require an initial guess of the cluster center coordinates, nor of the initial number of clusters. Five applicable examples will be tested to show the effectiveness of this new neural network.

Manuscript received September 15, 2003; revised December 18, 2004. This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC-92-2213-E-167-015.

The author is with the Institute of Information and Electrical Energy, National Chin-Yi Institute of Technology, Taichung, Taiwan, R.O.C. (e-mail: wangmh@chinyi.ncit.edu.tw).

Digital Object Identifier 10.1109/TNN.2005.853334

TABLE I
THREE DIFFERENT SORTS OF MATHEMATICAL SETS

Compared item	Crisp set	Fuzzy set	Extension set
Research objects	Data variables	Linguistic variables	Contradictory problems
Model	Mathematics model	Fuzzy mathematics model	Matter-element model
Descriptive function	Transfer function	Membership function	Correlation function
Descriptive property	Precision	Ambiguity	Extension
Range of set	$C_A(x) \in (0,1)$	$\mu_A(x) \in [0,1]$	$K_A(x) \in [-\infty,\infty]$

II. SUMMARY OF EXTENSION THEORY

In the crisp set, an element either belongs to or does not belong to a set, so the range of the truth-values is $\{0, 1\}$, which can be used to solve a two-valued problem. In contrast to the crisp set, the fuzzy set allows for the description of concepts in which the boundary is not explicit. It concerns not only whether an element belongs to the set but also to what degree it belongs. The range of membership function is [0, 1] in the fuzzy set. The extension theory was originally created by Cai to solve contradictions and incompatibility problems in 1983 [15]. The extension set extends the fuzzy set from [0, 1] to $[-\infty, \infty]$. This means that an element belongs to any extension set to a different degree. Define the membership function by $K(x)$ to represent the degreeto which an element belongs to a set. A degree between zero and one corresponds to the normal fuzzy set. When $K(x) < 0$, it describes the degree to which x does not belong to a set, which is not defined in a fuzzy set. When $-1 < K(x) < 0$, this means that the element x still has a better chance to be included into the set if the set is adjusted. $K(x) < -1$ implies that the element x has no chance to belong to the set. It is also to represent the degree of an element not belonging to a set. The extension theory tries to solve incompatibility or contradiction problems by the transformation of the matter element. Comparisons of crisp sets, fuzzy sets, and extension sets are shown in Table I. Some definitions of extension theory are introduced in the next section.

A. 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, a matter-element in extension theory can be described as follows:

$$R = (N, c, v) \tag{1}$$

where $N, c,$ and v are called the three fundamental elements of the matter-element. For example, R = (John, Weight, 80 kg) can be used to state that John'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:

$$R = (N, c, v) = \left(N, c, \langle w^L, w^U \rangle\right) \tag{2}$$

where w^L and w^U are the lower bound and upper bound of a classical domain, respectively.

2) Multidimensional Matter-Element: Assuming $R = (N, C, V)$ a multidimensional 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 multidimensional matter-element is defined as

$$R = (N, C, V) = \begin{bmatrix} N, c_1, v_1 \\ c_2, v_2 \\ \cdots\cdots \\ c_n, v_n \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \cdots \\ R_n \end{bmatrix} \tag{3}$$

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

$$R = (N, C, V) = \begin{bmatrix} \text{workpiece } A, \text{length}, [21 \text{ cm}, 22 \text{ cm}] \\ \text{diameter}, [4 \text{ cm}, 6\text{cm}] \\ \text{weight}, [20 \text{ kg}, 30 \text{ kg}] \end{bmatrix} \tag{4}$$

is a three-dimensional matter element.

3) Divergence of Matter-Element: A matter may have many characteristics; the same characteristics and values may also belong to some other matter. In extension theory, there are some formulations to express these points as follows.

Definition 1: If a matter has many characteristics, it can be written as

$$N \dashv (N, c, v) \dashv \{(N, c_1, v_1), (N, c_2, v_2), \ldots, (N, c_n, v_n)\} \tag{5}$$

The symbol "$\dashv$" means the extension of matter.

Definition 2: If some other matter has the same characteristic, it can be written as

$$(N, c, v) \dashv \{(N_1, c, v_1), (N_2, c, v_2), \ldots, (N_n, c, v_n)\}. \tag{6}$$

Definition 3: If some matter has the same value, it can be written as

$$(N, c, v) \dashv \{(N_1, c_1, v), (N_2, c_2, v), \ldots, (N_n, c_n, v)\}. \tag{7}$$

B. Extension Set Theory

1) Definition of Extension Set: Let U be the universe of discourse. Then an extension set A on U is defined as a set of ordered pairs as follows:

$$A = \{(x, y) | x \in U, y = K(x) \in (-\infty, \infty)\} \tag{8}$$

where $y = K(x)$ is called the membership function for extension set A. $K(x)$ maps each element of U to a membership grade between $-\infty$ and ∞. The higher the degree, the more the element belongs to the set. Under a special condition, when $0 \le K(x) \le 1$, it corresponds to a normal fuzzy set. $K(x) \le -1$ implies that the element x has no chance to belong to the set. When $-1 < K(x) < 0$, it is called an extension domain, which means that the element x still has a chance to become part of the set.

2) Definition of Correlation Function: If $X_o = \langle a, b\rangle$ and $X = \langle c, d\rangle$ are two intervals in the real number field, and $X_o \subset$

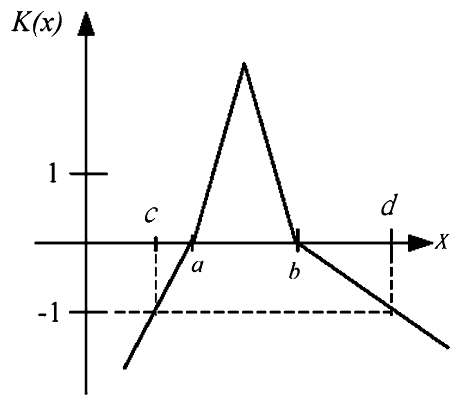

Fig. 1. The extended membership function.

X, then the correlation function in the extension theory can be defined as follows:

$$K(x) = \begin{cases} -\rho(x, X_o) & x \in X_o \\ \frac{\rho(x,X_o)}{\rho(x,X)-\rho(x,X_o)} & x \notin X_o \end{cases} \tag{9}$$

where $\rho(x, X_o)$ is defined as the extension distance between x and X_o and written as

$$\rho(x, X_o) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2} \tag{10}$$

$$\rho(x, X) = \left| x - \frac{c+d}{2} \right| - \frac{d-c}{2}. \tag{11}$$

The correlation function $K(x)$ can be used to calculate the membership grade between x and X_o as shown in Fig. 1. When $K(x) \geq 0$, it indicates the degree to which x belongs to X_o. When $K(x) < 0$, it describes the degree to which x does not belong to X_o, which is not defined in fuzzy set theory. When $-1 < K(x) < 0$, it is called the extension domain, which means that the element x still has a chance to become part of the set if conditions change.

III. Extension Neural Network Type 2

The structure of ENN-2 is almost same as ENN proposed in our earlier paper [14]; the proposed ENN-2 is a combination of the neural network and the extension theory. The extension theory introduces a novel distance measurement for classification processes, and the neural network can embed the salient features of parallel computation power and learning capability. As in the statement mentioned in the introduction, the proposed ENN-2 also simultaneously has stability and plasticity characteristics; it permits an adaptive process for significant and new information, and keeps the old information in memory.

A. Architecture of the ENN-2

The schematic architecture of the ENN-2 is depicted in Fig. 2. It comprises an input layer and an 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 jth input node and the mth output node are w_{mj}^L and w_{mj}^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 learning algorithm of the ENN-2 is discussed in the next section.

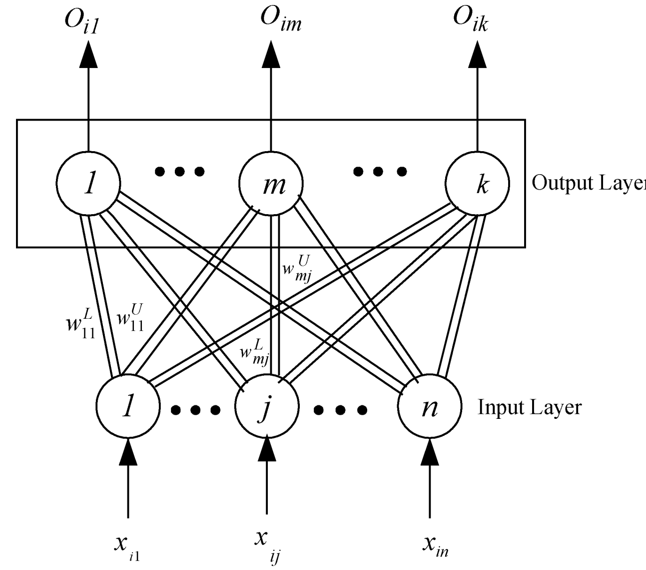

Fig. 2. The structure of an extension neural network.

B. Unsupervised Learning Algorithm of the ENN-2

The learning of the ENN-2 can be seen as unsupervised learning. This algorithm implements a follow-the-leader approach. It does not require the initial number of clusters, nor an initial guess of the cluster center coordinates. The ENN-2 uses a threshold called the distance parameter (DP) λ and a novel ED function to control the clustering process. DP λ is used to measure the distance between the cluster center and the desired boundary. First, a pattern is selected as the center of the first cluster, and the initial weights of the first cluster can be computed from the center with desired distance parameter λ. Then the next pattern is compared to the first cluster. It is clustered with the first if its distance is smaller than the vigilance parameter (i.e., it just equals the number of features). Otherwise, it forms the center of a new cluster. This process is repeated for all patterns until a stable cluster formation occurs. Before the learning, several variables have to be defined as follows:

X_i	ith pattern;
x_{ij}	jth feature of the$i$th input pattern;
Z_k	center of the cluster k;
λ	distance parameter;
N_p	total number of the input pattern;
n	number of the features;
k	number of existing clusters;
M_k	number of patterns belonging to cluster k.

The detailed unsupervised learning algorithm of the ENN-2 can be described as follows.

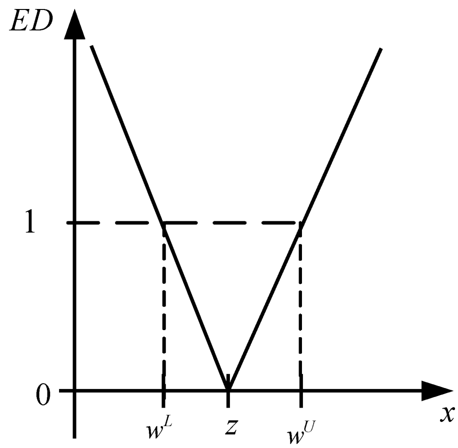

Fig. 3. The proposed ED.

Step 1) Set the desired DP λ. The DP λ is used to measure the distance between the cluster center and the desired boundary. It is a user-defined parameter that must be judiciously determined from an engineering knowledge of the system requirements.

Step 2) Produce the first pattern, and $M_1 = 1$. Then the center coordinates and weights of the first cluster are calculated as

$$k = 1 \tag{12}$$

$$Z_k = X_k \Rightarrow \{z_{k1}, z_{k2}, \ldots, z_{kn}\} = \{x_{k1}, x_{k2}, \ldots, x_{kn}\} \tag{13}$$

$$w_{kj}^L = z_{kj} - \lambda \quad \text{for} \quad j = 1, 2, \ldots, n \tag{14}$$

$$w_{kj}^U = z_{kj} + \lambda \quad \text{for} \quad j = 1, 2, \ldots, n. \tag{15}$$

Step 3) Read the input pattern vectors by letting $i = 2$, and go to next step

Step 4) Read the ith input pattern $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ and calculate the extension distance ED_m between X_i and the existing mth cluster center as

$$ED_m = \sum_{j=1}^{n} \left[\frac{|x_{ij} - z_{mj}| - \frac{(w_{mj}^U - w_{mj}^L)}{2}}{\left| \frac{(w_{mj}^U - w_{mj}^L)}{2} \right|} + 1 \right] \quad \text{for}\, m = 1, 2, \ldots, k. \tag{16}$$

This proposed ED is a modification extension distance of (16). It can be graphically presented as Fig. 3; it can describe the distance between the x and a real interval $\langle w^L, w^U \rangle$. The concept of distance, the position relation between a point and an interval, can be precisely expressed by means of the quantitative form. When a point lies in the interval, the distance is considered as zero by the classical math, but different positions of a point in the interval can be described with the distance value by the proposed extension distance. Fig. 3 clearly shows that the distance equals one when a point appears on the lower bound and upper bound of interval.

Step 5) Find the p for which

$$ED_p = \min\{ED_m\} \quad \text{for m} = 1, 2, \ldots, k. \tag{17}$$

Step 6) *Step 6*: If $ED_p > n$, then create a new cluster center. According to the definition of proposed extension distance, when the x lies in the interval, the distance is considered as smaller than 1, thus if $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ has n features in the clustering process, when $ED_p > n$ expresses that X_i does not belong to p-th cluster. Then a new cluster center will be created,

$$k = k + 1 \tag{18}$$

$$Z_k = X_i \Rightarrow \{z_{k1}, z_{k2}, \ldots, z_{kn}\} = \{x_{i1}, x_{i2}, \ldots, x_{in}\} \tag{19}$$

$$w_{kj}^L = z_{kj} - \lambda \quad \text{for}\ j = 1, 2, \ldots, n \tag{20}$$

$$w_{kj}^U = z_{kj} + \lambda \quad \text{for}\ j = 1, 2, \ldots, n \tag{21}$$

$$M_k = 1 \tag{22}$$

else, the pattern X_i belongs to the cluster p, and update the weights and center of cluster p

$$w_{pj}^{U(\text{new})} = w_{pj}^{U(\text{old})} + \frac{1}{M_p + 1}\left(x_{ij} - z_{pj}^{\text{old}}\right) \tag{23}$$

$$w_{pj}^{L(\text{new})} = w_{pj}^{L(\text{old})} + \frac{1}{M_p + 1}\left(x_{ij} - z_{pj}^{\text{old}}\right) \tag{24}$$

$$z_{pj}^{\text{new}} = \frac{w_{pj}^{U(\text{new})} + w_{pj}^{L(\text{new})}}{2} \quad \text{for}\ j = 1, 2, \ldots, n \tag{25}$$

$$M_p = M_p + 1 \tag{26}$$

Step 7) If input pattern X_i changes from cluster "o" (the old one) to "k" (the new one), then the weights and center of cluster "o" are modified as

$$w_{oj}^{U(\text{new})} = w_{oj}^{U(\text{old})} - \frac{1}{M_o}\left(x_{ij} - z_{oj}^{\text{old}}\right) \tag{27}$$

$$w_{oj}^{L(\text{new})} = w_{oj}^{L(\text{old})} - \frac{1}{M_o}\left(x_{ij} - z_{oj}^{\text{old}}\right) \tag{28}$$

$$z_{oj}^{\text{new}} = \frac{w_{oj}^{U(\text{new})} + w_{oj}^{L(\text{new})}}{2} \quad \text{for}\ j = 1, 2, \ldots, n \tag{29}$$

$$M_o = M_o - 1. \tag{30}$$

The result of tuning two clusters' weights is shown in Fig. 4, clearly indicating the change of ED_o and ED_k. The cluster of pattern x_{ij} is changed from cluster "o" to "k" due to $ED_o' > ED_k'$. From this step, we can clearly see that the learning process is only to adjust the weights of the oth and the kth clusters. Therefore, the proposed method has a speed advantage over other unsupervised learning algorithms and can quickly adapt to new and important information. It should be noted that the clustering process of the proposed ENN-2 keeps stability and plasticity characteristics at the same time.

Step 8) Set $i = i + 1$ and repeat Steps 4)–8) until all the patterns have been compared with the existing clusters,

Step 9) If the clustering process has converged, end; otherwise, return to Step 3).

According to the proposed unsupervised learning algorithm, ENN-2 permits fast adaptive processes for significant and new information. It is easy to acquire knowledge and maintain the classification database. Moreover, the proposed method has a

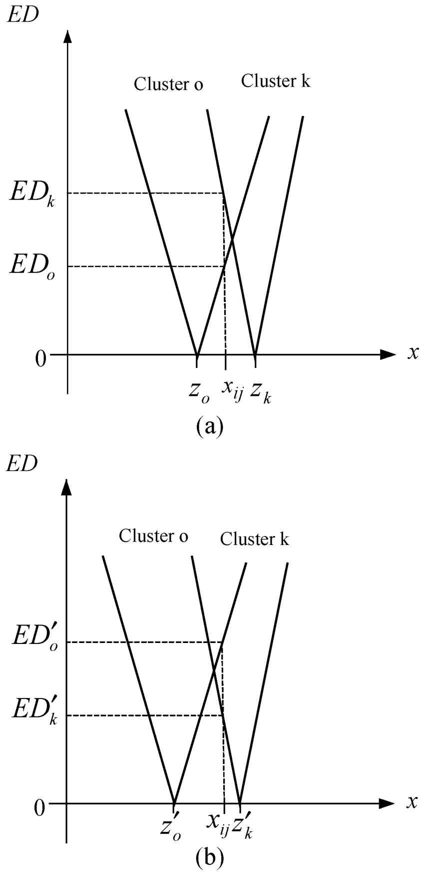

Fig. 4. The results of tuning cluster weights: (a) original condition; (b) after tuning.

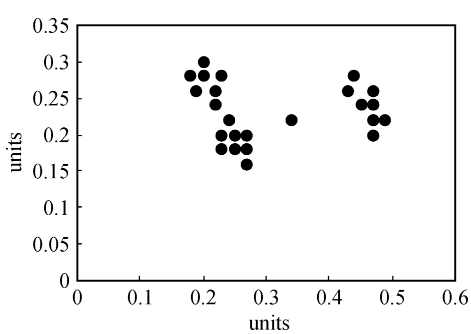

Fig. 5. The Simpson data set.

speed advantage over the traditional unsupervised learning algorithms.

IV. Experimental Results

To verify the effectiveness of the proposed ENN-2, three benchmark data sets and two practical industry examples are used to illustrate the applications of the proposed ENN-2. It should be noted that epoch, as used in this paper, is defined as one presentation of whole data vectors in the data set.

A. Simpson Data Set

The Simpson data set, a benchmark-tested data set, is a two-dimensional data set consisting of 24 points as shown in Fig. 5 [16]. This is a perceptual grouping problem in vision, which deals with the detection of the right partition of an image into subsets [17]. At lower spatial resolution, three cluster partitions may be perceived. To compare the learning capability, Table II shows the comparison of the experimental results of the proposed ENN with other typical ART neural networks [18] on the three-cluster portions. It can be seen from Table II that the proposed ENN has a shorter learning time than the traditional neural networks because the learning of ENN is only to tune the low bound and upper bound of the excited connections. On the other hand, the average confusion number of the proposed ENN is zero. It can completely partition the Simpson data set into three clusters by setting the distance parameter λ of 0.0615. In opposition, the other neural networks may produce confusion on the three-cluster partitions.

TABLE II
Comparison of the Classification Performance of Various Neural Networks

Model	Learning epoches	Average confusion number
Fuzzy-ART [18]	10	5.5
S-Fuzzy ART [18]	10	7.5
ENN-2	1	0.0

Fig. 6 shows the clustering results of the proposed ENN-2 with the different distance parameter λ. The number of clusters created ranges from one to four, and the learning times of every clustering process are also only one epoch. The distance parameter may be considered a constant prescribed by the user and determined through engineering judgment. Usually, smaller distance parameters can produce a higher degree of coherence patterns in a cluster, and more clusters will be produced. On the other hand, it may be viewed as a function of self-organized data structures. If labeled pattern attributes take discrete values (e.g., good/fair/bad), the clustering process is reiterated within a particular cluster until a clear distinction among attributes is achieved. In the continuous case, one can define intervals of the range of attribute values; automatic adjustment of distance parameters then occurs. To illustrate the stability and plasticity capability of the proposed ENN-2, the clustering results with different new data are shown in Fig. 7. If the new data is near the center of cluster 3 and $EP < n$ for a determined distance parameter (e.g. 0.05), then this new data will be included in cluster 3 and ENN-2 will only tune the weights between input layer and cluster 3 (i.e. the third node of output layer). If the new data are far from all cluster centers and $EP > n$ for a determined distance parameter (e.g. 0.05), then these new data will create a new cluster and the original weights of old clusters will be retained.

B. WINE Benchmark

The WINE recognition data were taken from the UCI repository of machine learning databases [19], [20]. These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of

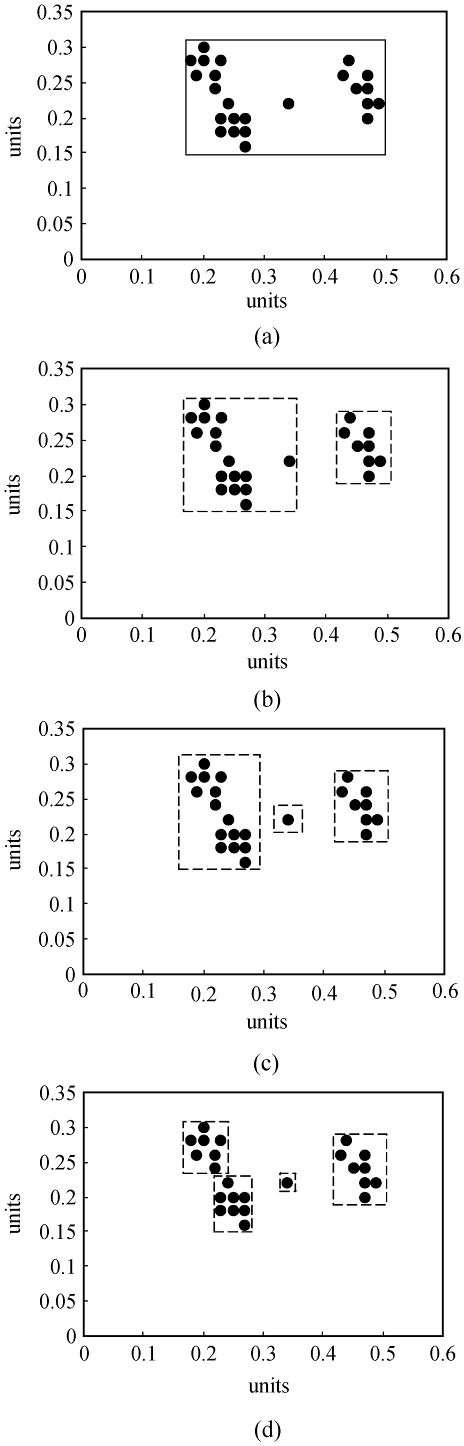

Fig. 6. A demonstration of the clustering performance of the proposed ENN-2 with four different distance parameters. (a) $\lambda = 0.15$; (b) $\lambda = 0.1$; (c) $\lambda = 0.0615$; (d) $\lambda = 0.05$.

the three types of wines. One hundred seventy-eight data vectors are given distributed by 59, 71, and 48 in three wine types. Because the proposed ENN-2 is an unsupervised learning pattern clustering neural network, so the clustering performance of the proposed ENN-2 is compared with the other unsupervised learning algorithms. Table III summarizes the classification results by different clustering methods that include ART network [4], k-means algorithm [21], and fuzzy-C-means algorithm [22]; all clustering results are the average of ten random trials. It is

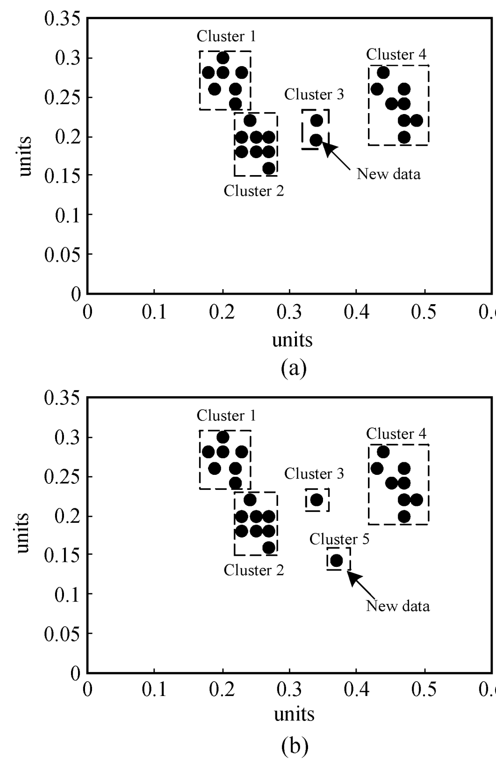

Fig. 7. The stability and plasticity test of the proposed ENN-2.

TABLE III
RECOGNITION RESULTS FOR THE WINE BENCHMARK BY VARIOUS METHODS

Method	Average Learning epoches	Average Right on Testing (%)
ART	11.4	94.64
K-Means	12.3	92.74
Fuzzy-C-Means	10.6	96.45
ENN-2	2.5	96.62

clear that the proposed ENN-2 has a shorter learning time than the other unsupervised learning algorithms. Also, the accuracy rates are quite high, with about 96.62% for this recognition data.

C. Breast Cancer Benchmark

This breast cancer databases was obtained from the University of Wisconsin Hospitals, Madison, from Dr. W. H. Wolberg [19], [23]. The aim was the correct identification of the type of breast cancer in one of two classes from a vector of ten attributes. Six hundred ninety-nine data vectors are given distributed by 458 (65.5%) benign and 241 (34.5%) malignant cancers. There are 16 instances that contain a single missing (i.e., unavailable) attribute value; thus only 683 data vectors are tested in this paper. Table IV summarizes the classification results by different clustering methods that include ART network, k-means algorithm, and fuzzy-C-means algorithm; all clustering results are the average of ten random trials. It is clear that the proposed ENN-2 has a shorter learning time and higher accuracy rates than the other unsupervised learning algorithms.

TABLE IV
RECOGNITION RESULTS FOR THE BREAST CANCER BENCHMARK BY VARIOUS METHODS

Method	Average Learning epoches	Average Right on Testing (%)
ART	13.5	92.43
K-Means	14.6	90.47
Fuzzy-C-Means	12.6	94.55
ENN-2	2.7	95.59

TABLE V
TESTED DATA OF GENERATOR SETS

Gen. No.	Input data <0.4f	0.4f~0.5f	1f	2f	3f	>3f	Fault types
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

D. Vibration Diagnosis of Generator Sets

To demonstrate the practical applications of the ENN-2, a vibration diagnosis data set [24], [25], consisting of 15 patterns, is introduced as shown in Table V. The vibration diagnosis of the 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 [14], [24], [25], the typical six attributes of vibration frequency (amplitude of <0.4 f, 0.4 f $\sim$ 0.5 f, f, 2 f, 3 f, and >3 f) are selected for vibration fault diagnosis, and the detailed training data are shown in Table V.

To compare diagnosis performance, the diagnosis results with two different classification methods, i.e., multilayer perceptron (MLP) [24] and adaptive wavelets network (AWN) [25], are shown in Table VI. The two traditional neural networks were capable of pointing toward faults, but both need to learn about 2561 and 900 epochs before fault diagnosis. Contrarily, the proposed ENN-2 only needs one epoch to learn with equivalent accuracy, and the structure of the proposed ENN-2 is simpler than the other neural networks, with only nine nodes and 36 connections needed. Fig. 8 shows the clustering results with different distance parameters. The number of clusters created ranges from 1 (when the λ is the largest) to 15 (when the λ is the smallest). It is clear that while the distance parameters are set between 0.8 and 2.1, the ENN-2 always clusters this six-dimension data set of three clustering groups as its fault types in Table V. When the ENN-2 is finished clustering, the matter-element models of three fault types can be produced as shown in Table VII. Where R_i is the matter-element of the ith fault type, where $F = \{F_1, F_2, F_3,\}$ is the fault set, F_i is the ith

TABLE VI
LEARNING RESULTS USING DIFFERENT NEURAL NETWORKS

Classifiers	MLP [24]	AWN [25]	ENN-2
Structure	6-13-3	6-13-3	6-3
No. of connections	117	117	36
Learning speed (epochs)	2561	900	1
Accuracy	100%	100%	100%

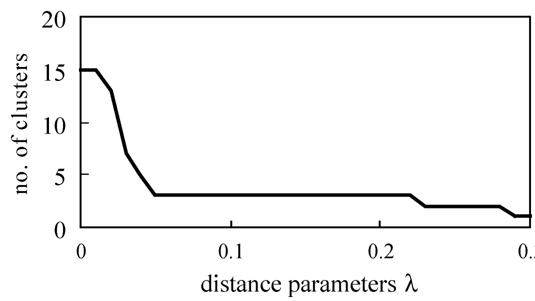

Fig. 8. The number of clusters with different distance parameters.

TABLE VII
FAULT MATTER-ELEMENT MODELS OF GENERATOR SETS

Fault types	Matter-element models
F_1: Oil-resonance fault	$R_1 = (F_1, C, V_1) = \left\{\begin{array}{lr} F_1, < 0.4f, & \langle 1.7, 7.1\rangle \\ 0.4f \sim 0.5f, & \langle 45, 50.4\rangle \\ f, & \langle 9.7, 15.1\rangle \\ 2f, & \langle 0.1, 5.3\rangle \\ 3f, & \langle 0., 4.5\rangle \\ > 3f, & \langle 0., 5.2\rangle \end{array}\right\}$
F_2: Imbalance fault	$R_2 = (F_2, C, V_2) = \left\{\begin{array}{lr} F_2, < 0.4f, & \langle 0.1\ 4.\rangle \\ 0.4f \sim 0.5f, & \langle 0.2, 4.8\rangle \\ f, & \langle 48.9\ 54.3\rangle \\ 2f, & \langle 3.1, 8.5\rangle \\ 3f, & \langle 0.8, 5.1\rangle \\ > 3f, & \langle 0.1, 4.7\rangle \end{array}\right\}$
F_3: Misalignment fault	$R_3 = (F_3, C, V_3) = \left\{\begin{array}{lr} F_3, < 0.4f, & \langle 0.3, 3.78\rangle \\ 0.4f \sim 0.5f, & \langle 0.2, 4.1\rangle \\ f, & \langle 22.3, 27.7\rangle \\ 2f, & \langle 21.4, 26.7\rangle \\ 3f, & \langle 15, 20.3\rangle \\ > 3f, & \langle 7.8, 13.2\rangle \end{array}\right\}$

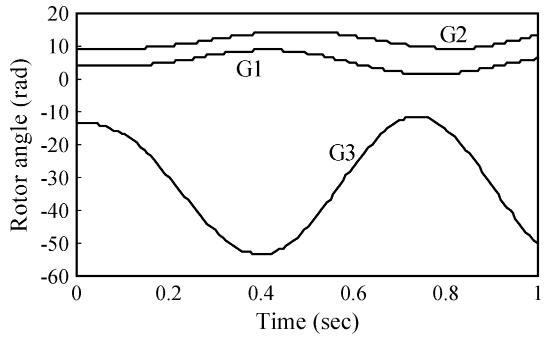

Fig. 9. The typical swing curves of the generators.

fault type. The up-bound and low-bound of every feature can be taken from the weights of ENN-2. It is of great importance to recognize incipient faults in generator sets, so that maintenance engineers can switch them safety and improve the reliability of power supplies. It should be noted that the proposed ENN-2 can produce meaningful output after the learning because the classified boundaries of the features can be self-organized by tuning the weights of ENN-2.

E. Coherency Identification of Large Power Systems

In the dynamic study of large power systems, it is necessary to model the external system by dynamic equivalents to improve the solution speed and to reduce the problem to solvable size. One approach to build up a dynamic equivalent is to identify generators in the external system with high coherency. A group of generators in the external system is said to be coherent if each pair of generators in the group is similar in terms of dynamic behavior. Basically, the coherent measurement is dependent on the tolerance of the rotor angle deviations of the generator. Fig. 9 shows the swing curves of three generators G1, G2, and G3. It is clear that the generator pair (G1, G2) is higher in coherence than the generator pairs (G1, G3) and (G2, G3). To verify the effectiveness of the proposed ENN-2 on large-scale systems, a comprehensive test at Taiwan power systems (Taipower) was conducted. Taipower is the only power system in Taiwan; it has a longitudinal structure covering a distance of 400 km from north to south. This system contains 34 generators, and the highest transmission system voltage is 345 kV. It is divided into three areas: north, central, and south, as shown in Fig. 10. In agreement with our past study [26], the typical three values (rotor speed $\omega(t_c)$ at the instant of fault clearing, $\omega(t_c + 0.2)$, and $\omega(t_c + 0.4)$) are selected for coherency identification.

Fig. 11 shows the number of coherent groups with different distance parameters in Taipower system while a three-phase fault at bus #53 occurs and the fault is cleared in 0.2 s. Obviously, the set of the resulting clusters can only decrease as the distance parameters increase in value. It should be noted that the presented method always converges with the maximum number of iterations not exceeding two, which is also less than our past results [26]. According to our past results [26], the k-means algorithm and ART neural network method need about four iterations. Table VIII shows typical coherent groups with three different distance parameters, and Figs. 12–14 show the different degrees of coherent groups with three different distance parameters. Usually, smaller distance parameters can produce a higher degree of coherence generator groups. The distance parameter is a user-made parameter that must be judiciously chosen on the basis of engineering experience.

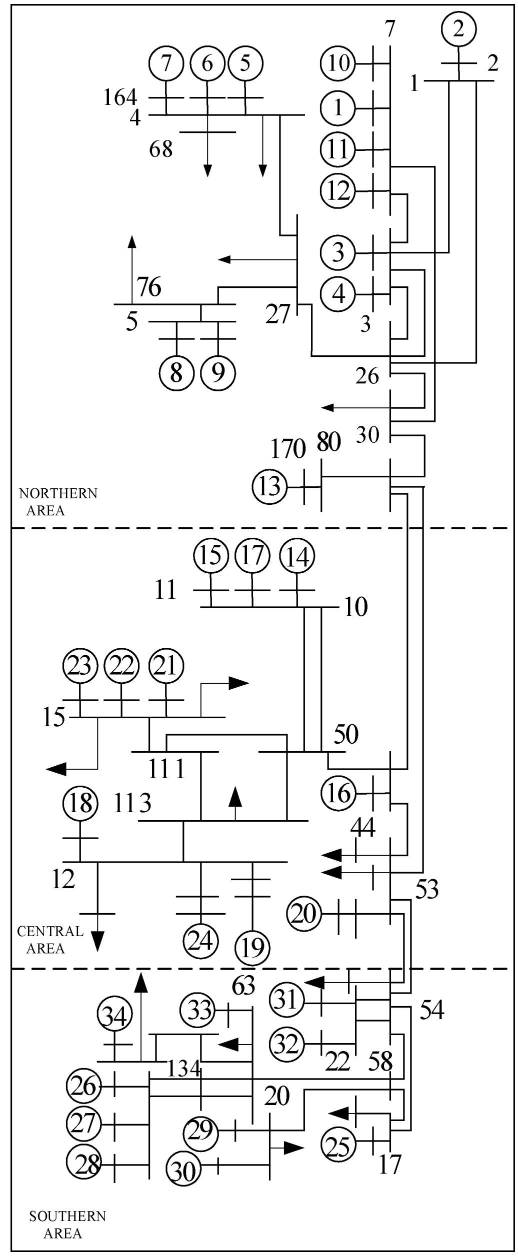

Fig. 10. The one-line diagram of the Taipower system.

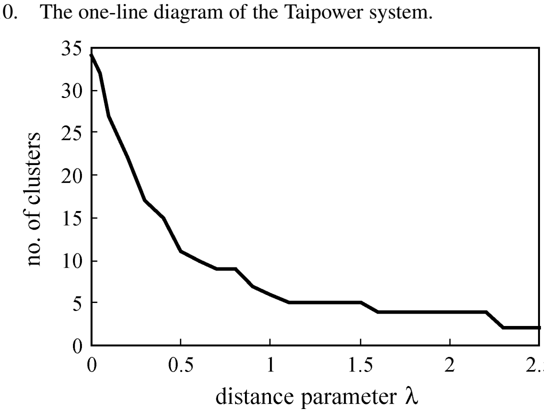

Fig. 11. The number of clusters with different distance parameters in Taipower system.

TABLE VIII
IDENTIFIED COHERENT GROUPS WITH DIFFERENT DISTANCE PARAMETERS

Distance parameters λ	Coherent groups
0.3	Group1: G1-G4, G11, G12, G21,G23
	Group 2: G5
	Group 3: G6, G26, G27, G34
	Group 4: G7, G8, G10, G30
	Group 5: G9, G29, G33
	Group 6: G13
	Group 7: G14
	Group 8: G15
	Group 9: G16, G19
	Group 10: G17
	Group 11: G18
	Group 12: G20
	Group 13: G22, G32
	Group 14: G24
	Group 15: G25
	Group 16: G28
	Group 17: G31
0.6	Group 1: G1-G4, G7, G8, G10-G12, G21, G23, G30
	Group 2: G5
	Group 3: G6, G26, G27, G31, G34
	Group 4: G13
	Group 5: G14-G17, G20
	Group 6: G16, G18, G19
	Group 7: G22, G32
	Group 8: G24
	Group 9: G25
	Group 10: G9, G28, G29, G33
1.0	Group 1: G1-G4, G6-G12, G21-G23, G25-G34
	Group 2: G5
	Group 3: G13
	Group 4: G14, G15, G17, G20
	Group 5: G16, G18, G19
	Group 6: G24

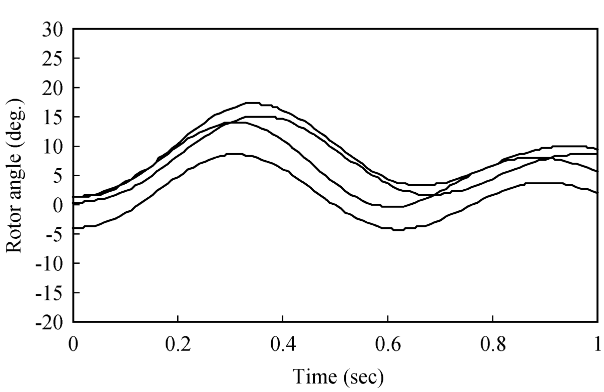

Fig. 12. Coherent group 3 of the Taipower system while distance parameter equals 0.3.

V. CONCLUSIONS

A novel extension neural network type 2 based on the author's earlier research is proposed in this paper. Compared with traditional neural networks and other traditional clustering methods,

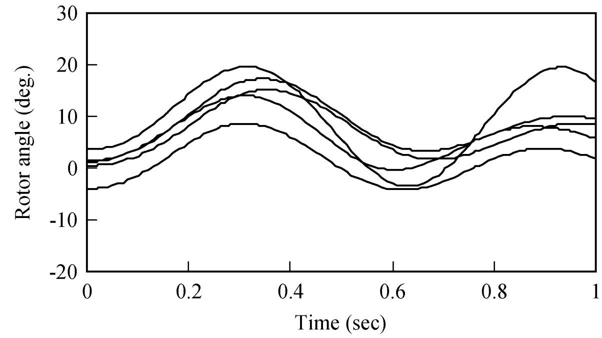

Fig. 13. Coherent group 3 of the Taipower system while distance parameter equals 0.6.

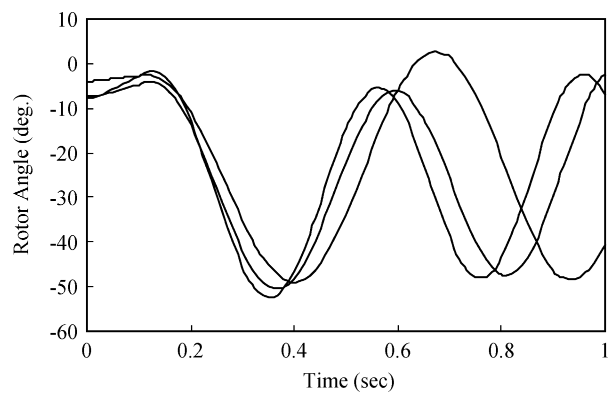

Fig. 14. Coherent group 5 of the Taipower system while distance parameter equals 1.

it permits an adaptive process for significant and new information, and can keep stability and plasticity characteristics at the same time. The proposed ENN-2 can produce meaningful output after learning because the classified boundaries of the features can be clearly determined by tuning the weights of ENN-2, so the matter-element models of the clustering problem can be easy built. Moreover, due to the simpler structure of the proposed ENN-2, it can keep the system low in computation cost and in memory consumption, which is a significant advantage in the modern commodities. From the tested examples, the proposed ENN-2 has been proved to have the advantage of less learning time, higher accuracy, and less memory consumption. It is also more efficient in engineering applications.

REFERENCES

[1] T. Kohonen, *Self-Organization and Associative Memory*, 2nd ed. Berlin, Germany: Springer-Verlag, 1988.
[2] D. F. Specht, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," *IEEE Trans. Neural Netw.*, vol. 1, pp. 111–121, 1990.
[3] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comput. Vision, Graph., Image Process.*, vol. 37, pp. 54–115, 1987.
[4] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition," *Neural Netw.*, vol. 4, pp. 493–504, 1991.
[5] F. Y. Shih, J. Moh, and F. Chang, "A new ART-based neural architecture for pattern classification and image enhancement without prior knowledge," *Pattern Recognit.*, vol. 25, pp. 533–542, 1992.
[6] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing Explorations in the Microstructure of Cognition*. Cambridge, MA: Bradford Books/MIT Press, 1986.

[7] K. C. Tan and H. J. Tang, "New dynamical optimal learning for linear multilayer FNN," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1562–1570, 2004.
[8] R. A. Felix, E. N. Sanchez, and G. Chen, "Reproducing chaos by variable structure recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1450–1457, 2004.
[9] G. Wallis, "Stability criteria for unsupervised temporal association networks," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 301–311, 2005.
[10] V. Petridis and V. G. Kaburlasos, "Fuzzy lattice network (FLNN): A hybrid model for learning," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, pp. 877–890, Sep. 1998.
[11] V. G. Kaburlasos and V. Petridis, "Fuzzy lattice neurocomputing (FLN) models," *Neural Netw.*, vol. 13, pp. 1145–1169, 2000.
[12] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, pp. 762–767, 1989.
[13] Z. Uykan and H. N. Koivo, "Analysis of augmented input layer RBFNN," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 364–369, 2005.
[14] M. H. Wang and C. P. Hung, "Extension neural network and its applications," *Neural Netw.*, vol. 16, pp. 779–784, 2003.
[15] W. Cai, "The extension set and incompatibility problem," *J. Sci. Explor.*, vol. 1, pp. 81–93, 1983.
[16] P. K. Simpson, "Fuzzy min-max neural networks-part 2: clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 32–45, 1993.
[17] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 888–905, Aug. 2000.
[18] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks-part I," *IEEE Trans. Neural Netw.*, vol. 13, pp. 645–661, 2002.
[19] C. J. Merz and P. M. Murphy, *UCI Repository of Machine Learning Databases*. Irvine: Dept. of Information and Computer Science, Univ. of California, 1996.
[20] J. Zhang, "Selecting typical instances in instance-based learning," in *Proc. 9th Int. Machine Learning Conf.*, Aberdeen, Scotland, 1992, pp. 470–479.
[21] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[22] H. J. Zimmerman, *Fuzzy Set Theory and Its Application*. Dordrecht, Holland: Kluwer, 1991.
[23] H. W. William and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in *Proc. Nat. Acad. Sci.*, vol. 87, Dec. 1990, pp. 9193–9196.
[24] H. Li, C. X. Sun, R. J. Liao, W. G. Chen, and X. S. Hu, "Improved BP algorithm in vibration fault diagnosis of steam turbine-generator set," *J. Chongqing Univ.*, vol. 22, pp. 47–52, 1999.
[25] H. Li, C. X. Sun, X. S. Hu, G. Yue, N. F. Tang, and K. Wang, "Study of method of adaptive wavelets network for vibration fault diagnosis of steam turbine-generation set," *Trans. China Electrotech. Soc.*, vol. 15, pp. 52–57, 2000.
[26] M. H. Wang and H. C. Chang, "Novel clustering method for coherency identification using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 9, pp. 2056–2062, 1994.

Mang-Hui Wang (S'90–M'94) was born in Taiwan, R.O.C., on June 22, 1963. He received the M.S. and Ph.D. degrees from the Electrical Engineering Department, National Taiwan University of Science and Technology, in 1990 and 1994, respectively.

He is currently a Professor of Institute of Information and Electrical Energy at the National Chin-Yi Institute of Technology, Taiwan. His major areas of research include power system stability and AI application to power systems.

Prof. Wang is a member of TAAI and CAAI.