A Fuzzy-Possibilistic Neural Network to Clustering

Shao-Han Liu and Jzau-Sheng Lin

Department of Electronic Engineering, National Chin-Yi Institute of Technology,

No. 35, Lane 215, Sec. 1, Chung-Shan Rd., Taiping, Taichung, Taiwan, R.O.C.

TEL: +886-4-3924505 ext 7311 FAX: +886-4-3926610

E-MAIL: jslin@chinyi.ncit.edu.tw

Abstract

Fuzzy Clustering has been proven to be advantageous over crisp clustering in some applications such as pattern recognition, image segmentation, and compression. In this paper, a new Hopfield-model net based on fuzzy possibilistic reasoning is proposed to clustering problem. The main purpose is to modify the Hopfield network and embed Fuzzy Possibilistic Fuzzy C-Means (FPCM) method to construct a classification system named Fuzzy-Possibilistic Hopfield Net (FPHN). The classification system is paradigms for the implementation of fuzzy logic systems in neural network architecture. Instead of one state in a neuron for the conventional Hopfield nets, each neuron occupies 2 states called membership state and typicality state in the proposed PFHN. The proposed network not only solves the noise sensitivity fault of Fuzzy C-Means (FCM) but also overcomes the simultaneous clustering problem of Possibilistic C-Means (PCM) strategy. In addition to the same characteristics as the possibilistic fuzzy c-means algorithm, the designed neural-network-based approach is self-organized structure that is highly interconnected and can be implemented in a parallel manner. The experimental results show that the proposed FPHN can obtain promising solutions.

Key words: *possibilistic c-means, fuzzy-possibilistic c-means, Hopfield neural network*

1. Introduction

Clustering has been an indispensable paradigm to unsupervised pattern recognition. Generally speaking, conventional methods such as *K*-means (*C*-means) [1] and ISODATA [2] are traditional clustering methods in which each sample belonging only one cluster. FCM [2]-[5], PFCM [6]-[7] and CFCM [8] are called fuzzy clustering methods in which every sample belonging all clusters with different degrees of membership. Every sample belongs all clusters with different degrees of possibility in possibilistc clustering algorithm [9-10]. Fuzzy-possibilistic c-means [11] solves the noise sensitivity fault of fuzzy c-means and the simultaneous clustering problem of possibilistic c-means strategy with membership and typicality.

In the application of optimization problem, neural networks exploit the massive parallelism of neurons. To update the performance, fuzzy reasoning algorithms have been added into neural network to construct fuzzy-neural systems [8, 12]. Kanstein *et al*. [13] embedded the possibilistic reasoning into a competitive learning network to clustering problem. Lin *et al*. [12] combined the penalized fuzzy c-means and competitive learning network to apply the multi-spectral image segmentation. Lin [8] also embedded the compensated fuzzy c-means into Hopfield net and applied to clustering. In this paper, the FPCM is added into Hopfield network to construct the FPHN to clustering. Additionally preserving the performance of fuzzy reasoning strategy, the FPHN not only solves the noise sensitivity fault of FCM but also overcomes the simultaneous clustering of the PCM. The FPHN can obtain promising solutions in clustering shown in experimental results.

The remainder of this paper is organized as follows. Section 2 reviews the fuzzy cluster technique. Possibilistic clustering techniques are presented in Section 3. Section 4, proposes fuzzy-possibilistic c-means strategy; Section 5 presents the Fuzzy Possibilistic Hopfield network (FPHN). Section 6 shows several experimental results; Finally, Section 7 gives the discussion and conclusions.

2. Fuzzy Clustering Techniques

Fuzzy clustering strategies are mathematical tools for detecting similarities between members of a collection of samples. The theory of fuzzy logic provides a mathematical framework to capture the uncertainties associated with the human cognition processes. Unlike the hard c-means method, in fuzzy c-means each training sample belongs to every cluster with some degree of membership. The purpose of the FCM approach, like the conventional clustering techniques, is to group data into clusters of similar items by minimizing a least squared error measure. For $c \geq 2$ and m>1, the algorithm chooses $\mu_x : Z \rightarrow [0,1]$ so that $\sum_x \mu_x = 1$ and $\varpi_i \in R^d$ for i=1, 2, ..., c to minimize the objective function

$$J_{FCM} = \frac{1}{2}\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i})^m \| z_x - \varpi_i \|^2 \tag{1}$$

where $\mu_{x,i}$ is the value of the ith membership grade on the xth sample z_x. The cluster centers $\varpi_1,...,\varpi_j,...,\varpi_c$ can be regarded as prototypes for the clusters represented by the membership grades. For the purpose of minimizing the objective function, the cluster centers and membership grades are chosen so that a high degree of membership occurs for samples close to the corresponding cluster centers. The membership grades and cluster centers are iteratively updated by the following formulas

$$\mu_{x,i} = \left(\sum_{\lambda=1}^{c} \frac{(\| z_x - \varpi_i \|^2)^{1/(m-1)}}{(\| z_x - \varpi_\lambda \|^2)^{1/(m-1)}} \right)^{-1}; \quad x = 1,2,...,n; \quad i = 1,2,...,c. \tag{2}$$

and

$$\varpi_i = \frac{1}{\sum_{x=1}^{n}(\mu_{x,i})^m} \sum_{x=1}^{n}(\mu_{x,i})^m z_x \tag{3}$$

The value $m \in$ (1, ∞) is the fuzzification parameter (or exponential weight). This parameter reduces the sensitivity of the class centers to noise in the data.

3. Possibilistic Clustering Techniques

The theory of fuzzy logic provides a mathematical environment to capture the uncertainties same as human cognition processes. The fuzzy clusters are generated by partition the training samples in accordance with the membership functions matrix $U = [\mu_{x,i}]$. The component $\mu_{x,i}$ denotes the grade of membership that a training sample belongs to a cluster. The fuzzy c-means algorithms use the probabilistic constraint to make the memberships of a training sample across clusters must sum to 1 that means the different grades of a training sample shared by distinct clusters but not as degrees of typicality. In contrast, each component generated by the possibilistic c-means (PCM) corresponds to a dense region in the data set. Each cluster is independent of the other clusters in the PCM strategy. The PCM strategy was proposed by Krishnapuram *et al.* [9-10] for unsupervised clustering. The objective function of the PCM can be formulated as

$$J_{PCM} = \frac{1}{2}\sum_{x=1}^{n}\sum_{i=1}^{c}(t_{x,i})^\eta \| z_x - \varpi_i \|^2 + \sum_{x=1}^{n}\beta_i \sum_{i=1}^{c}(1 - t_{x,i})^\eta \tag{9}$$

where

$\beta_i = \frac{\sum_{x=1}^{n} t_{x,i}^\eta \| z_x - \varpi_i \|^2}{\sum_{x=1}^{n} t_{x,i}^\eta}$, is the scale parameter at the *ith* cluster.

$t_{x,i} = \frac{1}{1 + \left(\| z_x - \varpi_i \|^2 / \beta_i \right)^{1/(\eta-1)}}$, Possibilistic typicality value of training sample z_x belonging to the cluster i.

$\eta \in [1, \infty)$, is a weighting factor called the possibilistic parameter.

4. Fuzzy-Possibilistic C-Means

Memberships and typicalities are both important for correct feature of data substructure in clustering problem. If a training sample been classified to a suitable cluster, membership is a better constraint for which the training sample is closest to this cluster. On the other word, typicality is an important factor for unburdening the undesirable effects of outliers to compute the cluster centers. In accordance with reference [11], typicality is related to the mode of the cluster and can be calculate based on all n training samples. Thus an objective function in the fuzzy-possibilistic c-means (FPCM) can depend on both of memberships and typicalities and be defined as

$$J_{FPCM} = \frac{1}{2}\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}{}^m + t_{x,i}^\eta) \| z_x - \varpi_i \|^2$$

where memberships, typicalities, and centroids are

$$\mu_{x,i} = \left(\sum_{\lambda=1}^{c} \frac{(\| z_x - \varpi_i \|^2)^{1/(m-1)}}{(\| z_x - \varpi_\lambda \|^2)^{1/(m-1)}} \right)^{-1}; \quad (10)$$

$$x = 1,2,\ldots,n; \quad i = 1,2,\ldots,c.$$

$$t_{x,i} = \left(\sum_{y=1}^{n} \frac{(\| z_x - \varpi_i \|^2)^{1/(\eta-1)}}{(\| z_y - \varpi_i \|^2)^{1/(\eta-1)}} \right)^{-1}; \quad (11)$$

$$y = 1,2,\ldots,n; \quad i = 1,2,\ldots,c.$$

and

$$\varpi_i = \frac{1}{\sum_{y=1}^{n} (\mu_{x,i}{}^m + t_{x,i}^\eta)} \sum_{x=1}^{n} (\mu_{x,i}{}^m + t_{x,i}^\eta) z_x \quad (12)$$

In the FPCM, membership $\mu_{x,i}$. is a function of training sample and all c cluster centers while the typicality $t_{x,i}$ is a function of training sample z_x and cluster center ϖ_i . Thus typicality $t_{x,i}$ does just depend on the location of the cluster center ϖ_i .

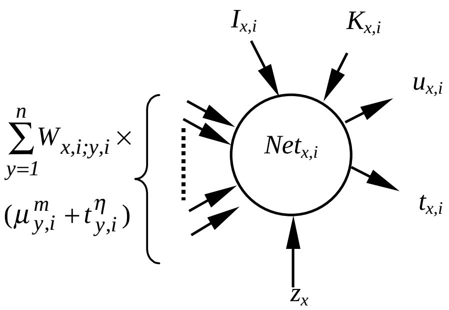

Figure 1. Architecture of the neuron (x, i) in a 2-D FPHN

5. Fuzzy-Possibilistic Hopfield Neural Network

The Hopfield-model neural networks [14-16] have been studied extensively. The features of this network are simple architecture and clear potential for parallel implementation. In order to update the performance in the application of optimal problems, modified Hopfield networks [17-20] have been proposed. Lin *et al.* [8, 17-19] proposed different fuzzy Hopfield networks to the applications of clustering problem and medical image segmentation. Cheng *et al.* [20] presented a possibilistic Hopfield network on CT brain hemorrhage image segmentation. These modified Hopfield networks base either fuzzy reasoning or possibilistic learning. For the purpose of solving the noise sensitivity fault of fuzzy reasoning and the simultaneous clustering problem of possibilistic learning, the fuzzy-possibilistic c-means strategy is embedded into Hopfield network to construct the FPHN. In the FPHN, shown in Figure 1, each neuron occupies 2 states named membership state based on all c cluster centers and typicality state based on all n training samples individually. Thus the total weighed input for neuron (x,i) and Lyapunov energy function in the FPHN can be modified as

$$Net_{x,i} = \left| z_x - \sum_{y=1}^{n} W_{x,i;y,i} (\mu_{y,i}^m + t_{y,i}^\eta) \right|^2 + I_{x,i} + K_{x,i} \quad (13)$$

and

$$E = \frac{1}{2} \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^m + t_{x,i}^\eta) \left| z_x - \sum_{y=1}^{n} W_{x,i;y,i} (\mu_{y,i}^m + t_{y,i}^\eta) \right|^2 - \sum_{x=1}^{n} \sum_{i=1}^{c} (I_{x,i} \mu_{x,i}^m + K_{x,i} t_{x,i}^\eta) \quad (14)$$

where $\sum_{y=1}^{n} W_{x,i;y,i} (\mu_{y,i}^m + t_{y,i}^\eta)$ is the total weighed input received from the neuron (y,i) in column i, m and η are fuzzification and typicality parameters, $\mu_{x,i}$ and $t_{x,i}$ are membership state and typicality state at neuron (x,i), and $I_{x,i}$, $K_{x,i}$ are input biases for membership and typicality states at neuron (x,i) respectively. The network reaches an equilibrium state when the modified Lyapunov energy function is minimized. The objective function for clustering problem in the 2-D FPHN is defined as follows:

$$E = \frac{A}{2} \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^m + t_{x,i}^\eta) \times \left| z_x - \sum_{y=1}^{n} \frac{1}{\sum_{h=1}^{n} (\mu_{h,i}^m + t_{h,i}^\eta)} z_y (\mu_{y,i}^m + t_{y,i}^\eta) \right|^2 + \frac{B}{2} \left\{ \left[\sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i} + t_{x,i}) \right] - n - c \right\}^2 \quad (15)$$

where E is the objective function that accounts for the energies of all training samples in the same class, and z_x , z_y are the training samples at rows x and y in the FPHN, respectively.

The first term in Eq. (15) defines the Euclidean distance between the training samples in a cluster and that cluster's centers over c clusters with membership grade and typicality degree.

The second term guarantees that n training samples in **Z** are distributed among these c clusters. More specifically, the second term (the constrained term), imposes constraints on the objective function, and the first term minimizes the intra-class Euclidean distance from training samples to the cluster centers.

All the neurons in the same row compete with one another to determine the training sample represented by that row belongs to all clusters with membership grades and typicality degree respectively. In other words, the summation of the membership states in the same row equals 1 and the summation of the typicality states in the same column also equals 1. That is the total sum of membership states in all n rows equal n and the total sum of typicality states in all c columns equal c. This assures that all n samples will be classified into c classes. The objective function in these networks can be further simplified as

$$E = \frac{1}{2}\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^{m} + t_{x,i}^{\eta})\times \left| z_x - \sum_{y=1}^{n}\frac{1}{\sum_{h=1}^{n}(\mu_{h,i}^{m} + t_{h,i}^{\eta})} z_y (\mu_{x,i}^{m} + t_{x,i}^{\eta}) \right|^2 \quad (16)$$

By using Eq. (16), the minimization of E is greatly simplified, since Eq. (16) contains only one term, removing the need to find the weighting factors A and B. Comparing Eq. (16) with the modified Lyapunov function Eq. (14), the synaptic interconnection weights and the bias input for the proposed can be obtained as

$$W_{x,i;y,i} = \frac{1}{\sum_{h=1}^{n}(\mu_{h,i}^{m} + t_{h,i}^{\eta})} z_y , \quad (17)$$

$$I_{x,i} = 0 , \quad (18)$$

and

$$K_{x,i} = 0 \quad (19)$$

By introducing Eqs. (17), (18), and (19) into Eq. (13), the input of neuron (*x*,*i*) can be expressed as

$$Net_{x,i} = \left| z_x - \sum_{y=1}^{n}\frac{1}{\sum_{h=1}^{n}(\mu_{h,i}^{m} + t_{h,i}^{\eta})} z_y (\mu_{y,i}^{m} + t_{y,i}^{\eta}) \right|^2 . \quad (20)$$

Consequently, the neuron states at neuron (*x*,*i*) are given by

$$\mu_{x,i} = \left[\sum_{j=1}^{c} \left(\frac{Net_{x,i}}{Net_{x,j}} \right)^{1/m-1} \right]^{-1} \quad \text{for all i,} \quad (21)$$

and

$$t_{x,i} = \left[\sum_{y=1}^{n} \left(\frac{Net_{x,i}}{Net_{y,i}} \right)^{1/\eta-1} \right]^{-1} \quad \text{for all i.} \quad (22)$$

Directly mapping training samples to the two-dimensional neuron array, the FPHN is trained to update all neuron states in order to classify the input samples into feasible clusters when the defined energy function converges to near global minimum.

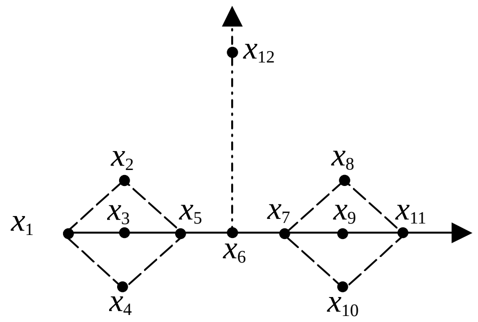

Figure 2. Coordinates of the data set

6. Experimental Results

To show the performance of the FPHN, a data set proposed by Pal *et al.* [11] and real multi-spectral images are used for simulation in an IBM compatible personal Pentium computer. The data set, shown in Figure 2, consists of 12 points on a 2-D coordinate given in Table 1. Initially, the states of neurons $\mu_{x,i}$ and $t_{x,i}$ are randomly set during 0 to 1. These two states for all neurons are modified iteratively to stable solutions as the defined Lyapunov energy function converging to a near-global minimum value. The cluster centers associated the run shown in Table 1 with *c*=2 are [(-3.19, 0.31), (3.19, 0.31)] and [(-3.20, 0.27), (3.20, 0.27)] for FCM and FPHN respectively. From these results, the centroids resulted by FPHN are more weakly influenced by point 12 than FCM. Table 2 shows the indices of the 12 points sorted by typicality values in each cluster. From Tables 1 and 2, the FPHN can also get the same promising results as FPCM. Same as the results in the FPCM, points 1-5 are most typical to cluster 1 and points 7-11 are also most typical to cluster 2. Points 6 and 12 with equal typicality values to both clusters, but point 12 is an order of magnitude smaller than the typicality value for point 6 that means point 6 belongs to both clusters with proper grades more strongly than point 12. This also means that the FPHN can prune outliers from the data to reduce the effects of noise.

The other example is multi-spectral image classification in MR head images of a patient diagnosed with cerebral infarction shown in Figure 3. These real images are acquired with T_2-weighted sequences for channel images CH = 1, 2, 4, and 5 and T_1-weighted signal for channel image 3. The acquisition parameters with different repetition time (TR) and echo time (TE) are TR_1/TE_1 = 2500 ms / 25ms, TR_2 / TE_2=2500ms / 50ms, TR_3 / TE_3=500ms / 20ms, TR_4/TE_4=2500ms / 75ms, and TR_5/TE_5=2500ms / 100ms respectively. Figure 4 shows the classified abnormal region with cerebral infarction. Experts indicated that the more promising result is obtained using the FPHN than those yielded by the fuzzy Hopfield neural network in reference [18].

Table 1. The membership grades and typicality degrees for FCM and FPHN

	Data set		FCM (m=3)		FPHN (m=3, η =3)			
x	p1	p2	$\mu_{x,1}$	$\mu_{x,2}$	$\mu_{x,1}$	$\mu_{x,2}$	$t_{x,1}$	$t_{x,2}$
1	-3.34	0.00	0.95	0.05	0.95	0.05	0.0227	0.0012
2	-3.34	1.67	0.96	0.04	0.96	0.04	0.0368	0.015
3	-3.34	0.00	1.00	0.00	1.00	0.00	0.8664	0.0016
4	-1.67	-1.67	0.92	0.08	0.92	0.08	0.0178	0.0014
5	-1.67	0.00	0.91	0.09	0.91	0.09	0.0287	0.0031
6	0.00	0.00	0.50	0.50	0.50	0.50	0.0067	0.0067
7	1.67	0.00	0.09	0.91	0.09	0.91	0.0028	0.0301
8	3.34	1.67	0.04	0.96	0.04	0.96	0.0015	0.0385
9	3.34	0.00	0.00	1.00	0.00	1.00	0.0016	0.8654
10	3.34	-1.67	0.08	0.92	0.07	0.93	0.0014	0.0193
11	5.00	0.00	0.05	0.95	0.05	0.95	0.0010	0.0210
12	0.00	10.00	0.50	0.50	0.50	0.50	0.0005	0.0005
Class center		(-3.19, 0.31) (3.19, 0.31)			(-3.20, 0.27) (3.20, 0.27)			

Table 2. The indices of the 12 points corresponding to a sort on $t_{x,1}$ and $t_{x,2}$

Typicality order	
$t_{x,1}$	$t_{x,2}$
3	9
2	8
5	7
1	11
4	10
6	6
7	5
9	3
8	2
10	4
11	1
12	12

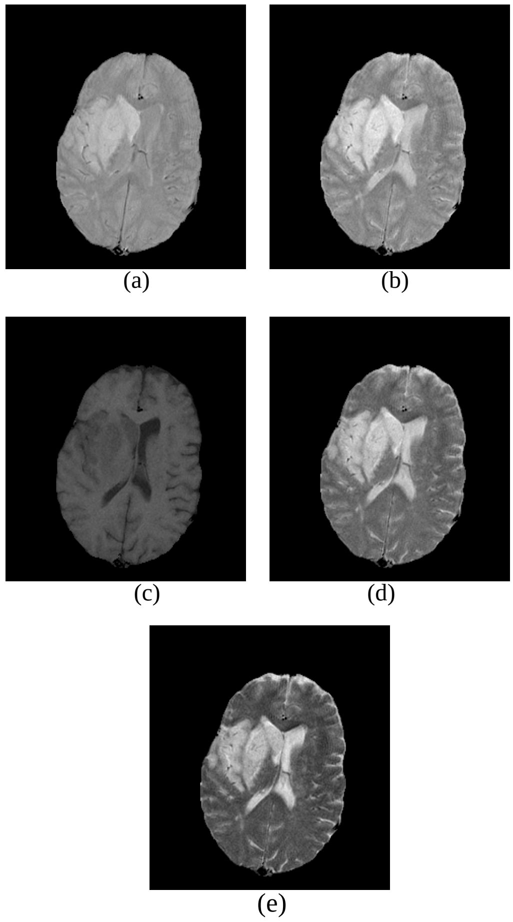

Figure 3. The multi-spectral MR head images with cerebral infarction: (a) TR_1/TE_1 = 2500 ms / 25ms; (b) TR_2 / TE_2=2500ms / 50ms; (c) TR_3 / TE_3=500ms / 20ms; (d) TR_4/TE_4=2500ms / 75ms; (e) TR_5/TE_5=2500ms / 100ms.

Figure 4. The classified image using the proposed FPHN in channel 2 with TR_2 / TE_2=2500ms / 50ms.

7. Discussion and Conclusions

A Modified Hopfield-net model called Fuzzy Possibilistic Hopfield Net (FPHN) embedded fuzzy possibilistc c-means strategy with 2 neuron states, membership state and typicality state, is proposed to clustering problem. Not only solves the noise sensitivity fault of fuzzy c-means but also overcomes the simultaneous clustering problem of possibilistic c-means strategy using the proposed FPHN. Therefore, the FPHN can prune outliers from the data to reduce the effects of noise. Moreover, the designed FPHN neural-network-based approach is a self-organized structure that is highly interconnected and can be implemented in a parallel manner. It can also be designed for hardware devices to achieve very high-speed implementation.

Acknowledgements

This work is supported by the National Science Council, ROC, under the Grant NSC89-2218–E -167-001.

References

[1] G. H. Ball and D. J. Hall, "A clustering technique for summarizing multivariate data," *Behav. Sci.*, vol. 12, pp. 153-155, 1967.

[2] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern*., vol. l.3, pp. 32-57, 1974.

[3] H. J. Zimmermann, *Fuzzy set theory and its application*, Cluwer:Boston, 1991.

[4] M. A. Ismail and S. Z. Selim, Fuzzy c-mean: optimality of solutions and effective termination of the algorithm, *Pattern Recognition*, vol. 19, pp. 481-485, 1986.

[5] J. C. Bezdek, *Fuzzy Mathematics in pattern classification*, in Ph.D. Dissertation, Applied Mathematics, Cornell University, Ithaca, New York, 1973.

[6] M. S. Yang, On a class of fuzzy classification maximum likelihood procedures, *Fuzzy Sets and Systems*, vol. 57, pp. 365-375, 1993.

[7] M. S. Yang, and C. -F. Su, On parameter estimation for normal mixtures based on fuzzy clustering algorithms, Fuzzy *Sets and Systems*, vol. 68, pp. 13-28, 1994.

[8] Jzau-Sheng Lin, "Fuzzy clustering using a compensated fuzzy Hopfield network," *Neural Processing Letters*, vol. 10, pp. 35-48, 1999.

[9] R. Krishnapuram and J. M. Keller, "A Possibilistic Approach to clustering," *IEEE Trans. On Fuzzy Systems*, vol. 1, pp. 98-110, 1993.

[10] R. Krishnapuram and J. M. Keller, "The Possibilistic C-Means Algorithm: Insights and Recommendations," *IEEE Trans. On Fuzzy Systems*, vol. 4, pp. 385-393, 1996.

[11] N. R. Pal, K., Pal, and J. C. Bezdek, "A Mixed c-Means Clustering Model," *IEEE International Conf. On Fuzzy Systems*, vol. 1, pp. 11-21, 1997.

[12] Jzau-Sheng Lin, K. –S. Cheng, and C. –W. Mao, "Segmentation of MultiSpectral Magnetic Resonance Image Using Penalized Fuzzy Competitive Learning Network," *Computers and Biomedical Research*, vol. 29, pp. 314-326, 1996.

[13] A. Kanstein, M. Thomas, and K. Goser, "Possibilistic Reasoning Neural Network," IEEE International Conference on Neural Networks, vol. 4, pp.2541-2545, 1997.

[14] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci.*, USA, vol. 79, pp. 2554-2558, 1982.

[15] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, pp. 141-152, 1985.

[16] Jzau-Sheng Lin and S. –H. Liu, "A Competitive Continuous Hopfield Neural Network for Vector Quantization in mage Compression," Eng. Appli. of Artif. Intell., vol. 12, pp. 111-118, 1999.

[17] Jzau-Sheng Lin, K. -S. Cheng, and C. –W. Mao, "A fuzzy Hopfield neural network for medical image segmentation," *IEEE Trans. Nucl. Sci.*, vol.43, pp.2389-2398, 1996.

[18] Jzau-Sheng Lin, K. -S. Cheng, and C. –W. Mao, "Multispectral magnetic resonance images segmentation using fuzzy Hopfield neural network," *J. Biomed. Comput.*, vol.42, pp.205-214, 1996.

[19] Jzau-Sheng Lin, "Segmentation of medical Images through a Penalized Fuzzy Hopfield Network with Moments Preservation," to be appeared in *J. of the Chinese Institute of Engineer*s.

[20] D. –C. Cheng, Q. Pu, K. –S. Cheng, and H. Burkhardt, " Possibilistic Hopfield Neural Network on CT Brain Hemorrhage Image Segmentation," Proc. of The 4th Asian Conf. On Computer Vision, vol. II, pp. 871-876, 2000.