An Efficient Fuzzy Classifier Based on Hierarchical Fuzzy Entropy

Cheng-Jian Lin[1], Chi-Yung Lee[2], and Shang-Jin Hong[1]

[1]Department of Computer Science and Information Engineering
Chaoyang University of Technology, Taichung, Taiwan 413, R.O.C.
[2]Department of Computer Science and Information Engineering
Nankai Institute of Technology, Nantou, Taiwan 542, R.O.C.

Abstract

In an earlier work, Lee *et al*. [1] presented a simple and fast fuzzy classifier that employed fuzzy entropy to evaluate pattern distribution information in a pattern space. In this paper, we extend his work to propose a new fuzzy classifier based on hierarchical fuzzy entropy (FC-HFE). We retained the main parts of the original structure and modified some methods (e.g., decision of the number of intervals on each dimension and class label assignment). Furthermore, the hierarchical fuzzy entropy is proposed for partitioning the decision region. The proposed FC-HFE can improve the classification accuracy and overcome some of the drawbacks in the Lee *et al* method. Finally, the FC-HFE is applied to evaluate the classification performance for iris and spiral databases. The simulation results show that the classification rate of the proposed FC-HFE is better than earlier methods.

Keywords: Hierarchical fuzzy entropy, fuzzy classifier, classification, iris and spiral.

I. Introduction

There are many classification problem methods at present. However, these classifiers are very complex and require much time to adjust and compute. In recent years, the fuzzy entropy applications have been widely adopted by some researches [2], [4]-[6]. In [1], Lee *et al*. presented an efficient fuzzy classifier with feature selection based on fuzzy entropy. They used fuzzy entropy measure instead of fuzzy rules to reflect the actual classification pattern distribution by computing the fuzzy entropy of each feature dimension. The non overlapping decision regions were then partitioned from the input feature space by fuzzy entropy. The decision regions were then automatically tuned according to the fuzzy entropy measure. Moreover, fuzzy entropy is employed to select the relevant features with good separability for the classification task. The computation speed of this classifier is much faster because the decision regions do not overlap. Both the computational load and complexity of the classifier are reduced and a feature selection method based on the fuzzy entropy increases the classification accuracy by discarding noise-corrupted, redundant, and unimportant features.

In this paper, we construct a new fuzzy classifier based on hierarchical fuzzy entropy. First, we redefine the class label assignment and add a shortest Euclid distance measure. Second, we redefine the condition that determines the number of intervals in each dimension. We also propose a new concept called hierarchical fuzzy entropy that details especially with the high fuzzy entropy decision regions in pattern space, and improves the classification accuracy further. The details for this concept are provided in Section 3.

The rest of this paper is organized as follows. In Section 2, we review briefly the method of Lee *et al*. [1]. A new fuzzy classifier based on hierarchical fuzzy entropy is proposed in section 3. In Section 4, some experimental results are shown that the accuracy classification of the proposed classifier is better than one of earlier classifier. Finally, conclusion is presented in Section 5.

II. Format Requirement

In [1], Lee *et al*. presented a simple and fast fuzzy classifier called Fuzzy Entropy-Based Fuzzy Classifier (FEBFC). It uses fuzzy entropy to divide the pattern space into decision regions enclosed by the surfaces produced from each dimension. The surfaces of each subspace are parallel to each dimension, i.e., the surfaces are extended from the membership function boundaries on each dimension. If only these decision regions are determined, these decision regions can be applied to classify the unknown patterns.

In the FEBFC method, the number of intervals (fuzzy set) in each dimension is very important because this has a profound effect on the learning efficiency and classification accuracy. The steps used in selecting the interval number for each dimension are described as follows:

step 1. Set the initial number of intervals I = 2.

step 2. Use a simple clustering algorithm to locate the center of each interval.

step 3. In order to apply fuzzy entropy to measure the pattern distribution information in an interval, a membership function must be assigned to each interval. In FEBFC, a triangular membership function is adopted for each interval.

step 4. Compute the total fuzzy entropy of all intervals for I and I-1 intervals.

step 5. If the total fuzzy entropy of I intervals is less than that of I-1 intervals, (i.e., partitioning this dimension into I intervals will be more “order” than into I-1 intervals), then partition again (I = I + 1) and go to *step2*; else stop to increase intervals on this dimension and decide the number of intervals on the next dimension.

Once the interval locations on each dimension are determined, the decision regions are divided simultaneously. The fuzzy entropy of the decision regions can be obtained via fuzzy entropy summation of the individual intervals in each dimension. A class label with the lowest fuzzy entropy is then assigned to each decision region.

After these five steps are executed, the fuzzy entropy of each feature dimension is computed. These feature dimensions that are helpful for classification speed and accuracy are selected using a backward elimination procedure. The backward elimination procedure selects the relevant features. The termination criterion is based on the classification rate. Because features with higher fuzzy entropy are less relevant to our classification goal, features that have the highest fuzzy entropy are removed if doing so does not decrease the classification rate. The above steps are repeated until all “irrelevant” features are removed. Finally, the remaining features serve as features for classification.

However, this method presents the following problems:

(1). In class label assignment for each decision region is performed according to the fuzzy entropy summation of the individual intervals in each dimension. The decision region is then assigned to the class with the lowest fuzzy entropy in this region. In terms of the fuzzy entropy axiom, the fuzzy entropy will be zero whether the matching degree of each class is zero or one. Therefore, the class assignment will make a mistake when the number of classes in which fuzzy entropy equals zero is more than one in this region.

(2). In FEBFC, increasing the number of intervals on each dimension is performed according to the rule that the total entropy of I (the number of intervals on a specified dimension) intervals is less than that for I-1 intervals. This rule sometimes will prevent the number of intervals from increasing, leading to incorrect classification. For example, the data are presented as an even intersecting distribution, like a spiral data base. As the fuzzy entropy of this data type is initially high, the data should be further divided to decrease the fuzzy entropy. Therefore, the stop condition for increasing the number of intervals will result in a limitation on increasing of number of intervals on each dimension.

(3). The pattern space was partitioned into non-overlapping decision regions using grid partition. A class label cannot be assigned to the decision region that has no training data. If a test pattern falls into this decision region, the system does not know how to classify it?

The above problems affect the FEBFC classification rate. We propose a new structure for solving the above problems. The difference between the new structure and FEBFC is explained in Section 3.

III. A New Method for Classification

The FEBFC problems explained in Section 2 have not been solved. We propose a fuzzy classifier based on hierarchical fuzzy entropy (FC-HFE) to solve these problems. We retain the main FEBFC structure and modify certain methods (ex. decision of the number of intervals on each dimension and class label assignment). A new method called hierarchical fuzzy entropy is used to partition the decision region.

We modify original method for determining the number of intervals first. The new method maintains the original condition for stopping the increase in the number of intervals when the total fuzzy entropy of **I** intervals is more than one of **I**-1 interval. The other condition for stopping the increase in intervals if the total fuzzy entropy of **I** intervals is more than one of **I**-1 interval and total fuzzy entropy of **I-1** intervals is less than the threshold φ(i.e., if the total fuzzy entropy of **I** intervals is less than one of **I**-1 interval or the total fuzzy entropy of **I** interval is more than the threshold φ, then **I= I** +1). **T**he value of threshold φ was obtained using the following equation:

$$\varphi = -N_{class} * \frac{1}{N_{class}} * \log_2(\frac{1}{N_{class}}) * (\mathrm{I}-1) * \theta = -\log_2(\frac{1}{N_{class}}) * (\mathrm{I}-1) * \theta . \quad \textbf{(1)}$$

where N_{class} is the number of data classes, and θ is a percentage of maximum total fuzzy entropy of **I-1** intervals. θ can be tuned using different classification problem. Therefore, we can produce enough intervals for to evenly intersect the data distribution using the proposed method. The fuzzy entropy of each dimension must be less than a specific value.

Another phenomenon was found in the FEBFC that value of fuzzy entropy for some decision region is always very high (i.e., the data distribution of decision region is very confused, or the data which is classified incorrectly are very much.) whether the pattern space is partitioned into how many decision regions by fuzzy entropy, preventing the classification rate from being increased.

We propose a method to deal properly with high fuzzy entropy decision regions called hierarchical fuzzy entropy. The structure is illustrated in Fig. 1. The 2D pattern space in Fig. 1 is partitioned unequally into 53 decision regions (i.e. sum of decision regions of each layer is 12+14+23+4=53) by hierarchical fuzzy entropy. The 1 to 4 decision region fuzzy entropy of layer 1 is very high. We partitioned these decision regions further. For example, decision region 1 was partitioned into eight decision regions the second time. The sub decision regions I and II were partitioned into four decision regions respectively the third time. Therefore, decision region 1 was partitioned into fourteen decision regions. The fuzzy entropy of most decision regions becomes low after the pattern space is partitioned using hierarchical fuzzy entropy.

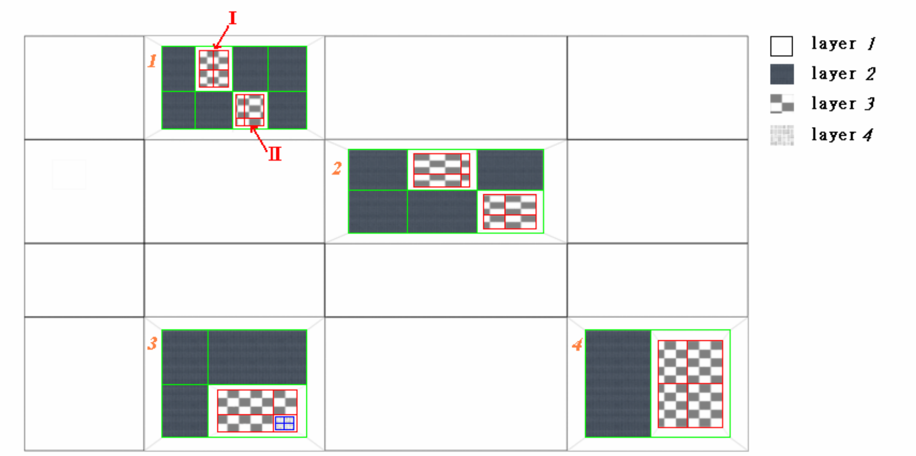


Fig. 1. A bird's eye view of decision regions partitioned using hierarchical fuzzy entropy.

Furthermore, we modify the method which to compute the fuzzy entropy of decision region. If there are no training data of a certain class on an interval, and then set the fuzzy entropy of this class equals one to avoid that the fuzzy entropy of certain class equals zero more than one for computing the lowest fuzzy entropy of a region. Following steps will describe the FC-HFE method:

step 1. ~ ***step 4.*** are the same with ones of FEBFC in selecting the number of intervals for each dimension in current pattern space .

step 5. If the total fuzzy entropy of **I** intervals is less than that of **I**-1 interval or the total fuzzy entropy of **I** interval is more than the threshold φ then partition again (**I** = **I** + 1) and go to ***step2***; Otherwise stop increase of interval on this dimension and decide the number of intervals on next dimension**.**

step 6. Once the intervals for each dimension are determined, the decision regions for the current pattern space are divided. The mean fuzzy entropy is computed using the following equation for all decision regions:

$$MFE_i = \sum_{j=1}^{N_i} FE_{ij} \Big/ N_i \,, \quad j = 1....N_i \,. \tag{2}$$

where N_i is the number of decision regions for the ith layer, FE_{ij} is the jth decision region for the ith layer and MFE_i is the mean fuzzy entropy of the ith layer.

step 7. The fuzzy entropy for every decision region in the current pattern space is compared with

MFE. If the fuzzy entropy decision regions are greater than mean fuzzy entropy, the region becomes an independent sub space and ***step 1****~* ***step 7*** are employed again to determine the decision regions for each sub space. Otherwise the decision region will be assigned a class label.

Above process is like a recursion function that is executed repeatedly for each sub space until all sub spaces cannot be partitioned anymore. When the above steps are completed, the test pattern is classified using the decision regions that have a class label. When testing pattern in a no class label region, the test pattern is computed using the short Euclid distance between the neighboring training data and test pattern, classified into the nearest training data class.

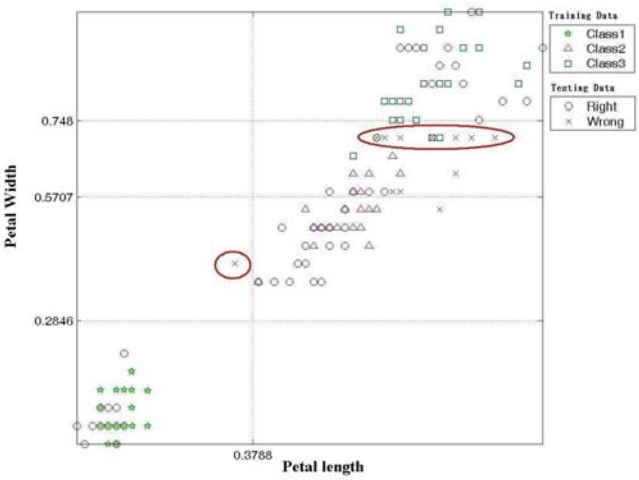
IV. Experimental Results

In this section, some experimental results present the effectiveness of the FC-HFE using the Iris and Spiral databases. The two databases were used for comparing the FC-HFE with FEBFC, under a series of random trials.

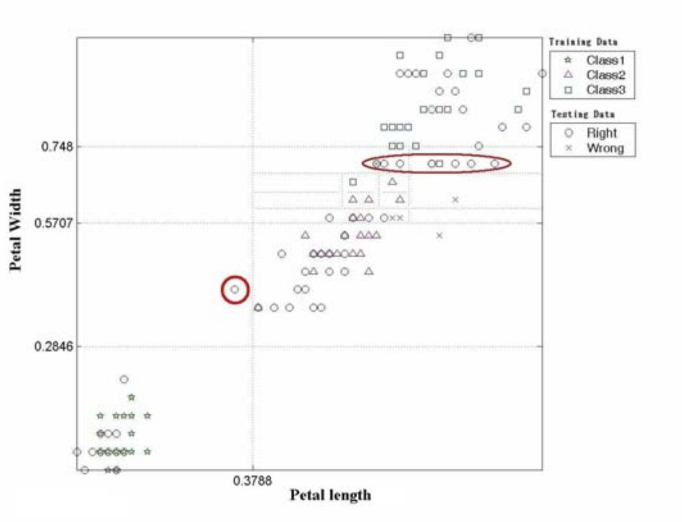
A. *Iris database*

The Iris database created by Fisher includes three classe. This database has four continuous features consisting of 150 instances: 50 for each class. One half of the 150 instances were randomly selected as the training set. The remaining patterns were used as the test set. Table 1 shows the result using the FEBFC and FC-HFE classifiers. The FC-HFE with $\theta = 0.6$ was used in this experiment. The FC-HFE and FEBFC experimental results were evaluated under two different circumstances: without feature selection (using the four original features) and with feature selection (using two selected features).

Furthermore, we will prove that the classifier recognition rate can be increased by hierarchical fuzzy entropy. The classification results with two features (petal length and petal width) were shown in Figs. 2 (a) and (b). The number of test error for the FEBFC is 11 in Fig. 2(a). The number of test error for the FC-HFE was decreased from 11 to 5 in Fig. 2(b). Because the region was partitioned into eleven decision regions by hierarchical fuzzy entropy and using the method of short Euclid distance measure, the data in the red circle were classified correctly.



(a)



(b)

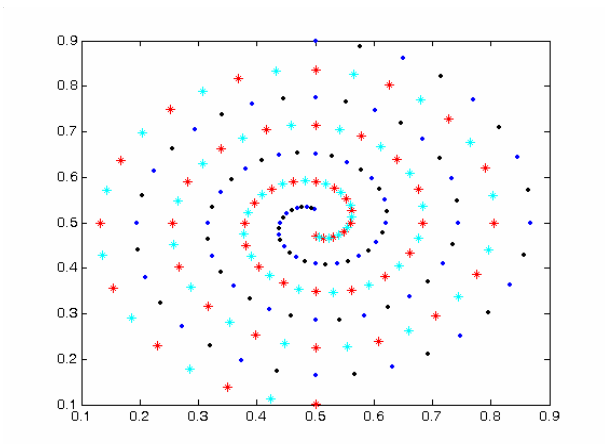
Fig. 2. (a) The Classification result of FEBFC with two features. (b) The Classification result of FC-HFE with two features.

Table 1. Classification Results on iris database

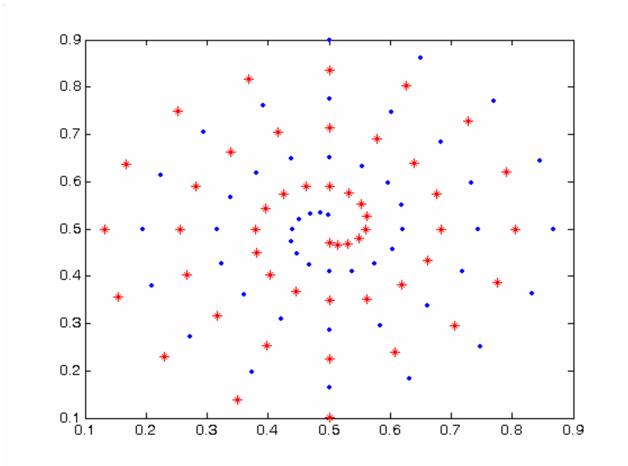
Models	**FEBFC** (4 feature)	**FC-HFE** (4 feature)	**FEBFC** (2 feature)	**FC-HFE** (2 feature)
Testing Recognition Rate(Average)	**95.99%**	**97.07%**	**97.33%**	**98%**

B. *Spiral database*

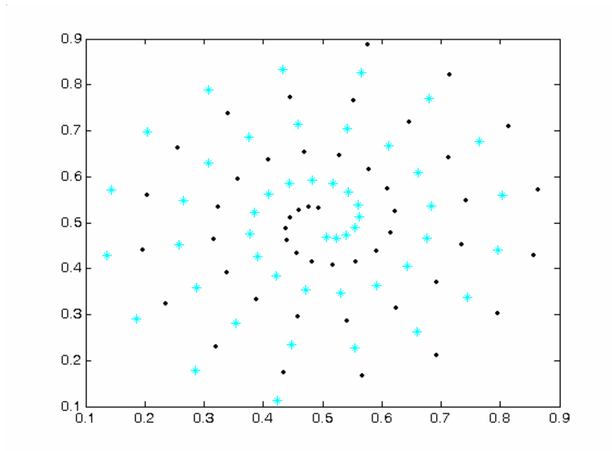
The input data points for the two-spiral problem [3] are shown in Fig. 3 (a). They are partitioned into a training set and a test set as shown in Figs. 3 (b) and (c), respectively. Table 2 shows result using the FEBFC and FC-HFE classifier, and FC-HFE with θ = 0.6 was used in this experiment. The classification results show that the FEBFC could not classify the spiral because the spiral data distribution is an evenly intersecting distribution. However, our proposed FC-HFE could provide correct classification. A diagram of the FEBFC and FC-HFE classification results is shown in Figs. 4(a) and (b).



(a)

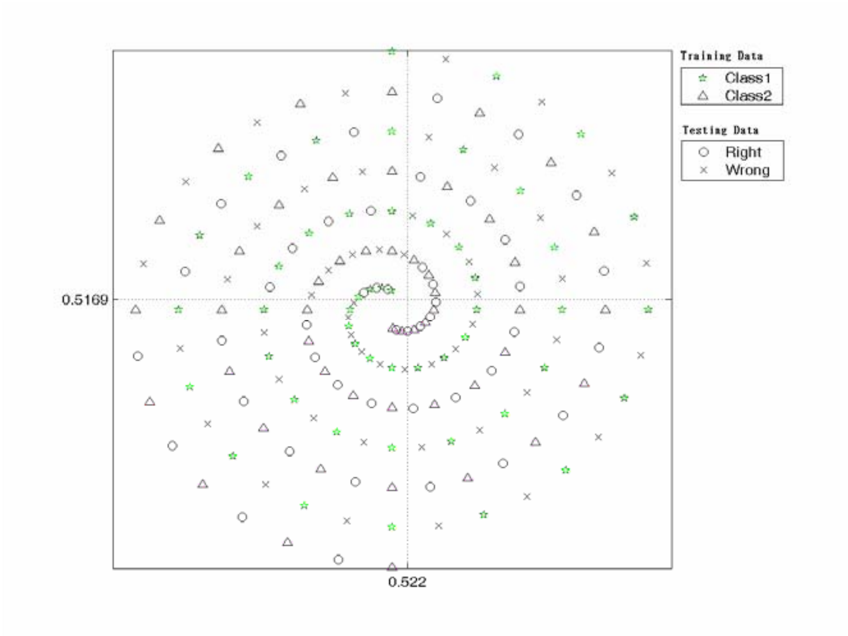


(b)

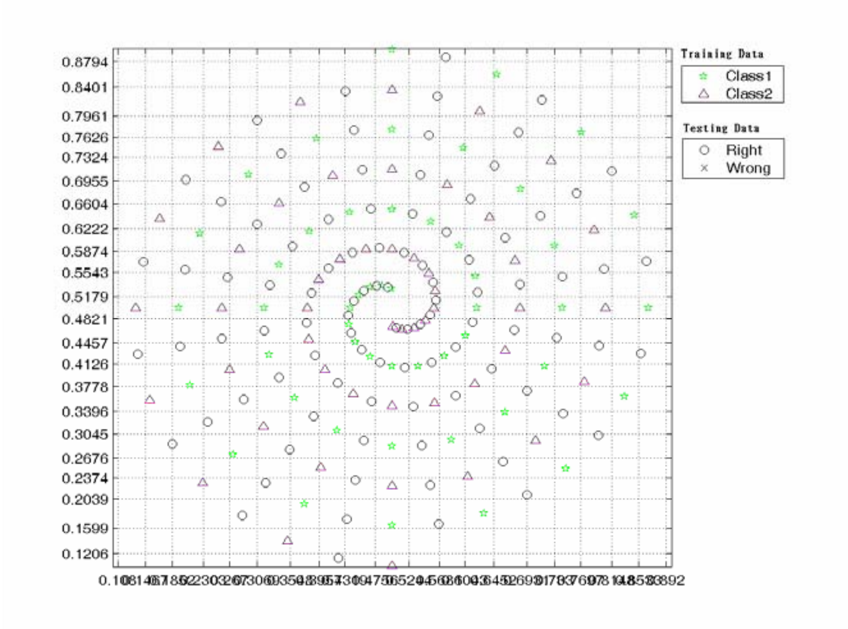


(c)

Fig. 3. (a) Input examples. (b) Training examples. (c) Testing examples.



(a)



(b)

Fig. 4. (a) The classification result of FEBFC for spiral. (b) The classification result of HFCFE for spiral.

Table 2. Classification Results on spiral database

Models	FEBFC	FC-HFE
Testing Recognition Rate	**49.47%**	**100%**

V. Conclusion

We proposed a new fuzzy classifier based on hierarchical fuzzy entropy (FC-HFE). It remain the structure of FEBFC that employs fuzzy entropy to evaluate the pattern distribution information in the pattern space, select the pattern space features, and increase some new methods for improving the classification rate and solving the drawbacks in the FEBFC.

We solve the problem involving not enough decision regions to evenly intersect data distribution in each dimension by redefining the condition in which the number of intervals in each dimension is determined. The high fuzzy entropy regions in the pattern space become clearer using hierarchical fuzzy entropy.

Our experimental results show that the FC-HFE has better average performance than the FEBFC. In addition, the FC-HFE classification rate exceeds the FEBFC rate very much in spiral database. This proves that the FC-HFE is an efficient classifier for various databases.

Acknowledgement

This research is supported by the National Science Council of R.O.C. under grant NSC 94-2213-E-324- 004.

References

[1] H.-M. Lee, C.-M Chen, J.-M. Chen and Y.-L. Jou, "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy", IEEE Trans. on Systems, Man and Cybernetics Part B, Vol. 31, No.3, pp. 426-432, June 2001.

[2] R. A. Fisher, "The use of multiple measurements in taxonomic problem", Ann Eugenics, Vol. 7, No. 2, pp. 179-188, 1936.

[3] H.C. Chua, , J. Jia, L. Chen, , Y. Gong, , "Solving the two-spiral problem through input data encoding", Electronics Letters, Vol.31, No. 10, pp.813-814, 11 May 1995.

[4] I. J. Rudas and M. O. Kaynak, "Entropy-Based Operations on Fuzzy Sets", IEEE Trans. on Fuzzy Systems, Vol. 6, No.1, pp. 33-40, 1998.

[5] N. R. Pal and S. K. Pal, "Entropy: A New Definition and It's Applications", IEEE Trans. on Systems, Man and Cybernetics, Vol.21, No.5, pp. 1260-1269, 1991.

[6] S. Di Zenzo, L. Cinque, and S. Levialdi , "Image Thresholding Using Fuzzy Entropies", IEEE Trans. on Systems, Man and Cybernetics Part B, Vol. 28, No.1, pp. 15-23, 1998.

Biography

Cheng-Jian Lin received the B.S. degree in electrical engineering from Ta-Tung University, Taiwan, R.O.C., in 1986 and the M.S. and Ph.D. degrees in electrical and control engineering from the National Chiao-Tung University, Taiwan, R.O.C., in 1991 and 1996. From April 1996 to July 1999, he was an Associate Professor in the Department of Electronic Engineering, Nan-Kai College, Nantou, Taiwan, R.O.C. Since August 1999, he has been with the Department of Computer Science and Information Engineering, Chaoyang University of Technology. Currently, he is a Professor of Computer

Science and Information Engineering Department, Chaoyang University of Technology, Taichung, Taiwan, R.O.C. He served as the chairman of Computer Science and Information Engineering Department from 2001 to 2005. His current research interests are neural networks, fuzzy systems, pattern recognition, intelligence control, bioinformatics, and FPGA design. He has published more than 100 papers in the referred journals and conference proceedings. Dr. Lin is a member of the Phi Tau Phi. He is also a member of the Chinese Fuzzy Systems Association (CFSA), the Chinese Automation Association, the Taiwanese Association for Artificial Intelligence (TAAI), the IEICE (The Institute of Electronics, Information and Communication Engineers), and the IEEE Computational Intelligence Society. He is an executive committee member of the Taiwanese Association for Artificial Intelligence (TAAI). Dr. Lin currently serves as the Associate Editor of International Journal of Applied Science and Engineering.

Chi-Yung Lee received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taiwan, R.O.C., in 1981 and the M.S. degrees in industrial education from the National Taiwan Normal University, Taiwan, R.O.C., in 1989. From August 1989 to July 2004, he was a lecturer in the Department of Electronic Engineering, Nan-Kai Institute of Technology, Nantou, Taiwan, R.O.C. Since August 2004, he has been with the Department of Computer Science and Information Engineering, Nan-Kai Institute of Technology. Currently, he is a lecturer of Computer Science and Information Engineering Department, Nan-Kai Institute of Technology, Nantou, Taiwan, R.O.C. His current research interests are neural networks, fuzzy systems, intelligence control, and FPGA design. He has published more than 20 papers in the referred journals and conference proceedings.

Shang-Jin Hong received the B.S. and M.S. degrees in computer science and information engineering from the Chaoyang University of Technology, Taiwan, R.O.C., in 2003 and 2005. His current research interests are fuzzy systems, pattern recognition, intelligence control, and evolutionary computation.