A Simple Method to Extract Fuzzy Rules by Measure of Fuzziness

Jieh-Ren Chang[*†], *Nai-Jian Wang*[*]
[*]*Department of Electrical Engineering, National Taiwan University of Science and Technology, Taiwan, ROC*
[†]*Department of Electronic Engineering, National Ilan University, Taiwan, ROC*
jrchang@niu.edu.tw, njwang@mail.ntust.edu.tw

ABSTRACT

To easily construct an efficient rule-based system for pattern classification problem is an important research topic. In this paper, we use a simple method to build up the fuzzy rules directly from numerical input-output data.

First, we extract fuzzy rules from different class regions which are confined by training data. These confined regions construct the basic element-nodes in the second layer for our proposed neural network structure. When these confined regions are overlapped, a recursive process is applied possibly to set up additive fuzzy rules in these uncertainty-overlap regions. The stop criterion is referred in a measure of fuzziness to keep creating the efficient rules for the recursive process. All additive fuzzy rules are added into this fuzzy-neural network. Finally, the method is compared with other algorithms using the Fisher iris data and a set of pseudo iris data for performance evaluation.

1: INTRODUCTIONS

In general, the methods of classification problem are divided into four groups as the following descriptions. 1) Statistical method: It was used in early classifier such as linear discriminate, quadratic discriminate, nearest neighbor, Bayes' independence and Bayes' second order methods. The Bayes' classifier was well known that attains best classification rate. However, it is not practical in solving classification problem in a real word since we need to know the probability density function of data as a priori [2]. 2) Neural network: It is a system that is deliberately constructed to make use of some organizational principles resembling those of human brain. In [3]-[5], it has good tasks for many applications. 3) Fuzzy inference engine: By querying experts' experience or other techniques directly from training data, the knowledge fuzzy rule database is created. It is applied to identify different pattern with fuzzy inference engine [6]-[11],[16]-[17]. 4) Hybrid neural-fuzzy technique: It combines the fuzzy inference and neural network theory to computer-based pattern recognition [12]-[15][18]-[19].

Although there are so many approaches to solve classification problem, conventional fuzzy-rule-based technique is still an easy method. Usually, a fuzzy system consists of several typical procedures such as fuzzification, defuzzification, rule creation and inference process. Hong and Lee have presented the drawbacks of most fuzzy controllers and fuzzy expert systems that they need to predefine membership functions and fuzzy rules to map numerical data to linguistic terms so as to make fuzzy reasoning work [9]. They proposed a method based on the fuzzy clustering technique to setup the decision tables. By these decision tables, the fuzzy variable membership functions and fuzzy rules can be derived from numerical data. However, they still need to determine the scaling in input variable. It usually takes more computation time to construct decision tables and merge operations as the attribute number or data scale becomes large.

Wu and Chen have a fuzzy learning algorithm [10] based on the α-cuts of equivalence relations and the α-cuts of fuzzy sets to construct the membership functions of the input/output variables of fuzzy rules and to induce the fuzzy rules from numerical training data set. By experiment on Iris data, it shows the algorithm reaches a higher average classification ratio and generates fewer rules than the existing algorithms. By this algorithm, we must decide many α-values to create the input-output-value subsets for each linguistic label of each linguistic variable. But, it didn't tell us how to select the α-values. In general, to describe the system character more precisely, we need more sample data. However, the dimension of equivalence relation matrix is augmented and the computation complexity increases when the training data increase. It will take a lot of cost in computation.

Hong and Chen propose the other fuzzy method to find relevant attributes which results in fewer dimensions in computation process [11]. They also predefine the number of fuzzy linguistic label for each attribute by a statistical formula. The purpose of the methods mentioned above is to decrease the computation time. But it still generates many rules and takes very much time on the computation process when the training data increase.

Some other approaches based on neural networks or fuzzy theory to extract knowledge or rules directly from training data are proposed, since it is not easy to elicit experienced knowledge from experts. P. K. Simpson defined Min-Max hyper-box [14] to set up the fuzzy rules by an expansion-contraction process, but it usually generated too many hyper-boxes that mean too many rules to be concerned. S. Abe and M-S. Lan extract the fuzzy rules in each activation hyper-boxes from numerical data by resolving overlaps between two classes recursively [1][8]. They said that it could decrease the computation burden due to large-scale problems in the learning process. But, there are still some drawbacks of this method in the following points. 1) It needs more computation time to resolve overlaps recursively by selecting two classes successively when the original data include many classes. 2) It sometimes can't be resolved in some critical condition. 3) It wastes much time to generate many meaningless fuzzy rules as the data are distributed uniformly in the regions of the feature space, but it can't improve the classification rate. Our purpose is to decrease the computation time and to extract more efficiency fuzzy rules by measure of fuzziness.

We know that even an expert could not make correct decision in some critical situation. Thus, we don't need to create too many rules in some uncertainty regions. It is our purpose to decrease the computation time and to extract more efficient fuzzy rules by measure of fuzziness. The proposed method is described in the following steps. 1) Finds the premised parts of fuzzy rule that extract directly from numerical training data. Each premised parts of fuzzy rule is named activation hyper-box as in [1]. 2) Finds uncertainty overlap regions among different activation hyper-boxes. 3) Extracts fuzzy rules from overlap uncertainty region recursively. A stop criterion is made by the measure of fuzziness. 4) Construct an easy and efficient neural network to solve the pattern classification problem.

In this paper, we construct an efficiency classifier by using a variable neural-fuzzy network structure. In section II, we describe the definition of measure of fuzziness for a fuzzy set, and the measure of fuzziness of a fuzzy rule be defined in the fuzzy system that will be used to evaluate the neural-fuzzy network node making sense. The variable neural-fuzzy network structure will be showed in section III. The activation hyper-box, membership functions and uncertainty overlap region are defined. Then, we discuss the learning algorithm to get all parameters in this fuzzy-neural network in section IV. Finally, we show the performance for this classifier to compare with other method in section V. We also make some conclusions in section VI.

2: MEASURE OF FUZZINESS

Characterization and Quantification of uncertainty are important issues that affect the management of uncertainty in many system models and designs. Two categories of uncertainty on data information can be recognized: vagueness and ambiguity. In general, vagueness is the uncertainty associated with difficulty of making a sharp or precise boundary in grouping objects of interest, while ambiguity is the uncertainty associated with choice, that is, difficulty in making a choice between two or more alternatives. Measures of uncertainty related to vagueness are referred to measures of fuzziness.

2.1: MEASURE OF FUZZINESS FOR A FUZZY SET

In general, a measure of fuzziness is a function $f : \tilde{P}(x) \to R$, where $\tilde{P}(x)$ denotes the set of all fuzzy subsets of a universe of discourse X, R is the real line, and the function f satisfies the following axioms:[20]

Axiom 1: $f(A) = 0$ if only if A is a crisp set.

Axiom 2: $A \prec B$, $f(A) \le f(B)$. Where $A \prec B$ denotes that A is shaper than B.

Axiom 3: $f(A)$ assumes the maximum values if only if A is maximally fuzzy.

Axiom 1 indicates that a crisp set has zero degree of fuzziness. Axiom 2 is based upon a particular definition of the relation "sharper than" on the set $\tilde{P}(x)$. By any different definitions, it should satisfy Axiom 2. Axiom 3 states the degree of fuzziness must keep in the maximum value only for a fuzzy set that is described as maximally fuzzy set in $\tilde{P}(x)$. Since there are different definitions of "shaper" in Axiom 2 and "maximally fuzzy" in Axiom 3, several different measures of fuzziness exist in the literature. In this paper, the degree of fuzziness of a fuzzy set will be expressed, in terms of the alike entropy form. This measure of fuzziness is defined by this function:

$$f(A) = -\sum_{x \in A} \{\mu_A(x) \log_2[\mu_A(x)] + [1-\mu_A(x)] \log_2[1-\mu_A(x)]\} \quad (1)$$

where $\mu_A(.)$ is the membership function of A and $\mu_A(x)$ is the grade of membership of x in A. Its normalized version of this measure of fuzziness is given by the formula:

$$\hat{f}(A) = \frac{f(A)}{|X|} \quad (2)$$

where $|X|$ denotes the cardinality of the universal set X. This measure of fuzziness can be considered as the entropy of a fuzzy set.

2.2: MEASURE OF FUZZINESS FOR A FUZZY RULE OF FUZZY SYSTEM

Without loss of generality, we consider multi-input-single-output fuzzy logic system, since a multi-output system can always be decomposed into a group of single-output system. In this section, we define a classification system by a sequence of multi-input-single-output fuzzy rules as following:

R_k : If (x_1 is A_{1k} and x_2 is A_{2k} ………and x_n is A_{nk}) then the class is C_k, $k = 1,2,....c$,

where n is the number of attribute of the classification system, i.e., the dimension of system, c is the number of class of the system, A_{ik} is the linguistic label, $i = 1,2,....n$.

By the T-norm operator with min operation, we can rewrite the system by the other symbols as following description:

$$R_k = \int_{X_1 \times X_2 \times ... \times X_n} \mu_{A_{1k}} \wedge \mu_{A_{2k}} \wedge ... \mu_{A_{nk}} /(x_1, x_2, ... x_n)$$

where X_i is the universal set of discourse on the i th feature in classification system, and $x_i \in X_i$. So, the membership value of this rule R_k can be represented as:

$$\mu_{R_k}(x_1, x_2, ..., x_n) = \min(\mu_{A_{1k}} \wedge \mu_{A_{2k}} \wedge ... \wedge \mu_{A_{nk}})$$

According to the definition of measure of fuzziness in equation (2), we can define the measure of fuzziness of the rule R_k in the fuzzy rule system as:

$$f(R_k) = -\sum_{(x_1, x_2, ..., x_n) \in X_1 \times X \times ... X_n} [(\frac{\mu_{R_k}(.)}{\mu_{R_k}(.) + \mu_{R_{k'}}(.)} \log_2 \frac{\mu_{R_k}(.)}{\mu_{R_k}(.) + \mu_{R_{k'}}(.)}) + (\frac{\mu_{R_{k'}}(.)}{\mu_{R_k}(.) + \mu_{R_{k'}}(.)} \log_2 \frac{\mu_{R_{k'}}(.)}{\mu_{R_k}(.) + \mu_{R_{k'}}(.)})] \quad (3)$$

where $k = 1,2,....c$, $\mu_{R_k}(\cdot)$ represents $\mu_{R_k}(x_1, x_2, ..., x_n)$, $(x_1, x_2, ..., x_n) \in (X_1 \times X_2 \times ... \times X)$ and satisfies $\{0 \le \mu_{R_k}(\cdot) \le 1\}$. The normalized result is

$$\hat{f}(R_k) = \frac{f(R_k)}{|X_1 \times X_2 \times ... \times X_n|} \quad (4)$$

where $|X_1 \times X_2 \times ... \times X_n|$ denotes the cardinality of the universes of discourse $X_1 \times X_2 \times ... \times X_n$. The $\hat{f}(R_k)$ can evaluates the degree of vagueness of fuzzy rule R_k is in a fuzzy rule system. It shows us whether the rule R_k is worth to exist in this rule-based system or not as extracting the efficient rules for our fuzzy system. The redundant rules are not necessary in our decision policy. If a rule attains a high measure of fuzziness of a rule in the fuzzy system, it means too much uncertain for this rule. So, we could discard this highly uncertain rule. In the next section, we will construct a suitable fuzzy rule structure for the pattern classifier by this concept.

3: A VARIABLE NEURAL-FUZZY NETWORK

First, let us review the general fuzzy min-max neural network structure in [14]. This structure has 3 layers: the first layer supports n attributes for testing data input; the second layer is configured by many hyper-boxes nodes which is created from an expansion-overlap-test-contraction algorithm; the third layer is output layer that has c nodes for different c classes. In fact, using the algorithm proposed by [14] and [15] to set up hyper-boxes, we find that each hyper-box must be cut very clearly in this algorithm. Also, some unnecessary cuts result in poor recognitions. This structure usually uses so many hyper-box nodes in second layer that the cost of implementation for this system is high. Our purpose is to construct a suitable number of hyper-boxes for nodes in the second layer of this network and conquer the above drawback.

Our proposed variable structure is shown in Fig. 1. The structure of input and output layer shown in Fig. 1 is the same as in [14]. But, there are c groups of hyper-box nodes in the second layer which are generated from many different uncertainty-overlap regions recursively. Each class pattern corresponds to a group of hyper-box-nodes where the function of each node is very efficient and useful. Each hyper-box node corresponds to a fuzzy rule. If the rule is more crisp and worth to exist, then the corresponding node could be included in the second layer of network. So, the number of network nodes in the second layer is variable and it depends on the setting measure of fuzziness of a fuzzy rule in the fuzzy system. We shall explain how to find the suitable number of hyper-boxes in each group based on the measure of fuzziness in the next section. This structure will reduce the cost on hardware implementation, because the redundant second layer nodes are eliminated. We also save the computation time on software implementation, since it just needs less computation unit than in [14].

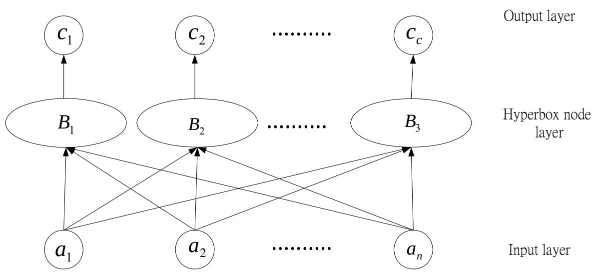

Fig 1 A three layer neural network that implements with variable structure in the second layer

The kth group of hyper-box of second layer in this network structure is presented in Fig. 2. It actually includes two sub-layers inside: the first sub layer is configured by the hyper-box nodes which are created from our proposed algorithm; the second sub-layer is a maximum-operation node, which takes the maximum values of inputs from the first sub-layer. The first sub layer nodes consist of fuzzy rules and they calculate the degrees of membership for an input pattern. The output of the second sub-layer is the degree of membership generated by resolving each uncertainty-overlap region. If the class k is not overlapped with other classes, the number of nodes will be reduced to one in the first sub-layer of the kth group of hyper-box nodes. The qth hyper-box node in the kth group of second layer is denoted as $B_{k,q}^{l,r}$ that extracts the kth class pattern from the rth overlap region of level l .

A general tree relationship between each uncertainty-overlap regions is shown in Fig. 3. For the original data distribution, it is thought as a lump overlap region on level 1. The c activation hyper-boxes are extracted in the first level and several overlap regions are found among these hyper-boxes that are called uncertainty overlap regions on level 2. Some activation hyper-boxes may be set up for each overlap region on level 2. We may find some overlap regions on level 3. By this procedure, we construct the whole overlap regions relation in a tree structure.

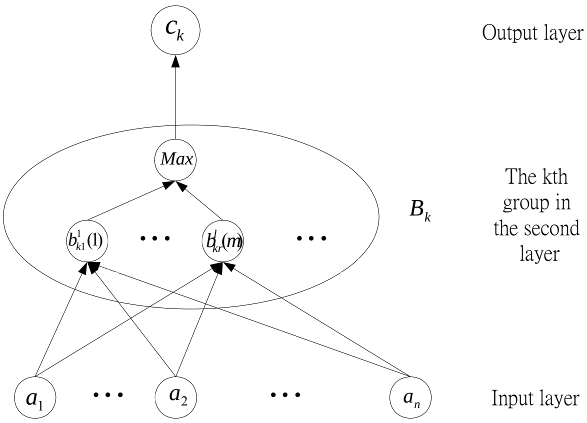

Fig. 2 The kth group of the second layer structure

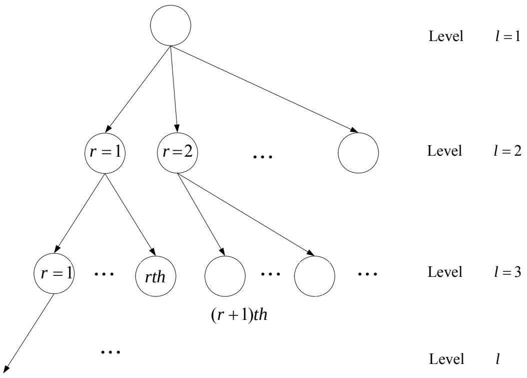

Fig 3 A architecture of the overlap regions tree relationship

To implement the variable fuzzy-neural network, it could exploit the parallel nature in the efficient hardware implementations. This structure is adjusted using the learning algorithm process and the required specification of the measure of fuzziness. The input layer has n processing elements, one for each of n attributes of input pattern. The node transfer function of second layer is the hyper-box membership function that is described by each fuzzy rule. The connections between the jth group of hyper-box of the second layer and the kth node of third layer are binary value. The equation for assigning the values is

$$u_{j,k} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

The outputs node presents the degree of which the input data matches within the class k. In our pattern classification problem, a winner-take-all strategy is adopted. So, the outputs node with the highest value represents the closest pattern class.

4: LEARNING ALGORITHM

The whole process of this method is described as follows:

Step 1: set level $l = 1$.

Step 2: Calculate the minimum point, maximum point and average point of each class for each attribute in the whole attribute space. It will be used to set up the first activation hyper-box and its own membership function for each class. The features of every activation hyper-box correspond to each node of the second layer in our proposed fuzzy-neural network structure.

Step 3: Find the overlap uncertainty-regions among the activation hyper-boxes of level l , then $l = l + 1$.

Step 4: Extract activation hyper-boxes and set up its own feature as in Step 1 from each overlap uncertainty-overlap regions of level l .

Step 5: Calculate the measure of fuzziness of each extracted fuzzy rule by equation (3) and (4). If it is bigger than a setting threshold, we discard this rule.

Step 6: If none of activation hyper-box exist in all uncertainty overlap regions in Step 4, then stop the learning process, else go to Step 2.

Step 7: Simplify some fuzzy linguistic labels in the premise part of some fuzzy rule by checking the measure of fuzziness between two fuzzy linguistic labels.

By this learning algorithm, we can discard some meaningless fuzzy rules in the fuzzy system to decrease the computational cost. The following subsections describe the detail of the main step of this algorithm.

5: PERFORMANCE EVALUATION

We develop a MATLAB program on a Pentium III PC to evaluate the performance. Two kinds of experiment have been performed to illustrate the operation and performance of this variable fuzzy-neural network in our computer simulation. First, we use Fisher's iris data [24] that is a standard data set in pattern classification study for a long time to evaluate the performance of the proposed method. There are three kinds of flowers in this data set, i.e. Setosa, Versicolor, and Verginica. Each flower can be identified by the four kinds of input attributes: sepal length, sepal width, petal length, and petal width. The Iris data includes 150 data samples. Second, we create a set of pseudo Iris data to show the efficiency of our proposed method.

5.1: IRIS DATA

In order to compare our method with others, we randomly choose 50% of the Iris data as the training data set and the remaining 50% as the testing data. The setting value is 0.7 for γ and β to decide if the rule exists in the uncertainty overlap region and if the fuzzy linguistic labels be merged into one fuzzy set or not. The average accuracy rate of the proposed method after 200 runs is listed in Table I. A comparison of the experiment results with some other methods is listed in Table II. It shows our proposed method requires fewer computation time and more suitable number of rules.

Table 1: The classification rate of the proposed method for the original Iris data

Setosa	Versicolor	Virginica	Average
100%	93.65%	95.82%	96.49%

Table 2: A comparison of the number of fuzzy rules and computation time for different algorithm

	Number of rules	Computation time	Average accuracy
The proposed method	4.74	5.057 sec	96.49%
Wu-and-Chen's method	3	54.742 sec	96.21%
Hong-and-Lee's method	6.21	659 sec	95.57%
Hong-and-Chen's method	7.10	353 sec	92.13%

Table 3: A comparison of the number of fuzzy rules and computation time with pseudo Iris data by using different algorithm

	Number of rules	Computation time	Average accuracy
The proposed method	3	31.36 sec	96.86%
Wu-and-Chen's method	3	64023 sec	92.42%
Hong-and-Lee's method	151.2	5593 sec	90.63%
Hong-and-Chen's method	24.6	3227 sec	94.3%

5.2: PSEUDO-IRIS DATA

For the sake of showing our method can efficiently discard the meaningless fuzzy rules in the chaotic uncertainty region, additional 550 data patterns are randomly generated for each class of Iris flower. Assume the Iris flower has the property that their attribute value is distributed uniformly between minimum and maximum value of the original sample data. Also, we assume four attributes of Iris flower are independent. First, 550 pseudo pattern data of Setosa are randomly generated from 4.4cm to 5.8cm for sepal length, from 2.9cm to 4.4cm for sepal width, from 1.0cm to 1.9cm for petal length, from 0.1cm to 0.6cm for petal width; 550 pseudo pattern data of Versicolor are randomly generated from 5.0cm to 7.0cm for sepal length, from 2.0cm to 3.4cm for sepal width, from 3.0cm to 5.1cm for petal length, from 1.0cm to 1.8cm for petal width; 550 pseudo pattern data of Virginica are randomly generated from 4.9cm to 7.7cm for sepal length, from 2.5cm to 3.8cm for sepal width, from 4.8cm to 6.9cm for petal length, from 1.4cm to 2.5cm for petal width. Then, the experiment data include the original Iris data and the pseudo Iris data. Next, 50% of the experiment data are randomly chosen as the training data and the remaining 50% as the testing data. After 200 runs of experiment, a comparison of the results with some other methods is listed in Table III. It shows that we can discard the rules in the uncertain region. Therefore, it creates rules more efficiently in our structure and reduces more computation time.

6: CONCLUTION

In this paper, we have presented a variable fuzzy-neural network structure based on the measure of fuzziness of the fuzzy rule. The basic idea is created from the concept that even the expert can't distinguish classes in some vagueness regions. By this proposed method, we can find more efficient fuzzy rules for the neural network structure. Furthermore, we also apply the proposed method to the Iris data and the pseudo Iris data in classification problem. The experimental results are compared with several different algorithms. The advantages of our proposed algorithm over other methods are described as follows:

1. It can get a good recognition rate compared with other methods in [9][10][11][14].
2. It generates fewer fuzzy rules than other methods [9][10][11][14].
3. It avoids a huge matrix computation [9] so its computation time decreases.
4. It provides a simple recursive process and stopping criteria to extract the fuzzy rules in the uncertainty-overlap region. Thus, the network structure is simple and easy to implement.
5. The classifier can be generated even for a large scale of data pattern.
6. The classifier can be generated by more efficient rules, since the rules are created by the measure of fuzziness.

REFERENCES

[1]. S. Abe and M-S. Lan, "A Method for Fuzzy Rules Extraction Directly from Numerical Data and Its Application to Pattern Classification," IEEE Trans. Fuzzy Systems, vol. 3, No. 1, pp. 18-28, Feb 1995.

[2] .K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic Press, 1972.

[3]. J. C. Bezdek, "Computing with uncertainty", IEEE Commun. Mag., pp. 24-36, Sept. 1992.

[4] .M. H. Hassoun, Fundamentals of Artificial Neural Networks. Cambridge, MA:MIT Press,1995.

[5] .P. K. Simpson, Artificial Neural systems: Foundations Paradigms, Applications, and Implementations. New York:Pergamon,1990.

[6] L. A. Zadeh, Fuzzy sets, Inform. And Control 8 , pp.

338-353, 1965.
[7] L. -X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst., Man, Cybern., vol. 22, pp. 1414-1427,Nov./Dec. 1992.
[8] S. Abe and M-S. Lan "Fuzzy rules extraction directly from numerical data for function approximation," IEEE Trans. Syst., Man, Cybern., vol. 25, pp. 119-129, Jan. 1995.
[9] T.P. Hong and C.Y. Lee, "Induction of Fuzzy rules and membership functions from training examples," Fuzzy Sets Syst., Vol. 84, no. 1, pp.33-47, 1996.
[10] Tzu-ping Wu and Shiyi-Ming Chen, "A new method for constructing membership functions and fuzzy rules from training examples," IEEE Trans.on Systems , Man, and Cybernetics-Part B: Cybernetics Vol 29, No 1, Feb. 1999.
[11] T.P. Hong and J.B. Chen, "Finding relevant attributes and membership functions," Fuzzy Sets Syst., Vol. 103, pp. 389-404, 1999.
[12] F. Uebele, S .Abe, and M-S. Lan, "A Neural Network-Based Fuzzy Classifier,", IEEE Tran. Systems,Man,and Cybernetics, vol. 25, no. 2, pp.353-361, 1995.
[13] S. Abe, Neural Networks and Fuzzy Systems: Theory and Applications, Kluwer Academic, Boston, 1996.
[14] P. K. Simpson, "Fuzzy min-max neural networks-Part 1 General Classification", IEEE Trans. Neural Network, vol. 3, pp. 776-786, Sept. 1992.
[15] B. Gabrys and A. Bargiela,, "General fuzzy min-max neural network for clustering and classification," IEEE Trans. Neural Network, vol. 11, pp.769-783, May. 2000.
[16] A. Grauel and H. Mackenberg, "Mathmatical analysis of the Sugeno controller leading to general design rules," Fuzzy Sets Syst. , Vol. 85, no. 2, pp. 165-175, 1997.
[17] F. Klawobnn and R. Kruse, "Constructing a fuzzy controller from data", Fuzzy Sets Syst. , Vol. 85, no. 2, pp. 177-193, 1997.
[18] C. T. Lin and C. S. George Lee, Neural-network-based fuzzy logic control and decision system," IEEE Trans. Comput, vol. 40,pp. 1320-1336, 1991.
[19] H. Ishibuchi and H. Tanaka, "Neural network that learn from fuzzy if-then rules," IEEE Trans. Fuzzy Systems, vol. 1,pp. 85-97, 1993.
[20] C. T. Lin and C. S. George Lee, Neural Fuzzy System-A Neuro-Fuzzy Synergism to Intelligent System, Prentice Hall, 1996.
[21] J. J. Shann and H. C. Fu, "A fuzzy neural network for rule acquiring on fuzzy control systems," Fuzzy Sets Syst. , Vol. 71, no. 3, pp. 345-357, 1995.
[22] R. Langari and L. Wang, "Fuzzy Models, modular networks, and hybrid learning," Fuzzy Sets Syst. , Vol. 79, no. 2, pp. 141-150, 1996.
[23] N. K. Kasabov, "Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid system," Fuzzy Sets Syst., Vol. 82, no. 2, pp. 135-149, 1996.
[24] R. Fisher, "The use of multiple measurements in taxonomic problems", Ann. Eugenics, vol. 7, pp 179-188, 1936.