

Finding Fuzzy Rules for IRIS by Neural Network with Weighted Fuzzy Membership Function

Joon Shik Lim

College of Software, Kyungwon University, Sungnam 461-701

Abstract

Fuzzy neural networks have been successfully applied to analyze/generate predictive rules for medical or diagnostic data. However, most approaches proposed so far have not considered the weights for the membership functions much. This paper presents a neural network with weighted fuzzy membership functions. In our approach, the membership functions can capture the concentrated and essential information that affects the classification of the input patterns. To verify the performance of the proposed model, well-known Iris data set is performed. According to the results, the weighted membership functions enhance the prediction accuracy. The architecture of the proposed neural network with weighted fuzzy membership functions and the details of experimental results for the data set is discussed in this paper.

Key Words : Fuzzy neural network, rule extraction, weighted membership function

1. INTRODUCTION

Neural network and fuzzy set theory can be effectively used for pattern classification and predictive rule generation tool. The goal of pattern classification is to partition the feature space into decision regions. Artificial neural networks have been successfully used in many pattern classification problems [3] [6]. Fuzzy set theory was introduced by Zadeh [23] as a means of representing and processing data by allowing partial set membership rather than crisp set membership or non-membership. As a combined approach of neural network and fuzzy set theory, fuzzy neural networks (FNN) have been proposed as an adaptive decision support tool [2] [8] [11][12] [13] [14] [17] [20]. Various architectures of FNN have been introduced along with algorithms for learning, adaptation and rule extraction [7] [10] [15] [16]. In pattern classification and predictive rule generation, high accuracy of acceptable regular patterns with simple types of knowledge representation in linguistic terms is desirable. In addition, the way of obtaining the optimal structure for fast adaptation without sacrificing the accuracy and performance is needed. The capability of fuzzy if-then rule extraction is one of the advantages of FNN for medical decision-making. For this, various approaches has been proposed. Among those approaches, fuzzy neural networks with self-organizing systems were developed by [9] [19] [21] to extract knowledge from a given set of training data. Setnes [16] presented a compact and accurate fuzzy rule-based model using a genetic algorithm. However, most approaches proposed so far have not considered the weights for the membership functions much.

In this paper, we present a new neural network model with weighted fuzzy membership functions (NNWFM) to classify IRIS data [4]. In our approach, three membership functions,

each of which small, medium, and large functions with weights, are initially defined and trained to generate a combined membership function using the bounded sum function. The weighted fuzzy membership functions can preserve the disjunctive fuzzy information and characteristic input patterns in the pattern space, which can result in reducing the number of hyperboxes and rules. We also present our neural network architecture and experimental results in the following sections. The effectiveness of NNWFM is first validated using the popularly used data sets, Iris [4] for benchmarking of pattern classification.

2. NEURAL NETWORK MODEL with WEIGHTED FUZZY MEMBERSHIP FUNCTION (NNWFM)

2.1 The Structure of NNW

The structure of NNWFM is illustrated in Fig. 1. The NNWFM comprises three layers, namely input, hyperbox, and class layer. The input layer contains n input nodes for n featured input patterns. The hyperbox layer consists of m hyperbox nodes. Each hyperbox node B_i to be connected to a class node contains n fuzzy sets for n input nodes. The output layer is composed of p class nodes. Each class node is connected to one or more hyperbox nodes. An h th pattern can be recorded as $I_h = \{A_h - (a_1, a_2, \dots, a_n), \text{diagnosis}\}$, where *diagnosis* is the result of diagnosis and A_h is the pattern on n different features. 'Unknown' features are represented as NULL. The i th fuzzy set of B_i denoted by B_i^j , has three weighted membership functions as shown in Fig. 2. These three weighted fuzzy membership functions are adjusted by the learning algorithm, which will be discussed in Section 2.2. The hyperbox nodes in the hyperbox layer are controlled by the adjust operations. After learning, the rules for classification can be extracted by

the hyperbox layer. The detail of rule extraction processes is described in the Section 2.3.

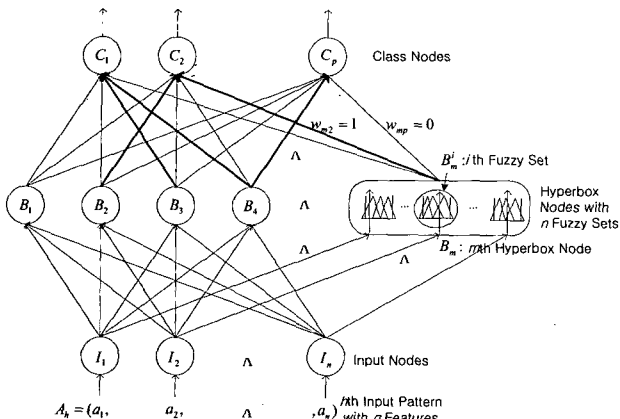


Fig. 1. Structure of NNWFM

2.2 Definitions and Operations

1) w_{hi} : The connection weight between a hyperbox node B_i and a class node C_i is represented by w_{hi} , which is initially set to 0. If there is a connection from a hyperbox node B_i to a class node C_i , the w_{hi} will be set to 1 from 0. C_i can have more than one connection from hyperbox nodes, whereas B_i is restricted to have one connection to a class node.

2) v_i : The $v_1, v_2,$ and v_3 represent the center vertices of the small, medium, and large membership functions respectively in Fig. 2. The center vertices can be adjusted during learning, while the other vertices v_0 and v_4 are fixed. It is assumed that the input feature value a_i ranges from v_{min} to v_{max} as shown in Fig. 2.

3) μ_j and W_j : μ_{js} are i th set of membership functions of a hyperbox node B_i for $j=1,2,3$, representing small, medium, and large respectively. The shape of each membership function μ_j is triangular, which is characterized by three vertices (v_{j-1}, v_j, v_{j+1}) with its membership function weight W_j ($0 \leq W_j \leq 1$, random weights in the range of $0.45 \leq W_j \leq 0.55$ are initially set) that represents the strength of the membership function determined through learning. The shaded triangles in Fig. 2, called weighted fuzzy membership functions, can be formed by (v_{j-1}, W_j, v_{j+1}) .

4) $Len(\mu_j)$: To measure the size of a μ_j , a length function, Len is defined as follow:

$$Len(\mu_j) = ((v_{j+1} - v_{j-1}) / 2) W_j \tag{2.1}$$

This function is similar to the length function, $Len(\mu_j)$, defined by Lee et al. [13] except that the $Len(\mu_j)$ is multiplied by the weights W_j .

5) $Adjust(B_i)$: This operation adjusts membership functions and their weights for n fuzzy sets in the hyperbox node B_i according to the input $A_k=(a_1, a_2, \dots, a_n)$. For each i th set of membership functions with the input a_i , the v_j s and W_j s of the membership functions are adjusted by the value of $\mu_j(a_i)$,

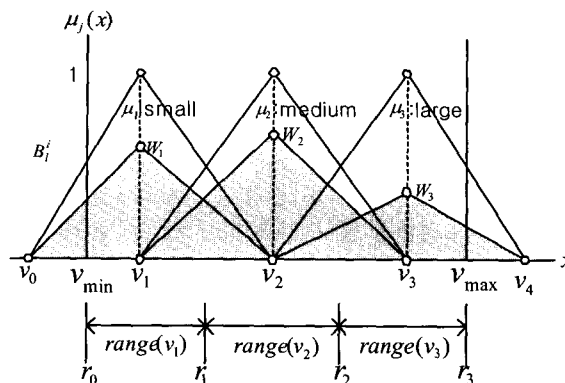


Fig. 2. An I th Set of Weighted Membership Functions of B_i

where $j=1, 2, 3$. As a result of $Adjust(B_i)$ operation, the new vertices $new(v_j)$ s and new weights $new(W_j)$ s are set by the following expression:

$$new(v_j) = v_j \pm \square E_j \mu(a_i) W_j \tag{2.2}$$

$$new(W_j) = W_j + \square (\mu(a_i) - W_j) \tag{2.3}$$

In these expressions, the \square and \square are the learning rate in the range from 0 to 1, and the variable E_j is the difference between v_j and input a_i . If E_j is bigger than the adjacent $E_{j\pm 1}$, the smaller one is selected. The detail process of the $Adjust(B_i)$ operation is described in the learning algorithm in the next section. Fig. 3 shows the result of the $Adjust(B_i)$ operations for the input a_i and the i th set of fuzzy weighted membership functions in B_i .

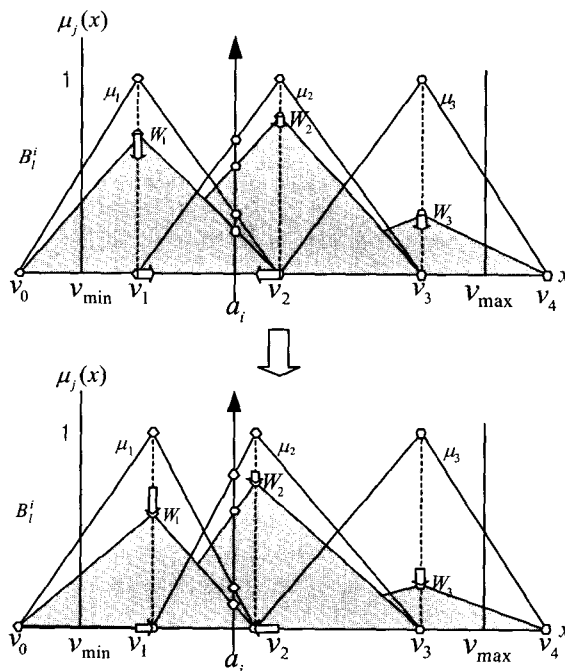


Fig. 3. Example of Before And After $Adjust(B_i)$ Operation

As shown in Fig. 3, the weights and the center of each membership functions are adjusted by the $Adjust(B_i)$ operation, e.g., $W_1, W_2,$ and W_3 are moved down, v_1 and v_2 are moved

toward a_i , and v_3 remains in the same location.

6) $Random(B_i)$: This operation makes the hyperbox node B_i with I (number of inputs) sets of randomly distributed membership functions with their random weights ranging from 0.45 to 0.55. The random center vertices v_{is} should be in their ranges such that

$$r_{i-1} \leq v_i < r_i, \text{ where } i = 1, 2, 3$$

while v_0 and v_4 are fixed as in Fig. 2. Also, the connection weights w_{is} are set to 0.

7) $Output(B_i)$: For the h th input $A_h = (a_1, a_2, \dots, a_n)$ with n features to the hyperbox B_i , output of the B_i is calculated by

$$Output(B_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 B_i^j(\mu_j(a_i)) W_j. \quad (2.4)$$

2.3 Learning Algorithm for NNWFM

This section describes the learning algorithm for NNWFM. The algorithm uses the $Learning(B_i, C_i)$ procedure to adjust the locations of vertices and weights, and to connect the hyperbox nodes to class nodes.

Algorithm NNWFM

```

1 While (result is satisfied)
1.1 for  $l=1$  to  $m$  //  $m$  is number of hyperboxes,
    // usually start from number input
1.1.1 Random( $B_i$ );
1.1.2 for  $j=1$  to  $p$  //  $p$  is number of class nodes
1.1.2.1  $w_{ij}=0$ ; // initial connection weight
    // between  $B_i$  and  $C_j$ 
1.2 for  $k=1$  to  $h$  //  $h$  is number of input patterns
1.2.1 find  $B_l$  that has the maximum value of
EnhOutput( $B_l$ ) // see the formula 2.6 among  $m$  hyperbox nodes from the input  $A_k$ 
// input vector:  $I_k = \{A_k = (a_1, a_2, \dots, a_n), \text{diagnosis}\}$ 
1.2.2 Learning( $B_i, C_i$ ); //  $C_i$  is a diagnosis in  $I_k$ 
    
```

Procedure $Learning(B_i, C_i)$;

```

//  $m$  is number of hyperboxes
1 Case 1:  $\forall m, w_{mi} = 0$ , where  $m \neq l$ 
1.1  $w_{ii} = 1$ ;
1.2 Adjust( $B_i$ );
    
```

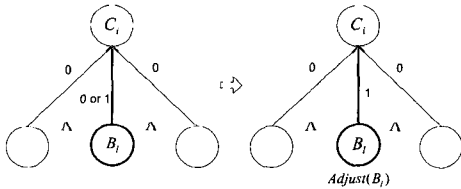


Fig. 4. Case 1: Adjust

```

2 Case 2:  $\exists m$  satisfying  $w_{mi} = 1$ , where  $m_l$ 
2.1  $w_{ii} = 1$ ;
2.2 Adjust( $B_i$ );
    
```

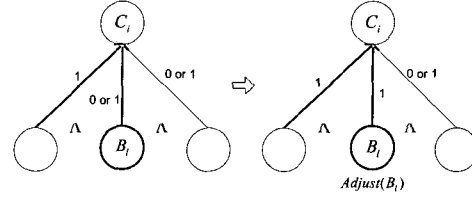


Fig. 5. Case 2: Adjust

Procedure $Adjust(B_i)$;

```

1 for  $i=1$  to  $n$ 
// for each  $i$ th set of membership function in  $B_i$ 
1.1 for  $j=1$  to 3 // for each membership function
1.1.1 if  $v_{j-1} \leq a_i < v_j$  // for left side of  $\mu_j$ 
1.1.1.1  $E_j = \min(|v_j - a_i|, |v_{j-1} - a_i|)$ ;
1.1.1.2  $new(v_j) = v_j - \alpha E_j \mu_j(a_i) W_j$ 
1.1.2 else  $v_j \leq a_i \leq v_{j+1}$  // for right side of  $\mu_j$ 
1.1.2.1  $E_j = \min(|v_j - a_i|, |v_{j+1} - a_i|)$ ;
1.1.2.2  $new(v_j) = v_j + \alpha E_j \mu_j(a_i) W_j$ 
1.1.3  $new(W_j) = W_j + \beta (\mu_j(a_i) - W_j)$ ;
    
```

2.4 Fuzzy Rule Extraction

The learned NNWFM can be used for fuzzy rule extraction in if-then form to classify input patterns. After learning, each fuzzy set in hyperbox node B_i contains three weighted fuzzy membership functions (WFMs, grey membership functions in Fig. 6).

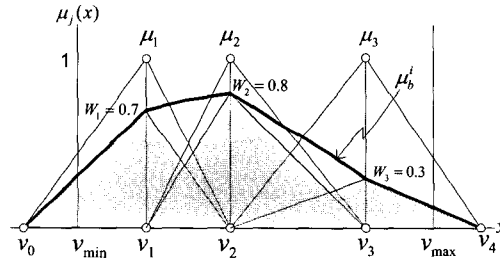


Fig. 6. Example of Bounded Sum of the Three Weighted Fuzzy Membership Functions (BSWFM, Bold Line)

The rules can be extracted directly from the WFMs. We suggest a rule extraction strategy as described below.

1) The bounded sum (one of operations on fuzzy set) of WFMs (BSWFM) in the i th fuzzy set of $B_i^l(x)$, denoted as $\mu_b^i(x)$ (bold line in Fig. 6), is defined by

$$\mu_b^i(x) = \sum_{j=1}^3 B_i^j(\mu_j(x)). \quad (2.5)$$

The bounded sum of WFMs combines the fuzzy characteristics of three WFMs in Fig. 6.

2) We use the Takagi and Sugeno[18] model for an inference mechanism that performs the reasoning process to derive an $Output(B_i)$ (2.4). In addition, we suggest a heuristic of $Len(\mu_j)$ to enhance the reasoning capability such that:

$$EnhOutput(B_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 B_i^j(\mu_j(a_i)) W_j / Len(\mu_j) \quad (2.6)$$

The bold lines in Fig. 6 are enhanced BSWFM using (2.6). The heuristic of $Len(\mu_i)$ is based on the concept that smaller $Len(\mu_i)$ has more concentrated information that is essential for classification of the given pattern.

3. EXPERIMENTAL RESULTS

In this section, we present the experimental results for the Iris data set to evaluate the accuracy and rule extraction capability of the proposed NNWFM.

The data set we used to evaluate the performance of the proposed model is Fisher's Iris data [4], which is one of the popularly used data sets in pattern classification studies. The data set consists of 150 four-dimensional feature vectors (sepal length, sepal width, petal length, and petal width) in three classes (Setosa, Versicolor, and Virginica 50 instances for each class). Nauck [15] uses a neuro-fuzzy method to learn fuzzy classifiers presenting seven rules with three misclassifications. Ishibuchi [7] obtains four fuzzy rules with four error classifications by weighted voting method. Setnes [16] proposes a genetic algorithm-based model, called GA-fuzzy modeling that produces two or three rules with only one error classification.

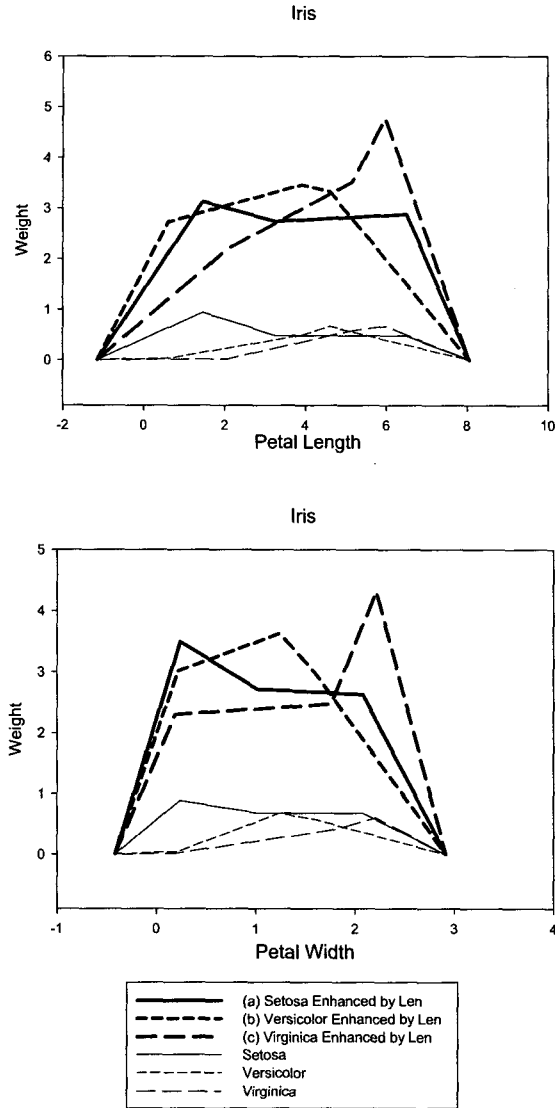
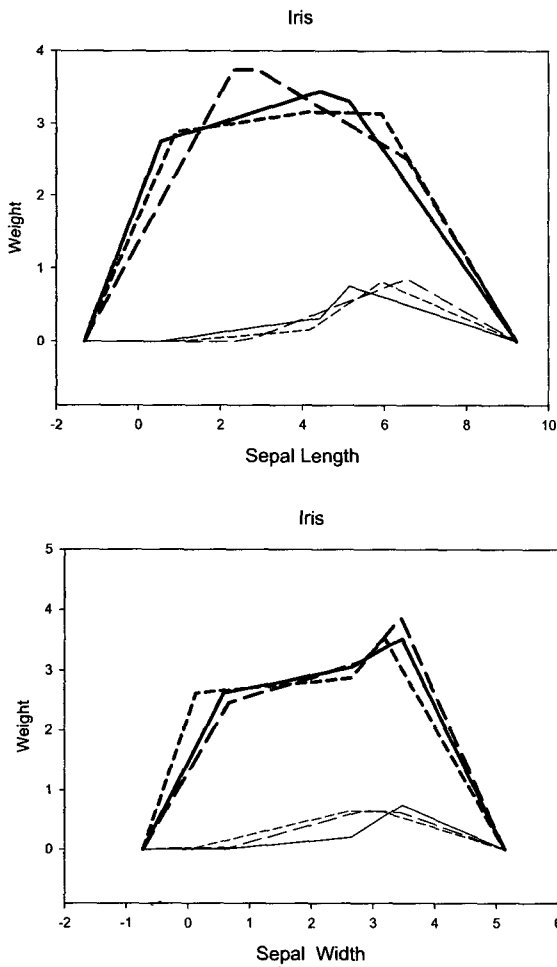


Fig. 7. BSWFMs and Enhanced BSWFMs (bold lines) for Iris Data Set

The three prediction rules in Table 1 for Iris data are extracted from the enhanced BSWFMs in Fig. 7. In Fig. 7, each graph represents trained membership functions for three classes: Setosa, Versicolor, Virginica, for each feature in the Iris data set. In this experiment, we used the membership functions shown by the bold lines, which are Enhanced BSWFMs based on BSWFMs. In each graph, the enhanced BSWFMs (the bold lines) are obtained by the division of $Len(\mu_i)$ function defined in the section 2.4 (see the equation 2.6). The $Len(\mu_i)$ function measures the width of a membership function. Heuristically, we can consider that the narrower the width of the membership function, the more concentrated information it may contain. The comparisons of their performance results on Iris data are shown in Table 2. The complete 150 cases in the data set are used as a training set. The results show that the proposed model, NNWFM, has the best classification rate with three rules.

TABLE 1
Rules for Iris Data

Rule	sepal length	sepal width	petal length	petal width	Class
rule 1	(a)	(a)	(a)	(a)	Setosa
rule 2	(b)	(b)	(b)	(b)	Versicolor
rule 3	(c)	(c)	(c)	(c)	Virginica

In each graph in Fig. 7, line (a), (b), and (c) (Enhanced BSWFM, bold line) represents each rule for classification.

TABLE 2
Performance Results on Iris Data

Algorithm	No. of Misclassifications	No. of Rules
Nauck [15]	3/150	7
Ishibuchi [7]	4/150	4
Setnes [16]	1/150	2 - 3
NNWFM	0/150	3

4. CONCLUDING REMARKS

In the proposed neural network model, a fuzzy set in a hyperbox node implies information of small, medium, and large weighted fuzzy membership functions. After learning, the enhanced bounded sums of weighted fuzzy membership functions (enhanced BSWFMs) in hyperbox nodes contain characteristics of input patterns. This feature gives the following advantages.

- 1) The enhanced BSWFMs in a hyperbox can maintain complementary information which handles local maxima problems, ambiguous subsets of feature space, and disjunctive fuzzy information in the pattern space.
- 2) Rules can be extracted easily since enhanced BSWFMs are direct interpretation of fuzzy rules.
- 3) The number of hyperbox nodes can be reduced, since the enhanced BSWFMs represent the characteristics of input patterns by preserving the fuzziness.
- 4) The enhanced BSWFMs graphs show visually important feature values.

By the above advantageous properties of NNWFM, we could extract human comprehensive three rules for IRIS data set to classify with 100% accuracy.

REFERENCES

[1] C.L Blake and C.J. Merz, "UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mllearn/MLRepository.html>]", Irvine, CA: University of California, Department of Information and Computer Science, 1998.
 [2] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural

network," *Neural Networks*, Vol. 4, pp. 565-588, 1991.
 [3] R.O. Duda and P.E. Hart, "Pattern Classification and Scene Analysis", New York: Wiley, 1973.
 [4] R. A. Fisher, "The Use of Multiple Measurements in taxonomic Problem," *Annals of Eugenics*, Vol. 7, No. 2, pp.179-188, 1936.
 [5] A. F. Gomez-Skarmeta, M. V. F. Jimenez, J. G. Marin-Blazques, "Approximative Fuzzy Rules Approaches for Classification with Hybrid-GA Techniques," *Information Sciences*, Vol. 136, pp.193-214, 2001.
 [6] S. Haykin, "Neural Networks, a comprehensive foundation", Prentice Hall, New Jersey, 1999.
 [7] H. Ishibuchi and T. Nakashima, "Voting in Fuzzy Rule-Based Systems for Pattern Classification Problems," *Fuzzy Sets and Systems*, Vol. 103, pp. 223-238, 1999.
 [8] R. Jang, "ANFIS: Adaptive network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, Vol. 23, pp. 665-685, May-June 1993.
 [9] C. F. Juang and C. T. Lin, "An On-Line Self-Constructing Neural Fuzzy Inference Network and Its Applications," *IEEE Trans. Fuzzy Systems*, Vol. 6(1), pp. 12-32, 1998.
 [10] N. Kasabov, *Foundation of Neural Networks, Fuzzy Systems and Knowledge Engineering*, The MIT Press, Cambridge, MA, 1996.
 [11] T. Kasuba, "Simplified Fuzzy ARTMAP," *IEEE AI Expert*, pp. 19-25, Nov. 1993.
 [12] B. Kosko, *Neural networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Englewood Cliffs, NJ:Prentice-Hall, 1992.
 [13] H.-M. Lee, K.-H. Chen, and I.-F. Jiang, "A Neural Networks with Disjunctive Fuzzy Information," *Neural Networks*, Vol. 11, pp. 1113-1125, 1998.
 [14] C. T. Lin and C. S. George Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Computers*, Vol. 40, No. 12, Dec., 1991.
 [15] D. Nauck and R. Kruse, "A Neuro-Fuzzy Method to Learn Fuzzy Classification Rules from Data," *Fuzzy Sets and Systems*, Vol. 89, pp. 277-288, 1997.
 [16] M. Setnes and H. Roubos, "GA-Fuzzy Modeling and Classification: Complexity and Performance," *IEEE Trans. Fuzzy Systems*, Vol. 8(5), pp. 509-522, 2000.
 [17] P. Simpson, "Fuzzy min-max neural networks-Part 1: Classification," *IEEE Trans. Neural Networks*, Vol. 3, pp. 776-786, 1992.
 [18] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Application to Modeling and Control," *IEEE Trans. Syst., Man, Cybern.*, Vol. 15, pp. 116-132, 1985.
 [19] K. Tanaka, M. Sano, and H. Watanabe, "Modeling and Control of Carbon Monoxide Concentration Using a Neuro-Fuzzy technique," *IEEE Trans. Fuzzy Systems*, Vol. 3, pp. 271-279, June, 1995.
 [20] C. Z. Ye, J. Yang, D. Y. Geng, Y. Zhou, N. Y. Chen, "Fuzzy Rules to Predict Degree of Malignancy in Brain Glioma," *Medical and Biological Engineering and Computing*, Vol.40, 2002.
 [21] J. S. Wang and C. S. G. Lee, "Self-Adaptive Neuro-Fuzzy

Inference System for Classification Applications," IEEE Trans. Fuzzy Systems, Vol. 10(6), pp. 790-802, 2002.

- [22] W. Wolberg, O. Mangasarian, "Multisurface Method of . Pattern for Medical Diagnosis Applied to Breast Cytology," Proc. National Academy of Sciences, Vol. 87, pp. 9193-9166, 1990.
- [23] L. Zadeh, "Fuzzy sets. Information and Control," Vol. 8, pp. 338-353, 1965. E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the Earth's atmosphere," Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (420-46)-3, Nov. 1988.
-



Joon Shik Lim

received the B.S. and M.S. degrees in computer science from Inha University, Korea, University of Alabama at Birmingham, and Ph.D. degree from Louisiana State University, Baton Rouge, Louisiana, in 1986, 1989, and 1994, respectively.

He is currently a Professor of College of Software at Kyungwon University, Korea. His research focuses on neuro-fuzzy systems, biomedical prediction systems, and human-centered systems. He has authored three textbooks Computer Practice and Applications (Jungik Press, 1998), Artificial Intelligence Programming (Green Press, 2000), and Javaquest (Green Press, 2003).

Phone : 031)750-5330

Fax : 031)750-5662

E-mail : jslim@kyungwon.ac.kr