CONSTRUCTING FUZZY CLASSIFICATION SYSTEMS FROM WEIGHTED TRAINING PATTERNS

**Tomoharu Nakashima, Yasuyuki Yokota
Hisao Ishibuchi**

College of Engineering
Osaka Prefecture University
Gakuen-cho 1-1, Sakai, Osaka 599-8531, Japan
{nakashi, yokota, hisaoi}@ie.osakafu-u.ac.jp

Andrzej Bargiela

Department of Computing
The Nottingham Trent University
Burton Street, Nottingham NG1 4BU, U.K.
andre@doc.ntu.ac.uk

KEYWORDS

Fuzzy rule-based systems, pattern classification, data mining, classification cost

ABSTRACT

In this paper, we propose a fuzzy rule-generation method for pattern classification problems. We consider a situation where each training pattern has a weight. The weight is considered as a cost of misclassification/ rejection of classification. Our fuzzy classification system consists of a set of fuzzy if-then rules. The antecedent part of fuzzy if-then rules linguistically specifies a subarea of a pattern space. Thus, fuzzy if-then rules are linguistically interpretable. The main aim of this paper is to construct fuzzy classification systems that reflect weighted training patterns. That is, we make the cost of misclassification and rejection of classification for unseen patterns as small as possible. In computer simulations, we also propose two methods for assigning appropriate weights from the distribution of given training patterns. The performance of the proposed fuzzy rule-generation method is examined for several real-world pattern classification systems that have been used in literature.

INTRODUCTION

Fuzzy rule-based systems have been applied mainly to control problems (Lee 1990, Leondes 1999, Sugeno 1985). One advantage of a fuzzy rule-based system is its interpretability. Recently fuzzy rule-based systems have also been applied to pattern classification problems. There are many approaches to the automatic generation of fuzzy if-then rules from numerical data for pattern classification problems. Genetic algorithms have also been used for generating fuzzy if-then rules for pattern classification (Ishibuchi and Nakashima 1999a, Ishibuchi Nakashima 1999b, Ishibuchi et al. 1995, Yuan and Zhan 1996).

In this paper, we propose a fuzzy rule-generation method for pattern classification problems where a weight is assigned to each given training pattern. We assume that weights of training patterns are given before the actual fuzzy classification sysetms are constructed. The proposed method incorporates the weights in generating fuzzy if-then rules as well as the compatibility of training patterns.

We also propose weight assignment techniques for our computer simulations as in many benchmark problems weights of training patterns are not available.

A series of computer simulations examine the advantage of our proposed method over the conventional method. We also show then one of the weight assignment techniques is useful for obtaining fuzzy classification systems with high generalization ability.

FUZZY RULE-BASED CLASSIFICATION

Pattern Classification Problems

Various methods have been proposed for fuzzy classification (Nozaki et al. 1996, Klir and Yuan 1995, Ishibuchi et al. 1992, Grabisch 1996, Grabisch and Dispot 1992, Grabixch and Nicolas 1994). Let us assume that our pattern classification problem is an n-dimensional problem with C classes and m given training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$. Without loss of generality, each attribute of the given training patterns is normalized into a unit interval $[0,1]$. That is, the pattern space is an n-dimensional unit hypercube $[0,1]^n$ in our pattern classification problem.

In this study we use fuzzy if-then rules of the following type as a base of our fuzzy rule-based classification systems:

Rule R_j: If x_1 is A_{j1} and ... and x_n is A_{jn}

then Class C_j with CF_j, $j = 1, 2, \ldots, N$, $\qquad$ (1)

where R_j is the label of the j-th fuzzy if-then rule, $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets on the unit interval $[0,1]$, C_j is a consequent class (i.e. one of the C given classes), CF_j is the grade of certainty of the fuzzy if-then rule R_j, and N is the total number of fuzzy if-then rules. As antecedent fuzzy sets, we use triangular fuzzy sets as in Fig. 1 where we show various partitions of the unit interval into a number of fuzzy sets.
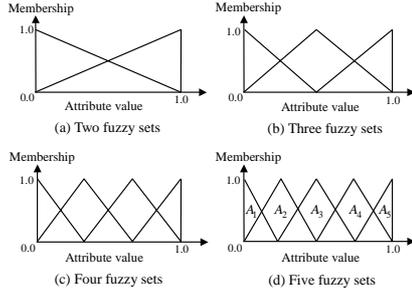
Fig. 1 An example of antecedent fuzzy

Generating Fuzzy If-Then Rules

In our fuzzy rule-based classification systems, we specify the consequent class and the grade of certainty of each fuzzy if-then rule from the given training patterns (Ishibuchi et al. 1992). In (Ishibuchi and Nakashima 2001) it is shown that the use of the grade of certainty in fuzzy if-then rules allows us to generate comprehensible fuzzy rule-based classification systems with high classification performance.

The consequent class C_j and the grade of certainty CF_j of fuzzy if-then rule are determined in the following manner:

[Generation Procedure of Fuzzy If-Then Rule]

Step 1: Calculate $\beta_{\text{Class } h}(R_j)$ for Class h as

$$\beta_{\text{Class } h}(R_j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_{j1}(x_{p1}) \cdot \ldots \cdot \mu_{jn}(x_{pn}),$$
$$h = 1, 2, \ldots, C, \qquad (2)$$

where $\mu_{ji}(\cdot)$ is the membership function of the fuzzy set A_{ji}.

Step 2: Find Class $\hat{h}$ that has the maximum value of $\beta_{\text{Class } h}(R_j)$:

$$\beta_{\text{Class } \hat{h}}(R_j) = \max\{\beta_{\text{Class } 1}(R_j), \ldots, \beta_{\text{Class } C}(R_j)\}. \quad (3)$$

If two or more classes take the maximum value, the consequent class C_j of the rule R_j can not be determined uniquely. In this case, specify C_j as $C_j = \phi$. If a single class takes the maximum value, let C_j be Class $\hat{h}$. The grade of certainty CF_j is determined as

$$CF_j = \frac{\beta_{\text{Class } \hat{h}}(R_j) - \overline{\beta}}{\sum \beta_{\text{Class } h}(R_j)}, \qquad (4)$$

where

$$\overline{\beta} = \frac{\sum_{h \neq \hat{h}} \beta_{\text{Class } h}(R_j)}{c - 1}. \qquad (5)$$

The number of fuzzy if-then rules in a fuzzy rule-based classification system depends on how each attribute is partitioned into fuzzy subsets. For example, when we divide each attribute into three fuzzy subsets in a ten-dimensional pattern classification problem, the total number of fuzzy if-then rules is $3^{10} = 59049$. This is what is called *the curse of dimensionality*. The grade of certainty CF_j can be adjusted by a learning algorithm (Nozaki et al. 1996).

Fuzzy Reasoning

By the rule generation procedure in the last subsection, we can generate N fuzzy if-then rules in (1). After both the consequent class C_j and the grade of certainty CF_j are determined for all N rules, a new pattern $\mathbf{x}$ is classified by the following procedure:

[Fuzzy reasoning procedure for classification]

Step 1: Calculate $\alpha_{\text{Class } h}(\mathbf{x})$ for Class h, $j = 1, 2, \ldots, C$, as

$$\alpha_{\text{Class } h}(\mathbf{x}) = \max\{\boldsymbol{\mu}_j(\mathbf{x}) \cdot CF_j \mid C_j = \text{Class } h\},$$
$$h = 1, 2, \ldots, C, \qquad (6)$$

where

$$\boldsymbol{\mu}_j(\mathbf{x}) = \mu_{j1}(x_1) \cdot \ldots \cdot \mu_{jn}(x_n). \qquad (7)$$

Step 2: Find Class h' that has the maximum value of $\alpha_{\text{Class } h}(\mathbf{x})$:

$$\alpha_{\text{Class } h'}(\mathbf{x}) = \max\{\alpha_{\text{Class } 1}(\mathbf{x}), \ldots, \alpha_{\text{Class } C}(\mathbf{x})\}. \qquad (8)$$

If two or more classes take the maximum value, then the classification of $\mathbf{x}$ is rejected (i.e. $\mathbf{x}$ is left as an unclassifiable pattern), otherwise assign $\mathbf{x}$ to Class h'.

PROPOSED METHOD

In this section we propose a fuzzy rule-generation method for pattern classification problems where a weight is given for each training pattern. First we explain the role of weights and then we present how fuzzy if-then rules are generated from weighted training patterns.

Concept of Weight

Let us consider an example of classification problems in which the idea of weighted traininng patterns is necessary. In the medical diagnosis of cancer two kinds of misclassification have to be taken into account. One is the case where a person is diagnosed as having cancer while he/she does not. The other case is that a person with cancer is classified as not having cancer. Although

the misclassiffccation should be as small as possible in both cases, the latter misclassification should be treated more seriously than the former case. In this paper we use the concept of weights to tackle this problem.

Cost Function

The weight of training patterns can be viewed as the importance of the patterns. We place more emphasis on those patterns with large weights than on those with small weights. The weight of misclassified/rejected patterns is considered as a cost of misclassification or rejection. Thus, our objective in this paper is to construct a fuzzy classification system S that minimizes the following cost function :

$$Cost(S) = \sum_{p=1}^{m} w_p \cdot z_p(S) , \qquad (9)$$

where $Cost(S)$ is the cost of misclassification/rejection made by a fuzzy classification system S, m is the number of training patterns, w_p is the weight of the training pattern $\mathbf{x}_p$, and $z_p(S)$ is a binary variable that is determined according to the classification result of the training pattern $\mathbf{x}_p$ by S: $z_p(S) = 0$ if $\mathbf{x}_p$ is correctly classified by S, and $z_p(S) = 1$ otherwise.

Generating Fuzzy If-Then Rules from Weighted Training Patterns

Let us assume that we have m training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$, and we also assume that a weight is given a priori for all training patterns. We modify the fuzzy rule generation procedure presented in the last section to accomodate the weighted training patterns :

[Generation Procedure of Fuzzy If-Then Rule]
Step 1: Calculate $\beta_{\text{Class } h}(R_j)$ for Class h ($h = 1, \ldots, C$) as

$$\beta_{\text{Class } h}(R_j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_{j1}(x_{p1}) \cdot \ldots \cdot \mu_{jn}(x_{pn}) \cdot w_p ,$$

$$h = 1, 2, \ldots, C . \qquad (10)$$

Step 2: Find Class $\hat{h}$ that has the maximum value of $\beta_{\text{Class } h}(R_j)$:

$$\beta_{\text{Class } \hat{h}}(R_j) = \max\{\beta_{\text{Class } 1}(R_j), \ldots, \beta_{\text{Class } C}(R_j)\} . \qquad (11)$$

This procedure is the same as the one that is used in the last section except for (10). In order to focus on those training patterns with large weights, we include the weight in the calculation of the compatibility of training patterns with the antecedenet part of fuzzy if-then rules when we dermine the consequent class of the fuzzy if-then rule.

WEIGHT ASSIGNMENT

In the next section we examine the performance of the proposed method by computer simulation. We use real-world pattern classification problems that are commonly used in literature. All the classification problems are available from the UCI machine learning repository. Since weights for training patterns are not included in these pattern classification sets, we propose two methods for weighting training patterns in order to make a synthetic situation where a weight is given to each training pattern. One is a class-bassed weighting method and the other is an overlap-based weighting method.

Class-Based Weighting Method

The aim of the class-based weighting method is to make a bias toward the classification of patterns from a particular class. For example, if the bias is toward the classification of Class 1 patterns, classification systems are expected to correctly classify Class 1 patterns even if the number of misclassification/rejection is large for other classes. Medical diagnoses is example appropriate for this weighting method.

In this weighting method, a weight for the pattern $\mathbf{x}_p$ is determined by the following equation :

$$w_p = \begin{cases} 1.0 & \text{if class of } \mathbf{x}_p \text{ is to be emphasized,} \\ 0.5 & \text{otherwise.} \end{cases} \qquad (12)$$

Overlap-Based Weighting Method

The aim of the overlap-based weighting method is to focus on overlapped areas between multiple classes. In order to determine the weights of given training patterns, we count the number of patterns from the same class in their neighborhood. Let us denote the neighborhood size as N_{size}. We examine N_{size} nearest patterns from each of given training patterns for determining the value of the weight. In the overlap-based weighting method we use the following equation to determine the weight of the p-th given pattern w_p :

$$w_p = \frac{N_p^{\text{same}}}{N_p^{\text{nearest}}} \qquad (13)$$

where N_p^{nearest} is the number of nearest training patterns to $\mathbf{x}_p$ and N_p^{same} is the number of given patterns from the same class as the p-th given pattern. The weight w_p of the p-th given pattern can be viewed as a measure of overlaps. That is, if the value of w_p is large, there are many patterns from the same class as p-th training pattern. On the other hand, the p-th given pattern is possibly an outlier if the value of w_p is low.

COMPUTER SIMULATIONS

We examined the performance of the proposed method for eight real-world pattern classification problems that are available from the UCI machine learning repository. We show the details of the nine classification problems in Table 1.

Table 1 Classification problems

Data set	Attributes	Classes	Patterns
Balance scale	4	3	625
Breast cancer	9	2	683
CMC	9	3	1473
Glass	9	7	214
Haberman	3	2	306
Hayes roth	4	3	132
Iris	4	3	150
Wine	13	3	178

In the following subsections we show the comparison of performance between conventional fuzzy classification systems and the proposed fuzzy systems. As discussed the difference between the conventional and the proposed method is that fuzzy if-then rules are generated by using (2) in the case of the conventional fuzzy classificaiton systems and by (10) in the case of the proposed one. In both cases, we partition each attribute of classification problems into three fuzzy sets (see Fig.1(b)). That is, the total number of fuzzy if-then rules generated for an n-dimensional pattern classification problem is 3^n.

Random Weighting

In a preliminary experiments we examined the performance of the proposed method under the situation that weihghts of training patterns are randomly determined. A weight was determined by a uniform random number in the interval of [0, 1]. That is, we randomly assign the importance of classification of training patterns in this subsection.

From these randomly weighted training pattterns we generate fuzzy if-then rules by using (10) to construct a fuzzy classification system. The classification ability of the system was examined for all given training patterns. We iterated the procedure 100 times. Note that we did not change the attribute values of training patterns but the value of their weights. That is, we examined the classification ability of fuzzy classification systems with 100 different sets of weights for training patterns. It should also be noted that the classification ability of the fuzzy classification that are generated by the conventional method (i.e., using (2)) is constant for the 100 iterations as the fuzzy rule-generation process does not consider the weights of training patterns.

We show the classification results of both the proposed and the conventional fuzzy classification sysetms in Table 2. Table 3 shows the cost of misclassification/rejection. From Table 2 and Table 3, we can see that the cost of misclassification/rejection is reduced by the proposed method while the number of correctly classified patterns by the proposed method is smaller than that by the conventional method. The reason of the reduction of the number of correctly classified patterns is that the fuzzy classification system constructed by the proposed method focuses on important training patterns with large weights.

Table 2 Classification results (Random weights)

Data set	Proposed	Conventional
Balance scale	91.5%	91.2%
Breast cancer	98.2%	98.2%
CMC	54.2%	55.4%
Glass	71.0%	72.0%
Haberman	73.8%	74.2%
Hayes roth	81.1%	86.4%
Iris	93.2%	94.0%
Wine	99.0%	98.9%

Table 3 Costs (Random weights)

Data set	Proposed	Conventional
Balance scale	24.5	27.3
Breast cancer	5.1	6.0
CMC	318.2	327.8
Glass	29.6	30.3
Haberman	39.8	39.5
Hayes roth	10.5	9.1
Iris	5.1	4.6
Wine	0.8	1.0

Class-Based Weighting

In this subsection we examine the case where the weights of training patterns are determined by using the class-based weighting method. That is, in this weighting method it is assumed that there is a priority of classification. We use the Harberman and Breast cancer data sets because they are related to medical diagnosis and are hence suitable for this weighting method. Both are two-classification problems. First, we focus on Class 1 training patterns. That is, we set $w_p = 1.0$ for Class 1 training patterns and $w_p = 0.5$ for Class 2 patterns. Table 4 and Table 5 show the classification results and the cost of misclassification /rejection, respectively.

We show in Table 6 and Table 7 the performance of fuzzy classification sysetms in the case where Class 2 training patterns are more focused on than Class 1 training patterns. From Tables 4-7, we can see that the cost of misclassification/rejection is reduced by the proposed method.

Table 4 Classification results (Class 1 focused)

Data set	Proposed	Conventional
Breast cancer	98.4%	98.2%
Haberman	73.9%	74.2%

Table 5 Costs (Class 1 focused)

Data set	Proposed	Conventional
Breast cancer	7	9
Haberman	40	40.5

Table 6 Classification results (Class 2 focused)

Data set	Proposed	Conventional
Breast cancer	98.4%	98.2%
Haberman	73.9%	74.2%

Table 7 Costs (Class 2 focused)

Data set	Proposed	Conventional
Breast cancer	7.5	9
Haberman	62.5	78.9

Overlap-Based Weighting

We examined the performance of our fuzzy classification sysetms for the case where overlap-based weighting method is used for training patterns. For each pattern classification problem, we specify the number of nearest training patterns as $N_p^{\text{nearest}} = 50$. That is, the proportion of the number of patterns from the same class to its 50 nearest patterns is used as the weight of the training pattern.

We examined the performance of the proposed method and the conventional one for all eight patern classification problems from Table 1. Classification results and costs of misclassification/reject are given in Table 8 and Table 9. From these tables, we can see that the performance of the proposed method is not better than that of the conventional one in terms of both classification results and the cost of misclassification/rejection. This is because many important training patterns with large weights are near the boundary area and it is difficult to correctly classify all important training patterns.

Table 8 Classification results (Overlap weighting)

Data set	Proposed	Conventional
Balance scale	90.9%	91.2%
Breast cancer	97.8%	98.2%
CMC	58.7%	60.8%
Glass	66.4%	72.0%
Haberman	74.2%	74.2%
Hayes roth	85.6%	86.4%
Iris	93.3%	94.0%
Wine	98.3%	98.9%

Table 9 Costs (Overlap weighting)

Data set	Proposed	Conventional
Balance scale	9.0	8.7
Breast cancer	2.3	2.4
CMC	189.6	184.2
Glass	17.8	15.9
Haberman	23.3	23.0
Hayes roth	7.1	6.8
Iris	4.4	3.8
Wine	1.7	1.3

Next, we examined the performance for test patterns. We used only 20% of the weighted training patterns to construct the fuzzy classification system. The other 80% were used as test patterns. That is, these test patterns are used to check the performance of the system for correctly classifying unseen patterns and for minimizing the cost function for unseen patterns. We conducted this kind of computer simulations 100 times, i.e. 100 different 20%-80% partitions of weighted patterns were used. We show the result of these computer simulations in Table 10 and Table 11. From these tables we can see that the performance of the proposed fuzzy rule-generation method for test patterns was improved.

Table 10 Classification results for test patterns

Data set	Proposed	Conventional
Balance scale	85.3%	84.3%
Breast cancer	93.8%	93.5%
CMC	45.3%	42.8%
Glass	57.4%	56.5%
Haberman	73.7%	73.6%
Hayes roth	43.0%	41.6%
Iris	91.1%	92.1%
Wine	89.1%	87.5%

Table 11 Costs for test patterns

Data set	Proposed	Conventional
Balance scale	23.9	27.5
Breast cancer	19.7	21.5
CMC	219.5	240.7
Glass	20.3	21.7
Haberman	18.1	18.6
Hayes roth	23.3	24.2
Iris	5.2	4.6
Wine	9.7	12.0

CONCLUSIONS

In this paper we proposed a fuzzy rule-generation method from weighted training patterns. Weights of patterns can be viewed as the grade of importance in the classification. The weights were treated as the costs of training misclassification/rejection. We formulate the problem of constructing classification problems from weighted training patterns as minimization of a cost function. Fuzzy if-then rules are generated by first

specifying the antecedent fuzzy sets and then determining the consequent class and grade of certainty. The determination of the consequent class and the grade of certainty of a fuzzy if-then rule is done by using compatible training patterns with the antecdent part of the fuzzy if-then rule. The proposed method considers the weight of training patterns in this process while the conventional one does not.

We also proposed two methods for weighting training patterns. One is a class-based weighting method where the weights of training patterns are determined by their class. This method can be used when different misclassifications should be distinguished or when patterns from some particular class should be focused on. The other method for weighting patterns is a overlap-based weighting method. This method can be used to focus on vague patterns that are placed near boundary areas and to obtain high generalization ability.

In computer simulations, we examined the performance of several fuzzy classification problems. We compared the performance of the proposed method and that of the conventional method. It was shown that the cost of misclassification/rejection was reduced by the proposed method.

Future work will incorporate a learning method of the grade of certainty. Since the determination of the grade of certainty is heuristically determined, it has to be optimized. We can consider two kinds of optimization for the learning of the grade of certainty. One is the optimization of the classification ability. That is, we can modify the grade of certainty so that the classification rate is maximized. In this case, the optimization function is formulated based on the number of correctly classified training patterns. This method is related to the learning method for the conventional fuzzy rule-generation method (see Nozaki et al. 1996). The other is the optimization of the cost of misclassification/rejection. That is, the modification of the grade of certainty is performed to minimize the cost. The decision of which optimization is used is dependent on the users' choice in real-application.

Another future work is how to optimize the grade of certainty. There are two main streams in the learning of fuzzy rules. One is a local learning method where optimization is performed by considering just one single fuzzy rule to be optimized. In this case, the optimal value of the grade of certainty is easily calculated from the formulation of optimization problem. However, this optimal value does not guarantee the globally optimum performance of fuzzy classification systeem as a whole. The other is a global learning method where the modification of the grade of certainty is performed from the classification results. This method is likely to produce the optimum fuzzy classification method, however, it is difficult to formulate the optimization problem mathmatically and to calculate the optimal value of the grades of certainty.

REFERENCES

M. Grabisch. 1996, "The representation of importance and interaction of features by fuzzy measures", *Pattern Recognition Letters*, Vol. 17:567-575.

M. Grabisch, F. Dispot. 1992,"A comparison of some methods of fuzzy classification on real data", *Proc. of 2nd Intl. Conf. on Fuzzy Logic and Neural Networks*, 659-662.

M. Grabisch, and J.-M. Nicolas. 1994, "Classification by fuzzy integral: performance and tests", *Fuzzy Sets and Systems*, Vol. 65, No. 2/3:255-271.

H. Ishibuchi, T. Nakashima. 1999a, "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems", *IEEE Trans. on Syst., Man, Cybernetics*, Part B Vol. 29:601-618.

H. Ishibuchi, T. Nakashima. 1999b, "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes", *IEEE Trans. on Industrial Electronics*, Vol. 46, No. 6:1057- 1068.

H. Ishibuchi, T. Nakashima. 2001, "T. Effect of rule weights in fuzzy rule-based classification systems", *IEEE Trans. on Fuzzy Systems*, Vol. 9, No. 4:506-515.

H. Ishibuchi, K. Nozaki, H. Tanaka. 1992, "Distributed representation of fuzzy rules and its application to pattern classification", *Fuzzy Sets and Systems*, Vol. 52, No. 1:21-32.

H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka. 1995, "Selecting fuzzy if-then rules for classification problems using genetic algorithms", *IEEE Trans. on Fuzzy Systems*, Vol. 3, No. 3:260-270.

G.J. Klir, B. Yuan. 1995, *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall.

C.C. Lee. 1990, "Fuzzy Logic in Control Systems: Fuzzy Logic Controller Part I and Part II", *IEEE Trans. Syst., Man, Cyberetics*, Vol. 20:404-435.

C.T. Leondes. 1999, "Fuzzy theory Systems", *Techniques and Applications. Academic Press*, San Diego, Vol.1-4, 1999.

T. Nakashima, H. Ishibuchi, A. Bargiela. 2004, "Constructing Fuzzy Classification Systems from Weighted Training Patterns, " *Proc. of 2004 IEEE International Conference on Systems, Man and Cybernetics*, pp. 2386-2391.

K. Nozaki, H. Ishibuchi, H. Tanaka. 1996, "Adaptive fuzzy rule-based classification systems", *IEEE Trans. on Fuzzy Systems*, Vol. 4, No. 3:238-250.

M. Sugeno. 1985, "An Introductory Survey of Fuzzy Control", *Information Science*, Vol. 30, No. 1/2:59-83.

Y. Yuan, H. Zhang. 1996, "A genetic algorithms for generating fuzzy classification rules", *Fuzzy Sets and Systems*, Vol. 84, No. 1:1-19.