Mathware & Soft Computing 5 (1998) 321-332

Analyzing the Reasoning Mechanisms in Fuzzy Rule Based Classification Systems *

Oscar Cordón[1], María José del Jesus[2], Francisco Herrera[1]
[1]Dept. of Computer Science and Artificial Intelligence
University of Granada, 18071 - Granada, Spain.
[2]Dept. of Computer Science
University of Jaén, 23071 - Jaén, Spain.
e-mail: ocordon, herrera@decsai.ugr.es, mjjesus@ujaen.es

Abstract

Fuzzy Rule-Based Systems have been succesfully applied to pattern classification problems. In this type of classification systems, the classical Fuzzy Reasoning Method classifies a new example with the consequent of the rule with the greatest degree of association. By using this reasoning method, we do not consider the information provided by the other rules that are also compatible (have also been fired) with this example.

In this paper we analyze this problem and propose to use FRMs that combine the different rules that have been fired by a pattern. We describe the behaviour of a general reasoning method and analyze two kinds of models, the first one using all the fired rules and the second one using partial information due to the fact that the rules with a lower association degree are not considered.

1 Introduction

Fuzzy Rule-Based Systems (FRBSs) combine the precision of the prediction with a high level of interpretability, which makes them very suitable for designing Classification Systems in real problems.

As is well known in Fuzzy Rule Based Classification Systems (FRBCSs), the classical Fuzzy Reasoning Method (FRM), maximum matching, classifies a new example with the consequent of the rule with the greatest association degree ([1, 4, 10, 11, 13, 14]). Using this inference method, we lose the information provided by the other fuzzy rules with different linguistic labels which also represent the value in the pattern attribute, although probably to a lesser degree.

*This work has been supported by CICYT under Project TIC96-0778

On the other hand, it is well known that in other FRBSs as the fuzzy logic controllers the best performance is obtained when we use defuzzification methods that operate on the fuzzy subsets obtained from the fuzzy rules fired, taking in consideration all of them for obtaining the output value via the defuzzification method ([5]).

The aim of this paper is to study in depth this problem, under the idea of taking into consideration the information available in FRBCSs for performing inference, i.e., to consider all the fired rules for making decisions as it happens in fuzzy logic controllers. We propose to use FRMs that combine the information given by the different rules fired by a pattern. We describe the behaviour of a general reasoning method for fuzzy rules with a class and a certainty degree associated to the classification of that class in the consequent, and analyze two kinds of inference models:

- the first one using all the information available, thus making all rules fired by a pattern participate in the classification of such pattern, and
- the second one using partial information consisting on selecting the fired rules to be considered in the inference process. The rules with a lower association degree are not used. This selection is modeled through operators from the OWA family [17, 18] under the fuzzy majority concept.

To achieve this aim, the paper is organized as follows. Section 2 describes the FRBCS structure and the general model of FRMs for fuzzy rules with a class and a certainty degree associated to the classification of that class in the consequent; Section 3 presents the said two Inference models, FRMs integrating all fuzzy rules and FRMs selecting a subset of fuzzy rules; Section 4 shows the results obtained using them in the Pima database, considering Rule Bases generated by means of two different learning methods. Finally, Section 5 introduces some concluding remarks.

2 Fuzzy Rule-Based Classification Systems

In an FRBCS, two components are distinguished: 1) The Knowledge Base (KB), composed of a Rule Base (RB) and a Data Base (DB), which is specific for a given classification problem, and 2) an FRM.

The FRBCS design implies finding both components, and this process is carried out through a supervised learning process, which starts with a set of correctly classified examples (training examples) and whose ultimate objective is to design a Classification System, assigning class labels to new examples with a minimum error. Finally, the system performance on the test data is computed to gain an estimate about the FRBCS real error. This process is described in Figure 1 and the structure of the FRBCS components is briefly described in the following subsections.

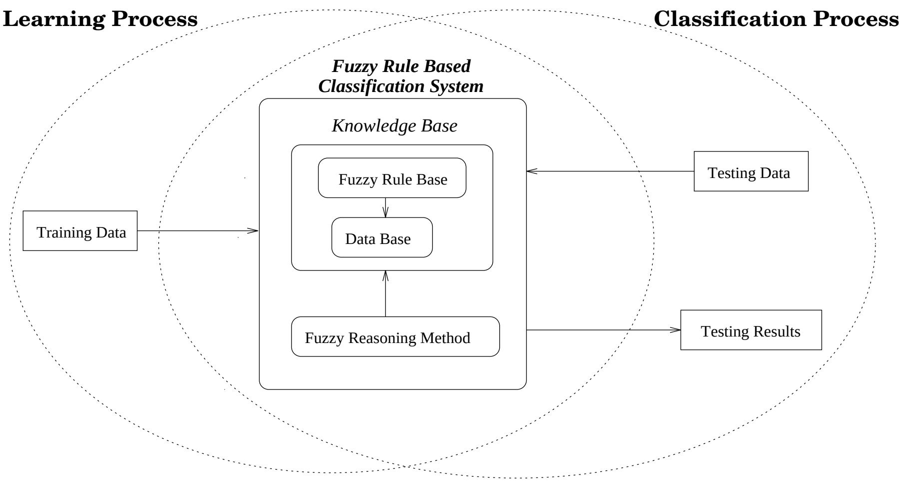

Figure 1: Design of an FRBCS (learning/classification)

2.1 Knowledge base

The KB is composed of the RB and the DB. In the specialized literature, several rule types have been used. The difference among the three existing types of fuzzy classification rules lies on the composition of the consequent: a class ([1, 10]), a class and a certainty degree associated to the classification of that class ([11]), and the certainty degree associated to the classification of each one of the possible classes ([14]).

In [7], a genetic learning method was developed for RBs of each type mentioned, and it was shown that the fuzzy rules having a class and a certainty degree in the consequent represent better the knowledge over the corresponding space zone. In this work, we will consider FRBCSs composed of an RB of the following type:

$$R_k: \text{ If } x_1 \text{ is } A_1^k \text{ and } \ldots \text{ and } x_N \text{ is } A_N^k \text{ then } Y \text{ is } C_j \text{ with } r^k$$

where:

- $x_1, \ldots, x_N$ are the selected features for the classification problem,
- $A_1^k, \ldots, A_N^k$ are linguistic labels used to discretize the continuous variable domain,
- Y is the class $C_j \in \{C_1, \ldots, C_M\}$ to which the example belongs, and
- r^k is the classification certainty degree in the class C_j for an example belonging to the fuzzy subspace defined by the rule antecedent.

The DB contains the definition of the fuzzy sets associated to the linguistic terms used in the RB. This mapping is common for all the rules in the RB to maintain the linguistic nature of the FRBCS.

2.2 Fuzzy reasoning methods

An FRM is an inference procedure, which derives conclusions from a fuzzy rule set and an example. The use of a reasoning method that combines the information of the rules fired with the pattern to be classified can improve the generalization capability of the Classification System.

In [6], we introduced a general reasoning model, that in this paper is particularized to an RB composed of rules with a class and its certainty degree in the consequent. This model is described in the following.

In the classification of an example $E^t = (e_1^t, \ldots, e_N^t)$, the RB $R = \{R_1, \ldots, R_L\}$ is divided into M subsets according to the class indicated by its consequent,

$$R = R_{C_1} \cup R_{C_2} \cup \ldots \cup R_{C_M}$$

and the next scheme is followed:

1. **Compatibility degree.** The compatibility degree of the antecedent with the example is computed for all the rules in the RB, applying a t-norm ([2, 9]) over the membership degree of the values of the example (e_i^t) to the corresponding fuzzy subsets.
$$R^k(E^t) = T(\mu_{A_1^k}(e_1^t), \ldots, \mu_{A_N^k}(e_N^t)), \quad k = 1, \ldots, L$$
2. **Association degree.** The association degree of the example E^t with the M classes is computed according to each rule in the RB.
$$\begin{array}{l} b_i^k = h(R^k(E^t), r^k), \quad k = 1, \ldots, |R_{C_i}| \\ i = 1, \ldots, M \end{array}$$
3. **Weighting function.** The values obtained are weighted by means of a function g. An expression which promotes the highest values and penalizes the smallest ones seems to be the most adequate choice for this function.
$$\begin{array}{l} B_i^k = g(b_i^k), \quad k = 1, \ldots, |R_{C_i}| \\ i = 1, \ldots, M \end{array}$$
4. **Pattern classification soundness degree for all classes.** To compute this value, an aggregation operator is used ([2, 9]) which combines, for each class, the positive association degrees computed in the previous step
$$\begin{array}{l} Y_i = f(B_i^k, \ k = 1, \ldots, |R_{C_i}| \ \text{ and } \ B_i^k > 0), \\ i = 1, \ldots, M \end{array}$$
with f being an aggregation operator that returns a value between the minimum and the maximum.

   It is clear that if we select f as the maximum operator, we have the classical FRM.

5. **Classification.** A decision function F is applied to the classification degrees of the example. This function will return the class label corresponding to the maximum value.

$$C_l = F(Y_1, \ldots, Y_M) \text{ such that } Y_l = \max_{j=1,\ldots,M} Y_j$$

3 Two Kinds of Inference Models

The classical FRM, called *maximum matching*, considers the rule with the highest association degree. It classifies the pattern with the class of this rule. Graphically, this method could be seen as shown in Figure 2, where the rule R_k would show the highest association degree.

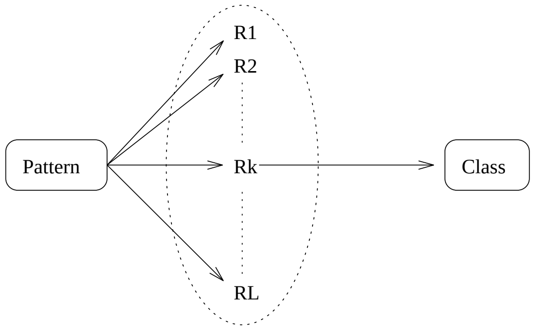

Figure 2: Fuzzy reasoning method that uses only the winner rule

The expression of function f_0 represents the classical FRM,

$$f_0(a_1, \ldots, a_s) = \max_{i=1,\ldots,s} a_i$$

with $a_1, \ldots, a_s$ being the values to aggregate for an example E^t with respect to a class C_j.

By using this reasoning method, we do not consider the information provided by the other rules that are also compatible (have also been fired) with this example. We propose to use FRMs that combine the information given by the different rules fired by a pattern.

According to the general reasoning model, we propose two kinds of inference models. The difference is based on the use of the function $f(\cdot)$ in step 4, and we distinguish between FRMs integrating all fuzzy rules and FRMs selecting a subset of fuzzy rules (the classical FRM is a particular case of this group in which this subset is composed of a single rule, the one with the highest association degree). They are analysed in the two following subsections.

A common proposal for some of the operators implied in the general model of reasoning is described in Table 1 ([6]).

1. Compatibility degree $R^k(E^t) = \min_{i=1,\ldots,N} \mu^k_{A_i}(e^t_i)$	*3. Weighting functions* $g_1(x) = x, \quad \forall x \in [0,1]$
2. Association degree $h(R^k(E^t), r^k_j) = R^k(E^t) \cdot r^k_j$	$g_2(x) = \begin{cases} x^2, & \text{if } x < 0.5 \\ \sqrt{x}, & \text{if } x \geq 0.5 \end{cases}$

Table 1: Proposals for the general model of reasoning operators

3.1 FRMs integrating all fuzzy rules

The first group of FRMs being an alternative to the classical one is formed by those that use all fired fuzzy rules for deriving conclusions from a set of fuzzy if-then rules and a pattern. This idea is represented graphically in Figure 3.

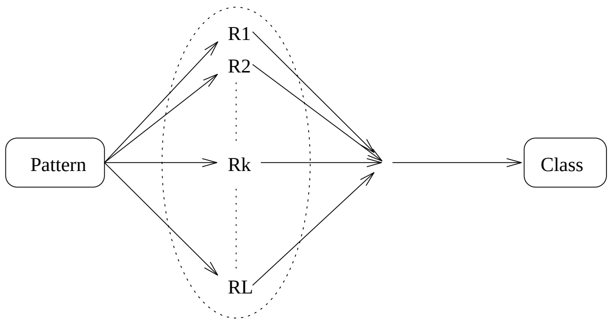

Figure 3: FRM integrating all fuzzy rules

Some proposals for the function f in FRBCSs belonging to this family are described in Table 2 ([6]).

4. Aggregation functions	
Normalized Sum $f_1(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i}{f_{1_{max}}}$ $f_{1_{max}} = \max_{j=1,\ldots,M} \sum_{i=1}^s a_i$	**Sowa and-like** $f_4(a_1,\ldots,a_s) = \alpha \cdot a_{min} + (1-\alpha)\frac{1}{s}\sum_{i=1}^s a_i$ $\alpha \in [0,1], \quad a_{min} = min\{a_1,\ldots,a_s\}$
Arithmetic Mean $f_2(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i}{s}$	**Sowa or-like** $f_5(a_1,\ldots,a_s) = \alpha \cdot a_{max} + (1-\alpha)\frac{1}{s}\sum_{i=1}^s a_i$ $\alpha \in [0,1], \quad a_{max} = max\{a_1,\ldots,a_s\}$
Quasiarithmetic Mean $f_3(a_1,\ldots,a_s) = H^{-1}\left[\frac{1}{s}\sum_{i=1}^s H(a_i)\right]$ $H(x) = x^p, \quad p \in \mathrm{R}$	**Badd** $f_6(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i^{\alpha+1}}{\sum_{i=1}^s a_i^{\alpha}}, \quad \alpha \in \mathrm{R}$
with $a_1,\ldots,a_s$ being the values to aggregate for an example E^t with respect to a class C_j	

Table 2: Different proposals for aggregating the association degrees

The function f_1 accumulates the association degree between the pattern and the class for all the rules in the RB. This sum is then divided by the maximun sum for all classes, to obtain a normalized value. Bardossy et al. studied this FRM as a method for combining fuzzy rule responses, called *additive combination* ([3]). This operator was also presented by Chi et al. in [4] like a defuzzification method to produce a classification result for an FRBCSs called *maximum accumulated matching*. Ishibuchi et al. used it in [12] as well, where the inference result is given by the voting of the fuzzy if-then rules that are compatible with the pattern, it was called *fuzzy reasoning method based on the maximum vote*.

The arithmetic (f_2) and quasiarithmetic (f_3) means are two operators with a compensation degree between the minimum and the maximum that combine the information given by each local classifier and obtain an average degree that considers the quality of the rules in the inference process. The functions f_4 and f_5 (Sowa and-like and Sowa or-like operators) give us a parameterized aggregation value lying between the minimum and the maximum as well. The function f_6 has not only a behaviour between the same extents, but also considers a weighting of the values to aggregate by its definition, as function f_3.

3.2 FRMs selecting a subset of fuzzy rules

In the general inference model, the functions proposed as aggregation operators (Table 2) consider the information given by all the rules compatible with the example. This behaviour could mean that, in large RBs, many rules presenting a low association degree between the example and the class could have more influence than those having a higher degree, leading the system towards a mistaken classification.

The incorporation of an FRM that considers only the information provided by the most suitable rule subset for the example in each new prediction may improve the system performance. This selection can be carried out by means of operators belonging to the OWA family under the fuzzy majority concept.

The OWA operators (Ordered Weighted Averaging) ([17, 18]) constitute an aggregation operator class completely covering the interval between the minimum and the maximum operators.

Their most important characteristic is the use of weights in the values to be aggregated, which are not associated to their particular values, but rather to their ordered position. This property allows us to associate, in the inference process, weights to the association degrees of the example, in an ordered way, with different rules, and therefore to enhance the most predictive rules for that example.

To define the OWA operator behaviour in the inference process environment, we consider that $(a_1, \ldots, a_s)$ are the values to be aggregated in the process in which the classification degree of an example in a class C_j is obtained. Let $W = (w_1, \ldots, w_s)$ be a vector of weights verifying that

$$\begin{array}{ll} 1) & \omega_i \in [0,1], \quad \forall i = 1, \ldots, s \\ 2) & \sum_{i=1}^{s} \omega_i = 1 \end{array}$$

The OWA aggregation operation related to the vector of weights W is the function

$$f_7(a_1, \ldots, a_s) = \sum_{i=1}^{s} \omega_i \cdot b_i$$

where $(b_1, \ldots, b_s)$ is a permutation of the vector $(a_1, \ldots, a_s)$ in which the elements are sorted in a decreasing way.

The parameterized behaviour of the OWA operator class is determined by the vector of weights W, which could be basically obtained in two different ways: the first one implies the use of a learning mechanism, considering a set of examples, their arguments and the aggregated values, and trying to fit the weights to this data collection through any kind of regression model. The second choice, on which this research is centred, is based on giving a semantic content to the weights. To carry out this task, they could be computed using linguistic quantifiers, which represent the fuzzy majority concept.

The fuzzy majority is a relaxed concept of majority, commonly represented by means of a fuzzy quantifier ([19]), that constitutes an attempt to establish a bridge between the formal systems and the human discourse, as well as providing a more flexible knowledge representation tool, and whose semantic can be represented by a fuzzy subset.

We will work using non-decreasing proportional quantifiers Q, defined in such a way that for any $r \in [0, 1]$, $Q(r)$ shows the degree to which the proportion r is compatible with the meaning represented by the quantifier. The membership function of a non-decreasing proportional quantifier can be defined as

$$Q(r) = \begin{cases} 0, & \text{if } r < a \\ \frac{r-a}{b-a}, & \text{if } a \le r \le b \\ 1, & \text{if } r > b \end{cases}$$

with $a, b, r \in [0, 1]$,

A possible way to compute the OWA weights using a non-decreasing proportional fuzzy quantifier Q is as follows ([17, 18]):

$$\omega_i = \begin{cases} Q\left(\frac{i}{s}\right) - Q\left(\frac{i-1}{s}\right), & i = 2, \ldots, s \\ Q\left(\frac{i}{s}\right), & i = 1 \end{cases}$$

with s being the number of values to be aggregated.

The OWA operator and this way of computing the array of related weights allow us to determine the proportion of rules, which are more compatible with the example, that we want to be involved in the inference process. To do so, we only have to specify the parameters a and b, which define the fuzzy quantifier, and compute the array of weights for the aggregation process, in which this quantifier is underlying. In this way, we are defining an inference method in which not all the rules are involved, but the "majority", and this concept of fuzzy majority is represented by a fuzzy quantifier. The resulting inference model shows similarities with classification schemes frequently used by human experts who, in many cases, neither classify according to only one rule (classical FRM), nor according to all of them, but decide depending on the majority of the rules established.

This idea is represented graphically in Figure 4.

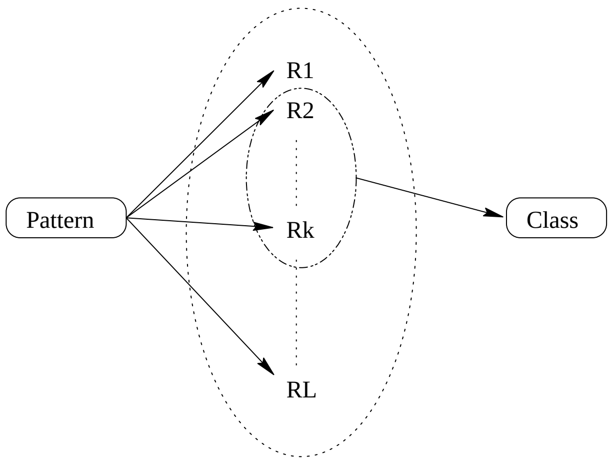

Figure 4: FRM selecting a subset of fuzzy rules

4 Experiments

To analyze the behaviour of the proposed methods, we have generated RBs for the Pima example base, with two different supervised learning methods:

- the extension of Wang and Mendel's algorithm ([15]) to classification ([4]), and
- the genetic learning method developed by the authors in [7].

Pima is a set of 768 cases involving the diagnosis of diabetes where eight variables are taken into account, with two possible classes (having or not having the complaint).

To estimate the real error made by the FRBCS we have used the random resampling method ([16]) with five random partitions of the example base into training and test examples (70 % and 30 %, respectively). The results we show are means of correct classification percentages for the five training and test sets (noted as Tra and Test, respectively) and the number of rules (described by NR).

In the following tables, the FRBCS obtained from the extension of Wang and Mendel's algorithm will be noted as FRBCS-T1, and those generated by the genetic learning method as FRBCS-T2. In the same way, FRMs will be described with a numbered function corresponding to the aggregation function which they are based on, according to the notation shown in Table 2 and in the previous subsection. When the FRM includes the weighting function g_2 in its definition, we will label it with a W. When the function used is g_1, it will be no labeled.

In Table 3, the best results obtained for the Pima example base are shown, using an FRBCS composed of an RB generated with the extension of Wang and Mendel's algorithm. From the FRMs analyzed (FRMs presented in Table 2 and FRMs based in OWA operators), we have selected those that have the best behaviour for the problem considering this RB ([6, 8]).

Table 4 shows the best results from the same experiments with RBs obtained by the genetic learning method ([7, 8]).

FRM	Tra	Test	NR
f_0	85.81	73.23	420.2
f_3 (p=20)	85.78	73.43	420.2
f_5 (α=0.9)	85.81	73.22	420.2
f_6 (p=10)	85.85	73.53	420.2
f_7 (a=0, b=0.3)	83.97	73.12	420.2
f_7 (a=0, b=0.5)	83.37	73.12	420.2

Table 3: Pima. FRBCS-T1

FRM	Tra	Test	NR
f_0	81.81	74.06	147.4
f_1 W	83.27	75.81	197.6
f_3 (p=20) W	78.14	74.17	154.8
f_5 α=0.7	80.87	74.68	160.4
f_7 (a=0, b=0.3)	84.21	76.20	185.2
f_7 (a=0, b=0.3) W	83.86	75.82	177

Table 4: Pima. FRBCS-T2

The FRMs present different behaviour for every learning method. In both cases, the best result is obtained by an FRM beloging to one of the proposals. We may conclude that the new reasoning methods improve the performance of the FRBCSs.

We should point out that the genetic learning process includes a multiselection process that gives several KB definitions with the minimum subsets of rules that cooperate with the FRMs. Due to this multiselection process, the number of rules is lesser than the number of rules obtained by the other learning method. This fact allows this learning method to present better results than the extension of Wang and Mendel's algorithm.

The results of all the experiments developed with the FRMs are shown in [6, 7, 8].

5 Concluding Remarks

In this paper we have presented the use of FRMs that combine the different rules that have been fired by a pattern in FRBCSs. We have described the behaviour of a general reasoning model for fuzzy rules with a class and a certainty degree associated to the classification of that class in the consequent and two kinds of models have been analyzed, the first one using all the fired rules and the second one using partial information since the rules with a lower association degree are not considered.

Finally, we must point out that some additional studies are necessary for analyzing all the possibilities of the new FRMs for increasing the generalisation capability of the FRBCSs. They are those associated to: the use of different weighting functions, the use of selection criteria for eliminating rules with a low association degree, the use of alternative methods for getting the weights in the OWA operator, and the possible fusion of both approaches integrating an aggregation operator acting over the selected set of rules.

References

[1] S. Abe, R. Thawonmas, A fuzzy classifier with ellipsoidal regions, IEEE Transactions on Fuzzy Systems **5:3** 358-368 (1997).

[2] C. Alsina, E. Trillas, L. Valverde, On some logical connectives for fuzzy set theory, J. Math. Anal. Appl. **93** 149-163 (1983).

[3] A. Bárdossy, L. Duckstein, Fuzzy rule-based modeling with applications to geophysical, biological and engineering systems, CRC Press (1995).

[4] Z. Chi, H. Yan, T. Pham, Fuzzy algorithms with applications to image processing and pattern recognition, World Scientific (1996).

[5] O. Cordón, F. Herrera, A. Peregrín, Applicability of the fuzzy operators in the design of fuzzy logic controllers, Fuzzy Sets and Systems **86** 15-41 (1997).

[6] O. Cordón, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, International Journal of Approximate Reasoning (1999). To appear.

[7] O. Cordón, M.J. del Jesus, F. Herrera, Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods, International Journal of Intelligent Systems **13** (1998) 1025-1053.

[8] O. Cordón, M.J. del Jesus, F. Herrera. Reasoning methods based on OWA operators under fuzzy majority in fuzzy rule-based classification systems. Technical Report DECSAI-98121, Dept. Computer Science and Artificial Intelligence, University of Granada, Granada, Spain, July 1998.

[9] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives, Information Sciences **36** 85-121 (1985).

[10] A. González, R. Pérez, Completeness and consistency conditions for learning fuzzy rules, Fuzzy Sets and Systems **96** 37-51 (1998).

[11] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and Systems **52** 21-32 (1992).

[12] H. Ishibuchi, T. Morisawa, T. Nakashima, Voting schemes for fuzzy-rule-based classification systems, Proc. Fifth IEEE Int. Conference on Fuzzy Systems 614-620, New Orleans (1996).

[13] L. I. Kuncheva, On the equivalence between fuzzy and statistical classifiers, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **4:3** 245-253 (1996).

[14] D.P. Mandal, C.A. Murthy, S.K. Pal, Formulation of a multivalued recognition system, IEEE Trans. on Systems, Man, and Cybernetics **22:4** 607-620 (1992).

[15] L.-X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Transactions on Systems, Man, and Cybernetics **22:6** 1414-1427 (1992).

[16] S.M. Weiss, C.A. Kulikowski, Computer systems that learn, Morgan Kaufmann Publishers (1991).

[17] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. on Systems, Man and Cibernetics **18:1** 183-190 (1988).

[18] R.R. Yager, Families of OWA operators, Fuzzy Sets and Systems **59** 125-148 (1993).

[19] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Comps. and Maths. with Appls. **9** 149-184 (1983).