A new fuzzy 3-rules pattern classifier with reject options based on aggregation of membership degrees

Hoel Le Capitaine **Carl Frélicot**
MIA Laboratory, University of La Rochelle, France
{hlecap01,cfrelico}@univ-lr.fr

Abstract

In this paper, we address the problem of fuzzy rule-based pattern recognition with reject options. These options are made possible thanks to simple rules whose satisfaction level is expressed by the value of dedicated operators that aggregate degrees of typicality. Results obtained with the proposed classifier on articicial and real data are given.
Keywords: Pattern recognition, reject options, fuzzy rules, aggregation.

1 Introduction

In many decision-making systems, we face the problem of aggregating collections of numerical or ordinal data to obtain a typical value. Aggregation operators are used to obtain an overall score for each alternative, which is exploited to establish a decision. In the context of pattern recognition, such a decision consists in assigning objects to a given class. Since the publication of L. A. Zadeh's paper on fuzzy sets, this theory has evolved into powerful tools for managing uncertainty in decision-making systems. Many research works have been carried out for applications to pattern recognition, e.g. fuzzy rule-based classifiers such as Takagi-Sugeno-Kang ones that approximate classification boundaries [9]. Fuzzy classifiers generally have faster training capabilities and comparable generalization abilities to other ones. This paper deals with the design of fuzzy rules whose inputs are the membership degrees of objects to the classes at hand instead of features describing them like in classical fuzzy if-then classifiers. By including possible rejection of extraneous and ambiguous objects, we allow to significantly improve the performance of such a classification or decision-making system.
The paper is organized as follows. In section 2, we recall the formal definition of aggregation operators and some special functions. In section 3, a brief overview of fuzzy classifier design and fuzzy rule-based classifiers is given. Section 4 briefly describes the principles of pattern rejection and the existing strategies leading to the different options: exclusive classification, ambiguity or distance rejection. Then, the new approach for obtaining classification boundaries using simple fuzzy rules is presented and discussed. Section 5 present results obtained on both artificial and real data sets. Concluding remarks and ideas for future work are finally given in section 6.

2 Aggregation Operators

The aggregation problem is of major importance in decision-making systems, where values to be aggregated are generally defined on a finite real interval or on ordinal scales. In this paper, we assume with no loss of generality that they come from the unit interval. If not, a simple transformation can be found to make this assumption true. Among the frequently used aggregation operators, in addition to the mean operators, one can cite: triangular norms [10], OWA (*Ordered Weighted Averaging*) operators [16], γ-operators [19], or fuzzy integrals [14]. These operators are divided in several categories, depending on the way the values are aggregated: conjunctives, disjunctives, compensatory, compen-

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 473–480
Torremolinos (Málaga), June 22–27, 2008

sative, and weighted operators. An aggregation operator $\mathcal{A}$ on the unit interval is said to be conjunctive if $\mathcal{A}(x_1, \cdots, x_n) \leq min(x_1, \cdots, x_n)$. The values are combined by a fuzzy logical *AND*, which means that the overall score is high if and only if all partial scores are high. If we add properties of non decreasingness, commutativity and associativity, we obtain the family of the triangular norms (t-norms). A triangular norm is a commutative, associative and monotone function $T : [0,1]^2 \rightarrow [0,1]$ having for neutral element 1, i.e. $T(x,1) = x$ for $x \in [0,1]$. The minimum operator is a t-norm and $T(x,y) \leq x \wedge y$, with $\wedge = min$, so that the minimum is the greatest t-norm. An aggregation operator on the unit interval is said to be disjunctive if $\mathcal{A}(x_1, \cdots, x_n) \geq max(x_1, \cdots, x_n)$. The values are combined by a fuzzy logical *OR*, which means that the overall score is low if and only if all partial scores are low. If we add properties of non decreasingness, commutativity and associativity, we obtain the family of the triangular conorms (t-conorms). A triangular conorm is a commutative, associative and monotone function $S : [0,1]^2 \rightarrow [0,1]$ having for neutral element 0, i.e. $S(x,0) = x$ for $x \in [0,1]$. The maximum operator is a t-conorm and $S(x,y) \geq x \vee y$, with $\vee = max$, so that the maximum is the weakest t-conorm. Table 1 shows the three basic t-norms ($\top$) and t-conorms ($\perp$), but there exists a large panel of triangular norms (indeed infinite since the combination of two t-norms is a t-norm), see [8] for a survey.

Table 1: Basic t-norms and t-conorms couples

Standard	$a \top_S b = \min(a,b)$
	$a \perp_S b = \max(a,b)$
Algebraic	$a \top_A b = a\, b$
	$a \perp_A b = a + b - a\, b$
Lukasiewicz	$a \top_L b = \max(a+b-1, 0)$
	$a \perp_L b = \min(a+b, 1)$

An aggregation operator on the unit interval is said to be compensatory if $min(x_1, \cdots, x_n) \leq \mathcal{A}(x_1, \cdots, x_n) \leq max(x_1, \cdots, x_n)$. Here, a high value (resp. low) can be compensated by a low value (resp. high). If we add properties of non decreasingness and idempotency, we obtain the family of mean operators. Conjunctives, disjunctives and compensatory operators form a large part of aggregation operators on $[0,1]$, but one can find some operators that do not belong to any of these categories, e.g. the symmetric sum [13] and compensative operators [19]. Both compensative and compensatory operators tend to express a compromise within the values, but the output of the former operators do not necessarily lie between the minimum and the maximum values whereas the latter do.

To conclude this section, let us mention the existence of weighted operators. Multi-criteria decision often need to establish the importance of each evaluated criterion, which implies an extension of the usual non weighted operators. These weights cause the loss of neutrality from the decision system, but could allow to perform better results. OWA operators are used to adjust the terms AND and OR, and allow an easier semantic interpretation of the linguistic quantifiers. Fuzzy integrals compute the mean value of a given function with respect to a fuzzy measure and thus can be seen as aggregation operators in the discrete case.

3 Fuzzy Pattern Classification

3.1 Classifier Design

Let $x = {}^t(x_1 \;\; x_2 \; ... \; x_p)$ be a pattern in a feature space, to be classified with respect to a set $\Omega = \{\omega_1, \cdots, \omega_c\}$ of c classes. A conventional hard classifier is a rule aiming at assigning an unknown pattern x to a particular class ω_i, thanks to the aggregation of class-labels $u_j(x)$, e.g. posterior probabilities that x belongs to the classes or membership degrees to fuzzy sets associated with the classes. We do not address the labelling problem in this paper, so we will use a measure of typicality:

$$u_i(x) = \frac{\alpha}{\alpha + d^2(x, p_i)} \quad (1)$$

where α is a user-defined parameter, d a distance, and p_i a prototype of ω_i obtained from a learning set of patterns. The most popular agregation operator is the standard t-conorm, defining the so-called *max classifier* (MC):

1. compute class-labels $u_j(x)$ $(j = 1, c)$
2. aggregate labels $\mathcal{A}(u_1, \cdots, u_c)$
3. rule: if $\mathcal{A}$ is u_i then assign x to ω_i

Such an exclusive classification rule is not so efficient because it supposes that classes do not significantly overlap (*separability*) and that Ω is exhaustively defined (*closed-world*). These assumptions are generally not valid in practice.

3.2 Fuzzy Rule-Based Classifiers

Fuzzy systems are meant to be models understandable for the end-user. They use if-then rules and a mechanism which should correspond to the expert knowledge for a given problem. A fuzzy if-then classifier consists in a model with fuzzy rules of the form:

$$\text{if } A_i^1 \text{ AND } A_i^2 \text{ AND} \cdots \text{ AND } A_i^n \text{ then } B_i \quad (2)$$

where A_i^k is a fuzzy set with membership function $a_i^k : \mathbb{R} \to [0,1]$, $i = 1, \cdots, m$, $k = 1, \cdots, p$ and $B_i \in \mathbb{R}$. Among the most popular models, the Takagi-Sugeno-Kang (*TSK*) model [15] is characterized by:

1. a set of m fuzzy rules.
2. a connective operator $\mathcal{A}$ whose output provides the satisfaction τ_i, or firing strength, of the rule i:
$$\tau_i(x) = \mathcal{A}(a_i^1(x_1), \cdots, a_i^p(x_p)) \quad (3)$$
3. a defuzzification method allowing the final assignment.

For instance, by choosing the product for the connective operator and the COA (*Center Of Area*) defuzzification method, we obtain a TSK2 [9] classifier for the object x:

$$C(x) = \frac{\sum_{i=1}^m B_i \prod_{k=1}^p a_i^k(x_k)}{\sum_{i=1}^m \prod_{k=1}^p a_i^k(x_k)} \quad (4)$$

Let us consider a simple two-classes problem as first example. We generated 100 two-dimensional samples equally arising from two normal distributions $\omega_1 \sim \mathcal{N}\left(p_1 = [1,3]^T, I\right)$ and $\omega_2 \sim \mathcal{N}\left(p_2 = [5,1]^T, I\right)$ where I is the identity matrix. Two rules are enough to define a TSK2 classifier for this problem:

- rule R1:
  if x_1 is about 1 AND x_2 is *about 3*, then b_1
- rule R2:
  if x_1 is about 5 AND x_2 is *about 1*, then b_2

where *about c* can be modelled by the following membership function:

$$a^k(x_k) = \exp\left(-(x_k - c)^2/2\right) \quad (5)$$

Thus, the satisfaction τ_i ($i = 1, 2$) of the rules are:

$$\tau_i(x) = \exp\left(-(x - p_i)^T(x - p_i)/2\right)$$

If we set b_1 to 1 and b_2 to 0, we obtain:

$$C(x) = \frac{\tau_1}{\tau_1 + \tau_2} \quad (6)$$

and the decision areas shown in Fig. 1 that vary from black, corresponding to ω_2 ($b_2 = 0$), to white corresponding to ω_1 ($b_1 = 1$). The classification boundary is given by $\tau_1 = \tau_2$, whose solution is the following hyperplan equation:

$$x_2 = \frac{3}{2}x_1 - 7 \quad (7)$$

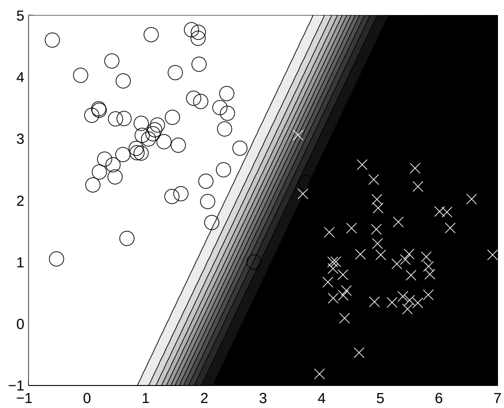

Figure 1: A simple two-classes $(\circ, \times)$ problem with normal distribution in $\mathbb{R}^2$

Many membership functions from $\mathbb{R}$ to $[0,1]$ can be used, then too much flexibility arises, leading to untractable learning. So we restrict the study to positive definite functions. Since classes may have ellipsoidal shapes, Abe and Thawonmas introduced a fuzzy classifier, based on neural networks and fuzzy rules, which provides ellipsoidal decision areas [1]. They use the following membership function:

$$a_i(x) = \exp(-h_i^2(x))$$
$$h_i^2(x) = \frac{d_i^2(x)}{\alpha_i}$$
$$d_i^2(x) = (x - p_i)^T \Sigma_i^{-1}(x - p_i)$$

where Σ_i is the covariance matrix of the class ω_i and α_i is a user-defined parameter.
Furthermore, it has been proved [9] that some TSK classifiers with adequate membership functions are equivalent to non-parametrical classifiers:

- TSK3 with $a_i(x) = \exp\Big(-\frac{1}{2}(x-p_i)^2\Big)$ is equivalent to the nearest neighbor classifier,
- TSK4 with $a_i(x) = \exp\Big(-\frac{1}{2h^2}(x-p_i)^2\Big)$ is equivalent to a Parzen classifier.

In [7], the authors proved that a monotonic function $f(x)$ can model the classification boundary for two classes in $\mathbb{R}^2$, and they propose to use other t-norms and t-conorms than minimum and maximum because they fail in $p > 2$-dimensional problems, even for linear separable problems.

4 The New Classifier

4.1 Reject Options

We said in subsection 3.1 that the separability and the closed-world assumptions are generally not valid in practice. Reject options have been proposed to overcome these difficulties and to reduce the misclassification risk. The first one, called distance reject option, is dedicated to outlying patterns. If x is far from all the class-prototypes, the option allows to assign it to no class. The second one allows to assign inlying patterns to several or all the classes. If x is close to two or more prototypes, it is associated with the corresponding classes. Finally, a pattern is exclusively classified when its maximum membership degree is significantly higher than the others. Different strategies can be adopted to handle the options at hand [4, 5] but they all lead to a three types decision system: exclusive classification, ambiguity rejection, distance rejection.

4.2 Reject Options Using Fuzzy Rules

In this section, we detail the construction of reject operators with the help of fuzzy rules. Step by step, by combining triangular norms, we define what the suitable operator should satisfy. In [18], Yager and Rybalov showed that t-norms and t-conorms provide downward and upward reinforcement, respectively. Downward (respectively upward) reinforcement corresponds to the fact that if the values are all low (respectively high), then they reinforce each other and the overall value will be low (respectively high). From these properties, they build a full reinforcement operator by the use of a fuzzy system modeling technique. We follow this approach and define new rules. According to the rules defined, the operator will have a conjunctive, disjunctive or a compensative behavior.

- rule $R1$: there is only one high value,
  (so exclusive classification is possible)
- rule $R2$: several values are high,
  (so ambiguity rejection is needed)
- rule $R3$: all the values are low
  (so distance rejection is needed)

We propose to formalize this set of rules.

4.3 The Fuzzy 3-Rules based Classifier with Reject Options (3-RCRO)

Let L, respectively H, be the fuzzy subset defined on $[0,1]$ corresponding to the concept *low*, respectively *high*. Furthermore, let (T,S) be any dual couple (t-norm,t-conorm). As mentioned in the second section, the t-conorm S is the fuzzy equivalent of the logical operator OR. We define $H(u_i) = u_i$ using a linear membership function and $L(u_i) = 1 - u_i$ as the negation of H. In [6], the authors built a fuzzy exclusive OR operator that extends the crisp XOR operator to the fuzzy context:

$$\underline{\perp}_{i=1,c} u_i = \left(\overset{1}{\underset{i=1,c}{\perp}} u_i\right) \top \left(\overline{\overset{2}{\underset{i=1,c}{\perp}} u_i \Big/ \overset{1}{\underset{i=1,c}{\perp}} u_i}\right) \tag{8}$$

where $$\overset{k}{\underset{i=1,c}{\perp}} u_i = \underset{A\in\mathcal{P}_{k-1}}{\top}\left(\underset{j\in C\setminus A}{\perp} u_j\right) \tag{9}$$

with $\mathcal{P}$ the powerset of $C = \{1, 2, ..., c\}$ and $\mathcal{P}_k = \{A \in \mathcal{P} : |A| = k\}$ where $|A|$ denotes the cardinality of subset A. Assusming u to be a sorted c-tuple, i.e. $u_1 \geq u_2 \cdots \geq u_c$, we defined in [12] an operator based on triangular norms and the Sugeno integral which quantifies the similarity of the block of values $\{u_j, \ldots, u_k\}$:

$$\Phi_{j,k}(u) = \begin{cases} \dfrac{\displaystyle\bot_{i=\frac{k+j}{2}}^{k} u_i \top \mathcal{K}_\lambda(i,k)}{\displaystyle\bot_{i=\frac{k+j}{2}}^{j} u_i \top \mathcal{K}_\lambda(i,j)} & \text{if } k-j \text{ is even} \\ \dfrac{\displaystyle\bot_{i=\frac{k+j+1}{2}}^{k} u_i \top \mathcal{K}_\lambda(i,k)}{\displaystyle\bot_{i=\frac{k+j-1}{2}}^{j} u_i \top \mathcal{K}_\lambda(i,j)} & \text{if } k-j \text{ is odd} \end{cases} \tag{10}$$

where $\mathcal{N}_\lambda(i,l)$ is a gaussian kernel defined by:

$$\mathcal{N}_\lambda(i,l) = \exp \frac{-(i-l)^2}{\lambda} \tag{11}$$

The new fuzzy 3-rules classifier with reject options we propose derive as follows:

- rule $R1$: if u_1 is high XOR u_2 is high XOR $\cdots$ XOR u_c is high, then $x \mapsto \omega_{\text{argmax}_j(u_j)}$
- rule $R2$: if $\Phi_{1,2}$ is high OR $\cdots$ OR $\Phi_{1,c}$ is high, then reject x for ambiguity
- rule $R3$: if u_1 is low AND u_2 is low AND $\cdots$ AND u_c is low, then reject x for distance.

We define the firing strengths of the 3 rules by:

$$\tau_1(x) = \bot_{i=1,c} H(u_i) = \bot_{i=1,c} u_i \tag{12}$$

$$\tau_2(x) = \bot_{i=2,c} H(\Phi_{1,i}(u)) = \bot_{i=2,c} \Phi_{1,i}(u) \tag{13}$$

$$\tau_3(x) = \top_{i=1,c} L(u_i) = \top_{i=1,c} \overline{u_i} \tag{14}$$

Finally, a simple *winner takes all* strategy applied to the triplet $\{\tau_1(x), \tau_2(x), \tau_3(x)\}$ activates the corresponding rule which gives the classification result. It is worthnoting that the proposed classifier does not involve any threshold, compared to most of other classifiers with reject options. Depending on his interest, the user can select only the reject rules, and choose a conjunction, for instance a triangular norm, for the aggregation process of these firing strengths. In this case, a threshold can be applied on $\top_{i=1,3} \tau_i(x)$. By changing the characteristics, various discrimination procedures are possible, see section 5 for examples.

Examples of label vectors $u(x)$ for a $c = 3$ classes problem and resulting firing strengths are given in Table 2 for different dual couples: standard (S), algebraic (A), and the parameterized Hamacher family (H_γ) defined by:

$$\top_{H,\gamma}(a,b) = \frac{a\,b}{\gamma + (1-\gamma)\,(a+b-a\,b)}$$
$$\bot_{H,\gamma}(a,b) = \frac{a+b+(\gamma-2)\,a\,b}{1-(1-\gamma)\,a\,b}$$

which reducts to $(\top_A, \bot_A)$ when $\gamma = 1$.

The given label vectors are representative of various situations: ambiguity between three and two classes, exclusive classification, and distance rejection, respectively. Whatever the dual couple, the Winning Firing Strength (*WFS*) gives the correct classification result. However, $\tau_2(x)$ does not exhibit selective ambiguity rejection, e.g. ambiguity between three or two classes in the Table. This refinment, corresponding to the k-order ambiguity concept, can obviously be obtained by indexing $\tau_2(x)$ by k instead of c in (13).

Table 2: Examples of firing strengths

$u(x)$		$\tau_1(x)$	$\tau_2(x)$	$\tau_3(x)$
$(0.85, 0.90, 0.75)^T$	S :	0.05	0.94	0.10
	A :	0.07	0.99	0.00
	H_0 :	0.16	0.95	0.05
$(0.85, 0.90, 0.10)^T$	S :	0.05	0.94	0.1
	A :	0.21	0.96	0.01
	H_0 :	0.20	0.94	0.06
$(0.15, 0.90, 0.10)^T$	S :	0.83	0.16	0.10
	A :	0.72	0.30	0.07
	H_0 :	0.70	0.32	0.09
$(0.15, 0.00, 0.10)^T$	S :	0.15	0.66	0.85
	A :	0.19	0.40	0.80
	H_0 :	0.17	0.45	0.81

5 Experimental Results

5.1 Artificial Data

This example aims at showing the classfication boundaries that can result from the triplet of firing strengths $\{\tau_1(x), \tau_2(x), \tau_3(x)\}$ using a particular dual couple: $(\top_A, \perp_A)$, but others give similar results. The data set consists of four classes composed of sixty points in $\mathbb{R}^2$ each. Classes slightly overlap in such a way that k-order ambiguity areas can appear. Membership degrees in the feature space are computed by equation (1) with $\alpha = 3$ and

$$d^2(x, p_i) = (x - p_i)^T \Sigma_i^{-1} (x - p_i)$$

where the mean vector p_i and covariance matrix Σ_i of each class are estimated from the data set.

First, we choose to define the 3-RCRO with no distance reject option. Thus, it simply consists in ambiguity rejecting patterns for which $\top_{i=1,3} \tau_i(x) = \tau_1(x)\,\tau_2(x)\,\tau_3(x)$ is higher than a threshold t, else exclusively assigning them to the class of maximum degree. Fig. 2 shows the contour plot of $\top_{i=1,3} \tau_i(x)$ varying from high values (white) to low ones (black).

The second result we give is obtained by allowing the three rules to be activated by the *WFS* classifier. As mentioned in the previous section, no threshold is needed. Classification boundaries are shown in Fig. 3, where white, grey ands black areas correspond to distance rejection, ambiguity rejection and exclusive classification respectively. Such a classification procedure results in a a very low error rate (nearly zero) and a high reject rate, therefore is more suitable for applications where the cost of misclassification is very high. Remind that other combinations of part of all the firing strengths leading to a dedicated 3-RCRO could allow the user to tune the different rates, but a threshold should be used.

5.2 Real Data Sets

In this final subsection, we present some results that show the classification performance of the 3-RCRO on well-known real data. The iris data set [3] contains $n = 150$ observations from $c = 3$ four-dimensional classes (iris species) of 50 points each. It is one of the most used benchmarks in pattern recognition, especially for cluster validity because two classes, numbered 2 and 3, present a substantial overlap in the feature space the class 1 is well separated from the others. For visualization purpose, only the third and fourth

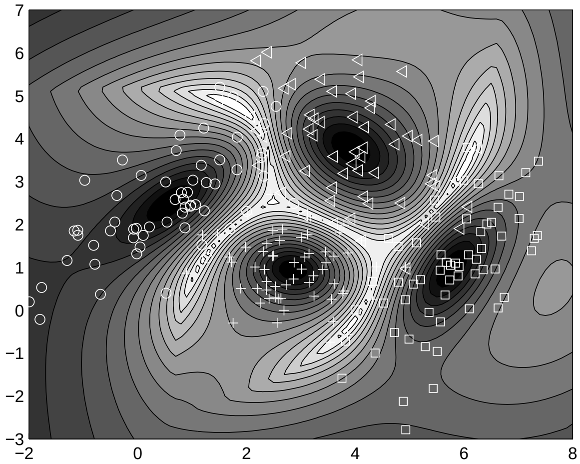

Figure 2: Contour plot of $\tau_1(x) \top \tau_2(x) \top \tau_3(x)$ with algebraic norms $(\top_A, \perp_A)$ – artificial data

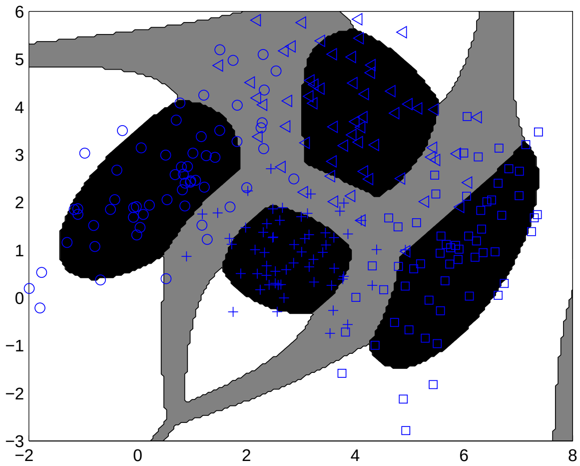

Figure 3: Classification areas using the winning firing strength – artificial data.

features were considered, as many authors do [2]. The membership degrees were computed using the same equations than for the artificial data.

Contour plot of $\top_{i=1,3} \tau_i(x)$ using $(\top_A, \perp_A)$ is shown in Fig. 4 varying from high values (white) to low ones (black). We obtained similar results with other dual couples. Table 3 shows the error rates obtained with classical classifiers using a resubstitution procedure: Quadratic Bayes (*QB*), Nearest Neighbor (*NN*) and the Max Classifier (*MC*, see section 3.1).

The performance of the *WFS* classifier and the 3-RCRO for different values of the threshold on $\top_{i=1,3} \tau_i(x)$ using $(\top_A, \perp_A)$ are reported in Table

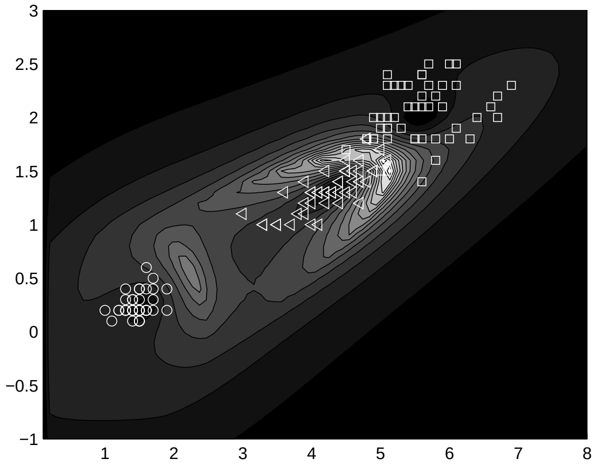

Figure 4: Contour plot of $\tau_1(x)\top\tau_2(x)\top\tau_3(x)$ with algebraic norms $(\top_A, \perp_A)$ – iris data

Table 3: Error rates without reject options – iris data

%	*QB*	*NN*	*MC*
error	2.00	4.67	2.67

4. Depending on the different strategies the user has in mind, the threshold can be set in order to minimize the error rate (e.g. $t = 0.08$), maximize the correct classification rate with a low error rate (e.g. $t = 0.15$), or maximize the correct classification rate with no reject options (e.g. $t = 1$) – this setting making the 3-RCRO coincide with the *MC* classifier as one could expect. Obviously, the reject rate decreases as t increases and its tuning is a keypoint as for every classifier involving a threshold. Note that rejected patterns were all rejected for ambiguity with $t = 0.15$ and not surprisingly some patterns were distance rejected with $t = 0.08$.

Table 4: Performance rates – iris data

%	*WFS*	$t = 0.08$	$t = 0.15$	$t = 1$
error	0.00	0.00	1.33	2.67
reject	25.33	12.00	2.00	0.00
correct	74.67	88.00	96.67	97.33

The second data set is the Forest Cover Type obtained from the UCI repository [3]. This is a very large GIS data set representing the forest cover type of a region, which contains $n = 581,012$ observations and $c = 7$ classes described by 54 attributes. Following [11], we only consider the $p = 10$ numeric valued attributes and the 495,141 points, belonging to classes 1 and 2. These two classes are equally distributed and have a significant overlap in the feature space. Error rates with usual supervised classifiers are shown in Table 5 and performance results of the proposed method are reported in Table 6. Compared to classical classifiers such as QB and NN without reject options, the proposed method reduces the error rate.

Table 5: Error rates without reject options – forest data

%	*QB*	*NN*	*MC*
error	25.56	29.83	24.91

Table 6: Performance rates – forest data

%	*WFS*	$t = 0.08$	$t = 0.15$	$t = 1$
error	9.58	3.18	19.14	24.91
reject	34.13	75.99	13.91	0.00
correct	56.29	20.82	66.94	75.09

6 Conclusion

In this paper, we propose a new approach to the classication problem including reject options. It is based on three fuzzy rules whose inputs are the membership degrees of objects to the classes at hand instead of features describing them and aggregation operators. These operators, based on combination of triangular norms and the Sugeno integral, are especially designed to give one of the three possible results. According to the firing strength of each rule, the pattern is either classified in a single class, or ambiguity rejected between several classes or distance rejected from all the classes. Two solutions have been proposed for the different decision boundaries. The first one consists in taking the most satisfied rule (*WFS*) and the second one consists in thresholding the conjonction of the three firing strengths (*3-RCRO*) or only part of them. Other combinations are under investigation. Experimental results show that the proposed approach is able to detect patterns that must be rejected and therefore gives satisfactory decison boundaries. We also

proposed to extend the approach to k-order ambiguity rejection by indexing the second rule in a proper way.
Among the future works we have in mind, let us cite the use of compensative and compensatory operators via combination of different couples for pattern classifiers with reject options, the use of uninorms instead of triangular norms couples as universal approximation [17] of fuzzy systems in this context.

References

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

[2] J. Bezdek, J. Keller, R. Krishnapuram, and N. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic, 1999.

[3] C. Blake and C. Merz. Uci repository of machine learning databases, 1998. `http://www.ics.uci.edu/mlearn/MLRepository.html`.

[4] B. Dubuisson and M. Masson. A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition*, 26(11):155–165, 1993.

[5] C. Frélicot and L. Mascarilla. A third way to design pattern classifiers with reject options. In *21th International Conference of the North American Fuzzy Information Processing Society*, 2002.

[6] C. Frélicot, L. Mascarilla, and M. Berthier. A new cluster validity index for fuzzy clustering based on combination of dual triples. In *IEEE International Conference on Fuzzy Systems*, 2006.

[7] F. Klawonn and E. P. Klement. Mathematical analysis of fuzzy classifiers. *Lecture Notes on Computer Sciences*, 1280:359–370, 1997.

[8] E.P Klement and R. Mesiar. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.

[9] L. I. Kuncheva. How good are fuzzy if-then classifiers? *IEEE Transactions on Systems, Man and Cybernetics*, 30(4):501–509, 2000.

[10] K. Menger. Statistical metrics. *Proc. National Academy of Science USA*, 28:535–537, 1942.

[11] P. Mitra, C. A. Murthy, and S.K. Pal. A probabilistic active support vector learning algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):413–418, 2004.

[12] H. Le Capitaine, T. Batard, C. Frélicot, and M. Berthier. Blockwise similarity in [0,1] via triangular norms and sugeno integrals - application to cluster validity. In *IEEE International Conference on Fuzzy Systems*, pages 835–840, London, UK, 2007.

[13] W. Silvert. Symmetric summation: a class of operations on fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 9(10):659–667, 1979.

[14] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[15] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[16] R. R. Yager. Ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.

[17] R. R. Yager and V. Kreinovich. Universal approximation theorem for uninorm-based fuzzy systems modeling. *Fuzzy Sets and Systems*, 140(2):331–339, 2003.

[18] R. R. Yager and A. Rybalov. Full reinforcement operators in aggregation techniques. *IEEE Transactions on Systems, Man and Cybernetics*, 28(6):757–769, 1998.

[19] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4(1):37–51, 1980.