



HAL
open science

A k-order fuzzy OR operator for pattern classification with k-order ambiguity rejection

Laurent Mascarilla, Michel Berthier, Carl Frelicot

► To cite this version:

Laurent Mascarilla, Michel Berthier, Carl Frelicot. A k-order fuzzy OR operator for pattern classification with k-order ambiguity rejection. *Fuzzy Sets and Systems*, 2008, 159 (15), pp.2011-2029. 10.1016/j.fss.2008.02.019 . hal-00289367

HAL Id: hal-00289367

<https://hal.science/hal-00289367>

Submitted on 12 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A K -order Fuzzy OR Operator for Pattern Classification with K -order Ambiguity Rejection

Laurent Mascarilla, Michel Berthier, Carl Frélicot

Lab. Mathématiques, Images et Applications - Université de La Rochelle, Avenue Michel Crépeau, 17042 La Rochelle Cedex 1, France {lmascari,mberthie,cfrelico}@univ-lr.fr

Abstract

In pattern recognition, the membership of an object to classes is often measured by labels. This article mainly deals with the mathematical foundations of labels combination operators, built on t -norms, that extend previous ambiguity measures of objects by dealing not only with 2 classes ambiguities but also with k classes, k lying between 1 and the number of classes c . Mathematical properties of this family of combination operators are established and a weighted extension is proposed, allowing to give more or less importance to a given class. A classifier with reject options built on the proposed measure is presented and applied on synthetic data. A critical analysis of the results led to derivate some new operators by aggregating previous measures. A modified classifier is proposed and applied to synthetic data as well as to standard real data.

Key words: triangular norms, ambiguity measure, classification with reject options

1. Introduction

The aim of this article is to define an ambiguity measure within the framework of pattern recognition, and more specifically supervised and unsupervised classification. For this purpose, we propose a new operator based on triangular norms. This family of operators essentially soften the concept of k -th highest component of a vector. We provide here well established mathematical definitions, check and prove properties of these operators. Some synthetic and real examples are exhibited.

The core of this paper is divided into three parts. In the first part we recall some basic properties of aggregation functions as well as previous works on ambiguity measurement. In the second part we define a class of operators based on triangular norms that combine labels, which are also called membership functions $\mu(\mathbf{x})$: the k -order fuzzy OR, here denoted $fOR-k$ for the sake of brevity, which is suitable for k -order ambiguity measurement (*i.e.* ambiguity between k classes). Classical properties of such aggregation operators are checked, others are proved. Weighted operators based on weighted t -norms are proposed as well. In the third part, we define a generic way for adding reject options, specially ambiguity rejection, to usual supervised classifiers. The resulting (generic) classifier is said to follow a "mixture first" strategy [9,17] based on this new ambiguity measure.

Various experiments on synthetic data sets show the good properties of the approach but reveal some drawbacks related to thresholds setting. This lead us to define other operators on top of the k -order fuzzy

OR and consequently a new classification algorithm. Their application to classical real data sets is given with quite satisfactory results.

It worth noting that, although we do not develop here other problematic than rejection in classification application, we have done such a work in case of $k = 2$ (fOR-2 operator). For example, we have investigated *variable selection* in various image processing and pattern recognition frameworks: see [18] where it was associated to a Sequential Forward Floating Search (SFFS) algorithm and [3] for a real-time dynamic color space selection for improving non-rigid objects tracking based on a Particle Filter approach. This special case also has been used to define a new *cluster validity measure* for fuzzy clustering [11]. All these applications have provided good performance results either on synthetic data or standard (UCI) data. This paper is focused on classification with reject options using the generalization to higher orders (fOR- k).

2. Background

2.1. Basic operators

An aggregation function is a mapping $\Phi : I^n \rightarrow \mathfrak{R}$, where I is a real interval. In many fields, e.g. decision making, values to be aggregated often are in $[0, 1]$ and consequent families $\Phi_{[0,1]} : [0, 1]^n \rightarrow [0, 1]$ have been defined. Among them, triangular norms (briefly t-norms) and conorms (briefly t-conorms) play an important role in fuzzy logic as multi-valued extensions of crisp sets' intersection \cap and union \cup operators, as well as boolean logic AND, OR connectives. A t-norm is a mapping $\top : [0, 1]^2 \rightarrow [0, 1]$ satisfying the following four axioms: $\forall x, y, z \in [0, 1]$

$$x \top y = y \top x \tag{1}$$

$$y \leq z \Rightarrow x \top y \leq x \top z \tag{2}$$

$$x \top (y \top z) = (x \top y) \top z \tag{3}$$

$$x \top 1 = x \tag{4}$$

The dual t-conorm \perp is defined as:

$$x \perp y = \overline{\bar{x} \top \bar{y}} \tag{5}$$

One usually takes $\bar{x} = 1 - x$. A t-conorm then satisfies axioms (1), (2), (3) as well as:

$$x \perp 0 = x \tag{6}$$

Axioms (2), (4) and (6) imply:

$$x \top y \leq x \tag{7}$$

$$x \leq x \perp y \tag{8}$$

It ensues:

- i) 0 is an absorbing element for any t-norm,
- ii) *min* is the largest t-norm,
- iii) 1 is an absorbing element for any t-conorm,
- iv) *max* is the smallest t-conorm.

Four basic t-norms and t-conorms are reported in Table 1 but there exist many others, in particular parametrized families, here exemplified by the Hamacher's one (see [16] and [15]).

2.2. Previous work

Suppose that values to be aggregated are labels $\mu_i = \mu_i(\mathbf{x}), \in [0, 1], i = 1, c$ and assume for writing convenience that they are sorted in decreasing order ($\mu_1 \geq \dots \geq \mu_c$). In a context of decision-making,

Table 1
Basic t-norms (\top) and t-conorms (\perp)

Standard	$\top_S = \min(x, y)$
	$\perp_S = \max(x, y)$
Algebraic	$\top_A = xy$
	$\perp_A = x + y - xy$
Łukasiewicz	$\top_L = \max(0, x + y - 1)$
	$\perp_L = \min(x + y, 1)$
Drastic	$\top_D = 0$ if $(x, y) \in [0, 1[^2$, $\min(x, y)$ otherwise
	$\perp_D = 1$ if $(x, y) \in]0, 1]^2$, $\max(x, y)$ otherwise
Hamacher $0 \leq \gamma \leq +\infty$	$\top_{H,\gamma} = \frac{xy}{\gamma + (1-\gamma)(x+y-xy)}$
	$\perp_{H,\gamma} = \frac{x+y-xy-(1-\gamma)xy}{1-(1-\gamma)xy}$

ambiguity straightforward relates to the comparison of μ_1 and μ_2 . In [13], we proposed an operator, based on the combination of a dual couple (\top, \perp), suitable for such a comparison. It was defined by:

$$\bigoplus_{i=1,c}^2 \mu_i = \bigtop_{i=1,c} \left(\bigoplus_{j \neq i} \mu_j \right) \quad (9)$$

This so-called *Fuzzy OR-2* operator (briefly *fOR-2*) have some mathematical properties resulting from those of \top and \perp , specifically:

- i) boundary conditions,
- ii) monotony,
- iii) symmetry,
- iv) idempotency of \top ¹,
- v) continuity with respect to each operand if \top is continuous.

Moreover, it has been shown that:

- if (\top_S, \perp_S) are taken, then

$$\bigoplus_{i=1,c}^2 \mu_i = \mu_2 \quad (10)$$

- whatever $c \geq 2$,

$$\bigtop_{i=1,c} \mu_i \leq \bigoplus_{i=1,c}^2 \mu_i \leq \bigoplus_{i=1,c} \mu_i \quad (11)$$

Referring to the compensation property of some aggregation functions, generically called Φ :

$$\min(x_1, \dots, x_c) \leq \Phi(x_1, \dots, x_c) \leq \max(x_1, \dots, x_c)$$

we called (11) the *weak compensation* property.

3. The k-order Fuzzy OR Operator

The natural extension of what we can call the *2-order ambiguity* concept leads to the definition of a *k-order fOR* operator. Let \mathcal{P} be the powerset of $C = \{1, \dots, c\}$ and $\mathcal{P}_k = \{A \in \mathcal{P} : |A| = k\}$ where $|A|$ denotes the cardinality of subset A . We propose to define the *k-order fuzzy OR* operator (briefly *fOR-k*) as:

$$\bigoplus_{i=1,c}^k \mu_i = \bigtop_{A \in \mathcal{P}_{k-1}} \left(\bigoplus_{j \in C \setminus A} \mu_j \right) \quad (12)$$

¹ \top_S and \perp_S are the only idempotent t-norm and t-conorm

Obviously, this definition agrees with (9) for $k = 2$. Let us review the mathematical properties fOR- k checks. Some of them result from those of \top and \perp , proofs being straightforward:

boundary conditions:

$$\bigperp_{i=1,c}^k 0 = 0 \quad \text{and} \quad \bigperp_{i=1,c}^k 1 = 1 \quad (13)$$

monotony: if $\lambda_i \leq \mu_i, \forall i = 1, c$, then:

$$\bigperp_{i=1,c}^k \lambda_i \leq \bigperp_{i=1,c}^k \mu_i \quad (14)$$

symmetry: for any permutation σ of C

$$\bigperp_{i=1,c}^k \mu_{\sigma(i)} = \bigperp_{i=1,c}^k \mu_i \quad (15)$$

Moreover, the particular role of 0 can be emphasized:

$$\bigperp^k(\mu_1, \dots, \mu_{c-1}, 0) = \left[\bigtop_{\substack{A \in \mathcal{P}_{k-1} \\ c \in A}} \left(\bigperp_{j \in C \setminus A} \mu_j \right) \right] \top \left[\bigtop_{\substack{A \in \mathcal{P}_{k-1} \\ c \notin A}} \left(\bigperp_{j \in C \setminus A} \mu_j \right) \right] \quad (16)$$

and if we denote $C' = \{1, \dots, c-1\}$ and \mathcal{P}'_k the set of subsets of C' with k elements:

$$\bigperp^k(\mu_1, \dots, \mu_{c-1}, 0) = \left[\bigtop_{B \in \mathcal{P}'_{k-2}} \left(\bigperp_{\substack{j \in C' \\ (B \cup \{c\})}} \mu_j \right) \right] \top \left[\bigtop_{A \in \mathcal{P}'_{k-1}} \left(\bigperp_{j \in C' \setminus A} \mu_j \right) \right] \quad (17)$$

$$= \left[\bigtop_{B \in \mathcal{P}'_{k-2}} \left(\bigperp_{\substack{j \in C' \\ B}} \mu_j \right) \right] \top \left[\bigperp^k(\mu_1, \dots, \mu_{c-1}) \right] \quad (18)$$

$$= \left[\bigperp^{k-1}(\mu_1, \dots, \mu_{c-1}) \right] \top \left[\bigperp^k(\mu_1, \dots, \mu_{c-1}) \right] \quad (19)$$

Since fOR- k aggregates c ($= |C|$) values, a sequence $\{\bigperp^1, \bigperp^2, \dots, \bigperp^c\}$ of c different fOR- k can always be constructed. From (12), we can derive simpler forms for \bigperp^1 as:

$$\begin{aligned} \bigperp_{i=1,c}^1 \mu_i &= \bigtop_{A \in \mathcal{P}_0} \left(\bigperp_{j \in C \setminus A} \mu_j \right) \\ &= \bigtop_{\emptyset} \left(\bigperp_{j \in C} \mu_j \right) \\ &= \bigperp_{j \in C} \mu_j \end{aligned} \quad (20)$$

and \bigperp^c as:

$$\begin{aligned}
\bigoplus_{i=1,c}^c \mu_i &= \bigvee_{A \in \mathcal{P}_{c-1}} \left(\bigoplus_{j \in C \setminus A} \mu_j \right) \\
&= \bigvee_{\substack{C \setminus \{i\} \\ i=1, \dots, c}} \mu_i \\
&= \bigvee_{i \in C} \mu_i
\end{aligned} \tag{21}$$

Now, let us show what properties (10-11) become for the fOR- k operator.

Property 1: Let $\{\mu_1, \dots, \mu_c\}$ be a set of numbers in $[0, 1]$. If (\top_S, \perp_S) are taken, then $\bigoplus_{i=1,c}^k \mu_i = \mu_k$, the k -th highest value.

Proof:

We write $\bigoplus_{i=1,c}^k \mu_i$ defined by (12) as:

$$\bigoplus_{i=1,c}^k \mu_i = \left[\bigoplus_{j \in C \setminus \kappa} \mu_j \right] \top \left[\bigvee_{A \in \mathcal{P}_{k-1} \setminus \kappa} \left(\bigoplus_{j \in C \setminus A} \mu_j \right) \right]$$

where $\kappa = \{1, \dots, k-1\}$.

If $A \in \mathcal{P}_{k-1} \setminus \kappa$ there exists i_0 in $\kappa \cap (C \setminus A)$, therefore

$$\bigoplus_{j \in C \setminus A} \mu_j \geq \mu_{i_0} \geq \mu_k.$$

As

$$\bigoplus_{j \in C \setminus \kappa} \mu_j = \mu_k$$

and

$$\bigvee_{A \in \mathcal{P}_{k-1} \setminus \kappa} \left(\bigoplus_{j \in C \setminus A} \mu_j \right) \geq \mu_k$$

this concludes the proof.

Property 2: Let $b = \frac{c}{2} + 1$ if c is even and $b = \frac{c+3}{2}$ if c is odd. We have

$$\bigoplus_{i=1,c} \mu_i = \bigoplus_{i=1,c}^1 \mu_i \geq \bigoplus_{i=1,c}^2 \mu_i \geq \dots \geq \bigoplus_{i=1,c}^b \mu_i. \tag{22}$$

Proof:

Letting $2 \leq k \leq b$, we are going to prove that $\bigoplus_{i=1,c}^{k-1} \mu_i \geq \bigoplus_{i=1,c}^k \mu_i$. By definition (12), we write:

$$\bigoplus_{i=1,c}^{k-1} \mu_i = \bigvee_{A \in \mathcal{P}_{k-2}} \left(\bigoplus_{j \in C \setminus A} \mu_j \right)$$

$$\bigoplus_{i=1,c}^{k-1} \mu_i = \bigvee_{\{i_1, \dots, i_{c-k+2}\}} (\mu_{i_1} \perp \mu_{i_2} \perp \dots \perp \mu_{i_{c-k+2}})$$

and

$$\bigoplus_{i=1,c}^k \mu_i = \bigvee_{\{j_1, \dots, j_{c-k+1}\}} (\mu_{j_1} \perp \dots \perp \mu_{j_{c-k+1}})$$

We know that $\binom{c}{k-2} \leq \binom{c}{k-1}$. Using a partition of \mathcal{P} into symmetric chains [5], we construct an injection τ of \mathcal{P}_{c-k+2} into \mathcal{P}_{c-k+1} such that $\tau(A) \subset A$ for all A in \mathcal{P}_{c-k+2} . We write $\tau(\{i_1, \dots, i_{c-k+2}\}) = \{l_1, \dots, l_{c-k+1}\}$. From (8) and (2) we have:

$$\prod_{i=1,c}^{k-1} \mu_i \geq \prod_{\{l_1, \dots, l_{c-k+1}\}} (\mu_{l_1} \perp \dots \perp \mu_{l_{c-k+1}})$$

and from (7) we obtain

$$\prod_{\{l_1, \dots, l_{c-k+1}\}} (\mu_{l_1} \perp \dots \perp \mu_{l_{c-k+1}}) \geq \prod_{i=1,c}^k \mu_i$$

and that concludes the proof.

Keeping in mind (20-21), it is finally worthnoting that if (\top_S, \perp_S) are taken, (22) becomes:

$$\mu_1 = \prod_{i=1,c} \mu_i = \prod_{i=1,c}^1 \mu_i \geq \prod_{i=1,c}^2 \mu_i \geq \dots \geq \prod_{i=1,c}^c \mu_i = \prod_{i=1,c} \mu_i = \mu_c$$

In this particular case the weak compensation property holds not only up to a bound b for k .

3.1. Weighted operators

For some applications, it can be of interest to introduce unequal importance between the various label μ_i by means of associated weights w_i . This can be seen as a way to adjust the cost of misclassifications between classes (e.g. if the cost of misclassification on a class is high, a higher weight could be put on it). That for, one has to modify the t-norms (and t-conorms) definitions. In the general case, no unified approach is found in the literature. However, in the particular case of strict archimedean operators some possibilities are enumerated in [8]. For instance, in the case of Hamacher operators we can define:

$$\prod_{i=1,c} \{\mu_i, w_i\} = \frac{1}{1 + c \sum_{i=1,c} w_i \frac{1-\mu_i}{\mu_i}} \quad (23)$$

$$\prod_{i=1,c} \{\mu_i, w_i\} = \frac{1}{1 + \frac{1}{c \sum_{i=1,c} w_i \frac{\mu_i}{1-\mu_i}}} \quad (24)$$

with the constraint $\sum_{i=1,c} w_i = 1$. Of course, unweighted operators (see Table 1) are found in case of uniform weighting, $\forall i = 1, c, \{\mu_i, 1/c\}$. Standard t-norms (min, max) do not belong to this framework and the commonly used variants are [8]:

$$\prod_{i=1,c} \{\mu_i, w_i\} = \min_{i=1,c} (\max(1 - w_i, \mu_i)) \quad (25)$$

$$\prod_{i=1,c} \{\mu_i, w_i\} = \max_{i=1,c} (\min(w_i, \mu_i)) \quad (26)$$

with the constraint $\max_{j=i,c} w_i = 1$. In case of uniform weighting, $\forall i = 1, c, \{\mu_i, 1\}$, we find again the unweighted operators (1). However, these operators have to be combined to produce a fOR- k and no well established strategies to calculate the necessary weights are available. We propose to use the following weighting which ensures to find unweighted fOR- k when using uniform weights:

$$\prod_{i=1,c}^k \{\mu_i, w_i\} = \prod_{A \in \mathcal{P}_{k-1}} \left\{ \left(\prod_{j \in C \setminus A} \{\mu_j, w'_j\} \right), w''_A \right\} \quad (27)$$

where w'_j et w''_A are normalized weights (*i.e.* sum equal to 1) calculated from initial weights w_i :

$$w'_j = \frac{w_j}{\sum_{j \in C \setminus A} w_j} \quad (28)$$

$$w''_A = \frac{\sigma_A}{\sum_{A \in \mathcal{P}_{k-1}} w_A}, \text{ and } \sigma_A = \sum_{l \in A} w_l \quad (29)$$

One can see the effect of such weighting in Fig. 1 for a four classes two-dimensional synthetic example. It can be observed that the class ω_1 with highest weight provides greatest fOR-1, while other fOR-k, for this same class, are lowered compared to the unweighted case, then revealing less ambiguity. This is consistent because patterns belonging to the class with the highest weight will be more easily exclusively classified (due to the great values of fOR-1) and consequently less easily ambiguity rejected.

A special case is the one of the standard t-norm, for which the different normalization constraint leads to:

$$w'_j = \frac{w_j}{\max_{j \in C \setminus A} w_j} \quad (30)$$

$$w''_A = \frac{\pi_A}{\max_{A \in \mathcal{P}_{k-1}} w_A}, \text{ and } \pi_A = \max_{l \in A} w_l \quad (31)$$

4. K-order Ambiguity in Pattern Classification

4.1. Classifier design

Let \mathbf{x} be a pattern described by p features, and $\Omega = \{\omega_1, \dots, \omega_c\}$ be a set of c classes. The aim of classifier design is to define rules (*classifiers*) that can associate an unknown pattern vector \mathbf{x} from a feature space with one class of Ω . It generally consists in defining a couple (H, L) composed of a *labeling function* and a *hardening function*:

$$\begin{aligned} - L : \mathfrak{R}^p &\rightarrow L_{\bullet c}, \\ \mathbf{x} &\mapsto \mu(\mathbf{x}) = {}^t(\mu_1(\mathbf{x}), \dots, \mu_c(\mathbf{x})) \end{aligned}$$

depending on the mathematical framework the classifier relies on (probabilistic, fuzzy or possibilistic). If $\mu_i(\mathbf{x})$ is the *posterior* probability that \mathbf{x} belongs to ω_i or a membership function to a fuzzy set associated with ω_i , then $L_{\bullet c} = L_{fc} = \{\mu(\mathbf{x}) \in [0, 1]^c : \sum_{i=1}^c \mu_i(\mathbf{x}) = 1\}$. Otherwise, $\mu_i(\mathbf{x})$ measures the typicality of \mathbf{x} to class ω_i , e.g. by:

$$\mu_i(\mathbf{x}) = \frac{1}{1 + d(\mathbf{x}, \mathbf{m}_i)} \quad (32)$$

where \mathbf{m}_i is a class-prototype obtained from a learning set of patterns, therefore $L_{\bullet c} = L_{pc} = [0, 1]^c$. For illustration purposes this is the function of choice in the remaining of this paper. Let us notice that, in the following, we only use the $\mu_i(\mathbf{x})$ measures as an input for the hardening function, which is in charge for decision making (namely exclusive classification or rejection) and described in the next item. The intrinsic quality of the various $\mu_i(\mathbf{x})$ is beyond the scope of this article (see for example classical book like [10]).

$$\begin{aligned} - H : L_{\bullet c} &\rightarrow L_{hc} \\ \mu(\mathbf{x}) &\mapsto l(\mathbf{x}) = {}^t(l_1(\mathbf{x}), \dots, l_c(\mathbf{x})): \\ l_i(\mathbf{x}) &\in \{0, 1\} \text{ and } \sum_{i=1}^c l_i(\mathbf{x}) = 1. \end{aligned}$$

The common function reduces to the class of maximum label selection.

Such an exclusive classification rule is not efficient in practice because it supposes that:

- i) Ω is exhaustively defined (*closed-world* assumption),
- ii) classes do not overlap (*separability* assumption)

4.2. Reject options

In order to overcome the above mentioned limits and to reduce the misclassification risk, reject options can be used. Two kinds of rejection have been defined. The first one, called *distance rejection* [9] is dedicated to outlying patterns and allows to associate a vector \mathbf{x} to no class. The second one, called *ambiguity rejection*,

allows \mathbf{x} to be classified in several or all the classes [7,14] ; it deals with inlying patterns. Formally, including reject options consists in modifying the hardening function H such that $l(\mathbf{x})$ can take any value of the set of vertices of the unit hypercube $L_{hc}^c = \{0, 1\}^c$. Thus, given its label vector $\mu(\mathbf{x})$, an unknown pattern \mathbf{x} will be either:

- distance rejected when $l(\mathbf{x}) = 0$, or
- exclusively classified when $l(\mathbf{x}) \in L_{hc} \subset L_{hc}^c$, or
- ambiguity rejected between a subset of classes $\omega = \{\omega_j : l_j(\mathbf{x}) = 1\}$ when $l(\mathbf{x})$ is any of the other vertices.

It must be noticed that, in case the membership function is normalized (*i.e.* $\sum_{i=1}^c \mu_i(\mathbf{x}) = 1$), no distance rejection is possible. In the general case, the design of classifiers with reject options can be made according to well-identified strategies operating in two sequential steps (H_1, H_2) [12], where H_1 is a function focusing on one of the three possible classification results at hand. In particular, when H_1 deals with ambiguity rejection, the classifier is said to follow a "mixture-first" strategy.

4.3. The proposed generic classifier

The fOR- k (12) can be used to define a generic classifier following a modified "mixture-first" strategy as shown in Algorithm 1, where c_{max} is predefined, μ_k and t_k ($k = 1, c_{max}$) are user-specified thresholds. Note that c_{max} can be set such that the weak compensation property holds, *i.e.* $c_{max} \leq b$ in (22), but not necessarily. One must notice that we do not discuss the labeling function L here. The choice of the method to compute the membership values is left to the user choice (see *e.g.* [10] for some introduction) and is out of interest for now.

Algorithm 1: *hardening function* $H = (H_1, H_2) : L_{pc} \rightarrow L_{hc}^c, \mu(\mathbf{x}) \mapsto l(\mathbf{x})$

```

// H1 : ambiguity rejection ?
for  $k = c_{max}$  down to 2 do
     $\perp := \bigwedge_{i=1, c}^k \mu_i(\mathbf{x})$ 
    if  $\perp > t_k$  then
        select  $J = \{j \in C : \mu_j(\mathbf{x}) \geq \mu_k\}$ 
        let  $l(\mathbf{x})$  such that  $l_i(\mathbf{x}) = \begin{cases} 1 & \text{if } i \in J \\ 0 & \text{otherwise} \end{cases}$ 
        exit
// H2 :  $\mathbf{x}$  is not ambiguity rejected
if  $\perp > t_1$  then // exclusive classification
    let  $l(\mathbf{x})$  such that  $l_i(\mathbf{x}) = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases}$ 
else // distance rejection
     $l(\mathbf{x}) = 0$ 

```

Examples of label vectors $\mu(\mathbf{x}) \in [0, 1]^c$ for a $c = 6$ -classes problem and resulting classification vectors $l(x) \in \{0, 1\}^c$ are given in Table 2 for different dual couples: $S = (\top_S, \perp_S)$, $A = (\top_A, \perp_A)$ and the parametrized family due to Hamacher, with its γ parameter set to 0 $H = (\top_{H,0}, \perp_{H,0})$.

Thresholds were set to $t_1 = 0.7$, $t_2 = 0.5$, $t_3 = 0.3$, and $\mu_k = 0.7, k \in \{1, 2, 3\}$ in order to give the same classification results whatever (\top, \perp) . To allow all the reported values to appear properly in the table, we set $c_{max} = 3 \leq b = 4$, where b is the one of the property 2. That is to say we only report values for which the order is always preserved whatever the t-norm.

Table 2

Classification examples with t-norms Standard, Algebraic and Hamacher ($\gamma = 0$)

$\mu(\mathbf{x})$	(\top, \perp) and $\perp^k (k = 1, 3)$	$l(x)$	$\mu(\mathbf{x})$	(\top, \perp) and $\perp^k (k = 1, 3)$	$l(x)$
0.10		0	0.10		0
0.75	$\perp^1 \perp^2 \perp^3$	1	0.25	$\perp^1 \perp^2 \perp^3$	0
0.20	$S : 0.85 \ 0.80 \ 0.75$	0	0.20	$S : 0.85 \ 0.80 \ 0.25$	0
0.85	$A : 0.99 \ 0.91 \ 0.48$	1	0.85	$A : 0.98 \ 0.77 \ 0.17$	1
0.80	$H : 0.93 \ 0.63 \ 0.34$	1	0.80	$H : 0.91 \ 0.57 \ 0.22$	1
0.05		0	0.05		0
0.10		0	0.10		0
0.25	$\perp^1 \perp^2 \perp^3$	0	0.25	$\perp^1 \perp^2 \perp^3$	0
0.20	$S : 0.80 \ 0.25 \ 0.20$	0	0.20	$S : 0.25 \ 0.20 \ 0.15$	0
0.15	$A : 0.91 \ 0.33 \ 0.07$	0	0.15	$A : 0.56 \ 0.01 \ 0.00$	0
0.80	$H : 0.83 \ 0.32 \ 0.10$	1	0.00	$H : 0.48 \ 0.11 \ 0.04$	0
0.05		0	0.05		0

4.4. Classification areas of an artificial data set

The following example aims at showing that such a classifier based on the ambiguity measure \perp^k can result in defining well-adapted classification boundaries. A learning set of $c = 4$ classes of 25 artificial gaussian two-dimensional samples each was generated such that three classes slightly overlap while the fourth one is well-separated from the others:

- $\omega_1 \sim \mathcal{N}(\begin{smallmatrix} -1 & -1 \end{smallmatrix}, I)$,
- $\omega_2 \sim \mathcal{N}(\begin{smallmatrix} -1 & 1 \end{smallmatrix}, I)$,
- $\omega_3 \sim \mathcal{N}(\begin{smallmatrix} 1 & 1 \end{smallmatrix}, I)$ and
- $\omega_4 \sim \mathcal{N}(\begin{smallmatrix} -3 & 3 \end{smallmatrix}, I)$.

Contour plots of \perp^2, \perp^3 and \perp^4 using $S = (\top_S, \perp_S)$ are shown in Fig. 1 (left column). Areas are as dark as \perp^k is high. To simplify the choice of the thresholds t_k they were set according to a predetermined fraction of the actual maximum value of the corresponding \perp^k computed on the data samples. In this example, we took 80% of each of the \perp^2, \perp^3 and \perp^4 maximum values and 20% of the \perp^1 maximum values, such that classification areas were good looking. Corresponding classification areas are plotted in Fig. 2: exclusive classification in one of the four classes in blue, the 3 kind of ambiguities reject areas in yellow, orange and brown, and distance reject area in white.

Let us notice that this choice of thresholds reveals an ambiguity reject area between 4 classes inside an exclusive classification area (see Fig. 2). It is obvious that this area should be eliminated by a better thresholds choice. However, we chose such thresholds to visualize the relative positions of the reject areas and show their relevance. In a real application, the erroneous area would be easily removed because, in this case, the values of the \perp^4 being significantly low relatively to the others \perp^k .

A last example on these 2-dimensional data is made of the plotting of the fOR- k ($k = 1, 3$) based on Hamacher t-norms with $\gamma = 0$ (see Fig. 1, middle column). As one can observe, fOR-1 reveals peaks around class centers mostly in the same way as the first order operator based on standard column. More surprisingly, the higher order operators (3 and beyond) tend to put ambiguity area not only between class centers but also close to them. This is due to the way our operators are made: by aggregating the whole label vector values it takes into account not only values of rank k and above but also the highest value, related to fOR-1. Hence, fOR- k , $k \leq 2$ become greater as fOR-1 increases. To exemplify this tendency let us take

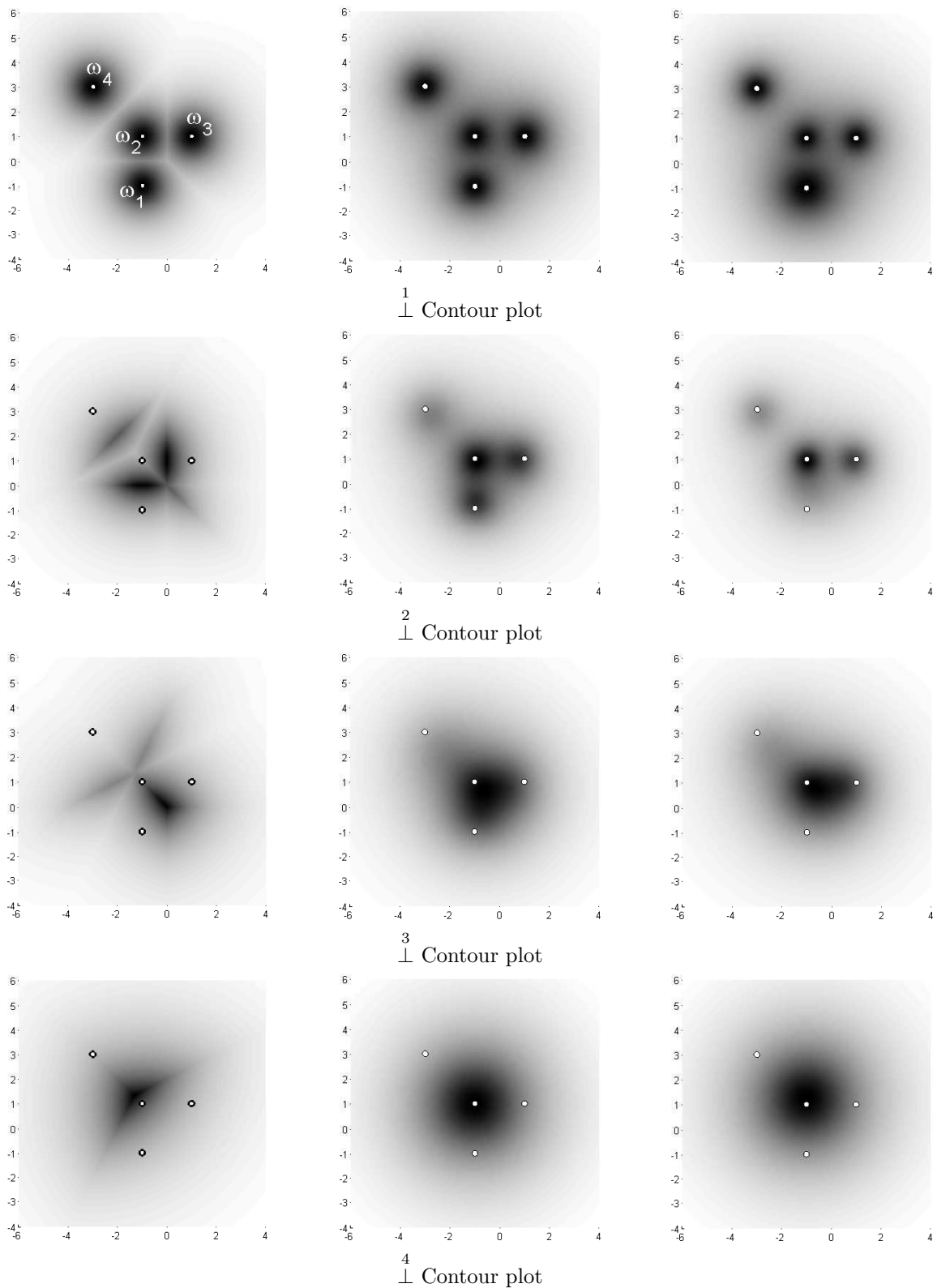


Fig. 1. Some $fOR-k$, with Euclidean distance based membership functions (Equation 32), on the data set: (\mathbb{T}_S, \perp_S) based (left column), $(\mathbb{T}_{H,0}, \perp_{H,0})$ based: unweighted (middle column) and weighted $w = {}^t \left(\frac{1}{2} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \right)$ (right column).

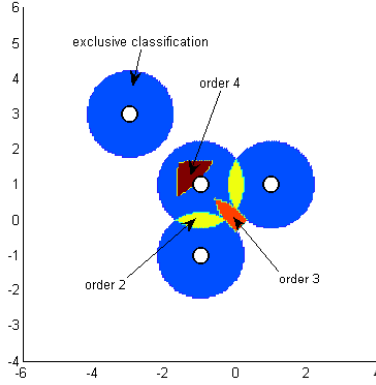


Fig. 2. Standard \perp^k classification result (Class centers are represented by white circles).

$\perp^2(0.5, 0.1, 0.01) = 0.1$. The result is numerically consistent in being close to the second value of the vector. However, by increasing the highest value of the vector $\perp^2(0.9, 0.1, 0.01) = 0.2$ one gets a greater value which is still numerically consistent but influenced by the greatest value. Thus, to some extent the second maximum value is attracted by the first one. More generally, \perp^1 attracts \perp^2, \dots, \perp^c toward class centers. A visual example can be given in taking a level curve along the line drawn in Fig. 3(a).

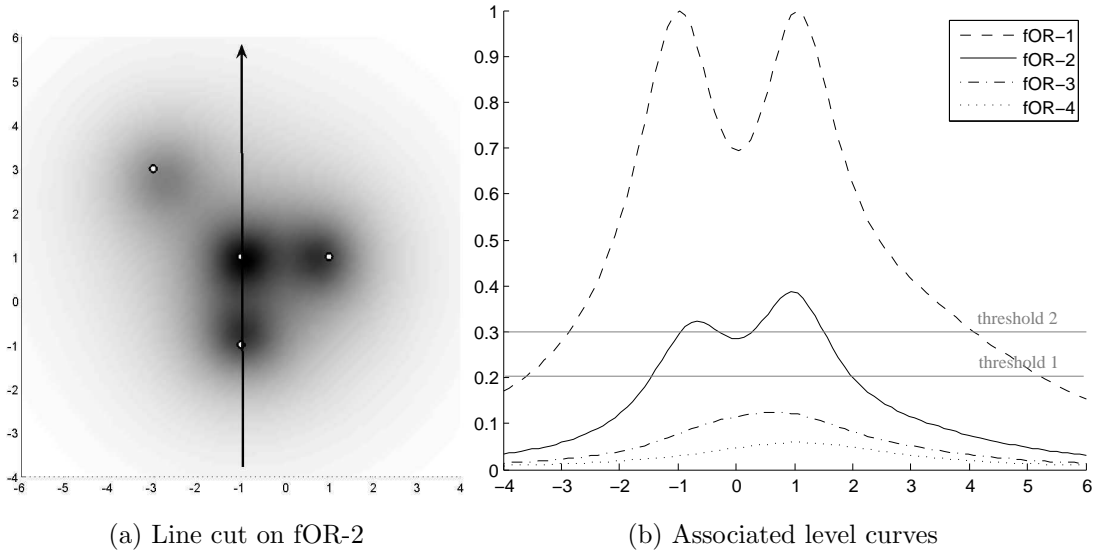


Fig. 3. Illustration of the difficulty of thresholds' setting with fOR- k based on Hamacher t-norm ($\gamma = 0$).

It can be checked, Fig. 3(b), that any threshold on fOR-2 will either place ambiguity areas around class centers, as with threshold 1 sets to 0.2, or completely ignore areas between classes, as with threshold 2 sets to 0.3. As such, this is not problematic but, from our point of view, which is thresholding fOR- k values to ensure classification or reject decisions, this behavior clearly introduces some bias toward class centers. We show in the next section some way to overcome this difficulty.

Finally, the last column of Fig. 1 exhibits weighting results on the fOR- k based on Hamacher norms ($\gamma = 0$). We choose to reinforce the class ω_1 by increasing its associated weight: $w = {}^t(\frac{1}{2} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6})$.

One can observe that the \perp^1 of class ω_1 is favored compared to the one of the other classes: its maximal value is higher than the other ones and its spread is wider. Patterns belonging to ω_1 will clearly be more easily exclusively classified. On the contrary, for orders higher than two, ω_1 is to some extent isolated: it

has less ambiguity than the others. This behavior is consistent as this class is more easily classified and shares less ambiguity with the others. As a consequence, ambiguity areas move toward the other classes centers. The next section presents the method we develop to ensure good reject areas even in the case of strict Archimedean t-norm, e.g. the Hamacher one.

4.5. Practical use: reducing from k thresholds to a single one.

The \perp^k operator we have defined exhibits well established properties. In particular, they are sorted in decreasing order from \perp^2 to \perp^b (see Equation (22)). From a practical point of view, reject areas determination is made complicated:

- generally, by the *number of thresholds* to be tuned (potentially as many as the classes number for ambiguity rejection plus a threshold for distance rejection),
- specifically, for some t-norms, like Hamacher’s one, by:

- *thresholds dynamic*: values can be small while difference between the various \perp^k is non linear,
- *reject areas are peaked around class centers* as shown in previous examples.

This is why we aim at providing a simple way of thresholding by use of a single threshold instead of the k thresholds values that would be used when directly using the fOR- k .

To counter the problems presented in the start of this section we chose to define operators based on the fOR- k with an approach close to the one followed for the definition of the fXOR in [17]: we will combine our operators to define more relevant reject areas that also are easier to obtain by a single thresholding. Let us recall that a *fuzzy eXclusive OR operator* (briefly *fXOR*) extends the crisp *XOR* operator to the fuzzy context. Such an operator has been shown to be suitable for revealing if μ_1 is significantly high and μ_2 is significantly low while remaining greater than lower degrees. It is a mapping from $[0, 1]^c$ to $[0, 1]$, defined by:

$$\perp_{i=1,c} \mu_i = \left(\perp_{i=1,c} \mu_i \right) \top \left(\overline{\perp_{i=1,c} \mu_i / \perp_{i=1,c} \mu_i} \right) \quad (33)$$

There exist other combinations of dual triples $(\top, \perp, \bar{\cdot})$, where $\bar{\cdot}$ denotes a fuzzy complement, that are fXOR operators, e.g. in [17]. The term at the right hand side of \top in (33) acts as a penalty one which weakens the fOR operator. Thanks to the monotony property and neutral elements of \top and \perp , we have whatever (\top, \perp) :

$$\top_{i=1,c} \mu_i \leq \perp_{i=1,c} \mu_i \leq \perp_{i=1,c} \mu_i \quad (34)$$

For instance, when the standard norms (\top_S, \perp_S) and the usual complement: $\bar{u} = 1 - u$, are used, fXOR reduces to $\min\{\mu_1, 1 - \mu_2/\mu_1\}$ and equals 0 whenever $\mu_1 = \mu_2$. This fXOR applied to unsupervised classification for the cluster validity problem, favorably compares to well established criteria (see [11] for a deeper inspection). The first stage of the procedure we propose consists in defining an operator a_k which determines if a pattern \mathbf{x}_j to be classified has a label vector $\mu((x))$ which is ambiguous for the order k and for this order only. This is done by aggregating by one fuzzy AND two terms which fulfill these roles respectively: the first one, which is the ratio of fOR- k on fOR-1, determines if fOR- k is large relatively to the fOR-1 ; the second one, by the negation of a_{k+1} removes ambiguity with higher orders. Let us notice that this last term makes it possible to concentrate the ambiguity areas on a fixed order. In particular it avoids a mixture between, on the one hand, areas related to ambiguity of order 2 and higher, and on the other hand, areas of exclusive classification as characterized by a high value of fOR-1. These terms are built in an iterative way according to the following scheme:

for all $\mathbf{x} \in \mathcal{R}^p$ to be classified

$$\left\{ \begin{array}{l} a_{c+1}(\mathbf{x}) = 0 \\ a_c(\mathbf{x}) = \left(\frac{c}{\perp \mu(\mathbf{x}) / \perp \mu(\mathbf{x})} \right) \top \overline{a_{c+1}(\mathbf{x})} = \left(\frac{c}{\perp \mu(\mathbf{x}) / \perp \mu(\mathbf{x})} \right) \\ \vdots \\ a_k(\mathbf{x}) = \left(\frac{k}{\perp \mu(\mathbf{x}) / \perp \mu(\mathbf{x})} \right) \top \overline{a_{k+1}(\mathbf{x})} \\ \vdots \\ a_2(\mathbf{x}) = \left(\frac{2}{\perp \mu(\mathbf{x}) / \perp \mu(\mathbf{x})} \right) \top \overline{a_3(\mathbf{x})} \end{array} \right.$$

The second stage in the construction of our measure consists in aggregating by one fuzzy OR the various a_k to find a measure which characterizes the ambiguities of order k or higher. Here, our aim is to be able to fix a possibly unique global threshold to isolate all the possible ambiguity areas:

$$\left\{ \begin{array}{l} \alpha'_c(\mathbf{x}) = a_c(\mathbf{x}) \\ \vdots \\ \alpha'_k(\mathbf{x}) = \frac{\perp}{i=k,c} a_i(\mathbf{x}) \\ \vdots \\ \alpha'_2(\mathbf{x}) = a_2(\mathbf{x}) \perp a_3(\mathbf{x}) \dots \perp a_c(\mathbf{x}) \end{array} \right.$$

Let us notice that, by (8), we have $\forall k, \alpha'_k \leq \alpha'_{k-1}$.

The last stage is a kind of normalization relatively to order 2. Indeed, thanks to the previous property, the α'_k are stacked in descending order, and thus a possible strategy to fix reject thresholds is to give a value to the threshold on the fOR-2 and decreasing values for the higher orders. Even if it provides an operational framework, there remain nevertheless $c - 1$ thresholds to be fixed to reveal all ambiguity areas. In order to reduce this choice to a single threshold, we propose the following heuristics, amongst many other candidates: the various fOR- k , $k \geq 2$ are translated so that the highest computed values of these fOR- k touch fOR-2. More formally:

let $j_{max} = \operatorname{argmax}_{j=1,n} (\alpha'_2(\mathbf{x}_j))$, where the \mathbf{x}_j 's are the n patterns to be classified,

$$\left\{ \begin{array}{l} \alpha_k(\mathbf{x}) = \alpha'_k(\mathbf{x}) + (\alpha'_2(\mathbf{x}_{j_{max}}) - \alpha'_k(\mathbf{x}_{j_{max}})), \\ \vdots \\ \alpha_2(\mathbf{x}) = \alpha'_2(\mathbf{x}_j) \end{array} \right.$$

Then, the practical use of α_k is the following one: as α_k are made comparable by normalization, a single threshold on α_2 is enough to isolate ambiguity areas of order 2 or higher. Then, with the same threshold on α_3 one can isolate the ambiguity areas of order 3 or higher and so on until order k . The corresponding classifier with reject options, according to the "mixture-first" strategy, is given in Algorithm 2. Only two threshold t_1 , for distance rejection, and t_2 for ambiguity rejection, of any order, is required. As for the previous Algorithm 1, it can reduce to an exclusive classifier (with not reject option) by an appropriate choice of thresholds, e.g. here $t_1 = t_2 = 1$.

By taking again the previous four classes example, one can visually check in Fig. 4 that areas associated to ambiguities are now correctly located between classes and no more toward centers.

It can be checked on Fig. 5 that the ambiguity area is correctly placed between classes and no more around class centers as in Fig. 3.

Algorithm 2: *hardening function* $H = (H_1, H_2) : L_{pc} \rightarrow L_{hc}^c, \mu(\mathbf{x}) \mapsto l(\mathbf{x})$

```

// H1 : ambiguity rejection ?
for k = 2 to cmax do
  αk := αk(x)
  if αk > t2 then
    select J = {j ∈ C : μj(x) ≥ μk}
    let l(x) such that li(x) =  $\begin{cases} 1 & \text{if } i \in J \\ 0 & \text{otherwise} \end{cases}$ 
    exit
// H2 : x is not ambiguity rejected
if  $\frac{1}{\perp} > t_1$  then // exclusive classification
  let l(x) such that li(x) =  $\begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases}$ 
else // distance rejection
  l(x) = 0

```

4.6. Real cases study

To validate the proposed operator and classification procedure we made some experiments on two real data sets. They are both taken from the well known UCI database [4]. The first one is the three classes Iris data set and the second one the six classes Glass Identification data set.

4.6.1. Iris data set

The iris data set [1] contains $n = 150$ observations from three 4-dimensional classes (*iris virginica*, *iris versicolor* and *iris setosa*) of 50 points each. It is one of the most used benchmarks in pattern recognition, and is known to have two classes, numbered 2 and 3, with a substantial overlap in the feature space while the first one is well separated from the others. Therefore, one can expect to mostly find ambiguity of order 2 in some area and very few, if any, ambiguity of higher order. For visualization purposes, and following [2], only the third and fourth two features (*petal width* and *petal length*) were considered in this analysis. As the classes are known to have a hyperellipsoidal shape, we used the Mahalanobis based distance instead of the Euclidean distance in the membership function (32). The three mean vectors and covariance matrices are estimated on the data set. Although we did not aim at discussing here the relative merits of the various classification approaches (probabilistic, fuzzy or possibilistic), we compare the results of our simple classifier (Poss) without rejection to the ones obtained by two classical methods: the Quadratic Bayes classifier (QB) with Maximum A Posteriori classification and the k-Nearest Neighbor (k-NN). For this last classifier best results are obtained using a single neighbor (1-NN), on these data. We provide results obtained by a resubstitution (Table 3(a)) and by a leave-one-out ² procedure (Table 3(b)) using Duda *et al.* classical book [10] associated MATLAB[®] software.

Our simple Possibilistic classifier based on Equation (32) labeling function provides resubstitution results between the ones of the Quadratic Bayes and of the 1-NN classifiers. These results ensure the effectiveness of our basic classifier, however we must point out that this choice of labeling function is not imposed by our approach. Any membership function can be used as well in the L part of our design for classifiers with reject options (see section 4.1 and Algorithm 1). Having clarified this point, we now deal with ambiguity rejection, *i.e.* the H part. Results obtained with the Hamacher t-norm ($\gamma = 0$) are provided in Fig. 6. Fig.6(a) gives typicality degrees centered around class prototypes. Fig. 6(b) reveals order 2 ambiguity area between class 2 and 3 thanks to FOR-2. Fig. 6(c) exhibits one ambiguity area setting in front of class 1 in the direction of

² It must be noticed that in the case of a 1-NN classifier results of a leave-one-out procedure are obviously the same as in the resubstitution case. Both results are given here to facilitate the reading.

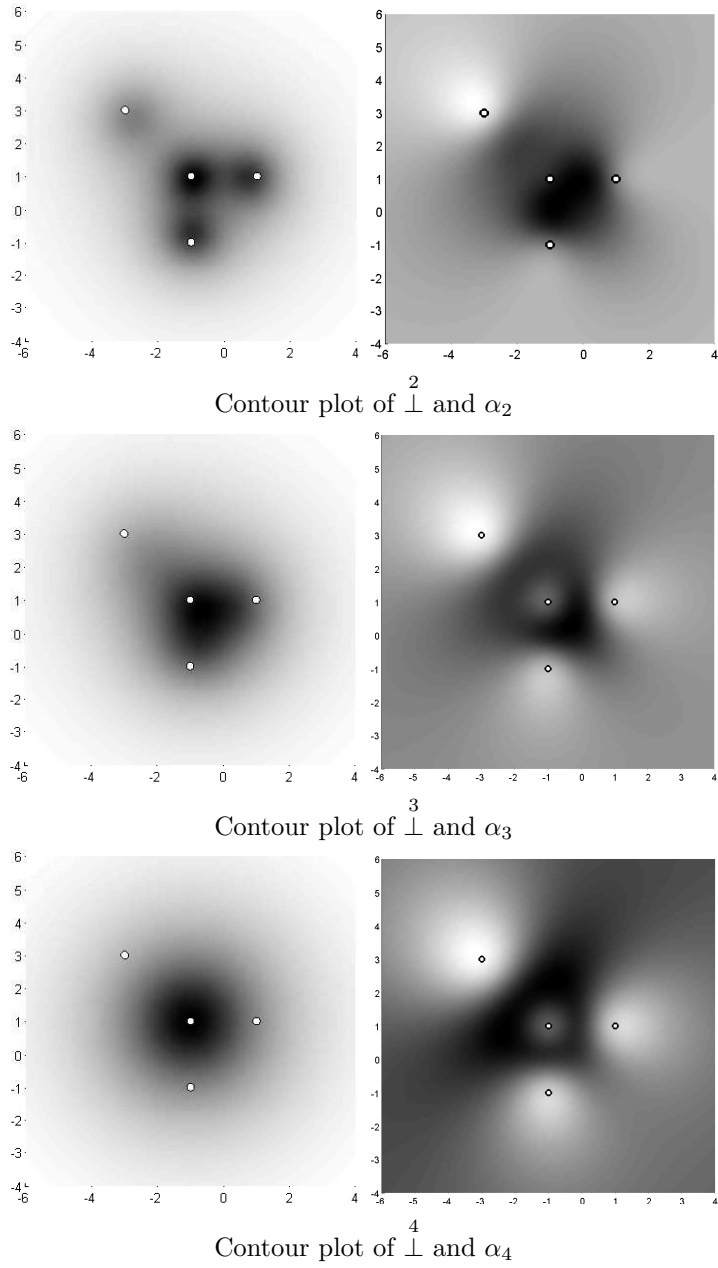


Fig. 4. $\text{FOR-}k$ (left) versus α_k (right) - Hamacher case ($\gamma = 0$).

Table 3
Iris data - Classification results without ambiguity rejection.

	correct	error	reject
Poss	146	4	0
QB	147	3	0
1-NN	143	7	0

(a) resubstitution

	correct	error	reject
Poss	145	5	0
QB	145	5	0
1-NN	143	7	0

(b) leave-one-out

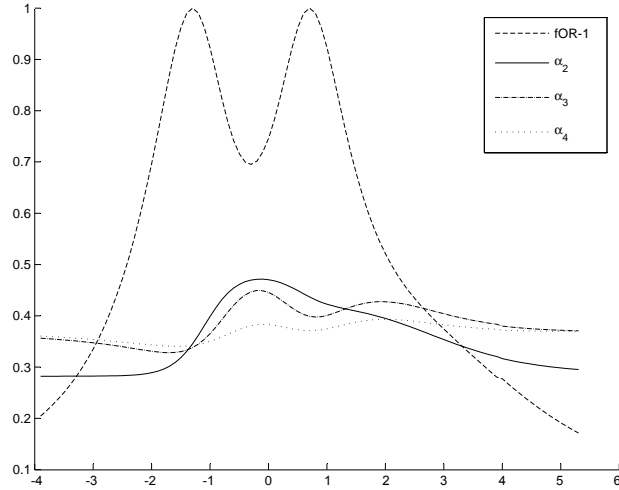


Fig. 5. α_k based on Hamacher t-norm ($\gamma = 0$) for the same line cut as in Fig.3.

the 2 other classes. All these ambiguities are geometrically well placed and order 3 ambiguity is of very small value, thus hardly selected by sensitive thresholds. However, it can be checked that these areas are attracted by classes centers. Thus, as expected, poorly tuned thresholds can lead to erroneously reject patterns with high membership values.

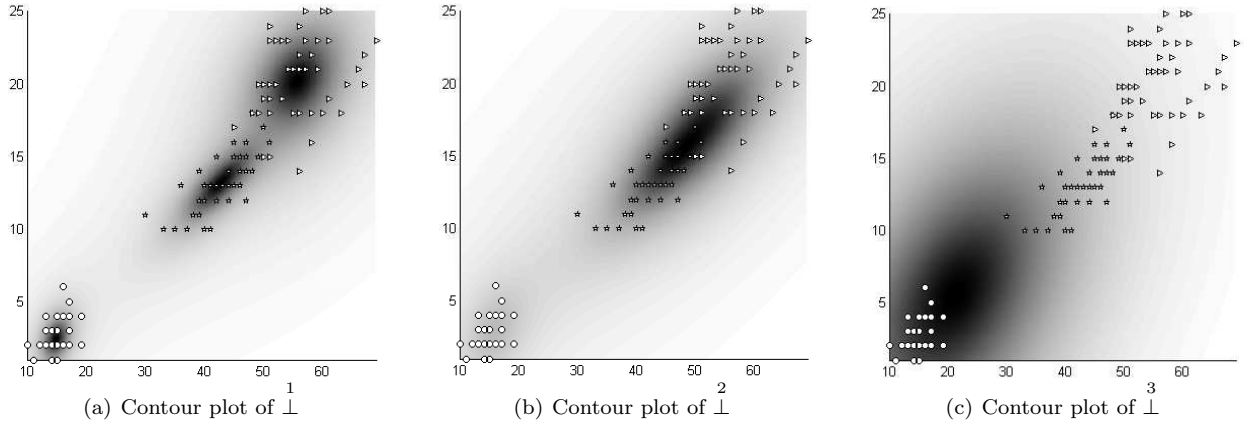
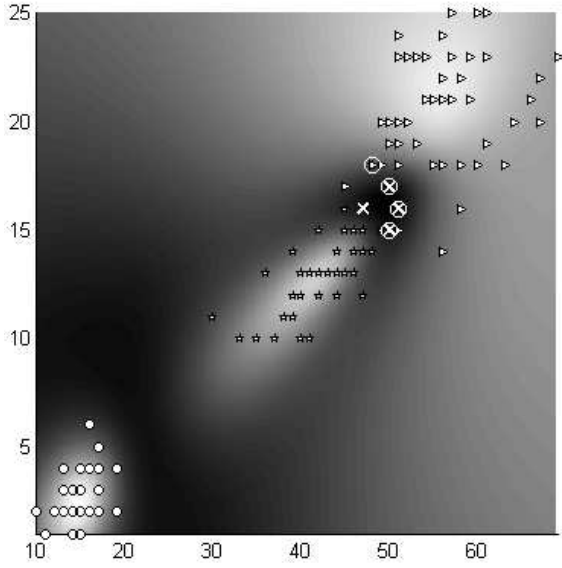


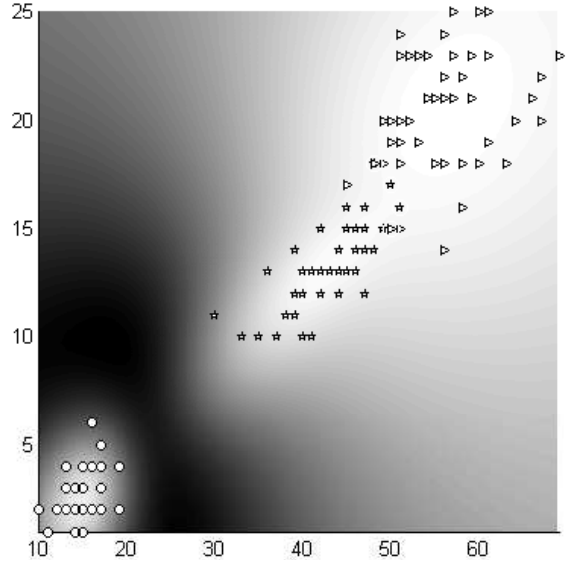
Fig. 6. Some α_k (based on Hamacher t-norm, with $\gamma = 0$) on the Iris dataset - Classes 1 to 3 are respectively represented by \circ , $*$ and \triangleleft symbols .

Therefore we used α_k instead of fOR- k to reveal ambiguity areas. It can be checked on Fig. 7(a) that ambiguity of order 2, as calculated with Hamacher t-norms, is well placed relatively to class centers. Meanwhile, Fig. 7(b) shows ambiguity of order 3 far away from class centers, thus leading not to reject between 3 classes. When applying the Algorithm 2 to these data, we choose to make no distance rejection, as this is not the core of the paper. By setting t_2 threshold to 0.28 and t_1 to 1, we ensured good ambiguity reject results without distance rejection. In Fig. 7(a), the four rejected points are darkened while the erroneous ones are circled.

Classification rates, obtained by applying the Algorithm 2 are shown in Table 4 as well as the various values for t_2 thresholds ; t_1 was set to 1 in order to disable distance rejection and to focus on ambiguity rejection. In the case of Hamacher t-norms ($\gamma = 0$), four points were rejected at the order 2, and none at the order 3. Three of the rejected points are ill-classified points, thus only one erroneous point remains. These results are quite good: roughly speaking for two rejected points we avoid one error point. Furthermore, we



(a) Contour plot of α_2 . The four ambiguity rejected points are crossed while the four erroneously classified points are circled.



(b) Contour plot of α_3 . No ambiguity rejection found.

Fig. 7. Contour plots of $\alpha_i, i = 2, 3$ with erroneously classified and rejected points.

found that, in accordance to visual inspection, no order 3 ambiguity is found. Rejection using standard t-norm behave in the same way (see Table 4). However, when using resubstitution, best results are obtained with Hamacher t-norm (4 rejected points save 3 error points).

Finally, we can check on Fig. 6(b) that the rejected points are geometrically reasonably well placed.

Table 4
Iris data - Classification results with ambiguity rejection.

	thresholds	classification		
		correct	error	reject
Poss	no rejection	146	4	0
Standard	$\alpha_2 = 0.75$	145	2	3
	$\alpha_3 = 0.75$	146	4	0
Hamacher	$\alpha_2 = 0.28$	145	1	4
$\gamma = 0$	$\alpha_3 = 0.28$	146	4	0

(a) resubstitution

	thresholds	classification		
		correct	error	reject
Poss	no rejection	145	5	0
Standard	$\alpha_2 = 0.75$	144	3	3
	$\alpha_3 = 0.75$	145	5	0
Hamacher	$\alpha_2 = 0.28$	142	3	5
$\gamma = 0$	$\alpha_3 = 0.28$	145	5	0

(b) leave-one-out

4.6.2. Glass Identification data set

The Glass Identification Database contains $n = 214$ observations from six 9-dimensional classes. The classes are unevenly distributed and the number of observations are 70,76,17,13,9,29 for classes from 1 to 6 respectively. These data are known to be highly non linearly separable. Our simple possibilistic prototype-based classifier using Mahalanobis distance leads to the results shown in Table 5(a) and Table 5(b) when using the resubstitution and leave-one-out method respectively. Let us notice that because of a ill-conditioned covariance matrix for class 5 (represented by only 9 points for 9 features) we must use Tikhonov regularization [19] to enable distance computation. Results provided by our simple Possibilistic classifier are a little better than the ones of the Quadratic Bayes classifier as well as the ones of the 1-Nearest Neighbor classifier when making resubstitution classification.

Due to the non linear separability of the data set, the 1-NN unsurprisingly outperforms other classifiers in the leave-one-out procedure: first two classifiers are prototype-based unlike the third one.

Table 5
Glass data - Classification results without rejection.

	correct	error	reject
Poss	161	53	0
BQ	156	58	0
1-NN	157	57	0

(a) resubstitution

	correct	error	reject
Poss	131	83	0
BQ	127	87	0
1-NN	157	57	0

(b) leave-one-out

For the rejection part, we again disabled distance rejection ($t_1 = 1$) to focus on ambiguity rejection. Results using resubstitution are provided in Table 6(a). A leave-one-out validation leads to the results given in Table 6(b), parameters being the same as for the resubstitution method. It can be checked that for the two different t-norms results are comparable but not equivalent:

- both detect few, if any, order 3 ambiguity (2 points for Standard t-norm in resubstitution, 1 point for Hamacher in leave-one-out);
- classification rates are better for Standard in resubstitution (5 rejected points save 4 error points, for the same result Hamacher needs to reject 8 points).

Table 6
Glass data - Classification results with ambiguity rejection.

	classification thresholds	correct	error	reject
Poss	no rejection	161	53	0
Standard	$\alpha_2 = 0.9$	160	49	5
	$\alpha_3 = 0.9$	161	51	2
Hamacher	$\alpha_2 = 0.25$	157	49	8
$\gamma = 0$	$\alpha_3 = 0.25$	161	53	0

(a) resubstitution

	classification thresholds	correct	error	reject
Poss	no rejection	131	83	0
Standard	$\alpha_2 = 0.9$	130	81	3
	$\alpha_3 = 0.9$	131	83	0
Hamacher	$\alpha_2 = 0.25$	127	81	6
$\gamma = 0$	$\alpha_3 = 0.25$	131	82	1

(b) leave-one-out

It can be shown that the choice of the t-norm make some influence on the results but in a sensible way: they agree on ambiguity order but some t-norms seem to be better adapted to some data sets. As an example, best results are obtained with Hamacher t-norm on the Iris data set, while Standard t-norm performed better on the Identification Glass data set.

5. Conclusion

In this paper, we have proposed a k -order Fuzzy OR operator, $\text{fOR-}k$, built on t-norms which is a straightforward generalization of a previously defined second order fuzzy OR operator. We have exhibited its main mathematical properties that make it suitable to define k -order ambiguity measures for classification application according to a chosen t-norm. Among these properties, we have studied the so-called weak compensation property (Equation 22) which appears of major importance. We proved that it is satisfied for the standard norms $S = (\top_S, \perp_S)$ and up to a bound b in the general case. This does not seem to be a real problem for many of the applications we have in mind, *i.e.* pattern recognition related applications. Since the special case of the $\text{fOR-}2$ has been already successfully studied in the frameworks of variable selection and cluster validity, we chose to focus on another pattern recognition application: classification with reject options, especially ambiguity rejection.

In this context, $\text{fOR-}k$ exhibits some particular behavior: the higher orders operators are influenced by the lower ones. To encounter this difficulty, we have introduced an aggregation scheme built on top of the $\text{fOR-}k$, according to the fuzzy eXclusive OR operator idea. This scheme allows us to propose an *ad-hoc*

solution to the multiple thresholds setting problem one has to face in the classification algorithm. Thus, a *single* threshold value is required to deal with all the ambiguity reject areas.

Since $\text{fOR}-k$ generalizes the notion of the k^{th} highest value, we argue that it measures the similarity between the $\text{fOR}-j$ operator values for $1 \leq j \leq k$. So it is a $(1, k)$ -blockwise similarity operator [6]. Investigating more general (j, k) -blockwise similarity operators, say $\Phi_{j,k}$, is promising. These operators will extend the one proposed in this paper ($\text{fOR}-k$). This leads us to investigate the definition of blockwise similarity and what kind of mathematical properties must be taken into account.

This and other extensions of the present work to image processing aspects, such as edge detection and mathematical morphology, are left for further study.

References

- [1] E. Anderson, The irises of the gaspe peninsula, *Bulletin of the American Iris Society* 59 (1935) 2–5.
- [2] J. Bezdek, J. Keller, R. Krishnapuram, N. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, The Handbooks of Fuzzy Sets Series (Dubois, D. and Prade, H., Eds), Kluwer Academic Publishers, Norwell, Massachusetts, 1999.
- [3] E. Bichot, Filtrage particulière fondé sur les informations de mouvement et de couleur pour le suivi d’objets non rigides en milieu perturbé. application au projet aqu@thèque, Ph.D. thesis, Université de La Rochelle (2006).
- [4] C. Blake, C. Merz, UCI repository of machine learning databases, dept. of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/~mlearn/MLRepository.html> (1998).
- [5] N. D. Bruijn, C. van Ebbenhorst Tengbergen, D. Kruyswijk, On the set of divisors of a number, *Nieuw Arch. Wiskunde* 23 (2) (1951) 191–193.
- [6] H. L. Capitaine, T. Batard, C. Frélicot, M. Berthier, Mesure de similarité par blocs via les normes triangulaires et l’intégrale de sugeno - application à la détection de contours, in: *Logique Floue et Applications (LFA)*, Nîmes, 2007.
- [7] C. Chow, On optimum recognition error and reject tradeoff, *IEEE Trans. on Information Theory* IT-16 (1970) 41–46.
- [8] D. Dubois, H. Prade, A review of fuzzy sets aggregation connectives, *Information Sciences* 36 (1985) 85–121.
- [9] B. Dubuisson, M.-H. Masson, A statistical decision rule with incomplete knowledge about classes, *Pattern Recognition* 26 (1) (1993) 155–165.
- [10] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification* (2nd edition), Wiley-Interscience, 2001, iISBN 0-471-05669-3.
- [11] C. Frélicot, M. Berthier, L. Mascarilla, A new cluster validity index for fuzzy clustering based on combination of dual triples, in: *Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, Vancouver, Canada, 2006.
- [12] C. Frélicot, L. Mascarilla, A third way to design pattern classifiers with reject options, in: : 21th Int. Conf. of the North American Fuzzy Information Processing Society (NAFIPS), New-Orleans, Louisiana, 2002.
- [13] C. Frélicot, L. Mascarilla, A. Fruchard, An ambiguity measure for pattern recognition problems using triangular-norms combination, *WSEAS Transactions on Systems* 8 (3) (2004) 2710–2715.
- [14] T.-M. Ha, The optimum class-selective rejection rule, *IEEE Trans. on Pattern Analysis and Machine Intelligence* PAMI-19 (6) (1997) 608–615.
- [15] E. Klement, R. M. (Eds), *Logical, algebraic, analytic and probabilistic aspects of triangular norms*, Elsevier, Amsterdam, The Netherlands, 2005.
- [16] G. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, NJ, USA, 1995.
- [17] L. Mascarilla, C. Frélicot, A class of reject-first possibilistic classifiers based on dual triples, in: *joint 9th Int. Fuzzy Systems Association World Congress (IFSA) and 20th Int. Conf. of the North American Fuzzy Information Processing Society (NAFIPS)*, Vancouver, Canada, 2001.
- [18] D. Semani, C. Frélicot, P. Courtellemont, Feature selection using an ambiguity measure based on fuzzy OR-2 operators, *International Fuzzy Systems Association, IFSA’2003*.
- [19] A. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math Dokl* 4 (1963) 1035–1038 English translation of *Dokl Akad Nauk SSSR* 15,1963, pp 501-504.