

Stratégies de rejet en classification supervisée : une synthèse par opérateurs de De Morgan

Reject strategies for supervised classification with De Morgan operators : a review

C. Frélicot, L. Mascarilla

Laboratoire Informatique Image Interaction (L3i), UPRES EA 2218, Université de La Rochelle, Avenue Michel Crépeau, 17042 La Rochelle, France cedex 1. Téléphone : 05 46458234 Fax : 05 46458242 E-mail : carl.frelicot, laurent.mascarilla@univ-lr.fr

Manuscrit reçu le 15 novembre 2002

Résumé et mots clés

Dans cet article, nous proposons une synthèse des stratégies mises en œuvre pour la conception de discriminateurs avec options de rejet opérant en deux étapes séquentielles. Outre l'approche classique dite « accepte d'abord », nous avons récemment défini des classes générales qui suivent deux approches différentes dites « rejette d'abord » [Fré98a, MF01b] et « mélange d'abord » [SFM02]. Ces trois approches diffèrent par la nature, et l'ordre, des tests effectués pour produire la sortie du discriminateur. La première consiste à tester en premier lieu le rejet de distance, puis seulement si nécessaire à tester l'affectation exclusive contre le rejet d'ambiguïté, la deuxième et la troisième, quant à elles, débutent, respectivement, par un test pour le classement exclusif et un test pour le rejet d'ambiguïté à opposer aux alternatives correspondantes. Nous unifions ici ces trois familles de discriminateurs par l'utilisation d'opérateurs flous fondés sur des opérateurs de De Morgan (t-norme, t-conorme, complément). Les comportements des différentes approches sont illustrées sur des exemples synthétiques.

Classification supervisée, rejet, opérateurs de De Morgan.

Abstract and key words

In this article we review strategies used in the design of two-folded rejection-based classifiers. Beside the so-called classical "accept-first" strategy we have recently proposed very general families built on two different approaches, namely the "reject-first" [Fré98a, MF01b] and "mixture-first" [SFM02] reject schemes. These three approaches differ by the kind, as well as the order, of the tests leading to the classifier final output. While the first one starts by testing for distance rejection and, if necessary, finishes by testing for exclusive classification or ambiguity rejection respectively, the two others start respectively by testing for exclusive classification and ambiguity rejection, and then finish by the remaining alternatives. We unify the three schemes by defining fuzzy operators built on De Morgan operators (t-norms, t-conorms, complement). Behaviours of such different classifiers are illustrated on artificially generated examples.

Classification, reject options, De Morgan operators.

1. Introduction

Cet article aborde le problème de classification supervisée dont l'objectif est de définir des règles d'affectation pour un objet, ou forme, \mathbf{x} inconnue à reconnaître optimales, au sens de la séparation d'un ensemble fini de c classes d'objets mutuellement exclusives $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$, ensemble défini *a priori*. La définition de l'objet \mathbf{x} requiert un espace de représentation. Lorsque \mathbf{x} est caractérisé par p attributs numériques, on le considère comme un vecteur, ou un point $\mathbf{x} = (x_1, x_2, \dots, x_p)^t$ de \mathfrak{R}^p , son espace de représentation. Reconnaître une forme \mathbf{x} consiste à l'associer à l'une des classes de Ω . Une fonction d'affectation ou règle de classement est donc le plus souvent exclusive. Formellement, elle peut être définie comme une application $D : \mathfrak{R}^p \rightarrow \Omega$, $\mathbf{x} \mapsto \omega_i$. Un autre formalisme consiste à définir D par : $\mathfrak{R}^p \rightarrow \mathcal{L}_{hc}$ la base canonique de \mathfrak{R}^c dont chaque vecteur (unitaire) est associé à une classe ω_i . Autrement dit, $D : \mathfrak{R}^p \rightarrow \mathcal{L}_{hc}$, $\mathbf{x} \mapsto l(\mathbf{x}) = (l_1(\mathbf{x}), l_2(\mathbf{x}), \dots, l_c(\mathbf{x}))^t$, où $\mathcal{L}_{hc} = \{l \in \mathfrak{R}^c : l_i \in \{0, 1\}, \sum_{i=1,c} l_i = 1\}$. Certains auteurs considèrent qu'une fonction de classement n'est qu'une fonction d'étiquetage non forcément binaire et qu'alors D est définie par [BRLA96]:

$$D : \mathfrak{R}^p \rightarrow \mathcal{L}_{\bullet c} \quad (1)$$

$$\mathbf{x} \mapsto l(\mathbf{x}) = (l_1(\mathbf{x}), l_2(\mathbf{x}), \dots, l_c(\mathbf{x}))^t \quad (2)$$

où $\mathcal{L}_{\bullet c}$ représente l'un des espaces suivants :

- l'espace des vecteurs d'étiquettes possibilistes (au sens degré de typicalité) :

$$\mathcal{L}_{pc} = \left\{ l \in \mathfrak{R}^c : l_i \in [0, 1], \sum_{i=1,c} l_i \neq 0 \right\} = [0, 1]^c \setminus \mathbf{0} \quad (3)$$

- l'espace des vecteurs d'étiquettes floues et probabilistes :

$$\mathcal{L}_{fc} = \left\{ l \in \mathcal{L}_{pc} : \sum_{i=1,c} l_i = 1 \right\} \quad (4)$$

- l'espace des vecteurs d'étiquettes binaires :

$$\mathcal{L}_{hc} = \{ l \in \mathcal{L}_{fc} : l_i \in \{0, 1\} \} \quad (5)$$

Notons que ces espaces sont emboîtés ($\mathcal{L}_{hc} \subset \mathcal{L}_{fc} \subset \mathcal{L}_{pc}$) et que seul \mathcal{L}_{hc} induit un classement exclusif. Dans la pratique, par exemple en reconnaissance de chiffres manuscrits pour la lecture automatique de codes postaux, l'hypothèse d'exclusivité des classes est rarement vérifiée et une forme \mathbf{x} à reconnaître peut se situer à l'intersection de plusieurs classes. Les espaces \mathcal{L}_{fc} et \mathcal{L}_{pc} non binaires permettent de gérer cette situation d'ambiguïté entre les classes. Mais peut-on alors parler de reconnaissance ? Contrairement à ces auteurs [BRLA96], nous pensons qu'une règle de classement doit être définie à valeurs dans un espace de vecteurs d'étiquettes binaires, mais pas forcément unitaire afin

de formaliser de telles situations qualifiées de *rejet d'ambiguïté*. Selon les règles et les auteurs, on distingue souvent le *rejet d'ambiguïté total* correspondant à une affectation à toutes les classes, par exemple tous les chiffres, et le *rejet d'ambiguïté sélectif* correspondant à une affectation à un sous-ensemble des classes, par exemple les '3' et les '8'. Le premier est rarement intéressant du point de vue du praticien car il correspond à une non-reconnaissance. Le second est plus intéressant car, selon le contexte applicatif, il pourra peut-être mettre en œuvre un traitement plus fin permettant de lever cette ambiguïté. Une autre hypothèse rarement vérifiée en pratique rend ces espaces inadéquats : l'exhaustivité des classes. La forme \mathbf{x} à reconnaître peut se situer loin de toute classe, soit parce que toutes les classes n'ont pas été définies, comme des lettres pour un code postal, soit parce que l'acquisition de la forme a été perturbée au point de la rendre aberrante. Pour gérer ces situations qualifiées de *rejet de distance*, une règle de classement doit être définie à valeurs dans un espace de vecteurs d'étiquettes binaires incluant l'origine, c'est-à-dire le vecteur nul $\mathbf{0}$. Pour ces raisons, nous incluons $\mathbf{0}$ à \mathcal{L}_{pc} qui devient l'hypercube $[0, 1]^c$ et définissons l'espace $\mathcal{L}_{hc}^c \supset \mathcal{L}_{hc}$ par :

$$\mathcal{L}_{hc}^c = \{ l \in \mathcal{L}_{pc} : l_i \in \{0, 1\} \} \quad (6)$$

Cet espace est l'ensemble des sommets de l'hypercube unité. Les différents espaces (3-6) sont représentés sur la figure 1 dans le cas de $c = 3$ classes. Une partition stricte de \mathcal{L}_{hc}^c permet de définir les trois actions possibles selon la valeur du vecteur d'étiquettes $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$ associé à un point \mathbf{x} à classer ou, de manière équivalente, selon la valeur de $s(\mathbf{x}) = \sum_{i=1,c} l_i(\mathbf{x})$:

- action α_i de classement exclusif si $l(\mathbf{x}) \in \mathcal{L}_{hc} \Leftrightarrow s(\mathbf{x}) = 1$
- action α_d de rejet de distance si $l(\mathbf{x}) = \mathbf{0} \Leftrightarrow s(\mathbf{x}) = 0$
- action α_0 de rejet d'ambiguïté si $l(\mathbf{x}) \in \mathcal{L}_{hc}^c \setminus \{\mathbf{0}, \mathcal{L}_{hc}\} \Leftrightarrow s(\mathbf{x}) > 1$, incluant le cas particulier du rejet d'ambiguïté total si $l(\mathbf{x}) = \mathbf{1} \Leftrightarrow s(\mathbf{x}) = c$.

Nous avons par ailleurs établi dans une tentative d'unification des règles de classement avec options de rejet [Fré98a] qu'il est plus logique de définir une règle de classement D par un couple de fonctions bien distinctes (L, H) composé :

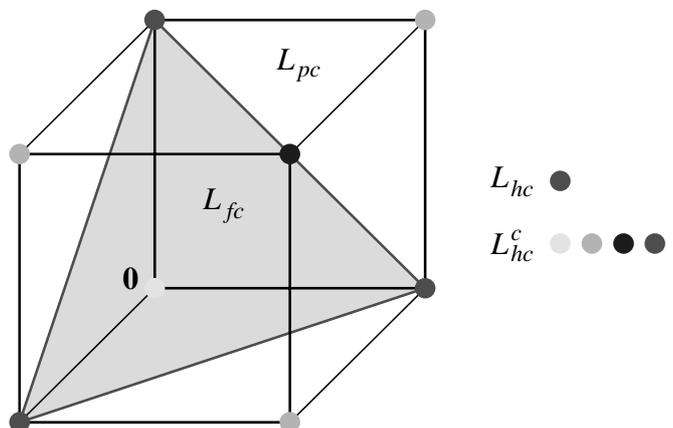


Figure 1. Différents espaces d'étiquettes

- d'une fonction d'étiquetage $L: \mathcal{R}^p \rightarrow \mathcal{L}_{\bullet c}$ (\mathcal{L}_{pc} ou \mathcal{L}_{fc}),

$$\mathbf{x} \mapsto \mu(\mathbf{x}) = (\mu_1(\mathbf{x}), \mu_2(\mathbf{x}), \dots, \mu_c(\mathbf{x}))^t$$

- d'une fonction de classement $H: \mathcal{L}_{\bullet c} \rightarrow \mathcal{L}_{hc}^c$,

$$\mu(\mathbf{x}) \mapsto l(\mathbf{x}) = (l_1(\mathbf{x}), l_2(\mathbf{x}), \dots, l_c(\mathbf{x}))^t$$

En effet, bon nombre de règles de classement avec ou sans option de rejet ne se distinguent que par la fonction d'étiquetage L et partagent la même fonction de classement H , par exemple: $i = \operatorname{argmax}_j \mu_j(\mathbf{x})$, $l_j(\mathbf{x}) = \delta_{ij}$ ($\forall j = 1, c$) où δ_{ij} représente le symbole de Kronecker. La réciproque est aussi vraie. C'est au niveau de L qu'une hypothèse sur le modèle mathématique sous-jacent à une règle de classement est en général faite. Dans le cas d'un modèle reposant sur la théorie des probabilités, L est à valeurs dans \mathcal{L}_{fc} et les $\mu_i(\mathbf{x})$ sont les probabilités a posteriori que \mathbf{x} soit issu de la classe ω_i , notées $P(\omega_i|\mathbf{x})$. Rappelons que le cadre probabiliste permet de bien représenter l'incertitude de l'information mais pas son imprécision et conduit souvent à confondre les deux. Un modèle reposant sur la théorie des ensembles flous induit lui aussi une fonction L à valeurs dans \mathcal{L}_{fc} et les $\mu_i(\mathbf{x})$ sont souvent des fonctions d'appartenance mesurant le degré d'appartenance de \mathbf{x} à la classe ω_i . Le cadre flou permet de bien modéliser l'imprécision, la gestion de l'incertitude étant alors reportée sur la fonction H par l'intermédiaire des options de rejet. Précisons que la contrainte de normalisation incluse dans la définition de \mathcal{L}_{fc} sur les étiquettes probabilistes est un axiome de Kolmogorov. L'équivalent pour les étiquettes floues est généralement attribué à Ruspini [Rus69]:

$$\sum_{j=1}^c \mu_j(\mathbf{x}) = 1 \tag{7}$$

Relacher cette contrainte permet de définir des fonctions L à valeurs dans \mathcal{L}_{pc} . Les étiquettes sont dites possibilistes en ce

sens qu'elles traduisent un degré de typicalité de \mathbf{x} à la classe ω_i . Ce cadre beaucoup plus souple (d'aucuns diront trop !) permet de bien appréhender à la fois l'imprécision et l'incertitude de l'information d'appartenance à une classe. Il est alors plus facile de définir une fonction H avec double option de rejet. Ceci est aussi possible dans un cadre probabiliste mais plus difficile dans le cas flou car les deux types de rejet sont parfois indiscernables, comme le montre l'exemple représenté sur la figure 2. Nous avons reporté à gauche les valeurs des étiquettes floues des points situés sur l'hyperplan séparateur entre $c = 2$ classes $\omega_1(\circ)$ et $\omega_2(\star)$, hyperplan représenté à droite par un trait plein vertical. Toutes les étiquettes ont une valeur proche de $\frac{1}{c} = \frac{1}{2}$, que les points soient proches ou non des classes. Pour rejeter en distance les points les plus éloignés par seuillage des étiquettes, il faudrait utiliser un seuil supérieur à $\frac{1}{2}$, par exemple celui représenté à gauche en pointillés. Mais dans ce cas, les points situés au lieu où les classes se chevauchent le plus seraient aussi rejetés en distance; or s'ils doivent être rejetés, ils devraient plutôt l'être en ambiguïté.

En classification floue ou possibiliste, il est commun d'utiliser comme fonction d'étiquetage associée à une classe ω_i une fonction d'appartenance $\mu_i(\mathbf{x})$ définie à partir d'un ou plusieurs prototypes de cette classe. Un prototype est un représentant de la classe qui peut avoir été obtenu par expertise ou estimé sur un ensemble d'apprentissage. Par exemple, le centre de gravité \mathbf{p}_i de ω_i est un prototype usuel. Alors, un certain nombre de fonctions monotones décroissantes d'une mesure de dissimilarité entre le point \mathbf{x} à classer et le prototype ont été définies (voir [Pal91]). Parmi celles-ci, citons:

- la fonction de type PCM (Possibilistic C-Means, [KK93])

$$\mu_i(\mathbf{x}) = \frac{1}{1 + \left(\frac{d^2(\mathbf{x}, \mathbf{p}_i)}{\eta_i}\right)^{1/(m-1)}} \tag{8}$$

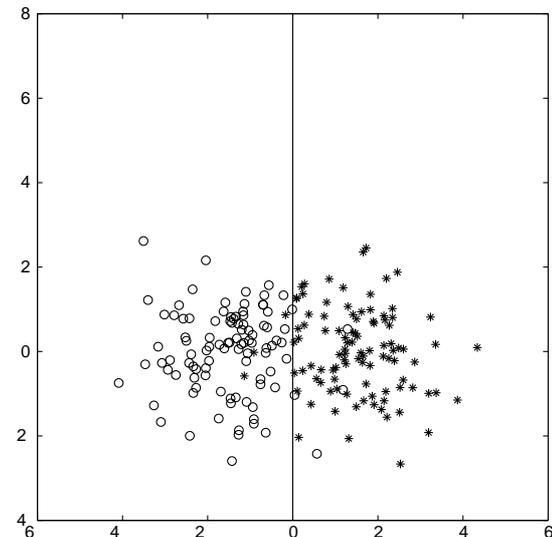
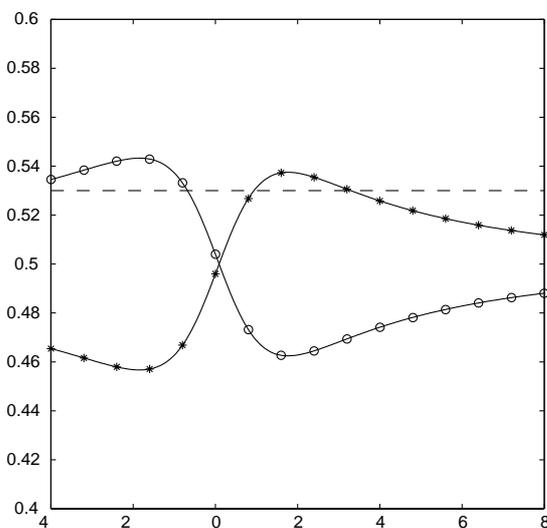


Figure 2. Seuillage (—) d'étiquettes floues (droite) de points situés sur l'hyperplan séparateur (—) entre deux classes $\omega_1(\circ)$ et $\omega_2(\star)$ (gauche)

où m , appelé exposant de flou, est un paramètre à fixer dans l'intervalle $]1, +\infty[$, η_i est un paramètre de résolution de la classe ω_i et classiquement d est la distance de Mahalanobis :

$$d^2(\mathbf{x}, \mathbf{p}_i) = (\mathbf{x} - \mathbf{p}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{p}_i) \quad (9)$$

où Σ_i est la matrice de covariance de la classe considérée. Cette fonction d'étiquetage est *a priori possibiliste* en ce sens qu'elle est à valeurs dans $[0, 1]$. Sauf mention contraire, nous utiliserons cette fonction (8) dans les exemples présentés avec $m = 2$ et $\eta_i = 1, \forall i = 1, c$.

- la fonction de type FCM (*Fuzzy C-Means*, [Bez87]) à valeurs normalisées au sens de (7)

$$\mu_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^c \left(\frac{d^2(\mathbf{x}, \mathbf{p}_i)}{d^2(\mathbf{x}, \mathbf{p}_j)} \right)^{1/(m-1)}} \quad (10)$$

L'objectif de cet article est de présenter une synthèse des différentes stratégies qui peuvent être mises en œuvre pour la conception de règles de classement H avec rejet à valeurs dans tout ou partie de l'espace \mathcal{L}_{hc}^c défini par (6). Nous nous limitons aux approches qui procèdent par seuillage et combinaison des c étiquettes fournies par une fonction d'étiquetage L . Les approches consistant à ajouter une étiquette supplémentaire pour modéliser une classe dédiée au rejet de distance, que ce soit dans un cadre probabiliste [Smy94], flou [Dav91] ou possibiliste [MDF96] ne sont pas abordées. Dans ce contexte, la section 2 dresse un panorama des différents types de règles selon le sous-espace de \mathcal{L}_{hc}^c mis en jeu :

- sans rejet (classement exclusif) : $l(\mathbf{x}) \in \mathcal{L}_{hc}$,
- avec rejet de distance : $l(\mathbf{x}) \in \mathcal{L}_{hc} \cup \mathbf{0} = \mathcal{L}_{hc}^0$,
- avec rejet d'ambiguïté total : $l(\mathbf{x}) \in \mathcal{L}_{hc} \cup \mathbf{1} = \mathcal{L}_{hc}^1$,
- avec rejet d'ambiguïté sélectif : $l(\mathbf{x}) \in \mathcal{L}_{hc}^c \setminus \{\mathbf{0}\} = \mathcal{L}_{hc}^{c \setminus 0}$,
- avec rejet de distance et rejet d'ambiguïté total : $l(\mathbf{x}) \in \mathcal{L}_{hc} \cup \mathbf{0} \cup \mathbf{1} = \mathcal{L}_{hc}^{01}$,
- avec rejet de distance et rejet d'ambiguïté sélectif : $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$.

Nous montrons que lorsque les deux types de rejet sont pris en compte, ils le sont séquentiellement. Ceci permet de définir différentes stratégies de rejet selon l'ordre de prise en compte. À la section 3, nous proposons une synthèse originale de ces stratégies pour le dernier type de règles. Selon la stratégie, un opérateur privilégié de combinaison des étiquettes, fondé sur des opérateurs de De Morgan, est proposé. Chaque opérateur permet de définir une classe de règles de classement avec double option de rejet que nous illustrons, sans perte de généralité, par des exemples artificiels volontairement simples, montrant ainsi que ces règles permettent d'obtenir des frontières de classement judicieuses. Nous concluons à la section 4 en mettant l'accent sur les problèmes ouverts et présentons quelques perspectives.

2. Classement et options de rejet

2.1. Classement exclusif

Dans le cadre probabiliste, la théorie bayésienne de la décision fournit la solution optimale au problème de classification. Si on note α_i l'action consistant à affecter un point \mathbf{x} à la classe ω_i , on peut associer un coût $\mathcal{C}(\alpha_i | \omega_j)$ à cette action α_i alors que la classe à laquelle appartient \mathbf{x} est ω_j . L'espérance mathématique du coût correspondant à l'action α_i , appelée risque conditionnel, est définie par :

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1}^c \mathcal{C}(\alpha_i | \omega_j) P(\omega_j | \mathbf{x}) \quad (11)$$

où $P(\omega_j | \mathbf{x})$ représente la probabilité *a posteriori* de classer le point \mathbf{x} dans la classe ω_j , probabilité donnée par la formule de Bayes. En considérant l'ensemble de toutes les valeurs possibles de \mathbf{x} , on définit le risque moyen par :

$$R(\alpha) = \int_{\mathcal{R}^p} R(\alpha(\mathbf{x}) | \mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (12)$$

où $f(\mathbf{x})$ représente la densité au point \mathbf{x} .

La règle de Bayes minimise $R(\alpha)$. L'action qui minimise $R(\alpha)$, minimise $R(\alpha_i | \mathbf{x})$. Dans le cas particulier de coûts $\{0, 1\}$, c'est-à-dire $\mathcal{C}(\alpha_i | \omega_j) = 1 - \delta_{ij}$, le risque conditionnel devient :

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1; j \neq i}^c P(\omega_j | \mathbf{x}) \quad (13)$$

$$= 1 - P(\omega_i | \mathbf{x}) \quad (14)$$

En conséquence, la règle de Bayes consiste à choisir l'action α_i qui maximise $P(\omega_i | \mathbf{x})$; c'est pourquoi on l'appelle la règle du *Maximum a posteriori* (MAP). Elle est optimale au sens de la minimisation de la probabilité d'erreur de classement $R(\alpha)$. Avec nos notations, elle se résume comme suit, où $l(\mathbf{x}) \in \mathcal{L}_{hc}$:

Tableau 1. Règle du maximum a posteriori (MAP)

$L :$	$\mu_j(\mathbf{x}) = P(\omega_j \mathbf{x})$
$H :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \delta_{ij} \quad (\forall j = 1, c)$

En pratique, les probabilités *a posteriori* $P(\omega_i | \mathbf{x})$ sont inconnues et estimées sur un ensemble d'apprentissage pour lequel on connaît l'appartenance des éléments aux différentes classes. Parmi les méthodes communément utilisées, citons celles permettant d'estimer les densités conditionnelles comme les *noyaux de Parzen* [DH73], les *k-PPV* (k-Plus Proches Voisins) [Fuk90], les fonctions potentielles, les réseaux de neurones [Bis95].

Si les étiquettes sont floues ou possibilistes, on obtient exactement la même règle de classement, même s'il convient de se souvenir que l'interprétation du résultat est fondamentalement différente. Introduite dans [PM77], on l'appelle aussi la règle du MAP (*Maximum d'Appartenance*).

2.2. Rejet d'ambiguïté

La notion de rejet, introduite par Chow [Cho57], a pour but de ne pas classer un point \mathbf{x} pour lequel plusieurs probabilités *a posteriori* sont du même ordre de grandeur. Compte tenu de la contrainte de normalisation (7), il est évident que cette situation intervient si la probabilité *a posteriori* maximale est faible. Dans le cadre bayésien, on peut formaliser ceci en introduisant une action supplémentaire α_0 à laquelle on associe un coût constant $\mathcal{C}(\alpha_0|\omega_j) = \mathcal{C}_0 \geq 0$ quelle que soit la classe ω_j ($j = 1, c$). On obtient alors la règle présentée en tableau 2 où $l(\mathbf{x}) \in \mathcal{L}_{hc}^1$:

Tableau 2. Règle de Chow

$L :$	$\mu_j(\mathbf{x}) = P(\omega_j \mathbf{x})$
$H :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \begin{cases} 1 & \text{si } \mu_i(\mathbf{x}) < 1 - \mathcal{C}_0 \\ \delta_{ij} & \text{sinon} \end{cases} \quad (\forall j = 1, c)$

Ce mécanisme permet de rejeter tout point pour lequel le risque d'erreur est important. Dans de nombreuses applications, il est possible d'appliquer après rejet un traitement plus attentif à un tel point dans le but de lever l'ambiguïté à moindre risque. La performance d'une telle règle est caractérisée par un compromis erreur-rejet [Cho70]. En fait, la règle de Chow minimise la probabilité d'erreur pour une probabilité de rejet fixée et inversement ; c'est en ce sens qu'elle est optimale. Les probabilités *a posteriori* étant normalisées, on montre que la possibilité de rejet disparaît dès que $\mathcal{C}_0 > \frac{c-1}{c}$. Dans ce cas, la règle est identique à la règle du MAP présentée en tableau 1. Le rejet d'ambiguïté ainsi défini est global ($l(\mathbf{x}) \in \mathcal{L}_{hc}^1$) puisqu'il est uniquement fondé sur la probabilité *a posteriori* maximale sans que la valeur des autres probabilités joue un rôle quelconque. Afin de raffiner ce rejet, on utilise la notion de *rejet d'ambiguïté sélectif* introduite dans la section 1 permettant d'accepter \mathbf{x} dans le sous-ensemble de classes $\omega \subseteq \Omega$ (de cardinalité > 1) dont les probabilités *a posteriori* dépassent \mathcal{C}_0 . Notre notation permet de formuler ce genre de règle aisément. Ha a proposé une règle optimale dans le cadre bayésien [Ha96] où $l(\mathbf{x}) \in \mathcal{L}_{hc}^{c,0}$:

Tableau 3. Règle de Ha

$L :$	$\mu_j(\mathbf{x}) = P(\omega_j \mathbf{x})$
$H :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \begin{cases} (\mu_j(\mathbf{x}) > \mathcal{C}_0) & \text{si } \mu_i(\mathbf{x}) < 1 - \mathcal{C}_0 \\ \delta_{ij} & \text{sinon} \end{cases} \quad (\forall j = 1, c)$

La règle de Ha minimise la probabilité d'erreur pour un nombre moyen de classes sélectionnées fixé et inversement [Ha97]. Elle étend en quelque sorte celle de Chow, le risque conditionnel (14) étant remplacé par :

$$R(\alpha(\mathbf{x})|\mathbf{x}) = 1 - \sum_{\omega_j \in \omega} P(\omega_j|\mathbf{x}) \tag{15}$$

On montre que si \mathcal{C}_0 n'appartient pas à l'intervalle $[0, 1/2]$ la possibilité de rejet disparaît et la règle se réduit à la règle du MAP. Notons que cette règle s'applique de manière directe si les étiquettes $\mu_j(\mathbf{x})$ sont floues ou possibilistes. Cependant, les étiquettes $\mu_j(\mathbf{x})$ n'étant pas normalisées dans le cas possibiliste, le coût \mathcal{C}_0 n'a plus de valeur limite et devient un simple seuil sur les étiquettes. Frélicot a appliqué ce principe de rejet sélectif dans [Fré97] afin de généraliser une règle proposée initialement dans [FD92], puis reprise dans [FMD95]. Il est bien évident que l'hypothèse de connaissance des probabilités *a posteriori* $P(\omega_j|\mathbf{x})$ assurant l'optimalité est rarement réaliste. Lorsqu'on utilise des estimations $\hat{P}(\omega_j|\mathbf{x})$, il a été proposé d'utiliser des coûts différents selon les classes $\mathcal{C}_{0,j} = \mathcal{C}(\alpha_0|\omega_j) \neq \mathcal{C}_0$ [FRG00]. Sous l'hypothèse selon laquelle les erreurs affectent plus les probabilités estimées $\hat{P}(\omega_j|\mathbf{x})$ dans les zones de mélange des classes, les auteurs ont montré que, pour un taux de rejet fixé, on peut trouver un ensemble de coûts $\mathcal{C}_{0,j}$ ($j = 1, c$) tel que la probabilité de classement correct conditionnellement à l'acceptation soit au moins égale à celle donnée par la règle de Chow [FR99]. C'est ce principe de seuillage différent selon les classes considérées qui a été appliqué pour définir la règle possibiliste décrite dans [Fré97].

2.3. Rejet de distance

Par opposition au précédent, le deuxième type de rejet introduit par Dubuisson et Masson [DM93] concerne les points \mathbf{x} qui se situent en dehors des zones denses, c'est-à-dire ceux pour lesquels la densité $f(\mathbf{x})$ est faible. On peut formaliser ceci en introduisant une action supplémentaire α_d à laquelle on associe une constante \mathcal{C}_d qui n'est pas un coût comme précédemment, mais un seuil sur la densité $f(\mathbf{x})$. On obtient une règle ne présentant plus le caractère d'optimalité précédent où $l(\mathbf{x}) \in \mathcal{L}_{hc}^0$:

Tableau 4. Règle de Dubuisson et Masson

$L :$	$\mu_j(\mathbf{x}) = P(\omega_j \mathbf{x})$
$H :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \begin{cases} 0 & \text{si } f(\mathbf{x}) < \mathcal{C}_d \\ \delta_{ij} & \text{sinon} \end{cases} \quad (\forall j = 1, c)$

Cette règle souffre de plusieurs problèmes majeurs dus à sa définition. D'abord, on ne peut l'appliquer que si l'on dispose d'une densité $f(\mathbf{x})$ à seuiller. Par conséquent, cette règle ne s'applique que dans un cadre probabiliste. Ensuite, le paramètre \mathcal{C}_d est relativement difficile à régler en pratique car, s'appliquant à la densité, il induit un rejet de distance défini au niveau global.

Enfin, pour la même raison, la règle peut conduire à des effets non désirés lorsque les classes en présence ont des distributions variées. Afin de pallier ces problèmes, il suffit de considérer la notion de rejet de distance localement, au niveau de chaque classe. La règle précédente a été modifiée en introduisant des seuils différents par classe $\mathcal{C}_{d,j} = \mathcal{C}(\alpha_d|\omega_j) \neq \mathcal{C}_d$ [MYP98]. Ces seuils sont appliqués aux composantes du mélange, c'est-à-dire aux densités conditionnelles pondérées par les probabilités *a priori*. L'extension de ce principe aux cadres non probabilistes consiste tout simplement à seuiller les étiquettes elles-mêmes. Soit $\mu_d = (\mu_{d,1}, \mu_{d,2}, \dots, \mu_{d,c})^t$ un vecteur de seuils et $s_d(\mathbf{x})$ la somme d'étiquettes booléennes définie par :

$$s_d(\mathbf{x}) = \sum_{j=1}^c (\mu_j(\mathbf{x}) \geq \mu_{d,j}) \tag{16}$$

De manière naturelle, l'action α_d sera appliquée si $s_d(\mathbf{x}) = \mathbf{0}$. On obtient la règle suivante, dite du *Rejet d'Appartenance* (RAP) où $l(\mathbf{x}) \in \mathcal{L}_{hc}^0$:

Tableau 5. Règle du rejet d'appartenance (RAP)

$L :$	$\mu_j(\mathbf{x})$	
$H :$	$s_d(\mathbf{x})$	éq. (16)
	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$	
	$l_j(\mathbf{x}) = \begin{cases} 0 & \text{si } s_d(\mathbf{x}) = \mathbf{0} \\ \delta_{ij} & \text{sinon} \end{cases}$	$(\forall j = 1, c)$

Le rejet de distance est ainsi défini de manière centrée sur chaque classe. Il faudra veiller à choisir les seuils $\mu_{d,j}$ avec attention afin ne pas engendrer d'effets de bord comme celui consistant à affecter le point \mathbf{x} à la classe d'appartenance maximale ω_i alors que $\mu_i(\mathbf{x})$ est inférieur à $\mu_{d,i}$. Un moyen de contourner ce problème est de forcer l'affectation exclusive à la classe d'où le point n'est pas rejeté en distance si $s_d(\mathbf{x}) \geq 1$. Cette règle n'est pas adaptée au cas où les étiquettes $\mu_j(\mathbf{x})$ sont floues à cause de la contrainte de normalisation (7), comme nous l'avons montré sur l'exemple de la figure 2 ; la contrainte doit être relâchée. C'est en ce sens que la notion de typicalité portée par des étiquettes possibilistes est fondamentale pour la discrimination avec option de rejet de distance. Dans la suite de cet article, nous considérerons que les étiquettes sont par défaut possibilistes.

2.4. Double option de rejet

Les deux types de rejet définis dans les sections précédentes sont bien évidemment complémentaires. La stratégie classique consiste à définir la fonction H selon deux étapes séquentielles permettant de tester les deux types de rejet indépendamment :

- H_1 : Rejet de distance ou non ?
- H_2 : Sinon, classement exclusif ou rejet d'ambiguïté ?

Insistons sur le fait que si un point \mathbf{x} est rejeté en distance lors de l'étape H_1 , il ne peut ni être classé de manière exclusive ni rejeté pour ambiguïté ; l'étape H_2 n'est alors pas considérée. Cette formulation trouve son intérêt dans des applications où la détection de points aberrants doit être rapide, par exemple dans le cas de panne de capteurs en diagnostic, de reconnaissance d'empreintes digitales ou de visages en biométrie, de caractères...

Dans [Fré98a], Frélicot a unifié la présentation des règles avec double option de rejet suivant cette stratégie dans les cadres probabiliste, flou et possibiliste. À titre d'illustration, entre autres règles, l'auteur a utilisé la combinaison des règles présentées en tableau 4 (H_1) et tableau 3 (H_2) pour le cas probabiliste. Si les étiquettes sont possibilistes, il a proposé de définir H_1 à partir de la règle présentée en tableau 5 et par exemple fonder H_2 sur un seuillage des étiquettes. Soit $\mu_0 = (\mu_{0,1}, \mu_{0,2}, \dots, \mu_{0,c})^t$ un vecteur de seuils et $s_0(\mathbf{x})$ la somme des étiquettes booléennes définie par :

$$s_0(\mathbf{x}) = \sum_{j=1}^c (\mu_j(\mathbf{x}) \geq \mu_{0,j}) \tag{17}$$

On obtient la règle dite du *Rejet d'Appartenance et d'Ambiguïté* (RAA) [Fré97] présentée en tableau 6 où $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$. Les seuils μ_d et μ_0 permettent de contrôler les taux de rejet de distance et d'ambiguïté respectivement. Plus on choisit une valeur élevée de μ_d , plus le nombre de points rejetés en distance est important. À la limite, tous les points sont rejetés en distance si $\mu_d = \mathbf{1}, \forall \mu_0$. À μ_d fixé, le taux de rejet en distance est constant et μ_0 permet de contrôler les taux respectifs de classement exclusif et de rejet d'ambiguïté. Plus μ_0 est choisi faible, plus on rejette de points en ambiguïté vers des sous-ensembles de classes à cardinalité de plus en plus élevée et moins de points sont classés. À la limite, $l(\mathbf{x})$ est dans \mathcal{L}_{hc}^{01} si $\mu_0 = \mathbf{0}$. La règle se réduit à la règle sans rejet du MAP ($l(\mathbf{x}) \in \mathcal{L}_{hc}$) lorsque $\mu_d = \mathbf{0}$ et $\mu_0 = \mathbf{1}$.

Tableau 6. Règle du rejet d'appartenance et d'ambiguïté (RAA)

$L :$	$\mu_j(\mathbf{x})$	
$H_1 :$	$s_d(\mathbf{x})$	éq. (16)
	$l_j(\mathbf{x}) = 0$ si $s_d(\mathbf{x}) = \mathbf{0}$	$(\forall j = 1, c)$
$H_2 :$	$s_0(\mathbf{x})$	éq. (17)
	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$	
	$l_j(\mathbf{x}) = \begin{cases} (\mu_j(\mathbf{x}) \geq \mu_{0,j}) & \text{si } s_0(\mathbf{x}) \neq \mathbf{0} \\ \delta_{ij} & \text{sinon} \end{cases}$	$(\forall j = 1, c)$

Les règles de classement que nous avons présentées dans cette section nécessitent comme toutes les autres d'être validées à partir d'un ensemble \mathfrak{N} de données étiquetées, c'est-à-dire pour lesquelles la classe est connue. Le principe commun à toute méthode de validation repose sur la séparation de \mathfrak{N} en sous-ensembles de conception et de test. L'ensemble de conception est utilisé pour estimer les paramètres des classes intervenant dans les fonctions d'étiquetage L et l'ensemble de test est utilisé

pour qualifier la performance de la règle. La probabilité d'erreur de classement est la mesure de performance privilégiée [JDM00]. Lorsqu'on introduit des options de rejet, un compromis erreur-rejet doit être réalisé et des probabilités de rejet estimées [Du01]. C'est à ce niveau que le réglage des seuils de rejet intervient manuellement ou par balayage des intervalles de définition des seuils et sélection de valeurs optimales vis-à-vis du compromis. Notons que ce compromis rend difficile la comparaison des règles sans et avec rejet, en particulier rejet de distance, sur des jeux de données réelles comme ceux disponibles en [BM98] et nous invitons le lecteur intéressé à consulter les références relatives à chaque règle pour obtenir des éléments de comparaison. En ce qui concerne les seuils de rejet de distance $\mu_{d,j}$ (16) et d'ambiguïté $\mu_{0,j}$ (17) de la règle RAA, Frélicot a proposé d'utiliser des statistiques d'ordre afin de les estimer de manière automatique à partir d'ensembles \aleph (étiquetés ou non) contenant de fortes proportions de points distribués uniformément sensés représenter des points à rejeter [Fré97].

3. Stratégies de rejet par opérateurs de De Morgan

3.1. Opérateurs de fusion

Dans la section 2.4, nous avons montré comment les règles de classement avec double option de rejet (distance, ambiguïté) pouvaient être définies à partir de deux fonctions H_1 et H_2 permettant de tester les deux types de rejet indépendamment. Ces fonctions reposent toujours sur la fusion des étiquettes $\mu_j(\mathbf{x})$. La logique floue offre des opérateurs mathématiques de combinaison de ces étiquettes adaptés à la définition de telles fonctions de classement avec rejet. Detyniecki présente ces opérateurs d'agrégation de manière très complète dans [Det00]. Le classement avec double option de rejet étant fondé sur l'appartenance ou non à une ou plusieurs classes, nous avons choisi d'utiliser :

- les normes triangulaires (ou t-normes), notées \top
- les conormes triangulaires (ou t-conormes), notées \perp
- les compléments, notés $\bar{\cdot}$

Ces opérateurs peuvent être vus comme l'extension au cas multi-valué des opérateurs d'intersection \cap , d'union \cup et de complément \neg de la théorie classique des ensembles que représentent les classes. Ils étendent aussi les connecteurs logiques ET, OU et NON de la logique Booléenne. Les t-normes et t-conormes sont commutatives, associatives, monotones croissantes et possèdent un élément neutre. Rappelons aussi que certaines combinaisons de ces normes sont duales relativement à l'opérateur de complément et satisfont les lois de De Morgan généralisées¹. Le lecteur intéressé peut trouver une syn-

thèse de ces triplets $(\top, \perp, \bar{\cdot})$ dans [DP85], [KY95]. Nous ne donnons ici que ceux que nous avons utilisés dans cet article et rappelons quelques caractéristiques :

- *normes Standard*: t-norme supérieure, t-conorme inférieure, les seules idempotentes

$$\mu_1 \top \mu_2 = \min(\mu_1, \mu_2) \quad (18)$$

$$\mu_1 \perp \mu_2 = \max(\mu_1, \mu_2) \quad (19)$$

$$\bar{\mu}_1 = 1 - \mu_1 \quad (20)$$

- *normes de Lukasiewicz*: t-norme satisfaisant la loi de non-contradiction, t-conorme satisfaisant la loi du tiers-exclus

$$\mu_1 \top \mu_2 = \max(0, \mu_1 + \mu_2 - 1) \quad (21)$$

$$\mu_1 \perp \mu_2 = \min(1, \mu_1 + \mu_2) \quad (22)$$

$$\bar{\mu}_1 = 1 - \mu_1 \quad (23)$$

- *normes de Hamacher*: famille paramétrée par $\gamma \geq 0$, normes archimédiennes strictes et continues²

$$\mu_1 \top \mu_2 = \frac{\mu_1 \mu_2}{\gamma + (1 - \gamma)(\mu_1 + \mu_2 - \mu_1 \mu_2)} \quad (24)$$

$$\mu_1 \perp \mu_2 = \frac{\mu_1 + \mu_2 - \mu_1 \mu_2 - (1 - \gamma)\mu_1 \mu_2}{1 - (1 - \gamma)\mu_1 \mu_2} \quad (25)$$

$$\bar{\mu}_1 = 1 - \mu_1 \quad (26)$$

lorsque $\gamma = 1$, on retrouve les t-norme et t-conorme probabilistes appelées aussi *produit* et *somme probabiliste*.

Dans le cas probabiliste, on peut jouer sur les probabilités *a priori* pour pondérer les densités conditionnelles au sein d'un mélange. Par exemple, en reconnaissance de caractères manuscrits dans des textes écrits en français, on voudra privilégier la classe des «e». Un mécanisme semblable est possible lorsqu'on fusionne des étiquettes μ_j en introduisant des poids w_j associés aux étiquettes. Il faut alors modifier les couples (\top, \perp) pour les prendre en compte. Il n'existe pas, à notre connaissance, dans la littérature d'approche unifiée pour cela. L'approche suggérée dans [DP88] pour les opérateurs archimédiens strictement monotones donne par exemple dans le cas des normes de Hamacher avec $\gamma = 0$, les versions pondérées simples suivantes :

$$\top_{j=1,c} \{\mu_j, w_j\} = \frac{1}{1 + c \sum_{j=1,c} w_j \frac{1 - \mu_j}{\mu_j}} \quad (27)$$

$$^1 \overline{\mu_1 \top \mu_2} = \bar{\mu}_1 \perp \bar{\mu}_2$$

$$\overline{\mu_1 \perp \mu_2} = \bar{\mu}_1 \top \bar{\mu}_2$$

$$^2 \forall \mu \in]0, 1[, \mu \top \mu < \mu < \mu \perp \mu$$

$$\forall \mu_2 > \mu_1, \mu \top \mu_1 < \mu \top \mu_2 \text{ et } \mu \perp \mu_1 < \mu \perp \mu_2$$

$$\bigcap_{i=1,c} \{\mu_i, w_j\} = \frac{1}{1 + \frac{1}{c \sum_{j=1,c} w_j \frac{\mu_j}{1 - \mu_j}}} \quad (28)$$

sous la contrainte $\sum_{j=1,c} w_j = 1$. Dans le cas équipondéré $\{\mu_j, 1/c\}$, on retrouve les opérateurs non pondérés. C'est pour cette simplicité que nous avons choisi $\gamma = 0$ dans les exemples, valeur la plus fréquemment rencontrée dans la littérature ; la t-norme est alors appelée *produit de Hamacher*.

3.2. stratégie « Accepte d'abord »

Nous qualifions l'approche suivie par les règles avec double option de rejet (distance, ambiguïté) présentées à la section précédente de stratégie *accepte d'abord* pour deux raisons :

- tester le rejet de distance ou non revient à tester de manière complémentaire l'acceptation ou non,
- nous avons proposé deux stratégies radicalement différentes, présentées dans les sections suivantes.

On peut reformuler les deux étapes séquentielles permettant de tester les deux types de rejet qu'une fonction de classement H doit comporter afin de fournir un vecteur d'étiquettes $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$ comme ceci :

- H_1 : Acceptation ou non ?
- H_2 : Si oui, classement exclusif ou rejet d'ambiguïté ?

L'étape caractérisant cette stratégie est H_1 . Nous rappelons aussi que si un point \mathbf{x} n'est pas accepté lors de l'étape H_1 , il ne peut ni être classé de manière exclusive ni rejeté pour ambiguïté ; l'étape H_2 n'est alors pas considérée. La nécessité d'accepter un point \mathbf{x} peut être mise en évidence par une valeur élevée de la combinaison de ses étiquettes $\mu_j(\mathbf{x})$ par un opérateur OU

(inclusif) flou défini par :

$$\bigcap_{j=1,c} \mu_j \quad (29)$$

La Figure 3 montre les isosurfaces de sortie du OU flou fondé sur la norme de Hamacher dans le cas non pondéré à gauche et pondéré à droite avec $w = (0.1 \ 0.9)^t$. Les zones claires, qui traduisent une valeur élevée de $\mu_1 \perp \mu_2$, suggèrent d'appliquer les actions α_1, α_2 ou α_0 . Inversement, les zones sombres traduisant une valeur faible de $\mu_1 \perp \mu_2$ suggèrent plutôt l'action de rejet de distance α_d . Un poids élevé, comme w_2 ici, devant jouer en faveur de l'action correspondante α_2 , déplace logiquement la zone correspondant à une valeur faible du OU flou vers des valeurs plus faibles de μ_2 et plus élevées de μ_1 . Il suffit donc de comparer $\bigcap_{j=1,c} \mu_j(\mathbf{x})$ à un seuil d'acceptation t_a pour définir H_1 . Si \mathbf{x} est accepté, il faut tester s'il peut être classé de manière exclusive ou rejeté en ambiguïté. N'importe quelle fonction H_2 définie selon les principes évoqués à la section 2.4 convient, en particulier celle de la règle RAA (voir tableau 6). Nous obtenons alors la règle présentée en tableau 7, où $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$.

Tableau 7. Règle accepte d'abord généralisée (ADAG)

$L :$	$\mu_j(\mathbf{x})$	
$H_1 :$	$l_j(\mathbf{x}) = 0$	si $\bigcap_{j=1,c} \mu_j(\mathbf{x}) \leq t_a \quad (\forall j = 1, c)$ éq. (29)
$H_2 :$	$s_0(\mathbf{x})$	éq. (17)
	$i = \arg \max_j \mu_j(\mathbf{x})$	
	$l_j(\mathbf{x}) = \begin{cases} (\mu_j(\mathbf{x}) \geq \mu_{0,j}) & \text{si } s_0(\mathbf{x}) \neq 0 \\ \delta_{ij} & \text{sinon} \end{cases}$	$(\forall j = 1, c)$

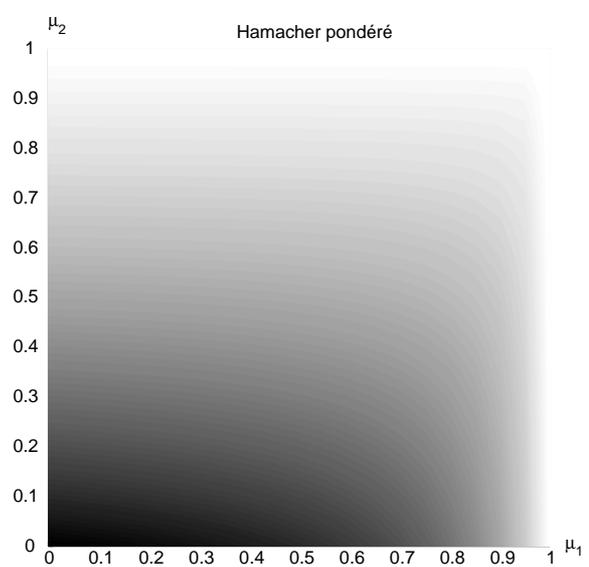
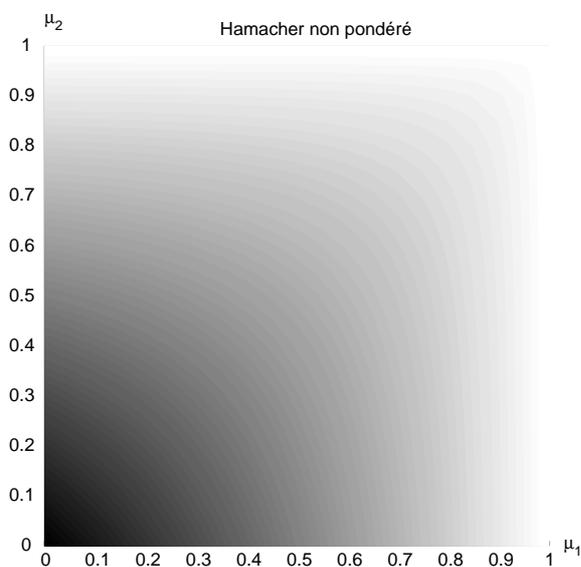


Figure 3. Isosurfaces d'un OU flou de Hamacher non pondéré (gauche) et pondéré (droite)

Les valeurs des seuils de rejet t_a et μ_0 de cette famille de discriminateurs permettent bien évidemment de contrôler les taux de rejet de distance, de classement exclusif et de rejet d'ambiguïté. Plus on choisit une valeur t_a faible, plus on accepte et moins on rejette en distance. À la limite, il n'y a pas de rejet de distance possible si $t_a = 0, \forall \mu_0$. Inversement, aucun point \mathbf{x} n'est accepté si $t_a = 1$. Pour une valeur fixée de t_a , le taux de rejet de distance est constant et c'est μ_0 qui permet de contrôler les taux respectifs de classement exclusif et de rejet en ambiguïté sélectif. Plus μ_0 est choisi faible, plus on rejette de points en ambiguïté vers des sous-ensembles de classes à cardinalité de plus en plus élevée et moins de points sont classés. À la limite, $l(\mathbf{x})$ est dans \mathcal{L}_{hc}^{01} si $\mu_0 = \mathbf{0}$. La règle se réduit à la règle sans rejet ($l(\mathbf{x}) \in \mathcal{L}_{hc}$) du *Maximum d'Appartenance* lorsque $t_a = 0$ et $\mu_0 = \mathbf{1}$.

Pour un réglage intermédiaire de ces seuils dans le cadre d'une application réelle, nous renvoyons le lecteur aux éléments de discussion donnés à la fin de la section 2.4. Nous nous contentons d'illustrer cette approche et les suivantes sur deux exemples synthétiques. Le premier consiste en un jeu de trois vecteurs jeux d'étiquettes possibilistes $\mu(\mathbf{x})$ que nous appelons jeu ETIQ. La sortie des OU flous non pondérés et les vecteurs d'étiquettes $l(\mathbf{x})$ fournis par la règle ADAG énoncée en tableau 7 pour les triplets Standard, de Hamacher et de Lukasiewicz sont présentés en tableau 8. Les seuils de rejet ont été fixés à $t_a = 0.5$ et $\mu_{0,j} = \mu_0 = 0.5, \forall j = 1, c$. Les résultats sont tout à fait conformes à l'intuition, quelles que soient les normes.

Le deuxième exemple est un jeu de données artificielles composé de $c = 3$ classes de 25 individus générés selon des distributions gaussiennes bi-dimensionnelles. Dans ce jeu que nous

appelons C3D2, deux classes ω_1 (\circ) et ω_2 (\star) se chevauchent notablement alors qu'une autre classe ω_3 (∇) est bien séparée des deux premières. Les isosurfaces du OU flou fondé sur la conorme standard (18) et les frontières de décision associées issues de la règle ADAG sont données à la figure 4. Les vecteurs d'étiquettes sont de type possibiliste. Ils ont été calculés selon (8-9) avec $\eta_i = \eta = 1$ ($\forall i = 1, c$) et $m = 2$. Précisons que ces vecteurs prendront toujours ces mêmes valeurs jusqu'à la fin de l'article. Les seuils de rejet ont été fixés à $t_a = 0.25$ et $\mu_{0,j} = \mu_0 = 0.25$ ($\forall j = 1, c$). Les régions de classement sont conformes à l'intuition: rejet de distance α_d en blanc, affectations exclusives α_i ($\forall j = 1, 3$) en gris clair, rejet d'ambiguïté sélectif $\alpha_0 = \{\alpha_1, \alpha_2\}$ en gris plus foncé.

Tableau 8. Exemples de classement par la règle accepte d'abord généralisée (ADAG)

$\mu(\mathbf{x})$	$\bigoplus_{i=1,c} \mu_i(\mathbf{x})$	$l(\mathbf{x})$
$(0.1, 0.8, 0.7)^t$	Std 0.80 Łuk. 1.00 Ham. 0.87	$(0, 1, 1)^t$
$(0.1, 0.05, 0.2)^t$	Std 0.20 Łuk. 0.35 Ham. 0.29	$(0, 0, 0)^t$
$(0.1, 0.05, 0.7)^t$	Std 0.70 Łuk. 0.85 Ham. 0.71	$(0, 0, 1)^t$

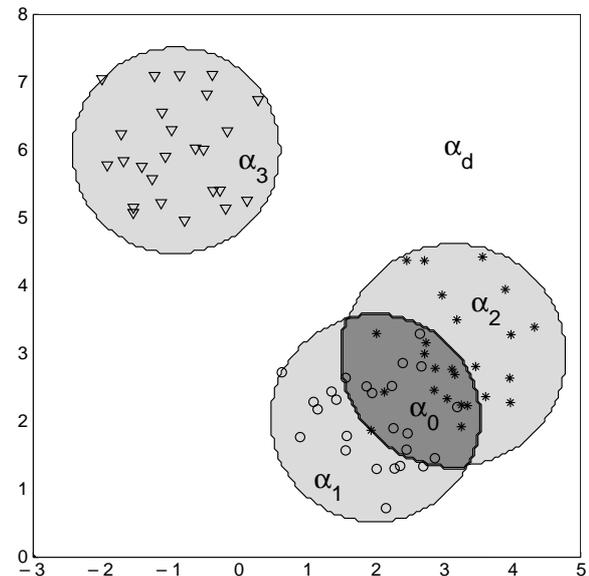
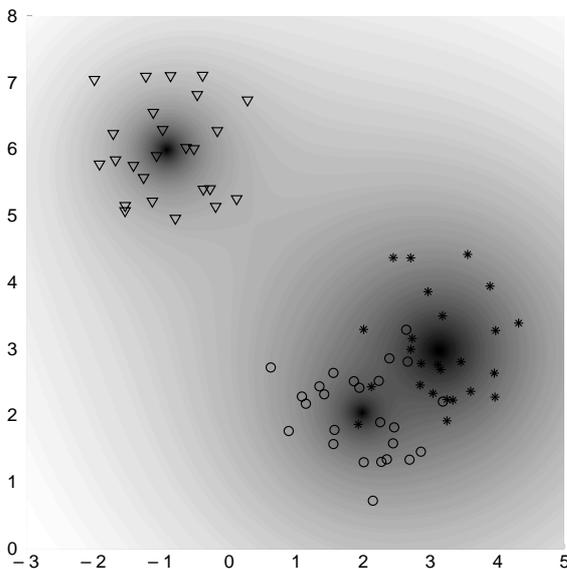


Figure 4. Isosurfaces du OU flou standard et régions de classement ($\omega_1 \circ, \omega_2 \star, \omega_3 \nabla$)

3.3. Stratégie « Rejette d'abord »

Frélicot *et al.* ont été les premiers à proposer une règle de classement reposant sur une stratégie de rejet radicalement différente de la précédente [FMD95]. L'idée est de tester d'abord si un point \mathbf{x} peut être classé de manière exclusive ou s'il doit être rejeté sans distinguer le type de rejet. C'est dans une deuxième étape que l'on teste, si besoin, le type de rejet. Cette stratégie est qualifiée *rejette d'abord* pour la démarquer des autres. Une règle H de classement suivant une telle stratégie peut se formuler de la manière suivante :

- H_1 : Rejet ou non ?
- H_2 : Si oui, rejet de distance ou d'ambiguïté ?

L'étape caractérisant cette stratégie est, là encore, H_1 . Si un point \mathbf{x} n'est pas rejeté lors de l'étape H_1 , il est classé de manière exclusive; l'étape H_2 n'est alors pas considérée. Récemment, Mascarilla et Frélicot ont proposé de généraliser cette stratégie [MF01a] en traduisant le classement exclusif d'un point \mathbf{x} par une valeur élevée de la combinaison de ses étiquettes possibilistes $\mu_j(\mathbf{x})$. Pour cela, les auteurs ont proposé d'utiliser un opérateur OU exclusif flou, noté \perp , et défini par :

$$\perp_{i=1,c} \mu_i = \perp_{i=1,c} (\mu_i \top \prod_{j=1,c; j \neq i} \bar{\mu}_j) \quad (30)$$

où $(\top, \perp, \bar{\cdot})$ représente n'importe quel triplet de De Morgan.

Les t-normes et t-conormes pondérées définies en (27-28) ne permettent pas de pondérer l'opérateur OU exclusif flou car il met en œuvre simultanément les opérateurs \top et \perp . En l'absence de solutions clairement reconnues, les mêmes auteurs ont proposé la solution la plus simple permettant de retrouver (30) dans le cas équipondéré, qui consiste à pondérer la conorme [MF01b] :

$$\perp_{i=1,c} \{\mu_i, w_i\} =$$

$$\perp_{i=1,c} \left(\left\{ \mu_i, 1 - \frac{c}{2} w_i \right\} \top \left\{ \prod_{j=1,c; j \neq i} \bar{\mu}_j, \frac{c}{2} w_i \right\} \right) \quad (31)$$

À titre d'illustration, la figure 5 montre les isosurfaces de sortie d'un OU exclusif flou fondé sur le triplet de Hamacher $(\top, \perp, \bar{\cdot})$ dans le cas non pondéré à gauche et pondéré à droite avec $w = (0.1 \ 0.9)^t$. Les zones claires traduisent une valeur élevée de $\mu_1 \perp \mu_2$, ce qui suggère d'appliquer les actions α_1 ou α_2 , le choix devant se porter sur celle correspondant à l'étiquette maximale. Inversement, les zones sombres suggèrent plutôt l'action de rejet, soit de distance α_d , soit d'ambiguïté α_0 selon que les étiquettes sont faibles ou élevées. Un poids élevé, comme w_2 ici, doit jouer en faveur de l'action d'affectation exclusive correspondante α_2 . La zone correspondant à une valeur élevée du OU exclusif flou est alors déplacée vers des valeurs plus faibles de μ_2 . Il suffit donc de comparer $\perp_{j=1,c} \mu_j(\mathbf{x})$ à un seuil de rejet t_r pour définir H_1 . Si \mathbf{x} est rejeté, il faut décider entre les deux types de rejet. Plusieurs fonctions H_2 sont envisageables, dont celle consistant à comparer à un seuil t_0 l'opérateur utilisé dans la fonction H_1 de la règle *Accepte D'Abord Généralisée* (ADAG) présentée en tableau 7, opérateur fondé sur la t-conorme correspondante. On obtient alors une famille de règles à la fois cohérente et très générale présentée en tableau 9 où $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$.

Tableau 9. Règle rejette d'abord généralisée (RDAG)

$L :$	$\mu_j(\mathbf{x})$
$H_1 :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \delta_{ij} \quad \text{si } \perp_{j=1,c} \mu_j(\mathbf{x}) \geq t_r \quad (\forall j = 1, c) \quad \text{éq. (30)}$
$H_2 :$	$l_j(\mathbf{x}) = \begin{cases} (\mu_j(\mathbf{x}) \geq t_0) & \text{si } \perp_{j=1,c} \mu_j(\mathbf{x}) > t_0 \\ 0 & \text{sinon} \end{cases} \quad (\forall j = 1, c)$

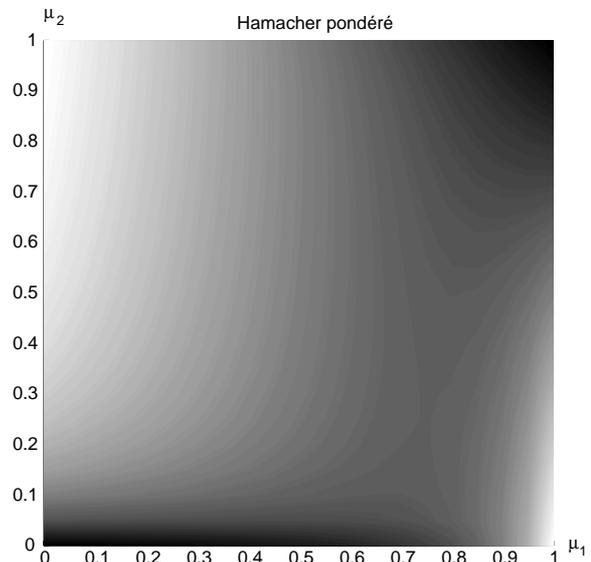
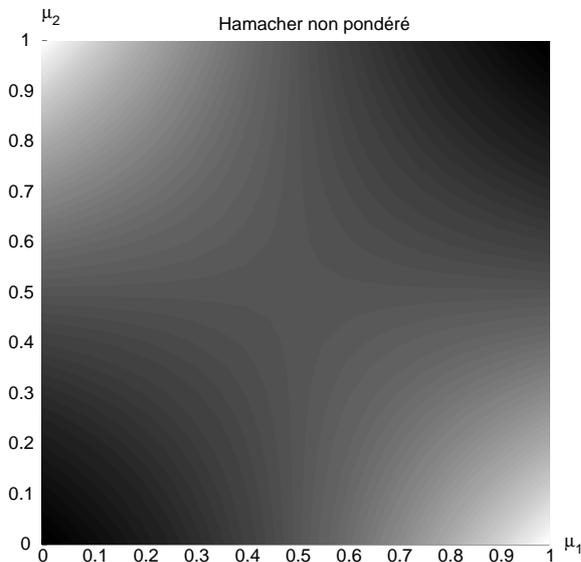


Figure 5. Isosurfaces d'un OU exclusif flou de Hamacher non pondéré (gauche) et pondéré (droite)

Le réglage des seuils permet de contrôler les taux de classement exclusif et de rejet. Plus on choisit une valeur de t_r élevée, moins il y a de classement exclusif et plus il y a de rejet. Pour une valeur fixée de t_r , le taux de classement exclusif est constant et c'est t_0 qui permet de contrôler les taux respectifs de rejet en distance et en ambiguïté sélective. Pour une valeur fixée de t_0 , le taux de rejet en distance est constant si $t_0 < t_r$ et t_r contrôle le rapport entre le taux de rejet en ambiguïté et le taux de classement exclusif. Les valeurs limites des seuils de rejet permettent de réduire cette famille de discriminateurs à des règles de classement sans rejet lorsque $t_r = 0, \forall t_0$. Dans ce cas, \mathbf{x} n'est jamais rejeté et par conséquent associé de manière exclusive à la classe d'appartenance maximale. On retrouve alors la règle du *Maximum d'Appartenance* sans rejet où $l(\mathbf{x}) \in \mathcal{L}_{hc}$. Au contraire, \mathbf{x} sera toujours rejeté lorsque $t_r = 1, \forall t_0$.

Tableau 10. Exemples de classement par la règle rejetée d'abord généralisée (RDAG)

$\mu(\mathbf{x})$	$\bigoplus_{i=1,c} \mu_i(\mathbf{x})$	$l(\mathbf{x})$
$(0.1, 0.8, 0.7)^t$	Std 0.30 Łuk. 0.00 Ham. 0.40	$(0, 1, 1)^t$
$(0.1, 0.05, 0.2)^t$	Std 0.20 Łuk. 0.05 Ham. 0.29	$(0, 0, 0)^t$
$(0.1, 0.05, 0.7)^t$	Std 0.70 Łuk. 0.55 Ham. 0.65	$(0, 0, 1)^t$

Le tableau 10 présente les sorties des OU flous exclusifs et les vecteurs d'étiquettes $l(\mathbf{x})$ fournis par la règle RDAG présentée en tableau 9 pour les triplets Standard, de Hamacher et de Łukasiewicz sur le jeu ETIQ. Les seuils de rejet ont été fixés à $t_r = 0.5$ et $t_0 = 0.5$. Les sorties des OU flous intervenant dans l'étape H_2 ont été données en tableau 8. Les résultats sont conformes à l'intuition. Les isosurfaces du OU exclusif flou (impliqué dans H_1) et du OU flou (impliqué dans H_2) fondés sur le triplet de Łukasiewicz (21) pour le jeu de données C3D2 sont données à la figure 6. Les zones foncées correspondent à des valeurs élevées des OU.

Telle que H_2 est définie, il se peut qu'une seule étiquette $\mu_i(\mathbf{x})$ soit supérieure à t_0 . On a alors deux options possibles :

- forcer l'action d'affectation α_i , ou
- définir une action de rejet d'ambiguïté supplémentaire, action intermédiaire entre le classement exclusif et le rejet de distance que nous notons $\{\alpha_i, \alpha_d\}$.

Ces deux possibilités sont illustrées à la figure 7, respectivement à gauche et à droite, sur le jeu de données C3D2. Les seuils de rejet ont été fixés à $t_r = 0.25$ et $t_0 = 0.25$. Les régions particulières correspondant aux actions $\{\alpha_i, \alpha_d\}$ apparaissent en gris foncé. Cette deuxième option peut s'avérer intéressante en pratique puisqu'elle apporte une information plus fine. L'influence d'une pondération (31) sur les isosurfaces OU exclusif flou (impliqué dans H_1) et les frontières de décision est illustrée à la figure 8 dans le cas du triplet de Hamacher (26) avec des poids égaux à $w = (1/6 \ 3/6 \ 2/6)^t$. Les seuils de rejet ont été fixés à $t_r = 0.56$ et $t_0 = 0.24$. L'option consistant à mettre en œuvre des actions $\{\alpha_i, \alpha_d\}$ est présentée à droite. Évidemment, plus un poids w_j est élevé, comme w_2 ici, plus le OU exclusif flou joue en faveur de l'action exclusive correspondante α_j et étend la zone correspondante, α_2 ici.

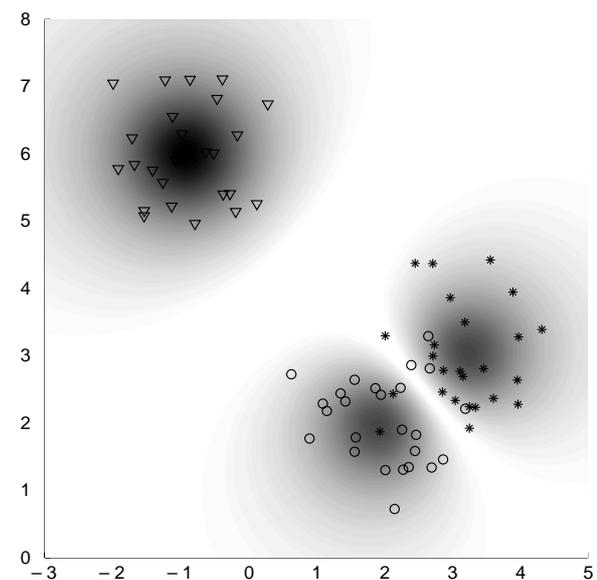
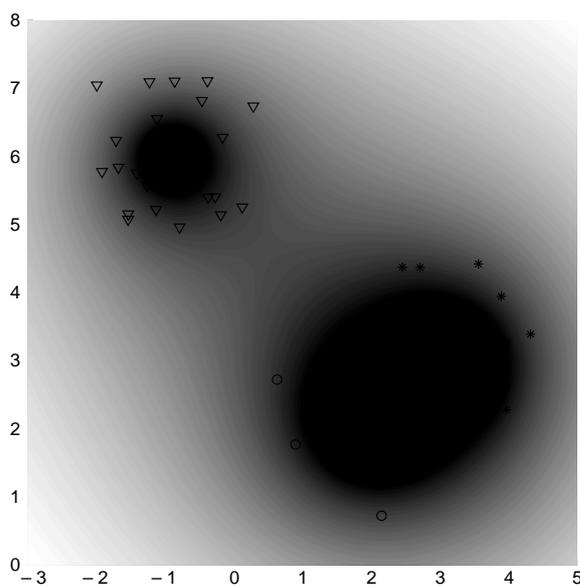


Figure 6. Isosurfaces du OU flou (gauche) et du OU exclusif flou (droite) de Łukasiewicz

Cette approche constitue une généralisation de la première règle fondée sur une stratégie *rejette d'abord* proposée dans la littérature [FMD95], [Fré98b]. On peut montrer que cette dernière est la règle RDAG avec les normes de Lukasiewicz dans le cas de $c = 2$ classes, en posant $\mu_{d,j} = t_a, (\forall j = 1, c)$.

3.4. Stratégie «Mélange d'abord»

Il y a trois types d'actions $(\alpha_i, \alpha_0, \alpha_d)$ associées à une règle de classement avec double option de rejet. Les stratégies précédentes constituent deux manières différentes de les combiner en deux étapes séquentielles. La troisième manière vient d'être proposée dans [SFM02]:

- | H_1 : Rejet d'ambiguïté ou non ?
- | H_2 : Sinon, classement exclusif ou rejet de distance ?

Là encore, l'étape caractérisant cette stratégie est H_1 . C'est pourquoi nous l'avons qualifiée de stratégie *mélange d'abord*. Si un point \mathbf{x} est rejeté en ambiguïté sélective lors de l'étape H_1 , l'étape H_2 n'est alors pas considérée. L'opérateur de combinaison des étiquettes $\mu_i(\mathbf{x})$ permettant de définir H_1 doit être capable de traduire la notion d'ambiguïté au sens du classement avec double option de rejet. Les contraintes associées sont :

- au moins deux classes concernées,
- étiquettes correspondantes élevées, sinon c'est le rejet de distance qui doit être mis en œuvre.

Les auteurs ont défini un opérateur d'étiquettes $\mu_i(\mathbf{x})$ traduisant bien cette situation, appelé OU-2 flou :

$$\bigoplus_{i=1,c}^2 \mu_i = \bigoplus_{i=1,c} (\mu_i \top \bigoplus_{j=1,c; j \neq i} \mu_j) \quad (32)$$

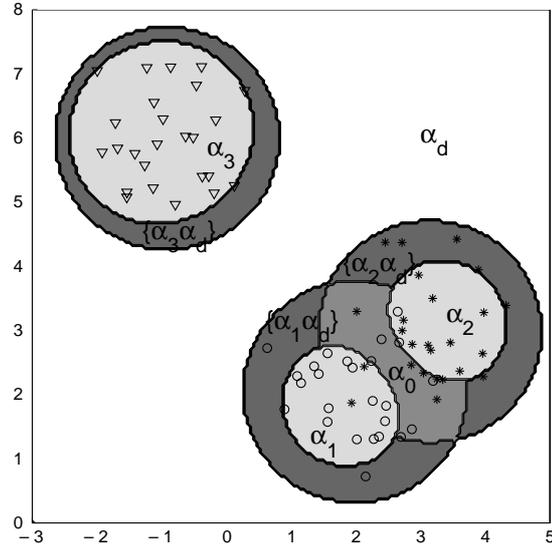
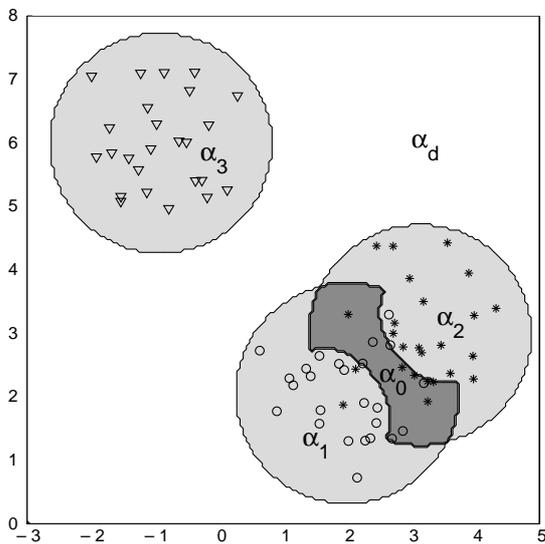


Figure 7. Régions de classement $(\omega_1 \circ, \omega_2 \star, \omega_3 \nabla)$; deux options

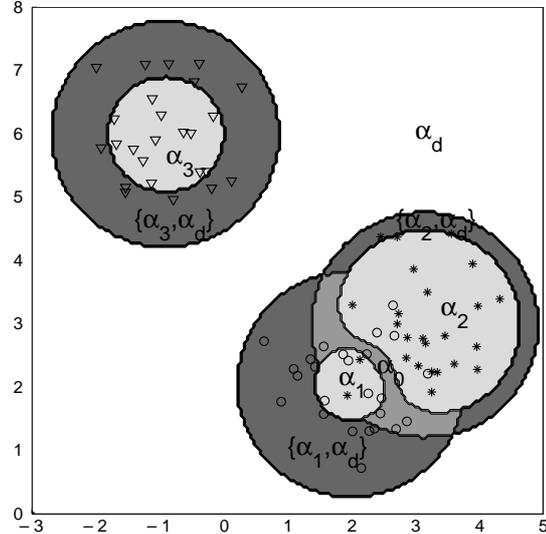
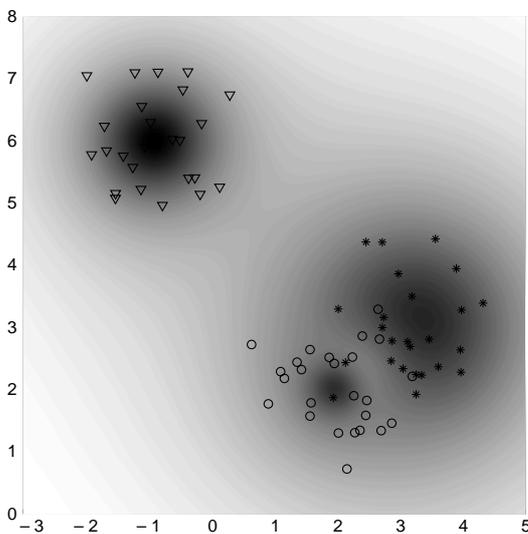


Figure 8. Isosurfaces d'un OU exclusif flou de Hamacher pondéré (gauche) et régions de classement $(\omega_1 \circ, \omega_2 \star, \omega_3 \nabla)$ (droite)

où (\top, \perp) est n'importe quel couple (t-norme, t-conorme). Dans le cas des normes standard *min* et *max*, il prend la valeur de la 2ème plus grande étiquette, d'où son nom.

Tout comme pour la stratégie précédente, les normes pondérées définies en (27-28) ne permettent pas de pondérer l'opérateur OU-2 flou. Frélicot et Mascarilla ont proposé de pondérer la conorme de la façon la plus simple permettant de retrouver (32) dans le cas équi pondéré [FM02]:

$$\bigoplus_{i=1,c}^2 \{\mu_i, w_i\} = \bigoplus_{i=1,c} \left(\left\{ \mu_i, 1 - \frac{c}{2} w_i \right\} \top \left\{ \bigoplus_{j=1,c; j \neq i} \mu_j, \frac{c}{2} w_i \right\} \right) \quad (33)$$

La figure 9 montre les isosurfaces de sortie d'un OU-2 flou fondé sur le couple (\top, \perp) de Hamacher dans le cas non pondéré à gauche et pondéré à droite avec $w = (0.1 \quad 0.9)^t$. Les zones claires traduisent une valeur élevée de $\mu_1 \perp \mu_2$, ce qui suggère d'appliquer l'action α_0 . Inversement, les zones sombres suggèrent plutôt les actions de classement exclusif α_1 ou α_2 ou celle de rejet de distance α_d . Un poids élevé, comme w_2 ici, doit jouer en faveur de l'action correspondante α_2 . En conséquence la zone correspondant à une valeur élevée du OU-2 flou, suggérant d'appliquer l'action α_0 , est déplacée vers des valeurs plus élevées de μ_1 et plus faibles de μ_2 . Il suffit donc de comparer $\bigoplus_{j=1,c}^2 \mu_j(\mathbf{x})$ à un seuil de rejet d'ambiguïté t_0 pour définir H_1 . Si \mathbf{x} doit être rejeté en ambiguïté, on peut, soit forcer toutes les étiquettes $l_j(\mathbf{x})$ à 1 et opérer un rejet pour ambiguïté totale, soit opérer un rejet pour ambiguïté sélective en comparant le vecteur $\mu(\mathbf{x})$ à un vecteur de seuils μ_0 . Si \mathbf{x} n'est pas rejeté pour ambiguïté sélective, il faut choisir l'une des deux autres actions possibles, à savoir α_i ou α_d . Plusieurs fonctions H_2 sont envisageables comme le seuillage de l'opérateur utilisé de la règle *Accepte D'Abord Généralisée* (ADAG) présentée en tableau 7, opérateur fondé sur la t-conorme correspondante, à savoir un

Tableau 11. Règle mélange d'abord généralisée (MDAG)

$L :$	$\mu_j(\mathbf{x})$
$H_1 :$	$l_j(\mathbf{x}) = (\mu_j(\mathbf{x}) \geq \mu_{0,j})$ si $\bigoplus_{j=1,c}^2 \mu_j(\mathbf{x}) \geq t_0 (\forall j = 1,c)$ éq. (32)
$H_2 :$	$i = \operatorname{argmax}_j \mu_j(\mathbf{x})$ $l_j(\mathbf{x}) = \begin{cases} \delta_{ij} & \text{si } \mu_i(\mathbf{x}) \geq t_d \\ 0 & \text{sinon} \end{cases} \quad (\forall j = 1,c)$

OU flou (29). Comme le max est la plus petite des t-conormes, nous proposons tout simplement de seuiller l'étiquette maximale $\mu_i(\mathbf{x})$. On obtient alors la famille de règles présentée en tableau 11, appelée *Mélange D'Abord Généralisée* (MDAG) [SFM02]. Si le rejet d'ambiguïté est sélectif, alors $l(\mathbf{x}) \in \mathcal{L}_{hc}^c$, sinon $l(\mathbf{x}) \in \mathcal{L}_{hc}^{01}$. En pratique, nous avons constaté que fixer les seuils $\mu_{0,j} = t_0$ ($\forall j = 1,c$) donne des résultats très satisfaisants. Les seuils de rejet t_0 et t_d permettent bien évidemment de contrôler les taux de rejet d'ambiguïté, de classement exclusif et de rejet distance. Plus on choisit une valeur t_0 petite, plus le rejet en ambiguïté est important, $\forall t_d$. Il passe même d'une ambiguïté à faible cardinalité (2 classes) à l'ambiguïté totale (c classes) si $t_0 = 0$. Pour une valeur fixée de t_0 , le taux de rejet pour ambiguïté sélective est constant et c'est t_d qui permet de contrôler les taux respectifs de rejet de distance et de classement exclusif. Plus t_d est choisi élevé, plus on rejette en distance et moins on classe de manière exclusive, et inversement. La règle se réduit à la règle du *Maximum d'Appartenance* (sans rejet) pour les valeurs limites $t_0 = 1$ et $t_d = 0$ de ses paramètres. À titre d'illustration, les sorties des OU-2 flous et les vecteurs d'étiquettes $l(\mathbf{x})$ fournis par la règle MDAG pour les triplets Standard, de Hamacher et de Łukasiewicz sur le jeu ETIQ sont présentés en tableau 12. Les seuils de rejet ont été fixés à $\mu_{0,j} = t_0 = 0.5$ et $t_d = 0.5$. Les vecteurs $l(\mathbf{x})$ sont, de manière évidente, conformes à l'intuition.

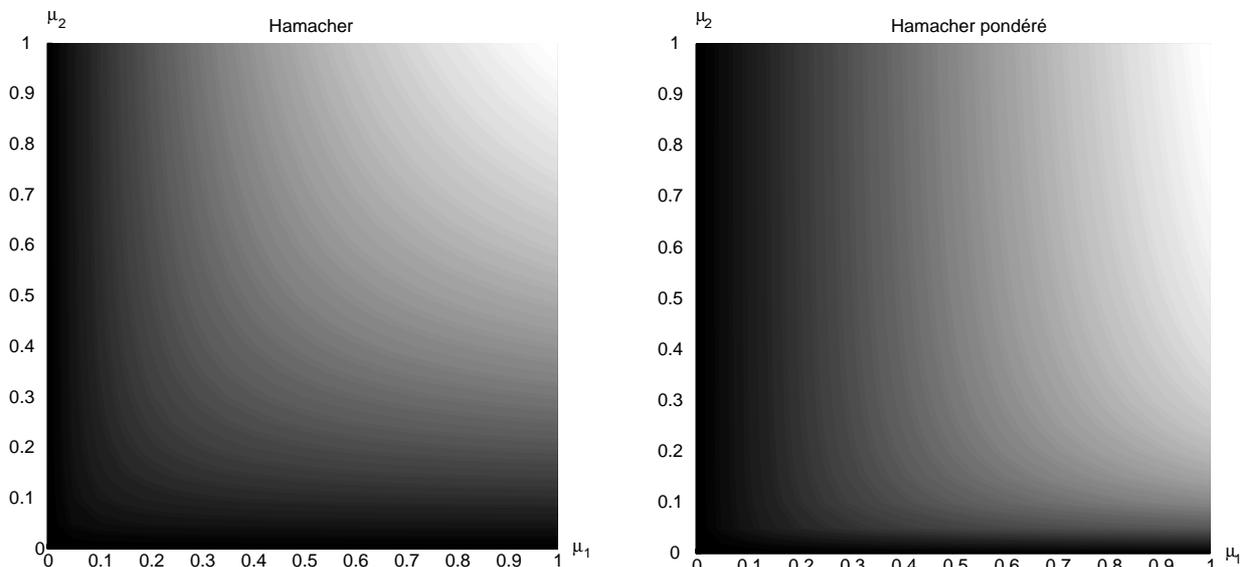


Figure 9. Isosurfaces d'un OU-2 flou de Hamacher non pondéré (gauche) et pondéré (droite)

Tableau 12. Exemples de classement par la règle mélange d'abord généralisée (MDAG)

$\mu(\mathbf{x})$	$\bigoplus_{i=1,c}^2 \mu_i(\mathbf{x})$	$l(\mathbf{x})$
$(0.1, 0.8, 0.7)^t$	Std 0.70 Łuk. 1.00 Ham. 0.75	$(0, 1, 1)^t$
$(0.1, 0.05, 0.2)^t$	Std 0.10 Łuk. 0.00 Ham. 0.16	$(0, 0, 0)^t$
$(0.1, 0.05, 0.7)^t$	Std 0.10 Łuk. 0.00 Ham. 0.21	$(0, 0, 1)^t$

La Figure 10 montre les isosurfaces du OU-2 flou fondé sur le couple de Hamacher (\top, \perp) non pondéré (26) à gauche et pondéré avec des poids égaux à $w = (1/6 \ 3/6 \ 2/6)^t$ pour le jeu de données C3D2. Un poids plus faible, comme w_1 ici, joue en faveur d'action correspondante α_1 . En conséquence, la zone où le OU-2 flou est élevé est décalée vers la classe ω_1 .

Tout comme pour la stratégie précédente, telle que H_1 est définie, il se peut qu'une seule étiquette $\mu_i(\mathbf{x})$ soit supérieure à $\mu_{0,i}$. On a alors deux options possibles :

- forcer l'action d'affectation α_i , ou
- définir une action de rejet d'ambiguïté supplémentaire, action intermédiaire entre rejet d'ambiguïté et classement exclusif que nous notons $\{\alpha_i, \alpha_0\}$.

Ces deux possibilités sont illustrées à la Figure 11, respectivement à gauche et à droite, sur le jeu de données C3D2 dans le cas du OU-2 flou de Hamacher non pondéré. Les seuils de rejet ont été fixés à $\mu_{0,j} = t_0 = t_d = 0.27$. Les régions particulières correspondant aux actions $\{\alpha_i, \alpha_d\}$ apparaissent en gris foncé. Là encore, cette deuxième option peut apporter une information plus fine en pratique.

Le cas pondéré est présenté à la Figure 12, avec des poids égaux à $w = (1/6 \ 3/6 \ 2/6)^t$. Le poids plus fort donné à ω_2 décale les zones d'ambiguïté vers ω_1 .

4. Conclusion

Dans cet article, nous avons défini des stratégies originales pour l'implantation des mécanismes de rejet de distance et d'ambiguïté en classification supervisée pour la reconnaissance des formes, et la définition de familles de règles de classement associées. Notre souci permanent a été de :

1. définir un cadre unifié pour la définition des règles de classement avec double option de rejet en séparant l'étiquetage du classement proprement dit,

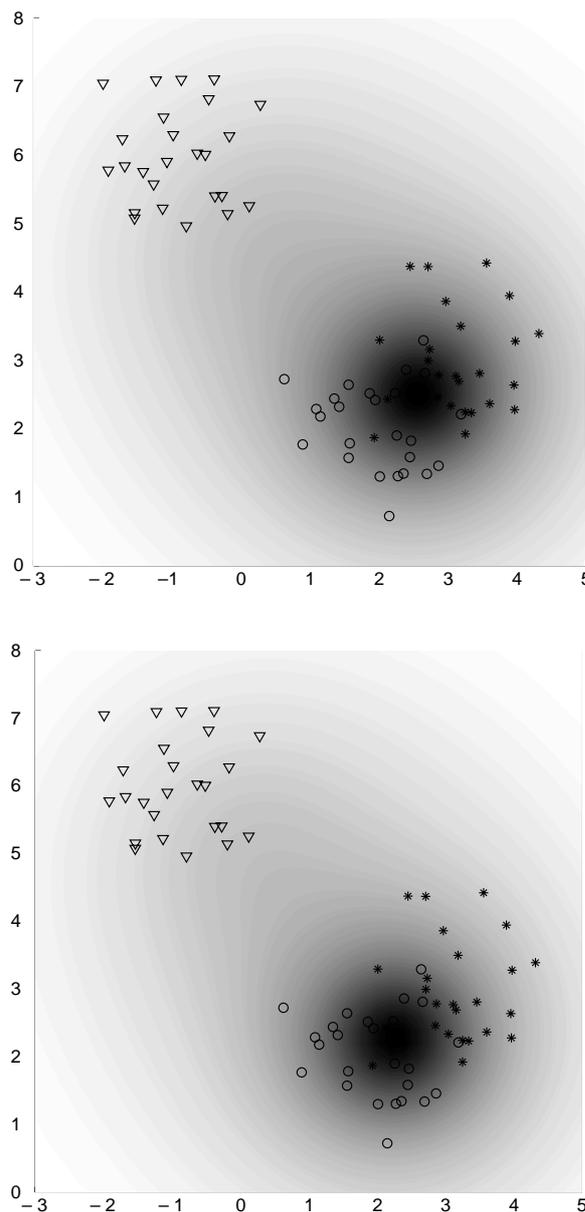


Figure 10. Isosurfaces du OU-2 flou de Hamacher non pondéré (haut) et pondéré (bas)

2. envisager toutes les stratégies possibles dans l'ordonnement des trois actions possibles que sont l'affectation exclusive, le rejet de distance et le rejet d'ambiguïté sélectif,
3. proposer des opérateurs de fusion d'étiquettes génériques et pondérables par l'intermédiaire de la combinaison de normes et conormes triangulaires,
4. proposer des règles de classement suivant les stratégies.

Sur ce dernier point, nous avons conçu des règles au comportement très satisfaisant vis-à-vis des frontières de classement, présentant deux caractéristiques que nous estimons fondamentales pour des règles avec options de rejet :

- rejet d'ambiguïté sélectif,
- réduction à une règle sans rejet (du type MAP) pour un réglage particulier de leurs paramètres (seuils).

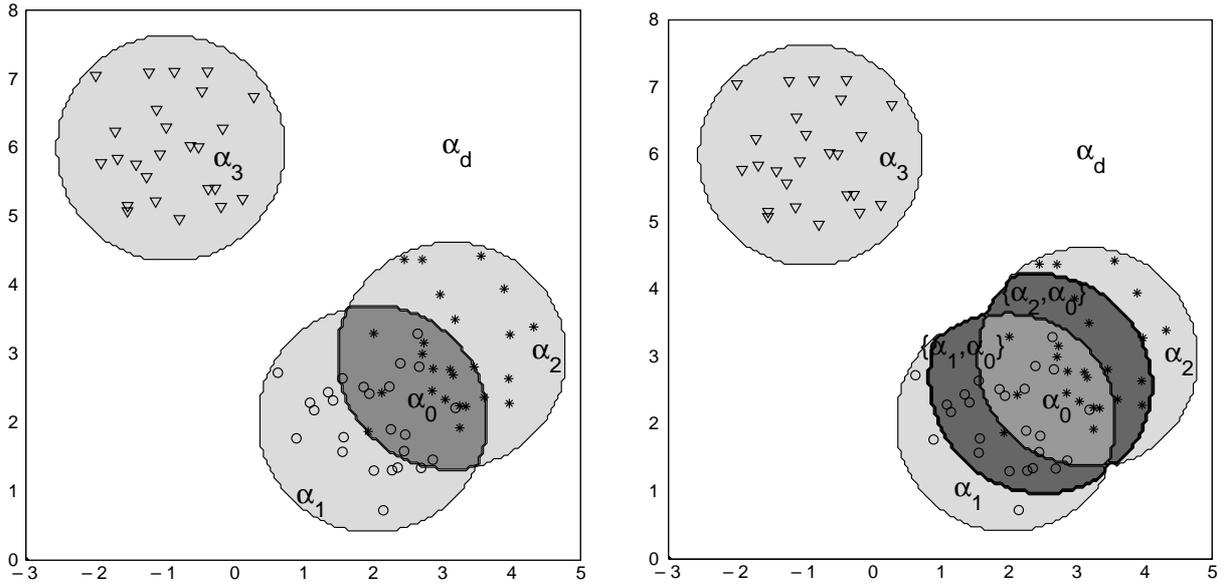


Figure 11. Régions de classement ($\omega_1 \circ, \omega_2 \star, \omega_3 \nabla$), Hamacher non pondéré

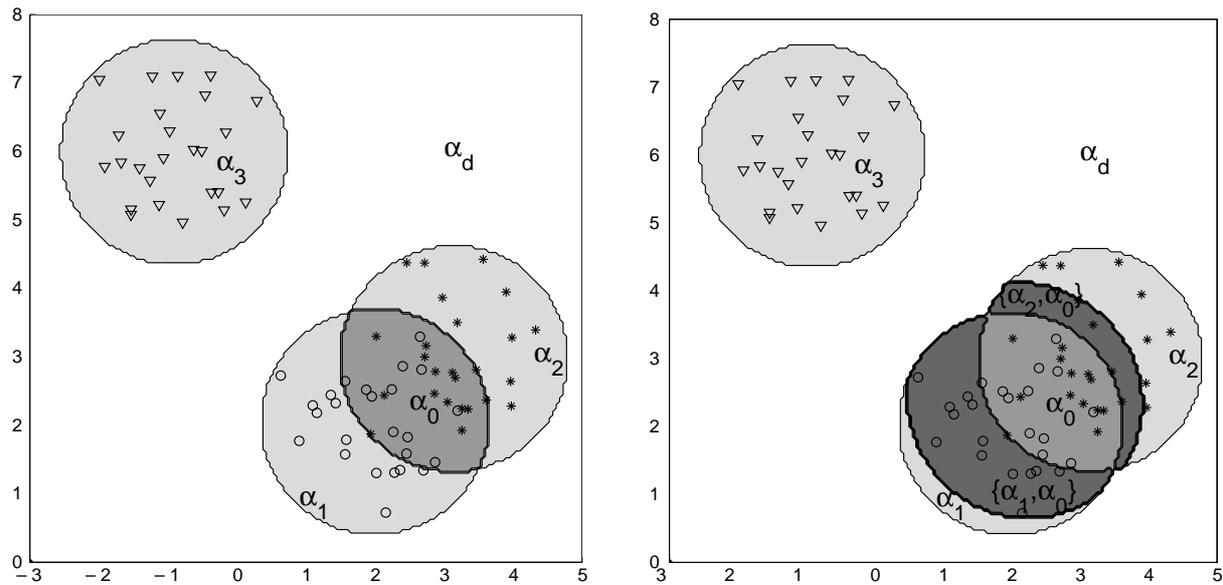


Figure 12. Régions de classement ($\omega_1 \circ, \omega_2 \star, \omega_3 \nabla$), Hamacher pondéré

Des problèmes ouverts demeurent. Tout d'abord, quelle stratégie utiliser, pour une application donnée, parmi les trois proposées ? Nous n'avons, pour l'heure, que des éléments de réponse. Il est clair que chacune des stratégies (accepte d'abord, rejette d'abord, mélange d'abord) étant focalisée sur une des trois actions possibles (rejet de distance, classement exclusif, rejet en ambi-güité), la nature des données à traiter (très mélangées ou non, contaminées par des données aberrantes ou non, etc) peut conduire à privilégier une des stratégies. Effectivement, si toutes les stratégies produisent finalement les trois mêmes types de régions de classement, ces régions ne sont pas équivalentes même si on utilise les mêmes triplets de De Morgan. Par exemple, l'opérateur OU-2 se focalisant sur le rejet d'ambiguïté

est plus efficace pour traiter des données très mélangées que ne le sont les fonctions H2 des règles ADAG ou RDAG, de même pour le rejet de distance et le classement exclusif avec les opérateurs OU exclusif et OU.

Ensuite, quels sont les triplets les mieux adaptés aux opérateurs de fusion proposés ? Le choix d'une t-norme parmi la douzaine de familles usuelles n'est pas un problème spécifique au classement mais beaucoup plus général qu'on rencontre entre autres dans les systèmes experts flous, les contrôleurs flous, l'agrégation et la décision multicritère. Nous pensons cependant qu'une étude mathématique des opérateurs de fusion et leurs propriétés, que nous avons commencée, peut apporter des réponses. En effet, chaque famille de normes a des propriétés intrinsèques

dont il serait intéressant d'étudier l'impact sur nos opérateurs, y compris selon que les étiquettes qu'elles combinent sont normalisées ou non.

Une dernière perspective concerne le choix des seuils qui, comme nous l'avons déjà dit, est un problème qui n'est pas lié aux approches proposées et par conséquent déborde largement du cadre de cet article. Néanmoins, nous envisageons d'étudier la liaison fonctionnelle qui peut exister entre les opérateurs de fusion d'étiquettes que nous avons proposés pour définir les différentes stratégies de rejet et les seuils de rejet qui sont des paramètres à régler. Les stratégies et les règles induites pourraient ainsi gagner en cohérence, ce qui permettrait de simplifier la tâche du praticien en guidant ses choix.

Références

- [Bez87] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Functions*, Plenum Press, New-York, NY, second edition, 1987.
- [Bis95] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
- [BM98] C.L. Blake and C.J. Merz, UCI repository of machine learning databases, 1998, Dept. of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [BRLA96] J.C. Bezdek, T.R. Reichherzer, G. Lim and Y. Attikiouzel, Classification with multiple prototypes, In: *5th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, volume 1, pages 626-632, New-Orleans, Louisiana, August, 1996.
- [Cho57] C.K. Chow, An optimum character recognition system using decision functions, *IRE Trans. on Electronic Computers*, EC-6, pp. 247-253, December, 1957.
- [Cho70] C.K. Chow, On optimum recognition error and reject tradeoff, *IEEE Trans. on Information Theory*, IT-16, pp.41-46, 1970.
- [Dav91] R. Davé, Characterization and detection of noise in clustering, *Pattern Recognition*, 12(11), pp. 657-664, 1991.
- [Det00] M. Detyniecki, *Mathematical aggregation operators and their application to video querying*, PhD thesis, Université de Paris 6, November, 2000.
- [DH73] R. Duda and P. Hart, *Pattern Recognition and Scene Analysis*, John Wiley and Sons, 1973.
- [DM93] B. Dubuisson and M-H. Masson, A statistical decision rule with incomplete knowledge about classes, *Pattern Recognition*, 26(1), pp. 155-165, 1993.
- [DP85] D. Dubois and H. Prade, A review of fuzzy sets aggregation connectives, *Information Sciences*, 36, pp. 85-121, 1985.
- [DP88] D. Dubois and H. Prade, *Théorie des possibilités – Application à la représentation des connaissances en informatique*, Masson, Paris, 1988.
- [Dub01] B. Dubuisson, *Diagnostic, intelligence artificielle et reconnaissance des formes*, Traité IC2 – Information, Commande, Communication, Hermès, Paris, ISBN 2-7462-0249-2, 2001.
- [FD92] C. Frélicot and B. Dubuisson, A multi-step predictor of membership function as an ambiguity reject solver in pattern recognition, In: *4th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, pp. 709-715, Palma de Mallorca, Spain, July, 1992.
- [FM02] C. Frélicot and 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)*, pp. -, New-Orleans, Louisiana, June 25-27, 2002.
- [FMD95] C. Frélicot, M-H. Masson and B. Dubuisson, Reject options in fuzzy pattern classification rules, In: *3rd European Congress on Intelligent Techniques and Soft Computing (EUFIT)*, volume 3, pp. 1459-1464, Aachen, Germany, June, 1995.
- [FR99] G. Fumera and F. Roli, Multiple reject thresholds for improving classification reliability, Technical report, University of Cagliari, pp. 1-13, 1999.
- [Fré97] C. Frélicot, Learning rejection thresholds for a class of fuzzy classifiers from possibilistic clustered noisy data, In: *7th Int. Fuzzy Systems Association World Congress (IFSA)*, volume 3, pp. 111-116, Prague, Czech Republic, June, 1997.
- [Fré98a] C. Frélicot, *On unifying probabilistic/fuzzy and possibilistic rejection-based classifiers*, pp. 736-745, In: *Lecture Notes in Computer Science 1451: Advances in Pattern Recognition* (Amin A., Dori D., Pudil P., Freeman H., Eds), Springer-Verlag, Berlin, 1998.
- [Fré98b] C. Frélicot, A rejection-based possibilistic classifier and its para-meters learning, In: *7th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, pp. 1423-1428, Anchorage, Alaska, May, 1998.
- [FRG00] G. Fumera, F. Roli and G. Giacinto, , Reject options with multiple thresholds, *Pattern Recognition*, 33(12), pp. 2099-2101, 2000.
- [Fuk90] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Boston, MA, second edition, 1990.
- [Ha96] T-M. Ha, An optimum class-selective rejection rule for pattern recognition, In: *13th Int. Conf. on Pattern Recognition (ICPR)*, volume 2, pp. 75-80, Vienna, Austria, August 25-30, 1996.
- [Ha96] T-M. Ha, The optimum class-selective rejection rule, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-19(6), pp. 608-615, June, 1997.
- [JDM00] A.K. Jain, R.P.W. Duin and J. Mao, Statistical pattern recognition: a review, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-22(1), pp. 4-37, January, 2000.
- [KK93] R. Krishnapuram and J.M. Keller, A possibilistic approach to clustering, *IEEE Trans. on Fuzzy Systems*, TFS-1, pp. 98-110, 1993.
- [KY95] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, NJ, USA, 1995.
- [MDF96] M-H. Masson, B. Dubuisson and C. Frélicot, Conception de systèmes de diagnostic par reconnaissance des formes floues, *Journal Européen des Systèmes Automatisés*, 30, pp. 319-341, 1996.
- [MF01a] L. Mascarilla and 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)*, pp. 743-747, Vancouver, Canada, July, 2001.
- [MF01b] L. Mascarilla and C. Frélicot, Une famille de discriminateurs possibilistes avec rejets fondée sur des triplets de De Morgan, In: *11^{èmes} Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, pp. 31-38, Mons, Belgium, Novembre, 2001.
- [MYP98] R. Muzzolini, Y-H. Yang and R. Pierson, Classifier design with incomplete knowledge, *Pattern Recognition*, 31(4), pp. 345-369, 1998.
- [Pal91] S.K. Pal, Fuzzy tools in the management of uncertainty in pattern recognition, image analysis, vision and expert systems, *International Journal Of Systems Sciences*, 22(3), pp. 511-549, 1991.
- [PM77] S.K. Pal and D. Majumber, Fuzzy sets and decision making approaches in vowel and speaker recognition, *IEEE Trans. on Systems, Man and Cybernetics*, SMC-7, pp.625-629, 1977.
- [Rus69] E. Ruspini, A New approach to clustering, *Information and Control*, 8, pp. 338-353, 1969.

[SFM02] D. Semani, C. Frélicot and L. Mascarilla, Discrimination possibiliste avec options de rejet: une nouvelle approche, In: *12^{èmes} Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, pp. 27-34, Montpellier, France, Octobre, 2002.

[Smy94] P. Smyth, Markov monitoring with unknown states, *IEEE Journal on Selected Areas in Communications*, 12(9), pp. 1600-1612, 1994.



Carl Frélicot

Carl Frélicot a obtenu un titre d'ingénieur en Informatique et un Doctorat en Contrôle des Systèmes de l'Université de Technologie de Compiègne, en 1988 et 1992 respectivement. Depuis 1993, il est Maître de Conférences au Département Informatique de l'Université de La Rochelle. Habilité à Diriger des Recherches en 2002, il anime la thématique Données-Formes-Interprétation du Laboratoire d'Informatique, Image, Interaction (L3I). Ses recherches portent sur différents aspects de la Reconnaissance des Formes, en particulier la classification avec rejet par des méthodes statistiques, prétopologiques, floues ou possibilistes.



Laurent Mascarilla

Laurent Mascarilla a obtenu son doctorat en Informatique à l'Université Paul Sabatier (Toulouse III) en 1996. Il est, depuis 1997, Maître de Conférences au Département Informatique de l'Université de La Rochelle, et membre du L3I (Laboratoire Informatique, Image, Interaction). Ses recherches portent sur la Reconnaissance des Formes et les applications de la logique floue au Traitement d'Images.

