

ARNAUD DEVILLEZ

PATRICE BILLAUDEL

GÉRARD VILLERMAIN LECOLIER

**Une méthode possibiliste de discrimination adaptée aux classes de forme complexe**

*Mathématiques et sciences humaines*, tome 147 (1999), p. 71-85

[http://www.numdam.org/item?id=MSH\\_1999\\_\\_147\\_\\_71\\_0](http://www.numdam.org/item?id=MSH_1999__147__71_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1999, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## UNE MÉTHODE POSSIBILISTE DE DISCRIMINATION ADAPTÉE AUX CLASSES DE FORME COMPLEXE

Arnaud DEVILLEZ<sup>1</sup>, Patrice BILLAUDEL<sup>1</sup>, Gérard VILLERMAIN LECOLIER<sup>1</sup>

**RÉSUMÉ** – *Notre équipe travaille sur la classification de données provenant des secteurs industriels et médicaux, dans le but de développer des systèmes de diagnostic et d'aide à la décision. Dans cet article, nous proposons une modification de la méthode floue du «pattern matching», pour classer des données comportant des classes de forme complexe. Nous décrivons la méthode de base avant de montrer ses limites, lorsque les classes ne sont pas convexes. Ensuite, nous en proposons une amélioration par l'introduction d'une approche multiprototype. Nous présentons un exemple industriel, qui consiste à trier automatiquement des bouteilles plastiques en vue de leur recyclage. Enfin, nous comparons les résultats obtenus par cette méthode avec ceux donnés par la méthode floue des k-plus proches voisins, sur trois types de données : bouteilles plastiques, iris et formes d'ondes.*

**MOTS-CLÉS** – Classification, Logique Floue, Théorie des possibilités, Fuzzy Pattern Matching, Décision multicritère.

**SUMMARY** – *A possibilistic discrimination method adapted to classes of complex shape. Our team works on the classification of data coming from industrial and medical sectors, in order to develop decision making and diagnosis systems. In this paper we propose to modify the fuzzy method of pattern matching, in order to classify data including classes of complex shape. We describe the basic method before showing its limits when classes are not convex. Then, we propose to improve the method by introducing a multiprototype approach. We present an industrial example, which consists in sorting automatically plastic bottles in order to recycle them. Finally, we compare the results obtained by this method with those given by the fuzzy k-nearest neighbours method, using three types of data : plastic bottles, iris and waveform data.*

**KEYWORDS** – Classification, Fuzzy Logic, Possibility theory, Fuzzy Pattern Matching, Multicriteria decision.

---

<sup>1</sup> Laboratoire d'Automatique et de Micro-électronique, Faculté des Sciences - Moulin de la Housse, BP 1039, 51687 Reims Cedex 2, e-mail : [arnaud.devillez@univ-reims.fr](mailto:arnaud.devillez@univ-reims.fr), [patrice.billauDEL@univ-reims.fr](mailto:patrice.billauDEL@univ-reims.fr), [gerard.villermain-lecolier@univ-reims.fr](mailto:gerard.villermain-lecolier@univ-reims.fr)

## INTRODUCTION

Les problèmes de diagnostic sont couramment rencontrés dans les domaines industriels et médicaux. Face à une situation donnée, l'opérateur, l'ingénieur et le médecin ont pour mission de prendre une décision suivant plusieurs critères, en faisant appel à toute leur expérience. Pour cela, ils peuvent soit avoir recours à un système d'aide à la décision que l'on rencontre souvent dans le secteur médical, soit utiliser des méthodes automatiques de décision comme les superviseurs de processus industriels.

Les objectifs de notre équipe sont le développement et la mise au point des systèmes d'aide à la décision et de supervision de processus industriels. Nous nous sommes orientés vers les méthodes de classification supervisée en supposant qu'une décision est prise par rapport à une situation déjà vécue, c'est à dire une situation classée dans l'ensemble des événements connus, que l'on appelle l'expérience. Le principe de ces méthodes est d'affecter un élément à une classe, en se référant à des éléments d'apprentissage, dont la classe est connue [7]. Ces méthodes, présentées puis comparées dans [11], comportent une phase d'apprentissage puis une phase de classement au cours de laquelle la décision d'affecter un objet à une classe est prise.

Considérant que cette décision comporte une part d'imprécision et d'incertitude, tant au niveau des critères qui permettent de la prendre, que des conséquences qu'elle engendre, nous avons choisi d'utiliser les techniques floues et possibilistes. Les méthodes retenues sont la méthode floue des  $k$ -plus proches voisins [17], la méthode des règles floues distribuées [15, 16] et la méthode floue du pattern matching [12]. Une étude comparative [2] nous a permis de retenir la méthode floue du pattern matching pour nos applications. Elle donne les meilleures performances pour les temps de calcul les plus courts, tant que les classes sont parfaitement séparables, que leur forme est convexe et qu'elles sont orientées parallèlement aux axes de représentation. Nous nous sommes aperçus que, si les nuages comportent des classes de formes non convexes ou obliques, c'est à dire des classes dont l'axe principal n'est pas parallèle à un des axes du repère, les résultats sont nettement moins satisfaisants.

L'objet de cet article est l'adaptation de la méthode floue du pattern matching aux nuages comportant des classes de forme quelconque. Après une présentation de la méthode de base, nous montrons les limites de l'algorithme classique. Puis nous proposons une amélioration basée sur une approche multiprototype, permettant de résoudre les problèmes dus à la forme des classes. Enfin, nous comparons les résultats que nous avons obtenus avec ceux d'autres méthodes en utilisant les données provenant d'une étude sur le tri des bouteilles plastiques, les données iris et les données de formes d'ondes.

## 1. METHODE FLOUE DU PATTERN MATCHING

### 1.1. PROBLEMATIQUE ET NOTATIONS

Les données sont décrites par  $a$  caractères et l'ensemble d'apprentissage comporte  $n$  points. La méthode floue du pattern matching, présentée dans [12, 18], utilise la définition de  $c$  profils ( $P^1, \dots, P^c$ ) pour les  $c$  classes. Chaque profil est représenté sur

chaque caractère par un sous-ensemble flou qui exprime la plage des valeurs typiques de ce caractère. Dans notre cas, il s'agit des densités de possibilité estimées et notées  $\pi^{ik}$  pour la classe  $C^i$  et le  $k^{\text{ème}}$  caractère. Le problème consiste à affecter un point  $x$  de l'espace de représentation  $\mathbf{R}^a$ , dont les valeurs pour les différents caractères sont  $(x^1, \dots, x^a)$ , à une des classes.

## 1.2. ESTIMATION DES DENSITES DE POSSIBILITES

Les profils sont les densités de possibilité estimées pour chaque caractère et chaque classe [23, 9]. Leur construction est basée sur la théorie des possibilités, introduite par Zadeh en 1978, pour prendre en compte l'incertitude et l'imprécision dans la description linguistique des caractères, des événements ou des situations [23]. Dans la méthode floue du pattern matching, l'apprentissage consiste à construire ces densités, pour chaque caractère et pour chaque classe, à partir des échantillons de référence. Les distributions de possibilité peuvent être calculées à partir des distributions de probabilité, qui sont elles mêmes estimées à partir des histogrammes des données.

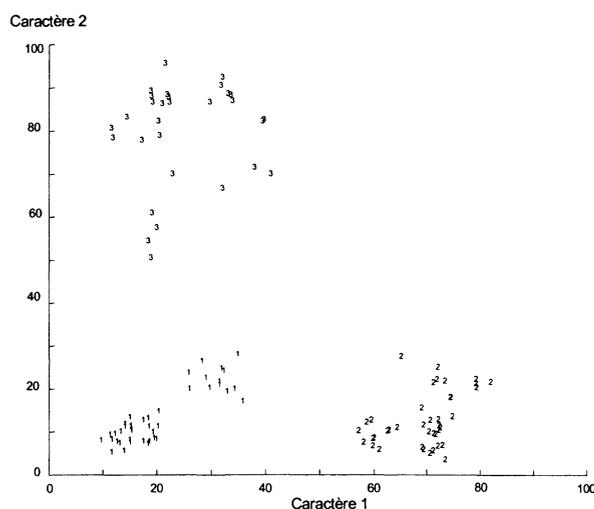


Figure 1. Exemple de nuages de points à trois classes

Les histogrammes des données sont établis à partir de l'ensemble d'apprentissage. Les bornes inférieures et supérieures de l'histogramme d'une classe  $C^i$  suivant un caractère, sont généralement les valeurs minimales et maximales des composantes des points d'apprentissage de la classe  $C^i$ , suivant ce caractère. Toutefois il arrive que ces bornes soient déterminées manuellement par l'utilisateur. L'intérêt est d'élargir la base des histogrammes de manière à agrandir le champ d'action du classifieur. Le nombre de barres  $h$  des histogrammes est un paramètre de la méthode. Nous discuterons son influence sur les performances dans la partie 1.4. Il permet de calculer la largeur des intervalles et de déterminer les bornes intermédiaires des histogrammes. La hauteur de chaque barre est l'effectif des points d'apprentissage de la classe  $C^i$ , situés dans l'intervalle considéré. L'exemple, présenté figure 1, est un nuage à trois classes dans un espace de dimension 2. L'apprentissage requière le calcul de six histogrammes. A titre d'illustration, nous avons représenté, sur la figure 2, les trois histogrammes calculés pour le caractère 1. Le passage des histogrammes aux distributions de probabilité nécessite deux transformations. Chaque barre est tout d'abord représentée par le centre

de l'intervalle noté  $y_i$ . Puis, le rapport entre la hauteur de la barre et le nombre total de points d'apprentissage, noté  $p(y_i)$ , détermine la probabilité associée au centre de l'intervalle.

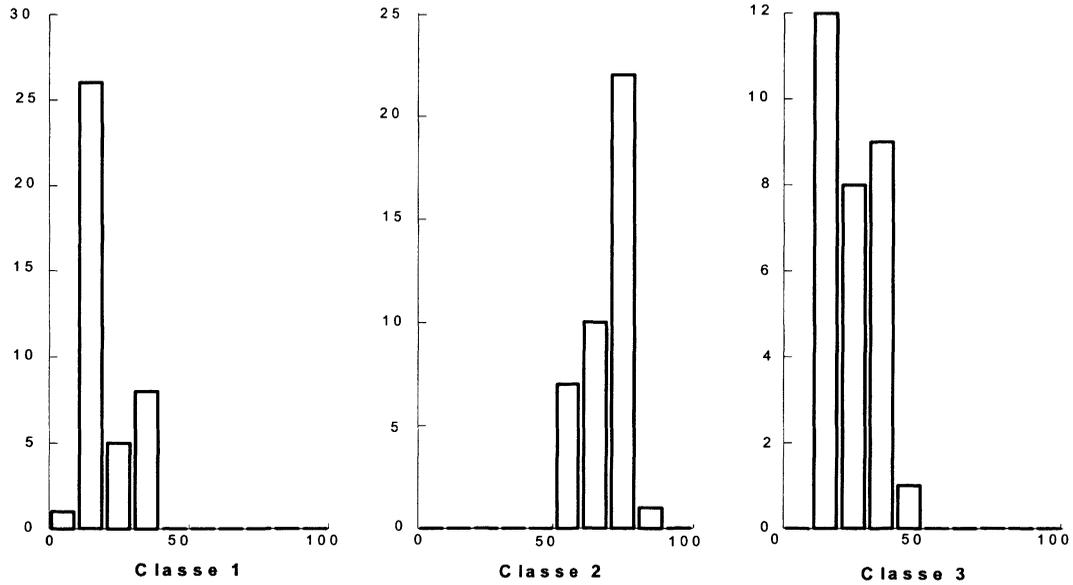


Figure 2. *Histogrammes des données pour le caractère 1 et pour chaque classe*

Les valeurs de probabilités  $\{p(y_i)|i=1, \dots, h\}$  des centres d'intervalles  $y_i$  sont classées dans l'ordre décroissant  $p(y_1) \geq p(y_2) \geq p(y_3) \geq p(y_4) \geq \dots \geq p(y_h)$ . La distribution de possibilité  $\{\pi(y_i)|i=1, \dots, h\}$  est déduite de la distribution de probabilité par la transformation bijective introduite par Dubois et Prade [9] :

$$\pi(y_i) = \sum_{j=1}^h \min[p(y_i), p(y_j)] = i \cdot p(y_i) + \sum_{j=i+1}^h p(y_j).$$

Cette expression vérifie l'hypothèse de la théorie des possibilités  $\pi(y_1)=1$ . La relation inverse s'écrit :

$$p(y_i) = \sum_{j=i}^h \frac{1}{h} [\pi(y_j) - \pi(y_{j+1})], \text{ où } \pi(y_1) \geq \pi(y_2) \geq \dots \geq \pi(y_h) \geq 0,$$

avec la convention  $\pi(y_{h+1}) = 0$ .

Le passage à la densité de possibilité est réalisé par une interpolation linéaire de la distribution. Les densités possibilistes pour chaque classe et chaque caractère, relatives à notre exemple, sont présentées figure 3.

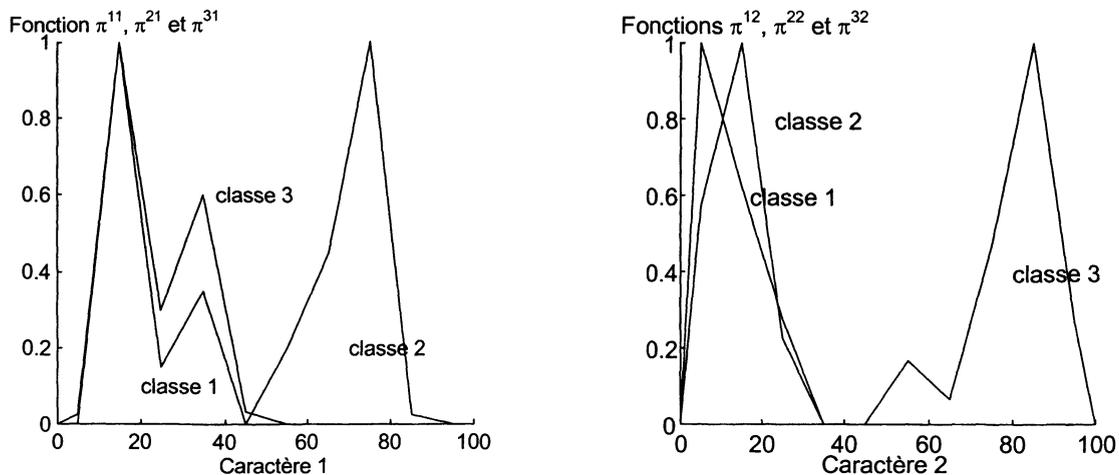


Figure 3. Profils des classes 1, 2 et 3 suivant les 2 caractères

1.3. REGLE DE CLASSEMENT

Le classement d'un nouveau point  $x$ , dont les valeurs des différents caractères sont  $(x^1, \dots, x^a)$ , s'effectue en deux étapes :

- détermination de la similitude entre  $x^k$  et les différents profils  $P^i$ . Le degré de similitude  $\pi^{ik}(x^k)$  entre le point et la représentation du profil  $P^i$  sur le  $k^{\text{ème}}$  caractère est obtenu par lecture du profil pour la valeur  $x^k$ .
- fusion, pour chaque profil, de tous les degrés de similitude  $(\pi^{i1}(x^1), \dots, \pi^{ia}(x^a))$  par un opérateur  $H$  :

$$v^i(x) = H[\pi^{i1}(x^1), \dots, \pi^{ia}(x^a)].$$

Le résultat  $v^i(x)$  de cette fusion représente le degré de ressemblance entre le point  $x$  et le profil  $P^i$ . Dans la méthode floue du pattern matching, cette valeur est considérée comme la possibilité d'appartenance du point  $x$  à la classe  $C^i$ . L'opérateur de fusion peut être un produit, un minimum, une moyenne, une intégrale floue [4, 12, 13], ou encore l'opérateur OWA [21, 22]. Le point est affecté à la classe pour laquelle il a la plus forte possibilité d'appartenance.

1.4. INFLUENCE DU PARAMETRE  $h$

Le paramètre  $h$  désignant le nombre de barres de l'histogramme a une forte influence sur les performances de la méthode. Il contrôle la taille de l'intervalle de discrétisation et il est indépendant du nombre de classes. Toutefois, sa valeur est liée aux effectifs des classes d'apprentissage. Si elle est trop élevée, il est possible que le nombre de points présents dans certains intervalles soit nul. Un "trou" apparaît alors dans les densités de possibilité, provoquant une diminution des valeurs d'appartenance, même si cet intervalle est situé au centre d'une classe. Le classement d'un point apparaissant dans cette zone sera par conséquent erroné. Si sa valeur est trop faible, la répartition des points à l'intérieur des classes n'est pas respectée puisque les densités de possibilité sont nivelées. La détermination du paramètre  $h$  est semblable à la détermination de la

dimension de la fenêtre de lissage dans la méthode des noyaux [6]. L'étude de la méthode floue du pattern matching a montré qu'une valeur comprise entre 5 et 15, donnait souvent les performances optimales, quelles que soient les données à classer [2].

## 2. LIMITES DE LA METHODE CLASSIQUE

Pour illustrer les limites de la méthode, nous nous sommes intéressés aux classes de formes non convexe et oblique, telles que celles présentées sur les figures 4a et 5a. Les densités de possibilités calculées pour ces nuages sont présentées sur les figures 4b et 5b.

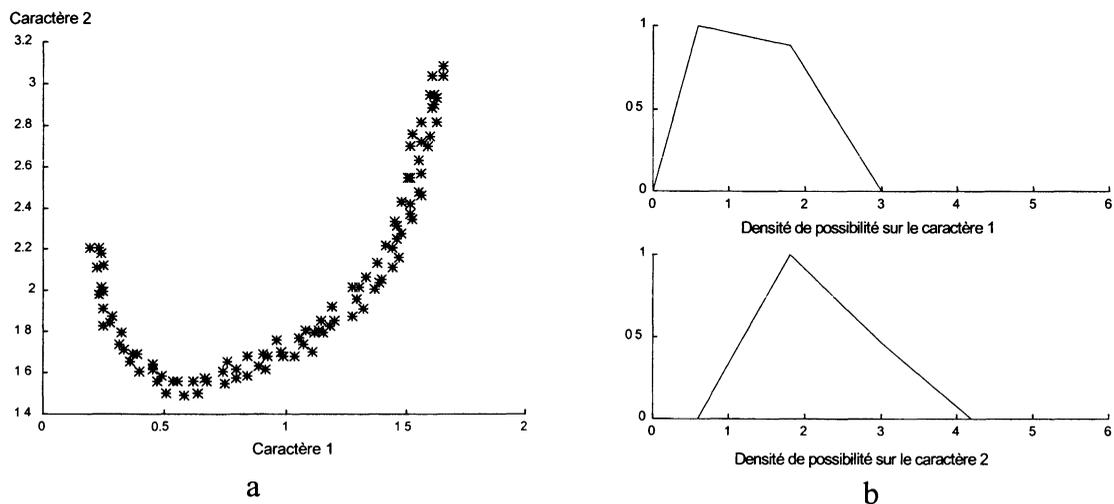


Figure 4. *Classe non convexe et densités de possibilité*

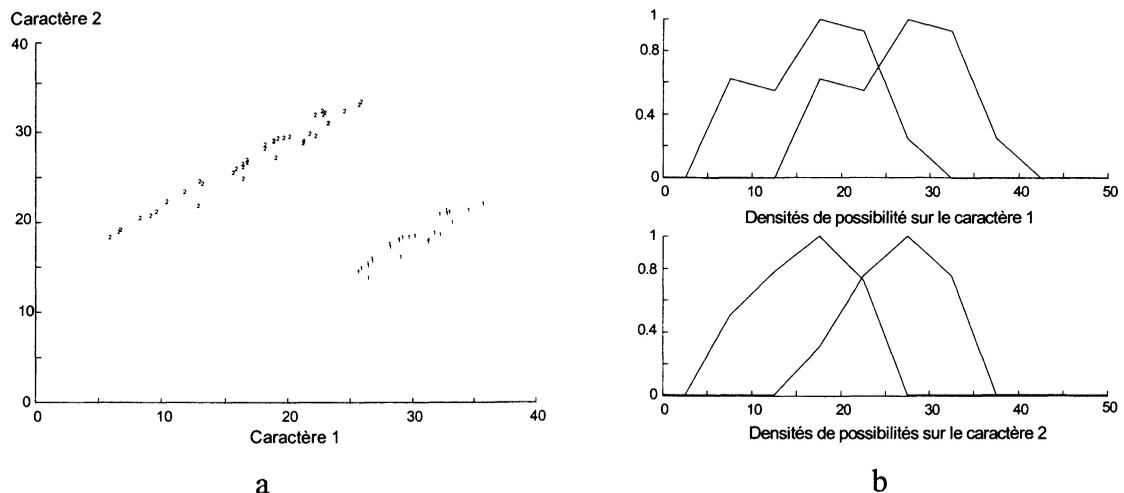


Figure 5. *Classes obliques et densités de possibilité*

La figure 6, représentant les courbes de niveaux d'appartenance, permet de visualiser le pouvoir de discrimination de la méthode. Elle montre que la méthode floue du pattern matching, associée à des opérateurs d'agrégation simples comme le minimum, ne fonctionne pas pour ces deux types de classes très caractéristiques. En effet, les courbes de niveaux d'appartenance ne respectent pas la forme des classes. Ce phénomène

provient de l'estimation, caractère par caractère, des densités de possibilité. Il engendre des problèmes de classement puisqu'un point situé dans la concavité de la parabole est affecté à la classe alors qu'il ne lui appartient pas. Dans le cas des classes obliques présentées figure 6b, nous assistons à un recouvrement des courbes de niveaux d'appartenance des deux classes, qui provoque des erreurs de classement. L'utilisation de la méthode floue du pattern matching est impossible pour des classes de formes non convexes et obliques, lorsqu'elle s'appuie sur une simple représentation des classes. C'est pour cette raison que nous proposons une amélioration basée sur une approche multiprototype.

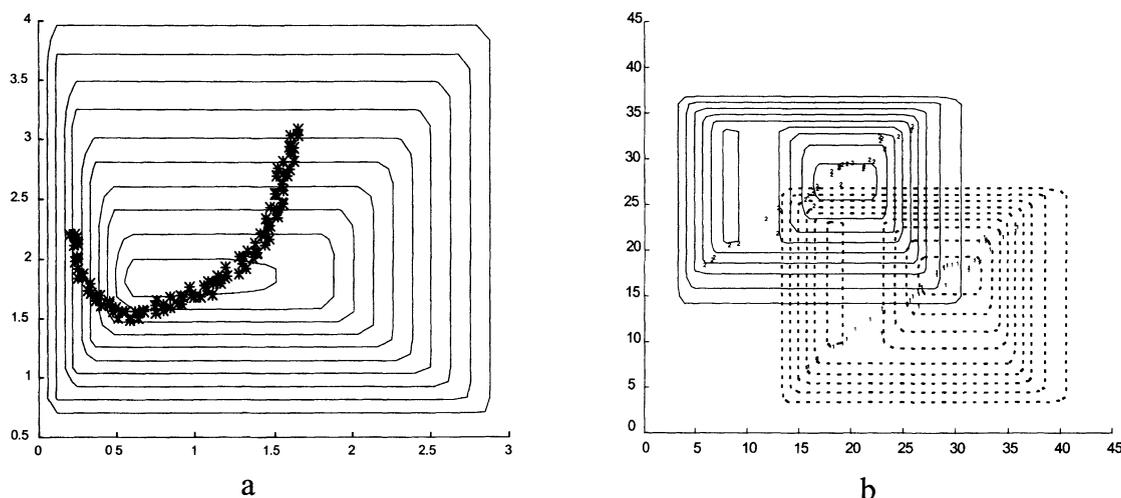


Figure 6. Courbes de niveaux d'appartenance obtenues en utilisant l'opérateur minimum

### 3. ADAPTATION DE LA METHODE

#### 3.1. APPROCHE MULTIPROTOTYPE

Chaque classe  $C^i$ , comportant  $n_i$  points d'apprentissage, est partagée en  $s_i$  sous-classes  $C_j^i$  par l'algorithme des  $c$ -moyennes floues [1]. Chacune de ces sous-classes est représentée par un prototype : son centroïde  $g_j^i$ . Une classe  $C^i$  est par conséquent représentée par  $s_i$  prototypes. En cela notre méthode est comparable à la méthode des  $k$ -plus proches voisins condensés, qui sélectionne les prototypes parmi les points présents dans les zones de forte densité de la classe [10]. Pour chaque point de la classe  $C^i$ , l'algorithme des  $c$ -moyennes floues délivre un coefficient d'appartenance à chaque sous-classe. Ces coefficients composent la matrice d'appartenance  $U^i$  de dimension  $s_i \times n_i$  où l'élément  $u_{jk}^i$  représente le coefficient d'appartenance du point  $x_k$  de la classe  $C^i$ , à la sous-classe  $C_j^i$ . Le lecteur trouvera les expressions du coefficient  $u_{jk}^i$  et du centroïde  $g_j^i$  dans l'ouvrage de Bezdek, consacré aux méthodes floues de coalescence [1].

Pour chaque point  $x$  à classer, une fonction exponentielle pour chaque sous-classe  $C_j^i$  est calculée selon l'expression suivante [19] :

$$\mu_j^i(x) = \exp[-\lambda_j^i \cdot d(x, g_j^i)].$$

L'expression de la fonction exponentielle est semblable à celle du noyau gaussien utilisé dans la méthode des noyaux de Parzen [7]. La notion d'appartenance graduée, issue de la théorie des sous-ensembles flous, apparaît dans l'expression de la distance.  $\lambda_j^i$  est une constante relative à la sous-classe  $C_j^i$  et  $d(x, g_j^i)$  représente la distance de Mahalanobis floue entre le point  $x$  et le centroïde de la sous-classe  $C_j^i$  dans l'espace  $\mathbf{R}^a$ . Elle utilise la matrice de covariance floue et est définie par l'expression suivante [14] :

$$d(x, g_j^i) = (x - g_j^i)^t \cdot \Sigma_j^{i-1} \cdot (x - g_j^i).$$

La connaissance des coefficients d'appartenance des points d'apprentissage permet le calcul de la matrice de covariance floue  $\Sigma_j^i$  de chaque sous-classe  $C_j^i$ , dont l'expression est :

$$\Sigma_j^i = \frac{\mathbf{S}_j^i}{\sum_{k=1}^{n_i} (u_{jk}^i)^2}.$$

$\mathbf{S}_j^i$  est la matrice de dispersion floue définie par :

$$\mathbf{S}_j^i = \sum_{k=1}^{n_i} (u_{jk}^i)^2 \cdot (x_k - g_j^i) \cdot (x_k - g_j^i)^t.$$

L'introduction de cette matrice dans le calcul de la distance permet de respecter la forme elliptique des sous-classes, en tenant compte de la répartition des niveaux d'appartenance. Pour chaque classe  $C^i$ , nous déterminons une fonction d'appartenance globale par la fusion  $F$  des  $s_i$  fonctions exponentielles, définies sur les  $s_i$  sous-classes  $(C_1^i, \dots, C_{s_i}^i)$  [18] :

$$\mu^i(x) = F(\mu_1^i, \dots, \mu_{s_i}^i) = \min \left[ 1, \sum_{j=1}^{s_i} \mu_j^i(x) \right]. \quad (1)$$

### 3.2. AGREGATION GLOBALE

Le degré d'appartenance  $v^i(x)$  d'un point  $x$  à chaque classe  $C^i$ , est obtenu par agrégation de la valeur de la fonction exponentielle globale et des valeurs de possibilités sur chaque caractère. Cette agrégation, dont l'expression est donnée ci-dessous, est réalisée par l'opérateur minimum [8] :

$$v^i(x) = \min[\pi^{i1}(x^1), \pi^{i2}(x^2), \pi^{i3}(x^3), \dots, \pi^{ia}(x^a), \mu^i(x)]. \quad (2)$$

Elle permet d'atténuer les zones de forte appartenance données par la fonction  $\mu^i(x)$  autour de chaque prototype. L'opérateur minimum appartient à la classe t-normes, dont la caractéristique essentielle est l'intolérance [24]. Aucun effet compensatoire n'apparaît avec cet opérateur. Si une seule des valeurs agrégées est faible, la valeur globale d'appartenance après agrégation s'en trouve aussitôt diminuée. Cette propriété est recherchée lors du classement puisque la présence d'une valeur très faible, parmi les possibilités et la fonction  $\mu^i(x)$ , signifie que le point  $x$  n'est pas dans la classe. Il est tout à fait cohérent que sa valeur d'appartenance à la classe, après agrégation, soit faible.

Nous constatons sur la figure 7 que les courbes de niveaux d'appartenance respectent la forme des classes. Ces dernières ont été décomposées en 5 sous-classes. L'approche multiprototype autorise l'utilisation de la méthode floue du pattern matching pour les classes de forme non convexe ou oblique.

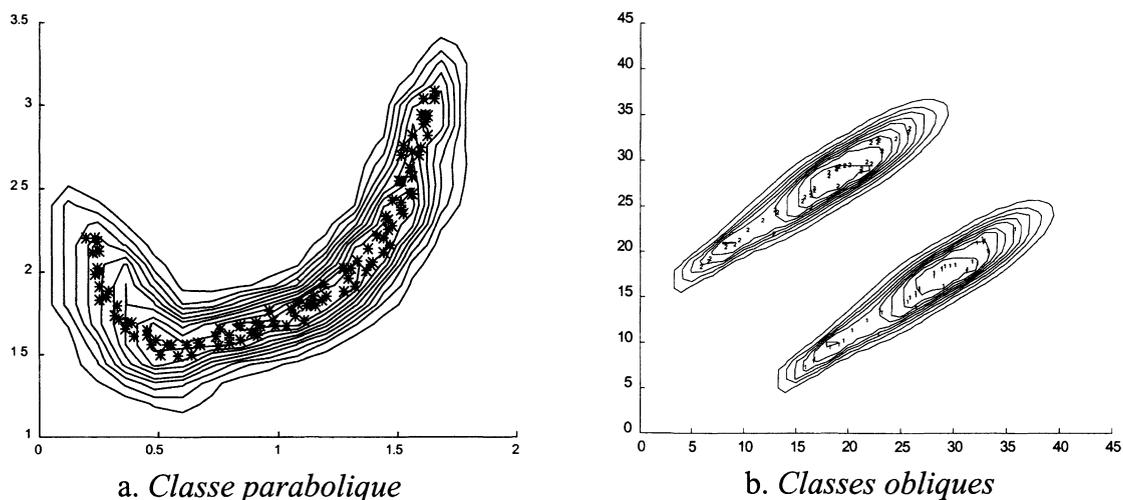


Figure 7. Courbes de niveaux d'appartenance obtenues par agrégation des possibilités et de la fonction exponentielle globale

### 3.3. INFLUENCE DU NOMBRE DE SOUS-CLASSES ET DES PARAMETRES $\lambda_j^i$

Les performances de l'algorithme dépendent du nombre  $s_i$  de sous-classes créées à l'intérieur de chaque classe  $C^i$ . Ce nombre peut être différent d'une classe à une autre. Les critères de validité des classes habituellement utilisés pour la recherche de la partition optimale dans un nuage de points, permettent de déterminer ce paramètre [3]. Parmi les critères les plus utilisés, nous pouvons citer l'hypervolume flou [1], la compacité-séparabilité [20] et l'entropie de classification [1]. Le critère de compacité-séparabilité donne généralement de très bons résultats. Un nombre de sous-classes trop important provoque une augmentation des temps de calculs sans accroissement des performances. Une valeur trop faible a pour effet de dégrader les contours d'appartenance, ce qui provoque une diminution des performances en classement.

Pour chaque sous-classe  $C_j^i$ , la fonction  $\mu_j^i(x)$  peut être assimilée à une fonction d'appartenance. Les valeurs  $\lambda_j^i$  permettent d'étaler les niveaux de cette fonction par

rapport aux centroïdes des sous-classes. Nous avons choisi d'utiliser la même valeur  $\lambda$  pour toutes les sous-classes. Selon les hypothèses définies dans [19], elle est déterminée pour que la fonction  $\mu_j^i(x)$  prenne une valeur proche de 0.5 pour la distance moyenne entre tous les points de la sous-classe et son centroïde. La valeur que nous avons utilisée pour l'exemple de la figure 7.a est 0.01. Ce paramètre a peu d'influence sur les résultats de classement. Toutefois il doit être déterminé de façon à ce que les courbes de niveaux d'appartenance enveloppent les points d'apprentissage, si une procédure de rejet en appartenance [10] doit être incluse dans l'algorithme.

### 3.4. APPRENTISSAGES INITIAL ET INCREMENTAL, REGLE DE CLASSEMENT

La phase d'apprentissage comprend le calcul des densités de possibilités, des centroïdes et des matrices de covariance floues des sous-classes. Elle peut aussi intégrer la détermination des paramètres de la méthode : le nombre  $h$  de barres des histogrammes, le facteur d'étalement  $\lambda$  de la fonction  $\mu^i(x)$  et le nombre de sous-classes  $s_i$  créées dans chaque classe.

Le classement d'un point  $x$  nécessite le calcul des possibilités d'appartenance sur chaque densité ainsi que le calcul des fonctions  $\mu_j^i(x)$ . L'agrégation de ces valeurs par les expressions (1) et (2), donne la possibilité d'appartenance du point à chaque classe. Ce dernier est affecté à celle pour laquelle il a la plus forte possibilité d'appartenance.

Il est possible d'intégrer une procédure d'apprentissage incrémental. Lorsqu'un point est classé, il est pris en compte dans les histogrammes dont la mise à jour entraîne une modification des densités de possibilité. L'adaptation des sous-classes n'est effectuée qu'après le classement de plusieurs points, puisque leur calcul demande un temps important.

## 4. APPLICATIONS

### 4.1. NUAGE COMPORTANT DES CLASSES DE FORME NON CONVEXES

La méthode que nous avons développée, nous a permis de résoudre le problème des classes de forme non elliptique exposé figure 8. La classe parabolique que nous avons présentée figure 7, a été décomposée en 5 sous-classes. La seconde classe parabolique et la classe elliptique ont été respectivement décomposées en 4 et 3 sous-classes. Nous pouvons constater que la forme des courbes de niveaux d'appartenance respecte celle des classes. En terme de pouvoir de discrimination, notre méthode donne des résultats similaires à la méthode des  $k$ -plus proches voisins. A titre d'exemple, nous présentons figure 8b les frontières données par la méthode du 1-plus proche voisin.

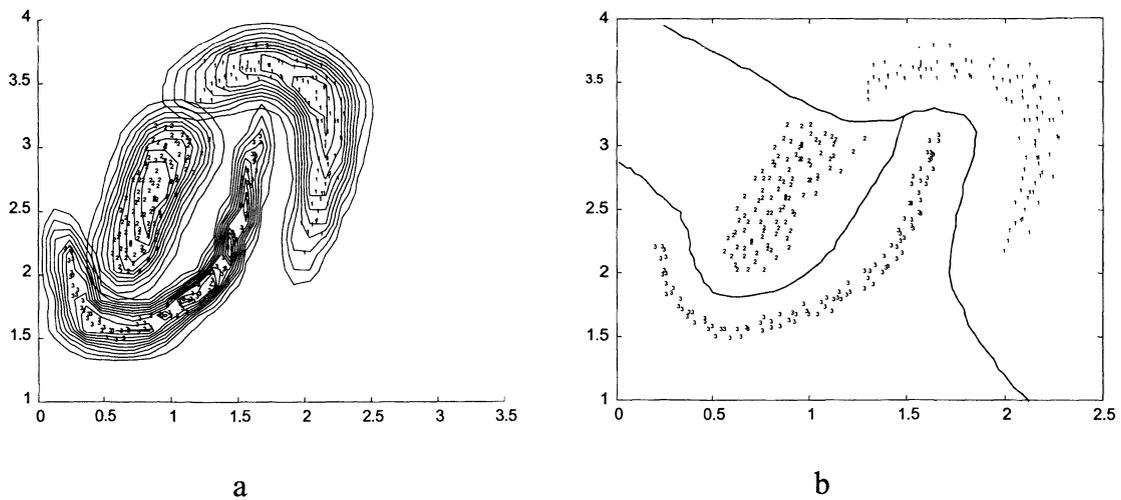


Figure 8. Courbes de niveaux d'appartenance pour des classes de forme parabolique (a) et frontière de discrimination obtenue par la règle du 1-ppv (b).

4.2. TRI DES BOUTEILLES PLASTIQUES

4.2.1. Présentation des données

Constituées soit de polychlorure de vinyle (PVC), soit de polyéthylène thérèphtalate (PET), soit de polyéthylène haute densité (PEHD), les bouteilles plastiques sont aujourd'hui au centre d'un enjeu technique et industriel important pour le recyclage. L'analyse du spectre infrarouge de chaque matière, accompagnée d'une étude statistique, nous a permis de sélectionner deux longueurs d'onde pour représenter les classes. Nous possédons 120 valeurs de transmission pour chacune des longueurs d'onde. Ces valeurs se répartissent en 3 classes de 40 points correspondant aux trois polymères, comme le montre la figure 9.

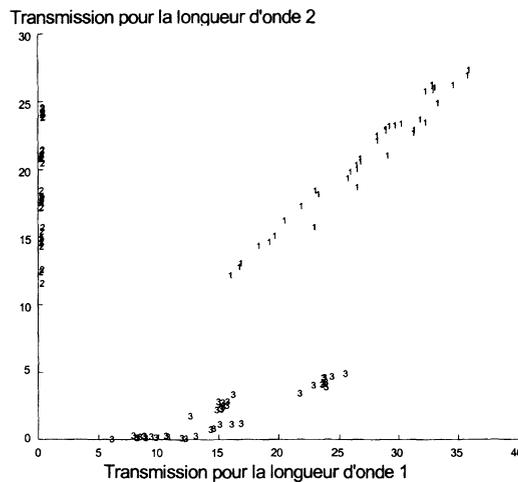


Figure 9. Nuage de points obtenu pour les matières plastiques, 1 : Polychlorure de Vinyle, 2 : Polyéthylène Thérèphtalate, 3 : Polyéthylène Haute Densité

#### 4.2.2. Courbes de niveaux d'appartenance et taux de bon classement

L'application de la méthode floue du pattern matching, dans sa version classique, avec le minimum comme opérateur d'agrégation, donne les courbes de niveaux d'appartenance présentées figure 10a. Nous constatons sur cette figure, que la forme oblique de la classe 1 n'est pas respectée. Ceci peut engendrer des problèmes de classement. En effet, certains points peuvent être affectés à cette classe alors qu'ils ne lui appartiennent pas. L'introduction de l'approche multiprototype, avec une décomposition de chaque classe en 4 sous-classes, donne les courbes de niveaux d'appartenance présentées figure 10b. Nous constatons que les courbes respectent la forme des classes. Pour cette application, nous avons obtenu un taux de bon classement de 100 % avec 20 points d'apprentissage par classe. Les paramètres  $h$  et  $\lambda$  ont été respectivement fixés à 5 et 0.2.

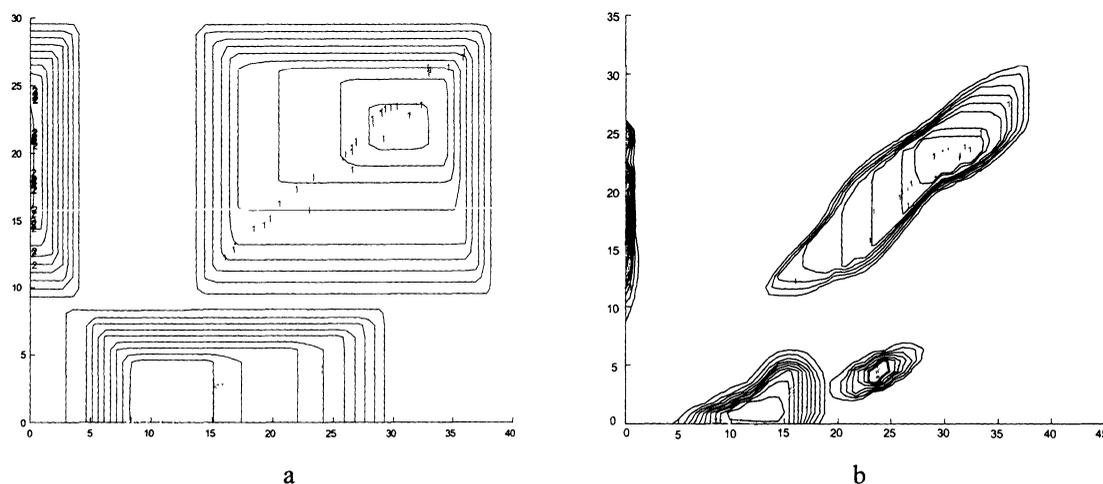


Figure 10. Courbes de niveaux d'appartenance obtenues avec l'opérateur minimum (a) et avec adjonction des fonctions exponentielles (b)

#### 4.3. CLASSEMENT DES DONNEES IRIS

Les données Iris sont largement utilisées pour la comparaison des classifieurs. Elles se composent de 3 classes de fleurs : Iris Sétosa, Versicolor, et Virginia. Chacune de ces classes est constituée de 50 échantillons caractérisés par 4 caractères : longueur et largeur de sépales, longueur et largeur de pétales. Le taux de bon classement donné par l'algorithme flou du pattern matching classique est de 94 % avec 25 points d'apprentissage par classe et le paramètre  $h$  fixé à 5. L'adjonction des prototypes avec la paramètre  $\lambda$  égal à 0.2 et une décomposition de chaque classe en 3 sous-classes nous a permis d'obtenir un taux de 97 %. Ce taux est identique au taux donné par la méthode des 10-plus proches voisins.

#### 4.4. DONNEES DE FORMES D'ONDES

Ces données sont issues de l'étude des méthodes de classement réalisée par Breiman [5]. Le nuage, composé de 5000 points, comporte 3 classes représentées par 21 caractères. Conformément aux études réalisées sur ces données par d'autres auteurs [6], nous avons utilisé 300 points d'apprentissage, soit 100 par classe, et nous avons calculé

le taux de bon classement sur les 5000 points. Les paramètres  $h$  et  $\lambda$  ont été respectivement fixés à 5 et 0.01. Les classes d'apprentissage ont été décomposées chacune en 3 sous-classes. Le taux de bon classement que nous obtenons est de 81 %. Nous avons testé plusieurs combinaisons pour les paramètres  $h$  et  $\lambda$  et il s'avère que les valeurs données ci-dessus peuvent être considérées comme les valeurs optimales, donnant le taux de bon classement le plus élevé. L'apprentissage intègre l'estimation des densités de possibilité, la décomposition des classes ainsi que le calcul des matrices de covariance floue. Le taux que nous obtenons peut être comparé au taux de 81.7 % donné par la méthode des 30-plus proches voisins [11]. La méthode floue du pattern matching modifiée peut être considérée comme très performante pour discriminer ces données.

## 5. COMPARAISON AVEC LA METHODE FLOUE DES $k$ -PLUS PROCHES VOISINS

Les critères que nous avons retenus pour comparer la méthode floue des 5-plus proches voisins [17] à la notre, sont le taux de bon classement et le temps de classement pour un point. Les résultats sont présentés dans le tableau 1. Les temps de calcul sont donnés pour un micro-ordinateur de type pentium cadencé à 120 Mhz.

Nous avons effectué 100 essais avec tirage aléatoire des points d'apprentissage, puis classement de la totalité des points. Le taux de bon classement donné dans le tableau 1 est la moyenne des taux de bon classement obtenus pour les 100 essais. Cette procédure permet d'atténuer l'influence de l'ensemble d'apprentissage sur les performances de l'algorithme.

Nous constatons que la méthode floue du pattern matching modifiée est plus rapide que la méthode floue des 5-plus proches voisins dans tous les cas. De plus elle utilise un nombre de points d'apprentissage moins important. En ce qui concerne les données de formes d'ondes, notre algorithme donne un taux de bon classement supérieur à celui obtenu par la méthode floue des 5-plus proches voisins.

Tableau 1. *Résultats comparatifs*

		<b>Bouteilles plastiques</b>	<b>Iris</b>	<b>Formes d'ondes</b>
<b>Méthode floue du pattern matching modifiée</b>	Nombre de points d'apprentissage par classe	20	25	100
	Taux de bon classement	100 %	97 %	81 %
	Temps de classement pour un point	0.017 s	0.053 s	0.13 s
<b>Méthode floue des 5 plus proches voisins</b>	Nombre de points d'apprentissage par classe	25	35	100
	Taux de bon classement	100 %	97 %	78 %
	Temps de classement pour un point	0.089 s	0.13 s	0.5 s

## CONCLUSION

L'amélioration de la méthode floue du pattern matching, que nous proposons, permet son application à des nuages comportant des classes de forme quelconque. Elle requière la détermination de trois paramètres : le nombre de barres  $h$  des histogrammes, le nombre  $n_i$  de sous-classes créées à l'intérieur de chaque classe  $C^i$  et le paramètre  $\lambda$  caractérisant l'étalement des fonctions exponentielles. L'expérience a montré qu'une valeur de  $h$  comprise entre 5 et 15 garantit de bonnes performances. Quelques critères permettent la détermination du nombre de sous-classes et du paramètre  $\lambda$ .

Nous avons comparé la méthode floue du pattern matching, ainsi modifiée, avec la méthode floue des 5-plus proches voisins, sur trois jeux de données. Les deux méthodes donnent des taux de bon classement similaires. Cependant notre méthode donne les temps de calcul les plus courts.

L'utilisation de cette méthode pour des problèmes tels que le tri des bouteilles plastiques est tout à fait envisageable même s'il est possible d'utiliser des méthodes plus simples telles que l'analyse discriminante linéaire. L'intégration des procédures de rejets en appartenance et en ambiguïté [10] étant aisée dans cet algorithme, nous envisageons son utilisation pour résoudre des problèmes de diagnostics, tant dans le domaine médical qu'industriel.

## BIBLIOGRAPHIE

- [1] BEZDEK J.C., *Pattern recognition with fuzzy objective function algorithms*, New-York, Plenum Press, 1981.
- [2] BILLAUDEL P., DEVILLEZ A., VILLERMAIN LECOLIER G., "Performance evaluation of fuzzy classification methods designed for real time application", *International journal of approximate reasoning* 20, (1999), 1-20.
- [3] BILLAUDEL P., DEVILLEZ A., VILLERMAIN LECOLIER G., "An unsupervised fuzzy classification algorithm for the non elliptic classes", *EURISCON'98*, (1998), Athens, Greece.
- [4] BOUCHON MEUNIER B., *La logique floue et ses applications*, Paris, Collection Vie Artificielle, Addison-Wesley, 1995.
- [5] BREIMAN L., FRIEDMAN J. H., OLSHEN R. A., STONE C. J., "Classification and regression trees", *Wadsworth international group : Belmont, California* (1984), 43-49.
- [6] CARAUX G., LECHEVALLIER Y., "Règles de décision de Bayes et méthodes statistiques de discrimination", *Revue d'intelligence artificielle* 10 (2,3), (1996) 219-283.
- [7] CELEUX G., DIDAY E., GOVAERT G., LECHEVALLIER Y., RALAMBONDRAIN H., *Classification automatique des données*, Paris, Dunod, 1989.
- [8] DEVILLEZ A., BILLAUDEL P., VILLERMAIN LECOLIER G., "Use of the Fuzzy Pattern Matching in a diagnosis module based on the pattern recognition", *CESA '98 IEEE Systems, man and cybernetics* 4, (1998), 902-907.
- [9] DUBOIS D. et PRADE H., *Théorie des possibilités Application à la représentation des connaissances en informatique*, Paris, Masson, 1987.

- [10] DUBUISSON B., *Diagnostic et reconnaissance des formes*, Traité des Nouvelles Technologies, série Diagnostic et Maintenance, Paris, Hermès, 1990.
- [11] GASCUEL O. et al., "Twelve numerical, symbolic and hybrid supervised classification methods", *International journal of pattern recognition and artificial intelligence* 12 (5), (1998) 517-571.
- [12] GRABISCH M. et SUGENO M., "Multi-attribute classification using fuzzy integral ", *Proc. of fuzzy IEEE* ,(1992), 47-54.
- [13] GRABISCH M., "The application of fuzzy integral in multicriteria decision making", *European journal of operational research* 89, (1996), 445-456.
- [14] GUSTAFSON D.E., KESSEL W.C., "Fuzzy clustering with a fuzzy covariance matrix ", *Proceedings of IEEE Conference on decision and control*, (1979), 761-766.
- [15] ISHIBUCHI H., NOZAKI K., TANAKA H., "Distributed representation of fuzzy rules and its application to pattern classification" *Fuzzy sets and systems* 52, (1992), 21-32.
- [16] ISHIBUCHI H., NOZAKI K., TANAKA H., "Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms", *IEEE Transactions on fuzzy systems* 3, (1995).
- [17] KELLER J.M., GRAY M.R, GIVENS J.A. JR, "A fuzzy k-nearest neighbor algorithm", *IEEE Transactions on systems, man and cybernetics*, 15 (4), (1985), 580-585.
- [18] *La logique floue*, Observatoire Français des techniques Avancées, série ARAGO n°14, Masson, 1994.
- [19] MASSON M. H., DUBUISSON B., FRELICOT C., "Conception d'un module de reconnaissance de formes floues pour le diagnostic", *RAIRO-APII-JESA* 30 (2, 3), (1996), 319-341.
- [20] XIE X. L., BENI G., "A validity measure for fuzzy clustering", *IEEE Transactions on pattern analysis and machine intelligence* 13 (8), (1991) 841-847.
- [21] YAGER R. R., "On ordered weighted averaging aggregation operators in multicriteria decision making", *IEEE Transactions on systems, man and cybernetics* 18 (1), (1988), 183-190.
- [22] YAGER R. R., "Aggregation operators and fuzzy systems modeling", *Fuzzy sets and systems* 67, (1994), 129-145.
- [23] ZADEH L. A., "Fuzzy sets as a basis for a theory of possibility", *Fuzzy sets and systems* 1, (1978), 3-28.
- [24] ZADEH L. A., "Fuzzy sets", *Informations and control* 8, (1965), 338-353.