

Le rejet d'ambiguïté dans les Fc+2M : extension à toutes les classes composées

The fuzzy c+2 means: solving the extended ambiguity reject in clustering

par Michel MÉNARD

Avenue Marillac 17042 La Rochelle Cedex 1 (france)
e-mail : michel.menard@l3i.univ-lr.fr

résumé et mots clés

Cet article concerne le problème de la classification dont l'objectif consiste à calculer une partition d'un ensemble de formes en classes disjointes. La méthode que nous proposons s'écrit sous la forme d'un problème de minimisation d'un critère de classification dont la solution dépend d'une fonction de contrainte dans laquelle la notion de rejet est introduite. Deux types de rejet ont été inclus : le rejet d'ambiguïté qui concerne les individus situés sur la frontière de deux ou plusieurs classes et le rejet de distance gérant ceux qui sont loin de toutes les classes. La notion de rejet permet de gérer l'incertitude due à l'imprécision et à la définition incomplète des classes. Pour calculer ces rejets, nous proposons une extension de l'algorithme des c -moyennes floues (FcM) de Bezdec [2]. Cet algorithme est appelé $c+2$ -moyennes floues (Fc+2 M). Ces avantages sont : (1) le degré d'appartenance aux classes de rejet d'un individu est déterminé par un calcul itératif durant la phase de coalescence; (2) l'algorithme inclut une modélisation de l'hésitation ou de l'ambiguïté, et des fonctions d'appartenance sont affectées à tous les sous-ensembles de classes de 2^Ω plutôt qu'aux éléments de Ω seulement (Ω étant l'ensemble des classes). De plus, le degré d'appartenance à la classe de rejet d'ambiguïté totale d'un individu est calculé explicitement; (3) en introduisant les rejets d'ambiguïté partielle, nous créons une discontinuité entre les fonctions d'appartenance affectées aux classes de Ω , ce qui permet de rendre ces fonctions d'appartenance indépendantes.

Reconnaissance des formes, théorie des ensembles flous, coalescence flou, c -moyennes floues, rejet d'ambiguïté et rejet de distance.

abstract and key words

We address in this paper a clustering problem whose goal consists of computing a partition of a family of patterns into disjoint classes. The method that we propose is formulated as a constrained minimization problem, whose solution depends on a cost function in which rejection options are introduced. Two types of rejection have been included : the ambiguity rejection which concerns patterns lying near the class boundaries and the distance rejection dealing with patterns that are far away from all the classes. To compute these rejections, we propose an extension of the fuzzy c -means (FcM) algorithm of Bezdec ([2]). This algorithm is called the fuzzy $c+2$ -means (Fc+2M). These measures allow to manage uncertainty due both to imprecise and incomplete definition of the classes. The advantages of our method are (1) the degree of membership to the rejection classes for a pattern x_k are learned in the iterative clustering problem; (2) the partial ambiguity rejections introduce a discounting process between the classical FcM membership functions in order to avoid the memberships to be spread across the classes; (3) the membership functions are more immune to noise and correspond more closely to the notion of compatibility.

Pattern recognition, fuzzy sets, clustering, fuzzy c -means, distance rejection, ambiguity rejection.

1. introduction

Cet article concerne le problème de la coalescence dont l'objectif est l'obtention d'une partition d'un ensemble d'individus non étiquetés en sous-ensembles ou classes disjoints deux à deux, tous non vides et dont l'union produit l'ensemble original.

Ce problème se définit de la façon suivante :

- Soit $Z = (z_k)_{k \in [1, n]}$ l'ensemble des individus où $z_k = (z_{k1}, z_{k2}, \dots, z_{kp})^t$ est un vecteur décrit par p caractéristiques;
- Soit $\Omega = (\omega_i)_{i \in [1, c]}$ un ensemble de classes.

Les individus appartenant à une même classe partagent des propriétés communes qui les distinguent des individus s'inscrivant dans d'autres classes. Dans la littérature, les algorithmes de coalescence se divisent en deux groupes :

- ceux qui créent une partition nette. Ces algorithmes assignent à chaque vecteur caractéristique une seule étiquette en prenant comme hypothèse qu'un vecteur ne peut appartenir à d'autres classes;
- ceux qui créent une partition floue. Ces algorithmes considèrent chaque classe comme un sous-ensemble flou. On associe une fonction d'appartenance à chaque individu mesurant son appartenance à chaque classe. Ainsi chaque individu peut être assigné à plusieurs classes avec des degrés de certitudes différents. La coalescence floue constitue donc une généralisation de la coalescence classique.

Le regroupement net ou flou d'individus peut être représenté par une matrice d'appartenance nette / floue $\mu = [\mu_{ik}]$ appelée partition nette / floue.

Les matrices de tailles $c \times n$ définissant une partition nette sont notées M_{cn} et sont définies par :

$$M_{cn} = \{ \mu \in \mathcal{R}^{c \times n} \mid \sum_{i=1}^c \mu_{ik} = 1, 0 < \sum_{k=1}^n \mu_{ik} < n, \mu_{ik} \in \{0, 1\}; 1 \leq i \leq c; 1 \leq k \leq n \}$$

Les matrices de tailles $c \times n$ définissant une partition floue sont notées M_{fcn} et sont définies par :

$$M_{fcn} = \{ \mu \in \mathcal{R}^{c \times n} \mid \sum_{i=1}^c \mu_{ik} = 1, 0 < \sum_{k=1}^n \mu_{ik} < n, \mu_{ik} \in [0, 1]; 1 \leq i \leq c; 1 \leq k \leq n \}$$

Dans le premier cas, la coalescence est souvent effectuée par un algorithme itératif, les c -moyennes HcM (Hard c -Means), qui permet de trouver une famille de c classes partitionnant l'ensemble $Z = (z_k)_{k \in [1, n]}$ (voir [29], [1]).

Le critère de coalescence utilisé pour définir les classes est :

$$J_1^{hcm}(\mu, v) = \sum_{i=1}^c \sum_{k=1}^n \mu_i(z_k) d_{ik}^2 \quad (1)$$

où $\mu \in M_{cn}$; d_{ik} est une mesure de la distance de l'individu z_k au centre de la $i^{\text{ème}}$ classe $v_i \in V = [v_1, \dots, v_c]$. V est la matrice des centres des classes. Les méthodes de coalescence classique ne sont pas très satisfaisantes lorsque l'ensemble Ω n'est pas structuré en classes clairement séparables, mais en classes chevauchantes, car la fonction d'appartenance est exclusive [20], [34]. Cet article traite plus particulièrement du second cas. Le critère généralement utilisé est le suivant :

$$J_m^{fcm}(\mu, v) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m d_{ik}^2 \quad (2)$$

où $\mu \in M_{fcn}$ est une matrice définissant une partition floue; $m > 1$ est le facteur de flou, $\mu_{ik} = \mu_i(z_k)$ et $d_{ik} = \|z_k - v_i\|_G$, est une mesure de la distance de l'individu z_k au centre de la $i^{\text{ème}}$ classe $v_i \in V = [v_1, \dots, v_c]$, G est une norme.

La partition floue ainsi générée apporte une mesure du degré d'appartenance de chaque individu à chaque classe donnée. Ces méthodes sont moins sensibles aux minima locaux que les algorithmes classiques puisqu'à chaque itération aucune décision binaire n'est prise. Ruspini, dans [34], explique l'échec des algorithmes classiques par le fait que la fonction d'appartenance à une classe est une relation transitive alors que la notion de similarité sur laquelle la classification devrait être basée ne l'est pas nécessairement. La plupart des fonctions d'appartenance sont définies dans ce but ([36], [35], [25]). Le premier algorithme de coalescence floue a été développé en 1969 par Ruspini [34]. L'algorithme de référence, utilisé aujourd'hui dans de nombreux domaines, est celui des c -moyennes floues (FcM) [4]. Il a été développé pour résoudre le problème de la minimisation du critère (2). Il consiste à choisir une partition initiale aléatoire $\mu^{(0)}$ et à itérer les deux équations suivantes :

$$v_i^{(t+1)} = \frac{\sum_{k=1}^n \mu_{ik}^{(t)m} z_k}{\sum_{k=1}^n \mu_{ik}^{(t)m}} \quad (3)$$

$$\mu_{ik}^{(t+1)} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}^{(t+1)}}{d_{jk}^{(t+1)}} \right)^{\frac{2}{m-1}}}$$

Cet algorithme converge sous certaines conditions [4] vers un minimum local de (2). Dans différents domaines, tels que la reconnaissance des formes, l'analyse de données ou bien le traitement d'images, l'algorithme des c -moyennes floues est très utilisé et obtient de bons résultats. De plus, un grand nombre d'algorithmes est dérivé de celui-ci, utilisant des métriques différentes ou des prototypes de dimensions supérieures. Citons les FcV (fuzzy

c -variétés [2], FcE [2] (fuzzy c -elliptotypes), AFc [14] (adaptive fuzzy c -elliptotypes), FMLE [3] (fuzzy maximum likelihood estimation), FcL (fuzzy c -lines : FCV, $r = 1$), FcRM [23] (fuzzy c -regressions means). Cependant, comme les fonctions d'appartenance sont générées sous la contrainte probabiliste due à Ruspini [34] ($\forall z_k \in Z, \sum_{i=1}^c \mu_i(z_k) = 1$ i.e. la somme est égale à 1 sur chaque colonne de μ), ces algorithmes souffrent de plusieurs inconvénients :

- Les degrés d'appartenance sont des degrés relatifs. Autrement dit, l'appartenance d'un individu à une classe dépend de l'appartenance de cet individu aux autres classes. Les fonctions d'appartenance construites sont donc interdépendantes. Aussi, les estimations des centres des classes ne correspondent pas aux centres réels ou typiques; dans le cas possibiliste, les centres générés représentent de manière intrinsèque les propriétés stables d'une classe [27]. L'algorithme FcM et les algorithmes dérivés rencontrent donc un problème sérieux lorsque l'on cherche à générer des fonctions d'appartenance à partir d'un ensemble d'apprentissage : la fonction d'appartenance μ_i ne représente pas l'aspect « typique » de la classe ω_i ;
- les points aberrants peuvent avoir des valeurs d'appartenance élevées et ils peuvent affecter de façon significative l'estimation des centres des classes;
- ces algorithmes modélisent dans la phase de classement l'hésitation ou l'ambiguïté entre classes à partir de règles de décisions floues fixées *a priori* [20]. Dans la phase de coalescence, les points ambigus ou non représentatifs des classes peuvent alors affecter la position des centres.

Pour pallier aux deux premiers problèmes, Krishnapuram et Keller [27] ont proposé un nouvel algorithme de classification appelé possibilistic c -means (PcM), où la contrainte est relâchée. Dans ce cas, les fonctions d'appartenance représentent l'aspect « typique » des classes. Par contre, le résultat n'est pas une partition floue ou probabiliste (les valeurs d'appartenance sont difficiles à interpréter lors de la phase de classement). De plus, l'algorithme PcM est très sensible aux initialisations, et les centres des classes peuvent coïncider. Les distributions possibilistes sont très dépendantes du choix des paramètres. Le rejet d'ambiguïté n'est pas introduit.

D'autres méthodes déterminent une estimation des centres plus robuste que celle donnée par l'approche FcM. Elles proposent de pondérer la participation des points dans l'estimation par leur pertinence ([37], [9]). Cependant la partition résultante est difficile à interpréter puisqu'un point non fiable peut avoir un degré d'appartenance important (la contrainte probabiliste est conservée). La pondération, associée à chaque point de l'espace de représentation, est introduite à partir d'informations supplémentaires. Celles-ci sont disponibles dans le cadre d'applications où, par exemple, à chaque source d'informations est attribuée un degré de fiabilité. L'ambiguïté ne peut être prise en compte par ce type de pondération.

Pour remédier aux problèmes des FcM, nous reformulons le problème de la coalescence floue en incluant la notion de rejet.

Les avantages sont les suivants :

- 1) contrairement aux travaux de Dubuisson *et al.*, ces rejets sont calculés dans l'étape de coalescence et non dans l'étape de décision [17], [18]; les données aberrantes ou ambiguës sont alors moins prises en compte dans la détermination des centres;
- 2) l'algorithme inclut une modélisation de l'hésitation ou de l'ambiguïté, et des fonctions d'appartenance sont affectées à tous les sous-ensembles de classes de 2^Ω plutôt qu'aux éléments de Ω seulement. De plus, le degré d'appartenance à la classe de rejet d'ambiguïté totale d'un individu z_k est calculé explicitement;
- 3) les notions de rejet conduisent à attribuer une fonction de pondération différente à chaque individu selon que celui-ci appartient clairement à une classe, à aucune ou bien à plusieurs classes de façon égale;
- 4) nous préservons l'approche probabiliste de la classification des c -moyennes floues puisque la contrainte de Ruspini est vérifiée; nous obtenons en effet une partition floue. En introduisant les rejets d'ambiguïté partielle, nous créons une discontinuité entre les fonctions d'appartenance affectées aux classes de Ω , ce qui permet de rendre ces fonctions d'appartenance indépendantes. Celles-ci représentent alors l'aspect « typique » des classes.

Cet article est organisé de la façon suivante. Dans la section 2, nous définissons les rejets d'ambiguïté partielle. Nous précisons, à l'aide de la théorie des ensembles rugueux, le concept de *frontière d'une classe* dans une partition nette. Puis nous étendons ce concept aux partitions floues et proposons une mesure sur les *frontières floues*. Nous considérons le problème de la coalescence dans le cadre de la théorie de la régularisation, et proposons un critère de classification prenant en compte les discontinuités de façon équivalente au modèle énergétique de Pavlidis [31]. Puis nous dérivons de la minimisation de ce critère les formules de mise à jour des fonctions d'appartenance et des centres. La preuve est détaillée dans l'annexe A. Dans la section 3, nous améliorons la robustesse de notre algorithme vis-à-vis du bruit en introduisant simultanément le rejet d'ambiguïté et celui de distance défini dans [15] dans le but de rejeter un individu quand il est trop loin du centre de la classe : à chaque classe ω_i est associée une classe de rejet notée $\bar{\omega}_i$. Ceci diffère donc de l'algorithme NC de Davé (Noise cluster) [12], [13] où une seule classe de rejet de distance est définie. La preuve de ce théorème est présentée en annexe B. Une comparaison de l'algorithme Fc + 2M avec les approches PcM, NC et statistiques sur la gestion du rejet de distance et sur la notion de robustesse est également décrite dans cette section. La section 4 illustre notre approche sur différents exemples dans le but de montrer l'intérêt du partitionnement conditionné par les mesures de rejet. Dans le premier exemple, nous comparons les résultats obtenus par les c -moyennes floues FcM [2] (algorithme sans rejet), les c -moyennes floues possibilistes FPcM [30] (algorithme avec valeurs de typicalité), les $c + 1$ -moyennes floues F(c + 1)M [15] (algorithme avec rejet

de distance seul), et les $c + 2$ -moyennes floues F(c + 2)M sur l'ensemble de données *diamonds*. Un second exemple montre le comportement des F(c + 2)M sur un ensemble de données 2-D concernant 3 classes et montre l'intérêt d'étendre le rejet d'ambiguïté à l'ensemble des 4 classes composées. De plus, nous comparons le comportement des fonctions d'appartenance en fonction des deux paramètres ρ_α et ρ , contrôlant les rejets de distance et d'ambiguïté. Enfin un troisième exemple évalue la robustesse de la méthode relativement à des points aberrants.

2. coalescence floue avec rejet d'ambiguïté

Pour réduire un taux d'erreurs excessif dû au bruit et aux autres facteurs d'incertitude inhérents aux mesures réelles, le rejet en ambiguïté apporte une solution intéressante. Dans la littérature, les règles permettant de rejeter un individu situé sur la frontière de deux classes sont proposées dans la phase de décision, et sont le plus souvent paramétrées par des valeurs de seuil.

Le rejet d'ambiguïté a été introduit par Chow dans le processus de décision ([10], [11]). L'objectif est de mesurer la proximité de z_k aux bords des classes. Dans le cas de la règle de Bayes définie à partir des probabilités *a posteriori* $P(\omega_i/z_k)$, la méthode revient à définir une nouvelle classe ω_0 , appelée classe de rejet, associée à un coût de rejet constant $C_r = C(0/j) \forall j \in [1, c]$:

$$\begin{cases} e(z_k) = i \text{ if } P(\omega_i/z_k) = \max_{j \in [1, c]} (P(\omega_j/z_k)) > 1 - C_r \\ e(z_k) = 0 \text{ autrement} \end{cases} \quad (4)$$

Le terme C_r , qui contrôle le taux de rejet, doit être choisi entre 0 et $\frac{c-1}{c}$. Cette règle partage l'espace de représentation en $(c+1)$ classes. Dans [17], Dubuisson précise que ce rejet est un rejet d'ambiguïté puisque la région correspondante dans l'espace de représentation est toujours localisée entre les classes.

Le principal inconvénient de ces définitions et de leurs utilisations est que ces rejets sont définis à partir de seuils fixés *a priori*, sans prendre en compte la forme des classes. Les individus rejetés lors de la phase décisionnelle ont été pris en compte avec la même importance que les autres dans le calcul des centres des classes, alors qu'en définitive on les considère non représentatifs de ces classes. Une extension a été apportée par Ha dans [22].

Dans cette section, nous construisons un critère de classification approprié dans le but de prendre en compte dès la phase de coalescence le rejet d'individus en ambiguïté.

Afin de tenir compte de l'hésitation et de l'ambiguïté entre classes, des fonctions d'appartenance sont affectées à tous les éléments de 2^Ω plutôt qu'aux éléments de Ω seulement ($\Omega = \{\omega_1, \dots, \omega_c\}$ étant l'ensemble des classes ω_i). Puisqu'il y a 2^c sous-ensembles dans

un ensemble de c éléments, nous cherchons à partitionner l'espace des représentations en $2^c - 1$ classes simples et composées, excluant l'ensemble vide, dans un problème à c classes. Il y a donc c classes simples, $(\{\omega_i\})_{i \in [1, c]}$, et pour déterminer le rejet d'ambiguïté nous associons une classe d'ambiguïté à chaque sous-ensemble de classes $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1, c]}\}$ (i.e. $A \subseteq \Omega, |A| > 1$). Dans la figure 1, il y a $2^3 - 1$ classes correspondant aux sous-ensembles $\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \omega_{12} = \{\omega_1, \omega_2\}, \omega_{13} = \{\omega_1, \omega_3\}, \omega_{23} = \{\omega_2, \omega_3\}$ and $\omega_{123} = \{\omega_1, \omega_2, \omega_3\}$.

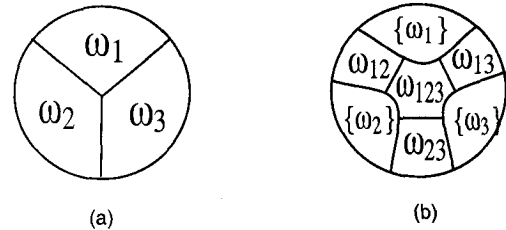


Figure 1. - La figure (a) illustre la partition de l'espace des représentations en trois classes simples. La figure (b) illustre la partition de l'espace des représentations lorsque le concept d'ambiguïté est appliqué. Nous associons une classe d'ambiguïté à chacun des sous-ensembles de classes $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1, 3]}\}$.

Définition 1

Soit $\Omega = \{\omega_1, \dots, \omega_c\}$ l'ensemble des classes ω_i . Nous associons une classe d'ambiguïté à chacun des sous-ensembles de classes $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1, c]}\}$ (i.e. $A \subseteq \Omega, |A| > 1$). Nous considérons qu'il existe un ensemble de $\sum_{k>1} \binom{c}{k}$ classes d'ambiguïté disjointes dans un problème à c classes.

2.1. le concept d'ambiguïté : frontière nette et floue

Cette section concerne le problème de la définition de la frontière de plusieurs classes dans un objectif de régularisation en classification. Nous utilisons la théorie des ensembles rugueux de Pawlak [32] pour préciser le concept d'ambiguïté et définir le critère de classification à minimiser.

Nous faisons la distinction entre deux types de frontière : le premier est défini par la théorie des ensembles rugueux, le second est obtenu en généralisant ce concept dans le cadre de la théorie des sous-ensembles flous. Le premier utilise les relations d'indiscernabilité, l'autre des fonctions à valeurs réelles. A partir de ces définitions, nous étendons le concept de frontière au concept de frontières floues de classes d'une partition floue.

Un système d'information est une paire $\mathcal{A} = (U, A)$, où U est un ensemble fini, non vide appelé l'univers et A , un ensemble fini, non vide d'attributs, i.e. $a : U \rightarrow V_a \forall a \in A$, où V_a est l'ensemble des valeurs de a . Par V nous notons l'ensemble $\cup \{V_a : a \in A\}$. Les éléments de U sont appelés objets et peuvent représenter par exemple des états, des patients, des observations.

Les attributs peuvent être interprétés comme des caractéristiques, variables...

A chaque sous-ensemble d'attributs $B \subseteq A$, une relation appelée B-indiscernable notée $IND_A B$ est associée et est définie par :

$$IND_A B = \{(u, u') \in U^2 : a(u) = a(u') \forall a \in B\}$$

Les objets u, u' satisfaisant la relation $IND(B)$ sont indiscernables selon B car ils ont les mêmes valeurs relativement aux attributs de B .

Définition 2

On appelle ensemble élémentaire (par rapport à B) un ensemble $[u]_B$ d'objets qui sont tous indiscernables par rapport à B .

Définition 3

Un ensemble définissable (par rapport à B) est une union finie d'ensembles élémentaires (par rapport à B).

Si $\mathcal{A} = (U, A)$ est un système d'information, $B \subseteq A$ est un ensemble d'attributs et $Z \subseteq U$ est un ensemble d'objets, alors les ensembles :

$$\{u \in U : [u]_B \subseteq Z\}$$

et

$$\{u \in U : [u]_B \cap Z \neq \emptyset\}$$

sont appelés respectivement la plus basse, respectivement la plus haute approximation de Z dans A . Ils sont notés \underline{BZ} et \overline{BZ} . Le premier est le plus grand ensemble définissable inclus dans Z (i.e. l'union de tous les ensembles élémentaires dont chacun est inclus dans Z); le second est le plus petit ensemble définissable contenant Z (c'est l'union des ensembles élémentaires dont chacun a une intersection non vide avec Z). L'ensemble $BN_B(Z) = \overline{BZ} - \underline{BZ}$ est appelé frontière de Z . L'ensemble \underline{BZ} est l'ensemble de tous les éléments de U qui peuvent être classés de façon certaine comme élément de Z , avec la connaissance représentée par les attributs de B . \overline{BZ} , est l'ensemble de tous les éléments qui appartiennent de façon possible à Z . L'ensemble $BN_B(Z)$ est constitué par les éléments que l'on ne peut classer ni dans Z ni dans \overline{Z} avec la connaissance de B . Une illustration de ces définitions est représentée sur la figure 2. L'ensemble U de tous les objets est représenté par le grand rectangle et les ensembles élémentaires sont représentés par de petits rectangles. Z représente l'ensemble à approximer.

Les informations sur les individus sont souvent disponibles sous la forme de tables. Dans l'exemple suivant, afin de préciser la notion de frontière entre classes, nous interprétons le résultat d'une partition nette dans le cadre de la théorie des ensembles rugueux. Chaque case dans le tableau 1 représente un individu z_k dans l'espace U ; il est noté dans le tableau par z_i^I .

Soit $\Omega = \{\omega_1, \omega_2, \omega_3\}$ un ensemble de classes. Dans cet exemple, U est partitionné en ensembles élémentaires : à chaque individu z_i^I est attribué un label $I \subseteq \{1, 2, 3\}$ indiquant avec certitude son appartenance à un sous-ensemble de 2^Ω ; il y a donc 4

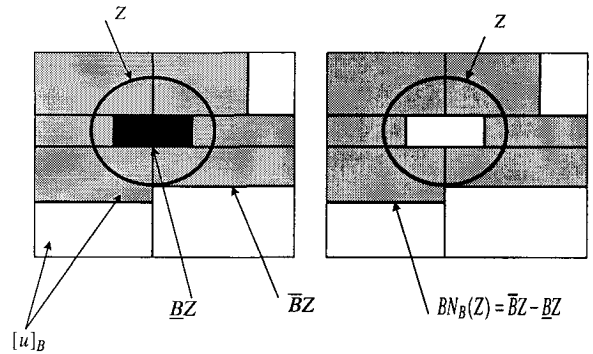


Figure 2. – Exemple de frontière, notée $BN_B(Z)$, définie par la théorie des ensembles rugueux.

Tableau 1. – Soit $\Omega = \{\omega_1, \omega_2, \omega_3\}$ un ensemble de classes. Le tableau définit une partition sur U en ensembles élémentaires : à chaque individu $z_i^I \in U$ est attribué un label $I \subseteq \{1, 2, 3\}$ indiquant avec certitude son appartenance à un sous-ensemble de 2^Ω . Dans cet exemple, l'attribut est le label attribué.

$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	ω_{12}	ω_{13}	ω_{23}	ω_{123}
z_1^1	z_1^2	z_1^3	z_1^{12}	z_1^{13}	z_1^{23}	z_1^{123}
z_2^1	z_2^2	z_2^3	z_2^{12}	z_2^{13}	z_2^{23}	z_2^{123}
z_3^1	z_3^2	z_3^3	z_3^{12}	z_3^{13}	z_3^{23}	z_3^{123}
z_4^1	z_4^2	z_4^3	z_4^{12}	z_4^{13}	z_4^{23}	z_4^{123}
z_5^1	z_5^2	z_5^3	z_5^{12}	z_5^{13}	z_5^{23}	z_5^{123}

classes correspondant aux classes ambiguës $\omega_{12}, \omega_{13}, \omega_{23}, \omega_{123}$ et 3 classes simples $\{\omega_1\}, \{\omega_2\}$ et $\{\omega_3\}$. La partition est donc définie par les ensembles élémentaires :

- $\{\omega_1\} = \{z_1^1, z_2^1, z_3^1, z_4^1, z_5^1\}$,
- $\{\omega_2\} = \{z_1^2, z_2^2, z_3^2, z_4^2, z_5^2\}$,
- $\{\omega_3\} = \{z_1^3, z_2^3, z_3^3, z_4^3, z_5^3\}$,
- $\{\omega_1, \omega_2\} = \{z_1^{12}, z_2^{12}, z_3^{12}, z_4^{12}, z_5^{12}\}$,
- $\{\omega_1, \omega_3\} = \{z_1^{13}, z_2^{13}, z_3^{13}, z_4^{13}, z_5^{13}\}$,
- $\{\omega_2, \omega_3\} = \{z_1^{23}, z_2^{23}, z_3^{23}, z_4^{23}, z_5^{23}\}$,
- $\{\omega_1, \omega_2, \omega_3\} = \{z_1^{123}, z_2^{123}, z_3^{123}, z_4^{123}, z_5^{123}\}$.

Maintenant, considérons la réalisation suivante : pour $z_i^I \in \omega_I$, $I \subseteq \{1, 2, 3\}$, $i \in [1, 5]$ l'attribut $a \in A$ est « oui » si $1 \in I$ et « non » autrement. Ceci conduit à :

- $\forall z_k \in E_1 = \{\omega_1\} \cup \{\omega_1, \omega_2\} \cup \{\omega_1, \omega_3\} \cup \{\omega_1, \omega_2, \omega_3\}$, $a(z_k) = \text{« oui »}$;
 - $\forall z_k \in E_2 = \{\omega_2\} \cup \{\omega_3\} \cup \{\omega_2, \omega_3\}$, $a(z_k) = \text{« non »}$
- $\{E_i : i = 1, 2\}$ représente une autre partition de U .

Aussi, les individus appartenant à E_1 sont indiscernables selon l'attribut a . Donc, l'attribut a génère deux ensembles élémentaires E_1 et E_2 .

Les individus $z_i^1 \forall i \in [1, 5]$ possèdent des informations qui nous permettent de les classer avec certitude comme appartenant à la classe ω_1 . La plus basse approximation de l'ensemble des individus appartenant à la classe ω_1 est l'ensemble $\{z_i^1 | i \in [1, 5]\} = \{z_k \in \omega_1\}$ et la plus haute approximation de cet

ensemble est l'ensemble $E_1 = \{z_k \in \omega_I : \forall \omega_I \in 2^\Omega \setminus \emptyset \text{ et } 1 \in I\}$. La frontière est alors définie par l'ensemble $E_1 - \{z_i^1 \mid i \in [1, 5]\} = \{z_k \in \omega_I : \forall \omega_I \in 2^\Omega \setminus \{\emptyset, \{\omega_1\}\} \text{ et } 1 \in I\}$.

Cet exemple apporte une définition simple de la frontière d'une classe :

Définition 4

Soit $\Omega = (\omega_i)_{i \in [1, c]}$ un ensemble de classes. Soit $Z = (z_k)_{k \in [1, n]}$ l'ensemble des individus où $z_k = (z_{k1}, z_{k2}, \dots, z_{kp})^t$ est un individu décrit par p caractéristiques (i.e. $z_k \in R^p$). Soit une partition nette déterminée sur 2^Ω . Nous définissons la frontière d'une classe ω_i avec $i \in [1, c]$ comme :

$$F_{\omega_i} = \{z_k \in A : \forall A \in 2^\Omega \setminus \{\emptyset, \{\omega_i\}\} \text{ et } \omega_i \in A\}$$

Dans une partition floue, chaque individu z_k peut appartenir avec un certain degré aux frontières de toutes les classes. Aussi, en nous appuyant sur la définition précédente, nous proposons une mesure de la taille de la frontière floue de la classe ω_i :

$$L_{\omega_i}(\mu, v) = \sum_{k=1}^n \sum_{A \in 2^\Omega \setminus \{\emptyset, \{\omega_i\}\}, \omega_i \in A} \mu_{Ak}^m d_{Ak}^2$$

$$= \sum_{k=1}^n \sum_{A \ni \omega_i, |A| > 1} \mu_{Ak}^m d_{Ak}^2$$

où d_{Ak} est la distance de l'individu z_k à la classe d'ambiguïté $A \subseteq \Omega$, m est un facteur de flou et μ_{Ak} est le degré d'appartenance de l'individu z_k à la classe d'ambiguïté. Le choix de d_{Ak} sera discuté dans la partie 2.3. Pour des hautes valeurs de d_{Ak} , si μ_{Ak} prend des valeurs importantes alors la taille de la frontière floue de la $i^{\text{ème}}$ classe doit être importante.

La définition suivante précise la frontière entre plusieurs classes :

Définition 5

Soit $\Omega = (\omega_i)_{i \in [1, c]}$ un ensemble de classes. Soit $Z = (z_k)_{k \in [1, n]}$ l'ensemble des individus où $z_k = (z_{k1}, z_{k2}, \dots, z_{kp})^t$ est une forme décrite par p caractéristiques (i.e. $z_k \in R^p$). Soit une partition nette déterminée sur 2^Ω . Nous définissons la frontière entre les $t > 1$ classes simples qui constituent $\Omega_t = \{\omega_1, \dots, \omega_t\} \subseteq \Omega$ comme :

$$F_{\Omega_t} = \{z_k \in A : A \supseteq \Omega_t\}$$

Ceci peut se généraliser au cas d'une partition floue. Nous proposons de définir la taille d'une frontière floue entre les $t > 1$ classes simples qui constituent $\Omega_t = \{\omega_1, \dots, \omega_t\} \subseteq \Omega$ par :

$$L_{\Omega_t}(\mu, v) = \sum_{k=1}^n \sum_{A \supseteq \Omega_t} \mu_{Ak}^m d_{Ak}^2$$

Cette définition est symétrique par rapport à $\{\omega_i \in \Omega_t\}$. Cette propriété est importante dans un objectif de reconnaissance de formes. La définition précédente est à rapprocher de la définition de l'adjacence floue définie entre deux sous-ensembles flous [6].

Parce qu'une frontière est commune à plusieurs classes, nous définissons la taille totale des frontières floues dans une partition floue comme :

$$J_m^{inter}(\mu, v) = \sum_{k=1}^n \sum_{A, |A| > 1} \mu_{Ak}^m d_{Ak}^2.$$

2.2. régularisation dans le problème de la coalescence

Du point de vue de la régularisation, le problème de la coalescence est dit « mal posé » s'il ne satisfait pas un ou plus des critères suivants : la solution existe, est unique et dépend continuellement des données. Des hypothèses *a priori* additionnelles doivent être imposées sur la solution pour convertir un problème mal posé en un problème bien posé. Lorsque l'on utilise la théorie des champs de Markov en traitement d'images, l'hypothèse de lissage est incorporée dans la fonction d'énergie par laquelle le coût de la solution est défini. L'utilisation de fonctions constantes par morceaux utilisées dans ces modèles permet de tenir compte des discontinuités, notamment lors des changements de niveaux de gris importants. Par exemple, Pavlidis propose dans [31] une fonction d'énergie dans laquelle est prise en compte, à travers la variable binaire $L_i(z)$, la frontière (i.e. discontinuité) de la région O_i :

$$E_\lambda = \sum_{i=1}^c \left(\sum_{z \in O_i} (u_i - g(z))^2 + \lambda L_i(z) \right)$$

où $g(z)$ est le niveau de gris de z ; c est le nombre de régions et u_i est le label associé à la région O_i . La frontière F_{O_i} est définie par : $F_{O_i} = \{z \in O_i : L_i(z) = 1\}$. Ce modèle assure un moyennage par morceau. Le lissage ne s'applique pas aux points où l'amplitude du gradient excède une certaine valeur. Chaque discontinuité est pénalisée par la quantité λ . Si $(u_i - g(z))^2 < \lambda$ alors il est moins coûteux de payer le prix $(u_i - g(z))^2$ et de mettre $L_i = 0$; autrement il est plus économique de mettre $L_i = 1$ pour insérer une discontinuité avec le coût λ .

Ce concept de régularisation peut être étendu au problème de la coalescence floue. Nous avons défini précédemment à partir du concept d'ambiguïté, et ceci pour une partition floue sur 2^Ω , les frontières floues entre classes. Nous proposons d'ajouter au critère de classification de l'algorithme FcM, un terme analogue à $\lambda L_i(z)$ et correspondant à une mesure de la taille de la frontière floue d'une classe. Aussi, nous définissons le critère de classification comme :

$$J_m^1(\mu, v) = J_m^{fcM}(\mu, v) + J_m^{inter}(\mu, v)$$

$$= \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m \|z_k - v_i\|^2 + \sum_{k=1}^n \sum_{A, |A| > 1} \mu_{Ak}^m d_{Ak}^2$$

$(u_i, g(z))$ dans l'expression de E_λ correspond à (v_i, z_k) dans $J_m^{fcm}(\mu, v)$ avec $\mu_{ik} = 1$ ou 0 (classification nette). Les discontinuités mesurées par $\sum_{i=1}^c \sum_{z \in O_i} \lambda L_i(z)$ correspondent aux frontières floues définies à partir des classes d'ambiguïté et mesurées par $J_m^{inter}(\mu, v) = \sum_{k=1}^n \sum_{A, |A|>1} \mu_{Ak}^m d_{Ak}^2$. Ce modèle est donc équivalent au modèle énergétique de Pavlidis. Comme nous le verrons par la suite, cette approche permet d'ajuster les degrés d'interactions entre classes simples à l'aide de fonctions d'appartenance définies sur les classes ambiguës.

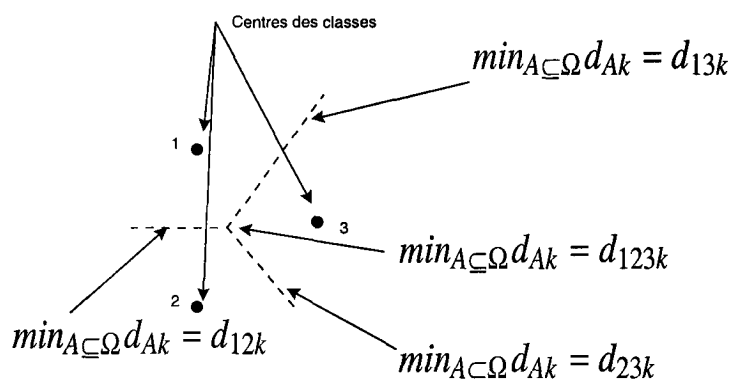


Figure 3. – Lieu où la distance aux classes d'ambiguïté est minimale. $d_{\{i,\dots,j\}k}$ représente la distance de l'individu z_k à la classe d'ambiguïté définie par $\{\omega_i, \dots, \omega_j\} = \omega_{i..j}$.

2.3. distances aux classes d'ambiguïté

Un individu est considéré comme ambigu s'il appartient à la frontière de deux ou plusieurs classes. Nous définissons la distance de l'individu z_k à la classe d'ambiguïté $A \subseteq \Omega, |A| > 1$ par :

$$d_{Ak} = \frac{\alpha_A \left(\sum_{\omega_i \in A} d_{ik} \right)^2}{|A|^2 \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}} \quad (5)$$

où, le terme α_A est un nombre positif. d_{ik} est la distance classique utilisée dans l'algorithme des FcM. Cette définition vérifie les propriétés suivantes :

• $d_{Ak} = \alpha_A \frac{(\langle d_{ik} \rangle_T^A)^2}{\langle d_{ik} \rangle_G^A}$ où $\langle d_{ik} \rangle_T^A$ est la moyenne arithmétique des $d_{ik}, \omega_i \in A$ définie comme :

$$\langle d_{ik} \rangle_T^A = \frac{1}{|A|} \sum_{\omega_i \in A} d_{ik}$$

et $\langle d_{ik} \rangle_G^A$ est la moyenne géométrique des $d_{ik}, \omega_i \in \Omega$ définie comme :

$$\langle d_{ik} \rangle_G^A = \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}$$

Ces moyennes vérifient les propriétés suivantes vis-à-vis des opérateurs $\langle d_{ik} \rangle_{\min} = \min_i d_{ik}$ et $\langle d_{ik} \rangle_{\max} = \max_i d_{ik}$:

$$\langle d_{ik} \rangle_{\min} \leq \langle d_{ik} \rangle_G \leq \langle d_{ik} \rangle_T \leq \langle d_{ik} \rangle_{\max}$$

La figure 3, pour $\alpha_A = 1 \forall A \subseteq \Omega$, illustre le comportement de cette distance dans le cas de 3 classes. Les lignes en pointillés représentent le lieu d'ambiguïté totale vérifiant :

$$\min_{A, |A|>1} d_{Ak} \leq d_{ik} \quad \forall i \in \{1, \dots, c\} \quad (6)$$

• d_{Ak} est homogène à une distance : $d_{Ak} = K \langle d_{ik} \rangle_T^A$, avec K rapport de deux distances.

• on ne considère pas le cas où $|A| = 1$, une classe ne pouvant être ambiguë avec elle-même. Cependant on retrouve $d_{Ak} = d_{ik}$ dans le cas où $\alpha_A = 1$ et lorsque l'individu est proche de la classe ω_i : l'opérateur $\langle d_{jk} \rangle_G$ se comportant comme l'opérateur $\min_j d_{jk}$.

Nous considérons les régions ambiguës comme des discontinuités, puisqu'il s'agit de régions de transitions ou de doutes entre deux ou plusieurs classes caractérisées chacune par des fonctions d'appartenance différentes. La détermination des classes d'ambiguïté permet de modéliser ces transitions. d_{Ak} joue le rôle du gradient en traitement d'images détectant les frontières entre objets. Par analogie avec les modèles des membranes [5], le terme α_A^{-1} représente le facteur de discontinuité et joue sur l'interdépendance des fonctions d'appartenance.

2.4. algorithme des c-moyennes floues étendues

Afin d'introduire le concept d'ambiguïté dans le problème de la coalescence floue, nous définissons un nouvel algorithme fondé sur la minimisation du critère de classification suivant :

$$\begin{aligned}
 J_m^1(\mu, v) &= J_m^{fcm}(\mu, v) + J_m^{inter}(\mu, v) \\
 &= \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m d_{ik}^2 + \sum_{k=1}^n \sum_{A \subseteq \Omega, |A|>1} \mu_{Ak}^m d_{Ak}^2 \quad (7)
 \end{aligned}$$

qui s'écrit encore :

$$J_m^1(\mu, v) = \sum_{k=1}^n \sum_{A \subseteq \Omega} \mu_{Ak}^m d_{Ak}^2 \quad (8)$$

avec les contraintes suivantes :

$$\forall k \in [1, n] \left\{ \begin{array}{l} \sum_{i=1}^c \mu_{ik} + \sum_{A, |A|>1} \mu_{Ak} = \sum_{A \subseteq \Omega} \mu_{Ak} = 1 \\ \frac{\sum_{A, |A|>1} \mu_{Ak}}{n} = \rho_\alpha \\ 0 \leq \mu_{Ak} \leq 1 \quad \forall A \subseteq \Omega \end{array} \right. \quad (9)$$

Pour les classes simples, la partition générée satisfait

$$\sum_{i=1}^c \mu_{ik} \leq 1 \quad \forall k.$$

2.4.1. effet régularisant de la fonction objective

Le critère de classification s'écrit en fonction des distances des individus aux classes simples et ambiguës. Si la distance à la classe d'ambiguïté d_{Ak} est importante, il est nécessaire pour minimiser J_m^{inter} que μ_{Ak} soit faible. Inversement si la distance à la classe d'ambiguïté est faible μ_{Ak} peut prendre des valeurs importantes (sous respect des contraintes $\mu_{Ak} < 1$ et $\sum_{A \subseteq \Omega} \mu_{Ak} = 1$). La

minimisation de J_m^{inter} conduit à rechercher des prototypes qui ne sont pas des centres de classes mais qui représentent les frontières entre régions. Minimiser J_m^{inter} conduit donc à minimiser les frontières entre classes. Dans le cas où $\alpha_A = \alpha$, on peut réécrire le critère de la façon suivante :

$$J_m^1(\mu, v) = J_m^{fcm}(\mu, v) + \alpha J_m^{inter'}(\mu, v)$$

avec

$$\begin{aligned} J_m^{inter'}(\mu, v) &= \sum_{k=1}^n \sum_{A, |A|>1} \mu_{Ak}^m d_{Ak}^2 \text{ et } d'_{Ak} \\ &= \frac{d_{Ak}}{\alpha} \quad (J_m^{inter'} \text{ ne dépend plus de } \alpha) \end{aligned}$$

Si $d_{Ak}^2 > d_{ik}^2$ et $\alpha = 1$, il est plus intéressant d'avoir $\mu_{ik} > \mu_{Ak}$ (de classer l'individu z_k dans une classe simple). Inversement, si $d_{Ak}^2 < d_{ik}^2$, il est plus économique de considérer z_k comme ambigu avec un coût μ_{Ak}^m .

α permet de pondérer le rapport entre J_m^{fcm} et $J_m^{inter'}$, le coût d'ambiguïté est alors $(\alpha \mu_{Ak}^m)$.

2.4.2. classe d'ambiguïté globale

La première contrainte impose aux degrés d'appartenance de se situer sur l'hyperplan défini par $\sum_{A \in 2^\Omega \setminus \emptyset} \mu_{Ak} = 1$.

Nous définissons le degré d'appartenance $\mu_{(c+1)k}$ de l'individu z_k à la classe d'ambiguïté globale notée $\omega_{(c+1)}$ comme l'agrégation,

notée F , des degrés d'appartenance de cet individu à chaque classe composée :

$$\mu_{(c+1)k} = \frac{F}{\sum_{A, |A|>1} \mu_{Ak}}$$

L'opérateur d'agrégation F peut être choisi à partir d'un grand nombre de normes et de conormes triangulaires réalisant les trois comportements classiques : conjonctif, disjonctif et de compromis [8]. Nous considérons le rejet d'ambiguïté comme un processus de rejet sélectif [22]. Les notions de rejet sont désirables dans des applications où il est plus coûteux de prendre une mauvaise décision que de refuser d'en prendre. Le rejet sélectif ou rejet d'ambiguïté est donc une extension du rejet simple : lorsqu'un individu ne peut être assigné à une des c classes de façon certaine, nous ne rejetons pas cet individu de toutes les classes mais seulement de certaines. Par exemple, pour un individu proche du plan de séparation des classes 1 et 2, et très loin du centre de la troisième classe, nous rejetons seulement la troisième classe et déclarons qu'il appartient au groupe composé de la première et seconde classe (i.e. z_k appartient à la classe d'ambiguïté ω_{12}). Une forme est donc ambiguë si et seulement si elle est *partiellement ambiguë* entre $t > 1$ classes simples.

Aussi, pour définir la classe d'ambiguïté globale, nous utilisons un opérateur de disjonction : l'opérateur addition. Nous pouvons remarquer que le critère s'écrit dans ce cas :

$$J_m^1(\mu, v) = \sum_{i=1}^{c+1} \sum_{k=1}^n \mu_{ik}^m d_{ik}^2 \quad (10)$$

Dans le but de paramétrer l'algorithme proposé, nous définissons

la deuxième contrainte $\frac{\sum_{k=1}^n \mu_{(c+1)k}}{n} = \rho_\alpha$ où $\rho_\alpha \in [0, 1]$ correspond au taux de rejet d'ambiguïté globale sur la totalité des individus. Cette valeur est plus simple à déterminer qu'une valeur de seuil puisqu'elle correspond à un pourcentage d'individus à rejeter que l'opérateur choisit.

Parce que les mesures d'appartenance sont normalisées, nous avons : $\mu_{(c+1)k} = \sum_{A, |A|>1} \mu_{Ak} \leq 1$

2.4.3. regroupement hiérarchique

Le nombre de classes recherchées dans l'algorithme Fc+2M est 2^Ω où Ω représente l'ensemble des classes simples. Ceci peut être prohibitif pour certaines applications où un grand nombre de classes est imposé. Par exemple en reconnaissance de caractères, le nombre total de classes est de 62. Pour les caractères chinois ou persans [39], la taille est encore bien plus importante. De plus, pour les ensembles flous la relation entre la dimension de l'espace métrique, n , et la dimension de la frontière d'un voisinage ouvert qui existe en classification nette n'est plus vérifiée. La frontière entre plusieurs classes peut être de dimension n . Cependant en pratique le nombre d'individus appartenant à

la frontière de t classes décroît rapidement lorsque $t \gg n$. Dans le cas d'un problème de classification où il existe un nombre important de classes, une grande partie d'entre elles n'a pas de frontières communes. Il est donc inutile d'appliquer l'algorithme Fc + 2M directement sur une telle structure de données. Nous proposons donc d'utiliser dans ce cas une méthode de regroupement hiérarchique décrite par les phases suivantes :

- subdiviser l'espace de représentation en méta-classes en utilisant la méthode des nuées dynamiques par exemple ou une de ses nombreuses variantes [16]. Cet algorithme nécessite cependant de partir de conditions initiales proches de la solution, et en particulier d'avoir une bonne estimation du nombre de méta-classes dans les données. Dans le cas où cette information n'est pas disponible, il est recommandé d'initialiser l'algorithme des nuées dynamiques après analyse d'une taxinomie numérique [33], [24];
- appliquer la méthode de coalescence des Fc + 2M sur chaque méta-classe. Lors des itérations successives, il est possible de supprimer les classes d'ambiguïté qui ne vérifient pas, par exemple, le critère suivant : $\sum_{k=1}^n \mu_{Ak} > s$ où s est un seuil.

2.5. théorème de l'algorithme des c-moyennes floues étendues

La minimisation du critère (8) avec les contraintes (9) donne les solutions suivantes (la preuve de ce théorème se trouve en annexe A) :

$$\forall k \in [1, n], \forall A \subseteq \Omega \quad \mu_{Ak}^* = \frac{1}{\sum_{B \subseteq \Omega} \left(\frac{d_{Ak}}{d_{Bk}} \right)^{\frac{2}{m-1}}}$$

$$\forall k \in [1, n], \forall A \subseteq \Omega \quad v_A^* = \frac{\sum_{k=1}^n \mu_{Ak}^m x_k}{\sum_{k=1}^n \mu_{Ak}^m}$$

où $\forall A = \{\omega_i\}, i \in [1, c]$ $d_{Ak} = \|x_k - v_A\|_G$ est la distance classique utilisée dans l'algorithme des c-moyennes floues, m est un exposant contrôlant la quantité de flou de la partition résultante, et $\forall A \subseteq \Omega, |A| > 1$

$$d_{Ak} = \frac{\alpha \left(\sum_{\omega_i \in A} d_{ik} \right)^2}{|A|^2 \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}} \quad (11)$$

où, le terme α est un nombre positif. Il joue le rôle d'un facteur d'échelle. La détermination de α à partir du taux de rejet d'ambiguïté globale est définie dans la preuve.

Soit $d_{(c+1)k}$ la distance à la classe d'ambiguïté globale, nous définissons :

$$\left(\frac{1}{d_{(c+1)k}} \right)^{\frac{2}{m-1}} = \sum_{A, |A| > 1} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}}$$

3. le rejet d'ambiguïté et de distance simultanément

Nous proposons une approche parallèle pour la gestion des deux types de rejet qualifiés rejet de distance et d'ambiguïté. Ils sont considérés simultanément et définissent deux classes supplémentaires.

En reconnaissance des formes, un individu x_k localisé trop loin d'une classe doit posséder une faible valeur d'appartenance vis-à-vis de cette classe. Dans le but de réduire le risque de mauvaise classification, le rejet de distance est utilisé pour éviter de classer cet individu. Dans la littérature, les règles proposées dans le but de rejeter ces individus, sont le plus souvent basées sur des valeurs de seuil lors de l'étape de décision ou classement (une valeur de seuil pour chaque classe par exemple) et non dans l'étape de classification. Aussi, pour les algorithmes basés sur les FcM, où les degrés d'appartenance de x_k sont normalisés, la classe de rejet ne peut être définie. En effet, si x_k se situe trop loin des centres alors il est localisé à la même distance de toutes les classes. Aussi $\mu_{ik} \cong \frac{1}{c} \forall i \in [1, c]$. (cf.figure 4).

Pour éviter ce problème, Davé propose un algorithme de coalescence robuste, appelé NC (noise cluster), modifiant le critère de classification utilisé dans l'approche FcM en incluant une classe de bruit [12], [13]. Dans cette approche, tous les individus sont

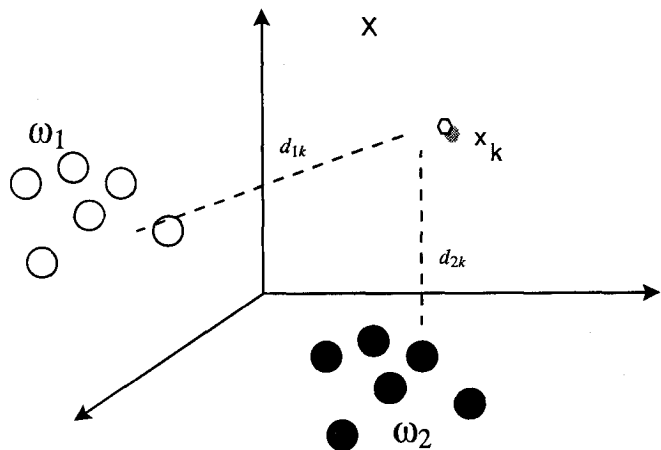


Figure 4. - x_k ne peut être rejeté en distance dans l'algorithme des FcM au niveau de l'étape de décision car les degrés d'appartenance sont normalisés.

correspond au taux de rejet d'ambiguïté globale fixé également sur la totalité des individus. Les autres contraintes permettent une interprétation probabiliste des fonctions d'appartenance.

3.1. théorème des c+2-moyennes floues

La minimisation du critère (13) avec les contraintes (14) donne les solutions suivantes (le paramètre r a disparu : il est égal au produit mc pour une solution simple du problème. La preuve de ce théorème figure en annexe B) :

$$\forall k \in [1, n], \forall A \in 2^\Omega \cup \Theta \setminus \emptyset \quad \mu_{Ak}^* = \frac{1}{\sum_{B \in 2^\Omega \cup \Theta \setminus \emptyset} \left(\frac{d_{Ak}}{d_{Bk}} \right)^{\frac{2}{m-1}}}$$

$$\forall k \in [1, n], \forall A \in 2^\Omega \setminus \{\emptyset, \Theta\} \quad v_A^* = \frac{\sum_{k=1}^n \mu_{Ak}^m x_k}{\sum_{k=1}^n \mu_{Ak}^m}$$

où $\forall A = \{\omega_i\}, i \in [1, c] \quad d_{Ak} = \|x_k - v_A\|_G$ est la distance classique utilisée dans l'algorithme des FcM, m est un exposant contrôlant la quantité de flou de la partition résultante, et $\forall A \subseteq \Omega, |A| > 1$

$$d_{Ak} = \frac{\alpha \left(\sum_{\omega_i \in A} d_{ik} \right)^2}{|A|^2 \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}} \quad (15)$$

où, le terme α est un nombre positif. Il joue un rôle de facteur d'échelle.

$d_{\Theta k}$ est défini par :

$$d_{\Theta k} = \frac{T(\rho)^2}{\left(\prod_{i=1}^c d_{ik} \right)^{\frac{1}{c}}} \quad (16)$$

où $T(\rho)$ est le seuillage qui peut être, soit fixé, soit déterminé par le taux de rejet global ρ . La détermination de α et de T , en fonction des taux de rejet de distance et d'ambiguïté sur la totalité des individus, est décrite dans l'annexe B (voir aussi [15] pour plus de détails). Soit $d_{(c+1)k}$ la distance à la classe d'ambiguïté globale, nous définissons :

$$\left(\frac{1}{d_{(c+1)k}} \right)^{\frac{2}{m-1}} = \sum_{A, |A| > 1} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}}$$

3.2. comparaison avec les approches PcM et NC

La fonction d'appartenance ou la distribution de possibilité affectée à la classe de rejet de distance est gérée différemment dans les trois algorithmes PcM, NC et Fc+2M. Dans l'approche NC, l'appartenance d'un individu x_k à la classe de rejet de distance est égale au complément de la somme des degrés d'appartenance de x_k à toutes les classes. De plus, la distance de tous les individus à la classe de rejet est constante et vaut δ .

Pour l'algorithme PcM, il y a une classe de rejet par classe ω_i ; le degré d'appartenance de x_k à la classe de rejet est alors le complément du degré d'appartenance à la classe ω_i .

Pour l'algorithme Fc+2M, il est défini, comme pour l'algorithme PcM, une classe de rejet par classe simple ω_i . Le degré d'appartenance de l'individu x_k à la classe $\bar{\omega}_i$ est défini par une fonction f_i ($f_i(\mu_{ik}) = \mu_{ik}$) où f_i est une fonction monotone décroissante. La distance d_{ik} n'est plus constante, elle est calculée comme l'inverse de la distance à la classe simple ω_i . Une autre différence avec les algorithmes précédents vient du fait que l'on calcule le degré d'appartenance d'un individu à la classe de rejet global de distance comme une agrégation des degrés d'appartenance de cet individu aux c classes $(\bar{\omega}_i)_{i \in [1, c]}$ (un individu est rejeté s'il est globalement rejeté par toutes les classes). Ceci semble plus conforme au raisonnement. De plus, contrairement aux algorithmes NC et PcM, le degré d'appartenance à la classe de rejet global de distance s'écrit de façon identique à celui des autres classes :

$$\mu_{\Theta k} = \frac{1}{\sum_{B \in 2^\Omega \cup \Theta \setminus \emptyset} \left(\frac{d_{\Theta k}}{d_{Bk}} \right)^{\frac{2}{m-1}}}$$

3.3. discussion sur la robustesse de l'algorithme

Les méthodes statistiques robustes, comme par exemple le M estimateur (maximum likelihood type estimator [38]), apportent des solutions intéressantes pour les problèmes dans lesquels les hypothèses sous-jacentes (distributions indépendantes, symétriques ...) sont inexactes. Un algorithme robuste doit être insensible au non respect de ces hypothèses dû aux bruits et aux artefacts. Soit $h(\eta)$, la fonction de pondération s'appliquant sur les erreurs $\eta_k = x_k - v$, ($k = 1, \dots, n$) liées aux observations x_k ($k = 1, \dots, n$)

pour estimer le vecteur de paramètres v : $v^* = \frac{\sum_k h(\eta_k) x_k}{\sum_k h(\eta_k)}$. Une condition nécessaire de robustesse est [38] :

$$\lim_{n \rightarrow \infty} |2\eta h(\eta)| = c \quad (17)$$

où $c \in [0, \infty)$ est une constante. On retrouve cette condition sur les modèles de régularisation en segmentation d'images s'adaptant

aux discontinuités entre régions [28]. On souhaite de plus que $h(\eta)$ soit petit pour des grandes valeurs de $|\eta|$ et tende vers 0 lorsque $|\eta|$ tend vers l'infini. Une condition souhaitable et qui est souvent demandée à un bon estimateur est que la fonction de pondération soit monotone décroissante en fonction de l'erreur ($h'(\eta) < 0$).

Davé et Krishnapuram, dans [17], montrent que les fonctions de pondération utilisées dans le M estimateur peuvent être comparées aux fonctions d'appartenance de l'algorithme NC1 (algorithme NC appliqué à la recherche d'une seule classe à la fois) ou aux distributions de possibilités données par l'algorithme PCM. Ces deux algorithmes conduisent au modèle suivant d'appartenance du point x_k au centre de la classe ω_1 :

$$\mu_{1k} = \frac{1}{1 + \left(\frac{d_{1k}^2}{\delta}\right)^{\frac{1}{m-1}}} \quad (18)$$

où δ est un nombre positif. Ce modèle déjà choisi expérimentalement par Zimmerman et Zysno dans [41] s'adapte aux classes mal définies et vagues et vérifie les conditions précédentes. Dans ce cas, la fonction de pondération s'écrit : $h(\eta_k) = (\mu_{1k})^m$.

De même, nous montrons ici, l'équivalence en terme de robustesse de l'algorithme Fc+2M dans le cas où $c = 1$ (i.e. regroupement des individus dans 1 seule classe à la fois, ω_1 , et rejet en distance) avec l'approche PcM : le degré d'appartenance à la classe est alors (voir partie 3.1 avec $\rho_\alpha = 0$) :

$$\mu_{1k} = \frac{1}{\left(\frac{d_{1k}}{d_{1k}}\right)^{\frac{2}{m-1}} + \left(\frac{d_{1k}}{d_{\theta k}}\right)^{\frac{2}{m-1}}} = \frac{1}{1 + \left(\frac{d_{1k}}{d_{\theta k}}\right)^{\frac{2}{m-1}}} \quad (19)$$

où ω_Θ est la classe de rejet en distance. Or $d_{\theta k} = \frac{T(\rho)^2}{d_{1k}}$ et $T(\rho)$ est un paramètre dépendant du taux de rejet choisi. Nous pouvons donc écrire :

$$\mu_{1k} = \frac{1}{1 + \left(\frac{d_{1k}^2}{T(\rho)^2}\right)^{\frac{1}{m'-1}}} \quad (20)$$

avec $m' = \frac{m+1}{2}$ et $T(\rho)^2$, équivalent à δ , constant. La fonction de pondération s'écrit $h(\eta_k) = (\mu_{1k})^m$ et vérifie les propriétés précédentes. L'algorithme est donc une version robuste de l'algorithme FcM et les propriétés des fonctions d'appartenance données par (18), dans le cas où une seule classe est recherchée à la fois, expliquent le comportement robuste de l'algorithme Fc+2M vis-à-vis du bruit.

3.4. intérêt du rejet d'ambiguïté

La robustesse est également renforcée par le terme de régularisation J_m^{inter} portant sur les classes d'ambiguïté. En effet, pour l'algorithme FcM, les régions entre classes considérées comme

des régions de doute ou d'ambiguïté sont modélisées par le facteur de flou m . Si l'on veut, dans ce cas, modéliser correctement cette région, il est nécessaire d'avoir une valeur de m importante ($m = 2$). Par conséquent, la forme de la fonction d'appartenance n'est pas souhaitable : la distance pour laquelle le degré d'appartenance est de 0.5 est trop importante (voir [26] et conditions précédentes), ce qui ne permet pas de séparer correctement les points bruits des individus. Dans le cas de l'algorithme Fc+2M, les régions de doute sont modélisées par les classes d'ambiguïté, aussi, il est possible d'utiliser des valeurs plus faibles de m ($m = 1.5$, par exemple, correspond à la valeur habituellement choisie pour l'algorithme PcM). Par conséquent, les fonctions d'appartenance tendent à décroître plus rapidement lorsque la distance augmente.

L'apport essentiel de l'algorithme Fc+2M vient du fait qu'il attribue une fonction de pondération différente à chaque individu selon que celui-ci appartient clairement à une classe, à aucune ou bien à plusieurs classes de façon égale. Il permet d'inclure une modélisation de l'ambiguïté et de l'ignorance en affectant des fonctions d'appartenance à tous les sous-ensembles de classes de $2^\Omega \cup \Theta$ plutôt qu'aux éléments de Ω seulement. Il s'agit d'une approche particulièrement riche et souple, qui est capable de s'adapter à un très grand nombre de situations. Nous citerons plus particulièrement la segmentation d'images IRM décrite dans [7] qui se heurte au problème de l'initialisation des masses pour la classification des pixels par la règle orthogonale de Dempster-Shafer. Nous pensons que l'algorithme Fc+2M peut résoudre ce problème de la façon suivante.

La classification est effectuée seulement d'après les niveaux de gris, et la fusion est effectuée au niveau du pixel. Les degrés d'appartenance générés par l'algorithme Fc+2M peuvent s'interpréter comme des degrés de croyance [40]. Soit $X = (x_k)_{k \in [1,n]}$ l'ensemble des données où x_k représente le niveau de gris du pixel k . Nous construisons, à partir de la partition floue obtenue sur X par l'algorithme Fc+2M, une structure de croyance sur le pixel k :

$$\begin{cases} m^k(\emptyset) = 0 \\ m^k(A) = \mu_{Ak} \quad \forall A \in 2^\Omega \cup \Theta \setminus \emptyset \\ \sum_{A \in 2^\Omega \cup \Theta} m^k(A) = 1 \end{cases} \quad (21)$$

Nous nous plaçons dans le cas du monde fermé, i.e. hypothèse d'exhaustivité de Ω , mais nous introduisons de façon explicite une classe supplémentaire qui est la classe de rejet de distance ω_Θ représentant l'ensemble des classes inconnues de Ω . Posons $\alpha = 1$ dans (15), l'hypothèse d'ambiguïté totale (i.e. entre toutes les classes de Ω) du pixel x_k implique : $\mu_{Ak} = a \quad \forall A \in 2^\Omega \setminus \emptyset$ avec a constant. Nous obtenons donc l'initialisation des masses suivante : $m_A^k = a \quad \forall A \in 2^\Omega \setminus \emptyset$, $m(\Theta) = \mu_{\Theta k}$. Ce résultat peut s'étendre aisément au cas des ambiguïtés partielles entre les $t > 1$ classes qui constituent $\Omega_t = \{\omega_1, \dots, \omega_t\} \subseteq \Omega$. Nous trouvons alors : $m_A^k = a' \quad \forall A \in 2^{\Omega_t} \setminus \emptyset$, $m(\Theta) = \mu_{\Theta k}$; la masse affectée aux autres hypothèses est donnée par $\mu_{Ak} < a'$, (i.e. les classes correspondant à ces hypothèses souffrent de moins d'ambiguïté).

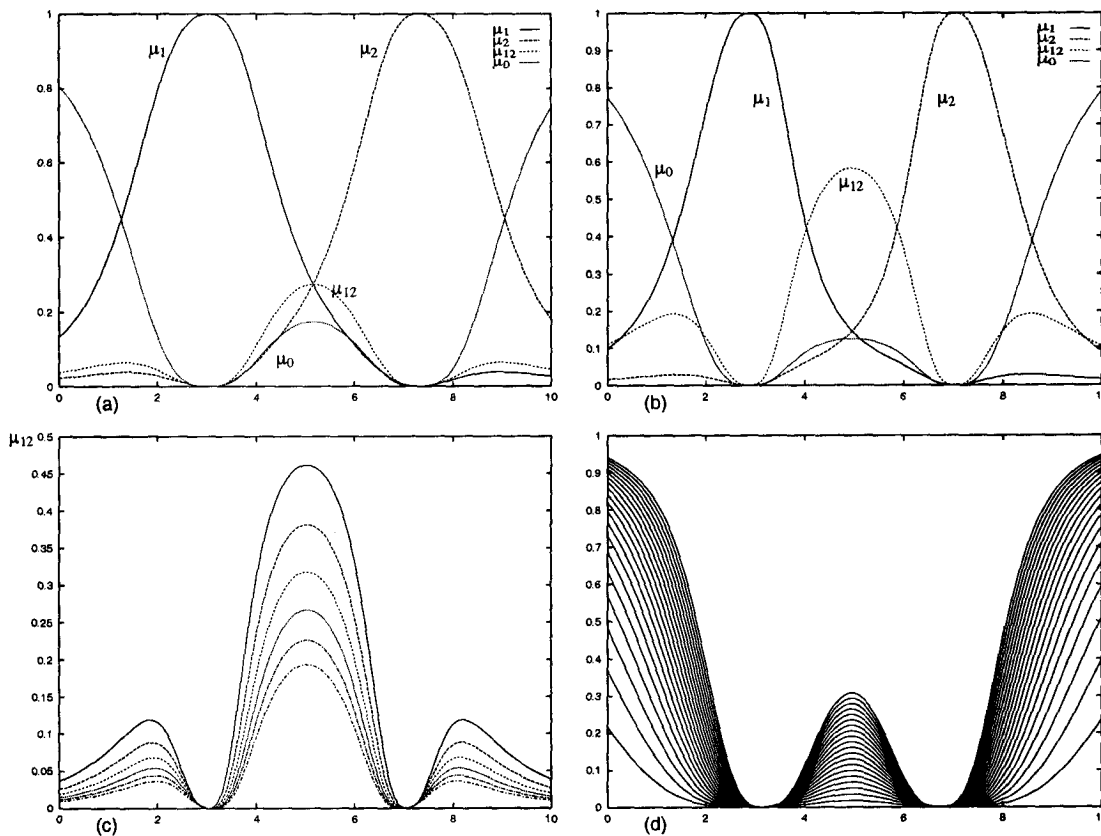


Figure 5. – Cas de deux classes gaussiennes normalisées. (a) fonctions d'appartenance obtenues avec $\alpha = 1$ and $\rho = 0.10$; (b) fonctions d'appartenance obtenues avec $\alpha = 0.5$ et $\rho = 0.10$; (c) évolution de $\mu_{12}(x)$ en fonction de $\alpha \in [0.5, 1]$ et $\rho = 0.4$; (d) évolution de $\mu_0(x)$ en fonction de $\rho \in [0, 0.02]$.

Ce résultat est analogue à celui rencontré dans [7] fondé sur une approche par raisonnement plus difficile à automatiser. En ce sens, l'algorithme Fc+2M permet une approche riche et flexible.

4. exemples numériques

4.1. cas 1-D : 2 classes, 1 classe de rejet en ambiguïté, 1 classe de rejet de distance

Nous présentons dans cette section un exemple simple montrant l'intérêt du rejet d'ambiguïté. Dans un premier temps, nous discutons de la forme des fonctions d'appartenance dans le cas d'un espace de représentation à une dimension. Dans un second temps, nous discutons de la robustesse de la méthode. Nous représentons sur la figure 5a, les fonctions d'appartenance obtenues ($\mu_\emptyset, \mu_{12}, \mu_1, \mu_2$) dans le cas d'individus appartenant à deux classes gaussiennes normalisées de moyenne respective 3 et 7. 200 points ont été générés. Les résultats ont été déterminés pour

une seule initialisation aléatoire sur $\mu^{(0)}$ obtenue à partir d'une distribution uniforme sur $[0, 1]$. Le nombre maximum d'itération est de 100 et le critère d'erreur est fixé à $\epsilon = 0.00001$. La gestion de la région de doute, située entre les classes, étant maintenant faite par les fonctions d'appartenance associées aux classes d'ambiguïté, m doit être plus faible que la valeur classique utilisée ($m = 2$). Nous prendrons $m = 1.5$ pour avoir une forme de fonction d'appartenance acceptable dans un but de régularisation [26]. Nous utilisons la norme euclidienne.

La partition est obtenue avec un taux de rejet global de distance de 10% et pour une valeur de $\alpha = 1$. Nous notons μ_{12} la fonction d'appartenance à la classe d'ambiguïté ω_{12} associée à la classe composée $\{\omega_1, \omega_2\}$. Les fonctions d'appartenance μ_\emptyset et μ_{12} possèdent la forme attendue. Nous notons un point particulier qui est totalement ambigu caractérisé par l'égalité : $\mu_{12k} = \mu_{1k} = \mu_{2k}$. Ceci s'explique de la façon suivante. De la relation

$$d_{ijk} = \alpha \frac{(d_{1k} + d_{2k})^2}{4 \left(\prod_{u=1}^c d_{uk} \right)^{\frac{1}{c}}}$$

et d'après la définition 1, un individu totalement ambigu vérifie $d_{1k} = d_{2k}$, aussi $\left(\prod_{u=1}^c d_{uk} \right)^{\frac{1}{c}} = d_{1k}$

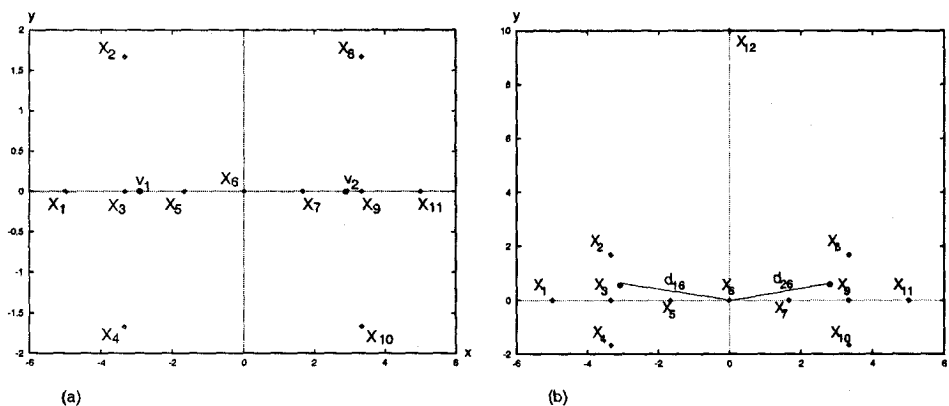


Figure 6. – Ensemble de données : (a) X_{11} ; (b) X_{12} .

Alors ($\alpha = 1$)

$$d_{ijk} = \frac{(d_{1k})^2}{(d_{1k})} = d_{1k}$$

ce qui signifie que le point est situé à la même distance des deux classes et de la classe d'ambiguïté ω_{12} , ce qui est bien représenté sur la figure 5a.

La figure 5b présente le cas où $\alpha = 0.5$ avec le même taux de rejet que dans le cas précédent. Comme montré précédemment, un point situé à la même distance des 2 classes donne : $\mu_{1k} \leq \mu_{12k}$ et $\mu_{1k} = \mu_{2k}$. Nous remarquons que la classe de rejet ω_{Θ} n'est pas sensible aux modifications de α .

La figure 5c illustre le comportement de la fonction d'appartenance à la classe de rejet d'ambiguïté pour différentes valeurs de α et ρ fixé. Le rejet d'ambiguïté introduit une discontinuité entre les fonctions d'appartenance aux classes ω_1 et ω_2 . La dépendance entre ces fonctions d'appartenance est fonction de α . Plus α est faible, plus les centres générés représentent de manière intrinsèque les propriétés stables d'une classe indépendamment des autres classes.

Pour comparaison, la figure 5d issue de [15] présente le comportement de la fonction d'appartenance à la classe de rejet global de distance en fonction de ρ .

Pour tester la robustesse de notre algorithme vis-à-vis de l'algorithme FcM, nous avons calculé la moyenne sur 5 essais des erreurs d'estimation des prototypes dans le cas des deux distributions précédentes. Pour l'algorithme Fc+2M, nous avons obtenu une erreur moyenne de 0.05, pour l'algorithme FcM une erreur de 0.08, ceci pour un rejet de distance de 0.05 et $\alpha = 1$. Ceci s'explique par le fait que la forme des fonctions d'appartenance dans le cas Fc+2M est plus réaliste que dans le cas FcM : à l'infini la valeur de $\mu_{\Theta k}$ tend vers 1 et celles des $(\mu_{ik})_{i \in \{1, \dots, c\}}$ tendent vers 0. Deux biais sur la position des centres sont introduits par l'algorithme FcM : le premier induit par la forme des fonctions d'appartenance à l'infini tend à éloigner artificiellement les centres des classes, le second introduit par la dépendance entre les fonctions d'appartenance tend à les rapprocher. Malgré leur importance,

ces deux effets ont tendance à s'annihiler dans le cas 1-D, ce qui explique la faible erreur obtenue par l'algorithme FcM. Ces deux effets n'existent pas dans l'algorithme Fc+2M. Les résultats ont été obtenus avec une seule initialisation aléatoire pour chaque essai.

4.2. le cas des données diamonds

Dans cet exemple, nous utilisons l'ensemble des données bi-dimensionnelles présenté par Pal *et al* dans [30] (cf. figure 6a et 6b) dont les coordonnées sont données dans le tableau 2.

Ces données sont séparées en deux sous-ensembles : X_{11} et $X_{12} = X_{11} \cup \{x_{12}\}$ où x_{12} représente un point éloigné des deux classes. Ces données représentent deux classes en forme de *diamants* composées de cinq points chacun. Pour montrer l'intérêt du rejet en ambiguïté, nous avons appliqué les mêmes tests effectués dans [30] en utilisant les $c + 2$ -moyennes floues sur les deux ensembles de données X_{11} et X_{12} . Dans un but de comparaison, nous reprenons le protocole expérimental décrit dans [30].

Tableau 2. – Fonctions d'appartenance aux deux classes ($m = 2, \rho = 0$).

Pt	Données x y	FcM sur X_{11}		FcM sur X_{12}	
		μ_{1k}	μ_{2k}	μ_{1k}	μ_{2k}
1	-5.00 0.00	0.95	0.05	0.93	0.07
2	-3.34 1.67	0.94	0.06	0.97	0.03
3	-3.34 0.00	1.00	0.00	0.99	0.01
4	-3.34 -1.67	0.94	0.06	0.90	0.10
5	-1.67 0.00	0.91	0.06	0.92	0.08
6	0.00 0.00	0.50	0.50	0.50	0.50
7	1.67 0.00	0.09	0.91	0.08	0.92
8	3.34 1.67	0.06	0.94	0.03	0.97
9	3.34 0.00	0.00	1.00	0.01	0.99
10	3.34 -1.67	0.06	0.94	0.10	0.90
11	5.00 0.00	0.05	0.95	0.06	0.94
12	0.00 10.00			0.50	0.50

4.2.1. limite des FcM

Les tableaux 2 et 3 montrent le mauvais comportement des c -moyennes floues illustré par les valeurs d'appartenance x_6 et x_{12} . Bien que ces deux points soient éloignés d'une distance différente des deux centres, ils sont caractérisés par la même valeur d'appartenance. En particulier, le tableau 3 montre l'ensemble des points triés selon leur valeur d'appartenance dans le cas où $\rho = 0$ (i.e. cas FcM). La localisation médiane, dans ce tableau, des points x_6 et x_{12} , est conforme à celle attendue parce que ces points sont éloignés des deux classes. Cependant, le fait que ces points soient situés à des distances différentes vis-à-vis des classes 1 et 2 (cf. figure 6b) n'est pas pris en compte par les c -moyennes floues. Nous montrons dans le tableau 4 (respectivement 5) les résultats obtenus avec les $c + 1$ -moyennes floues pour $\rho = 0.05$ (respectivement $\rho = 0.4$).

Le point x_3 (respectivement x_9) appartient fortement à la classe 1 (respectivement, à la classe 2). x_6 et x_{12} ont globalement le même

comportement, i.e. leur fonction d'appartenance à la classe de rejet croît quand ρ croît. La différence essentielle entre ces deux points est l'amplitude de la dérivée de la fonction d'appartenance à la classe de rejet par rapport à ρ , i.e. x_{12} est plus rapidement rejeté que x_6 .

Aussi x_{12} est rejeté, x_6 non : ce point appartient de façon similaire aux deux classes simples. En effet x_6 n'appartient pas particulièrement à une des deux classes. Cela montre l'intérêt de construire une quatrième classe gérant l'ambiguïté de ce point.

Le tableau 6 présente les résultats obtenus dans [30] issus de l'algorithme des c -moyennes floues possibilistes. Les points sont triés par valeurs de typicalité, ainsi x_{12} est le moins représentatif des deux classes. Il s'agit d'un résultat similaire à celui obtenu avec les $c + 1$ -moyennes floues (voir tableau 5). En effet, celui-ci affecte à x_{12} la valeur d'appartenance la plus faible à chacune des deux classes. Ce point est donc le moins représentatif des deux classes. Ce qui est confirmé par la troisième colonne de la table : x_{12} est le plus rejeté.

Tableau 3. – Tri selon les fonctions d'appartenance aux deux classes ($m = 2$). Algorithme FcM.

	Classe 1	Classe 2
-	9	3
	8	2
	11	1
	7	5
	10	4
	6	6
	12	12
	4	10
	5	7
	1	11
	2	8
+	3	9

Tableau 5. – Tri selon les fonctions d'appartenance aux deux 2 classes et une classe de rejet ($m = 2, \rho = 0.4$). Algorithme F(c+1)M

	Classe 1	Classe 2	Classe de rejet
-	12	12	3
	9	3	9
	11	1	7
	10	2	5
	8	4	6
	7	5	2
	6	6	8
	1	11	4
	4	10	10
	2	8	1
	5	7	11
+	3	9	12

Tableau 4. – Tri selon les fonctions d'appartenance aux deux classes et une classe de rejet ($m = 2, \rho = 0.05$). Algorithme F(c+1)M

	Classe 1	Classe 2	Classe de rejet
-	9	3	3
	11	1	9
	8	2	7
	10	4	5
	7	5	2
	12	12	8
	6	6	4
	5	7	10
	4	10	6
	1	11	1
	2	8	11
+	3	9	12

Tableau 6. – Tri selon les valeurs de typicalité aux 2 classes ($m = 2, \eta = 2.0$) Algorithme FPcM.

	Classe 1	Classe 2
-	12	12
	11	1
	10	4
	8	2
	9	3
	7	5
	6	6
	4	10
	1	11
	5	7
	2	8
+	3	9

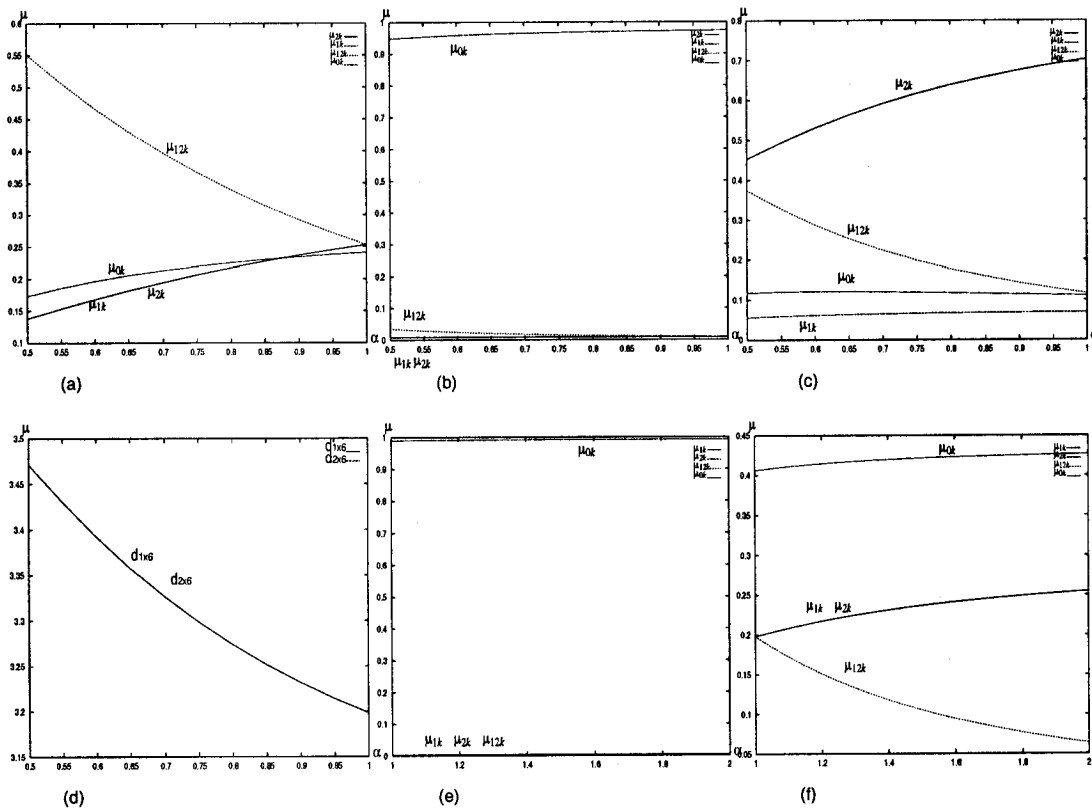


Figure 7. – Ensemble de données X_{12} : comportement des individus en fonction de α . (a) comportement de l'individu x_6 avec $\rho = 0.25$; (b) comportement de l'individu x_{12} avec $\rho = 0.25$; (c) comportement de l'individu x_5 avec $\rho = 0.25$; (d) distances aux centres des classes avec $\rho = 0.25$; (e) comportement de l'individu x_{12} avec $\rho = 0.4$; (f) comportement de l'individu x_6 avec $\rho = 0.4$.

4.2.2. introduction de la classe d'ambiguïté : les $c + 2$ -moyennes floues.

La figure 7a montre les différentes valeurs d'appartenance de x_6 pour $\rho = 0.25$ et $\alpha \in [0.5, 1]$. Nous pouvons noter que pour $\alpha = 1$, nous obtenons $\mu_{12k} = \mu_{1k} = \mu_{2k}$. Pour un taux de rejet important (faible valeur de α), la valeur d'appartenance à la classe d'ambiguïté est prépondérante. Finalement, notons que l'appartenance à la classe de rejet $\mu_{\Theta k} = \mu_{0k}$ n'est pas très sensible aux variations de α .

La figure 7b montre le bon comportement de notre algorithme pour le point x_{12} ($\rho = 0.25$). Ici aussi, la valeur d'appartenance à la classe de rejet n'est pas très sensible à la variation de α . Nous remarquons que la valeur d'appartenance du point x_{12} à la classe d'ambiguïté est supérieure à celle des deux classes 1 et 2 (comme la figure 6b le montre).

La figure 7c présente l'appartenance de x_5 aux 4 classes en fonction de α . Nous pouvons faire la remarque suivante : $\mu_{\Theta x_5}$ n'est pas sensible à α . Les variations des mesures d'appartenance vis-à-vis de α se répercutent sur μ_{2x_5} et μ_{12x_5} .

La figure 7d montre un des principaux intérêts de l'algorithme $F(c+2)M$ vis-à-vis des centres des classes. Elle décrit l'influence

de α sur la localisation des centres des classes, représentée par leur distance au point x_6 , centre du repère. Les points définis comme ambigus pour $\alpha = 0.5$ sont de plus en plus pris en compte pour le calcul des centres quand α croît, ce qui est représenté par des distances de plus en plus faibles (*i.e.* les centres se rapprochent de x_6).

Les figures 7e et 7f présentent le comportement des fonctions d'appartenance des points x_{12} et x_6 avec un fort taux de rejet. Une fois encore, nous remarquons que $\mu_{1k} = \mu_{2k} = \mu_{12k}$ quand $\alpha = 1$. $\mu_{\Theta k}$ n'est pas très sensible à α . Pour x_6 , la décroissance de μ_{12} se répercute de façon équivalente entre μ_1 et μ_2 .

Notre algorithme peut distinguer entre un individu modérément atypique et un individu hautement atypique. L'algorithme Fc+2M donne des degrés d'appartenance faibles pour des individus aberrants (haute valeur d'appartenance pour la classe de rejet de distance). De plus l'introduction des fonctions d'appartenance affectées aux classes d'ambiguïté apporte une discontinuité entre les fonctions d'appartenance affectées aux classes simples. Celles-ci représentent l'aspect « typique » des classes, tout en fournissant une partition floue apportant une approche probabiliste de la classification.

4.3. les IRIS de Fisher

4.3.1. données IRIS non perturbées. Performance de l'algorithme

Nous avons testé notre méthode sur l'ensemble des données IRIS [19] qui est très largement utilisé pour évaluer les performances des algorithmes de reconnaissance des formes. Nous comparons nos résultats avec ceux obtenus dans [30] et [15]. Cet ensemble de données est composé de 150 vecteurs composés de 4 attributs divisés en 3 classes physiques représentant différentes sous-espèces de fleurs. Chaque classe de fleurs contient 50 vecteurs. Une des trois classes est clairement séparée des deux autres, tandis que ces dernières admettent un chevauchement. Les résultats présentés dans cette section ont été obtenus après avoir testé différentes initialisations. Pour chaque test, les meilleurs résultats obtenus sont donnés. Le nombre maximum d'itération est de 100 et le critère d'erreur est fixé à $\epsilon = 0.0001$.

Tableau 7. – Nombre d'erreurs de substitution sur les données IRIS pour les algorithmes FPcM et FcM.

m	δ	$E(H_{MT}(T))$ FPcM	$E(H_{MM}(U))$ FPcM	$E(H_{MM}(U))$ FcM
1.5	1.5	17	17	17
1.5	3.0	15	17	17
2.0	2.0	15	16	16
2.0	5.0	15	16	16
3.0	3.0	14	14	15
5.0	2.0	12	14	15
5.0	5.0	12	15	15

Tableau 8. – Erreur de substitution et nombre de données rejetées pour $m = 3$, $\rho = 0.08$ et $\rho_\alpha \in [0.19, 0.27]$ (norme euclidienne). Le résultat donnant 0 vecteur rejeté, tout en donnant une erreur de substitution faible (comparée à celle des algorithmes FPcM et FcM), souligne l'intérêt de prendre en compte les notions de rejet durant la phase de partitionnement plutôt que pendant la phase de décision. Les individus sont globalement rejetés en distance ou en ambiguïté avec un taux égal à ρ et ρ_α mais sont cependant classés dans des classes simples (les fonctions d'appartenance aux classes composées restent inférieures). Les rejets permettent dans ce cas un meilleur positionnement des centres.

ρ_α	$E(H_{MM}(U))$	$R_D(H_{MM}(U))$	$R_A(H_{MM}(U))$
0.19	11	0	0
0.20	11	0	0
0.21	11	0	0
0.22	11	0	0
0.23	11	0	0
0.24	11	0	0
0.25	11	0	0
0.26	11	0	0
0.27	11	0	0

Dans les tableaux 7, 8, 9, 10 et 11, nous présentons les résultats en terme de nombre d'erreurs de substitution. Celles-ci sont basées sur la comparaison des versions nettes de la matrice U pour une partition nette correcte des fleurs. Le tableau 7 montre les résultats obtenus dans [30]. $E(H_{MT}(T))$, respectivement $E(H_{MM}(U))$, représente l'erreur de substitution lorsque la partition nette est déterminée en prenant le maximum des valeurs de typicalités, respectivement des degrés d'appartenance. δ est une constante utilisée dans l'algorithme FPcM et définie par l'utilisateur. Elle détermine la distance à laquelle la valeur de typicalité d'un individu dans la classe ω devient égale à 0.5. Dans les autres tableaux, $E(H_{MM}(U))$ représente l'erreur de substitution des $c + 2$ -moyennes floues, $R_D(H_{MM}(U))$ représente le nombre d'individus rejetés en distance en utilisant une norme Euclidienne ou de Mahalanobis, et $R_A(H_{MM}(U))$, le nombre d'individus rejetés en ambiguïté pour les deux normes précédentes. Nous détaillons ces résultats dans les paragraphes suivants.

Tout d'abord, on peut noter les remarques générales suivantes :

Tableau 9. – Erreur de substitution et nombre de données rejetées pour $m = 1.2$, $\rho = 0.02$ and $\rho_\alpha \in [0.005, 0.04]$ (norme de Mahalanobis). Les résultats sont bien meilleurs que ceux des algorithmes FcM, F(c+1)M, et FPcM; par exemple le meilleur résultat obtenu par les FcM avec la même norme est de 7 erreurs, où F(c+2)M donne $E(H_{MM}) = 1$. Les cas ($\rho_\alpha \in [0.02, 0.025]$) montrent clairement les bons résultats obtenus avec les F(c+2)M, où les points mal classés (par les FcM), sont maintenant soit bien classés (meilleure localisation des centres, les points les plus significatifs sont mieux pris en compte), soit rejetés en distance ou en ambiguïté.

ρ_α	$E(H_{MM}(U))$	$R_D(H_{MM}(U))$	$R_A(H_{MM}(U))$
0.005	3	3	0
0.01	3	3	0
0.015	2	3	2
0.02	1	3	3
0.025	1	3	3
0.03	1	3	4
0.035	1	3	4
0.04	2	3	5

Tableau 10. – Valeurs de R_A , pour $m = 1.1$ et $\rho_\alpha = 0.025$: elles semblent peu influencées par rapport aux variations de ρ (norme de Mahalanobis).

ρ	$E(H_{MM}(U))$	$R_D(MM(U))$	$R_A(MM(U))$
0.020	1	3	3
0.025	1	4	3
0.030	1	4	3
0.035	1	5	3
0.040	1	5	3
0.045	1	5	3
0.050	1	7	3
0.055	1	7	5
0.060	1	9	4
0.065	1	9	4
0.070	1	9	4

Tableau 11. – Valeurs de R_D , pour $m = 1.1$ et $\rho = 0.020$: elles semblent peu influencées par rapport aux variations de ρ_α (norme de Mahalanobis).

ρ_α	$E(H_{MM}(U))$	$R_D(H_{MM}(U))$	$R_A(H_{MM}(U))$
0.005	3	3	0
0.01	2	3	1
0.015	2	3	2
0.02	1	3	3
0.025	1	3	3
0.03	1	3	4
0.035	1	3	4
0.04	1	3	4
0.045	1	3	7
0.05	1	3	7
0.055	1	3	9
0.06	1	3	10
0.065	0	3	10
0.070	0	3	11

- le résultat donnant 0 vecteur rejeté (cf. tableau 8), tout en donnant une erreur de substitution faible (comparée à celle des algorithmes FPcM et FcM), souligne l'intérêt de prendre en compte les notions de rejet durant la phase de partitionnement plutôt que pendant la phase de décision. Les individus sont globalement rejetés en distance ou en ambiguïté avec un taux égal à ρ et ρ_α mais sont cependant classés dans des classes simples (les fonctions d'appartenance aux classes composées restent inférieures). Les rejets permettent dans ce cas un meilleur positionnement des centres.

- pour des valeurs particulières de ρ, α (qui est globalement pris en compte par ρ_α) et m , on obtient 7 individus mal classés ou non classés, distribués de la façon suivante : $E(H_{MM}(U)) = 1, R_D(H_{MM}(U)) = 3$ et $R_A(H_{MM}(U)) = 3$. Ceci montre l'intérêt d'utiliser à la fois les rejets d'ambiguïté et de distance dans l'étape de classification. Ainsi, le taux de mal classé est faible. Les centres finaux produits par les c -moyennes floues et les $c+2$ -moyennes floues sont différents. Dans le second cas, ils sont plus proches des centres des trois classes physiques.

Nous montrons sur le tableau 8, le nombre d'erreurs de substitution et le nombre d'individus rejetés pour $m = 3, \rho = 0.08$ et pour $\rho_\alpha \in [0.19, 0.27]$ lorsque nous utilisons une norme euclidienne dans le calcul des distances. Il est intéressant de noter que les performances de l'algorithme proposé sont significativement meilleures en regard du taux de classification correcte que ceux de la tableau 7. Les rejets en distance et en ambiguïté permettent de modifier les centres des classes (par rapport à ceux obtenus avec les c -moyennes floues) parce que les données rejetées ou ambiguës sont moins prises en compte dans le calcul des centres.

Le tableau 9 montre, pour les valeurs particulières de $m = 1.2$ et $\rho = 0.02$, le nombre d'erreurs de substitution et le nombre d'individus rejetés en utilisant la norme de Mahalanobis prise en compte dans l'algorithme de Gustafson et Kessel [21]. Dans

la lumière des exemples précédents, la classe de rejet, ainsi que les quatre classes de rejet d'ambiguïté permettent de localiser de façon plus précise les centres, ainsi, $E(H_{MM}(U))$, décroît comparé avec les algorithmes FcM, F(c+1)M, et FPcM. Les résultats sont largement meilleurs; par exemple le meilleur résultat obtenu par les FcM avec la même norme est de 7 erreurs, où F(c+2)M donne $E(H_{MM}) = 1$. Les cas ($\rho_\alpha \in [0.02, 0.025]$) montrent clairement les bons résultats obtenus avec les F(c+2)M, où les points mal classés (par les FcM), sont maintenant soit bien classés (meilleure localisation des centres, les points les plus significatifs sont mieux pris en compte), soit rejeté en distance ou en ambiguïté. De plus, notre algorithme est capable de préciser à quelle classe composée appartient ces données.

Le tableau 10 montre le comportement correct des valeurs de $R_A(H_{MM}(U))$ en fonction de $\rho \in [0.02, 0.070]$. Elle montre une faible variation de $R_A(H_{MM}(U))$ vis-à-vis de ce paramètre. Le taux de rejet de distance ρ semble peu influencé par le taux de rejet en ambiguïté et vice versa (tableau 11).

Dans la figure 8a (respectivement 8d), est montrée, dans le principal plan factoriel, la représentation graphique du maximum des fonctions d'appartenance des trois classes obtenues en utilisant la norme de Mahalanobis (respectivement la norme euclidienne). La projection de cette fonction en terme de valeurs d'iso-appartenance produit, comme attendu, des classes ellipsoïdales.

La figure 8b (respectivement 8e), représente le maximum des fonctions d'appartenance aux 4 classes de rejet d'ambiguïté des données IRIS en utilisant la norme de Mahalanobis (respectivement la norme euclidienne). Les classes de rejet d'ambiguïté ω_{13}, ω_{23} et ω_{123} sont relativement peu importantes par rapport à la classe ω_{12} .

La figure 8c (respectivement 8f), représente la fonction d'appartenance à la classe de rejet global de distance des données IRIS en utilisant la norme de Mahalanobis (respectivement norme euclidienne). Les valeurs d'appartenance des individus à cette classe sont importantes lorsque ces individus sont situés loin des classes.

En résumé, l'amélioration des performances de classification vient du fait que l'algorithme Fc+2M attribue une fonction de pondération définie sur des distances différentes selon que l'individu appartient clairement à une classe, à aucune ou bien à plusieurs classes de façon égale. Cette fonction de pondération fait correspondre à l'individu un niveau de confiance adapté à ses caractéristiques sans apport supplémentaire d'information. Le niveau de confiance est décrit par : $\text{conf}_k = 1 - (\mu_{(c+1)k} + \mu_{0k})$. Plus le niveau de confiance est élevé, plus la contribution de ce point dans l'estimation des centres est significative (i.e. la différence $1 - \text{conf}_k$ est distribuée sur les classes simples).

4.3.2. données IRIS perturbées : robustesse de la méthode relativement à des points aberrants

Afin de comparer l'algorithme proposé avec celui défini par les FcM, en termes de robustesse et de capacité à rejeter en ambiguïté, nous utilisons le jeu de données IRIS que nous perturbons en

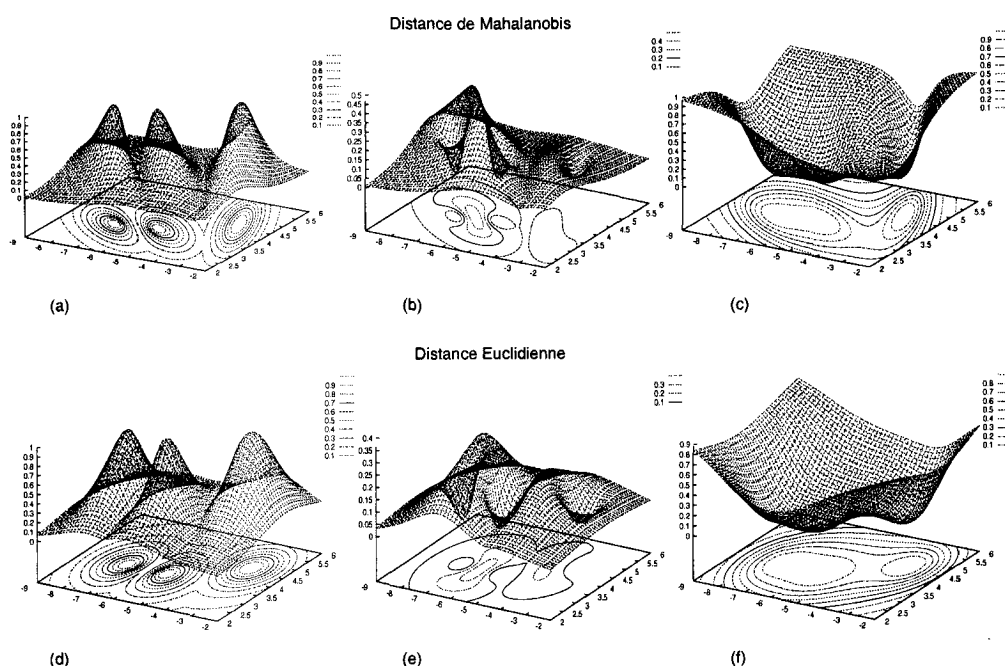


Figure 8. – Le maximum des fonctions d'appartenance des données IRIS sur le premier plan factoriel. (a) et (d) aux 3 classes; (b) et (e) aux 4 classes d'ambiguïté; (c) et (f) à la classe de rejet. (a), (b) et (c) : distance de Mahalanobis avec les paramètres suivants $m = 1.5$ et $\rho = 0.08$, $\rho_\alpha = 0.05$, erreur=0.01; (d), (e) et (f) : distance Euclidienne avec les paramètres suivants $m = 3$ et $\rho = 0.08$, $\rho_\alpha = 0.20$, erreur = 0.01.

ajoutant 50 points issus d'une distribution gaussienne de vecteur moyenne [6.24, 2.87, 4.90, 1.67]. Pour les deux premiers tests la distribution est normalisée, pour le troisième test la variance est multipliée par 2. Les données IRIS ainsi que les points aberrants sont représentés sur la figure 9 dans le plan factoriel principal. La quatrième classe correspond aux points aberrants. Les résultats présentés sur le tableau joint à la figure sont obtenus pour une seule initialisation aléatoire et selon le protocole expérimental suivant :

- la norme choisie est celle de Mahalanobis;
- $m = 1.1$ (Pour cette norme, des m faibles donnent de meilleurs résultats);
- $\rho = 0.25$ (ceci correspond à 25% des individus à rejeter);
- $\rho_\alpha = 0.02$ pour une distribution normalisée (test 1 et 2) et $\rho_\alpha = 0.01$ pour une distribution de variance double (test 3);
- Le nombre maximum d'itération est de 100;
- le critère d'erreur est fixé à $\epsilon = 0.0001$.

Dans la table, $E(H_{MM}(U))$ représente l'erreur de substitution des FcM ou Fc + 2M, $R_D(H_{MM}(U))$ représente le nombre d'individus rejetés en distance et $R_A(H_{MM}(U))$ le nombre d'individus rejetés en ambiguïté. $E^i(H_{MM}(U))$ représente l'erreur de substitution sur les données IRIS seulement, $E^b(H_{MM}(U))$ représente l'erreur de substitution sur les points aberrants seulement.

Les résultats obtenus illustrent bien les remarques énoncées précédemment :

- performances en classification. L'algorithme FcM donne respectivement 59, 64 et 68 erreurs pour les tests 1, 2 et 3. L'algorithme Fc+2M permet de rejeter des individus en distance ou en ambiguïté. Le taux de mauvaise classification est donc nettement plus faible, respectivement 6, 6 et 3. Dans les deux premiers tests, 5 individus aberrants restent cependant mal classés.

- performances en rejet. Les résultats montrent que les paramètres ρ et ρ_α contrôlent correctement les rejets :

- en distance : $\rho = 0.25$ donne 25% d'individus rejetés soit 50;
- en ambiguïté une valeur de ρ_α divisée par 2 (test 3 par rapport aux tests 1 et 2) permet de tenir compte de l'évolution de la variance du bruit.

Ces résultats témoignent de la robustesse de l'algorithme vis-à-vis de données perturbées tant du point de vue des performances de classification que de la capacité à rejeter en ambiguïté.

5. conclusion

Nous commençons cet article en justifiant l'intérêt de trouver une partition floue et d'introduire deux types de classes de rejet dans le processus de coalescence : le rejet de distance gère les individus qui sont loin des classes et le rejet d'ambiguïté concerne ceux situés sur la frontière de plusieurs classes. La méthode que nous

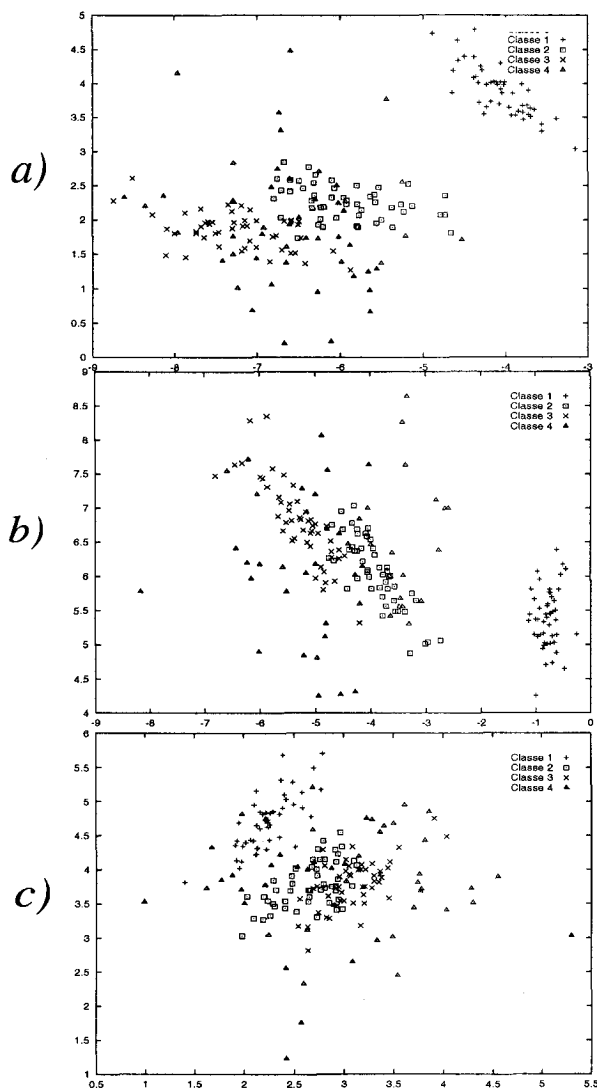


Figure 9. – Etude expérimentale de la robustesse de l’algorithme.

proposons est formalisée comme un problème de minimisation d’un critère de classification sous contrainte dont la solution dépend d’une fonction de contrainte dans laquelle la notion de rejet est introduite.

Pour remédier aux problèmes des FcM et des algorithmes dérivés, nous précisons le concept d’ambiguïté en définissant la notion de frontières floues entre classes dans une partition floue. Puis nous considérons le problème de la coalescence dans le cadre de la théorie de la régularisation, et proposons un critère de classification prenant en compte les discontinuités (i.e. frontières floues) de façon équivalente au modèle énergétique de Pavlidis. Nous améliorons la robustesse de notre algorithme vis-à-vis du bruit en complétant le traitement de l’ambiguïté avec celui du rejet partiel de distance défini dans [15]. Les avantages sont les suivants :

	<i>FcM</i>	<i>Fc + 2M</i>
	$E(H_{MM}(U)) = 59$	$E(H_{MM}(U)) = 6$
	$E^i(H_{MM}(U)) = 9$	$R_A(H_{MM}(U)) = 3$
	$E^b(H_{MM}(U)) = 50$	$R_D(H_{MM}(U)) = 50$
		$E^i(H_{MM}(U)) = 1 \quad E^b(H_{MM}(U)) = 5$
	<i>FcM</i>	<i>Fc + 2M</i>
	$E(H_{MM}(U)) = 64$	$E(H_{MM}(U)) = 6$
	$E^i(H_{MM}(U)) = 14$	$R_A(H_{MM}(U)) = 6$
	$E^b(H_{MM}(U)) = 50$	$R_D(H_{MM}(U)) = 49$
		$E^i(H_{MM}(U)) = 1 \quad E^b(H_{MM}(U)) = 5$
	<i>FcM</i>	<i>Fc + 2M</i>
	$E(H_{MM}(U)) = 68$	$E(H_{MM}(U)) = 3$
	$E^i(H_{MM}(U)) = 18$	$R_A(H_{MM}(U)) = 2$
	$E^b(H_{MM}(U)) = 50$	$R_D(H_{MM}(U)) = 50$
		$E^i(H_{MM}(U)) = 2 \quad E^b(H_{MM}(U)) = 1$

- 1) ces rejets sont calculés dans l’étape de coalescence et non dans l’étape de décision; les données aberrantes sont alors moins prises en compte dans la détermination des centres;
- 2) l’algorithme inclut une modélisation de l’hésitation ou de l’ambiguïté, et des fonctions d’appartenance sont affectées à tous les sous-ensembles de classes de 2^Ω plutôt qu’aux éléments de Ω seulement. De plus, le degré d’appartenance à la classe de rejet d’ambiguïté totale d’un individu est calculé explicitement;
- 3) les notions de rejet conduisent à attribuer une fonction de pondération différente à chaque individu selon que celui-ci appartient clairement à une classe, à aucune ou bien à plusieurs classes de façon égale;
- 4) nous préservons l’approche probabiliste de la classification des c -moyennes floues puisque nous préservons la contrainte de Ruspini; nous obtenons en effet une partition floue. En introduisant

les rejets d'ambiguïté partielle, nous créons une discontinuité entre les fonctions d'appartenance affectées aux classes de Ω , ce qui permet de les rendre indépendantes. Les fonctions d'appartenance générées représentent alors l'aspect « typique » des classes;

5) les définitions que nous avons choisies pour les classes de rejet de distance et d'ambiguïté peuvent s'appliquer sur des algorithmes utilisant des métriques différentes ou des prototypes de dimensions supérieures, comme par exemple les algorithmes FcRM et AFc.

Les premières expérimentations sont encourageantes et comparées de façon favorable avec les résultats d'autres méthodes comme les algorithmes FcM, FPcM et F(c+1)M sur les mêmes ensembles de données. L'algorithme utilise deux paramètres supplémentaires : ρ le taux de rejet de distance qui a été introduit dans l'algorithme des $c + 1$ -moyennes floues et ρ_α le taux de rejet d'ambiguïté. L'influence de ces paramètres est le suivant : plus le taux de rejet de distance est important, plus grand est le nombre d'individus rejetés et plus l'algorithme est robuste au bruit. Plus le taux de rejet en ambiguïté est important, moins les classes ont une influence les unes sur les autres et plus la probabilité de mauvaise classification est faible. Pour $\rho_\alpha \rightarrow 0$ et $\rho \rightarrow 0$, l'algorithme Fc+2M est équivalent à l'algorithme des c -moyennes. Notre algorithme est capable de préciser à quelle classe composée, l'individu appartient. Il apporte donc une solution à l'initialisation des masses sur les hypothèses composées en classification d'images par exemple. La position des centres tient compte de la répartition des individus vis-à-vis des classes. Les futures investigations porteront sur la recherche de critères permettant d'obtenir les paramètres ρ_α et ρ .

6. annexe

6.1. annexe A : détail de la preuve du théorème des c -moyennes floues étendues

En additionnant la contrainte $\forall k \in [1, n], \sum_{A \subseteq \Omega} \mu_{Ak} = 1$ au critère (8) avec les multiplicateurs de Lagrange $(\lambda_k)_{k \in [1, n]}$, on obtient :

$$\bar{J}_m^1(\mu, v, \lambda) = \sum_{A \subseteq \Omega} \sum_{k=1}^n \mu_{Ak}^m d_{Ak}^2 + \sum_{k=1}^n \lambda_k \left(1 - \sum_{A \subseteq \Omega} \mu_{Ak} \right)$$

Calcul de $\mu_{Ak}^* \forall A \subseteq \Omega$

En dérivant $\bar{J}_m^1(\mu, v, \lambda)$ par rapport à μ_{Ak} , on obtient :

$$\frac{\partial \bar{J}_m^1(\mu, v, \lambda)}{\partial \mu_{Ak}} = m \mu_{Ak}^{m-1} d_{Ak}^2 - \lambda_k$$

$\frac{\partial \bar{J}_m^1(\mu, v, \lambda)}{\partial \mu_{Ak}} = 0$ implique que le degré d'appartenance de l'individu z_k à la classe A est :

$$\mu_{Ak}^* = \left(\frac{\lambda_k^*}{m d_{Ak}^2} \right)^{\frac{1}{m-1}}$$

En sommant sur A :

$$\sum_{A \subseteq \Omega} \mu_{Ak}^* = \lambda_k^* \sum_{A \subseteq \Omega} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}} \quad (22)$$

où $\lambda_k^* = \left(\frac{\lambda_k^*}{m} \right)^{\frac{1}{m-1}}$.

Avec la première contrainte, nous obtenons :

$$\lambda^k \left(\sum_{A \subseteq \Omega} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}} \right) = 1$$

qui donne l'expression de $\mu_{Ak}^*, \forall A \subseteq \Omega, \forall k \in [1, n]$:

$$\mu_{Ak}^* = \frac{1}{\sum_{B \subseteq \Omega} \left(\frac{d_{Ak}}{d_{Bk}} \right)^{\frac{2}{m-1}}}$$

Pour déterminer le paramètre α tel que (seconde contrainte) :

$$\sum_{k=1}^n \mu_{(c+1)k}^* = \sum_{k=1}^n \sum_{A \subseteq \Omega, |A|>1} \mu_{Ak}^* = \rho_\alpha n$$

Nous choisissons (voir section 2.3) :

$$d_{Ak} = \frac{\alpha_A \left(\sum_{\omega_i \in A} d_{ik} \right)^2}{|A|^2 \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}} \quad (23)$$

Nous obtenons (dans ce qui suit, nous fixons $\alpha_A = \alpha$ constant) :

$$\sum_{k=1}^n \frac{1}{\alpha^{\frac{2}{m-1}} \sum_{B \subseteq \Omega} \left(\frac{1}{d_{Bk}} \right)^{\frac{2}{m-1}} + \sum_{B \subseteq \Omega, |B|>1} \left(\frac{1}{d_{Bk}^*} \right)^{\frac{2}{m-1}}} = \rho_\alpha n$$

avec $d_{Ak}^* = \frac{1}{\alpha} d_{Ak}$

La relation précédente peut s'exprimer comme :

$$\sum_{k=1}^n \frac{D}{\alpha^{\frac{2}{m-1}} E(\rho) + D} = \rho_\alpha n \quad (24)$$

où $E(\rho) = \sum_{B \subseteq \Omega} \left(\frac{1}{d_{Bk}} \right)^{\frac{2}{m-1}}$ et $D = \sum_{B \subseteq \Omega, |B|>1} \frac{1}{d_{Bk}^* \frac{2}{m-1}}$.

Le système défini par la formule 24 se résout par la méthode de Newton-Raphson pour $\alpha > 0$.

A ce point, une remarque est indispensable. Le cas spécial $\rho_\alpha = 0$ (pas de rejet du tout) implique $\alpha \rightarrow +\infty$ ($\mu_{(c+1)k} = 0$), L'algorithme proposé se réduit à l'algorithme FcM.

6.2. annexe B : détail de la preuve du théorème des c+2-moyennes floues

En additionnant la contrainte $\forall k \in [1, n], \sum_{A \in 2^\Omega \cup \emptyset} \mu_{Ak} = 1$ au critère (13) avec les multiplicateurs de Lagrange $(\lambda_k)_{k \in [1, n]}$, on obtient :

$$\bar{J}_m^{fc+2m}(\mu, v, \lambda) = \sum_{A \subseteq \Omega} \sum_{k=1}^n \mu_{Ak}^m d_{Ak}^2 + \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^r d_{ik}^2 + \sum_{k=1}^n \lambda_k \left(1 - \sum_{A \in 2^\Omega \cup \emptyset} \mu_{Ak} \right)$$

En dérivant $\bar{J}_m^{fc+2m}(\mu, v, \lambda)$ par rapport à μ_{Ak} , on obtient (c.f. Annexe A) :

$$\sum_{A \subseteq \Omega} \mu_{Ak}^* = \lambda'_k \sum_{A \subseteq \Omega} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}} \quad (25)$$

où $\lambda'_k = \left(\frac{\lambda_k^*}{m} \right)^{\frac{1}{m-1}}$.

Calcul de $\mu_{\emptyset k}^*$

Nous reprenons le calcul effectué dans [15].

En dérivant $\bar{J}_m^{fc+2m}(\mu, v, \lambda)$ par rapport à μ_{ik} , on obtient :

$$\frac{\partial \bar{J}_m^{fc+2m}(\mu, v, \lambda)}{\partial \mu_{ik}} = r \mu_{ik}^{r-1} d_{ik}^2 - \lambda_k \prod_{j=1, j \neq i}^c \mu_{jk}$$

$\frac{\partial \bar{J}_m^{fc+2m}(\mu, v, \lambda)}{\partial \mu_{ik}} = 0$ impose que :

$$\mu_{ik}^* = \left(\frac{\lambda_k^* \prod_{j=1, j \neq i}^c \mu_{jk}^*}{r d_{ik}^2} \right)^{\frac{1}{r-1}}$$

En multipliant sur i :

$$\prod_{i=1}^c \mu_{ik}^* = \left(\frac{\lambda_k^*}{r} \right)^{\frac{c}{r-1}} \prod_{i=1}^c \left(\frac{1}{d_{ik}} \right)^{\frac{2}{r-1}} \prod_{i=1}^c \mu_{ik}^{* \frac{c-1}{r-1}}$$

En écrivant $d_{\emptyset k} = \sqrt{c} \left(\prod_{i=1}^c d_{ik} \right)^{\frac{1}{c}}$, $\mu_{\emptyset k} = \prod_{i=1}^c \mu_{ik}^*$ peut s'exprimer comme :

$$\mu_{\emptyset k} = \left(\frac{m \lambda_k^{m-1}}{r} \right)^{\frac{r-c}{r-c}} \left(\frac{\sqrt{c}}{d_{\emptyset k}} \right)^{\frac{2c}{r-c}} \quad (26)$$

Deux voies peuvent être poursuivies :

- en mettant r égal à mc , nous pouvons dériver explicitement la formule de mise à jour de la matrice de partition [15];
- pour les autres valeurs de r différentes de mc , la résolution est complexe.

(26) devient :

$$\mu_{\emptyset k} = \left(\frac{1}{c} \right)^{\frac{1}{m-1}} \lambda'_k \left(\frac{\sqrt{c}}{d_{\emptyset k}} \right)^{\frac{2}{m-1}} = \lambda'_k \left(\frac{1}{d_{\emptyset k}} \right)^{\frac{2}{2m-1}} \quad (27)$$

Avec la première contrainte (voir 14), nous obtenons :

$$\lambda'^k \left(\sum_{A \subseteq \Omega} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}} + \left(\frac{1}{d_{\emptyset k}} \right)^{\frac{2}{m-1}} \right) = 1$$

ou bien encore :

$$\lambda'^k \left(\sum_{A \in 2^\Omega \cup \emptyset} \left(\frac{1}{d_{Ak}} \right)^{\frac{2}{m-1}} \right) = 1$$

qui donne l'expression de μ_{Ak}^* , $\forall A \in 2^\Omega \cup \emptyset \setminus \emptyset, \forall k \in [1, n]$:

$$\mu_{Ak}^* = \frac{1}{\sum_{B \in 2^\Omega \cup \emptyset} \left(\frac{d_{Bk}}{d_{Ak}} \right)^{\frac{2}{m-1}}}$$

Pour déterminer le paramètre α tel que (troisième contrainte) :

$$\sum_{k=1}^n \mu_{(c+1)k}^* = \sum_{k=1}^n \sum_{A, |A|>1} \mu_{Ak}^* = \rho_\alpha n$$

Nous choisissons (voir section 2.3) :

$$d_{Ak} = \frac{\alpha_A \left(\sum_{\omega_i \in A} d_{ik} \right)^2}{|A|^2 \left(\prod_{\omega_i \in \Omega} d_{ik} \right)^{\frac{1}{c}}} \quad (28)$$

Nous obtenons (nous fixons $\alpha_A = \alpha$ constant) :

$$\sum_{k=1}^n \frac{1}{\alpha^{\frac{2}{m-1}} \sum_{B \in \Omega \cup \emptyset} \left(\frac{1}{d_{Bk}} \right)^{\frac{2}{m-1}} + \sum_{B \subseteq \Omega, |B|>1} \left(\frac{1}{d_{Bk}^*} \right)^{\frac{2}{m-1}}} = \rho_\alpha n$$

avec $d_{Ak}^* = \frac{1}{\alpha} d_{Ak}$

Cette relation précédente peut s'exprimer comme :

$$\sum_{k=1}^n \frac{D}{\alpha^{\frac{2}{m-1}} E(\rho) + D} = \rho_\alpha n \quad (29)$$

$$\text{où } E(\rho) = \sum_{B \in \Omega \cup \emptyset} \left(\frac{1}{d_{Bk}} \right)^{\frac{2}{m-1}} \text{ et } D = \sum_{B \subseteq \Omega, |B| > 1} \frac{1}{d_{Bk}^* \frac{2}{m-1}}.$$

Le dernier pas consiste à calculer la valeur de $(d_{\Theta k})_{k \in [1, n]}$, tel que

$$\sum_{k=1}^n \mu_{\Theta k} = \rho n \text{ (seconde contrainte) :}$$

$$\begin{aligned} \sum_{k=1}^n \mu_{\Theta k} &= \rho n \\ \sum_{k=1}^n \frac{1}{\sum_{B \in 2^{\Omega \cup \emptyset} \setminus \emptyset} \left(\frac{d_{\Theta k}}{d_{Bk}} \right)^{\frac{2}{m-1}}} &= \rho n \\ \sum_{k=1}^n \frac{1}{1 + \sum_{B \subseteq \Omega} \left(\frac{d_{\Theta k}}{d_{Bk}} \right)^{\frac{2}{m-1}}} &= \rho n \end{aligned}$$

En utilisant $d_{\Theta k} = \sqrt{c} \left(\prod_{i=1}^c d_{ik} \right)^{\frac{1}{c}}$ et $d_{ik} = \frac{T_i^2}{d_{ik}}$ (cf. [15]), nous obtenons :

$$\sum_{k=1}^n \frac{1}{1 + \sum_{B \subseteq \Omega} \left(\frac{\sqrt{c} \left(\prod_{i=1}^c \frac{T_i^2}{d_{ik}} \right)^{\frac{1}{c}}}{d_{Bk}} \right)^{\frac{2}{m-1}}} = \rho n \quad (30)$$

En mettant $\sqrt{c} \left(\prod_{i=1}^c T_i^2 \right)^{\frac{1}{c}} = T(\rho)$, (30) devient :

$$\sum_{k=1}^n \frac{1}{1 + \sum_{B \subseteq \Omega} \left(\frac{T(\rho)}{d_{Bk} \left(\prod_{i=1}^c d_{ik} \right)^{\frac{1}{c}}} \right)^{\frac{2}{m-1}}} = \rho n \quad (31)$$

Cette relation peut s'exprimer comme :

$$\sum_{k=1}^n \frac{1}{1 + T(\rho)^{\frac{2}{m-1}} E'(\rho_\alpha)} = \rho_\alpha n \quad (32)$$

$$\text{où } E'(\rho_\alpha) = \sum_{B \subseteq \Omega} \left(\frac{1}{d_{Bk} \left(\prod_{i=1}^c d_{ik} \right)^{\frac{1}{c}}} \right)^{\frac{2}{m-1}}.$$

Le système défini par les formules 29 et 32 peut être résolu par la méthode de Newton-Raphson pour $\alpha > 0$ et $T(\rho) > 0$. Pour appliquer cette méthode, nous initialisons ρ_0 , en résolvant l'équation donnée par les fuzzy c+1-means [15]. Cette valeur d'initialisation est proche de la solution cherchée et assure une bonne convergence de la méthode de Newton-Raphson.

A ce stade, une remarque est indispensable. Le cas particulier $\rho_\alpha = 0$ et $\rho = 0$ (pas de rejet du tout) implique $\alpha \rightarrow +\infty$ et $T(\rho) \rightarrow +\infty$ ($\mu_{(c+1)k} = 0$ et $\mu_{\Theta k} = 0$), l'algorithme proposé se réduit à l'algorithme des FcM.

BIBLIOGRAPHIE

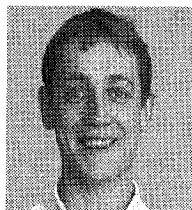
- [1] M.R. Andenberg. *Cluster Analysis for Applications*, Academic Press, New York, 1973.
- [2] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New-York 1981.
- [3] J.C. Bezdek, and J.C. Dunn. Optimal fuzzy partition : a heuristic for estimating the parameters in a mixture of normal distributions, *IEEE Trans. on Computers*, c-24, 835-838, 1975.
- [4] J.C. Bezdek, R.J. Hathaway, M.J. Sabin and W.T. Tucker. Convergence Theory for Fuzzy c-Means : Counter examples and Repairs, 17(5) : 873-877, September/October 1987.
- [5] A. Black and A. Zisserman. Visual reconstruction, In *Cambridge, MA : MIT Press*, 1987.
- [6] I. Bloch. Fuzzy spatial relationships : a few tools for model-based pattern recognition in aerial images. In *Image and Signal Processing for remote sensing III, SPIE, Vol 2955*, page 141-150, Taormina, Italie, 1996.
- [7] I. Bloch. Fusion de données, ensembles flous et morphologie mathématique en traitement d'images. *Rapport d'habilitation. 95D 007. Telecom Paris*, 1995.
- [8] I. Bloch and H. Maître. Fuzzy mathematical morphologies : A comparative study, *Pattern Recognition*, 28(9) : 1341-1387, 1995.
- [9] C. Pedrycz and P. Waletzky. Fuzzy clustering with partial supervision, *IEEE Transactions on systems, Man, and Cybernetics*, 27(5) : 788-795, 1997.
- [10] C.K. Chow. On optimum Recognition Error and Reject Tradeoff. In *IEEE Trans. Information Theory. Vol 16*, pages 41-46, January 1990.
- [11] C.K. Chow. Recognition Error and Reject Trade-off. In *Proc. Third Ann. Symp. Document Analysis and Information Retrieval. Vol.13.*, pages 1-8, Univ. of Nevada, Las Vegas, April 1994.
- [12] R. Dave. Characterization and detection of noise in clustering, 1991. In *vol.12, pp. 657-664, Pattern recognition letters*, 1991.
- [13] R. Dave. Robust fuzzy clustering algorithms, *IEEE International Conference on Fuzzy Systems. San Francisco*, 2 : 1281-1286, 1993.
- [14] R.N. Dave. Boundary detection through fuzzy clustering. *Proc. 1st IEEE International Conference on Fuzzy Systems, San Diego, CA*, 1992. 1992
- [15] C. Demko, P. Loonis and M. Ménard. Les c+1 moyennes floues : introduction du rejet en classification. *SGC'98, Montpellier*, septembre 1998.
- [16] E. Diday. Optimisation en classification automatique, *INRIA Ed.*, 1-2, 179.
- [17] B. Dubuisson. Decision with reject option. In *European Signal Processing Conference, SPAIN*, 1990.
- [18] B. Dubuisson and M.H. Masson. A statistical decision rule with incomplete knowledge about classes, *Pattern Recognition*, 26 : 155-165, 1993.
- [19] R.A. Fisher. The statistical utilisation of multiple measurements, (*Ann. Eugen.*, 8, 1938.
- [20] C. Frélicot, B. Dubuisson and M.H. Masson. Reject options in fuzzy pattern classifications rules, *Proc. Eufit'95, Aachen, Germany*, 1995.

- [21] Donald E. Gustafson and William C. Kessel. Fuzzy clustering with a fuzzy covariance matrix, In *Proc. IEEE CDC*, pages 761–766, San Diego, CA, January 10-12 1979.
- [22] T.M. Ha. The Optimum Class-Selective Rejection Rule. In *IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.19, N° .6*, 1997, pages 608–614, Yokohama, Japan, march 1997.
- [23] R.J. Hathaway and J.C. Bezdek. Switching Regression Models and Fuzzy Clustering. In *IEEE Transactions on fuzzy systems, Vol.1, N° .3*, pages 195–203, 1993.
- [24] A.K. Jain and R.C. Dubes. Algorithms for clustering data, *Prentice Hall*, 1988.
- [25] J.C.Bezdek. Pattern recognition with fuzzy objective function algorithms. In *Plenum Press*, New-York, 1987.
- [26] R. Krishnapuram and J.M. Keller. The Possibilistic C-Means Algorithm : Insights and Recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3) : 148-158, August 1996.
- [27] Raghu Krishnapuram. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2) : 98-110, 1993.
- [28] S.Z. Li. Markov Random Field Modelin in Computer Vision, *Computer Science Workbench. Springer*, 16 : 260 pp, 1995.
- [29] J. MacQueen. Some methods of classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symposium on Math., Stat., and Prob.*, U. California Press, Berkeley, CA, 1967.
- [30] Nikhil R. Pal, Kuhu Pal and James C. Bezdek. A Mixed c-Means Clustering Model. In *Proceedings of the sixth IEEE International Conference on Fuzzy Systems*, apges 11-21, Barcelona, July 1997.
- [31] T. Pavlidis. A critical survey of image analysis methods. pages 502-511, *ICPR*, 1986.
- [32] Z. Pawlack. Rough Sets. Theoretical Aspects of Reasoning about Data., *Kluwer Academic Publishers. Dordrecht, Boston, London*, 1991.
- [33] A. Ribert, A. Ennaji, E. Stocker, and Y. Lecourtier. Structuration de données : application à la construction d'un classifieur neuronal distribué, *RFIA'98*, 1998, 2 : 145-151, janvier 1998.
- [34] E. Ruspini. A new approach to clustering, *Information and Control*, 15 : 22-32, 1969.
- [35] S.K. Pal. Fuzzy tools in the management of uncertainty in pattern recognition, image analysis, vision and expert systems. *International Journal of Systems Sciences. Vol. 22*, pages 511-549, 1991.
- [36] S.K.Pal and D Dutta Majumber. Fuzzy sets and decision making approaches in vowel and speaker recognition, *I.E.E.E. Transactions on Systems, Man and Cybernetics. Vol.7*, pages 625-629, 1977.
- [37] N. Tarel, J-Ph. et N. Boujema. Une approche floue du recalage 3D, généricité et robustesse. *10ème Congrès de Reconnaissance des Formes et Intelligence artificielle (RFIA'96)*, 1 : 267–276, 1996.
- [38] J.W. Tukey. Explortary Data Analysis. *Reading, MA : Addison-Wesley*, 1977.
- [39] M.H. Yektae, E.L. Zahzah and M. Menard. Automatic segmentation of printed Persian, *SCIA'97, 10th Scandinavian Conference on Image Analysis. Finland*, pages 767–772, june 1997.
- [40] E. Zahzah. *Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites*. PhD Thesis Université Paul Sabatier, Toulouse, 1992.
- [41] H.J. Zimmerman, and P. Zysno. Quantifying vagueness in decision models. *European J.Operational Res.*,22 : 148-158, 1985.

Manuscrit reçu le 2 avril 1998.

L' AUTEUR

Michel MÉNARD



Michel Ménard est né en 1966. Docteur en Electronique de l'Université de Poitiers en 1993, il mène des recherches en traitement et en analyse d'images avec une attention particulière pour la reconnaissance d'objets naturels. Les domaines étudiés sont la modélisation floue d'objets et la fusion multi-sources au sein du L3i.