

# ECM : Algorithme évidentiel des $c$ -moyennes

## ECM : Evidential $c$ -means algorithm

M.-H. Masson<sup>1</sup>

T. Dencœur<sup>2</sup>

<sup>1</sup> Université de Picardie Jules Verne

<sup>2</sup> Université de Technologie de Compiègne

Laboratoire Heudiasyc, UMR CNRS 6599

BP 20529 60205 Compiègne

mmasson@hds.utc.fr

### Résumé :

Une nouvelle méthode de classification automatique, intitulée ECM (Evidential  $c$ -means), est proposée. Elle se fonde sur le concept de partition crédale introduit dans le cadre de la théorie des fonctions de croyance. Ce nouveau concept généralise ceux de partition nette, floue, probabiliste ou possibiliste. La partition crédale est obtenue grâce à un algorithme semblable à celui des  $c$ -moyennes floues. Plusieurs outils d'interprétation sont proposés. Un indice de validité permet de guider le choix du nombre de classes. Les tests sur des données synthétiques montrent l'intérêt de la méthode pour l'analyse exploratoire de données.

### Mots-clés :

Classification automatique, théorie des fonctions de croyance, validité de partition, robustesse, estimation inférieure et supérieure d'une classe.

### Abstract:

A new clustering method for object data, called ECM (Evidential clustering) is proposed, in the theoretical framework of belief functions. It is based on the concept of credal partition, extending those of hard, probabilistic, fuzzy and possibilistic ones. An algorithm similar to the fuzzy  $c$ -means is introduced. Several tools for interpreting a credal partition are proposed and a validity index may help the user to choose a proper number of classes. Experiments with synthetic data sets illustrate the usefulness of the method for exploratory data analysis.

### Keywords:

Clustering, belief functions theory, cluster validity, robustness, upper and lower estimations of a cluster.

## 1 Introduction

Les méthodes de classification automatique sont des techniques classiques d'analyse exploratoire de données. Elles consistent à regrouper  $n$  individus en  $c$  classes  $\omega_1, \dots, \omega_c$ . La similarité des objets est soit donnée directement (on parle alors de données relationnelles), soit calculée à partir d'un tableau individus-

variables. On trouve de nombreuses méthodes de classification automatique dans la littérature. Les méthodes nettes supposent que les individus appartiennent à une classe unique. Au contraire, dans les méthodes floues (ou probabilistes), chaque individu a la possibilité d'appartenir à plusieurs classes avec des degrés d'appartenance différents. On cherche alors, en minimisant par une méthode itérative un critère approprié, à déterminer une matrice de partition floue  $U = (u_{ik})$  où  $u_{ik} \in [0; 1]$  désigne le degré d'appartenance de l'individu  $i$  à la classe  $\omega_k$ , sous la contrainte de normalisation  $\sum_k u_{ik} = 1$ . Les algorithmes les plus utilisés sont celui de Bezdek [3] intitulé FCM (fuzzy  $c$ -means) pour les données individus-variables et son équivalent relationnel RFCM (Relational Fuzzy  $c$ -means) [8]. Certains auteurs, observant des résultats contre-intuitifs en présence de données bruitées ou aberrantes, ont proposé de relâcher la contrainte de normalisation imposée aux appartenances. C'est ainsi qu'ont été proposés l'algorithme possibiliste de Krishnapuram et Keller (PCM pour possibilistic  $c$ -means) [10] et celui de Davé [4] (NC pour Noise-clustering).

Dans cet article nous introduisons une nouvelle approche basée sur le concept de partition crédale, défini dans le cadre de la théorie des fonctions de croyance. Une *partition crédale* généralise les notions de partition nette, floue, probabiliste et possibiliste. Des tra-

vaux antérieurs avaient permis de déterminer une telle partition à partir de données relationnelles [6]. Dans cet article, nous proposons une nouvelle version, inspirée des algorithmes précédents, qui est applicable à des données individus-variables.

L'article est organisé comme suit. Dans un premier temps, nous rappelons certains concepts fondamentaux de la théorie des fonctions de croyance. Ensuite, différents algorithmes sont passés en revue et commentés. L'algorithme évidentiel des  $c$ -moyennes est présenté dans le paragraphe 4. Nous détaillons l'algorithme, proposons plusieurs outils d'interprétation ainsi qu'un indice de validité de classe. Le paragraphe 5 conclut l'article.

## 2 Théorie des fonctions de croyance

Considérons une variable  $x$  prenant sa valeur dans un ensemble fini ordonné  $\Omega$ . Une connaissance partielle quant à la valeur de  $x$  peut être représentée par une *masse de croyance* [13, 14], définie comme une fonction  $m$  de  $2^\Omega$  vers  $[0, 1]$ , vérifiant :

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (1)$$

Les sous-ensembles de  $\Omega$  qui reçoivent une masse non nulle sont appelés des *éléments focaux* de  $m$ . L'absence d'information sur  $x$  est représentée par  $m(\Omega) = 1$ , alors qu'une connaissance parfaite conduit à l'allocation de la totalité de la masse à un singleton de  $\Omega$  (on parle alors de masse certaine). Lorsque tous les éléments focaux sont des singletons,  $m$  est équivalente à une distribution de probabilité. Une masse de croyance telle que  $m(\emptyset) = 0$  est dite *normale*. Si la masse est sous-normale, on travaille alors dans l'hypothèse de *monde ouvert*, ce qui signifie que l'on accepte l'hypothèse que la valeur de  $x$  puisse être à l'extérieur de  $\Omega$  [14]. Une masse de croyance peut être, de façon équivalente, représentée par une fonction de *plausibilité*  $pl : 2^\Omega \mapsto [0, 1]$ , définie par :

$$pl(A) \triangleq \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega. \quad (2)$$

Il faut noter que  $pl$  est équivalente à une mesure de probabilité lorsque les éléments focaux sont des singletons et à une mesure de possibilité si les éléments focaux sont emboîtés.

La connaissance disponible étant représentée par une fonction de croyance, il est souvent nécessaire de prendre une décision quant à la valeur de  $x$ . Plusieurs possibilités existent. Une première solution consiste à choisir la valeur de plus grande plausibilité. Une autre approche consiste à transformer les masses en une distribution de probabilité grâce à la *transformation pignistique* proposée par Smets [14]. Elle est définie par :

$$\text{BetP}(\omega) \triangleq \sum_{A \in \mathcal{A}} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega, \quad (3)$$

où  $|A|$  désigne le cardinal de  $A \subseteq \Omega$ . Si la masse est sous-normale ( $m(\emptyset) \neq 0$ ), il faut procéder à une normalisation préalable de  $m$ , par exemple en transférant la masse de l'ensemble vide à  $\Omega$  [15].

## 3 Partitions floues et possibilistes

Soient  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  des vecteurs de  $\mathbb{R}^p$  décrivant les  $n$  individus. On fixe  $c$  ( $2 \leq c < n$ ) le nombre de classes. Chaque classe est représentée par un centre  $\mathbf{v}_k \in \mathbb{R}^p$ . Soit  $V$  la matrice de dimension  $(c \times p)$  des coordonnées des centres des classes telle que  $V_{kq}$  est la  $q^{\text{ième}}$  composante du centre  $\mathbf{v}_k$ . L'algorithme FCM cherche une matrice de partition  $U = (u_{ik})$  (avec  $u_{ik} \in [0, 1] \quad \forall i, k$ ) de dimension  $(n \times c)$  et la matrice  $V$  par minimisation du critère suivant :

$$J_{\text{FCM}}(U, V) \triangleq \sum_{i=1}^n \sum_{j=1}^c u_{ij}^\beta d_{ij}^2, \quad (4)$$

sous la contrainte de normalisation de la partition. Dans la fonction (4),  $\beta > 1$  est un coefficient réglant la dureté de la partition et  $d_{ij}$  désigne la distance Euclidienne entre  $\mathbf{x}_i$  et le centre  $\mathbf{v}_j$ . Le critère est minimisé grâce à un algorithme itératif, optimisant alternativement la position des centres et les degrés d'appartenance. Les formules de mise à jour de  $U$  et  $V$

sont obtenues par introduction d'un multiplicateur de Lagrange associé à la contrainte de normalisation et en annulant les dérivées partielles du Lagrangien par rapport aux paramètres du modèle.

Dans la version possibiliste des  $c$ -moyennes introduite par Krishnapuram and Keller [10], la contrainte de normalisation n'est plus imposée ; pour éviter la solution triviale  $u_{ij} = 0$  pour tout  $i$  et  $j$ , un terme de pénalité est ajouté à la fonction objective :

$$J_{\text{PCM}}(U, V) \triangleq \sum_{i=1}^n \sum_{j=1}^c u_{ij}^\beta d_{ij}^2 + \sum_{j=1}^c \eta_j \sum_{i=1}^n (1 - u_{ij})^\beta, \quad (5)$$

où les  $\eta_i$  sont des poids positifs fixés par l'utilisateur, modulant les effets opposés des deux termes dans  $J_{\text{PCM}}$ . Une procédure d'optimisation alternée, similaire à FCM, peut être aussi déduite des conditions nécessaires d'optimalité pour  $U$  et  $V$ . Un des intérêts de l'algorithme réside dans le fait que le degré d'appartenance d'un individu à une classe donnée ne dépend que de sa distance à cette classe, reflétant par conséquent un degré de typicalité plutôt qu'un degré relatif d'appartenance. Cette propriété permet à l'algorithme de détecter des individus aberrants.

Une autre approche, présentée comme robuste aux individus atypiques, est celle de Davé [4]. Elle consiste à ajouter aux  $c$  classes initiales une classe de bruit (supposée à une distance fixe  $\delta$  de tous les individus). Le paramètre  $\delta$  contrôle le taux d'individus considérés comme aberrants. L'appartenance  $u_{i*}$  d'un individu  $i$  à la classe de bruit est donnée par :

$$u_{i*} = 1 - \sum_{k=1}^c u_{ik}. \quad (6)$$

Le critère minimisé est le suivant :

$$J_{\text{NC}}(U, V) \triangleq \sum_{i=1}^n \sum_{j=1}^c u_{ij}^\beta d_{ij}^2 + \sum_{i=n}^c \delta^2 \left( 1 - \sum_{j=1}^c u_{ij} \right)^\beta. \quad (7)$$

L'algorithme itératif qui en découle est similaire à FCM et PCM.

## 4 C-moyennes évidentielles

### 4.1 Algorithme

Dans des travaux récents sur la classification automatique de données relationnelles, nous avons proposé de représenter une connaissance partielle sur la classe d'un individu  $i$  par une masse de croyance  $m_i$  définie sur l'ensemble  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ . Cette représentation permet de modéliser toutes les situations, de l'ignorance totale à la connaissance la plus précise. L'ensemble des masses  $M = (m_1, \dots, m_n)$  forme une *partition crédale* qui généralise les partitions nette, floue, et possibiliste. Le nombre de degrés de liberté plus important que dans les autres types de partition permet d'envisager une meilleure modélisation et une description plus riche de données complexes.

**Exemple 1** Considérons un ensemble de  $n = 5$  individus et  $c = 3$  classes. On suppose que les masses décrivant la répartition des individus dans les classes sont données dans le tableau 1. Cet exemple a pour but d'illustrer comment une partition crédale permet de représenter des situations variées : la classe de l'individu 2 est connue avec certitude, alors que l'incertitude pour l'individu 4 est totale ; le cas de l'individu 3 correspond à une situation de connaissance partielle (notons que  $m_3$  est une masse Bayésienne) ; enfin, la masse  $m_1(\emptyset) = 1$  indique que la classe l'individu 1 n'appartient pas à  $\Omega$  (l'individu 1 est un individu atypique).

Tableau 1 – Exemple de partition crédale

$A$	$m_1(A)$	$m_2(A)$	$m_3(A)$	$m_4(A)$
$\emptyset$	1	0	0	0
$\{\omega_1\}$	0	0	0.2	0
$\{\omega_2\}$	0	1	0.4	0
$\{\omega_1, \omega_2\}$	0	0	0	0
$\{\omega_3\}$	0	0	0.4	0
$\{\omega_1, \omega_3\}$	0	0	0	0
$\{\omega_2, \omega_3\}$	0	0	0	0
$\Omega$	0	0	0	1

Déterminer une partition crédale à partir de données individus-variables revient à chercher, pour chaque individu  $i$ , les quantités  $m_{ij} =$

$m_i(A_j)$  ( $A_j \neq \emptyset, A_j \subseteq \Omega$ ) de sorte que  $m_{ij}$  soit d'autant plus grande que la distance  $d_{ij}$  entre  $i$  et l'élément focal  $A_j$  est faible. Il s'agit donc de définir la distance entre un individu et tout sous-ensemble de  $\Omega$ . Comme pour les algorithmes FCM, PCM et NC, on choisit de représenter chaque classe  $\omega_k$  par un centre  $\mathbf{v}_k \in \mathbb{R}^p$ . Nous proposons d'associer à chaque sous-ensemble  $A_j$  de  $\Omega$  le barycentre  $\bar{\mathbf{v}}_j$  des centres des classes composant  $A_j$ . Plus précisément, en introduisant la notation

$$s_{kj} = \begin{cases} 1 & \text{si } \omega_k \in A_j \\ 0 & \text{sinon,} \end{cases} \quad (8)$$

nous calculons le barycentre  $\bar{\mathbf{v}}_j$  associé à  $A_j$  par :

$$\bar{\mathbf{v}}_j = \frac{1}{c_j} \sum_{k=1}^c s_{kj} \mathbf{v}_k, \quad (9)$$

où  $c_j = |A_j|$  désigne le cardinal de  $A_j$ . La distance  $d_{ij}$  est alors définie par :

$$d_{ij}^2 \triangleq \|\mathbf{x}_i - \bar{\mathbf{v}}_j\|^2. \quad (10)$$

D'autre part, comme dans la méthode de Davé, un traitement différent est adopté pour l'ensemble vide. Assimilé à la classe de bruit, il est supposé à une distance fixe de tous les individus. Nous proposons alors de minimiser le critère suivant :

$$J_{\text{ECM}}(M, V) \triangleq \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (11)$$

sous les contraintes :

$$\sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1, n, \quad (12)$$

où  $m_{i\emptyset}$  désigne  $m_i(\emptyset)$ . Le critère  $J_{\text{ECM}}$  est similaire à  $J_{\text{NC}}$  à un coefficient supplémentaire  $c_j^\alpha$  près, dont le rôle est de pénaliser les sous-ensembles de  $\Omega$  de cardinalité importante; l'exposant  $\alpha$  permet de contrôler le degré de pénalisation. Les paramètres  $\beta$  et  $\delta$  ont la même signification que dans la méthode de Davé. Pour établir l'algorithme d'optimisation, on procède

de la même manière que pour FCM et NC. On considère dans un premier temps que  $V$  est fixe. Pour résoudre le problème de minimisation sous contraintes par rapport à  $M$ , on introduit  $n$  multiplicateurs de Lagrange  $\lambda_i$  :

$$\mathcal{L}(M, \lambda_1, \dots, \lambda_n) = J_{\text{ECM}}(M, V) - \sum_{i=1}^n \lambda_i \left( \sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} - 1 \right). \quad (13)$$

En dérivant le Lagrangien par rapport aux  $m_{ij}$ ,  $m_{i\emptyset}$  et  $\lambda_i$  et en annulant les dérivées partielles, on obtient :

$$\frac{\partial \mathcal{L}}{\partial m_{ij}} = \beta c_j^\alpha m_{ij}^{\beta-1} d_{ij}^2 - \lambda_i = 0, \quad (14)$$

$$\frac{\partial \mathcal{L}}{\partial m_{i\emptyset}} = \beta \delta^2 m_{i\emptyset}^{\beta-1} - \lambda_i = 0, \quad (15)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} - 1 = 0. \quad (16)$$

En adoptant une méthode de résolution similaire à celle de Davé, on obtient finalement l'équation suivante de mise à jour des masses :

$$m_{ij} = \frac{c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} c_k^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (17)$$

et

$$m_{i\emptyset} = 1 - \sum_{A_j \neq \emptyset} m_{ij} \quad \forall i = 1, n. \quad (18)$$

On considère maintenant que  $M$  est fixe. La minimisation de  $J_{\text{ECM}}$  par rapport à  $V$  est un problème d'optimisation non contraint. Les dérivées partielles de  $J_{\text{ECM}}$  par rapport aux centres des classes sont données par :

$$\frac{\partial J_{\text{ECM}}}{\partial \mathbf{v}_l} = \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^\alpha m_{ij}^\beta \frac{\partial d_{ij}^2}{\partial \mathbf{v}_l}. \quad (19)$$

$$\frac{\partial d_{ij}^2}{\partial \mathbf{v}_l} = 2(s_{lj})(\mathbf{x}_i - \bar{\mathbf{v}}_j) \left(-\frac{1}{c_j}\right). \quad (20)$$

De (19) et (20) on tire :

$$\frac{\partial J_{\text{ECM}}}{\partial \mathbf{v}_l} = -2 \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} (\mathbf{x}_i - \bar{\mathbf{v}}_j) \quad (21)$$

$$= -2 \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} \left( \mathbf{x}_i - \frac{1}{c_j} \sum_k s_{kj} \mathbf{v}_k \right) \quad (22)$$

$\forall l = 1, c.$

En annulant ces dérivées, on trouve  $l$  équations linéaires par rapport à  $\mathbf{v}_k$  :

$$\sum_i \mathbf{x}_i \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} \quad (23)$$

$$= \sum_k \mathbf{v}_k \sum_i \sum_{A_j \neq \emptyset} c_j^{\alpha-2} m_{ij}^\beta s_{lj} s_{kj} \quad l = 1, c.$$

Soit  $B$  une matrice de dimension  $(c \times p)$  définie par :

$$B_{lq} = \sum_{i=1}^n x_{iq} \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} \quad (24)$$

$$= \sum_{i=1}^n x_{iq} \sum_{A_j \ni \omega_l} c_j^{\alpha-1} m_{ij}^\beta \quad l = 1, c \quad q = 1, p,$$

et  $H$  une matrice de dimension  $(c \times c)$  définie par :

$$H_{lk} = \sum_i \sum_{A_j \neq \emptyset} c_j^{\alpha-2} m_{ij}^\beta s_{lj} s_{kj} \quad (25)$$

$$= \sum_i \sum_{A_j \ni \{\omega_k, \omega_l\}} c_j^{\alpha-2} m_{ij}^\beta \quad k, l = 1, c.$$

Avec ces notations,  $V$  est solution du système linéaire suivant :

$$HV = B. \quad (26)$$

L'algorithme complet est décrit dans le tableau 2.

## 4.2 Outils d'interprétation

Suivant ce qui est attendu de l'analyse, plusieurs outils peuvent aider l'utilisateur à interpréter les résultats fournis par l'algorithme

Tableau 2 – Algorithme ECM

Entrées	$\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ : $n$ vecteurs de $\mathbb{R}^p$
Param.	$c$ : nombre de classes $1 < c < n$ $\alpha \geq 0$ : (défaut 1) $\beta > 1$ : (défaut 2) $\delta > 0$ : distance à l'ens. vide $\epsilon$ : seuil d'arrêt (défaut $10^{-3}$ )
Init.	Choix aléatoire de $V_0$
	$t \leftarrow 0$
	<b>Répéter</b>
	$t \leftarrow t + 1$
	• Calcul de $M_t$ grâce à (10), (9), (17), (18), et $V_{t-1}$ ;
	• Calcul de $H_t$ et $B_t$ grâce à (24), (25), et $M_t$ ;
	• Résolution de $H_t V_t = B_t$ ;
	<b>jusqu'à</b> $\ V_t - V_{t-1}\  < \epsilon$

ECM. En premier lieu, une partition crédale peut être convertie en une partition possibiliste en calculant pour chaque masse  $m_i$  la plausibilité  $\text{pl}_i$  des différentes classes :

$$\text{pl}_i(\{\omega_k\}) = \sum_{A \cap \{\omega_k\} \neq \emptyset} m_i(A). \quad (27)$$

La valeur  $\text{pl}_i(\{\omega_k\})$  représente la plausibilité (ou la possibilité) que l'individu  $i$  appartienne à la classe  $k$ . De la même façon, une partition floue peut être obtenue en calculant les probabilités pignistiques  $\text{BetP}_i(\{\omega_k\})$  induite par chaque  $m_i$  et en interprétant cette valeur comme le degré d'appartenance de l'individu  $i$  à la classe  $k$ . Si nécessaire, il est facile d'obtenir une partition nette en affectant chaque individu à la classe la plus probable. Notons que les individus associés à des masses importantes allouées à l'ensemble vide peuvent être rejetés avant une affectation aux classes.

De manière alternative, il est possible de suivre un point de vue similaire à celui de Lingras [11], qui se place dans le cadre de la théorie des ensembles approximatifs. Il propose d'associer à chaque classe deux ensembles d'individus : l'estimation inférieure de la classe  $k$ , notée

$\omega_k^L$ , rassemble les individus qui peuvent être classés sans ambiguïté dans  $\omega_k$ . L'estimation supérieure, notée  $\omega_k^U$ , désigne l'ensemble des individus qui appartiennent potentiellement à  $\omega_k$  mais qui peuvent également appartenir à une autre classe. Formellement, soit  $X(A_j)$  l'ensemble des individus dont la masse la plus importante est portée sur  $A_j$  :

$$X(A_j) = \{i/m_i(A_j) = \max_{A_k \in 2^\Omega \setminus \emptyset} m_i(A_k)\}. \quad (28)$$

On définit les ensembles inférieur et supérieur associés à chaque classe  $\omega_k$  par :

$$\omega_k^L = X(\{\omega_k\}), \quad (29)$$

et

$$\omega_k^U = \bigcup_{\{j/\omega_k \in A_j\}} X(A_j). \quad (30)$$

**Exemple 2** Pour illustrer ce point, considérons le jeu de données synthétiques bidimensionnelles suivant : quatre classes de 200 points chacune ont été générées à partir de lois de Student multivariées à 5 degrés de liberté de centres respectifs  $[0;0]$ ,  $[0;4]$ ,  $[4;0]$  et  $[4;4]$ . Ce jeu de données est représenté sur la figure 1. L'algorithme a été chargé de rechercher une partition crédale en quatre classes avec les paramètres suivants :  $\alpha = 1$ ,  $\beta = 2$ ,  $\delta^2 = 20$  and  $\epsilon = 10^{-3}$ . La figure 2 représente l'affectation des individus à l'élément focal de  $\Omega$  qui reçoit la masse la plus élevée (notons que  $\omega_{ij}$  désigne  $\{\omega_i, \omega_j\}$  et  $\omega_{ijk}$  désigne l'ensemble  $\{\omega_i, \omega_j, \omega_k\}$ ). Chaque sous-ensemble est représenté par son enveloppe convexe. Le centre des classes est repéré par une croix. Les individus de  $X(\emptyset)$ , pour lesquels la masse la plus importante est attribuée à l'ensemble vide, sont signalés par des carrés. On voit qu'une partition satisfaisante est trouvée et que les individus aberrants sont correctement détectés. Les figures 3 and 4 montrent les estimations inférieures et supérieures des classes, représentées également par leur enveloppe convexe.

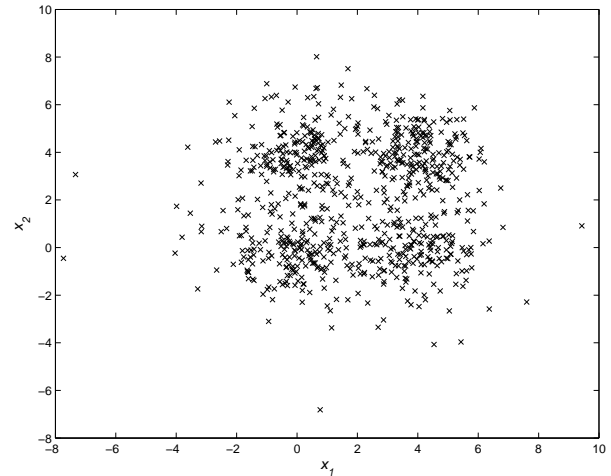


Figure 1 – Jeu de données synthétique

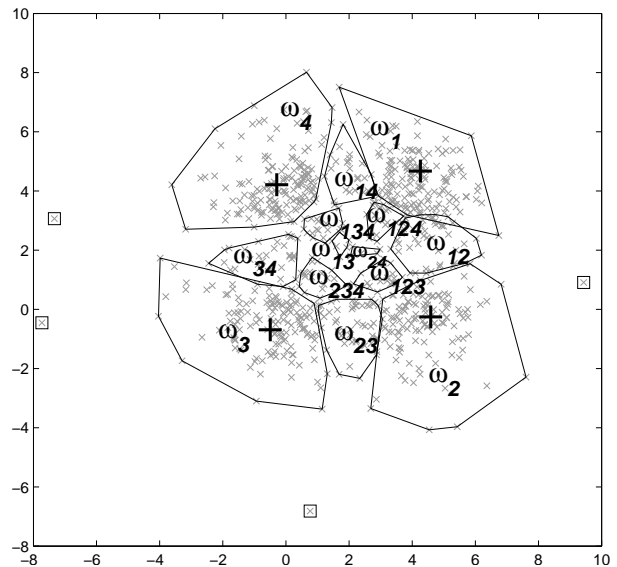


Figure 2 – Partition calculée à partir de  $m$

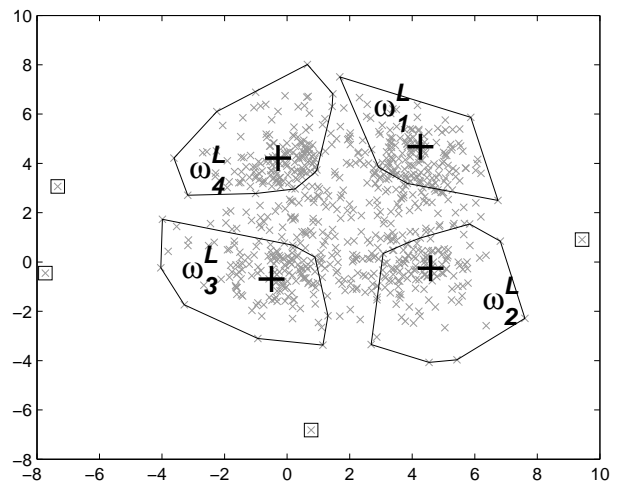


Figure 3 – Estimations inférieures des 4 classes

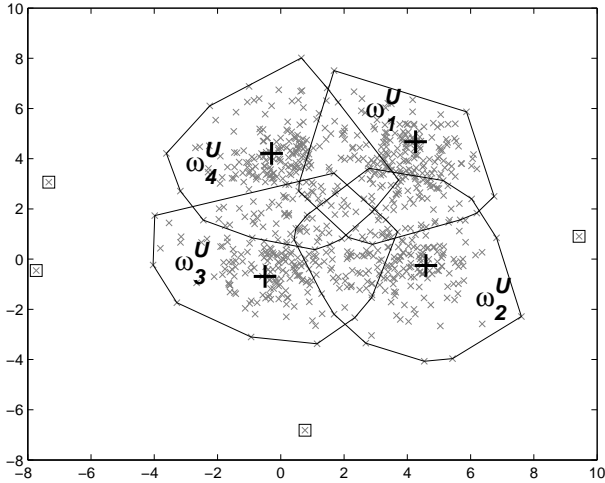


Figure 4 – Estimations supérieures des 4 classes

### 4.3 Limiter la complexité

Pour chaque individu, l'algorithme ECM distribue une fraction de la masse unité à chaque élément de  $2^\Omega$ . Par conséquent, le nombre de paramètres varie de façon exponentielle par rapport au nombre de classes (et linéaire par rapport au nombre d'individus). Pour un nombre de classes limité (inférieur à 10), les temps de calculs restent acceptables. A titre illustratif, pour le jeu de données synthétiques composé de 800 individus, le tableau 3 rapporte les temps moyens d'exécution (obtenus sur 100 initialisations différentes) en fonction du nombre de classes désiré.

Tableau 3 – Temps d'exécution moyens sur le jeu de données synthétiques

$c$	2	3	4	5	6
$t$	3s	11.8s	12.2s	63.6	185s

Si nécessaire, il est cependant possible de réduire la complexité de l'algorithme en imposant un nombre restreint d'éléments focaux. Par exemple, on peut se limiter à  $\Omega$ , et des éléments focaux de cardinalité inférieure ou égale à 2. On réduit ainsi considérablement la complexité, tout en gardant une certaine flexibilité.

### 4.4 Validité de la partition

Un des problèmes centraux des méthodes de classification automatique réside dans le choix *a priori* du nombre de classes. On parle de validité de la partition. La plupart des méthodes proposées dans la littérature consistent à calculer un indice de validité à partir de partitions obtenues pour différentes valeurs de  $c$  et à chercher pour quelle valeur de  $c$  on observe soit un minimum, soit un maximum, soit un changement brusque du critère. C'est le cas par exemple pour le coefficient de partition et ses variantes [2] [5], et l'entropie de partition [1].

Dans le cas d'une partition crédale, on sent intuitivement que si le nombre de classes est correctement choisi, les centres vont être situés dans les zones de forte densité et la majeure partie de la masse sera allouée aux singletons de  $\Omega$ . Au contraire, si  $c$  est trop faible ou trop élevé, la masse risque d'être distribuée à des sous-ensembles de plus forte cardinalité ou à l'ensemble vide. En d'autres termes, plus une partition crédale sera *spécifique*, mieux elle représentera la structure des données. Ces considérations amènent à l'utilisation de mesures d'entropie définies pour les fonctions de croyance.

Klir and Wierman [9, p. 51] ont défini la *non-spécificité* d'une masse sous-normale  $m$  par :

$$N(m) \triangleq \sum_{A \in 2^\Omega \setminus \emptyset} m(A) \log_2 |A| + m(\emptyset) \log_2 |\Omega|, \quad (31)$$

où  $0 \leq N(m) \leq \log_2(|\Omega|)$ . Cette mesure tend à être faible lorsque la masse est allouée à peu d'éléments focaux ( $\neq \emptyset$ ), et de faible cardinalité. Nous proposons de définir un indice de validité pour une partition crédale comme la moyenne sur l'ensemble des  $m_i$  des non-spécificités normalisées :

$$N^*(c) \triangleq \frac{1}{n \log_2(c)} \sum_{i=1}^n N(m_i), \quad (32)$$

où  $0 \leq N^*(c) \leq 1$ . Cet indice doit être minimisé. Son application est illustrée sur le jeu

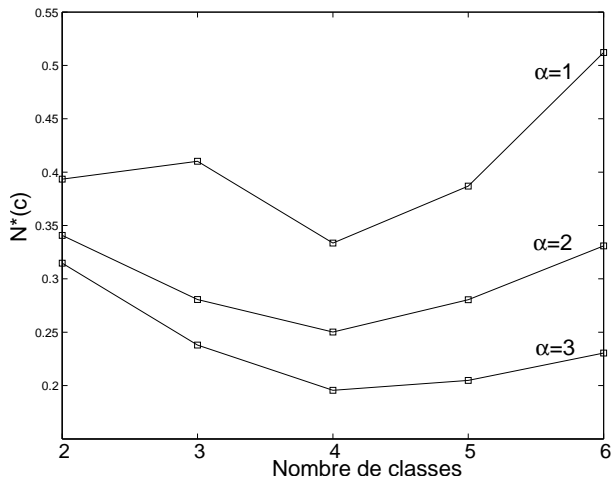


Figure 5 – Indice de validité

de données précédent. Les paramètres d'ECM sont fixés à  $\beta = 2$ ,  $\delta^2 = 20$  et  $\epsilon = 10^{-3}$ . Pour différentes valeurs de  $\alpha$  (1, 2 and 3), on fait varier le nombre de classes entre 2 et 6. Les résultats obtenus sont présentés sur la figure 5. On voit que le minimum est toujours obtenu pour  $c = 4$  classes, ce minimum étant moins prononcé à mesure que  $\alpha$  augmente, ou, de façon équivalente, à mesure que les éléments focaux de grande cardinalité sont de plus en plus pénalisés. Cet exemple simple suggère que  $N^*$  peut constituer un bon indice de validité d'une partition crédale.

## 5 Conclusion

Nous avons proposé dans cet article une nouvelle méthode de classification automatique fondée sur la théorie des fonctions de croyance. Les concepts de partitions floue, possibiliste et nette sont présentés comme des cas particuliers d'un concept plus général, celui de partition crédale. Pour calculer une telle partition à partir d'un tableau individus-variables, un algorithme, intitulé ECM, similaire à FCM, a été proposé. Il est fondé sur un schéma de minimisation alterné, avec, dans un premier temps, la détermination des centres des classes, et, dans un second temps, l'allocation de masses aux différents sous-ensembles de classes. L'interprétation de la partition est facilitée par la

possibilité d'associer à chaque classe ses estimations inférieure et supérieure, notions liées respectivement à la nécessité et la possibilité qu'un individu appartienne à une classe donnée. Enfin un indice de validité de classe, fondé sur la non-spécificité d'une fonction de croyance, a été proposé et illustré sur un exemple simple. Ce travail ouvre un certain nombre de perspectives, au nombre desquelles une version à base de noyaux et un algorithme plus efficace pour les données relationnelles.

## Références

- [1] J. C. Bezdek. Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3, 58-73, 1974.
- [2] J. C. Bezdek. *Pattern Recognition with fuzzy objective function algorithms.* Plenum Press, New-York, 1981.
- [3] J. C. Bezdek, J. Keller, R. Krishnapuram, and N. R. Pal. *Fuzzy models and algorithms for pattern recognition and image processing.* Kluwer Academic Publishers, Boston, 1999.
- [4] R.N. Davé. Clustering relational data containing noise and outliers. *Pattern Recognition Letters*, 12, 657-664, 1991.
- [5] R.N. Davé. Validating fuzzy partition obtained through  $c$ -shell clustering. *Pattern Recognition Letters*, 17, 613-623, 1996.
- [6] T. Denceux, and M.-H. Masson. EVCLUS : Evidential CLUSTERing of proximity data. *IEEE Transactions on Systems, Man and Cybernetics Part B*, 34(1), 95-109, 2004.
- [7] D. Dubois and H. Prade. *Possibility Theory : An approach to computerized processing of uncertainty.* Plenum Press, New-York, 1988.
- [8] R.J. Hathaway, J.C. Bezdek, and J.W. Davenport. On relational versions of  $c$ -means algorithm. *Pattern Recognition Letters*, 17, 607-612, 1996.
- [9] G.J. Klir, and M.J. Wierman. *Uncertainty-based Information. Elements of Generalized Information Theory.* Springer-Verlag, New-York, 1998.
- [10] R. Krishnapuram, and J. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2), 98-110, 1993.
- [11] P. Lingras. Unsupervised rough set classification using GAs. *Journal of Intelligent Information Systems*, 16, 215-228, 2001.
- [12] A. Ramer. Uniqueness of information measure in the theory of evidence. *Fuzzy Sets and Systems*, 24, 183-196, 1987.
- [13] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.
- [14] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66, 191-243, 1994.
- [15] R. R. Yager. On the normalization of fuzzy belief structure. *International Journal of Approximate Reasoning*, 14, 27-153, 1996.