

CECM : Algorithme évidentiel des c -moyennes avec contraintes

CECM : Constrained evidential c -means algorithm

V. Antoine^{1,3}

B. Quost^{1,3}

M.-H. Masson^{2,3}

T. Dencœur^{1,3}

¹ Université de Technologie de Compiègne

² Université de Picardie Jules Verne, IUT de l'Oise

³ Laboratoire Heudiasyc, UMR CNRS 6599

BP 20529 60205 Compiègne

violaine.antoine@hds.utc.fr

Résumé :

Un nouvel algorithme de classification automatique basé sur ECM et intitulé CECM (Constrained Evidential c -means) est proposé. Il se fonde sur deux concepts : la partition crédale reposant sur la théorie des fonctions de croyance et la classification sous contraintes. La notion de partition crédale généralise celles de partition nette, floue, probabiliste ou possibiliste. La classification sous contraintes est une approche qui utilise peu de connaissances a priori. Elle est représentée dans cet article par des contraintes *Must-link* et *Cannot-link* entre des paires d'individus. Ce concept permet de diriger l'algorithme vers une solution et ainsi d'améliorer les performances. Nous introduisons dans ECM la métrique de Mahalanobis et nous modifions le critère minimisé par ECM pour prendre en compte les contraintes. Les performances sont illustrées sur plusieurs jeux de données.

Mots-clés :

Classification automatique, théorie des fonctions de croyance, contraintes par paires, classification par contraintes.

Abstract:

A new clustering algorithm for object data derived from ECM, called CECM (Constrained Evidential c -means) is proposed. It is based on two concepts : the credal partition relying on the theoretical framework of belief functions, and constrained clustering. The notion of credal partition extends those of hard, probabilistic, fuzzy and possibilistic ones. Constrained classification is an approach which uses few additional knowledge using *Must-link* or *Cannot-link* constraints between two points. This concept enables to direct the algorithm towards a solution and improves the robustness of the results. We introduce in ECM the Mahalanobis distance and we modify the criterion minimized by ECM to take into account the constraints. The performances are illustrated on several databases.

Keywords:

Clustering, belief functions theory, pairwise constraints, constrained clustering.

1 Introduction

La classification non supervisée est une méthode classique d'analyse de données qui consiste à regrouper des individus ou objets en classes en se fondant uniquement sur une notion de similarité entre leurs descripteurs. Cependant, il existe de nombreuses applications pour lesquelles des informations connues ou peu coûteuses à collecter pourraient être prises en compte par l'algorithme de classification pour permettre une amélioration des résultats de la classification. Ainsi Wagstaff [1] a introduit des contraintes spécifiant si deux objets appartiennent au même groupe (contrainte *Must-link*) ou si deux objets sont d'un groupe différent (contrainte *Cannot-link*). Ces couples contraints ont été étudiés et intégrés dans de nombreux algorithmes non supervisés tels que les c -moyennes (CM) et les c -moyennes floues (FCM) [2]. Ils sont pris en compte de diverses manières, le plus souvent en ajoutant un terme de pénalité dans la fonction objectif de l'algorithme [3, 4], ou en modifiant les distances entre objets à l'aide des contraintes [5].

L'algorithme FCM est une méthode associant à chaque individu des degrés d'appartenance à chacune des classes. Ces degrés d'appartenance sont rassemblés dans une matrice de partition floue $\mathbf{U} = (u_{ik})$ calculée en minimisant une fonction objectif appropriée, où u_{ik} représente la valeur d'appartenance de l'élément i à la

classe ω_k . Les contraintes suivantes doivent être respectées :

$$u_{ik} \geq 0 \quad \forall i, k, \quad (1)$$

$$\sum_k u_{ik} = 1. \quad (2)$$

L'inconvénient de la méthode est qu'elle aboutit parfois à des résultats contre-intuitifs en présence de données bruitées ou aberrantes. C'est pourquoi certains auteurs ont proposé des algorithmes possibilistes [6, 7] et des algorithmes définis dans le cadre des fonctions de croyances [8, 9, 10]. Ainsi l'algorithme ECM [9, 10] permet d'obtenir une partition crédale des données, notion qui généralise celles des partitions nettes, floues, probabilistes et possibilistes. Dans cet article nous proposons d'ajouter des couples de contraintes à l'algorithme ECM pour ainsi combiner les avantages de l'ajout de connaissances a priori et du cadre des fonctions de croyance. Nous présentons également une formulation de ECM prenant en compte la distance de Mahalanobis.

L'article est organisé comme suit. Dans une première partie, nous rappelons les concepts fondamentaux de la théorie des fonctions de croyance et son utilisation pour l'algorithme ECM. Ensuite nous définissons un nouvel algorithme baptisé CECM en décrivant tout d'abord les contraintes utilisées puis la formulation de la fonction objectif obtenue. Nous proposons une formulation d'ECM prenant en compte la distance de Mahalanobis. Les résultats de l'algorithme sont présentés dans la troisième partie. Enfin, nous concluons sur quelques perspectives du travail.

2 Fonctions de croyance et classification

2.1 Théorie des fonctions de croyance

La théorie de Dempster-Shafer [11] (ou théorie des fonctions de croyance) est utilisée comme un formalisme pour la modélisation de

connaissances incertaines. Seuls les principaux concepts sont rappelés dans cette partie.

Considérons un ensemble fini $\Omega = \{\omega_1, \dots, \omega_c\}$ appelé cadre de discernement dans lequel une variable y prend une valeur. La connaissance partielle de la valeur de y peut être modélisée par une fonction de masse m de $2^\Omega \rightarrow [0, 1]$ vérifiant :

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (3)$$

Les sous-ensembles A de Ω tels que $m(A) > 0$ sont appelés des éléments focaux de m . Ainsi la quantité $m(A)$ s'interprète comme la quantité de croyance placée dans A et qui ne peut pas être affectée à un sous-ensemble de A . Une absence totale d'information sur la variable y est représentée par la fonction de masse vide : $m(\Omega) = 1$. A l'inverse, une fonction de masse peut allouer toute la masse de croyance à un singleton de Ω : $m(\{\omega_k\}) = 1$. Notons que si tous les éléments focaux sont des singletons alors m est similaire à une distribution de probabilité. Sous l'hypothèse d'un monde ouvert, une masse $m(\emptyset) > 0$ représente la croyance que la valeur prise par y n'appartient pas à Ω [12].

Etant donnée une fonction de masse m , il est possible de définir une fonction de plausibilité $pl : 2^\Omega \rightarrow [0, 1]$ et une fonction de croyance $bel : 2^\Omega \rightarrow [0, 1]$:

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega, \quad (4)$$

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega. \quad (5)$$

La première fonction désigne la quantité de croyance qui pourrait être allouée à A sous réserve d'informations supplémentaires, la seconde, le degré total de croyance spécifique et justifiée en A . On a :

$$pl(A) = 1 - bel(\bar{A}). \quad (6)$$

Quand une décision doit être prise pour la valeur de y , on peut transformer la fonction de

masse en probabilité *pignistique* [13] :

$$BetP(\omega) = \sum_{\omega \in A} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega, \quad (7)$$

où $|A|$ représente le cardinal de $A \subseteq \Omega$. Dans le cas où $m(\emptyset) \neq 0$, une normalisation de m est nécessaire avant de procéder à la transformation pignistique, par exemple en transférant $m(\emptyset)$ à Ω [14].

2.2 Partitions floues et crédales

Soient $\{\mathbf{x}_1 \dots \mathbf{x}_n\}$ des individus dans \mathbb{R}^p à classer dans l'ensemble $\Omega = \{\omega_1 \dots \omega_c\}$. Chaque classe ω_k , $k \in \llbracket 1; c \rrbracket$, est définie par un centre de coordonnées $\mathbf{v}_k \in \mathbb{R}^p$. Soit V la matrice des centres des classes ; une matrice $U = (u_{ik})$ est définie comme étant une matrice de partition floue. L'algorithme FCM [2] recherche les matrices U et V permettant la minimisation du critère suivant :

$$J_{FCM}(U, V) = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^\beta d_{ik}^2, \quad (8)$$

sous les contraintes (1) et (2). Dans (8), d_{ik} représente la distance euclidienne entre l'élément \mathbf{x}_i et le centre \mathbf{v}_k et $\beta > 1$ est un coefficient réglant la dureté de la partition. La fonctionnelle est minimisée en alternant l'optimisation de la position des centres et l'optimisation des degrés d'appartenance. Plus de détails peuvent être trouvés dans l'article [2].

Dans la version crédibiliste des c -moyennes introduite par Masson et Denoeux [9, 10], la matrice de partition floue U est remplacée par une matrice de fonctions de masse M , nommée partition crédale. Cette dernière généralise les partitions nettes, floues et possibilistes. Elle est constituée de masses de croyance définies sur tous les sous-ensembles possibles de Ω pour chaque individu. Ainsi m_{ij} correspond au degré de croyance dans le fait que la classe de \mathbf{x}_i appartient à l'ensemble $A_j \subseteq \Omega$. L'algorithme cherche de la même manière que FCM les matrices M et V permettant la minimisation d'un

critère :

$$J_{ECM}(M, V) = \sum_{i=1}^n \sum_{A_k \neq \emptyset} |A_k|^\alpha m_{ik}^\beta d_{ik}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (9)$$

sous les contraintes :

$$\sum_{k/A_k \subseteq \Omega, A_k \neq \emptyset} m_{ik} + m_{i\emptyset} = 1 \quad \forall i \in \llbracket 1; n \rrbracket, \quad (10)$$

où $m_{i\emptyset}$ représente la masse de l'élément i affectée à l'ensemble vide. Cet ensemble est considéré comme une classe de bruit et donc est traité séparément des autres sous-ensembles. Le paramètre δ désigne une distance fixe entre tous les individus et l'ensemble vide. Un coefficient supplémentaire $|A_k|^\alpha$ a été ajouté pour pénaliser les sous-ensembles de Ω dont la cardinalité est importante : l'exposant α permet le contrôle de cette pénalisation.

Chaque classe ω_l est représentée comme dans FCM par un centre \mathbf{v}_l . Le centre d'un élément focal $A_k \subseteq \Omega$ est calculé comme l'isobary-centre des centres associés aux classes composant A_k :

$$\bar{\mathbf{v}}_k = \frac{1}{|A_k|} \sum_{k=1}^c s_{lk} \mathbf{v}_l, \quad (11)$$

avec

$$s_{lk} = \begin{cases} 1 & \text{si } \omega_l \in A_k, \\ 0 & \text{sinon.} \end{cases} \quad (12)$$

La distance d_{ik}^2 entre \mathbf{x}_i et l'ensemble focal A_k est alors définie par :

$$d_{ik}^2 = \|\mathbf{x}_i - \bar{\mathbf{v}}_k\|^2. \quad (13)$$

3 ECM avec contraintes

3.1 Formulation des contraintes

Considérons deux objets \mathbf{x}_i et \mathbf{x}_j décrits par deux fonctions de masses m_i et m_j . En se plaçant dans $\Omega^2 = \Omega \times \Omega$ il est possible de calculer une fonction de masse quantifiant la croyance sur l'appartenance conjointe des individus \mathbf{x}_i et \mathbf{x}_j . Cette fonction de masse notée

$m_{i \times j}$ s'obtient en étendant m_i et m_j sur Ω^2 , puis en les combinant [13, 8]. Son expression est donnée par :

$$m_{i \times j}(A \times B) = m_i(A) m_j(B) \quad (14)$$

$$\forall A, B \subseteq \Omega, A \neq \emptyset, B \neq \emptyset.$$

A partir de $m_{i \times j}$, il est possible de calculer la plausibilité que les objets \mathbf{x}_i et \mathbf{x}_j appartiennent ou non à la même classe. Dans Ω^2 , l'évènement " \mathbf{x}_i et \mathbf{x}_j appartiennent à la même classe" est le sous-ensemble de Ω^2 : $S = \{(\omega_1, \omega_1), (\omega_2, \omega_2), \dots, (\omega_c, \omega_c)\}$ et l'évènement " \mathbf{x}_i et \mathbf{x}_j n'appartiennent pas à la même classe" est le complémentaire \bar{S} de S dans Ω^2 . Les plausibilités correspondantes sont obtenues par :

$$pl_{i \times j}(S) = \sum_{A \cap B \neq \emptyset} m_i(A) m_j(B), \quad (15)$$

$$pl_{i \times j}(\bar{S}) = 1 - bel_{i \times j}(S), \quad (16)$$

$$= 1 - \sum_{A \subseteq \Omega / |A|=1} m_i(A) m_j(A). \quad (17)$$

Prenons par exemple trois individus à classer dans deux classes. Supposons que l'on dispose d'une partition crédale définie par le tableau 1.

Tableau 1 – Exemple de partition crédale

A	$m_1(A)$	$m_2(A)$	$m_3(A)$
\emptyset	0	0	0
$\{\omega_1\}$	1	1	0
$\{\omega_2\}$	0	0	1
Ω	0	0	0

On déduit de cette partition crédale les fonctions de masses suivantes :

$$m_{1 \times 2}(\{(\omega_1, \omega_1)\}) = 1,$$

$$m_{1 \times 2}(A \times B) = 0 \quad \forall (A \times B) \neq (\{(\omega_1, \omega_1)\}),$$

$$m_{1 \times 3}(\{(\omega_1, \omega_2)\}) = 1,$$

$$m_{1 \times 3}(A \times B) = 0 \quad \forall (A \times B) \neq (\{(\omega_1, \omega_2)\}). \quad (18)$$

Ainsi pour l'exemple on aura :

$$pl_{1 \times 2}(S) = 1, \quad pl_{1 \times 2}(\bar{S}) = 0,$$

$$pl_{1 \times 3}(S) = 0, \quad pl_{1 \times 3}(\bar{S}) = 1. \quad (19)$$

On remarque que la plausibilité d'avoir les individus 1 et 2 dans une classe différente est nulle : une contrainte *Must-link* peut donc être établie entre eux. A l'inverse, la plausibilité que les individus 1 et 3 soient dans la même classe est nulle : une contrainte *Cannot-link* peut donc être établie entre eux.

A l'inverse, supposons que la partition crédale est inconnue (cas de l'algorithme ECM), alors si des contraintes sur les individus sont disponibles, il est possible de les traduire sous forme de plausibilités $pl_{i \times j}$ et de les utiliser pour trouver une partition crédale. Si l'on sait que \mathbf{x}_i et \mathbf{x}_j appartiennent à la même classe (noté $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}$), on impose $pl_{i \times j}(\bar{S}) = 0$, et si l'on sait qu'ils appartiennent à des classes différentes (noté $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}$), on fixe $pl_{i \times j}(S) = 0$.

3.2 Formulation quadratique

Nous proposons donc d'intégrer une pénalité dans la fonction objectif J_{ECM} :

$$J_{CECM}(M, V) = J_{ECM}(M, V) + \gamma \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}} pl_{i \times j}(\bar{S}) + \eta \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}} pl_{i \times j}(S), \quad (20)$$

sous les contraintes (10). Le second terme de l'équation (20) est le coût de violer une contrainte *Must-link* tandis que le dernier terme est le coût de violer une contrainte *Cannot-link*. Les coefficients γ et η permettent de contrôler le compromis entre la fonction objectif de ECM et le respect des contraintes.

Si l'on fixe $\beta = 2$, la minimisation de (20) par rapport à M est un problème quadratique à contraintes linéaires : la convergence de l'algorithme est assurée en un temps raisonnable. La minimisation de (20) par rapport à V se fait de la même manière que dans ECM puisque les centres n'interviennent pas dans les contraintes.

3.3 Choix des distances

La distance d_{ik}^2 entre l'individu \mathbf{x}_i et le centre $\bar{\mathbf{v}}_k$ est supposée Euclidienne dans la version de base de ECM. Les classes sont donc sup-

posées être sphériques, cependant l'utilisation de la distance de Mahalanobis dans le cas où les classes sont ellipsoïdales pourrait être intéressante. Chaque classe ω_k est associée à une ellipse d'orientation et de taille spécifiques, définie par une matrice de covariance Σ_k . De ce fait, $\Sigma = \{\Sigma_1 \dots \Sigma_k \dots \Sigma_c\}$ correspond à un nouveau paramètre à optimiser dans la nouvelle fonction objectif $J_{CECM}(M, V, \Sigma)$. Seul le premier terme de cette fonction utilise les distances donc lui seul est modifié. En supposant que le volume des ellipses est identique pour chaque classe ($|\Sigma_k| = 1$ pour $k = \llbracket 1; c \rrbracket$), et en s'inspirant de [15, 4], nous calculons la matrice de covariance associée à ω_l par :

$$\Sigma_l = \frac{\sum_{i=1}^n (m_{il})^2 (\mathbf{x}_i - \mathbf{v}_l)(\mathbf{x}_i - \mathbf{v}_l)^\top}{\sum_{i=1}^n (m_{il})^2}. \quad (21)$$

De la même manière que le calcul des centres (11)(12), la matrice de covariance associée à $A_k \subseteq \Omega$ est calculée comme la moyenne pondérée des matrices associées aux classes composant A_k :

$$\bar{\Sigma}_k = \frac{1}{|A_k|} \sum_{l=1}^c s_{lk} \Sigma_l. \quad (22)$$

Le calcul des distances d_{ik} devient alors :

$$d_{ik}^2 = |\bar{\Sigma}_k|^{\frac{1}{p}} (\mathbf{x}_i - \bar{\mathbf{v}}_k)^\top \bar{\Sigma}_k^{-1} (\mathbf{x}_i - \bar{\mathbf{v}}_k), \quad (23)$$

4 Expérimentations

4.1 Méthode d'évaluation

Dans les expériences réalisées, nous avons utilisé des jeux de données pour lesquelles la partition réelle P est initialement connue : il est donc possible d'évaluer la qualité de la partition \hat{P} calculée par l'algorithme. Soit a (respectivement, b) le nombre de couples d'individus classés dans la même classe (respectivement, dans des classes différentes) simultanément par P et \hat{P} . L'indice de Rand permet de mesurer le degré de concordance global de P et \hat{P} :

$$Rand(P, \hat{P}) = \frac{2(a + b)}{n(n - 1)}. \quad (24)$$

4.2 Données synthétiques

Pour observer l'intérêt d'ajouter des contraintes, nous avons créé un jeu de données *Toys* en deux dimensions supposées être constitué de deux classes, chacune étant un mélange de deux lois normales (cf tableau 2 et figure 1).

Tableau 2 – Construction de *Toys*

μ	\mathcal{N} σ	Nb objets	classe
(0, 0)		100	1
(0, 7)	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$	100	1
(7, 0)		100	2
(7, 7)		100	2

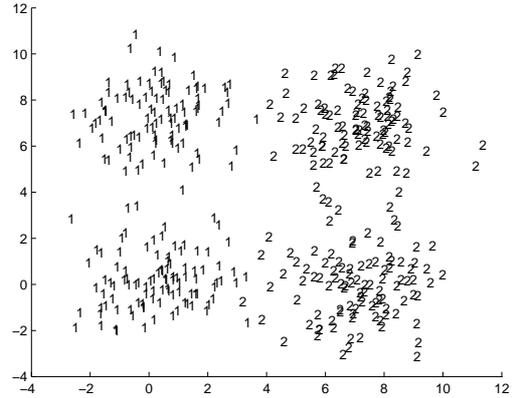


Figure 1 – Jeu de données *Toys*

En utilisant la distance Euclidienne avec $\alpha = 1$ et $\delta^2 = 100$, l'algorithme ECM trouve une frontière diagonale entre les classes. La direction de la diagonale (gauche ou droite) dépend de l'initialisation des centres. Le résultat est présenté sur la figure 2. Chaque point est affecté à l'élément focal de masse la plus élevée. Les centres sont représentés par les croix de grande taille. Après une transformation pignistique des fonctions de croyances obtenues par ECM, et après l'affectation des individus aux classes à

l'aide probabilités les plus fortes, nous obtenons un indice de Rand de 0.56.

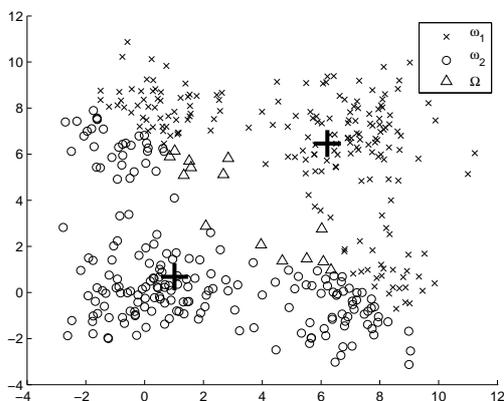


Figure 2 – Partition crédale obtenue par ECM

La distance Euclidienne permet d'obtenir uniquement des classes sphériques. Cette distance n'est pas adaptée à ce type de données. Si l'on choisit d'utiliser la distance de Mahalanobis en gardant le paramétrage précédent, on obtient le résultat présenté figure 3.

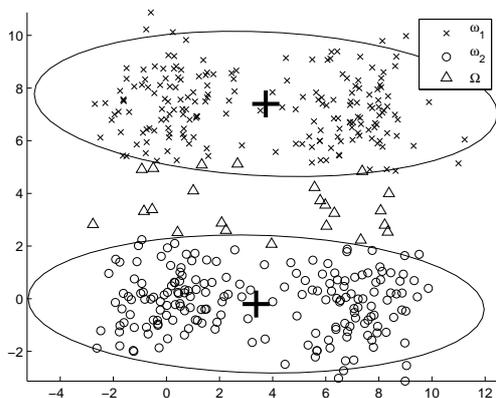


Figure 3 – Partition obtenue avec ECM en utilisant une distance de Mahalanobis

Selon l'initialisation des centres, les résultats peuvent être deux classes horizontales ou deux classes verticales, ce qui donne un indice de Rand soit de 0.5, soit de 1. L'ajout de contraintes permet de converger vers la solution

désirée. Par exemple, en utilisant seulement 10 contraintes, CECM trouve les classes désirées, comme le montre la figure 4.

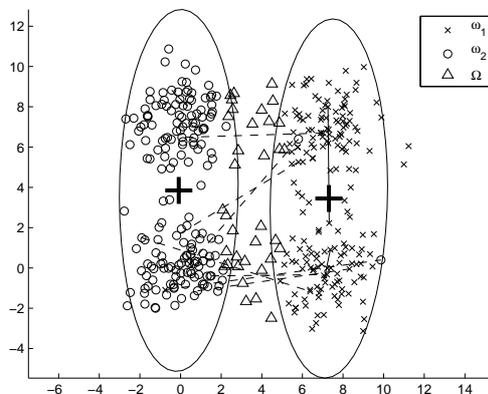


Figure 4 – CECM avec une distance de Mahalanobis et 10 contraintes

Notons pour la figure 4 qu'un segment continu entre deux points correspond à une contrainte *Must-link* entre deux individus et qu'un segment en pointillés a une contrainte *Cannot-link*.

4.3 Données Iris

La base de données Iris est une base couramment employée dans le cadre de la classification. Elle est composée de trois classes représentant différentes espèces d'iris : setosa, versicolor et virginica. Pour chaque espèce, 50 spécimens sont observés. Les caractéristiques mesurées sont la longueur et la largeur d'un sépale ainsi que la longueur et la largeur d'un pétale, en cm. Il faut noter qu'une seule des classes est linéairement séparable des autres classes. De plus les classes ont une distribution non sphérique, nous avons donc choisi d'utiliser la distance de Mahalanobis. Les expériences sont effectuées avec $\alpha = 1$ et $\delta^2 = 1000$.

Il faut savoir qu'il est possible d'obtenir avec une connaissance a priori de moins bons résultats que sans connaissance a priori. C'est un problème reconnu dans le cas de couples contraints [16]. Les résultats sont donc sensibles au paramétrage de γ et η . Nous supposons

que $\gamma = \eta$, c'est à dire que les contraintes *Must-link* et *Cannot-link* ont la même importance. Si γ et η sont trop élevés, alors le respect des contraintes prévaut sur la minimisation des distances intraclasses. La cohérence de la classification peut en être très affectée. A l'inverse si ces coefficients sont trop faibles les contraintes ne sont pas respectées : l'ajout de contraintes serait inutile. La figure 5 présente l'indice de Rand moyen obtenu sur 100 exécutions avec sélection aléatoire de contraintes. Pour permettre une comparaison de différentes valeurs des paramètres γ et η en fonction du nombre de contraintes, γ est divisé par le nombre de contraintes *Must-link* et η par le nombre de contraintes *Cannot-link*.

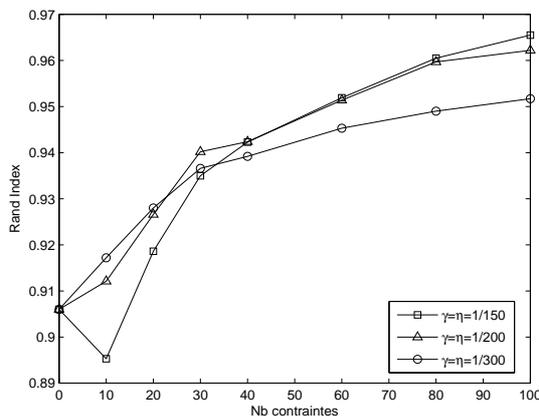


Figure 5 – CECM : indice de Rand moyen

Ainsi on peut voir que l'ajout progressif de contraintes améliore logiquement l'indice de Rand. Des valeurs de coefficients élevés ($\gamma = \eta = \frac{1}{150}$) entraînent une forte prise en compte des contraintes par l'algorithme. Ceci représente un inconvénient quand il existe peu de contraintes, car elles peuvent aisément être en contradiction avec la partition initiale. Cependant, si il existe un grand nombre de contraintes, peu de données non contraintes sont à classer et l'indice de Rand est élevé. Dans le cas où $\gamma = \eta = \frac{1}{300}$, une faible quantité de contraintes produit de bons résultats car les contraintes trop contradictoires avec la partition ne sont pas prises en compte. Cependant,

lorsque la quantité de contraintes augmente, leur importance relative diminue. Les valeurs $\gamma = \eta = \frac{1}{200}$ permettent un bon compromis entre l'attache aux contraintes et le respect des distances. Comme dans [4], les résultats sont d'autant meilleurs que le nombre de contraintes est fort.

Les figures 6 et 7 présentent deux résultats de l'algorithme : le premier est sans contraintes (l'indice de Rand est 0.90) et le second avec 30 contraintes (l'indice de Rand est 1.00). Chaque point est affecté à la classe de maximum de probabilité pignistique à postériori. Pour une raison de lisibilité, on ne représente pas les contraintes.

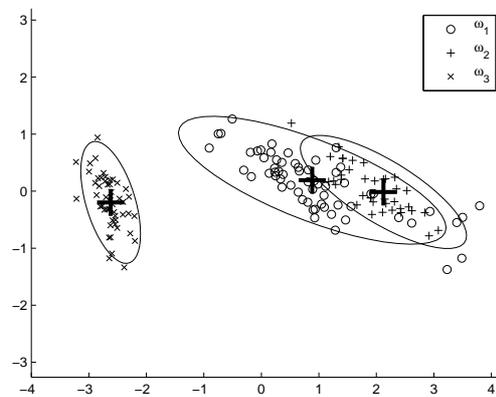


Figure 6 – ECM

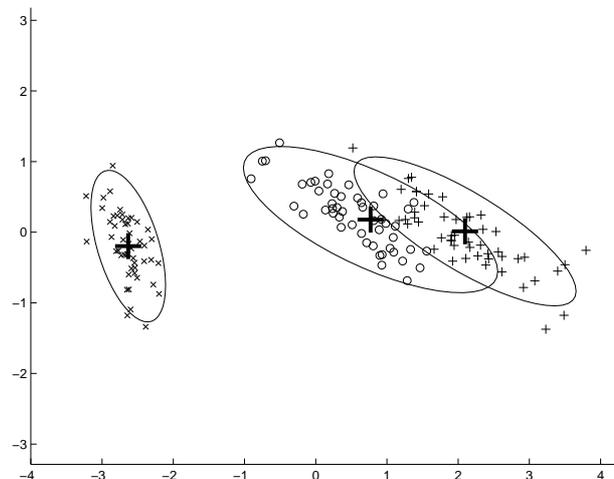


Figure 7 – CECM : 30 contraintes

Ces figures permettent de visualiser l'évolution des ellipses. Notons que la solution dépend des contraintes. Par ailleurs, la solution représentée à la figure 7 est parfaite.

5 Conclusion

Nous avons proposé dans cet article une nouvelle méthode de classification automatique intitulée CECM. Il s'agit d'une extension d'ECM, un algorithme de classification évidentiel. Notre contribution est double : nous avons proposé, d'une part, une manière d'introduire des contraintes *Must-link* et *Cannot-link*. Nous avons, d'autre part, introduit la métrique de Mahalanobis dans l'algorithme. Cette métrique, plus générale que la métrique Euclidienne, permet de traiter des classes non sphériques et de s'adapter automatiquement à l'ajout de contraintes.

Des expériences ont été réalisées pour évaluer l'impact des paramètres de l'algorithme sur ses performances. Il en résulte qu'un bon paramétrage dépend en partie du nombre de contraintes existantes. Une fois ce paramétrage trouvé, l'algorithme produit de bons résultats comparé à des algorithmes utilisant le même type de contraintes [4]. L'ajout de contraintes entraîne parfois une diminution des performances de la méthode [16]. En effet, une seule instance de l'algorithme est exécutée et cette instance peut être, dans le pire des cas, de moins bonne qualité qu'une instance sans contrainte. De plus, l'obtention d'excellents résultats exige un nombre de contraintes important, ce qui peut être difficile à obtenir. Une perspective est donc d'utiliser l'apprentissage actif [3, 16, 4], c'est à dire de choisir automatiquement, lors de l'exécution de l'algorithme, les paires de contraintes paraissant les plus utiles à la classification.

Références

- [1] K. Wagstaff, C. Cardie, S. Rogers, S. Schroedl, Constrained k-means clustering with background knowledge. *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, pp. 577-584.
- [2] J.C Bezdek. *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer Academic Publishers, 1999.
- [3] S. Basu, A. Banerjee, R.J. Mooney, Active semi-supervision for pairwise constrained clustering. *Proceedings of the SIAM International Conference on Data Mining, 2004*, pp. 333-344.
- [4] N. Grira, M. Crucianu, N. Boujemaa, Active semi-supervised fuzzy clustering. *Pattern Recognition, 2008*, pp. 1851-1861.
- [5] D. Klein, S.D. Kamvar, C.D. Manning, From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering. *Machine Learning - International Workshop - , 2002*, pp. 307-314.
- [6] R. Krishnapuram, J.M Keller, A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems, 1993*, pp. 98-110.
- [7] S. Sen, R.N. Davé, Clustering of relational data containing noise and outliers. *Fuzzy Systems Proceedings, 1998*, pp. 98-110.
- [8] T. Denœux, M.-H. Masson, EVCLUS : evidential clustering of proximity data. *IEEE Trans. Systems, Man and Cybernetics : B, 2004*, pp. 95-109.
- [9] M.-H. Masson, T. Denœux, ECM : Algorithme évidentiel des c-moyennes. *Actes des Rencontres Françaises sur la Logique Floue et ses Applications, 2007*, pp. 17-24.
- [10] M.-H. Masson, T. Denœux, ECM : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition, 2008*, pp. 1384-1397.
- [11] G. Shafer. *A mathematical theory of evidence*. Princeton university press, Princeton, NJ, 1976.
- [12] P. Smets, The transferable belief model for quantified belief representation. *Handbook of Defeasible Reasoning and Uncertainty Management Systems, 1998*, pp. 267-301.
- [13] P. Smets, R. Kennes, The transferable belief model. *Artificial Intelligence, 1994*, pp. 191-234.
- [14] R. R. Yager, On the normalization of fuzzy belief structures. *International Journal of Approximate Reasoning, 1996*, pp. 127-153.
- [15] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix. *IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, 1978*, pp. 761-765.
- [16] I. Davidson, K. Wagstaff, S. Basu, Measuring constraint-set utility for partitional clustering algorithms. *Lecture Notes in Computer Science, 2006*, pp. 115-126.