

Extraction de concepts descriptifs avec exceptions par classification non supervisée hybride

Descriptive concept extraction with exceptions by hybrid clustering

Marie-Jeanne Lesot
Bernadette Bouchon-Meunier
Laboratoire d'Informatique de Paris VI, Université Pierre et Marie Curie
8 rue du capitaine Scott, 75 015 Paris, France
{Marie-Jeanne.Lesot, Bernadette.Bouchon-Meunier}@lip6.fr

Résumé :

La modélisation de concepts naturels procède le plus souvent par interrogation d'experts, pour établir un lien entre les descriptions numériques disponibles et les descriptions linguistiques utilisées naturellement. Nous proposons ici d'exploiter directement des ensembles de données non étiquetés, et d'extraire les concepts qui permettent une description sémantique de ces données. La méthode proposée consiste d'abord à identifier les sous-groupes correspondant aux concepts, puis à les représenter sous la forme de sous-ensembles flous. L'identification doit conserver les exceptions présentes dans la base, qui peuvent être associées à des concepts minoritaires mais caractérisent l'ensemble de données ; elle est basée sur la combinaison, dans une procédure itérative, d'un algorithme de clustering hiérarchique et des c -moyennes floues. La description est faite par des fonctions d'appartenance de type gaussiennes généralisées. L'adéquation des résultats de la méthode avec des classifications spontanées est illustrée sur des ensembles de données artificiels et réels.

Mots-Clés :

Extraction de concepts, Traitement des exceptions, c -moyennes floues, Classification ascendante hiérarchique.

Abstract:

Natural concepts modelling aims at numerically representing semantic knowledge ; generally, experts are asked to provide examples of linguistic terms associated with numerical data descriptions. We propose to exploit directly non labelled databases to extract the concepts that enable a semantic description of the data. The proposed method consists in identifying the subgroups which correspond to the concepts and then representing them as fuzzy subsets. The identification must preserve the outliers, which are associated with minor concepts ; it is based on a conjugate iterative use of the hierarchical clustering algorithm and the fuzzy c -means algorithm. The description builds membership functions as generalized gaussians. The adequacy of the results with spontaneous descriptions is illustrated on artificial and real databases.

Keywords:

Concept extraction, outlier handling, fuzzy c -means, agglomerative hierarchical clustering.

1 Introduction

La modélisation de concepts naturels vise à établir une relation entre des données décrites numériquement et des connaissances d'ordre sémantique, exprimées par des termes linguistiques tels que « grand », « jeune » ou « chaud » ; généralement, cette modélisation est effectuée par le biais de sous-ensembles flous. Pour les construire, on utilise différents protocoles d'interrogation d'experts [1] ou des techniques d'inférence de partitions floues [6, 7, 14], qui recherchent des subdivisions du domaine numérique corrélées avec les étiquettes connues des données.

On peut considérer une tâche analogue dans un contexte non supervisé, en se basant uniquement sur des ensembles de valeurs numériques, sans disposer des étiquettes linguistiques qui leur sont associées : la distribution des données elle-même contient des informations sur les concepts qui sont nécessaires à une description sémantique. On peut par exemple chercher à apprendre si des données nécessitent la définition des concepts « faible » et « élevé », ou si elles font aussi appel à un concept « intermédiaire », et déterminer leurs caractérisations sous forme de sous-ensembles flous. Le but de la tâche d'extraction de concepts descriptifs est donc de construire les fonctions d'appartenance qui correspondent à un profil simplifié et fidèle de l'ensemble des données.

Cet objectif de description impose des contraintes sur le traitement des exceptions,

c'est-à-dire des données nettement différentes du reste de la base : le plus souvent, elles sont vues comme des aberrations dues au bruit, et ne sont pas modélisées. A l'inverse, dans un contexte descriptif, elles sont considérées comme caractéristiques de l'ensemble traité, et représentent des concepts minoritaires, mais significatifs, associés à des termes linguistiques du type « anormalement faible ». Elles doivent donc être traitées comme les groupes de données plus importants.

Nous proposons une méthode qui extrait les concepts associés à une description sémantique des données, indépendamment de la taille des groupes qui les représentent. Elle est composée de deux étapes : la première consiste à identifier les données associées à un même concept, et utilise un algorithme combinant les c -moyennes floues avec un algorithme de clustering hiérarchique ; la seconde décrit les concepts par des fonctions d'appartenance définies à partir d'une famille paramétrique de fonctions.

La partie 2 présente le traitement des exceptions effectué par les méthodes de détection d'exceptions et de clustering. La partie 3 présente le mode d'identification des concepts proposé et la partie 4 la méthode de construction des sous-ensembles flous représentatifs associés. La partie 5 illustre les résultats obtenus sur des bases d'apprentissage artificielles et réelles.

2 Traitement des exceptions

2.1 Détection d'exceptions

Certaines méthodes de fouille de données ont pour but de mettre en évidence les exceptions qui diffèrent nettement de la base considérée. La plupart se placent dans un cadre statistique : elles se basent sur une hypothèse de loi de distribution et définissent des tests de déviation, qui dépendent de la loi considérée, du nombre d'exceptions attendues et de leur type [2].

Knorr *et al.* [13] définissent comme exceptions

les données x telles qu'une proportion fixée des données est à une distance de x supérieure à un seuil fixé ; ce paramètre global ne prend pas en compte la répartition locale des données, ce qui n'est pas satisfaisant. Breunig *et al.* [3] définissent un degré relatif d'isolement, basé sur la densité locale des données, pour indiquer à quel point une donnée est une exception.

Ces techniques permettent de mettre en évidence des exceptions, mais elles ne cherchent pas la présence éventuelle de groupes de données homogènes identifiables, elles ne traitent donc que l'un des aspects de l'extraction de concepts descriptifs.

2.2 Clustering par partitionnement

La classification non supervisée [11], ou *clustering*, vise à extraire d'un ensemble de données des sous-groupes pertinents, c'est-à-dire à la fois homogènes et distincts. Nous présentons ici le traitement des exceptions qui les différencie des concepts permettant une description sémantique.

L'algorithme des c -moyennes floues (*fuzzy c-means*, ou *fcm*) est proche de l'algorithme des k -moyennes mais présente l'avantage d'une sensibilité à l'initialisation plus faible. Il fournit une partition de l'ensemble des données en sous-ensembles flous en optimisant la fonction

$$J = \sum_{i=1}^n \sum_{r=1}^c u_{ri}^m \|x_i - w_r\|^2 \quad (1)$$

$$\text{sous la contrainte : } \forall i \quad \sum_{r=1}^c u_{ri} = 1 \quad (2)$$

où $X = \{x_i, i = 1..n\}$ désigne l'ensemble des données, c est le nombre de clusters cherchés, fixé *a priori* ; w_r est le centre, ou prototype, du cluster r ; u_{ri} désigne le degré d'appartenance de la donnée x_i au cluster r et m est un hyperparamètre, généralement fixé à 2. L'influence des exceptions est illustrée sur la figure 1 : un observateur humain détecte un profil constitué de trois groupes, interprétés par exemple comme « élevé », « faible » et « anormalement faible », l'adverbe « anormalement »

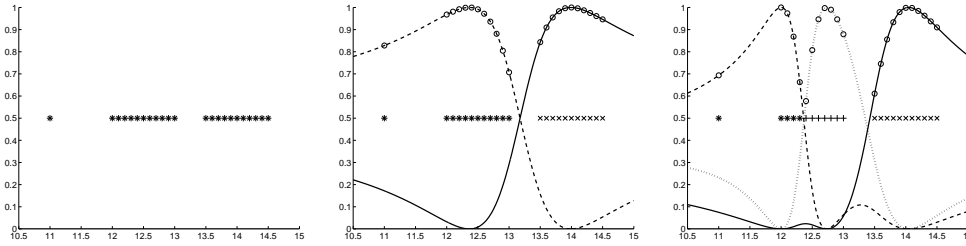


Figure 1 – Algorithme des *fcm* : à gauche, les données ; au centre, $c = 2$; à droite, $c = 3$.

rendant compte de la spécificité de la donnée la plus à gauche. La solution à deux clusters est la plus satisfaisante, car elle correspond à une partition intuitive, bien qu'elle n'isole pas l'exception, et efface sa particularité en l'affectant avec un degré élevé au cluster de gauche. Le regroupement obtenu avec trois clusters qui seul peut mettre l'exception en évidence, est particulièrement contre-intuitif, en raison de la scission du cluster central qu'il produit, et que l'on observe quelle que soit l'initialisation.

Ce comportement est dû à deux particularités des *fcm* [10]. D'une part, elles ne prennent pas directement en compte la séparabilité des groupes : le coût (1) pénalise le manque de compacité (*i.e.* une forte variance) des groupes constitués, et non la ressemblance entre données de groupes différents. De plus, les *fcm* tendent à produire des groupes de même taille, qui équilibrent les influences des données sur les positions des prototypes. Aussi, elles ont des difficultés à représenter des groupes de petites tailles, et donc les exceptions. Nous mentionnons ci-dessous quelques variantes des *fcm* proposées pour contrer ces effets.

Certaines variantes visent à réduire l'influence des points d'appartenance faible aux clusters, afin qu'ils n'interviennent pas dans la définition des prototypes correspondants : elles autorisent des degrés d'appartenance à valeurs dans $\{0, 1\}$ pour des données distinctes des centres des clusters [16], pénalisent les degrés proches de 0.5 [9] ou utilisent une métrique plus robuste que la métrique euclidienne [17]. On pourrait attendre de ces approches qu'elles permettent de stabiliser un prototype sur des données

isolées, sans qu'il soit attiré vers des clusters de taille supérieure. Cependant, l'effet obtenu est que les prototypes ne sont pas attirés vers les données nettement différentes, et représentent mieux les groupes de taille moyenne. Ces méthodes répondent donc à un objectif de robustesse [5], mais n'isolent pas les exceptions.

Les *c*-moyennes possibilistes [12] (*possibilistic c-means*, ou *pcm*) diffèrent des *fcm* par l'interprétation des degrés u_{ri} (cf. éq. (1)) : les *pcm* n'imposent pas la contrainte de normalisation (2) et considèrent u_{ri} comme des degrés de typicalité : une donnée excentrée a alors un faible degré de typicalité pour tous les groupes construits. Un post-traitement pourrait permettre de mettre en évidence les exceptions ; toutefois, les *pcm* sont instables par rapport à l'initialisation et sensibles aux valeurs des hyperparamètres ; enfin, elles peuvent conduire à des clusters confondus qui ne reflètent pas la structure de l'ensemble de données [10].

L'algorithme *noise clustering* [4] traite simultanément les exceptions et les groupes plus importants : un « cluster-bruit » est ajouté, pour regrouper les données mal représentées par les clusters existants. Il n'est pas caractérisé par un prototype, mais directement par la distance aux données, δ , qui est la même pour toutes les données, et qui est définie comme

$$\delta^2 = \lambda \frac{1}{cn} \sum_{i=1}^n \sum_{r=1}^c \|x_i - w_r\|^2 \quad (3)$$

Il est ensuite inclus dans le coût J (éq. (1)) comme un cluster classique [4]. λ est un hyperparamètre qui contrôle la distance d'isolement à partir de laquelle une donnée est considérée

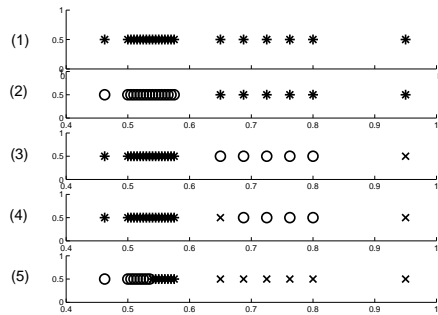


Figure 2 – Algorithme *noise clustering* : (1) Données ; (2) $\lambda = 1$; (3) $\lambda = 0.5$; (4 et 5) $\lambda = 0.05$ avec des initialisations différentes. Chaque symbole correspond à un cluster différent, la croix \times est associée au « cluster-bruit ».

comme exception : plus il est élevé, plus la proportion d’exceptions est faible (cf. fig. 2). Cette méthode pose cependant deux problèmes : certaines valeurs de λ conduisent à une sensibilité à l’initialisation, comme le montrent les graphes 4 et 5 (fig. 2), voire à des partitions contre-intuitives (graphe 5). De plus, la méthode utilise une définition globale des exceptions, basée sur l’unique paramètre δ , et ne considère pas la densité locale : il paraît naturel d’identifier sur la figure 2 deux groupes principaux et deux exceptions, mais la donnée la plus à gauche apparaît toujours comme trop proche du centre du cluster de gauche par rapport à la distance moyenne aux centres, pour être isolée comme exception.

2.3 Clustering hiérarchique

Les algorithmes hiérarchiques construisent une hiérarchie de regroupements imbriqués différant par le nombre de clusters [11]. Les algorithmes agglomératifs (*classification ascendante hiérarchique*, ou CAH) partent d’une configuration où chaque donnée est un cluster, et opèrent par fusions successives. Le critère de regroupement consiste par exemple à fusionner les groupes les plus proches, la distance étant du type $d(A, B) = op(\{d(a, b), a \in A, b \in B\})$: *op* est un opérateur comme *min* (stratégie de chaînage simple), *max* (chaînage complet), ou encore la moyenne ; le critère de Ward fusionne les groupes qui minimisent l’augmentation de

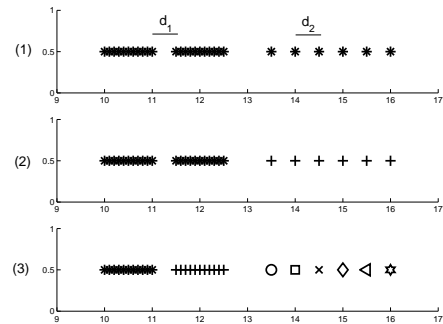


Figure 3 – Algorithme hiérarchique avec chaînage simple : (1) Ensemble de données, $d_1 = d_2 = 0.5$; (2) $s = 0.9$; (3) $s = 0.4$.

variance intraclasse. Chaque critère impose des contraintes différentes : le chaînage complet minimise le diamètre des clusters et mène à des groupes compacts, le critère de Ward construit des groupes homogènes, de variance faible, le chaînage simple, basé sur la distance minimale entre données affectées à des clusters différents, est sensible à la séparabilité des clusters. Il permet donc de détecter de petits groupes de données isolés, ce qui en fait un candidat pour l’extraction de concepts descriptifs.

Toutefois, il a tendance à créer des clusters allongés, en particulier lorsque des groupes naturels sont reliés par une chaîne de données peu distantes les unes des autres : c’est le problème du chaînage [11] (*chaining effect*). De plus, il n’est pas adapté aux données de densité variable : l’arrêt du processus de regroupement est contrôlé par un seuil s , tel que seules les fusions de groupes distants de moins de s ont lieu. Cette définition ne prend pas en compte la densité locale des données, comme l’illustre la figure 3 où il semble naturel de distinguer trois groupes. Cependant, les distances d_1 et d_2 sont égales et sont donc traitées de manière identique : si $s > d_1$ (graphe 2), les regroupements correspondants sont effectués, pour former deux clusters, sinon, on obtient huit groupes (graphe 3). Notons que d’autres critères (chaînage complet ou critère de Ward) gèrent ce cas de manière satisfaisante, mais ils n’ont pas le comportement désiré pour l’isolement des exceptions.

3 Extraction de concepts

L'extraction de concepts descriptifs constitue une tâche de regroupement avec conservation de l'information portant sur les exceptions, celles-ci étant définies à une échelle locale. Les algorithmes de clustering ne permettent pas de répondre à ce double objectif ; on pourrait effectuer successivement une étape de détection d'exceptions puis une étape de clustering. Cependant, étant donné les propriétés différentes des algorithmes, il apparaît possible d'atteindre les deux buts simultanément, en couplant des méthodes de classification : nous proposons de combiner dans une procédure itérative l'algorithme hiérarchique avec chaînage simple et les *fc*m, pour exploiter leurs capacités à construire des clusters respectivement bien séparés, pour CAH, et compacts, pour les *fc*m. La procédure itérative permet de définir plusieurs seuils de fusion pour CAH, et donc de s'adapter à la densité locale des données. De plus, elle permet d'utiliser un critère d'arrêt qui ne nécessite pas de fixer *a priori* le nombre de clusters recherchés.

3.1 Couplage itératif

La procédure itérative que nous proposons considère un groupe G et envisage une subdivision suivant l'algorithme le plus adapté à ses caractéristiques : si G contient des sous-groupes clairement séparés, il est soumis à CAH. S'il est peu séparable, mais peu compact, il correspond à un cas de chaînage, et on lui applique les *fc*m. Puis on réitère le processus sur les sous-groupes obtenus. Les critères de choix évaluent le gain apporté par chacune des deux méthodes.

Le critère de gain de CAH est défini comme $C_{CAH} = D_G/\delta_G$ où D_G est la distance maximale de fusion et δ_G la distance minimale entre données distinctes¹, mesurées dans G . Il évalue la présence de ruptures dans la distribution des données et compare G à un groupe de données uniformément réparties, référence de groupe à

¹ou le maximum entre cette valeur et une distance significative minimale δ définie par l'utilisateur.

ne pas subdiviser, pour lequel $C_{CAH} = 1$. δ_G donne une échelle de valeurs qui détermine si D_G est localement significative, elle permet donc de prendre en compte la densité locale des données.

Le critère de gain des *fc*m évalue la qualité de la subdivision optimale par *fc*m (pour le choix de c , voir partie 3.3). Il est défini comme

$$C_{fcm}(c) = \frac{\frac{1}{c} \sum_{r=1}^c \sigma(C_r)}{\frac{1}{c} \sigma(G)} \quad (4)$$

où $\sigma(C)$ désigne l'écart-type d'un ensemble de données C et C_r , $r = 1..c$, les sous-groupes construits par les *fc*m. Ce critère intervient dans la mesure proposée par Rezaee [15] pour sélectionner le nombre optimal de clusters. La condition $C_{fcm} < 1$ considère que la subdivision est justifiée si la moyenne des écarts-type est inférieure à l'écart-type obtenu par une subdivision en sous-groupes réguliers.

Nous appliquons un critère heuristique supplémentaire, C_{diam} , pour le choix d'utilisation des *fc*m, qui permet d'incorporer des connaissances *a priori* : les *fc*m ne sont envisagées que si le diamètre de G est supérieur à un seuil fourni par l'utilisateur (ou défini par défaut en fonction du diamètre de l'ensemble des données). Ce seuil indique une taille en dessous de laquelle les groupes construits, s'ils ne sont pas justifiés par une séparabilité importante mise en évidence par C_{CAH} , risquent de ne plus être sémantiquement interprétables.

3.2 Choix du seuil de regroupement

Le clustering hiérarchique nécessite le réglage d'un seuil s , tel que les fusions entre groupes ont lieu si les distances correspondantes sont inférieures à s . Nous le définissons en fonction de la distribution du groupe considéré pour permettre une adaptation locale :

$$s = \bar{d} + \alpha\sigma(d) \quad (5)$$

d est le vecteur des distances de fusions envisagées, \bar{d} et $\sigma(d)$ sa moyenne et son écart-type ;

Tableau 1 – Algorithme d'extraction de concepts (voir les détails des critères partie 3).

Initialisation

$G = X$

- choisir la proportion d'exception α
- le diamètre seuil
- la distance minimale significative

Algorithme

- si G est séparable au sens de C_{CAH}
 - calculer le seuil s^* défini par l'éq. (5)
 - subdiviser G par CAH avec le seuil s^*
- sinon, si G est peu compact au sens de C_{diam}
 - calculer le nombre de clusters c^* par C_{stab}
 - si la subdivision est justifiée par $C_{fcm}(c^*)$
 - subdiviser G par fcm avec la valeur c^*
- itérer sur les sous-groupes obtenus

α est un paramètre qui permet de déterminer le taux de regroupement souhaité. Ainsi, en supposant que d suit une distribution gaussienne, $\alpha = 2$ conduit à effectuer 95,5% des regroupements envisagés. Ce paramètre correspond donc grossièrement à la proportion d'exceptions attendues, la notion d'exception étant définie en fonction de la répartition locale des données.

3.3 Choix de c

Il existe de nombreux critères pour le choix du nombre de clusters c ; nous le déterminons en utilisant la propriété de stabilité des fcm : lorsque c est inférieur ou égal au nombre de clusters naturels présents dans les données, les fcm convergent vers la même configuration associée à une même valeur du coût J (eq. (1)), indépendamment de l'initialisation aléatoire; lorsque c est supérieur, les fcm exploitent les degrés de liberté non contraints par les données et on observe une instabilité de la position des centres et de la valeur de coût. Pour un c fixé, la stabilité est mesurée par $C_{stab} = \sigma(J)/\bar{J}$, quotient entre l'écart-type de l'erreur et sa moyenne, lorsqu'on varie l'initialisation; la valeur de c choisie est la plus grande valeur permettant un comportement stable.

4 Représentation de concepts

L'algorithme de clustering précédent, résumé dans le tableau 1, identifie les concepts en déterminant les données qui leur sont associées, quelle que soit leur dimension. L'étape suivante consiste à construire les sous-ensembles flous qui les décrivent; nous nous plaçons dans le cas classique d'une description attribut par attribut.

Dans le cadre de description associé à l'extraction de concepts, les fonctions d'appartenance construites doivent représenter les groupes identifiés, alors que dans les cas d'inférence de partitions floues, elles doivent couvrir la totalité du domaine des données: les fcm conduisent à des fonctions « très larges », et non concentrées sur les données observées (cf. fig. 1); de même, la méthode de Gyenesei [8] construit les représentations uniquement en fonction des positions des centres et d'un taux de recouvrement fixé *a priori*, sans considérer les groupes eux-mêmes. Pour la description, les fonctions d'appartenance doivent dépendre directement des données constituant les clusters.

Nous les construisons en déterminant les paramètres de la famille des gaussiennes généralisées, proposée dans [10], d'autres fonctions pourraient être envisagées de même. Elle est définie comme

$$\mu(x) = \exp \left[- \left(\frac{\|x - \bar{x}(G)\|}{a} \right)^b \right] \quad (6)$$

et correspond à une gaussienne si $b = 2$. $\bar{x}(G)$ désigne la moyenne du groupe considéré: ayant été obtenu par clustering, G peut être considéré comme suffisamment compact pour que $\bar{x}(G)$ en constitue un représentant pertinent. Le paramètre a influence principalement la largeur du plateau des valeurs élevées, et b la vitesse de décroissance de la fonction; ces deux paramètres conjoints contrôlent le noyau et le support de la fonction d'appartenance construite. Il faut noter qu'à la différence des fonctions trapézoïdales, le noyau est réduit à un point, le centre du groupe; toutefois, la caractérisation des clusters est plus souple.

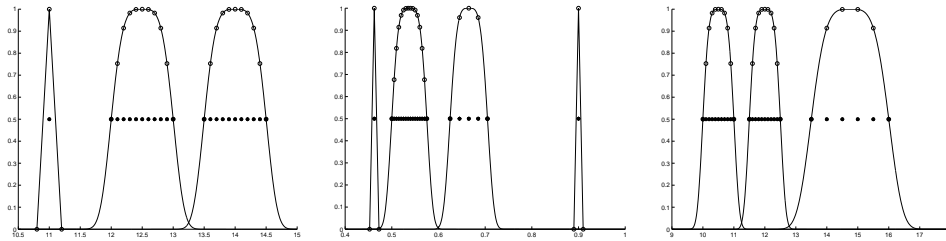


Figure 4 – Profils des bases représentées sur les figures 1, 2 et 3, obtenus avec la méthode hybride.

Une représentation probabiliste du groupe, associée à $a = \sigma(G)$ et $b = 2$ ne constitue pas une description satisfaisante car elle affecte aux données les plus distantes du centre des degrés trop faibles. Des tests montrent que $b = 4$ conduit à des fonctions de forme convenable ; la valeur de a est fixée en fonction des caractéristiques du groupe, de telle sorte que les données les plus distantes aient un degré d'appartenance égal à 0.5. Ainsi tous les points affectés au groupe vérifient $\mu \geq 0.5$.

Cette définition a l'inconvénient d'imposer une fonction d'appartenance symétrique ; pour plus de souplesse, nous décrivons le groupe par deux fonctions de cette famille, ayant même valeur et même dérivée en $x = \bar{x}(G)$, et représentant respectivement les données supérieures et inférieures au centre, en définissant deux valeurs a^+ et a^- .

Dans le cas de groupes réduits à une valeur, les choix précédents ne peuvent s'appliquer. Aussi ceux-ci sont décrits par une fonction d'appartenance triangulaire, dont le support est défini en fonction de la distance minimale significative.

5 Résultats

Nous avons appliqué la méthode proposée à des données unidimensionnelles, artificielles et réelles ; les figures représentent les fonctions d'appartenance obtenues, les données étant représentées par le symbole *, décalé en ordonnée en cas de multiples occurrences.

La figure 4 montre les résultats obtenus avec les bases des figures 1 à 3 : les clusters construits correspondent à des regroupement in-

tuitifs qui peuvent être associés à des significations sémantiques (comme « élevé », « faible » et « anormalement faible » pour le graphe de gauche). Les graphes au centre et à droite montrent que la définition des exceptions prend en compte une information locale. Dans les trois cas, les ensembles flous construits constituent des profils représentatifs des bases initiales.

La figure 5 représente les profils obtenus à partir des évaluations réelles de deux promotions d'étudiants à une même épreuve. On peut interpréter ces profils, et constater que leur structure principale est similaire, même si le groupe de droite est moins homogène. Il contient deux étudiants en difficulté et présente une tête de classe plus importante. L'extraction de concepts permet donc d'obtenir des connaissances d'ordre sémantique sur le niveau des classes.

Nous avons appliqué la méthode proposée à des données provenant du site web² de l'Institut National de la Statistique et des Études Économiques (INSEE) concernant la population des départements métropolitains français. Le profil obtenu (fig. 6) est plus complexe que les profils précédents ; il met en évidence l'existence de 4 départements « surpeuplés » (correspondant au Rhône, Bouches du Rhône, Paris et Nord), un groupe important « fortement peuplé », 5 catégories intermédiaires, un groupe important « peu peuplé » et enfin une exception « très peu peuplée » (correspondant à la Lozère). Comme précédemment, elle permet d'extraire une information sémantique et d'interpréter les valeurs numériques observées.

²<http://www.insee.fr>

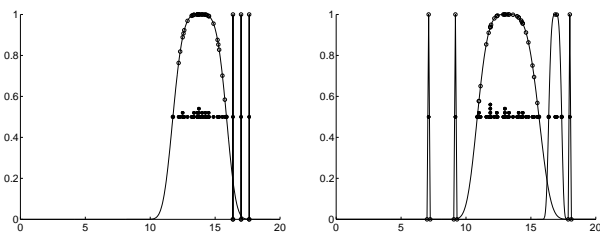


Figure 5 – Profils de deux promotions d'étudiants décrits par leur note à une même épreuve.

6 Conclusion

L'extraction automatique de concepts descriptifs permet de définir un profil décrivant des données numériques ; elle constitue un processus d'abstraction qui fournit une caractérisation sémantique des données. Si un expert associe une étiquette linguistique aux concepts identifiés, la méthode permet d'acquérir une représentation des catégories mentales, sans que l'expert n'ait à étiqueter chaque donnée.

La méthode proposée est basée sur la combinaison dans une procédure itérative des c -moyennes floues avec un algorithme de classification hiérarchique. Elle permet de traiter simultanément et de manière identique les concepts classiques et les exceptions, et d'utiliser une définition locale des exceptions.

Au-delà de la constitution des groupes, la méthode itérative fournit une information plus riche, par le biais d'un historique de constitution des groupes : les perspectives de ce travail incluent l'exploitation de cet historique, dans le but de déterminer à quel point une donnée est une exception et de définir des degrés d'« exceptionnalité ».

Références

- [1] N. Aladenise, B. Bouchon-Meunier. Acquisition de connaissances imparfaites : mise en évidence d'une fonction d'appartenance. *Revue Internationale de systémique*, 11(1) : 109-127, 1997.
- [2] V. Barnett, T. Lewis. *Outliers in statistical data*. Wiley and Sons, 1994.
- [3] M. Breunig, H. Kriegel, R. Ng, J. Sander. LOF : Identifying density-based local outliers. *Proc. of*

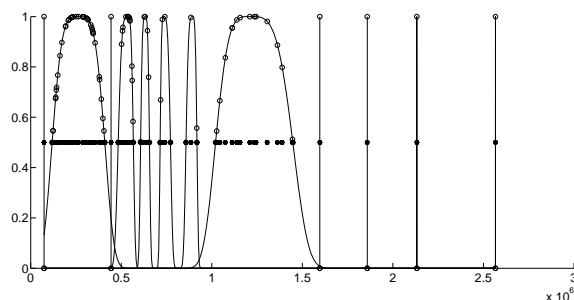


Figure 6 – Profil des départements français caractérisés par leur population en 2001.

ACM SIGMOD, 2000, pp. 93-104.

- [4] R. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12 : 657-664, 1991.
- [5] R. Davé, R. Krishnapuram. Robust clustering methods : a unified view. *IEEE Trans. on fuzzy systems*, 5(2) : 270-293, 1997.
- [6] T. Van de Merckt. Decision trees in numerical attribute spaces. *Proc. of the 13th Int. Joint Conf. on Artificial Intelligence*, 1993, pp. 1016-1021.
- [7] J. Dougherty, R. Kohavi, M. Sahami. Supervised and unsupervised discretization of continuous features. *Proc. of ICML*, 1995, pp. 194-202.
- [8] A. Gyenesei. Fuzzy Partitioning of Quantitative Attribute Domains by a Cluster Goodness Index. Technical Report 368, Turku Centre for Computer Science, 2000.
- [9] F. Höppner, F. Klawonn. A New Approach to Fuzzy Partitioning. *Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, 2001, pp. 1419-1424.
- [10] F. Höppner, F. Klawonn, R. Kruse, T. Runkler. *Fuzzy Cluster Analysis, Methods for classification, data analysis and image recognition*. Wiley, 2000.
- [11] A. Jain, M. Murty, P. Flynn. Data clustering : a review. *ACM Computing survey*. 31(3) : 264-323, 1999.
- [12] R. Krishnapuram, J. Keller. A possibilistic approach to clustering. *IEEE Transactions on fuzzy systems*, 1 : 98-110, 1993.
- [13] E. Knorr, R. Ng, V. Tucakov. Distance Based outliers : algorithms and applications. *Vldb Journal*, 8(3-4) : 237-253, 2000.
- [14] C. Marsala, B. Bouchon-Meunier. Fuzzy partitioning using mathematical morphology in a learning scheme. *Proc. of FUZZ'IEEE*, 1996, pp. 1512-1517.
- [15] M. Rezaee, B. Lelieveldt, J. Reiber. A new cluster validity index for the fuzzy c -means. *Pattern Recognition Letters*, 19 : 237-246, 1998.
- [16] P. Rousseeuw, E. Trauwaert, L. Kaufman. Fuzzy clustering with high contrast. *Journal of Computational and Applied Mathematics*, 64 : 81-90, 1995.
- [17] K. Wu, M. Yang. Alternating c -means clustering algorithms. *Pattern recognition*, 35 : 267-2278, 2002.