La Revue de Modulad

CLASSIFICATION FLOUE NON HIERARCHIQUE : LE PROGRAMME FUZZY

Jean-Luc DUPOUEY

Centre de Recherches forestières
I.N.R.A.
54280 CHAMPENOUX

Résumé : Le programme FUZZY effectue une classification floue non hiérarchique en un nombre de classes fixé *a priori*. Les individus ne sont plus affectés à l'une ou l'autre de ces classes, mais on leur attribue une valeur d'appartenance continue entre 0 et 1 à chacune des classes. L'algorithme est basé sur l'optimisation itérative d'un critère de variance intra-classe, utilisant la distance euclidienne. Un indice de validité de la classification est fourni. L'opportunité de ce type original de classification est discutée.

Mots-Clés : classification floue, classification valuée, ensemble flou.

1. Introduction : les ensembles flous.

Dès 1965, ZADEH introduit en mathématiques la notion de "fuzzy sets"[1] KAUFFMANN la développe en Europe (1977). Celle-ci repose sur le fait que : "Most real-world classes are fuzzy in nature - in the sense that the transition from membership to nonmembership in such classes is gradual rather than abrupt. Thus, given an object x and a class F, the real question in most cases is not wether x is or is not a member of F, but the degree to which x belongs to F"[2] (ZADEH, 1977)

Dans la théorie classique des ensembles, la fonction d'appartenance d'un ensemble E prend ses valeurs dans la paire {0,1} : un individu appartient ou non à l'ensemble. Les ensembles flous sont par contre définis par une fonction d'appartenance u telle que :

$$u : E \rightarrow [0,1]$$
$$u : x \mapsto u(x)$$

qui à chaque élément x associe un **degré d'appartenance** u(x) compris entre 0 et 1. Pour les valeurs 0 et 1, on est ramené à l'appartenance et à la non appartenance classiques.

On peut ainsi distinguer différents types d'individus selon leurs valeurs d'appartenance (fig 1) :

- les individus "centraux" (a) qui ont une forte valeur d'appartenance à un sous-ensemble donné et qui, placés au centre de ce sous-ensemble, en sont des représentants typiques ;

- les individus "périphériques" (b) qui ont une valeur moyenne d'appartenance à une classe donnée car situés sur ses bords ;

- les individus "hybrides" (c), intermédiaires entre plusieurs classes ;

- les individus "étrangers" enfin (d), de faible valeur d'appartenance.

De nombreux objets d'étude de l'analyse de données sont des ensembles flous : les catégories socio-professionnelles, les horizons d'un sol, les pathologies humaines... De nombreux champs d'applications utilisent déjà ce concept et les techniques afférentes. On peut citer, de façon non exhaustive, l'informatique, l'économie, les sciences de la terre, la médecine, l'écologie (DUPOUEY, 1985)... On en trouvera une synthèse partielle dans DUBOIS & PRADE (1983). En informatique, les bases de données et l'intelligence artificielle (reconnaissance des formes, systèmes experts), sont des domaines privilégiés d'applications.

---

1 La traduction littérale en est : "ensembles flous". Les travaux francophones dans ce domaine ont donné lieu à des terminologies diverses : "sous-ensembles flous" (KAUFFMANN, 1977) ou "partitions valuées" (LIBERT & ROUBENS, 1982).

2 "La plupart des classes du monde réel sont floues par nature dans le sens où la transition de l'appartenance à la non-appartenance est plutôt graduelle que brusque. Ainsi, étant donné un objet x et une classe F, la vraie question dans la plupart des cas n'est pas de savoir si x est ou n'est pas élément de F, mais le degré d'appartenance de x à F".

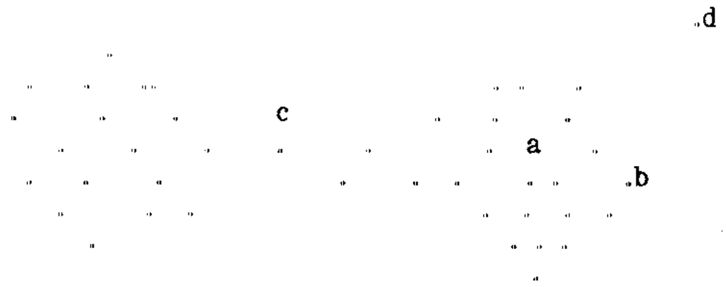

Figure 1 : types de points différenciés en classification floue.

2. Application de la théorie des ensembles flous en classification automatique.

Dès 1969, RUSPINI comprend l'intérêt des ensembles flous pour la classification automatique. Depuis, diverses méthodes métriques et non métriques ont été élaborées. Nous utilisons ici celle proposée par BEZDEK (1974) et DUNN (1974).

Une **partition floue** de E en K sous-ensembles est définie par la donnée de K fonctions u_k telles que :

$$\forall x \in E \qquad \begin{array}{l} u_k : E \to [0,1] \\ \sum_k u_k(x) = 1 \end{array}$$

Le but de la classification floue est donc de construire une telle partition floue, K étant fixé *a priori*.

2.1. L'algorithme de classification floue.

Celui-ci est basé sur l'optimisation itérative d'un critère de variabilité intra-classe :

$$\min_{u_k(x)} \left[\sum_k VAR(k) \right]$$

où $\quad VAR(k) = \sum_x [u_k(x)]^r . d[x, CG(k)]$

$CG(k) = \sum_x [u_k(x)]^r x \,/\, \sum_x [u_k(x)]^r$ (centre de gravité de la classe k)

d étant un indice de dissimilarité sur E et r R, $r \geq 1$

avec les contraintes :

$$\forall x \in E, \ 0 \leq u_k(x) \leq 1 \quad \text{et} \quad \sum_k u_k(x) = 1$$

La solution du programme non linéaire associé à ce critère est donnée par :

$$u_k(x) = d[x,CG(k)]^{1/(1-r)} / \sum_i d[x,CG(i)]^{1/(1-r)}$$

en supposant connus les CG(i), pour i=1,...,K.

L'élévation à la puissance r des poids que représentent les valeurs d'appartenance a pour conséquence d'amplifier les différences entre points éloignés et points centraux. Concrètement, on passe progressivement d'une classification ordinaire (r=1) à une classification où tous les individus sont équirépartis entre toutes les classes (théoriquement pour r tendant vers l'infini, pratiquement dès que r=2).

On trouvera une description détaillée de l'algorithme d'optimisation employé dans BOCK (1979). Celui-ci se résume aux étapes suivantes :

(1) - donnée du nombre de classes K et de la valeur du coefficient r ;

(2) - lecture ou tirage aléatoire d'une partition initiale ;

(3) - calcul des centres de gravité de chaque classe ;

(4) - calcul de nouvelles valeurs d'appartenance pour chaque individu (moyenne harmonique des distances de l'individu aux centres de gravité de toutes les classes)

(5) - retour en (3) jusqu'à obtention du critère minimum.

2.2. Critère de validité des classes obtenues.

Le problème du nombre de classes optimal est courant. Divers indices qui mesurent la validité d'une partition floue ont été proposés. Nous avons retenu l'indice proposé par LIBERT & ROUBENS (1982). Il est défini par :

$$I = \{ K [\sum_x u_{max}(x) / M + \min_x (u_{max}(x))] / 2 - 1 \} / (K-1)$$

avec M : nombre d'individus
et $u_{max}(x) = \max_k [u_k(x)]$

Cet indice, qui varie entre 0 et 1, ne dépend que légèrement du nombre de classes K et indique la netteté de la séparation entre classes. Intuitivement, on peut le justifier par le fait que lorsqu'on mesure la qualité d'une partition, on s'intéresse d'abord aux maxima des valeurs d'appartenance. On considère donc le rapport $\Sigma\, u_{max}(x)/M$, fonction linéaire des $u_{max}(x)$. L'ajout du plus petit des maxima met en évidence la séparation entre classes. La division par 2 fait varier l'indice entre 0 et 1 et la pondération par K permet de réduire l'influence du nombre de classes. On trouvera dans LIBERT & ROUBENS (1982) une discussion des mérites comparés de cet indice et d'autres proposés antérieurement.

3. Considérations pratiques.

3.1. Choix d'une distance.

Le programme FUZZY utilise la distance euclidienne classique. Les limitations inhérentes à ce choix et les diverses solutions pour s'en affranchir ont été présentées par ROUX (1989) dans cette revue. Les variables doivent être quantitatives et de mêmes unités ou avoir subi une normalisation préalable. Les données qualitatives sont utilisables après analyse factorielle des correspondances ou par transformation conduisant à une distance du khi-2.

Le calcul des distances dans le programme source est de toute façon écrit dans un sous-programme indépendant, DIST, permettant à l'utilisateur d'introduire toute mesure de dissimilarité d, c'est-à-dire vérifiant :

$$\forall\ (x,y) \in E^2, \qquad \begin{array}{l} d(x,x)=0 \\ d(x,y)\geq 0 \\ d(x,y)=d(y,x) \end{array}$$

Notons que LIBERT & ROUBENS (1982) ont développé une méthode de classification floue applicable à des données non métriques, pour lesquelles on ne disposerait que d'une matrice de dissimilarité.

3.2. Les entrées du programme.

Outre le fichier de données, le programme lit en paramètres le format de ces données, le nombre d'individus et de variables, le nombre de classes finales de la partition et le coefficient r.

L'algorithme part d'une partition floue initiale de l'ensemble des individus en K classes. Le résultat final peut dépendre du choix de cette partition. On peut donc dans FUZZY spécifier cette partition initiale, même partiellement, si l'on a une idée *a priori* sur l'affectation de certains individus. Les individus non affectés sont équirépartis automatiquement entre toutes les classes (appartenance 1/K à chacune des K classes).

3.3. Les sorties du programme.

On obtient en sortie :

- la description de la partition initiale,
- le nombre d'itérations effectuées avant convergence du critère d'optimisation,
- la mesure de la validité de la classification obtenue selon le critère défini précédemment,
- le tableau des valeurs d'appartenance de chaque individu à chacune des classes (valeur entre 0 et 1),
- les coordonnées des centres de gravité de chaque classe.

3.4. Choix du nombre de classes et du coefficient r.

Lorsqu'on ne connaît pas *a priori* le nombre K de classes finales, le critère I de validité peut constituer un second critère à optimiser dans ce but. On fera donc varier K autour d'une valeur choisie au vu des données, et on recherchera le nombre de classes pour lequel I est maximum, le coefficient r restant fixe.

LIBERT et ROUBENS (1982) donnent à titre indicatif l'espérance de I calculée sous l'hypothèse d'une distribution uniforme pour chaque individu des valeurs d'appartenance. On a, en fonction du nombre de classes :

K	2	3	4	5	6
E(I)	0.250	0.208	0.181	0.160	0.145

Dans les conditions plus réalistes d'une distribution non uniforme, la dépendance de I par rapport à K disparaît.

Le choix du coefficient r est laissé à l'appréciation de l'utilisateur. Il permet d'imposer un degré de flou plus ou moins important à la partition finale. Lorsqu'on le fait croître à partir de 1, on fait progressivement apparaître, à partir d'une classification ordinaire, les individus centraux, périphériques ou hybrides. Puis ces distinctions s'estompent jusqu'à n'obtenir qu'une équirépartition de tous les individus dans toutes les classes. On utilise communément des valeurs autour de r=1.2.

4. Discussion-conclusion.

Cette méthode, basée sur la minimisation d'une pseudo-variance intra-groupe, est une généralisation aux ensembles flous des méthodes classiques de classification ISODATA et k-moyennes. En particulier, pour r=1, on travaille sur le même critère d'optimisation qu'avec ces méthodes. La plupart des logiciels de classification de cette famille fournissent d'ailleurs les coordonnées des centres de gravité et souvent la distance des individus d'une classe à son centre de gravité. On trouve même dans le programme des k-moyennes de BMDP un graphique donnant la distance à leur centre de gravité des individus d'une classe, suivi du graphique des distances des autres individus. Notre programme constitue une généralisation de ces procédures.

En fait, la classification floue se distingue principalement par le fait que l'on ne s'intéresse pas à l'affectation proprement dite des individus, mais à leur distance au centre de gravité des différents classes.

On peut aussi constater que cette méthode permet de généraliser la procédure de recherche des "formes fortes" employée dans la méthode des nuées dynamiques (DIDAY, 1970). En effet, ces formes stables retrouvées à partir de partitions initiales différentes correspondent aux individus centraux en classification floue, ayant une forte valeur d'appartenance à un groupe donné.

Finalement, cette méthode de classification fournit une information beaucoup plus détaillée et riche des relations entre individus qu'une partition classique. En contrepartie, elle est à employer avec discernement. On obtient en effet en sortie, pour M individus à classer, non plus M informations de classement comme avec les méthodes classiques (les classes d'affectation de ces M individus), mais K x M valeurs d'appartenance ! Il faut donc bien être dans le cas où l'on donne la primauté à ces valeurs d'appartenance, et non au regroupement en sous-ensembles disjoints. La méthode peut valablement venir en complément d'une ordination ou d'une classification classique où, après avoir défini des classes (ou leur noyau caractéristique), on veut connaître le degré de liaison des individus à ces classes.

5. Bibliographie.

BEZDEK J. (1974) - Numerical taxonomy with fuzzy sets. *J. Math. Biol.*, 1, 57-71.

BOCK H.H. (1979) - Fuzzy clustering procedures, *Analyse de données et informatique*, AMIRCHANY M. & NEEL D. Eds, INRIA, Paris, 205-218.

DIDAY E. (1970) - La méthode des nuées dynamiques et la reconnaissance des formes, *Rev. Stat. appl.*, 19, 2.

DUBOIS D. & PRADE H. (1983) - The development of fuzzy set theory - applications outside of pure mathematics -, *Proceed. 5th Int. Seminar on Fuzzy Set Theory*, J Kepler Univ., Linz, Austria, Sept. 5-9, E.P. KLEMENT Ed.

DUNN J. (1974) - A fuzzy relative to the ISODATA process and its use in detecting compact, well-separated clusters. *J. Cybernetics*, 3 (3), 32-57.

DUPOUEY J.L. (1985) - Intérêt de la notion d'ensemble flou en phytosociologie forestière. Application à la classification des relevés de végétation. *Colloques Phytosociologiques*, 14, "Phytosociologie et foresterie, 43-53.

KAUFFMANN A. (1977) - Introduction à la théorie des sous-ensembles flous. I- Elements théoriques de base, Masson, Paris, 424 p.

LIBERT G. & ROUBENS M (1982) - Méthodes non métriques valuées en classification automatique et détermination du nombre optimal de classes. *Actes du Colloque Statistique et Analyse des données*, Pointe-à-Pitre, 13-18 Déc., INRIA, 71-84

ROUX M. (1989) - Construction ascendante hiérarchique rapide : le programme CAHVR, *Revue de Modulad*, 3, 1-6

RUSPINI E.H. (1969) - A new approach to clustering. *Inf. Control*, 15, 22-32.

ZADEH L.A. (1965) - Fuzzy sets, *Inf. Control*, 8, 338-353