PUBLICATIONS MATHÉMATIQUES ET INFORMATIQUES DE RENNES

JACQUES LEBBE

Classification conceptuelle de descripteurs

Publications de l'Institut de recherche mathématiques de Rennes, 1993, fascicule 3
« Fascicule de didactique des mathématiques », , p. 41-59
<http://www.numdam.org/item?id=PSMIR_1993___3_41_0>

CLASSIFICATION CONCEPTUELLE DE DESCRIPTEURS

Jacques LEBBE
Université de Paris 6
IRISA — Université de Rennes 1

1. Une vision de l'Analyse des Données

Schématiquement l'Analyse des Données est représentée par deux pôles, un pôle opératoire (AOD) où l'on cherche à répondre à des questions désignées par l'utilisateur (Analyse Discriminante, Régression, ...) et un pôle exploratoire où l'on cherche une présentation démonstrative des données susceptible de permettre à l'utilisateur de formuler les bonnes questions (Analyse des Correspondances, Classification Automatique, Sériation, ...).

L'Analyse Exploratoire de Données (AED) consiste en la recherche d'une redescription résumée et structurée des données. Dans cette définition deux notions sont importantes, d'une part l'idée de résumé, car l'on désire occulter un aspect des données pour mieux en discerner la part pertinente, d'autre part l'idée de structure, car c'est en déterminant quel type de structure peut avoir le résumé cherché que l'utilisateur d'une méthode d'AED définit ce qui est signifiant pour lui.

Selon les types de structures auxquelles on s'intéresse plusieurs formes d'AED existent. Si ces structures sont formées de classes on parle alors de Classification Automatique (CA) (Jambu, 1978), (Celeux et coll., 1989).

2. Classification Automatique

Une méthode de CA cherche à structurer un ensemble d'items en un ensemble de classes. Plus formellement :

Soit I, l'ensemble des items à classifier.

Soit $C = \mathcal{P}(I)$, l'ensemble des classes de I.

Soit $K = \mathcal{P}(C)$, l'ensemble des classes de C.

L'ensemble $S \subset K$ est un type de structure classificatoire ssi :

$\forall\, k \in S, \bigcup \{c \in k\} = I$

$\forall\, k \in S, \varnothing \notin k$

$\exists\, k \in S$ tel que $I \in k$

$\exists\, k \in S$ tel que $\forall\, i \in I, \{i\} \in k$.

On appelle un élément $k \in S$ une structure classificatoire de type S ou plus simplement une classification. Ainsi S est un type de structure classificatoire si toutes ses classifications sont formées de classes non vides qui recouvrent I, s'il existe une classification de S contenant I et une contenant les singletons de I.

Un algorithme de CA consiste, étant donné un ensemble I d'items à classifier et une structure classificatoire S, à trouver une classification $k \in S$.

3. Types de structures classificatoires

Les types de structures classificatoires se différencient par les propriétés de leurs classifications.

Un des plus simples est l'ensemble des partitions $P \subset K$ qui est caractérisé par le fait que ses classifications ne contiennent que des classes qui n'ont pas d'élément en commun. C'est-à-dire :

$$\forall p \in P, \forall c, c' \in p, c \cap c' = \emptyset.$$

Un autre type classique est l'ensemble des hiérarchies $H \subset K$ qui est caractérisé par le fait que ses classifications contiennent toutes I et les singletons de I et que toutes les paires de classes sont, soit incluses l'une dans l'autre, soit n'ont pas d'élément commun. C'est-à-dire :

$$\forall h \in H, I \in h \wedge \forall i \in I, \{i\} \in h \wedge \forall c, c' \in h, c \cap c' \in \{\emptyset, c, c'\}.$$

Si $c \cap c' = c$ (respectivement c'), on dira que la classe c' est plus générale que la classe c (respectivement plus spécifique).

$h_0 = \{I, \{i\}, \{i'\}, \{i''\}, \ldots\}$ est appelée hiérarchie triviale.

Exemple :

Soit $I = \{1, 3, 13, 21\}$,

$h = \{\{1, 3, 13, 21\}, \{1, 3, 21\}, \{3, 21\}, \{1\}, \{3\}, \{13\}, \{21\}\}$ est une hiérarchie.

De nombreux autres types de structures classificatoires ont été définis par exemple les multipartitions (Brossier, 1979), les recouvrements, les pyramides (Diday, 1986) etc. De plus un type peut lui même se diviser en sous-types selon des données supplémentaires associées aux classes, par exemple les hiérarchies munies d'un ordre sur leurs classes sont dites hiérarchies stratifiées, de même si a chaque classe d'une hiérarchie est associée un réel on obtient une hiérarchie indicée (Caillez et Pagès, 1976).

4. Algorithmes de Classification Hiérarchique

Un algorithme de Classification Hiérarchique est un algorithme de classification où le type de structure classificatoire est H, c'est-à-dire l'ensemble des hiérarchies.

La plupart des algorithmes de classification hiérarchique ont en commun un même cadre qui consiste à partir de la hiérarchie triviale puis à la modifier pas à pas tant que cela est possible.

Algorithme général :

> Soit $h_0 = \{I, \{i\}, \{i'\}, \{i''\}, \ldots\}$, la hiérarchie triviale,
>
> Posons $h = h_0$
>
> faire tant que op(h) existe
>
> $\quad h = \text{op}(h)$
>
> fin

où op: $H \to H$ est un opérateur de modification de hiérarchie.

Dans les deux types d'algorithmes les plus utilisés (ascendant et descendant) les modifications consistent à rajouter des classes à la hiérarchie courante tout en respectant des conditions qui garantissent l'obtention d'une hiérarchie à chaque étape.

La Classification Descendante Hiérarchique (CDH) consiste à rajouter à la hiérarchie courante deux classes formant une partition d'une classe divisible de cette hiérarchie. Une classe est dite divisible si ce n'est pas un singleton et si elle n'inclut aucune autre classe de cardinal supérieur à 1 de la hiérarchie.

Exemple :

Soit $I = \{1, 3, 13, 20, 21\}$.

Soit $h = \{\{1, 3, 13, 20, 21\}, \{13, 20, 21\}, \{1\}, \{3\}, \{13\}, \{20\}, \{21\}\}$.

La seule classe divisible de h est $\{13, 20, 21\}$.

La Classification Ascendante Hiérarchique (CAH) consiste à rajouter à la hiérarchie courante une classe résultante de l'union de deux classes agrégeables. Une classe est dite agrégeable si elle n'est incluse dans aucune autre classe de la hiérarchie à part l'ensemble I lui-même.

Exemple :

Dans l'exemple précédent, les seules classes agrégeables de h sont :
$\{1\}$, $\{3\}$ et $\{13, 20, 21\}$.

5. Algorithme de Classification Ascendante Hiérarchique

Soit a: $H \to K$ l'application qui calcule l'ensemble des classes agrégeables d'une hiérarchie,

$$\text{a}(h) = \{c \in h - \{I\} \mid \forall\, c' \in h - \{c\} - \{I\}, c \not\subset c'\}.$$

Soit g: $H \to K$ l'application qui calcule l'ensemble des agrégations (paires de classes agrégeables) admissibles d'une hiérarchie,

$$\text{g}(h) = \{\{c, c'\} \mid c, c' \in \text{a}(h), c \neq c'\}.$$

Soit m: $H \to C$ la fonction qui calcule la meilleure classe à rajouter à une hiérarchie. Le plus souvent m(h) est déterminée par une fonction q: $C \times C \to \mathbf{R}$ de calcul de la qualité d'une agrégation, les classes aggrégées c et c' maximisant $q(c, c')$. Quelque soit la fonction m, pour que la hiérarchie soit déterminée par l'algorithme, la meilleure classe doit être nécessairement unique.

$$m(h) = c \cup c',\ \{c, c'\} \in g(h) \text{ et } q(c, c') = \max_{\{c'', c'''\} \in g(h)} q(c'', c''').$$

On remarquera que cette qualité ne fait pas intervenir l'ensemble de la hiérarchie où la classe va être rajoutée, ainsi la CAH peut n'être seulement qu'une heuristique d'optimisation locale.

Algorithme de CAH :

> Soit $h_0 = \{I, \{i\}, \{i'\}, \{i''\}, \ldots\}$, la hiérarchie triviale,
>
> Posons $h = h_0$
>
> faire tant que $g(h) \neq \varnothing$
>
> $\quad h = h \cup m(h)$
>
> fin

Exemple de CAH

Soit $I = \{1, 2, 11, 18\}$.

Soit la fonction $q(c, c') = \min_{i \in c,\, i' \in c'} (i - i')$.

$h = \{\{1, 2, 11, 18\}, \{1\}, \{2\}, \{11\}, \{18\}\}$.

$g(h) = \{\{1\}, \{2\}\}, \{\{1\}, \{11\}\}, \{\{1\}, \{18\}\}, \{\{2\}, \{11\}\}, \{\{2\}, \{18\}\}, \{\{11\}, \{18\}\}\}$.

$m(h) = \{1, 2\}$, $q(\{1\}, \{2\}) = 1$.

$h = \{\{1, 2, 11, 18\}, \{1, 2\}, \{1\}, \{2\}, \{11\}, \{18\}\}$.

$g(h) = \{\{\{1, 2\}, \{11\}\}, \{\{1, 2\}, \{18\}\}, \{\{11\}, \{18\}\}\}$.

$m(h) = \{11, 18\}$, $q(\{11\}, \{18\}) = 7$.

$h = \{\{1, 2, 11, 18\}, \{1, 2\}, \{11, 18\}, \{1\}, \{2\}, \{11\}, \{18\}\}$.

$g(h) = \varnothing$.

6. Catégories d'items à classifier

L'ensemble I des items peut être de nature quelconque. Deux cas de CAH sont classiques en AED lors de l'étude des tableaux de données : d'une part l'étude des objets décrits et d'autre part l'étude des descripteurs sur ces objets (Anderberg, 1973) (Lerman, 1981).

Soit O un ensemble d'objets.

Soit D un ensemble de descripteurs.

Soit V un ensemble de valeurs.

Soit t: $D \times O \rightarrow V$,
un tableau de données qui associe à chaque couple (d, o) une valeur dite description relativement à d de l'objet o.

On appellera l: $O \rightarrow V^{\mathrm{card}(D)} = (\mathrm{t}(d, o), \mathrm{t}(d', o), \mathrm{t}(d'', o), \ldots)$,
la fonction associant à chaque objet o sa description complète dans le tableau t.

Dans le cas classique, quelque soit l'ensemble I des items à classifier, la fonction q de calcul de la qualité d'une agrégation dépend de la fonction r: $I \times I \rightarrow \mathbf{R}$ qui mesure, grâce à l'information contenue dans le tableau t, la ressemblance (similarité ou dissimilarité) entre deux items (objets ou descripteurs).

Dans le cas de deux objets, leur ressemblance sera d'autant plus grande que leurs descriptions complètes se ressemblent. La fonction r est alors calculée à partir de la comparaison des valeurs $\mathrm{t}(d, o)$ et $\mathrm{t}(d, o')$ des deux objets à comparer pour chacun des descripteurs.

Dans le cas de deux descripteurs, leur ressemblance sera d'autant plus grande qu'ils contribuent de la même manière à la ressemblance entre les objets. C'est-à-dire que deux descripteurs sont d'autant plus ressemblants que les ressemblances entre objets relativement à chacun des descripteurs sont fortement corrélées. La fonction r est alors calculée à partir de la comparaison des différences entre les valeurs $\mathrm{t}(d, o)$ et $\mathrm{t}(d, o')$ d'une part et des différences entre les valeurs $\mathrm{t}(d', o)$ et $\mathrm{t}(d', o')$ d'autre part des deux descripteurs à comparer pour chaque paire d'objets. On parle souvent d'indice d'association pour désigner de telles mesures de ressemblance entre descripteurs. On remarquera que le deuxième problème ne revient pas au premier après échange des objets et des descripteurs et transposition du tableau.

7. Intérêts de la Classification de Descripteurs

La classification de descripteurs, dans une approche d'AED, permet de résoudre trois catégories de problèmes.

Premièrement, les descripteurs peuvent être les véritables objets de l'étude. Par exemple, si dans une enquête l'on cherche à connaître l'opinion des français (O = échantillon de français) sur les prochaines élections présidentielles, les descripteurs seront les réponses à des questions comme "Accepteriez-vous X comme président ?" (D = présidentiables) avec comme valeurs possibles V = {"oui", "non"}. Dans ce cas les classes de descripteurs qui rassemblent les présidentiables qui ont grâce aux yeux des mêmes gens seront plus informatives sur l'opinion des français que des classes de gens. Cette approche est la plus classique en AED.

Deuxièmement, on peut chercher si les descripteurs sont homogènes quant à leur manière de structurer les objets, ce afin d'évaluer les conséquences sur les objets de la présence de tel ou tel ensemble de descripteurs dans le tableau. Contrairement au cas précédent, on s'intéresse à la structuration des objets du tableau de données et l'on cherche à déterminer si tous les descripteurs concourent à la même structuration ou si au contraire il existe des groupes de descripteurs structurant les objets de manières différentes. En effet dans les méthodes d'AED la structure finale sur les objets résulte du consensus entre les structures induites par chaque descripteur. Ainsi deux ensembles de descripteurs fortement structurants mais non consensuels peuvent masquer les relations entre les objets. Par exemple si on désire comparer les villes de France (O = villes) selon leur pluviométrie (V = pluviométrie) relevée journellement (D = jours), on peut se demander s'il est pertinent de comparer les villes à partir de l'ensemble total des jours de l'année et si certaines périodes (hiver, été, ...) ne regroupent pas les villes de manières différentes. Un moyen simple consiste à faire plusieurs analyses en introduisant divers ensembles de descripteurs. La structuration des descripteurs permet d'avoir automatiquement le même résultat. Classiquement en AED, ce genre de problème est abordé grâce aux Analyses Factorielles. Celles-ci permettent d'évaluer les corrélations entre descripteurs ce qui peut suggérer l'existence de classes de descripteurs. Mais on peut aussi souhaiter obtenir ces classes automatiquement.

Troisièmement, on peut chercher à réduire le nombre de descripteurs d'un tableau de données. En effet une classification de descripteurs ayant été obtenue, si une classe contient un ensemble de descripteurs fortement ressemblants, il est raisonnable de penser qu'un d'entre eux (éventuellement muni d'un poids plus grand) pourrait remplacer les autres. Par exemple, si on dispose d'un ensemble de caractères manuscrits (O = caractères) décrits par divers paramètres comme le nombre de boucles ou de points de rebroussement (D = paramètres) on peut chercher à réduire l'ensemble D tout en permettant de respecter la discrimination des lettres que les caractères représentent. Si le nombre de boucles et le nombre de points de rebroussement sont proches dans la classification car il discriminent les mêmes lettres, il est acceptable de supprimer l'un des deux. Cette dernière approche rejoint les problèmes classiques de sélection de descripteurs en AOD (Analyse Discriminante pas à pas, Discrimination par Arbre, ...) (Celeux, 1990) mais à la différence de ces méthodes on cherche ici, non pas à découvrir un unique ensemble de descripteurs, mais plutôt à exhiber à l'utilisateur, par une approche exploratoire, une structure des descripteurs permettant de guider son choix.

8. Classification conceptuelle

La définition originale d'une méthode de Classification Conceptuelle (CC) est la suivante :

> *L'étiquette de Classification Conceptuelle peut être donnée à toute méthode qui détermine une structure dans un ensemble d'objets et dans laquelle les noeuds représentent des "concepts" caractérisant les classes correspondantes et les arcs (entre les noeuds) représentent des relations entre les concepts. Par "concept" nous désignons une description adaptée à l'homme dans laquelle interviennent des propriétés des objets et des relations entre eux (Michalski et Stepp, 1981 (b)).*

Nous préférons la définition suivante :

> *Une méthode de classification conceptuelle est un algorithme de classification où les classes formées sont contraintes à se rapprocher le plus possible des classes descriptibles dans un langage donné. Ces classes seront appelées "concepts".*

La CC a été l'objet de nombreux travaux et comparaison avec les méthodes classiques en CA (Michalski, 1980), (Michalski et Stepp, 1981 (a), 1983), (Michalski et coll., 1982), (Dale, 1985).

La justification de la CC en AED est claire si l'on observe des utilisateurs de la CA. Après qu'un algorithme ait produit un résultat l'utilisateur passe toujours par une phase d'interprétation où il tente de retrouver une explication rationnelle des classes produites. Cette phase d'interprétation qui n'est pas classiquement automatisée en Analyse des Données conduit à deux résultats : premièrement une sélection des classes qui semblent les plus pertinentes (souvent évaluées comme telles car facilement explicables) et deuxièmement à la sélection parmi l'ensemble des explications des classes pertinentes d'une explication particulière (souvent choisie car concise).

Le but de la CC est d'automatiser autant que faire se peut la génération d'une structure classificatoire directement interprétée en cherchant à restreindre les classes aux seules dont la forme des explications est suffisamment proche de celle qui sera facilement acceptée par l'utilisateur pour former un nouveau concept. Bien entendu, on ne traite pas des concepts directement (concept = conçu par un esprit) mais d'une forme possible de leur représentation, désignée par représentation conceptuelle, qui espère-t-on fournira de bonnes explications des classes aux yeux de l'utilisateur.

9. Classification Conceptuelle Hiérarchique

Cette idée de Classification Conceptuelle Hiérarchique (CCH) a été principalement utilisée pour la recherche de partitions ou de hiérarchies (mais aussi de pyramides (Brito, 1991) ou de hiérarchies de recouvrement (Bisson, 1993)). Mais rien n'empêche de l'appliquer à la recherche d'autres types de structures classificatoires.

Soit R l'ensemble des représentations des classes. Les représentations possibles dépendent des applications et seront envisagées plus loin.

Soit ρ: $C \rightarrow R$, l'application de calcul de la représentation d'une classe.

Soit κ: $R \times C \rightarrow$ {Vrai, Faux}, l'application qui détermine si une représentation r est une représentation conceptuelle de la classe c. Par raccourci, $\kappa(\rho(c), c)$ sera noté $\kappa(c)$.

Etant donné un ensemble I d'items à classifier et un ensemble R de représentations, la CCH revient à trouver $h \in H$ tel que :

$$\forall\, c \in h,\ \kappa(c) = \text{Vrai}$$

c'est-à-dire une hiérarchie dont toutes les classes ont une représentation conceptuelle.

Du fait de la définition d'une hiérarchie, pour que le problème ait une solution il faut que :

$$\kappa(I) = \text{Vrai et } \forall\, i \in I,\ \kappa(\{i\}) = \text{Vrai}.$$

10. Classification Conceptuelle Ascendante Hiérarchique

Pour devenir une Classification Conceptuelle Ascendante Hiérarchique (CCAH), un algorithme de CAH doit modifier la fonction g de calcul des agrégations admissibles :

$$\mathrm{g}(h) = \{\{c, c'\} \mid c, c' \in \mathrm{a}(h), c \neq c', \kappa(c \cup c') = \text{Vrai}\}.$$

D'autre part on exige que la fonction q qui calcule la qualité d'une agrégation soit une fonction de la seule représentation associée au résultat de l'agrégation. Ainsi,

$$q(c, c') = f(\rho(c \cup c')),$$

où f: R $\rightarrow$ $\mathbf{R}$ est une application qui calcule la qualité d'une représentation conceptuelle.

On remarquera que du fait de cette dernière particularité, le critère d'agrégation classique en CAH qui permet de calculer la qualité d'une agrégation quelconque à partir des qualités des agrégations par paire de tous les singletons n'existe pas nécessairement en CCAH.

On remarquera que d'autre part l'existence d'une application f de calcul de la qualité d'une représentation conceptuelle n'est pas toujours nécessaire. En effet il suffit de définir un ordre sur les représentations conceptuelles noté $>_R$. Si $r >_R r'$ on dit que r est meilleure que r'. Pour déterminer de manière unique la meilleure agrégation à effectuer cet ordre doit être strict et total.

Dans ce cas la fonction m qui calcule la meilleure classe à rajouter à une hiérarchie devient :

$$m(h) = c \cup c',$$

$$\text{tel que } \{c, c'\} \in g(h), \forall \{c'', c'''\} \in g(h), \rho(c \cup c') >_R \rho(c'' \cup c''').$$

Exemple de CCAH

Soit $I = \{1, 2, 11, 18\}$.

Soit $R = \mathbf{R} \times \mathbf{R}$, l'ensemble des couples de réels.

Soit la fonction $\rho(c) = (\Sigma_{i \in c}\, i \mid \text{card}(c), \text{card}(c))$, qui associe à une classe un couple formé de sa moyenne et de son cardinal.

Soit $\kappa(c)$ = Vrai ssi $\rho(c) = (m, n)$ et $m \in \mathbf{N}$ c'est-à-dire si la moyenne de la classe est entière.

L'ordre entre représentations conceptuelles est l'ordre lexicographique inverse :

$$r = (m, n) >_R r' = (m', n') \text{ ssi } m < m' \text{ ou si } m = m' \text{ et } n < n'$$

On vérifie que toutes les classes de $h_0 = \{\{1, 2, 11, 18\}, \{1\}, \{2\}, \{11\}, \{18\}\}$ ont une représentation conceptuelle, la moyenne de leurs éléments étant entière. Par exemple $\rho(I) = (8, 4)$ et $\rho(\{2\}) = (2, 1)$.

$h = h_0$.

$g(h) = \{\{\{1\}, \{11\}\}, \{\{2\}, \{18\}\}\}$.

$m(h) = \{1, 11\}$, $\rho(\{1, 11\}) = (6, 2) >_R \rho(\{2, 18\}) = (10, 2)$.

$h = \{\{1, 2, 11, 18\}, \{1, 11\}, \{1\}, \{2\}, \{11\}, \{18\}\}$.

$g(h) = \{\{\{1, 11\}, \{18\}\}, \{\{2\}, \{18\}\}\}$.

$m(h) = \{2, 18\}$, $\rho(\{2, 18\}) = (10, 2) >_R \rho(\{1, 11, 18\}) = (10, 3)$.

$h = \{\{1, 2, 11, 18\}, \{1, 11\}, \{2, 18\}, \{1\}, \{2\}, \{11\}, \{18\}\}$.

$g(h) = \varnothing$.

Agrégation des représentations conceptuelles

Il est souhaitable que la fonction ρ soit décomposable, c'est-à-dire qu'il existe une application α: $R \times R \to R$ d'agrégation des représentations conceptuelles telle que :

$$\alpha(\rho(c), \rho(c')) = \rho(c \cup c').$$

Cela permet de simplifier le calcul de la qualité d'une agrégation. En effet dans ce cas la fonction q devient :

$$q(c, c') = f(\rho(c \cup c')) = f(\alpha(\rho(c), \rho(c'))).$$

Cela permet de définir une fonction f ': $R \times R \to \mathbf{R}$ qui permet de calculer la qualité d'une agrégation directement à partir des représentations des classes de l'agrégation, avec :

$$q(c, c') = f'(\rho(c), \rho(c')).$$

Exemple

Dans le cas de l'exemple précédent une fonction α existe, en effet :

si $\rho(c) = (m, n)$ et $\rho(c') = (m', n')$ alors

$$\alpha((m, n), (m', n')) = ((mn + m'n') / (n + n'), n + n') = \rho(c \cup c').$$

Si par contre dans cet exemple, la représentation conceptuelle n'était constituée que de la moyenne il serait nécessaire de revenir aux représentations des singletons pour construire la représentation d'une classe.

Lien avec la CAH classique

L'originalité de la CCAH par rapport à l'algorithme classique de CAH réside en trois points.

- Une généralisation : car tout élément de C n'est pas une classe acceptable.
- Une propriété originale : car à toute classe de la hiérarchie est associée une représentation conceptuelle calculée et utilisée pendant la construction de la hiérarchie.
- Une spécification : car la qualité d'une agrégation est évaluée seulement à partir de la qualité de la classe résultant de l'agrégation (par l'intermédiaire d'une qualité de la représentation conceptuelle associée) et sans tenir compte des classes agrégées.

Mais dans quelle mesure la CCAH est une approche nouvelle de la classification ?

En fait la CAH avec le critère d'agrégation du lien maximum est un cas particulier de CCAH si l'on considère que :

$R = C$, avec $\rho(c) = c$,

et que $\rho(c) >_R \rho(c')$ si $\max_{i,\, i' \in c} \delta(i, i') < \max_{i'',\, i''' \in c'} \delta(i'', i''')$,

où δ: $I \times I \to \mathbf{R}$ est une dissimilarité sur I.

Cette méthode revient à une CAH avec l'indice d'agrégation par le lien maximum car si $\{c, c'\} \in g(h)$ alors :

$$\max_{i \in c,\, i' \in c'} \delta(i, i') = \max_{i'',\, i''' \in c \cup c'} \delta(i'', i''').$$

La CAH avec le critère d'agrégation par le lien moyen n'est par contre pas une CCAH, car contrairement au cas précédent l'indice d'agrégation n'est pas calculable à partir du seul résultat de l'agrégation. En effet, il faut nécessairement pour calculer la distance moyenne entre deux items des classes à agréger connaître ces deux classes.

On remarquera que dans ces deux cas précédents toute classe possède une représentation conceptuelle ; il n'y a pas de contraintes sur les classes acceptables. Par contre, il existe aussi un domaine classique de l'AED, la classification sous contrainte, où toutes les classes ne sont pas acceptables. La plus connue est la classification sous contrainte de contiguïté où l'ensemble I des items est muni d'un ordre total et strict et où seuls les intervalles de I pour cet ordre sont acceptées (Murtagh, 1982) (Perruchet, 1982). De manière plus générale, il existe aussi la classification sous contrainte de graphe où par exemple seules les classes formées des sommets des parties connexes du graphes sont acceptées (Diday et coll., 1979). Une méthode de CAH sous contrainte utilisant le critère d'agrégation du lien maximum est très proche d'une CCAH.

Mais on remarquera que dans tous les cas cités les représentations conceptuelles des classes sont très pauvres puisque limitées à la classe elle-même. Ces méthodes utilisées seules, sans aide supplémentaire pour l'interprétation, ne présentent que peu de capacités explicatives. Par contre, il est classique en AED d'accompagner les classifications produites par CAH de résultats complémentaires, comme des descriptions statistiques des classes. Ces représentations des classes permettent d'interpréter plus facilement, d'expliquer et de justifier auprès de l'utilisateur le résultat de l'algorithme et visent donc au même but pragmatique que les "concepts" de la CC. Mais ces représentations sont construites *a posteriori* alors qu'en CCAH elles sont construites pendant le déroulement de l'algorithme et c'est sur la base de ces représentations, c'est là l'originalité principale, que les agrégations à accomplir sont choisies.

11. Classification Conceptuelle Ascendante Hiérarchique d'Objets

La forme la plus habituelle de représentation des classes en CC d'objets, appelée Classification Conceptuelle Ascendante Hiérarchique d'Objets (CCAHO), est la représentation conjonctive caractéristique (ou caractérisation conjonctive).

Soit t: $D \times O \rightarrow V$ un tableau de données.

L'ensemble des représentations conjonctives est $R = (\mathcal{P}(V) - \{\varnothing\})^{\mathrm{card}(D)}$.

Soit la fonction e: $C \times D \rightarrow \mathcal{P}(V) - \{\varnothing\}$ qui associe à chaque classe c l'ensemble des valeurs du descripteur d qu'au moins un objet de la classe prend, c'est-à-dire :

$$e(c, d) = \{v \in V \mid \exists\, o \in c,\ t(d, o) = v\}$$

La fonction de représentation est alors $\rho(c) = r = (e(c, d), e(c, d'), e(c, d''), \ldots)$. On notera $r(d) = e(c, d)$ l'ensemble des valeurs correspondant au descripteur d dans la représentation conjonctive r.

Exemple :

Soit $O = \{o, o', o'', o'''\}$.

Soit $D = \{d, d', d'', d'''\}$.

Soit $V = \{0, 1, 2\}$.

Soit $\mathrm{l}(o) = (1, 1, 1, 2)$,

$\mathrm{l}(o') = (1, 1, 2, 0)$,

$\mathrm{l}(o'') = (2, 0, 1, 2)$,

$\mathrm{l}(o''') = (2, 0, 1, 1)$.

$\rho(\{o, o'\}) = (\{1\}, \{1\}, \{1, 2\}, \{0, 2\})$

La fonction $\kappa(r, c)$ prend la valeur Vrai si r est une représentation conjonctive caractéristique de la classe c, c'est-à-dire ssi :

$\forall\, o \in c, \forall\, d \in D, \mathrm{t}(d, o) \in \mathrm{r}(d)$ et

$\forall\, o \in O - \{c\}, \exists\, d \in D, \mathrm{t}(d, o) \notin \mathrm{r}(d)$.

La première condition explique la dénomination de représentation conjonctive car l'appartenance d'une valeur de l'objet à la partie de la représentation correspondante est exigée pour tous les descripteurs. Par construction cette condition est vérifiée si r est la représentation de la classe c et donc il suffit de tester si :

$\forall\, o \in O - \{c\}, \exists\, d \in D, \mathrm{t}(d, o) \notin \mathrm{e}(c, d)$,

C'est-à-dire qu'une représentation conjonctive est conceptuelle si tous les contre-exemples (objets de $O - \{c\}$) se différencient des objets de la classe par au moins un descripteur.

Exemple :

En reprenant les données de l'exemple précédent.

$\kappa(\rho(\{o, o'\})) = \text{Vrai}$

car ni o'' ni o''' ne présente la valeur 1 pour le premier descripteur

Parmi les fonctions d'évaluation de la qualité d'une représentation conjonctive on peut prendre :

$\mathrm{f}(r) = 1 / \prod_{d \in D} \mathrm{card}(\mathrm{r}(d))$.

On montre que si $\rho(c)$ est caractéristique alors $\mathrm{f}(\rho(c))$ est maximale parmi les caractérisations de c. On montre aussi que si $\rho(c)$ n'est pas caractéristique alors il n'existe pas de caractérisation conjonctive de c.

Exemple :

En reprenant les données de l'exemple précédent.

$h_0 = \{\{o, o', o'', o'''\}, \{o\}, \{o'\}, \{o''\}, \{o'''\}\}$.

On vérifie que toutes les classes de h_0 ont une représentation conceptuelle : $\rho(I) = (V, V, V, V)$ est caractéristique car il n'y a pas de contre-exemple, de même tous les objets ont une représentation conceptuelle car leur description complète est unique dans le tableau.

$h = h_0$.

$g(h) = \{\{\{o\}, \{o'\}\}, \{\{o\}, \{o'''\}\}, \{\{o'\}, \{o''\}\}, \{\{o''\}, \{o'''\}\}\}$.

L'ensemble $\{\{o'\}, \{o''\}\}$ n'est pas une agrégation admissible, c'est-à-dire $\kappa(\rho(\{o', o''\})) =$ Faux, car il n'existe pas de caractérisation conjonctive de $\{o', o''\}$. En effet $\rho(\{o', o''\}) = (\{1, 2\}, \{0, 1\}, \{1, 2\}, \{0, 2\})$ et contient pour chaque descripteur la valeur de la description de o.

$m(h) = \{o'', o'''\}$, $\rho(\{o'', o'''\}) = (\{2\}, \{0\}, \{1\}, \{1, 2\})$, $f(\rho(\{o'', o'''\})) = 1 / 2$.

etc.

12. Classification Conceptuelle Ascendante Hiérarchique de Descripteurs

L'idée de CC a été uniquement appliquée à la classification des objets mais rien n'empêche de l'étendre à la classification de descripteurs. L'intérêt de la Classification Conceptuelle Ascendante Hiérarchique de Descripteurs (CCAHD) est de permettre de résoudre les trois types de problèmes déjà cités pouvant motiver la recherche d'une classification de descripteurs mais de manière explicative.

Les difficultés de mise en oeuvre résident dans les choix des représentations à associer aux classes, des représentations considérées comme conceptuelles et des moyens d'évaluer la qualité de ces représentations.

Représentation par graphe de discrimination

Si on dispose d'une fonction Δ_d: $V \times V \rightarrow \{0, 1\}$ qui indique si deux valeurs du même descripteur d sont suffisamment différentes pour considérer que deux objets qui les présentent sont discriminés, alors tout descripteur d'un tableau de données définit un graphe sur les objets dont les arêtes réunissent les couples d'objets discriminés. De même on peut associer à une classe de descripteurs un graphe dont l'ensemble des arêtes est l'union des arêtes des graphes associés aux descripteurs de la classe, c'est-à-dire le graphe de discrimination de la classe de descripteurs. L'intérêt d'une telle représentation est que la CCAHD peut alors être considérée comme la construction pas à pas d'un ensemble de descripteurs virtuels (un par classe de descripteurs) qui auraient pour valeurs le produit cartésien des valeurs des descripteurs de la classe correspondante.

Dans ce cas une représentation peut être considérée comme conceptuelle si le graphe de discrimination possède des propriétés particulières, tels les split-graphs pour lesquels il existe un ensemble de sommets reliés à tous les autres et aucune autre arête dans le graphe, ou encore les graphes formés d'une seule clique. Dans ces deux cas cela permet d'exprimer le graphe associé à chaque classe de descripteurs avec une grande économie de moyens : le sous-ensemble d'objets discriminés de tous les autres dans le premier cas, discriminés entre eux dans le deuxième. La qualité des représentations peut être évaluée en fonction du nombre d'arêtes du graphe.

Représentation par dissimilarités

Il est aussi possible d'associer aux ensembles de descripteurs une matrice de dissimilarités (un graphe complet valué) sur les objets calculée à partir d'une méthode d'Analyse de Données classique.

Dans ce cas la qualité de la représentation associée à la classe de descripteurs peut être choisie pour le fait que les descripteurs structurent les objets d'une manière particulièrement intéressante pour l'utilisateur. Par exemple on peut évaluer la qualité d'une représentation par le meilleur pourcentage d'inertie expliquée par une partition en deux des objets, ou par le meilleur pourcentage d'inertie expliquée par le premier axe factoriel d'une Analyse en Composantes Principales, etc. Cela permet de regrouper des descripteurs s'ils sont fortement consensuels dans leur manière de structurer les objets.

Exemple :

Cet exemple de CCAHD, calculable à la main, a pour but de montrer comment cette technique permet de construire des classes de descripteurs structurant les objets de manière particulière.

Soit $I = \{o_1, o_2, ..., o_{20}\}$, $V = \{0, 1\}$, $D = \{d_1, d_2, d_3, d_4\}$.

Soit le tableau de données de la figure 1.

Soit $R = 2^I$, l'ensemble des partitions en deux classes des objets.

Soit $r \in R$, $r = \{c_O, c'_O\}$ et soit $f(r) = \mathrm{card}(c_O) \cdot \mathrm{card}(c'_O)$.

Les représentations sont des partitions en deux des objets qui seront considérées comme d'autant meilleures que le produit des cardinaux de leurs classes est plus élevé ou, ce qui revient au même, qu'elles sont bien équilibrées.

Les partitions seront obtenues par une procédure de classification qui est elle-même "conceptuelle". Si la classe de descripteurs est un singleton la partition qui la représente est celle induite par le descripteur.

Si la classe est formée de deux descripteurs d et d' la représentation est la meilleure des partitions, au sens de la fonction f, que l'on peut former par des conjonctions des descripteurs de départ (formant des descripteurs virtuels) c'est-à-dire soit $d = 0 \wedge d' = 0$ soit $d = 1 \wedge d' = 0$ soit $d = 0 \wedge d' = 1$ soit $d = 1 \wedge d' = 1$. Par exemple, en consultant le tableau de contingence (Figure 2) correspondant aux descripteurs d_3 et d_4 on voit que la meilleure partition correspond à la conjonction $d_3 = 1 \wedge d_4 = 0$, sa qualité est $8\ (5 + 5 + 2) = 96$. Le descripteur virtuel résultant est donné figure 1 dans la colonne correspondant à la classe de descripteurs.

Suivant le même principe, si la classe est formée de plus de deux descripteurs, la représentation est construite en substituant, par la procédure précédente, à la paire de descripteurs ayant la meilleure représentation, un descripteur virtuel, et ce jusqu'à la formation d'un unique descripteur virtuel pour toute la classe.

	$\{d_1\}$	$\{d_2\}$	$\{d_3\}$	$\{d_4\}$	$\{d_1, d_2\}$	$\{d_3, d_4\}$	$\{d_1, d_2, d_3, d_4\}$
o_1	0	0	0	1	0	1	0
o_2	1	1	1	0	1	0	1
o_3	1	1	0	1	1	1	1
o_4	0	0	1	0	0	0	1
o_5	0	0	1	1	0	1	0
o_6	0	0	0	0	0	1	0
o_7	0	0	0	0	0	1	0
o_8	0	1	1	0	1	0	1
o_9	1	0	1	0	1	0	1
o_{10}	0	0	1	0	0	0	1
o_{11}	0	1	0	1	1	1	1
o_{12}	1	1	1	0	1	0	1
o_{13}	1	0	0	1	1	1	1
o_{14}	0	0	0	0	0	1	0
o_{15}	0	1	0	0	1	1	1
o_{16}	0	0	0	1	0	1	0
o_{17}	1	1	1	0	1	0	1
o_{18}	1	1	0	0	1	1	1
o_{19}	0	0	1	1	0	1	0
o_{20}	0	0	1	0	0	0	1

Figure 1 : Tableau de données binaires.

		$\{d_1\}$	
		0	1
$\{d_2\}$	0	10	2
	1	3	5

		$\{d_1\}$	
		0	1
$\{d_3\}$	0	7	3
	1	6	4

		$\{d_1\}$	
		0	1
$\{d_4\}$	0	8	5
	1	5	2

		$\{d_2\}$	
		0	1
$\{d_3\}$	0	6	4
	1	6	4

		$\{d_2\}$	
		0	1
$\{d_4\}$	0	7	6
	1	5	2

		$\{d_3\}$	
		0	1
$\{d_4\}$	0	5	8
	1	5	2

		$\{d_1, d_2\}$	
		0	1
$\{d_3\}$	0	5	5
	1	5	5

		$\{d_1, d_2\}$	
		0	1
$\{d_4\}$	0	6	7
	1	4	3

		$\{d_1, d_2\}$	
		0	1
$\{d_3, d_4\}$	0	3	5
	1	7	5

Figure 2 : Tableaux de contingences calculés à partir du tableau de la figure 1.

Voici les étapes de la construction :

$h_0 = \{\{d_1, d_2, d_3, d_4\}, \{d_1, d_2\}, \{d_1\}, \{d_2\}, \{d_3\}, \{d_4\}\}$

$h = h_0$

$g(h) = \{\{\{d_1\}, \{d_2\}\}, \{\{d_1\}, \{d_3\}\}, \{\{d_1\}, \{d_4\}\}, \{\{d_2\}, \{d_3\}\}, \{\{d_2\}, \{d_4\}\}, \{\{d_3\}, \{d_4\}\}\}$

$m(h) = \{d_1, d_2\}$

$\rho(\{d_1, d_2\}) = \{$
$\{o_1, o_4, o_5, o_6, o_7, o_{10}, o_{14}, o_{16}, o_{19}, o_{20}\},$
$\{o_2, o_3, o_8, o_9, o_{11}, o_{12}, o_{13}, o_{15}, o_{17}, o_{18}\}$
$\}$

$q(\{d_1\}, \{d_2\}) = 100$

$h = \{\{d_1, d_2, d_3, d_4\}, \{d_1, d_2\}, \{d_1\}, \{d_2\}, \{d_3\}, \{d_4\}\}$

$g(h) = \{\{\{d_1, d_2\}, \{d_3\}\}, \{\{d_1, d_2\}, \{d_4\}\}, \{\{d_3\}, \{d_4\}\}\}$

$m(h) = \{d_3, d_4\}$

$\rho(\{d_3, d_4\}) = \{$
$\{o_2, o_4, o_8, o_9, o_{10}, o_{12}, o_{17}, o_{20}\},$
$\{o_1, o_3, o_5, o_6, o_7, o_{11}, o_{13}, o_{14}, o_{15}, o_{16}, o_{18}, o_{19}\}$
$\}$

$q(\{d_3\}, \{d_4\}) = 96$

$h = \{\{d_1, d_2, d_3, d_4\}, \{d_1, d_2\}, \{d_3, d_4\}, \{d_1\}, \{d_2\}, \{d_3\}, \{d_4\}\}$

$g(h) = \varnothing$

On remarquera qu'il est inutile pour construire la représentation associée à I de recommencer sa construction depuis le début ; à chaque étape il suffit d'utiliser les représentations des classes précédentes. Ainsi :

$\rho(I) = \{$
$\{o_1, o_5, o_6, o_7, o_{14}, o_{16}, o_{19}\},$
$\{o_2, o_3, o_4, o_8, o_9, o_{10}, o_{11}, o_{12}, o_{13}, o_{15}, o_{17}, o_{18}, o_{20}\}$
$\}$

$q(I) = 91$

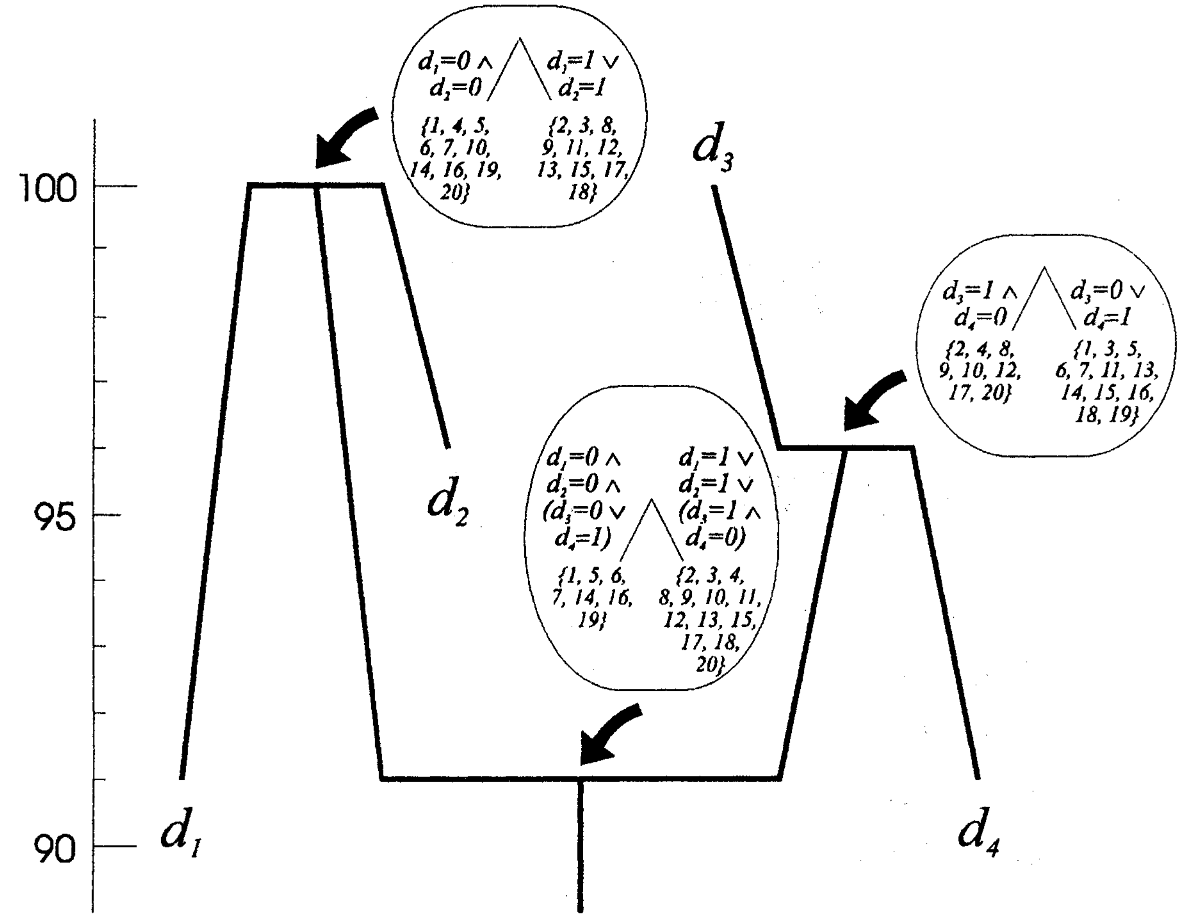

Figure 3 : Dendrogramme résultat. Chaque classe de descripteurs est positionnée en ordonnée selon la qualité de sa représentation. La représentation est elle-même décrite dans un cartouche par les conjonctions correspondant aux deux valeurs du descripteur virtuel correspondant et par les indices des objets des deux classes de la partition.

13. Le problème du blocage en CCAH

L'application générale du modèle "conceptuel" en CAH pose de nombreux problèmes pratiques et théoriques. Un problème particulièrement important est celui du blocage possible de l'algorithme avec certains choix de représentation si aucune paire de classes agrégeables ne possède de représentation conceptuelle. Plusieurs attitudes sont envisageables face à ce problème crucial à résoudre en CCAHD.

Arrêter la construction

Le plus simple est d'arrêter la construction en cas de blocage en considérant que ce dernier est une propriété intéressante des données liée aux choix de représentation. L'inconvénient d'un tel choix est que parfois la production de l'algorithme peut n'être que la hiérarchie triviale.

Autoriser les agrégations de plus de deux classes

Face à un blocage on peut tenter de chercher si un ensemble de trois classes agrégeables ou plus ne possède pas de représentation conceptuelle. Bien entendu cette solution n'est réaliste qu'avec un petit jeu de données ou en toute fin de construction de la hiérarchie. Dans le même esprit on peut tenter d'utiliser un algorithme glouton, en prenant par exemple la paire de classes la plus proche de réaliser un concept puis en lui rajoutant la classe telle que l'agrégation des trois est la plus proche de réaliser un concept, et ce jusqu'à l'éventuelle construction d'un concept etc.

Il est aussi possible de rechercher pour chaque paire de classes agrégeables s'il existe un ensemble de classes incluant cette paire dont la représentation soit conceptuelle. Dans ce cas cet ensemble de classes peut être agrégé. Hélas cette solution ne permet pas de lever tous les blocages en respectant une construction hiérarchique car un tel ensemble de classes ne contient pas nécessairement que des classes agrégeables.

Accepter que toutes les représentations ne soient pas conceptuelles pendant la construction

Une autre solution simple consiste à accepter que les hiérarchies que construit successivement l'algorithme ne soient pas toutes "conceptuelles" à condition que la dernière le soit. Comme les deux solutions précédentes cela ne produit pas nécessairement une hiérarchie saturée mais cela permet de ne jamais être bloqué. Dans ce cas, si une agrégation ayant une représentation conceptuelle existe on la crée en priorité, sinon, on réalise l'agrégation dont la représentation est la plus proche d'une représentation conceptuelle. Une fois la hiérarchie finale obtenue, on élimine si nécessaire les classes dont la représentation n'est pas conceptuelle.

Choisir des représentations plus générales

Si on définit R tel que $\forall\, c \in C$, $\kappa(c)$ = Vrai, alors bien entendu il n'y a plus de risque de blocage. C'est le cas de la CAH par lien maximum. C'est aussi le cas en CCAHO si on accepte des disjonctions dans les représentations ou si on utilise un ensemble important de formules de la logique des prédicats dans le cas de la description d'objets structurés (Bisson, 1993). Dans ces deux cas, les représentations conceptuelles sont assez générales pour que toute classe soit caractérisable ; par contre la représentation associée à une classe n'est plus unique dans le cas général ce qui demande pour choisir celle retenue l'emploi d'heuristiques de généralisation dont les propriétés déterminent la hiérarchie construite.

Adapter la contrainte à la construction hiérarchique

On peut aussi définir κ en fonction de R de telle manière qu'un blocage ne puisse apparaître que si plus aucune classe ayant une représentation conceptuelle ne peut être rajoutée à la hiérarchie. Cela parait surprenant mais cela est possible même si toutes les classes n'ont pas de représentation conceptuelle. Par exemple dans la CAH avec contrainte de contiguïté à une ou deux dimensions on ne bloque jamais (cela est faux pour plus de deux). En effet dans un ensemble d'intervalles (respectivement de rectangles) il est toujours possible de trouver un intervalle (respectivement un rectangle) en recouvrant deux sans empiéter sur aucun des autres. De manière analogue en CCAHO si les objets sont décrits par des descripteurs ayant chacun seulement deux valeurs possibles un blocage ne peut apparaître qu'après la construction de la dernière classe et ce quelque soit le nombre de descripteurs (Guénoche, 1989).

Accepter que la construction ne soit pas hiérarchique à chaque étape

Une solution non encore mise en oeuvre consisterait à ne pas se limiter à la structure hiérarchique, au moins durant la construction, en acceptant d'agréger des paires de classes dont seule l'une d'elles est agrégeable mais dont l'intersection est vide. Cet algorithme permet de construire une structure plus riche que les hiérarchies ou que les pyramides mais tout de même moins riche que les hiérarchies de recouvrement. En augmentant le nombre d'agrégations admissibles cela permet de diminuer fortement les blocages dans les cas où R est beaucoup plus petit que C sans augmenter trop la complexité de l'algorithme. Dans ce cas la fonction g devient :

$$g(h) = \{\{c, c'\} \mid c \in a(h), c' \in h, c \cap c' = \emptyset, \kappa(c \cup c') = \text{Vrai}\}.$$

Si une hiérarchie est nécessaire, celle-ci peut être facilement extraite de la structure construite en ordonnant les classes (par exemple dans l'ordre inverse d'ajout à la structure) et en les rajoutant successivement à la hiérarchie triviale dans cet ordre à condition que la classification obtenue reste hiérarchique à chaque étape.

14. Conclusion

Ce texte a pour but de présenter la CC et son intérêt en AED afin de montrer l'originalité de ce domaine mais aussi les liens existants avec des méthodes classiques. L'approche "conceptuelle" fournit un modèle applicable de manière large à de nombreux problèmes en CA. Son application à la classification hiérarchique de descripteurs pour aider un utilisateur à choisir et à formuler des descripteurs à retenir dans une étude semble particulièrement prometteuse.

15. Références

ANDERBERG M.R., 1973. *"Cluster analysis for application"*. Academic Press, New York. 359 pp.

BISSON G., 1993. *"KBG : Construction de bases de connaissances en logique des prédicats"*. Thèse de l'Université Orsay Paris Sud.

BROSSIER G., 1979. *"Bi-classe et bi-classification"*. In Classification Automatique et perception par ordinateur, INRIA, Le Chesnay: 159-162.

CAILLEZ F. et PAGES J.P., 1976. *"Introduction à l'Analyse des Données"*. Société de Mathématiques Appliquées et de Sciences Humaines, Paris.

CELEUX G. et coll., 1989. *"Classification automatique pour l'analyse des données"*. Dunod, Paris.

CELEUX G., 1990. *"Analyse discriminante sur variables continues"*. INRIA, Le Chesnay.

DALE M.B., 1985. *"On the Comparison of Conceptual Clustering and Numerical Taxonomy"*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-7, N°2: 241-244.

DE PINHO DE BRITO DUARTE SILVA M.P., 1991. *"Analyse de Données Symboliques, Pyramides d'Héritage"*. Thèse de l'Université Paris IX Dauphine.

DIDAY E. et coll., 1979. *"Classification sous contraintes"*. In Optimisation en Classification Automatique, INRIA, Le Chesnay, Vol. 2 : 677-696.

DIDAY E., 1986. *"Une représentation visuelle des classes empiétantes : les pyramides"*. R.A.I.R.O., Vol. 20, N°5: 475-526.

GUENOCHE A., 1989. *"Generalization and Conceptual Classification : Indices and an Algorithm"*. In Proceedings of the Conference on Data Analysis, Antibes, Learning Symbolic and Numerical Knowledge, E. Diday ed., Nova Science Publishers, Inc., New York: 503-510.

JAMBU M., 1978. *"Classification automatique pour l'analyse des données"*. Dunod, Paris.

LERMAN I.C., 1981. *"Classification et analyse ordinale des données"*. Dunod, Paris, 740 pp.

MICHALSKI R.S. and STEPP R.E., 1981 (a). *"An application of AI techniques to structuring objects into an optimal conceptual hierarchy"*. Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada.

MICHALSKI R.S. and STEPP R.E., 1981 (b). *"Revealing conceptual structure in data by inductive inference"*. Machine Intelligence, Vol. 10: 173-196.

MICHALSKI R.S. and STEPP R.E., 1983. *"Automated construction of classifications: Conceptual Clustering versus numerical taxonomy"*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-5, 396-410.

MICHALSKI R.S., 1980. *"Knowledge Acquisition Though Conceptual Clustering: A Theoretical Framework and an Algorithm for Partitioning Data into Conjunctive Concepts"*. International Journal of Policy Analysis and Information Systems, Vol. 4, N°3: 219-244.

MICHALSKI R.S., DIDAY E. and STEPP R.E., 1982. *"A recent advance in Data Analysis: Clustering objects into classes characterized by conjunctive concets"*. In Progress in Pattern Recognition, Vol 1, eds. Kanal L. et Rosenfeld A.

MURTAGH F., 1982. *"Fondements théoriques de la classification hiérarchique sous contrainte de contiguité continue"*. In Actes regroupés des Journées de Classification de Toulouse (mai 1980) et de Nancy (juin 1981), Lerman I.C. ed., IRISA, Rennes : 177-191.

PERRUCHET C., 1982. *"Classification sous contrainte de contiguité continue"*. In Actes regroupés des Journées de Classification de Toulouse (mai 1980) et de Nancy (juin 1981), Lerman I.C. ed., IRISA, Rennes : 192-207