LES CAHIERS DE L'ANALYSE DES DONNÉES

M. BRUYNOOGHE

Classification ascendante hiérarchique des grands ensembles de données : un algorithme rapide fondé sur la construction des voisinages réductibles

Les cahiers de l'analyse des données, tome 3, n° 1 (1978), p. 7-33

<http://www.numdam.org/item?id=CAD_1978__3_1_7_0>

Les Cahiers de l'Analyse des Données
Vol. III - 1978 - n° 1 - p. 7-33

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE DES GRANDS ENSEMBLES DE DONNÉES : UN ALGORITHME RAPIDE FONDÉ SUR LA CONSTRUCTION DES VOISINAGES RÉDUCTIBLES [CLAS. RAP.]

par M. Bruynooghe (1)

1 *Introduction : le traitement des grands ensembles de données* :

Avec l'effectif n de l'ensemble I dont on recherche la structure, croît nécessairement le temps requis pour l'analyse des données. Si par exemple on doit considérer toutes les permutations de l'ensemble I, le temps croîtra comme n! , i.e. selon la formule de Stirling comme $(2\pi n)^{1/2}(n/e)^n$: une telle croissance est tout à fait prohibitive. Depuis quelques années, la diagonalisation des matrices carrées symétriques a connu des progrès spectaculaires (cf e.g. TII B n° 12 § 4 ; TII D) : avec l'algorithme SYMQR le temps de diagonalisation d'une matrice n x n croît comme n^2. Tel est également l'ordre de croissance pour l'algorithme de E. Diday (nuées dynamiques = agrégation autour de centres variables), pourvu que le nombre des variables décrivant les individus à classer ait été préalablement réduit par l'analyse factorielle. Mais sous sa forme originelle, l'algorithme de classification ascendante hiérarchique (cf TI B n° 4 [C.A.H.] et *infra* § 2) requiert un temps qui croît comme n^3 : ce qui, avec les moyens de calculs disponibles en 1977, ne permet pas de dépasser un effectif de quelques centaines.

Cependant, on sait aujourd'hui construire un arbre de longueur minima qui croît comme n^2 : or (on l'a rappelé dans ces cahiers : cf [Squelette Arborescent] Vol I n° 4 pp 441 sqq.) cette construction équivaut à réaliser une classification ascendante hiérarchique particulière : avec agrégation suivant le saut minimum. Cette procédure d'agrégation est toutefois en butte à l'effet de chaînage (formation de classes filiformes) qui en restreint beaucoup l'utilité. Il était donc très souhaitable de réaliser dans des temps d'ordre n^2 une classification ascendante avec d'autres critères : notamment celui d'agrégation suivant la variance. C'est ce qui est fait ici : disons tout de suite que traitant sur l'ordinateur I B M 370-168 du CNRS un ensemble de 1561 objets caractérisés chacun par 7 facteurs (ceux-ci obtenus par l'analyse d'un tableau de correspondance 1561 x 70), nous avons obtenu en 38 secondes la hiérarchie totale binaire fondée sur le critère de la variance.

Dans la suite, nous montrons d'abord (§ 2) comment la méthode des *voisinages réductibles*, permet de ne pas considérer à chaque étape toutes les paires de noeuds ou individus, mais seulement une faible part d'entre elles, parmi lesquelles se trouve certainement celle réalisant l'écart minimum. Au § 3 cette méthode est appliquée à la construction de l'arbre de longueur minima, en suivant pas à pas sur un exemple simple le cheminement de l'algorithme. Au § 4 on présente de même une application à l'agrégation suivant la variance. Le § 5 montre les performances

(1) *Université d'Aix-Marseille II ; § C.R.E.T. : Centre de recherche en économie des transports ; Avenue Gaston Berger : 13100 Aix-en Provence.*

du nouvel algorithme. Quant à l'exemple auquel on a fait allusion ci-dessus - la classification des 1561 communes de la région Languedoc-Roussillon d'après leurs distances aux équipements urbains - il est publié à part dans ce même cahier (cf [COMMUNES] pp 35-46).

2 *La méthode des voisinages réductibles* :

2.1 *L'algorithme de C.A.H.* : rappelons en figurant un exemple, les notations usuelles et la marche de l'algorithme. L'ensemble I = {1,2,3, 4,5,6,7} est muni d'une distance (ou écart positif ne satisfaisant pas à

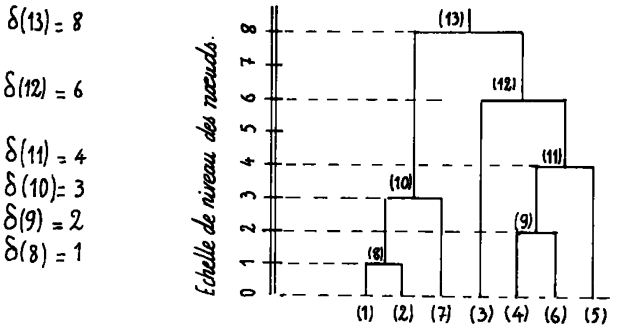


Figure 2-1: exemple de classification ascendante hiérarchique.

tous les axiomes d'une distance) $\delta(i,i')$; et pour calculer l'écart $\delta(u, v)$ entre deux parties quelconques u et v de I, on a fait choix d'une formule qui règle la procédure d'agrégation (sur la diversité des formules possibles, on s'arrêtera au § 2.2). Chacun des noeuds n de la classification est placé sur la figure à une altitude, ou niveau $\delta(n)$, qui n'est autre que l'écart entre les deux classes (ou individus) par agrégation desquels le noeud n a été créé : ainsi on a $\delta(8) = \delta(1,2) = \delta(\{1\},\{2\}) = 1$ (écart entre les individus 1 et 2 ; ou, ce qui revient au même entre les parties de I, {1} et {2}, réduites chacune à l'un de ces individus) ; $\delta(12) = \delta(3,11) = \delta(\{3\}, \{4,6,5\})$.

Pour décrire brièvement l'algorithme de classification ascendante nous utiliserons la notation d'*arbre non connexe* et d'ensemble de *sommets*. On note A l'arbre total :

$$A = \{1,2,3,4,5,6,7,8,9,10,11,12,13\} ,$$

(ici on a écrit en bref, 1 pour {1} = classe réduite à l'individu 1 ; 2 pour {2} , etc.)

En ôtant de A ceux de ses noeuds dont le niveau dépasse δ, on obtient un sous-arbre noté $A(\delta)$, lequel est non-connexe si $\delta < \delta(13)$.

$$A(\delta) = \{a \mid a \in A \ ; \ \delta(a) \leq \delta\} ;$$

e.g. dans notre exemple :

$A(0) = \{1,2,3,4,5,6,7\}$, ensemble des classes réduites à un élément ;

$A(3) = \{1,2,3,4,5,6,7,8,9,10\}$.

l'arbre A(0) n'est évidemment pas connexe ; il est formé de 7 branches isolées chacune réduite à un individu-classe ; l'arbre A(3) comprend 4 branches connexes 10,9,3,5, dont les deux dernières sont réduites à un

individu-classe, comme il est figuré ici. Si l'on appelle sommet le

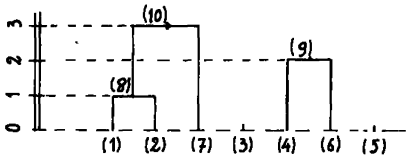


Figure 2.2 : exemple d'arbre non connexe $A_{10} = A(3)$.

noeud ou individu le plus haut d'une branche isolée on a pour ensemble des sommets des arbres non-connexes considérés ici :

$$\text{Som } A(0) = \{1,2,3,4,5,6,7\} \ ;$$

$$\text{Som } A(3) = \{3,5,9,10\}.$$

Enfin, on note encore A_n l'arbre non-connexe $A(\delta(n))$ obtenu en ôtant de A les noeuds dont le niveau dépasse $\delta(n)$, niveau du noeud n : (e.g. $A_{10} = A(\delta(10)) = A(3)$; et de même on écrira A_7 pour $A(0)$.

Ceci posé on peut formuler avec concision la construction ascendante de A à partie de $A_7 = A(0)$; en passant par A_8 , A_9 ..., jusqu'à $A_{13} = A$. On a :

$$\delta(h) = \inf\{\delta(s,s') \mid s,s' \in \text{Som } A_{h-1} \ ; \ s \neq s'\}$$

et en notant s_h, s'_h deux éléments de Som A_{h-1} réalisant ce minimum $\delta(h)$ de l'écart entre sommets de A_{h-1} , on a :

$$A_h = A_{h-1} \cup \{(s_h \cup s'_h)\} \ ;$$

$$\text{Som } A_h = (\text{Som } A_{h-1} \cup (s_h \cup s'_h)) - \{s_h, s'_h\} \ ;$$

i.e. A_h diffère de A_{h-1} en ce qu'est créé un nouveau sommet $s_h \cup s'_h$, obtenu par réunion de s_h et s'_h ; (lesquels étaient des sommets pour A_{h-1} , mais ne le sont plus pour A_h) ; e.g. A_{11} diffère de A_{10} par la création du sommet $11 = 9 \cup 5$; et l'on a $\text{Som } A_{11} = \{3,10,11\}$ tandis que $\text{Som } A_{10} = \{3,5,9,10\}$.

Pour passer de A_{h-1} à A_h , il faut reconnaître une paire de sommets s_h , s'_h réalisant dans Som A_{h-1} le minimum de l'écart ; et pour cela avoir calculé tous les écarts de ces sommets deux à deux ; à moins qu'on ne soit capable d'éliminer d'un seul coup un ensemble aussi grand que possible de paires par lesquelles ce minimum n'est certainement pas réalisé : c'est justement la construction, expliquée ci-dessous, des voisinages réductibles.

2.2 *L'axiome de réductibilité* : l'algorithme de classification ascendante requiert qu'à chaque itération on calcule l'écart du nouveau sommet créé, $s_h \cup s'_h$ aux sommets préexistants de Som A_{h-1}. Les écarts (ou indices de dissimilarité) communément utilisés dans les procédures d'agrégation satisfont à une formule de récurrence qui permet de calculer l'écart entre une classe t et la classe obtenue par réunion disjointe

des classes s et s'. On peut comprendre toutes les formules usuelles dans une formule générale

$$\delta(t,s \cup s') = \alpha(s)\ \delta(t,s) + \alpha(s')\ \delta(t,s') + \beta\delta(s,s') + \gamma|\delta(t,s) - \delta(t,s')| \ ;$$

formule dont les coefficients $\alpha(s)$, β, γ sont donnés ici pour quelques stratégies classiques ; dans le tableau on a noté n_s, $n_{s'}$, n_t l'effectif ou parfois la masse des classes s, s', t.

Stratégie	$\alpha(s)$	β	γ	références
saut minimal	1/2	0	-1/2	Sokal & Sneath 1963
diamètre min.	1/2	0	1/2	ibid Mc Quitty 1964
distance moy.	$n_s/(n_s+n_{s'})$	0	0	Sokal & Michener 1958 ; Mc Quitty
écart entre c. de g.	$n_s/(n_s+n_{s'})$	$-n_s n_{s'}/(n_s+n_{s'})^2$	0	Sokal § Michener 1958 ; Gower, 1967
variance min.	$(n_t+n_s)/(n_s+n_{s'}+n_t)$	$-n_t/(n_s+n_{s'}+n_t)$	0	Benzécri 1965 ; Wishart 1969 ; Anderson 1971 .

Tableau : formule de récurrence du calcul des écarts.

Ceci étant rappelé, on peut dire que notre méthode requiert en bref que les écarts faibles qui apparaissent par application de la formule donnant $\delta(t,s \cup s')$ proviennent d'écarts $\delta(t,s)$ $\delta(t,s')$ dont l'un au moins est faible : ainsi, si l'on a circonscrit au pas h-1 un ensemble de couples de sommets où sont tous les écarts faibles, cet ensemble fournira encore les écarts faibles du pas h pourvu que l'on substitue à s et s' supprimés leur union $s \cup s'$. Ceci sera précisé au § 2.4 : auparavant énonçons et vérifions un axiome qui précise comment les écarts faibles se combinent en des écarts faibles.

Axiome de réductibilité :

$$\forall\ s, s', t : \delta(s,s') \leq \inf\{\delta(s,t),\ \delta(s',t)\}$$

$$\Rightarrow \qquad \inf\{\delta(s,t),\ \delta(s',t) \leq \delta((s \cup s'),t) \ ;$$

l'axiome est vérifié pour les stratégies du saut minimal, du diamètre de la distance moyenne et de la variance. Faisons par exemple le calcul pour cette dernière ; on a

$$\delta(s \cup s',t) = (n_s+n_{s'}+n_t)^{-1}\left((n_t+n_s)\ \delta(s,t) + (n_t+n_{s'})\ \delta(s',t) - n_t\delta(s,s')\right) \ ;$$

en notant $r = \inf\{\delta(s,t),\ \delta(s',t)\}$ le plus petit des deux écarts $\delta(s,t)$ et $\delta(s',t)$, on voit que les trois termes (deux positifs, un négatif) dont se compose $\delta(s \cup s',t)$ peuvent être remplacés par des termes proportionnels à r qui leur sont inférieurs ou égaux *en valeur algébrique*, d'où :

$$\delta(s \cup s',t) \geq (n_s+n_{s'}+n_s)^{-1}\left((n_t+n_s)r + (n_t+n_{s'})r - n_t r\right) = r \ ;$$

ce qui est bien dans l'énoncé de l'axiome. On notera au passage que l'axiome de réductibilité assure que la stratégie ne présente pas d'inversion : i.e. que l'on n'aura jamais entre $s_h \cup s'_h$ et un sommet t

préexistant un écart moindre que celui entre s_h et s'_h eux-mêmes, qui viennent d'être agrégés pour former le nouveau sommet (en effet on a certes $\delta(s,s') \leq \inf\{\delta(s,t), \delta(s',t)\}$, puisque s et s' ont été agrégés ; sinon on aurait agrégé s ou s' à t ; d'où $\delta(s \cup s',t) \geq$ inf). En particulier les stratégies d'agrégation selon la distance entre centres de gravité des classes ou selon la distance angulaire ne satisfont pas à l'axiome, car elles peuvent engendrer une hiérarchie binaire munie d'un indice de niveau qui présente des inversions (pour la distance angulaire cf [C.A.H.] § 2.4.3).

2.3 *Voisinage d'ordre ρ* : Soit S un ensemble muni d'une distance (ou simplement d'un écart) que nous noterons δ (puisque nous avons en vue l'écart entre parties de l'ensemble I des individus à classer ; plus précisément entre les sommets d'un arbre non-connexe). On dit qu'un ensemble U de paires d'éléments de S (une telle paire sera encore appelée arête) est un voisinage d'ordre ρ ou ρ-voisinage si U contient toute arête (s,s') telle que $\delta(s,s') <_s \rho$ et ne contient aucune arête telle que $\rho <_s \delta(s,s')$ (ici $<_s$ signifie inférieur strictement ; il n'importe pas que les paires (s, s') avec $\delta(s,s') = \rho$ soient ou non dans U). Le terme de ρ-voisinage s'explique parce que par les arêtes de U chaque élément de S est relié à tous les autres éléments qui voisinent avec lui à une distance inférieure à ρ. Dans la suite on notera X l'ensemble des sommets (éléments de S) intervenant dans l'une ou l'autre des arêtes de U ; on notera G = (X , U), comme un graphe le système des sommets et des arêtes du ρ-voisinage et G sera appelé parfois *graphe de similarité*. En formules on peut écrire :

$$\forall\, s,s' \in S : \quad \delta(s,s') <_s \rho \Rightarrow (s,s') \in U$$

$$\delta(s,s') >_s \rho \Rightarrow (s,s') \notin U$$

$$X = \{s \mid s \in S \ ; \ \exists\, s' : (s,s') \in U\}.$$

Pratiquement, la capacité de la mémoire centrale de l'ordinateur utilisé, nous contraindra à ne considérer que des ρ-voisinages tels que

$$\text{Card } X \leq N \quad ; \quad \text{Card } U \leq L$$

Les entiers N et L désignant des bornes fixées. Etant donné S (muni de l'écart ρ), on choisira donc un nombre ρ aussi grand que possible, tel qu'existe un ρ-voisinage G = (X,U) satisfaisant aux contraintes Card X ≤ N , Card U ≤ L. Un tel choix requiert principalement le calcul des distances entre éléments de S, et la constitution d'une sorte d'histogramme de celles-ci : c'est en bref une opération dont l'ordre de durée est $(\text{Card } S)^2$ (sur ce choix, cf infra, § 3.3. Phase 1 et § 4.3 phase 1).

On voit aisément que le minimum de l'écart entre éléments de S est réalisé pour une arête au moins de tout ρ-voisinage non-vide (i.e. tel que $U \neq \emptyset$). En effet soit (s,s') ∈ U : si δ(s,s') est le min. de l'écart entre points de S , ce minimum est bien réalisé dans U : et sinon, U contient certainement (de par sa définition) toute arête de S strictement plus courte que (s,s'), donc toutes celles réalisant le min.. Ainsi si l'on connaît un ρ-voisinage de S , la recherche d'une arête minimale, étape indispensable à la construction ascendante d'une hiérarchie, s'effectuera plus vite : au lieu de considérer toutes les arêtes de S on se bornera à celles de U. L'axiome de réductibilité du § 2.2 nous permettra de ne considérer dans le processus ascendant que des ρ-voisinages dont la plupart sont construits très simplement comme on l'explique ci-dessous.

2.4 *L'algorithme des voisinages réductibles* : L'Algorithme de [C.A.H.] rappelé au § 2.1 procède itérativement en une suite d'étapes dont l'indice h varie de Card I + 1 à 2 Card I - 1 (de 8 à 13 dans l'exemple). A l'étape h on considère l'ensemble S_{h-1} = Som A_{h-1} des sommets de l'arbre (non-connexe) A_{h-1} : on détermine une paire (s_h , s'_h) réalisant le min. de l'écart, d'où construction d'un nouveau noeud $s_h \cup s'_h$ qui reçoit le numéro h. Dans l'ordinateur les informations à véhiculer sont les suivantes. Chaque noeud est décrit par les numéros (A[h] et B[h]dans les notations de [C.A.H.]) des classes s_h et s'_h par réunion desquelles il est obtenu ; ainsi que le niveau $\delta(h) = \delta(s_h , s'_h)$ auquel s_h et s'_h ont été agrégés. De plus il faut tenir à jour un tableau DIS donnant les écarts deux à deux de tous les sommets, ou éléments de l'ensemble S_{h-1}. C'est ce tableau DIS qui est le plus encombrant, et dont le traitement (choix de la paire s_h , s'_h) est le plus coûteux.

Dans l'algorithme des voisinages réductibles, on considère à l'étape h non S_{h-1} tout entier avec toutes ses arêtes (dont les longueurs sont dans le tableau DIS) mais seulement un ρ-voisinage $G_{h-1} = (X_{h-1} , U_{h-1})$ lequel on l'a dit au § 2.3, contient certainement une arête (s_h , s'_h) convenable : ce qui constitue à la fois une économie de place (U_{h-1} est moins encombrant que le tableau DIS) et une économie de temps (la recherche de (s_h , s'_h) est accélérée). Cette économie serait toutefois illusoire s'il fallait à chaque étape h construire le voisinage G_{h-1} sans bénéficier des traitements antérieurs : ici intervient l'axiome de réductibilité qui permet de construire simplement G_h à partir de G_{h-1}. Voici comment .

Soit $G_{h-1} = (X_{h-1} , U_{h-1})$ un ρ-voisinage de S_{h-1} : alors on prend pour arêtes de G_h toutes les arêtes (s,s') de U_{h-1} dont les extrémités sont distinctes de s_h et s'_h ; et de plus à toute arête de la forme (s_h , t) ou (s'_h , t) on substitue l'arête $(s_h \cup s'_h , t)$ si toutefois celle-ci ne dépasse pas ρ. Il est clair qu'on a bien ainsi un ρ-voisinage sur S_h : en effet si une arête plus courte que ρ a des extrémités distinctes de s_h et s'_h elle est déjà dans U_{h-1} et subsiste dans U_h ; et si une nouvelle arête $(s_h \cup s'_h , t)$ est plus courte que ρ l'axiome de réductibilité nous assure que l'une au moins des arêtes (s_h , t) et (s'_h , t) est inférieure à ρ, donc figure dans U_{h-1} , et qu'à partir de celle-ci on a inclus $(s_h \cup s'_h)$ dans U_h. On dit que U_h est construit à partir de U_{h-1} par réduction : d'où le terme de voisinage réductible.

Il y a cependant un cas où ce processus de récurrence tombe en défaut : il se peut que U_h soit vide ! en effet à chaque pas Card U diminue : car d'une part disparaît l'arête (s_h , s'_h) d'autre part souvent pour deux arêtes (s_h , t) et (s_h , t') figurant dans U_{h-1} il n'y en a plus qu'une $(s_h \cup s'_h , t)$ dans U_h. Quand le U_h déterminé à partir de U_{h-1} est vide, il faut pour construire un voisinage sur S_h procéder directement, ce qui est plus coûteux (comme on l'a dit au § 2.3).

En somme l'algorithme procède comme suit. Au départ $S = S(0) = S_{Card I-1}$ ($S(0) = S_7$ dans l'exemple de la figure 2.1) n'est autre que I lui-même, pour lequel on doit déterminer directement un ρ-voisinage non vide G. Le choix

de ρ qu'on appellera *seuil de stratification* est délicat : on y a fait allusion au § 2.3 ; on y reviendra dans les applications § 3.3 phase 1 et § 4.3 phase 1.

A partir de là, on construit aisément par réduction la suite des triplets $\{S_h , U_h , X_h\}$, jusqu'à aboutir à un U_h vide. De façon précise cela se produit lorsque le prochain noeud à construire a_{h+1}, est à un niveau $\delta(a_{h+1}) = \delta(s_{h+1} , s'_{h+1})$ supérieur à ρ : il est clair qu'alors la paire (s_{h+1} , s'_{h+1}) ne peut être dans U_h ; on comprend pourquoi ρ a été appelé *seuil de stratification*. Il faut réinitialiser l'algorithme itératif en déterminant directement un nouveau ρ-voisinage. L'efficacité de l'algorithme fait l'objet du § 5 : mais on peut dès maintenant en donner un exemple : dans le cas déjà cité au § 1 de la classification des 1561 communes du Languedoc-Roussillon, l'algorithme des voisinages réductibles a calculé $1{,}59\ 10^6$ écarts entre sommets pour édifier une hiérarchie selon le critère de la variance : pour édifier la même hiérarchie, l'algorithme de C.A.H. en aurait calculé $0{,}56\ 10^9$, soit 350 fois plus !

Dans les deux paragraphes qui suivent, on suit pas à pas sur des exemples simples le cheminement de l'algorithme des voisinages réductibles.

3 *Application à la classification suivant le critère du saut minimum et à l'arbre de longueur minimale*

3.1 *Arbre de longueur minimum et agrégation suivant le saut minimum*

L'arbre de longueur minimum (i.e. le sous-graphe sur I ayant tout point de I pour sommet et dont les arêtes ont une longueur totale minimum) est parfois utilisé pour schématiser la structure d'un nuage de points muni d'une distance (ROSS 1969). On peut figurer l'arbre de longueur minimum sur les graphiques plans issus de l'analyse factorielle des correspondances, lorsque le nombre d'individus représentés graphiquement ne dépasse pas quelques dizaines. Dans le cas d'un grand ensemble d'individus, après avoir obtenu une partition de l'ensemble en ne conservant que la partie supérieure de l'arbre de classification,on associera à chaque classe un point dans l'espace des facteurs, et on représentera l'arbre de longueur minimum qui relie l'ensemble de ces points. L'image simplifiée de la structure des données, fournie par l'arbre de longueur minimum peut confirmer l'interprétation des facteurs de l'analyse factorielle qui construit une représentation spatiale des données multidimensionnelles.

De plus, comme on l'a rappelé au § 1 la construction de l'arbre de longueur minimum équivaut à celle de la classification hiérarchique avec pour critère le saut minimum. Les progrès réalisés dans l'une de ces constructions servent donc naturellement à l'autre.

Gower et Ross (1969) ont montré que l'information nécessaire à l'analyse hiérarchique selon le critère du lien minimum est contenue dans l'arbre de longueur minimum du graphe de similarité. L'utilisation de l'algorithme de PRIM permet alors l'analyse arborescente d'un grand ensemble de données, selon le critère du lien minimum, en un temps de calcul qui varie comme le carré du nombre d'éléments à classer et non plus comme le cube avec l'algorithme classique de classification ascendante hiérarchique.

De plus l'algorithme de PRIM, n'utilise qu'une seule fois chaque

indice de distance entre deux objets et il n'est donc pas nécessaire de mettre en mémoire la matrice des distances entre objets, dont la taille est proportionnelle au carré du nombre d'objets à classifier. Il suffit de calculer les distances entre objets à partir du tableau des données ou du tableau des facteurs et l'espace mémoire nécessaire est proportionnel au nombre d'éléments à classer et non plus au carré comme dans le cas de l'algorithme classique de classification ascendante hiérarchique.

Nous verrons ici d'abord en général (§ 3.2) puis sur un exemple (§ 3.3) comment la méthode des voisinages réductibles permet d'accélérer encore ces constructions.

3.2 *Construction par l'algorithme des voisinages réductibles*

L'application de l'algorithme des voisinages réductibles, au cas particulier de la stratégie d'agrégation selon le saut minimal, permet de construire rapidement l'arbre de longueur minimum du graphe de similarité complet, dont les arêtes représentent tous les couples d'objets de l'ensemble à classifier. Les arêtes de l'arbre minimum sont choisies dans le même ordre qu'avec l'algorithme de KRUSKAL et l'algorithme proposé est plus rapide que celui de KRUSKAL puisqu'il n'est pas nécessaire de vérifier à chaque étape que l'on ne forme pas de cycles avec les arêtes déjà choisies.

La méthode des voisinages réductibles engendre une hiérarchie selon le critère du saut minimal et un arbre de longueur minimum du graphe de similarité, en construisant une suite(*) d'arbres binaires A_0 , A_1 ... A_h ... $A_{|I|-1}$, une suite de graphes G_0 , G_1 , ... G_h , ... $G_{|I|-1}$ et une suite d'arbres T_0 , T_1 , ... T_h , ... $T_{|I|-1}$, tels que :

* $A_{|I|-1}$ = A (A est la hiérarchie totale binaire de parties sur l'ensemble I des objets à classer selon le critère du saut minimal)

* $T_{|I|-1}$ est l'arbre de longueur minimum du graphe de similarité.

Initialisation Faire h = 0

Soit $A_0 = I$, $T_0 = \emptyset$

Pas 0 : Détermination d'un seuil de stratification ρh et construction du graphe de similarité $G_h = (X_h , U_h)$ tel que $|X_h| \leq N$ et $|U_h| \leq L$

$\forall s \in X_h , s' \in X_h , (s, s') \in u_h$ noter $u(s,s') = (i,i')$ l'arête de longueur minimale qui relie les classes s et s' dans le graphe de similarité initial sur l'ensemble I des objets à classifier.

Pas 1 : Recherche des deux sommets les plus proches

$h = h + 1$

$\delta(s_h , s'_h) = \inf \{\delta(s,s') | s,s' \in X_{h-1} , (s,s') \in U_{h-1}\}$

Pas 2 : Construction de l'arbre binaire A_h à partir de l'arbre binaire A_{h-1} et de l'arbre T_h à partir de l'arbre T_{h-1}.

$a_h = s_h \cup s'_h$

$\tau(a_h) = \delta(s_h , s'_h)$; (ici et dans la suite $\tau(a_h)$ désigne le niveau du noeud a_h).

(*) *Au § 2.4, le numérotage est fait, comme dans [C.A.H.] de Card I à 2 Card I - 1 ; ici on va de 0 à |I|-1 = Card I - 1.*

$A_h = A_{h-1} \cup \{a_h\}$; $Som(A_h) = Som(A_{h-1}) \cup \{a_h\} - \{s_h , s'_h\}$

$(i_h , i'_h) = u(s_h , s'_h)$ et $T_h = T_{h-1} \cup \{(i_h , i'_h)\}$

Pas 3 : Test d'arrêt

Si $h = |I| - 1$ FIN
Sinon, aller au Pas 4

Pas 4 : Construction du graphe G_h à partir du graphe G_{h-1}

Faire $\rho_h = \rho_{h-1}$;
$X_h = X_{h-1} + \{a_h\} - \{s_h , s'_h\}$;

Notons $X(a_h) = \{t \mid t \in X_{h-1} - \{s_h , s'_h\}$; $(s_h , t) \in U_{h-1}$ ou $(s'_h , t) \in U_{h-1}\}$

(i.e. $X(a_h)$ est l'ensemble des t reliés à s_h ou s'_h dans G_{h-1}).

$U_h = U_{h-1} \cup \{(a_h , t) \mid t \in X(a_h)\}$

$-\{(t',t) \mid (t',t) \in U_{h-1}$; $(t \in \{s_h , s'_h\}$ ou $t' \in \{s_h , s'_h\})\}$;

(on supprime les arêtes issues de s_h ou s'_h ; on introduit les arêtes reliant a_h à $t \in X(a_h)$) .

si $U_h = \emptyset$ aller au Pas O.

Calculer $\delta(a_h , t)$ pour $\forall\, t \in X(a_h)$ suivant la formule :

$\delta(a_h , t) = \inf\{\delta(s_h , t) ; \delta(s'_h , t)\}$;

où l'on convient de poser $\delta(s_h , t)$ (ou $\delta(s'_h , t)$) infini, si (s_h , t) (ou (s'_h , t)) n'est pas dans U_{h-1} ; car alors $\delta(s_h , t)$ (ou $\delta(s'_h , t)$) excède ρ_h.

$$u(t,a_h) = \begin{cases} u(t,s_h) & \text{si } \delta(t,a_h) = \delta(t,s_h) \\ & \text{; pour } \forall\ t \in X(a_h). \\ u(t,s'_h) & \text{si } \delta(t,a_h) = \delta(t,s'_h) \end{cases}$$

Aller ensuite au pas 1

L'algorithme présenté ci-dessus ne calcule pas les nouveaux indices de dissimilarité $\delta(t,a_h)$ à partir du tableau des données, mais à partir des indices définis dans le graphe de similarité construit après la dernière modification du seuil de stratification.

L'intérêt de la méthode des voisinages réductibles pour construire une classification hiérarchique selon le critère du lien minimum, est de ne nécessiter ni la mise en mémoire du tableau triangulaire des indices de dissimilarité entre objets, ni celle du tableau des données ou du tableau des facteurs.

La justification de l'algorithme de construction de l'arbre de longueur minimum fait l'objet des lemmes suivants : (M. Bruynooghe 1977-a)

Lemme 1 : L'indice de dissimilarité entre deux sommets t et t' du graphe $G_h = (X_h , U_h)$ est égal à la longueur de la plus petite arête qui relie les deux classes t et t' dans le graphe de similarité complet sur l'ensemble I des objets.

Lemme 2 : Pour tout $h \in [1,2,\ldots|I|-1]$, $(i_h , i'_h) = u(s_h , s'_h)$ est l'arête la plus courte qui ne forme pas de cycles avec les arêtes déjà choisies.

Lemme 3 : $T_{|I|-1} = \{(i_h , i'_h) \mid h = 1,2,\ldots|I|-1\}$ est l'arbre de longueur minimum du graphe de similarité complet sur I.

La longueur de l'arbre minimum est égale à :

$L(T_{|I|-1}) = \Sigma\{\delta(i_h , i'_h) \mid h = 1,2,\ldots|I|-1\}$

3.3 *Un exemple numérique simple*

Soit un ensemble de 9 objets représentés dans un plan par des points dont les coordonnées sont dans le tableau suivant :

Point	Abscisse	Ordonnée
	F_1	F_2
a	-23	-16
b	-41	- 7
c	-29	7
d	- 1	14
e	4	- 6
f	18	19
g	20	4
h	19	-14
i	33	- 1

D'où un tableau de distances, calculé avec la métrique euclïdienne classique

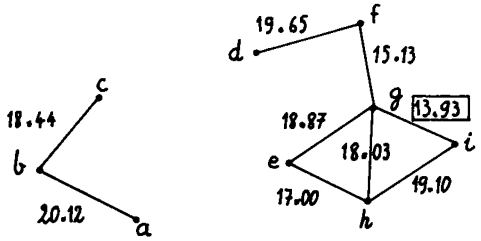
	a	b	c	d	e	f	g	h	i
a	0								
b	20.12	0							
c	23.77	18.44	0						
d	37.20	45.18	28.86	0					
e	28.79	45.00	35.47	20.62	0				
f	53.91	64.47	48.51	19.65	28.65	0			
g	47.42	61.98	49.09	23.26	18.87	15.13	0		
h	42.04	60.41	52.39	34.41	17.00	33.01	18.03	0	
i	57.97	74.24	62.51	37.16	29.43	25.00	13.93	19.10	0

PHASE I CHOIX DU SEUIL DE STRATIFICATION INITIAL

Soit v(i) le voisin le plus proche de l'objet i. On pose
$\rho_0 = \sup \{ \delta(i,v(i)) \mid i \in I \}$

Objet	Voisin le plus proche	Indice de dissimilarité
a	b	20.12
b	c	18.44
c	b	18.44
d	f	19.65
e	h	17.00
f	g	15.13
g	i	13.93
h	e	17.00
i	g	13.93

On a donc $\rho_0 = \delta(a,b) = 20.12$

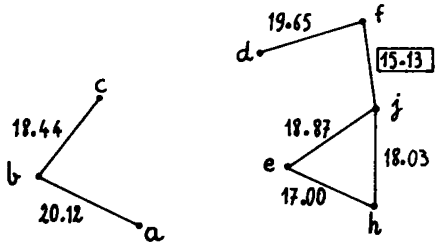


Graphe $G_0 = (X_0 , U_0)$

ETAPE I : Agrégation de g et i

$j = \{g\} \cup \{i\}$; on a en notant $\alpha(j)$ et $\beta(j)$ les deux descendants de j (aîné et benjamin) ; $\tau(j)$ le niveau de j ; et u(g,i) l'arête de longueur minimale reliant les deux classes g et i (ici toutes deux réduites à un point !)

$\alpha(j) = g \quad \beta(j) = i \quad \tau(j) = 13.93 \quad u(g,i) = (g,i)$

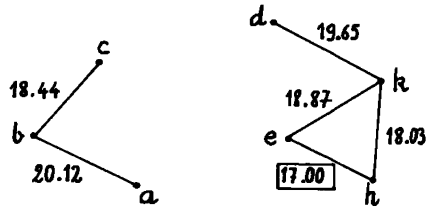


Graphe $G_1 = (X_1 , U_1)$

ETAPE II : Agrégation de f et j

$k = \{f\} \cup \{j\}$

$\alpha(k) = f \quad \beta(k) = j \quad \tau(k) = 15.13 \quad u(f,j) = (f,g)$

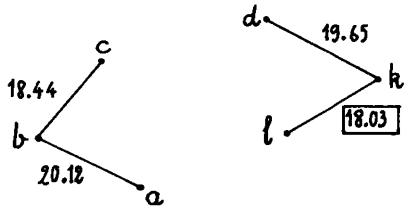


Graphe $G_2 = (X_2 , U_2)$

ETAPE III : Agrégation de e et h

$\ell = \{e\} \cup \{h\}$

$\alpha(\ell) = e \quad \beta(\ell) = h \quad \tau(\ell) = 17.00 \quad u(e,h) = (e,h)$

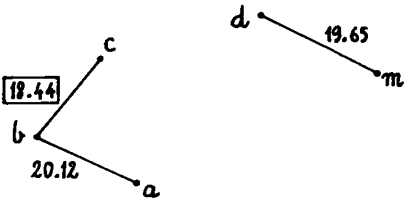


Graphe $G_3 = (X_3 , U_3)$

ETAPE IV : Agrégation de k et ℓ

$m = \{k\} \cup \{\ell\}$

$\alpha(m) = k \quad \beta(m) = \ell \quad \tau(m) = 18.03 \quad u(k,\ell) = (g,h)$

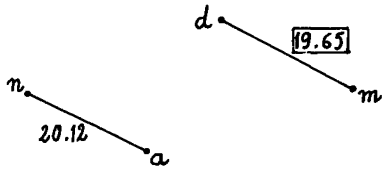


Graphe $G_4 = (X_4 , U_4)$

ETAPE V : Agrégation de b et c

$n = \{b\} \cup \{c\}$

$\alpha(n) = b \quad \beta(n) = c \quad \tau(n) = 18.44 \quad u(b,c) = (b,c)$



Graphe $G_5 = (X_5 , U_5)$

ETAPE VI : Agrégation de d et m

$O = \{d\} \cup \{m\}$

$\alpha(O) = d \quad \beta(O) = m \quad \tau(O) = 19.65 \quad u(d,m) = (d,f)$

Graphe $G_6 = (X_6 , U_6)$

ETAPE VII : Agrégation de a et n

$p = \{a\} \cup \{n\}$

$\alpha(p) = a \quad \beta(p) = n \quad \tau(p) = 20.12 \quad u(a,n) = (a,b)$

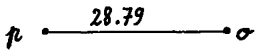
A la fin de cette étape, l'ensemble des arêtes est vide et il est nécessaire de redéfinir la valeur du seuil de stratification.

PHASE II DEFINITION DU SECOND SEUIL DE STRATIFICATION

On pose $\rho_7 = \text{Sup}\{\delta(s,v(s)) \mid s \in X_7\}$

avec $X_7 = \{O,p\}$

Soit $\rho_7 = \delta(O,p) = 28.79$ (longueur de l'arête la plus courte qui relie les classes $O = \{d,e,f,g,h,i\}$ et $p = \{a,b,c\}$)



Graphe $G_7 = (X_7 , U_7)$

Etape VIII : Agrégation de O et p

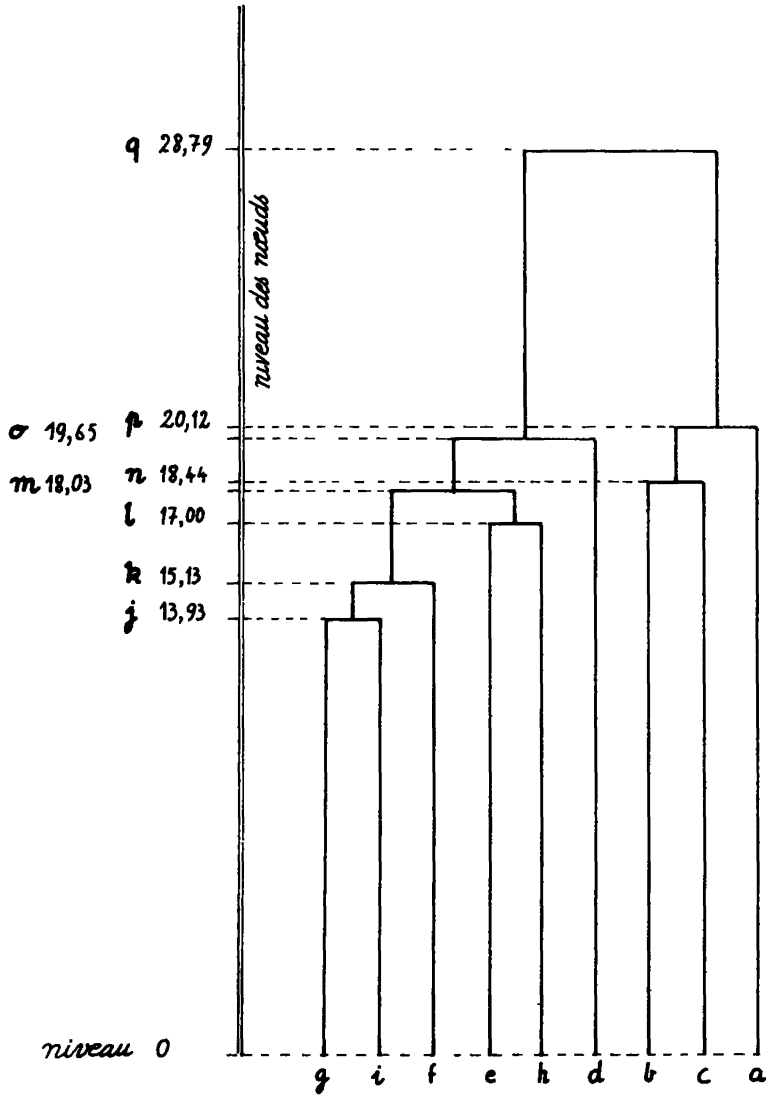
$q = \{O\} \cup \{p\}$

$\alpha(q) = O \quad \beta(q) = p \quad \tau(q) = 28.79 \quad u(O,p) = (a,e)$

q

Graphe $G_8 = (X_8 , U_8)$
(réduit au seul point isolé q)

Le tableau suivant récapitule les résultats obtenus au cours de la procédure de classification, selon le critère du saut minimal.



Hiérarchie engendrée selon le critère du saut minimal.

4 Classification selon le critère de la variance, par la méthode des voisinages réductibles

4.1 Analyse factorielle et classification hiérarchique

Supposons que l'ensemble sur lequel est édifiée la classification hiérarchique binaire soit un nuage de points munis de la distance distributionnelle du χ^2, soit k_{IJ} le tableau de description des éléments à classer ; chaque élément i étant identifié à son profil f^i_J muni de la masse f_i, la distance prise entre deux éléments i et i' est la distance distributionnelle du χ^2 de centre f_J entre les profils f^i_J et $f^{i'}_J$

$$d^2(i,i') = \| f^i_J - f^{i'}_J \|^2_{f_J} = \Sigma\{(f^i_j - f^{i'}_j)^2/f_j \mid j \in J\}$$

La distance du χ^2 entre les éléments i et i' peut se calculer à partir des facteurs suivant la formule :

$$d^2(i,i') = \Sigma\{(F_n(i) - F_n(i'))^2 \mid n = 1,2,\ldots, p\}$$

Cette formule est mathématiquement exacte si la somme est étendue à l'ensemble de tous les facteurs F_α, qui sont dans R_J muni de la métrique du χ^2 un système de coordonnées orthonormées (cf J.P. Benzécri,1973). Mais on peut avantageusement se borner aux premiers facteurs qui ont été trouvés significatifs. Ainsi, en classification automatique, on ne gardera en mémoire qu'une description réduite des objets, puis des centres de gravité des classes, limitée aux premiers facteurs. On calculera alors, à la demande, les valeurs des distances à partir du tableau des facteurs.

Dans l'algorithme de [C.A.H.] la matrice des écarts qui figure en mémoire est mise à jour à chaque étape h par une formule itérative variant suivant la procédure d'agrégation comme il a été rappelé au § 2.2 (cf tableau). Il est cependant possible de calculer les écarts relatifs aux nouveaux noeuds directement pourvu que l'on connaisse les coordonnées de ceux-ci (cf TI B n° 5 [Inf. Tab.] § 2.3.2 Remarque) : c'est ce qu'il faut faire quand on applique la méthode des voisinages réductibles. On posera donc, dans le cas de l'agrégation suivant la variance :

$$\delta(s,s') = (m_s\, m_{s'}/(m_s + m_{s'}))\, \| s - s' \|^2 \ ;$$

dans cette formule s et s' sont les vecteurs des centres des classes : vecteurs dont les composantes sont les coordonnées F (s) sur les axes factoriels (ou plus généralement des coordonnées orthonormées d'origine quelconque).

4.2 L'algorithme rapide de classification hiérarchique selon le critère de la variance

Donnons maintenant la description du nouvel algorithme, particularisé au cas de la méthode d'agrégation selon la variance, la métrique de base étant celle du χ^2, les distances entre objets étant calculées à partir du tableau des facteurs issus d'une analyse factorielle des correspondances. Toutes les phases de l'algorithme rapide de classification selon le critère de la variance sont détaillées ci-après, étant donné l'intérêt de la stratégie d'agrégation selon la variance.

Initialisation Faire h = 0

$A_0 = \{\{i\} \mid i \in I\}$

$\forall\, i \in I : \tau\{i\} = 0$

Pas 0 : Détermination d'un seuil de stratification ρ_h et construction du graphe de similarité $G_h = (X_h , U_h)$ tel que $|X_h| \leq N$ et $|U_h| \leq L$

L'indice de dissimilarité (ou écart) entre deux classes s et s' de $Som(A_h)$ est calculé par la formule suivante :

$$\delta(s,s') = (f_s \, f_{s'} / (f_s + f_{s'})) \, \Sigma\{(F_n(s) - F_n(s'))^2 \mid n = 1,2,\ldots,p\};$$

où p est le nombre de facteurs significatifs (ou plus généralement des coordonnées) utilisés pour édifier la classification ascendante hiérarchique.

Le calcul de $\delta(s,s')$ est interrompu dans la sommation sur l'ensemble des p facteurs, dès que la somme partielle calculée jusqu'à un certain rang $p_o \leq p$ est supérieur au seuil de stratification ρ_h : en effet on ne conservera que les arêtes inférieures ou égales à ρ_h.

Aller à la phase 1.

Pas 1 : Agrégation des deux parties disjointes les moins dissemblables

$h = h + 1$

$$\delta(s_h , s_{h'}) = \inf\{\delta(s,s') \mid s,s' \in X_{h-1} , (s,s') \in U_{h-1}\}$$

Pas 2 : Construction de l'arbre A_h à partir de l'arbre A_{h-1}

$a_h = s_h \cup s'_h$;

$A_h = A_{h-1} \cup \{a_h\}$;

$Som(A_h) = Som(A_{h-1}) \cup \{a_h\} - \{s_h , s'_h\}$;

Calcul des masses et facteurs (pour n = 1,..., p) :

$f_{a_h} = f_{s_h} + f_{s'_h}$;

$F_n(a_h) = [f_{s_h} . F_n(s_h) + f_{s'_h} . F_n(s'_h)] / f_{a_h}$;

le nouveau sommet créé a_h a été placé sur les axes factoriels comme barycentre des deux sommets dont il est la réunion disjointe.

Pas 3 : Test d'arrêt

Si $h = |I| - 1$, alors $A_{|I|-1} = A$ FIN

Sinon, aller à la phase 4.

Pas 4 : Construction du graphe G_h à partir du graphe G_{h-1}

$\rho_h = \rho_{h-1}$

$X_h = X_{h-1} + \{a_h\} - \{s_h , s'_h\}$

Calculer $\delta(t,a_h)$ pour tout sommet t de l'ensemble $X(a_h)$ défini comme au Pas 4 du § 3.2 (i.e. ensemble des t reliés à s_h ou à s'_h dans G_{h-1}) :

Comme au Pas 0, le calcul de l'indice de dissimilarité $\delta(t,a_h)$ est interrompu dès que la somme partielle calculée jusqu'à un certain rang $p_o \leq p$ est supérieure au seuil ρ_h.

Déterminer $V(a_h , \rho_h) = \{t | t \in X(a_h) \quad , \delta(t,a_h) \leq \rho_h\}$
et $U_h = U_{h-1} - \{(t,s_h) | t \in V(s_h , \rho_h)\} - \{(t,s'_h) | t \in V(s'_h , \rho_h)\} + \ldots$
$+ \ldots\{(t,a_h) | t \in V(a_h , \rho_h)\}$

soit $U_h = U_{h-1} \cup \{(a_h , t) \mid t \in X(a_h) \; ; \; \delta(a_h , t) \leq \rho_h\}$
$- \{(t,t') | (t,t') \in U_{h-1} \; ; \; (t \in \{s_h , s'_h\}$ ou $t' \in \{s_h , s'_h\})\}$;

i.e. on supprime les arêtes issues de s_h ou s'_h ; on introduit les arêtes issues de a_h et aboutissant à un t relié dans G_{h-1} à s_h ou à s'_h (i.e. selon la notation déjà posée au § 3.2 Pas 4 : $t \in X(a_h)$) pourvu que l'écart $\delta(a_h ,t)$ n'excède pas ρ_h. Comme au Pas 0, le calcul de $\delta(a_h,t)$ est interrompue dès que la somme partielle

$$\delta(a_h ,t) = (f_t \; f_{ah}/(h + f_{ah})) \; \Sigma\{(F_n(a_h) - F_n(t))^2 | n = 1, \ldots, p\},$$

excède ρ_h.

Si $U_h = \emptyset$ Aller au Pas 0
Sinon, aller au Pas 1

4.3 *Un exemple numérique simple*

Soit l'ensemble des 9 objets, représentés par des points situés dans un plan muni de la métrique euclidienne, considéré précédemment au § 3.3 pour illustrer la procédure de construction d'une hiérarchie selon le critère du lien minimal, par la méthode des voisinages réductibles.

Supposant que les masses des objets soient toutes égales à 2, le tableau suivant donne les écarts entre les points de l'ensemble à classifier selon le critère de la variance : $\delta(i,i') = (m_i m_{i'}/(m_i+m_{i'}))\|i-i'\|^2$; c'est-à-dire, puisque toutes les masses valent 2 , $\delta(i,i') = \|i - i'\|^2$.

	a	b	c	d	e	f	g	h	i
a	0								
b	405	0							
c	565	340	0						
d	1384	2041	833	0					
e	829	2026	1258	425	0				
f	2906	4157	2353	386	821	0			
g	2249	3842	2410	541	356	229	0		
h	1768	3649	2745	1184	289	1090	325	0	
i	3361	5512	3908	1381	866	625	194	365	0

PHASE I CHOIX DU SEUIL DE STRATIFICATION INITIAL

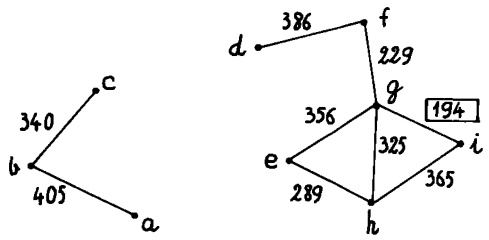
La hiérarchie engendrée par la procédure de classification est indépendante du choix des différents seuils de stratification. On peut envisager différentes variantes pour choisir un seuil de stratification initial ρ_0.

Ici, on calculera l'écart entre chaque objet i et son voisin le plus proche v(i) pour connaître l'ordre de grandeur des écarts entre objets susceptibles d'être réunis et on pose :

$$\rho_0 = \text{Sup}\{\delta(i,v(i)) \mid i \in I\}$$

Objet	Voisin le plus proche	Ecarts
a	b	405
b	c	340
c	b	340
d	f	386
e	h	289
f	g	229
g	i	194
h	e	289
i	g	194

On a donc $\rho_0 = \delta(a,b) = 405$



Graphe $G_0 = (X_0 , U_0)$

ETAPE I : Agrégation de g et i

$$j = \{g\} \cup \{i\}$$

$$\alpha(j) = g \qquad \beta(j) = i \qquad \tau(j) = 194$$

$$M^2(j) = 194 \ ; \ m_j = 4 \ ; \ V^2(j) = 48.5 \ ; \ V(j) = 6.96$$

(comme en [C.A.H.] § 2.5 on note M^2 le moment d'ordre 2 et V^2 la variance du noeud j ; i.e. du nuage formé des deux centres $\alpha(j)$ et $\beta(j)$) .

Calcul des coordonnées de f

$$F_1(j) = 26.50 \qquad F_2(j) = 1.50$$

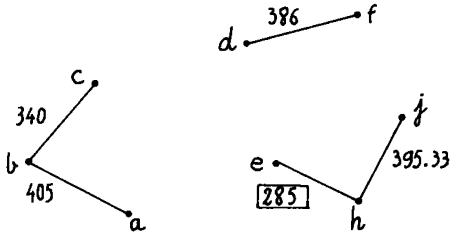
Calcul des écarts suivants :

$$\delta(j,e) = 750.00$$

$$\delta(j,f) = 504.66$$

$$\delta(j,h) = 395.33$$

Elimination des arêtes (j,e) et (j,f) dont les écarts sont supérieurs au seuil de stratification initial.



Graphe $G_1 = (X_1 , U_1)$

ETAPE II : Agrégation de e et h

$k = \{e\} \cup \{h\}$;

$\alpha(k) = h$; $\beta(k) = e$; $\tau(k) = 289$;

$M^2(k) = 289$; $m_k = 4$; $V^2(k) = 72.25$; $V(k) = 8.50$;

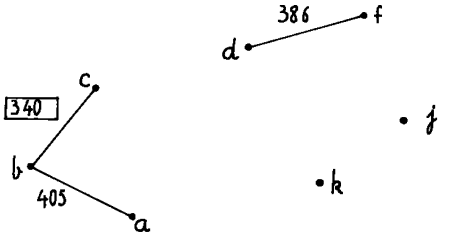
Calcul des coordonnées de k :

$F_1(k) = 11.50$ $F_2(k) = -10.00$

Calcul de l'écart suivant :

$\delta(k,j) = 714.50$;

Elimination de l'arête (k,j) :



Graphe $G_2 = (X_2 , U_2)$

ETAPE III : Agrégation de b et c

$l = \{b\} \cup \{c\}$;

$\alpha(l) = b$; $\beta(l) = c$; $\tau(l) = 340$;

$M^2(l) = 340$; $m_l = 4$; $V^2(l) = 85.00$; $V(l) = 9.22$

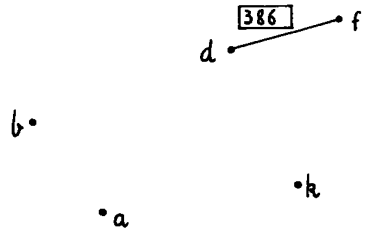
Calcul des coordonnées de l

$F_1(l) = -35.00$ $F_2(l) = 0.00$

Calcul de l'écart suivant :

$\delta(a,l) = 533.33$

Elimination de l'arête (a,ℓ)



Graphe $G_3 = (X_3 , U_3)$

ETAPE IV : Agrégation de d et f

$m = \{d\} \cup \{f\}$;

$\alpha(m) = d$; $\beta(m) = f$; $\tau(m) = 386$;

$M^2(m) = 386$; $m_m = 4$; $V^2(m) = 96.50$; $V(m) = 9.823$

Calcul des coordonnées de m

$F_1(m) = 8.50$ $F_2(m) = 16.50$

A la suite de l'agrégation des deux sommets d et f, l'ensemble des arêtes est vide et il est nécessaire de construire un nouveau graphe de similarité sur l'ensemble $X_4 = \{a,j,k,l,m\}$ après avoir augmenté la valeur du seuil de stratification.

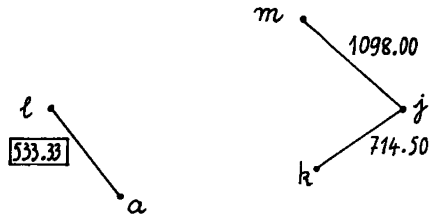
PHASE II CHOIX D'UN SECOND SEUIL DE STRATIFICATION

Après avoir calculé l'indice de dissimilarité entre chaque sommet s et son voisin le plus proche v(s), on pose :

$$\rho_4 = \text{Sup}\{\delta(s, v(s)) \mid s \in X_4\}$$

Sommet	Sommet le plus proche	Indice de dissimilarité
a	ℓ	533.33
j	k	714.50
k	j	714.50
ℓ	a	533.33
m	j	1098.00

On a donc $\rho_4 = \delta(m,j) = 1098.00$



Graphe $G_4 = (X_4 , U_4)$

ETAPE V : Agrégation de a et l

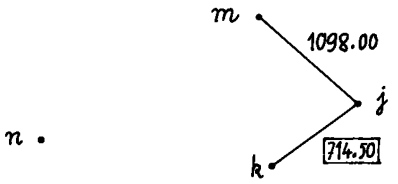
$n = \{a\} \cup \{l\}$

$\alpha(n) = a \quad \beta(n) = l \quad \tau(n) = 533.33$

$M^2(n) = 873.33 \; ; \; m_n = 6 \; ; \; V^2(n) = 145.55 \; ; \; V(n) = 12.065;$

Calcul des coordonnées de n :

$F_1(n) = -31.00 \quad F_2(n) = -5.33$



Graphe $G_5 = (X_5 , U_5)$

ETAPE VI : Agrégation de k et j

$O = \{k\} \cup \{j\}$

$\alpha(O) = k \quad \beta(O) = j \quad \tau(O) = 714.50$

$M^2(O) = 1197.50 \; ; \; m_0 = 8 \; ; \; V^2(O) = 149.69 \; ; \; V(O) = 12.235$

Calcul des coordonnées de O

$F_1(O) = 19.00 \quad F_2(O) = -4.25$

Calcul de l'écart suivant

$\delta(O,m) = 1422.16$

Elimination de l'arête (O,m)

A la fin de cette étape, l'ensemble des arêtes est vide et il est nécessaire de redéfinir la valeur du seuil de stratification.

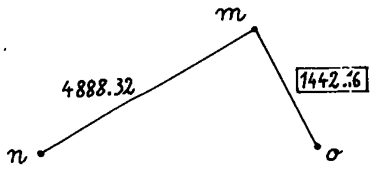
PHASE III DEFINITION D'UN TROISIEME SEUIL DE STRATIFICATION

On pose $\rho_6 = \text{Sup}\{ \delta(s,v(s)) \mid s \in X_6 \}$

avec $X_6 = \{m,n,o\}$

Sommet	Sommet le plus proche	Ecart
m	O	1442.16
n	m	4888.32
O	m	1442.16

On a donc $\rho_6 = \delta(m,n) = 4888.32$



Graphe $G_6 = (X_6 , U_6)$

ETAPE VII : Agrégation de m et o

$p = \{m\} \cup \{O\}$

$\alpha(p) = m \quad \beta(p) = o \quad \tau(p) = 1442.16$

$M^2(p) = 3025.66$; $m_p = 12$; $V^2(p) = 252.14$; $V(p) = 15.879$

Calcul des coordonnées de p

$F_1(p) = 15.50 \quad F_2(p) = 2.66$

Calcul de l'écart suivant :

$\delta(p,n) = 8905.00$

Elimination de l'arête (p,n)

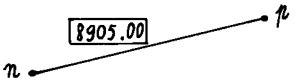
L'ensemble des arêtes est vide et il est nécessaire de redéfinir la valeur du seuil de stratification.

PHASE IV DEFINITION D'UN QUATRIEME SEUIL DE STRATIFICATION

On pose $\rho_7 = \text{Sup}\ \{\delta(s,v(s)) \mid s \in X_7\}$

avec $X_7 = \{n,p\}$

soit $\rho_7 = \delta(n,p) = 8905.00$



Graphe $G_7 = (X_7 , U_7)$

ETAPE VIII : Agrégation de n et p

$q = \{n\} \cup \{p\}$

$\alpha(q) = n \quad \beta(q) = p \quad \tau(q) = 8905$

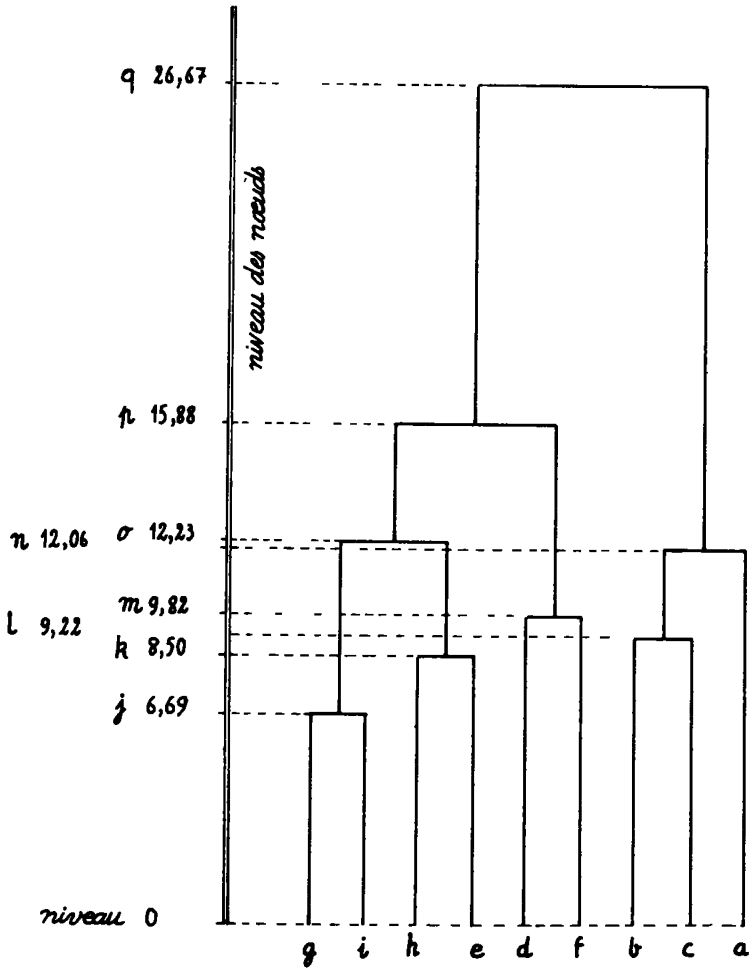
$M^2(q) = 12804$; $m_q = 18$; $V^2(q) = 711.33$; $V(q) = 26.671$

Graphe $G_8 = (X_8 , U_8)$

q (réduit au seul point isolé q)

Le tableau suivant récapitule les résultats obtenus au cours des différentes étapes de la classification ascendante hiérarchique, selon le critère de la variance :

Etape	Classe n	$\alpha(n)$	$\beta(n)$	$\tau(n)$	$M^2(n)$	m_n	$V^2(n)$	$V(n)$
I	j	g	i	194	194	4	48.50	6.694
II	k	h	e	289	289	4	72.25	8.500
III	l	b	c	340	340	4	85.00	9.220
IV	m	d	f	386	386	4	96.50	9.823
V	n	a	l	533.33	873.33	6	145.55	12.065
VI	o	k	j	714.50	1197.50	8	149.69	12.235
VII	p	o	m	1442.16	3025.66	12	252.14	15.879
VIII	q	p	n	8905.00	12804.00	18	711.33	26.671



Hiérarchie engendrée selon le critère de la variance.

5 Les performances du nouvel algorithme

5.1 Temps de calcul : Prenons l'exemple d'un grand ensemble de données que nous avons traité sur l'ordinateur 370-168 du CNRS : 1561 objets caractérisés chacun par 7 facteurs. La hiérarchie totale binaire a été construite en 38 secondes, selon le critère de la variance.

Le seuil de stratification a été redéfini sept fois lors de la construction de la hiérarchie et 371 entrées-sorties par blocs de 13030 octets entre la mémoire centrale et un disque de travail, ont été nécessaires pour munir l'ensemble des sommets supérieurs de l'arbre binaire d'une structure de graphe, après chaque variation du seuil de stratification.

La capacité de la mémoire centrale de l'ordinateur IBM 370-168 est égale à 4000K-octets. Seuls 560 K-octets ont été utilisés pour construire la hiérarchie binaire de parties sur l'ensemble des 1561 objets. En effet le nouvel algorithme ne nécessite que la mise en mémoire du tableau des facteurs et de la structure d'information qui représente le graphe des liaisons de contiguïté entre les objets ou classes à agréger.

Le temps de calcul total, égal à 40,70 secondes, se décompose de la manière suivante :

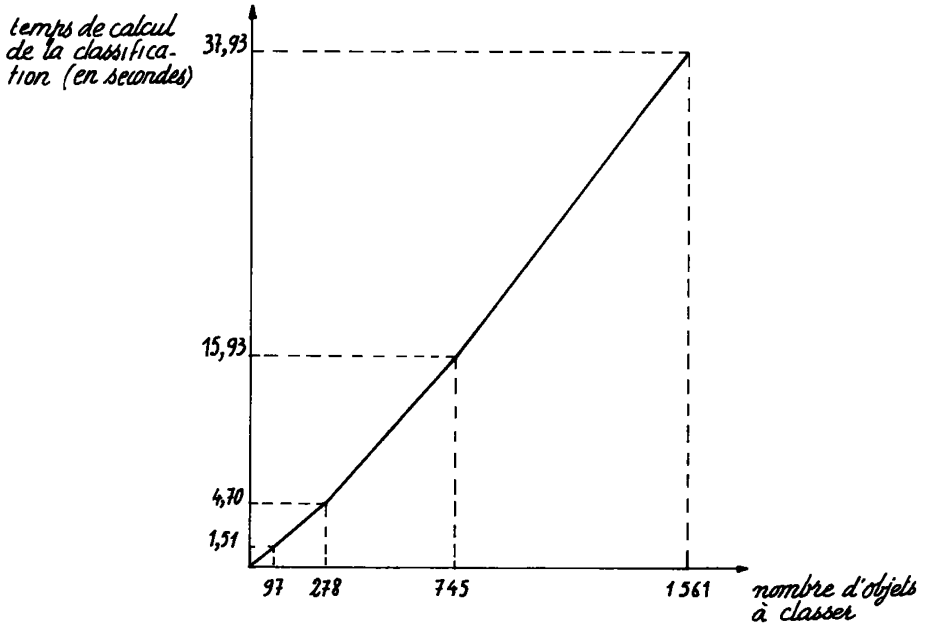
- Lecture des données : 1,09 s
- Opérations intermédiaires d'entrée-sortie : 1,68 s
- Classification ascendante hiérarchique : 37,93 s

Pour plus de détail, le tableau suivant donne la décomposition du temps de calcul de la classification arborescente des 1561 objets, selon les différentes phases de l'algorithme et selon les 7 itérations au début desquelles le graphe de similarité est construit après chaque variation du seuil de stratification.

Numéro d'itération	I	II	III	IV	V	VI	VII
Seuil de stratification	$0,8.10^{-4}$	$0,32.10^{-3}$	$0,13.10^{-2}$	$0,51.10^{-2}$	$0,205.10^{-1}$	$0,81.10^{-1}$	0,27
Nombre de sommets du graphe	1561	749	278	97	35	12	5
Nombre d'arêtes du graphe	6962	7007	2508	669	160	31	10
Nombre d'étapes de construction de la hiérarchie (par itération)	812	471	181	62	23	7	4
Nombre de distances calculées sur le graphe de similarité	8683	12066	4169	1079	249	36	0
TEMPS DE CALCUL (en secondes)							
distances entre objets ou classes (pour définir le graphe de contiguité)	16,87	5,66	1,08	0,23	0,10	0,06	0,06
définition du graphe de contiguité	1,50	1,54	0,63	0,26	0,14	0,08	0,05
détermination des sommets à agréger et mises à jour du graphe	3,63	4,03	1,48	0,39	0,11	0,027	0,006
Temps cumulés des trois calculs détaillés ci dessus.	22,00	11,23	3,19	0,88	0,35	0,17	0,11

Classification ascendante hiérarchique sur 1561 objets ; bilan détaillé par itération des 37,93 secondes dévolues à la classification proprement dite.

Le graphique suivant donne d'après d'autres exemples que nous avons traités la variation du temps de calcul de la hiérarchie binaire enfonction du nombre d'objets de l'ensemble sur lequel est édifiée la classification.



Variation du temps de calcul de la hiérarchie binaire en fonction du nombre d'objets de l'ensemble sur lequel est édifiée la classification

Les expériences réalisées sur ordinateur montrent donc que le temps de calcul du nouvel algorithme varie presque proportionellement au nombre d'objets à classer lorsque ce nombre est inférieur à 1600. Pour un ensemble composé de plusieurs milliers d'objets, le temps augmenterait approximativement comme le carré du nombre d'objets et non comme le cube, ce qui serait le cas avec l'algorithme de [C.A.H.] usuel.

5.2 *Comparaison avec la méthode des nuées dynamiques* :

Pour un ensemble composé de quelques milliers d'objets, il s'avère inutile d'éditer une hiérarchie exhaustive de toutes les parties : l'algorithme de classification par la "méthode des voisinages réductibles " permet d'obtenir rapidement un certain nombre de partitions de l'ensemble des objets par coupure de la hiérarchie binaire à différents niveaux. On comparera donc la méthode des v. r. à une autre méthode rapide communément utilisée pour édifier une partition sur un grand ensemble de données, la méthode des nuées dynamiques (DIDAY 1971).

Bornons-nous a une version simplifiée : l'agrégation autour de centres variables dans un espace euclidien E. En bref on choisit dans l'espace E où sont représentés les individus à classer, n centres d'agrégation $C_1, \ldots, C_q, \ldots, C_n$ et chaque individu i est attaché au centre dont il est le plus proche ; ainsi se trouvent constituées n classes. On prend alors pour nouveaux centres d'agrégation les centres de gravité de ces classes et

on répartit les individus en les affectant chacun au centre le plus proche. Et ainsi de suite jusqu'à stabilisation des centres. L'intérêt de la méthode vient de ce qu'on peut montrer qu'à chaque itération la variance intérieure aux classes (intraclasse) décroît, tandis que complémentairement s'accroît la variance entre les classes (variance du nuage des centres ; ou v. interclasse). L'algorithme des nuées dynamiques n'aboutit toutefois qu'à un optimum local dépendant du choix initial des centres ; et c'est pourquoi se font généralement plusieurs essais à partir de centres diversement choisis.

Puisque la quantité critère - la variance - est essentiellement la même qu'en classification hiérarchique, on comparera les partitions fournies par ces deux méthodes : sans perdre de vue que tandis que la méthode des n. dyn. ne donne qu'une seule partition à la fois, la classification hiérarchique en donne d'une coupe une suite emboitée (selon le niveau auquel est tronqué l'arbre). Afin de donner à l'algorithme des n. dym. les meilleures données initiales on prendra pour centres c_q initiaux les centres des classes obtenues par classification ascendante ; et l'on observera dans quelle mesure l'algorithme des n. dyn. améliore ce résultat (qui est déjà vraisemblablement supérieur à la plupart des optimum relatifs auxquels aboutirait la méthode des n. dyn. à partir de centres tirés au hasard).

Dans l'exemple des 1561 individus, la partition en 120 classes, obtenue par coupure de l'arbre hiérarchique, extrait 89,6% de la variance du nuage de points (i.e. la variance interclasse est de 89,6% de la v. totale). Après convergence de la méthode d'agrégation autour de 120 centres variables, on obtient une partition localement optimale qui extrait 90,2% de l'inertie totale du nuage de points. La faible augmentation de l'inertie expliquée montre que l'algorithme de classification par la méthode des voisinages réductibles permet de construire une partition satisfaisante d'un grand ensemble de données.

Le tableau suivant donne l'augmentation du pourcentage d'inertie expliqué, après convergence de la "méthode des centres variables", pour plusieurs essais effectués en coupant l'arbre hiérarchique plus ou moins haut.

Nombre de classes de la partition	Taux d'explication initial	Taux d'explication après convergence	Augmentation du taux d'explication
100	87,44	88,20	0,76
110	88,10	88,72	0,62
120	88,68	89,30	0,62
130	89,23	89,79	0,56
140	89,73	90,24	0,51
150	90,19	90,67	0,48

Ainsi, l'algorithme de classification par la "méthode des voisinages réductibles" permet de traiter de grands ensembles de données. Les résultats obtenus ont la finesse propre à la classification hiérarchique ; et pour le coût du calcul ou la qualité des partitions offrent les mêmes avantages que la méthode des nuées dynamiques. L'utilisateur a toute facilité de sectionner l'arbre hiérarchique à divers niveaux jugés pertinents au cours de l'interprétation. Après avoir initialement défini une centaine de classes, et représenté l'arbre supérieur construit sur ces classes élémentaires, on peut retenir quelques grandes classes; subdiviser certaines d'entre elles afin d'en expliciter le contenu et d'en éclairer la disposition le long des axes factoriels ; etc.

BIBLIOGRAPHIE

BENZECRI J.P. (1964-1973). Leçons sur les classifications. Cours 3° cycle.

BENZECRI J.P. (1973). L'Analyse des Données. T. I, lLa Taxinomie ; T. II, L'Analyse des Correspondances, Paris, DUNOD .

BENZECRI J.P. (1976).Histoire et Préhistoire de l'Analyse des Données. Les Cahiers de l'analyse des données, Vol. I n° 1, 2, 3, 4.

BRUYNOOGHE M. (1976).Un algorithme de classification ascendante hiérarchique d'un grand ensemble de données. Communication au Congrès européen des statisticiens, Grenoble, 6-10 Sept. 1976.

BRUYNOOGHE M. Un algorithme rapide de classification automatique utilisant la notion de voisinage : la "méthode des graphes de contiguïté". Communication présentée aux Premières Journées Nationales de Classification (organisées par l'Association des Statisticiens Universitaires), Vannes, 25-27 Mai 1977.

BRUYNOOGHE M. (1977-b)Classification automatique d'un grand ensemble de données par la "méthode des graphes réductibles". Application à l'analyse de la distance à la ville en Languedoc-Roussillon. Symposium on Data Analysis and Informatics. Paris, 7-9 Septembre 1977 et Colloque International de Taxinomie Numérique Orsay, 6 Septembre 1977.

CORMACK R. M. (1971). A review of classification, J. R. Statist. Soc., A, 134, Part 3, 321-367.

DIDAY E. (1971).Une nouvelle méthode en classification automatique et reconnaissance des formes. Revue de statistique Appliquée, Vol. XIX, n° 2.

GOWER J.C.and ROSS G.J.S. (1969). Minimum spanning trees and single linkage cluster analysis. Appl. Statist., 18, 54-64.

JAMBU M. (1972). Techniques de classification automatique appliquées à des données "Sciences Humaines", Thèse de doctorat de 3° cycle, Laboratoire de Statistique de l'Université Paris VI.

JARVIS R.A. and PATRICK E.A. (1973. Clustering using a similitarity measure based on shared near neibhors. IEEE Transactions on Computers, Vol. C-22, n° 11, 1025-1034.

ROUX M. (1968). Un algorithme pour construire une hiérarchie particulière. Thèse de doctorat de 3° cycle. Laboratoire de Statistique de l'Université Paris VI.

SOKAL R.R. and SNEATH P.H.A. (1963). Principles of Numerical taxonomy. London Freeman.

WARD J.H. (1963). Hierarchical grouping to optimise an objective function. J. Am. Statist. Ass., 59, 236-244.

WISHART D. (1969). An algoritm for hierarchical classifications. Biometrics 25, 165-170.