RAIRO. RECHERCHE OPÉRATIONNELLE

HENRI RALAMBONDRAINY
EDWIN DIDAY

Optimisation d'une classification hiérarchique

RAIRO. Recherche opérationnelle, tome 18, nº 3 (1984), p. 279-295
<http://www.numdam.org/item?id=RO_1984__18_3_279_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 18, n° 3, août 1984, p. 279 à 295)

OPTIMISATION D'UNE CLASSIFICATION HIÉRARCHIQUE (*)

par Henri RALAMBONDRAINY et Edwin DIDAY (1)

Résumé. — *Les méthodes de classification hiérarchiques habituelles fournissent des hiérarchies qui optimisent à chaque étape un critère d'agrégation. Ces méthodes n'optimisent pas de critère global.*

La première partie de ce rapport présente un algorithme qui recherche une hiérarchie optimale sur toute la population au sens du critère de l'inertie.

La deuxième partie traite des problèmes d'inversions des hiérarchies indicées, un résultat original est démontré concernant ces problèmes, puis appliqué à l'algorithme d'optimisation.

Mots clés : Hiérarchies; inversions.

Abstract. — *The methods of hierarchical classification usually applied produce hierarchies which optimize the "aggregation criterion" at each stage. These methods do not optimize a global criterion.*

The first part of this report presents an algorithm which is intended to establish an optimal hierarchy for the whole population with reference to the "inertia criterion".

The second part is concerned with the inversion problem of hierarchies, an original result is presented, and applied to the optimisation algorithm.

Keywords: Hierarchies; inversions.

1. INTRODUCTION

Les méthodes de classification hiérarchique habituelles optimisent généralement à chaque étape un critère d'agrégation. Ces méthodes n'optimisent pas de critère global.

Dans cet article, nous avons adopté une approche différente. Nous cherchons à obtenir des structures hiérarchiques qui soient optimales selon un critère global.

D'autres auteurs ont également posé les problèmes sous cette forme. Signalons les travaux de Caroll et Pruzanski [4] et Chandon [5] : ils recherchent une ultramétrique la plus proche au sens des moindres carrés de l'indice de distance de départ.

(*) Reçu en octobre 1982.
(1) I.N.R.I.A., Domaine de Voluceau, Rocquencourt, 78150 Le Chesnay.

R.A.I.R.O. Recherche opérationnelle/Operations Research, 0399-0559/84/03 279 17/$ 3.70

La première partie de ce travail présente un algorithme qui cherche une hiérarchie optimale sur un ensemble E au sens du critère de l'inertie, c'est-à-dire dont la somme des inerties des classes est minimale [12].

C'est l'un des critères proposés par Diday dans [9]. L'intérêt de cette méthode est de fournir une hiérarchie ayant globalement des nœuds de plus faibles inerties que par les méthodes habituelles.

L'étude des hiérarchies optimales selon d'autres critères nous a amené à nous intéresser aux problèmes d'inversions des hiérarchies indicées. La deuxième partie présente un résultat original concernant ces problèmes.

L'un des intérêts de ce travail est de montrer que la méthode des nuées dynamiques n'est pas seulement une méthode de partitionnement mais qu'elle peut être utilisée avec d'autres structures inter-classes comme celle des hiérarchies.

2. HIÉRARCHIES LOCALEMENT OPTIMALES

2.1. Introduction

Les algorithmes de classification automatique hiérarchiques fournissent des hiérarchies de parties d'un ensemble donné. Ces hiérarchies ne sont en général pas optimales au sens d'un critère global. On se propose en utilisant la Méthode des Nuées Dynamiques (M.N.D.) de trouver des hiérarchies optimales selon un critère relatif à l'inertie.

2.2. Le problème d'optimisation

Soit $E \subset R^p$ l'ensemble des objets à classer de cardinal égal à n. On suppose E muni d'une distance quadratique d et à chaque point $x \in E$ est associé un poids $p_x \geqq 0$.

On appelle hiérarchie dichotomique définie sur une ensemble fini E, tout ensemble H de parties de E tel que :

(i) $\forall x \in E$; $\{x\} \in H$;

(ii) $E \in H$;

(iii) $B \cap B' \in \{B, B', \varnothing\}$ pour tout $B \in H$ et tout $B' \in H$;

(iv) $\forall B \in H$; card $B > 1 \Rightarrow \exists (B_1, B_2) \in H^2$; $B_1 \overset{c}{\neq} B$; $B_2 \overset{c}{\neq} B$ et $B_1 \cup B_2 = B$.

En suivant les principales étapes de la M.N.D. telles qu'elles ont été décrites dans [10] on définit :

— *l'espace de recouvrement S :*

$S = \mathbb{H}$ ensemble des hiérarchies dichotomiques définies sur E;

— *l'espace de représentation* $\mathbb{L}$:

$\mathbb{L}=[\mathbb{R}^p]^N$ avec $N=2n-1$ nombre des nœuds d'une hiérarchie dichotomique définie sur E;

— *le critère à optimiser :*

$$\underset{\substack{H \in \mathbb{H} \\ L \in \mathbb{L}}}{\text{Min}}\ W(H, L)=\sum_{\substack{B \in H \\ l \in L}} I(B, l),$$

avec $I(B, l)=\sum\{p_x d^2(x, l), x \in B\}$ l sera le centre de gravité du nœud B de la hiérarchie H. On recherche donc parmi toutes les hiérarchies dichotomiques définies sur E, celle dont la somme des inerties des nœuds est minimale. L'intérêt de ce critère est qu'il exprime, si H est meilleure que H', que globalement la hiérarchie H a des nœuds de plus faibles inerties que H'.

2.3. L'algorithme

L'algorithme consiste à partir d'une hiérarchie initiale H_0 puis d'optimiser le critère W par itérations successives en utilisant alternativement une fonction de représentation g et une fonction d'affectation f.

— *La fonction de représentation g :*

On définit une fonction g de $\mathbb{H}$ dans $\mathbb{L}$:

$$g : \mathbb{H} \to \mathbb{L},$$
$$H \xrightarrow{g} g(H)=L=(g_1, \ldots, g_N) \in [\mathbb{R}^p]^N,$$

g_i est défini comme le centre de gravité du nœud B_i.

— *La fonction d'affectation f :*

f est une application de $\mathbb{L} \times \mathbb{H}$ dans $\mathbb{H}$:

$$f : \mathbb{L} \times \mathbb{H} \to \mathbb{H},$$
$$(L, H) \xrightarrow{f} H'=f(L, H),$$

la fonction f permet, à partir d'un couple (L, H), c'est-à-dire d'une hiérarchie et l'ensemble des centres de gravité de ses nœuds de générer une nouvelle hiérarchie $H' \in \mathbb{H}$ de la manière suivante :

Soit $H=\{B_i \mid i \in I\}$ où $I=[1, N]$ et B_{i_0} un nœud de H; soit $S_{B_{i_0}}$ l'ensemble des antécédents de B_{i_0}, autrement dit :

$$S_{B_{i_0}}=\{B_i \in \mathrm{H} \mid \mathrm{B}_{i_0} \subset B_i \text{ et } \mathrm{B}_{i_0} \neq \mathrm{B}_i\}.$$

Si $B_{i_1} \notin S_{B_{i_0}}$ on définit $S_{B_{i_0}-B_{i_1}} = S_{B_{i_0}} \cap \complement S_{B_{i_1}}$ (ensemble des antécédents de B_{i_0} ne contenant pas B_{i_1}).

Exemple : Soit la hiérarchie $H = \{ B_i, i = 1, \ldots, 9 \}$ (cf. *fig.* 1), les ensembles d'antécédents de B_1 et B_6 sont respectivement :

$$S_{B_1} = \{ B_3, B_7, B_9 \} \qquad \text{et} \qquad S_{B_6} = \{ B_7, B_9 \},$$

alors l'ensemble des antécédents de B_1 ne contenant pas B_6 est $S_{B_1 - B_6} = \{ B_3 \}$.

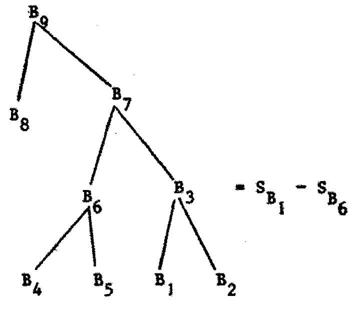


Figure 1

Soient :

$$\delta_1 (B_{i_0}, S_{B_{i_1} - B_{i_0}}) = \sum_{B_i \in S_{B_{i_1}} - B_{i_0}} \frac{p(B_{i_0}) p(B_i)}{p(B_{i_0}) + p(B_i)} d^2 (g_{B_{i_0}}, g_{B_i})$$
$$+ I(B_{i_0}) \operatorname{card} S_{B_{i_1} - B_{i_0}} + I(B_{i_0} \cup B_{i_1}) \qquad \text{où} \quad p(\mathrm{B}) = \{ \textstyle\sum p_x \mid x \in B \}.$$

δ_1 exprime le gain d'inertie après affectation de B_{i_0} à B_{i_1} :

$$\delta_2 (B_{i_0}, S_{B_{i_0} - B_{i_1}}) = \sum_{\substack{B_i \in S_{B_{i_1}} - B_{i_0} \\ B_i \neq B_{i_1}}} \frac{p(B_{i_0}) p(B_i)}{p(B_i) - p(B_{i_0})} d^2 (g_{B_{i_0}}, g_{B_i})$$
$$+ I(B_{i_0}) (\operatorname{Card} S_{B_{i_0} - B_{i_1}} - 1) + I(B_{i_1}).$$

δ_2 exprime la perte d'inertie après affectation de B_{i_0} à B_{i_1}, alors :

$$\Delta_L (B_{i_0}, B_{i_1}) = \delta_1 (B_{i_0}, S_{B_{i_1} - B_{i_0}}) - \delta_2 (B_{i_0}, S_{B_{i_0} - B_{i_1}}).$$

La fonction f sera définie comme la composés de deux fonctions φ_1 et φ_2 (cf. *fig.* 2) définies de la manière suivante :

Δ_L est une fonction de $H \times H$ dans R. L'ensemble H étant fini, il existe deux sommets disctints $(i_0, i_1) \in H$ *réalisant le minimum de* Δ_L :

$$\min_{i,\, i' \in H} \Delta_L(B_i, B_{i'}) = \Delta_L(B_{i_0}, B_{i_1}).$$

On suppose i_0, i_1 unique, alors la fonction φ_1 de $\mathbb{H} \times \mathbb{L}$ dans $I \times I$ est définie par $\Phi_1(H, L) = (i_0, i_1)$.

En effet, on peut toujours se ramener à un choix unique de (i_0, i_1) lorsque plusieurs couples réalisent le minimum, il suffit par exemple de prendre le couple qui est pour l'ordre lexicographique le plus petit.

La fonction Φ_2 est une application de $I \times I$ dans H définie comme suit :

– Si $B_{i_0} \cup B_{i_1} \notin H$ et $\Delta_L(B_{i_0}, B_{i_1}) < 0$ alors :

$$\varphi_2(\mathrm{i}_0, \mathrm{i}_1) = H' = (H'_1 \cup H'_2 \cup H'_3 \cup B_{i_0} \cup B_{i_1}),$$

où H'_1 est le complémentaire dans H de $S_{B_{i_0} - B_{i_1}} \cup S_{B_{i_1} - B_{i_0}}$ et :

$$H'_2 = \{ B'_i = B_i - B_{i_0} \mid B_i \in H,\ B_i \in S_{B_{i_0} - B_{i_1}} - B_{i_2} \},$$

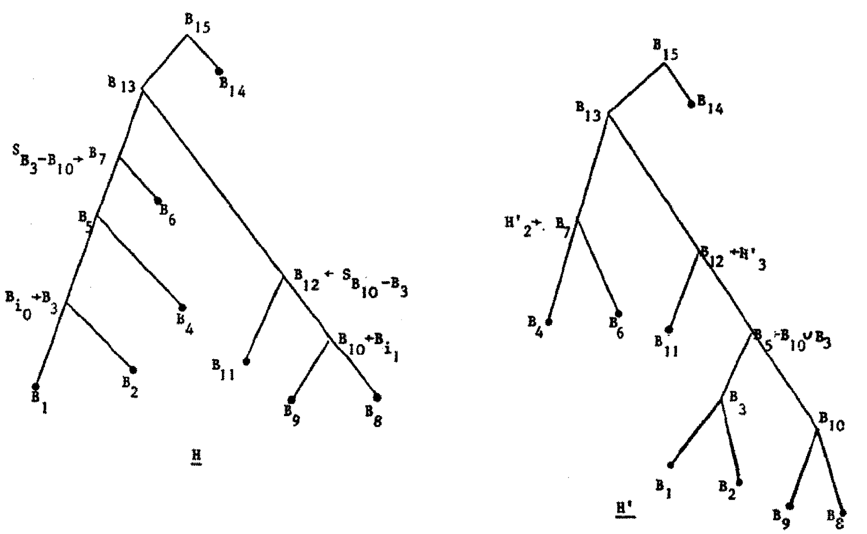


Figure 2

où B_{i_2} tant le successeur immédiat de B_{i_0}, $B_{i_2} \in S_{B_{i_0}-B_{i_1}}$:

$$H'_3 = \{ B'_i = B_i \cup B_{i_0} \mid B_i \in H,\ B_i \in S_{B_{i_1}-B_{i_0}} \}.$$

– Sinon $\Phi_2(i_0, i_1) = H$.

Exemple : Soit la hiérarchie H donnée dans la figure 2, les nœuds réalisant le minimum négatif de $\Delta_L(B_i, B'_i)$ étant $B_3\,(i_0=3)$ et $B_{10}\,(i_1=10)$, le successeur immédiat de B_3 dans H est $B_5\,(i_2=5)$, on a alors :

$$S_{B_3-B_{10}} = \{ B_5, B_7 \} \qquad \text{et} \qquad S_{B_{10}-B_3} = \{ B_{12} \}$$

et:

$$\Phi_2(3,10) = H' = (H'_1 \cup H'_2 \cup H'_3 \cup B_3 \cup B_{10}) \qquad \text{avec} \quad H'_2 = \{ B_7 \},$$
$$H'_3 = \{ B_{12} \},\ H'_1 = \{ H - B_5, B_7, B_{12} \} \}.$$

2.4. Propriétés de l'algorithme

Afin de démontrer la convergence de l'algorithme nous allons établir le théorème suivant :

THÉORÈME : *les valeurs des critères $W(H', L')$ et $W(H, L)$ sont liés par l'expression suivante :*

$$W(H', L') = W(H, L) + \Delta_L(B_{i_0}, B_{i_1}).$$

Démonstration : on a :

$$W(H, L) = \sum_{B_i \in H} I(B_i);$$

$$W(H', L') = \sum_{B_{i'} \in H} I(B'_i) \qquad \text{et} \qquad H' = f(H, L).$$

Décomposons $W(H, L)$:

$$W(H, L) = \sum_{B_i \notin S_{B_{i_0}-B_{i_1}} \cup S_{B_{i_1}-B_{i_0}}}$$
$$+ \sum_{B_i \in S_{B_{i_0}-B_{i_1}}-B_{i_2}} I(B_i) + I(B_{i_2}) + \sum_{B_i \in S_{B_{i_1}-B_{i_0}}} I(B_i),$$

d'après la définition de la fonction d'affectation f:

$$W(H', L') = \sum_{B_i \in S_{B_{i_0}-B_{i_1}} \cup S_{B_{i_1}-B_{i0}}} I(B_i)$$
$$+ \sum_{B_i \in S_{B_{i_0}-B_{i_1}}-B_{i_2}} I(B_i - B_{i_0}) + I(B_{i_0} \cup B_{i_1}) + \sum_{B_i \in S_{B_{i_1}-B_{i_0}}} I(B_i \cup B_{i_0}),$$

d'après le théorème de Huygens :

$$I(B_i \cup B_{i_0}) = I(B_i) + I(B_{i_0}) + \frac{p(B_{i_0})\,p(B_i)}{p(B_i)+p(B_{i_0})} d^2(g_{B_{i_0}}, g_{B_i}),$$

$$I(B_i - B_{i_0}) = I(B_i) - I(B_{i_0}) - \frac{p(B_{i_0})\,p(B_i)}{p(B_i)-p(B_{i_0})} d^2(g_{B_{i_0}}, g_{B_i}).$$

On montre alors facilement que :

$$W(H', L') - W(H, L) = \delta_1(B_{i_0}, S_{B_i - B_{i_0}}) - \delta_2(B_{i_0}, S_{B_{i_0} - B_{i_1}}) = \Delta_L(B_{i_0}, B_{i_1}).$$

D'après la définition de f, $\Delta(B_{i_0}, B_{i_1}) < 0$ par suite on a :

PROPOSITION 1 : *la suite u_n converge en décroissant.*

Démonstration : on a :

$$u_n = W(H^n, L^n) < W(H^{n-1}, L^{n-1}) = u_{n-1},$$

car :

$$W(H^n, L^n) - W(H^{n-1}, L^{n-1}) = \Delta(B_{i_0}, B_{i_1}) < 0.$$

La suite u_n est décroissante et minimisée par 0 donc converge.

PROPOSITION 2 : *la suite v_n converge en atteignant sa limite.*

On sait que toute suite convergente sur un ensemble fini atteint sa limite or c'est le cas de la suite u_n puisque l'espace est fini.

Supposons que la convergence soit atteinte à l'itération M donc :

$$u_{M+1} = u_M,$$

d'où :

$$W(v_{M+1}) = W(v_M).$$

On a alors $v_{M+1} = v_M$ c'est-à-dire $L^{M+1} = L^M$, $H^{M+1} = H^M$; en effet, l'affectation ne se fait que s'il existe B_{i_0} et B_{i_1} tels que $\Delta_L M(B_{i_0}, B_{i_1}) < 0$, à la convergence, il n'existe pas de $\Delta_L M(B_{i_0}, B_{i_1}) < 0$ donc : $H^M = H^{M+1}$.

2. Généralisation

On peut voir facilement que l'algorithme s'applique :

— aux hiérarchies non totales c'est-à-dire dont les éléments terminaux ne sont pas singletons, cela permet de se limiter à améliorer par exemple, les parties hautes d'une hiérarchie;

— aux k-hiérarchies $\bar{H}=(H_1, H_2, \ldots, H_k)$ où chaque H_i est une hiérarchie totale. Ceci présente un intérêt dans le traitement de grands ensembles de données où l'on désire avoir un ensemble de hiérarchies localement optimales par rapport à un critère global.

2.6. Présentation du programme

Le programme nécessite en entrée les paramètres concernant les données : nombre d'individus, nombre de variables, format de lecture, distance sur les individus, les paramètres concernant la hiérarchie initiale : critère d'agrégation, format de lecture.

Le programme édite la valeur du critère à chaque itération, les descriptions et les arbres correspondant à la hiérarchie initiale et optimale.

La hiérarchie optimale peut être évidemment interprétée par les aides à l'interprétation habituelles.

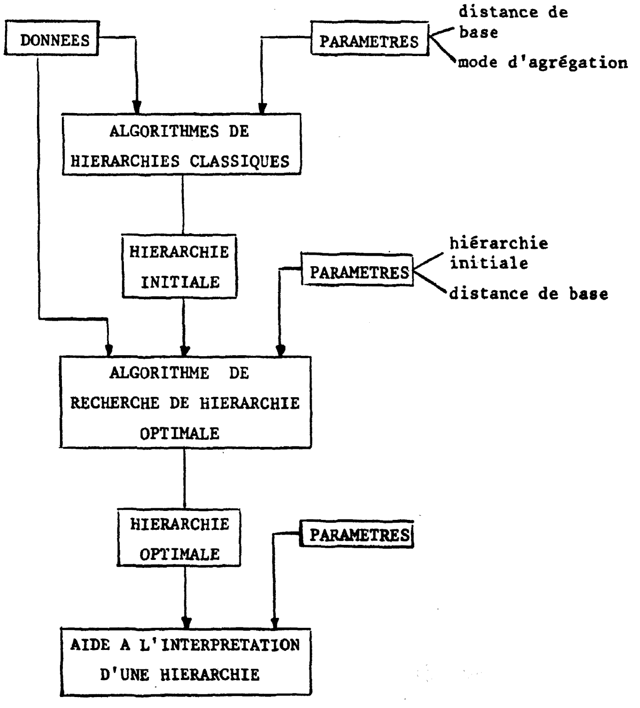


Figure 3. — Étapes de traitements pour la recherche d'une hiérarchie optimale.

2.7. Exemples

Les données traitées sont relatives à la comptabilité informatique du centre de calcul de l'I.N.R.I.A. Il s'agit de 25 utilisateurs caractérisés par 13 variables de consommations ressources machines mensuelles.

On a constitué quatre jeux de données, en tirant trois sous-populations d'effectifs 10, 15 et 20.

Dans le cas $n=10$, la figure 4 présente la hiérarchie $H1$ obtenue par la méthode « minimisation de l'inertie des classes » et la figure 5 la hiérarchie $H2$ obtenue après convergence de l'algorithme au bout d'une itération par l'affectation des nœuds 2 et 15.

i	d(i)	a(i)	b(i)	i(i)	description des classes de la hierarchie
11	0.0951	10	7	0.2000	1604 1303
12	0.2212	6	4	0.2000	1210 1202
13	0.5229	8	11	0.3000	1304 1604 1303
14	0.7039	12	5	0.3000	1210 1202 1208
15	2.2251	3	1	0.2000	1104 1101
16	2.5517	13	14	0.5000	1304 1604 1303 1210 1202 1208
17	5.2950	9	2	0.2000	1307 1103
18	10.5267	16	17	0.8000	1304 1604 1303 1210 1202 1208 1307 1103
19	21.9133	18	15	1.0000	1304 1604 1303 1210 1202 1208 1307 1103 1104 1101

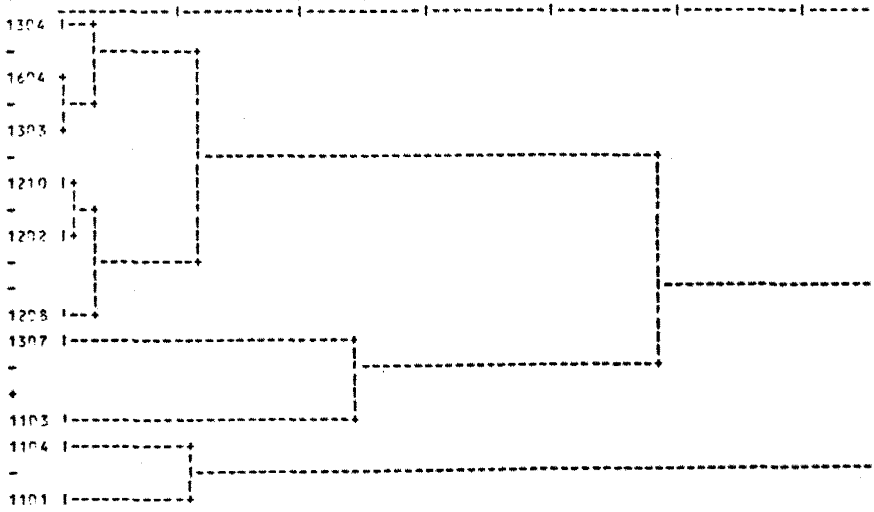
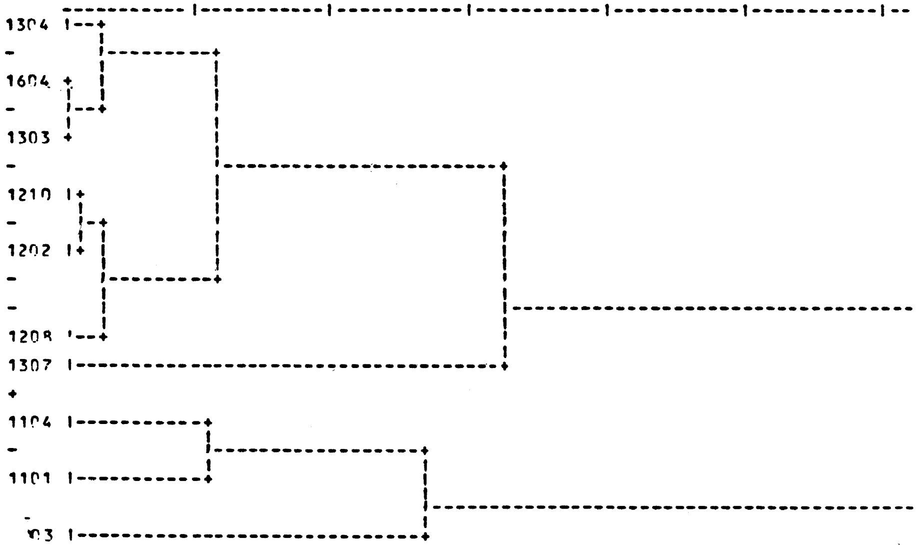


Figure 4. — Hiérarchie H_1, méthode de minimisation de l'inertie des classes.
d, indices de la hiérarchie; *a*, *b*, ainés et benjamins; *i*, poids du palier.

i	d(i)	a(i)	b(i)	i(i)	description des classes de la hierarchie
11	0.0961	10	7	0.2000	1604 1303
12	0.2212	6	4	0.2000	1210 1202
13	0.5929	8	11	0.3000	1304 1604 1303
14	0.7039	12	5	0.3000	1210 1202 1208
15	2.2251	3	1	0.2000	1104 1101
16	2.5517	13	14	0.6000	1304 1604 1303 1210 1202 1208
17	5.7251	15	2	0.3000	1104 1101 1103
18	7.1656	16	9	0.7000	1304 1604 1303 1210 1202 1208 1307
19	21.9103	18	17	1.0000	1304 1604 1303 1210 1202 1208 1307 1104 1101 1103



Hierarchie : H_2 : Hiérarchie optimale pour le critère de l'inertie

critère H_1 = 44,12

critère H_2 = 41,19

Figure 5. — Hierarchie H_2, hiérarchie optimale pour le critère de l'inertie.

Les différents essais consignés dans le tableau 2.7.1 amènent les remarques suivantes :

1. Les hiérarchies classiques et en particulier la méthode « minimisation de l'inertie d'une classe » ne sont pas optimales pour le critère de l'inertie et peuvent être améliorées par l'algorithme.

TABLEAU 2.7.1

Nombre d'individus (n)	Méthode	Critère = somme des inerties des nœuds		Inerties intra des partitions des niveaux supérieurs Nombre de Classes				
				2	3	4	5	6
25	HC HO	43,5 42,8	−1,7%	13,84 13,69−	12 9,91−	8,24 8,11−	7,06 6,42−	5,36 5,04−
20	HC HO	28,4 25,5	−10 %	8,69 8,26−	6,02 6,23+	5,27 4,52−	3,25 3,08−	2,71 2,66−
15	HC HO	47,05 43,93	−6,6%	14,22 14,39−	12,27 11,79−	7,97 8,17+	5,37 5,88+	3,96 3,95−
10	HC HO	44,12 41,19	−6,6%	12,75 12,89+	10,07 8,28−	4,78 4,78	3,52 3,52	1,30 1,30

H C, Hiérarchie classique : méthode minimisation de l'inertie d'une classe.
H O, Hiérarchie optimisant la somme des inerties des nœuds.

2. Les inerties entre des niveaux supérieurs de la hiérarchie optimale sont en général meilleures (signe − dans le tableau 2.7.1) que celles de la hiérarchie classique. Il ne faut cependant pas oublier que l'objectif de la méthode est l'amélioration d'un critère global et non la recherche d'une partition optimale à chaque étape.

3. LE PROBLÈME DE L'INVERSION

La représentation visuelle d'une hiérarchie H est plus facilement interprétable si elle est indicée de façon à ce que la hauteur de chaque palier corresponde à la valeur prise par l'indice d'agrégation pour les deux paliers qui l'ont formé. Autrement dit, si f est choisi à partir de δ de la façon suivante :

$$\left.\begin{array}{l} f: H \to \mathbb{R}^+ \quad \text{telle que} \quad f(h)=\delta(h_1, h_2) \\ \text{pour tout} \quad h_1, h_2 \text{ et } h=h_1 \cup h_2 \quad \text{dans H.} \end{array}\right\} \qquad (1)$$

Un tel choix de f peut conduire à des inversions pour une hiérarchie créée par une méthode quelconque. La proposition suivante permet d'énoncer deux conditions nécessaires et suffisantes à satisfaire par δ pour que ce ne soit pas le cas pour les méthodes classiques. La première de ces conditions est valable pour la méthode proposée.

TABLEAU CENTRÉ-RÉDUIT DES DONNÉES : 2.7.2

	1	2	3	4	5	6	7	8	9	10	11	12	13
:sys	5.27	−0.34	−0.25	−0.51	−0.48	−0.38	−0.67	−0.61	−0.15	−0.44	−0.28	−0.40	2.09
1101	0.62	5.23	5.96	−0.29	−0.46	−0.38	−0.34	1.84	−0.39	1.55	0.28	2.90	0.32
1103	1.28	−0.00	−0.10	0.86	−0.07	−0.27	0.27	0.43	0.87	1.27	0.26	4.65	−0.72
1104	0.26	3.03	1.62	0.07	−0.44	−0.38	−0.44	4.01	1.77	1.49	0.39	−0.34	−0.52
1202	0.02	−0.25	−0.19	1.16	1.55	0.93	1.85	−0.25	−0.28	−0.25	−0.09	−0.41	0.59
1208	−0.64	−0.36	−0.26	1.18	2.04	1.59	0.60	−0.70	0.38	−0.92	−0.55	−0.41	−0.99
1210	0.12	−0.13	−0.15	0.99	0.41	−0.01	1.51	−0.48	−0.39	0.59	−0.30	0.01	−0.43
1303	−0.51	−0.33	−0.24	−0.49	−0.27	−0.16	−0.25	−0.68	−0.39	−0.66	−0.55	−0.41	0.59
1304	−0.23	−0.10	−0.14	0.12	0.59	0.57	−0.12	0.39	−0.39	−0.38	−0.22	−0.41	1.90
1307	−0.36	−0.32	−0.24	1.39	4.77	5.57	3.45	−0.65	−0.39	−0.38	−0.44	−0.41	1.15
1604	−0.62	−0.36	−0.25	−0.71	−0.32	−0.23	−0.54	−0.70	−0.39	−0.88	−0.55	−0.41	−0.72
1609	0.23	−0.05	−0.11	0.78	0.43	0.20	0.44	2.66	1.07	1.31	−0.38	−0.41	0.59
2121	0.23	−0.20	−0.22	0.91	0.71	0.35	1.59	0.02	−0.28	−0.52	1.73	−0.41	0.59
3100	−0.55	−0.30	−0.23	−0.90	−0.56	−0.40	−0.77	−0.32	−0.39	−0.55	0.32	−0.41	0.27
3110	−0.61	−0.36	−0.25	−0.81	−0.55	−0.40	−0.72	−0.62	−0.39	−0.67	−0.36	−0.41	−0.94
3120	0.18	0.29	−0.09	0.40	−0.28	−0.33	−0.11	0.03	−0.39	3.11	4.93	−0.18	−1.00
3130	−0.32	−0.25	−0.24	−0.59	−0.50	−0.38	−0.52	−0.15	−0.39	−0.19	−0.20	0.82	−0.16
4104	0.04	−0.21	−0.21	0.79	−0.31	−0.33	−0.31	0.85	−0.39	−0.20	1.06	−0.41	−0.34
4105	−0.50	−0.15	−0.14	−1.08	−0.57	−0.41	−0.80	−0.20	−0.39	−0.59	−0.55	−0.41	−0.62
4106	−0.58	−0.32	−0.25	−1.08	−0.57	−0.41	−0.80	−0.67	−0.39	−0.54	−0.22	−0.41	1.34
4107	−0.64	−0.36	−0.26	−1.02	−0.55	−0.39	−0.75	−0.70	−0.39	−0.93	−0.55	−0.41	−1.21
5113	0.60	−0.20	−0.21	0.94	−0.13	−0.25	0.51	0.92	4.94	0.44	0.57	0.61	−0.53
5114	0.44	−0.21	−0.20	−0.43	−0.53	−0.40	−0.43	−0.28	0.24	−0.15	0.75	1.01	−0.80
5211	0.15	0.00	−0.13	1.32	0.03	−0.24	1.09	0.61	0.14	1.07	−0.41	0.00	0.41
5213	−0.64	−0.36	−0.26	−1.08	−0.57	−0.41	−0.80	−0.66	−0.39	−0.92	−0.55	−0.41	−1.00

Il s'agit de 25 utilisateurs du centre de calcul de l'I.N.R.I.A. caractérisés par 13 variables de consommation ressources machines mensuelles.

Notons que Milligan [2] présente un cas particulier de ce théorème en donnant des conditions suffisantes de non inversion que doivent satisfaire les coefficients de la formule de Lance et Williams, 1967; on trouvera des conditions nécessaires et suffisantes de non-inversion dans E. Diday, 1982 [10].

NOTATIONS : Soit H la hiérarchie obtenue à l'aide de l'algorithme de la C.A.H. muni de l'indice d'agrégation δ et f l'application définie par (1). On note P_{h_i} la partition qui précède la formation de $h_i = h_{i-1} \cup h_{i-2}$ dans le déroulement de l'algorithme général (voir *fig.* 6) :

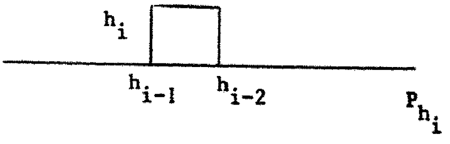


Figure 6

On peut alors énoncer le résultat suivant :

PROPOSITION 1 : *Les trois conditions suivantes sont équivalentes :*

(1) *(H, f) est une hiérarchie indicée au sens large.*

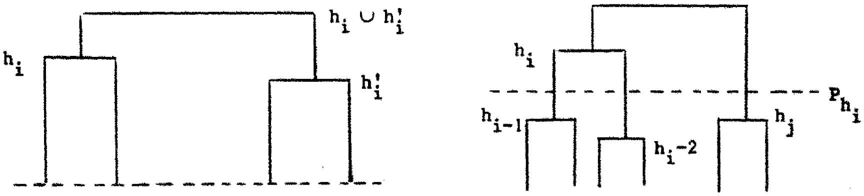
(2) *Pour tout $h_i \in H$, $h'_i \in H$ avec $h_i \cup h'_i$ dans H on a :*

$$f(h_i \cup h'_i) \geqq \mathrm{Max}\{f(h_i), f(h'_i)\}.$$

(3) *Pour tout $h_j \in P_{h_i}$ avec $h_i = h_{i-1} \cup h_{i-2}$ et $h_j \neq h_{i-1}$, $h_j \neq h_{i-2}$ on a :*

$$\delta(h_i, h_j) \geqq f(h_i).$$

REMARQUE : pour fixer les idées, la condition (2) peut être appelée « condition locale » et la condition (3) « condition globale » (voir *fig.* 7) :



(2) Condition locale

$f(h_i \cup h'_i) \geqq \mathrm{Max}(f(h_i), f(h'_i))$

(3) Condition globale

$\delta(h_i, h_j) \geqq f(h_i) \quad \forall h_j \in P_{h_i}$ et $h_j \neq h_i$.

Figure 7

Démonstration : Nous allons successivement montrer que : (1) $\Rightarrow$ (3), (3) $\Rightarrow$ (2), (2) $\Rightarrow$ (1).

Montrons d'abord que (1) $\Rightarrow$ (3). On peut faire le raisonnement suivant pour tout $h_i = h_{i-1} \cup h_{i-2} \in H$ et $h_j \in P_{h_i}$ différent de h_{i-1} et h_{i-2}. Soit h_l la première classe qui se forme après h_i au cours de l'algorithme (voir *fig.* 8).

On a nécessairement $f(h_l) \geqq f(h_i)$ par construction de H. Si $h_l \neq h_j \cup h_i$ (avec possibilité d'avoir $h_l = h_j \cup h_{l-1}$) on a $\delta(h_i, h_j) \geqq f(h_l)$ puisque h_i et h_j appartiennent à P_{h_l} et donc $\delta(h_i, h_j) \geqq f(h_i)$. Si $h_l = h_i \cup h_j$ d'après (1) on a $h_i \subset h_l \Rightarrow \delta(h_i, h_j) = f(h_l) \geqq f(h_i)$.

On a donc dans tous les cas $\delta(h_i, h_j) \geqq f(h_i)$, ce qui prouve que (1) implique (3).

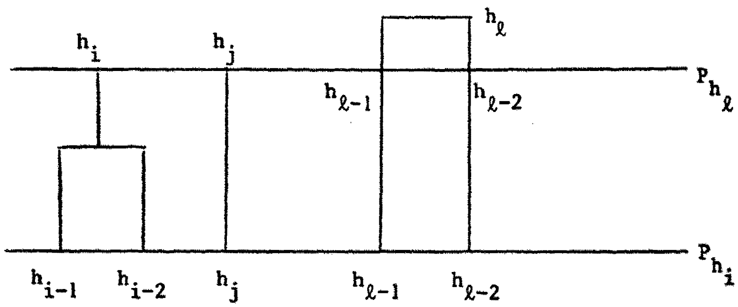


Figure 8

Montrons maintenant que (3) $\Rightarrow$ (2).

Pour tout $h_i \in H$ et $h'_i \in H$, tels que $h_i \cup h'_i \in H$ on peut faire le raisonnement suivant : supposons qu'au cours de l'algorithme h'_i soit créé avant h_i; h'_i appartient à P_{h_i} puisqu'au moment où h_i est créé, h'_i existe encore nécessairement, sinon on ne pourrait pas créer le palier $h_i \cup h'_i$ (voir *fig.* 9); d'après (3) on a donc $\delta(h'_i, h_i) \geqq f(h_i)$.

La condition (3) et le fait que h'_i soit construit avant h_i implique que $f(h_i) \geqq f(h'_i)$; en effet, on peut construire une suite de paliers $h_{i-1} \supset \ldots \supset h_{i-n}$ (voir *fig.* 9) *où* h_{i-n} est formé de deux éléments de $P_{h'_i}$ distincts de h'_{i-1} et h'_{i-2}; d'où $f(h_{i-n}) \geqq f(h'_i)$: en appliquant (3) dans $P_{h_{i-1}}, \ldots, P_{h_{i-n}}$ on a donc :

$$f(h'_i) \leqq f(h_{i-n}) \leqq \ldots \leqq f(h_{i-1}) \leqq f(h_i).$$

On a donc finalement $\delta(h'_i, h_i) \geqq f(h_i) \geqq f(h'_i)$ d'où :

$$\delta(h'_i, h_i) \geqq \operatorname{Max}(f(h_i); f(h'_i)),$$

qui prouve (3).

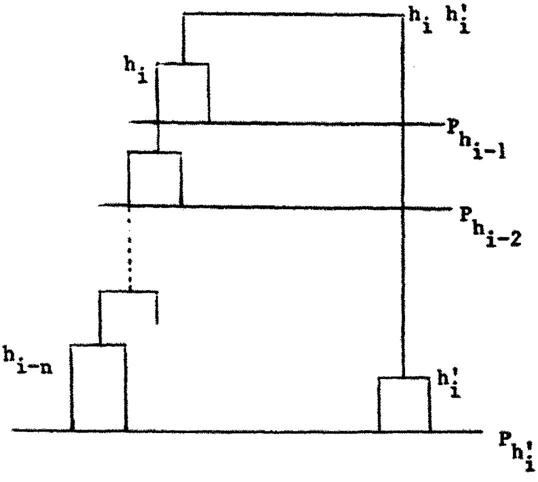


Figure 9

Il reste à montrer que (2) $\Rightarrow$ (1). Soit $h \subset h'$ avec h et h' dans H; il existe une suite finie $h_i, \ldots, h_n$ (voir *fig.* 10) avec $h = h_i \subset h_{i+1} \subset \ldots \subset h_n = h'$ où les h_j sont de taille croissante avec j et sont tous contenus dans h_n sinon (par définition d'une hiérarchie) ils seraient d'intersection vide avec h_n, ce qui n'est pas possible puisque ils contiennent h.

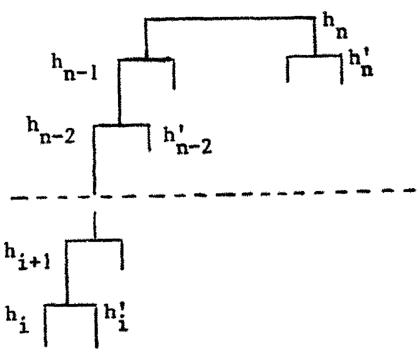


Figure 10

Si (2) est vrai, on a :

$$\frac{\begin{gathered} f(h_n) \geqq \operatorname{Max}\{f(h_{n-1}), f(h'_{n-1})\}, \\ f(h_{n-1}) \geqq \operatorname{Max}\{f(h_{n-2}), f(h'_{n-2})\} \end{gathered}}{f(h_{i+1}) \geqq \operatorname{Max}\{f(h_i), f(h'_i)\}},$$

d'où $f(h_n) \geqq f(h_i) \Rightarrow f(h') \geqq f(h)$ donc (1) est vrai. $\square$

REMARQUE : seule la condition (2) est applicable dans le cas de notre algorithme; en effet, la condition (3) suppose l'existence de la partition P_{h_i}.

Ainsi l'on effectera B_{i_0} à B_{i_0} (*cf.* 2.3) que si les conditions suivantes sont vérifiées :

$$f(B_{i_1} \cup B'_{i_0}) \geqq \operatorname{Max}[f(B_{i_0}), f(B_{i_1})]$$

et

$$f(B_{i_2}) \geqq f(B_{10} \cup B_{i_1}),$$

où B_{i_2} est le successeur immédiat de $B_{i_0} \cup B_{i_1}$.

4. CONCLUSION

L'approche qui a été présentée est intéressante du point de vue théorique car elle pose les problèmes de recherche de structures hiérarchiques en termes d'optimisation d'un critère global. Le prolongement naturel serait l'utilisation d'autres critères : par exemple, critère exprimant l'adéquation entre indice de distance de base et l'ultramétrique induite par la hiérarchie et les l'algorithmes correspondants. Dans [6] sont présentés l'étude d'autres critères et algorithmes.

Le théorème concernant le problème d'inversion sera utile dans la recherche de telles hiérarchies optimales indicées.

BIBLIOGRAPHIE

1. J. P. BENZECRI, *L'analyse des Données*, Dunod, 1973.
2. G. MILLIGAN, *Ultrametric Hierarchical Clustering Algorithms*, Psychometrica, vol. 44, n° 3, 1979.
3. M. BRUNOOGHE, *Classification automatique d'un grand tableau de données par la méthode des graphes réductibles*, Colloque I.R.I.A., septembre 1977.
4. J. D. CAROLL et S. PRUZANSKY, *Fitting of Hierarchical Tree Structure (H.T.S.) Models, Mixtures of H.T.S. Models and Hybrid Models via Mathematical Programming and Alterning Least Squares*, Bell Laboratoires, Murray Hill, New Jersey, 1974.
5. J. L. CHANDON, *Construction de l'ultramétrique la plus proche au sens des moindres carrés*, Note interne I.A.E., mai 1978.
6. R. CHIFFLET, *Structures hiérarchiques locales optimales* (*Thèse de 3^e cycle*, Paris, 1979).
7. R. M. CORMACK, *A Review of Classification*, J.R. Statistic Sol., vol. 134, part 3, 1971.

8. E. DIDAY, *Classification automatique séquentielle pour grands tableaux*, R.A.I.R.O., mars 1975.
9. E. DIDAY, *Problems of Clustering and Recent Advances*, Invited paper at the 11th European Congress of Statistics, Oslo, 1978. (Disponible à l'INRIA).
10. E. DIDAY, *Inversions en classification hiérarchique : application à la construction adaptative d'indices d'agrégation*, Revue de Statistiques Appliquées, vol. XXXI, n° 1, 1983.
11. E. DIDAY, J. LEMAIRE, J. POUGET et F. TESTU, *Éléments d'analyse des données*, Dunod, 1982.
12. RALAMBONDRAINY, et R. CHIFFLET, *Optimisation de classifications hiérarchiques*, Rapport de Recherche I.N.R.I.A., n° 70.