



HAL
open science

Analyse logique, combinatoire et statistique de la construction d'une hiérarchie binaire implicative ; niveaux et noeuds significatifs

Israël-César Lerman

► **To cite this version:**

Israël-César Lerman. Analyse logique, combinatoire et statistique de la construction d'une hiérarchie binaire implicative ; niveaux et noeuds significatifs. [Research Report] PI 1827, 2006, pp.45. inria-00115826v2

HAL Id: inria-00115826

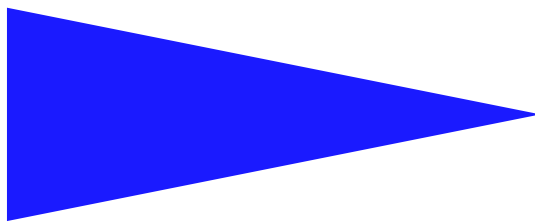
<https://inria.hal.science/inria-00115826v2>

Submitted on 24 Nov 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PUBLICATION
INTERNE
N° 1827



ANALYSE LOGIQUE, COMBINATOIRE ET STATISTIQUE DE LA
CONSTRUCTION D'UNE HIÉRARCHIE BINAIRE IMPLICATIVE ;
NIVEAUX ET NOEUDS SIGNIFICATIFS.

ISRAËL-CÉSAR LERMAN

Analyse logique, combinatoire et statistique de la construction d'une hiérarchie binaire implicative ; niveaux et noeuds significatifs.

Israël-César LERMAN*

Systèmes biologiques
Projet Symbiose

Publication interne n° 1827 — Novembre 2006 — 45 pages

Résumé : Nous reprenons ici d'une façon nouvelle et systématique l'étude d'un type spécifique d'analyse des données fondé sur la classification ascendante hiérarchique binaire. L'ensemble organisé est un ensemble d'attributs de description, généralement booléens. La relation évaluée de similarité représentée est de nature implicative et l'arbre obtenu est un arbre "implicatif". Ce type d'analyse des données a été introduit et développé par R. Gras et ses collaborateurs. Nous en étudions les fondements et proposons une axiomatique nouvelle qui conduit à des aspects constructifs et énumératifs. Une interprétation différente de celle des auteurs mentionnés, des résultats de ce type d'analyse des données est proposée et justifiée. Les propriétés mathématiques de la construction statistique de l'arbre implicatif sont étudiées. Enfin, nous analysons de façon précise et complète l'adaptation de nos critères de reconnaissance des niveaux et noeuds les plus "significatifs" d'un arbre de classification classique au cas d'un arbre binaire implicatif.

Mots clés : Analyse implicative, classification ascendante hiérarchique orientée, arbre implicatif, niveaux et noeuds "significatifs"

(Abstract: pto)

* Irisa - Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cédex — lerman@irisa.fr

Foundations, construction and significant nodes of an implicative binary hierarchy.

Abstract: Implicative data analysis is a specific method of data analysis, employing asymmetrical similarities and based on a construction of an ascendant binary hierarchical classification. This method has been introduced and developed by R. Gras and collaborators. Statistical descriptive attributes (generally, boolean attributes) are organized according to the latter structure, which is called : a binary implicative tree. In this technical report we reconsider in a new and systematic manner this approach. The analysis of its foundations leads us to a new axiomatic and to a new constructive and combinatorial properties. We give and justify a new interpretation of the results obtained by this type of data analysis method. The mathematical properties of the statistical construction of the binary implicative tree are studied. Finally, we analyze in a precise and complete manner our criteria for recognizing the most “significant” levels and the most “significant” nodes of the implicative binary tree.

Key-words: Implicative analysis, oriented ascendant hierarchical classification, implicative tree, “significant” levels, “significant” nodes

PLAN

1. Introduction

2. Expressions formelles et sémantiques de la hiérarchie implicative

- Hiérarchie de parties et arbre de classification associé; le cas binaire
- Hiérarchie de fourches, arbre binaire indicé de classification associé et expression ou propriétés formelles d'une hiérarchie implicative (orientée)
- Inadéquation entre la théorie de la démonstration et le sens formel d'une hiérarchie implicative

3. Construction statistique d'une hiérarchie implicative

- Introduction; indice d'implication entre classes d'attributs booléens
- Propriétés algorithmiques d'une Classification Ascendante Hiérarchique Implicative

4. Niveaux et noeuds "significatifs" d'une hiérarchie implicative

- Introduction
- Critère d'adéquation global
- Critère d'adéquation local
- Un nouveau critère local

5. Conclusion

1 Introduction

On suppose l'existence d'un univers \mathcal{U} d'objets sur lequel se trouve défini un ensemble \mathcal{A} d'attributs booléens. \mathcal{U} peut être assimilé à un ensemble potentiel, très généralement de très grande taille, sinon infinie et pratiquement inaccessible. Nous représentons un attribut booléen a au niveau de \mathcal{U} par son extension $\mathcal{U}(a)$ qui est le sous ensemble de \mathcal{U} où a est à *VRAI*.

a et b étant deux attributs booléens de \mathcal{A} , une règle $a \rightarrow b$, qui signifie que a *VRAI* implique b *VRAI*, s'exprime par l'inclusion $\mathcal{U}(a) \subset \mathcal{U}(b)$. Ne retenir que des inclusions logiques de cette forme pour comparer de façon relative l'ensemble des couples d'attributs de \mathcal{A} paraît particulièrement réducteur pour "comprendre" la sémantique des interrelations entre attributs de \mathcal{A} . Car des situations telles que $\mathcal{U}(a)$ est "presque" entièrement inclus dans $\mathcal{U}(b)$, ou celles où $\mathcal{U}(a)$ est "suffisamment" inclus dans $\mathcal{U}(b)$, ne sont pas distinguées. On voudrait pouvoir reconnaître des règles $a \Rightarrow b$ pouvant comprendre des exceptions [1] où a *VRAI* conduit à une tendance que b soit *VRAI*, tendance qu'il s'agira alors d'évaluer. D'ailleurs, même dans le cas de l'inclusion totale $\mathcal{U}(a) \subset \mathcal{U}(b)$, on ne peut considérer équivalents tous les cas de figure, indépendamment des cardinaux de $\mathcal{U}(a)$ et de $\mathcal{U}(b)$.

De toute façon, l'évaluation comparée des tendances implicatives entre les attributs de \mathcal{A} ne peut se faire au niveau de l'univers potentiel \mathcal{U} ; mais à celui d'un ensemble d'apprentissage (on dit encore échantillon) \mathcal{O} issu de \mathcal{U} . Il se fait que dans le domaine du "Data Mining", l'ensemble \mathcal{O} est le plus fréquemment de taille très élevée (e.g. de l'ordre de plusieurs centaines de milliers, sinon plus). Cette circonstance rend précise l'évaluation comparée des implications relativement à la définition d'un indice numérique adéquat pour mesurer une implication imprécise ou partielle sur \mathcal{U} [24, 4].

Précisément, l'indice qui est apparu le premier dans le domaine affiché de l'Extraction de Connaissances dans les Données (ECD) est celui dit de la "confiance" [1]. Il s'agit, relativement à la propension de l'implication $a \Rightarrow b$, de la proportion conditionnelle - calculée - au niveau de l'ensemble \mathcal{O} - $p(b/a) = \text{card}(\mathcal{O}(a \wedge b)) / \text{card}(\mathcal{O}(a))$, où $\mathcal{O}(\alpha)$ est le sous ensemble des objets de \mathcal{O} où l'attribut α est à *VRAI* et où $a \wedge b$ désigne la conjonction entre les deux attributs a et b .

Il paraît naturel et légitime que l'indice d'implication ait un caractère dissymétrique. Pourtant, se référant - de façon implicite ou explicite - à la notion générale de corrélation, certains indices proposés ont un caractère parfaitement symétrique par rapport au couple (a, b) . Sous l'appellation "mesure de la qualité d'une règle d'association", différents indices à caractère dissymétrique ou symétrique ont pu être proposés dans la littérature. L'analyse de l'élaboration et du fondement de tels indices est un problème majeur en vue de la détection et de l'organisation des règles pertinentes [38, 20, 8, 18, 30].

En effet, devant le très grand nombre de règles extraites, une telle organisation selon une structure de forme fixée permet de synthétiser l'information issue des données en vue de l'extraction de connaissances. Essentiellement, deux types de structures ont été considérées : les graphes valués et les arbres; surtout ceux de classification que nous nous limiterons à étudier ici. En retenant une notion symétrique de la similarité entre attributs, on peut à titre d'illustration citer les graphes de similitude de Cl. Flament [5] et les arbres de classification de la méthode AVL (Analyse de la Vraisemblance des Liens)[22, 28, 37].

Dans le cas d'une similarité implicative telle que celle, probabiliste pouvant être construite selon le principe général de la méthode AVL, R. Gras [6, 32, 22] a bâti des graphes d'implication, qu'il a alors appliqué dans le domaine de la didactique des mathématiques. Ce graphe est orienté et valué. Il est construit sur un ensemble \mathcal{A} d'attributs booléens où la valuation numérique portée sur un arc (a, b) est définie par la similarité implicative de a vers b . La fixation à partir de considérations de significativité statistique d'un seuil, permet d'obtenir un graphe discret sans circuit qu'on peut, par l'algorithme de Marimont décomposer en niveaux. De nombreuses applications dans divers domaines ont pu montrer tout l'intérêt de cette structure de condensation. On peut certes trouver des graphes de "causalité" dans d'autres contextes où c'est l'indice défini par la proportion (ou probabilité) conditionnelle qui est mis en oeuvre (e.g. les graphes bayésiens [3]); mais, ce qui distingue les graphes d'implication se trouve dans le sens qui sous-tend la nature statistique de la valuation. Ce qui, en particulier, permet de définir de façon uniforme et sans arbitraire un seuil qui permet à partir du graphe valué de déterminer un graphe discret.

En considérant une similarité symétrique sur \mathcal{A} , la condensation fournie par un arbre de classification est plus synthétique que celle d'un graphe valué discrétisé à partir d'un seuil. C'est que la structure d'arbre est plus spécifique. Elle induit naturellement une forme d'interprétation de résultats globaux que ne recouvre pas le graphe mentionné ci-dessus. On "découvre" des classes d'attributs qui s'emboîtent ou sont disjointes et dont chacune définit un profil de comportement. Dans ces conditions et dans le cas d'une similarité implicative, il peut être intéressant de développer face à un graphe implicatif, le versant "arbre implicatif de classification" qui correspondrait à un système implicatif de classes et sous-classes. C'est précisément la structure originale qui a été proposée par R. Gras et ses collaborateurs [13, 7, 11] sous l'appellation de "hiérarchie implicative".

Il s'agit de par sa construction d'un arbre binaire incomplet (sur une partie de \mathcal{A}) où chaque fusion (binaire) est orientée de gauche à droite. Par rapport à la clarté de la démarche constructive, nous donnerons quant à nous au paragraphe 2.2 la définition formelle de ce type de structure qui, à notre connaissance, n'a pas encore été exprimée. Nous donnerons également dans ce paragraphe une propriété caractéristique d'une forme d'ultramétrie en termes de triangles isocèles, mais "orientés". Ces définitions et propriétés résultent d'un souci de transfert du contexte d'une hiérarchie associée à un arbre binaire de classification dans le cas classique où le lien est symétrique. C'est pour cette raison que nous donnerons au paragraphe 2.1 une définition et un développement adaptés de cette dernière structure.

L'objet affiché de la recherche d'une hiérarchie implicative serait la découverte de "métarègles" telles que celle par exemple $a \rightarrow (b \rightarrow c)$, faisant ainsi un parallèle avec la théorie de la démonstration [11]. Le paragraphe 2.3 se veut l'analyse de cette affirmation. Nous chercherons à mettre clairement en évidence une inadéquation de nature entre le formalisme de la théorie de la démonstration et l'interprétation des résultats associée à l'obtention d'une hiérarchie implicative.

À la manière d'un arbre binaire de classification, la hiérarchie implicative est bâtie à partir d'une mesure de similarité implicative, donc orientée, entre couples de classes d'attributs. Le point de départ consiste en la définition d'une telle mesure entre classes singletons; laquelle étant directement déduite de la mesure de similarité implicative entre attributs. La condensation de l'ensemble des attributs sur une telle structure est sensée révéler des tendances implicatives entre classes ou sous classes d'attributs. Nous reprenons en 3 les choix adoptés dans [11]. Par rapport à ces choix nous établissons les correspondants de propriétés mathématiques ou algorithmiques qu'il est important de considérer dans la construction d'arbres de classification (monotonie, formule de réactualisation).

Il est également très important de pouvoir transférer à cette nouvelle structure d'arbre implicatif la méthode de reconnaissance des niveaux et noeuds intéressants que nous disons les plus "significatifs" d'un arbre de classification. Cette méthode fait partie intégrante de la classification ascendante hiérarchique AVL. Cette méthode de reconnaissance est basée sur le comportement le long de la suite des niveaux de l'arbre, d'une statistique non paramétrique normalisée, mesurant l'adéquation entre la partition produite à un niveau donné et la similarité dont se trouve muni l'ensemble organisé [22, 23, 31]. L'élaboration de cette mesure statistique se fait au niveau relationnel. Il s'agit de comparer la relation de partition et celle valuée par la similarité entre éléments de l'ensemble concerné. Deux versions de ce critère ont été développés. Pour la première, on retient une structure ordinale de la similarité et pour la seconde, une structure numérique. C'est relativement à la structure ordinale de la similarité ou "préordonnance", qu'une adaptation du critère est proposée [11]. Cette dernière est cohérente mais nous a semblé un peu rapide par rapport à l'analyse des structures à confronter. L'objet du paragraphe 4 consiste précisément à analyser finement la nature des structures à associer pour une définition conforme du critère d'adéquation. Diverses déclinaisons sont alors considérées où on retrouvera bien sûr celle correspondante à [11]. Mais surtout, on considérera une conception à partir de la prise en compte directe d'une similarité numérique au lieu de celle ordinale associée; ce qui permet un gain important en complexité calcul et une plus grande discrimination dans la prise en compte de la notion de ressemblance. Notre recherche nous conduit également à considérer une version globale d'un critère d'adéquation entre une hiérarchie classificatoire et une similarité [23, 29, 34] qui peut permettre selon le même principe général mais de façon plus analytique, la détermination de niveaux et noeuds "significatifs" d'un arbre de classification associé à une similarité binaire symétrique ou implicative.

Signalons pour terminer cette introduction que l'objet premier de cette étude était défini par le paragraphe 4. Il a été stimulé par une question qui nous avait été posée par P. Kuntz (École Polytechnique de l'Université de Nantes) lors des "Rencontres Francophones de Classification" (Bordeaux 2004). Pour répondre pleinement

à cette question il nous a fallu préciser les structures manipulées aussi bien sur les plans logique, formel que métrique et statistique; ce qui nous a conduit à l'ensemble du développement. Ce dernier reprend d'une manière spécifique la plupart des aspects méthodologiques liés à la construction d'une hiérarchie implicative.

2 Expressions formelles et sémantiques de la hiérarchie implicative

2.1 Hiérarchie de parties et arbre de classification associé; le cas binaire

Désignons ici par E l'ensemble organisé, dont on indiquera par n le cardinal. Un premier but fondamental de ce paragraphe est de définir ce qu'est formellement une hiérarchie implicative. Compte tenu du mode de sa construction ascendante [13, 11, 16], le concept se réfère à un arbre binaire de classification; mais où la relation entre deux classes est orientée de gauche à droite. Nous allons dans ces conditions partir de la notion classique de hiérarchie de parties sur un ensemble E pour ensuite préciser comment construire un arbre de classification associé compatible, comportant une indéxation ordinale de ses niveaux. L'ensemble de tels arbres est énumérable. Nous distinguerons le cas binaire, puisque c'est à partir de la formalisation spécifique de ce dernier que nous pourrions donner une expression formelle de la définition d'une hiérarchie implicative.

Définition 2.1 Une hiérarchie de parties $\mathcal{H}(\mathcal{E})$ de E est un ensemble de parties de E telles que :

1. $E \in \mathcal{H}(\mathcal{E})$
2. $\forall x \in E, \{x\} \in \mathcal{H}(\mathcal{E})$
3. $\forall (X \in \mathcal{H}(\mathcal{E}), Y \in \mathcal{H}(\mathcal{E})), X \cap Y = X$ ou $X \cap Y = Y$ ou $X \cap Y = \emptyset$

Nous dirons de plus qu'une telle hiérarchie de parties est *binnaire* si toute partie non réduite à un singleton peut être obtenue comme la réunion d'exactly deux parties disjointes de la hiérarchie; Nous notons $\mathcal{H}_b(E)$ une telle hiérarchie.

À une hiérarchie de parties $\mathcal{H}(\mathcal{E})$ nous pouvons associer et de différentes façons, un arbre de classification indexé par la fonction ordinale de *niveau* de formation d'un noeud. Les différentes façons correspondent aux différentes indéxations possibles des noeuds de l'arbre et dont chacun sous-tend précisément un élément de la hiérarchie $\mathcal{H}(\mathcal{E})$. Mais, commençons par définir la notion d' *arbre de classifications indicé par la notion de niveau*.

Définition 2.2 Un arbre de classification sur E indicé par la fonction de niveau est la représentation en termes de graphe d'une suite de partitions sur E de la forme $\{P_k \mid 0 \leq k \leq l+1\}$, totalement ordonnée par la relation de finesse. Dans cette suite, P_0 est la partition discrète à n classes singletons et P_{l+1} , est celle grossière à une seule classe E . P_{k+1} est moins fine que P_k et s'en déduit par la réunion de classes de P_k , $0 \leq k \leq l$. P_k occupe le k -ème niveau de l'arbre des classifications. Ainsi, P_0 occupe le niveau des feuilles de l'arbre et P_{l+1} , celui de la racine de l'arbre. Chaque classe de P_{k+1} - notons la E_j^{k+1} - est représentée au niveau $k+1$ par un noeud qui se trouve à la jonction de branches issues des noeuds du niveau k , correspondants aux classes dont la réunion est E_j^{k+1} ; la disposition latérale de ces derniers noeuds étant telle que ces branches ne se croisent pas, $0 \leq k \leq l$.

Cette dernière condition est algorithmiquement aisée à satisfaire.

Nous dirons qu'un arbre de classification est *binnaire* si et seulement si chaque réunion de classes permettant de passer de la partition P_k à la suivante P_{k+1} comporte exactement deux arguments et cela pour tout k , $0 \leq k \leq l$. Cependant, le passage entre P_k et P_{k+1} , plusieurs paires de classes sans composante commune peuvent fusionner "en même temps". Ainsi pour être explicite, si on considère par exemple que la partition P_k comporte cinq classes :

$$P_k = \{E_1^k, E_2^k, E_3^k, E_4^k, E_5^k\} \quad (1)$$

une partition P_{k+1} telle que :

$$P_{k+1} = \{E_1^k \cup E_2^k, E_3^k, E_4^k \cup E_5^k\} \quad (2)$$

résulte d'une fusion *binaires* ; mais, il n'en est pas de même de celle :

$$P'_{k+1} = \{E_1^k \cup E_2^k, E_3^k \cup E_4^k \cup E_5^k\} \quad (3)$$

puisqu'il existe une fusion comportant trois arguments.

Bien que nous ayons pu traiter le cas général [26], nous allons nous limiter aux arbres de classification binaires. En effet, c'est à partir de la formalisation de ces derniers que nous pourrions déduire celle des hiérarchies implicatives. Ainsi, nous allons montrer comment à partir d'une hiérarchie binaire de parties $\mathcal{H}_b(\mathcal{E})$, on peut construire et de différentes façons un arbre compatible de classification binaire. Plus précisément,

Définition 2.3 *Un arbre de classification sur E est compatible avec une hiérarchie de parties $\mathcal{H}(\mathcal{E})$ si et seulement si chacune des parties de $\mathcal{H}(\mathcal{E})$ se retrouve sous-tendant un noeud, une feuille ou la racine de l'arbre des classifications.*

Soit donc, comme justifié ci-dessus, une hiérarchie binaire de parties $\mathcal{H}_b(\mathcal{E})$. Nous définirons :

Définition 2.4 *Une chaîne complète de parties, éléments de $\mathcal{H}_b(\mathcal{E})$, est une suite $(X_0, X_1, \dots, X_g, X_{g+1}, \dots, X_{h+1})$ strictement croissante au sens de l'inclusion de parties, éléments de $\mathcal{H}_b(\mathcal{E})$, telle que X_0 est un singleton, $X_h = E$ et où aucun élément de $\mathcal{H}_b(\mathcal{E})$ peut correspondre à une partie strictement comprise entre deux parties X_g et X_{g+1} , $0 \leq g \leq h$.*

On peut voir que pour tout g , $0 \leq g \leq h$, on a nécessairement :

$$X_{g+1} = X_g \cup Y \quad (4)$$

où Y est une partie de E qui est un élément de $\mathcal{H}_b(\mathcal{E})$. En effet X_{g+1} est la réunion d'exactly deux parties, éléments de la hiérarchie. X_g en est une et $X_{g+1} - X_g = Y$ en est l'autre.

X_0 va être représenté par une feuille de l'arbre et X_{h+1} , par la racine. Autrement, chaque partie X_g ($1 \leq g \leq h$) donnera lieu à un noeud de l'arbre, où celui associé à X_{g+1} surplombe celui associé à X_g .

Chaque singleton $\{x\}$ ($x \in E$), va donner lieu à une chaîne complète de parties qui peut être construite récursivement. Relativement à une partie X_g de la suite construite à partir de $\{x\}$, la partie suivante X_{g+1} est définie comme étant celle de $\mathcal{H}_b(\mathcal{E})$ "la plus petite" contenant strictement X_g et qui est de la forme (4) ci-dessus. On peut désigner par :

$$\mathcal{C} = \{C_x \mid x \in E\} \quad (5)$$

l'ensemble des chaînes complètes obtenues à partir des différents singletons.

Maintenant, deux chaînes complètes se "rejoignent" nécessairement ; c'est-à-dire, que leurs sections finissantes sont identiques. Plus précisément, soient :

$$\begin{aligned} C_x &= (\{x\}, X_1, \dots, X_g, X_{g+1}, \dots, X_{h+1}) \\ C_y &= (\{y\}, Y_1, \dots, Y_j, Y_{j+1}, \dots, Y_{k+1}) \end{aligned} \quad (6)$$

deux chaînes complètes issues de x et y ($x \neq y$). Il existe nécessairement un indice g_0 et un indice j_0 à partir desquels :

$$(X_{g_0+1}, \dots, X_h) = (Y_{j_0+1}, \dots, Y_k) \quad (7)$$

En effet, car de toute façon $X_{h+1} = Y_{k+1} = E$. À titre d'illustration, considérons l'exemple suivant où E comprend 15 éléments codés de 1 à 15. Les deux chaînes complètes suivantes :

$$\begin{aligned}
C_1 &= (\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \\
&\quad \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}) \\
C_5 &= (\{5\}, \{5, 6\}, \{5, 6, 7\}, \{5, 6, 7, 8\}, \{5, 6, 7, 8, 9\}, \{5, 6, 7, 8, 9, 1, 2, 3, 4\}, \\
&\quad \{5, 6, 7, 8, 9, 1, 2, 3, 4, 10, 11, 12, 13, 14\}, \{5, 6, 7, 8, 9, 1, 2, 3, 4, 10, 11, 12, 13, 14, 15\})
\end{aligned} \tag{8}$$

se rejoignent sur la partie $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. On a $g_0 = 4$ et $j_0 = 5$.

Définition 2.5 *Une chaîne complète C' est moins fine qu'une chaîne complète C si et seulement si la suite des parties, non compris le singleton de départ, est une section finissante de C .*

Ainsi, la chaîne complète :

$$\begin{aligned}
&(\{4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \\
&\quad \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\})
\end{aligned}$$

est moins fine que celle C_1 ci-dessus.

Il s'agit à présent à partir de l'ensemble des n chaînes complètes de montrer comment on peut spécifier un arbre binaire indicé de classification, compatible avec une hiérarchie binaire de parties $\mathcal{H}_b(\mathcal{E})$ sur E . Dans cette construction, on mettra en évidence et de proche en proche l'énumération de tels arbres. Nous allons dans ces conditions exprimer de façon informelle l'algorithme de construction et l'énumération qui en découle.

2.1.1 Algorithme

L'état initial est défini par une chaîne complète C et bien que cela ne soit pas obligé; mais pour des raisons de rapidité calcul, on partira de la chaîne complète la plus longue. Notons la sous la forme :

$$C = ((X_0 = \{x\}, X_1, \dots, X_g, X_{g+1}, \dots, X_h, X_{h+1} = E))$$

X_0 et X_{h+1} définiront respectivement une feuille de l'arbre des classifications et sa racine. X_1, X_2, \dots, X_h définiront des noeuds propres de l'arbre des classifications. La première valeur attribuée aux niveaux de formation de ces noeuds se trouve définie par les rangs de la suite ci-dessus; il s'agira ainsi de l'entier g pour le noeud surplombant la partie X_g .

Écartons les chaînes complètes comparables à C (moins fines) et dans l'ensemble des autres chaînes complètes considérons celle la plus longue que nous notons :

$$D = (Y_0 = \{y\}, Y_1, \dots, Y_j, Y_{j+1}, \dots, Y_k = Y_{k+1} = E)$$

Imaginons que la jonction des deux chaînes se fasse pour $g = g_0 + 1$ et $j = j_0 + 1$; soit :

$$(X_{g_0+1}, \dots, X_h) = (Y_{j_0+1}, \dots, Y_k)$$

Il s'agit dans ces conditions d'inscrire la formation des noeuds surplombant Y_1, Y_2, \dots, Y_{j_0} par rapport à ceux, déjà établis et concernant X_1, \dots, X_{g_0} . Il est entendu que dans cette insertion, un même niveau peut être associé soit à un X_g seul, soit à un Y_j seul, soit enfin à un X_g et à un Y_j . En effet, comme nous l'avons mentionné ci-dessus, l'arbre est binaire mais plus d'une fusion peut apparaître à un niveau donné.

$Y_0 = \{y\}$ va occuper le niveau 0 des feuilles. Nous allons, de façon récursive et de proche en proche étudier le positionnement des noeuds associés aux Y_j par rapport à ceux établis pour les X_g ($1 \leq j \leq j_0, 1 \leq g \leq g_0$) et la transformation de la valeur des niveaux qui en résulte.

Relativement aux noeuds associés aux $X_g, 1 \leq g \leq g_0$, considérons la subdivision :

$$(1, 2, \dots, g, g+1, \dots, g_0)$$

Le noeud associé à Y_1 peut être placé soit dans un intervalle strict de la forme $]g, g+1[$ ($0 \leq g \leq g_0$ où $g = 0$ (resp. g_0)) correspond à l'intervalle avant 1 (resp., après g_0)) soit sur un niveau d'indice g .

Dans le cas où Y_1 est placé strictement entre g et $g + 1$, on attribue au noeud associé à Y_1 le niveau $g + 1$ et les niveaux associés à la suite croissante des noeuds relative à

$$X_{g+1}, \dots, X_{g_0}, X_{g_0+1}, \dots, X_h, X_{h+1} = E)$$

se trouvent augmentés d'une unité. Dans ce cas à chaque niveau il se crée une fusion binaire.

Dans le cas où Y_1 est placé au niveau g , on attribue au noeud associé à Y_1 le niveau g et les niveaux associés à la suite croissante des noeuds ci-dessus, restent invariants. Dans ce cas, au niveau g il se produit deux fusions binaires surplombant respectivement X_g et Y_1 .

Dans le premier cas où le niveau du noeud Y_1 est $g + 1$, le choix de Y_2 s'effectue de la même manière que celle de Y_1 ; mais par rapport à la nouvelle subdivision

$$(g + 2, g + 3, \dots, g_0 + 1)$$

Dans le second cas où le niveau du noeud Y_1 est g , le choix de Y_2 s'effectue de la même manière que celle de Y_1 ; mais par rapport à la nouvelle subdivision

$$(g + 1, g + 2, \dots, g_0)$$

On peut remarquer que l'ensemble X_{g_0} défini avant le placement de Y_1 , sous-tend après ce placement, le noeud $g_0 + 1$ dans le premier cas et celui g_0 dans le second.

Le placement de Y_{j+1} après celui de Y_j suit exactement le même processus, $1 \leq j \leq j_0 - 1$. La hauteur totale de l'arbre binaire obtenu jusqu'à la racine est comprise entre $h + 1$ et $j_0 + h + 1$. La valeur $h + 1$ concerne le cas où Y_1, Y_2, \dots, Y_{j_0} se placent tous à des niveaux déjà occupés par des noeuds existants associé aux X_g . Celle $j_0 + h + 1$, concerne le cas où Y_1, Y_2, \dots, Y_{j_0} se placent tous à des niveaux différents de ceux déjà occupés par un noeud.

Donnons ici une définition avant de poursuivre.

Définition 2.6 C, C' et C'' étant trois chaînes complètes et mutuellement incomparables au sens de la définition 2.5, C est dite plus proche de C' que de C'' si le premier noeud de C à se retrouver dans C' est de rang plus petit que celui à se retrouver dans C'' .

Ainsi, relativement à l'exemple précédent, en prenant

$$C = (\{13\}, \{13, 14\}, \{10, 11, 12, 13, 14\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \\ \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\})$$

$$C' = (\{10\}, \{10, 11\}, \{10, 11, 12\}, \{10, 11, 12, 13, 14\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \\ \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\})$$

et

$$C'' = (\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \\ \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\})$$

C est plus proche de C' que de C'' ; en effet, le rang du premier noeud de C à se retrouver dans C' est 3, alors que le rang du premier noeud de C à se retrouver dans C'' est 4.

Dans ces conditions, imaginons qu'on se trouve après la i -ème étape de l'algorithme et qu'on ait constitué un arbre indicé par la fonction de niveau à partir de la suite de chaînes complètes

$$\{C_j \mid 1 \leq j \leq i\}$$

On pourra remarquer qu'à partir d'un certain niveau et avant que la construction ascendante de l'arbre ne soit achevée, la partie finale de cet arbre est unaire et correspond à une section finissante de la chaîne complète la plus longue (voir l'exemple ci-dessus).

Le choix de la $(i + 1)$ -ème chaîne et son "accrochage" à l'arbre construit se fait comme suit :

- dans l'ensemble des chaînes restantes, mutuellement incomparables et incomparables avec les i premières chaînes, on détermine la chaîne la plus longue qu'on désigne par C_{i+1} ;
- dans l'ensemble des i premières chaînes on détermine celle, d'indice j_0 dont C_{i+1} est la plus proche ;
- on effectue la jonction entre C_{i+1} et C_{j_0} de la même façon que la jonction de D à C ci-dessus a été effectuée, mais à une différence près que nous allons préciser.

En se situant avant la jonction entre C_{i+1} et C_{j_0} , si on considère deux noeuds consécutifs, ou bien une feuille et un noeud qui lui succède, de C_{j_0} de niveaux respectifs r et s ($r < s$) et le placement d'un noeud de C_{i+1} par rapport à l'intervalle $[r, s[$. Ce noeud peut être placé à l'un des niveaux $r, r + 1, \dots, s - 1$ et alors, l'affectation acquise des niveaux pour l'ensemble des noeuds installés reste invariable. Mais, si ce noeud est placé strictement entre deux noeuds consécutifs de rangs $r + u$ et $r + u + 1$ ($\leq s$), alors le noeud se voit attribué le niveau $r + u + 1$ et tous les niveaux de valeur supérieure ou égale à $r + u + 1$ voient leurs valeurs augmentées d'une unité.

2.1.2 Énumération

Le problème que nous nous posons maintenant concerne le nombre de façons dont on peut réaliser l'acrochage de C_{i+1} à C_{j_0} en tenant compte de la construction déjà acquise l'arbre des classifications.

On pourra commencer par remarquer que, compte tenu du mode de construction de l'arbre, la "longueur" (nombre de noeuds) de C_{i+1} est inférieure à celle de C_{j_0} . Désignons par N le nombre de noeuds de C_{i+1} strictement compris entre le singleton de départ et le noeud de jonction avec C_{j_0} . Considérons à présent l'intervalle commençant des indices des noeuds acquis dans la formation de l'arbre jusqu'au noeud de C_{j_0} qui se retrouve dans C_{i+1} et qui correspond donc à la jonction entre C_{i+1} et C_{j_0} . Indiquons cet intervalle sous la forme :

$$A = [1, 2, \dots, n, \dots, \nu, \nu + 1] \tag{9}$$

Il s'agit maintenant d'affecter des rangs aux noeuds de C_{i+1} qui se situent entre le singleton et le noeud de jonction avec C_{j_0} . Un noeud de C_{i+1} peut être situé à un niveau déjà acquis entre 2 et ν et dans ce cas la suite (9) reste invariante. Il peut être situé entre deux noeuds consécutifs n et $n + 1$, $0 \leq n \leq \nu$ ($n = 0$ correspond au singleton et $n = \nu + 1$, au noeud de jonction). Dans ce dernier cas, le noeud affecté reçoit la valeur $n + 1$ de son indice et tous les indices consécutifs de (9) sont augmentés d'une unité. Compte tenu de la binarité de l'arbre, deux noeuds consécutifs de C_{i+1} ne peuvent recevoir la même valeur n de la suite A , alors qu'une suite de plusieurs noeuds consécutifs de C_{i+1} peut se placer dans l'intervalle $]n, n + 1[$ ($0 \leq n \leq \nu$). Représentons cet intervalle par son milieu $(2n + 1)/2$ et considérons la suite C :

$$C = \left[\frac{1}{2}, 1, \frac{3}{2}, \dots, n, \frac{2n + 1}{2}, n + 1, \dots, \nu, \frac{2\nu + 1}{2} \right] \tag{10}$$

Cet ensemble peut se décomposer en deux parties complémentaires; celle A ci-dessus et

$$B = \left[\frac{1}{2}, \frac{3}{2}, \dots, \frac{2n + 1}{2}, n + 1, \dots, \frac{2\nu + 1}{2} \right] \tag{11}$$

qui est l'ensemble des $\nu + 1$ premiers multiples impairs de $\frac{1}{2}$.

L'ensemble des manières d'accrocher C_{i+1} à C_{j_0} pour une augmentation de la formation de l'arbre, peut ainsi être mis en correspondance bijective avec l'ensemble des applications *croissantes* f de

$$E = [1, 2, \dots, e, \dots, N] \quad (12)$$

dans C , telles que chaque valeur de A est atteinte au plus une fois; soit :

$$(\forall a \in A), \quad \text{card}(f^{-1}(a)) = 0 \text{ ou } 1 \quad (13)$$

Théorème 2.1 *Le nombre de telles applications est donné par la formule :*

$$\sum_{m=0}^{\nu} C_{\nu}^m \times C_{\nu+N-m}^{\nu} \quad (14)$$

Pour établir cette formule nous allons considérer le cas plus général de l'énumération de l'ensemble \mathcal{F} des applications croissantes de E dans un ensemble

$$C = \{1, 2, \dots, c, \dots, M\}$$

dont les éléments de l'un de ses sous ensembles fixé, sont atteints au plus une fois. Plus précisément, considérons une décomposition de C en deux ensembles complémentaires A et B de cardinaux respectifs K et L ($K+L = M$) tels que chaque élément de A est atteint au plus une fois. A est fixé. Une application f de \mathcal{F} est telle qu'on ait la condition (13) ci-dessus.

Associons à f l'application g de C dans E définie comme suit :

$$\forall c \in C, c \mapsto g(c) = \text{card}(f^{-1}(c)) \quad (15)$$

On a bien sûr

$$\forall a \in A, g(a) = 0 \text{ ou } 1 \quad (16)$$

et

$$\sum_{c \in C} g(c) = N \quad (17)$$

En désignant par \mathcal{G} l'ensemble de telles applications, il y a une correspondance bijective entre \mathcal{F} et \mathcal{G} .

Commençons la construction de g par celle g_A de sa restriction à A . À cette fin désignons par I_{α} le sous ensemble de A , choisi de taille α , où g vaut 1. Relativement à la restriction g_B de g sur B , que nous noterons :

$$B = \{b_l \mid 1 \leq l \leq L\}$$

avec $b_1 < b_2 < \dots < b_l < \dots < b_L$, on a :

$$\sum \{g_B(b) \mid b \in B\} = N - \alpha \quad (18)$$

La fonction cumulée que nous noterons g'_B de B dans $E_{\alpha} = \{1, 2, \dots, N - \alpha\}$:

$$g'_B(b) = \sum \{g(b') \mid b' \leq b, b, b' \in B\} \quad (19)$$

est une fonction bijectivement associée à g_B , croissante au sens large.

Pour α fixé, le nombre d'applications g_A est C_K^{α} puisque chacune correspond au choix de l'un des ensembles I_{α} . D'autre part, le nombre d'applications croissantes (au sens large) d'un ensemble de cardinal L dans

un ensemble de cardinal $N - \alpha$; mais où, pour le dernier élément b_L de B , $g'(b_L) = N - \alpha$, correspond à l'énumération des fonctions croissantes de $\{b_1, b_2, \dots, b_{L-1}\}$ dans E_α . Il s'agit donc de :

$$C_{L+N-\alpha-1}^{L-1} = C_{L+N-\alpha-1}^{N-\alpha} \quad (20)$$

Il en résulte, en considérant les différentes valeurs possibles de α ($0 \leq \alpha \leq K$) que

$$\text{card}(\mathcal{F}) = \sum_{\alpha=0}^K C_K^\alpha \times C_{L+N-\alpha-1}^{L-1} \quad (21)$$

Remarquons que dans le cas précédent $K = \nu$ et $L = \nu + 1$. L'expression ci-dessus concernée peut se mettre sous la forme :

$$\sum_{m=0}^{\nu} \frac{(\nu + N - m)!}{m!(\nu - m)!(N - m)!}$$

qui est une somme de coefficients multinomiaux.

2.2 Hiérarchie de fourches, arbre binaire indicé de classification associé et expression ou propriétés formelles d'une hiérarchie implicative (orientée)

L'intérêt du paragraphe précédent concernant l'introduction des hiérarchies binaires et la construction - à notre connaissance nouvelle - d'arbres indicés de classification compatibles se justifie par lui-même. Cependant, si nous avons introduit ces objets ou développements, c'est que nous avons en vue une formalisation claire de la notion de hiérarchie implicative ou orientée. Notre formalisation n'est pas incompatible avec l'expression [16], elle aura toutefois, un accent très différent.

Par rapport à ce qui existe dans la littérature concernant une relation symétrique sur E (notion de hiérarchie de parties), la notion de hiérarchie binaire est la plus proche de celle visée. Néanmoins, il faut définir une autre notion de hiérarchie binaire de parties qui conduira directement à la notion de hiérarchie implicative. Il s'agira de ce que nous appellerons une "hiérarchie de fourches".

Définition 2.7 Une fourche de E est une paire $\{X, Y\}$ de parties disjointes de E , où X et Y sont des parties propres de E : $X \subset E, Y \subset E, X \neq \emptyset, Y \neq \emptyset$, et $X \cap Y = \emptyset$.

Une fourche est figurée par un sommet représentant $X \cup Y$ duquel sont issues de façon descendante deux branches reliant le sommet précédent à deux sommets représentant respectivement X et Y , à la manière de la figure suivante :

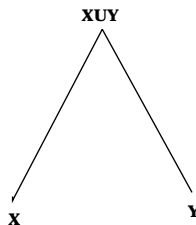


FIG. 1 – Fourche non orientée

Nous allons maintenant définir la notion de *hiérarchie de fourches* et celle associée, de *hiérarchie complète de fourches* sur E . Nous notons $\mathcal{HF}(\mathcal{E})$ la première et $\mathcal{H}_c\mathcal{F}(\mathcal{E})$, la seconde.

Définition 2.8 Une hiérarchie de fourches $\mathcal{HF}(\mathcal{E})$ est un ensemble de fourches telles que si $\{X, Y\}$ est une fourche et si X (resp., Y) n'est pas un singleton, c'est qu'il s'agit du sommet d'une fourche. D'autre part, si $\{X, Y\}$ et $\{Z, T\}$ sont deux fourches pour lesquelles - sans restreindre la généralité - $\text{card}(X \cup Y) \leq \text{card}(Z \cup T)$, on a l'un des cas suivants :

1. $(X \cup Y) \cap (Z \cup T) = \emptyset$
2. $X \cup Y \subset Z$ ou $X \cup Y \subset T$
3. Pour tout noeud Z sommet d'une fourche, il existe exactement une fourche $\{X, Y\}$ telle que $Z = X \cup Y$.
Si de plus, on a les conditions :
4. pour tout singleton $\{x\} (x \in E)$, il existe au moins une fourche dont l'une des composantes est ce singleton
5. Il existe exactement une fourche $\{X, Y\}$ pour laquelle $X \cup Y = E$;
alors, il s'agit d'une hiérarchie complète de fourches $\mathcal{H}_c\mathcal{F}(\mathcal{E})$.

Associons à une hiérarchie binaire $\mathcal{H}_b(\mathcal{E})$ celle qu'on obtient en enlevant les singletons (parties réduites à un élément) et que nous notons $\mathcal{H}_b^{-s}(\mathcal{E})$:

$$\mathcal{H}_b^{-s}(\mathcal{E}) = \mathcal{H}_b(\mathcal{E}) - \{\{x\} \mid x \in E\} \quad (22)$$

On peut de façon naturelle associer à une telle hiérarchie binaire de la forme $\mathcal{H}_b^{-s}(\mathcal{E})$, une hiérarchie complète de fourches que nous notons (comme ci-dessus) $\mathcal{H}_c\mathcal{F}(\mathcal{E})$. En effet, par définition, chaque partie \mathcal{Z} de E ($\mathcal{Z} \subset E$), non réduite à un singleton, appartenant à $\mathcal{H}_b(\mathcal{E})$, est exactement la réunion de deux parties disjointes X et Y (sous ensembles propres de E). Ainsi, nous associons à \mathcal{Z} , la fourche $\{X, Y\}$ dont \mathcal{Z} représente le sommet et X et Y les noeuds pendants.

Inversement, à une hiérarchie complète de fourches de la forme $\mathcal{H}_c\mathcal{F}(\mathcal{E})$, on peut associer une hiérarchie binaire de la forme $\mathcal{H}_b^{-s}(\mathcal{E})$. En effet, on associe à chaque sommet de $\mathcal{H}_c\mathcal{F}(\mathcal{E})$, la partie \mathcal{Z} qu'il représente. On a bien que \mathcal{Z} peut se mettre sous la forme

$$\mathcal{Z} = X \cup Y \quad (23)$$

où X (resp., Y) est le sommet d'une fourche s'il ne s'agit pas d'un singleton (voir Définition 2.8). En désignant par \mathcal{F} la correspondance ainsi définie entre l'ensemble $\mathcal{H}_b^{-s}(\mathcal{E})$ et celui $\mathcal{H}_c\mathcal{F}(\mathcal{E})$ qui lui correspond par construction, on a :

Théorème 2.2 *La correspondance \mathcal{F} est bijective.*

Maintenant, nous avons vu comment construire un arbre binaire indicé associé à une hiérarchie binaire, pouvant comporter plusieurs noeuds à un niveau donné. Nous avons surtout à partir de la reconnaissance de la nature du problème, énuméré l'ensemble des solutions possibles. La correspondance bijective établie ci-dessus entre une hiérarchie binaire de la forme $\mathcal{H}_b^{-s}(\mathcal{E})$ et une hiérarchie complète de fourches de la forme $\mathcal{H}_c\mathcal{F}(\mathcal{E})$, autorise la même construction.

Comme nous l'avons annoncé, c'est la notion de hiérarchie de fourches qui va être la plus directement étendue à la notion dissymétrique de hiérarchie implicative.

Définition 2.9 *Une fourche orientée de E est définie par la donnée d'un couple (A, B) de parties propres et disjointes de E remplissant les conditions suivantes :*

1. $A \subset E, B \subset E$;
2. $A \neq \emptyset, A \neq E, B \neq \emptyset, B \neq E$;
3. $A \cap B = \emptyset$.

Un tel couple traduit l'implication $A \rightarrow B$, où A est situé à gauche de B , ayant ainsi une jonction orientée où A est concaténée à B , à sa gauche. La fourche définie par (A, B) peut être figurée comme suit (voir FIG. 2) :

Le sommet de la fourche définit la jonction orientée de A vers B . Ce sommet représente $A \cup B$ dès lors qu'il définit le noeud pendant d'une fourche qui succède de façon ascendante à cette dernière.

L'axiomatique qui régit une hiérarchie implicative (orientée) *complète* est en tout point analogue à celle d'une hiérarchie complète de fourches $\mathcal{H}_c\mathcal{F}(\mathcal{E})$ à cela près qu'au lieu de s'agir d'un ensemble de *paires* de parties de E , il s'agit d'un ensemble de *couples* de parties de E . Nous allons dans ces conditions reprendre la définition 2.8 pour l'adapter au cas orienté. Nous noterons $\mathcal{HI}_c(\mathcal{E})$ une hiérarchie implicative complète.

Définition 2.10 *Une hiérarchie implicative complète $\mathcal{HI}_c(\mathcal{E})$ sur E est un ensemble de fourches orientées (A, B) de E telles que si (X, Y) est une fourche orientée et si X (resp., Y) n'est pas un singleton, c'est qu'il s'agit du sommet d'une fourche. D'autre part, si (X, Y) et (Z, T) sont deux fourches pour lesquelles - sans restreindre la généralité - $\text{card}(X \cup Y) \leq \text{card}(Z \cup T)$, on a l'un des cas suivants :*

(A \rightrightarrows B, AUB)

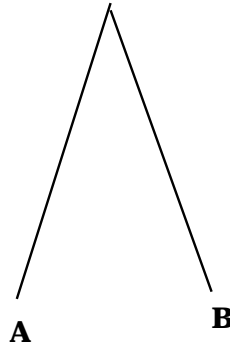


FIG. 2 – Fourche orientée

1. $(X \cup Y) \cap (Z \cup T) = \emptyset$
2. $X \cup Y \subset Z$ ou $X \cup Y \subset T$
3. Pour tout noeud Z de la hiérarchie, il existe exactement une fourche orientée (X, Y) telle que $X \cup Y = Z$.
On a de plus les conditions suivantes :
4. Pour tout singleton $\{x\} (x \in E)$, il existe au moins une fourche dont l'une des composantes est ce singleton.
5. Il existe exactement une fourche orientée (X, Y) pour laquelle $X \cup Y = E$.

Maintenant, en omettant les deux dernières conditions 4 et 5, on aura défini la notion de hiérarchie implicative non nécessairement complète sur E et que nous noterons $\mathcal{HI}(\mathcal{E})$.

Revenons ici à la notion de hiérarchie binaire de parties (voir Définition 2.1) pour une extension au cas où cette hiérarchie est incomplète. Pour une telle hiérarchie que nous notons $\mathcal{H}_{bi}(\mathcal{E})$, la condition 1 n'est pas requise. La définition axiomatique devient :

Définition 2.11 Une hiérarchie binaire incomplète de parties $\mathcal{H}_{bi}(\mathcal{E})$ est un ensemble de parties de E tel que :

1. $(\forall x \in E), \{x\} \in \mathcal{H}_{bi}(\mathcal{E})$
2. $\forall (X \in \mathcal{H}_{bi}(\mathcal{E})), X \cap Y = X$ ou $X \cap Y = Y$ ou $X \cap Y = \emptyset$
3. Toute partie non réduite à un singleton peut être obtenue comme la réunion d'exactly deux parties disjointes de la hiérarchie.

La donnée d'une hiérarchie binaire incomplète de parties $\mathcal{H}_{bi}(\mathcal{E})$ induit la donnée d'une partition :

$$P = \{E_1, E_2, \dots, E_j, \dots, E_k\} \quad (24)$$

de E , telle que la restriction de la hiérarchie $\mathcal{H}_{bi}(\mathcal{E})$ sur chacune des classes E_j , non réduite à un singleton, est une hiérarchie binaire complète $\mathcal{H}_b(E_j)$. Certaines classes E_j peuvent être réduites à un singleton, auquel cas $\mathcal{H}_b(E_j)$ n'est pas définie.

Inversement, à partir des différentes hiérarchies complètes $\mathcal{H}_b(E_j)$, $1 \leq j \leq k$, on peut obtenir une hiérarchie binaire incomplète compatible de la forme $\mathcal{H}_{bi}(\mathcal{E})$, en construisant une fonction surjective de l'ensemble des noeuds des différentes hiérarchies de la forme $\mathcal{H}_b(E_j)$, dans un intervalle commençant $[1, 2, \dots, m]$ de l'ensemble des entiers. Cette fonction est telle qu'elle est strictement croissante sur chaque chaîne complète de chaque $\mathcal{H}_b(E_j)$. Cela suppose bien sûr que m est supérieur ou égal à la longueur de la plus longue chaîne complète. L'énumération de telles fonctions suit un raisonnement déjà considéré au paragraphe 2.1 précédent.

Introduisons ici la hiérarchie binaire incomplète; mais singletons non compris que nous notons $\mathcal{H}_{bi}^{-s}(\mathcal{E})$ et qui est définie de la même manière que $\mathcal{H}_{bi}(\mathcal{E})$ à partir de $\mathcal{H}_b(\mathcal{E})$:

$$\mathcal{H}_{bi}^{-s}(\mathcal{E}) = \mathcal{H}_{bi}(\mathcal{E}) - \{\{x\} \mid x \in E\} \quad (25)$$

Maintenant, concernant une hiérarchie implicative $\mathcal{HI}(\mathcal{E})$, en associant à chaque sommet de l'une de ses fourches orientées (A, B) , la partie $A \cup B$ de E , on obtient une correspondance bijective avec une hiérarchie binaire incomplète de parties, à laquelle on aura ôté les singletons, donc de la forme $\mathcal{H}_{bi}^{-s}(\mathcal{E})$.

2.2.1 Ultramétrie

Cette propriété a été considérée dans [12]. Elle est relative à l'indice de dissimilarité entre classes adopté dans la construction ascendante d'une hiérarchie implicative. L'indice de dissimilarité entre deux éléments de l'ensemble organisé est égal à l'indice de dissimilarité entre les deux classes composant une fourche et auxquelles appartiennent respectivement les deux éléments. Le caractère monotone de cet indice le long de la suite des niveaux assure son caractère ultramétrique. Notre approche est d'une autre nature, elle est liée à notre formalisme et est purement ordinale, sans référence à un critère de dissimilarité particulier entre classes. Dans notre cas le caractère ultramétrique mis en évidence est une conséquence immédiate de la correspondance entre une hiérarchie binaire dans le cas classique symétrique et une hiérarchie binaire dans le cas implicatif.

Relativement à une hiérarchie implicative ou orientée $\mathcal{HI}(\mathcal{E})$ sur E , nous avons associé à chacune de ses fourches orientées (A, B) le noeud d'une hiérarchie binaire représentant $A \cup B$ qui correspond au sommet de la fourche. Une telle hiérarchie est généralement incomplète et a été notée $\mathcal{H}_{bi}(\mathcal{E})$ (voir Définition 2.11). Nous avons pu associer à une telle hiérarchie plusieurs fonctions ordinales compatibles de l'ensemble de ses noeuds dans un intervalle commençant $[1, 2, \dots, m]$ de l'ensemble des entiers. Fixons en une, quelconque, que nous notons ν . Cette dernière se présente sous la forme :

$$\nu : \mathcal{H}_{bi}^{-s}(\mathcal{E}) \rightarrow [1, 2, \dots, m] \quad (26)$$

elle attribue à chacun des noeuds, un rang définissant le niveau de sa formation.

Soit un noeud N de $\mathcal{H}_{bi}(\mathcal{E})$ surplombant un sous ensemble de la forme $A \cup B$ où A et B sous-tendent chacun un noeud ou un singleton de $\mathcal{H}_{bi}(\mathcal{E})$. La valeur de la fonction δ symétrique que nous allons définir sur $E \times E$ est telle que :

$$(\forall (a, b) \in A \times B), \delta(a, b) = \delta(b, a) = \nu(N) \quad (27)$$

Plus précisément, la fonction δ :

$$\delta : E \times E \rightarrow [1, 2, \dots, m, \infty]$$

associe à chaque couple d'éléments (x, y) de $E \times E$, le rang du niveau du premier noeud où la *paire* d'éléments $\{x, y\}$ se trouve réunie. Dans le cas où l'un des deux éléments x ou y reste seul dans sa classe, ce qui correspond alors à un singleton, c'est la valeur ∞ qui est attribuée par δ .

Il est classique et d'ailleurs aisé de voir que δ définit une distance ultramétrique sur E caractérisée par la propriété :

$$(\forall x, y, z \in E), \delta(x, y) \leq \max\{\delta(x, z), \delta(y, z)\} \quad (28)$$

En effet, le niveau de première réunion de x et y ne peut être strictement supérieur à chacun, de x et z d'une part et de y et z d'autre part. À partir de (27) on démontre que tout triangle $\{x, y, z\}$ est isocèle, la base étant le plus petit des côtés. Plus précisément, si on suppose, sans restreindre la généralité, que :

$$\delta(x, y) \leq \delta(x, z) \leq \delta(y, z) \quad (29)$$

on établit aisément que :

$$\delta(x, z) = \delta(y, z)$$

Néanmoins, ce qu'on a de plus compte tenu du caractère binaire de l'arbre, est le caractère strict de la première des inégalités de (29), soit :

$$\delta(x, y) < \delta(x, z) = \delta(y, z)$$

Inversement, la donnée d'une distance ultramétrique à valeurs dans $[1, 2, \dots, m, \infty]$ et pour laquelle tout triangle est isocèle avec une base strictement plus petite que chacun des deux côtés, permet de dériver une hiérarchie binaire incomplète $\mathcal{H}_{bi}(\mathcal{E})$; ainsi d'ailleurs qu'un arbre binaire indicé de classification compatible qui lui est associé.

À cet égard, on partira de la notion de chaîne complète des parties $\mathcal{H}_{bi}(\mathcal{E})$ qui reprend la définition 2.4

en enlevant la condition $X_h = E$. $\mathcal{H}_{bi}(\mathcal{E})$ s'obtient en regroupant en tant que ses éléments, l'ensemble des parties de E qui définissent des composantes de chaînes complètes mutuellement incomparables et qui sont directement obtenues à partir de la donnée de la distance ultramétrique δ . L'algorithme exprimé de façon informelle se déroule comme suit :

1. Choisir un élément quelconque x de E et établir le vecteur des similarités ultramétriques à x :

$$\{\delta(x, y) \mid y \in E - \{x\}\}. \quad (30)$$

2. Trier la famille de ces valeurs et établir la suite strictement croissante des valeurs atteintes que nous notons :

$$0 < \delta_1^x < \delta_2^x < \dots < \delta_k^x < \dots < \delta_l^x. \quad (31)$$

3. Former la suite croissante des parties qui représentent des cercles "ultramétriques" :

$$S^x = \{C_k^x = \{y \mid \delta(y, x) \leq \delta_k^x\} \mid 1 \leq k \leq l\}, \quad (32)$$

cette suite de parties est une chaîne complète de $\mathcal{H}_{bi}(\mathcal{E})$ où chaque partie C_k^x représente un noeud de la hiérarchie dont le niveau est δ_k^x .

4. Ayant bâti un ensemble

$$\{S^x \mid x \in F \subset E\}$$

des chaînes, on choisira - tant que $E - F$ n'est pas vide - un élément de $E - F$, pour construire la chaîne de parties S^y dont on abandonnera la section finissante à partir de $C_{k_0}^y$ si l'ensemble défini par $C_{k_0}^y$ a déjà été précédemment rencontré dans l'un des S^x , pour un x de F .

Nous avons pu voir, après avoir interprété une hiérarchie binaire incomplète $\mathcal{H}_{bi}(\mathcal{E})$ comme une famille de fourches (non orientées), qu'une hiérarchie implicative $\mathcal{HI}(\mathcal{E})$ pouvait s'interpréter comme une famille de fourches orientée (voir Définition 2.9 et Fig. 2)

Maintenant, nous voulons avoir le correspondant adéquat de la notion d'ultramétrie et la propriété correspondante du caractère isocèle des triangles avec une base strictement plus petite que chacun des deux côtés égaux. Nous allons considérer le cas d'une hiérarchie implicative complète $\mathcal{HI}_c(\mathcal{E})$; la généralisation au cas d'une hiérarchie implicative quelconque se comprenant assez bien à partir de la restriction complète à chacune des classes E_j d'une partition P de E .

Théorème 2.3 *La donnée d'une hiérarchie implicative complète $\mathcal{HI}_c(\mathcal{E})$ définit un ordre total sur E , ce dernier correspond à celui de gauche à droite des feuilles de la hiérarchie.*

Considérons la relation binaire R sur E définie comme suit :

$$(\forall (x, y) \in E \times E), xRy \Leftrightarrow \exists (A, B) \in \mathcal{HI}_c(\mathcal{E}), (x, y) \in A \times B \quad (33)$$

(A, B) est une fourche orientée (voir Définition 2.10). La relation R est une relation d'ordre total et strict :

$$(\forall (x, y) \in E \times E), xRy \Rightarrow \neg yRx \quad (34)$$

où \neg désigne la négation.

En effet, compte tenu de la condition 3 de la définition 2.10, le couple (B, A) n'appartient pas à la hiérarchie implicative (n'en est pas une fourche orientée). D'autre part, pour tout sous ensemble B' de B , respectivement A' de A , (B', A') ne peut appartenir à la hiérarchie, car autrement, avec (A, B) la condition 2 de la définition 2.10 s'en trouve violée.

$$(\forall (x, y, z) \in E \times E \times E), xRy \text{ et } yRx \Rightarrow xRz \quad (35)$$

En effet, il existe deux fourches orientées (A, B) et (C, D) telles que :

$$(x, y) \in A \times B \text{ et } (y, z) \in C \times D$$

Comme $B \cap C \neq \emptyset$ en raison de l'appartenance commune de y à B et à C , on a, en vertu de 2 de la définition 2.10 :

$$A \cup B \subset C$$

ou

$$C \cup D \subset B$$

On obtient dans le premier cas que $(x, z) \in C \times D$; alors que dans le second, $(x, z) \in A \times B$. C.Q.F.D.

Il est à remarquer que la relation binaire sur E définie par une hiérarchie implicative complète est beaucoup plus riche et d'ailleurs de nature différente de la relation ci-dessus d'ordre total R qu'elle induit. En effet, si à une hiérarchie implicative complète sur E correspond exactement un ordre total sur E ; inversement, il est aisé de voir qu'à un ordre total sur E correspond exactement $(N-1)!$ hiérarchies implicatives binaires sur E . Imaginons pour cela la suite ordonnée de gauche à droite $(e_1, e_2, \dots, e_i, e_{i+1}, \dots, e_N)$ des éléments de E . Le premier pas de la construction d'une hiérarchie implicative correspond au choix d'un intervalle de la forme (e_i, e_{i+1}) conduisant à la fusion orientée $(e_i \rightarrow e_{i+1})$ et au noeud $\{e_i, e_{i+1}\}$. $(N-1)$ choix sont possibles. À la j -ème étape de la construction, en considérant conformément à l'ordre latéral de gauche à droite la suite des feuilles ou noeuds formés au niveau $j-1$, $(N-j-1)$ intervalles consécutifs sont délimités et il s'agit de choisir l'un d'entre eux pour établir une fusion orientée. j variant de 0 jusqu'à $(N-2)$, on obtient le résultat annoncé.

Relativement à l'ordre total R , une hiérarchie implicative $\mathcal{H}_{bi}(\mathcal{E})$ apparaît comme une famille \mathcal{C} de segments connexes qui vérifient les conditions :

1. $\emptyset \in \mathcal{C}$;
2. \mathcal{C} contient tous les segments singletons comprenant un seul élément;
3. Si C' et C'' appartiennent à \mathcal{C} , $C' \cap C''$ se réduit à l'un ou l'autre des trois éléments \emptyset, C' ou C'' ;
4. Pour chaque segment C , non réduit à un seul élément, il existe un couple de segments disjoints et adjacents (C', C'') tel que $C = C' \cup C''$; c'est-à-dire, dont C est la concaténation de gauche à droite.

Ainsi, en partant d'une axiomatique en termes d'un système de fourches orientées, on se trouve conduit; mais comme conséquence, à une définition qui reprend celle 2.2 de [16].

La relation d'ordre total R va permettre de définir un support à la relation binaire valuée sur E que constitue une hiérarchie implicative complète. Sans risque d'ambiguïté, désignons aussi par R l'ensemble des couples (x, y) de $E \times E$ pour lesquels xRy ; c'est-à-dire, pour lesquels, la feuille correspondante à x se trouve à gauche de celle correspondante à y :

$$R = \{(x, y) \mid (x, y) \in E \times E, xRy\} \quad (36)$$

Dans ces conditions, définissons sur E comme suit la notion de distance que nous pouvons qualifier d'"orientée" :

$$(\forall (x, y) \in E \times E), \eta(x, y) = \begin{cases} \infty & \text{si } (x, y) \notin R \\ \text{valeur du niveau de fusion orientée entre } x \text{ et } y & \end{cases}$$

Introduisons ici la notion de triangle "orienté" relativement à l'ordre total R .

Définition 2.12 (x, y, z) est un triangle orienté relativement à l'ordre total R si on a :

$$xRy \text{ et } yRz$$

Ainsi, pour une hiérarchie implicative dont l'ordre induit est R , l'ordre des feuilles de gauche à droite est x, y et z . Désignons par \mathcal{T} l'ensemble de tels triangles, leur énumération conduit à l'expression suivante :

$$\text{card}(\mathcal{T}) = \binom{N}{3} = \frac{(N-2) \times (N-1) \times N}{6}$$

où $N = \text{card}(E)$. En effet, il y a une correspondance bijective entre l'ensemble des parties à 3 éléments de E et l'ensemble des triangles orientés.

Dans ces conditions et en tenant compte de la binarité, deux situations sont possibles quant à la restriction de la hiérarchie implicative à l'ensemble $\{x, y, z\}$. Elles sont figurées ci-dessous :

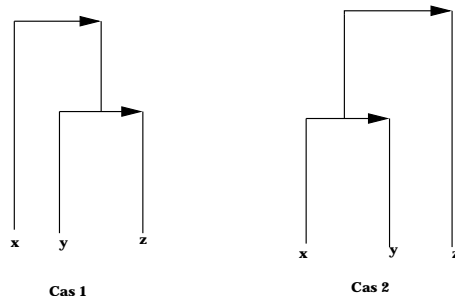


FIG. 3 – Deux formes implicatives fondamentales

Pour la première situation la première jonction orientée a lieu de y vers z , elle est suivie de celle de x vers le noeud représentant $\{x, y\}$. Pour la deuxième situation, la première jonction a lieu de x vers y , le noeud résultant $\{x, y\}$ est relié de gauche à droite vers z . On a dans le premier cas :

$$\eta(x, y) = \eta(x, z) > \eta(y, z)$$

Dans le second cas on a :

$$\eta(x, y) < \eta(x, z) = \eta(y, z)$$

Dans chacun des deux cas l'inégalité ultramétrique que nous dirons "orientée" se trouve satisfaite. Elle s'exprime dans le contexte de l'ensemble \mathcal{T} des triangles orientés. La notion de segment non orienté est remplacée par celle de vecteur dont l'origine est à gauche de l'extrémité. Ainsi, l'inégalité ultramétrique orientée s'exprime comme suit :

$$\begin{aligned} \eta(x, y) &\leq \max\{\eta(x, z), \eta(y, z)\} \\ \eta(x, z) &\leq \max\{\eta(x, y), \eta(y, z)\} \\ \eta(y, z) &\leq \max\{\eta(x, y), \eta(x, z)\} \end{aligned} \tag{37}$$

On en déduit la propriété du caractère isocèle des triangles "orientés".

Théorème 2.4 *Tout triangle (x, y, z) de \mathcal{T} est isocèle, la base étant strictement le plus petit des côtés.*

On a xRy et yRz . Ainsi et de gauche à droite on a x puis y , puis z . Dans ces conditions, on a nécessairement :

$$\begin{aligned} \eta(x, y) &\leq \eta(x, z) \\ \text{et } \eta(y, z) &\leq \eta(x, z) \end{aligned} \tag{38}$$

Ces inégalités peuvent d'ailleurs être obtenues à partir des inégalités ultramétriques (37). On ne peut pas en effet avoir :

$$\eta(x, y) > \eta(x, z) \tag{39}$$

car dans ce cas, en vertu de la première inégalité ultramétrique, on a :

$$\eta(x, z) < \eta(x, y) \leq \eta(y, z) \tag{40}$$

La troisième inégalité ultramétrique permet d'obtenir

$$\eta(x, z) < \eta(x, y) = \eta(y, z) \quad (41)$$

mais alors, l'ordre induit est tel que z est entre x et y , ce qui viole l'ordre dont on est parti où c'est y qui est intermédiaire entre x à sa gauche et z à sa droite.

De la même façon on se trouve conduit à la même conclusion si

$$\eta(x, z) < \eta(y, z)$$

Sinon d'ailleurs, on aurait un croisement entre le vecteur de jonction de la classe A de x vers celle C de z , avec la branche ascendante issue de la classe B de y .

Compte tenu du caractère binaire de la hiérarchie implicative à laquelle nous avons associé la fonction η , nous avons nécessairement l'un des deux cas suivants, où les inégalités sont strictes :

$$\eta(x, y) < \eta(y, z) \text{ ou } \eta(y, z) < \eta(x, y) \quad (42)$$

Si $\eta(x, y) < \eta(y, z)$, la deuxième inégalité ultramétrique ci-dessus (37) donne :

$$\eta(x, z) \leq \eta(y, z) \quad (43)$$

qui avec (38), ou d'ailleurs, la dernière inégalité ultramétrique donne :

$$\eta(y, z) = \eta(x, z) \quad (44)$$

D'où

$$\eta(x, y) < \eta(x, z) = \eta(y, z) \quad (45)$$

et le triangle (x, y, z) est isocèle avec une base \vec{xy} strictement plus petite que chacun des deux côtés \vec{xz} et \vec{yz} .

Si $\eta(y, z) < \eta(x, y)$, la deuxième inégalité ultramétrique [voir (37)] donne :

$$\eta(x, z) \leq \eta(x, y) \quad (46)$$

qui avec (38) ou d'ailleurs, la première inégalité ultramétrique (37), donne :

$$\eta(x, z) = \eta(x, y) \quad (47)$$

D'où

$$\eta(y, z) < \eta(x, y) = \eta(x, z) \quad (48)$$

et le triangle (x, y, z) est isocèle avec une base \vec{yz} strictement plus petite que chacun des deux côtés \vec{xy} et \vec{xz} .

Nous avons montré la construction d'une hiérarchie binaire de classification $\mathcal{H}_b(E)$ associée à une distance ultramétrique sur E pour laquelle tout triangle est isocèle avec une base strictement plus petite que chacun des deux côtés égaux. De façon quelque peu analogue, nous allons montrer la construction d'une hiérarchie complète $\mathcal{HI}(E)$ dès lors qu'on se donne un ordre total R sur E ainsi qu'une distance ultramétrique orientée η sur R .

L'algorithme de construction, exprimé de façon informelle se déroule comme suit :

1. ordonner les éléments de E de gauche à droite conformément à R : $x_{(1)} < x_{(2)} < \dots < x_{(N)}$

2. prendre l'élément le plus à gauche $x_{(1)}$ et établir la table des dissimilarités ultramétriques orientées à $x_{(1)}$:

$$\{\eta(x_{(1)}, y) \mid (x_{(1)}, y) \in R\}$$

3. la suite de ces valeurs est non décroissante, établir la suite strictement croissante des valeurs atteintes que nous notons :

$$0 < \eta_1^1 < \eta_2^1 < \dots < \eta_k^1 < \dots < \eta_l^1$$

4. former la suite croissante des intervalles commençants qui représentent des cercles ultramétriques "orientés" :

$$S^1 = \{C_k^1 = \{y \mid \eta(x_{(1)}, y) \leq \eta_k^1\} \mid 1 \leq k \leq l\}, \quad (49)$$

cette suite d'intervalles représente une chaîne complète de la hiérarchie implicative $\mathcal{HI}(E)$, chaque intervalle représente un noeud de la hiérarchie implicative.

5. ayant bâti un ensemble de chaînes

$$\{S^j \mid 1 \leq j \leq I\} \quad (50)$$

on considérera x^{I+1} tant que $I \leq N - 2$, pour construire la chaîne associée, la construction s'arrêtant dès lors qu'on rencontre un noeud déjà trouvé dans la chaîne qui la précède directement.

2.3 Inadéquation entre la théorie de la démonstration et le sens formel d'une hiérarchie implicative

L'ensemble désigné de façon générale par E ci-dessus devient ici - comme c'est le cas le plus fréquent - un ensemble \mathcal{A} d'attributs booléens :

$$\mathcal{A} = \{a^j \mid 1 \leq j \leq p\}. \quad (51)$$

Telle que présentée et développée par ses initiateurs [13, 12, 16], la hiérarchie implicative est définie comme une organisation hiérarchisée de règles et de - selon la terminologie des auteurs cités - "métarègles". À un niveau donné d'un arbre hiérarchique binaire, c'est une règle de degré 1 qui peut apparaître correspondante à une implication entre deux attributs de la forme $a^j \rightarrow a^k$, ($1 \leq j \neq k \leq p$), ou bien c'est une "métarègle" qui apparaît. Cette dernière correspondrait à une suite emboîtée d'implications. Le noeud associé définirait alors une "métarègle" ou une R -règle. L'ordre de cette dernière est défini par le nombre d'implications qu'elle comprend (voir [12, 10]). Ainsi, dans l'exemple suivant (voir FIG. 4) :

le noeud du niveau 5 est présenté comme correspondant à une R -règle d'ordre 4 :

$$(a_1 \rightarrow (a_6 \rightarrow a_4)) \rightarrow (a_3 \rightarrow a_5) \quad (52)$$

Cependant, sur le plan de la logique, si on admet qu'une *métarègle* est une règle ou une propriété sur un ensemble de règles, des formules telles que celle (52) définissent tout simplement des *règles* ; pouvant certes être plus ou moins complexes. D'ailleurs ces R règles ne peuvent traduire dans leur interprétation le mode de construction statistique de la hiérarchie implicative (voir paragraphe 3). Alors que ce mode de construction est parfaitement compatible avec l'expression formelle et sémantique d'une hiérarchie implicative telle qu'elle a été formalisée au paragraphe 2.2.

Considérons les trois situations logiques suivantes d'un arbre implicatif, respectivement sur trois, trois et quatre attributs. Nous les noterons (I), (II) et (III) (voir FIG. 5) :

En retenant pour la comparaison des diverses implications une valuation ordinaire - pouvant par exemple correspondre à un degré de nécessité - l'interprétation d'une hiérarchie implicative, telle qu'elle résulte de l'analyse du paragraphe 2.2 donne pour chacune des situations (I), (II) et (III) ce qui suit :

Pour (I) :

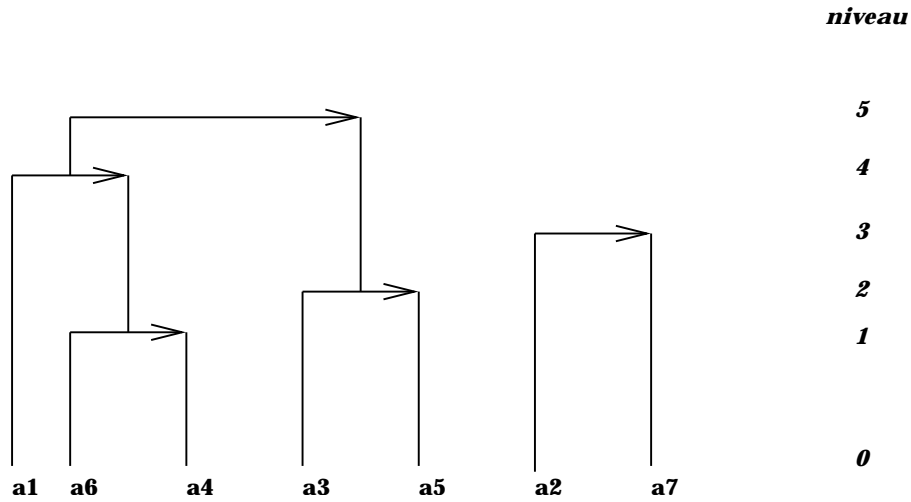


FIG. 4 – Un arbre implicatif

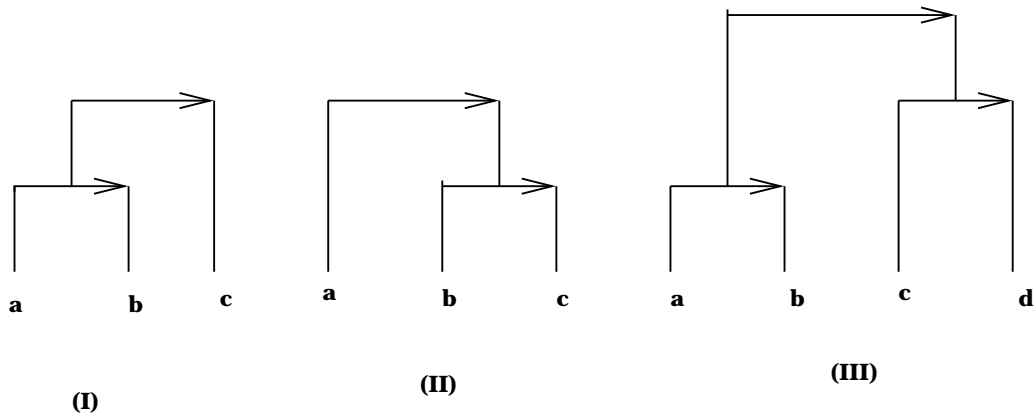


FIG. 5 – Trois cas fondamentaux

$$(a \rightarrow b) \& (a \rightarrow c) \& (b \rightarrow c)$$

et $deg(a \rightarrow b) > deg(a \rightarrow c) = deg(b \rightarrow c)$ (53)

où $deg(\alpha \rightarrow \beta)$ définit la valuation de l'implication $\alpha \rightarrow \beta$.

Pour (II) :

$$(a \rightarrow b) \& (a \rightarrow c) \& (b \rightarrow c)$$

et $deg(a \rightarrow b) = deg(a \rightarrow c) < deg(b \rightarrow c)$ (54)

Pour (III) :

$$(a \rightarrow b) \& (c \rightarrow d) \& (a \rightarrow c) \& (a \rightarrow d) \& (b \rightarrow c) \& (b \rightarrow d)$$

et $deg(a \rightarrow b) > deg(c \rightarrow d) = deg(a \rightarrow d) = deg(b \rightarrow c) = deg(b \rightarrow d)$ (55)

Selon [12, 10], les formules implicatives traduites par les arbres (I), (II) et (III) sont respectivement :

$$((a \rightarrow b) \rightarrow c) \tag{56}$$

$$(a \rightarrow (b \rightarrow c)) \tag{57}$$

et $((a \rightarrow b) \rightarrow (c \rightarrow d))$ (58)

Nous allons établir que les formules logiques 56, 57 et 58 ont des sens logiques très différents de ceux que nous avons attribués et qui sont respectivement définis dans (53), 54 et (55).

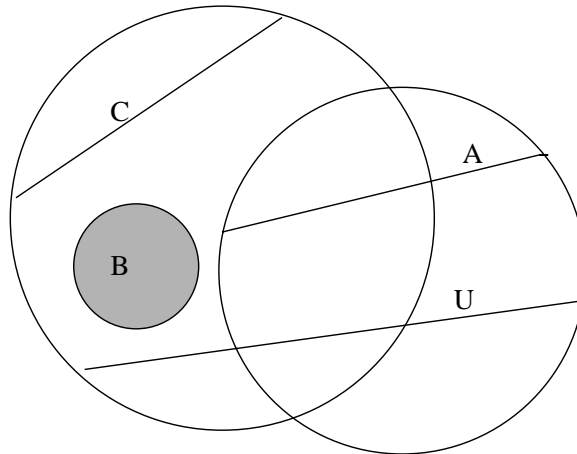
Commençons par préciser que si α est un attribut booléen et x un élément de l'univers \mathcal{U} des objets concernés, l'expression $\alpha(x)$ signifie que l'attribut α est à *VRAI* sur x . Dans ces conditions $((a \rightarrow b) \rightarrow c)$ s'exprime comme suit :

$$\begin{aligned}
 \forall x \in \mathcal{U}, [(a(x) \rightarrow b(x)) \rightarrow c(x)] &\equiv \\
 \forall x \in \mathcal{U}, [\neg a \vee b(x) \rightarrow c(x)] &\equiv \\
 \forall x \in \mathcal{U}, [(\neg(\neg a \vee b) \vee c(x))] &\equiv \\
 \forall x \in \mathcal{U}, [(a \wedge \neg b) \vee c(x)] &\equiv \\
 \forall x \in \mathcal{U}, [(a \vee c(x)) \wedge (\neg b) \vee c(x)] &\equiv \\
 (\forall x \in \mathcal{U}, (\neg a(x) \rightarrow c(x))) \wedge (\forall x \in \mathcal{U}, (b(x) \rightarrow c(x))) &\equiv
 \end{aligned} \tag{59}$$

Par conséquent :

$$(\neg a \rightarrow c) \& (b \rightarrow c) \tag{60}$$

ce qui ne correspond nullement à (53) ci-dessus. D'ailleurs, (60) peut être représenté de façon ensembliste comme suit :



$$[(a \rightarrow b) \rightarrow c]$$

FIG. 6 – $(a \text{ imp } b) \text{ imp } c$

où A , B et C représentent les sous ensembles de \mathcal{U} où, respectivement a , b et c sont à *VRAI*. De façon particulière, \mathcal{U} correspond dans cette figure à $A \cup B$; mais cela n'a rien d'obligé.

On se rend ainsi compte que de la formule implicative $((a \rightarrow b) \rightarrow c)$, on ne peut dériver $a \rightarrow b$ ou $a \rightarrow c$ et c'est $\neg a \rightarrow c$ qui est valide. Maintenant, si on considère comme dans [12, 10, 16] qu'on a la condition préalable $(a \rightarrow b)$; soit :

$$(\forall x \in \mathcal{U}), \neg a \vee b(x) \tag{61}$$

alors $((a \rightarrow b) \rightarrow c)$ n'est pas informant, car quelque soit l'attribut c , on a, ν désignant la fonction de vérité :

$$(\forall x \in \mathcal{U}), \nu(\neg(\neg a \vee b(x)) \wedge c(x)) = FAUX \tag{62}$$

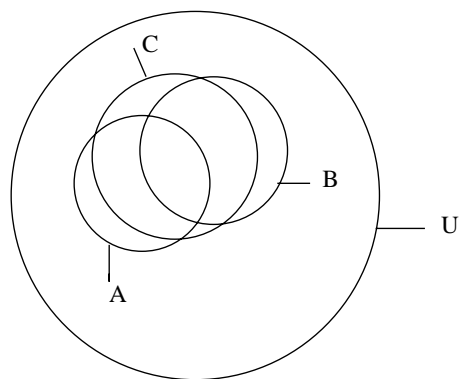
Considérons maintenant la formule implicative :

$$(a \rightarrow (b \rightarrow c)) \tag{63}$$

Elle s'exprime par :

$$\begin{aligned} (\forall x \in \mathcal{U})[a(x) \rightarrow \neg b(x) \vee c(x)] &\equiv \\ (\forall x \in \mathcal{U})[\neg a(x) \vee \neg b(x) \vee c(x)] &\equiv \\ (\forall x \in \mathcal{U})[\neg(a \wedge b(x)) \vee c(x)] &\equiv \\ a \wedge b \rightarrow c & \end{aligned} \tag{64}$$

ce qui au niveau ensembliste peut s'exprimer comme suit : " la partie de B qui est dans A est incluse dans la partie de C qui est dans A "



$$(a \rightarrow (b \rightarrow c))$$

FIG. 7 – a imp (b imp c)

Ainsi, de la formule logique $(a \rightarrow (b \rightarrow c))$ on ne peut dériver ni $b \rightarrow c$, ni $a \rightarrow b$, ni $a \rightarrow c$. Dans le cas où on exige au préalable $b \rightarrow c$, on ne peut pas non plus dériver ni $a \rightarrow b$, ni $a \rightarrow c$.

Considérons enfin la formule logique

$$((a \rightarrow b) \rightarrow (c \rightarrow d)) \quad (65)$$

elle s'exprime par :

$$\begin{aligned} & (\forall x \in \mathcal{U})[(a(x) \rightarrow b(x)) \rightarrow (c(x) \rightarrow d(x))] \equiv \\ & (\forall x \in \mathcal{U})[\neg(\neg a(x) \vee b(x)) \vee (\neg c(x) \vee d(x))] \equiv \\ & (\forall x \in \mathcal{U})[(a(x) \wedge \neg b(x)) \vee (\neg c(x) \vee d(x))] \equiv \\ & (\forall x \in \mathcal{U})[a(x) \vee (\neg c(x) \vee d(x)) \wedge (\neg b(x) \vee (\neg c(x) \vee d(x)))] \end{aligned} \quad (66)$$

Ce qui correspond à :

$$\begin{aligned} & \neg a \rightarrow (c \rightarrow d) \\ & \text{et } b \rightarrow (c \rightarrow d) \end{aligned} \quad (67)$$

Dans ces conditions, on n'a ni $a \rightarrow b$, ni $c \rightarrow d$, ni $b \rightarrow c$ ni $b \rightarrow d$. Une configuration ensembliste correspondante peut être la suivante (voir FIG. 8).

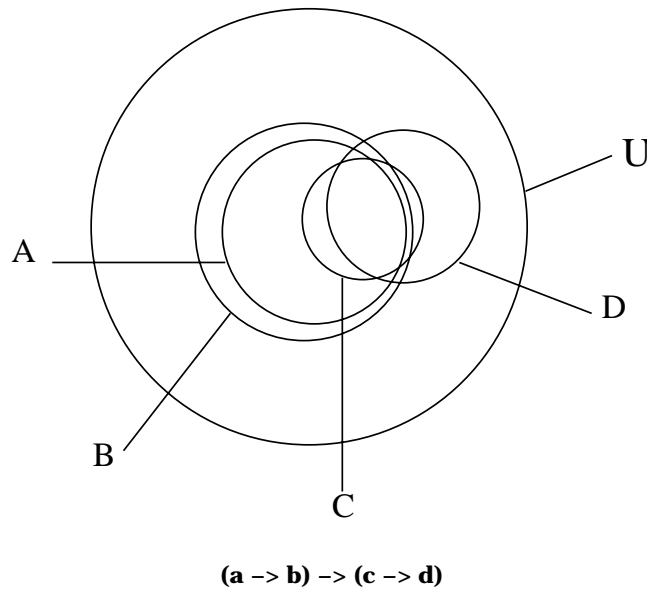


FIG. 8 – $(a \text{ imp } b) \text{ imp } (c \text{ imp } d)$

3 Construction statistique d'une hiérarchie implicative

3.1 Introduction ; indice d'implication entre classes d'attributs booléens

Le principe général de la construction d'une hiérarchie implicative binaire sur un ensemble \mathcal{A} est celui de la CAH (Classification Ascendante Hiérarchique) [22, 28, 36]. Lorsque cette dernière porte sur un ensemble E d'entités, on suppose que E est muni d'un indice de proximité p , associant à chaque paire $\{x, y\}$ d'éléments de E , un nombre positif $p(x, y)$ "mesurant" la proximité (ou la ressemblance) entre x et y . la première étape conceptuelle d'une CAH consiste à étendre la notion de proximité p entre éléments de E à celle \mathcal{P} entre parties disjointes de E . À partir de là, la construction ascendante hiérarchique d'un arbre de classification se déroule comme suit :

- État initial : Partition discrète dont chaque classe contient un seul élément, niveau := 0 ;
- Progression : Déterminer une nouvelle partition en fusionnant les paires de classes les plus proches au sens de \mathcal{P} , niveau := niveau + 1 ;
- Règle d'arrêt : La partition formée est grossière et comporte une seule classe, niveau := niveau + 1 .

L'arbre obtenu est un arbre à niveaux. Le niveau 0 est celui des feuilles de l'arbre et le dernier niveau atteint est celui de la racine de l'arbre. À chaque niveau se trouve associée une partition qui est moins fine que la partition du niveau précédent. L'arbre obtenu n'est pas nécessairement binaire. Il l'est si à chacun des niveaux la valeur maximale de l'indice de proximité \mathcal{P} est atteinte, soit sur une seule paire de classes du niveau, soit sur différentes paires de classes, mutuellement sans composante commune. Si tel n'est pas le cas [26] on peut toujours se ramener à une forme binaire de l'arbre des classifications et ce, en lui adjoignant une indéxation. À chaque niveau on attachera la valeur maximale de l'indice de proximité entre classes du niveau précédent. Dans ces conditions, une agrégation multiple sera décomposée en une suite d'agrégations binaires (dans n'importe quel ordre) où chacune donne naissance à un nouveau niveau ; mais, avec la même valeur de l'indice.

Dans l'expression ci-dessus, l'ensemble E peut tout aussi bien être un ensemble d'attributs qu'un ensemble d'objets. Dans le cas d'une hiérarchie implicative, l'ensemble E est défini par un ensemble \mathcal{A} d'attributs booléens. Comme nous l'a montré l'analyse ci-dessus (paragraphe 2.2) (voir aussi [16]), ce qui change de façon essentielle c'est la notion de mesure d'association entre attributs ou classes d'attributs, qui n'est plus symétrique mais orientée, disons de gauche à droite pour fixer les idées.

Relativement à l'ensemble $\mathcal{A} = \{a^j \mid 1 \leq j \leq p\}$ des attributs booléens de description d'un univers \mathcal{U} d'objets, la première étape consiste, relativement à un couple (a, b) d'attributs faisant partie de \mathcal{A} , d'évaluer - sur la base d'un ensemble d'apprentissage \mathcal{O} de taille n , issu de \mathcal{U} - la propension que l'attribut b soit à *VRAI* sachant que a est à *VRAI*. En d'autres termes, il s'agit de mesurer au moyen d'un indice statistique le degré d'implication $(a \rightarrow b)$. Un tel indice porte le nom dans la littérature internationale de "Mesure d'intérêt d'une règle d'association". Compte tenu du caractère fondamental du choix d'une telle "mesure", une littérature abondante et riche a été consacrée à ce sujet [14]. Ce dernier comporte des aspects autant logiques, formels que statistiques. Nous ne pouvons ici nous étendre sur un tel sujet. Signalons néanmoins que l'approche *AVL* (Analyse de la Vraisemblance des Liens) [22, 28] a nourri un certain nombre d'indices [8, 30, 17]. La base de leur construction se réfère à la philosophie de la théorie de l'information ; mais, au niveau des relations mutuelles entre attributs. Relativement à un couple (a^j, a^k) d'attributs et à la mesure de l'implication $a^j \rightarrow a^k$, il s'agit de gérer la probabilité définissant le degré d'in vraisemblance de la petitesse du nombre observé de contre-exemples ; c'est-à-dire, de $n(a^j \neg a^k)$ qui représente au niveau de \mathcal{O} le nombre d'objets où a^j est à *VRAI* et où a^k est à *FAUX*. Cette mesure probabiliste peut être relative à un sous ensemble filtré de couples d'attributs de \mathcal{A} [30]. Pour l'obtenir on se situe par rapport à une hypothèse d'absence de liaison ou d'indépendance associant à l'ensemble \mathcal{A} des attributs un ensemble \mathcal{A}^* d'attributs aléatoires mutuellement indépendants respectant les distributions marginales des attributs de \mathcal{A} . On associe à la mesure de probabilité un coefficient de similarité implicative discriminant sur l'ensemble des règles sélectionnées *a priori* comme potentielles. On peut d'ailleurs supposer un tel coefficient compris entre 0 et 1 et le poser égal à 0 pour toute règle non admissible *a priori*. Tout en gardant à l'esprit que différents coefficients peuvent être pris en considération, notons comme dans [12, 16] par c un tel coefficient. Cela, d'autant plus que nous allons reprendre la construction ascendante d'une hiérarchie implicative [10, 16] avec la même expression de la comparaison orientée entre classes. Nous y apporterons des compléments méthodologiques importants qu'il est classique de considérer en CAH. Le premier concerne la propriété de monotonie de l'indice orienté de comparaison entre classes et le second, concerne la formule dite de réactualisation des similarités entre classes après une agrégation, ici binaire [19, 15, 26].

Dans ces conditions, on suppose établie une matrice des similarités implicatives se référant à une échelle $[0, 1]$, qu'on note :

$$\{c(a^j, a^k) \mid (a^j, a^k) \in \mathcal{A} \times \mathcal{A}\} \quad (68)$$

étant entendu que $c(a^j, a^k)$ est posé égal à 0 dès lors que la règle $a^j \rightarrow a^k$ n'est pas potentiellement admissible [11, 12]. En désignant par $n(a)$ le nombre des objets où a est à *VRAI* et dans la mesure où pour un couple (a^j, a^k) d'attributs tel que $n(a^j) < n(a^k)$, si $a^j \rightarrow a^k$ est potentiellement admissible, il n'en est pas de même de $(a^k \rightarrow a^j)$.

Nous avons vu (cf. paragraphe 2.2) que relativement à une hiérarchie implicative sur un ensemble E , l'ensemble des feuilles de la classe sous-tendant un noeud donné, se trouve muni d'un ordre total directement associé à la structure ultramétrique et orientée de la construction de la hiérarchie. Lorsque cette construction est statistique, pour une classe formée C_j de taille k_j :

$$C^j = \{a^{j_l} \mid 1 \leq l \leq k_j\} \quad (69)$$

où l'ordre induit est :

$$a^{j_1} < a^{j_2} < \dots < a^{j_l} < \dots < a^{j_{k_j}} \quad (70)$$

on définit la cohésion [11, 12] de la classe C^j par la moyenne géométrique des valeurs de l'indice d'implication c sur l'ensemble des couples $(a^{j_l}, a^{j_{l'}})$ du graphe de la relation d'ordre définie ci-dessus, nommément :

$$c(C^j) = \left(\prod \{c(a^{j_l}, a^{j_{l'}}) \mid 1 \leq l < l' \leq k_j\} \right)^{\frac{2}{k_j \times (k_j - 1)}} \quad (71)$$

Maintenant, considérons deux classes C^j et $C^{j'}$ présentes à un niveau donné de la hiérarchie implicative. De la même façon que pour C^j , on pose pour $C^{j'}$:

$$C^{j'} = \{a^{j'_m} \mid 1 \leq m \leq k_{j'}\} \quad (72)$$

avec

$$a^{j'_1} < a^{j'_2} < \dots < a^{j'_m} < \dots < a^{j'_{k_{j'}}} \quad (73)$$

L'indice de jonction orientée $(C^j \rightarrow C^{j'})$ proposé dans [12, 10] est défini par :

$$c(C^j \rightarrow C^{j'}) = \left(c(C^j)^{\binom{k_j}{2}} \times c(C^{j'})^{\binom{k_{j'}}{2}} \times \prod \{c(a^{j_l}, a^{j'_m}) \mid 1 \leq l \leq k_j, 1 \leq m \leq k_{j'}\} \right)^{\frac{2}{r \times (r-1)}} \quad (74)$$

où $r = k_j + k_{j'}$

Relativement à l'union orientée que nous notons $C_j \vec{\cup} C_{j'}$ (voir aussi [10]) sous-tendue par l'ordre total :

$$a^{j_1} < \dots < a^{j_l} < \dots < a^{j_{k_j}} < a^{j'_1} < \dots < a^{j'_m} < \dots < a^{j'_{k_{j'}}} \quad (75)$$

L'indice de jonction $c(C_j \rightarrow C_{j'})$ représente exactement l'indice de cohésion $c(C_j \vec{\cup} C_{j'})$ relativement à l'ensemble totalement ordonné par la relation ci-dessus [cf. (75)] des attributs

$$\{a^{j_l} \mid 1 \leq l \leq k_j\} \cup \{a^{j'_m} \mid 1 \leq m \leq k_{j'}\} \quad (76)$$

qui est de cardinal $r = k_j + k_{j'}$.

Pour une classe C qui sous-tend un noeud d'une hiérarchie implicative nous avons introduit l'ensemble C , mais muni de l'ordre total induit par la construction de la hiérarchie implicative. On peut noter $o(C)$ un tel ensemble totalement ordonné. L'indice d'association orienté $c(C_j \rightarrow C_{j'})$ de C_j vers $C_{j'}$ n'est autre que la moyenne géométrique du coefficient c sur l'ensemble des couples ordonnés de $o(C_j \vec{\cup} C_{j'})$ où C_j est placé à gauche de $C_{j'}$ et où C_j et $C_{j'}$ sont totalement ordonnés. $c(C_j \rightarrow C_{j'})$ jouera le rôle du critère de fusion binaire de classes de la CAH (voir au début du paragraphe). Or dans le contexte de la CAH différents critères de fusion de classes sont proposés. On comprend dans ces conditions qu'il pourra en être ainsi dans le cadre de la Classification Ascendante Hiérarchique Implicative.

3.2 Propriétés algorithmiques d'une Classification Ascendante Hiérarchique Implicative

L'algorithme défini dans [10] est exactement l'algorithme de base de la CAH (voir au début du paragraphe 3.1) à cela près que chaque classe se trouve accompagnée d'un ordre total sur l'ensemble des éléments compatibles avec la construction algorithmique de l'arbre. Cependant, cet arbre ne va pas jusqu'à une racine regroupant tous ses éléments. Sa formation s'arrête dès lors que devant tous les couples $(C_j, C_{j'})$ de classes dernièrement formées, munies chacune d'un ordre total, la valeur maximale de l'indice $c(C_j \rightarrow C_{j'})$ est nulle.

Illustrons une fois de plus de façon imagée la progression d'un tel algorithme. On suppose un axe ordinal portant la suite ordonnée des attributs de façon compatible avec l'ordre latéral des feuilles de gauche à droite de la construction en cours des classifications. À une étape donnée, définie par un niveau de la hiérarchie, chaque classe C_j se trouve représentée par un segment (un intervalle) de l'axe ordinal. J désignant l'ensemble des indices des classes ou segments formés et $J^{[2]}$ l'ensemble des couples d'indices distincts de J , si $(C_j, C_{j'})$, pour $(j, j') \in J^{[2]}$ est le couple de classes ou l'un des couples de classes réalisant une valeur maximale positive de $c(C_j \rightarrow C_{j'})$, on placera le segment C_j directement à gauche de $C_{j'}$ pour former $C_j \dot{\cup} C_{j'}$.

Théorème 3.1 *Le critère c est monotone*

Cette propriété se dit également de non inversion du critère de fusion - ici orientée - entre classes. Elle exprime que la valeur maximale du critère de fusion entre classes décroît au fur et à mesure de la suite des fusions opérées, donc décroît par rapport au niveau de la hiérarchie. Avec des notations qu'on comprend, il s'agit de montrer que :

$$c(C \cup C' \rightarrow C'') < c(C, C') \quad (77)$$

$$\text{sachant que } c(C, C'') < c(C, C') \quad (78)$$

$$\text{et } c(C', C'') < c(C, C') \quad (79)$$

Le membre de gauche de (77) représente la moyenne géométrique générale sur l'ensemble des couples ordonnés (a^j, a^k) du segment ordinal associé à l'union orientée $C \dot{\cup} C' \dot{\cup} C''$. Relativement à cette dernière, introduisons les ensembles de couples ordonnés d'attributs :

$$C^{(2)}, C'^{(2)}, C''^{(2)}, C \times C', C \times C'' \text{ et } C' \times C'' \quad (80)$$

où un ensemble de la forme $B^{(2)}$ représente les couples ordonnés d'éléments de B conformément à l'ordre total dont se trouve muni B . On introduira de même, relativement aux unions orientées $C \dot{\cup} C''$ et $C' \dot{\cup} C''$ (voir les membres de gauche de (78) et (79) les ensembles :

$$C^{(2)}, C''^{(2)} \text{ et } C \times C'' \quad (81)$$

$$\text{et ceux } C'^{(2)}, C''^{(2)} \text{ et } C' \times C'' \quad (82)$$

La moyenne (ici géométrique et que nous pouvons noter *moyg*) sur une suite quelconque d'ensembles est égale à la moyenne des moyennes sur les différents ensembles composants ; ces derniers n'étant pas nécessairement disjoints. Par rapport à la famille des ensembles (80) adjoignons pour l'indexation de la moyenne géométrique les ensembles $C^{(2)}$, $C'^{(2)}$ et $C''^{(2)}$; de sorte qu'au total chacun de ces ensembles intervient deux fois. Comme $C^{(2)}$, $C'^{(2)}$ et $C''^{(2)}$ ont été formés préalablement, la moyenne géométrique sur la suite étendue des couples s'en trouve augmentée.

La suite des couples d'indexation peut maintenant se décomposer en la suite des couples :

$$C^{(2)} \cup C'^{(2)} \cup C \times C' = (C \dot{\cup} C')^{(2)} \quad (83)$$

suivie de la suite des couples :

$$C^{(2)} \cup C''^{(2)} \cup C \times C'' = (C \dot{\cup} C'')^{(2)} \quad (84)$$

et suivie de la suite des couples :

$$C'^{(2)} \cup C''^{(2)} \cup C' \times C'' = (C' \dot{\cup} C'')^{(2)} \quad (85)$$

Interprétons alors la nouvelle moyenne géométrique étendue et globale comme la moyenne géométrique des trois moyennes géométriques sous-tendues par (83), (84) et (85). En vertu des inégalités (78) et (79), chacune de ces moyennes est inférieure ou égale (pour (83)) à $c(C, C')$. Il en est donc de même de la moyenne globale (étendue à (83), (84) et (85)) et donc *a fortiori* de celle étendue à la suite des ensembles de couples (80); ce qui démontre (77). C.Q.F.D.

Formule de réactualisation

Soient C , C' et C'' trois classes orientées en présence, imaginons la fusion orientée de gauche à droite $C \rightarrow C'$, notée également $C \dot{\cup} C'$. L'objet de la formule de réactualisation est d'exprimer $c((C \rightarrow C') \rightarrow C'')$ qu'on peut également noter $c((C \dot{\cup} C') \dot{\cup} C'')$ en fonction de $c(C)$, $c(C')$, $c(C'')$, $c(C \dot{\cup} C')$, $c(C \dot{\cup} C'')$ et $c(C' \dot{\cup} C'')$. On désignera ci-dessous par k , k' et k'' les cardinaux de C , C' et C'' . D'autre part, $\binom{l}{2}$ note pour l entier, un coefficient binomial qui vaut $\frac{l \times (l-1)}{2}$.

Théorème 3.2

$$c((C \rightarrow C') \rightarrow C'') = \left\{ c(C \dot{\cup} C')^{\binom{k+k'}{2}} \times c(C'')^{\binom{k''}{2}} \times \left(\frac{c(C \dot{\cup} C'')^{\binom{k+k''}{2}}}{c(C)^{\binom{k}{2}} c(C'')^{\binom{k''}{2}}} \right) \times \left(\frac{c(C' \dot{\cup} C'')^{\binom{k'+k''}{2}}}{c(C')^{\binom{k'}{2}} c(C'')^{\binom{k''}{2}}} \right) \right\}^{\frac{1}{\binom{k+k'+k''}{2}}} \quad (86)$$

En effet, de par la définition du critère de fusion orientée on a :

$$c((C \rightarrow C') \rightarrow C'') = \left\{ c(C \dot{\cup} C')^{\binom{k+k'}{2}} \times c(C'')^{\binom{k''}{2}} \times \prod \left\{ c(a_h, a_{j''}) \mid a_h \in C \dot{\cup} C', a_{j''} \in C'' \right\} \right\}^{\frac{1}{\binom{k+k'+k''}{2}}} \quad (87)$$

Le produit peut se mettre sous la forme :

$$\prod \left\{ c(a_j, a_{j''}) \mid a_j \in C, a_{j''} \in C'' \right\} \times \prod \left\{ c(a_{j'}, a_{j''}) \mid a_{j'} \in C', a_{j''} \in C'' \right\} \quad (88)$$

Maintenant, relativement à une union orientée $D \rightarrow E$ où $\text{card}(D) = s$ et $\text{card}(E) = t$ avec $u = s + t$ et où on a :

$$c(D \rightarrow E) = \left(c(D)^{\binom{s}{2}} \times c(E)^{\binom{t}{2}} \times \prod \left\{ c(d, e) \mid (d, e) \in D \times E \right\} \right)^{\frac{1}{\binom{u}{2}}} \quad (89)$$

On obtient :

$$\prod \left\{ c(d, e) \mid (d, e) \in D \times E \right\} = \frac{c(D \dot{\cup} E)^{\binom{u}{2}}}{c(D)^{\binom{s}{2}} \times c(E)^{\binom{t}{2}}} \quad (90)$$

En exprimant de la sorte chacun des produits de (88) dans (87) on obtient la formule (86) de réactualisation annoncée.

La formule de réactualisation permet de considérer une algorithmique de construction ascendante de l'arbre implicatif fondée sur la suite des états de la matrice des coefficients d'association orientée entre classes formées. En notant ces dernières $\{C_j \mid 1 \leq j \leq J\}$, on accompagnera cette matrice de la suite des cohésions de ces classes :

$$\{c(C_j) \mid 1 \leq j \leq J\} \quad (91)$$

- État initial : La matrice est celle du coefficient c dit de "cohésion" entre couples d'attributs booléens (voir par exemple [11])
- Progression : Déterminer la valeur maximale du coefficient de fusion orientée $c(C_j \rightarrow C_{j'})$ entre classes formées et les couples de classes $(C_j, C_{j'})$, concernés par cette valeur maximale. Le plus souvent il y en a un seul tel couple. Néanmoins, nous considérerons les différents cas présentés ci-dessous. À cette fin on introduit au préalable la variable *niveau* qui indique le niveau atteint de l'arbre implicatif. Cette variable vaut 0 pour le niveau des feuilles de l'arbre. Plaçons nous dans ces conditions à un niveau de l'arbre où la partition en classes orientées et construites ultramétriquement

$$\{C_j \mid 1 \leq j \leq J\}$$

vient d'être obtenue.

Cas

- La valeur maximale de $c(C_j \rightarrow C_{j'})$, $1 \leq j \neq j' \leq J$, est atteinte sur un seul couple (j_1, j'_1) où $j_1 \neq j'_1$. La classe $C_{j_1} \cup C_{j'_1}$ occupera le niveau suivant dont la valeur résulte de l'incréméntation d'une unité de la dernière valeur atteinte de la variable *niveau* :

$$\text{niveau} := \text{niveau} + 1$$

On réactualise les valeurs de $c(C_{j_1} \cup C_{j'_1} \rightarrow C_{j''})$ et de $j'' \notin \{j_1, j'_1\}$ et de $c(C_{j''} \rightarrow C_{j_1} \cup C_{j'_1})$, grâce à la formule de réactualisation (86).

- La valeur maximale de $c(C_j \rightarrow C_{j'})$ est atteinte sur plusieurs couples de classes de la forme $(C_{j_m} \rightarrow C_{j'_m})$, $1 \leq m \leq M$, qui sont mutuellement sans composante commune. Ainsi, le nombre total de classes qui interviennent est $2M$. Dans ces conditions, pour chacun des couples $(C_{j_m}, C_{j'_m})$, $1 \leq m \leq M$, on procède à la
 1. fusion orientée $C_{j_m} \cup C_{j'_m}$;
 2. réactualisation des similarités implicatives c entre la nouvelle classe formée et les autres classes en présence au moyen de la formule (86).

Une fois toutes les nouvelles classes formées (chacune par fusion binaire) on associe à chacune de ces dernières un noeud de l'arbre implicatif où différents noeuds placés au *même* niveau de l'arbre implicatif défini comme ci-dessus par :

$$\text{niveau} := \text{niveau} + 1$$

- La valeur maximale de $c(C_j \rightarrow C_{j'})$, $1 \leq j \neq j' \leq J$ est atteinte sur plusieurs couples de classes de la forme $C_{j_m} \rightarrow C_{j'_m}$, $1 \leq m \leq M$, qui partagent des composantes communes. Il en résulte que le nombre total de classes qui interviennent est strictement inférieur à $2M$. Considérons dans ces conditions le graphe orienté obtenu sur l'ensemble des classes qu'on note

$$\{C_{j_m} \mid 1 \leq m \leq M\} \cup \{C_{j'_m} \mid 1 \leq m \leq M\} \quad (92)$$

et qui est associé à la relation binaire $c(C_j \rightarrow C_{j'})$ maximal.

Dans ce dernier cas, nous proposons de nous ramener au cas précédent. À cette fin, relativement à une même composante connexe \mathcal{C} , on cherchera à supprimer le plus petit nombre d'arcs de façon à rendre la restriction du graphe sur \mathcal{C} biparti et cela, aux sommets célibataires près créés. Une idée algorithmique peut être la suivante :

On commence par rendre le graphe \mathcal{C} symétrique en remplaçant chaque arc par une arête. Maintenant, relativement à une arête $\{X, Y\}$, on définit son degré de centralité au moyen de la somme $(n(X) + n(Y))$ où $n(X)$ (resp., $n(Y)$) est le nombre de sommets de \mathcal{C} pouvant être accessibles par un chemin ne passant pas par Y à partir de X (resp., à partir de Y ne passant pas par X). On coupe l'arête $\{X, Y\}$ (ou l'une des arêtes en cas d'exaequos) pour laquelle la valeur maximale de $(n(X) + n(Y))$ est atteinte. On poursuit de même. Le processus est terminé lorsque pour toute arête restante $\{X, Y\}$, $(n(X) + n(Y)) = 0$. Dans ce cas et pour les arêtes restantes, on restitue l'orientation et ainsi, les arcs initiaux concernés.

4 Niveaux et noeuds “significatifs” d’une hiérarchie implicative

4.1 Introduction

Un des points distinctifs les plus importants de la méthode de classification ascendante hiérarchique *AVL* sous-tendue par le programme *CHAVLH* (Classification Hiérarchique par Analyse de la Vraisemblance des Liens en cas de variables Hétérogènes) [37] est relatif à la méthode de détection des *niveaux* et *noeuds* les plus “significatifs” d’un arbre de classification [22, 23, 31]. À la base d’une telle détection est l’élaboration d’un coefficient statistique de nature combinatoire et non paramétrique entre une information σ de type *Similarité* sur l’ensemble E à organiser et une partition π sur E . Notons $C(\sigma, \pi)$ ce coefficient.

Chacun des niveaux de l’arbre des classifications sur E définissant une partition sur E , c’est le comportement de la distribution observée du coefficient $C(\sigma, \pi)$ le long de la suite croissante des niveaux qui permet la reconnaissance de ceux les plus “significatifs”. Ces derniers correspondent aux maxima locaux d’une telle distribution. Dans ce contexte nous appelons le coefficient $C(\sigma, \pi)$ “Statistique globale des niveaux”.

Dans notre approche la reconnaissance des *noeuds* les plus “significatifs” de l’arbre des classifications utilise le taux de variation du critère $C(\sigma, \pi)$ entre deux niveaux consécutifs. En réservant un niveau j pour chaque nouveau noeud ν_j produit par fusion binaire de deux branches, on définit :

$$C(\nu_j) = \Delta_j(C(\sigma, \pi)) = C(\sigma, \pi_j) - C(\sigma, \pi_{j-1}) \quad (93)$$

où π_l est la partition obtenue au niveau l , $0 \leq l \leq |E| - 1$.

$C(\nu_j)$ est appelé dans le contexte “Statistique locale des niveaux”. C’est le comportement de sa distribution observée le long de la suite des niveaux qui permet de détecter les noeuds les plus “significatifs”. Il s’agit des noeuds ν_j qui correspondent aux maxima locaux de $C(\nu_j)$ le long de la suite des niveaux.

Nous définirons également et au préalable une forme de critère local normalisé dont le point de départ est un indice brut d’association entre la fusion conduisant au noeud ν_j et la similarité σ .

Pour des raisons de simplicité et pour fixer les idées, nous nous limiterons ci-dessous au cas des arbres binaires où exactement une seule fusion de deux branches s’opère. L’extension au cas d’agrégations multiples se produisant à un même niveau peut être envisagée et l’a d’ailleurs été dans le cas de la classification ascendante hiérarchique symétrique ([26, 37]).

Notre propos va s’organiser comme suit. Le paragraphe 4.2 ci-dessous est réservé à l’expression du critère global entre une information de type *Similarité* et une partition. Nous rappellerons cette expression selon qu’il s’agit d’une similarité ordinale ou numérique. Nous étudierons ensuite l’adaptation du critère au cas de la structure associée à un niveau d’un arbre implicatif. Le paragraphe 4.3 suivra le même déroulement ; mais concernera le critère local d’évaluation d’un noeud, issu de la variation du critère global. Nous terminerons ce paragraphe en discutant un critère local proposé dans [9]. Le dernier paragraphe 4.4 a un caractère prospectif. On y étudie la conception d’un critère d’adéquation qui, pour un niveau donné, tient compte de tout le passé de l’arbre, depuis la racine jusqu’au niveau concerné.

4.2 Critère d’adéquation global

4.2.1 Cas d’une similarité symétrique

4.2.1.1. Cas d’une similarité ordinale Désignons par E l’ensemble à organiser et par n son cardinal. $F = P_2(E)$ désignera l’ensemble des paires ou parties à deux éléments de E . On suppose que E est muni d’une similarité ordinale qu’on appelle *préordonnance* et qui est un préordre total sur F qu’on notera ω . Ce dernier est établi de telle façon que :

$$(\forall (p, q) = (\{x, y\}, \{z, t\}) \in F \times F) \ p <_{\omega} q \Leftrightarrow \begin{array}{l} \text{les deux composantes } z \text{ et } t \text{ de } q \\ \text{sont plus ressemblantes} \\ \text{que les deux composantes } x \text{ et } y \text{ de } p. \end{array} \quad (94)$$

La donnée d'un préordre total ω sur un ensemble fini F est équivalente à la donnée d'une partition sur F et d'un ordre total et strict sur l'ensemble des classes. On peut noter

$$\pi(F) = \{F_l \mid 1 \leq l \leq m\} \quad (95)$$

une telle partition où la valeur de l'indice l est conforme à l'ordre croissant des classes, soit - pour l'ordre quotient -

$$F_1 < F_2 < \dots < F_l < \dots < F_m \quad (96)$$

Ainsi et dans notre contexte, pour un couple (p, q) de paires, si p et q appartiennent à une même classe F_l

$$p <_\omega q \text{ et } q <_\omega p \quad (97)$$

et si $(p, q) \in F_l \times F_{l'}$ avec $l < l'$,

$$p <_\omega q \text{ et } \neg q <_\omega p \quad (98)$$

Désignons par f_l le cardinal de la classe F_l , $1 \leq l \leq m$. Trois notions de rang peuvent être associées au préordre total ω , la notion de rang au sens large, celle au sens strict et celle dite du rang moyen. Nous les noterons respectivement rgl , rgs et rgm . Si q est un élément de F_l , on a :

$$rgl(q) = \text{card}\{p \mid p <_\omega q\} = \sum_{k=1}^l f_k \quad (99)$$

$$rgs(q) = \text{card}\{p \mid p <_\omega q \text{ et } \neg q <_\omega p\} = \sum_{k=1}^{l-1} f_k \quad (100)$$

$$rgm(q) = \sum_{k=1}^{l-1} f_k + \frac{f_l + 1}{2} \quad (101)$$

où $f = \text{card}(F)$.

Considérons le cas de référence où ω est un ordre total et strict sur F . Considérons la représentation de ω par son graphe strict au niveau de $F \times F$:

$$grs(\omega) = \{(p, q) \mid p <_\omega q \text{ et } \neg(q <_\omega p)\} \quad (102)$$

Afin de bâtir une mesure d'adéquation entre ω et une partition $\pi(E) = \{E_1, E_2, \dots, E_j, \dots, E_k\}$ de E , nous représentons $\pi(E)$ au même niveau; c'est-à-dire, au moyen d'un sous ensemble de $F \times F$. La donnée de $\pi(E)$ est équivalente à la donnée d'une partition en deux classes de F ; $S(\pi)$ et $R(\pi)$. $S(\pi)$ est l'ensemble des paires d'objets de E séparées par la partition $\pi(E)$ et où $R(\pi)$ est l'ensemble des paires d'objets de E réunis par la partition $\pi(E)$. Le préordre total sur F associé est défini en posant

$$S(\pi) < R(\pi) \quad (103)$$

de $F \times F$.

La construction suit une démarche caractéristique de la méthode *AVL* pour l'élaboration des coefficients d'association. On introduit l'indice brut

$$s(\omega, \pi) = \text{card}[gr(\omega) \cap (S(\pi) \times R(\pi))] \quad (104)$$

En notant par ϵ la fonction indicatrice de $R(\pi)$ dans F , on démontre [22, 23] que $s(\omega, \pi)$ se met sous la forme :

$$s(\omega, \pi) = \sum_{p \in F} \epsilon(p) \times l(p) - \frac{r \times (r + 1)}{2} \quad (105)$$

où $l(p)$ est ce que nous avons désigné ci-dessus par $rgl(p)$; c'est-à-dire, le rang au sens large de p et où $r = r(\pi)$ désigne le cardinal de $R(\pi)$.

On introduit à des fins de normalisation statistique une hypothèse d'absence de liaison ou d'indépendance associant à la partition π une partition aléatoire π^* dans l'ensemble muni d'une probabilité uniforme, de toutes les partitions de même type que π . Le type d'une partition est la suite ordonnée des cardinaux de ses classes. Cette forme de l'hypothèse d'absence de liaison est plus naturelle que celle duale où on fixe la partition π et on associe à l'ordre total ω un ordre aléatoire ω^* dans l'ensemble muni d'une probabilité uniforme, de tous les ordres totaux sur F . Il y en a $f!$ ($f = \text{card}(F)$). L'espérance mathématique de l'indice aléatoire $s(\omega, \pi^*)$ a une forme particulièrement simple

$$\mathcal{E}[s(\omega, \pi^*)] = \frac{r \times s}{2} \quad (106)$$

où $s = s(\pi) = \text{card}(S(\pi))$ est le nombre de paires séparées par la partition π ($r + s = f$).

La moyenne et la variance de la variable $l(p)$ sont $\frac{f+1}{2}$ et $\frac{f^2-1}{12}$. Introduisons celle associée, centrée et réduite :

$$\lambda(p) = \frac{l(p) - \frac{f+1}{2}}{\sqrt{\frac{f^2-1}{12}}} \quad (107)$$

Dans ces conditions, l'indice brut centré $s(\omega, \pi) - \mathcal{E}[s(\omega, \pi^*)]$, que nous pouvons noter $sc(\omega, \pi)$, se met - au coefficient multiplicatif près - $\sqrt{\frac{f^2-1}{12}}$, sous la forme :

$$\sum_{p \in F} \epsilon(p) \times \lambda(p) \quad (108)$$

Ainsi, au coefficient $\frac{f^2-1}{12}$ près, la variance de la variable aléatoire

$$\sum_{p \in F} \epsilon^*(p) \times \lambda(p) \quad (109)$$

où ϵ^* code la partition aléatoire π^* , est celle de l'indice aléatoire $s(\omega, \pi^*)$. Cette variance se met sous la forme [22, 23] :

$$\mathcal{V}_\epsilon = \left(\sum_{i=1}^k \pi_i^2 \right) \left(\sum_{p \in F} \lambda(p)^2 \right) + \left(\sum_{i=1}^k \pi_i^3 \right) \left(\sum_G \lambda(p) \times \lambda(p') \right) + \left(\sum_{i=1}^k \pi_i^2 \right)^2 \left(\sum_H \lambda(p) \times \lambda(p') \right) \quad (110)$$

Cette expression déjà simplifiée suppose la condition non contraignante que les n_i sont assez grands pour confondre n_i , $n_i - 1$, $n_i - 2$ et $n_i - 3$. $\pi_i = \frac{n_i}{n}$, $1 \leq i \leq k$. G est l'ensemble des couples de paires de la forme $(\{x, y\}, \{x, z\})$ ayant une seule composante commune et H est l'ensemble des couples de paires de la forme $(\{x, y\}, \{z, t\})$ sans composante commune. On en déduit dans ces conditions la statistique d'adéquation $S(\omega, \pi)$ de π à ω en considérant l'indice brut $s(\omega, \pi)$ centré et réduit :

$$S(\omega, \pi) = \frac{s(\omega, \pi) - \mathcal{E}[s(\omega, \pi^*)]}{\sqrt{\text{var}(s(\omega, \pi^*))}} \quad (111)$$

Ce dernier peut se mettre sous la forme :

$$S(\omega, \pi) = \frac{\sum_{p \in F} \epsilon(p) \times \lambda(p)}{\sqrt{V_\epsilon}} \quad (112)$$

il correspond à l'une des versions du critère que nous avons ci-dessus appelée $\mathcal{C}(\sigma, \pi)$. Cette version correspond au cas où σ est un ordre total et strict ω sur F .

Considérons maintenant la normalisation statistique par rapport à la forme duale de l'hypothèse d'absence de liaison où on fixe la partition π et où on associe à l'ordre total ω , un ordre aléatoire ω^* dans l'ensemble, muni d'une probabilité uniforme de tous les ordres totaux sur F . $\mathcal{E}[s(\omega^*, \pi)]$ a la même valeur que $\mathcal{E}[s(\omega, \pi^*)]$; cependant, $\text{var}[s(\omega^*, \pi)]$ a une expression notablement plus simple que $\text{var}[s(\omega, \pi^*)]$.

On a :

$$\mathcal{E}[s(\omega^*, \pi)] = \frac{r \times s}{2} \text{var}[s(\omega^*, \pi)] = \frac{r \times s \times (f+1)}{12} \quad (113)$$

Cette dernière expression peut constituer une approximation de $\text{var}[s(\omega, \pi^*)]$. La forme obtenue du critère $\mathcal{C}(\sigma, \pi)$ est alors :

$$S'(\omega, \pi) = \frac{s(\omega, \pi) - \frac{r \times s}{2}}{\sqrt{\frac{r \times s \times (f+1)}{12}}} \quad (114)$$

En fait, nous avons longtemps travaillé avec cette forme du critère. Cependant, actuellement et depuis déjà longtemps, c'est une forme plus rigoureuse du critère correspondante à $S(\omega, \pi)$ dans le cas d'une information σ numérique qui est prise en compte (programme CHAVLH) [37] (voir paragraphe ci-dessous 2.1.3.).

C'est la forme $S'(\omega, \pi)$ qui a été prise en considération dans une adaptation au cas d'une hiérarchie implicative [10].

4.2.1.2. Cas d'une préordonnance totale Reprenons l'expression du critère brut $s(\omega, \pi)$ dans le cas où ω est un ordre total et strict (cf. (104)); mais, en utilisant la notion de rang strict que nous notons ici s :

$$(\forall p \in F), l(p) = s(p) + 1; \text{ on a :} \\ s(\omega, \pi) = \sum_{p \in F} \epsilon(p) s(p) - \frac{r \times (r-1)}{2} \quad (115)$$

Cette expression se généralise au cas où ω est un préordre total, qu'on notera pour distinguer ϖ . Dans ce cas et avec les notations adoptées ci-dessus, on obtient :

$$s(\varpi, \pi) = \sum_{p \in F} \epsilon(p) s(p) - \sum_{1 \leq l < l' \leq m} r_l \times r_{l'} \quad (116)$$

où r_l est le nombre de paires réunies par la partition π qui se trouvent dans la classe F_l de la préordonnance totale ϖ . S'il est tout à fait envisageable d'effectuer l'étude distributionnelle de

$$\sum_{p \in F} \epsilon^*(p) s(p)$$

il paraît par contre sensiblement plus difficile de considérer cette étude pour la globalité de l'indice aléatoire $s(\varpi, \pi^*)$.

Nous allons plutôt partir de l'extension de la notion de rang au sens large à partir de la notion de rang

moyen [23] (cf. 101), noté rgm ; mais que nous noterons ici m pour simplifier. La moyenne de m est $\frac{f+1}{2}$, alors que la variance peut se mettre sous la forme :

$$var(m) = \sum_{1 \leq k \leq m} f_k \times (f_1 + f_2 + \dots + f_{k-1} + \frac{f_k + 1}{2})^2 - \left(\frac{f+1}{2}\right)^2 \quad (117)$$

Dans ces conditions, nous partirons de l'expression formelle ci-dessus (108) comme indice brut entre ϖ et π :

$$sc_m(\varpi, \pi) = \sum_{p \in F} \epsilon(p) \lambda(p) \quad (118)$$

où ici

$$\lambda(p) = \frac{m(p) - \frac{f+1}{2}}{\sqrt{var(m)}} \quad (119)$$

L'espérance mathématique de l'indice brut aléatoire $s_m(\varpi, \pi^*)$ est nulle. Sa variance peut se mettre sous l'une de deux formes. La première est celle (110) exprimée ci-dessus où $\lambda(p)$ est défini par (119). La deuxième forme résulte d'une formule due à Mantel [35]. Pour l'exprimer introduisons l'ensemble $E^{[2]}$ des couples d'objets à composantes distinctes ($card(E^{[2]}) = n(n-1)$) et étendons λ et ϵ à $E^{[2]}$ en posant

$$\lambda(x, y) = \lambda(y, x) \text{ et } \epsilon(x, y) = \epsilon(y, x)$$

pour tout $(x, y) \in E^{[2]}$. Considérons alors les expressions suivantes :

$$\begin{aligned} A_1 &= \left(\sum_{E^{[2]}} \lambda(x, y) \right)^2, \quad A_2 = \sum_{x \in E} \left(\sum_{y \in E - \{x\}} \lambda(x, y) \right)^2, \quad A_3 = \sum_{E^{[2]}} \lambda(x, y)^2 \\ B_1 &= \left(\sum_{E^{[2]}} \epsilon(x, y) \right)^2 = \left(\sum_{1 \leq i \leq k} n_i(n_i - 1) \right)^2 \\ B_2 &= \sum_{x \in E} \left(\sum_{y \in E - \{x\}} \epsilon(x, y) \right)^2 = \sum_{1 \leq i \leq k} n_i(n_i - 1)^2 \\ B_3 &= \sum_{E^{[2]}} \epsilon(x, y)^2 = \sum_{1 \leq i \leq k} n_i(n_i - 1) \end{aligned} \quad (120)$$

L'expression de la variance de $s_m(\varpi, \pi^*)$ s'écrit alors :

$$\mathcal{V}_\epsilon = \frac{1}{2n^{[2]}} A_3 B_3 + \frac{1}{n^{[3]}} (A_2 - A_3)(B_2 - B_3) + \frac{1}{4n^{[4]}} (A_1 - 4A_2 + 2A_3)(B_1 - 4B_2 + 2B_3) - \frac{1}{4(n^{[2]})^2} A_1 B_1 \quad (121)$$

où $n^{[j]}$ indique le j -ème moment factoriel :

$$n^{[j]} = n \times (n-1) \times \dots \times (n-j+1) \quad (122)$$

Dans la mesure où les n_i sont suffisamment "grands" on peut proposer l'expression approchée

$$\begin{aligned} \mathcal{V}_\epsilon &= \frac{1}{2} A_3 \sum_{1 \leq i \leq k} \pi_i^2 + (A_2 - A_3) \sum_{1 \leq i \leq k} \pi_i^3 + \frac{1}{4} (A_1 - 4A_2 + 2A_3) \left(\sum_{1 \leq i \leq k} \pi_i^2 \right)^2 - \frac{1}{4} A_1 \left(\sum_{1 \leq i \leq k} \pi_i^2 \right)^2 \\ &= A_3 \times \frac{1}{2} \sum_{1 \leq i \leq k} \pi_i^2 + (A_2 - A_3) \sum_{1 \leq i \leq k} \pi_i^3 - (A_2 - \frac{1}{2} A_3) \left(\sum_{1 \leq i \leq k} \pi_i^2 \right)^2 \\ &= \frac{1}{2} A_3 \left(\sum_{1 \leq i \leq k} \pi_i^2 \right) \left(1 + \left(\sum_{1 \leq i \leq k} \pi_i^2 \right) \right) + (A_2 - A_3) \sum_{1 \leq i \leq k} \pi_i^3 - A_2 \left(\sum_{1 \leq i \leq k} \pi_i^2 \right)^2 \end{aligned} \quad (123)$$

En effet, les termes en A_1 se simplifient ; et déjà, dans le contexte de la définition de λ , $A_1 = 0$.

Le critère d'adéquation prend dans ces conditions la forme :

$$S_m(\varpi, \pi) = \frac{sc_m(\varpi, \pi)}{\mathcal{V}_\epsilon} \quad (124)$$

4.2.1.3. Cas d'une similarité numérique La détermination d'une préordonnance totale sur l'ensemble E à organiser par la classification, est très généralement associée au choix d'un indice numérique de similarité entre éléments de E [21, 22]. L'idée de retenir une information ordinale de la ressemblance a pu répondre à deux soucis. Le premier concerne les propriétés d'invariance [2] ou de stabilité [21, 22] et le second, relié d'ailleurs au premier, à vouloir comparer deux structures combinatoires de mêmes types sur E , à savoir deux préordres totaux sur $F = P_2(E)$. Or il s'avère que la stabilité n'est acquise que dans des conditions particulières [21, 22]. Et, dans la mesure où l'indice numérique de similarité est bien fondé sur les plans formel et statistique, ne tenir compte que du préordre total associé, entraîne une perte de richesse de l'information *Similarité*. D'autre part et c'est très important, l'accroissement de la complexité calcul est celle d'un tri sur un ensemble de $n(n-1)/2$ valeurs (donc de l'ordre de $n^2 \log n$, en moyenne), ce qui devient lourd avec l'accroissement de la taille de l'ensemble E .

Maintenant, si on considère un indice tel que $sc_m(\varpi, \pi)$ auquel nous avons été conduits et qui résulte du codage par la notion de rang moyen de la similarité brute, on peut affiner très sensiblement un tel indice en substituant au rang moyen une similarité numérique suffisamment étudiée et qui tient compte de la description formelle et statistique de E . Désignons par c un tel indice. On suppose que c est normalisé de la même façon que λ ci-dessus ; c'est-à-dire :

$$\sum_{p \in F} c(p) = 0 \text{ et } \frac{1}{f} \sum_{p \in F} c(p)^2 = 1 \quad (125)$$

Signalons ici que l'indice développé dans *AVL* suppose une normalisation d'abord "locale" ne tenant compte que des deux éléments à comparer dans leurs descriptions statistiques respectives. Ensuite, intervient une normalisation "globale" où les conditions (125) deviennent satisfaites [25, 27, 33].

Dans ces conditions et par rapport à (118), on posera pour le cas numérique l'indice centré

$$sc_\nu(c, \pi) = \sum_{p \in F} \epsilon(p)c(p) \quad (126)$$

La valeur approchée de la variance \mathcal{W}_ϵ de l'indice aléatoire $sc_\nu(c, \pi^*)$ prend une forme analogue à \mathcal{V}_ϵ (cf. (123)). On aura juste à remplacer dans les expressions de A_2 et A_3 $\lambda(x, y)$ par $c(x, y)$ pour $\{x, y\} \in F$. C'est le coefficient

$$S_\nu(c, \pi) = \frac{sc_\nu(c, \pi)}{\sqrt{\mathcal{W}_\epsilon}} \quad (127)$$

qui définit dans le contexte du programme CHAVLH (Classification Hiérarchique par *AVL* en cas de données Hétérogènes) [37] le critère d'adéquation entre une partition et une similarité numérique. $S_\nu(c, \pi)$ est la version du critère général $\mathcal{C}(\sigma, \pi)$ que nous avons adoptée pour la détection des niveaux les plus significatifs (voir l'introduction ci-dessus).

Cependant, la forme ordinale du critère peut avoir un intérêt propre, notamment lorsqu'on souhaite certaines propriétés d'invariance. D'autre part, conceptuellement, on peut se rendre compte comment la forme ordinale du critère nous a conduit à la forme numérique.

4.2.2 Cas d'une similarité orientée (implicative)

4.2.2.1. Cas d'une similarité ordinale On se trouve ici dans le contexte de la construction d'une hiérarchie implicative (cf. paragraphes 2 et 3). Revenons à ce contexte où c'est un ensemble \mathcal{A} d'attributs booléens qui est

organisé. Sur \mathcal{A} est donnée une similarité implicative que nous continuerons comme au paragraphe 3 à noter c , bien que le sens ne soit pas celui du paragraphe qui précède directement. Soit donc

$$\left\{ c(a^j, a^k) \mid (a^j, a^k) \in \mathcal{A}^{[2]} \right\} \quad (128)$$

où $\mathcal{A}^{[2]}$ est l'ensemble des couples à composantes distinctes de \mathcal{A} .

Dans [11, 16] on considère une transposition directe du critère pour une donnée ordinale sticte et c'est la forme simplifiée (114) qui est adoptée. Du préordre total ϖ sur $\mathcal{A}^{[2]}$ associé à c , on retient le graphe strict :

$$grs(\varpi) = \left\{ \left((a^i, a^j), (a^k, a^l) \right) \in \mathcal{A}^{[2]} \times \mathcal{A}^{[2]} \mid c(a^i, a^j) < c(a^k, a^l) \right\} \quad (129)$$

qu'on confronte avec $R'(\pi) \times S'(\pi)$ où $R'(\pi)$ est l'ensemble des couples d'attributs d'une même classe de la partition π et où $S'(\pi)$ est l'ensemble des couples d'attributs séparés par π . La forme du critère adoptée est alors directement celle de $S'(\omega, \pi)$ (114) à cela près que l'ordre total ω sur l'ensemble F des paires est remplacé par la partie stricte du préordre total ϖ sur l'ensemble $\mathcal{A}^{[2]}$ des couples. Cependant, la normalisation simplifiée (moyenne $r \times s/2$ et variance $r \times s \times (f+1)/12$) n'est justifiée que dans le cas d'origine et pour un modèle aléatoire associant pour une partition π fixée, à l'ordre total ω , un ordre total aléatoire ω^* . Le modèle dual et d'ailleurs plus adéquat, où ω est fixé et où à la partition π on associe une partition aléatoire π^* , se trouve moins concerné pour la valeur *exacte* de la variance par l'expression simplifiée ($r \times s \times (f+1)/12$).

Dans cette adaptation du critère tout se passe comme si, ce qu'il y a lieu de retenir d'un niveau de l'arbre implicatif était de même nature que dans le cas classique (non orienté); c'est-à-dire, une partition de l'ensemble \mathcal{A} des attributs distinguant les couples d'attributs dont les deux composantes sont dans une même classe et ceux, dont les deux composantes se trouvent dans deux classes distinctes.

Dans le cadre d'une approche distincte que nous pourrions discuter ci-dessous [9], les auteurs cherchent à tenir compte d'une donnée de nature ordinale. Très clairement, par rapport à notre analyse (cf. paragraphe 2.2), il s'agit pour une classe C de retenir l'ordre total et strict déduit de la construction ultramétrique et orientée de la classe C . Notons $o(C)$ cet ordre total et désignons par

$$a^{j_1} < a^{j_2} < \dots < a^{j_i} < \dots < a^{j_m} \quad (130)$$

la suite ordonnée des attributs de C conformément à $o(C)$. En notant $\mathcal{O}(a^{j_i})$ le sous ensemble des objets où a^{j_i} est à *VERAI*, on n'a pas nécessairement comme il est assuré dans [9]

$$card(\mathcal{O}(a^{j_1})) \leq \dots \leq card(\mathcal{O}(a^{j_i})) \leq \dots \leq card(\mathcal{O}(a^{j_m})) \quad (131)$$

(voir ci-dessous)

Désignons par $\pi^o = \{C^h \mid 1 \leq h \leq k\}$ la partition en classes totalement ordonnées, produite - dans notre contexte - à un niveau donné de la hiérarchie implicative. Sur \mathcal{A} se trouve défini un ordre partiel résultant des différents segments orientés C^h , $1 \leq h \leq k$. Plus précisément, le graphe d'un tel ordre est :

$$\left\{ (a^i, a^j) \mid \exists h, a^i, a^j \in C^h \text{ et } a^i < a^j \text{ pour } o(C^h) \right\} \quad (132)$$

Nous indiquerons par η sa fonction indicatrice. Pour des raisons de simplicité et d'efficacité et, compte tenu de l'analyse précédente, nous retiendrons la forme numérique (108) ou (118) de la similarité ordinale. L'indice brut d'adéquation se met alors sous la forme :

$$so_\nu(c, \pi^o) = \sum_{q \in \mathcal{A}^{[2]}} \eta(q) \times c(q) \quad (133)$$

À la partition en classes orientées π^o nous associons une partition aléatoire orientée π^{o^*} qu'on notera

$$\pi^{\sigma^*} = \{C^{h^*} \mid 1 \leq h \leq k\} \quad (134)$$

où la classe aléatoire C^{h^*} correspond à un segment orienté de la forme

$$a^{j_1^*} < a^{j_2^*} < \dots < a^{j_i^*} < \dots < a^{j_m^*} \quad (135)$$

qui correspond à une permutation aléatoire prise selon un modèle uniforme, sur une partie de cardinal m de l'ensemble \mathcal{A} des attributs.

L'expression générale de l'espérance mathématique est [22, 27] :

$$\mathcal{E}[so_\nu(c, \pi^{\sigma^*})] = \frac{1}{n^{[2]}} \left(\sum_{J^{[2]}} \eta(i, j) \right) \left(\sum_{J^{[2]}} c(i, j) \right) \quad (136)$$

où $J = \{1, 2, \dots, j, \dots, p\}$ indexe l'ensemble (A) des attributs.

En désignant par m_h le cardinal de la classe C_h , $1 \leq h \leq k$, le second membre de (136) se met sous la forme :

$$\begin{aligned} & \frac{1}{2} \left(\sum_{1 \leq h \leq k} \frac{m_h(m_h - 1)}{n(n-1)} \right) \left(\sum_{J^{[2]}} c(i, j) \right) \\ & \simeq \frac{1}{2} \left(\sum_{1 \leq h \leq k} \pi_h^2 \right) \left(\sum_{J^{[2]}} c(i, j) \right) \end{aligned} \quad (137)$$

où $\pi_h = \frac{m_h}{n}$, $1 \leq h \leq k$.

Pour ce qui est de la variance de l'indice brut aléatoire, elle ne peut être déduite de l'expression de Mantel. En effet, cette dernière n'est valable que si les deux relations à comparer sont soit symétriques, soit antisymétriques, ce qui n'est pas le cas ici. Il nous faut nous référer à une formule que nous avons établie et qui a un caractère analytique [22, 27]. On y introduit par rapport à J les ensembles d'indexation :

$$\begin{aligned} G_1 &= \{[(i, j), (i, k)]\}, \quad G'_1 = \{[(i, j), (h, i)]\} \\ G_2 &= \{[(i, j), (h, j)]\}, \quad G'_2 = \{[(i, j), (j, k)]\} \\ &\text{et } H = \{[(i, j), (h, k)]\} \end{aligned} \quad (138)$$

où des lettres différentes indiquent des indices distincts. On montre dans ces conditions que la variance de l'indice brut aléatoire $so_\nu(c, \pi^{\sigma^*})$ se met sous la forme :

$$\begin{aligned} \mathcal{V}_\eta &= \frac{1}{n^{[2]}} \left(\sum_{1 \leq h \leq k} m_h(m_h - 1) \right) \left(\sum_{J^{[2]}} c(i, j)^2 \right) \\ &+ \frac{1}{n^{[3]}} \left(\sum_{1 \leq h \leq k} m_h(m_h - 1)(m_h - 2) \right) \times \\ &\quad \left\{ \sum_{G_1} c(i, j)c(i, k) + \sum_{G'_1} c(i, j)c(h, i) + \sum_{G_2} c(i, j)c(h, j) + \sum_{G'_2} c(i, j)c(j, k) \right\} \\ &+ \frac{1}{n^{[4]}} \left(\sum_{1 \leq h \leq k} m_h(m_h - 1)(m_h - 2)(m_h - 3) + \sum_{1 \leq h \neq h' \leq k} m_h(m_h - 1)m_{h'}(m_{h'} - 1) \right) \times \\ &\quad \left\{ \sum_H c(i, j)c(h, k) \right\} - \frac{1}{4} \left\{ \left(\sum_{1 \leq h \leq k} \frac{m_h(m_h - 1)}{n^{[2]}} \right) \left(\sum_{J^{[2]}} c(i, j) \right) \right\}^2 \end{aligned} \quad (139)$$

Le critère global d'adéquation se met comme précédemment sous la forme :

$$S_{o_\nu}(c, \pi^o) = \frac{s_{o_\nu}(c, \pi^o) - \mathcal{E}[s_{o_\nu}(c, \pi^{o^*})]}{\sqrt{\mathcal{V}_\eta}} \quad (140)$$

où \mathcal{V}_η vient d'être exprimé ci-dessus et étant entendu que tout moment factoriel $m_h^{[s]}$ pour $s = 2, 3$ ou $s = 4$ qui apparaît sous forme multiplicative est posé égal à 0 dès lors que $m_h < s$.

4.3 Critère d'adéquation local

4.3.1 Cas d'une similarité symétrique

4.3.1.1. Critère obtenu à partir d'un indice brut local; forme ordinale puis numérique Considérons la construction ascendante hiérarchique d'un arbre *binnaire* de classification tel que celui de la méthode *AVL* issu d'une similarité symétrique. Considérons deux niveaux consécutifs $h - 1$ et h d'un tel arbre et notons comme suit les partitions de ces niveaux :

$$\pi_{h-1} = \{C_1^{h-1}, C_2^{h-1}, \dots, C_j^{h-1}, \dots, C_{p-h+1}^{h-1}\} \quad (141)$$

et

$$\pi_h = \{C_1^h, C_2^h, \dots, C_j^h, \dots, C_{p-h}^h\} \quad (142)$$

où π_h se déduit de π_{h-1} par la fusion d'exactly deux classes. Sans restreindre aucunement la généralité, supposons qu'il s'agit des deux classes C_1^{h-1} et C_2^{h-1} dont la fusion donne C_1^h . On a :

$$C_1^h \leftarrow C_1^{h-1} \cup C_2^{h-1}$$

et pour $2 \leq j \leq h$, $C_j^h \leftarrow C_j^{h-1}$ (143)

La première conception d'un critère local a consisté à procéder selon la démarche générale où un indice brut de comparaison est normalisé par rapport à une hypothèse d'absence de liaison [22]. Cette conception s'est située dans un contexte ordinal. Cependant, nous verrons ici son extension dans le contexte d'une similarité numérique.

On suppose donc comme au paragraphe 4.2.1.1. ci-dessus, que la similarité est ordinale et notée comme ci-dessus ω . On en retient le graphe strict $grs(\omega)$ [cf.(102)]. Du passage ci-dessus entre π_{h-1} et π_h , on distingue au niveau de l'ensemble F des paires d'éléments l'ensemble - ici noté - $r(\pi_h)$ des paires qu'on vient de réunir par rapport à l'ensemble $S(\pi_h)$ des paires laissées séparées. $X * Y$ désignant l'ensemble des paires $\{x, y\}$ où $x \in X$ et $y \in Y$, on a :

$$r(\pi_h) = C_1^{h-1} * C_2^{h-1}$$

$$S(\pi_h) = (C_1^{h-1} \cup C_2^{h-1}) * \left(\sum_{3 \leq j \leq p-h+1} C_j^{h-1} \right)$$

$$+ \sum_{3 \leq j \leq p-h+1} C_j^{h-1} * C_{j'}^{h-1} \quad (144)$$

(somme ensembliste)

Dans ces conditions, on confronte $S(\pi_h) \times r(\pi_h)$ avec $grs(\omega)$ au moyen de l'indice brut :

$$s(\omega, \pi_h - \pi_{h-1}) = \text{card}[grs(\omega) \cap (S(\pi_h) \times r(\pi_h))] \quad (145)$$

La notation $\pi_h - \pi_{h-1}$ veut indiquer la variation de la partition produite entre les niveaux $h - 1$ et h .

La normalisation considérée est de même type que celle adoptée pour $s(\omega, \pi)$ [cf.(114)]. Elle se justifie

dans le cas où à ω on associe de façon uniforme un ordre aléatoire ω^* sur l'ensemble $S(\pi_h) \cup r(\pi_h)$ que nous notons ici G_h . Le coefficient normalisé se présente dans ces conditions sous la forme :

$$S(\omega, \pi_h - \pi_{h-1}) = \frac{s(\omega, \pi_h - \pi_{h-1}) - \frac{r'_h \times s_h}{2}}{\sqrt{r'_h \times s_h \times (r'_h + s_h + 1)}/12} \quad (146)$$

où $r'_h = \text{card}(r(\pi_h))$ et $s_h = \text{card}(S(\pi_h))$. Plus précisément,

$$\begin{aligned} r'_h &= m_1^{h-1} \times m_2^{h-1} \\ \text{et } s_h &= (m_1^{h-1} + m_2^{h-1}) \times \sum_{3 \leq j \leq p-h+1} m_j^{h-1} + \sum_{3 \leq j \leq p-h+1} m_j^{h-1} \times m_{j'}^{h-1} \end{aligned} \quad (147)$$

Compte tenu de l'analyse effectuée dans le cas du critère global on peut considérer une adaptation simplifiée d'un tel critère dans le contexte numérique. À cet effet, on posera un indice brut de la forme :

$$sc_\nu(c, \pi_h - \pi_{h-1}) = \sum_{p \in C_1^{h-1} * C_2^{h-1}} c(p) \quad (148)$$

qui représente la somme des similarités des paires qu'on vient de fusionner entre π_{h-1} et π_h .

Le modèle permutational sur G_h conduit à la moyenne et à la variance suivantes de l'indice brut aléatoire $sc_\nu(c^*, \pi_h - \pi_{h-1})$ associé :

$$\mathcal{E}[sc_\nu(c^*, \pi_h - \pi_{h-1})] = \frac{r'_h \times \text{moy}_{G_h}(c)}{g_h} \quad (149)$$

$$\text{var}[sc_\nu(c^*, \pi_h - \pi_{h-1})] = \frac{g_h^2}{g_h - 1} \times \frac{r'_h}{g_h} \times \left(1 - \frac{r'_h}{g_h}\right) \text{var}_{G_h}(c) \quad (150)$$

où $g_h = \text{card}(G_h)$ et où $\text{moy}_{G_h}(c)$ et $\text{var}_{G_h}(c)$ désignent la moyenne et la variance de l'indice c sur G_h . Le critère local prend dans ces conditions la forme :

$$Sc_\nu(c, \pi_h - \pi_{h-1}) = \frac{sc_\nu(c, \pi_h - \pi_{h-1}) - (r'_h \times \text{moy}_{G_h}(c))/g_h}{\sqrt{\frac{g_h^2}{g_h - 1} \times \frac{r'_h}{g_h} \times \left(1 - \frac{r'_h}{g_h}\right) \text{var}_{G_h}(c)}} \quad (151)$$

Ce modèle aléatoire de normalisation aurait bien pu être considéré dans le cas du critère global. Mais nous avons préféré une forme plus rigoureuse et plus pertinente du modèle aléatoire. Cette dernière fait appel à un modèle permutational ; mais au niveau de l'ensemble E des éléments à organiser. Elle ne peut être considérée ici où l'ensemble G_h ne correspond pas à l'ensemble des paires d'une partie de E .

4.3.1.2. Critère local obtenu à partir de la variation d'un critère global; forme ordinale ou numérique de la similarité Nous avons désigné par $S_m(\varpi, \pi)$ (cf. (124)) le critère global d'adéquation entre une préordonnance totale ϖ et une partition π et par $S_\nu(c, \pi)$ un tel critère dans le cas où il s'agit d'une similarité numérique c . Désignons par $\mathcal{C}(\sigma, \pi)$ l'un ou l'autre de ces deux critères. À partir de $\mathcal{C}(\sigma, \pi)$, un critère local permettant le passage entre les partitions π_{h-1} et π_h des niveaux $h-1$ et h , produisant le noeud ν_h d'un arbre binaire de classification, est fourni par le taux d'accroissement de $\mathcal{C}(\sigma, \pi)$; soit :

$$\mathcal{C}(\nu_h) = \Delta_h(\mathcal{C}(\sigma, \pi)) = \mathcal{C}(\sigma, \pi_h) - \mathcal{C}(\sigma, \pi_{h-1}) \quad (152)$$

Ce critère ainsi que celui qui le précède (151) permettent avec des logiques proches la détection des noeuds les plus "significatifs" de l'arbre des classifications. Ces derniers correspondent au maxima locaux de la distribution de $\Delta_h(\mathcal{C}(\sigma, \pi))$ le long de la suite des niveaux de l'arbre des classifications. Il est intéressant de comparer les ensembles de noeuds retenus par l'un ou l'autre des différents critères exprimés.

4.3.2 Cas d'une similarité orientée

Le traitement de ce cas va suivre une ligne parallèle à celui précédent à cela près qu'il y aura lieu de tenir compte du caractère orienté de la similarité et de l'arbre d'organisation classificatoire.

4.3.2.1. Critère obtenu à partir d'un indice brut local; forme ordinale puis numérique Désignons par π_{h-1}^o et π_h^o les deux partitions en classes qu'on peut continuer à écrire au moyen des expressions (141), étant entendu que chacune des classes C_j^{h-1} ($1 \leq j \leq p-h+1$) est sous-tendue par un ordre total déduit de la construction ultramétrique de la classe. $C_1^{h-1} \times C_2^{h-1}$ est l'ensemble des couples qu'on vient de fusionner dans le passage entre π_{h-1}^o et π_h^o . L'ordre total dont se trouve muni C_1^h résulte de la concaténation de gauche à droite de l'ordre total dont se trouve muni C_1^{h-1} et de l'ordre total dont se trouve muni C_2^{h-1} . Désignons par $r(\pi_h^o)$ l'ensemble des couples ordonnés qu'on fusionne dans le passage de π_{h-1}^o à π_h^o . Ceux qui restent séparés sont définis par l'ensemble :

$$S(\pi_h^o) = (C_1^{h-1} + C_2^{h-1})() \left(\sum_{3 \leq j \leq p-h+1} C_j^{h-1} \right) + \sum_{3 \leq j < j' \leq p-h+1} C_j^{h-1}()C_{j'}^{h-1} \quad (153)$$

(somme ensembliste) où, relativement à deux ensembles X et Y , nous avons noté $X()Y$ l'ensemble des couples ordonnés $(X \times Y) + (Y \times X)$ (somme ensembliste).

Relativement à une préordonnance totale définissant un préordre total ω , cette fois-ci sur l'ensemble des couples ordonnés d'éléments distincts (il s'agit de $\mathcal{A}^{[2]}$ dans notre cas), l'indice brut prend la même forme (145) que ci-dessus :

$$s(\omega, \pi_h^o - \pi_{h-1}^o) = \text{card}[grs(\omega) \cap (S(\pi_h^o) \times r(\pi_h^o))] \quad (154)$$

Nous laissons ici le soin au lecteur de poursuivre le même type de développement que ci-dessus réservé au cas symétrique. À cet égard et pour la normalisation, l'ensemble des paires est remplacé par l'ensemble des couples à composantes distinctes.

Dans le cas numérique l'indice brut devient :

$$sc_\nu(c, \pi_h^o - \pi_{h-1}^o) = \sum_{q \in C_1^{h-1} \times C_2^{h-1}} c(q) \quad (155)$$

qui représente la somme des similarités implicatives des couples qu'on vient de fusionner entre π_{h-1}^o et π_h^o . Les correspondants des formules (149) et (150) pour l'espérance mathématique et la variance de l'indice brut aléatoire $sc_\nu(c^*, \pi_h^o - \pi_{h-1}^o)$ s'obtient en substituant à l'ensemble des paires, l'ensemble des couples à composantes distinctes.

4.3.2.2. Critère local obtenu à partir de la variation d'un critère global; forme ordinale ou numérique de la similarité Ici encore nous allons directement retenir la forme numérique. Dans celle ordinale, on remplacera la similarité numérique par son rang au niveau de l'ensemble $\mathcal{A}^{[2]}$ des couples. Nous reprenons donc le critère $S_{o_\nu}(c, \pi^o)$ (cf. 140) que nous noterons ici $\mathcal{C}(c, \pi^o)$ et définir comme dans (152) le taux de variation $\Delta_h(\mathcal{C}(c, \pi^o))$ entre deux niveaux consécutifs :

$$\Delta_h(\mathcal{C}(c, \pi^o)) = \mathcal{C}(c, \pi_h^o) - \mathcal{C}(c, \pi_{h-1}^o) \quad (156)$$

Comme ci-dessus, ce sont les maxima locaux de la distribution de $\Delta_h(\mathcal{C}(c, \pi^o))$ le long de la suite des niveaux de l'arbre implicatif des classifications, qui permettent de détecter les noeuds les plus "significatifs".

4.4 Un nouveau critère local

Dans leur article [9] les auteurs proposent un nouveau critère local issu de la comparaison entre les deux classes C_k et C_{k+1} sous-tendues par deux noeuds consécutifs ν_k et ν_{k+1} ; mais avec un coefficient multiplicatif d'adéquation de nature ordinale concernant la classe C_{k+1} . Ce critère est mis sous la forme :

$$co(C_{k+1}) = \frac{c(C_{k+1})}{c(C_k)} \times o(C_{k+1}) \quad (157)$$

Pour préciser les différents arguments de cette formule, introduisons relativement à une même classe C_h deux préordres totaux $\omega_t(C_h)$ et $\omega_o(C_h)$. $\omega_t(C_h)$ est le préordre total associé à la suite croissante du nombre d'occurrences des différents attributs de C_h . $\omega_o(C_h)$ est l'ordre total sur C_h déduit de la construction ultramétrique de l'arbre implicatif restreint à C_h . $c(C_h)$ est la valeur de l'indice de cohésion de la classe $c(C_h)$:

$$c(C_h) = \sum \left\{ c(a^i, a^j) \mid (a^i, a^j) \in \omega_o(C_h) \right\} \quad (158)$$

Enfin $o(C_h)$ est un indice probabiliste normalisant la similarité entre $\omega_t(C_h)$ et $\omega_o(C_h)$ en termes de la relative petitesse du nombre d'inversions entre $\omega_o(C_h)$ et $\omega_t(C_h)$.

Compte tenu de son caractère particulier il importe de comparer le comportement du critère $co(C_{k+1})$ avec ceux des critères de conception conforme à l'approche *AVL*, présentés ci-dessus.

4.5 Critères ultramétriques

Relativement à un arbre binaire de classification hiérarchique (symétrique ou orienté), deux types de critères ont été présentés. Le premier dit "global" prend en considération un seul niveau k de la hiérarchie. Dans le second dit "local" c'est le passage du niveau k au niveau $k+1$ qui est pris en considération. Alors que les niveaux k et $k+1$ définissent deux partitions consécutives où, dans le cas orienté, les classes sont munies d'ordres totaux; nous allons considérer ici des critères qui prennent en considération toute la structure ultramétrique de la section commençante de l'arbre jusqu'à un niveau donné h .

Désignons ici une nouvelle fois par E l'ensemble organisé et par $F = P_2(E)$ (resp., $G = E_{[2]}$) l'ensemble de ses paires (resp. de ses couples à composantes distinctes). La structure ultramétrique mentionnée peut être représentée par une préordonnance ultramétrique [21, 22] qui est une préordonnance totale ϖ_u sur F dans le cas d'une similarité symétrique (resp., sur G dans le cas d'une similarité implicative (orientée)). Une même classe du préordre total ϖ_u est formée de l'ensemble des paires (resp., des couples) qu'on vient de relier à un niveau donné de l'arbre des classifications. Si C_j et $C_{j'}$ sont les deux classes qu'on vient de fusionner à ce niveau, il s'agit de $C_j * C_{j'}$ dans le cas symétrique et de $C_j \times C_{j'}$ dans le cas orienté (voir ci-dessus pour les notations). Le préordre total ϖ_u est supposé établi de gauche à droite de façon croissante avec la ressemblance ultramétrique. Ainsi, si

$$\pi(E) = \left\{ C_j \mid 1 \leq j \leq k \right\} \quad (159)$$

est la partition du dernier niveau retenu, la première classe du préordre ϖ_u est formée de l'ensemble

$$\sum_{1 \leq j < j' \leq k} C_j * C_{j'} \quad (160)$$

dans le cas symétrique et de l'ensemble

$$\sum_{1 \leq j < j' \leq k} (C_j \times C_{j'} + C_{j'} \times C_j) \quad (161)$$

dans le cas orienté.

Nous coderons ce préordre total en utilisant la notion déjà vue de "rang moyen" et que nous noterons ici rgm_u . L'indice brut d'association entre la section commençante de l'arbre et une similarité σ sur E qui s'exprime de façon numérique, se met dans ces conditions sous la forme :

$$s(\varpi_u, \sigma) = \sum_{p \in F} rgm_u(p)\sigma(p) \quad (162)$$

dans le cas symétrique et

$$s_o(\varpi_u, \sigma) = \sum_{q \in G} rgm_u(q)\sigma(q) \quad (163)$$

dans le cas orienté.

Selon toujours le même principe général la normalisation statistique suppose l'association à ϖ_u , selon un modèle aléatoire permutatif sur E , une préordonnance ultramétrique aléatoire ϖ_u^* . Cette dernière correspond en fait à une permutation aléatoire des feuilles de l'arbre. Les calculs de l'espérance mathématique et de la variance de l'indice brut aléatoire

- $(\mathcal{E}[s(\varpi_u^*, \sigma)] \text{ et } \text{var}[s(\varpi_u^*, \sigma)])$ dans le cas symétrique,
- $(\mathcal{E}[s_o(\varpi_u^*, \sigma)] \text{ et } \text{var}[s_o(\varpi_u^*, \sigma)])$ dans le cas orienté,

se font selon des formules analogues à celles considérées ci-dessus [voir (110), (123), (136), (136), (139)].

Néanmoins, il faudra tenir compte de la nature particulière de la structure comparée avec σ qui n'est plus une partition mais une préordonnance ultramétrique.

Considérons le développement commençant de l'arbre jusqu'au niveau k et désignons par $\mathcal{C}_u(k)$ (resp., $\mathcal{C}_u^o(k)$) le critère brut $s(\varpi_u, \sigma)$ (resp., $s_o(\varpi_u, \sigma)$) normalisé dans le cas symétrique (resp. orienté). Ce critère mesure l'adéquation globale entre la structure ultramétrique de l'arbre commençant et la similarité σ . Le taux de variation d'un tel critère entre deux niveaux consécutifs $h-1$ et h , évalue la pertinence du noeud ν_h qui vient de se former au niveau h par la fusion de deux classes. En notant $\mathcal{C}_u(\nu_h)$ (resp., $\mathcal{C}_u^o(\nu_h)$) ce taux de variation dans le cas symétrique (resp., orienté), on a :

$$\mathcal{C}_u(\nu_h) = \mathcal{C}_u(h) - \mathcal{C}_u(h-1) \quad (164)$$

et

$$\mathcal{C}_u^o(\nu_h) = \mathcal{C}_u^o(h) - \mathcal{C}_u^o(h-1) \quad (165)$$

L'évolution le long de la suite des niveaux de l'arbre de $\mathcal{C}_u(\nu_h)$, dans le cas symétrique (resp., $\mathcal{C}_u^o(\nu_h)$ dans le cas orienté) fournira une nouvelle version de la détection des noeuds les plus "significatifs" de l'arbre des classifications.

5 Conclusion

Comme nous l'avons mentionné à la fin de notre introduction, la motivation première de ce travail a correspondu au souci de répondre à une question de Pascale Kuntz (École Polytechnique de l'Université de Nantes) sur l'adaptation adéquate de nos critères de détection des niveaux et noeuds les plus "significatifs" dans le cas des arbres de classification orientés représentant des hiérarchies implicatives. À cette question, déjà mentionnée dans l'introduction, nous avons répondu au paragraphe 4 ci-dessus. Cependant, il nous a fallu pour cela ou plutôt, nous avons senti la nécessité, de reprendre l'ensemble du formalisme d'une façon nouvelle et conséquente recouvrant tout autant les aspects axiomatiques ou définitionnels que constructifs et énumératifs d'une hiérarchie implicative. Ainsi, l'objet de ce travail est devenu beaucoup plus général. Les arbres organisant les liens implicatifs (orientés) sont binaires. Ils correspondent à un enrichissement structurel des arbres binaires

organisant des liens symétriques. C'est pour cette raison que nous avons établi soigneusement un parallèle entre les deux types de structure agrégative. Ainsi, nous avons commencé par reprendre la notion de hiérarchie de parties d'un ensemble fini E et, soucieux de nous limiter au degré de généralité nécessaire, on s'est restreint au caractère binaire de la hiérarchie de parties. À partir d'une telle définition nous avons établi un procédé de construction algorithmique d'un arbre ordinalement indicé de classification, compatible. Une énumération de tels arbres a été proposée.

Pour obtenir une définition axiomatique applicable dans le cas des hiérarchies implicatives, on passe dans [16] par la donnée d'une permutation totale sur l'ensemble E où la notion de partie est remplacée par celle de segment. Dans notre cas l'axiomatique est basée sur la notion de fourche (symétrique ou orientée) à composantes disjointes sur l'ensemble des parties de E . Ainsi, nous serrons au plus près la structure binaire de l'arbre. Le développement de l'axiomatique en termes de fourches orientées dans le cas implicatif, permet d'obtenir la permutation mentionnée ci-dessus comme une résultante. Il en résulte également un indice de niveau ordinal ultramétrique et une notion de triangle isocèle orienté dont la base est le plus petit des côtés, au sens de cet indice. De plus, la donnée de l'indice ultramétrique permet de bâtir l'arbre implicatif associé.

Les propriétés algorithmiques de la construction statistique de l'arbre implicatif pour l'indice même qui a été proposé pour son élaboration [11, 12] ont été étudiées au paragraphe 3 (monotonie, formule de réactualisation). La logique de cette construction suit le sens formel d'une hiérarchie implicative qui est développé au paragraphe 2.3. Nous y démontrons l'inadéquation de cette logique par rapport au formalisme de la théorie de la démonstration. Il en résulte une différence de nature dans l'interprétation des résultats. Nous allons chercher à l'illustrer en nous permettant d'emprunter les résultats de l'intéressante application relative à une enquête de l' "Association des Professeurs de Mathématiques de l'Enseignement Public" auprès de professeurs de mathématiques de classes terminales. L'objectif de l'enquête concerne leur évaluation pédagogique sur l'enseignement des mathématiques et ses objectifs [12]. Une des sections commençantes de l'arbre implicatif est définie par la structure (I) de la figure FIG. 5 où **a** (noté "N") définit "le caractère essentiel de l'objectif de l'acquisition de savoir faire" et **b** noté ("A") celui de l' "acquisition de connaissances". **c** (noté "OP6") correspond à l'attitude " préférer des programmes bien définis indiquant ce que je dois faire et ce que je ne dois pas faire". Nous interpréterons la structure (I) comme suit :

Considérer essentiel l'acquisition de savoir faire, implique de considérer essentiel l'acquisition de connaissances. D'autre part et à un degré moindre, considérer essentiel l'acquisition de savoir faire (resp., de connaissances), implique de préférer des programmes bien définis indiquant ce qu'il y a lieu de faire et ce qu'il y a lieu de ne pas faire.

Considérons maintenant un deuxième exemple où c'est la structure (II) de la figure FIG. 5 qui est représentée. Dans ce cas **a** (noté "OP2") correspond à l'attitude de "préférer au Bac un grand problème avec plusieurs parties plutôt qu'un ensemble de petits problèmes indépendants". **b** (noté "OP5") correspond à considérer que "la démonstration est la seule façon rigoureuse de faire des mathématiques" et **c** (noté "OP4") correspond à l'attitude "Quand je corrige, j'aime bien un barème très détaillé sur les résultats à obtenir". On a deux degrés d'implication. La plus forte est illustrée par $b \rightarrow c$; et la moins forte est illustrée deux fois par $(a \rightarrow b) \& (a \rightarrow c)$.

Références

- [1] R. AGRAWAL, T. IMIELIENSKY, and A. SWAMI. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216. AAAIPress, 1993.
- [2] J.P. BENZECRI. Analyse factorielle des proximités. *Publications de l'Institut de Statistique de l'Université de Paris*, (XIII et XIV), 1964-1965.
- [3] E. CHARNIAK. Bayesian networks without tears : making bayesian networks more accessible to the probabilistically unsophisticated. In Issue 4 Volume 12, editor, *AI Magazine archive*, pages 50–63. American Association for Artificial Intelligence, 1991.
- [4] F. DAUDÉ. *Analyse et justification de la notion de ressemblance entre variables qualitatives dans l'optique de la classification hiérarchique par AVL*. PhD thesis, Université de Rennes 1, juin 1992.

- [5] C. FLAMENT. Similarity analysis : a technique for research into social representations. In S. Moscovici, editor, *Social Representations*, 1979.
- [6] R. GRAS. *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*. PhD thesis, Doctorat d'État, Université de Rennes 1, octobre 1979.
- [7] R. GRAS. *L'implication statistique*. La pensée sauvage, 1996.
- [8] R. GRAS, R. COUTURIER, J. BLANCHARD, H. BRIAND, P. KUNTZ, and P. PETER. Quelques critères pour une mesure de qualité de règles d'association. In H. Briand, M. Sebag, R. Gras, and F. Guillet, editors, *Mesures de qualité pour la fouille des données*, pages 3–31. Cépaduès, 2004.
- [9] R. GRAS and P. KUNTZ. Significativité des niveaux d'une hiérarchie orientée en analyse statistique implicative. In M. Chavent and M. Langlais, editors, *Classification et fouille des données*, pages 39–50. Cépaduès, 2004.
- [10] R. GRAS and P. KUNTZ. Discovering r-rules with a directed hierarchy. *Soft Computing*, (5) :453–460, March 2006.
- [11] R. GRAS, P. KUNTZ, and H. BRIAND. Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille des données. *Mathématiques et Sciences Humaines*, (154-155) :9–29, 2001.
- [12] R. GRAS, P. KUNTZ, and H. BRIAND. Hiérarchie orientée de règles généralisées en analyse implicative. In M-S. Hacid, Y. Kodratoff, and D. Boulanger, editors, *Extraction et gestion des connaissances EGC 2003*, pages 145–158. Hermes, 2003.
- [13] R. GRAS and A. LARHER. L'implication statistique, une nouvelle méthode d'analyse des données. *Mathématiques et Sciences Humaines*, (120) :5–31, 1993.
- [14] Eds F. GUILLET and H. HAMILTON. *Quality Measures in Data Mining*. SPRINGER, 2006, in press.
- [15] M. JAMBU. *Classification automatique pour l'analyse des données, tome 1*. DUNOD, 1978.
- [16] P. KUNTZ. Classification hiérarchique orientée en ASI. In R. Gras, F. Spagnolo, and J. David, editors, *Troisièmes Rencontres Internationales - A.S.I. Analyse Statistique Implicative*, pages 53–62. Università degli Studi di Palermo, 2005.
- [17] S. LALLICH, P. LENCA, and B. VAILLANT. Variations autour de l'intensité d'implication. In R. Gras, F. Spagnolo, and J. David, editors, *Troisièmes Rencontres Internationales - A.S.I. Analyse Statistique Implicative*, pages 237–246. Università degli Studi di Palermo, 2005.
- [18] S. LALLICH and O. TEYTAUD. Évaluation et validation de l'intérêt des règles d'association. In H. Briand, M. Sebag, R. Gras, and F. Guillet, editors, *Mesures de qualité pour la fouille des données*, pages 193–217. Cépaduès, 2004.
- [19] G.N. LANCE and W.T. WILLIAMS. A general theory of classification sorting strategies : 1=hierarchical systems, 2=clustering systems. *Computer Journal*, (9-10) :373–380, 1967.
- [20] P. LENCA, B. VAILLANT, P. MEYER, and S. LALLICH. Association rule interestingness measures : experimental and theoretical studies. In F. Guillet and H. Hamilton, editors, *Quality Measures in Data Mining*, pages 51–76. Springer, 2006, in press.
- [21] I.C. LERMAN. *Les bases de la classification automatique*. Gauthier-Vilars, 1970.
- [22] I.C. LERMAN. *Classification et analyse ordinaire des données*. Dunod, 1981.
- [23] I.C. LERMAN. Sur la signification des classes issues d'une classification automatique. In J. Felsenstein, editor, *Numerical Taxonomy*, pages 179–198. Springer-Verlag, 1983.
- [24] I.C. LERMAN. Justification et validité statistique d'une échelle $[0,1]$ de fréquence mathématique pour une structure de proximité sur un ensemble de variables observées. *Publications de l'Institut de Statistique des Universités de Paris*, pages 27–57, 1984.
- [25] I.C. LERMAN. Construction d'un indice de similarité entre objets décrits par des variables d'un type quelconque. application au problème de consensus en classification. *Revue de Statistique Appliquée*, (XXXV(2)) :39–60, 1987.
- [26] I.C. LERMAN. Formules de réactualisation en cas d'agrégations multiples. *RAIRO, série RO*, (23) :151–163, 1989.
- [27] I.C. LERMAN. Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles I et II. *Mathématique Informatique et Sciences Humaines*, (118 et 119) :35–52 et 75–100, 1992.

-
- [28] I.C. LERMAN. Likelihood linkage analysis (LLA) classification method (around an example treated by hand). *Biochimie, Elsevier editions*, (75) :379–397, 1993.
- [29] I.C. LERMAN. Comparing classification tree structures : a special case of comparing q-ary relations. *RAIRO-Operations Research*, (33) :339–365, september 1999.
- [30] I.C. LERMAN and J. AZÉ. A new probabilistic measure of interestingness for association rules, based on the likelihood of the link. In F. Guillet and H. Hamilton, editors, *Quality Measures in Data Mining*, pages 207–236. Springer, 2006, in press.
- [31] I.C. LERMAN and N. GHAZZALI. What do we retain from a classification tree. In E. Diday and Y. Lechevallier, editors, *Symbolic-Numeric data analysis and learning*, pages 27–42. Nova Science, 1991.
- [32] I.C. LERMAN, R. GRAS, and H. ROSTAM. Élaboration et évaluation d'un indice d'implication pour des données binaires I et II. *Revue Mathématique et Sciences Humaines*, 74 and 75 :5–35, 5–47, 1981.
- [33] I.C. LERMAN and Ph. PETER. Indice probabiliste de vraisemblance du lien entre objets quelconques ; analyse comparative entre deux approches. *Revue de Statistique Appliquée*, (LI(1)) :5–35, Octobre 2003.
- [34] I.C. LERMAN and F. ROUXEL. Comparing classification tree structures : a special case of comparing q-ary relations II. *RAIRO-Operations Research*, (33) :251–281, july/september 2000.
- [35] N. MANTEL. Detection of disease clustering and a generalized regression approach. *Cancer Research*, (vol. 27, n° 2) :209–220, 1967.
- [36] J-P. NAKACHE and J. CONFAIS. *Approche pragmatique de la classification*. TECHNIP, 2005.
- [37] P. PETER, H. LEREDDE, and I.C. LERMAN. Notice du programme CHAVLH (Classification Hiérarchique par Analyse de la Vraisemblance des Liens en cas de variables Hétérogènes). Dépôt APP (Agence pour la Protection des Programmes) IDN.FR.001.240016.000.S.P.2006.000.20700, Université de Rennes 1, Décembre 2005.
- [38] P-N. TAN, V. KUMAR, and J. SRIVASTAVA. Selecting the right interestingness measure for association patterns. In *Proceedings of the 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2002.