BIBLIOGRAPHIE

[1] J.C. BEZDEK et J.C. HARRIS
Fuzzy partitions and relations ; an axcomatic basis for clustering.
Fuzzy sets and systems (1978) 111 - 127.

[2] M. DEFAYS
Ultramétriques et relations floues.
Institut de psychologie des Sciences de l'éducation.
SART TILMAN B 4000 LIEGE I Belgique.

[3] M. GONDRAND et M. MINOUX
Graphes et algorithmes. Eyrolles, Paris,1979.

[4] F. CAILLIEZ et J. P. PAGES
Introduction à l'analyse des données. SMASH, Paris, 1976.

[5] M. RICHETIN
Analyse structurale des systèmes complexes.
Thèse d'Etat, Université Paul Sabatier, juillet 1975.

Data Analysis and Informatics
E. Diday et al. (eds.)
© North-Holland Publishing Company, 1980

INFORMATION PREORDINALE ET ANALYSE DES PREFERENCES

H. Abdi , J.P. Barthelemy , X. Luong

In this paper we use some concepts derived from information theory in order to construct metrics on spaces of preferences.

Ce travail porte sur l'utilisation de notions issues de la théorie de l'information pour construire des distances entre des données de préférences.

INTRODUCTION

La démarche qui consiste à utiliser des concepts issus de la théorie de l'information en analyse des données est classique. Cependant, elle n'est développée que dans les cas suivants :

1) Classification (Boyd [4] , Arabie-Boormann [1])
2) Tableaux présence-absence (Comyn [6]).

Dans ce travail, nous examinons le cas de données de préférences.

Les travaux de Kampé de Feriet [9] et Losfeld [10] montrent que les deux points évoqués ci-dessus relèvent de l'information sur un ensemble ordonné E. Comyn - Losfeld [7] proposent alors une "distance en information" qui n'est -en fait- qu'un cheminement minimal dans le graphe (valué) non orienté de couverture de E ; le calcul de cette distance est -en général- malaisé. Comyn-Van Dorpe [8] (dans le cas latticiel) et Barthelemy [2] (dans le cas général) ont indiqué une condition portant sur la mesure d'information pour que ce calcul puisse s'effectuer à l'aide d'une "formule". Une difficulté surgit lorsqu'on cherche à appliquer ceci à des données de préférences ; les conditions évoquées ci-dessus ne sont pas -en général- vérifiées pour les préordres totaux (cette question est étudiée en détail dans [3]) et il faut enrichir ces derniers par des matchs. En d'autres termes pour "cheminer" entre deux préordres totaux on devra passer par des intermédiaires non nécessairement transitifs.

Tenant compte de cette remarque, nous construisons, dans ce texte, un certain

Laboratoire de Psychologie, Faculté des Lettres, 25030 Besançon Cedex
E N S M M, 25030 Besançon Cedex
Laboratoire de Statistique, Faculté des Lettres - 25030 Besançon Cedex

nombre de distances, issues de la théorie de l'information, entre des données relationnelles et évoquons une application à la psychologie sociale.

RELATIONS BINAIRES - INFORMATION - DISTANCES

Soit X un ensemble fini et soit p une mesure de probabilité strictement positive et partout définie (toute partie non vide de X est un évènement de probabilité non nulle). Une "préférence" sur X, obtenue par des comparaisons par paires, sera représentée par une relation binaire r $\subset$ X x X. La section $r(x) = \{y; (x,y) \in r\}$ est l'évènement "être préféré à x". Si I est une mesure d'information sur X. (I est une application de l'ensemble des parties de X dans $\mathbb{R}^+$ telle que $I(A) < I(B)$ lorsque $B \subset A$), on définit l'information locale, en x, pour r, par :

$$I(r ; x) = I(r(x))$$

(quantité d'information apportée par l'évènement "être préféré à x". L'information moyenne de r est l'espérance $\mathcal{I}(r)$ de la variable aléatoire $I(r ; x)$.

$$\mathcal{I}(r) = \sum_{x \in X} I(r ; x)\, p_x = \sum_{A \leftarrow r} I(r(A))\, p(A).$$

Dans cette formule $A \leftarrow r$ signifie que A est une classe d'équivalence pour : $x \equiv y$ si et seulement si $r(x) = r(y)$, r(A) est alors la section par un élément quelconque de A.

Considérons maintenant une distance d sur l'ensemble des parties de X et définissons la dissemblance locale en x entre deux relations binaires r et s par :

$$d(r,s ; x) = d(r(x), s(x)).$$

Un calcul facile montre que l'espérance de d(r,s ; x) est encore une distance (distance moyenne entre r et s) :

$$D(r,s) = \sum_{x \in X} d(r,s ; x)\, p_x = \sum_{A \leftarrow r, B \leftarrow s} p(A \cap B)\, d(r(A), s(B))$$

Un cas particulièrement intéressant est celui où d est définie par la mesure d'information I. Si I vérifie :

(H^-) $I(A) + I(B) \geqslant I(A \cap B) + I(A \cup B)$,

$d^-(A,B) = 2\, I(A \cap B) - I(A) - I(B) = I(A/B) + I(B/A)$ est une distance. Dualement, si I vérifie :

(H^+) $I(A) + I(B) \leqslant I(A \cap B) + I(A \cup B)$,

$d^+(A,B) = I(A) + I(B) - 2\, I(A \cup B)$ est une distance.

La dissemblance locale correspondante vaut -par exemple sous l'hypothèse (H^+)- :

$d^+(r,s ; x) = I(r ; x) - I(r \cup s ; x) + I(s ; x) - I(r \cup s ; x)$, la quantité

$I(r ; x) - I(r \cup s ; x)$ évalue la perte d'information quand on passe de l'évènement "être préféré à x par r" à l'évènement "être préféré à x par r ou par s". De plus, les hypothèses (H^-) et (H^+) s'étendent aux relations binaires et si I vérifie (H^-) resp.(H^+), il en sera de même pour l'information moyenne $\mathcal{I}$; la distance définie par $\mathcal{I}$ valant, toujours en prenant l'exemple de l'hypothèse

(H^+)- :

$$D^+(r,s) = \sum_{A \leftarrow r,\ B \leftarrow s} d^+(r(A), r(B))\, p(A \cap B) = \mathcal{I}(r) + \mathcal{I}(s) - 2\mathcal{I}(r \cup s).$$

INFORMATION ET DISTANCES DE GRAPHES

Considérons un ensemble R de relations binaires. Soit G(R) le graphe dont les sommets sont les éléments de R u_{rs} étant un arc si et seulement si s couvre r ($r \subset s$ et il n'existe pas de relation $t \in R$ telle que $r \subset t \subset s$). Soit une mesure d'information sur R ordonné par inclusion ([10]). La perte d'information sur l'arc u_{rs} vaut $\mathcal{I}(r) - \mathcal{I}(s)$.
Si $c : u_1 \ldots u_n$ est une chaîne de G(R) reliant les sommets $r_1 \ldots r_n$, la variation de l'information le long de c vaut :

$\tilde{\mathcal{I}}(c) = \sum_{i=1}^{n} |\mathcal{I}(r_i) - \mathcal{I}(r_{i-1})|$. C(r,s) désignant l'ensemble des chaînes entre r et s, lorsque G(R) est connexe, la quantité

$$\Delta_R(r,s) = \min_{c \in C(r,s)} \tilde{\mathcal{I}}(c)$$

est une distance.

Dans [8] et [2] sont indiquées des conditions pour calculer Δ_R à l'aide d'une "formule" (donc sans algorithme de cheminement minimal) et -dans ces cas- Δ_R coïncide avec l'une des distances D^- ou D^+. Pratiquement, ces conditions sont les hypothèses (H^-) ou (H^+) -ou leur généralisation lorsque R n'est pas stable par intersection ou par réunion-. Ainsi si l'on part de la mesure d'information $I = -c \operatorname{Log} \frac{1}{p}$ (l'entropie) ou $I = 1 - \frac{1}{p}$ (l'information hyperbolique), on trouve pour deux préordres totaux r et s : $D^+(r,s) = \Delta_M(r,s)$, M désignant l'ensemble des matchs (relations totales et réflexives) de X.

Pour $I = 1-p$, $D^+(r,s) = D^-(r,s) = \Delta_P(r,s) = \Delta_M(r,s)$, P désignant l'ensemble des préordres de X. D'une manière générale on peut vérifier qu'il n'existe aucune information moyenne telle que $\Delta_{P'} = D^+$ ou $\Delta_{P'} = D^-$ (P' désignant l'ensemble des préordres totaux de X). On sait cependant construire ([3]) d'autres types d'information vérifiant la première condition (c'est alors la borne supérieure de deux préordres totaux qui interviendra dans D^+).

LE PROGRAMME PREØRDRE

Ce programme calcule des distances "moyennes" sur une ensemble de préordres totaux de X. En particulier il résout des problèmes de "cheminement" minimal (au sens des chaînes) sur le graphe -valué- G(M). Sa conception est simple : après lecture des paramètres et données on génère les sections s(A) (sous programme) puis on effectue des "opérations ensemblistes" du type $|A \cap B|$, $|s(A) \cup s(B)|$,... ce qui permet de calculer la distance que l'on désire.

COMPARAISON DE PREORDRES TOTAUX ISSUS D'UN QUESTIONNAIRE D'OPINION

Remarquons tout d'abord que, lorsqu'on étudie des préordres totaux admettant "peu" de classes (par exemple issus de notes de 0 à 5 attribuées à "beaucoup" d'individus), c'est l'information locale qui recevra une interprétation pertinente. L'information moyenne (qui -dans le cas d'une distribution uniforme

de probabilités- est maximale, parmi les préordres totaux, sur les ordres totaux) est trop limitée par les "conditions de l'expérience" (six classes, au plus). Cette information ne sera qu'un intermédiaire pour le calcul d'une distance. En revanche, lorsqu'aucune contrainte n'est imposée quand au nombre de classes, l'information moyenne pourra être comprise comme la "quantité d'indifférence" contenue dans la préférence.

Evoquons rapidement une application "concrète". Il s'agit à partir d'un "questionnaire" de contribuer à l'étude de la fonction des injures (c.f. Maget[11], Chastaing [5]). Ce questionnaire est constitué d'un certain nombre d'opinions sur la fonction des injures que les sujets enquêtés doivent noter de 0 à 6. Le problème posé est d'établir une typologie sur ces opinions en ne tenant compte que des relations de préférence qu'elles induisent sur les enquêtés (un enquêté i est préféré à j par une opinion O si i note O au moins aussi bien que j). Pour ce faire on calcule (à l'aide du programm:e PREØRDRE) une distance moyenne entre ces préordres totaux, donc entre les opinions.

La matrice de distances obtenue a mis en évidence l'existence de deux pôles d'attraction : il s'agit de deux opinions très proches de toutes les autres (et d'ailleurs fort proches l'une de l'autre). Ces deux pôles supprimés, le même phénomène se produit et ainsi de suite, mettant en évidence l'existence d'une sorte de "noyau central" et montrant qu'une tentative de classification serait vaine. l'étude des opinions "éloignées" conduit à des résultats encore plus pertinents : les candidats à l'originalité se dégagent et on constate qu'on retrouve -en fait- des "têtes de chapitre". Ces quelques remarques montrent que la méthode utilisée a -essentiellement- permis une réflexion critique sur la construction même du questionnaire. Indiquons enfin qu'un certain nombre de faits mis en évidence par cette analyse (noyau central, têtes de chapitres) se sont trouvés difficilement décelables par l'A.F.C. que nous avons avons effectué parallèlement.

References

1 P. Arabie, S.A. Boorman : Structural measures and the method of sorting, in Multidimensional scaling, Seminar Press, New-York, 1972

2 J.P. Barthélemy : Remarques sur les propriétés métriques des ensembles ordonnés, Math. Sci. Hum., n° 61, 1978, p.39-60

3 J.P. Barthélemy : Propriétés métriques des ensembles ordonnés. Comparaison et agrégation des relations binaires. Thèse,1979

4 J.P. Boyd : Information distance for discrete structures, in Multidimensional scaling, Seminar Press, New-York, 1972

5 Chastaing : La psychologie des jurons, Journal Psycho. Normale et pathologique, 2, 1976

6 G. Comyn : Applications de l'information généralisée à l'analyse des données, Colloques IRIA, 1977 (premières journées analyse des données et informatique).

7 G. Comyn, J. Losfeld : Information et préordres, Journées Lyonnaise questionnaires, 1975, Public. du Labo. Structure de l'information, n° 1, p. 45-70

8 G. Comyn, I. Cl. Van-Dorpe : Valuation et semi-modularité dans les demi-treillis, Math. Sci. Hum., n° 56, 1976, p.63-73

9 J. Kampé de Feriet : La théorie généralisée de l'information et la mesure subjective de l'information, Lect. Notes in Math, 398, Springer, 1974

10 J. Losfeld : Information généralisée et relation d'ordre, Lect. Notes in Math, 398, Springer, 1974

11 M. Maget : Fonctions de la forme technique et architecture, 1946

OTHER TOPICS