MATHÉMATIQUES ET SCIENCES HUMAINES

CLAUDE FLAMENT

Le traitement des ex æquo en analyse de similitude : la réunion des arbres maximaux ou RAM

Mathématiques et sciences humaines, tome 114 (1991), p. 35-40
<http://www.numdam.org/item?id=MSH_1991__114__35_0>

Math. Inf. Sci. hum. (29e année, n° 114, *1991*, p.35-40)

LE TRAITEMENT DES EX ÆQUO EN ANALYSE DE SIMILITUDE : LA RÉUNION DES ARBRES MAXIMAUX OU RAM

Claude FLAMENT [1]

RÉSUMÉ — *On considère un graphe complet dont les arêtes sont totalement préordonnées. En analyse de similitude, plutôt que de procéder à un ordonnancement des arêtes ex æquo par une méthode lexicographique sur leurs intitulés, l'auteur propose de rechercher la réunion des arbres maximaux (RAM).*

ABSTRACT — Tied Edges in Similarity Analysis : The Union of the Maximum Spanning Trees.
The complete graph endowed with a complete preorder on its edges is considered. In similarity analysis, one often researches all the maximum spanning trees (MSTs) by using a lexicographic method on the labels of the tied edges. Instead of that, the author suggests to directly determine the union of the MSTs (the RAM in the text).

L'analyse de similitude, telle qu'elle est classiquement pratiquée (cf. DEGENNE et VERGES, 1973), recherche l'*arbre maximum de similitude* - ce qui suppose que la similitude constitue un ordre *strict* sur les arêtes. C'est rarement le cas dans les applications à des données réelles. La plupart des algorithmes transforment le *préordre* de similitude en ordre strict par ordonnancement des ex æquo par une méthode lexicographique sur les intitulés des arêtes, ce qui est toujours artificiel, et parfois artéfactuel (les intitulés pouvant être liés à la signification des variables).

Rappelons un algorithme de recherche de l'arbre maximum (lorsque l'ordre est strict). Un *cocycle* d'un graphe est l'ensemble des arêtes associées à une bipartition des variables, c'est-à-dire, l'ensemble des arêtes allant d'une partie de la bipartition à l'autre. Un résultat de Rosenstiehl (1967) montre que l'arbre maximum n'est autre que l'ensemble des arêtes maximums dans au moins un cocycle. D'où la méthode dite des *plus proches voisins* : on considère les cocycles associés à chaque variable séparément ; l'arête maximum se lit comme la valeur maximum dans la ligne du tableau de similitude.

[1] C.A.M.S., Vieille Charité - MARSEILLE.

Exemple :

	A	B	C	D	E	F
A	\	. 53	. 44	. 64	. 45	. 82
B	. 53	\	. 62	. 78	. 61	. 37
C	. 44	. 62	\	. 45	. 75	. 72
D	. 64	. 78	. 45	\	. 65	. 69
E	. 45	. 61	. 75	. 65	\	. 55
F	. 82	. 37	. 72	. 69	. 55	\

Tableau 1

On obtient alors le graphe des plus proches voisins, qui, ici, n'est pas convexe (figure 1).

A ——— F   B ——— D   C ——— E

Figure 1

On va alors définir de nouveaux cocycles en tenant compte des composants connexes ; par exemple {A,F} contre l'ensemble des autre composantes {B,C,D,E} :

	B	C	D	E
A	. 53	. 44	. 64	. 45
F	. 37	. 72	. 69	. 55

Tableau 2

d'où :

A ——— F ——— C ——— E   B ——— D

Figure 2

On recommence jusqu'à ce qu'on obtienne l'arbre maximum :

	A	C	E	F
B	. 53	. 62	. 61	. 37
D	. 54	. 45	. 65	. 69

Tableau 3

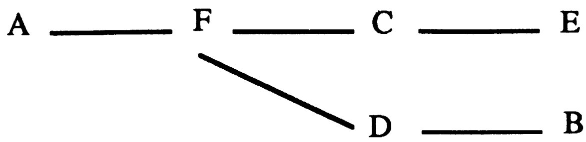

Figure 3

Remarque : La similitude du tableau 1 n'est pas un ordre strict, puisque deux arêtes ont pour valeur . 45. Mais ce couple d'ex æquo n'intervient jamais dans les manipulations ultérieures.

Il n'en va pas toujours ainsi, et alors il n'existe pas d'arbre maximum, malgré une pratique abusive.

Exemple :

	A	B	C	D	E
A	\	. 65	. 62	. 37	. 80
B	. 65	\	. 77	. 48	. 60
C	. 62	. 77	\	. 65	. 41
D	. 37	. 48	. 65	\	. 72
E	. 80	. 60	. 41	. 72	\

Tableau 4

A ——— E ——— D   B ——— C

Figure 4

	A	D	E
B	. 65	. 48	. 60
C	. 62	. 65	. 41

Tableau 5

Au tableau 5, on a deux arêtes maximales ; on peut privilégier (AB) au titre de l'ordre lexicographique, mais c'est arbitraire ; et si on retient les deux arêtes maximales, ce qu'on obtient n'est pas un arbre (figure 5).

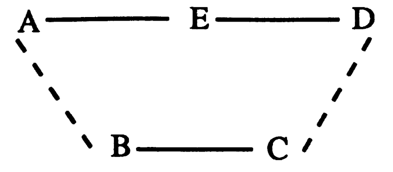

Figure 5

De quoi s'agit-il ? Reprenons le préordre de similitude, affiné par l'ordre lexicographique ; pour cet ordre strict, la figure 6 représente l'arbre maximum :

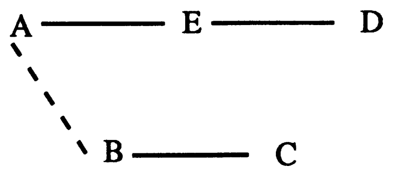

Figure 6

Mais on peut tout aussi bien prendre l'ordre lexicographique *inverse*, et l'arbre maximum est alors celui de la figure 7 :

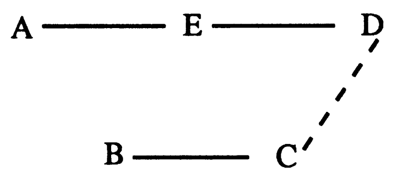

Figure 7

En fait, ces deux arbres (figures 6 et 7) sont équivalents par rapport au préordre de similitude ; ce sont deux *arbres maximaux*, et la figure 5 présente leur réunion : *la réunion des arbres maximaux*, ou *RAM* (cf. FLAMENT, 1975).

Nous proposons de modifier l'analyse de similitude habituelle, qui recherche un arbre maximum qui n'existe peut-être pas, en procédant à la recherche systématique de la RAM (qui se réduit bien évidemment à l'arbre maximum, s'il existe).

Des précautions sont à prendre pour l'algorithme de recherche de la RAM.

Exemple :

	A	B	C	D	E	F
A	\	. 60	. 30	. 42	. 75	. 57
B	. 60	\	. 72	. 49	. 35	. 50
C	. 30	. 72	\	. 70	. 59	. 58
D	. 42	. 49	. 70	\	. 60	. 80
E	. 75	. 35	. 59	. 60	\	. 37
F	. 57	. 50	. 58	. 80	. 37	\

Tableau 6

Les plus proches voisins sont (figure 8)

A ——— E   B ——— C   D ——— F

Figure 8

On va opposer la première composante connexe {A,E} à l'ensemble des autres {B,C,D,E} :

	B	C	D	F
A	. 60	. 30	. 42	. 57
E	. 35	. 59	. 60	. 37

Tableau 7

Le tableau 7 fait apparaître deux arêtes maximales, d'où l'arbre de la figure 9 :

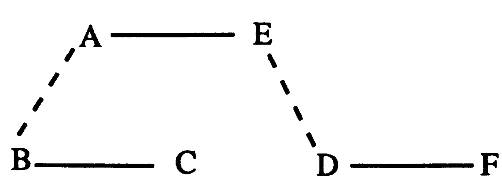

Figure 9

Mais cet arbre n'est ni maximum ni maximal ; en effet, l'examen du tableau 6 montre l'existence d'une arête (C,D) à . 70, supérieure aux arêtes maximales du tableau 7.

Cette arête serait apparue si, au lieu d'opposer la première composante (première lexicographiquement) au reste des variables, on avait par exemple opposé la composante (B,C) au reste :

	A	D	E	F
B	.60	.49	.35	.50
C	.30	.70	.59	.58

Tableau 8

D'où, le graphe non connexe de la figure 10 :

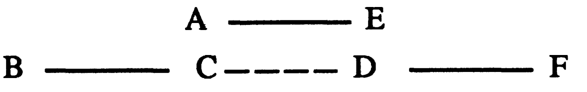

Figure 10

En opposant {A,E} au reste, comme dans le tableau 7 et la figure 9, on obtient un graphe qui n'est pas un arbre (figure 11), mais qui est la RAM de notre exemple.

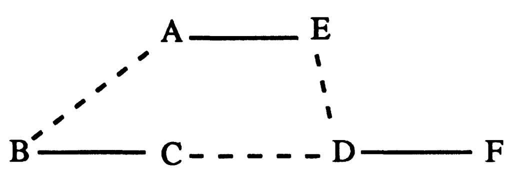

Figure 11

Pour être sûr qu'on a rien oublié, on opposera la composante {D,F,} au reste.

L'algorithme est alors simple :
Première phase : les plus proches voisins
Deuxième phase (si nécessaire), on considére *tous* les cocycles définis par l'opposition de chaque composante connexe avec le reste des variables, et à chaque fois on retient la ou les arêtes maximales. (FLAMENT, 1975).

Il convient donc de ne plus parler d'arbre maximum de similitude, mais de la RAM de similitude.

BIBLIOGRAPHIE

DEGENNE, A, VERGES, P., 1973, "Introduction à l'analyse de similitude", *Revue Française de Sociologie, 14*, 1973, 471-512.
FLAMENT, C., 1975, "Arêtes maximales des cocycles d'un graphe préordonné", *Mathématiques et Sciences humaines, 51*, 1975, 5-12.
ROSENSTIEHL, P, 1967, "L'arbre minimum d'un graphe", *in* P. Rosenstiehl (Ed.), *Théorie des graphes,* Rome, I.C.C., Paris, Dunod, 357-368.