SUR UNE Q-TECHNIQUE A DEUX PHASES BASEE SUR LA NOTION DE S-COMPARAISON .

ABDELWAHEB REBAI
Département des Méthodes Quantitatives
Fac. des Sciences Economiques et de Gestion
B.P n° 69 SFAX 3028, TUNISIE

ABSTRACT : The concepts of (α, β)-likes of a prototype X and (α, β)-bunch of objects introduced, here, as specific subsets of the 1-cut of a so-called S-comparison relation which is a very specific kind of tolerance relations , are bottomed on the distinction between fuzzy sets similarity and nondissimilarity pointed out in [5] . These two-threshold based concepts are, to some extent, the respective analogues of the concepts of r-sphere of center X and radius r and r-clique found, for instance, in [3] . But unlike the latter two concepts, the former two ones are not based on T-distances associated with *-relations which possess some well-known metrical properties. In this paper, we introduce the basic concepts and present some immediate results related to the simultaneous usage of fuzzy sets similarity and nondissimilarity scalar indices in cluster analysis, in addition to a two-phase, bunch-based Q-technique which is , in our opinion, a far cry from conventional clustering techniques .

MOTS CLES : Agglomérat, Coalescence, ξ-coupe, Indice de Similarité, Indice de Non-Dissimilarité, Opérateur de maximalisation, Partition Floue, Q-technique , Quasi-Partition .

1- INTRODUCTION

Etant donnés un univers d'objets-acteurs (fini ou infini) $\mathbf{U}$, une partie finie Ω de $\mathbf{U}$, un ensemble $\mathfrak{D}$ de k descripteurs co-classifiants [2] , il est possible d'associer à $\mathbf{U}$ une texture au moyen de fonctions standardisantes.[1] Par texture associée à $\mathbf{U}$, il faut entendre un ensemble de k applications $A_1, A_2, \dots, A_k$ de $\mathbf{U}$ dans l'intervalle [0,1] [7], définissant un profil flou pour chaque objet-acteur. Désormais $\underline{X} = \sum_{d \in \mathfrak{D}} \mu_{\underline{X}}(d)/d$, désigne le profil flou de X, correspondant à une texture appropriée . Des indices scalaires de similarité entre objets-acteurs peuvent être définis en utilisant les indices scalaires de similarité entre ensembles flous . Or, étant donné un indice scalaire, S, il est possible de construire, entre autres, un indice $\mathfrak{S}$ de non-dissimilarité entre X et Y , en posant :

$\mathfrak{S}(\underline{X}, \underline{Y}) = S(\overline{\underline{X}}, \overline{\underline{Y}})$ où $\overline{\underline{X}}$ et $\overline{\underline{Y}}$ sont les complémentaires flous respectifs de $\underline{X}$ et $\underline{Y}$. Les indices S et $\mathfrak{S}$ ne fournissent pas, en

---

[1] Des exemples intéressants de fonctions standardisantes peuvent être trouvés dans Zeleny [9]

général, la même information [5] , ce qui rend leur usage simulané plus significatif pour l'analyse typologique .

Dans cet article, une Q-technique qui tient compte des différentes remarques suivantes, est introduite :
1°) la similarité et la non-dissimilarité entre ensembles flous ne sont pas, en général, des synonymes ;
2°) L'homogénéité des agglomérats à former est déterminée au moyen de deux seuils donnés à l'avance, un seuil pour la similarité et un seuil pour la non-dissimilarité;
3°) cette Q-technique va permettre de construire des agglomérats non nécessairement disjoints et d'induire des agglomérats flous ;
4°) le nombre d'agglomérats à former n'est pas fixé a priori, mais au contraire il est déterminé par la Q-technique.

2- DEFINITIONS ET RESULTATS

2.1- NOTION DE S-COMPARAISON

Soient, S, un indice scalaire de similarité d'ensembles flous, et $\mathcal{N} = (T, \tau_1, \tau_2)$ un triplet d'opérateurs où T est une t-norme définie sur l'intervalle [0,1] [6] et τ_1, τ_2 les deux opérateurs de maximalisation associés, respectivement à deux t-normes données T_1 et T_2 [4], alors la relation, Σ, définie pour tout couple $\eta = (\alpha, \beta)$ de [0,1] x [0,1], par l'équation :

$$(\forall \quad X, Y \in \mathcal{U}) \ : \ \mu_\Sigma(X,Y) = T(\ S(\underline{X},\underline{Y})\ \tau_1\ \alpha, \widetilde{S}(\underline{X},\underline{Y})\ \tau_2 \beta) \quad (1)$$

est une relation binaire floue reflexive, symètrique et intransitive qui **synthétise** et véhicule une information sur la **satisfaction** des seuils α et β, respectivement, par la **similarité** et la non-dissimilarité des objets-acteurs X et Y.

DEFINITION 2.1.1 La relation, Σ, définie par l'équation (1) sera dite relation de **S-comparaison** sur $\mathcal{U}$ et $\mathcal{K} = \{ \eta, S, \mathcal{N}, Tx \}$ sera son **ensemble clé** .

η sera le couple des niveaux de la S-comparaison, S l'indice scalaire de similarité, $\mathcal{N}$ le triplet d'opérateurs de synthèse et Tx la matrice des profils relative à l'ensemble des descripteurs co-classifiants $\mathcal{D}$.

Exemple 2.1.1 : Si $T(x, y) = T_W(x, y)$ (drastic product) , $T_1(x, y) = T_2(x, y) = x \wedge y$ alors, Σ sera définie par :

$$\mu_{\Sigma}(X,Y) = \begin{cases} 1 \text{ si } S(\underline{X,Y}) \geq \alpha \text{ et } \tilde{S}(\underline{X,Y}) \geq \beta \\ \tilde{S}(\underline{X,Y}) \text{ si } S(\underline{X,Y}) \geq \alpha \text{ et } \tilde{S}(\underline{X,Y}) < \beta \\ S(\underline{X,Y}) \text{ si } S(\underline{X,Y}) < \alpha \text{ et } \tilde{S}(\underline{X,Y}) \geq \beta \\ 0 \quad \text{sinon} \end{cases} \qquad (2)$$

RESULTAT 2.1.1:

$$(\forall X,Y \in U) : \mu_{\Sigma}(X,Y) = 1 \Leftrightarrow S(\underline{X,Y}) \geq \alpha \text{ et } \tilde{S}(\underline{X,Y}) \geq \beta$$

CONSEQUENCE : La 1-coupe des relations de S-comparaison est la même indépendamment, du triplet d'opérateurs de synthèse choisi pour l'ensemble clé .

2.2- (α,β)-S-« LIKES » et (α,β)-S-« BUNCH »

DEFINITION 2.2.1 : Soit $X \in \Omega$, on définit le **(α,β)-S-« Likes » du prototype** X dans Ω, par : $L(X) = \{ Y \in \Omega \ / \ \mu_{\Sigma}(X,Y) = 1 \}$

DEFINITION 2.2.2 : Soit $\mathfrak{B} \subseteq \Omega$, si $(\forall \ X,Y \in \mathfrak{B}) : \mu_{\Sigma}(X,Y) = 1$, alors $\mathfrak{B}$ sera dit **(α,β)-S-« bunch »** de Ω .

DEFINITION 2.2.3 : Un (α,β)-S-« bunch » $\mathfrak{B}$ sera dit **maximal** dans Ω, s'il n'existe pas un (α, β)-S-« bunch » $\mathfrak{B}' \subseteq \Omega$ tel que $\mathfrak{B} \subset \mathfrak{B}'$.

Soit $X \in \Omega$, intéressons nous, maintenant, à l'extraction d'un « bunch » maximal de plus grand cardinal à partir de $L(X)$ dit P.G.B (plus grand bunch) de $L(X)$.

2.2.4-PROCEDURE D'EXTRACTION D'UN P.G.B

Les différentes étapes de la procédure d'extraction se présentent, alors, comme suit :

ETAPE 1 - Construire la matrice d'une S-comparaison des éléments de $L(X)$ (indiquer les positions des "1" uniquement) . Aller à l'étape 2.

ETAPE 2 - Si cette matrice est pleine, $L(X)$ constitue un « bunch » sinon identifier l'objet-acteur dont la ligne (ou colonne) comporte le plus petit nombre de "1" et barrer sa ligne et sa colonne . Aller à l'étape 3;

ETAPE 3 - Reprendre l'étape 2 pour la nouvelle matrice obtenue à l'étape 2 ; puis aller à l'étape 4;

ETAPE 4 - Répéter la démarche précédente jusqu'à l'obtention d'une matrice carrée pleine . Les objets-acteurs qui n'ont pas été éliminés forment, alors, un « bunch » .

RESULTAT 2.2.1 Soit $X \in \Omega$, alors, Le « bunch » formé à partir de $\mathcal{L}(X)$ par la procédure d'extraction est un P.G.B de $\mathcal{L}(X)$.

RESULTAT 2.2.2 Soit $X \in \Omega$, si $\mathfrak{B}$ est un P.G.B de $\mathcal{L}(X)$, alors, $\mathfrak{B}$ est maximal dans Ω .

3- UNE Q-TECHNIQUE A DEUX PHASES ($\mathcal{L}$ n' $\mathfrak{B}$ technique)

DEFINITION 3.1 : On définit le **(α,β)-S-Scatter** de Ω par :

$$\mathrm{Scatt}(\Omega) = \{ X \in \Omega \,/\, \mathcal{L}(X) = \{X\} \}.$$

Un objet-acteur X appartient à Scatt(Ω) si et seulement si l'ensemble $\mathcal{L}(X)$ de ses semblables dans Ω est **trivial** (i.e. un singleton) et donc donne lieu à un « bunch » trivial. L'ensemble $\Omega^* = \Omega - \mathrm{Scatt}(\Omega)$ sera le support de la coalescence (bunching support) i.e. l'ensemble des objets-acteurs qui vont constituer les « bunches » non triviaux.

REMARQUE 3.1: Des « bunches » non triviaux existent si et seulement si $\mathrm{Scatt}(\Omega) \neq \Omega$.

La Q-technique qui va suivre, consiste à induire une **partition floue** [8] à partir d'une **Quasi-partition** de Ω^* [1] .

3.2- LES DEUX PHASES DE LA Q-TECHNIQUE

Les deux phases de cette Q-technique, se présentent comme suit :

PHASE 1: Obtenir une Quasi-partition $\{\mathfrak{B}_1, \mathfrak{B}_2, ..., \mathfrak{B}_c\}$ de Ω^* par des « bunches » maximaux , de la façon suivante :

ETAPE 1 : Construire la matrice d'une S-comparaison des objets-acteurs de Ω^*. Si cette matrice est pleine, Ω^* forme un P.G.B, sinon, aller à l'étape 2,

ETAPE 2 : Extraire un P.G.B à partir du « Likes » du prototype ayant le moins de semblables dans Ω^* . Soit $\mathfrak{B}_1$ ce « bunch » . Aller à l'étape 3;

ETAPE 3 : Extraire un P.G.B à partir du « Likes » du prototype ayant le plus de semblables dans Ω^* ; Soit $\mathfrak{B}_2$ ce « bunch » . Si $\mathfrak{B}_1 \cup \mathfrak{B}_2$ couvre Ω^* , STOP , sinon aller à l'étape 4.

ETAPE 4 : Si l'ensemble de travail $\mathcal{W} = \Omega^* - \mathfrak{B}_1 \cup \mathfrak{B}_2$ est réduit à un seul objet-acteur, extraire un P.G.B à partir de son « Likes » , sinon extraire un P.G.B à partir du « Likes » du prototype appartenant à $\mathcal{W}$ et ayant le moins de semblables dans Ω^* . Soit $\mathfrak{B}_3$ ce « bunch » . Si $\mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{B}_3$ couvre Ω^* , STOP , sinon aller à l'étape 5

ETAPE 5 : Si l'ensemble de travail $\mathfrak{W} = \Omega' - \mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{B}_3$ est réduit à un seul objet-acteur, extraire un P.G.B à partir de son « Likes », sinon extraire un P.G.B à partir du « Likes » du prototype appartenant à $\mathfrak{W}$ et ayant le plus de semblables dans Ω'. Soit $\mathfrak{B}_4$ ce « bunch ».

Si $\mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{B}_3 \cup \mathfrak{B}_4$ couvre Ω', STOP , sinon aller à l'étape 6 .

ETAPE 6 : Continuer avec les étapes 4 et 5 , en actualisant , chaque fois , l'ensemble de travail , $\mathfrak{W}$, jusqu'au recouvrement de Ω' .

PHASE 2 : Obtenir une partition floue $\{ \mathfrak{F}_1, \mathfrak{F}_2, \ldots, \mathfrak{F}_c \}$ de Ω' en sorte que La fonction d'appartenance de chaque $\mathfrak{F}_j$ pour $j=1,2,\ldots,c$, soit déterminée à partir des $\mathfrak{B}_j$ en posant, pour tout objet-acteur X de Ω' et pour $j=1,2,\ldots,c$, :

$$\mu_{\mathfrak{F}_j}(X) = \begin{cases} 0 & \text{si } X \notin \mathfrak{B}_j \\ \dfrac{\sum_{Y \in \mathfrak{B}_j} \{ S(\underline{X,Y}) + \tilde{S}(\underline{X,Y}) \}}{\sum_{\mathfrak{B}_r \in \mathfrak{B}(X)} \sum_{Z \in \mathfrak{B}_r} \{ S(\underline{X,Z}) + \tilde{S}(\underline{X,Y}) \}} & \text{si } X \in \mathfrak{B}_j \end{cases} \qquad (3)$$

où $\mathfrak{B}(X)$ est l'ensemble des « bunches » contenant X. L'avantage de ces deux phases est qu'elles nous assurent que pour tout $j=1,2,\ldots,c$ et pour tout ξ de $[0,1]$, la ξ-coupe de $\mathfrak{F}_j$ est un (α, β) -« bunch »

<u>RESULTAT 3.2</u> : Soient, Ω, un ensemble d'objets-acteurs, et $\mathcal{K} = \{ (\alpha_0, \beta_0), S, \mathcal{N}, Tx \}$, l'ensemble clé d'une $\mathcal{S}$-comparaison sur Ω , alors il existe deux seuils $\alpha_1(\alpha_0, \beta_0, S, Tx) \geq \alpha_0$ et $\beta_1(\alpha_0, \beta_0, S, Tx) \geq \beta_0$ tels que la structure des (α_0, β_0) -S-« bunches » obtenus, reste stable quand on augmente α_0 (resp.β_0) à un niveau quelconque $\alpha \leq \alpha_1$(resp. à un niveau quelconque $\beta \leq \beta_1$) .

4- EXEMPLE ILLUSTRATIF :

Supposons pour illustrer la Q-technique à deux phases décrite ci-dessus, que nous désirons établir une typologie des minorités américaines en prenant le chômage pour but classificatoire . Onze catégories sociales vont constituer les objets-acteurs et les taux de chômage des années 1973, 1975 et 1980 vont constituer les descripteurs classifiants. Les données relatives à cet exemple figurent à la table 1, où mb veut dire masculin blanc , mn masculin noir , mh masculin d'origine hispanique , fb féminin blanc, fn féminin noir, fh féminin d'origine hispanique .

Les classes d'âge vont de16 à 19 ans et de 20 à 24 ans . L'ensemble des objets-acteurs est Ω = { mb16-19, mn16-19, fn16-19, mh16-19, fh16-19, mb20-24, fb20-24, mn20-24, fn20-24, mh20-24, fh20-24 }

TABLE 1

	chom 73	chom 75	chom 80
mb 16- 19	12.30	18.30	16.20
mn 16 -19	27.70	38.10	37.40
fn 16 -19	35.90	41.00	39.90
mh 16 -19	19.00	27.60	21.70
fh 16 -19	20.70	27.90	23.70
mb 20 -24	6.50 *	13.20	11.10
fb 20 - 24	7.00	11.20*	8.50*
mn 20 - 24	12.80	24.70	23.80
fn 20 - 24	18.30	24.30	23.40
mh 20 - 24	8.20	16.30	12.20
fh 20 - 24	9.00	17.20	11.20

*** indique la valeur de référence (anchor value)**
Source : Le Travail dans le Monde 1, Bureau International Du Travail, Genève 1984 , page 48

Dans ce cas, les scores les plus bas sont les meilleurs car les descripteurs classifiants sont des taux de chômage, donc la fonction standardisante sera de la forme : $f(x) = \frac{x^*}{x}$, où x^* désigne le minimum observé pour chaque descripteur; et si l'on choisit l'indice de similarité d'ensembles flous défini par (4) :

$$S(\underset{\sim}{X}, \underset{\sim}{Y}) = \frac{\sum_{d \in D} [\mu_{\underset{\sim}{X}}(d) \wedge \mu_{\underset{\sim}{Y}}(d)]}{\sum_{d \in D} [\mu_{\underset{\sim}{X}}(d) \vee \mu_{\underset{\sim}{Y}}(d)]} \qquad (4)$$

et la S-comparaison définie par (2), nous obtenons, en prenant comme seuils $\alpha = 0.70$ pour la similarité et $\beta = 0.70$ pour la non-dissimilarité , Scott(Ω) = { mb20-24, fb20-24 }. Si l'on choisit fh20-24 parmi les prototypes ayant le moins de semblables dans Ω', on obtient la matrice suivante :

	mh20-24	fh20-24
mh20-24	1	1
fh20-24	1	1

qui est pleine , donc $\mathfrak{B}_1$ = { mh20-24, fh20-24 } constitue un « bunch ». Si l'on choisit fh16-19 parmi les prototypes ayant le plus de semblables dans Ω', on obtient la matrice suivante :

	~~mn16-19~~	mh16-19	fh16-19	mn20-24	fn20-24
~~mn16-19~~	1		1		
mh16-19		1	1	1	1
fh16-19	1	1	1	1	1
mn20-24		1	1	1	1
fn20-24		1	1	1	1

où l'élimination de mn16-19 donne lieu à une matrice pleine et donc au « bunch » $\mathfrak{B}_2$ = { mh16-19, fh16-19, mn20-24, fn20-24 }.Mais Ω' n'est pas encore couvert par $\mathfrak{B}_1$ et $\mathfrak{B}_2$, fn16-19 est choisi comme prototype appartenant à $W = \Omega' - \mathfrak{B}_1 \cup \mathfrak{B}_2$ et ayant le moins de semblables dans Ω' . La matrice obtenue :

	mn16-19	fn16-19
mn16-19	1	1
fn16-19	1	1

est pleine, donc on a un bunch $\mathfrak{B}_3$ = { mn16-19, fn16-19 }. Ω' n'étant pas couvert par $\mathfrak{B}_1$, $\mathfrak{B}_2$ et $\mathfrak{B}_3$, W est réduit à l'unique prototype mb16-19 qui donne lieu à la matrice pleine suivante :

	mb16-19	mn20-24	fn20-24
mb16-19	1	1	1
mn20-24	1	1	1
fn20-24	1	1	1

et donc au « bunch » $\mathfrak{B}_4$ = { mb16-19, mn20-24, fn20-24 } . Ceci achève le recouvrement de Ω'. La Q-technique décrite ci-dessus , a donc donné
à la première phase : Scatt(Ω) = { fb 20-24, mb 20-24 }
et les « bunches » empiétantssuivants :
$\mathfrak{B}_1$ = { mh 20-24, fh 20-24 };
$\mathfrak{B}_2$ = { mh 16-19, fh 16-19, mn 20-24, fn 20 - 24 };
$\mathfrak{B}_3$ = { mn 16-19 , fn 16-19 }; et
$\mathfrak{B}_4$ = { mb 16-19, mn 20-24, fn 20 - 24 };
et **à la seconde phase** : Scatt(Ω) = { fb 20-24, mb 20-24 }

et les « bunches » flous suivants :

$\mathcal{F}_1$ = 1/mh 20-24 + 1/fh 20-24 ;

$\mathcal{F}_2$ = 1/mh16-19 + 1/fh 16-19 + 0.57/mn 20-24 + 0.59/fn 20 - 24 ;

$\mathcal{F}_3$ = 1/mn 16-19 + 1/fn 16-19 ; et

$\mathcal{F}_4$ = 1/mb 16-19 + 0.43/mn 20-24 + 0.41/fn 20 - 24 ;

CONCLUSION

La littérature sur la Q-analyse abonde de Q-techniques dont la plupart sont des techniques d'agglomération dites conventionnelles. Or, pour de nombreuses raisons mathématiques et pratiques il est devenu plus avantageux de considérer des Q-techniques donnant lieu à des agglomérats flous ou, à la limite, empiétants . Nous avons, alors essayé de donner à la Q-technique à deux phases présentée dans cet article toutes les caractéristiques possibles d'une technique d'agglomération non conventionnelle .

REFERENCES

[1] Z. ABID, Contribution à L'analyse Structurale des Systèmes Complexes Au Moyen des Notions de Recouvrements et de Semi-valuation; Thèse d'Etat , Université Claude Bernard Lyon-I(Mai 1986)31-39

[2] C. ARNAL, A. BATTINI et J. P. MARCIANO, Le Problème Des Descripteurs Classifiants : Une Illustration Sur La Structure du Tiers Monde, Math et SCi. hum. n°91 (1985) 23-55 .

[3] B. BOUCHON et G. COHEN, On Fuzzy Relations and Partitions, dans : P.WANG, Ed., Advances in Fuzzy Sets, Possibility Theory and Applications (Plenum Press, NY, 1983) 97-105.

[4] L. BOUR et M. LAMOTTE, Existence et Propriétés d'un Opérateur de Maximalisation, BUSEFAL 37 (Hiver 1988) 34-41 ; L.S.I , Université Paul Sabatier, Toulouse.

[5] D. DUBOIS et H. PRADE, A Unifying View of Comparison Indices In A Fuzzy Set-Theoretic Framework , dans : Ronald R. Yager, Ed., Recent Developments In Fuzzy Set And Possibility Theory (Pergamon Press, NY,1982) 3-13.

[6] H. PRADE, Quantitative Methods in Approximate and Plausible Reasoning : The State of The Art , Rapport L.S.I n° 185 (Novembre 1983) 43 , Université Paul Sabatier, Toulouse

[7] G.R. SMITH, Textured Sets : An Approach to Aggregation Problems with Multiple Concerns, IEEE Trans on Syst, Man, and Cybernetics, Vol, SMC-10, N°4 (1980) 202-204.

[8] L. ZADEH, Fuzzy Sets and Their Application to Pattern Classification and Clustering Analysis, dans : Van Ryzin, Ed., Classification and Clustering (Academic Press, Inc., NY, 1977) 251-299

[9] M. ZELENY, Multiple Criteria Decision Making (McGraw-Hill, Inc, NY, 1985) 159-160.