MATHÉMATIQUES ET SCIENCES HUMAINES

PIERRE MICHAUD
FRANÇOIS MARCOTORCHINO

Modèles d'optimisation en analyse des données relationnelles

Mathématiques et sciences humaines, tome 67 (1979), p. 7-38
<http://www.numdam.org/item?id=MSH_1979__67__7_0>

Math. Sci. hum. (*17*^{e} année, n°67, *1979*, p.7-38)

MODELES D'OPTIMISATION EN ANALYSE DES DONNEES RELATIONNELLES

Pierre MICHAUD François MARCOTORCHINO

1. INTRODUCTION

Le problème général que nous considérons dans cet article est le suivant: il s'agit de "l'agrégation" ou de "l'approximation" de plusieurs relations binaires quelconques par une relation binaire particulière.
Les problèmes tels que les problèmes d'agrégation de classements ou de préférences ainsi que les problèmes d'agrégation de partitions ou de similarités en sont des cas particuliers.

Dans le premier de ces deux cas, un certain nombre de relations binaires sont à approximer par une relation d'ordre total (classement). Dans le second cas elles sont à approximer par une relation d'équivalence (classification).

Dans le cas de l'agrégation de préférences (classement automatique), il n'existe à l'heure actuelle que des méthodes d'énumération implicite pour résoudre le problème de façon exacte; or ces méthodes sont très onéreuses en temps de calcul et pratiquement ne permettent pas de classer plus de 20 objets.

Dans le cas de l'agrégation de similarités (classification automatique) , il existe des algorithmes performants, comme la méthode des "nuées dynamiques" de E.Diday; mais cette méthode , comme les autres (en particulier l'algorithme des transferts de S.Régnier),n'est qu'une méthode heuristique ne garantissant qu'un optimum local du problème considéré.

Dans le présent article nous montrerons que pour ces problèmes, et d'autres problèmes voisins, il existe une modélisation linéaire.

Par cette approche linéaire, nous traiterons ces deux problèmes d'agrégation de préférences et d'agrégation de similarités qui sont les plus importants pour les applications pratiques. Nous généraliserons cette approche au cas de l'agrégation de relations binaires par d'autres types de relations, telles que les relations de préordre ou de quasi-ordre etc...

Cette modélisation repose sur les deux principes suivants:
1) La prise en compte des données sous forme relationnelle.
2) L'utilisation d'une règle de majorité.

En plus de la résolution des problèmes d'agrégation de préférences et de similarités, la linéarisation des modèles permet en outre d'obtenir la réponse à diverses questions ((3), (6), (11)) restées ouvertes.

2. PRESENTATION DU PROBLEME

Le problème général que nous traiterons dans le présent article est le suivant: il s'agit de "l'agrégation" ou de "l'approximation" de plusieurs relations binaires par une relation binaire particulière.
les problèmes évoqués dans l'introduction, tels que le problème de l'agrégation de préférences ou le problème de l'agrégation de similarités en sont des cas particuliers.

Dans le premier de ces deux cas, un certain nombre de relations binaires sont à approximer par une relation d'ordre total. Dans le second cas elles sont à approximer par une relation d'équivalence.
Les relations d'ordre total et d'équivalence ont en commun la propriété suivante: ce sont des relations transitives particulières.
Mais il existe d'autres types de relations transitives utilisables.

Le problème général de l'agrégation de relations binaires, que nous considérons, consiste précisemment à approximer des relations binaires par l'une ou l'autre de ces relations transitives particulières.
L'approximation de m relations binaires données par l'une de ces relations transitives sera réalisée par une approche générale valable, à quelques modifications près, pour l'ensemble des cas considérés.

3. PRESENTATION DES DONNEES

3.1. PRESENTATION RELATIONNELLE DES DONNEES

Chacune des m relations binaires de départ, définissant les données individuelles d'un "juge" ou "critère" numéro k, sera représentée par un tableau C^k. Ce tableau, que nous appellerons tableau d'opinion ou de comparaisons, est un tableau en (0,1) de terme général:

$c^k_{ij} = 1$ Si i "est en relation" avec j pour le juge k.
$c^k_{ij} = 0$ Sinon

Par convention la valeur c_{ij} avec i=j ne sera jamais considérée.

Suivant le type de problème à traiter et la nature des données à considérer, la signification des valeurs c_{ij} sera différente.

Par exemple:

- Si la relation binaire C^k représente les préférences individuelles d'un juge k, la signification du tableau C^k sera la suivante:

  $c^k_{ij} = 1$ si l'objet i est préféré à l'objet j pour le juge k
  $c^k_{ij} = 0$ sinon.

- Dans le cas où les tableaux C^k représentent des similarités entre différents objets, la signification du tableau C^k est la suivante:

  $c^k_{ij} = 1$ si l'objet i est semblable ou similaire à l'objet j, pour le juge k.
  $c^k_{ij} = 0$ sinon.

A chaque relation binaire représentée par un tableau C^k, on associe un autre tableau, noté C'^k ; ce nouveau tableau, à valeurs (-1,+1), est défini de la façon suivante:

$c'^k_{ij} = 1$ si l'objet i "est en relation " avec l'objet j, pour le juge k.
$c'^k_{ij} = -1$ sinon

l'expression de c'^k_{ij} en termes de c^k_{ij} est donnée par

$$c'^k_{ij} = 2\, c^k_{ij} - 1 .$$

L'expression de c^k_{ij} en termes de c'^k_{ij} est donnée par:

$$c^k_{ij} = (c'^k_{ij} + 1)/ 2$$

L'obtention des tableaux C^k ou C'^k peut se faire directement ou par l'intermédiaire de tableaux de données annexes.

En sommant tous les tableaux individuels C^k, représentant les relations binaires de départ, on obtient un tableau collectif C, qui représente l'opinion de l'ensemble des m juges.

On a donc : $C = \sum_k C^k$ et $c_{ij} = \sum_k c^k_{ij}$

Le terme général c_{ij} du tableau collectif C représente d'une façon générale le nombre de relations binaires où i est en relation avec j; c'est le nombre de juges ayant mis i en relation avec j.

De la même façon le tableau collectif C' est défini par:

$$C' = \sum_k C'^k \quad \text{et} \quad c'_{ij} = \sum_k c'^k_{ij} .$$

Le terme général c'_{ij} du tableau collectif C' est égal au nombre de juges ayant mis i en relation avec j moins le nombre de juges n'ayant pas mis i en relation avec j.

3.2. GENERALISATIONS POSSIBLES

- On peut combiner les tableaux collectifs entre eux. L'addition d'un tableau collectif C1, représentant l'opinion de m1 juges, et d'un tableau collectif C2, représentant l'opinion de m2 juges, donne un tableau collectif C , représentant l'opinion de m juges; avec C=C1+C2 et m=m1+m2.

- On peut diviser les tableaux C ou C' par m ,le nombre de juges. On obtient alors des tableaux de "fréquences".

Remarque: Ces tableaux de fréquences ne doivent pas être combinés directement. L'addition d'un tableau C1/m1 et d'un tableau C2/m2 définit une opinion collective dans laquelle le groupe des m1 juges a autant de poids que le groupe des m2 juges (il faudrait alors diviser par deux le résultat pour obtenir un tableau de fréquences).

- On peut attribuer à un juge k une importance ou un pouvoir particuliers en représentant son opinion non pas par un tableau individuel mais par un tableau collectif; par exemple:

a) Pondérations
Donner à un juge k une importance p_k .
C'est par définition représenter l'opinion du juge k par une opinion collective C. Cette opinion collective est l'opinion

de p_k juges ayant la même opinion que le juge k. On a $C = p_k C^k$ et $m = p_k$.
Le tableau collectif d'opinions C du juge k est défini par:

$c_{ij} = p_k$ si i "est en relation avec" j pour le juge k.
$c_{ij} = 0$ sinon

b) "Pouvoir"
Donner à un juge k le pouvoir de p_k juges.
C'est par définition représenter l'opinion du juge k par une opinion collective C. Cette opinion collective est l'opinion de p_k juges que le juge k représente; les p_k juges n'ont pas forcément la même opinion que le juge k, et peuvent très bien avoir chacun une opinion différente.
On a $C = \Sigma C^k$ et $m = p_k$.
Le tableau collectif C d'opinion du juge k est défini par:

c_{ij} = le nombre de juges que k représente ayant mis i en relation avec j.

Cette quantité est comprise entre 0 et p_k, mais elle peut prendre toutes les valeurs intermédiaires; ce qui n'est pas le cas pour une pondération.

4. PROPRIETES DES TABLEAUX D'OPINIONS

4.1. PROPRIETES DES TABLEAUX INDIVIDUELS

Définissons les propriétés suivantes valables pour un tableau de relation binaire C^ℓ ou C'^ℓ quelconques. Les résultats correspondants aux tableaux C'^ℓ seront obtenus en remplaçant c^ℓ_{ij} par $(c'^\ell_{ij} + 1)/2$.

4.1.1. Définition de propriétés

1) Transitivité
elle est définie par:

$$c^\ell_{ij} = 1 \text{ et } c^\ell_{jk} = 1 \text{ impliquent } c^\ell_{ik} = 1$$

Que l'on peut écrire linéairement :

$$c^\ell_{ij} + c^\ell_{jk} - c^\ell_{ik} \leq 1 \qquad \text{pour } i \neq j, j \neq k, i \neq k$$

2) Antisymétrie
Elle est définie par:

$$c^\ell_{ij} = 1 \text{ entraine } c^\ell_{ji} = 0$$

qui se met sous la forme linéaire suivante:

$c^{\ell}_{ij} + c^{\ell}_{ji} \leq 1$ pour $i \neq j$

3) Symétrie
Elle est définie par:

$c^{\ell}_{ij} = 1$ entraine $c^{\ell}_{ji} = 1$

qui se met sous la forme linéaire suivante:

$c^{\ell}_{ij} - c^{\ell}_{ji} = 0$ pour $i \neq j$

4) Totalité
Elle est définie par:

Pour tout couple (i,j) l'un au moins des deux cas $c^{\ell}_{ij} = 1$ ou $c^{\ell}_{ji} = 1$ est vrai.

Ce qui se met sous la forme linéaire suivante:

$c^{\ell}_{ij} + c^{\ell}_{ji} \geq 1$ pour $i \neq j$

5) Intermédiarité
Elle est définie par:

$c^{\ell}_{ij} = 1$ et ($c^{\ell}_{jk} = 0$ et $c^{\ell}_{kj} = 0$) et $c^{\ell}_{ks} = 1$ impliquent $c^{\ell}_{is} = 1$

Ce qui s'exprime par:
i"préféré" à j , j et k non comparables, k "préféré" à s entrainent que i est "préféré" à s. En notant P la préférence et I l'incomparabilité cette propriété est quelquefois donnée par certains auteurs sous la forme: $PIP \subset P$.

Ce qui se met sous la forme linéaire suivante(Michaud):

$c^{\ell}_{ij} - (c^{\ell}_{jk} + c^{\ell}_{kj}) + c^{\ell}_{ks} - c^{\ell}_{is} \leq 1$ pour $i \neq j \neq k \neq s$

Cette inégalité est équivalente à:

$c^{\ell}_{ij} - (c^{\ell}_{jk} + c^{\ell}_{kj}) + c^{\ell}_{ks} - 1 \leq c^{\ell}_{is}$

L'expression de gauche ne peut dépasser la valeur 1. Elle n'atteint la valeur 1 que si l'on a simultanément i "préféré" à j, j et k non comparables et k "préféré" à s; d'où le résultat.

6) Cohérence
Elle est définie par:

$c^{\ell}_{ij} = 1$ et $c^{\ell}_{jk} = 1$ impliquent que les 2 expressions ($c^{\ell}_{is} = 0$ et $c^{\ell}_{si} = 0$) et ($c^{\ell}_{sk} = 0$ et $c^{\ell}_{ks} = 0$) ne doivent être simultanément vérifiées.

Ce qui s'exprime par i "préféré" à j et j "préféré" à k , ainsi que i "incomparable" à s et s "incomparable " à k est interdit. Avec les mêmes notations qu'en 5), certains auteurs expriment cette propriété par: $P^2 \cap I^2 = \emptyset$.

Ce qui se met sous la forme linéaire suivante (Michaud):

$$(c^{\ell}_{is} + c^{\ell}_{si}) + (c^{\ell}_{sk} + c^{\ell}_{ks}) \leq c^{\ell}_{ij} + c^{\ell}_{jk} - 1 \qquad i,j,k,s \text{ étant tous différents}$$

La terme de droite ne peut dépasser la valeur 1 qui est atteinte lorsque i est "préféré" à j et j "préféré" à k. Dans ce cas le terme de gauche doit être supérieur ou égal à 1; et, i et s d'une part et k et s d'autre part ne peuvent être simultanément non comparables.

Les formules linéaires précédentes étant des définitions caractérisent complétement les propriétés associées. En combinant les formules correspondant à certaines des propriétés de définition précédentes, on en déduit de nouvelles formules. Attention dans ce cas les nouvelles formules obtenues ne caractérisent ni ne définissent la conjonction de ces différentes propriétés, elles n'en sont que la déduction.

4.1.2.Combinaisons de propriétés

Proprietés déduites de :

7) Transitivité, antisymétrie
En combinant les propriétés de transitivité et d'antisymétrie, on déduit l'expression linéaire suivante:

$$c^{\ell}_{ij} + c^{\ell}_{jk} + c^{\ell}_{ki} \leq 2 \qquad \text{pour } i,j,k \text{ tous différents}$$

8) Transitivité, totalité
En combinant les propriétés de transitivité et de totalité on déduit l'expression linéaire suivante:

$$c^{\ell}_{ij} + c^{\ell}_{jk} + c^{\ell}_{ki} \geq 1 \qquad \text{pour } i,j,k \text{ tous différents}$$

9) Antisymétrie, totalité
En combinant les propriétés d'antisymétrie et de totalité on déduit l'expression linéaire suivante:

$$c^{\ell}_{ij} + c^{\ell}_{ji} = 1 \qquad i \neq j$$

10) Transitivité, antisymétrie, totalité
Comme cette propriété contient les propriétés 7) et 8) on vérifie simultanément:

$$1 \leq c^{\ell}_{ij} + c^{\ell}_{jk} + c^{\ell}_{ki} \leq 2 \quad \text{pour } i,j,k \text{ différents}$$

En combinant les trois propriétés de transitivité d'antisymétrie et de totalité on obtient les inégalités doubles suivantes:

$$c^{\ell}_{ij} + c^{\ell}_{jk} - 1 \leq c^{\ell}_{ik} \leq c^{\ell}_{ij} + c^{\ell}_{jk}$$

Remarque:
Toutes les expressions, de 1) à 10), caractérisant soit la définition de propriétés (4.1.1) soit la combinaison de propriétés (4.1.2) sont des formules linéaires en termes des différents coefficients du tableau de comparaisons C^{ℓ} (ou du tableau C'^{ℓ} si l'on remplace c^{ℓ}_{ij} par $(c'^{\ell}_{ij} + 1)/2$).

4.1.3.Définition de relations

Définissons quelques relations transitives classiques:

a)Relation de préordre:
C'est une relation, correspondant à un tableau C^{ℓ}, qui n'est que transitive.
Elle est définie par la formule donnée en 1).

b)Relation de préordre total:
C'est une relation qui est à la fois transitive et totale .
Elle est définie par les formules données en 1) et 4).

c)Relation d'ordre :
C'est une relation qui est à la fois transitive et antisymétrique.
Elle est définies par les formules données en 1) et 2)

d)Relation d'ordre total:
C'est une relation qui est à la fois transitive antisymétrique et totale.
Elle est définie par les formules données en 1), 2) et 4).

e)Relation d'équivalence:
C'est une relation qui est à la fois transitive et symétrique.
Elle est définie par les formules données en 1) et 3).

f)Relation d'équivalence totale:
C'est une relation qui est à la fois transitive symétrique et totale.
Elle est définie par les formules données en 1), 3) et 4).

Cette relation n'a aucun intérêt pratique et ne sera pas considérée par la suite; elle correspond à l'unique configuration où tous les éléments sont équivalents et appartiennent à une seule et même classe.

g)Relation d'ordre d'intervalles:
C'est une relation qui est à la fois transitive antisymétrique et qui possède la propriété d'intermédiarité.
Elle est définie par les formules données en 1), 2) et 5).

h)Relation de quasi-ordre (semiorder) :
C'est une relation qui est à la fois transitive antisymétrique et qui possède les propriétés d'intermédiarité et de cohérence.
Elle est définie par les formules données en 1), 2), 5) et 6).
On notera qu'un quasi-ordre est ici défini de façon antisymétrique, et correspond donc au terme anglais de semiorder, alors qu'il peut être aussi défini comme une relation totale, formalisation généralement utilisée dans (12).

Il existe également d'autres types de relations binaires classiques mais non transitives; citons par exemple:

i)Relation de tournoi:
C'est une relation qui est à la fois antisymétrique et totale.
Elle est définie par les formules données en 2) et 4) qui sont équivalentes à la formule donnée en 8).

j)Relation de match:
C'est tout simplement une relation totale.
Elle est définie par la formule 4).

Remarque:
Notons que toutes les relations binaires considérées ici sont définies exclusivement à partir des formules correspondant à la définition des propriétés présentées en (4-1-1). Comme nous l'avons vu ces formules sont toutes linéaires; il en résulte que ces relations binaires peuvent être définies en termes d'égalités ou d'inégalités linéaires, par rapport aux coefficients des tableaux de données C^k ou C'^k.

4.1.4.Obtention des tableaux C^k

A priori les tableaux d'opinions individuels C^k et C'^k peuvent n'avoir aucune autre propriété que celle de valoir 0 ou 1 pour c^k_{ij} et celle de valoir 1 ou -1 pour c'^k_{ij}.

Cependant dans la pratique l'obtention des tableaux C^k ou C'^k peut se faire directement ou par l'intermédiaire de tableaux de données annexes. Dans ce dernier cas les tableaux C^k ou C'^k auront souvent des propriétés particulières; donnons quelques exemples.

a) Cas de l'agrégation de préférences

1)Dans le cas le plus simple les données brutes se présentent sous forme de tableaux de classements (avec ou sans ex-aequo); le rang de l'objet i pour le juge k étant représenté par r_{ik}. Le tableau C^k correspondant sera obtenu de la façon suivante:

$$c_{ij}^k = 1 \quad \text{si} \quad r_{ik} \leq r_{jk}$$
$$c_{ij}^k = 0 \quad \text{sinon}$$

2) Si maintenant les données brutes se présentent sous la forme d'échelles ou de notes n_{ik} (note du juge k pour l'objet i), le tableau C^k associé pourra avoir l'une des formes suivantes selon les hypothèses:

$$c_{ij}^k = 1 \quad \text{si} \quad n_{ik} \geq n_{jk}$$
$$c_{ij}^k = 0 \quad \text{sinon}$$

Dans ce cas le tableau C^k représentera un préordre total.

$$c_{ij}^k = 1 \quad \text{si} \quad n_{ik} > n_{jk}$$
$$c_{ij}^k = 0 \quad \text{sinon}$$

Dans ce cas le tableau C^k représentera un ordre pouvant être partiel. Notons cependant que dans ce cas l'incomparabilité, correspondant à l'égalité des notes, est transitive.

$$c_{ij}^k = 1 \quad \text{si} \quad n_{ik} > n_{jk} + s$$
s étant un seuil positif ou nul

$$c_{ij}^k = 0 \quad \text{sinon}$$

Dans ce cas le tableau représente un ordre en général partiel. Notons que dans ce cas l'incomparabilité n'est pas en général transitive, mais vérifie les propriétés d'intermédiarité et de cohérence 5) et 6) de (4-1-1), il s'agit donc d'un tableau représentant un quasi-ordre.

b) Cas de l'agrégation de similarités

1) Le cas le plus simple correspond à la situation où un juge représente un "critère" à plusieurs modalités. Le tableau de données brutes associé est alors un tableau de modalités, où r_{ik} représente la valeur de la modalité attribuée à l'objet i par le juge ou critère k. Le tableau C^k est défini de la façon suivante:

$$c_{ij}^k = 1 \quad \text{si} \quad r_{ik} = r_{jk}$$
c'est à dire i et j ont même modalité pour le juge k

$$c_{ij}^k = 0 \quad \text{sinon}$$

Dans ce cas le tableau C^k représente une partition.

Dans le cas où le juge ou critère k représente une propriété, le tableau de données brutes associé est alors un tableau de présence-absence, où $r_{ik} = 1$ si l'objet i a la propriété k et $r_{ik} = 0$ si l'objet i n'a pas la propriété k. Dans ce cas le tableau C peut se définir de plusieurs façons:

$c^k_{ij} = 1$ si $r_{ik} . r_{jk} = 1$ c'est à dire si i et j ont tous les deux la propriété k
$c^k_{ij} = 0$ sinon

$c^k_{ij} = 1$ si $r_{ik} = r_{jk}$ c'est à dire si i et j ont ou n'ont pas simultanément la propriété k
$c^k_{ij} = 0$ sinon

Dans les deux cas le tableau C^k représente une partition.

2) Dans le cas où les données brutes sont représentées par un tableau de "proximités" P^k (représentant des distances, des dissimilarités, ou des mesures de proximités non nécessairement symétriques), le tableau C^k est défini de la façon suivante:

$c^k_{ij} = 1$ si $p^k_{ij} \leq s$, s étant un seuil donné
$c^k_{ij} = 0$ sinon

Le tableau C^k peut très bien ne pas représenter une partition, par exemple lorsque P^k n'est pas symétrique.

4.2. PROPRIETES DES TABLEAUX COLLECTIFS C ET C'

Comme nous l'avons vu précédemment les tableaux collectifs C et C' sont obtenus respectivement en sommant les tableaux individuels C^k et C'^k pour l'ensemble des m juges.

Au Chapitre 4-1 nous avons défini des relations et des propriétés individuelles associées aux tableaux C^k et C'^k, dans ce chapitre nous définirons cette fois des relations et des propriétés collectives associées aux tableaux C et C'
Pour ce faire nous utiliserons une règle de majorité " ma" qui nous permettra de définir une relation collective, à partir du tableau C (ou C').
La majorité ma est définie par ma = p m où p représente un pourcentage strictement positif; $0 < p \leq 1$. Sauf mention contraire p = 1/2, c'est la règle de la majorité de Condorcet. Mais toute autre valeur de ma est possible, en particulier si p = 1 on retrouve la règle de Pareto ou règle de l'unanimité.

Cette relation collective est alors définie de la façon suivante:

i "est en relation " avec j dans le relation collective , si et seulement si $c_{ij} \geq$ ma; c'est à dire s'il y a au moins ma juges ayant mis i en relation avec j.

En général la relation collective dépend de la valeur ma de la majorité.

Remarque: Si m = 1 , cas d'un tableau individuel C^k, ma = p. Dans ce cas c_{ij} ma est indépendant de ma et équivalent à $c_{ij} = 1$.
On retrouve la définition des relations individuelles. Mais dans ce cas, ces relations individuelles sont indépendantes de la valeur de la majorité.

4.2.1. Définition de propriétés sur la relation collective

La définition se fera à partir des tableaux collectifs C et C'

1) Transitivité
Elle est définie par:

$$c_{ij} \geq ma \quad \text{et} \quad c_{jk} \geq ma \quad \text{impliquent} \quad c_{ik} \geq ma$$

2) Antisymétrie
Elle est définie par:

$$c_{ij} \geq ma \quad \text{implique} \quad c_{ji} \geq ma$$

3) Symétrie
Elle est définie par:

$$c_{ij} = c_{ji}$$

4) Totalité
Elle est définie par:

Au moins l'une des inégalités $c_{ij} \geq$ ma ou $c_{ji} \geq$ ma est vérifiée.
Ce qui se traduit par $c_{ij} + c_{ji} \geq 2ma$

5) Intermédiarité
Elle est définie par:

$$c_{ij} \geq ma \text{ et } c_{jk} < ma \text{ et } c_{kj} < ma \text{ et } c_{ks} \geq ma \text{ impliquent } c_{is} \geq ma$$

6) Cohérence
Elle est définie par:

Les inégalités suivantes ne doivent pas être toutes simultanément vérifiées:

$c_{ij} \geq ma$ et $c_{jk} \geq ma$ et $c_{is} < ma$ et $c_{si} < ma$ et $c_{sk} < ma$ et $c_{ks} < ma$

Remarque:
Certains auteurs considèrent d'autres types de propriétés dites "floues" . Ces propriétés sont associées à un tableau "flou" qui serait à rapprocher du tableau de fréquences C/m. Attention ces propriétés floues ne sont pas définies nécessairement à partir de la règle de majorité précédemment définie.
Par exemple (les propriétés seront définies relativement au tableau C):

7) Transitivité floue
Elle est définie par:

$$c_{ik} \geq \mathrm{Min}(c_{ij}, c_{jk})$$

8) Transitivité stochastique
Elle est définie par:

$$c_{ij} \geq m/2 \text{ et } c_{jk} \geq m/2 \quad \text{impliquent} \quad c_{ik} \geq \mathrm{Max}(c_{ij}, c_{jk})$$

$$\text{c'est à dire } c_{ik} \geq c_{ij} \text{ et } c_{ik} \geq c_{jk}$$

9) Antisymétrie floue
Elle est définie par:

$$c_{ij} = c_{ji} \quad \text{implique} \quad c_{ij} = 0 \quad \text{soit } c_{ij} \geq 0 \text{ et } -c_{ij} \geq 0$$

10) Antisymétrie floue parfaite
Elle est définie par:

$$c_{ij} \geq c_{ji} \quad \text{implique } c_{ji} = 0 \text{ soit } c_{ji} \geq 0 \quad \text{et } -c_{ji} \geq 0$$

etc..

4.2.2.Linéarisation de ces propriétés (Michaud)

On a le résultat suivant:
Toutes les propriétés de 1) à 10) peuvent se mettre sous la forme d'un système d'égalités ou d'inégalités linéaires contenant les entiers c_{ij} ainsi que des variables logiques x_{ij}.
Les conditions 3)de symétrie et 4) de totalité étant d'ailleurs initialement linéaires,ne nécessiteront pas une telle mise en forme.

Ceci résulte des conditions suivantes:

Par la suite nous considérerons des expressions

$a_i \geq 0$ i variant de 1 à q a_i étant un entier relatif
et $b_j \geq 0$ j variant de 1 à r

Ces expressions peuvent dépendre de variables quelconques et varier arbitrairement. La seule hypothèse étant qu'elles sont toutes bornées inférieurement et supérieurement. Nous appellerons par la suite t la plus grande de ces bornes. Par hypothèse on a:

$t \geq a_i \geq -t$ i variant de 1 à q
et $t \geq b_j \geq -t$ j variant de 1 à r

CONDITION 1

La vérification simultanée des q inégalités du système $a_i \geq 0$ $i = 1...q$ est interdite.
Cette condition admet la représentation linéaire suivante:

(i) $(t+1)x_i \geq a_i + 1$ $i= 1....q$

(ii) $\Sigma\ x_i \leq q-1$

A chaque inégalité $a_i \geq 0$ est associée une variable logique x_i valant 0 ou 1. Pour $a_i < 0$ x_i peut valoir 0 ou 1 tandis que $a_i \geq 0$ entraine $x_i = 1$(relation (i)). (ii) exprime qu'il ne peut y avoir plus de q-1 x_i simultanément égaux à 1. D'où le résultat.
Le nouveau système contient le nombre d'inégalités et de variables initiales plus autant de variables logiques x_i que d'inégalités plus une nouvelle inégalité.

CONDITION 2

La vérification simultanée des q inégalités $a_i \geq 0$ pour i=1...q implique la vérification simultanée des r inégalités $b_j \geq 0$ pour j=1...r.
Cette condition admet la représentation linéaire suivante:

(i) $(t+1)x_i \geq a_i + 1$ i=1...q

(ii) $b_j \geq (t+1)(\ \Sigma\ x_i - q)$ j=1...r

Comme précédemment à chaque inégalité $a_i \geq 0$ est associée une variable logique x_i valant 0 ou 1 de telle sorte que $a_i \geq 0$ entraine $x_i=1$ (relations (i)). Lorsque les q inégalités $a_i \geq 0$ sont vérifiées $\Sigma\ x_i - q = 0$ et (ii) devient $b_j \geq 0$. Dans les autres cas l'expression $\Sigma\ x_i - q$ est strictement négative. Comme par définition $b_j \geq -(t+1)$ la relation est alors toujours vérifiée indépendamment de la valeur de b_j.
Le nouveau système contient le même nombre de variables et d'inégalités que le système initial plus q variables logiques x_i.

Remarque:

Les conditions 1 et 2 contiennent également les cas d'égalités. Dans ce cas il suffit de représenter $a_i = 0$ par 2 inégalités $a_i \geq 0$ et $-a_i \geq 0$, de même pour b_j.

En résumé:
-les propriétés 3) symétrie et 4) totalité sont linéaires initialement

-La propriété 7) transitivité floue a une structure spéciale

-Les propriétés 1) transitivité , 2) antisymétrie ,5) intermédiarité, 8) transitivité stochastique, 9) antisymétrie floue et 10) antisymétrie floue parfaite correspondent à la Condition 2.

-La propriété 6) cohérence correspond à la Condition 1

Toutes les propriétés correspondant à la Conditions 1 ou 2 sont donc linéaires.
Pour appliquer les Conditions 1 ou 2 il suffira de remplacer les inégalités $c_{ij} \geq ma$ et $c_{ji} < ma$ respectivement par $c_{ij} \geq m'$ et $c_{ji} < m''$; m' étant le plus petit entier supérieur ou égal à ma et m'' étant égal à $m'-1$.

Donnons deux exemples:

a) Transitivité
La propriété $c_{ij} \geq ma$ et $c_{jk} \geq ma$ impliquent $c_{ik} \geq ma$ se met sous la forme linéaire suivante:

$$(m+1)\ x_{ij} \geq (c_{ij} - m') + 1$$

$$(m+1)\ x_{jk} \geq (c_{jk} - m') + 1$$

$$c_{ik} - m' \geq (m+1)(x_{ij} + x_{jk} - 2)$$

Les variables logiques x_{ij} valant 0 ou 1

Les indices i,j,k étant tous différents.

b) Antisymétrie
La propriété: $c_{ij} \geq m'$ implique $c_{ji} < m''$ se met sous la forme linéaire suivante:

$$(m+1)\ x_{ij} \geq (c_{ij} - m') + 1$$

$$(m'' - c_{ji}) \geq (m+1)(x_{ij} - 1)$$

Pour terminer montrons directement que la propriété 7) transitivité floue peut se mettre sous forme linéaire.

La propriété: $c_{ik} \geq \mathrm{Min}\ (c_{ij}, c_{jk})$

se met sous la forme linéaire suivante:

(i) $mx_{ijk} \geq c_{ij} - c_{jk} \geq -m(1 - x_{ijk})$

(ii) $c_{ik} \geq c_{jk} - m(1 - x_{ijk})$

(iii) $c_{ik} \geq c_{ij} - mx_{ijk}$

$x_{ijk} = 0$ ou 1

D'après (i) $c_{ij} > c_{jk}$ entraine $x_{ijk} = 1$ dans ce cas la relation (ii) se réduit à $c_{ik} \geq c_{jk}$ (cette relation est toujours vérifiée si $x_{ijk} = 0$).

De la même façon $c_{ij} < c_{jk}$ entraine $x_{ijk} = 0$ et cette fois ci la relation (iii) se réduit à $c_{ik} \geq c_{ij}$.

Dans le cas où $c_{ij} = c_{jk}$ $x_{ijk} = 0$ ou 1 etc...

En définitive les propriétés sont donc toutes linéarisables.

4.2.3. Propriétés des tableaux C déduites de celles des tableaux C^k

Lorsque chacun des tableaux C^k vérifie l'une des propriétés définies en (4-1-1), les éléments des tableaux C et C' vérifient les inégalités ou égalités suivantes:

1) Transitivité pour chacun des C^k,
elle implique:

$c_{ij} + c_{jk} - m \leq c_{ik}$ (tableau C)
et pour i,j,k tous différents
$c'_{ij} + c'_{jk} - m \leq c'_{ik}$ (tableau C')

2) Antisymétrie pour chacun des C^k
elle implique:

$c_{ij} + c_{ji} \leq m$
et pour $i \neq j$
$c'_{ij} + c'_{ji} \leq 0$

3) Symétrie pour chacun des C^k
elle implique :

$c_{ij} = c_{ji}$
et pour $i \neq j$
$c'_{ij} = c'_{ji}$

4) Totalité pour chacun des C^k
elle implique:

$c_{ij} + c_{ji} \geq m$

$c'_{ij} + c'_{ji} = 0$ pour $i \neq j$

5) Intermédiarité pour chacun des C^k
elle implique:

$$c_{ij} - (c_{jk} + c_{kj}) + c_{ks} - c_{is} \leq m$$

et i,j,k,s tous différents

$$c'_{ij} - (c'_{jk} + c'_{kj}) + c'_{ks} - c'_{is} \leq 3m$$

6) Cohérence pour chaque tableau C^k
elle implique :

$$- m \leq - c_{ij} - c_{jk} + (c_{is} + c_{si}) + (c_{sk} + c_{ks})$$

et i,j,k,s tous différents

$$- 4m \leq - c'_{ij} - c'_{jk} + (c'_{is} + c'_{si}) + (c'_{sk} + c'_{ks})$$

7) Transitivité, antisymétrie pour chacun des C^k
elle implique:

$$c_{ij} + c_{jk} + c_{ki} \leq 2m$$
et i,j,k tous différents
$$c'_{ij} + c'_{jk} + c'_{ki} \leq m$$

8) Transitivité, totalité pour chaque tableau C^k
elle implique:

$$c_{ij} + c_{jk} + c_{ki} \geq m$$
et i,j,k tous différents
$$c'_{ij} + c'_{jk} + c'_{ki} \geq -m$$

9) Antisymétrie, totalité pour chaque tableau C^k
En combinant les propriétés d'antisymétrie et de totalité on déduit:

$$c_{ij} + c_{ji} = m$$
et i,j,k tous différents
$$c'_{ij} + c'_{ji} = 0$$

10) Transitivité, antisymétrie, totalité pour chaque tableau C^k
Comme cette propriété contient les propriétés 7) et 8), on en déduit:

$$m \leq c_{ij} + c_{jk} + c_{ki} \leq 2m$$
et i,j,k tous différents
$$-m \leq c'_{ij} + c'_{jk} + c'_{ki} \leq m$$

En combinant les trois propriétés de transitivité, d'antisymétrie et de totalité on déduit les doubles inégalités suivantes:

$$c_{ij} + c_{jk} - m \leq c_{ik} \leq c_{ij} + c_{jk}$$

et

$$c'_{ij} + c'_{jk} - m \leq c'_{ik} \leq c'_{ij} + c'_{jk} + m$$

5. DEFINITION DU PROBLEME ET PROPRIETES

On se propose dans cette nouvelle étape de trouver une relation collective transitive particulière "approximant au mieux" les m tableaux individuels C^k.
Cette relation transitive particulière est l'inconnue du problème et nous la noterons Y.

Dans tous les cas la relation binaire collective cherchée sera représentée, comme les tableaux C^k, par un tableau binaire Y, de terme général:

$y_{ij} = 1$ Si i "est en relation" avec j dans dans la relation collective
$y_{ij} = 0$ Sinon

L'expression "est en relation avec" dépend du contexte,et peut représenter l'une des 7 relations transitives définies en 4.1.3.
Bien évidemment le tableau Y peut aussi représenter un tournoi ou un match; dans ce cas, la solution du problème est évidente.
Par exemple $y_{ij} = 1$ signifie i est classé avant j dans le cas où Y représente un ordre total tandis qu'il signifie i est équivalent à j si Y représente une relation d'équivalence.

Il nous faut maintenant définir ce que nous entendons par "approximer au mieux".
Nous définirons la distance $d(Y, C^k)$ entre un tableau C^k et la relation inconnue Y.
Nous définirons ensuite ce que nous appellerons la "distance" entre l'ensemble des tableaux collectifs C^k et la relation Y.

Le problème consistera alors à rechercher la relation collective Y située à distance minimale de l'ensemble des m relations C^k.

5.1. LES DISTANCES ASSOCIEES

5.1.1. Distances entre Y et C^k

La distance de Hölder entre la relation binaire collective cherchée Y et un tableau de relation C^k, est donnée par:

$$d_\ell^\ell(Y, Ck) = \sum_i \sum_j |y_{ij} - c_{ij}^k|^\ell$$

Cette distance a les propriétés suivantes:

1) Cette distance est linéaire en terme de Y .
La linéarité est due à la séparabilité de l'expression précédente en fonction de Y .
Considérons, en effet, une fonction séparable arbitraire de Y, elle peut s'écrire sous la forme:

$$f(Y) = \sum_i \sum_j f_{ij}(y_{ij})$$

Mais puisque y_{ij} est égal à 0 ou à 1, cette précédente expression est égale à:

$$f(Y) = \sum_i \sum_j f_{ij}(0) + \sum_i \sum_j (f_{ij}(1) - f_{ij}(0))y_{ij}$$

qui est bien une expression linéaire en Y . Ce résultat s'applique à la distance $d_\ell^\ell(C^k, Y)$, qui est une fonction séparable de Y; on obtient donc:

$$d_\ell^\ell(C^k,Y) = \sum \sum |c_{ij}^k|^\ell + \sum \sum(|1 - c_{ij}^k|^\ell - |c_{ij}^k|^\ell)y_{ij}$$

2) Toutes les distances de Hölder sont indépendantes de la caractéristique 1 de la métrique.
Puisque c_{ij}^k est égal à 0 ou 1, $|c_{ij}^k|^\ell$ et $|1 - c_{ij}^k|^\ell$ sont égaux respectivement à c_{ij}^k et $1 - c_{ij}^k$; et l'expression de $d_\ell^\ell(C^k ,Y)$ est égale à :

$$\sum_i \sum_j c_{ij}^k + \sum_i \sum_j (1- 2c_{ij}^k) y_{ij}$$

finalement l'expression de $d_\ell^\ell(C^k, Y)$ est indépendante de ℓ; elle est linéaire par rapport aux y_{ij} et est donnée par:

$$\sum_i \sum_j c_{ij}^k - \sum_i \sum_j c_{ij}'^k y_{ij}$$

Remarque:
Dans le langage de la théorie des ensembles, $d_1(C^k, Y)$ représente la distance de la différence symétrique entre deux ensembles. Les deux ensembles sont C^k et Y, leurs élements sont représentés par des 1 dans chaque tableau. La différence symétrique entre deux ensembles est égale au nombre d'éléments dans un des ensembles mais pas dans l'autre, c'est précisemment la valeur donnée par :

$$\sum_i \sum_j c_{ij}^k - \sum_i \sum_j c_{ij}'^k y_{ij}$$

5.1.2."Distances" entre Y et tous les tableaux de relation C^k

a) Nous définirons la "distance" entre Y et tous les C^k par l'expression homogène suivante:

$$\sqrt[\ell]{\sum_k d_\ell^\ell (Y, C^k)}$$

C'est cette expression que la relation Y doit minimiser.

Or minimiser la précédente fonction est équivalent à minimiser (même solution) la fonction suivante:

$$\sum_{k=1}^{m} d_\ell^\ell (Y , C^k)$$

Cette fonction objectif ou économique, est indépendante de la caractéristique ℓ de la métrique de Hölder; elle est linéaire et égale à

$\sum_k d_1 (C^k,Y)$ pour $\ell = 1,2....$

En particulier la distance quadratique ou euclidienne est en fait linéaire.

L'expression développée de cette fonction économique est donnée par:

$$\sum_i \sum_j (\sum_k c_{ij}^k) - \sum_i \sum_j (\sum_k c'^k_{ij}) y_{ij}$$

C'est à dire:

$$\sum_i \sum_j c_{ij} - \sum_i \sum_j c'_{ij} y_{ij}$$

Minimiser cette expression est équivalent à maximiser :

$$\sum_i \sum_j c'_{ij} y_{ij}$$

Interprétons cette dernière fonction économique. Supposons que i soit en relation avec j dans la relation collective Y. Dans ce cas, d'après la définition de C, nous avons c_{ij} accords; c'est à dire c_{ij} juges ont mis i en relation avec j. De la même façon, c'_{ij} représente le nombre d'accords moins le nombre de non-accords; c'est à dire le nombre de juges ayant mis i en relation avec j moins le nombre de juges n'ayant pas mis i en relation avec j. Si nous sommons toutes les valeurs c'_{ij} y_{ij}, pour tous les couples (i,j) et $i \neq j$, on obtient le nombre total d'accords moins le nombre total de non-accords entre l'ensemble des m relations C^k et la relation Y.

En résumé, le problème consistant à trouver la relation collective Y située à "distance" minimale de l'ensemble des relations C^k est équivalent à trouver la relation collective Y , maximisant le nombre total d'accords moins le nombre total de "non-accords" avec l'ensemble des opinions émises.

b) On peut également utiliser la "distance" entre Y et tous les tableaux C^k définie par:

$$\max_k \; d(Y,C^k)$$

Cette "distance" est en fait la limite quand la caractéristique tend vers l'infini de la "distance" entre Y et les tableaux C^k définie précédemment à l'aide de la métrique de Holder.

Minimiser cette expression est équivalent à:

Minimiser z'

avec $z' \geq d(Y, C^k)$ pour $k=1,\ldots,m$

Ce qui est aussi équivalent à:

Maximiser z

avec $z \leq \sum_i \sum_j c'_{ij} \; y_{ij}$ pour $k=1,\ldots,m$

Il s'agit également d'un problème linéaire, cette fois en termes de z et des y_{ij}.

5.2. UNANIMITE ET LINEARITE
(Propriété paretienne)

1) Agrégation de tableaux individuels C^k transitifs par un tableau collectif Y transitif et total (préordre total).

On a le résultat suivant(Michaud):

Si j est en relation avec i pour l'ensemble des m juges et si i est en relation avec j pour un nombre de juges strictement inférieur à m, alors j est en relation avec i dans la solution collective optimale; c'est à dire le rang de j est inférieur ou égal au rang de i dans la solution optimale.

Appelons Y^1 la solution collective optimale. Dans ce cas le résultat précédent s'exprime par:

$c'_{ji} = m$ et $c'_{ij} < m$ entraine $y^1_{ji} = 1$

Démonstration:

Supposons que $c'_{ji} = m$ que $c'_{ij} < m$ mais que l'on ait $y^1_{ji} = 0$.
Comme Y^1 est total $y^1_{ij} = 1$ et i est classé strictement avant j dans le classement optimal Y^1.

Les n objets à classer se présentent comme suit dans le classement optimal Y^1:
Les objets strictement classés avant i (et avant j)
L'objet i
Les objets classés entre i et j (i et j exceptés)
L'objet j
Les objets classés strictement après j (et après i)

Appelons K l'ensemble des objets classés entre i et j (sans i et j). Cet ensemble se décompose en trois sous-ensembles disjoints définis de la façon suivante :
K_1 est l'ensemble des objets ex-aequo à i
K_2 est l'ensemble des objets classés strictement après i mais strictement avant j.
K_3 est l'ensemble des objets ex-aequo à j.

K_2 peut très bien ne pas contenir d'éléments.
De même pour K_1 et K_3; ce qui est le cas lorsque Y représente un ordre total.

La valeur de la fonction économique est donnée par:

$$F(C,Y^1) = Cte + c'_{ij} + \sum_{k\in K_1\cup K_2\cup K_3} c'_{ik} + \sum_{k\in K_1\cup K_2\cup K_3} c'_{kj} + \sum_{k\in K_1} c'_{ki} + \sum_{k\in K_3} c'_{jk}$$

Les termes contenant les indices i ou j sont indiqués, la partie constante contient tous les autres termes ou ni i ni j ne figurent.

Si l'on échange dans le préordre total Y^1 la position de i et de j on obtient un nouveau préordre total Y^2, où les ex-aequo de i deviennent les ex-aequo de j etc...
La valeur de la fonction économique correspondante est donnée par:

$$F(C,Y^2) = Cte + c'_{ji} + \sum_{k\in K_1\cup K_2\cup K_3} c'_{jk} + \sum_{k\in K_1\cup K_2\cup K_3} c'_{ki} + \sum_{k\in K_1} c'_{kj} + \sum_{k\in K_3} c'_{ik}$$

La constante est la même que précédemment. De plus les

expressions $\sum_{k\in K_1} c'_{ki}$, $\sum_{k\in K_3} c'_{jk}$, $\sum_{k\in K_1} c'_{kj}$ et $\sum_{k\in K_3} c'_{ik}$

apparaissent dans les deux fonctions économiques.

En retranchant $F(C,Y^1)$ de $F(C,Y^2)$ ces expresssions disparaissent et l'on obtient (en regroupant terme à terme):

$$F(C,Y^2)-F(C,Y^1)=c'_{ji} - c'_{ij} + \sum_{k\in K_1\cup K_2}(c'_{jk}-c'_{ik})+ \sum_{k\in K_1\cup K_2}(c'_{ki}- c'_{kj})$$

Comme par hypothèse chaque C'^kest transitif le tableau collectif C vérifie (4-2-2-1):

$$c'_{ji} + c'_{ik} - m \leq c'_{jk} \quad \text{d'une part}$$

$$\text{et } c'_{kj} + c'_{ji} - m \leq c'_{ki} \quad \text{d'autre part.}$$

Comme $c'_{ji} = m$ on a respectivement:

$$c'_{ik} \leq c'_{jk} \quad \text{et } c'_{kj} \leq c'_{ki}$$

Dans l'expression de $F(C,Y^2)- F(C,Y^1)$ les termes entre parenthèses sont donc tous positifs ou nuls; comme d'autre part $c'_{ji} - c'_{ij} > 0$ par hypothèse, il en résulte que l'expression $F(C,Y2)-F(C,Y1)$ est strictement positive et Y^1

n'est pas solution optimale, ce qui contredit l'hypothèse $y^1_{ji} = 0$. D'où le résultat.

2) Agrégation de tableaux individuels C^k transitifs par un tableau collectif Y antisymétrique transitif et total (ordre total)

Le résultat est identique au cas précédent ainsi que la démonstration. Dans ce cas K_1 et K_3 ne contiennent pas d'éléments.

3) Autres résultats (Michaud)

Les résultats précédents sont vrais également lorsque la distance entre Y et C^k est définie par:

$$d(C,Y) = \operatorname*{Max}_k d(C^k,Y)$$

Dans ce cas la solution optimale Y est définie par:

$$d(C,Y^1) = \operatorname{Min} \operatorname*{Max}_k d(C^k,Y)$$

Les résultats antérieurs sont vrais en prenant $m = 1$; c'est à dire lorsque le tableau collectif C représente l'un quelconque des tableaux individuels C^k.

Avec les mêmes notations que précédemment (Y^1 solution optimale) on a:

$$F(C^k,Y^2) - F(C^k,Y^1) \geq 0 \qquad \text{si } c'^k_{ji} = 1 \text{ et } c'^k_{ij} = 1$$

$$\text{et } F(C^k,Y^2) - F(C^k,Y^1) > 0 \qquad \text{si } c'^k_{ji} = 1 \text{ et } c'^k_{ij} = -1$$

(Le terme $c'^k_{ji} - c'^k_{ij}$ étant respectivement nul ou égal à 2 dans l'expression de $F(C^k,Y^2) - F(C^k,Y^1)$ définie précédemment en 1)).

Comme $d(C^k,Y) = \sum_i \sum_j c_{ij} - F(C^k,Y)$ on a :

$$d(Ck,Y^2) \leq d(C^k,Y^1) \qquad \text{si } c'^k_{ji} = 1 \text{ et } c'^k_{ij} = 1$$

$$\text{et} \quad d(C^k,Y^1) < d(C\ ,Y\ ) \qquad \text{si } c'^k_{ji} = 1 \text{ et } c'^k_{ij} = -1$$

De plus $c'_{ji} = m$ et c'_{ij} m impliquent que $c'^k_{ji} = 1$ et $c'^k_{ij} = -1$ pour au moins un indice k. Pour cet indice l'inégalité stricte est vérifiée.

Ces résultats sont valables pour n'importe quel tableau C^k.

On a donc $\operatorname*{Max}_k d(Ck,Y^2) < \operatorname*{Max}_k d(C^k,Y^1)$

ce qui contredit l'optimalité de Y^1. D'où le résultat.

4) Agrégation de tableaux individuels C^k transitifs par un tableau collectif Y transitif et symétrique (équivalence).

On a le résultat suivant:
Si j est en relation avec i pour l'ensemble des m juges et si i est en relation avec j pour l'ensemble des m juges, alors j est en relation avec i dans le solution collective optimale; c'est à dire j et i sont dans la même classe dans la solution optimale.

Appelons encore Y^1 la solution optimale. Ce résultat s'exprime par:

$$c'_{ij} = m \text{ et } c'_{ji} = m \quad \text{impliquent} \quad y^1_{ij} = 1$$

Démonstration:

Supposons que $c'_{ij} = m$ et $c'_{ji} = m$ mais que $y^1_{ij} = 0$. Comme Y^1 est symétrique $y^1_{ji} = 0$ c'est à dire i et j ne sont pas dans la même classe.

Appelons:
K_i les objets qui sont dans la même classe que i (i excepté)
K_j les objets qui sont dans la même classe que j (j excepté)
(Ces ensembles peuvent très bien être vides).

La valeur de la fonction économique est donnée par :

$$F(C,Y^1) = Cte + \sum_{k\in K_i} (c'_{ik} + c'_{ki})$$

Les termes contenant les indices i sont indiqués. La partie constante contient tous les termes où l'indice i ne figure pas.

Appelons Y^2 la solution collective dans laquelle i est mis dans la même classe que j, la valeur correspondante de la fonction économique est donnée par:

$$F(C,Y^2) = Cte + \sum_{k\in K_j} (c'_{ik} + c'_{ki}) + c'_{ij} + c'_{ji}$$

La constante étant la même que précédemment.

Posons:

$$S_i = \sum_{k\in K_i} (c'_{ik} + c'_{ki}) \quad (\text{Si } K_i = \emptyset,\ S_i = 0)$$

et

$$S_j = \sum_{k\in K_j} (c'_{ik} + c'_{ki}) \quad (\text{Si } K_j = \emptyset,\ S_j = 0)$$

a) $S_j \geq S_i$

Dans ce cas $F(C,Y^2) - F(C,Y^1) = S_j - S_i + c'_{ij} + c'_{ji} \geq 2m$ et Y^1 n'est pas solution optimale.

b) $S_i > S_j$

Dans ce cas écrivons la fonction économique associée à Y^1 sous la forme suivante :

$$F(C,Y^1) = Cte + \sum_{k \in K_j} (c'_{jk} + c'_{kj})$$

Les termes contenant les indices j, cette fois, sont indiqués. La partie constante contient tous les termes où l'indice j ne figure pas

Appelons Y^3 la solution collective dans laquelle j est mis dans la même classe que i, la fonction économique associée est donnée par:

$$F(C,Y^3) = Cte + \sum_{k \in K_i} (c'_{jk} + c'_{kj}) + c'_{ij} + c'_{ji}$$

La constante étant la même que dans la formule précédente.

Comme par hypothèse chaque tableau C'^k est transitif le tableau collectif C' verifie :

$c'_{ij} + c'_{jk} - m \leq c'_{ik}$ d'une part et $c'_{ji} + c'_{ik} - m \leq c'_{jk}$ d'autre part

Comme par hypothèse $c'_{ij} = m$ et $c'_{ji} = m$, on a respectivement:

$$c'_{jk} \leq c'_{ik} \quad \text{et} \quad c'_{ik} \leq c'_{jk} \quad \text{d'où} \quad c'_{ik} = c'_{jk}$$

De même de:

$c'_{kj} + c'_{ji} - m \leq c'_{ki}$ et $c'_{ki} + c'_{ij} - m \leq c'_{kj}$ on en déduit

$$c'_{ki} = c'_{kj}$$

Il en résulte que :

$$\sum_{k \in K_j} (c'_{jk} + c'_{kj}) = \sum_{k \in K_j} (c'_{ik} + c'_{ki}) = S_j$$

$$\sum_{k \in K_i} (c'_{jk} + c'_{kj}) = \sum_{k \in K_i} (c'_{ik} + c'_{ki}) = S_i$$

On a donc $F(C,Y^3) - F(C,Y^2) = S_i - S_j + c'_{ij} + c'_{ji} \quad 2m$

puisque $S_i - S_j > 0$ par hypothèse.

Dans ce cas également Y^1 n'est pas non plus solution optimale. D'où le résultat.

En résumé Y^1 ne peut être solution optimale, ce qui contredit l'hypothèse $y'_{ij} = 0$; d'où le résultat annoncé.

Remarque:
Les résultats présentés ici ne font aucune hypothèse, excepté la transitivité, quant à la nature des tableaux de données individuels.

Si par contre les tableaux de données individuels sont de même nature que le tableau collectif (préordres pour le cas 1) (et 3)) , ordres pour le cas 2) (et 3)) et partitions pour le cas 4), on retrouve des résultats connus donnés par J.P.Barthélemy (2) , J.Feldman (5) , B.Monjardet (10) et S.Régnier (11).

5.3. REPRESENTATION DE Y SOUS FORME DE CONTRAINTES LINEAIRES

Nous avons vu précédemment que la relation transitive cherchée Y était définie par un tableau binaire de terme général y_{ij}. Mais ces valeurs y ne sont pas arbitraires et suivant la signification donnée à Y (Relations de préordre, d'ordre, de quasi-ordre ou d'équivalence) les valeurs y_{ij} devront vérifier un certain nombre de propriétés (transitivité, antisymétrie, symétrie, totalité, intermédiarité, cohérence).

Ces propriétés ont été définies en (4-1-1) pour les tableaux C^k et s'appliquent bien entendu au tableau Y (substituer y_{ij} à c^k_{ij} dans les formules définissant les propriétés).
Comme nous l'avons vu précédemment chaque propriété correspond à un système d'égalités ou d'inégalités linéaires. Dans le cas du tableau Y les égalités ou inégalités seront linéaires par rapport aux y_{ij}.

Donnons quelques exemples:

a) Si Y est une relation d'ordre total, Y devra vérifier le système suivant

$$y_{ij} + y_{jk} - y_{ik} \leq 1 \qquad \forall \; (i\neq j, j\neq k, i\neq k)$$

$$y_{ij} + y_{ji} = 1 \quad \text{pour } i\neq j$$

$$y_{ij} = 0 \text{ ou } 1$$

b) Si Y est une relation d'équivalence , Y devra vérifier le système suivant:

$$y_{ij} + y_{jk} - y_{ik} \leq 1 \qquad \forall \; (i\neq j, j\neq k, i\neq k)$$

$$y_{ij} - y_{ji} = 0 \quad \text{pour } i\neq j$$

$$y_{ij} = 0 \text{ ou } 1$$

6. LES MODELES ASSOCIES

6.1. MODELES BIVALENTS

Notre problème, consistant à trouver une relation collective maximisant le nombre total d'accords moins le nombre total de non-accords, peut être représenté par un programme linéaire à variables bivalentes.
Ce programme est défini de la façon suivante:
Les variables 0-1 sont les valeurs y_{ij} de la relation Y.
La fonction économique à maximiser représente le nombre total d'accords moins le nombre total de non-accords.
Les contraintes représentent les différentes propriétés que doivent vérifier la relation Y. Il s'agit de la transitivité et selon le cas des propriétés d'antisymétrie, de symétrie , de totalité, d'intermédiarité ou de cohérence.
La fonction économique $\Sigma\Sigma\, c'_{ij}\, y_{ij}$ est linéaire par rapport aux variables y_{ij} ainsi que les contraintes représentant les différentes propriétés de la relation Y. Il s'agit donc bien d'un programme linéaire.

En fonction de la relation Y recherchée on obtient les modèles suivants:

a) Modèle pour une relation d'ordre total:

$$\text{Max} \sum_i \sum_{j\neq i} c'_{ij}\, y_{ij}$$

$$y_{ij} + y_{ji} = 1 \qquad i\neq j$$

$$y_{ij} + y_{jk} - y_{ik} \leq 1 \qquad i\neq j, j\neq k, i\neq k$$

$$y_{ij} = 0 \text{ ou } 1$$

Ce modèle a n(n-1) relations de totalité et antisymétrie, et n(n-1)(n-2) relations de transitivité.
Ce modèle peut être réduit; division par 2 du nombre de variables et approximativement par 3 du nombre de contraintes. (Voir (7) p 139)

Voici quelques propriétés de ce modèle:

1) Ce problème est équivalent dans le cas où Y est un ordre total à la maximisation du nombre d'accords. En effet $\Sigma\Sigma\, c'_{ij}\, y_{ij}$ est égal à $2\, \Sigma\, \Sigma\, c_{ij}\, y_{ij} - m\, \Sigma\, \Sigma\, y_{ij}$;et le terme $m\, \Sigma\, \Sigma\, y_{ij} = mn(n-1)/2$ puisque Y est un ordre total.

Ce problème est équivalent également à la minimisation du nombre de désaccords ($\Sigma\, \Sigma\, c_{ji}\, y_{ij}$), soit enfin à la maximisation du nombre d'accords moins le nombre de désaccords ($\sum_i \sum_{j\neq i} (c_{ij} - c_{ji})y_{ij}$). (Voir (7) p 137 ,où la notion de désaccord est définie).

2) Chacune de ces fonctions économiques peut être divisée par la quantité ($\Sigma\Sigma\, c_{ij}$); ce qui donne des fonctions économiques normalisées équivalentes à la fonction économique originale; on retrouve ici les notions de "coefficient d'adhésion" introduit par E.Jacquet-Lagrèze, de "taux" d'accords ou de désaccords etc... (voir (7) p 178).

3) Dans le cas où les tableaux C^k ne représentent pas des ordres , même le cas d'un unique tableau de comparaisons est intéressant, c'est le problème classique de la théorie des graphes où l'on recherche un ordre à distance minimale d'une relation binaire, d'un tournoi par exemple.

b) Modèle dans le cas d'une relation d'équivalence

$$\text{Max} \quad \sum_{i}\sum_{j\neq i} c'_{ij}\, y_{ij}$$

$$y_{ij} - y_{ji} = 0 \qquad i\neq j$$

$$y_{ij} + y_{jk} - y_{ik} \leq 1 \qquad i\neq j, j\neq k, k\neq i$$

$$y_{ij} = 0 \text{ ou } 1$$

Ce modèle a n(n-1) relations de symétrie et n(n-1)(n-2) relations de transitivité. Comme dans le cas précédent le modèle peut être réduit, division par 2 du nombre des variables et division par 2 du nombre des contraintes, (voir (8) p 45).

Indiquons succinctement quelques propriétés de ce modèle :

1) Contrairement au cas où Y est un ordre total, ce problème n'est pas équivalent à la maximisation du nombre total d'accords

$$\Sigma\,\Sigma\, c_{ij}\, y_{ij}$$

ou à la minimisation du nombre total de non-accords

$$\Sigma\,\Sigma\,(m - c_{ij})\, y_{ij}$$

bien que la fonction économique $\Sigma\,\Sigma\, c'_{ij} . y_{ij}$ soit par définition la différence de ces deux quantités.

Remarquons que lorsque Y est une partition , le problème de la maximisation du nombre d'accords a une solution évidente.

Il s'agit de la partition "grossière" où tous les éléments sont équivalents. Il en est de même du problème de la minimisation du nombre de non-accords qui lui aussi a une solution évidente. Il s'agit cette fois de la partition "triviale" où chaque élément forme une classe distincte.

Le critère de la partition centrale $c'_{ij} = c_{ij} + (c_{ij} - m)$ correspond à une fois le critère d'obtention de la partition grossière plus une fois le critère d'obtention de la partition triviale.
Notons (1,1) les coefficients de compromis linéaire pour la partition centrale . Par extension , on peut définir des partitions (a,b) centrales avec $c'_{ij} = a\, c_{ij} + b\, (c_{ij} - m)$ pour $a + b = 2$.

2) Les tableaux de similarités C^k peuvent représenter des partitions, dans ce cas particulier le problème est connu sous le nom du problème de la "partition centrale" posé par S. Régnier (voir (11)).

3) Lorsque les tableaux de similarités C^k ne représentent pas des partitions , même le cas d'un simple tableau de similarités est intéressant.C'est le problème de théorie des graphes où une relation binaire (le tableau C^k) est approximée par une relation d'équivalence (la relation Y).
Un cas particulier de ce problème est le problème de Zahn, où une relation symétrique est approximée par une relation d'équivalence.

c) Modèle pour une relation de quasi-ordre

$$\text{Max} \sum_i \sum_{j \neq i} c'_{ij}\, y_{ij}$$

$$y_{ij} + y_{ji} \leq 1 \qquad i \neq j$$

$$y_{ij} + y_{jk} - y_{ik} \leq 1 \qquad i,j,k \text{ différents}$$

$$y_{ij} - (y_{jk} + y_{kj}) + y_{k\ell} - y_{i\ell} \leq 1 \qquad i,j,k,\ell \text{ différents}$$

$$y_{ij} + y_{jk} - (y_{i\ell} + y_{\ell i}) - (y_{ik} + y_{ki}) \leq 1 \qquad i,j,k,\ell \text{ différents}$$

Ce modèle a $n(n-1)$ variables y_{ij} et un nombre de contraintes de l'ordre de n^4 .

d) Autres modèles

De la même façon on pourra tout aussi bien définir les modèles où la relation collective Y représente un préordre partiel, un préordre total, un ordre partiel, un ordre d'intervalles.

Contrairement aux cas de l'ordre total ou de l'équivalence, le nombre de variables ou de contraintes ne peut être réduit pour les autres modèles. Le nombre des variables reste égal à $n(n-1)$
Le nombre de contraintes est de l'ordre de n^3 (du aux relations de transitivité) sauf pour l'ordre d'intervalles où le nombre de contraintes est de l'ordre de n^4 (du aux relations d'intermédiarités).

6.2. PROGRAMMES LINEAIRES ASSOCIES ET RESOLUTION

Si nous remplaçons les conditions $y_{ij} = 0$ ou 1, par des conditions de variation continue $0 \leq y_{ij} \leq 1 \quad \forall \ i \neq j$ nous obtenons, à la place des modèles bivalents précédents, des programmes linéaires continus.
Ces programmes linéaires continus peuvent être résolus par programmation linéaire, par la méthode du simplexe par exemple.

Cependant, même la résolution du problème continu, sous sa forme initiale, parait difficile.
En effet la taille des bases, représentées par le nombre de contraintes, est de l'ordre de n^3 ou n^4 selon les cas; ce qui exclut toute possibilité pratique de résolution.
Par contre, la résolution du problème dual associé, (voir (7)), est possible; dans ce cas la taille des bases étant égale à $n(n-1)$ le nombre de variables y_{ij} du problème primal, il devient alors possible de résoudre des problèmes pour n relativement grand.

Disons simplement que d'un point de vue théorique, la résolution du problème dual donne la solution du problème primal continu.
Par contre le modèle bivalent et le modèle continu ne sont pas équivalents (voir (7)) et la possibilité d'obtenir une solution fractionnaire pour le programme linéaire associé existe. Mais ce phénomène est assez rare et ne semble pas se produire dans les problèmes pratiques.

Actuellement nous avons résolu par cette approche plus de 140 problèmes d'agrégation par un classement ou par une partition. Dans chaque cas, pour le problème continu, nous avons obtenu une solution optimale bivalente, donc solution optimale du problème bivalent. On trouvera quelques uns de ces exemples détaillés dans (1), (7), (9).

6.3. AUTRES MODELES

Lorsque l'on combine des tableaux individuels C^k on obtient un tableau collectif C.
Mais ce tableau ne conserve pas forcément les propriétés initiales que pouvaient avoir les tableaux C^k ; par exemple l'antisymétrie ou la transitivité ne sont pas nécessairement additives (effet Condorcet).

Il peut sembler intéressant d'approximer ce tableau collectif C par un tableau collectif inconnu Z. Ce tableau Z pouvant par exemple posséder les propriétés de transitivité, d'antisymétrie et de totalité collectives (définies en (4.2.1))(ordre total collectif). D'une façon générale on peut approximer le tableau collectif C par un tableau collectif Z vérifiant certaines des propriétés définies en (4.2.1).
Toutes ces propriétés qui représentent des contraintes sur le tableau Z sont des contraintes linéaires comme nous l'avons vu en (4.2.2). De même la fonction économique à minimiser est linéaire si elle est définie par:

$$\text{Min} \sum_i \sum_{j \neq i} |c_{ij} - z_{ij}|$$

qui se met sous la forme linéaire suivante:

$$\text{Min} \sum_i \sum_j u^+_{ij} + u^-_{ij}$$

avec

$$c_{ij} - z_{ij} = u^+_{ij} - u^-_{ij}$$

$$u_{ij} \geq 0$$

$$u_{ij} \geq 0$$

Les z_{ij} étant entiers.

La définition de ces problèmes sous forme de programmation linéaire en nombres entiers ne semble pas avoir été faite même dans le cas de la recherche d'un tableau collectif Z possédant la propriété de transitivité floue 7) donnée en (4.2.1). Par contre la résolution approchée de ce dernier problème est classique; elle est obtenue en recherchant la fermeture transitive de la relation collective C. Ce problème d'approximation se rencontre souvent en classification automatique sous le nom de recherche de l'ultramétrique sous-dominante lorsque le tableau C représente cette fois un tableau de dissimilarités.

7. REFERENCES

[1] AYACHE G., MARCOTORCHINO J.F., MICHAUD P. "Démocratie Française et la Presse, Techniques d'Analyse de la

Réception d'un Message", Revue Française de Communication, No 1, 30, (1978).

[2] BARTHELEMY J.P., "sur les Eloignements Symétriques et le Principe de Pareto", Mathématiques et Sciences Humaines, No 56, 97, (1976).

[3] BARTHELEMY J.P., MONJARDET B., "Ajustement et Résumé de Données Relationnelles: Les Relations Centrales", (Article présenté aux 2ème Journées Internationales d'Analyse des Données de l'IRIA, Versailles, (1979)).

[4] DIDAY E., Nouvelles Méthodes et Nouveaux Concepts en Classification Automatique et Reconnaissance des Formes, Thèse de Doctorat d'Etat, Université de Paris VI, (1972).

[5] FELDMAN J., "Pôles,Intermédiaires et Centres dans un Groupe d'Opinions", Mathématiques et Sciences Humaines, No 43, 39, (1973).

[6] JACQUET-LAGREZE E., La Modélisation des Préférences; Préordres, Quasi-Ordres et Relations Floues, Thèse de 3ème Cycle, Université Paris V, (1975).

[7] MARCOTORCHINO J.F., MICHAUD P., Optimisation en Analyse Ordinale des Données, Masson, Paris (1979).

[8] MICHAUD P., MARCOTORCHINO J.F., "Optimization in Ordinal Data Analysis", (article présenté pour une partie au Symposium International d'Optimisation Discrète, Vancouver (1977),) et disponible dans Rapport Technique No1,IBM France Scientific Centre, (1978).

[9] MICHAUD P., MARCOTORCHINO J.F., "Optimisation en Analyse des Données Relationnelles", (Article présenté aux 2ème Journées Internationales d'Analyse des Données de l'IRIA, Versailles (1979)).

[10] MONJARDET B., "Tournois et Ordres Médians pour une Opinion", Mathématiques et Sciences Humaines, No 43, 55, (1973).

[11] REGNIER S., "Sur quelques Aspects Mathématiques des Problèmes de Classification Automatique", I.C.C. Bulletin No 4, 175, (1965).

[12] "Modélisation des préférences et quasi-ordres", Math. Sci. hum., n°62 et 63 (1978).