APPLIED STOCHASTIC MODELS AND DATA ANALYSIS, VOL. 4, 273–282 (1988)

Clustering and classification

CLUSTERWISE AGGREGATION OF RELATIONS

WOLFGANG GAUL

Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe, Kaiserstr. 12, D-7500 Karlsruhe 1, West Germany

AND

MARTIN SCHADER

Institut für Informatik, Fachbereich WOW, Universität der Bundeswehr Hamburg, Holstenhofweg 85, D-2000 Hamburg 70, West Germany

SUMMARY

In this paper we handle the general problem of finding $q(>1)$ central relations on a set of objects which best fit the information contained in a finite number of given relations on that set. The proposed CAR (clusterwise aggregation of relations) algorithm allows one to consider the well-known situation of determining a single central relation as a special case ($q = 1$) and takes into account the fact that the representation of appropriately selected subsets of relations by different central relations can provide additional insights into whether different clusters or segments of relations exist in the given set of relations. Two examples demonstrate the usefulness of the suggested approach.

KEY WORDS Aggregation Cluster analysis Data analysis Distance Ordinal data Paired comparisons data Relations Social choice theory

INTRODUCTION

A familiar problem in data analysis and social choice theory is that of aggregating information contained in a finite number of relations on a set of objects into a single *central* relation on that set (see References 1 and 2 and the literature discussed there).

In order to find such a single central relation the usual approach is to define an appropriate distance on the set of (binary) relations on the set of objects—or a subset thereof—and to minimize the sum of distances between the given relations and the single central relation one is looking for. Depending on the type of result desired, this minimization has to be carried out with respect to certain constraints, e.g. the single central relation wanted has to fulfil conditions such as reflexivity, transitivity, symmetry and the like. For algorithmic aspects dealing with these problems the reader is referred to References 1–8; recent applications may be found in References 9–13.

In this paper we handle the more general problem of finding $q(>1)$ central relations which best fit the information contained in a finite number of given relations. The proposed CAR (clusterwise aggregation of relations) algorithm allows one to consider the above-mentioned situation of determining a single central relation as a special case ($q = 1$) and takes into account the fact that the representation of appropriately selected subsets of relations by different central

8755–0024/88/040273–10$05.00

Received 25 August 1987
Revised 7 June 1988

relations can provide additional insights as to whether different clusters or segments of relations exist in the given set of relations.

The next section gives an introduction into the problem of fitting q central relations to a given number of p relations and describes the CAR algorithm. The third section gives two examples which demonstrate the usefulness of the suggested approach. Finally, some concluding remarks are given.

A MODEL FOR CLUSTERWISE AGGREGATION OF RELATIONS

Let $I = \{1, \ldots, n\}$ be the set of objects under consideration, and denote by $R_1, \ldots, R_p$ the given relations on I where $M = \{1, \ldots, p\}$ describes the related index set. Different situations can be handled within this framework.

If M is a set of judging subjects then $R_1, \ldots, R_p$ could be individual relations which result from a *paired comparisons* experiment with respect to the elements of I.

Another possibility would be that $R_1, \ldots, R_p$ are obtained from complete orders or preorders —rankings, so to speak—on the elements of I.

On the other hand $R_1, \ldots, R_p$ might be derived from a *mixed data matrix*

$$\mathbf{A} = (a_{im})_{i \in I, m \in M}$$

where a_{im} is the value of variable m on object i. Here M denotes a set of variables used to describe the elements of I. In this case the relations $R_1, \ldots, R_p$ are usually defined by

$$\left.\begin{array}{ll} iR_m j :\Leftrightarrow a_{im} = a_{jm} & \text{for a nominal variable } m \\ iR_m j :\Leftrightarrow a_{im} \leqslant a_{jm} & \text{for an ordinal or a cardinal variable } m \end{array}\right\} (i, j \in I)$$

where R_m is an equivalence relation or a complete preorder on I.

Once relations have been established the next task is to compute a suitable distance function which assigns appropriate (dis)similarity values to pairs of relations. Different possibilities are known, e.g., one can define the graph of relation R by $G_R := \{\{(i, j) : i, j \in I \text{ and } iRj\}$ and—for two relations R, S—use the well-known distance function $d(R, S) := |G_R \cup G_S| - |G_R \cap G_S|$ (see References 14 or 15, for instance, for other distance functions).

In order to find a central relation S on I defining a classification (which implies that S is an equivalence relation) or a ranking (which implies that S is a complete order or preorder relation) of the objects the conventional approach is to compute the distance d and to minimize the sum $\Sigma_{m=1}^{p} d(R_m, S)$ under the constraints ensuring that S is of desired type.

Although obtaining exactly one resulting relation S may be imperative in some applications (e.g. in voting procedures) we feel that—in order to obtain a more profound description of the data—it will make sense to follow the suggestion of Lemaire[16] and to divide the set $M = \{1, \ldots, p\}$ into q clusters $M_1, \ldots, M_q$.

We therefore minimize the objective function

$$\sum_{j=1}^{q} \sum_{m \in M_j} d(R_m, S_j) \tag{1}$$

subject to the constraints that $\{M_1, \ldots, M_q\}$ is a partition of M and $S_1, \ldots, S_q$ are relations (on I) of some specified type(s). (An approach related to this problem which uses a least-squares deviation criterion in order to decompose a given similarity matrix into binary matrices, is described by Mirkin;[17] we also refer to the 'likelihood of links' method outlined by Lerman[18] which allows the hierarchical classification of descriptive variables or relations.)

Even for moderate numbers n, p and q minimization of (1) under the given restrictions is a discrete and combinatorial problem, the solution of which will not be possible in closed form, and, since in this context the relations S_j take on the role of cluster centers, we suggest the following CAR (clusterwise aggregation of relations) algorithm:

Set $t = 0$.
Choose $t_{\max} > 0$ and a starting partition $\{M_1^{(0)}, \ldots, M_q^{(0)}\}$.
Repeat

Calculate the corresponding central relations $S_1^{(t)}, \ldots, S_q^{(t)}$ according to

$$\sum_{m \in M_j^{(t)}} d(R_m, S_j^{(t)}) = \min_S \sum_{m \in M_j^{(t)}} d(R_m, S)$$

Calculate a minimal distance partition $\{M_1^{(t+1)}, \ldots, M_q^{(t+1)}\}$ using $m \in M_j^{(t+1)} :\Leftrightarrow d(R_m, S_j^{(t)}) = \min\{d(R_m, S_k^{(t)}) : k = 1, \ldots, q\}$. If this minimum is not unique take j to be the smallest k for which it is attained. Set $t = t + 1$.

until no more changes occur or $t > t_{\max}$.

The resulting partition is called an *end partition*. It is easy to see that during the iteration a monotonic reduction of the objective function in (1) is achieved. The end partition and the central relations, however, need not establish an optimal solution of (1). Furthermore, during the second iterative step one or more empty classes $M_j^{(t+1)}$ may be generated. These will remain empty until the end of the iterative process so that the end partition may have less than q classes. It should also be pointed out that the classical problem of minimizing $\Sigma_m\, d(R_m, S)$ must be solved q times during each performance of the first iterative step.

Obviously, the CAR algorithm belongs to the family of *wandering centroids*—or *nuées dynamiques*—methods which were stimulated among others by Sebestyen,[19] Régnier,[20] Gower[21] and Diday.[22] Its computational expenditure is dependent on the algorithm used to determine the solution of the $\min\{\Sigma_m\, d(R_m, S) : S\}$ problems. Various types of algorithms, such as branch and bound, cutting plane, dynamic programming or subgradient methods, and even powerful heuristics are available (see the references in the introduction).

Let us finally remark that, instead of tackling problem (1)—minimizing distances—it would be equally reasonable to maximize the association between relations (see Reference 23 for criteria of this type).

EXAMPLES

Two data sets known from previous research efforts to analyse structures on sets of objects are used for the illustration of the CAR algorithm.

An application to paired comparisons data

The first example is concerned with evaluations of what makes an advertising message for print ads of products a successful one.[24] A total of $p = 69$ persons participating in courses of continued education at the Chamber of Industry and Commerce of Karlsruhe and students of an introductory course of marketing at the University of Karlsruhe took part in the study. The main part of the survey consisted of a paired comparisons experiment of ten cognac print ads

Table I. The result of the paired comparison experiment

	111 / 234 . . . . . . 9 / 10
1	111111110000000100010010000010111110110010100
2	111000100010100001011000010000111100000000000
3	101101111000001101101111000110111111111000100
.	100000010000000001011110010010101111010011100
.	111011111000000101011110001010111101110000100
.	101111010000010001111110000100110111100000110
	111010101001100111111101000100111011100101000
	010000000110111100000000010100111110100000000
	111011110010100100010110010010111110000110100
	100010000111101110010000000001101110000111101
	101011010000000001011110001000111111000000100
	111111110000010000001110001010111111110000100
	101111001000000001111001000000110110000000111
	110111101000101000101111111101101010000100001
	100011000101110010011000111000110010000000111
	101010000000000001111101110000110000000000011
	101011110000010001001010001010111111010010100
	110111111000010000101000101111011011111000001
	100000000011100101110011010010100110000011110
	100110001001101110110101110101000010001101011
	110111111000000000011111111111111101111100100
	100111110001011100000110011000110111110010100
	110101001101010000111001111111100000000000011
	111001100011011110000000000000111011111001100
	101010000010100101010110000000110110110010110
	010000000101101010000000111111100000000001011
	100101011101110111111111111011110100011011110
	110011101101001000000100111101111010100101000
	011010011111111111000010110111001010111011001
	111011111000010000001000111111010001111000111
	111111111011101100100110000010100010111110100
	101010110000000000011110000100111110110010100
	001010000010010001011010010000110110000000111
	101010111000000001011010000100011100110111000
	110110111011101100000100100100101010111010110
	100010010010100101111111010010100100000011110
	100010010000000001110111001010101100110010100
	100010110000000001010111000010101100010011110
	100010010000100101011110010000111110010110100
	000000000000000000101101111101100011111000111
	010011100100111010000000111111111010000010100
	110111011000000000011000111010110111010010110
	001100000011100011110001000000010010001011111
	101010010010100101011110010000111110010010100
	101010010000000001111110000010111111010110100
	100001100100000000001100101101011011101100001
	100101110000000101001110101100011111111010110
	110111101000000000110100101100110010101000011
	101111111000000000111101110001000110001001011
	000001000000000001111111111001110011100000111
	111101111100001011000011011010010111111000001
	101000010000000001011011001010110101010000110
	111111111011101101000110000110101110010110100
	100010000010100101110110010000101110000010100
	101000010010000000000110000000101110110111100
	101010010010000101011110000010110110010010100
	111011110000011000001110001110111101010000100
	101111111000001001101110001000110111110000100
	101011111000010000011110011110110111110000100
	000000000011111111010000111111111010001100001
	000010000110100000011100110100111110000000001
	110110101111101010100100111101000010101101011
	001010000011111111110111010000100010000111111
	010000000111111111010000101001001110010010110
.	110111100000000000110001111100111001000000001
.	101011110010010101001110001010111111110110100
.	000011011000010000011010011010110101010000110
68	101011111000000101101111000010101110110010100
69	101010010010100101011110010010111100000110100

in which pairs of ads were shown to participants for three seconds each. The participants were asked to select that print ad which—according to the question 'Which of the two ads is more appealing to you'—was preferable. For each participant his/her individual paired comparisons matrix was recorded.

Here, $M = \{1, \ldots, 69\}$ is the set of participants and $I = \{1, \ldots, 10\}$ is the set of print ads. Each paired comparisons matrix determines a relation on I. The relations $R_1, \ldots, R_{69}$ are given in Table I. The columns in this table are numbered according to the pairs of elements $(1,2), (1,3), \ldots, (1,10), (2,3), (2,4), \ldots, (2,10), (3,4), (3,5), \ldots, (3,10), \ldots, (9,10)$; a 1 in row m and column (i,j) indicates that participant m has preferred ad i to ad j.

For a better understanding of the grand matrix in Table I we represent R_1, the preferences of participant 1, in Table II. In this 10×10 matrix a 1 in row i and column j indicates that ad i was preferred to ad j.

Of course, an interesting question is whether different clusters within the sample of 69 respondents react in different ways to the advertising exposures. To obtain an answer we applied the CAR algorithm where we used the subgradient-method (O) described by Schader and Tüshaus[25] to solve the min $\{\Sigma_m d(R_m, S) : S\}$ problems. The best values of the objective function (1) subject to the condition that $S_1, \ldots, S_q$ are complete orders are given in Table III for $q = 1, \ldots, 6$. The results were obtained using 1000 random starting partitions for each value of q.

The greatest improvement was reached for the segmentation of the sample set M into two classes. After a partition into four classes only minor changes—compared to the first improvements—of the objective function follow. Additionally, for marketing reasons a split into too many classes would cause problems in handling separate advertising campaigns for all the different target audiences. From the two-class solution we obtained $M_1 = \{1, 3, \ldots, 6, 9, \ldots, 13, 17, 19, 22, 24, 25, 27, 31, \ldots, 34, 36, \ldots, 39, 42, \ldots, 45, 47, 52, \ldots, 59, 63, 64, 66, \ldots, 69\}$, $M_2 = M - M_1$ and the complete orders

$$S_1 : 5 < 10 < 3 < 8 < 1 < 7 < 6 < 4 < 2 < 9$$

$$S_2 : 9 < 4 < 5 < 1 < 10 < 3 < 8 < 7 < 6 < 2$$

Keeping the obtained partition $\{M_1, M_2\}$ of M fixed, these orders are optimal with respect to (1). Optimality can be proved by controlling the gap between the objective function values of the primal and dual problems corresponding to min $\{\Sigma_m d(R_m, S) : S\}$ as is proposed in References 5 and 6.

Figure 1 gives the results of this two-segment case in terms of a probabilistic ideal point model representation (see Reference 24 for a detailed description).

Table II. The preferences of participant 1

	1	2	3	4	5	6	7	8	9	10
1	-	1	1	1	1	1	1	1	1	0
2	0	-	0	0	0	0	0	0	1	0
3	0	1	-	0	0	1	0	0	1	0
4	0	1	1	-	0	0	0	0	1	0
5	0	1	1	1	-	1	1	1	1	1
6	0	1	0	1	0	-	0	1	1	0
7	0	1	1	1	0	1	-	0	1	0
8	0	1	1	1	0	0	1	-	1	0
9	0	0	0	0	0	0	0	0	-	0
10	1	1	1	1	0	1	1	1	1	-

Table III. Best values of the objective function

q	1	2	3	4	5	6
$\sum_{j=1}^{q} \sum_{m \in M_j} d(R_m, S_j)$	2266	1830	1640	1496	1414	1348

Again, one can see that there is a clear distinction relative to the preference orderings of the two subgroups M_1 and M_2 where in a (probabilistic) ideal point model orderings are expressed by the distances of the different print ads from the (expected) ideal points $I1$ and $I2$. (A mathematical description of the probabilistic ideal point model and comparisons to other models for evaluating paired comparisons data is given in Reference 26. In our context, Figure 1 is inserted to indicate that on the basis of the CAR results M_1, M_2 other techniques can be applied, the results of which can be compared to the optimal central relations S_1 and S_2 of the CAR approach.)

An application to ordinal data

This second example analyses data describing the management organization in business firms having headquarters and several subordinate bodies (classified, for instance, according to regions or products). A total of $n = 36$ top-level managers were asked to report on $p = 37$ statements intended to describe the headquarters' business interactions with its subdivisions (s.d.s). The data of this survey, first communicated by Gabele and Niemeyer[27] have been studied by Dobbener[28] and Schader and Tüshaus[25] among others. Here, $M = \{1, \ldots, 37\}$ is the set of statements and $I = \{1, \ldots, 36\}$ is the set of managers. Some examples of statements to be answered by the managers are:

We decide on manpower requirements and personnel disposition in the s.d.s.
We determine prices, delivery and payment conditions for the s.d.s.

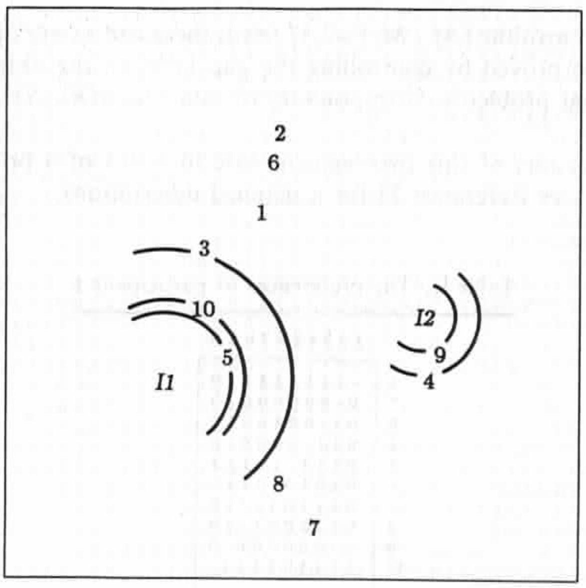

Figure 1. Graphical representation of probabilistic ideal points results

We project marketing mix and advertising strategy of the s.d.s.
We co-ordinate hiring and dismissal of executive personnel with the s.d.s.
...

For each of these statements the manager had to decide if it is 'absolutely incorrect', 'true to a minimum degree', ..., or 'absolutely correct' in reference to his organization. The data matrix is given in Table IV. Here, we encoded the seven possible answers by 1, 2, ..., 7 or 7, 6, ..., 1, respectively, so that small numbers indicate that responsibility is delegated to the s.d.s., whereas higher numbers show that planning decisions are passed in the headquarters.

In this case $R_1, \ldots, R_{37}$ are complete preorders on $I = \{1, \ldots, 36\}$. Using the same distance function d as in the previous example and searching for central complete preorders we obtained the following result for $q = 2$ classes:

$$M_1 = \{1, 2, \ldots, 8, 9, 12, 13, \ldots, 29, 30\} \quad \text{and} \quad M_2 = \{10, 11, 31, 32, \ldots, 37\}$$

are the subsets of statements which best could be represented by two complete preorders. It is easily possible to find reasons for this partition when considering the type of the corresponding statements. Whereas all the statements numbered 1, ..., 30 are of the type 'We decide ...', 'We determine ...', 'We project ...', statements 31, ..., 37 refer to an existing co-operation between headquarters and s.d.s., e.g.

Table IV. The data matrix **A**

	1 2 3 37
1	6 7 7 2 2 2 4 7 6 1 2 1 6 4 5 5 7 5 5 2 1 3 5 7 5 3 5 7 3 6 2 1 2 4 4 1 1
2	3 6 6 1 2 3 3 4 3 1 1 2 2 1 3 2 6 4 2 2 3 4 5 5 6 4 4 6 5 5 5 2 3 3 4 6 7
3	6 6 7 7 5 7 7 7 5 6 6 4 5 2 5 4 7 7 4 7 5 6 7 7 7 6 7 7 6 7 5 1 5 1 2 1 3
.	5 4 4 1 1 3 2 2 1 4 3 1 1 1 6 4 1 1 1 1 1 1 3 6 6 4 6 6 3 6 6 4 7 3 6 2 4
.	6 6 6 3 3 4 4 4 3 2 3 1 5 1 7 4 7 3 2 1 3 5 5 6 7 4 7 6 5 6 2 3 2 3 1 3 7
.	7 3 3 1 1 2 3 1 5 7 7 1 7 1 7 5 7 3 1 2 1 7 7 7 7 6 7 7 5 7 3 1 2 1 1 7 7
	6 6 6 3 6 7 7 5 5 2 2 6 5 3 6 6 6 4 3 3 4 5 7 6 6 6 6 6 6 7 1 2 2 2 1 2 3
	4 7 7 2 4 4 4 3 1 6 6 5 4 7 5 7 5 1 2 1 1 3 3 6 6 3 4 6 5 6 4 5 4 3 1 1 6
	6 2 6 2 5 3 5 5 5 5 4 2 4 2 6 3 6 4 3 1 2 2 4 6 4 6 3 6 4 7 2 7 7 7 1 6 6
	5 2 3 5 2 2 2 5 3 2 2 1 2 2 2 1 4 2 2 2 2 2 2 4 3 4 5 4 1 4 3 4 3 7 7 7 6
	5 6 3 7 2 1 1 3 1 7 7 1 3 7 7 2 3 2 3 1 3 4 5 5 7 1 2 6 2 5 4 6 7 3 1 1 3
	5 7 1 5 5 6 5 4 1 6 6 1 1 4 3 5 2 5 5 4 4 3 2 2 5 5 5 5 4 5 4 1 4 6 4 6 4
	7 5 4 1 1 1 1 1 1 1 1 1 5 4 4 1 6 1 1 1 1 1 4 7 7 3 7 7 4 7 2 1 1 1 4 4 7
	3 6 6 4 4 4 5 5 5 4 3 5 4 5 6 5 5 4 4 4 3 4 5 5 6 5 5 5 6 6 2 2 2 2 4 5 4
	6 1 1 3 1 2 1 4 2 7 2 1 2 1 4 3 5 6 1 1 2 4 6 5 6 5 6 7 2 6 4 1 4 3 7 7 7
	4 7 7 3 1 1 3 1 1 3 3 1 4 1 7 7 2 2 6 1 4 4 4 7 7 4 6 5 4 6 5 6 5 6 1 1 1
	7 5 5 5 6 4 7 7 4 3 2 5 7 5 7 4 7 5 5 3 4 6 7 7 7 5 6 6 7 6 2 1 5 2 1 1 5
	7 4 3 4 7 5 7 5 6 6 4 7 7 3 7 4 6 6 3 3 4 5 7 7 7 5 6 6 3 6 1 1 2 1 2 3 7
	2 2 3 4 5 5 5 3 1 1 1 2 3 5 4 2 2 5 1 4 1 2 4 5 6 4 4 4 5 4 2 4 2 3 4 3 5
	4 2 5 4 4 3 4 5 2 6 6 4 3 6 5 7 3 2 2 6 2 3 4 3 6 4 6 6 7 6 3 6 3 2 2 6 2
	5 5 1 4 3 3 4 4 3 2 2 3 4 4 6 3 5 4 4 3 3 5 7 6 2 4 5 5 5 5 4 4 3 4 3 6 6
	6 6 7 7 7 4 5 7 4 5 5 6 7 7 6 6 4 5 5 7 6 6 6 7 7 7 6 6 5 7 1 1 1 1 2 1 1
	5 7 7 7 6 4 5 6 4 5 4 3 7 5 6 2 7 6 4 6 3 5 6 7 5 7 7 7 7 7 2 1 1 1 2 1 3
	6 6 7 2 1 1 2 7 3 4 3 1 1 7 7 1 6 1 1 1 3 3 2 7 5 6 7 7 2 7 2 2 7 3 1 3 1
	3 7 7 2 2 4 2 4 4 1 1 2 4 4 4 4 6 7 1 4 2 3 4 7 7 4 7 7 5 5 5 1 1 4 1 7 4
	6 3 5 2 4 5 5 5 3 2 2 1 6 6 7 6 6 5 1 4 1 4 5 6 6 6 6 7 7 7 2 3 2 1 1 1 1
	3 2 3 3 1 1 1 2 1 1 4 1 3 2 5 4 3 1 1 1 1 3 3 7 7 5 5 7 4 7 7 6 4 6 1 1 1
	7 4 6 4 6 5 7 7 6 2 2 6 6 1 7 6 7 5 6 4 3 5 7 7 6 5 5 7 1 6 1 1 2 3 1 1 6
	3 7 7 1 3 1 2 4 6 7 2 5 6 5 7 1 4 4 1 5 1 2 3 7 7 5 7 7 5 7 3 1 4 3 1 7 5
	6 1 5 2 2 6 2 1 2 1 1 2 6 7 4 1 7 2 3 4 2 4 7 6 2 4 6 6 4 5 2 1 1 2 6 3 2
	2 7 7 1 1 2 2 4 3 7 7 1 3 4 5 2 4 3 1 4 1 3 5 6 7 4 5 7 5 6 7 3 5 2 1 1 5
.	4 6 7 1 2 2 3 2 2 2 1 2 2 2 2 5 5 3 2 1 1 2 4 6 6 3 5 6 6 6 5 2 2 4 6 2 3
.	3 7 7 7 3 3 3 6 4 3 2 2 4 7 4 2 3 3 5 2 2 3 3 6 2 4 4 4 6 6 5 1 3 6 2 2 2
.	5 3 6 5 2 2 1 2 4 7 7 2 5 3 5 1 6 4 2 5 6 6 6 7 7 5 5 6 5 5 3 2 3 3 2 6 3
35	6 3 3 1 5 5 7 4 2 7 4 1 6 1 7 7 7 4 1 1 1 5 7 7 7 5 4 6 4 7 5 5 1 1 1 1 1
36	6 6 7 7 5 5 4 4 4 5 5 3 4 1 7 6 7 5 3 3 2 3 7 7 5 7 7 6 1 6 2 1 1 6 1 1 6

31. The marketing strategy for the s.d.s. is determined in co-operation with the s.d.s.
32. We co-ordinate hiring and dismissal of executive personnel with the s.d.s.
33. We arrange the assignment of new projects to our s.d.s. together with the s.d.s.
34. Differences between s.d.s. are settled in joint meetings in the headquarters.
35. Costing and accounting axioms for the s.d.s. are determined in co-operation with the s.d.s.
36. Our overall budget is planned together with the s.d.s.
37. Apportionment of our general expenses is co-ordinated with the s.d.s.

Associated with the cluster M_1 is the complete preorder $S_1 : C_1 < C_2 < C_3 < C_4 < C_5$ with the classes $C_1 = \{4, 10, 11, 13, 15, 19, 27, 32\}$, $C_2 = \{1, 2, 5, 6, 8, 9, 12, 16, 20, 21, 24, 25, 29, 30, 31, 33, 34, 35\}$, $C_3 = \{14\}$, $C_4 = \{26\}$ and $C_5 = \{3, 7, 17, 18, 22, 23, 28, 36\}$.

In Figure 2 we plotted the median of the judgements of the statements in M_1 given by the members of the classes C_1, C_2 and $C_3 \cup C_4 \cup C_5$, respectively (the single-element classes C_3 and C_4 were united with C_5 for reasons of clarity).

We see that the classes are ordered according to increasing dependence of the s.d.s. on the firm's headquarters — there is only one variable with an inversion of this ordering.

A similar picture is given by the complete preorder S_2 for the cluster M_2. S_2 is defined by $C_1' < C_2' < C_3' < C_4' < C_5'$ with $C_1' = \{1, 7, 13, 17, 22, 23, 25, 26, 28, 30, 35, 36\}$, $C_2' = \{5, 14, 18, 19, 24, 32, 33\}$, $C_3' = \{2, 3, 6, 8, 10, 11, 12, 15, 16, 20, 21, 27, 29, 31, 34\}$, $C_4' = \{4\}$ and $C_5' = \{9\}$.

Again we plotted the median of C_1', C_2' and $C_3' \cup C_4' \cup C_5'$. The results are shown in Figure 3. It should be noted that—as in the first example—S_1 and S_2 are optimal with respect to (1) and that we solved the $\min\{\Sigma_m d(R_m, S) : S\}$ problems using the subgradient method (P) from Reference 6.

To give a short résumé: our results indicate that—according to managers' information—those firms which have a rigid management are also characterized by intensive consultation and co-operation with their subdivisions.

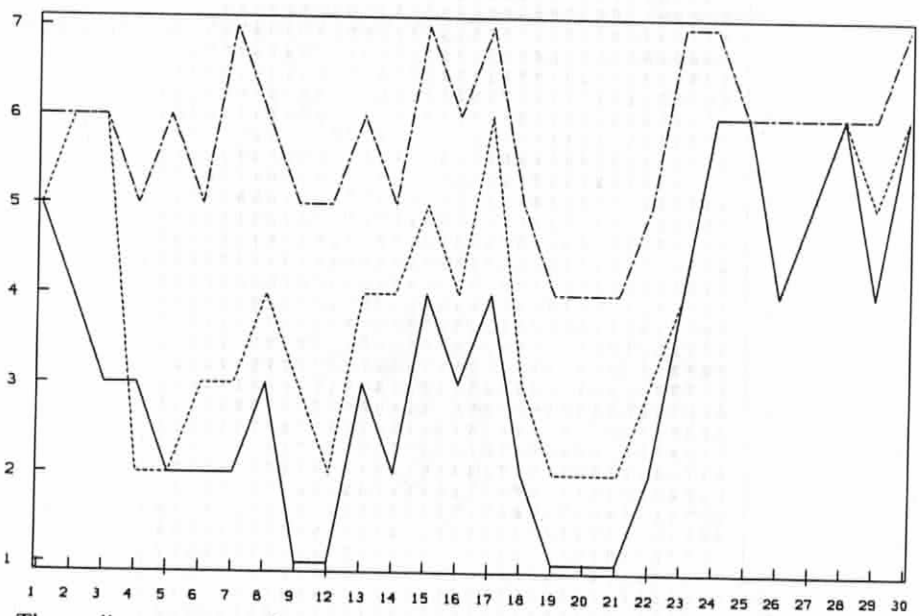

Figure 2. The median answers to the statements in M_1: ——— the firms in C_1; ----- the firms in C_2; —·—·—·— the firms in $C_3 \cup C_4 \cup C_5$

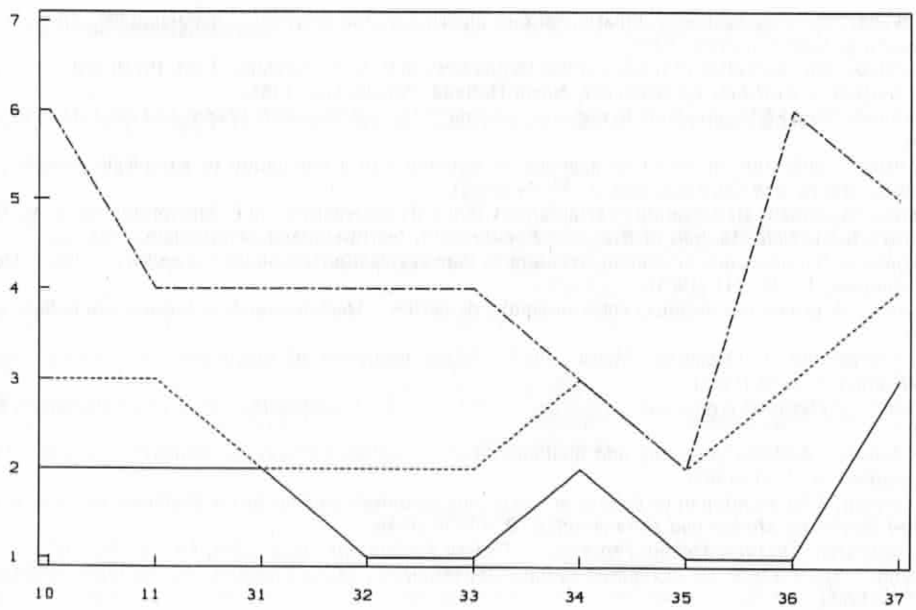

Figure 3. The median answers to the statements in M_2

CONCLUSION

On the basis of a given set of relations collected to provide information about an interesting set of objects the described CAR algorithm allows additional insights concerning the degree of homogeneity of the given data. Comparisons of the improvements of the objective function—obtained if the number of classes (of the set of given relations) is increased—allow the assessment of the number of central relations required to profoundly describe the essential information provided.

ACKNOWLEDGEMENT

We thank two anonymous referees for their careful scrutiny and helpful suggestions.

REFERENCES

1. J. P. Barthélemy and B. Monjardet, 'The median procedure in cluster analysis and social choice theory', *Mathematical Social Sciences*, **1**, 235–267 (1981).
2. J. P. Barthélemy and B. Monjardet, 'The median procedure in data analysis: new results and open problems', in H. H. Bock (ed.), *Classification and Related Methods of Data Analysis*, North-Holland, Amsterdam, 1988.
3. M. Grötschel and Y. Wakabayashi, 'A cutting plane algorithm for a clustering problem', *Report No. 9*, Institute of Mathematics, University of Augsburg, West Germany, 1987.
4. F. Marcotorchino and P. Michaud, *Optimisation en Analyse Ordinale des Données*, Masson, Paris, 1979.
5. P. Michaud, 'Agrégation à la majorité IV: un nouvel algorithme de résolution', *Technical Report No. 113*, Centre Scientifique IBM France, Paris, 1988.
6. M. Schader and U. Tüshaus, 'Subgradient methods for analyzing qualitative data', in W. Gaul and M. Schader (eds), *Classification as a Tool of Research*, North-Holland, Amsterdam, 1986.
7. M. Schader and U. Tüshaus, 'Analysis of qualitative data: a heuristic for finding a complete preorder', in H. H. Bock (ed.), *Classification and Related Methods of Data Analysis*, North-Holland, Amsterdam, 1988.

8. Y. Wakabayashi, 'Aggregation of binary relations: algorithmic and polyhedral investigations', *Thesis*, University of Augsburg, West Germany, 1986.
9. P. Michaud, 'The calibration of traffic-conflict techniques', in F. Marcotorchino, J.-M. Proth and J. Janssen (eds), *Data Analysis in Real Life Environment*, North-Holland, Amsterdam, 1985.
10. P. Michaud, 'Rankings aggregation in regional planning', *Applied Stochastic Models and Data Analysis*, **2**, 83–95 (1986).
11. P. Parisot, 'Application of similarity aggregation techniques to a population of paraplegic patients', *Applied Stochastic Models and Data Analysis*, **1**, 35–54 (1985).
12. G. Vescia, 'Automatic classification of cetaceans by similarity aggregation', in F. Marcotorchino, J.-M. Proth and J. Janssen (eds), *Data Analysis in Real Life Environment*, North-Holland, Amsterdam, 1985.
13. I. Warnesson, 'Optimization of semantic relations by data aggregation techniques', *Applied Stochastic Models and Data Analysis*, **1**, 121–141 (1985).
14. M. Besson, 'A propos des distances entre ensembles de parties', *Mathématiques et Sciences Humaines*, **11**, 17–35 (1973).
15. J. C. Gower and P. Legendre, 'Metric and Euclidean properties of dissimilarity coefficients', *Journal of Classification* **3**, 5–48 (1986).
16. J. Lemaire, 'Agrégation typologique de données de préférences', *Mathématiques et Sciences Humaines*, **58**, 31–50 (1977).
17. B. G. Mirkin, 'Additive clustering and qualitative factor analysis methods for similarity matrices', *Journal of Classification*, **4**, 7–31 (1987).
18. I. C. Lerman, 'Classification of medullary lesions among paraplegic patients by the likelihood of the links method', *Applied Stochastic Models and Data Analysis*, **2**, 13–30 (1986).
19. G. S. Sebestyen, *Decision-Making Processes in Pattern Recognition*, Macmillan, New York, 1962.
20. S. Régnier, 'Sur quelques aspects mathématiques des problèmes de classification automatique', *ICC Bulletin*, **4**, 175–191 (1965).
21. J. C. Gower, 'A comparison of some methods of cluster analysis', *Biometrics*, **23**, 623–637 (1967).
22. E. Diday, 'Une nouvelle méthode en classification automatique et reconnaissance des formes', *Revue de Statistique Appliquée*, **19**, 19–34 (1971).
23. F. Marcotorchino, 'Maximal association as a tool for classification', in W. Gaul and M. Schader (eds), *Classification as a Tool of Research*, North-Holland, Amsterdam, 1986.
24. W. Gaul and I. Böckenholt, 'Generalized Thurstonian scaling of advertising messages', *Working Paper*, Institute of Decision Theory and Operations Research, University of Karlsruhe (TH), 1987.
25. M. Schader and U. Tüshaus, 'Klassifikation ordinaler Daten—Eine Fallstudie', in I. Dahlberg and M. Schader (eds), *Automatisierung in der Klassifikation*, Indeks, Frankfurt, 1983.
26. I. Böckenholt and W. Gaul, 'Analysis of choice behaviour via probabilistic ideal point and vector models', *Applied Stochastic Models and Data Analysis*, **2**, 209–226 (1986).
27. E. Gabele and R. Niemeyer, 'Führungspraxis divisionalisierter Unternehmen—Einsatzbedingungen und Leistungsfähigkeit der Clusteranalyse in der empirischen Forschung', *Journal für Betriebswirtschaft*, **30**, 114–142 (1980).
28. R. Dobbener, 'Bewertung von Klassifikationsergebnissen im Hinblick auf natürliche Strukturen', Paper presented at the SIG-NK Meeting of the GfKl, Frankfurt, 1982.