CLASSIFICATION AS A TOOL OF RESEARCH
W. Gaul and M. Schader (Editors)

TWO-MODE HIERARCHICAL CLUSTERING AS AN INSTRUMENT FOR MARKETING RESEARCH

Espejo, E./Gaul, W.

Institute for Decision Theory and Operations Research
University of Karlsruhe, Karlsruhe, F.R.G.

Recently developed two-mode hierarchical clustering methodology can help to analyse square asymmetric and/or rectangular data matrices such as association matrices, brand-switching matrices, confusion-recognition matrices etc., which are often used to represent relevant marketing information. On the basis of data obtained from an experiment concerning measurement of the effectiveness of advertising messages different possibilities for the evaluation of the corresponding two-mode proximity matrices via hierarchical clustering methods are proposed and algorithms known from the literature are compared with an own development.

1. INTRODUCTION

Since the beginning of the sixties hierarchical clustering methods have been widely used to represent symmetric proximities between pairs of elements of a given set. The graphical representation of a hierarchical clustering solution in form of a dendrogram or rooted ultrametric tree has shown to be adequate in many cases and many algorithms have been proposed to generate such dendrogram solutions for the discovery of the relations between elements of a one-mode set.

In marketing research, however, square asymmetric and/or rectangular data matrices such as association matrices (describing e.g. relations between products and interesting properties of products), brand-switching matrices (revealing e.g. fluctuations in choice behavior), confusion-recognition matrices (showing e. g. misinterpretations of advertising messages), etc., are often collected and interpretations of relations between elements of different modes (the modes in the above marketing context are products and properties, first purchases and second purchases, advertising messages and advertised brands) are of interest.

Recently developed two-mode hierarchical clustering methodology can help to analyse this kind of data. The result of such an analysis is a dendrogram the structure of which reveals relationships, common characteristics, etc., for sets of elements mixed together from different modes. Furnas (1980) has stated conditions for the representability of two-mode proximity matrices in form of a dendrogram the terminal nodes of which are assigned to the elements of the sets of different modes, De Sarbo and De Soete (1984) and De Soete et. al. (1984) have proposed penalty approaches combining a least squares criterion with the penali-

zation of the violation of adequate constraints which force ultrametric (dendrogram) or additive tree representability.
Other models for analysing two-mode data have been developed, see e.g. the two-mode generalization PENCLUS (Both and Gaul (1985)) of the ADCLUS model (Shepard and Arabie (1979)) which is similar to GENNCLUS (De Sarbo (1982)) and allows for non-hierarchical overlapping two-mode clustering or the MDS- approach of Böckenholt (1985)) which both use the marketing example described in this paper for the presentation of the results obtained. From the other papers dealing with multi-mode multi-way proximity data only Carroll et. al. (1984), Harshman et. al. (1982) and Heiser and Meulman (1983) are explicitly mentioned because of space limitation, the interested reader will find further references in the literature cited in the papers referred to above.
In this paper the missing-values-aspect apparent in many problems in which certain data subsets are suspicious or even lacking is emphasized as every two-mode matrix can be viewed (after appropriate transformation) as part of a so-called grand matrix which describes the dissimilarity of the pairs of elements taken from the union of the sets of different modes. Here, missing data positions can be replaced by values derived from a priori information or from estimation procedures based on the available part of the data or can be just handled as missing values. This is, for instance, done by the MV (Missing Values) versions MVSL, MVCL, MVAL of the classical SL (Single Linkage), CL (Complete Linkage), AL (Average Linkage) algorithms which are described in the next section and are compared with algorithms known from the literature. The comparison is based on data obtained from an experiment concerning German cigarette brands advertising and shows that MVAL competes favourably with the other approaches.The paper concludes with some remarks for further research.

2. ALGORITHMIC ASPECTS OF TWO-MODE HIERARCHICAL CLUSTERING

Let be $P=(p_{ij})_{nxm}$ the given two-mode proximity matrix describing relations between the elements $i \varepsilon S_1=\{1,\ldots,n\}$ and $j \varepsilon S_2=\{1,\ldots,m\}$ where S_1, S_2 are the two sets of different modes with $S_1 \cap S_2= \emptyset$. After having converted P to a dissimilarity matrix $D=(d_{ij})_{nxm}$(if necessary) the problem consists in finding a so-called ultrametric grand matrix $G=(g_{i_1 i_2})_{(n+m)x(n+m)}$, that is a grand matrix which satisfies the ultrametric inequality

$$g_{i_1 i_2} \leq \max \{g_{i_1 k}, g_{i_2 k}\}, \; i_1,i_2,k \varepsilon \{1,\ldots,n+m\},$$

(a necessary and sufficient condition for representability of the relations given by $g_{i_1 i_2}$ via a dendrogram) and best reveals the relationships inherent in the given two-mode data P or D, respectively.

To solve the problem, one could proceed as follows:

(α) Starting from P one could view D as the known part of a grand matrix $\Delta=(\delta_{i_1 i_2})_{(n+m)x(n+m)}$ the missing values of which have to be optimally estimated before an ultrametric grand matrix G which best fits Δ is calculated via a one-mode hierarchical clustering procedure.

(β) Starting from P one could use D to build a matrix $T=(t_{ij})_{nxm}$ which satifies the two-class ultrametric inequality

$$t_{ij} \leq \max \{t_{ik}, t_{lk}, t_{lj}\} \ , \quad i,l \in S_1, \ j,k \in S_2,$$

(which according to Furnas (1980) is a necessary and sufficient condition to represent the relations between and within the sets S_1 and S_2 via a unique (up to the internal structure of one-mode subtrees) two-mode hierarchical clustering in form of a dendrogram. This procedure was adopted e.g. in the 2MLS (2 Modes Least Squares) approach of De Soete et. al. (1984).

(γ) Of course, starting from P one could view D as the known part of a grand matrix which needs no further specification because the missing-values-algorithms constructing the final ultrametric grand matrix G from D will surely change missing values to "best" fitted values and probably also change some of the matrix coefficients of D. This happens e.g. in the 1MLS (1 Mode Least Squares) approach of De Soete (1984) which allows for missing values and, obviously, can be applied to the grand-matrix-with-missing-values-case.

Although penalty approaches of constrained optimization (see e. g. Gaul and Hartung (1979), Rao (1978)) obtain an increasing importance within the data analysis area which allow to incorporate additional constraints of interest - as e. g. the (two class) ultrametric inequality - in data analysis problems one should also check whether classical well-known one-mode approaches could be adapted to the two-mode case.

Indeed, another way to handle the two-mode missing-values-problem consists in applying MV (Missing Values) versions of the well-known single, complete and average linkage algorithms to a grand matrix in which, initially, only the D positions are known. Here, in the ν-th step one has to update a relation $R^{(\nu)} \subset S_1 \cup S_2 \times S_1 \cup S_2$ (with $R^{(0)}=S_1 x S_2 \cup S_2 x S_1$) and $g^{(\nu)}_{i_1 i_2}$-values, $(i_1,i_2) \in R^{(\nu)}$ (with $g^{(0)}_{i_1 i_2} = d_{i_1 i_2}$, $(i_1,i_2) \in R^{(0)}$) according to the SL, CL and AL missing-values-rules to yield $R^{(\nu+1)}$ and $g^{(\nu+1)}$.

The algorithm stops when $G^{(\nu+1)}$ fulfils the ultrametric inequality. A more thorough description of these procedures is given elsewhere. Instead, Fig. 1 is inserted to help to clarify the different ways how an ultrametric grand matrix can be yielded via the (α), (β), (γ) - approaches.

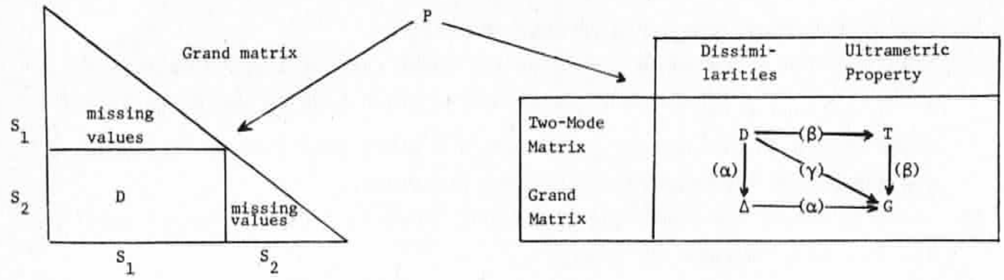

Fig. 1: How to get an ultrametric grand matrix from a two-mode proximity matrix P

3. EXAMPLE, RESULTS AND INTERPRETATION

On the basis of four two-mode proximity matrices obtained from an experiment concerning measurement of the effectiveness of advertising messages conducted with students from a marketing course at the University of Karlsruhe and described in more detail in Both and Gaul (1985) the just mentioned possibilities of two-mode hierarchical clustering have been applied and compared.

16 cigarette print ads of 8 cigarette brands (two ads for each brand) were used, for all print ads the brand names were masked. The judging subjects were randomly divided into two groups, for one of the groups, additionally, the slogans (if existing) on the print ads were masked, too.

13 properties (as sympathy, adventure, harmony, tradition, etc., see Fig. 2) and 12 cigarette brands (containing the 8 advertised brands and 4 brands for which no advertisements were shown, see Fig. 3 and Fig. 4) were used to build up two 16x13 association matrices (ads (with and without slogans) vs. properties) and two 16x12 confusion-recognition matrices (ads (with and without slogans) vs. brands). Tab. 1 gives the results of the application of the different two-mode hierarchical clustering algorithms mentioned above in terms of VAF (Variance Accounted For) and CCC (Cophenetic Correlation Coefficient).

Models / (Examples) data sets	2MLS		1MLS		MVAL	
	VAF	CCC	VAF	CCC	VAF	CCC
Confusion Data (with slogans)	0.9620	0.9842	0.9409	0.9741	0.9096	0.9538
Confusion Data (without slogans)	0.9018	0.9507	0.8548	0.9396	0.9475	0.9745
Association Data (with slogans)	0.3822	0.7921	0.4120	0.7532	0.6910	0.8316
Association Data (without slogans)	0.3614	0.7898	0.4721	0.7631	0.5830	0.7645

Tab. 1: Comparison of different two-mode hierarchical clustering algorithms

As can be seen from Tab. 1 in terms of VAF and CCC the MVAL algorithm completes favourably with the one-mode and two-mode least squares penalty approaches 1MLS and 2MLS. Generally, there might exist situations in which non-hierarchical approaches of two-mode clustering (see e.g. Both and Gaul (1985)) would allow for better interpretations as could be suspected in the case of the analysis of the association data where the VAF- and CCC-values indicate a fit not as good as for the confusion-recognition data. Here, MVAL outperfroms 1MLS and 2MLS in terms of VAF and CCC.

In Tab. 1 and Fig. 2, 3, 4 results of selected two-mode hierarchical clustering solutions yielded from the different algorithms are shown. The MVSL and MVCL results are omitted in all cases because MVAL did always better. In Fig. 2, 3, 4 brand name no. x means print ad no. x of the corresponding brand. Additionally, one can see that for DUNHILL, ERNTE 23, KIM and PHILIP MORRIS no print ads were shown at all. A comparison of the differences of the dendrograms obtained from different algorithms has to be left to the reader. Instead, some marketing aspects of German cigarette brands advertising should be mentioned.

From the marketing point of view it can be stated that - as can be seen from Fig. 3, 4 - incorporation of slogans does not lead to remarkable changings with respect to the CAMEL cluster (always including the WEST 1 print ad in which just a smoking man similar to the CAMEL-man was shown) as it should be for good "imagery-products" advertising. The MARLBORO cluster remains stable because MARLBORO can afford to do without slogans at all.

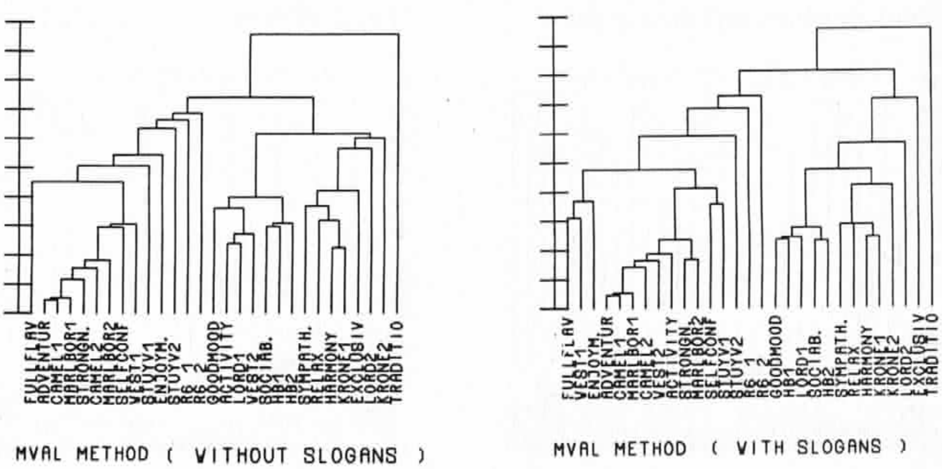

Fig. 2: Two- mode hierarchical clustering of association data of properties and print ads without and with slogans

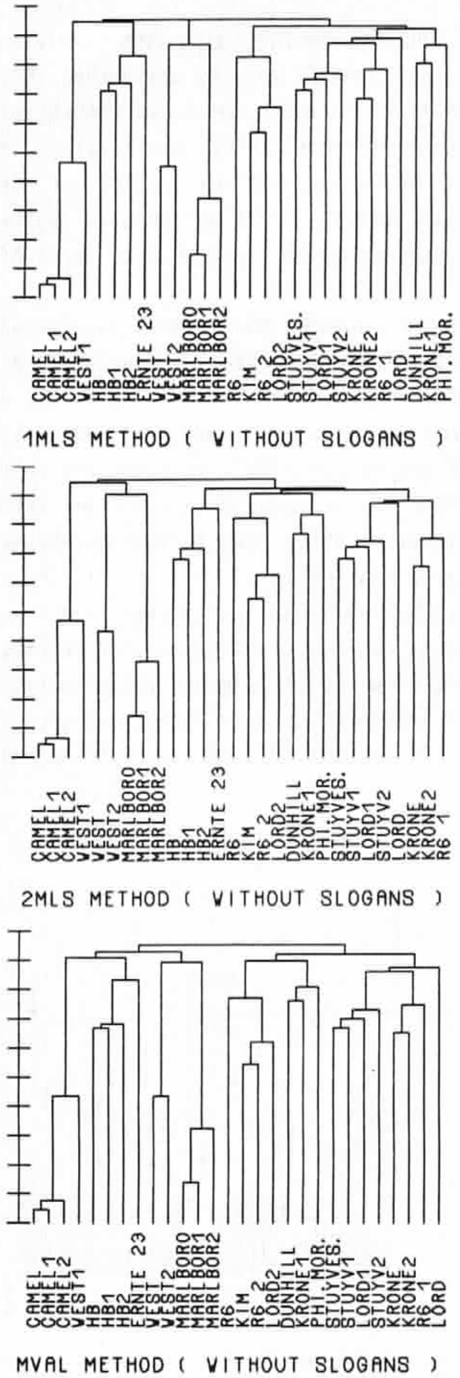

Fig. 3: Two-mode hierarchical clustering of confusion-recognition data of brands and print ads without slogans

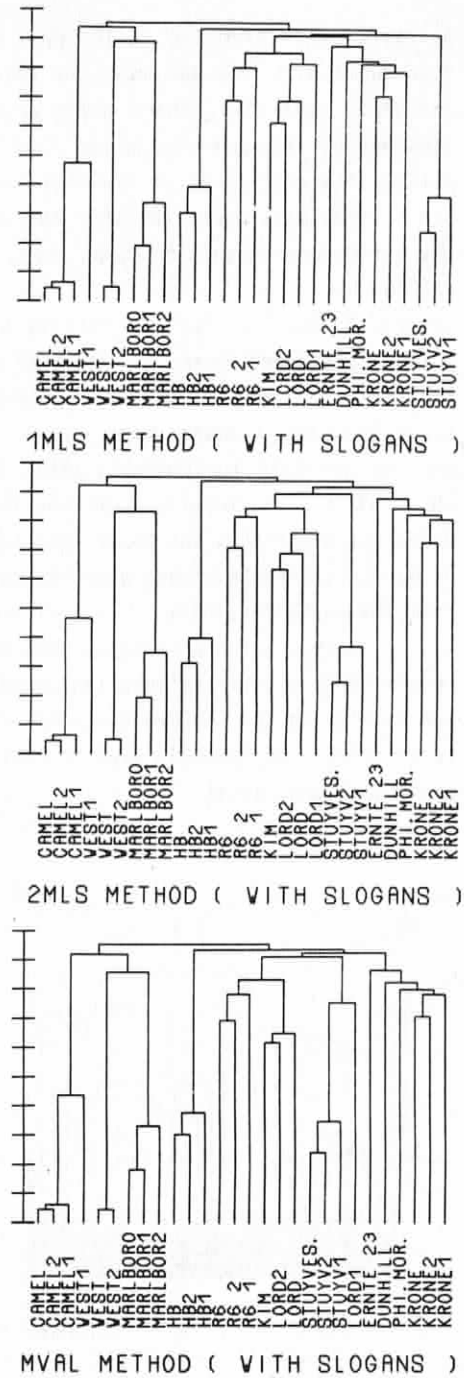

Fig. 4: Two-mode hierarchical clustering of confusion-recognition data of brands and print ads with slogans

Of course, for print ads the slogans of which contain the brand name (as for WEST "Let's go WEST" or for STUYVESANT " The world of Peter STUYVESANT") there must be a remarkable improvement in recognition of the brand the shown print ad advertises for although the brand name was, of course, masked within the slogan.
Without slogans the print ads LORD 2 and R6 2 which show smoking women are assigned to KIM a brand, especially, designed for the female target group. Slogans are needed to help to form different LORD and R6 clusters (but notice the differences within the different algorithms). No special comparisons concerning differences of relative heights of the subtrees which describe the corresponding clusters will be given because of space limitations.
In Fig. 2 the two-mode hierarchical clusterings of the association data show two big clusters (the adventure - enjoyment - full flavour - selfconfidence - strongness - cluster with CAMEL, MARLBORO, STUYVESANT and less clear-cut R6, WEST and the exclusiveness - good mood - harmony - relaxation - sociability - sympathy - cluster with HB, KRONE, LORD) where tradition seems to be least appropriate to describe cigarette brands advertising. Within these clusters one can recognize the different importance with which special properties are assigned to single brands via corresponding subclusters and one can also recognize that these assignments remain relatively stable whether a slogan is added or not as it should be for "imagery products" advertisement which has to convey emotions, feelings, etc., more via figurative information than via verbal information.

4. CONCLUSIONS

One-mode hierarchical clustering models are known to be appropriate for analysing marketing data. Also, two-mode hierarchical clustering will penetrate the marketing research area and, soon, come up with new marketing applications. Within the competing two-mode hierarchical clustering approaches the missing-values version of the classical average linkage procedure MVAL has, at least, the advantage of being simple to describe, easy to implement and fast to calculate its optimal solution. Of course, further comparisons of the different procedures are needed although the above example has already shown that MVAL competes favourably with some recently developed penalty approaches. Further research, e.g. combining missing values and penalty aspects, should be of interest.

REFERENCES:

[1] Böckenholt, I., Graphical Representation of Categorical Proximities: A Maximum Likelihood Method, in Gaul W., Schader M. (Eds.), Classification as a Tool of Research (North Holland, Amsterdam, 1986).

[2] Both, M. and Gaul W., PENCLUS (Penalty Clustering) for Marketing Applications, Working paper No. 82, Institut für Entscheidungstheorie und Unternehmensforschung, University of Karlsruhe (1985).

[3] Carroll, J. D., Clark, L. and DeSarbo, W. S., The Representation of Three-Way Proximity Data by Simple and Multiple Tree Structure Models, J. of Classification 1 (1984) 25-74.

[4] De Sarbo, W., GENNCLUS: New Models for General Nonhierarchical Clustering Analysis, Psychometrika 47 (1982) 449-475.

[5] De Sarbo, W. and De Soete, G., On the Use of Hierarchical Clustering for the Analysis of Nonsymmetrical Proximities, Journal of Consumer Research 11 (1984) 601-610.

[6] De Soete, G., Ultrametric Tree Representations of Incomplete Dissimilarity Data, J. of Classification 1 (1984) 235-242.

[7] De Soete, G., De Sarbo, W. S., Furnas, G. W. and Carroll, J. D., The Estimation of Ultrametric and Path Length Trees from a Rectangular Proximity Data, Psychometrika 49 (1984) 289-311.

[8] Furnas, G. W., Objects and their Features: the Metric Representation of Two Class Data, Unpublished Doctoral Dissertation, Stanford University, (February 1980).

[9] Gaul, W. and Hartung, J., A Barrier Method with Arbitrary Starting Point, Mathematische Operationsforschung und Statistik, Ser. Optimization 10 (1979) 317-323.

[10] Harshman, R., Green, P., Wind, Y. and Lundy, M., A Model for the Analysis of Asymmetric Data in Marketing Research, Marketing Science 1 (1982)

[11] Heiser, W. M. and Meulman J., Analizing Rectangular Tables by Joint and Constrained Multidimensional Scaling, Journal of Econometrics 22 (1983) 139-167.

[12] Rao, S. S. Optimization (Wiley Eastern Ltd, New Dehli, Bangalore, Bombay, 1978).

[13] Rao, V. R. and Sabavala, D. J., Inference of Hierarchical Choice Processes from Panel Data, Journal of Consumer Research 8 (1981) 85-96.

[14] Shepard, R. N. and Arabie Ph., Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties, Psychological Review 86 (1979) 87-123.