Journal of Classification 10:51-74 (1993)

An Error Variance Approach to Two-Mode Hierarchical Clustering

Thomas Eckes

Universität Wuppertal

Peter Orlik

Universität Saarbrücken

Abstract: A new agglomerative method is proposed for the simultaneous hierarchical clustering of row and column elements of a two-mode data matrix. The procedure yields a nested sequence of partitions of the union of two sets of entities (modes). A two-mode cluster is defined as the union of subsets of the respective modes. At each step of the agglomerative process, the algorithm merges those clusters whose fusion results in the smallest possible increase in an internal heterogeneity measure. This measure takes into account both the variance within the respective cluster and its centroid effect defined as the squared deviation of its mean from the maximum entry in the input matrix. The procedure optionally yields an overlapping cluster solution by assigning further row and/or column elements to clusters existing at a preselected hierarchical level. Applications to real data sets drawn from consumer research concerning brand-switching behavior and from personality research concerning the interaction of behaviors and situations demonstrate the efficacy of the method at revealing the underlying two-mode similarity structure.

Keywords: Clustering; Two-mode data; Ultrametric representation; Agglomerative algorithm; Heterogeneity index; Brand-switching; Behavior-situation congruence.

The authors would like to express their appreciation to Phipps Arabie and two anonymous referees for many helpful comments on earlier versions of the paper.

Authors' Addresses: Thomas Eckes, Fachbereich 1, Bergische Universität Wuppertal, Gaußstr. 20, D-5600 Wuppertal 1, FRG; Peter Orlik, Fachrichtung Psychologie, Universität des Saarlandes, Im Stadtwald, D-6600 Saarbrücken, FRG.

1. Introduction

Data indicating the relationships between two modes (Tucker 1964) or sets of entities (e.g., objects and features, stimuli and responses) are quite frequently collected in the behavioral and social sciences. The study of such data is usually restricted to the analysis of a single mode. For instance, some hierarchical clustering or multidimensional scaling model is employed to reveal the object-to-object similarity structure. But in many empirical studies, objects and features (or variables) can be considered entities with equal structural status; that is the similarity structure within one mode (e.g., objects judged) is of the same interest as the similarity structure within the other mode (e.g., judgment scales). Whenever there is no obvious a priori basis for prefering the analysis of one mode to the analysis of the other, it may be more useful to look for a common representation for both modes. Furthermore, a two-mode representation showing simultaneously the structure of two sets of entities and of their interrelationships may help in finding an adequate interpretation for some dimension or cluster.

Spatial or continuous representations of two-mode data, especially preference or dominance data, have been available since Coombs' (1950) original formulation of the unfolding model and its multidimensional generalization by Bennett and Hays (1960; for an excellent review of these and related models, see Carroll and Arabie 1980). More recently, DeSarbo and Rao (1984) presented a general set of multidimensional unfolding models and corresponding algorithms. Another prominent example of continuous models for two-mode data is correspondence analysis. This technique represents the rows and columns of a contingency table as points in high-dimensional space and then projects them onto a best-fitting subspace of lower dimensionality for ease of interpretation (see, e.g., Greenacre 1984; Greenacre and Hastie 1987). In contrast, until recently, there have been few nonspatial models allowing the simultaneous discrete representation of two modes (see Coppi and Bolasco 1989; Law, Snyder, Hattie, and McDonald 1984). Such discrete representations are the focus of the present paper. After a short review of existing two-mode clustering procedures, we propose a new agglomerative algorithm for constructing a two-mode hierarchical classification. Having introduced some pertinent terminology, we describe the rationale and the formal properties of this algorithm. Finally, we present two substantive applications from marketing and personality research.

2. Two-Mode Clustering Methods

Approaches for jointly clustering n row and m column elements in a two-mode data matrix $\mathbf{X}$ fall roughly into three classes. The first class is composed of "direct clustering methods." These methods do not refer explicitly to the notion of distance or similarity. They perform a reordering of rows and columns of the data matrix and yield (possibly overlapping) clusters which are interpretable directly on the input data. Examples are the "bond energy algorithm" originally proposed by McCormick, Schweitzer, and White (1972), the "modal block method" for categorical data developed by Hartigan (1975, 1976), and the hierarchical classes model introduced by De Boeck and Rosenberg (1988). The bond energy algorithm aims specifically at permuting the rows and columns of an input data matrix in such a way as to push the numerically larger matrix elements together. This is accomplished by maximizing the summed "bond strengths" (or "bond energy") over row and column permutations of the input matrix, where the "bond strength" between two nearest-neighbor elements is defined as their product. In recent years, the bond energy algorithm has been substantially improved (Arabie and Hubert 1990; Arabie, Schleutermann, Daws, and Hubert 1988) and investigated in the context of simulated annealing (Schleutermann, Arabie, Hubert, and Bronsard 1990) and blockmodeling (Arabie, Hubert, and Schleutermann 1990). In Hartigan's method, each block is described by a cluster of cases and a cluster of variables such that each variable in the block is constant over the cases in the block, except for cases that also belong to other blocks. With real data sets, many blocks will conform to this definition. Hence, selection of just a few blocks to represent the data is performed by finding in succession the patterns of variables which occur most frequently, ignoring those patterns found in previous steps (for a modification of Hartigan's algorithm, see Duffy and Quiroz 1991). De Boeck and Rosenberg's hierarchical classes model is restricted to object by attribute data matrices with binary entries of 0, 1. In such a model, objects with identical sets of attributes are grouped into disjoint object classes, each defined by a different set of attributes. These object classes are ordered hierarchically on the basis of their respective sets of attributes to reflect the subset/superset relations among them. Analogously, attributes are grouped into hierarchically ordered attribute classes. The hierarchy of object classes is linked to the hierarchy of attribute classes by means of a (symmetric) association relation. An object class is said to be associated to an attribute class if and only if the attribute class is in the set of attribute classes possessed by the given object class.

Methods belonging to the second class aim at fitting tree structures, that is ultrametric and/or additive trees, to two-mode data. Based on the pioneering work by Furnas (1980), least-squares procedures for estimating ultrametric and additive trees from two-mode proximity data have been proposed by DeSarbo and De Soete (1984) and De Soete, DeSarbo, Furnas, and Carroll (1984). The algorithms utilize a penalty function to enforce the ultrametric inequality generalized for the case of nonsymmetric or two-mode proximity matrices. More specifically, the algorithm for estimating an ultrametric tree consists of (a) transforming the $n \times m$ data matrix **X** into a matrix **T** best approximating **X** in a least-squares sense, where **T** satisfies the two-mode ultrametric inequality (defined below), (b) constructing a square $(n + m)$ by $(n + m)$ matrix **D** which satisfies the ordinary one-mode ultrametric inequality, and (c) using standard hierarchical clustering methods to obtain from **D** the ultrametric tree representation of both row and column elements. More recently, Espejo and Gaul (1986) developed a two-mode variant of the classical average linkage clustering method that compared favorably with the computationally complex penalty function algorithms.

The third class contains methods which are based on the ADCLUS (Shepard and Arabie 1979) model representing interobject proximities as combinations of discrete and possibly overlapping properties. A generalization of the ADCLUS model to the case of nonsymmetric or two-mode proximity data is the GENNCLUS methodology developed by DeSarbo (1982). The GENNCLUS procedure for the simultaneous clustering of both row and column elements utilizes a series of gradient-based techniques and combinatorial optimization methods in an alternating least-squares framework. Another generalization of the ADCLUS model is provided by PENCLUS (see Both and Gaul 1986). PENCLUS differs from GENNCLUS in implementing a penalty function approach to the construction of an overlapping or nonoverlapping two-mode classification.

The two-mode clustering procedure proposed here combines the advantages of the methods belonging to the first two classes outlined above. Similarly to the direct clustering methods, the procedure uses only that information contained within the input data matrix **X** for the simultaneous clustering of entities from both modes. It does not rely on the intervening construction of an $(n + m)$ by $(n + m)$ matrix having **T** (or **X**) as the $n \times m$ submatrix consisting of the last n rows and the first m columns. Stated differently, our procedure yields two-mode clusters which can be interpreted directly on the input data. Like the tree-fitting methods, it constructs an ultrametric tree representation of the two-mode data. It utilizes an agglomerative clustering criterion defining clusters with an average inter-mode relationship as strong as possible relative to a small within-cluster variance. In the following section some concepts relevant to our approach (called the "centroid effect method") are defined and details of the agglomerative process are presented.

3. The Centroid Effect Method

3.1 Definitions

Following Carroll and Arabie (1980) and Tucker (1964), a *mode* is defined as a particular set of entities. Modes will be denoted by capital letters A and B. Entities (e.g., objects, variables, experimental conditions) are denoted by subscripts; for example, A_i, $i = 1, \ldots, n$, could denote n objects (row elements), B_j, $j = 1, \ldots, m$, could denote m variables (column elements).

A *two-mode array* is defined as the Cartesian product $A \times B$ of two different modes A and B with pairs (cells) (A_i, B_j). A *two-mode data matrix* $\mathbf{X} = (x_{ij})$ is an assignment of numerical values x_{ij} (e.g., ratings, confusion frequencies, reaction times) to the elements (A_i, B_j) of a two-mode array. $\mathbf{X}$ contains nm elements. The set of row elements of $\mathbf{X}$ is $\{A_i\}$; the corresponding set of column elements is $\{B_j\}$.

Let $A' = \{A_{i'}\}$ be a subset of A, and $B' = \{B_{j'}\}$ be a subset of B. A *two-mode cluster* C_r is defined as the union of the two sets A' and B':

$$C_r = A' \cup B' = \{A_{i'}\} \cup \{B_{j'}\} . \tag{1}$$

A *two-mode submatrix* is an assignment of numerical values $x_{i'j'}$ to elements of $A' \times B' = \{(A_{i'}, B_{j'})\} \subset A \times B$. The two-mode submatrix corresponding to $A' \times B'$ is $\mathbf{X}_r = (x_{i'j'})$ with $n_r m_r$ elements, where n_r is the number of entities in A' and m_r is the number of entities in B'.

The *union of two clusters* $C_p = A' \cup B'$, and $C_q = A'' \cup B''$, where $A'' \subset A$, and $B'' \subset B$, $A' \cap A'' = \varnothing$, $B' \cap B'' = \varnothing$, is a cluster C_t defined as follows:

$$C_t = C_p \cup C_q = \{A' \cup B'\} \cup \{A'' \cup B''\} . \tag{2}$$

Let there be two clusters C_p and C_q at a particular hierarchical level. The assignment of numerical values to the elements of a two-mode array $A^* \times B^*$ with $A^* = A' \cup A''$ and $B^* = B' \cup B''$ yields a submatrix $\mathbf{X}_t$. This submatrix is decomposed into four submatrices or blocks, two of which correspond to the clusters $C_p = A' \cup B'$ and $C_q = A'' \cup B''$: $\mathbf{X}_p = (x_{i'j'})$ and $\mathbf{X}_q = (x_{i''j''})$, respectively; the other two blocks correspond to the sets $R_\alpha = A' \cup B''$, and $R_\beta = A'' \cup B'$: $\mathbf{X}_\alpha = (x_{i'j''})$ containing $n_p m_q$ elements, and $\mathbf{X}_\beta = (x_{i''j'})$ containing $n_q m_p$ elements, respectively. The decomposition of submatrix $\mathbf{X}_t$ into four blocks is illustrated in Figure 1.

$$
\begin{array}{cc}
 & \begin{array}{ccc} \quad B' \quad & \quad B'' \quad & \end{array} \\
\begin{array}{c} A' \\ \\ A'' \\ \end{array} & \left[\begin{array}{c:c:c} \underline{X}_p=(x_{i'j'}) & \underline{X}_\alpha=(x_{i'j''}) & \\ \hdashline \underline{X}_\beta=(x_{i''j'}) & \underline{X}_q=(x_{i''j''}) & \cdots \\ \hdashline \vdots & \vdots & \cdots \end{array}\right]
\end{array}
$$

Figure 1. Illustration of the Decomposition of Submatrix $\mathbf{X}_t$.

Let $M(A, B)$ be the set of all subsets of the union of set A and set B. A *two-mode hierarchical clustering system* is defined as a subset Ω of $M(A, B)$, which satisfies the following conditions:

(1) $A \cup B \in \Omega, \varnothing \notin \Omega$;
(2) $\{A_i\}, \{B_j\} \in \Omega$ for all $A_i \in A$ and $B_j \in B$;
(3) if $C_r, C_s \in \Omega$ with $C_r \cap C_s \neq \varnothing$, then $C_r \subseteq C_s$ or $C_s \subseteq C_r$.

3.2 The Algorithm

The centroid effect algorithm aims at constructing a two-mode hierarchical clustering system with clusters having minimal internal heterogeneity. Specifically, a two-mode cluster is said to have low internal heterogeneity to the extent that its elements show strong inter-mode relations with as small a variance of corresponding numerical values as possible.

In the following, the input data are presumed to be carefully scored or normalized such that larger entries indicate a stronger relationship between the corresponding row and column elements; that is, the data are interpreted as "similarities." In line with the above notion of cluster heterogeneity, the internal heterogeneity measure of a two-mode cluster C_r is expressed as

$$MSD_r = \frac{1}{n_r m_r} \sum_{A_{i'} \in A', B_{j'} \in B'} (x_{i'j'} - \mu)^2 , \tag{3}$$

where μ is the maximum entry in the input matrix $\mathbf{X}$, that is $\mu = \max_{i,j}(x_{ij})$. Thus, MSD_r is the mean squared deviation of entries $x_{i'j'}$ in the corresponding submatrix $\mathbf{X}_r$ from the maximum entry μ in $\mathbf{X}$.

In the definition of the *MSD* index, μ is chosen to be the maximum entry since it has the advantage of providing an easy-to-interpret measure of the internal heterogeneity of a newly formed two-mode cluster. Specifically, the present choice ensures that the cluster elements will show relatively strong, homogeneous inter-mode relations. In addition, it is in accordance with the traditional error variance rationale in one-mode hierarchical clustering; that is, the fusion value of the cluster formed at the lowest level, defined as the increase in the error variance within the cluster, is zero. Thus, our approach parallels Ward's (1963) one-mode agglomerative algorithm which starts with merging two entities whose fusion results in the minimum increase in the within-cluster error-sum-of-squares measure. Some issues concerning our choice of the maximum matrix entry will be dealt with in the final section.

Since

$$\frac{1}{n_r m_r} \sum_{A_{i'} \in A', B_{j'} \in B'} x_{i'j'}^2 = s_r^2 + \bar{x}_r^2 ,$$

where s_r^2 is the variance of entries in $\mathbf{X}_r$, and $\bar{x}_r^2$ is the corresponding squared mean, (3) can be written as

$$MSD_r = s_r^2 + (\bar{x}_r - \mu)^2 . \tag{4}$$

The squared difference between the mean entry in the submatrix corresponding to C_r and the maximum entry in the input matrix is called the "centroid effect" of this cluster. It can be seen that the problem of minimizing MSD_r is equivalent to finding a cluster C_r for which the sum of the variance of the corresponding numerical values and the centroid effect is minimum. Hence, the *MSD* index may be interpreted as a two-mode error variance term, analogous to the variance criterion of traditional hierarchical or nonhierarchical (partitioning) one-mode clustering techniques.

The fusion rule may now be specified as follows. At each step of the agglomerative algorithm, a particular subset of mode A is merged with a particular subset of mode B such that the increase in the internal heterogeneity measure of the resulting two-mode cluster is as small as possible. To accomplish this objective, several heuristic criteria are employed that are closely related to the *MSD* index. Which criterion will be used at any particular step in the agglomerative process depends on the subsets considered. Three general cases can be distinguished. In each case, those two subsets yielding the smallest criterion value will be merged.

Case I. A single-element subset $\{A_{i'}\}$ is to be merged with another single-element subset $\{B_{j'}\}$:

$$MSD_{i'j'} = (x_{i'j'} - \mu)^2 . \tag{5}$$

Case IIa. A single-element subset $\{B_{j'}\}$ is to be merged with an existing two-mode cluster C_r:

$$MSD_{\alpha} = \frac{1}{n_r} \sum_{A_{i'} \in A'} (x_{i'j'} - \mu)^2 . \tag{6}$$

Case IIb. A single-element subset $\{A_{i'}\}$ is to be merged with an existing two-mode cluster C_r:

$$MSD_{\beta} = \frac{1}{m_r} \sum_{B_{j'} \in B'} (x_{i'j'} - \mu)^2 . \tag{7}$$

Consider now the case in which two existing two-mode clusters are to be merged. According to the definition of the union of two clusters given above and schematically illustrated in Figure 1, the squared deviations are to be computed only for those entities of mode A and only for those entities of mode B which belong to the sets R_α and R_β, respectively. This case may be formally expressed as follows:

Case III. A two-mode cluster $C_p = A' \cup B'$ is to be merged with a two-mode cluster $C_q = A'' \cup B''$ into a new cluster C_t:

$$MSD_{\alpha\beta} = \frac{1}{n_p m_q + n_q m_p} \times \left[\sum_{A_{i'} \in A', B_{j''} \in B''} (x_{i'j''} - \mu)^2 + \sum_{A_{i''} \in A'', B_{j'} \in B'} (x_{i''j'} - \mu)^2 \right] . \tag{8}$$

As a conventional criterion for determining the number of two-mode clusters present in a given data matrix, a marked increase in fusion values can be considered indicative of the formation of a relatively heterogeneous cluster. Thus, a decision as to the number of clusters can be reached analogously to the fusion criterion in the one-mode error-sum-of-squares clustering method (Ward's [1963] method).

Since the number-of-clusters problem has no satisfactory solution based on a single criterion (see, e.g., Everitt 1979; Milligan and Cooper 1985), one may find it helpful to use another criterion related to the *MSD* index above. For a given cluster C_r, this criterion, which we call the "centroid effect ratio" (*CER*), is:

$$CER_r = \frac{\bar{x}_r^2}{s_r^2 + \bar{x}_r^2} . \tag{9}$$

The *CER* index measures the contribution of the mean "cluster effect size," expressed as the squared mean $\bar{x}_r^2$, to the total "cluster effect size," expressed as the sum of the within-cluster variance s_r^2 and $\bar{x}_r^2$. *CER* is directly related to the well-known "coefficient of variation" (V): $CER = 1/(1 + V)$. Thus, a cluster having a relatively low value of the *CER* index can be interpreted as having a small average inter-mode relation between entities relative to a large within-cluster variability. Preliminary evidence suggests that a value of the *CER* index smaller than 80% is indicative of a relatively low cohesiveness among the elements of the respective cluster. This cluster may then be excluded from further consideration.

As a final option, the algorithm allows constructing overlapping clusters to circumvent the problem characterizing hierarchical clustering methods in general: Once an element of the set or sets to be classified is a member of a cluster at some hierarchical level it can no longer be merged with any other cluster at the same level of the agglomerative process. The procedure is as follows: Having reached a decision as to the number of (disjoint) clusters, each row and column element of the input data matrix not already belonging to a given cluster is considered in turn a possible candidate for joining it. An element may be added if the corresponding criterion value (see Formulae (6) and (7) above) does not show a marked increase and/or the centroid effect ratio of the resulting cluster remains sufficiently high. Once again, values of the *CER* index above 80% may be considered a lower limit for including the respective element into the cluster. In any case, the decision as to the inclusion of a given row or column element into some cluster is left to the user. Although of an ad hoc nature, the present option may prove to be a useful adjunct to a disjoint two-mode hierarchical classification. An illustration of this procedure is given for the second empirical data set analyzed below.

To demonstrate the basic features of the algorithm, consider an example with row elements A_1, A_2, A_3, A_4, and column elements B_1, B_2, B_3. The artificial input data are shown in Table 1.

Presupposing that larger entries indicate stronger inter-mode relationships, elements A_2 and B_1 are joined together first. In the next step, elements

Table 1
Example Data

Mode A	Mode B B_1	B_2	B_3
A_1	4.1	1.9	3.1
A_2	6.8	2.3	2.4
A_3	1.7	3.6	6.1
A_4	3.0	5.8	3.2

Note. The higher the score, the stronger the relationship.

Table 2
Results for Example Data

Hierarchy Level	Increase in MSD Index	CER Index	Two-mode Cluster
1	0.00[a]	1.00	$\{A_2, B_1\}$
2	0.49[a]	1.00	$\{A_3, B_3\}$
3	1.00[a]	1.00	$\{A_4, B_2\}$
4	7.29[b]	0.94	$\{A_1, A_2, B_1\}$
5	11.60[c]	0.93	$\{A_3, A_4, B_2, B_3\}$
6	19.63[c]	0.83	$\{A_1, A_2, A_3, A_4, B_1, B_2, B_3\}$

Note. [a]See Formula (5). [b]See Formula (6b). [c]See Formula (7).

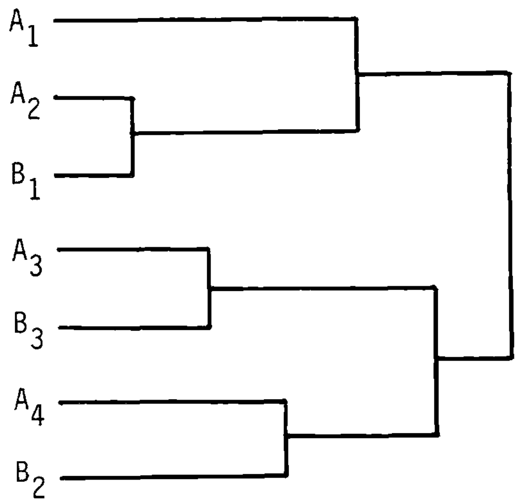

Figure 2. Tree Representation for the Example Data.

A_3 and B_3 are merged etc. The complete listing of the fusion process is given in Table 2. A graphical display of the two-mode hierarchical clustering system is provided by Figure 2.

4. Applications

In this section we present two applications of our approach to real data sets drawn from consumer research concerning brand-switching behavior and from personality research concerning the interrelationship between behaviors and situations. These applications are only suggestive of a range of substantive applications where a two-mode hierarchical clustering analysis seems to be the method of choice. Some further applications are discussed in Eckes (1991) and Eckes and Orlik (1991).

4.1 Soft Drinks Brand-Switching Data

The first application is concerned with a study from consumer research addressing the question of which brands of soft drinks chosen at one point in time will also be chosen at some later point in time (i.e., brand-switching probabilities). Since these data have been repeatedly analyzed in the two-mode clustering literature (see, e.g., Arabie, Schleutermann, Daws and Hubert 1988; DeSarbo 1982; De Soete and DeSarbo 1984), they lend themselves readily to evaluating the efficacy of our approach.

In a study by Bass, Pessemier and Lehmann (1972), 280 subjects were required to select a 12-ounce can of soft drink four days a week for three weeks from among eight brands. Table 3 presents the 8×8 nonsymmetric brand-switching matrix containing the observed probabilities of switching from one brand of soft drinks in period t to another brand in period $t + 1$. Prior to analysis, the matrix was normalized by dividing each cell by the product of the respective row and column arithmetic mean — a preprocessing step recommended in DeSarbo (1982) to dampen the effects of differences in market share.

Figure 3 displays the hierarchical tree structure and the corresponding sequence of fusion values derived from a centroid effect analysis of the normalized nonsymmetric proximity data.

The brands in capital letters represent the row items (in period t) and the other brands represent the column items (in period $t + 1$). Two major clusters emerge: one with non-diet items Pepsi, Coke, 7-Up, and Sprite, and the other with diet items Diet Pepsi, Tab, Like, and Fresca. The cluster of non-diet drinks is split into a cluster consisting of the market leaders, Coke and Pepsi, and a cluster containing the leading non-cola, lemon-lime drinks, Sprite and 7-Up. As can be seen, the diet drinks form a fairly compact group, the only exception being Fresca, which is identified as a cluster of its own (in each case, these cluster descriptions refer to both period t and period $t + 1$). Our results indicate that consumers are most likely to switch between Coke and Pepsi, between 7-Up and Sprite, and among the four diet drinks, Diet Pepsi, Tab, Like, and Fresca.

As noted above, there is some tradition of using the brand-switching data of Bass et al. to illustrate new methods of data analysis. Thus, it seems appropriate to compare the structural representation in Figure 3 with the results of previous studies. Using his GENNCLUS model, which requires the same number of clusters for both modes, DeSarbo (1982) identified three clusters. The only difference between the row clusters and the column clusters was that Fresca moved from a cluster containing 7-Up and Sprite (period t) to one containing Tab, Like, and Diet Pepsi (period $t + 1$). A high correspondence between DeSarbo's results and ours exists for his second row cluster comprising Tab, Like, and Diet Pepsi and our cluster of diet drinks (excluding Fresca). Since GENNCLUS allows for overlapping clusters, it is difficult to find any other simple correspondences. DeSarbo and De Soete (1984) fitted an ultrametric tree to the brand-switching data using a penalty function approach. Their tree representation and ours are in perfect agreement. This relationship even holds for the fusion levels at which the same row and column brands are joined. According to DeSarbo and De Soete (1984), these fusion levels may be considered inversely related to brand loyalty. For example, in both ultrametric trees the loyalty for Coke and Sprite

Table 3
Brand-Switching Data

	Period t + 1							
Period t	Coke	7-Up	Tab	Like	Pepsi	Sprite	Diet Pepsi	Fresca
Coke	.612 (24.87)	.107 (5.33)	.010 (1.19)	.033 (2.32)	.134 (5.61)	.055 (4.19)	.013 (1.36)	.036 (3.25)
7-Up	.186 (7.56)	.448 (22.33)	.005 (0.60)	.064 (4.50)	.140 (5.86)	.099 (7.53)	.012 (1.25)	.046 (4.15)
Tab	.080 (3.25)	.120 (5.98)	.160 (19.03)	.360 (25.32)	.080 (3.35)	.040 (3.04)	.080 (8.34)	.080 (7.21)
Like	.087 (3.54)	.152 (7.58)	.087 (10.35)	.152 (10.69)	.239 (10.01)	.043 (3.27)	.131 (13.66)	.109 (9.83)
Pepsi	.177 (7.19)	.132 (6.58)	.008 (0.95)	.030 (2.11)	.515 (21.57)	.075 (5.71)	.026 (2.71)	.037 (3.34)
Sprite	.114 (4.63)	.185 (9.22)	.029 (3.45)	.071 (4.99)	.157 (6.58)	.329 (25.04)	.029 (3.02)	.086 (7.75)
Diet Pepsi	.093 (3.78)	.047 (2.34)	.186 (22.13)	.093 (6.54)	.116 (4.86)	.093 (7.08)	.256 (26.68)	.116 (10.46)
Fresca	.226 (9.18)	.093 (4.64)	.053 (6.31)	.107 (7.53)	.147 (6.16)	.107 (8.14)	.067 (6.98)	.200 (18.03)

Note. Values in parantheses are normalized brand switching probabilities. From "The Theory of Stochastic Preference and Brand Switching" by F.M. Bass, 1974, *Journal of Marketing Research*, *11*, p. 19.

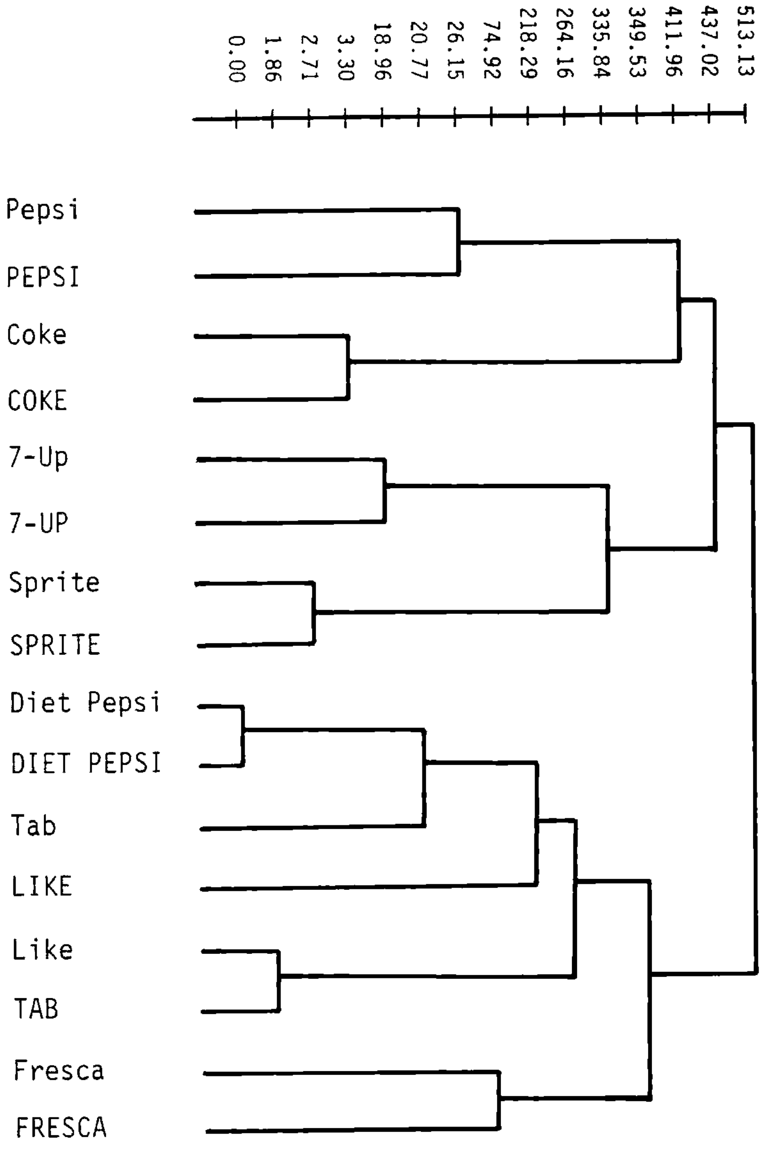

Figure 3. Tree Representation for the Brand-Switching Data.

seems to be much greater than that for Fresca and Tab. Finally, our results may be compared to Arabie, Schleutermann, Daws, and Hubert's (1988) application of an improved version of the bond energy algorithm. Although Arabie et al. did not use the normalization of brand-switching probabilities described in DeSarbo (1982), the three row clusters and the four column clusters they obtained reveal some marked similarities to our solution. The first row cluster containing Diet Pepsi, Tab, and Like can easily be identified with our first sub-cluster of diet drinks; the second row cluster agrees well with our non-diet coke cluster; the third row cluster consisting of 7-Up and Sprite corresponds to our non-diet non-cola cluster. The correspondences between the column clusters and our solution are less pronounced.

4.2 Behavior-Situation Congruence

The second application deals with a long-standing issue from personality and social psychology concerning the interrelationship between behaviors and situations. Viewed from an interactionist perspective (see Endler and Magnusson 1976), what is important in predicting social behavior is neither the determination of stable dispositions underlying the behavior nor the identification of salient properties of the situation in which the behavior occurs, but the analysis of the mutual relation between behavior and situation. In line with this reasoning, two-mode clustering seems to be the model of choice for depicting the structure of behavior-situation interactions. In the present case, the input matrix consists of mean ratings of the appropriateness of behaviors in everyday situations. Since these data are of some interest in their own right and since they are particularly well suited to illustrating the various facets of a centroid effect analysis, the theoretical and empirical context of this application is discussed in more detail.

Price (1974) addressed the question of whether classes of behaviors can be discovered which are uniquely appropriate in certain classes of situations. Fifty-two subjects were asked to rate, on a scale from 0 through 9, the appropriateness of 15 behaviors in each of 15 common situations (e.g., "sleep on a bus", "laugh at the movies", "eat at a job interview"). Behaviors and situations had been abstracted beforehand from detailed diaries kept by an independent sample of subjects for one entire day, recording each situation they found themselves in and what behavior they engaged in for each situation. The appropriateness ratings were averaged across all 52 subjects for each of the 225 behavior-situation items. Behaviors and situations were then arranged in a 15×15 matrix (see Table 4).

Before presenting the results of our analysis, we will give a short outline of how Price went about answering the above question. First, he performed two separate hierarchical cluster analyses (using Carlson's [1972]

Table 4

Behavior x Situation-Appropriateness Rating Data

	Behavior														
Situation	Run	Talk	Kiss	Write	Eat	Sleep	Mumble	Read	Fight	Belch	Argue	Jump	Cry	Laugh	Shout
Class	2.52	6.21	2.10	8.17	4.23	3.60	3.62	7.27	1.21	1.77	5.33	1.79	2.21	6.23	1.94
Date	5.00	8.56	8.73	3.62	7.79	3.77	3.12	2.88	3.58	2.23	4.50	4.42	3.04	8.00	3.79
Bus	1.44	8.08	4.27	4.87	5.48	7.04	5.17	7.17	1.52	2.15	4.17	3.12	3.08	7.10	3.00
Family dinner	2.56	8.52	4.92	2.58	8.44	2.29	2.54	3.96	1.67	2.50	3.25	2.29	3.21	7.13	1.96
Park	7.94	8.42	7.71	7.00	8.13	5.63	5.40	7.77	3.06	5.00	5.06	7.42	5.21	8.10	6.92
Church	1.38	3.29	2.38	2.85	1.38	1.77	3.52	3.58	0.62	1.42	1.92	1.71	3.13	2.60	1.33
Job interview	1.94	8.46	1.08	4.85	1.73	0.75	1.31	2.48	1.04	1.21	1.83	1.48	1.37	5.88	1.65
Sidewalk	5.58	8.19	4.75	3.38	4.83	1.46	4.96	4.81	1.46	2.81	4.08	3.54	3.71	7.40	4.88
Movies	2.46	4.98	6.21	2.73	7.48	4.08	4.13	1.73	1.37	2.58	1.71	2.31	7.15	7.94	2.42
Bar	1.96	8.25	5.17	5.38	7.67	2.90	6.21	4.71	1.90	5.04	4.31	3.75	3.44	8.23	4.13
Elevator	1.63	7.40	4.79	3.04	5.10	1.31	5.12	4.48	1.58	2.54	2.58	2.12	3.48	6.77	1.73
Restroom	2.83	7.25	2.81	3.46	2.35	2.83	5.04	4.75	1.77	5.12	3.48	3.65	4.79	5.90	3.52
Own room	6.15	8.58	8.52	8.29	7.94	8.85	7.67	8.58	4.25	6.81	7.52	6.73	8.00	8.17	6.44
Dorm lounge	4.40	7.88	6.54	7.73	7.19	6.08	5.50	8.56	2.40	4.00	4.88	4.58	3.88	7.75	3.60
Football game	4.12	8.08	5.08	4.56	8.04	2.98	5.23	3.69	2.04	3.85	4.98	7.12	4.31	7.90	7.94

Note. The higher the score, the more appropriate the behavior in the situation. From "Behavioral Appropriateness and Situational Constraint as Dimensions of Social Behavior" by R.H. Price and D.L. Bouffard, 1974, *Journal of Personality and Social Psychology*, *30*, p. 581.

algorithm), one of situations based on intercorrelations of situations across behaviors resulting in four situation clusters, and the other of behaviors based on intercorrelations of behaviors across situations resulting in four behavior clusters. Then he computed the mean appropriateness rating for the members of each behavior cluster with respect to those of each situation cluster. By plotting the cluster profiles for each of the situation clusters in terms of the behavior clusters, Price was able to demonstrate that each of the situation profiles had a distinctly different form and that there were some clusters of behaviors judged to be highly appropriate for some situations clustered together but not for others. For example, in his portrayal, the behaviors "eat", "laugh", "kiss", and "fight" appeared to be especially appropriate in the situations "date", "family dinner", and "movies".

Price's approach, based on techniques available at the time, is flawed for several reasons. First, similarity between behaviors is operationally defined according to similarity in the pattern of appropriateness ratings across situations. Hence, differences in the mean level of behavioral appropriateness are neglected, giving rise to some rather peculiar behavior clusters. For instance, "fight" and "kiss" belong to the same cluster; the corresponding mean ratings across situations are highly correlated (i.e., $r = .87$; see Price 1974, p. 576). However, the mean rating for "fight" is 1.96 (with a standard deviation of 0.95), the one for "kiss" is 5.00 (with a standard deviation of 2.22; see Price and Bouffard 1974, p. 582). This shows that "fight" is considered inappropriate in almost all situations studied, whereas "kiss" is considered appropriate in at least some situations — an information not retained in Price's representation of the data. Of course, the same reasoning applies to situational similarity using intercorrelations of situations across behaviors.

More importantly, the question Price tried to answer is of an inherently two-mode nature. Separate classifications of the set of behaviors and of the set of situations with post hoc efforts to relate the results of both classifications to each other cannot faithfully represent the intricate interrelationships between the entities of both modes. The latter goal can be better accomplished by a simultaneous hierarchic (and/or nonhierarchic) classification of behaviors and situations, as shown below.

Finally, Price's approach does not allow for a detailed consideration of particular behavior-situation relations. The congruence between behaviors and situations is only represented at the abstract level of clusters of behaviors or clusters of situations. Which behaviors of a given behavior cluster are judged to be highly appropriate (or highly inappropriate) in a specified situation remains an open question. In the following we show how a centroid effect analysis deals with these issues.

As a preliminary step, the data in Table 4 were centered by subtracting the midpoint of the rating scale (i.e., 4.50) from each entry; then each column (representing an element of the behavior mode) was reflected; that is entries were first duplicated in a columnwise fashion, and then the duplicated entries were rescored by multiplying with -1. This reflection of column elements ensured that behaviors highly appropriate in a given situation could be clustered separately from the same behaviors being highly inappropriate in some other situations. Put differently, in the final hierarchical solution each behavior is made to appear two times, once clustering with situations in which it is considered appropriate, and a second time clustering with situations in which it is considered inappropriate. The reason for this preparatory step is that exceptions to the rules of normal conduct for a particular situation (i.e., inappropriate behavior) may be at least as revealing about the underlying structure of consensual perceptions of the situation as is the most commonly exhibited behavior.

Figure 4 displays the hierarchical clustering solution constructed by the centroid effect analysis of the augmented 15×30 input matrix.

Situation labels are put in capital letters, and a minus sign in front of the label of a behavior indicates that the behavior is judged to be inappropriate in the situation(s) with which it is clustered. The succession of fusion values suggests the selection of four two-mode clusters, leaving four entities unclassified (i.e., "-talk", "-laugh", "fight", and "-write"). Two clusters comprise situations characterized exclusively by highly inappropriate behaviors (Clusters A and B), whereas the two other clusters consist of situations characterized exclusively by highly appropriate behaviors (Clusters C and D). Thus, in a job interview it is highly inappropriate to sleep, kiss, or belch (see Cluster B); in contrast, in one's own room he or she can feel free to sleep, talk, or read (see Cluster D). Our results show that situations like "job interview" have a fairly high degree of structure and definition, allowing for only a very small range in behavioral choices; situations like "own room" have substantially less structure and definition, allowing for a substantially greater range in behavioral choices (see Snyder and Ickes 1985).

In addition to visually representing the interrelationships between behaviors and situations, the hierarchical level at which a given behavior is for the first time joined together with a situation indicates the degree of judged (in)appropriateness of the behavior in that situation. For instance, as can be seen from the fusion values included in Figure 4, "sleep" is considered more inappropriate in a job interview than "cry", "jump", or "shout". Conversely, "talk" and "read" are considered more appropriate behaviors in one's own room than "cry", "mumble", or "argue".

As a final step, an overlapping cluster solution was constructed. Each row and column element of the input matrix was added to one or more of the four disjoint clusters if the resulting increase in fusion values was relatively low and the *CER* index did not fall below 80%. Table 5 gives a summary pre-

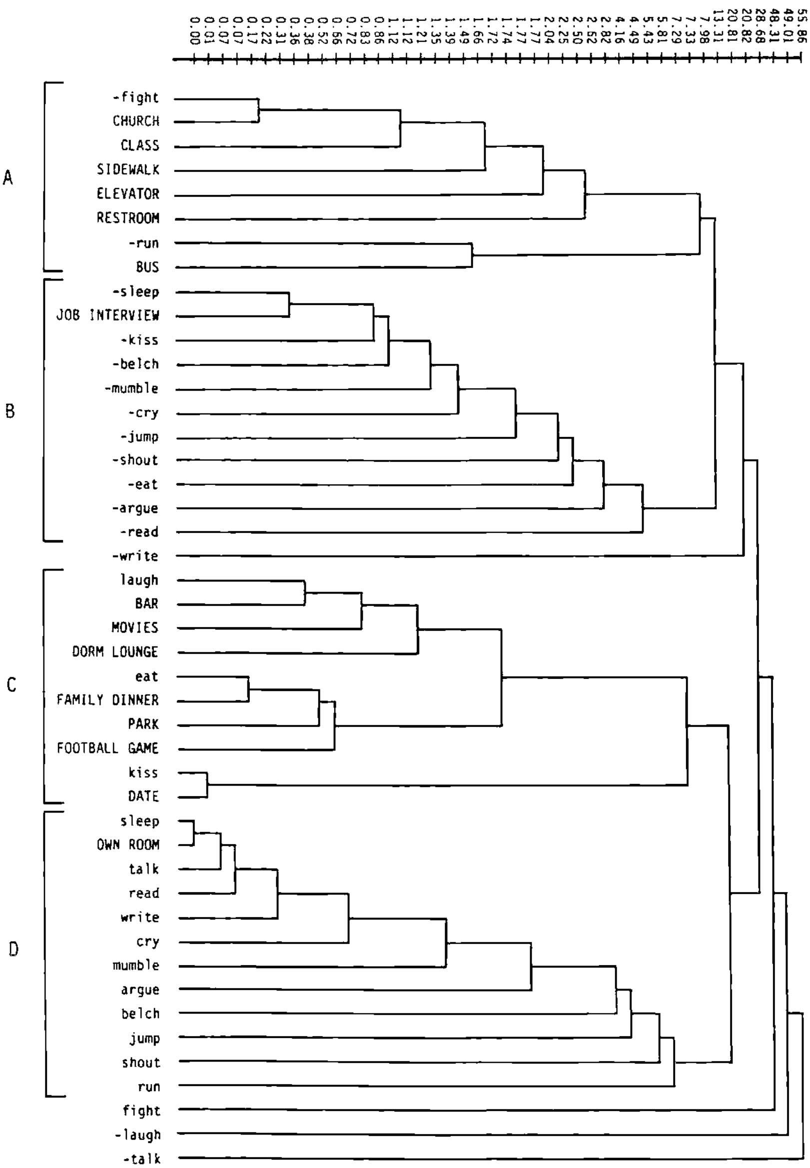

Figure 4. Tree Representation for the Behavior × Situation-Appropriateness Rating Data.

Table 5

Four Cluster Nonoverlapping and Overlapping Solution for Behavior x Situation-Appropriateness Rating Data

Original elements		Added elements	
Behaviors	Situations	Behaviors	Situations
	Cluster A		
-fight, -run	church, class, sidewalk, elevator, restroom, bus	---	job int., bar, movies, family d.
	Cluster B		
-sleep, -kiss, -belch, -mumble, -cry, -jump, -shout, -eat, -argue, -read	job int.	talk, -fight, -run	church
	Cluster C		
laugh, eat, kiss	bar, movies, dorm lou., family d., park, football, date	talk, -fight	own room, sidewalk, bus, elevator
	Cluster D		
sleep, talk, read, write, cry, mumble, argue, belch, jump, shout, run	own room	kiss, laugh, eat	park

Note. Behaviors with a minus sign are considered inappropriate in the respective situation(s).

presentation of the original four-cluster nonoverlapping solution and the corresponding overlapping solution.

Most interestingly, "talk" is now also contained within Cluster B, showing that this is the only behavior judged (highly) appropriate in a job interview. Furthermore, this behavior is joined with primarily social situations like "date", "family dinner", and "bar" (see Cluster C). Considering Cluster D, note that all the behaviors judged to be appropriate in one's own room are also judged to be appropriate in a park. Thus, "own room" and "park" seem to be situations highly similar to each other with respect to the behaviors they tend to elicit (see Frederiksen 1972). Taken together, the overlapping cluster solution provides a more complete and differentiated representation of the complex structural relationships between behaviors and situations, thereby leading to a deeper understanding of the behavior-situation congruence issue.

5. Limitations and Further Directions for Development

At the present stage of its development, certain limitations of our approach to two-mode hierarchical clustering should be noted. First, the *MSD* index and the fusion criteria are basically defined as (mean) squared differences between μ, the maximum entry in the input matrix, and entries corresponding to elements that are going to be merged when the new cluster is formed. We have chosen μ to be the maximum entry for pragmatic reasons, that is, the *MSD* index and the closely related fusion criteria have the advantage of being easy to interpret. However, since the internal heterogeneity measure depends on a single numerical value, it is prone to error to some extent. To examine what this dependence implies for the robustness of our algorithm, more detailed investigations using a large number of artificially created data sets have to be made. At a minimum, in the majority of applications the approach proposed here will necessitate some kind of normalization or standardization (Milligan and Cooper 1988) of the raw input data.

The present approach allows constructing overlapping clusters after a decision as to the number of disjoint clusters has been made. Each row and column element of the input matrix that has not already been classified is tentatively assigned to each cluster and retained only if the internal heterogeneity measure of the augmented cluster does not show a marked increase, or, alternatively, if the *CER* index does not fall below a predetermined threshold value. This procedure is limited insofar as overlapping clusters are constructed only after the number of disjoint clusters has been decided on. In addition, the criterion for determining whether a given row or column element is to be assigned to some existing cluster lacks objectivity. Future developments may improve the currently sequential, ad hoc procedure by incorporating the option for overlapping clusters at the very start of cluster formation.

These limitations notwithstanding, the approach presented here seems to be well suited to elucidating the two-mode similarity structure inherent in a data matrix. Applications of this method to two real data sets drawn from marketing research and from interactionist personality research revealed the interrelationships between the respective sets of entities. When theoretical considerations or hypotheses about the empirical data structure hint at the existence of strong unique associations between the elements of two different modes such as in the behavior-situation example, two-mode hierarchical clustering will prove a particularly useful tool of research.

References

ARABIE, P., and HUBERT, L.J. (1990), "The Bond Energy Algorithm Revisited," *IEEE Transactions on Systems, Man, and Cybernetics, 20,* 268-274.

ARABIE, P., HUBERT, L.J., and SCHLEUTERMANN, S. (1990), "Blockmodels from the Bond Energy Approach," *Social Networks, 12,* 99-126.

ARABIE, P., SCHLEUTERMANN, S., DAWS, J., and HUBERT, L. (1988), "Marketing Applications of Sequencing and Partitioning of Nonsymmetric and/or Two-Mode Matrices," in *Data, Expert Knowledge and Decisions*, Eds. W. Gaul, and M. Schader, Berlin: Springer-Verlag, 215-224.

BASS, F.M. (1974), "The Theory of Stochastic Preference and Brand Switching," *Journal of Marketing Research, 11,* 1-20.

BASS, F.M., PESSEMIER, E.A., and LEHMANN, D.R. (1972), "An Experimental Study of Relationships Between Attitudes, Brand Preference, and Choice," *Behavioral Science, 17,* 532-541.

BENNETT, J.F., and HAYS, W.L. (1960), "Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data," *Psychometrika, 25,* 27-43.

BOTH, M., and GAUL, W. (1986), "Ein Vergleich zweimodaler Clusteranalyseverfahren," *Methods of Operations Research, 57,* 593-605.

CARLSON, K.A. (1972), "A Method for Identifying Homogeneous Classes," *Multivariate Behavioral Research, 7,* 483-488.

CARROLL, J.D., and ARABIE, P. (1980), "Multidimensional Scaling," *Annual Review of Psychology, 31,* 607-649.

COOMBS, C.H. (1950), "Psychological Scaling Without a Unit of Measurement," *Psychological Review, 57,* 148-158.

COPPI, R., and BOLASCO, S. (Eds.) (1989), *Multiway Data Analysis*, New York: North-Holland.

DE BOECK, P., and ROSENBERG, S. (1988), "Hierarchical Classes: Model and Data Analysis," *Psychometrika, 53,* 361-381.

DESARBO, W.S. (1982), "GENNCLUS: New Models for General Nonhierarchical Clustering Analysis," *Psychometrika, 47,* 449-475.

DESARBO, W.S., and DE SOETE, G. (1984), "On the Use of Hierarchical Clustering for the Analysis of Nonsymmetric Proximities," *Journal of Consumer Research, 11,* 601-610.

DESARBO, W.S., and RAO, V.R. (1984), "GENFOLD2: A Set of Models and Algorithms for the GENeral UnFOLDing Analysis of Preference/Dominance Data," *Journal of Classification, 1*, 147-186.

DE SOETE, G., DESARBO, W.S., FURNAS, G.W., and CARROLL, J.D. (1984), "The Estimation of Ultrametric and Path Length Trees from Rectangular Proximity Data," *Psychometrika, 49*, 289-310.

DUFFY, D.E., and QUIROZ, A.J. (1991), "A Permutation-Based Algorithm for Block Clustering," *Journal of Classification, 8*, 65-91.

ECKES, T. (1991), "Bimodale Clusteranalyse: Methoden zur Klassifikation von Elementen zweier Mengen," *Zeitschrift für Experimentelle und Angewandte Psychologie, 38*, 201-225.

ECKES, T., and ORLIK, P. (1991), "An Agglomerative Method for Two-Mode Hierarchical Clustering," in *Classification, Data Analysis, and Knowledge Organization*, Eds. H.H. Bock, and P. Ihm, Berlin: Springer-Verlag, 3-8.

ENDLER, N.S., and MAGNUSSON, D. (1976), "Toward an Interactional Psychology of Personality," *Psychological Bulletin, 83*, 956-974.

ESPEJO, E., and GAUL, W. (1986), "Two-Mode Hierarchical Clustering as an Instrument for Marketing Research," in *Classification as a Tool of Research*, Eds. W. Gaul, and M. Schader, Amsterdam: North-Holland, 121-128.

EVERITT, B.S. (1979), "Unresolved Problems in Cluster Analysis," *Biometrics, 35*, 169-181.

FREDERIKSEN, N. (1972), "Toward a Taxonomy of Situations," *American Psychologist, 27*, 114-123.

FURNAS, G.W. (1980), "Objects and Their Features: The Metric Representation of Two-Class Data," Unpublished Doctoral Dissertation, Stanford University.

GREENACRE, M.J. (1984), *Theory and Applications of Correspondence Analysis*, London: Academic Press.

GREENACRE, M.J., and HASTIE, P. (1987), "The Geometric Interpretation of Correspondence Analysis," *Journal of the American Statistical Association, 82*, 437-447.

HARTIGAN, J.A. (1975), *Clustering Algorithms*, New York: Wiley.

HARTIGAN, J.A. (1976), "Modal Blocks in Definition of West Coast Mammals," *Systematic Zoology, 25*, 149-160.

LAW, H.G., SNYDER, C.W., HATTIE, J.A., and McDONALD, R.P. (Eds.) (1984), *Research Methods for Multimode Data Analysis*, New York: Praeger.

McCORMICK, W.T., SCHWEITZER, P.J., and WHITE, T.W. (1972), "Problem Decomposition and Data Reorganization by a Clustering Technique," *Operations Research, 20*, 993-1009.

MILLIGAN, G.W., and COOPER, M.C. (1985), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika, 50*, 159-179.

MILLIGAN, G.W., and COOPER, M.C. (1988), "A Study of Standardization of Variables in Cluster Analysis," *Journal of Classification, 5*, 181-204.

PRICE, R.H. (1974), "The Taxonomic Classification of Behaviors and Situations and the Problem of Behavior-Environment Congruence," *Human Relations, 27*, 567-585.

PRICE, R.H., and BOUFFARD, D.L. (1974), "Behavioral Appropriateness and Situational Constraint as Dimensions of Social Behavior," *Journal of Personality and Social Psychology, 30*, 579-586.

SCHLEUTERMANN, S., ARABIE, P., HUBERT, L.J., and BRONSARD, F. (1990), "Some Algorithms for "Bond Energy" Data Analysis, Including Simulated Annealing," in *Knowledge, Data and Computer-Assisted Decisions*, Eds. M. Schader, and W. Gaul, Berlin: Springer-Verlag, 139-152.

SHEPARD, R.N., and ARABIE, P. (1979), "Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties," *Psychological Review, 86*, 87-123.

SNYDER, M., and ICKES, W. (1985), "Personality and Social Behavior," in *Handbook of Social Psychology*, Vol.2, Eds. G. Lindzey, and E. Aronoson, New York: Random House, 883-947.

TUCKER, L.R. (1964), "The Extension of Factor Analysis to Three-Dimensional Matrices," in *Contributions to Mathematical Psychology*, Eds. N. Frederiksen, and H. Gulliksen, New York: Holt, Rinehart and Winston, 109-127.

WARD, J.H. (1963), "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association, 58*, 236-244.