PHIPPS ARABIE, J. DOUGLAS CARROLL, WAYNE DeSARBO, and JERRY WIND*

Most clustering techniques used in product positioning and market segmentation studies render mutually exclusive equivalence classes of the relevant products or subjects space. Such classificatory techniques are thus restricted to the extent that they preclude overlap between subsets or equivalence classes. An overlapping clustering model, ADCLUS, is described which can be used in marketing studies involving products/subjects that can belong to more than one group or cluster simultaneously. The authors provide theoretical justification for and an application of the approach, using the MAPCLUS algorithm for fitting the ADCLUS model.

Overlapping Clustering: A New Method for Product Positioning

Product positioning analysis (Wind 1977, 1980) of a given brand and its competitors has been based primarily on consumers' evaluations of similarities, perceptions, and importances of sets of attributes and various usage occasions. Such data traditionally have been analyzed by use of cross-tabulations, profile charts, and multidimensional scaling and/or cluster analysis. More recently, conjoint analysis has been used for assessing both the importance of the attributes and the perceived appropriateness of different brands on various attributes (Green and Srinivasan 1978). In addition to these direct approaches for assessing the positioning of a brand, some indirect approaches have been used occasionally: market share data, brand switching matrices, and brand vulnerability analysis (Wind 1977) which combines purchase patterns with overall attitudes toward the brand to assess the brand's degree of vulnerability and/or opportunities. The indirect approaches focus primarily on the identification of strengths and weaknesses in a brand's positioning, whereas the direct approaches attempt to explain the reasons for these strengths and weaknesses. Common to all these approaches is the identification of a single competitive set and positioning per brand.

Market segmentation analysis typically has included two sets of procedures: one for the identification of segments and one for the determination of the key discriminating characteristics of the various segments. Segment identification methods include simple sorting of respondents on some *a priori* basis for segmentation (e.g., users vs. nonusers) and a variety of clustering procedures for cases in which the segmentation cannot be determined *a priori* (Wind 1978). These clustering procedures have been applied to benefit, need, attitude, and lifestyle segmentation (Frank, Massy, and Wind 1972). To assess the key discriminating characteristics of the various segments, a variety of multivariate statistical methods such as multiple discriminant analysis, AID (Morgan and Sonquist 1963; Sonquist 1970), multiple regression, etc. have been used. Common to all the clustering-based segmentation analyses is the assignment of each consumer to a single segment.

Thus, cluster analysis has been used frequently in product positioning and market segmentation studies. In all these applications, a brand (in positioning studies) or a consumer (in segmentation studies) is identified as a member of one and only one cluster. This classification of brands and consumers into mutually exclusive and collectively exhaustive clusters, although methodologically elegant, is conceptually questionable. Brands can compete against more than

*Phipps Arabie is Professor of Psychology and Sociology at the University of Illinois, Champaign, and Resident Visitor at Bell Laboratories. J. Douglas Carroll and Wayne DeSarbo are Members of Technical Staff at Bell Laboratories, Murray Hill, New Jersey. Jerry Wind is Professor of Marketing, the Wharton School, University of Pennsylvania. The authors are indebted to Pamela Baker and Dan C. Knutson for the technical assistance provided. This research was supported in part by LEAA Grant 78-NI-AX-0142 and NSF Grant SOC 76-24512.

Journal of Marketing Research
Vol. XVIII (August 1981), 310-7

one competitive set (i.e., have multiple positioning); similarly, consumers can belong to more than one segment. Methodological constraints have precluded consideration of such cases and have restricted conventional positioning/segmentation analysis to the less realistic but technically simpler analysis.

The objective of our article is to describe a new clustering model and algorithm which relaxes the constraint (common to most other clustering algorithms) of clustering objects (brands) or subjects (consumers) into mutually exclusive and exhaustive categories, and which allows for the establishment of *overlapping* clusters. We first discuss the concept of overlapping clusters, then outline a specific overlapping clustering model and algorithm. The method is illustrated in the context of positioning and is compared with more conventional clustering procedures. We conclude with an outline of a possible application to market segmentation and a brief discussion of future applications of overlapping clusters.

THE CONCEPT OF OVERLAPPING CLUSTERS

Hierarchical clustering algorithms are among the most commonly used clustering analyses in marketing research. Users of these approaches, however, tend to discard much of the detail (e.g., levels of nesting for specific clusters) found in the dendrogram. The most commonly employed alternative is to obtain a partition of the set of entities being clustered. That is, the objects are segregated into mutually exclusive (and exhaustive) subsets. (As conventionally portrayed, each level or horizontal slice of a dendrogram is also a partition.) For a single partition, there is no nesting of subsets (or clusters—we use the terms interchangeably), so that representation of structure via a partition is necessarily nonhierarchical.

However, in reality, it is typically at least as easy to find examples in which the clusters should *overlap*, without the requirement found in hierarchical clustering that if two distinct clusters overlap, one must be a proper subset of the other. Such a restriction is often unrealistic. A person can belong to more than a single segment. For example, in benefit segmentation studies, a person can desire several different benefits from a particular product. The individual may desire both fresh breath *and* decay prevention (fluoride) in toothpaste. Similarly, a brand can compete in more than one cluster of products. Dentyne gum may compete with bubble gums as a candy substitute, but also may compete with mouthwash, toothpaste, etc. as a mouth freshener. Such classes (or clusters) of brands will show overlap.

If, rather than using a method of overlapping clustering one seeks to allow all possible patterns of overlap of a set of n entities, there are 2^{n-1} *clusters* to be considered (excluding the null set). As each of these clusters can in turn be present or absent in any given *clustering* solution, there are $2^{2^{n-1}} - 1$ possible clustering solutions when overlap is allowed. To look exhaustively at all possibilities for even moderate n is impossible. What is needed is an heuristic data-oriented approach that selects only those relatively few (potentially overlapping) clusters which contribute to goodness-of-fit and, one hopes, substantive interpretability.

Though methods of overlapping clustering have been available for decades (see reviews by Arabie 1977; Shepard and Arabie 1979), these methods (including the B_k method of Jardine and Sibson 1968, as well as that of Peay 1974) have been used much less frequently than hierarchical approaches. In addition, the use of these overlapping methods seems to have several drawbacks in practice. First, the methods generally produce too many clusters with too much overlap. In fact, arbitrary constraints often must be used to prevent excessive overlap (cf. Spilerman 1966). This problem results in part from the fact that most overlapping clustering methods, like most hierarchical methods, have not used a model or objective function to suggest which clusters are essential to the clustering representation. A second drawback is that if the clusters are too inclusive and/or have too much overlap, embedding them in a spatial solution (via multidimensional scaling techniques) is generally difficult. In such situations, the final result is simply a list of clusters (and, one hopes, their interpretations) without any visualizable graphic representation.

Conceptually, it is desirable to extend the concepts of product positioning and market segmentation to encompass the cases of overlapping clusters. Methodologically, this extension would require a clustering method for representing overlapping structure in data in a parsimonious manner, focusing only on necessary cluster overlaps that can be substantively justifed and often spatially presented.

THE METHOD

The ADCLUS Model

Assume n objects to be clustered, with input data of $M = n(n-1)/2$ entries constituting a two-way symmetric (or symmetrized) proximity matrix having no missing entries. Although the raw data may be in the form of either similarities or dissimilarities, we first transform them linearly to be similarities on the interval [0,1]. (Because the data are assumed to be on an interval scale, this transformation in no way affects the goodness-of-fit, but does allow for the standardization of various parameters in the method described hereafter.) S = $\|s_{ij}\|$ will refer to these transformed proximities, with which the fitted $\hat{S}$ matrix is being compared.

The basic equation underlying the ADCLUS (for ADditive CLUStering) model can be written as

$$\hat{s}_{ij} = \sum_{k=1}^{m} w_k p_{ik} p_{jk}, \tag{1}$$

where $\hat{s}_{ij}$ is the theoretically reconstructed similarity between objects i and j, w_k is a non-negative weight representing the salience of the property corresponding to subset k, and

$$p_{ik} = \begin{cases} 1, \text{ if object } i \text{ has property } k \\ 0, \text{ otherwise.} \end{cases}$$

Thus, we have a set of m subsets or clusters of the n objects, and these clusters (which are to be "recovered" or fitted by the clustering algorithm) are allowed to overlap, although there is no explicit requirement that they do so. We also associate with each of the m clusters a (typically non-negative numerical) weight, w_k (k = 1, ..., m). The rationale for the acronym ADCLUS is that the predicted similarity $\hat{s}_{ij}$ of any pair of objects is the sum of the weights of those clusters containing both objects i and j. These weights, which gauge the salience of their respective subsets, are an aspect of the ADCLUS model not found in other approaches to overlapping clustering. Note that the weights and the clusters are both fitted in applying the ADCLUS model. Moreover, the clustering is of course a discrete representation of the structure in the data, whereas the weights allow for continuous variation in the importance of the clusters. These features of the ADCLUS model, and their intimate relation to the objective function (discussed hereafter), qualify ADCLUS as a model for judgmental tasks in which proximities data are collected.

The MAPCLUS Algorithm

In matrix notation, the ADCLUS model is written

$$\hat{\mathbf{S}} = \mathbf{PWP}', \tag{2}$$

where $\hat{\mathbf{S}}$ is an $n \times n$ symmetric matrix of reconstructed similarities $\hat{s}_{ij}$ (with ones in the principal diagonal), $\mathbf{W}$ is an $m \times m$ diagonal matrix with the weights w_k (k = 1, ..., m) in the principal diagonal (and zeroes elsewhere), and $\mathbf{P}$ is the $n \times m$ rectangular matrix of *binary* values p_{ik}. Here, $\mathbf{P}'$ is the $m \times n$ matrix transpose of the matrix $\mathbf{P}$. Note that each column of $\mathbf{P}$ represents one of the m subsets, with the ones of that column defining the constituency of stimuli within the respective subset. Shepard and Arabie (1979) imposed the constraint that the m^{th} subset (and only that subset) was a column of all ones, whose weight was in effect an additive constant for equation 1, as required for the use of variance accounted for (VAF) as a measure of goodness-of-fit. For our usage, we prefer to express the model as

$$\hat{\mathbf{S}} = \mathbf{PWP}' + \mathbf{C}, \tag{3}$$

where $\mathbf{C}$ is an $n \times n$ matrix having zeroes in the principal diagonal and the (fitted) additive constant c in all the remaining entries. (That constant is simply the weight w_m fitted for the complete subset in equation 2.) Strictly, m in equation 2 and in the Shepard and Arabie (1979) description corresponds to $m - 1$ in equation 3. However, we believe that this inconsistency is clear enough to allow uniform references hereafter to m as the number of subsets, plus an (m + 1)st weight as the additive constant.

The Shepard and Arabie (1979) algorithm for fitting the ADCLUS model did not lead to a generally successful algorithm for fitting the model because of computational and numerical difficulties (see p. 118–19). Subsequently, Arabie and Carroll (1980) devised a different algorithm—MAPCLUS (for MAthematical Programming CLUStering)—for fitting the ADCLUS model.

The most obvious difference between the ADCLUS and MAPCLUS programs is that, for the latter, the (fairly small) number of subsets, m, is specified by the user at the beginning of an analysis and does not change throughout the computation. In practice, MAPCLUS has been able to obtain solutions acceptable in terms of both interpretability and goodness-of-fit, using considerably fewer clusters for various data sets than was possible with the ADCLUS program.

Because a detailed description of the MAPCLUS algorithm is given elsewhere (Arabie and Carroll 1980), we offer the following cursory overview of MAPCLUS. The $\mathbf{P}$ matrix is *initially* considered to have *continuously* varying p_{ik} in spite of the ultimate binary nature of $\mathbf{P}$. The *initial* values of $\mathbf{P}$ can be taken from any of several sources (detailed hereafter), and $\mathbf{W}$ is initially all zeroes. We use a gradient approach to minimize a loss function which is the weighted sum of an A- and a B-part. The former is simply a normalized measure of sum of squared error. The more novel B-part consists of a "penalty function" in the form of a polynomial designed to move all pairwise products $p_{ik}p_{jk}$ toward 0 or 1. Thus, the overall algorithm constitutes a "mathematical programming" approach to solving a discrete problem by treating it as a continuous problem with constraints allowing only a particular set of discrete values of parameters. Another way of describing this specific "penalty function" approach is that we attempt to approach the discrete solution by a sequence of increasingly close continuous approximations.

The subsets and weights are computed as follows. Given whatever estimates we have for the first subset $[p_{(1,1)}, ..., p_{(n,1)}]$, univariate regression is used to estimate w_1, and afterward the p_{i1} values are improved iteratively. Then, following an alternating least squares approach, we take residuals and fit them with a second subset, and so on, until the fit of the m^{th} subset has been iteratively improved and its weight estimated. We also apply multiple linear regression to improve our estimates of all the w_k (k = 1, ..., m) weights simultaneously. The whole procedure then is repeated with increases in the weight for the B-part relative to the A-part of the loss function, until asymptotically the (0,1) constraint holds essentially perfectly. When no further improvement in goodness-of-fit is forthcoming, we apply three additional techniques ("polishing," *de novo* iterations, and combinatorial optimization) to refine the fit still further.

In the present instance, we seek to maximize the variance accounted for, subject to the constraint that **P** is asymptotically binary. Our loss function takes the form

$$L_k(\alpha_k, \beta_k, \Delta, \mathbf{P}) = \alpha_k A_k + \beta_k B_k. \tag{4}$$

Considering first the left side of equation 4, note that the loss function is computed only for subset k. Moreover, we do not sum the penalty function over k. The reason is that we are using an alternating least squares approach (Wold 1966) which underlies the iterative fitting in turn of each subset p_{ik} ($i = 1, ..., n$) and its associated weight w_k. Wold has shown that for problems posed in a continuous form, the alternating least squares approach (NIPALS) will asymptotically lead to at least a local optimum (minimum) for the *overall* least squares problem for all m subsets of parameters in a model such as ours. (If there is only a single optimum this solution will, under very general conditions, be that global optimum, but this situation generally will not hold for the kind of highly nonlinear model we are fitting.) Because we are only fitting the k^{th} subset at any instant, Δ in equation 4 refers to the (centered) residuals computed for the remaining $m - 1$ subsets and the reader may wish to associate an implicit subscript k with Δ. For the present, the reader is asked only to note this statement of the procedure, as we believe the explanation is most easily presented in the next two subsections which explain in detail the alternating least squares implementation in MAPCLUS.

In the right side of equation 4, the term $\alpha_k A_k$ refers to the weight α_k applied to the normalized sum of squared error, A_k. Specifically,

$$A_k = \frac{a_k}{d_k}, \tag{5}$$

where:

$$a_k = \sum_{i>j}^{n} \sum^{n-1} (\delta_{ij} - w_k p_{ik} p_{jk})^2, \tag{6}$$

and

$$d_k = \frac{4 \sum_{i>j}^{n} \sum^{n-1} \delta_{ij}^2}{M}. \tag{7}$$

In equation 6, a_k is simply the sum of squared error, the minimization of which is equivalent to maximizing VAF. The denominator of A, d_k, is a normalization factor interpreted as the variance of the residuals, δ_{ij}, computed over the remaining subsets. The factor of 4 represents the maximum variance of 1/4 that could be obtained from the input data s_{ij} which, as noted before, are in the range [0,1].

If our loss function consisted only of the A term (i.e., the B weight $\beta = 0$), our problem would reduce to performing principal components analysis, for which well-known continuous methods are available. However, our model demands that $p_{ik} = 0,1$ and the B-part of the loss function is designed, by successively closer continuous approximations, to enforce this discrete constraint. We refer to this as a "mathematical programming" approach, as it entails optimizing a nonlinear function (A_k) with constraints on parameters imposed by use of the "penalty function" method.

Elaborating on the right side of equation 4, we have

$$B_k = \frac{u_k}{v_k}, \tag{8}$$

where:

$$u_k = \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} [(p_{ik} p_{jk} - 1) p_{ik} p_{jk}]^2 \tag{9}$$

and

$$v_k = \sum_{i>j}^{n} \sum^{n-1} (p_{ik} p_{jk} - T_k)^2, \tag{10}$$

where T_k is simply the mean of the pairwise products of $p_{ik} p_{jk}$, namely

$$T_k = \frac{1}{M} \sum_{i>j}^{n} \sum^{n-1} p_{ik} p_{jk}. \tag{11}$$

The numerator of B, u_k, is designed to force the pairwise products $p_{ik} p_{jk}$ to be 0,1. B is deliberately nonhomogeneous, because otherwise the p_{ik} could approach *any* relatively distinct pair of values instead of only 0,1. Products of *pairs* of the form $p_{ik} p_{jk}$ are emphasized to reduce the likelihood of singleton subsets, i.e., all but one of the p_{ik} ($i = 1, ..., n$) being zero. We deliberately include the diagonal terms in the numerator of B_k, but exclude them in the denominator for this reason.

Because MAPCLUS relies on a minimization procedure, we require the gradient of the loss function L_k, ∇L_k, with respect to p_{ik}. It is straightforward that the i^{th} component of ∇L_k ($\nabla_i L_k$) is

$$\nabla_i L_k = \frac{\partial L_k}{\partial p_{ik}} = \frac{\alpha}{d_k} \left[\frac{\partial a_k}{\partial p_{ik}} \right] + \beta \left[\frac{v_k \frac{\partial u_k}{\partial p_{ik}} - u_k \frac{\partial v_k}{\partial p_{ik}}}{v_k^2} \right] \tag{12}$$

where:

$$\frac{\partial a_k}{\partial p_{ik}} = -2 w_k \sum_{j \neq i} p_{jk} (\delta_{ij} - w_k p_{ik} p_{jk}), \tag{13}$$

$$\frac{\partial u_k}{\partial p_{ik}} = 2 p_{ik} \sum_{j} [(p_{ik} p_{jk} - 1)(p_{jk}^2)(2 p_{ik} p_{jk} - 1)], \tag{14}$$

and

$$(15) \quad \frac{\partial v_k}{\partial p_{ik}} = 2 \sum_{j \neq i} \left(p_{jk} - \frac{1}{M} \sum_{h \neq i} p_{hk} \right) (p_{ik} p_{jk} - T_k).$$

Finally, returning to the weights α_k and β_k in the loss function of equations 3 and 12, we use the constraint that $\alpha_k + \beta_k = 1$. We typically begin with $\alpha_k = \alpha_0 = .50$ and, as computation proceeds, increase the value of β_k relative to α_k (details given hereafter) to ensure that the final values of the $p_{ik} = 0,1$. This adjustment of α_k and β_k is done according to

$$(16) \quad \alpha_k' = \frac{\alpha_k}{\alpha_k + K_1 \beta_k} \quad \text{and} \quad \beta_k' = \frac{K_1 \beta_k}{\alpha_k + K_1 \beta_k}$$

where K_1 is currently defined as 2.0

We also note that the variance accounted for reported in the following application is given by

$$(17) \quad VAF = 1 - \frac{\sum_{i>j} (s_{ij} - \hat{s}_{ij})^2}{\sum_{i>j} (s_{ij} - \bar{s})^2}$$

and the $\hat{s}_{ij}$ are computed as in equation 1.

AN ILLUSTRATIVE APPLICATION TO PRODUCT POSITIONING

The MAPCLUS algorithm is illustrated on similarity data for 15 breakfast food items studied by Green and Rao (1972). The data constitute row conditional similarity judgments for each of 15 breakfast foods (listed in Table 1) given by 21 male Wharton School MBA students and their wives (a total of 42 respondents). The data for each subject were converted to a derived matrix of dissimilarities between all 15(14)/2 = 105 pairs of the food items. This preprocessing of the data was effected by use of the TRICON computer program (Carmone, Green, and Robinson 1968) to implement a procedure of Coombs (1964) that searches for the underlying complete pairwise rank order which reconstructs the preference ordering while minimizing patterns of intransitive preferences. As noted before, the output for TRICON is, for each subject, a triangular halfmatrix of "indirect" dissimilarities (in the form of ranks) for the 105 pairs of stimuli. The ranks were averaged over subjects to produce a two-way (food items by food items, averaged over respondents) matrix, as listed by Green and Rao (1972, p. 26).

MAPCLUS was applied to the data set, and various numbers of clusters ranging from 10 to 4 were specified. In using MAPCLUS, one finds the same tradeoff between interpretability and goodness-of-fit as in models of multidimensional scaling and factor analysis, in contrast to other clustering techniques. In changing from 8 to 5 clusters, trailing off of variance accounted for was slight, but the larger cluster solutions seemed much less interpretable than the 5-cluster solution. Specifically, most of the clusters in the 8-, 7-, and 6-cluster solutions appeared to be much too "large," i.e. too inclusive. The 4-cluster solution was altogether uninterpretable. The 5-cluster solution, accounting for 84.8% of the variance in the input data, was judged the best solution with respect to the tradeoff between the number of clusters and goodness-of-fit. The solution is listed in Table 2. The clusters, enclosed by contours, are embedded in a two-dimensional scaling solution (reproduced from Green and Rao, 1972, p. 31) in Figure 1.

Each of the clusters and its interpretation are listed in Table 2. The most heavily weighted subset consists of the pastry items. All of the items spread with either butter or margarine are in the second cluster, and all of the toasted items appear in the third cluster. The least sweet items (hard rolls and butter, toast and margarine, and buttered toast) are excluded from the fourth cluster, giving it a "sweet foods" characterization. The subset with the smallest weight consists of the relatively simple bread foods. The additive constant (0.125) may be interpreted as the weight for the complete set of stimuli. Alternatively, if that value were added to each of the 105 entries in the input matrix (an allowable transformation, under the assumption of interval-scale data), there would be no need for an additive constant (of zero) for the sixth or complete set of objects.

In discussing overlap, we argued for the substantive advantages of being able to portray this real-world aspect of structure in data. Figure 2 (taken from Green and Rao, 1972, p. 31) uses the same scaling solution as Figure 1 to embed clusters from the complete-link method (see Hartigan 1975; Hubert 1974; Johnson 1967) which is the method of hierarchical clustering

Table 1

BREAKFAST FOOD ITEMS USED BY GREEN AND RAO (1972)

Food item	*Plotting code used in figures*
1. Toast pop-up	TP
2. Buttered toast	BT
3. English muffin and margarine	EMM
4. Jelly donut	JD
5. Cinnamon toast	CT
6. Blueberry muffin and margarine	BMM
7. Hard rolls and butter	HRB
8. Toast and marmalade	TMd
9. Buttered toast and jelly	BTJ
10. Toast and margarine	TMn
11. Cinnamon bun	CB
12. Danish pastry	DP
13. Glazed donut	GD
14. Coffee cake	CC
15. Corn muffin and butter	CMB

Table 2
MAPCLUS REPRESENTATION OF THE 15 FOOD ITEMS

Rank by weight	Weight	Items in subset	Interpretation
1	.430	jelly donut, glazed donut, cinnamon bun, Danish pastry, coffee cake	pastries
2	.393	blueberry muffin and margarine, corn muffin and margarine, English muffin and margarine, hard rolls and butter, toast and margarine, buttered toast	foods spread with margarine or butter
3	.313	toast pop-up, cinnamon toast, toast and marmalade, buttered toast and jelly, buttered toast, toast and margarine	toasted foods
4	.264	all but hard rolls and butter, toast and margarine, buttered toast	sweet foods (compared with those excluded)
5	.203	English muffin and margarine, hard rolls and butter, toast and margarine, buttered toast, buttered toast and jelly, toast and marmalade	relatively simple bread foods

Note: The data are from Green and Rao (1972, p. 26). Variance accounted for = 84.8%, with five subsets and an additive constant of 0.125.

Figure 1
MAPCLUS SOLUTION FOR 15 FOOD ITEMS LISTED IN TABLES 1 AND 2 (VAF = 84.8%)

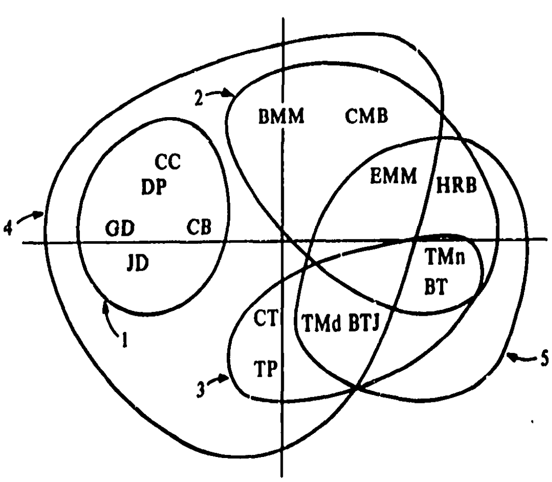
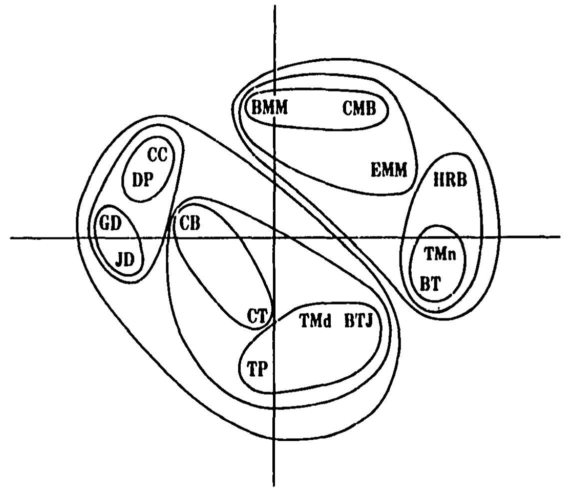


Figure 2
COMPLETE-LINK HIERARCHICAL CLUSTERING SOLUTION FOR THE 15 FOOD ITEMS
(from Green and Rao 1972, p. 31)



most commonly used in marketing research, psychology, and related areas. Although certain levels of embeddedness (see preceding discussion) were discarded by Green and Rao (1972), in their presentation of the complete-link solution, we regard their usage of that method as exemplary. Note, however, the substantive restrictions imposed by the hierarchical constraints. For instance, as soon as buttered toast (BT) is clustered with toast and margarine (TM), there can be no linkage between buttered toast and buttered toast and jelly (BTJ), except at the level of the complete or trivial clustering. In addition, note that although many of the MAPCLUS subsets are also found in the complete-link solution of Figure 2, others are not. For instance, the third most heavily weighted cluster, the toasted foods, does not appear at all in the complete-link representation, nor does the fifth MAPCLUS subset.

In addition, without the screening already performed by Green and Rao in removing uninterpretable levels of embedding or nestedness, the difficulties with the complete-link solution compared with Figure 1 (MAPCLUS solution) would be all the more apparent. If a cluster at a given level of nestedness in Figure 2 is eliminated (e.g., the [BMM, CMB] dyad is dropped so that only the [BMM, CMB, EMM] triad appears), there is no objective function against which the resulting change in overall goodness-of-fit can be evaluated. In contrast, for the MAPCLUS solution, the data analyst can add, delete, or modify clusters (to produce a "constrained" solution; see Shepard and Arabie 1979; Carroll and Arabie 1980), and use straightforward multiple linear regression to estimate weights for the new set of clusters and obtain the corresponding variance accounted for.

Finally, the complete-link clustering has no counterpart to the weights for the clusters in the MAPCLUS representation for the Green-Rao (1972) data. That is, the data analyst is given no indication of the relative salience (in the subjects' judgments) of the clusters in the complete-link solution. Although some promising work in this area has been done for hierarchical clustering (Hubert 1973), there is simply no counterpart in hierarchical clustering, or in other approaches to nonhierarchical clustering, to the weights that are such an integral part of both the ADCLUS model and the interpretation of resulting solutions from the MAPCLUS algorithm.

CONCLUDING REMARKS

The availability of an overlapping clustering model and algorithm (Shepard and Arabie 1979) makes feasible the clustering of objects (brands) or subjects (customers) in a way that does not require each object or subject to be a member of one and only one cluster, but rather allows for membership in several clusters. This development frees the researcher from the previous methodological limitations of clustering objects or subjects into mutually exclusive and exhaustive categories in spite of the conceptually obvious fact that a person can belong to more than a single segment (e.g., multiple role concept) and that a brand can compete in more than one cluster of products. The output of the MAPCLUS overlapping clustering algorithm, compared with the output of the more conventional hierarchical clustering algortihms, clearly illustrates the conceptual advantages of having the flexibility of clustering each brand in more than a single cluster. Similar advantages accrue if segment membership is not limited to a single segment.

Despite the popularity of clustering-based segmentation analysis which focuses on the grouping of respondents into mutually exclusive and exhaustive clusters, in some cases an overlapping segmentation option might be desirable. Overlapping clustering can be applied in market segmentation to cluster individuals in the same way the procedure has been applied in product positioning to cluster brands; i.e., instead of inputing a $k \times k$ matrix of brand similarity one can use any $n \times n$ matrix of subject similarity in the MAPCLUS algorithm. This $n \times n$ subject matrix can be similarity rating or ranking, or importance scores of desired benefits, needs, attitudes, etc.

In addition to this straightforward application, which would result in the clustering of individuals but not necessarily with a concomitant interpretable multidimensional space, one can use as input to the MAPCLUS algorithm data which allow the subjects to be clustered in some interpretable space such as brand or product feature space. This procedure can be carried out either in the more conventional manner, with ideal points in a brand space (see, for example, Green and Tull 1978) or, if more appropriate, with points in a benefit (product feature) space (see, for example, Green, Wind, and Claycamp 1975). Once the respondents are located graphically in an appropriate multidimensional space, they can be analyzed with MAPCLUS using the same input (after its adjustment to an $n \times n$ subject matrix), and the resulting configurations can be compared.

In any of these applications of overlapping clustering, one has a problem with the capability of the current version of the MAPCLUS algorithm. Currently, one cannot economically process a matrix larger than 30×30. Although work is underway to develop a more efficient algorithm, one can imagine using in the interim a two-step procedure for clustering of respondents.

1. Clustering of any size sample (100 respondents, for example) by any clustering algorithm (such as hierarchical clustering; Howard and Harris 1966; Johnson 1967; or others) and selecting ≤ 30 pseudosegments. To ensure stable results it might be desirable to use two clustering algorithms and compare their resulting segments.
2. Take the 30 or so pseudosubjects, develop for them a 30×30 matrix of intersubject similarities (based on their average scores on the variables of interest), and submit the data matrix to the MAPCLUS algorithm.

This procedure is feasible and can result in overlapping segments.

The reported developments in overlapping cluster models and algorithms offer new challenges to marketing researchers.

1. Algorithm developments to overcome the size and cost constraints of MAPCLUS.
2. Model developments to extend the overlapping clustering to three- (or higher) way matrices (Carroll and Arabie 1979).
3. Reexamination of the segmentation and positioning concepts to assess the implication of an overlapping clustering formulation for marketing strategy considerations. If multiple positioning is desired, i.e., positioning a brand against each of the clusters in which the brand competes, how can the strategy be executed in terms of promotion material and distribution?

The concept of overlapping segments is more complex because it can suggest: (1) the selection of several target segments, all of which share some common characteristic, and (2) the selection of a single target segment but the recognition that members of this segment share interest (loyalties?) with other segments as well.

Both interpretations suggest that given the likely presence of overlapping segments (i.e., additional evidence of the presence of segment heterogeneity), the selection of a target market segment *should not* be done in isolation from the positioning decision (whether single or multiple), but rather both should be undertaken simultaneously. This conclusion is consistent with the recent developments in positioning/segmentation research as seen in the approaches of flexible segmentation (Wind 1978) and componential segmentation (Green, Wind, and Claycamp 1978) and its more recent formulation in the context of the POSSE system (Carroll et al. 1980). Technically, we are urging that because the analyst begins with a two-mode, two-way (see Carroll and Arabie 1980) data matrix of brands by benefits, it is desirable to seek ADCLUS clusterings of the entities in each mode.

The acceptance of this concept, coupled with the high likelihood of the presence of both overlapping positioning and segments, suggests the need to examine explicitly the relation between the two sets of results. That is, one should select a target segment(s) (with or without overlap) and analyze the overlapping positioning of brands as perceived by members of these segments. Alternatively, one can first select a desirable positioning (with or without overlap) and analyze the various overlapping segments most responsive to the target positioning. Such analyses could better provide insights into the complexity of the marketplace and lead to thoughtful positioning/segmentation decisions. This sequential approach is not as accurate as a decision based on simultaneous analyses of the two decisions (such as the one proposed in the POSSE model). However, to the extent that management is relying on separate positioning and segmentation analysis, making these decisions separately or at best sequentially, it is desirable to consider the fact that a brand is likely to compete in more than a single market against different brands and to attract consumers with diverse benefit preferences. In these cases the methododology we propose is most applicable and could help advance the value of the positioning/segmentation concepts as guidelines for marketing strategy.

REFERENCES

Arabie, P. (1977), "Clustering Representations of Group Overlap," *Journal of Mathematical Sociology*, 5, 113-28.

——— and J. D. Carroll (1980), "MAPCLUS: A Mathematical Programming Approach to Fitting the ADCLUS Model," *Psychometrika*, 45, 211-35.

Carmone, F. J., P. E. Green, and P. J. Robinson (1968), "TRICON—an IBM 360/65 FORTRAN IV Program for the Triangularization of Conjoint Data," *Journal of Marketing Research*, 5 (May), 219-20.

Carroll, J. D. and P. Arabie (1979), "INDCLUS: A Three-way Approach to Clustering," paper presented at meeting of Psychometric Society, Monterey, California.

——— and P. Arabie (1980), "Multidimensional Scaling," in *Annual Review of Psychology*, M. R. Rosenzweig and L. W. Porter, eds. Palo Alto, California: Annual Reviews.

———, W. S. DeSarbo, S. Goldberg, and P. E. Green (1980), "A General Approach to Product Design Optimization via Conjoint Analysis," working paper, Bell Laboratories, Murray Hill, New Jersey.

Coombs, C. H. (1964), *A Theory of Data*. New York: John Wiley & Sons, Inc.

Frank, R., W. Massy, and Y. Wind (1972), *Market Segmentation*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Green, P. E. and V. R. Rao (1972), *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms*. New York: Holt, Rinehart and Winston.

——— and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Reserach*, 5, 103-23.

——— and D. S. Tull (1978), *Research for Marketing Decision*, 4th ed. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

———, Y. Wind, and H. J. Claycamp (1975), "Brand Features Congruence Mapping," *Journal of Marketing Research*, 12 (August), 306-13.

Hartigan, J. A. (1975), *Clustering Algorithms*. New York: John Wiley & Sons, Inc.

Howard, N. and B. Harris (1966), "A Hierarchical Grouping Routine, IBM 360/65 FORTRAN IV Program." Philadelphia: University of Pennsylvania Computer Center.

Hubert, L. J. (1973), "Monotone Invariant Clustering Procedures," *Psychometrika*, 38, 47-62.

——— (1974), "Some Applications of Graph Theory to Clustering," *Psychometrika*, 39, 283-309.

Jardine, N. and R. Sibson (1968), "The Construction of Hierarchic and Non-hierarchic Classifications," *Computer Journal*, 11, 177-84.

Johnson, S. C. (1967), "Hierarchical Clustering Schemes," *Psychometrika*, 32, 241-54.

Morgan, J. N. and J. A. Sonquist (1963), "Problems in the Analysis of Survey Data and a Proposal," *Journal of the American Statistical Association*, 58, 415-34.

Peay, E. R. (1974), "Hierarchical Clique Structures," *Sociometry*, 37, 54-65.

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

Sonquist, J. A. (1970), *Multivariate Model Building*. Ann Arbor: Survey Research Center, University of Michigan.

Spilerman, S. (1966), "Structural Analysis and the Generation of Sociograms," *Behavioral Science*, 11, 312-18.

Wind, Y. (1977), "The Perception of the Firm's Competitive Position," in *Behavioral Models of Market Analysis: Foundations for Marketing Action*, F. M. Nicosia and Y. Wind, eds. Hinsdale, Illinois: Dryden Press.

——— (1978), "Issues and Advances in Segmentation Research," *Journal of Marketing Research*, 15 (August), 317-37.

——— (1980), "Going to Market—New Twists for Some Old Tricks," *Wharton Magazine* (Spring).

——— and P. J. Robinson (1972), "Product Positioning: An Application of Multidimensional Scaling," in *Attitude Research in Transition*, R. I. Haley, ed. Chicago: American Marketing Association.

Wold, H. (1966), "Estimation of Principal Components and Related Models by Iterative Least Squares," in *Multivariate Analysis*, P. R. Krishnaiah, ed. New York: Academic Press.