PSYCHOMETRIKA—VOL. 56, NO. 1, 121–136
MARCH 1991

MULTICLUS: A NEW METHOD FOR SIMULTANEOUSLY PERFORMING MULTIDIMENSIONAL SCALING AND CLUSTER ANALYSIS

WAYNE S. DESARBO

DEPARTMENTS OF MARKETING AND STATISTICS
THE UNIVERSITY OF MICHIGAN

DANIEL J. HOWARD

MARKETING DEPARTMENT
E. L. COX SCHOOL OF BUSINESS
SOUTHERN METHODIST UNIVERSITY

KAMEL JEDIDI

MARKETING DEPARTMENT
GRADUATE SCHOOL OF BUSINESS
COLUMBIA UNIVERSITY

This paper develops a maximum likelihood based method for simultaneously performing multidimensional scaling and cluster analysis on two-way dominance or profile data. This MULTICLUS procedure utilizes mixtures of multivariate conditional normal distributions to estimate a joint space of stimulus coordinates and K vectors, one for each cluster or group, in a T-dimensional space. The conditional mixture, maximum likelihood method is introduced together with an E-M algorithm for parameter estimation. A Monte Carlo analysis is presented to investigate the performance of the algorithm as a number of data, parameter, and error factors are experimentally manipulated. Finally, a consumer psychology application is discussed involving consumer expertise/experience with microcomputers.

Key words: multidimensional scaling, cluster analysis, maximum likelihood estimation, consumer psychology.

1. Introduction

This paper develops a maximum likelihood based method for simultaneously performing multidimensional scaling and cluster analysis on a given set of two-way dominance/preference or profile data. This procedure, which we shall call MULTICLUS, utilizes mixtures of multivariate conditional normal distributions to estimate a joint space of stimulus coordinates and K vectors, one for each cluster or group, in a T-dimensional space. The next section presents the technical structure of the model as well as the E-M algorithm devised for estimating the model's parameters. Section three reports the results of a Monte Carlo analysis that examines the performance of the method as a number of data, parameter, and error factors are experimentally manipulated. Finally, an application is provided in consumer psychology utilizing the procedure.

We wish to thank the editor, associate editor, and three anonymous reviewers for their helpful comments on earlier versions of this manuscript.

Reprint requests should be sent to Wayne S. DeSarbo, Marketing and Statistics Depts., School of Business, The University of Michigan, Ann Arbor, MI 48109.

0033-3123/91/0300-9561 $00.75/0

2. The MULTICLUS Method

The Model

Let the indices i, j, k, and t denote, respectively, a particular object (e.g., subject), stimulus, cluster, and dimension; $1 \leq i \leq I$, $1 \leq j \leq J$, $1 \leq k \leq K$, and $1 \leq t \leq T$. Define:

- Δ_{ij}: the observed profile/dominance value of (column) stimulus j for (row) object i;
- b_{jt}: the t-th coordinate value for stimulus j;
- a_{kt}: the t-th coordinate for the vector terminus for cluster k;
- $\boldsymbol{\Sigma}_k$: the $J \times J$ variance-covariance matrix for cluster k.

We assume that the probability density function for the $1 \times J$ random row vector $\boldsymbol{\Delta}_i = (\Delta_{i1}, \ldots, \Delta_{iJ})$ is a finite mixture of conditional distributions:

$$g(\boldsymbol{\Delta}_i; \boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}) = \sum_{k=1}^{K} \lambda_k f_{ik}(\boldsymbol{\Delta}_i | \mathbf{a}_k, \mathbf{B}, \boldsymbol{\Sigma}_k), \tag{1}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{K-1})$ are the $K - 1$ independent mixing proportions of the finite mixture such that

$$0 \leq \lambda_k \leq 1, \tag{2}$$

and

$$\lambda_K = 1 - \sum_{k=1}^{K-1} \lambda_k. \tag{3}$$

Here, $\boldsymbol{\Sigma}$ is a $K \times J \times J$ array containing each $\boldsymbol{\Sigma}_k$ matrix, $\mathbf{A} = ((a_{kt}))$, $\mathbf{a}_k$ is the $1 \times T$ row vector of coordinates for the k-th row of $\mathbf{A}$, and $\mathbf{B} = ((b_{jt}))$. The distribution of each f_{ik} is specified as a conditional multivariate normal:

$$f_{ik}(\boldsymbol{\Delta}_i | \mathbf{a}_k, \mathbf{B}, \boldsymbol{\Sigma}_k) = (2\pi)^{-J/2} |\boldsymbol{\Sigma}_k|^{-1/2} \exp \{-1/2(\boldsymbol{\Delta}_i - \mathbf{a}_k\mathbf{B}')\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\Delta}_i - \mathbf{a}_k\mathbf{B}')'\}, \tag{4}$$

where $\mathbf{a}_k\mathbf{B}'$ represents the scalar products or linear projection of the stimulus points onto cluster k's vector. That is, we assume a random sample of $\boldsymbol{\Delta}_i$ drawn from a mixture of conditional multivariate normal densities of K underlying groups or clusters in unknown proportions $\lambda_1, \lambda_2, \ldots, \lambda_K$ (see McLachlan & Basford, 1988, concerning the use of such mixtures in previous approaches to pattern clustering, as well as Goodman, 1974, and Takane, 1976, for similarities with latent structure analysis and latent profile models).

Given a sample of independent objects, the likelihood has the form:

$$L = \prod_{i=1}^{I} \left[\sum_{k=1}^{K} \lambda_k (2\pi)^{-J/2} |\boldsymbol{\Sigma}_k|^{-1/2} \exp \{-1/2(\boldsymbol{\Delta}_i - \mathbf{a}_k\mathbf{B}')\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\Delta}_i - \mathbf{a}_k\mathbf{B}')'\} \right], \tag{5}$$

or

$$\ln L = \sum_{i=1}^{I} \ln \left[\sum_{k=1}^{K} \lambda_k f_{ik}(\boldsymbol{\Delta}_i | \boldsymbol{\Sigma}_k, \mathbf{a}_k, \mathbf{B}) \right]. \tag{6}$$

Given $\mathbf{\Delta}$, T, and K, the task is to estimate $\boldsymbol{\lambda}$, $\mathbf{\Sigma}$, $\mathbf{A}$, and $\mathbf{B}$ so as to maximize expressions (5) or (6), given the conditions specified in (2) and (3).

A number of relevant issues need to be discussed concerning this particular maximum likelihood framework. One, unlike finite mixtures of other types of density functions, the parameters of finite mixtures of normal densities are identified (see Teicher, 1961, 1963; Yakowitz, 1970; and Yakowitz & Spragins, 1968). Two, there exists no simple sufficient statistics for the parameters of such a normal mixture (Dynkin, 1961). Three, a potential problem of nonuniqueness in parameter values (e.g., $\mathbf{\Sigma}_k$, $\mathbf{a}_k$, $\mathbf{B}$) theoretically exists if $\lambda_k = 0$ for any k. Such a result might occur if an excessive number of clusters were extracted. In such cases, cluster k is eliminated from the solution and K reduced accordingly. Fourth, given the bilinear form of the scalar products $\mathbf{a}_k\mathbf{B}'$, scale and rotational indeterminacies exist with respect to these parameter values. Various rotational and normalization options are described later in the paper. Finally, note that once estimates of $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\lambda}$, and $\mathbf{\Sigma}$ are obtained through a maximum likelihood procedure, each object i may be assigned to each cluster k using the estimated posterior probability (applying Bayes' rule):

$$\hat{p}_{ik} = \frac{\hat{\lambda}_k f_{ik}(\mathbf{\Delta}_i|\hat{\mathbf{a}}_k, \hat{\mathbf{B}}, \hat{\mathbf{\Sigma}}_k)}{\sum_{k=1}^{K} \hat{\lambda}_k f_{ik}(\mathbf{\Delta}_i|\hat{\mathbf{a}}_k, \hat{\mathbf{B}}, \hat{\mathbf{\Sigma}}_k)}, \tag{7}$$

providing a "fuzzy" clustering of each object i into K clusters. Partitions could be formed, if desired, by simply assigning object i to the cluster whose $\hat{p}_{ik}$ was highest (note that hierarchical or overlapping clusters cannot be identified here).

The Algorithm

The maximum likelihood estimates of $\boldsymbol{\lambda}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{\Sigma}$, and $\mathbf{P} = ((p_{ik}))$ are found by initially forming an augmented Lagrangean to reflect the constraints on λ_k:

$$\mathbf{\Phi} = \sum_{i=1}^{I} \ln \left[\sum_{k=1}^{K} \lambda_k f_{ik}(\mathbf{\Delta}_i|\mathbf{a}_k, \mathbf{B}, \mathbf{\Sigma}_k) \right] - u \left(\sum_{k=1}^{K} \lambda_k - 1 \right), \tag{8}$$

where u is the corresponding multiplier. The resulting maximum likelihood stationary equations are obtained by equating the first order partial derivatives of the augmented log likelihood function in (8) to zero:

$$\frac{\partial \mathbf{\Phi}}{\partial u} = \sum_{k=1}^{K} \lambda_k - 1 = 0; \tag{9}$$

$$\frac{\partial \mathbf{\Phi}}{\partial \lambda_k} = \sum_{i=1}^{I} \frac{1}{\sum_k \lambda_k f_{ik}(\cdot)} f_{ik}(\cdot) - u = 0; \tag{10}$$

$$\frac{\partial \mathbf{\Phi}}{\partial \mathbf{a}_k} = \sum_{i=1}^{I} \frac{1}{\sum_k \lambda_k f_{ik}(\cdot)} \lambda_k f_{ik}(\cdot)(\mathbf{B}'\mathbf{\Sigma}_k^{-1}\mathbf{\Delta}_i' - \mathbf{B}'\mathbf{\Sigma}_k^{-1}\mathbf{B}\mathbf{a}_k') = \mathbf{0}; \tag{11}$$

$$\frac{\partial \Phi}{\partial \mathbf{B}} = \sum_{i=1}^{I} \frac{1}{\sum_k \lambda_k f_{ik}(\cdot)} \sum_{k=1}^{K} \lambda_k f_{ik}(\cdot) \mathbf{\Sigma}_k^{-1}(\mathbf{\Delta}_i - \mathbf{a}_k \mathbf{B}')' \mathbf{a}_k = \mathbf{0}; \tag{12}$$

$$\frac{\partial \Phi}{\partial \Sigma_k^{-1}(r, s)} = \sum_{i=1}^{I} \frac{\lambda_k f_{ik}(\cdot)}{\sum_k \lambda_k f_{ik}(\cdot)} (1 - \delta_{rs}/2)[\Sigma_k(r, s) - (\Delta_{ir} - \mathbf{a}_k \mathbf{B}_r')(\Delta_{is} - \mathbf{a}_k \mathbf{B}_s')'] = 0; \tag{13}$$

of if all covariance matrices are assumed equal,

$$\frac{\partial \Phi}{\partial \Sigma^{-1}(r, s)} = \sum_{i=1}^{I} \frac{1}{\sum_k \lambda_k f_{ik}(\cdot)} \sum_{k=1}^{K} \lambda_k f_{ik}(\cdot)(1 - \delta_{rs}/2) \cdot [\Sigma(r, s) - (\Delta_{ir} - \mathbf{a}_k \mathbf{B}_r')(\Delta_{is} - \mathbf{a}_k \mathbf{B}_s')'] = 0, \tag{14}$$

where $\Sigma_k^{-1}(r, s)$ denotes the rs element of $\mathbf{\Sigma}_k^{-1}$, $\Sigma_k(r, s)$ denotes the rs element of $\mathbf{\Sigma}_k$, and δ_{rs} is the Kronecker delta. To estimate u, we multiply both sides of (10) by λ_k and sum over k:

$$\sum_{i=1}^{I} \frac{\sum_k \lambda_k f_{ik}(\cdot)}{\sum_k \lambda_k f_{ik}(\cdot)} - u \sum_k \lambda_k = 0,$$

or

$$\hat{u} = I. \tag{15}$$

To estimate λ_k, we multiply both sides of (10) by λ_k and simplify:

$$\sum_{i=1}^{I} \frac{\lambda_k f_{ik}(\cdot)}{\sum_k \lambda_k f_{ik}(\cdot)} - \lambda_k u = 0, \tag{16}$$

or

$$\sum_{i=1}^{I} \hat{p}_{ik} - \lambda_k I = 0, \tag{17}$$

and

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{I} \hat{p}_{ik}}{I}. \tag{18}$$

To estimate $\mathbf{a}_k$, we expand (11) and substitute for $\hat{p}_{ik}$:

$$\sum_{i=1}^{I} \hat{p}_{ik}(\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_k^{-1}\mathbf{\Delta}_i' - \hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_k^{-1}\hat{\mathbf{B}}'\hat{\mathbf{a}}_k') = \mathbf{0}, \tag{19}$$

or

$$\hat{\mathbf{a}}_k' = \left(\sum_{i=1}^{I} \hat{p}_{ik}\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_k^{-1}\hat{\mathbf{B}}\right)^{-1}\left(\sum_{i=1}^{I} \hat{p}_{ik}\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_k^{-1}\mathbf{\Delta}_i'\right). \tag{20}$$

By expanding and substituting in (13) and (14), analytical expressions for $\hat{\mathbf{\Sigma}}_k$ or $\hat{\mathbf{\Sigma}}$ can be obtained using:

$$\hat{\mathbf{\Sigma}}_k = \frac{1}{I\hat{\lambda}_k} \sum_{i=1}^{I} \hat{p}_{ik}(\mathbf{\Delta}_i - \hat{\mathbf{a}}_k\hat{\mathbf{B}}')'(\mathbf{\Delta}_i - \hat{\mathbf{a}}_k\hat{\mathbf{B}}'), \tag{21}$$

or

$$\hat{\mathbf{\Sigma}} = \frac{1}{I} \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}(\mathbf{\Delta}_i - \hat{\mathbf{a}}_k\hat{\mathbf{B}}')'(\mathbf{\Delta}_i - \hat{\mathbf{a}}_k\hat{\mathbf{B}}'). \tag{22}$$

To estimate $\mathbf{B}$, the following stationary equation is derived from (12):

$$\sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\, \hat{\mathbf{\Sigma}}_k^{-1}(\mathbf{\Delta}_i' - \hat{\mathbf{B}}\hat{\mathbf{a}}_k')\hat{\mathbf{a}}_k = 0, \tag{23}$$

or

$$\sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{\Sigma}}_k^{-1}\mathbf{\Delta}_i'\hat{\mathbf{a}}_k = \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{\Sigma}}_k^{-1}\hat{\mathbf{B}}\hat{\mathbf{a}}_k'\hat{\mathbf{a}}_k. \tag{24}$$

Taking the transpose of both sides and postmultiplying by $\hat{\mathbf{\Sigma}}_k$,

$$\sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{a}}_k'\mathbf{\Delta}_i = \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{a}}_k'\hat{\mathbf{a}}_k\hat{\mathbf{B}}', \tag{25}$$

and

$$\hat{\mathbf{B}}' = \left(\sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{a}}_k'\hat{\mathbf{a}}_k\right)^{-1}\left(\sum_{i=1}^{I} \sum_{k=1}^{K} \hat{p}_{ik}\hat{\mathbf{a}}_k'\mathbf{\Delta}_i\right). \tag{26}$$

(The unique inverse exists for $K \geq T$ with $\hat{\lambda}_k \neq 0$ for all k. Note, since K vectors can always be perfectly fitted in a K-dimensional space, solutions with $T > K$ will not improve the log-likelihood over corresponding $T = K$ solutions, and thus will be ignored.) In this formulation, the maximum likelihood equation for estimating $\mathbf{\Sigma}_k$ (or $\mathbf{\Sigma}$), $\mathbf{a}_k$, and $\mathbf{B}$ are weighted averages of the maximum likelihood equations $\partial \log f_{ik}(\cdot)/\partial M = 0$, where M denotes the parameter of interest, arising from each component separately, and the weights are the posterior probabilities of membership of the objects in each cluster. This specific structure lends itself to the application of a two-stage E-M algorithm (Dempster, Laird, & Rubin, 1977) for the iterative estimation of these parameters. In the E stage, one estimates λ_k and p_{ik} by (7) and (18). In the M stage, one estimates $\mathbf{a}_k$ by (20), $\Sigma_k(\Sigma)$ by (21)((22)), and $\mathbf{B}'$ by (26). For specified initial values of these parameters, the expectation (E phase) and maximization (M phase) steps of this algorithm are alternated until convergence of a sequence of log-likelihood values is obtained. (See Sclove, 1977; and Titterington, Smith, and Makov, 1985, for similar applications of the E-M procedure for estimating the parameters of finite mixture distributions.)

Tests for T and K

The MULTICLUS method must be performed for various values of the dimensionality of the underlying space (T) as well as for varying numbers of underlying clusters (K). Assuming a concern to examine analyses for $T \le K = 1, \ldots, 4$, some 10 analyses must be undertaken. To identify the "best" values for T and K, one obvious way is to use the likelihood-ratio test statistic to test for the smallest values of T and K compatible with the data. However, according to McLachlan and Basford (1988), regularity conditions do not hold for the differences in $-2 \ln L$ for nested models to have their usual asymptotic null distribution of chi-squared with degrees of freedom equal to the difference in the number of model parameters. Here, there is the lack of an obvious natural saturated model for use in such a statistical test. Recently, Sclove (1983) and Bozdogan and Sclove (1984) have proposed using Akaike's (1974) information criterion (AIC) for the choice of the number of groups in mixture clustering models. Accordingly, in MULTICLUS, one would select K and T that minimizes:

$$\text{AIC}(K, T) = -2 \ln L + 2N(K, T), \tag{27}$$

where $N(K, T)$ is the number of free parameters for the full MULTICLUS model:

$$N(K, T) = KT + JT + K\left(\frac{J(J-1)}{2}\right) + K - 1 - T^2. \tag{28}$$

Although we shall adopt this AIC heuristic to select appropriate values of K and T, as pointed out by Titterington, Smith, and Makov (1985), Bozdogan (1983), and Sclove (1987), this AIC criterion here relies essentially on the same regularity conditions needed for differences in $-2 \ln L$ to have its usual asymptotic distribution under the null hypothesis. In addition, Bozdogan (1987) found that use of the AIC tended to result in overfitting the true dimensionality of certain models. As such, we regard such measures as "heuristic figures-of-merit." (McLachlan, 1987, has more recently proposed a bootstrapping procedure for selecting the number of components in a mixture that indeed looks promising for potential application here.) We also recommend the inspection of other goodness of fit measures. For example, one should also examine a variance-accounted-for (VAF) measure between $\mathbf{\Delta}_i$ and $\hat{\mathbf{\Delta}}_i = \Sigma_{k=1}^{k} \hat{p}_{ik}\hat{\mathbf{a}}_k\hat{\mathbf{B}}'$, across all i, for values of T and K, as is done in traditional metric multidimensional scaling and some forms of non-hierarchical cluster analysis.

Program Options

A number of options have been programmed in MULTICLUS. There are a number of preprocessing options (e.g., row/column centering, normalization, standardization, etc.) for $\mathbf{\Delta}$ prior to estimation. However, such preprocessing can have direct impact on the ability to estimate model parameters. For example, row centering or standardization for $I > J$ reduces the column rank of $\mathbf{\Delta}$ by one and we cannot estimate a nonsingular

TABLE 1

Independent Factors for Monte Carlo Analysis

Factor	Levels
X_1 : Number of Rows in $\underset{\sim}{\Delta}$ (I)	40 60
X_2 : Number of Columns in $\underset{\sim}{\Delta}$ (J)	14 22
X_3 : Number of Clusters (K)	4 6
X_4 : Number of Dimensions (T)	2 4
X_5 : Error in $\underset{\sim}{\Delta}$	$N(0, .1\sigma^2)$ $N(0, .2\sigma^2)$

σ^2 = variance of $\underset{\sim}{\Delta}$

Σ or Σ_k. There are also implications for (20) where generalized inverses must be employed for certain preprocessing options. The user can provide his/her own starting values, have them generated randomly, or use a singular value decomposition (SVD) to generate **B** (the default option). Options also exist for external or internal analyses where **A** and/or **B** can be given and held fixed throughout the analysis. For estimating the covariance matrices, one can hold Σ_k fixed, estimate one common covariance matrix Σ for all groups, constrain Σ_k to be the identity matrix, or estimate Σ_k freely. Options also exist for estimating diagonal Σ or Σ_k. Finally, given the well known scale and rotational indeterminacies of such bilinear vector models, a number of orthogonal rotations (e.g., simple structure rotations with respect to **B**) and normalization options are available.

3. MULTICLUS Monte Carlo Analysis

To examine the performance of the MULTICLUS methodology, a Monte Carlo analysis was performed. The objective was to investigate the performance of the MULTICLUS algorithm as a number of data, parameter, and error factors were experimentally manipulated. Table 1 presents a list of the five independent factors utilized in this Monte Carlo analysis. In addition to factors manipulating the size of the input data $\mathbf{\Delta}$ (e.g., I and J), factors influencing the number of parameters to estimate (T and K, where $K \geq T$), and error levels were also specified. Table 1 also provides labels $X_1, \ldots, X_5$ for these five independent factors. Some eleven dependent variables were collected to measure performance in terms of overall solution quality, parameter recovery, and computational effort: the number of major iterations required for convergence (NI), an overall variance-accounted-for statistic (VAF), an adjusted VAF statistic (ADJ) (adjusted for degrees of freedom), a root-mean-square between $\mathbf{P}$ and $\hat{\mathbf{P}}$

(RMS(**P**)) after appropriate column permutation, an overall correlation coefficient between **P** and $\hat{\mathbf{P}}$ (COR(**P**)) after appropriate column permutation, a root-mean-square between $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$ (RMS($\boldsymbol{\lambda}$)) after appropriate permutation, a correlation coefficient computed between $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$(COR($\boldsymbol{\lambda}$)) after appropriate permutation, a sums-of-squares-accounted-for statistic between **B** and $\hat{\mathbf{B}}$ after Procrustean transformation to $\hat{\mathbf{B}}$(SSAF(**B**)) (and corresponding adjustment made to **A**), a variance-accounted-for statistic between **B** and $\hat{\mathbf{B}}$(VAF(**B**)) after Procrustean transformation to $\hat{\mathbf{B}}$ (and corresponding adjustments made to **A**), a sum-of-squares-accounted-for statistic between **A** and $\hat{\mathbf{A}}$(SSAF(**A**)), and a variance-accounted-for statistic between **A** and $\hat{\mathbf{A}}$ (VAF(**A**)). Thus, measures were used to measure computational effort (NI), overall data recovery (VAF, ADJ), *P* recovery (RMS(**P**), COR(**P**)), $\boldsymbol{\lambda}$ recovery (RMS($\boldsymbol{\lambda}$), COR($\boldsymbol{\lambda}$)), **B** recovery (SSAF(**B**), VAF(**B**)), and **A** recovery (SSAF(**A**), VAF(**A**)).

A full factorial design was utilized with two replications per cell. An analysis of variance was performed for each of the 11 dependent measures. Note, four covariates were also identified and utilized in the analysis relating to estimates of the separation of the randomly generated cluster vectors ($\mathbf{a}_k$): the minimum angle between vectors (MIN(**A**)), the maximum angle between vectors (MAX(**A**)), the median angle between vectors (MED(**A**)), and the standard deviation of the angles between the vectors (SD(**A**)). These four variables were used as covariates in each analysis to control for the distribution of the randomly generated $\mathbf{a}_k$ vectors since, for example, if the *K* vectors in **A** were generated very close to each other, MULTICLUS performance could possibly suffer. Future Monte Carlo work on this may consider making such covariates actual factors in the design. Prior to discussing the results, two issues are noted. First, arcsine (of square roots) transformations were performed and ANOVA analyses done on those dependent measures ranging between 0-1 with somewhat similar results, and are thus not reported for sake of brevity. Finally, for discussing interaction effects, corresponding tables of means were calculated to examine crossover interactions, but are also not displayed. A verbal summary of the results follows.

For computational effort, the size of $\boldsymbol{\Delta}$ (*I* in particular) significantly affects the number of iterations required for convergence (i.e., larger *I* significantly increases NI). Concerning overall data recovery, both VAF and adjusted VAF measures are significantly affected by the error factor (X_5), where the higher error level leads to poorer recovery in each case. Recovery of both **P** and $\boldsymbol{\lambda}$ appears to be reasonably robust to these five independent factors, with only one interaction ($X_1X_3X_5$) affecting one of these four dependent measures—RMS($\boldsymbol{\lambda}$). Here, the mean RMS($\boldsymbol{\lambda}$) is slightly higher across all levels of X_1 and X_3 for the lower error condition, except for the ($X_1 = 1$, $X_3 = 0$) case where the RMS($\boldsymbol{\lambda}$) mean for the higher error level is nearly three times that for the lower error level. The recovery of the stimulus points (**B**) is only significantly affected by the $X_2X_3X_4$ interaction in both dependent measures here. Upon inspection of the tables of means, it appears that in both **B** recovery measures, recovery is nearly twice as high in the low *T* level versus high *T* level for the ($X_2 = X_3 = 0$) and ($X_2 = X_3 = 1$) cells, although they are about the same magnitude in the remaining X_2 and X_3 cells. Finally, with respect to **A** recovery, one sees the first evidence of a significant covariate, MIN(**A**), which indicates that recovery of **A** is in general better as the minimum angle among all the $\mathbf{a}_k$ vectors gets larger (i.e., they are spread further apart). The SSAF(**A**) dependent measure is significantly affected only by an $X_3X_4X_5$ interaction where only for the $X_3 = X_4 = 1$ cell does the higher error condition produce a slightly higher mean SSAF(**A**). An $X_1X_2X_5$ interaction is the only significant effect on VAF(**A**) where only for the ($X_1 = 1$, $X_2 = 0$) cell does the higher error level condition produce a slightly higher mean VAF(**A**).

In summary, the Monte Carlo analysis has provided some preliminary evidence concerning the somewhat robust performance of the MULTICLUS methodology over data, parameter, and error related independent factors. All else considered, larger data sets with more parameters to estimate involve higher computational effort. Higher levels of error added to $\mathbf{\Delta}$ will significantly detract from overall data recovery, although no main effects for error were observed with any of the parameter recovery measures. A few scattered significant higher order interaction terms occurred in the $\mathbf{A}$ and $\mathbf{B}$ recovery measures. Certainly such results must be deemed as preliminary given the modest scope of this endeavor. Future work in this area should involve additional complexities such as incorporating additional factors (e.g., type of start, internal versus external analysis, separation of $\mathbf{a}_k$), specifying more levels per factor (especially concerning the error factor), having more replications per cell, and so forth.

4. MULTICLUS Application

Study Background

Currently, one of the most active areas of research in consumer psychology concerns the influence of prior knowledge on differential evaluations of product characteristics. Alba and Hutchinson (1987), in a recent review, propose that a consumer's ability to differentiate relevant from irrelevant information about a product improves as prior knowledge about the product increases. This is often reflected in less knowledgeable consumers utilizing peripheral product information in formulating evaluative judgments, although those who are more knowledgeable focus on functional product characteristics (also, see Park & Lessig, 1981). Consumers with different knowledge bases may also utilize essentially the same product information when forming evaluations, but weigh that information differently with respect to its relative importance in any given task. Because they lack the ability to comprehend the nature and importance of the information with which they are confronted, less knowledgeable consumers often give disproportionately higher weight to product characteristics that are easily understood and/or frequently emphasized in promotional campaigns (also, see Bettman & Park, 1980; Sujan, 1985).

Our study will examine how different levels of consumer knowledge systematically affect the types of information that consumers judge to be important in evaluating microcomputers. Consistent with the above noted literature, we hypothesized that persons with high levels of prior knowledge with microcomputers would weigh functional attributes as more important, whereas those with low levels of prior knowledge would weigh attributes that are simple, more superficial, and easier to understand as more important in a purchase decision. In this study, we operationalized prior knowledge as a function of product usage experience and self-perceived familiarity/expertise.

Measures of prior knowledge that reflect respondents' perceptions of their familiarity and/or expertise in a domain have been frequently used in previous studies (see Alba & Marmorstein, 1987; Alba, 1983; Bettman & Sujan, 1987; Brucks, 1985; Hutchinson & Brady, 1982). Product usage experience has also been frequently used as an indicator of prior knowledge, most typically operationalized as extent of prior product usage and product ownership (see Bettman & Park, 1980; Jacoby, Chestnumt, & Fisher, 1978; Newman & Staelin, 1972; Park & Lessig, 1981). In our study, product ownership will be used as an indicator of prior knowledge. Alba and Hutchinson (1987) note that the experience of making a purchase decision for products that are relatively high in cost (e.g., microcomputers) is likely to be an effective means of bringing someone to differentiate between relevant and irrelevant product attributes. Extent of product usage will also be used as an indicator of prior knowledge and defined in terms of variability in usage experience. Alba and Hutchinson further note that experts are more likely than novices to search for and acquire new information in a domain. Thus, it seems reasonable to infer that higher levels of product knowledge will be associated with a more elaborate array of product usage experiences.

In summary, it was expected that an individual's level of prior knowledge concerning microcomputers will influence his/her evaluation of the attributes of a microcomputer. Specifically, those with higher levels of prior knowledge (as defined by ownership, variability in usage experience, and self-perceived familiarity/expertise) will weigh functional attributes (as discussed below) as more important in purchasing a microcomputer, and those with lower levels of prior knowledge will weigh attributes that are easy to understand, or allow simple quality inferences to be made, as more important in purchasing a microcomputer.

Study Description

Sixty-nine graduate and undergraduate students enrolled in the business school of a large Southwestern university were administered a survey instrument to measure their evaluations of the importance of various attributes/features of microcomputers. Based on the prior responses of a team of microcomputer experts (owners and heavy users) and novices (nonowners and nonusers), a list of 15 microcomputer product attributes were utilized summarizing the most important aspects of a microcomputer. This pretest group was also asked for judgments on whether the attributes were easy to understand and utilize as simple inferential cues concerning product quality, or whether they were characteristics one would come to appreciate/understand through extensive usage and the experiences associated with purchase/leasing decision making. Seven attributes were judged to be best described by the first classification and were expected to be weighted higher in importance by persons with low prior knowledge: (a) ease of use; (b) total package purchase/lease price; (c) length of warranty; (d) compatibility with IBM products/software; (e) the particular brand name or manufacturer; (f) physical styling of the computer; and (g) portability. Hereafter, these seven attributes will be referred to as "heuristic" product attributes. Eight additional attributes were judged to be best described by the alternative criterion noted above: (a) the type (PC, XT, AT) of computer; (b) the amount of internal RAM; (c) whether there is a hard disk; (d) size of the hard disk; (e) type of monitor/screen (e.g., monochrome, color, EGA); (f) number of floppy disk drives; (g) processing speed; and (h) software availability. Hereafter, these eight attributes will be referred to as "focal" product attributes. All 15 attributes were presented to the respondents in this study who were asked "regardless of whether you currently own/lease a microcomputer, indicate the relative importance of the below listed features to you in your purchasing a microcomputer." All items were scored on seven point scales (not at all important—very important).

MULTICLUS Analysis

Table 2 presents the statistical summary of the results of MULTICLUS analyses on the row standardized data (with the $\boldsymbol{\Sigma}_k = \mathbf{I}$ restriction since $\boldsymbol{\Sigma}_k \neq \mathbf{I}$ is not compatible with this preprocessing selection as previously discussed) for $T \leq K = 1, \ldots, 4$. As can be seen, the minimum AIC statistic occurs for $T = K = 2$ (VAF = 0.426), whose solution we will present below. The lower portion of Table 2 also provides some justification for this solution based on the variance-accounted-for measure, where a noticeable jump is seen in going from $T = 1$ to $T = 2$ and from $K = 1$ to $K = 2$. As noted earlier, the likelihood and VAF goodness of fit indicies do not improve for solutions with $T > K$; both the log likelihood and VAF goodness-of-fit measures for solutions

TABLE 2

MULTICLUS Statistical Results for the Microcomputer Evaluations Data

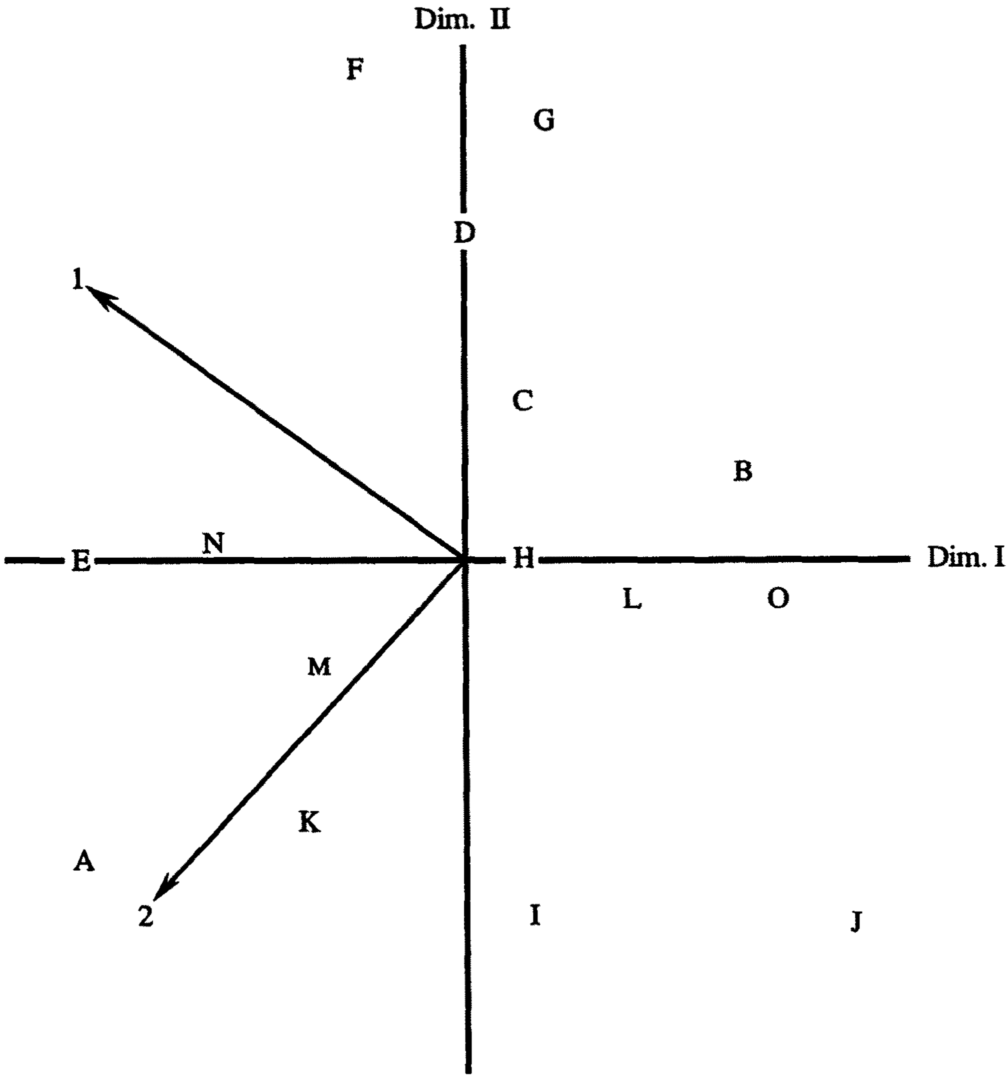
K	T	Number of Iterations	ln L	N(K,T)	AIC
1	1	2	-1341.3	15	2712.6
2	1	8	-1341.3	17	2716.6
2	2	7	-1300.5	31	2662.9*
3	1	9	-1341.3	19	2720.6
3	2	16	-1300.5	34	2668.9
3	3	53	-1293.2	47	2680.3
4	1	10	-1341.3	21	2724.6
4	2	15	-1300.5	37	2674.9
4	3	42	-1292.6	51	2687.2
4	4	33	-1287.7	63	2701.5

VAF

	K = 1	K = 2	K = 3	K = 4
T = 1	0.246	0.246	0.246	0.246
T = 2		0.426	0.426	0.426
T = 3			0.474	0.490
T = 4				0.511

*minimum AIC

with $T > K$ were identical to corresponding solutions with $T = K$ and thus are not reported. Figure 1 presents the MULTICLUS results for this solution ($\lambda_1 = 0.52$, $\lambda_2 = 0.48$) where the 15 attributes/features are designated by the letters A-O and the two cluster vectors by "1" and "2". The first dimension appears to separate attributes that relate specifically to user operating convenience, where software availability (E), IBM compatability (N), and ease of use (A) lie on the extreme left end of the dimension, from those attributes which can be easily inferred through visual observation where brand name (B), styling (O), and portability (J) lie on the extreme right side. The second dimension clearly distinguishes the "focal" attributes such as hard disk (F), its size (G), and internal RAM (D) from the "heuristic" attributes such as price (K), warranty length (I), ease of use (A), and portability (J). Of interest is the positioning of the two vectors for the two groups or clusters. In particular, the focal characteristics such as hard disk and size, internal RAM, software availability and processing speed project higher on the vector for cluster one, while ease of use, package price and warranty



A : ease of use
B : brand name/manufacturer
C : type of computer
D : amount of internal RAM
E : software availability
F : whether there is a hard disk
G : size of hard disk
H : type of monitor
I : length of warranty
J : portability
K : total package purchase/lease price
L : number of floppy drives
M : processing speed
N : compatibility with IBM products/system
O : styling

FIGURE 1
MULTICLUS configuration of attributes and cluster vectors for the microcomputer data.

TABLE 3

Mean Attribute Scores by Group Membership

Attribute	Group 1	Group 2	F
Ease of use	6.08	6.64	7.50
Total package purchase/lease price	5.34	6.02	4.29
Length of warranty	4.60	5.70	11.45
Compatibility with IBM products/software	5.15	6.08	5.44
Physical styling of the computer	4.14	4.44	0.86
Portability	3.54	4.97	13.72
Particular brand name or manufacturer	4.48	4.11	1.08
Type of computer	5.37	4.38	7.73
Amount of internal RAM	5.80	4.12	30.52
Whether there is a hard disk	6.28	3.73	58.83
Size of the hard disk	5.74	3.50	50.46
Type of monitor screen	5.11	4.76	1.18
Number of floppy disk drives	4.51	4.50	0.00
Software availability	6.54	5.85	8.21
Processing speed	5.57	5.17	0.13

length (heuristic characteristics) project higher on cluster two's vector. Thus, cluster one evaluates the more functional and technical (focal) features of microcomputers as most important, while cluster two appears to weigh the financial and ease of use (heuristic) aspects more heavily. Note, we ran the designated $T = K = 2$ solution in MULTICLUS some 100 times with different starts (using the SVD default option for $\mathbf{B}$) to examine the potential problems associated with obtaining locally optimum solutions. The solution reported above was recovered (up to rotation) in 99 of the 100 computer runs (this dropped to 89 out of 100 for pure random starts on all parameters). Twenty five such runs were also performed for the other 9 solutions with similar results. Thus, there is little variation in the quality of the solution recovered. The next subsection will attempt to investigate the basis for the cluster assignment in terms of prior knowledge and experience.

Explaining the Group Structure

Prior knowledge of microcomputers was operationalized through two indicators of product usage experience (microcomputer ownership and variability of usage) and a scale measuring self-perceived familiarity/expertise with microcomputers. Microcomputer ownership was measured with an item that asked respondents if they currently owned a microcomputer. Twenty-three persons stated "yes" and forty-one persons stated "no." Five persons indicated they currently leased a microcomputer and were placed in the ownership category. Variability in usage was measured with an item that asked respondents to indicate the typical uses they make of a microcomputer: word processing, communication with mainframe computer, spreadsheet/accounting calculations, games/hobbies, statistical packages, computer programming, accessing databases, graphics, other external communications (e.g., electronic mail). A score of 1 was

TABLE 4

Two Group Discriminant Analysis and Indicators of Prior Knowledge

Variable	Function Coefficient	Wilk's Lambda
Product ownership	0.38	0.66**
Self-reported familiarity/expertise	0.61	0.55**
Usage variability	0.45	0.51**

*p < .05
**p < .01

given for each usage indicated and then summed to yield a variability in product usage score. Scores ranged from 0 to 9.

Self-perceived familiarity/expertise was measured through eight items scaled strongly agree to strongly disagree using a five point scale:

1. When it comes to *purchasing* a microcomputer, I consider myself knowledgeable about the microcomputer market.
2. When it comes to *using* a microcomputer, I consider myself knowledgeable about microcomputers.
3. I often have anxiety about having to use a computer.
4. Compared to other students, I probably spend more time than most using a microcomputer.
5. Many fellow students often seek my advice about microcomputers.
6. I would rather use a microcomputer with fewer options and more simplicity than one with more options and more complexity.
7. I often read the latest computer magazines.
8. If I can, I prefer to do my work and/or school related assignments on a microcomputer.

As seen, the above items range from requests for respondents' direct impressions of their familiarity/expertise (e.g., Item 2) to those that indirectly indicate familiarity/expertise (e.g., Item 3). Cronbach's alpha reliability coefficient for the eight item scale was 0.90. Therefore, the eight items were summed to yield a score on self-perceived familiarity/expertise, which was used as the final indicator of prior knowledge. In summary, three independent variables were employed as indicators of prior knowledge (microcomputer ownership, usage variability, and a scale measuring self-perceived familiarity/expertise with microcomputers), and were used to explain group structure. The criterion variable was dichotomously defined. If the probability ($\hat{p}_{i1}$) of Group 1 membership for any subject was greater than 0.50, the subject was assigned to cluster one. If the probability of Group 1 membership for any subject was less than or equal to 0.50, the subject was assigned to Cluster 2. Table 3 provides the group means for the importance of each product attribute and a (pseudo) F-test (given the clusters were initially found on the basis of this data) of the mean difference. Notice how the means for the focal attributes are uniformly higher for Group 1 than for Group 2, while the opposite is true for the heuristic attributes.

A discriminant analysis was performed attempting to predict the dichotomous group membership as a function of the three indicators of prior product knowledge. The discriminant function was highly significant ($\chi^2 = 43.61$; $p < 0.001$) with each predictor significantly explaining variation in group membership (see Table 4). For each predictor, high product knowledge was found to be predictive of membership in the group (Group 1) that tended to judge the focal attributes as more important when purchasing a microcomputer. Use of the three indicators of prior knowledge resulted in 85.51% ($N = 59$) of the cases being correctly classified.

Thus, what we know, or think we know, about a product influences the importance that we attribute to different characteristics of that product. Consumer psychologists have traditionally considered prior product knowledge as a unidimensional construct (Brucks, 1985; Alba & Hutchinson, 1987). This study suggests that different operationalizations of prior knowledge can uniquely contribute to variations in product attribute evaluations. Thus, a multidimensional conceptualization of the prior product knowledge concept may be more appropriate as shown via MULTICLUS.

References

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control, 6*, 716–723.

Alba, J. W. (1983). The effects of product knowledge on the comprehension, retention and evaluation of product information. In R. P. Bagozzi & A. M. Tybout (Eds.), *Advances in consumer research* (Vol. 10, pp. 577–580). Ann Arbor, MI: Association for Consumer Research.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13*, 411–454.

Alba, J. W., & Marmorstein, H. (1987). The effects of frequency knowledge on consumer decision making. *Journal of Consumer Research, 14*, 14–25.

Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research, 7*, 234–248.

Bettman, J. R., & Sujan, M. (1987). Effects of framing on evaluation of comparable and noncomparable alternatives by expert and novice consumers. *Journal of Consumer Research, 14*, 141, 154.

Bozdogan, H. (1983). *Determining the number of component clusters in standard multivariate normal mixture models using model-selection criterion* (Technical Report VIC/DQM/A83-1). Washington, DC: Army Research Office.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345–370.

Bozdogan, H., & Sclove, S. L. (1984). Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistical Mathematics, 36*, 163–180.

Brucks, M. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research, 12*, 1–16.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society-B, 39*, 1–38.

Dynkin, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions. *Selected translations in mathematical statistics and probability* (pp. 17–40). Providence, RI: American Mathematical Society.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Hutchinson, J., & Brady, F. (1982). Dimensional validity, consistency of preference and product familiarity: An exploratory investigation of wine testing. In A. A. Mitchell (Ed.), *Advances in consumer research* (Vol. 9, pp. 398–401). St. Louis, MO: Association for Consumer Research.

Jacoby, J., Chestnut, R., & Fisher, W. (1978). A behavioral approach to information acquisition in nondurable purchasing. *Journal of Marketing Research, 15*, 532–544.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36*, 318–324.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. New York: Marcel Dekker.

Newman, J. W., & Staelin, R. (1972). Prepurchase information seeking for new cars and major household appliances. *Journal of Marketing Research*, *9*, 249–257.

Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, *8*, 223–230.

Sclove, S. C. (1977). Population mixture models and clustering algorithms. *Communication in Statistics*, *6*, 417–434.

Sclove, S. L. (1983). Application of the conditional population-mixture model to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-5*, 428–433.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343.

Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research*, *12*, 31–46.

Takane, Y. (1976). A statistical procedure for the latent profile model. *Japanese Psychological Research*, *18*, 82–90.

Teicher, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics*, *32*, 244–248.

Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics*, *34*, 1265–1269.

Titterington, D. M., Smith, A. F. M., & Makov, V. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.

Yakowitz, S. J. (1970). Unsupervised learning and the identification of finite mixtures. *IEEE Transactions Information Theory and Control*, *IT-16*, 330–338.

Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, *39*, 209–214.

Manuscript received 2/1/88
Final version received 2/15/90