PSYCHOMETRIKA—VOL. 48, NO. 2.
JUNE, 1983

INDCLUS: AN INDIVIDUAL DIFFERENCES GENERALIZATION OF THE ADCLUS MODEL AND THE MAPCLUS ALGORITHM

J. DOUGLAS CARROLL

BELL LABORATORIES

PHIPPS ARABIE

UNIVERSITY OF ILLINOIS AT CHAMPAIGN
and
BELL LABORATORIES

We present a new model and associated algorithm, INDCLUS, that generalizes the Shepard-Arabie ADCLUS (ADditive CLUStering) model and the MAPCLUS algorithm, so as to represent in a clustering solution individual differences among subjects or other sources of data. Like MAPCLUS, the INDCLUS generalization utilizes an alternating least squares method combined with a mathematical programming optimization procedure based on a penalty function approach to impose discrete (0,1) constraints on parameters defining cluster membership. All subjects in an INDCLUS analysis are assumed to have a common set of clusters, which are differentially weighted by subjects in order to portray individual differences. As such, INDCLUS provides a (discrete) clustering counterpart to the Carroll-Chang INDSCAL model for (continuous) spatial representations. Finally, we consider possible generalizations of the INDCLUS model and algorithm.

Key words: additive clustering, nonhierarchical clustering, combinatorial optimization, three-way clustering, individual differences clustering.

The ADCLUS (for ADditive CLUStering) model was described in detail by Shepard and Arabie [1979; also see references in that paper for earlier, brief presentations by those authors.]. More concretely, an ADCLUS representation of a stimulus domain consists of a set of m (possibly overlapping) subsets or clusters, each having an associated numerical weight, w_k (where $k = 1, \ldots, m$). For any pair of stimuli, the predicted similarity is simply the sum of the weights of those subsets containing the given pair of stimuli. Shepard and Arabie described the ADCLUS *model* and provided illustrative applications to several data sets. Arabie and Carroll [1980] provided an algorithm called MAPCLUS (for MAthematical Programming CLUStering), differing from that given by Shepard and Arabie [1979] for fitting the same ADCLUS *model.*

The ADCLUS model, as proposed by Shepard and Arabie [1979], and the algorithm for fitting it assumed a single two-way proximities matrix that was one-mode [Tucker, 1964], i.e., both the rows and the columns corresponded to the same set of stimuli (mode) being scaled. Carroll [Note 1, p. 6] suggested an individual differences generalization of the ADCLUS model, to be fitted to a three-way two-mode (subjects × stimuli × stimuli) proximities matrix. Preliminary results for the model and an algorithm for fitting it, both

We are indebted to Seymour Rosenberg for making available the data from Rosenberg and Kim [1975]. Also, this work has benefited from the observations of S. A. Boorman, W. S. DeSarbo, G. Furnas, P. E. Green, L. J. Hubert, L. E. Jones, J. B. Kruskal, S. Pruzansky, D. Schmittlein, E. J. Shoben, S. D. Soli, and anonymous referees.

This research was supported in part by NSF Grant SES82 00441, LEAA Grant 78-NI-AX-0142, and NSF Grant SES80 04815.

Authors' addresses: J. Douglas Carroll, Room 2C-553, Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974; Phipps Arabie, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820.

0033-3123/83/0600-5012$00.75/0

called INDCLUS (for INdividual Differences CLUStering) were given by Carroll and Arabie in 1979 [Note 2; also see Arabie & Carroll, 1980, p. 233].

Overview

Formally, the (two-way) ADCLUS model predicts interstimulus similarity $\hat{s}_{ij}$ (i, $j = 1, \ldots n$, where n is the number of stimuli being clustered) as

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

where w_k is the numerical weight (assumed to be nonnegative) of the k^{th} cluster ($k = 1, \ldots, m-1$); p_{ik} is unity if stimulus i is present in cluster k, otherwise zero. The additive constant c can alternatively be represented as the weight (not assumed to be nonnegative) of an m^{th} cluster comprising the complete set of the n stimuli (sometimes referred to as the "universal set").

The following qualifications to the preceding variables should be noted. The input data to which the model is fitted are assumed to be the $M = n(n-1)/2$ entries constituting a two-way symmetric (or symmetrized) row and column unconditional 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. (Since the data are assumed to be on an interval scale, the particular transformation chosen in no way affects the goodness-of-fit, but does allow for the standardization of various parameters in the program described below.) $\mathbf{S} \equiv \| s_{ij} \|$ will always refer to these transformed proximities, to which the fitted $\hat{\mathbf{S}}$ matrix is being compared. Turning to the $\mathbf{P}$ matrix, note that each column represents one of the $m-1$ subsets (or clusters—we use the terms interchangeably), with the ones of that column defining constituency of stimuli within the respective subset.

The INDCLUS model is written

$$s_{ij}^h \cong \sum_{k=1}^{m-1} w_{hk} p_{ik} p_{jk} + c_h \tag{2}$$

where s_{ij}^h is the similarity between stimuli i and j for subject (or other data source) h ($h = 1, \ldots, H$), w_{hk} is the weight for subject h on subset k, and c_h is the additive constant for subject h.

Thus, we have generalized the ADCLUS model in (1) so as to fit a series of H proximities matrices. All H subjects, conditions, or whatever the sources of data happen to be, have the same set of clusters (defined by the unities in each of the $m-1$ columns of $\mathbf{P}$), but the weights now vary as a joint function of which cluster *and* subject is being considered. (The additive constant c_h also varies over subjects.)

The Arabie and Carroll [1980] MAPCLUS approach to fitting the (two-way) ADCLUS model used a "penalty function" approach based on optimizing, at each stage in an overall iterative process, a composite loss function that combined an A-part (designed to maximize variance accounted for) and a B-part (comprising the penalty function designed to constrain the $\mathbf{P}$ matrix to be of the appropriate form, i.e., all entries either 0 or 1). Although these components of the loss function will be discussed below in detail, we note that because all subjects in the INDCLUS model are assumed to employ the same subsets, but with different patterns of weights for individual subjects, the B-part of the MAPCLUS composite loss function is unchanged as it appears in the INDCLUS composite loss function. In MAPCLUS, the A-part of the loss function was computed for a single input proximities matrix. For INDCLUS, the A-part of the loss function must be summed and normalized over the H sources of data.

INDCLUS can be viewed as a clustering counterpart to the INDSCAL [Carroll &

Chang, 1970] model. In the latter model, all subjects are assumed to employ a common set of continuous dimensions, but these dimensions are differentially emphasized by the subjects, so as to yield a weighted Euclidean space. INDCLUS is a discrete model in which subjects differentially weight a common set of clusters. In principle, INDCLUS is applicable to any data set suitable for INDSCAL, although as argued by Shepard and Arabie [1979], discrete models like INDCLUS would seem to be more appropriate for stimulus domains where the investigator is seeking a discrete portrayal of structure. There is some reason to believe that such discrete structures are most appropriate for "conceptual" (as opposed to purely "perceptual") domains of stimuli [see Pruzansky, Tversky, & Carroll, 1982]. However, even in domains for which a continuous spatial model is more appropriate, a discrete model such as ADCLUS or INDCLUS often provides a useful and informative complementary representation.

The analogy between INDCLUS and INDSCAL breaks down, however, in one interesting detail. The INDCLUS algorithm subsumes MAPCLUS [Arabie & Carroll, 1980] as a special case, whereas INDSCAL [Carroll & Chang, 1970] and related algorithms for fitting that weighted Euclidean model do *not* subsume two-way multidimensional scaling [e.g., as practiced in the KYST2A program of Kruskal, Young, & Seery, Note 3]. That is, if a data analyst specifies during a run of INDCLUS that there is $H = 1$ subject, the result is comparable to a MAPCLUS analysis. In contrast, such a declaration for INDSCAL results in an indeterminate and generally suboptimal solution unless special steps are taken, rather than a solution comparable to a (metric) KYST2A analysis, or to the "classical" two-way metric approach to multidimensional scaling [Torgerson, 1958].

The Algorithm

The Penalty Function Approach

As with MAPCLUS, the loss function for INDCLUS takes the form

$$L_k(\alpha_k, \beta_k, \Delta, \mathbf{P}) = \alpha_k A_k + \beta_k B_k. \tag{3}$$

Considering first the left side of (3), 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 [originally called NILES or NIPALS by Wold, 1966] which underlies the iterative fitting in turn of each subset p_{ik} $(i = 1, \ldots, n)$ and its associated weights w_{hk}. Since we are only fitting the k^{th} subset at any instant, Δ in (3) refers to the residuals (centered individually for each subject) computed for the remaining $m - 2$ subsets, so that the reader may wish to associate an implicit subscript "k" and superscript "h" with Δ.

In the right side of (3), the term $\alpha_k A_k$ is the product of the coefficient α_k and the normalized sum of squared error, A_k. Specifically,

$$A_k = \frac{a_k}{d_k}, \tag{4}$$

where

$$a_k = \frac{1}{H} \sum_{h=1}^{H} \sum_{i>j}^{n} \sum^{n-1} (\delta_{ij}^h - w_{hk} p_{ik} p_{jk})^2 \tag{5}$$

and

$$d_k = \frac{4 \sum_{h=1}^{H} \sum_{i>j}^{n} \sum^{n-1} (\delta_{ij}^h)^2}{MH}. \tag{6}$$

Minimizing a_k, the squared error summed over the entries in all subjects' matrices, is equivalent to maximizing the variance accounted for (VAF). The denominator of A_k, d_k, is a normalization factor proportional to the variance of the residuals, δ_{ij}^h, computed over the subsets other than k. (The residuals δ^h are so defined that their means $\bar{\delta}^h$ are all zero. Arabie and Carroll [1980] gave the technical explanation for the normalizing constant of 4 in (6).)

The B-part of the loss function (presented below) is designed to enforce the constraint that $p_{ik} = 0,1$ even though the p_{ik} are initially allowed to vary continuously. Elaborating on the right side of (3), we have

$$B_k = \frac{u_k}{v_k}, \tag{7}$$

and

$$u_k = \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n}[(p_{ik}p_{jk} - 1)p_{ik}p_{jk}]^2, \tag{8}$$

and

$$v_k = \sum_{i>j}^{n}\sum^{n-1}(p_{ik}p_{jk} - T_k)^2, \tag{9}$$

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

$$T_k = \frac{1}{M}\sum_{i>j}^{n}\sum^{n-1} p_{ik}p_{jk}. \tag{10}$$

The rationale for the particular form of the B-part, as well as some unsuccessful alternatives, is given in Arabie and Carroll [1980]. The crucial point to be made in the present instance is that B is in no way dependent on the h^{th} subject, since all subjects are assumed to have the *same* indicators of cluster membership p_{ik} ($i = 1, \ldots, n$; $k = 1, \ldots, m - 1$).

As noted below, INDCLUS uses a gradient procedure that employs the partial derivatives of L_k in (3). Since we require weights α and β of the A- and B-parts, respectively, to obey the normalizing condition that $\alpha + \beta = 1$ (with $\alpha, \beta > 0$) as we fit each subset in turn (see discussion of an "inner" iteration below), L_k is a convex combination of A_k and B_k. After the gradient procedure has maximized L_k for fixed initial values of α and β [see Arabie & Carroll, 1980 for details], then the value of β is increased, so as to place more emphasis on the B-part (or penalty function). By successively closer continuous approximations, the B-part enforces the discrete constraint and, in conjunction with the nonlinear function A_k, qualifies our approach as one of "mathematical programming." Ultimately, the entries in **P** are "polished" (rounded) to be exactly 0,1. Consistent with the alternating least squares approach, this maximization procedure is iteratively applied (conditionally) to each subset in turn.

As noted, INDCLUS uses a gradient procedure to minimize the objective function in (3). Since the B-part of that function is unchanged for the two-way (MAPCLUS) case, we refer the reader to (12), (14), and (15) in the 1980 paper for the relevant partial derivatives. However, the partial derivative for the A-part for INDCLUS is

$$\frac{\partial a_k}{\partial p_{ik}} = -\frac{2}{H}\sum_{h=1}^{H} w_{hk}\sum_{j\neq i} p_{jk}(\delta_{ij}^h - w_{hk}p_{ik}p_{jk}). \tag{11}$$

The weights α_k and β_k in the loss function of (3) are adjusted according to the description given in Arabie and Carroll [1980, p. 216], although the types of normalization used

in INDCLUS for matrix-unconditional as well as matrix-conditional input proximities data (see discussion below) produce a slightly different pattern of balance between the α and β, even though the formulae are unchanged from the MAPCLUS program.

The Alternating Least Squares Structure of INDCLUS

The hallmark of an alternating least squares procedure is, of course, iterative computing nested to several levels of depth. In the MAPCLUS paper [Arabie & Carroll, 1980, pp. 217–222], we distinguished between major, outer, and inner iterations. The last of the three is effectively the combustion chamber of the engine that drives INDCLUS and MAPCLUS. For the latter- and simpler-algorithm, an inner iteration consisted of obtaining an estimate via univariate linear regression of the weight w_k and the additive constant c_k specific to subset k, as well as moving each of the p_{ik} ($i = 1, \ldots, n$) elements by one step of the computed gradient.

For fitting a three-way data set used as input to INDCLUS, there are now H times as many weights and additive constants e_k^h ($h = 1, \ldots, H$) for the subsets, so that an INDCLUS "inner loop" includes an innermost loop for each of the H subjects (or other sources of data). For the equations describing an inner iteration in the MAPCLUS paper, if h is added as a superscript for the residuals δ_{ij} and as a subscript for the weights w_k and their additive constants c_k (now renamed as e_k), then the equations provide an appropriate description for INDCLUS' loop structure and are therefore not repeated in the present paper.

Estimating the Weights

Just as in MAPCLUS, regression can be invoked either (a) to fit weights w_{hk} ($k = 1, \ldots, m-1$; $h = 1, \ldots, H$) and the additive constants c_h to a user-supplied set of binary features (coded as the **P** matrix in (2)) to yield a "constrained solution" [cf. Carroll & Arabie, 1980], without recourse to any of the iterative computation described above, or (b) at the end of each outer iteration (viz., series of inner iterations, all applied to the same subset k), using the iteratively fitted values of p_{ik}.

Arabie and Carroll [1980, p. 223] noted that in the MAPCLUS (two-way) case, this least squares regression problem requires solving

$$\mathbf{s} = \mathbf{Q}\mathbf{w} \tag{12}$$

where $\mathbf{s}$ is the column vector of M elements with $s_{(ij)}$ as the general entry (with the concatenated subscript $(ij) = 1, \ldots, M$), $\mathbf{Q}$ is the $M \times (m+1)$ matrix with $q_{(ij)k}$ as the general entry, where $q_{(ij)k} = p_{ik}p_{jk}$, and $\mathbf{w}$ is the $(m+1)$-dimensional column vector with general entry w_k (where $w_{m+1} = c$).

For INDCLUS, (12) now becomes

$$\mathbf{s}^h = \mathbf{Q}\mathbf{w}^h \tag{13}$$

and it is noteworthy that we only have to obtain the pseudo-inverse of $\mathbf{Q}$ once in order to solve for the H sets of weights, $\mathbf{w}^h$. Thus, although there are H times as many weights to solve for with INDCLUS, compared to the two-way MAPCLUS case, the computational burden does not increase proportionally. On the other hand, the increased number of weights to be estimated allows more opportunities for negative weights to occur, and they have no simple interpretation in this model. In the case of MAPCLUS, negative weights were so infrequently encountered that we found them to be no cause for concern. However, for some three-way data sets, negative weights have been sufficiently annoying that we have included an option, at the user's discretion, of constraining the w_{hk} ($h = 1, \ldots, H$; $k = 1, \ldots, m-1$) to be nonnegative at the end of each outer iteration. The procedure for implementing the optional constraint is taken from Lawson and Hanson [1974, Ch. 23].

Normalizing the Input Proximities Matrices

The INDCLUS program allows the user to declare whether the input matrices are matrix unconditional or matrix-conditional [cf. Takane, Young, & de Leeuw, 1977]. In the former case, all matrices are assumed to be comparable (or, roughly speaking, "on the same scale"). Thus, such data (over all H matrices, simultaneously) are put on the interval [0,1]. (Then, to simplify the mechanics of regression, the mean $\bar{s}^h$ of each matrix is subtracted from all entries within the matrix.) In the case of matrix conditional data, each of the H matrices is separately standardized to have a mean of 0 and a standard deviation of 1.0. When the declaration of matrix-conditional data is made, the weights often tend to be considerably larger than those resulting from an unconditional analysis.

Other Computational Details

Arabie and Carroll [1980, pp. 224–226] described two strategies of combinatorial optimization in MAPCLUS, executed after the completion of the alternating least squares phase, and involving singleton and "doubleton" reversals of the constituent elements of each cluster in turn. (The more general strategy of "l-tuple" reversals was also developed but not implemented in the MAPCLUS program because of computational expense.) These same two strategies are incorporated in INDCLUS. Moreover, the equations used to describe the procedure in MAPCLUS only require adding a superscript to the residuals δ_{ij}^h and all functions of them, plus an extra summation over $h = 1, \ldots, H$ in the numerator of the variances in order to be appropriate for INDCLUS.

Just as with MAPCLUS, initial values of **P** for the alternating least squares phase of INDCLUS may be: (a) user-supplied, (b) the output from a random number generator, or (c) the output of a rational strategy. For INDCLUS, the third case consists of aggregating the H proximities matrices (after normalization) to form a (single) two-way matrix which then becomes input for the rational procedure used in MAPCLUS. An alternative and more expensive approach also available to the user is to execute a complete two-way (MAPCLUS) analysis for the aggregate two-way matrix and, after obtaining the *final* two-way solution, use it as the initial configuration for INDCLUS.

Applications

Examples of fitting the two-way case of the ADCLUS model have appeared in Arabie and Carroll [1980, 1983], Arabie, Carroll, DeSarbo, and Wind [1981], Eckes [1981], Rabin and Frank [Note 4], and Shepard and Arabie [1979]. Several of these papers observed that further information about individual differences in weighting of the clusters would be of interest. We now present two applications of the INDCLUS model, fitted by the algorithm we have just described. (Yet another application is found in K. Miller & Gelman [1983].)

Kinship Data of Rosenberg and Kim [1975]

The fifteen most commonly used kinship terms (given in Table 1) were printed on slips of paper for use in a sorting task [see G. A. Miller, 1969; Rosenberg, 1977, 1982 for details of the paradigm used by Rosenberg & Kim, 1975]. Eighty-five male and eighty-five female subjects were run in the condition where subjects gave (only) a single-sort of the fifteen terms. A different group of subjects (eighty males and eighty females) were told that after making their first sorts of the terms, they should give additional subjective partitioning(s) of these stimuli using "a different basis of meaning each time." Rosenberg and Kim [1975] used only the data from the first and second sortings for this group of subjects. Thus, we have six *conditions*, which will be our "subjects" for an INDCLUS analysis: females' single-sort, males' single-sort, females' first-sort, males' first-sort, females' second-

sort, and males' second-sort. Again note that the subjects in the first two conditions were distinct from the subjects in the last four conditions.

There is a further technical detail required in the present description of the Rosenberg and Kim [1975] data. The subjects' partitions of the stimuli comprise nominal scale data that do not immediately assume the form of a proximities matrix. Thus, some pre-processing is necessary to obtain such a matrix. If we form a stimuli × stimuli co-occurrence matrix for each experimental condition, with the $(i, j)^{th}$ entry defined as the number of subjects who placed stimuli i and j in the same group, and subtract that entry from the total number of subjects contributing to the matrix, then we have a measure of dissimilarity that Drasgow and Jones [1979; also see G. A. Miller, 1969] have called the S-measure. Alternatively, we can take the squared Euclidean distance between all rows/columns of the (symmetric) co-occurrence matrix to obtain an indirect measure of dissimilarity between all pairs of stimuli. Rosenberg and his colleagues [e.g., Rosenberg &

Table 1

INDCLUS Solution for Kinship Data from Rosenberg and Kim [1975]
(Matrix Unconditional Analysis Using S-measure)

Weights for Different Sources of Data							
Females' Single-Sort	Males' Single-Sort	Females' First Sort	Females' Second Sort	Males' First Sort	Males' Second Sort	Elements of Subset	Interpretation
.052	.143	.551	.241	.299	.295	brother, father, grandfather, grandson, nephew son, uncle	Male relatives, excluding cousin
.049	.146	.554	.246	.291	.306	aunt, daughter, granddaughter, grandmother, mother, niece, sister	Female relatives, excluding cousin
.552	.397	.283	.373	.340	.237	aunt, cousin, nephew, niece, uncle	Collateral (Romney & D'Andrade, 1964) relatives
.478	.372	.206	.322	.241	.219	brother, daughter, father, mother, sister, son	Nuclear family
.626	.449	.251	.385	.395	.253	granddaughter, grandfather, grandmother, grandson	Direct ancestors and descendants ±2 generations removed
.055	.075	.132	.158	.158	.207	Additive constants	
78.6%	68.8%	96.3%	78.9%	82.4%	71.7%	Variance accounted within condition	Overall VAF = 81.1%

Sedlak, 1972] refer to the latter measure of dissimilarity as the δ measure (not to be confused with the earlier use of that symbol in this paper). Extensive comparisons of the performance of the two measures when used as input to nonmetric multidimensional scaling have been conducted by L. E. Jones and his students [e.g., Drasgow & Jones, 1979]. In addition to these two different approaches to pre-processing the data, there is also the matrix-conditional/unconditional dichotomy. Thus, in summary, we have four different approaches to analyzing $H = 6$ proximities matrices.

By way of substantive motivation, Rosenberg and Kim [1975] noted the two straightforward bases of partitioning the kinship terms: (a) by family and generational considerations (e.g., nuclear family versus grandchildren) and, (b) by sex, which is inherent in all the terms except cousin. Given these two conflicting bases for organization, it is naturally of interest to look at differences among groups of subjects in the use of these two schemes of classification, and to see which experimental paradigms and methods of portraying structure in the data could faithfully depict such differences. In seeking a discrete representation of structure via hierarchical clustering, Rosenberg and Kim [1975, Fig. 2, p. 496] found it necessary to analyze various of their matrices *separately*, in order to recover evidence of the two bases of judgmental organization.

Table 1 presents a five-cluster solution from an INDCLUS analysis of the six matrices, using the S-measure (matrix-unconditional analysis), with 81.1% of the variance accounted for (over all six matrices). The clusters are easily interpreted. In the order listed, the first two are sex-defined, the third is the collateral relatives [in the terminology of Romney & D'Andrade, 1964; also see Boorman & Olivier, 1973], the fourth is the nuclear family, and the fifth consists of grandparents and grandchildren. The patterns of the weights also yield interesting results. For example, the statement of Rosenberg and Kim [1975, p. 489] that subjects restricted to a single-sort ignore sex as a basis of organization is strongly supported by the relatively low weights for the sex-defined clusters in the first two columns (especially for female subjects) of Table 1. For the multiple-sort conditions, it is interesting to note that female subjects emphasized sex in the first sorting (given that the two relevant clusters have much higher weights), whereas male subjects waited until the second sorting to emphasize the salience of sex as a factor in sorting the kinship terms. Across all conditions, females' data were better fitted to the model than were males' data. Also, data from the first sort were better fitted than for the second sort, for both females and males.

We noted earlier that an INDCLUS analysis using S-measure data assumed to be matrix-unconditional is only one of four possible approaches to the data of Rosenberg and Kim [1975]. For the remaining three, we obtained the same five clusters, with weights having nearly the same patterns as those given in Table 1, and thus corroborating the substantive points given in the preceding discussion. Overall variances accounted for were: 79.5% for S-measure conditional, 89.4% for δ-measure unconditional, and 88.1% for δ-measure conditional. The last two analyses each yielded one slightly negative weight ($-.0036$, and $-.014$, respectively).

Factions in the U.S. Supreme Court, 1975-1979 Terms

Data from U.S. Supreme Court decisions have been analyzed by various scaling and related techniques [Dawes, Brown, & Kaplan, Note 5; also see Stookey & Baer, 1976, and references therein, and Provine, 1980, for substantive background on the related problem of case selection], but not by overlapping clustering techniques. Since 1949, the *Harvard Law Review* has published an annual summary of Court activities for the preceding year. From the November issues of the *Harvard Law Review* [1976, 1977, 1978, 1979, 1980], we took the data on "voting alignments" for the same nine Justices during the 1975-1979 terms of the Court. (The late Justice Douglas was present infrequently during

his last term in 1975 and is not included in the present analysis.) The percent of concurring votes for each pair of Justices forms the entry in a proximities matrix for each of the $H = 5$ terms of the Court from 1975 through 1979.

Over all the five terms of the Court considered in the analysis presented in Table 2, the dyadic cluster of Justices Brennan and Marshall receives the highest weights. During the 1976 and 1977 terms, those liberal Justices concurred 93.6% of the time, the highest such agreement over all the entries in the five matrices. The second cluster (in the order listed in Table 2) comprises the four Nixon appointees to the Court, plus Justices Stewart and White, who were most often regarded as conservatives.

We have labeled the third cluster "Conservative-Centrist." According to the (controversial) popular press account of the Court offered in *The Brethren* [Woodward & Armstrong, 1981], Justice Powell was "the most moderate of the four Nixon appointees" [1981, p. 528], and Justices Powell, Stevens, and Stewart were the centrist group [1981, pp. 518–524] who prevailed during a series of verdicts involving capital punishment. Justice Rehnquist has been more conservative than centrist, but his frequent agreement with Chief Justice Burger presumably links him with the other Justices in that cluster.

The fourth cluster subsumes the two liberal Justices, Brennan and Marshall, from Cluster 1. They were in agreement with Justice Blackmun on abortion issues and civil rights. The latter issue in turn links all three to Justice White, a Kennedy appointee. Justice Stevens' "wild card" [Woodward & Armstrong, 1981, p. 508] voting record and frequent role as a swing voter qualify him as a centrist member of Cluster 4.

To elaborate on these last two clusters, we note that only Justice Stevens is common to both. The remaining eight Justices are evenly partitioned between the two clusters. While some Justices seem firmly entrenched in their membership in one of the two clusters as interpreted (e.g., Rehnquist and Stewart in Cluster 3; Brennan and Marshall in Cluster 4), it seems likely that other Justices (e.g., Powell and Blackmun) would be assigned with more variability if data over a longer period (and covering more issues) had been included. It is also of interest to note the greater variability of the weights of Centrist-Liberal cluster, as compared to the Conservative-Centrist cluster.

These four clusters (plus an additive constant for each matrix) accounted overall for

Table 2

INDCLUS Solution for Voting Alignments of U.S. Supreme Court Justices 1975-1979
(Matrix Unconditional Analysis)

Weights for Different Terms						
1975	1976	1977	1978	1979	Justices in Subset	Interpretation
.548	.640	.627	.427	.424	Brennan, Marshall	Liberals
.342	.314	.105	.278	.258	Blackmun, Burger, Powell, Rehnquist, Stewart, White	Conservatives
.142	.106	.144	.195	.169	Burger, Powell, Rehnquist, Stevens, Stewart	Conservative-Centrist
.089	.120	.096	.262	.163	Blackmun, Brennan, Marshall Stevens, White	Centrist-Liberal
.266	.240	.278	.212	.233	Additive constants	
84.7%	80.5%	62.3%	83.7%	71.1%	Variance accounted for	Overall VAF = 77.7%

77.7% of the variance. A matrix conditional analysis of the same data yielded the same clusters, with the same pattern of weights, and 76.4% VAF.

Discussion and Future Prospects

In the published version of his R. A. Fisher Memorial Lecture, W. H. Kruskal [1981, p. 511] notes the continuing importance of classification in statistics and gives humorous examples [1981, p. 511, footnote 3] of the difficulties that can arise from a classification with overlapping categories. Nonetheless, he points out that one of his examples was apparently of great utility and importance in the history of a subdiscipline of zoology. We have earlier cited examples of successful interpretations of fitting the two-way ADCLUS [Shepard & Arabie, 1979] model, using the two-way algorithm MAPCLUS. The three-way INDCLUS generalization raises several issues.

First, we note that in multidimensional scaling, the generalization from two- to three-way models and algorithms [Carroll & Chang, 1970; Tucker, 1972] seemed to answer statistical needs that had previously not been satisfied, as inferred from the heavy usage these three-way models of scaling have enjoyed [cf. Carroll & Arabie, 1980; Carroll & Wish, 1974]. To date, the only other *discrete* counterpart besides INDCLUS to three-way scaling methods has been the multiple hierarchical clustering approach of Carroll and Pruzansky [Carroll & Pruzansky, Note 6; 1980]. It is our hope that these clustering methods may be as useful for portraying *discrete* structure as three-way scaling methods (e.g., INDSCAL) have been for continuous structure. For example, recent developments in sociometry, emphasizing the different types of social ties that underlie social structure [Boorman & White, 1976], have demonstrated the need for a three-way discrete representation of social structure that depicts not only the actors involved but also how the different types of social ties (corresponding to the "subjects" mode in INDCLUS) are interrelated. [See Knoke, 1983, as well as Arabie & Carroll, 1983, for a substantive discussion of this point.]

Second, just as the advent of INDSCAL [Carroll & Chang, 1970] overcame some of the rotational problems found in two-way multidimensional scaling, INDCLUS offers a possible solution to one inelegant feature of the ADCLUS model. Specifically, Shepard and Arabie [1979, pp. 95–97] noted that three dyadic ADCLUS clusters (a, b), (b, c), (a, c), all with identical weights would have the same goodness-of-fit as the single cluster (a, b, c) having the same weight as as that shared by the three dyadic clusters. The principle of parsimony, however, was invoked to favor the representation having the least number of clusters, as a means of resolving this particular lack of uniqueness. Furnas [Note 7] has pointed out that for the INDCLUS model, unless all three dyadic clusters had tied weights for *each* subject, then the three clusters would clearly be discriminable from the composite cluster. Thus, we see INDCLUS as offering a solution to one potential aspect of nonuniqueness in the ADCLUS model (albeit one for which there is as yet no evidence of occurrence for empirical data).

Third, we note that DeSarbo's [1982] recent approach to fitting the (two-way) ADCLUS model could also be generalized to the three-way case. DeSarbo's GENNCLUS algorithm allows for symmetric *or* nonsymmetric input proximities data, clustering of both modes (if they are distinct) of a two-way two-mode matrix, and the optional imposition of various constraints not available in other approaches to fitting the ADCLUS model. As noted, these features could in principle be generalized to the three-way case.

While it is customary for "future prospects" sections of articles to be expansive and optimistic, we must depart from this tradition in one detail. An obvious next step for both the MAPCLUS and INDCLUS algorithms would be to incorporate monotone regression

[J. B. Kruskal, 1964a, 1964b] so that the input data could be assumed to be just ordinal instead of interval [cf. Shepard & Arabie, 1979, p. 118]. However, we have recently discovered a general theoretical degeneracy inherent in the fitting of discrete models like ADCLUS and INDCLUS, when the fitting entails optimizing a measure of badness-of-fit such as Kruskal's [1964a, 1964b] "stress."

Specifically, for a single two-way one-mode proximities matrix (e.g., input to MAPCLUS), assume that there is a unique largest pairwise similarity (or smallest dissimilarity) value, s_{ij}. Now let those two objects i and j form a dyadic cluster whose weight is simply $s_{ij} - c$, where c is the additive constant that corresponds to the weight for the complete set (implicitly present) of the n objects. Thus we have two clusters—(i, j) and the complete set—with weights of $(s_{ij} - c)$ and c, respectively. Now the predicted similarity between i and j will be $\hat{s}_{ij} = s_{ij}$, and for all other y,z (distinct from each other and from i and j), $\hat{s}_{yz} = c$. This "solution" yields a zero value of stress, for both stress formulae one and two [Kruskal & Carroll, 1969], and qualifies as "degenerate" since there are only two distinct fitted values $\hat{s}$, in contrast to an input matrix **S** typically having far more than two distinct values. If there is not a unique largest similarity value in the matrix, then a dyadic cluster may be required for each of the pairs of objects corresponding to the tied largest value. However, if the ties occur so as to render the set of corresponding objects a maximal complete subgraph, then one cluster will suffice.

For a three-way two-mode proximities matrix (suitable as input for INDCLUS), if the largest similarity value corresponds to different pairs of objects across subjects in a matrix conditional analysis, then a nonmetric version of INDCLUS might be feasible, if there are enough subjects. Thus, the generalization of ADCLUS/MAPCLUS to the three-way INDCLUS offers a possible answer to the problem of inherent degeneracies in fitting this nonmetric version of a discrete model.

Thus, just as we earlier [Arabie & Carroll, 1980, pp. 226-227] concluded that there is an inherent nonuniqueness in various discrete models such as ADCLUS and INDCLUS, we now note an inherent degeneracy lurking when traditional approaches to monotone regression are considered for fitting data to such models. A similar degeneracy can also be demonstrated for nonmetric approaches to fitting tree structures and some other discrete models, at least in the two-way cases.

REFERENCE NOTES

1. Carroll, J. D. *Handout for models for individual differences in similarities judgments.* Paper presented at Mathematical Psychology Meetings, Purdue University, West Lafayette, IN, August 25–27, 1975.
2. Carroll, J. D., & Arabie, P. *INDCLUS: A three-way approach to clustering.* Paper presented at meeting of Psychometric Society, Monterey, CA, 1979.
3. Kruskal, J. B., Young, F. W., & Seery, J. B. *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Murray Hill, N. J.: Bell Telephone Laboratories, 1973.
4. Rabin, M. D., & Frank, M. Unpublished work, University of Connecticut Health Center, August, 1982.
5. Dawes, R. M., Brown, M. E., & Kaplan, N. *The skewed hourglass: A configurational approach to constructing a Guttman scale when domination is unspecified.* Paper presented at annual convention of the Midwestern Psychological Association, Chicago, April 1965.
6. Carroll, J. D., & Pruzansky, S. *Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares.* Paper presented at the U.S.-Japan Seminar on Multidimensional Scaling, University of California at San Diego, La Jolla, California, Auguest 20–24, 1975.
7. Furnas, G. Personal communication, June, 1982.

REFERENCES

Arabie, P., & Carroll, J. D. MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika*, 1980, *45*, 211–235.

Arabie, P., & Carroll, J. D. Conceptions of overlap in social structure. In L. Freeman, A. K. Romney, & D. R. White (Eds.), *Methods of social network analysis.* Berkeley and Los Angeles: University of California Press, 1983.

Arabie, P., Carroll, J. D., DeSarbo, W., & Wind, J. Overlapping clustering: A new method for product positioning. *Journal of Marketing Research*, 1981, *18*, 310–317.

Boorman, S. A., & Olivier, D. C. Metrics on spaces of finite trees. *Journal of Mathematical Psychology*, 1973, *10*, 26–59.

Boorman, S. A., & White, H. C. Social structure from multiple networks. II. Role structures. *American Journal of Sociology*, 1976, *81*, 1384–1446.

Carroll, J. D., & Arabie, P. Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology.* Palo Alto, CA: Annual Reviews, 1980.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, 1970, *35*, 283–319.

Carroll, J. D., & Pruzansky, S. Discrete and hybrid scaling models. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice.* Bern: Hans Huber, 1980.

Carroll, J. D., & Wish, M. Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. II). San Francisco: Freeman, 1974.

DeSarbo, W. S. GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, 1982, *47*, 449–475.

Drasgow, F., & Jones, L. E. Multidimensional scaling of derived dissimilarities. *Multivariate Behavioral Research*, 1979, *14*, 227–244.

Eckes, T. *Formale Modelle zu Ähnlichkeitsstrukturen.* Doctoral Dissertation, University of the Saar. Saarbrücken, West Germany, 1981.

Harvard Law Review, 1976, *90*, 277.

Harvard Law Review, 1977, *91*, *296*.

Harvard Law Review, 1978, *92*, *328*.

Harvard Law Review, 1979, *93*, 276.

Harvard Law Review, 1980, *94*, 288.

Knoke, D. Organization sponsorship and influence: Representation of social influence associations. *Social Forces*, 1983, in press.

Kruskal, J. B. Multidimensional scaling by optimizating goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, *29*, 1-27. (a)

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, *29*, 115–129. (b)

Kruskal, J. B., & Carroll, J. D. Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis II:* New York: Academic Press, 1969.

Kruskal, W. Statistics in society: Problems unsolved and unformulated. *Journal of the American Statistical Association*, 1981, *76*, 505–515.

Lawson, C. L., & Hanson, R. J. *Solving least squares problems.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Miller, G. A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 1969, *6*, 169–191.

Miller, K., & Gelman, R. The child's representation of number: A multidimensional scaling analysis. *Child development*, 1983, in press.

Provine, D. M. *Case selection in the United States Supreme Court.* Chicago: University of Chicago Press, 1980.

Pruzansky, S., Tversky, A., & Carroll, J. D. Spatial versus tree representations of proximity data. *Psychometrika*, 1982, *47*, 3–24.

Romney, A. K., & D'Andrade, R. G. Cognitive aspects of English kin terms. *Transcultural studies in cognition. American Anthropologist Special Issue*, 1964, 66, (3, Pt. 2), 146–170.

Rosenberg, S. New approaches to the analysis of personal constructs in person perception. In J. K. Cole (Ed.), *Nebraska Symposium on Motivation* (Vol. 24), Lincoln: University of Nebraska Press, 1977.

Rosenberg, S. The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate methods in the social sciences: Applications.* Hillsdale, NJ: Erlbaum, 1982.

Rosenberg, S., & Kim, M. P. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 1975, *10*, 489–502.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Shepard, R. N., & Arabie, P. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 1979, *86*, 87–123.

Stookey, J. A., & Baer, M. A. A critique of Guttman scaling: With special attention to its application to the study of collegial bodies. *Quality and Quantity*, 1976, *10*, 251–260.

Takane, Y., Young, F. W., & de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 1977, *42*, 7–67.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Tucker, L. R The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston, 1964.

Tucker, L. R Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 1972, *32*, 3–27.

Wold, H. Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate analysis*. New York: Academic Press, 1966.

Woodward, B., & Armstrong, S. *The brethern*. New York: Avon, 1981.

Manuscript received 8/9/82
Final version received 11/9/82