L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

An L_1-norm procedure for fitting overlapping clustering models to proximity data

Anil Chaturvedi

AT&T Laboratories, USA

J. Douglas Carroll

Rutgers University, New Jersey, USA

Abstract: We describe a general approach to finding Least Absolute Deviation estimates of two-way and three-way overlapping clustering models called ADCLUS and INDCLUS. The suggested approach utilizes a combinatorial optimization approach that takes advantage of a separability property of this loss function for fitting these models. Our approach helps in robustifying the solutions in the presence of extreme outliers in the data.

Key words: Cluster analysis, overlapping clusters, combinatorics, trilinear models, least absolute deviation.

AMS subject classification: Primary 62H30; secondary 92J45.

1 Introduction

Shepard and Arabie (1979) first introduced an overlapping clustering model for classification based on similarity data called ADCLUS (for additive clustering). Subsequently, Arabie and Carroll (1980), provided a mathematical programming approach for fitting this model. Carroll and Arabie (1983), proposed an individual differences generalization of this model, which they called INDCLUS (for individual differences clustering), and also devised a procedure for fitting this three-way generalization, thus providing a methodology for three-way overlapping clustering. These algorithms optimize a least squares, or L_2-norm based, loss function. The theoretical significance and conceptual elegance of overlapping clustering can be seen in applications in many different substantive domains, as illustrated, for example, in Arabie, Carroll, DeSarbo, and Wind (1981) and Srivastava, Alpert, and Shocker (1984). In this paper, we present a procedure for fitting these models via an L_1-norm, which we call LADCLUS (for Least Absolute Deviation clustering). This procedure is a special case of a general approach (Carroll and Chaturvedi, 1995) for fitting a general multilinear model including both discrete and continuous parameters via L_1- or L_2-norms, or other L_p-norms. Our procedure is computationally significantly faster and can handle much larger data sets than Lakshmi-Ratan's (1985) L_1-norm based approach for fitting the two-way overlapping clustering problem. It is computationally as simple and tractable as a procedure proposed by Chaturvedi and Carroll (1994), which provides a more efficient algorithm than the earlier algorithms for fitting these models via an L_2-norm. The principal benefit of fitting these overlapping clustering models via an L_1-norm would be to robustify the estimation of model parameters vis-a-vis extreme outliers in the data. Fitting models via an L_1-norm tends to reduce the effects of extreme outliers, as compared to fitting via an L_2 (OLS)-norm (Hampel, Ronchetti, Rousseeuw and Stahel, 1986; Kaufman and Rousseeuw, 1990). The increased robustness of the L_1-norm procedure to the L_2-norm procedure is analogous to the illustration in Rousseeuw and Leroy (1987, pp. 10-11) in the context of linear regression, wherein the L_1-norm estimate is shown to be more robust to extreme values of the dependent variable. The derived solutions would be more robust to extreme values in the data (analogous to the dependent variable in the regression case).

2 The INDCLUS model

Assume that N objects are being clustered into R overlapping clusters. Then, the INDCLUS model (Carroll, 1975; Carroll and Arabie, 1983) is written as:

$$\mathbf{S}_k = \mathbf{P}\mathbf{W}_k\mathbf{P}' + \mathbf{C}_k + \text{error}, \tag{1}$$

where:

$\mathbf{S}_k$ is an $(N \times N)$ similarity matrix for the kth subject (or other source of data); $k = 1, ..., K$,

$\mathbf{W}_k$ is an $(R \times R)$ diagonal matrix of weights for the kth subject (or other source of data); $k = 1, ..., K$,

$\mathbf{P}$ is an $(N \times R)$ *binary* indicator matrix defining the possibly overlapping clusters, and

$\mathbf{C}_k$ is an $(N \times N)$ matrix, all of whose entries are c_k, which can be thought of as the weight for a universal cluster denoted by an $N \times 1$ unit vector $\mathbf{1}$, all of whose components are 1.

The diagonal entries in the $(N \times N)$ matrix $\mathbf{S}_k$ are usually not defined. The estimation problem is to determine the ordinary least squares (OLS) estimates of parameters $\mathbf{P}$, $\mathbf{W}_k$ and $\mathbf{C}_k$. The diagonal elements of the weight matrices $\mathbf{W}_k$ must be non-negative and elements of $\mathbf{P}$ must be constrained to be either 0 or 1. The ADCLUS model is the special case of the INDCLUS model in which $K = 1$. The ADCLUS procedure proposed by Shepard and Arabie (1979) combines a combinatorial algorithm with iterative estimation of the weights. The MAPCLUS procedure proposed by Arabie and Carroll (1980) fits the ADCLUS model via a penalty function based mathematical programming technique embedded in an overall alternating least squares procedure. The INDCLUS method proposed by Carroll and Arabie (1983) generalizes the MAPCLUS approach to the three-way INDCLUS model. These are all OLS procedures (using an L_2-norm based fit measure). Various other techniques have been developed for fitting these models, such as the Maximum Likelihood approach of Hiroshi Hojo (1983), the Qualitative Factor Analysis (QFA) procedure of Mirkin (1987), and the OLS procedure called SINDCLUS, of Chaturvedi and Carroll (1992, 1994). We now describe the LADCLUS algorithm for fitting these models via an L_1-norm.

3 The LADCLUS algorithm

The LADCLUS procedure uses a separability property of the LAD loss function for fitting the INDCLUS model defined in (1) via an alternating estimation procedure discussed below. This separability property has also been used in the SINDCLUS procedure (Chaturvedi and Carroll, 1994). While Mirkin (1990) also uses a one-cluster-at-a-time approach to estimating a bilinear model, his approach utilizes a different algorithm for estimating the parameters for a cluster. The two main differences of the approach presented in this paper with Mirkin's (1990) approach are: (a) Mirkin does not use the separability property, but a different approach to estimating the parameters for any given cluster, and (b) his approach does not yield overall L_1-norm based estimates for the model parameters, since he does not iterate over clusters (as we do) in order to obtain a globally optimal solution. This separability property can best be stated in the form of two procedures - the elementary discrete and elementary continuous LAD procedures. These are illustrated below.

A. The elementary discrete LAD procedure

Consider the following illustrative problem. Let

$$\mathbf{L} = \begin{bmatrix} 7 & 5 & 3 & 9 \\ 8 & 6 & 5 & 1 \\ 9 & 4 & 2 & 7 \\ 5 & 3 & 4 & 6 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \qquad \mathbf{r}' = \begin{bmatrix} 1 & 4 & 9 & 3 \end{bmatrix}$$

The estimation problem is to find the least absolute deviation (LAD) estimate of $\mathbf{x}$, where $\mathbf{L} = \mathbf{x}\mathbf{r}' +$ error and $\mathbf{x}$ is constrained to be binary (0 or 1). That is

$$\begin{bmatrix} 7 & 5 & 3 & 9 \\ 8 & 6 & 5 & 1 \\ 9 & 4 & 2 & 7 \\ 5 & 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1x_1 & 4x_1 & 9x_1 & 3x_1 \\ 1x_2 & 4x_2 & 9x_2 & 3x_2 \\ 1x_3 & 4x_3 & 9x_3 & 3x_3 \\ 1x_4 & 4x_4 & 9x_4 & 3x_4 \end{bmatrix} + error$$

If we let

$$\begin{aligned} f_1 &= \mid 7 - 1x_1 \mid + \mid 5 - 4x_1 \mid\mid 3 - 9x_1 \mid + \mid 9 - 3x_1 \mid, \\ f_2 &= \mid 8 - 1x_2 \mid + \mid 6 - 4x_2 \mid + \mid 5 - 9x_2 \mid + \mid 1 - 3x_2 \mid, \\ f_3 &= \mid 9 - 1x_3 \mid + \mid 4 - 4x_3 \mid + \mid 2 - 9x_3 \mid + \mid 7 - 3x_3 \mid, \text{ and} \\ f_4 &= \mid 5 - 1x_4 \mid + \mid 3 - 4x_4 \mid + \mid 4 - 9x_4 \mid + \mid 6 - 3x_4 \mid, \end{aligned}$$

then the sum of absolute errors is given by

$$F = f_1 + f_2 + f_3 + f_4.$$

Note that f_1 is a function only of x_1; f_2 is a function only of x_2; f_3 is a function only of x_3; and f_4 is a function only of x_4. Thus, F is *separable* in x_1, x_2, x_3, and x_4. To minimize F, one can separately minimize f_1 w.r.t x_1, f_2 w.r.t x_2, f_3 w.r.t x_3, and, f_4 w.r.t. x_4. To minimize, say, f_1 w.r.t. x_1, one can easily evaluate f_1 at $x_1 = 1$ and $x_1 = 0$. The x_1 yielding a minimum of these two possible values is then chosen. Thus, for I (0-1) variables, only $2I$ function evaluations and comparisons are needed, as compared to 2^I evaluations and comparisons for explicit enumeration.

B. The elementary continuous LAD procedure

Again, consider the illustrative problem given in the Elementary discrete LAD procedure of determining $\mathbf{x}$, where $\mathbf{x}$ is real. As in the case of the Elementary discrete LAD procedure, it can be shown that F is *separable* in x_1, x_2, x_3, and x_4. Thus, to minimize F, one can separately minimize f_1 w.r.t x_1, f_2 w.r.t x_2, f_3 w.r.t x_3, and, f_4 w.r.t. x_4. The minimization of, say, f_1 w.r.t. x_1, is equivalent to performing simple L_1 regression with a single independent variable. While the general multivariate L_1-regression problem can be formulated and solved as a constrained linear programming problem (Rousseeuw and Leroy 1987 pp 146), in our case since there is only one independent variable, the simplex solutions can easily be determined by evaluating the functions $f_1 - f_4$ at the respective corner points. To illustrate, in order to minimize, say, f_1 w.r.t. x_1, simply evaluate f_1 at the four corner points given by $x_1 = 7/1$, $5/4$, $3/9$ and $9/3$. The value of x_1 yielding the minimum is chosen as the optimum LAD estimate of x_1. In this specific case, f_1 is a minimum at $x_1 = 3/9$. More generally, to minimize the LAD criterion as a function of a single variable x_i, for the component f_i of the overall loss function

$$F = \sum_{i=1}^{I} f_i$$

where $f_i = \sum_{j=1}^{J} \mid l_{ij} - x_i r_j \mid$ is a function only of x_i, we simply evaluate f_i at the J values

$$x_i^{(j')} = \frac{l_{ij'}}{r_{j'}}$$

and then choose the $\widehat{x}_i$ to be the $x_i^{(j')}$ minimizing

$$fi^{(j')} = \sum_{j=1}^{J} \mid l_{ij} - x_i^{(j')} r_j \mid .$$

This can easily be shown to provide the x_i minimizing f_i, thus completing the Elementary Continuous LAD procedure.

The LADCLUS procedure utilizes an alternating least absolute deviation (L_1-norm) procedure in fitting the ADCLUS/INDCLUS models. The algorithm converges to at least a local optimum, as the objective function value decreases (or does not increase) at each stage of the algorithm, and there is a lower bound to the objective function.

The estimation problem in LADCLUS is to find LAD estimates of $\mathbf{P}$, $\mathbf{W}_k$, and $\mathbf{C}_k$, $(k = 1, ..., K)$ in the equation:

$$\mathbf{S}_k = \mathbf{P}\mathbf{W}_k\mathbf{P}' + \mathbf{C}_k + error, \tag{2}$$

where $\mathbf{P}$, $\mathbf{W}_k$, and $\mathbf{C}_k$, are as defined in equation (1). If $\mathbf{S}_k$ were considered to be a nonsymmetric $(N \times N)$ matrix, then the above equation can be generalized to

$$\mathbf{S}_k = \mathbf{P}\mathbf{W}_k\mathbf{Q}' + \mathbf{C}_k + error, \tag{3}$$

where $\mathbf{Q}$ is an $(N \times R)$ matrix, not necessarily the same as $\mathbf{P}$. In the *symmetric* case of INDCLUS, $\mathbf{P} = \mathbf{Q}$.

We use (3) to estimate the parameters for the symmetric case *without* imposing the constraint $\mathbf{P} = \mathbf{Q}$. Thus, we will first define the estimation problem for the nonsymmetric case, and then specialize it to the symmetric case. This is the same overall strategy as followed by Chaturvedi and Carroll (1994) in their SINDCLUS approach to OLS estimation of ADCLUS/INDCLUS. Defining the following symbols :

$\mathbf{p}_r = (N \times 1)$ binary vector for the rth cluster,

$\mathbf{w}_r = (K \times 1)$ vector of the weights for the rth cluster,

$\mathbf{q}_r = (N \times 1)$ binary vector for the rth cluster (not necessarily $= \mathbf{p}_r$),

$\mathbf{P}_{(-r)} = (N \times R)$ binary matrix including the universal cluster $\mathbf{1}$ but excluding the rth cluster,

$\mathbf{W}_{(-r)} = (R \times R)$ weight matrix for the kth subject or other source of data, including the weight for the universal cluster but excluding the weight for the rth cluster, and

$\mathbf{Q}_{(-r)} = (N \times R)$ matrix including the universal cluster $\mathbf{1}$ but excluding the rth cluster,

we can rewrite (3) as

$$\mathbf{S}_k = \mathbf{p}_r w_{kr} \mathbf{q}' + \mathbf{P}_{(-r)}\mathbf{W}_{k(-r)}\mathbf{Q}'_{(-r)} + error,$$

If we have estimates of all but the rth cluster, then we can define $\overline{\mathbf{S}}_k$ as

$$\overline{\mathbf{S}}_k = \mathbf{S}_k - \mathbf{P}_{(-r)}\mathbf{W}_{k(-r)}\mathbf{Q}'_{(-r)} \tag{4}$$

to get

$$\overline{\mathbf{S}}_k = \mathbf{p}_r w_{kr} \mathbf{q}'_r + error, \tag{5}$$

that is,

$$\overline{\mathbf{S}}_k = f\ (parameters\ for\ cluster\ r) + error \tag{6}$$

If we have K matrices $\mathbf{S}_k$ of order $(N \times N)$, we can use a procedure similar to the CANDECOMP based algorithm called the INDSCAL method for fitting the INDSCAL model formulated by Carroll and Chang (1970). We call this general procedure CANDCLUS (Carroll and Chaturvedi, 1995), which stands for CANonical Decomposition CLUStering. Let us assume that we have the parameter estimates for all clusters, except the ith cluster. Let $\mathbf{T}_1$ be a $(K \times N^2)$ matrix where the kth row has all the N^2 terms of the $N \times N$ matrix $\overline{\mathbf{S}}_k$, and $\mathbf{T}_2$ and $\mathbf{T}_3$ be ($N \times KN$ matrices). $\mathbf{T}_1$ has all N^2 elements of $\overline{\mathbf{S}}_k$ in its kth row. $\mathbf{T}_2$ is the supermatrix that has the jth row of matrices $\overline{\mathbf{S}}_1, ..., \overline{\mathbf{S}}_K$ in the jth row. Thus:

$$\mathbf{T}_2 = \left[\overline{\mathbf{S}}_1 \mid \overline{\mathbf{S}}_2 \mid ... \overline{\mathbf{S}}_k \mid ... \mid \overline{\mathbf{S}}_K\right]$$

Similarly,

$$\mathbf{T}_3 = \left[\overline{\mathbf{S}}'_1 \mid \overline{\mathbf{S}}'_2 \mid ... \ \overline{\mathbf{S}}'_k \mid ... \mid \overline{\mathbf{S}}'_K\right].$$

Assuming that estimates of $\mathbf{p}_r$ and $\mathbf{q}_r$ are known, the parameters for the rth cluster are estimated by iterating the following 3 Steps until *at least a local optimum* is reached.

- **Step 1. Estimating $\mathbf{w}_r$ conditionally**

Given current estimates, $\widehat{\mathbf{p}}_r$ and $\widehat{\mathbf{q}}_r$, of $\mathbf{p}_r$ and $\mathbf{q}_r$, let $\mathbf{g}_r$ be a vector of N^2 elements such that

$$\mathbf{g}_r = \widehat{\mathbf{p}}_r \otimes \widehat{\mathbf{q}}_r,$$

where $\otimes$ is the Kronecker product. Then, using the elementary continuous LAD procedure outlined earlier, one can find the LAD estimate $\widehat{\mathbf{w}}_r$ in the Equation

$$\mathbf{T}_1 = \mathbf{w}_r \mathbf{g}'_r + error.$$

Non-negativity constraints are imposed easily by simply setting all negative weights to zero as in Carroll, DeSoete, and Pruzansky (1989).

- **Step 2. Estimating $\mathbf{p}_r$ conditionally**

Given current estimates, $\widehat{\mathbf{w}}_r$ and $\widehat{\mathbf{q}}_r$, of $\mathbf{w}_r$ and $\mathbf{q}_r$, let $\mathbf{h}_r$ be a vector of KN elements such that

$$\mathbf{h}_r = \widehat{\mathbf{w}}_r \otimes \widehat{\mathbf{q}}_r.$$

Then, by using the elementary discrete LAD procedure in the equation

$$\mathbf{T}_2 = \mathbf{p}_r \mathbf{h}'_r + error,$$

one can find $\widehat{\mathbf{p}}_r$, the LAD estimates of $\mathbf{p}_r$.

• Step 3. Estimating $\mathbf{q}_r$ conditionally

Given current estimates, $\widehat{\mathbf{w}}_r$ and $\widehat{\mathbf{p}}_r$, of $\mathbf{w}_r$ and $\mathbf{p}_r$, let $\mathbf{j}_r$ be a vector of KN elements such that

$$\mathbf{j}_r = \widehat{\mathbf{w}}_r \otimes \widehat{\mathbf{p}}_r.$$

Then, by using the elementary discrete LAD procedure in the equation

$$\mathbf{T}_3 = \mathbf{q}_r \mathbf{j}_r' + error,$$

one can find $\widehat{\mathbf{q}}_r$, the LAD estimates of $\mathbf{q}_r$.

The LADCLUS algorithm starts off with random binary starting values for the matrices $\mathbf{P}$ and $\mathbf{Q}$, and random continuous values in the diagonal matrices $\mathbf{W}_k$. The LADCLUS procedure iterates up to a prespecified maximum number of major iterations or until the percentage reduction in the fit value is less than a prespecified criterion. The current default is 0.0001. The current fit value (M2) and the previous fit value (M1) are updated after each major iteration. Each major iteration of the LADCLUS algorithm involves determining at least locally optimal conditional L_1-norm estimates of the parameters for the $r+1$ clusters, using the one-cluster-at-a-time strategy. Thus, each major iteration comprises $r+1$ minor iterations, corresponding to the $r+1$ clusters.

The minor iteration of the LADCLUS procedure corresponding to the r^{th} cluster involves sequentially finding conditionally optimal parameter estimates $\mathbf{w}_r$, $\mathbf{p}_r$, and $\mathbf{q}_r$. This is achieved by first forming supermatrices $\mathbf{T}_1, \mathbf{T}_2$ and $\mathbf{T}_3$. The matrix $\mathbf{g}_r$ is then formed and $\widehat{\mathbf{w}}_r$ is estimated as described in Step 1 using the elementary continuous LAD procedure. This is followed by Steps 2 and 3 of estimating $\widehat{\mathbf{p}}_r$ and $\widehat{\mathbf{q}}_r$ respectively, using the elementary discrete procedure. The fit value is then computed (F2), and compared to the previous fit value (F1). This process is repeated until there is no improvement in fit (i.e. F2 = F1). At this point an (at least locally) optimal set of parameter estimates have been obtained for the r^{th} cluster, conditional on the fixed values of the $R-1$ other clusters (and the universal cluster).

It should be noted that in Steps 2 and 3 above, the $\widehat{\mathbf{p}}$ and $\widehat{\mathbf{q}}$ vectors cannot be all zero. Thus, each cluster must have at least one object in it. Since the matrices $\mathbf{S}_k$ usually do not have diagonals in the case of the INDCLUS model, Steps 1, 2 and 3 above need to be modified. In the case of undefined diagonals, we simply drop the corresponding columns from the $\mathbf{T}_1$ matrix and $\mathbf{g}$ vector in Step 1. Similarly, we don't consider the diagonal elements in Steps 2 and 3 for fitting the INDCLUS model. For estimating the weights for the universal cluster, where $\mathbf{p}$ and $\mathbf{q}$ are fixed, we just use Step 1, with $\mathbf{p} = \mathbf{q} = \mathbf{1}$.

While we do not impose the constraint that $\mathbf{P} = \mathbf{Q}$ in the estimation procedure, we find empirically that in the symmetric case of INDCLUS, the matrices $\mathbf{P}$ and $\mathbf{Q}$ are equal upon convergence at the global optimum. We conjecture that this will always occur under conditions of unique global optimality. Also, while the default option in LADCLUS is to estimate the INDCLUS model by treating the diagonals as missing data, LADCLUS also allows fitting of the INDCLUS model when diagonals are treated as non-missing. One major advantage of LADCLUS in addition to its greater computational efficiency, is its ability to handle arbitrary patterns of missing data satisfactorily. (In fact, the default option for diagonals simply is a special case of handling missing data). Since, at each stage of the algorithm, we are conditionally estimating one new "dimension" (e.g., the $\mathbf{p}_r$ or $\mathbf{q}_r$ vector in LADCLUS), through the use of the elementary discrete LAD procedure mentioned earlier, omission of data is accomplished by omitting the corresponding terms from the corresponding LAD loss function. (The treatment of missing data as described is a special case of weighted LAD fitting, with weights of zero for missing observations and one for those that are present.) The generalization of LADCLUS to *weighted* LAD, is also straightforward, since each of the three conditional LAD estimation stages can simply be replaced with an appropriately weighted LAD estimation procedure.

One final comment vis-a-vis LADCLUS is that no special case needs to be described for the fitting of the ADCLUS model, in which $K = 1$. ADCLUS is simply fit as a special case of the INDCLUS model, in which the third way, for subjects or other sources of data, has only one level.

4 Applications of LADCLUS to some real data

The LADCLUS procedure was applied to the Kinship data of Rosenberg and Kim published in Arabie, Carroll, and DeSarbo (1987). The application of SINDCLUS to this data set is described in detail in Chaturvedi and Carroll (1994). We present this application of LADCLUS to compare the solutions derived via LADCLUS and SINDCLUS.

The fifteen most commonly used kinship terms - Aunt, Brother, Cousin, Daughter, Father, Granddaughter, Grandfather, Grandmother, Grandson, Mother, Nephew, Niece, Sister, Son, and Uncle, were printed on slips of paper for use in a sorting task by Rosenberg and Kim (1975). Eighty-five male and eighty-five female subjects were run in a condition where subjects gave (only) a single-sort of the fifteen terms. A different group of subjects (eighty-five males and eighty-five 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 correspond to our subjects: females' single-sort, males' single-sort, females' first-sort, males' first-sort, females' second-sort, and males' second-sort. Again note that the "subjects" (or other sources of data) in the first two conditions were distinct from those in the last four conditions.

Since the subjects' partitions of the stimuli comprise nominal scale data that do not immediately assume the form of a proximity matrix, some pre-processing is necessary to obtain such a matrix. If we form a stimuli $\times$ stimuli co-occurrence matrix for each experimental condition, with the (i, j)th entry derived 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 what is called the S-measure (Arabie, Carroll, and DeSarbo, 1987). As in Arabie, Carroll, and DeSarbo (1987), the six matrices constructed using the S-measure were analyzed using LADCLUS via a matrix unconditional approach. A five cluster solution explaining 38.75 percent of absolute deviation in the data (around the grand median) was extracted. The optimal clusters derived using the LADCLUS procedure are identical to the clusters derived by Arabie, Carroll, and DeSarbo (1987). The five cluster solution is presented in Table 1, while the importance weights derived via LADCLUS and INDCLUS are presented in Tables 2 and 3.

The clusters are easily interpreted. In the order listed, the first two are sex-defined, the third is the collateral relatives, the fourth is the nuclear family, while the fifth consists of grandparents and grandchildren.

5 Conclusions

The LADCLUS procedure introduced least absolute deviations as an objective function for the ADCLUS/INDCLUS models. The enhanced robustness of LADCLUS to the presence of extreme outliers over existing least squares procedures was demonstrated in an extensive Monte-Carlo simulation. The LADCLUS procedure can be extended to fit other hybrid multivariate models that entail both continuous and discrete parameters, where the discrete parameters can take on any discrete values. We hope that the LADCLUS procedure further enhances the applicability of the ADCLUS and INDCLUS models. While in the current version LADCLUS can yield locally optimal solutions, we are investigating approaches that ameliorate these problems and yield globally optimal solutions, even for large data sets.

Table 1: LADCLUS & INDCLUS solutions for Rosenberg and Kim data.

Cluster	Items in Cluster	Interpretation
a	Brother, father, grandfather grandson, nephew, son, uncle	Male relatives excluding cousins
b	Aunt, daughter, granddaughter, grandmother, mother, niece, sister	Female relatives excluding cousins
c	Aunt, cousin, nephew, niece, uncle	Collateral relatives
d	Brother, daughter, father, mother, sister, son	Nuclear family
e	Granddaughter, grandfather, grandmother, grandson	Grandparents and Grandchildren
f	All objects	Universal cluster

Table 2: LADCLUS weights for the Rosenberg and Kim data.

Subject	a	b	c	d	e	Universal
F' single	0.11	0.11	0.44	0.29	0.48	0.04
M' single	0.2	0.19	0.29	0.28	0.28	0.05
F' first	0.57	0.57	0.27	0.17	0.21	0.05
F' second	0.25	0.27	0.33	0.25	0.32	0.09
M' first	0.33	0.33	0.28	0.17	0.29	0.08
M' second	0.31	0.32	0.11	0.15	0.13	0.16

Table 3: INDCLUS weights for the Rosenberg and Kim data.

Subject	a	b	c	d	e	Universal
F' single	.052	.049	.552	.478	.626	.055
M' single	.143	.146	.397	.372	.449	.075
F' first	.551	.554	.283	.206	.251	.132
F' second	.241	.246	.373	.322	.385	.158
M' first	.299	.291	.340	.241	.395	.158
M' second	.295	.306	.237	.219	.253	.207

References

[1] Arabie, P., and Carroll, J. D. (1980). MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika* **45,** 211-235.

[2] Arabie, P., Carroll, J. D., and DeSarbo, W. S. (1987). Three-way Scaling and Clustering. Sage, Newbury Park, California.

[3] Arabie, P., Carroll, J. D., DeSarbo, W. S., and Wind, J. (1981). Overlapping clustering: a new method for product positioning. *Journal of Marketing Research* **18,** 310-17.

[4] Carroll, J. D., (1975). Models for individual differences in similarities. Paper presented at the Eighth Annual Mathematical Psychology Meeting, Purdue University, West Lafayette, IN.

[5] Carroll, J. D. and Arabie, P. (1983). An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika* **48,** 157-169.

[6] Carroll, J. D., and Chang J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 283-319.

[7] Carroll, J. D., and Chaturvedi, A. D. (1995). A general approach to clustering and multidimensional scaling of two-way, three-way, or higher-way data. In *Geometric Representations of Perceptual Phenomena: Articles in Honor of Tarow Indow's 70th Birthday*, Eds. R. D. Luce, M. D'Zmura, D. D. Hoffman, G. Iverson, and A. K. Romney. In press.

[8] Carroll, J. D., DeSoete, G., and Pruzansky, S. (1989). Fitting of the latent class model via iteratively reweighted least squares CANDECOMP with nonnegativity constraints. In *Multiway Data Analysis*, Eds. R. Coppi and S. Bolasco, pp. 463-472. Amsterdam: North Holland.

[9] Carroll, J. D. and Pruzansky, S. (1975). Fitting of hierarchical tree

structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares. *Proc. of the U.S.-Japan Seminar on Multidimensional Scaling*, 9-19.

[10] Carroll, J. D. and Pruzansky, S. (1980). Discrete and hybrid scaling models. In *Similarity and Choice*, Eds. E. D. Lantermann and H. Feger, pp. 108-139. Bern: Hans Huber.

[11] Chaturvedi, A. D. (1993). Perceived product uniqueness in product differentiation and product choice. *9316296*, University Microfilms International, Ann Arbor, Michigan.

[12] Chaturvedi, A., D., and Carroll, J. D. (1992). ALCLUS: alternating least squares clustering. Presented at the Annual Meeting of The Classification Society of North America, East Lansing, Michigan, June 11-13.

[13] Chaturvedi, A. D., and Carroll, J. D. (1994). An alternating combinatorial optimization approach to fitting the INDCLUS and generalized INDCLUS models. *Journal of Classification* **11,** 155-170.

[14] Chaturvedi, A. D., Carroll, J. D., and Lakshmi-Ratan, R. A. (1993). MADCLUS: minimum absolute deviation clustering. Presented at the Annual Meeting of The Classification Society of North America, Pittsburgh, Pennsylvania, June 24-26.

[15] DeSarbo, W. S. (1982). GENNCLUS: new models for general nonhierarchical clustering analysis. *Psychometrika* **47,** 449-475.

[16] Hartigan, J. A. (1975). *Clustering Algorithms.* New York: Wiley.

[17] Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior* **8,** 176-184.

[18] Hiroshi, Hojo (1983). A maximum likelihood method for additive clustering and its applications. *Japanese Psychological Research* **25,** 191-201.

[19] Johnson, S. C. (1967). Hierarchical Clustering Schemes. *Psychometrika* **32,** 241-254.

[20] Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data.* New York: Wiley Inter-Science.

[21] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

[22] Lakshmi-Ratan, R. A. (1985). A linear mixed integer programming approach to overlapping cluster analysis. Working paper, University of Wisconsin, Madison.

[23] Mirkin, B. G. (1987). Additive clustering and qualitative factor analysis methods for similarity matrices. *Journal of Classification* **4,** 7-31.

[24] Mirkin, B. G. (1989). Erratum. *Journal of Classification* **6,** 271-272.

[25] Mirkin, B. G. (1990). A sequential fitting procedure for linear data analysis models. *Journal of Classification* **7,** 167-195.

[26] Rosenberg, S., and Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research* **10,** 489-502.

[27] Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

[28] Shepard, R. N. and Arabie, P. (1979). Additive custering representation of similarities as combinations of discrete overlapping properties. *Psychological Review* **86,** 87-123.

[29] Srivastava, R. K., Alpert, M. I., and Shocker, A. D. (1984). A customer-oriented approach for determining market structures. *Journal of Marketing* **48,** 32-45.

[30] Tucker, L. R., and Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika* **28,** 333-367.