Multidimensional Scaling of Sorting Data*

Yoshio Takane

McGill University

1. Introduction

In the stimulus sorting method the subjects are asked to sort a set of stimuli into as many sorting clusters as they wish so that the stimuli in a cluster are more similar to each other than those in different clusters. This method has enjoyed great popularity among social scientists as a quick and easy data collection method for similarities. In this paper we develop a multidimensional quantification method for the sorting data collected over a sample of subjects. Given multiple sets of sorting data this method finds, in a multidimensional Euclidian space, a configuration of points in such a way that a weighted sum of squared distances between cluster centroids averaged over the subjects is maximized under suitable normalization restrictions.

2. The Method

Let us assume that each of N individuals has sorted a set of n stimuli into N_k (k=1,...,N) clusters (groups) in terms of similarity among the stimuli. For each subject define an n by N_k matrix $\underset{\sim}{G}_k$ of dummy variables indicating a group to which each of the n stimuli belongs. That is,

$$(1) \qquad \underset{\sim}{G}_k = [g_{ir}^{(k)}], \ (i=1,\ldots,n;\ r=1,\ldots,N_k;\ k=1,\ldots,N),$$

where
$$g_{ir}^{(k)} = \begin{cases} 1, & \text{if stimulus i is classified into cluster r by subject k} \\ 0, & \text{otherwise.} \end{cases}$$

---

* The preparation of this manuscript was supported by A6394 to the author from the Natural Sciences and Engineering Council of Canada.

Let $\underline{X}$ denote an n by t matrix of stimulus coordinates common to all individuals, where t is the dimensionality of the representation space. Without loss of generality we assume that $\underline{X}$ is columnwise centered.
That is,

$$(2) \quad \underline{1}_n'\underline{X} = \underline{0}_t'$$

(where $\underline{0}_t$ is the t-component zero vector), or more indirectly,

$$(3) \quad \underline{\Pi}_n^{\perp}\underline{X} = \underline{X},$$

where $\underline{\Pi}_n^{\perp}$ is the centering matrix of order n.

We now define an n(n-1)/2 by n design matrix $\underline{A}$ for every possible pairwise comparison between stimulus coordinates. Let $\underline{D}$ be the matrix of Euclidian distances between stimuli.
We have

$$(4) \quad \sum_{i,j<i} d_{ij}^2 = \tfrac{1}{2}\mathrm{tr}\underline{D}^2 = \mathrm{tr}(\underline{A}\underline{X}\underline{X}'\underline{A}') = n\mathrm{tr}(\underline{X}\underline{\Pi}_n^{\perp}\underline{X}) = n\mathrm{tr}(\underline{X}'\underline{X})$$

(Takane, 1977). Note that $\underline{A}'\underline{A} = n\underline{\Pi}_n^{\perp}$. Similarly, the sum of squared Euclidian distances between cluster centroids for subject k is given by

$$(5) \quad \sum_{r,s<r} n_r^{(k)} n_s^{(k)} (\bar{d}_{rs}^{(k)})^2 = \mathrm{tr}(\underline{A}\underline{\Pi}_{G_k}\underline{X}\underline{X}'\underline{\Pi}_{G_k}\underline{A}')$$

$$= n\mathrm{tr}(\underline{X}'\underline{\Pi}_{G_k}\underline{\Pi}_n^{\perp}\underline{\Pi}_{G_k}\underline{X})$$

$$= n\mathrm{tr}(\underline{X}'\underline{\Pi}_{G_k}\underline{X}),$$

where $n_r^{(k)}$ and $n_s^{(k)}$ are the numbers of stimuli put in clusters r and s, respectively, by subject k, $\bar{d}_{rs}^{(k)}$ is the Euclidian distance between centroids of clusters r and s, and where

$$(6) \quad \underline{\Pi}_{G_k} = \underline{G}_k(\underline{G}_k'\underline{G}_k)^{-1}\underline{G}_k'.$$

We have an identity,

$$(7)\qquad \tfrac{1}{2}\mathrm{tr}\underline{D}^2 = n\mathrm{tr}(\underline{X}'\underline{X}) = n[\mathrm{tr}(\underline{X}'\underline{\Pi}_{G_k}\underline{X}) + \mathrm{tr}(\underline{X}'\underline{\Pi}^{\perp}_{G_k}\underline{X})],$$

where $\underline{\Pi}^{\perp}_{G_k} = \underline{I}_n - \underline{\Pi}_{G_k}$ for each k. The second term on the right hand side of the avobe identity represents the sum of squared Euclidian distances between stimulus points and their corresponding cluster centroids. If we divide both sides of (7) by n and take an average over subjects, we obtain

$$(8)\qquad \mathrm{tr}(\underline{X}'\underline{X}) = \frac{1}{N}\sum_{k=1}^{N}[\mathrm{tr}(\underline{X}'\underline{\Pi}_{G_k}\underline{X}) + \mathrm{tr}(\underline{X}'\underline{\Pi}^{\perp}_{G_k}\underline{X})]$$

$$= \mathrm{tr}[\underline{X}(\frac{1}{N}\sum_{k=1}^{N}\underline{\Pi}_{G_k})\underline{X}] + \mathrm{tr}[\underline{X}'(\frac{1}{N}\sum_{k=1}^{N}\underline{\Pi}^{\perp}_{G_k})\underline{X}]$$

$$= \mathrm{tr}(\underline{X}'\underline{B}\underline{X}) + \mathrm{tr}(\underline{X}'\underline{B}^{\perp}\underline{X}),$$

where $\underline{B} = \frac{1}{N}\sum_{k=1}^{N}\Pi_{G_k}$ and $\underline{B}^{\perp} = \frac{1}{N}\sum_{k=1}^{N}\underline{\Pi}^{\perp}_{G_k}$. We might determine $\underline{X}$ so that $\mathrm{tr}(\underline{X}'\underline{B}\underline{X})$ is maximized for a fixed value of $\mathrm{tr}(\underline{X}'\underline{X})$, say $\mathrm{tr}(\underline{X}'\underline{X}) = 1$. This is quite sensible, because $\mathrm{tr}(\underline{X}'\underline{B}\underline{X})$ represents the portion of $\mathrm{tr}(\underline{X}'\underline{X})$ which is strictly related to inter-cluster distances. However, when $t>1$ (the multidimensional case), we need an additional restriction on $\underline{X}$. It is convenient to require

$$(9)\qquad \underline{X}'\underline{X} = \underline{I}_t.$$

It is well known that the maximum of $\mathrm{tr}(\underline{X}'\underline{B}\underline{X})$ under this restriction is given by the matrix of normalized eigenvectors of $\underline{B}$ corresponding to its t dominant eigenvalues. However, $\underline{X}$ should also satisfy the centering restriction (2). Fortunately, this can be handled rather trivially, since $\underline{B}$ has an eigenvector proportional to $\underline{1}_n$ ($\underline{B}\underline{1}_n = \underline{1}_n$) and all other eigenvectors are orthogonal to this vector. We should simply avoid the constant eigenvector to be included in $\underline{X}$. This amounts to defining

$$(10)\quad \underline{B}^* = \underline{B} - \underline{1}_n\underline{1}_n'/n$$

and obtaining t eigenvectors of $\underline{B}^*$ (corresponding to its t dominant eigenvalues), assuming that $\underline{B}^*$ has at least t nonzero eigenvalues.

Once $\underline{X}$ is obtained, the cluster centroids for each subject can be obtained by

$$(11)\quad \underline{Y}_k = (\underline{G}_k'\underline{G}_k)^{-1}\underline{G}_k'\underline{X}, \ (k=1,\ldots,N).$$

This $\underline{Y}_k$ provides information concerning individual differences in sorting behavior.

The proposed method has a rather straightforward relationship to conventional dual scaling (Nishisato, 1980), also known as the type III quantification method (Hayashi, 1952) or correspondence analysis (Benzécri, et.al., 1973). This, as well as other details of the derivation of the method, is fully described in Takane (1980).

3. A Similar Method

The total sum of squared distances (4) can be decomposed in another interesting way. A distance is defined either between two stimuli in a same cluster or between two stimuli in different clusters. A set of distances can thus be partitioned uniquely into two mutually exclusive subsets. Let a diagonal matrix $\underline{D}_{B_k}$ of order $n\times(n-1)/2$ whose m^{th} diagonal element is one, if a pair of stimuli corresponding to the m^{th} row of matrix $\underline{A}$ are put into different clusters by subject k, and zero, otherwise, and $\underline{D}_{W_k} = \underline{I}_{n(n-1)/2} - \underline{D}_{B_k}$. Then we have

$$(12)\quad tr(\underline{X}'\underline{X}) = tr(\underline{X}'\underline{H}_k\underline{X}) + tr(\underline{X}'\underline{\tilde{H}}_k\underline{X}),$$

where $\underline{H}_k = \underline{A}'\underline{D}_{B_k}\underline{A}/n$ and $\underline{\tilde{H}}_k = \underline{A}'\underline{D}_{W_k}\underline{A}/n$, and, by averaging (12) over N subjects, we obtain

$$(13) \quad tr(\underline{X}'\underline{X}) = tr(\underline{X}'\underline{H}\underline{X}) + tr(\underline{X}'\underline{\tilde{H}}\underline{X})$$

where $\underline{H} = \sum_{k=1}^{N} \underline{H}_k/N = \underline{A}'[\frac{1}{N}\sum_{k=1}^{N}(\underline{D}_{B_k}/n)]\underline{A}$ and $\underline{\tilde{H}} = \sum_{k=1}^{N} \underline{\tilde{H}}_k/N = \underline{A}'[\frac{1}{N}\sum_{k=1}^{N}(\underline{D}_{W_k}/n)]\underline{A}$.

The $\underline{X}$ which maximizes $tr(\underline{X}'\underline{H}\underline{X})$ under restriction (9) can be obtained by solving for the eigenvectors of $\underline{H}$ corresponding to its A dominant eigenvalues.

The above procedure happens to be equivalent to Hayashi's (1952) type IV quantification method, which obtains eigenvectors of matrix $\underline{A}'\underline{D}_e\underline{A}$ (Takane, 1977), where $\underline{D}_e$ is any diagonal matrix of dissimilarity constructed in a manner similar to $\underline{D}_{B_k}$. In the present case we set $\underline{D}_e = \frac{1}{N}\sum_{k=1}^{N}(\underline{D}_{B_k}/n)$. Our experience indicates that this method also gives results similar to those obtained by the method developed in section 2.

4. Process Model for Sorting

The criterion of maximizing the weighted sum of squared distances between cluster centroids may be somewhat arbitrary from an empirical point of view. In particular it does not seem to have anything to do with the way in which the subjects perform the sorting task. It still remains an empirical question whether sorting clusters are actually conceived by the subjects as such. In what follows we develop a psychological model of sorting behavior.

Takane (1980) discuss a close relationship between the proposed method and Coombs' (1964) unfolding model. In constructing the psychological model for sorting we take the idea underlying the unfolding model as a point of departure. We assume that each subject has an ideal point for each sorting cluster, and that the ideal point for a cluster is given by the cluster centroid. We further assume that the closer a stimulus point to the ideal, the higher is the probability that the stimulus is classified into the cluster of the ideal. The probability should be small, if the stimulus is located far away from the ideal. Thus, we may explicitly state the probability of a particular stimulus put into a certain cluster as some decreasing function of the distances between the stimulus point and the ideal points of clusters. The problem of determining $\underline{X}$ then is to maximize the probability of obtaining observed clusters over N subjects. We use Luce's (1959) type choice model to specify the probability. Let $\tilde{d}_{ir}^{(k)}$ represent the distance between stimulus i and the centroid of cluster r by subject k. Then under this model the probability $p_{ir}^{(k)}$ that subject k puts stimulus i into cluster r is given by

$$(14) \quad p_{ir}^{(k)} = \frac{\exp(-\tilde{d}_{ir}^{(k)})}{\sum_{s=1}^{N_k} \exp(-\tilde{d}_{is}^{(k)})}.$$

A particular stimulus can be put into any one of N_k clusters by subject k. The multinomial distribution is appropriate for this kind of situation. Let $p_i^{(k)}$ denote the probability that stimulus i is put into a certain cluster by subject k. Then

$$(15) \quad p_i^{(k)} = \prod_{r=1}^{N_k} (p_{ir}^{(k)})^{g_{ir}^{(k)}} ,$$

and the likelihood function for the total set of observations, may now be stated as

$$(16) \quad L = \prod_{k,i} p_i^{(k)} .$$

We obtain $\underline{X}$ which maximizes the likelihood. The likelihood can be maximized numerically using the scoring method.

Again, our experience shows that this new criterion gives similar results to those obtained by the method developed in section 2, justifying the criterion employed previously.

7. Concluding Remarks

In this paper we proposed a method of multdimensional scaling for sorting data. This method is simple and has a special advantage when one wishes to obtain a quick multidimensional scaling solution from the sorting data. Unlike conventional multidimensional scaling procedures, it requires no prior conversions of the sorting data into similarity data. Furthermore, it permits a sort of individual differences analysis with the sorting data. The straightforward relationship of the method to dual scaling as well as to the unfolding model adds further credibility to the proposed method. It has been shown that it could provide reasonable approximations to a psychological process model of sorting behavior.

It might be interesting to extend the proposed method to three-way cases allowing individual differences in the stimulus configuration. If the conceived space is the weighted Euclidian, then CANDECOMP (Carroll and Chang, 1970) type algorithm should directly apply to this situation. Programs are now available for both the method proposed in this paper (Takane, 1981) and its three-way extension (Takane, 1983).

References

Benzécri, J.P., et.al. (1973) L'Analyse des donnees: I & II. Paris: Dunod.

Carroll, J.D., and Chang, J.J. (1970) "Analysis of individual differences in multidimensional scaling via an N-way generalization of the Eckart-Young decomposition." Psychometrika, 35, 283-319.

Hayashi, C. (1952) "On the prediction of phenomena from qualitative data and the quantification of wualitative data from mathematico-statistical point of view." Annals of the Institute of Statistical Mathematics, 3, 69-98.

Luce, R.D. (1959) Individual choice behavior. New York: Wiley.

Nishisato, S. (1980) Analysis of categorical data: dual scaling and its applications. Toronto: University of Toronto Press.

Takane, Y. (1977) "On the relations among four methods of multidimensional scaling." Behaviormetrika, 4, 29-43.

Takane, Y. (1980) "Analysis of categorizing behavior by a quantification method." Behaviormetrika, 8, 57-67.

Takane, Y. (1981) "MDSORT: A special-purpose multidimensional scaling program for sorting data." Behavior Research Methods & Instrumentation, 13, 698.

Takane, Y. (1982) "IDSORT: An individual differences multidimensional scaling program for sorting data." Behavior Research Methods & Instrumentation, 14, 546.