Behaviormetrika
1980, No. 8, 75-86

ANALYSIS OF CATEGORIZING BEHAVIOR BY A QUANTIFICATION METHOD

Yoshio Takane*

A quantification method was developed for 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 the sum of squared inter-cluster distances averaged over subjects is maximized under suitable normalization conditions. Examples were given to illustrate the use of the method and its relationship to other scaling methods was discussed.

1. Introduction

The stimulus sorting method has been very popular among social scientists as a quick and easy data collection method for similarities (Clark, 1968; Miller, 1969; Burton, 1972; Rapoport & Fillenbaum, 1972; Rosenberg & Sedlak, 1972; Steffire, 1972). In its standard format the method requires subjects to sort a set of stimuli into as many groups as they wish in terms of similarity between the stimuli. This method is deemed particularly appealing, 1) when the subjects are naive (the sorting task is very easy to perform), 2) when the number of stimuli is very large (more than 20), and 3) when individual differences in the perceptual structure of stimuli are unimportant, or at least not the subject matter of research.

Sorting data collected over subjects are usually analyzed in an aggregated form. They are first converted into a matrix of (dis) similarities, which are then subjected to an analysis by (nonmetric) multidimensional scaling (Kruskal, 1964), by hierarchical clustering schemes (Johnson, 1967) or by latent partition analysis (Wiley, 1967; Evans, 1970) to find out a structure underlying the similarity matrix. Numerous kinds of similarity indices have been devised for this purpose. These range from simple frequency counts to more sophisticated information theoretic measures of similarity (Burton, 1972; see also Rapoport and Fillenbaum, 1972). Unfortuntely those indices are all *ad hoc* in the sense that no explicit representation models of the data are taken into account in their derivation. However, if the data are ultimately to be represented by some model (e.g., a distance model in multidimensional scaling), it is certainly preferable that the initial data conversion porcess itself is in some sense consistent with the representation model of the data (Takane, Young & de Leeuw, 1977; Young, de Leeuw & Takane, 1980). In this paper we develop a multidimensional quantification method which meets this basic requirement.

The method we develop in this paper simultaneously scales and represents the sorting data. It finds a configuration of stimulus points in such a way that the sum of

Key Words and Phrases: sorting method, quantification method, dual scaling, unfolding model
* Department of Psychology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec H3A 1B1, Canada. The research reported in this paper was supported by Grant A6394 to the author by Natural Sciences and Engineering Research Council of Canada.

squared inter-cluster distances averaged over subjects is a maximum given suitable normalization restrictions on the configuration. The scaling aspect of the method permits a kind of individual differences analysis in categorizing behavior (Medin & Schaffer, 1978) despite the fact that the sorting data have been considered more appropriate for non-individual-differences type of analysis. (We will get back to this point later on.)

The general approach we pursue in this paper is very much in the spirit of the Guttman-Hayashi (Guttman, 1941; Hayashi, 1952) tradition of optimal scaling in the sense that it seeks to find a quantification of stimuli based on some optimality considerations. In a later section of this paper we discuss a relationship of the proposed method to a conventional method of optimal scaling (i.e., Quantification Method III by Hayashi, 1952, or Dual Scaling, Nishisato, 1980) as well as a relationship to one of the most representative methods of psychological scaling (the unfolding model by Coombs, 1964). A weakness of the current method as a descriptive model will be briefly discussed in the final section.

2. The method

Let us assume that each of N individuals has sorted a set of n stimuli into N_k ($k=1, \cdots, N$) clusters (groups) in terms of similarity among the stimuli. Most often the number of groups into which stimuli are classified is left to each subject's own discretion, though it is quite sensible to avoid $N_k=1$ or $N_k=n$. For each subject (indexed by k), define an n by N_k matrix $\boldsymbol{G}_k$ of dummy variables indicating a group to which each of the n stimuli belongs. That is,

$$\boldsymbol{G}_k = [g_{irk}], (i=1, \cdots, n; r=1, \cdots, N_k; k=1, \cdots, N), \tag{1}$$

where

$$g_{irk} = \begin{cases} 1, & \text{if stimulus } i \text{ is calssified into cluster } r \text{ by subject } k \\ 0, & \text{otherwise.} \end{cases}$$

We assume that matrix $\boldsymbol{G}_k$ is complete in the sense that every stimulus belongs to one and only one cluster; *i.e.*,

$$\boldsymbol{G}_k \boldsymbol{l}_{N_k} = \boldsymbol{l}_n, \quad (k=1, \cdots, N), \tag{2}$$

where $\boldsymbol{l}_{N_k}$ and $\boldsymbol{l}_n$ are, respectively, N_k- and n-component vectors of ones. Let $\boldsymbol{X}$ denote an n by A matrix of stimulus coordinates common to all individuals, where A is the dimensionality of the representation space. Without loss of generality we assume that $\boldsymbol{X}$ is columnwise centered. That is,

$$\boldsymbol{l}_n' \boldsymbol{X} = \boldsymbol{O}_A', \tag{3}$$

(where $\boldsymbol{O}_A$ is the A-component zero vector), or more indirectly,

$$\boldsymbol{\Pi}_n^{\perp} \boldsymbol{X} = \boldsymbol{X}, \tag{4}$$

where $\boldsymbol{\Pi}_n^{\perp}$ is the contering matrix of order n (*i.e.*, $\boldsymbol{\Pi}_n^{\perp}=\boldsymbol{I}_n-\boldsymbol{\Pi}_n$ where $\boldsymbol{I}_n$ is the identity matrix of order n and $\boldsymbol{\Pi}_n=\boldsymbol{l}_n\boldsymbol{l}_n'/n$).

We then define an $n(n-1)/2$ by n matrix A, which defines every possible pairwise difference between stimulus coordinates. Let D be the matrix of Euclidian distances between stimuli.

We have

$$\sum_{i,j<i} d_{ij}^2 = \frac{1}{2} trD^2 = tr(AXX'A') = n\, tr(X\Pi_n^{\perp}X) = n\, tr(X'X) \tag{5}$$

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

$$\begin{aligned}\sum_{r,s<r} n_r^{(k)} n_s^{(k)} (\bar{d}_{rs}^{(k)})^2 &= tr(A\Pi_{G_k}XX'\Pi_{G_k}A') \\ &= n\, tr(X'\Pi_{G_k}\Pi_n^{\perp}\Pi_{G_k}X) \\ &= n\, tr(X'\Pi_{G_k}X)\,, \end{aligned} \tag{6}$$

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

$$\Pi_{G_k} = G_k(G_k'G_k)^{-1}G_k'\,. \tag{7}$$

Notice that $\bar{d}_{rs}^{(k)}$ is weighed by the product of $n_r^{(k)}$ and $n_s^{(k)}$ before it is summed over pairs of clusters in order to reflect the size of clusters in the summation. Note also that Π_{G_k} is a similarity matrix whose elements are positive when a stimulus corresponding to a row and a stimulus corresponding to a column are classified into a same group, and are zero otherwise. The (i, j) element of G_kG_k' is one when stimuli i and j are put into a same group by subject k, and is zero otherwise. The $(G_n'G_n)^{-1}$ between G_k and G_k' has the effect of scaling nonzero elements of G_kG_k' by the size of groups. That is, if the (i, j) element of G_kG_k' is one, and the number of stimuli in the group to which both i and j belong is n_r, then the (i, j) element of Π_{G_k} is $1/n_r$; the similarity between two stimuli which are classified into a same group is inversely related to the number of stimuli in that group. This makes intuitive sense, since, for example, two stimuli are considered more similar to each other when they are put in a group consisting of those two stimuli alone than when they are put in a group of a thousand others.

We have an identity.

$$\frac{1}{2} trD^2 = n\, tr(X'X) = n[tr(X'\Pi_{G_k}X) + tr(X'\Pi_{G_k}^{\perp}X)]\,, \tag{8}$$

where $\Pi_{G_k}^{\perp}=I_n-\Pi_{G_k}$ for each k. The second term on the right hand side of the above identity, $tr(X'\Pi_{G_k}^{\perp}X)$, represents the sum of squared Euclidian distances between stimulus points and their corresponding cluster centroids. Thus, Eq. (8) states the basic decomposition of the total sum of squared Euclidian distances into two components, one related to inter-cluster distances and the other related to intra-cluster distances (Takeuchi & Yanai, 1972).

If we divide both sides of (8) by n and take an average over subjects, we obtain

$$tr(X'X) = \frac{1}{N}\sum_{k=1}^{N}[tr(X'\Pi_{G_k}X)+tr(X'\Pi_{G_k}^{\perp}X)]$$
$$= tr\Big[X'\Big(\frac{1}{N}\sum_{k=1}^{N}\Pi_{G_k}\Big)X\Big]+tr\Big[X'\Big(\frac{1}{N}\sum_{k=1}^{N}\Pi_{G_k}^{\perp}\Big)X\Big]$$
$$= tr(X'BX)+tr(X'B^{\perp}X)\,, \qquad (9)$$

where $B = \frac{1}{N}\sum_{k=1}^{N}\Pi_{G_k}$ and $B^{\perp} = \frac{1}{N}\sum_{k=1}^{N}\Pi_{G_k}^{\perp}$. Matrix B is an average similarity matrix. We might determine X so that $tr(X'BX)$ is maximized for a fixed value of $tr(X'X)$, say $tr(X'X)=1$. This is quite sensible, because $tr(X'BX)$ represents the portion of $tr(X'X)$ which is strictly related to inter-cluster distances. However, when $A>1$ (the multidimensional case), the normalization restriction on X ($tr\,X'X=1$) alone is not sufficient; some kind of linear independence restriction is necessary on the column vectors of X. It is convenient to require X to be columnwise orthonormal; *i.e.*,

$$X'X = I_A\,. \qquad (10)$$

It is well known that the maximum of $tr(X'BX)$ under this restriction is given by the matrix of normalzied eigenvectors of B corresponding to its A dominant eigenvalues. However, X should also satisfy the centering restriction (3). Fortunately, this can be handled rather trivially, since B has an eigenvector proportional to 1_n ($Bl_n=l_n$) and all other eigenvectors are orthogonal to this vector. We should simply avoid the constant eigenvector to be included in X. This amounts to defining

$$B^* = B-l_n l_n'/n \qquad (11)$$

and obtaining A eigenvectors of B^* (corresponding to its A dominant eigenvalues) instead of B, assuming that B^* has at least A nonzero eigenvalues.

Once X is obtained, the cluster centroids for each subject can be obtained by

$$Y_k = (G_k'G_k)^{-1}G_k'X\,, \quad (k=1,\cdots,N)\,. \qquad (12)$$

This Y_k provides information concerning individual differences in sorting behavior.

3. Illustrative examples

In this section we present some results of analysis performed by the proposed method. Two sets of stimuli were used in this study; both of them were extracted from Rapoport and Fillenbaum (1972). The first set of stimuli consisted of 24 color names and the second set of 29 HAVE words. (See Tables 2 and 4 for specific stimuli employed.) Stimuli were each printed on IBM cards, and given to the subjects in two separate decks. Ten university students (psychology majors at McGill) participated in the experiment. For each stimulus set the subject was asked to sort the stimuli into as many groups (clusters) as they wanted. An average subject took fifteen minutes in total to complete the task.

The data were analyzed by the method discussed in the previous section. Appropriate dimensionalities of solutions may be determined by Bartlett's chi square

correction formula (Nishisato, 1975, p. 183) for the test of significant eigenvalues. This formula can be written, under the present notation, as

$$\chi^2 = -\left[Nn-1-\frac{n+\sum_{k=1}^{N} N_k - N}{2}\right] ln(1-\lambda_j^2) ,$$

where λ_j is the j-th largest eigenvalue of B^*. The degree of freedom associated with the above chi square is $n+\sum_{k=1}^{N} N_k - N - 2j$. A large value of the chi square indicates a significant departure of λ_j from zero.

Table 1
Eigenvalues (A), chi squares (B), and the associated d.f. (C) for the color data

	(A)	(B)	(C)
1	0.834	235.573	80
2	0.813	213.775	78
3	0.724	147.284	76
4	0.602	89.287	74
5	0.557	73.394	72
6	0.451	44.917	70
7	0.357	26.933	68
8	0.242	11.972	66
9	0.224	10.191	64
10	0.186	6.969	62
11	0.151	4.562	60
12	0.142	4.046	58
13	0.123	3.011	56
14	0.104	2.133	54
15	0.090	1.623	52
16	0.064	0.823	50
17	0.043	0.373	48
18	0.036	0.255	46
19	0.034	0.226	44
20	0.016	0.050	42
21	0.007	0.011	40
22	0.000	0.0	38
23	0.000	0.0	36
24	0.000	0.0	34

Eigenvalues, chi squares and associated d.f. are shown in Table 1 for the color data. The 5% significance level leads to three significant eigenvalues. Stimulus coordinates corresponding to the significant eigenvalues are given in Table 2.

Evidently dimension 1 represents a contrast between brownish colors and others (red and blue) which are, in turn, distinguished on dimension 2, and dimension 3 represents a red-green contrast. While not statistically significant, a few subsequent dimensions are also interpretable: Dimension 4 (yellow vs others), dimension 5 (ivory, silver vs brown) and dimension 6 (purple vs blue). The general tendency is that, as we move from more dominant dimensions to less dominant ones, a fewer stimuli are apt to dominate the whole dimension.

Table 2
Derived stimulus configuration for the color data

		dim 1	dim 2	dim 3
1	Beige	0.221	−0.011	−0.175
2	Blue	−0.251	0.416	−0.212
3	Bronze	0.250	−0.064	−0.193
4	Brown	0.262	−0.061	−0.199
5	Chartreuse	0.097	0.149	0.339
6	Crimson	−0.250	−0.222	0.066
7	Gold	0.202	−0.066	−0.119
8	Green	0.081	0.259	0.420
9	Ivory	0.180	−0.021	−0.212
10	Khaki	0.152	0.175	0.322
11	Magenta	−0.238	−0.273	0.066
12	Mustard	0.101	−0.080	0.109
13	Olive	0.120	0.213	0.387
14	Orange	0.003	−0.142	0.081
15	Pink	−0.223	−0.269	0.070
16	Purple	−0.284	0.133	−0.131
17	Red	−0.258	−0.269	0.081
18	Rust	0.120	−0.118	−0.096
19	Scarlet	−0.258	−0.269	0.081
20	Silver	0.118	−0.028	−0.229
21	Tan	0.262	−0.044	−0.174
22	Turquoise	−0.230	0.421	−0.199
23	Violet	−0.279	0.252	−0.194
24	Yellow	0.101	−0.080	0.109

The same chi square criterion indicates two significant dimensions for the HAVE words. (See Table 3.) Stimulus coordinates for the two significant dimensions are displayed in Table 4.

It is clear that dimension 1 represents a contrast between the steady state of possession (*e.g.*, have) and the state of possession which is about to be lost (*e.g.*, give). Dimension 2, on the other hand, represents a distinction between the state of nonpossession which may not change for the time being (lack, need and want) and the state of nonpossession which is about to change (*e.g.*, receive). Figure 1 shows the plot of stimulus coordinates given in Table 4 along with sorting clusters (encircled) by one of the subjects. This subject sorted the 29 stimuli into four clusters. Those four clusters roughly correspond with the four states of possession mentioned above. It is interesting to note that for the stimuli in clusters 1 and 3 (two states of nonpossession), dimension 2 seems to represent the shades of sureness of change in the states. For example, "receive" is right at the bottom indicating that the state of nonpossession is most likely to change; "lack" is located at the top, which bears no information about the possibility of change in the state, "beg" is located in the middle, which indicates that an action is taken to change the state (the change is probable, but not certain).

In Figure 1 cluster centroids for the one subject are indicated by stars. By allowing to draw in sorting clusters and their centroids by different subjects on the plot of stimulus configuration, the present method may lead to important insight into the nature of individual differences in the sorting task.

Table 3
Eigenvalues (A), chi squares (B), and the associated d.f. (C) for the HAVE data

	(A)	(B)	(C)
1	0.732	187.146	89
2	0.684	153.747	87
3	0.596	106.984	85
4	0.544	85.473	83
5	0.468	60.186	81
6	0.418	46.737	79
7	0.357	33.151	77
8	0.298	22.612	75
9	0.276	19.352	73
10	0.255	16.362	71
11	0.233	13.650	69
12	0.203	10.256	67
13	0.184	8.384	65
14	0.146	5.216	63
15	0.137	4.587	61
16	0.115	3.267	59
17	0.100	2.459	57
18	0.091	2.026	55
19	0.075	1.389	53
20	0.066	1.047	51
21	0.057	0.786	49
22	0.050	0.607	47
23	0.043	0.444	45
24	0.026	0.167	43
25	0.022	0.117	41
26	0.012	0.037	39
27	0.008	0.015	37
28	0.003	0.003	35
29	0.000	0.0	33

4. Relation to other methods

The proposed method has a rather straightforward relationship to two of the most representative scaling methods to date.

Let

$$G = [G_1, \cdots, G_N] . \tag{13}$$

Dual scaling (Nishisato, 1980; Hayashi, 1952) of this binary matrix obtains a matrix X of row quantifications as eigenvectors of matrix

$$H = D_R^{-1/2}\left(GD_C^{-1}G' - \frac{D_R l_R l_R' D_R}{M}\right) D_R^{-1/2} \tag{14}$$

where D_R and D_C are diagonal matrices of row and column totals of G, l_R is the R-component vector of ones (where R is the number of rows in G, which is equal to n in the present case), and $M = l_R' D_R l_R$.

Since for G defined in (13) we have

Table 4
Derived stimulus configuration for the HAVE data

		dim 1	dim 2
1	Accept	0.078	−0.186
2	Beg	0.094	0.099
3	Belong	−0.302	0.105
4	Borrow	0.053	−0.029
5	Bring	0.110	−0.114
6	Buy	0.028	−0.096
7	Earn	−0.099	−0.197
8	Find	0.064	−0.199
9	Gain	−0.123	−0.167
10	Get	0.015	−0.206
11	Get rid of	0.247	0.122
12	Give	0.253	−0.018
13	Have	−0.347	0.042
14	Hold	−0.270	−0.060
15	Keep	−0.296	−0.039
16	Lack	0.083	0.479
17	Lend	0.192	0.024
18	Lose	0.242	0.058
19	Need	−0.056	0.470
20	Offer	0.212	−0.002
21	Own	−0.333	0.064
22	Receive	0.084	−0.223
23	Return	0.188	−0.115
24	Save	−0.249	−0.086
25	Sell	0.253	0.002
26	Steal	0.001	−0.081
27	Take	0.002	−0.087
28	Use	−0.118	−0.015
29	Want	−0.005	0.457

$$D_R = NI_n\,,$$

$$D_C = diag\,(G'G)\,,$$

and

$$M = Nl_n'l_n = Nn\,,$$

(14) reduces to

$$H = \frac{1}{N}G[diag\,(G'G)]^{-1}G' - l_n l_n'/n$$

$$= \frac{1}{N}\sum_{k=1}^{N} G_k(G_k'G_k)^{-1}G_k' - l_n l_n'/n$$

$$= B - l_n l_n'/n = B^*. \qquad (15)$$

Although the proposed method was derived from an entirely different principle, (15) shows that it is a special case of dual scaling. In dual scaling approach column quantification matrix Y is obtained by

$$Y = D_C^{-1}GX = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}, \qquad (16)$$

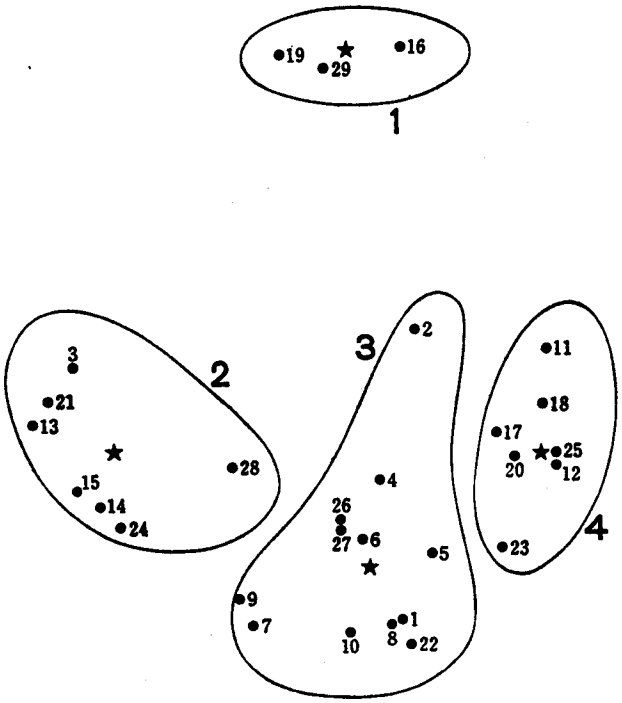

Fig. 1 Plot of stimulus configuration for the HAVE data, and clusters and cluster centroids for subject 6

which also agrees with $\boldsymbol{Y}_k$ ($k=1, \cdots, N$), the matrix of cluster centroids for subject k, given in (12).

The unfolding model (Coombs, 1964) postulates ideal points for clusters. Let $\boldsymbol{Y}_k$ denote the matrix of coordinates of ideal points for clusters by subject k. Define

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \tag{17}$$

where

$$\boldsymbol{Y}' = [\boldsymbol{Y}_1', \cdots, \boldsymbol{Y}_N'],$$
$$\boldsymbol{D}_G = \begin{bmatrix} \boldsymbol{G}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{G}_N \end{bmatrix}, \tag{18}$$

and

$$\boldsymbol{E} = \left[\begin{array}{c|c} \boldsymbol{I}_n & \\ \vdots & -\boldsymbol{D}_G \\ \boldsymbol{I}_n & \end{array}\right]. \tag{19}$$

Then *diag* $(\boldsymbol{EZZ'E'})$ is a diagonal matrix of squared distances between stimulus points and ideal points of clusters to which they belong. We wish to find $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that they minimize *tr* [*diag* $(\boldsymbol{EZZ'E'})$] under suitable normalization conditions on $\boldsymbol{X}$. Define

$$\begin{aligned} g &= tr\,[diag\,(\boldsymbol{EZZ'E'})] \\ &= tr\,(\boldsymbol{Z'E'EZ}) \\ &= N\,tr\,(\boldsymbol{X'X}) - 2\,tr\,(\boldsymbol{X'GY}) + tr\,(\boldsymbol{Y'D_G'D_GY}) \end{aligned} \tag{20}$$

Differentiating g with respect to $\boldsymbol{Y}$ and setting the result to zero gives

$$-\frac{1}{2}\frac{\partial g}{\partial \boldsymbol{Y}} = \boldsymbol{G'X} - \boldsymbol{D_G'D_GY} \equiv \boldsymbol{O},$$

which leads to

$$\boldsymbol{Y} = (\boldsymbol{D_G'D_G})^{-1}\boldsymbol{G'X} = [diag\,(\boldsymbol{G'G})]^{-1}\boldsymbol{G'X}\,. \tag{21}$$

Define g^* to be the minimum of g over $\boldsymbol{Y}$; *i.e.*,

$$g^* = \min_{\boldsymbol{Y}} g = N\,tr(\boldsymbol{X'X}) - tr[\boldsymbol{X'G}(\boldsymbol{D_G'D_G})^{-1}\boldsymbol{G'X}]\,. \tag{22}$$

Minimizing g^* over normalized $\boldsymbol{X}$ (10) is equivalent to maximizing the second term on the right hand side under the same normalization restriction on $\boldsymbol{X}$, since $tr\,(\boldsymbol{X'X})$ is constant in this case. Note that $\boldsymbol{G}(\boldsymbol{D'_GD_G})^{-1}\boldsymbol{G'} = \boldsymbol{NB}$. Thus, the proposed method can also be considered a special type of the unfolding model.

A less directly related method is the latent partition analysis by Wiley (1967; Evans, 1970). Although the latent partition analysis is designed for the same kind of sorting data as the present method is designed for, it assumes a discrete structure (a modal partition of stimuli) underlying the manifest sorting data as opposed to a continuous Euclidian space in the present method. It is interesting, however, to compare the usefulness of the two methods in the analysis of categorizing behavior.

5. Discussion and a prospect

We have seen a couple of examples of analysis by the proposed method. This method is simple (solutions can be obtained analytically), and has a special advantage when one wishes to obtain a quick multidimensional scaling solution from sorting data. The straightforward relationship of the method to other well-established scaling methods adds further credibility to the method. (At the same time the proposed method demonstrates yet another approach to the conventional scaling methods.)

Perhaps one of the major drawbacks of the present method is some arbitrariness in the choice of an optimization criterion. Although the criterion of maximizing the average sum of squared inter-cluster distances makes some intuitivie sense, its logical foundation, particularly as it pertains to descriptive relevance of the criterion, is not firmly established. We initially set out our discussion by stating that the scaling of original data should in some sense be consistent with the representation model of the data. This desideratum is met by the proposed method. However, our optimization criterion is still arbitrary in the sense that it has nothing to do with the way in which the subjects perform the sorting task. It is doubtful that the subjects actually conceptualize a set of stimuli in such a way that the sum of inter-cluster distances is a maximum. Preferably an optimization criterion is so constructed that it takes into account the actual (psychological) processes involved in a specific task situation

(Takane, 1980; Young, de Leeuw & Takane, 1980). The model of how the subjects perform the sorting task has to be an essential ingredient of the optimization criterion. We plan to develop such a procedure in the near future. The method proposed in this paper can be used as an initialization method to such a procedure.

6. Acknowledgement

The author is grateful to Dr. Nishisato for his helpful comments on an earlier draft of this paper.

References

Burton, M. (1972). Semantic dimensions of occupation names. In A.K. Romney, et. al. (Eds.), *Multidimensional scaling Vol. II*. New York: Seminar Press.

Clark , H.H. (1968). On the use and meaning of prepositions. *Journal of Verbal Learning and Verbal Behavior* **7** 421–431.

Coombs, C.H. (1964). *A theory of data*. New York: Wiley.

Evans, G.T. (1970). The analysis of categorizing behavior. *Psychometrika*, **35**, 367–382.

Guttman. L. (1941). The quantification of a class of attributes: a theory and method of scale construction. In P. Horst (Ed.) *The prediction of personal adjustment*. New York: Social Science Research Council.

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

Johnson, S.C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241–254.

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

Medin, D.L., and Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207–238.

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

Nishisato, S. (1975). *Oyo shinri shakudo hoseiho* (*Applied psychological scaling*). Tokyo: Seishin Shobo.

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

Rapoport, A., and Fillenbaum, S. (1972). An experimental study of semantic structures. In A.K. Romney, et. al. (Eds.), *Multidimensional scaling, Vol. II*. New York: Seminar Press.

Rosenberg, S., and Sedlak, A. (1972). Structural representations of perceived personality trait relationship. In A.K. Romney, et. al. (Eds.), *Multidimensional scaling, Vol. II*. New York: Seminar Press.

Steffire, V.J. (1972). Some application of multidimensional scaling to social science problems. In A.K. Romney, et. al. (Eds.), *Multidimensional scaling, Vol. II*. New York: Seminar Press.

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

Takane, Y. (1980). *Tajigen shakudoho* (*Multidimensional scaling*). Tokyo: University of Tokyo Press.

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

Takeuchi, K., and Yanai, H. (1972). *Tahenryo kaiseki no kiso* (*Foundations of multivariate analysis*). Tokyo: Toyokeizaishimpo.

Wiley, D.E. (1967). Latent partition analysis. *Psychometrika*, 32, 183–193.

Young, F.W., de Leeuw, J., and Takane, Y. (1980). Quantification of qualitative data. In E.D. Lantermann and H. Feger, (Eds.), *Similarity and choice*. Bern: Hans Huber.

(Received May, 1980)

ERRATA for Takane (Analysis of categorizing behavior by a quantification method. **Behaviormetrika**, 1980, 8, 75-88.)

The λ_j^2 in the chi square formula on the 3rd line of page 79 should be read as λ_j (without square). The squaring of λ_j has the effect of deflating the chi square value. Tables 1 and 3 should be replaced. (See the reverse side.) The model and the analytical method presented in the paper remain valid.

Table 1

Eigenvalues (A), chi squares (B), and the associated d.f. (C) for the color data

	(A)	(B)	(C)
1	0.834	355.669	80
2	0.813	331.536	78
3	0.724	255.167	76
4	0.602	182.654	74
5	0.557	161.001	72
6	0.451	118.557	70
7	0.357	87.323	68
8	0.242	54.920	66
9	0.224	50.208	64
10	0.186	40.740	62
11	0.151	32.394	60
12	0.142	30.375	58
13	0.123	25.952	56
14	0.104	21.633	54
15	0.090	18.752	52
16	0.064	13.182	50
17	0.043	8.788	48
18	0.036	7.240	46
19	0.034	6.802	44
20	0.016	3.179	42
21	0.007	1.464	40
22	0.000	0.0	38
23	0.000	0.0	36
24	0.000	0.0	34

Table 3

Eigenvalues (A), chi squares (B), and the associated d.f. (C) for the HAVE data

	(A)	(B)	(C)
1	0.732	320.942	89
2	0.684	280.683	87
3	0.596	220.864	85
4	0.544	191.257	83
5	0.468	153.660	81
6	0.418	131.762	79
7	0.357	107.443	77
8	0.298	86.084	75
9	0.276	78.777	73
10	0.255	71.655	71
11	0.233	64.747	69
12	0.203	55.278	67
13	0.184	49.505	65
14	0.146	38.312	63
15	0.137	35.765	61
16	0.115	29.871	59
17	0.100	25.721	57
18	0.091	23.238	55
19	0.075	19.097	53
20	0.066	16.498	51
21	0.057	14.234	49
22	0.050	12.468	47
23	0.043	10.622	45
24	0.026	6.461	43
25	0.022	5.389	41
26	0.012	3.027	39
27	0.008	1.901	37
28	0.003	0.814	35
29	0.000	0.0	33