Behaviormetrika
Vol.37, No.1, 2010, 15–32

TOTAL INFORMATION ANALYSIS: COMPREHENSIVE DUAL SCALING

Shizuhiko Nishisato* and José G. Clavel**

The traditional quantification procedure (e.g., dual scaling, correspondence analysis) is extended in order to tap into information which is typically ignored. Noting that the traditional symmetric scaling yields a visual image of distorted data structure and recalling that the widely used practice of looking at data in reduced space may also miss capturing rare but key-information in total space, a method, called total information analysis (TIA), is proposed to subject not only within-set but also between-set relations in total space. Numerical examples are used to explain why TIA offers partial solutions to some theoretical problems inherent in the current practice of multidimensional quantification analysis.

1. Introduction

The idea of capturing total information contained in given data was clearly in the mind of Andrews (1972) behind the proposal of his well-known Andrews curves. His proposal is ideal in the sense that the curves represent the entire information of data in a simple two-dimensional graph. At the same time, however, it is difficult to interpret the curves since they are sums of trigonometric functions, associated with components in the data.

The present study shares the same objective as Andrews' in analyzing the entire information in data, but proposes quite a different approach from the Andrews curves: our top priority is placed on the interpretation of the entire data structure. In this regard, our approach has something in common with that of Van Deun et al. (2007), which will be discussed at the end of the current paper.

To begin with, let us look at some possibly problematic aspects of the current practice in multidimensional analysis of categorical data, in particular, symmetric scaling, one-mode analysis and dimension reduction. It should be noted, however, that these are problematic at least from our point of view.

1.1 Symmetric Scaling and Alternatives

The first problem comes from the practical consideration of plotting the row variables and the column variables jointly in the same space (symmetric scaling), or plotting only row variables or column variables (one-mode analysis). The problem of

Key Words and Phrases: Symmetric scaling, the Young-Householder theorem, within-set and between-set distances, multidimensional analysis in reduced space, axis interpretation versus cluster interpretation, k-means clustering.

* University of Toronto, Toronto, Canada
** Universidad de Murcia, Spain
Correspondence concerning the paper should be addressed to S. Nishisato, 9 St. George's Road, Toronto, Ontario M9A 3S9, Canada.

symmetric scaling stems from the fact that the space for row variables and the space for column variables are generally not the same, making it difficult to justify plotting them in the same space. The problem of one-mode analysis lies in the fact that the main interest of analyzing a two-way table is in clarifying the relations between row variables and column variables, thus ignoring their relation in one-mode analysis defeats the purpose of analysis. If we look at many publications of factor analysis, we often see that analysis of a respondents-by-anxiety table, for example, results typically in the multidimensional configuration of anxiety items, rarely in the multidimensional configuration of respondents, and almost never in the relation between these two multidimensional configurations. Contrary to the typical practice of one mode analysis, the last rarest case of both configurations of anxiety items and subjects should be the focal point of factor analysis. The problems of symmetric scaling and one-mode analysis are closely related to each other, and can be resolved together as described below.

Let us use a small numerical example. Table 1 shows the data from Nishisato (1980) in which three teachers were evaluated in their teaching performance by twenty-nine students.

Table 1: Evaluation of Three Teachers by Twenty-nine Students

Teacher	Good	Aver.	Poor	Sum
A	1	3	6	10
B	3	5	2	10
C	6	3	0	9
Sum	10	11	8	29

The traditional mathematical decomposition can be described as analysis of joint frequencies of the contingency table into the following bilinear form:

$$f_{ij} = \frac{f_{i.}f_{.j}}{f_t}[1 + \rho_1 y_{i1} x_{j1} + \rho_2 y_{i2} x_{j2} + \ldots + \rho_K y_{iK} x_{jK}], \tag{1}$$

where $f_{i.}, f_{.j}$ are respectively marginals of row i and column j, ρ_k is the k-th singular value and K is typically the smaller of n and m of the $n \times m$ contingency table minus 1.

The above decomposition satisfies the dual relations:

$$y_{jk} = \frac{1}{\rho_k}\frac{\sum_{j=1}^{m} f_{ij} x_{jk}}{f_{i.}} \qquad \text{and} \qquad x_{ik} = \frac{1}{\rho_k}\frac{\sum_{i=1}^{n} f_{ij} y_{ik}}{f_{.j}} \tag{2}$$

y_{jk} and x_{ik} are called normed weights (Nishisato, 1980) or standard coordinates (Greenacre, 1984). When we multiply both sides by the singular value, we obtain the means of the weighted responses. In other words, they are the projection of one row variable onto the space of the column variable and the projection of one column variable onto the space of the row variables. Therefore, $\rho_k y_{jk}$ and $\rho_k x_{ik}$ are referred to as projected weights (Nishisato, 1980) or principal coordinates (Greenacre, 1984). As will be noted later, the Young-Householder theorem (Young and Householder, 1938) assures us that those projected weights are projections of points on orthogonal axes in the Euclidean space, and as such they can be treated as coordinates from which Euclidean distance can be calculated.

Using these formulas, the data in Table 1 can be decomposed into those in Table 2.

Table 2: Decomposition of Table 1

row		column		row		column	
Normed weight							
A	1.2320	Good	−1.0760	A	0.6182	Good	0.8615
B	−0.1228	Aver.	−0.0922	B	−1.3729	Aver.	−1.2759
C	−1.2325	Poor	1.4717	C	0.8386	Poor	0.6775
Projected weight							
A	0.7478	Good	−0.6531	A	0.1099	Good	0.1531
B	−0.0745	Aver.	−0.0560	B	−0.2440	Aver.	−0.2267
C	−0.7481	Poor	0.8933	C	0.1490	Poor	0.1204
$\rho_1 = 0.6070$				$\rho_2 = 0.1777$			

It is well known in quantification research that variate y and variate x do not span the same space. Thus, although the distance comparison between two y*'s* or between two x*'s* is appropriate, the comparison between y and x cannot be done in any justifiable way unless the singular value is 1. We describe this problem by saying that the within-set distances can be compared, but that the between-set distances cannot. However, we can always use the same space by projecting the column variables onto the row space or the row variables onto the column space. In other words, the set $(y_{jk}, \rho_k x_{ik})$ spans the same space, so does the set $(\rho_k y_{jk}, x_{ik})$. Plotting these quantities in the same space is referred to as non-symmetric scaling, and this is a geometrically correct plot. From the practical point of view, however, the norm of the projected vector is often much smaller than that of the normed vector, making the row-column comparison difficult when one is projected onto the other. From the logical point of view, it does not make sense to plot the normed weights since they do not contain the information associated with singular values. In consequence, most researchers have opted for the so-called symmetric scaling or French plot of graphing projected row variables and projected column variables as if they were in the same space. Researchers Lebart et al. (1984) have cautioned investigators saying that under symmetric scaling the distance between a row variable and a column variable in the joint graph is not accurate and thus cannot be compared in any rigorous way.

According to Nishisato and Clavel (2003), the angle of the discrepancy between row space and column space for solution k, θ_k, is given by

$$\theta_k = \cos^{-1} \rho_k \tag{3}$$

In our numerical example, the two singular values are $\rho_1 = 0.6070$ and $\rho_2 = 0.1777$, thus the angles of discrepancies are $\theta_1 = 52.6$ degrees and $\theta_2 = 79.8$ degrees. These

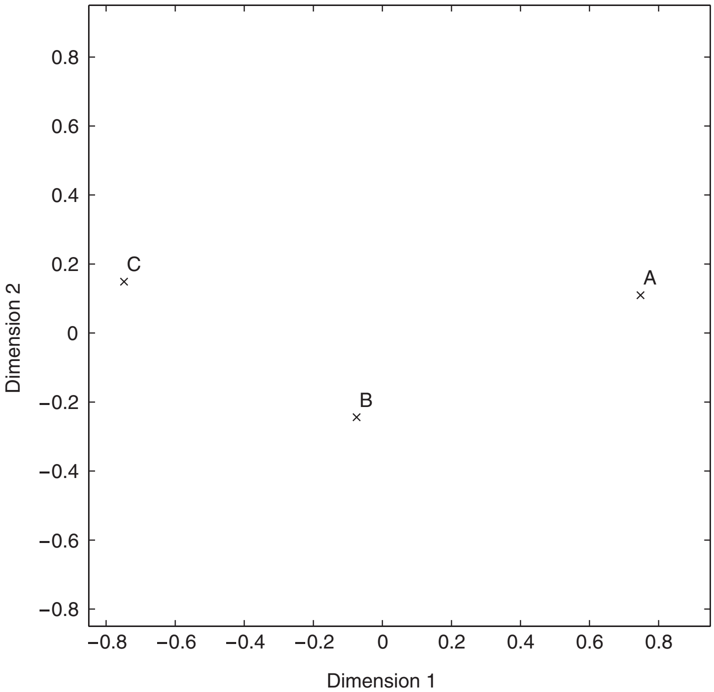

Figure 1: Plot of Teachers: Within-Row Relations

are surprisingly large and are definitely non-negligible space discrepancies, yet under symmetric scaling we treat them as if they were zero's.

Using the projected weights in Table 2, let us plot two-dimensional graphs of rows (Figure 1) and columns (Figure 2). Let us note that these two figures have different axis orientations. Recall that the first row principal axis and the first column principal axis cross at the origin with the angle of 52.6 degrees and the corresponding second principal axes with the angle of 79.8 degrees. In spite of these discrepancies, the currently most popular practice is symmetric scaling (Figure 3), which is nothing but an overlay of Figure 2 onto Figure 1. When we look at this example of symmetric scaling, we wonder why this has been a standard procedure for the joint graphical display. It does not make sense to overlay one configuration in one space over another in different space even if it provides an easily interpretable graph. Since dual scaling determines the coordinates of row and column variables so as to maximize singular values, we now see that one of its characteristics is to minimize the amount of space discrepancy in each dimension. Figure 4 is a geometrically correct non-symmetric joint display, in which variables in row space are projected onto column space. It is easy to see that this geometrically correct joint display does not show an association between teachers and categories as clearly as seen in symmetric scaling. The fact remain, however, that symmetric scaling is not an accurate representation of the relation between rows and columns of the data. Symmetric scaling makes the joint configuration easier to interpret since it places between-set variables closer than they actually are. However,

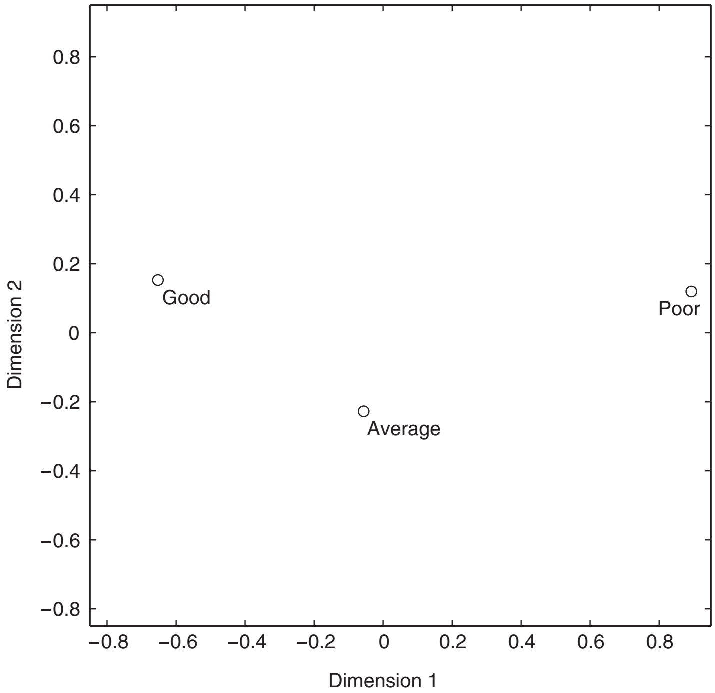

Figure 2: Plot of Categories: Within-Column Relations

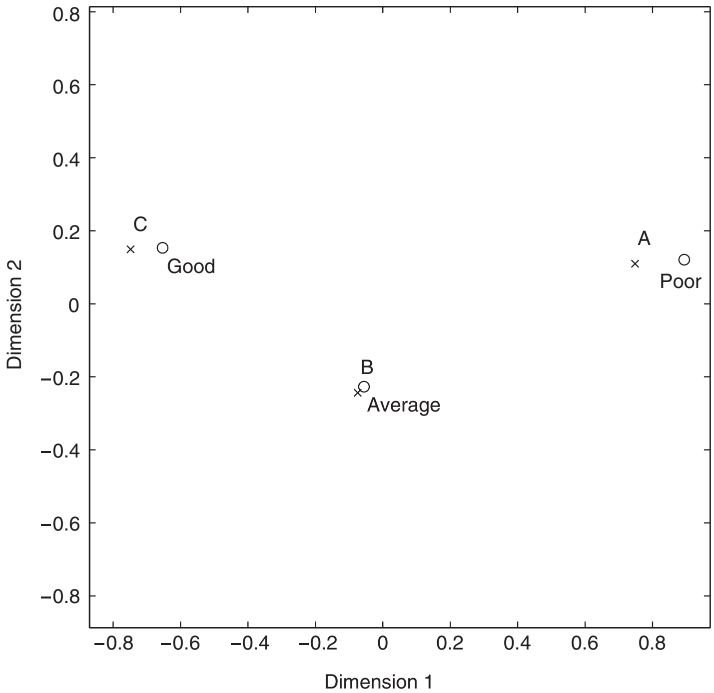

Figure 3: Joint Plot of Teachers and Categories: Symmetric Scaling

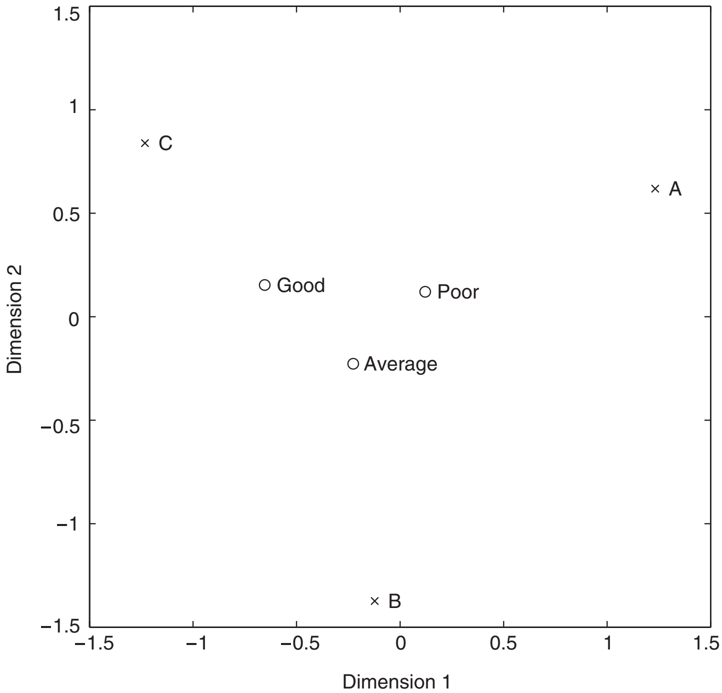

Figure 4: Geometrically Correct Non-Symmetric Plot

it is not a logically correct procedure, hence should be avoided.

Once we consider a different amount of discrepancy between row-space and column-space from dimension to dimension, we realize that we need four-dimensional space to accommodate the accurate coordinates of both row variables and column variables in Table 1, that is, two dimensions for each solution (component). For a comprehensive description of the joint data structure, it is clear that we need both within-set and between-set distances. Since the angle of space discrepancy in each solution is known, it is possible to calculate the between-set distances.

One can always question why a 3×3 matrix cannot be perfectly and completely graphed in two-dimensional space. It is true that the entire information is embedded in the 3×3 contingency table, that is, the information about the coordinates of row variables, those of column variables and singular values associated with the angles of space discrepancies. However, the key point is that the original representation of the data in the 3×3 contingency table is not directly amenable to the extraction of those coordinates in common space, but it must be transformed into the 6×6 table of both within-set and between-set distances, which yields a four-dimensional configuration of both row and column variables in common space. Thus, our proposal is to transform the original contingency data into the form amenable to identifying coordinates of both row variables and column variables in common space. In other words, our proposal is to overcome the problem of one-mode analysis and aim to attain two-mode analysis in a straightforward way.

1.2 Dimension Reduction

To describe data in the simplest form has been the traditionally accepted approach to data analysis. This approach has been a major part of the scientific principle, called the principle of parsimony, and it has taken such different forms as the concepts of least-squares, simple structure, the varimax rotation and data reduction. It is a time-honored approach that no one would normally question about its appropriateness. It is a practical backbone of the traditional inferential statistical approach to data analysis.

In the current paper, however, we would like to take a conservative way of dealing with data in total space, rather than in reduced space, because we typically have no information about how data were generated. This leads us to the belief that data in hand are all we can count on. Although it may not be obvious, quantification theory has a historical association with this conservative approach: when data are not quantitative, the best quantification strategy is to find their transforms which provide linear regression on the data. Many researchers in quantification theory have shown that such transforms provide a maximal information one can extract from the data. Therefore, we would like to adopt this starting stance of quantification theory that the entire information in the data is a basis for quantification. This approach is typically taken also in exploratory data analysis.

In addition to this basic premise of quantification, we have another reason why dealing with the total data space may benefit us. Suppose n standardized variables are distributed in 3-dimensional space. Then all the points lie on the surface of a three-dimensional sphere, each at a distance of one from the origin. Suppose we project those points in two-dimensional space, that is, we reduce the dimensionality of space by one. Then all the points are either on the circle of radius one or inside the circle, and none of the points are outside this circle. We can see with some imagination that the distance between two points in the original sphere are now generally closer to each other in two-dimensional space.

Nishisato and Clavel (2008) discussed that two data points lying in high-dimensional space look closer to each other than they actually are when they are projected onto low-dimensional space. This is so since the distance in K-dimensional space is the square root of the sum of the K squared quantities and $\Sigma_{k=1}^{K}(x_{ik}-x_{jk})^2 \geq \Sigma_{k=1}^{K-1}(x_{ik}-x_{jk})^2$. Yet it is a common practice in multidimensional data analysis (e.g., factor analysis, multidimensional scaling) that researchers look at and interpret data structure in reduced space, typically in two-dimensional space. In the current paper, we would like to remedy not only this possible distortion of the distance structure, but also the possibility that some obscure phenomena may be totally missed in the process of dimension reduction. For instance, in the behavioral sciences some rare phenomena such as behavioral correlates of autism and anorexia are of extreme importance to researchers, and we should therefore avoid the possibility of missing them in the process of dimension reduction. Under the circumstances, the current study takes the view that total space, rather than reduced space, be looked into to understand the data structure. This choice is likely to necessitate some mode of summarization of a complex outcome at the expense of information loss at the final stage of analysis.

2. Proposed Procedure

We have looked at two potential problems with the traditional procedure of quantification, one on symmetric scaling and the other on interpretations of results in reduced space. Based on preliminary work (Clavel and Nishisato, 2008; Nishisato and Clavel, 2003, 2008), we can now advance the idea of analyzing both within-set distances and between-set distances in total space. The first step is to carry out dual scaling of contingency table $\mathbf{F}$, that is, to decompose data matrix $\mathbf{F}$ as follows:

$$\mathbf{F} = \frac{1}{f_t}\mathbf{D}_n\mathbf{Y}\mathbf{\Phi}\mathbf{X}'\mathbf{D}_m \tag{4}$$

where $\mathbf{D}_n$ and $\mathbf{D}_m$ are diagonal matrices of row marginals and column marginals of $\mathbf{F}$, respectively, $\mathbf{Y}$ is the $n \times K$ matrix of eigenvectors associated with the rows, $\mathbf{X}$ is the $m \times K$ matrix of eigenvectors associated with the columns, and $\mathbf{\Phi}$ is the $K \times K$ diagonal matrix of singular values. The solution corresponding to the singular value of 1 is called the trivial solution and is discarded from analysis.

The second step is to calculate within-set and between-set distances. Since the discrepancy between row space and column space depends on each dimension (solution, component), the Nishisato-Clavel distance (Clavel and Nishisato, 2008; Nishisato and Clavel, 2003) is defined in terms of elements of the bilinear decomposition. The theoretical basis of this distance measure is the Young-Householder theorem (Young and Householder, 1938), which states that if the matrix $\mathbf{B}$ is positive definite and is factored as $\mathbf{A}\mathbf{A}'$ then the elements of $\mathbf{A}$ are the projections of points on real Euclidean space of dimensionality equal to the rank of $\mathbf{A}$. In quantification theory, the matrix $\mathbf{A}$ is the matrix of projected weights and the matrix $\mathbf{B}$ is

$$\mathbf{B} = \mathbf{D}_n{}^{-\frac{1}{2}}\mathbf{F}'\mathbf{F}\mathbf{D}_n{}^{-\frac{1}{2}} \text{ or } \mathbf{B} = \mathbf{D}_m{}^{-\frac{1}{2}}\mathbf{F}\mathbf{F}'\mathbf{D}_m{}^{-\frac{1}{2}} \tag{5}$$

From the above discussion, we now know that projected weights are coordinates in Euclidean space and that one can derive a Euclidean distance measure from the projected weights. The Nishisato-Clavel within-set distance (e.g., the distance between row i and row i') is given by

$$d_{ii'} = \sqrt{\sum_{k=1}^{K} \rho_k^2 \left(y_{ik} - y_{i'k}\right)^2} \tag{6}$$

The within-set distances are identical to the Euclidean distances, calculated for the rows of the n-by-K matrix $\mathbf{Y\Phi}$ and computed for the rows of the m-by-K matrix $\mathbf{X\Phi}$.

The Nishisato-Clavel between-set distance (i.e., the distance between row i and column j) is given by

Table 3: Distance Matrix Based on Two Solutions for Table 1: TIA

0	0.8952	1.4965	1.2682	0.8280	0.7537
0.8952	0	0.7798	0.6850	0.3080	0.9843
1.4965	0.7798	0	0.6562	0.7731	1.4830
1.2682	0.6850	0.6562	0	0.7077	1.5468
0.8280	0.3080	0.7731	0.7077	0	1.0108
0.7537	0.9843	1.4830	1.5468	1.0108	0

Table 4: Distance Matrix Based on Solution 1 for Table 1

0	0.8224	1.4960	1.2565	0.7831	0.7391
0.8224	0	0.6736	0.6108	0.0602	0.9404
1.4960	0.6736	0	0.6270	0.7155	1.4728
1.2565	0.6108	0.6270	0	0.5972	1.5465
0.7831	0.0602	0.7155	0.5972	0	0.9493
0.7391	0.9404	1.4728	1.5465	0.9493	0

$$d_{ij} = \sqrt{\sum_{k=1}^{K} \rho_k^2 \left(y_{ik}^2 + x_{jk}^2 - 2\rho_k y_{ik} x_{jk} \right)} \tag{7}$$

Notice that the above distance is calculated by taking into consideration the space discrepancy in the form of the cosine law. Let us define super-distance matrix $\mathbf{D}^*$, consisting of within-row ($\mathbf{D}_{\mathbf{r}}^*$), within-column ($\mathbf{D}_{\mathbf{c}}^*$), between-row-column ($\mathbf{D}_{\mathbf{rc}}^*$) and between-column-row ($\mathbf{D}_{\mathbf{cr}}^*$) distance matrices arranged as

$$\mathbf{D}^* = \begin{bmatrix} \mathbf{D}_{\mathbf{r}}^* & \mathbf{D}_{\mathbf{rc}}^* \\ \mathbf{D}_{\mathbf{cr}}^* & \mathbf{D}_{\mathbf{c}}^* \end{bmatrix} \tag{8}$$

For our example, this matrix can be expressed as:

$$\begin{bmatrix} \mathbf{D}^*_{\text{teachers}} & \mathbf{D}^*_{\text{teachers-categories}} \\ \mathbf{D}^*_{\text{categories-teachers}} & \mathbf{D}^*_{\text{categories}} \end{bmatrix}$$

In terms of the formulas, typical elements of this matrix $\mathbf{D}^*$ are

$$\begin{bmatrix} \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik} - y_{i'k})^2} & \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik}^2 + x_{jk}^2 - 2\rho_k y_{ik} x_{jk})} \\ \sqrt{\sum_{k=1}^{K} \rho_k^2 (x_{jk}^2 + y_{ik}^2 - 2\rho_k x_{jk} y_{ki})} & \sqrt{\sum_{k=1}^{K} \rho_k^2 (x_{jk} - x_{j'k})^2} \end{bmatrix} \tag{9}$$

In our example, the super-distance matrix, based on all solutions (i.e., K=2), can be calculated as in Table 3. If we calculate a super-distance matrix, using the coordinates associated with only the first solution (i.e., K=1), we obtain the distance matrix as in Table 4.

Notice the cell-wise differences between Table 3 and Table 4. Relatively small differences are due to the fact that the first solution accounts for 92 percent of the variance and the second one for only 8 percent. For a larger contingency table, the cell-wise differences between distances in reduced space and those in total space would

Table 5: Euclidean Distance Matrix Without Space Discrepancy for Table 1

0	0.8952	1.4965	1.4016	0.8714	0.1459
0.8952	0	0.7798	0.7017	0.0253	1.0342
1.4965	0.7798	0	0.0951	0.7876	1.6417
1.4016	0.7017	0.0951	0	0.7077	1.5468
0.8714	0.0253	0.7876	0.7077	0	1.0108
0.1459	1.0342	1.6417	1.5468	1.0108	0

Table 6: Eigenvalues of Cross-Product Matrices of Tables 3 and 5

Table 3	Table 5
1.8897	2.3503
0.4592	0.1825
0.1074	0
0.0749	0
0	0
0	0
2.5312	2.5328

be much larger than those in Tables 3 and 4. This observation will make us realize how different information we are looking at by examining the results in reduced space from the one based on all solutions, that is, the given data. Interpreting data in reduced space may therefore lead to an over-simplified conclusion on the structure of the data. To avoid looking at shortened distances in reduced space, our proposal is to use the distance matrix based on all solutions, that is, the distance matrix evaluated in total space. One may wonder that if we use all the solutions, thus the entire information from the data, why we should define the distance measure in terms of the elements of the bilinear decomposition, rather than the input matrix itself. Our justification is that the space discrepancy exists for each dimension and there is no simple way to derive the same distance measure that accounts for space discrepancy directly from the input matrix.

Let us calculate a super-distance matrix on the assumption of no space discrepancy. In other words, calculate the Euclidean distance matrix for the columns of matrix $[\mathbf{\Phi Y}', \mathbf{\Phi X}']$, that is,

$$\mathbf{D}_{\mathbf{0}}^{*} = \begin{bmatrix} \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik} - y_{i'k})^2} & \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik} - x_{jk})^2} \\ \sqrt{\sum_{k=1}^{K} \rho_k^2 (x_{jk} - y_{ik})^2} & \sqrt{\sum_{k=1}^{K} \rho_k^2 (x_{jk} - x_{j'k})^2} \end{bmatrix} \tag{10}$$

We consider that this distance matrix is what we expect from symmetric scaling, provided that we extend the analysis to both within-set and between-set distances. Distance matrix $\mathbf{D}_{\mathbf{0}}^{*}$ for our example is as given in Table 5. We can see that corresponding within-set distances in Tables 3 and 5 are identical and between-set distances different. An important difference between the two tables is the fact that they require different number of dimensions, namely four dimensions for Table 3 (TIA) and two dimensions for Table 5 (symmetric scaling). The eigenvalues of the cross-product matrices of Tables 3 and 5 are listed in Table 6.

Table 7: Clusters and Inter-Cluster Average Distances by K-mean Clustering of Table 3

Cluster	(A, Poor)	(B, Average)	(C, Good)
(A, Poor)	0	0.9295	1.4486
(B, Average)	0.9295	0	0.7364
(C, Good)	1.4486	0.7364	0

Table 8: Total Distance Matrix, Rearranged by Clusters

Cluster	A	Poor	B	Average	C	Good
A	0	0.7537	0.8952	0.8280	1.4965	1.2682
Poor	0.7537	0	0.9843	1.0108	1.4830	1.5468
B	0.8952	0.9843	0	0.3080	0.7798	0.6850
Average	0.8280	1.0108	0.3080	0	0.7731	0.7077
C	1.4965	1.4830	0.7798	0.7731	0	0.6562
Good	1.2682	1.5468	0.6850	0.7077	0.6562	0

The third step is to analyze the super-distance matrix. There are at least two ways to do the analysis, one to subject it to multidimensional scaling (MDS) and the other to carry out cluster analysis. Out of these two approaches, our proposal is to use cluster analysis since MDS is likely to lead us to the same problem of graphing the results in reduced space. Recall that the analysis of the super-distance matrix $\mathbf{D}^*$ now requires more dimensions than the traditional within-set analysis and if we were to maintain 'total information analysis', how can we interpret all MDS components without resorting to the traditional choice of only a few components?

In contrast, cluster analysis is unlikely to force us to deal with the reduced space strategy and offers us an ideal tool to look at clusters of row and column variables in multidimensional space. Here the main contrast of the two approaches is between looking at clusters on individual axes typically in reduced space by MDS and clusters in total space by cluster analysis. This contrast may be termed as that of "axis interpretation" (i.e., interpretation of coordinates of individual dimensions) versus "cluster interpretation" (i.e., interpretation based on closeness of variables in multidimensional space) of data structure. In cluster analysis, we do not wish to impose any constraints on the cluster structure, and therefore we propose the use of partitioning rather than hierarchical clustering.

Partitioning of variables in Table 3 by the k-means clustering yields the results of three clusters as summarized in Table 7. Up to this point, we have maintained the strategy of dealing with the entire information in data. This strategy, however, is suddenly abandoned here and is replaced with summarization of complex relations. This loss of information helps us to interpret an overall picture of data clouds. When the data set is as small as the current example, however, we can always present the original distance matrix in which the variables are re-arranged by clusters (Table 8). In this table, we can see the entire picture in a straightforward way. The proposed approach presents a stark contrast to symmetric scaling which involves the perennial space discrepancy problem and to nonsymmetric scaling which has the problem of different norms. Also notice one fundamental difference between MDS and cluster analysis approaches: Cluster analysis can identify any number of clusters, for example, in two-dimensional space, while MDS interprets the same data with two components associated with the dimensions.

Earlier we saw even tighter clusters of teachers and categories in symmetric scaling than the k-means clustering results, but symmetric scaling is theoretically wrong and does not offer any justifiable description of data. We also saw widely separated elements within a cluster in non-symmetric scaling, which does not help us see the data structure and therefore is of little use. Our conclusion is that we may now abandon and forget about these problematic approaches and that we may look at an alternative as proposed in the current paper.

When the data matrix is large, the method of k-means clustering would not present much difficulty in finding mutually exclusive clusters, but we will need some reasonable means of determining the number of clusters. This will be left as one of the future problems for the current approach. With an extremely large data set, it may even be necessary, as a second step of clustering, to subject a large inter-cluster distance matrix to the k-means clustering to further summarize the initial clustering results. This, however, is only a suggestion. In terms of the spirit of our endeavor towards total information analysis, we can always look at the distance matrix of those variables in one cluster at a time as in Table 8, thus without any loss of the input information. The main role of the k-means clustering is therefore to identify groupings of the variables,

In summary, our proposal is to expand the input contingency table of frequencies to the super-distance matrix, obtained from all the components of the original data matrix, and then to subject it to cluster analysis to identify groupings of row and column variables in total common space, hence the name total information analysis, abbreviated as TIA. In this approach, all the variables in rows and columns of the data matrix contribute to the final summary of the results. It should be noted, however, that we will lose some information at the final stage of summarizing distance relations by clustering.

3. Two Examples

3.1 Example 1

Garmize and Rycklak (1964) postulated particular relations between Rorschach responses and moods of the subjects, and collected a set of selected Rorschach responses under six types of experimentally induced moods, fear, anger, depression, ambition, security and love. Table 9 shows the joint frequencies of responses. Since the table is 11×6, the total number of non-trivial solutions is five. Singular values, the percentage of information accounted for by the component (δ), the angle of space discrepancy (θ) and normed weights are listed in Table 10. According to Nishisato (1994), traditional dual scaling, using all five solutions, yields associations of Rorschach responses with moods as shown in Table 11.

Table 9: Rorschach Responses and Induced Moods

Rorschach Responses	Induced Moods Fear	Anger	Depression	Ambition	Security	Love
Bat	33	10	18	1	2	6
Blood	10	5	2	1	0	0
Butterfly	0	2	1	26	5	18
Cave	7	0	13	1	4	2
Clouds	2	9	30	4	1	6
Fire	5	9	1	2	1	1
Fur	0	3	4	5	5	21
Mask	3	2	6	2	2	3
Mountains	2	1	4	1	18	2
Rocks	0	4	2	1	2	2
Smoke	1	6	1	0	1	0

Table 10: Dual Scaling Statistics and Normed Weights

Solution	1	2	3	4	5
Singular value	0.6807	0.5005	0.4128	0.3579	0.2686
δ (percent)	42.72	23.10	15.71	11.81	6.65
θ (degree)	47.10	59.97	65.62	69.03	74.42
Bat	−1.0345	−0.3255	0.3697	−0.9512	−0.3141
Blood	−1.2718	−0.7061	1.4514	−0.4929	0.2388
Butterfly	1.7122	−0.8759	0.4172	−0.4260	1.1794
Cave	−0.5493	0.5978	−1.0713	−0.9629	0.4187
Clouds	−0.3432	−0.1541	−1.8061	0.8272	0.4837
Fire	−0.6179	−0.5961	1.5231	1.6560	0.2500
Fur	1.1518	−0.1540	−0.2047	0.0346	−2.5014
Mask	−0.0757	0.0815	−0.5139	−0.1166	0.0927
Mountains	0.4978	3.0825	0.7608	−0.1074	0.4138
Rocks	0.1674	0.2898	0.3218	1.9600	−0.2914
Smoke	−0.8325	−0.0963	1.3773	3.4843	0.0175
Fear	−1.2719	−0.3360	0.9451	−1.3396	−0.1009
Anger	−0.6468	−0.4672	0.8509	2.0834	−0.0694
Depression	-0.5709	0.1728	−1.6384	0.0625	0.3563
Ambition	1.5134	−1.0105	0.3551	−0.3219	1.7976
Security	0.6236	2.5328	0.7044	0.0007	0.2034
Love	1.1110	−0.4681	−0.2147	−0.2107	−1.7501

Table 11: Associations Between Rorschach Responses and Moods by Traditional Dual Scaling

Rorschach Responses	Moods
Bat, Cave	Fear
Butterfly	Ambition
Clouds, Cave	Depression
Fur	Love
Mountains	Security
Smoke, Rock, Fire	Anger

Table 12: Distances Based on Two Solutions (Upper Triangle) and in Total Space (Lower Triangle)

0	0.25	1.89	0.57	0.48	0.31	1.49	0.68	2.00	0.87	0.18	0.67	0.56	0.57	1.66	1.71	1.36
0.56	0	2.03	0.82	0.69	0.45	1.67	0.90	2.25	1.10	0.43	0.76	0.72	0.78	1.80	1.90	1.52
1.94	2.09	0	1.71	1.45	1.59	0.53	1.31	2.15	1.20	1.78	1.90	1.55	1.54	1.01	1.79	0.94
0.85	1.34	1.84	0	0.40	0.60	1.22	0.41	1.43	0.51	0.40	0.79	0.57	0.41	1.49	1.36	1.15
1.22	1.58	1.78	0.81	0	0.29	1.02	0.22	1.72	0.41	0.33	0.74	0.39	0.32	1.29	1.44	0.95
1.10	0.89	1.83	1.54	1.44	0	1.22	0.50	1.99	0.69	0.29	0.71	0.44	0.48	1.42	1.64	1.12
1.66	1.96	1.16	1.53	1.48	1.70	0	0.84	1.68	0.71	1.35	1.52	1.15	1.10	0.89	1.43	0.65
0.84	1.22	1.40	0.57	0.68	1.17	1.10	0	1.55	0.20	0.52	0.85	0.48	0.36	1.19	1.33	0.83
2.04	2.27	2.16	1.65	2.05	2.11	1.90	1.64	0	1.42	1.83	1.98	1.82	1.64	2.03	1.46	1.77
1.36	1.49	1.53	1.31	1.07	0.87	1.17	0.85	1.62	0	0.71	0.99	0.62	0.49	1.12	1.25	0.76
1.65	1.49	2.32	1.93	1.66	0.72	2.06	1.60	2.25	1.00	0	0.65	0.47	0.43	1.55	1.58	1.23
0.90	1.05	2.02	1.16	1.39	1.28	1.78	1.10	2.08	1.44	1.71	0	0.43	0.54	1.93	1.93	1.62
1.13	1.23	1.81	1.28	1.25	1.06	1.57	1.02	2.01	1.08	1.40	1.30	0	0.32	1.50	1.73	1.20
1.02	1.34	1.75	0.84	0.90	1.34	1.45	0.73	1.86	1.14	1.68	1.30	1.30	0	1.54	1.43	1.19
1.77	1.95	1.15	1.68	1.64	1.73	1.31	1.32	2.10	1.45	2.14	2.04	1.81	1.79	0	1.87	0.39
1.77	1.99	1.85	1.54	1.73	1.83	1.62	1.40	1.50	1.46	2.08	1.99	1.89	1.73	1.93	0	1.54
1.48	1.73	1.16	1.36	1.34	1.53	0.97	0.98	1.88	1.16	1.94	1.80	1.58	1.45	1.06	1.67	0

The first two solutions account for 65.82 percent of the total information. Some researchers would plot a two-dimensional graph and would try to interpret the graph, which shows three clusters [Mountains; Security], [Butterfly, Fur; Ambition, Love] and [Bat, Blood, Cave, Clouds, Fire, Mask, Rocks, Smoke; Fear, Anger, Depression]. This is an example of interpretation of data in reduced space, and we see larger clusters than those in five-dimensional space. To see possible consequences of looking at data in reduced dimension, let us calculate two super-distance matrices, one based on the first two solutions and the other in total five-dimensional space. In Table 12, the distances based on two solutions are listed in the upper triangle section (i.e., above the main diagonal elements of 0's) and those based on five solutions (total space) in the lower triangle section. Large differences between the corresponding elements in the table suggest potential and substantial differences between the analysis in reduced space and that in total space.

As was done by traditional dual scaling (Table 11), the k-means clustering was used to form six clusters of associations between Rorschach responses and moods (Table 13). Both results are based on all five solutions. Let us now compare the results in the two tables. In traditional dual scaling, Cave appears twice, Mask is missing, and Fear is associated with Bat and Cave, rather than Bat and Blood in TIA, and Depression is associated with Cave and Clouds, rather than Cave, Clouds and Mask in TIA. Although this view may be biased, TIA results look more reasonable than those of dual scaling, and this view can be reinforced when we look at the clustering results from TIA (Table 14). We should note that the clustering was done to arrive at mutually exclusive clusters, while this was not the case with dual scaling. Another difference between the two approaches is that the partitioning process is objective and mechanical, and that dual scaling results need to be judged in terms of reasonable weights to be included in the association scheme. To save the space, the rearranged distance matrix as in Table 8 is not presented for the current example.

Table 13: Associations Between Rorschach Responses and Moods by K-Means Clustering

Cluster	Rorschach Responses	Moods
A	Bat, Blood	Fear
B	Butterfly	Ambition
C	Fur	Love
D	Cave, Clouds, Mask	Depression
E	Mountains	Security
F	Fire, Rock, Smoke	Anger

Table 14: Inter-Cluster Centroid Distances

Cluster	A	B	C	D	E	F
A	0	1.818	1.597	1.005	1.818	1.090
B	1.818	0	0.904	1.485	1.770	1.635
C	1.597	0.904	0	1.169	1.530	1.404
D	1.005	1.485	1.169	0	1.464	1.125
E	1.818	1.770	1.530	1.464	0	1.651
F	1.090	1.635	1.404	1.125	1.651	0

3.2 Example 2

Kretschmer was one of the most eminent psychiatrists during the first half of the 20th century. In psychology, his body-mind typology is well known. Table 15 is borrowed from Kretschmer (1925), where “pyknic” indicates a short stocky physique, “leptosomatic” means a slender, thin and frail body and “dysplastic” refers to abnormal unbalanced development. Dual scaling of this table yields two solutions. The singular values and space discrepancy angles are $\rho_1 = 0.5082$, $\rho_2 = 0.2611$, $\theta_1 = 59.46$ degrees and $\theta_2 = 74.86$ degrees. The joint graph by symmetric scaling is as shown in Figure 5. The results seem to show Kretschmer’s typology very well, namely the schizophrenic mental type is associated with the leptosomatic body type, the manic-depressive mental type with the pyknic body type, the epileptic mental type with the dysplastic body type, and the athletic body type lying between the schizophrenic and the epileptic body types. Since this data set can be explained exhaustively by two solutions, we typically consider that this symmetric graph represents 100 percent of information in the data. From the TIA point of view, however, this data set requires four dimensions, rather than two. The TIA super-distance matrix is shown in Table 16. Since there are only three mental types, three clusters were formed by the k-means clustering. The three clusters and the inter-cluster-centroid distances are shown in Table 17.

Let us compare Figure 5 (symmetric scaling) and Table 17 (TIA). In symmetric scaling, Athletic and Others are relatively close to each other, while in TIA clustering, Athletic belongs to the cluster [Epileptic, Dysplastic] and Others to cluster [Schizophrenic, Leptosomatic]. The cause of this difference is due to looking at the data in two-dimensional space (symmetric scaling) or in four-dimensional space (TIA clustering). Although the rearranged distance matrix by clusters is not presented,

Table 15: Kretschmer's Typology

Mental Type	Pyknic	Leptosomatic	Athletic	Dysplastic	Others	Total
Maniac-Depressive	879	261	91	15	114	1360
Schizophrenic	717	2632	884	549	450	5232
Epileptic	83	378	435	444	166	1506
Total	1679	3271	1410	1008	730	8098

Table 16: Distance Matrix Based on Two Solutions for Table 1: TIA

0	1.2734	1.6225	1.0346	1.2357	1.3093	1.5125	1.1309
1.2734	0	0.7539	1.0606	0.3361	0.4069	0.7207	0.2514
1.6225	0.7539	0	1.3670	0.7650	0.6417	0.7705	0.6650
1.0346	1.0606	1.3670	0	1.1903	1.2952	1.5442	1.0147
1.2357	0.3361	0.7650	1.1903	0	0.5004	0.8350	0.3960
1.3093	0.4069	0.6417	1.2952	0.5004	0	0.3447	0.2925
1.5125	0.7207	0.7705	1.5442	0.8350	0.3447	0	0.6091
1.1309	0.2514	0.6650	1.0147	0.3960	0.2925	0.6091	0

Table 17: Inter-Cluster Distances

Members	Cluster	A	B	C
Maniac-Depressive, Pyknic	A	0	1.014	1.305
Schizophrenic, Leptosomatic, Others	B	1.014	0	0.498
Epileptic, Dysplastic, Athletic	C	1.305	0.498	0

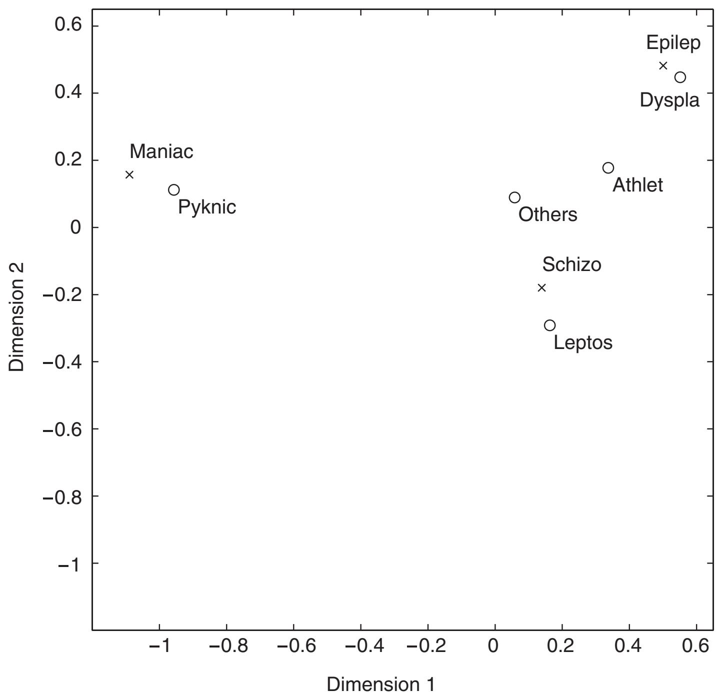

Figure 5: Joint Plot of Kretschmer's data: Symmetric Scaling

this example, too, shows clear TIA clusters, making the results comparatively easy to interpret. We should note that symmetric scaling presents a distorted relation between mental types and body types.

4. Concluding Remarks and Future Problems

As the two numerical examples suggest, an overall description of data structure by the traditional method of symmetric scaling may not be too far off from that of TIA. However, it is difficult to justify the use of symmetric scaling from the theoretical point of view. In contrast, TIA is based on the correct geometric description of the data and offers a sound framework and clear results. TIA as presented here, however, is not complete as yet, for some practical and theoretical problems are left unsolved, for example, how to handle possible influences on the results when the number of categories of one variable is vastly different from that of the other variables (currently these variables are equally scaled), how to handle the dimension of the super-distance matrix when the data matrix is relatively large, say 800×200, how to determine a reasonable number of clusters in the k-means clustering, and how to introduce a better alternative to the current k-means clustering such that discriminative information of variables is retained in groupings or graphics. As we apply TIA to other data sets, some more practical problems are likely to emerge. One possible extension of TIA is to apply it to other types of categorical data. Van Deun et al. (2007) considered analyzing rows and columns of unfolding data simultaneously, which is the same as our approach of analyzing within-set and between-set distances together. Since their approach is quite different from ours, interested readers are referred to their work. Another possible extension is to use other measures of distance such as the Hellinger distance and the Rajski distance. An immediate problem facing this extension is how to define between-set distances for these measures. This is a challenging problem, although interesting and important. This will be left for future research.

As a framework for multidimensional data analysis, TIA is promising since the same framework can be extended to analysis of quantitative data, where typically principal component analysis (PCA) is used. In PCA, too, the same discrepancy between row space and column space exists, yet we carry out most of time only one-mode analysis, typically finding a multidimensional configuration of only variables. As is the case with quantification problems, however, we should explore the relation between the multidimensional configuration of variables and that of subjects. Since we know the amount of space discrepancy in each dimension, we can construct a super matrix of within-set and between-set distances based on all possible components (solutions) as we carried out in the current paper. Thus the extension of TIA to quantitative data will be an immediate step to remedy a long-ignored space discrepancy problem in multivariate analysis. It is hoped that the current paper will serve as encouragement for further developments of both categorical and continuous multivariate data analysis.

Acknowledgements

We would like to thank the Editor and two anonymous reviewers whose insightful comments greatly helped us.

References

Andrews, D. (1972). Plots of high-dimensional data. *Biometrics*, 28:125–136.

Clavel, J. and Nishisato, S. (2008). Joint analysis of within-set and between-set distances. In Shigemasu, K. e. a., editor, *New Trends in Psychometrics*, pages 41–50, Tokyo. Universal Academy Press.

Garmize, L. and Rycklak, J. (1964). Role-play validation of socio-cultural theory of symbolism. *Journal of Consulting Psychology*, (28):107–115.

Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press.

Kretschmer, E. (1925). *Physique and Character*. Kegan, Paul, Trench and Trubner.

Lebart, L., Morineau, A., and Warwick, K. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley, New York.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications*. The University of Toronto Press, Toronto.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Lawrence Erlbaum Associates, Hilsdale, N.J.

Nishisato, S. and Clavel, J. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika*, (30):87–98.

Nishisato, S. and Clavel, J. (2008). Interpreting data in reduced space: A case of what is not what in multidimensional data analysis. In Shigemasu, K. e. a., editor, *New Trends in Psychometrics*, pages 357–366, Tokyo. Universal Academy Press.

Van Deun, K., Heiser, W., and Delbeke, L. (2007). Multidimensional unfolding by nonmetric multidimensional scaling of spearman distances in the extended permutation polytope. *Multivariate Behavioral Research*, 42(1):103–132.

Young, G. and Householder, A. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, (3):19–22.

(Received January 6 2009, Revised November 21 2009)