Author: Yoshio Takane

takane@takane2.psych.mcgill.ca

Keywords: scale levels, data transformations, distance models, nonmetric multidimensional scaling, additive models

Abstract:

A data matrix typically represents some kind of relationship between row and column entities. The relationship represented by the data may be described by a model presumed to have generated the data. Observed data, on the other hand, may be measured on one of a variety of scale levels: nominal, ordinal, interval, or ratio. In such cases we may attempt to do two things simultaneously: 1) we transform the data by a transformation appropriate for the scale level, and 2) we fit a model to the transformed data to account for the data. This process of simultaneous data transformations and data representations is called optimal scaling. In this article we briefly discuss some of the key features of optimal scaling.

Optimal scaling

Suppose that dissimilarity judgments are obtained between a set of stimuli. The dissimilarity between stimuli may be represented by a Euclidean distance model. However, it is rare to find the observed dissimilarity data measured on a ratio scale. It is more likely that the observed dissimilarity data satisfy only the ordinal scale level. That is, they are only monotonically related to the underlying distances. In such cases we may consider transforming the observed data monotonically, while simultaneously representing the transformed dissimilarity data by a distance model. This process of simultaneously transforming the data and representing the transformed data is called optimal scaling [2, 8].

Let δ_{ij} denote the observed dissimilarity between stimuli i and j measured on an ordinal scale. Let d_{ij} represent the underlying Euclidean distance between the two stimuli represented as points in an A-dimensional Euclidean space. Let x_{ia} denote the coordinate of stimulus i on dimension a. Then, d_{ij} can be written as $d_{ij} = \{\sum_{a=1}^{A}(x_{ia} - x_{ja})^2\}^{1/2}$. (We use X to denote the matrix of x_{ia} and sometimes write d_{ij} as $d_{ij}(X)$ to explicitly indicate that d_{ij} is a function of X.) Optimal scaling obtains the best monotonic transformation (m) of the observed dissimilarities (δ_{ij}) and the best representation ($d_{ij}(X)$) of the transformed dissimilarity ($m(\delta_{ij})$) in such a way that the squared discrepancy between them is as small as possible. Define the least squares criterion, $\phi = \sum_{i,j<i}(m(\delta_{ij}) - d_{ij}(X))^2$. We minimize this criterion with respect to m and X under suitable normalization restrictions on m or on X. This is called nonmetric multidimensional scaling [3], which played an important role in the development of ideas of optimal scaling.

We give an example of optimal scaling from nonmetric multidimensional scaling (MDS). Rothkopf [4] reported stimulus confusion data among 36 Morse code signals.

Shepard [5] analyzed his data by nonmetric MDS that allowed a monotonic transformation of the confusion probabilities and a representation of the transformed data in a multidimensional Euclidean space. Figure 1 displays the derived stimulus configuration. From this we may deduce that the process mediating confusions among the signals is two-dimensional; one is the total number of components in the signals, and the other is the mixture rate of two kinds (dots and dashes) of components. Signals having more components tend to be located toward the top, and those having more dots tend to be located toward the left of the configuration. Figure 2 displays the optimal inverse monotonic transformation of the confusion probabilities. The derived optimal transformation looks very much like a negative exponential function, $p_{ij} = a\exp(-d_{ij})$, or possibly a Gaussian, $p_{ij} = a\exp(-d_{ij}^2)$, typically found in stimulus generalization data.

***** Insert Figures 1 and 2 about here *****

In the above example, the data involved are dissimilarity data, for which the distance model may be an appropriate choice. Other kinds of data may also be considered for optimal scaling. For example, the data may reflect the joint effects of two or more underlying factors. In this case, an ANOVA-like additive model may be appropriate. As another example, preference judgments are obtained from a single subject on a set of objects (e.g., cars) characterized by a set of features (size, color, gas efficiency, roominess, etc.) In this case, a regression-like linear model that combines these features to predict the overall preference judgments may be appropriate. As a third example, preference judgments are obtained from a group of subjects on a set of stimuli. In this case, a vector model of preference may be appropriate, in which subjects are represented as vectors and stimuli as points in a multidimensional space, and subjects' preferences are obtained by projecting the stimulus points onto the subject vectors. This leads to a PCA-like (Principal Component Analysis) bilinear model [7]. Alternatively, subjects' ideal stimuli may be represented as (ideal) points, and it may be assumed that subjects' preferences are inversely related to the distances between subjects'ideal points and actual stimulus points. This is called unfolding (or ideal point) model [1].

Any one of the models described above may be combined with various types of data transformations depending on the scale level on which observed data are assumed to be measured. Different levels of measurement scale allow different types of transformations while preserving the essential properties of the information represented by numbers. In psychology, four major scale levels have traditionally been distinguished: nominal, ordinal, interval, and ratio [6]. In the nominal scale level, only the identity of numbers is considered meaningful (i.e., $x = y$ or $x \neq y$). Telephone numbers and gender (male and females coded as 1's and 0's) are examples of this level of measurement. In the nominal scale, any one-to-one transformation is permissible, since it preserves the identity (and non-identity) between numbers. (This is called admissible transformation.) In the ordinal scale level, an ordering property of numbers is also meaningful (i.e., for x and y such that $x \neq y$, either $x > y$ or $x < y$, but how much larger or smaller is not meaningful). An example of this level of measurement is the rank numbers given to participants in a race. In the ordinal scale, any monotonic (or order-preserving) transformation is admissible. In the interval scale level, the difference between two numbers is also meaningful. A difference in temperature measured on an interval scale can be meaningfully interpreted (e.g., the difference between yesterday's temperature and today's is such and such), but because the origin (0 point) in the scale is arbitrary as the temperature measured in Celsius or Fahrenheit, their ratio is not meaningful. In the interval scale, any affine transformation ($x' = ax + b$) is admissible. In the ratio scale level, a ratio between two numbers is also meaningful (e.g., temperature measured on the absolute scale where -273^oC is set as the zero point). In the ratio scale, any similarity transformation ($x' = ax$) is admissible. In optimal scaling, a specific transformation of the observed data is sought within each class of the admissible transformations consistent with the scale level on which observed data are assumed to be measured.

It is assumed that one of these transformations is tacitly applied in a data generation process. For example, if observed dissimilarity data are measured on an ordinal scale, the model prediction, d_{ij}, is assumed error-perturbed and then monotonically transformed to obtain the observed dissimilarity data, δ_{ij}. Optimal scaling reverses this operation by first transforming back δ_{ij} to the error-perturbed distance by m, which is then represented by the distance model, $d_{ij}(X)$.

References

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

[**2**] Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester: Wiley.

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

[**4**] Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired associate learning. *Journal of Experimental Psychology*, **53**, 94-101.

[**5**] Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, **5**, 33-48.

[**6**] Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In Stevens, S. S. (Eds.), *Handbook of experimental psychology* (pp. 1-49). New York: Wiley.

[**7**] Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In Guliksen, H., and Messick, S. (Eds.), *Psychological scaling: Theory and applications* (pp. 155-167). New York: Wiley.

[**8**] Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, **46**, 357-388.

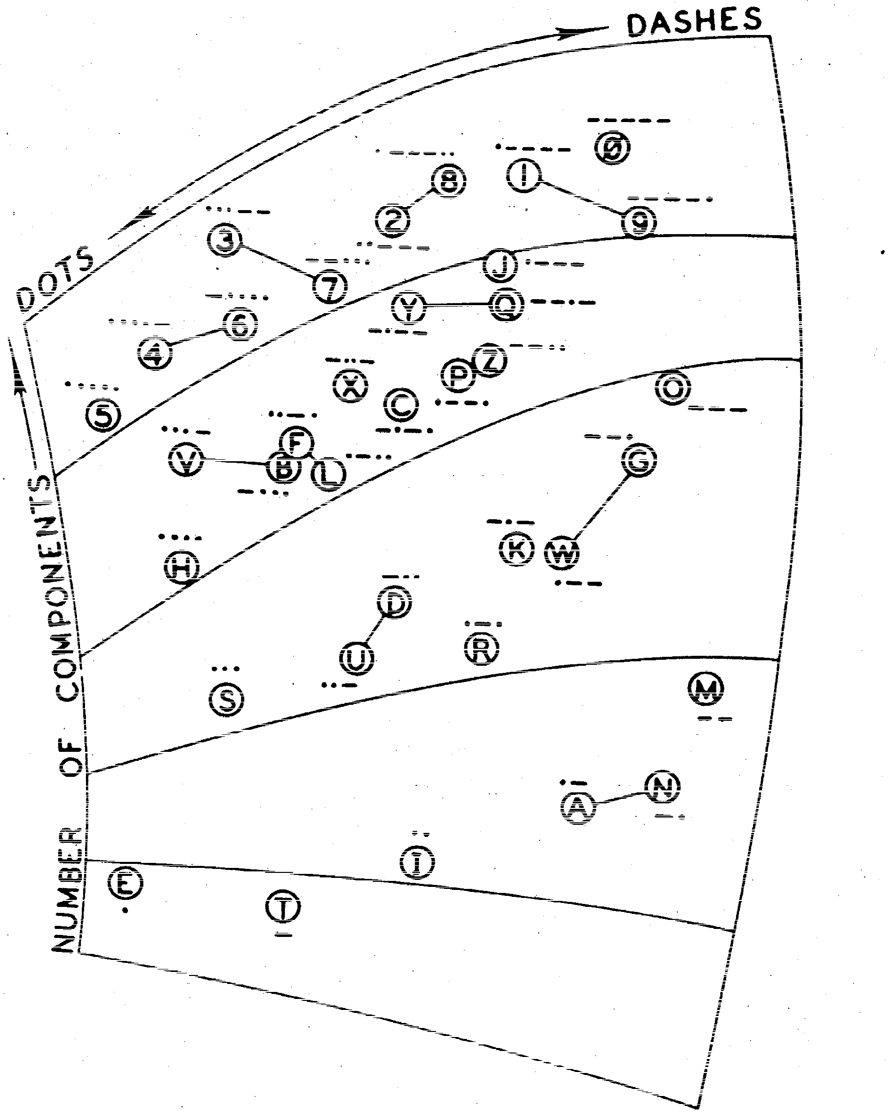

Figure 1: A stimulus configuration obtained by nonmetric multidimensional scaling of confusion data [4] among Morse code signals.

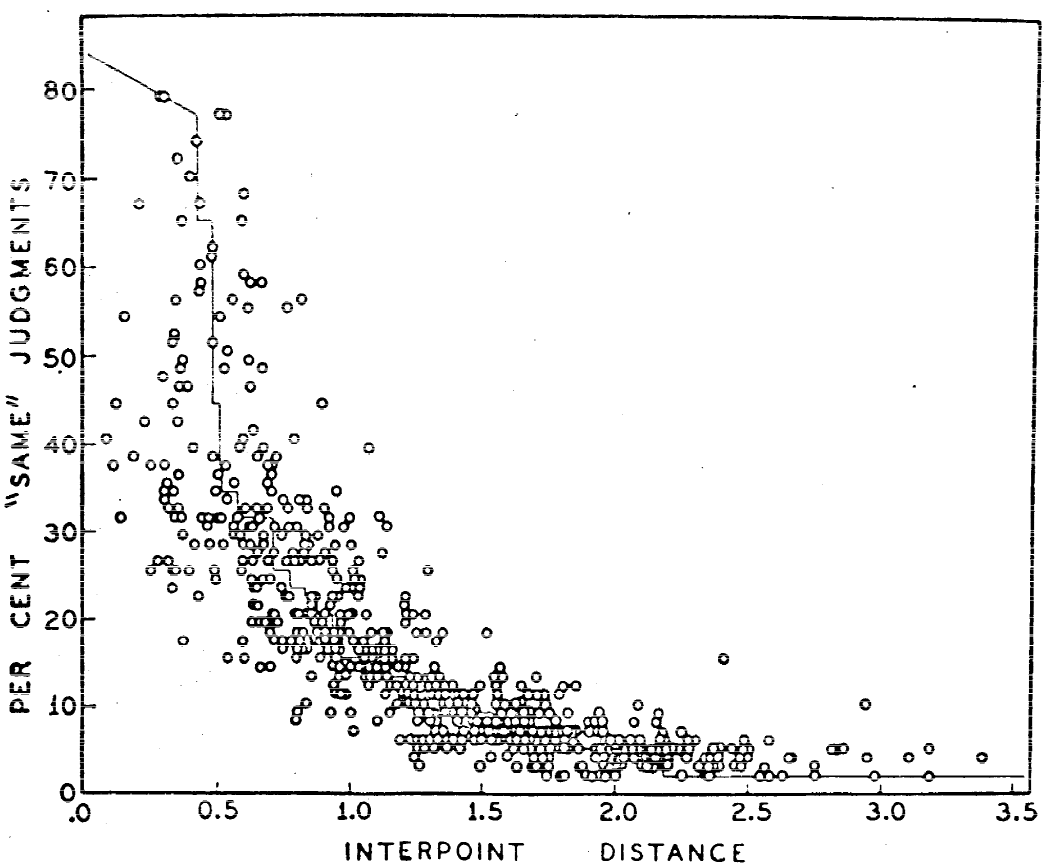

Figure 2: An optimal data transformation of the confusion probabilities.