ENCYCLOPEDIA
OF COGNITIVE SCIENCE
(to appear in 2002)

Multidimensional Scaling

Mark Steyvers

Stanford University

Assumptions of MDS

Multidimensional Scaling (MDS) describes a family of techniques for the analysis of proximity data on a set of stimuli to reveal the hidden structure underlying the data. The proximity data can come from similarity judgments, identification confusion matrices, grouping data, same-different errors or any other measure of pairwise similarity. The main assumption in MDS is that stimuli can be described by values along a set of dimensions that places these stimuli as points in a multidimensional space and that the similarity between stimuli is inversely related to the distances of the corresponding points in the multidimensional space. The Minkowski distance metric provides a general way to specify distance in a multidimensional space:

$$d_{ij} = \left[\sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right|^{r} \right]^{1/r},$$

where n is the number of dimensions, and x_{ik} is the value of dimension k for stimulus i. With r =2, the metric equals the Euclidian distance metric while r=1 leads to the city-block metric. A Euclidian metric is appropriate when the stimuli are composed of integral or perceptually fused dimensions such as the dimensions of brightness and saturation for colours. The city-block metric is appropriate when the stimuli are composed of separable dimensions such as size and brightness (Attneave, 1950). In practice, the Euclidian distance metric is often used because of mathematical convenience in MDS procedures.

MDS can be applied with different purposes. One is exploratory data analysis; by placing objects as points in a low dimensional space, the observed complexity in the original data matrix can often be reduced while preserving the essential information in the data. By a representation of the pattern of proximities in two or three dimensions, researchers can visually study the structure in the data.

It also has been used to discover the mental representation of stimuli that explains how similarity judgments are generated. Sometimes, MDS reveals the psychological dimensions hidden in the data that can meaningfully describe the data. The multidimensional representations resulting from MDS are also often useful as the representational basis for various mathematical models of categorization, identification, and/or recognition memory (Nosofsky, 1992) or generalization (Shepard, 1987).

We will illustrate some of the issues in MDS with the analysis of a face similarity judgement task. In Figure 1a, the average of a group of subjects' similarity ratings is shown for 10 faces shown in Figure 1c. The idea is to reveal some of the perceptual dimensions that subjects might have used when generating similarity judgments for these faces.

Techniques for MDS

There are many different MDS techniques to analyse proximity data and many issues in the analysis and interpretation of the results. First, there is the distinction between metric and nonmetric MDS. The goal of metric MDS is to find a configuration of points in some multidimensional space such that the interpoint distances are related to the experimentally obtained similarities by some transformation function (e.g., a linear transformation function). If the proximity data are generated with Euclidian distances for some stimulus configuration, then a procedure called classical metric MDS (Torgeson, 1965) can exactly recreate the configuration of points. Because a closed form solution exists to find such a configuration of points, classical metric MDS can be performed efficiently on large matrices. In nonmetric MDS (first devised by Shepard in 1962), the goal is to establish a monotonic relationship between interpoint distances and obtained similarities. The advantage of nonmetric MDS is that no assumptions need to be made about the underlying transformation function; the only assumption is that the data is measured at the ordinal level.

Kruskal (1964) proposed a measure for the deviation from monotonicity between the distances d_{ij} and the observed dissimilarities

o_{ij} called the *stress* function:

$$S = \sqrt{\frac{\sum_{ij} \left(d_{ij} - d_{ij}^*\right)^2}{\sum_{ij} d_{ij}^2}}$$

Note that the observed dissimilarities o_{ij} do not appear in this formula. Instead, the discrepancy between the predicted distances d_{ij} and the target distances d_{ij}^* are measured. Based on the current configuration of points, the target distances d_{ij}^* are found by monotonic regression and represent the distances that are monotonically related to the observed dissimilarities o_{ij}. Several iterative minimization algorithms exist to move the object points in a multidimensional space in order to minimize stress (see Borg & Groenen, 1997). In the face similarity example, Figure 1d displays what is known as the Shepard plot. It shows the relationship between predicted distances d_{ij} (for the two-dimensional scaling solution in Figure 1c) and observed dissimilarities as filled circles and can serve to understand what metric transformation would be appropriate to relate one to the other. The line in the plot shows the relationship between the target distances d_{ij}^* found by monotonic regression and observed dissimilarities. Kruskal stress essentially is a measure based on the sum of the squared deviations between the filled circles and the line along the abscissa.

Another distinction in MDS is between weighted MDS, replicated MDS and MDS on a single matrix (Young, & Hamer, 1994). In replicated MDS, several matrices of similarity data can be analysed simultaneously. The matrices are provided by different subjects or by a single subject tested at multiple times and a single scaling solution captures the similarity data of all matrices through separate metric or nonmetric relationships for each matrix. This approach can take individual differences in response bias into account. In weighted replicated MDS (e.g., INDSCAL, Carroll & Chang, 1970), the dimensions in the scaling solution can be weighted differently for each subject or subject replication to model differences in attention or sensitivity for the different dimensions.

Finally, there is the distinction between deterministic and probabilistic MDS. In deterministic MDS, each object is represented as a single point in multidimensional space (e.g., Borg & Groenen, 1997) whereas in probabilistic MDS (MacKay, 1989), each object is represented as a probability distribution in multidimensional space. In understanding the mental representation of objects, this last approach is useful when representation of objects is assumed to be noisy (i.e., the presentation of the same object on every trial gives rise to different internal representations).

An important issue in MDS is choosing the number of dimensions for the scaling solution. A configuration with a high number of dimensions achieves very low stress values but cannot easily be comprehended by the human eye, and is apt to be determined more by noise than by the essential structure in the data. On the other hand, a solution with too few dimensions might not reveal enough of the structure in the data.

A well known method to select the dimensionality is the scree test (also known as the elbow test) where stress (or other lack of fit measure) is plotted against the dimensionality. Ideally, this choice is visually obvious from the "elbow" in the scree plot where after a certain number of dimensions, the stress is not reduced substantially. However, in many data-sets, stress decreases smoothly with increasing dimensionality making the choice of appropriate dimensionality very difficult with this method. In Figure 1b, the filled circles shows the scree plot for the face similarity dataset. Note that a slight elbow is present at two dimensions which suggests that a two dimensional configuration might be appropriate. A more salient indicator for the appropriate dimensionality can be obtained by cross-validation. The idea is to test how the configuration optimised to model the proximity data for one group of subjects can generalize to the proximity data of a different group of subjects. In Figure 1b, the open circles show the stress value for a second group of subjects with a clear rise in stress value after two dimensions while the stress continues to decrease for the first group of subjects. In this case, it seems reasonable to conclude that a two dimensional configuration is appropriate because it can best generalize to other subjects. Lee (2001) has explored other techniques to determine dimensionality based on balancing the trade-off between model fit and model complexity.

Another important issue is the interpretation of the scaling solution resulting from MDS procedures. If the proximity data were generated by a function of the distances along some set of dimensions, then the resulting configuration of points in a scaling solution should reflect those dimensions. However, often Euclidian distances are used in scaling procedures so that the orientation of axes in the resulting configuration is arbitrary: any rotation of the axes would result in the same distances (and therefore stress). In such cases, the researcher can either visually scan the configuration in order to choose an orientation of axes that leads to interpretable results or apply less arbitrary procedures by multiple regression analyses. In such analyses, the idea is to regress meaningful variables on the coordinates for the different dimensions and rotate the solution as to maximize the interpretability. In Figure 1c, the two dimensional scaling solution is shown for the 10 faces. After visual inspection, the configuration can be interpreted as the perceptual dimensions of age and adiposity.

Advances in MDS

The success of the MDS approach arises in part from the simplicity of the underlying assumptions and the wide availability of computer software to create scaling solutions. Recent research has expanded the scope of the MDS approach in several directions. In Isomap (Tenenbaum, De Silva, & Lanford, 2000), stimuli are represented as points lying on a non-linear manifold in some multidimensional space. Similarity then is computed as the geodesic distance on the manifold (i.e., the shortest distance along the manifold) as opposed to Euclidian distance in MDS. This technique is capable of discovering the non-linear degrees of freedom that underlie complex data-sets.

A drawback of most MDS algorithm is that a N x N matrix of similarity judgments is needed to scale N objects. Therefore, the number of similarity ratings needed depends quadratically on the number of objects which leads to practical limitations (e.g., subject time, number of subjects) on the number of objects that can be used in scaling studies. With modified MDS procedures, the amount of data that needs to be collected might be reduced. In the anchor point method (e.g., Buja, Swayne, Littman, & Dean, 1998), subjects rate all similarity pairs involving N objects and a smaller number of K anchor points that provide a representative sample of N objects. A modified MDS procedure then analyses the N x K similarity matrix in order to scale N objects. Future research will have to show how small K can become relative to N in order for this technique to make drastic savings possible in the similarity data collection.

Another recent advance in MDS models is the feature mapping approach (Rumelhart & Todd, 1992; Steyvers, & Busey, 2000). In the traditional approach, the physical representation of the features comprising the stimuli is explicitly ignored. In such a purely top-down approach, the multidimensional representations are sometimes difficult to relate back to the physical stimulus by visual interpretation or regression analyses. In the feature mapping approach, in addition to the proximity data, additional physical measurements on the set of stimuli is available. The goal is to find a mapping between the set of physical measurements of an stimulus to the position of that stimulus in an abstract psychological space. With this approach, the multidimensional space is related directly to the physical dimensions of the stimuli.

Challenges for MDS

Tversky and Hutchinson (1986) have argued that for language-related stimuli that have conceptual as opposed to perceptual relations, geometric models based on MDS may fail to capture some aspects of the data and might therefore be inappropriate as a representational basis. For such stimuli, tree or graph-theoretic structures might be better suited than spatial/dimensional models based on MDS. Nevertheless, the geometric model of similarity has been applied to a wide variety of stimuli rich in conceptual structure. In Latent Semantic Analysis (Landauer & Dumais, 1997), words are placed in a high dimensional semantic space by a procedure related to MDS by analysing the co-occurrence statistics for words appearing in contexts in a large corpus. In this semantic space, words with similar meaning are placed in nearby regions of the space. While the Latent Semantic Analysis approach has been very successful in modelling human performance in a variety of semantic tasks, it remains to be seen to what degree a geometric model is appropriate to model language related processes.

Also, for some highly structured objects, the simple geometric model for similarity as used in MDS may not be particularly revealing for analysing the process of generating similarity judgments. In a geometric model, it is assumed that the set of objects can be described by a fixed collection of feature values where the process of generating similarity judgments is always based on the differences for the same set of features. However, for highly structured or complex objects, the features that play a role in the similarity judgments may differ depending on what objects are compared. In alignment models (Goldstone, 1994), similarity is assessed dynamically by an alignment process that analyses which features play corresponding roles in the objects that are compared.

References

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology* 63: 516–556.

Borg, I., & Groenen, P. (1997). *Modern Multidimensional Scaling, Theory and Applications*. New York, Springer-Verlag.

Buja, A., Swayne, D.F., Littman, M., & Dean, N. (1998). XGvis: Interactive Data Visualization with Multidimensional Scaling. *Journal of Computational and Graphical Statistics*, to appear.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckart-Young' decomposition. *Psychometrika, 35*, 238-319.

Goldstone, R.L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3-28.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27, 115-129.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Lee, M.D. (2001). Determining the dimensionality of multidimensional scaling models for cognitive modeling. *Journal of Mathematical Psychology, 45*, 149-166.

MacKay, D.B. (1989). Probabilistic multidimensional scaling: an anisotropic model for distance judgments. *Journal of Mathematical Psychology, 33*, 187-205.

Nosofsky, R.M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

Rumelhart, D.E., & Todd, P.M. (1992). Learning and connectionist representations. In D. Meyers and S. Kornblum (Eds.), *Attention and Performance*. Cambridge, MA: MIT Press.

Shepard, R.N. (1962). Analysis of proximities: Multidimensional scaling with an unknown distance function I & II. *Psychometrika, 27*, 125-140, 219-246.

Shepard, R.N. (1987). Towards a universal law of generalization for psychological. *Science, 237*, 1317:1323.

Steyvers, M., & Busey, T. (2000). Predicting Similarity Ratings to Faces using Physical Descriptions. In M. Wenger, & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Lawrence Erlbaum Associates.

Tenenbaum, J.B., De Silva, V., & Lanford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290* (5500): 22.

Torgeson, W.S. (1965). Multidimensional scaling of similarity. *Psychometrika, 30*, 379-393.

Tversky, A., & Hutchinson, J.W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review, 93*, 3-22.

Young, F.W., & Hamer, R.M. (1994). *Theory and applications of multidimensional scaling*. Hillsdale, NJ: Erlbaum.

Further Reading

Ashby, F.G. (1992). *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Erlbaum.

Edelman, S. (1998). Representation is representation of similarity. Behavioral and Brain Sciences,21, 449-498.

Medin, D.L., & Goldstone, R.L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*, 390-398.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

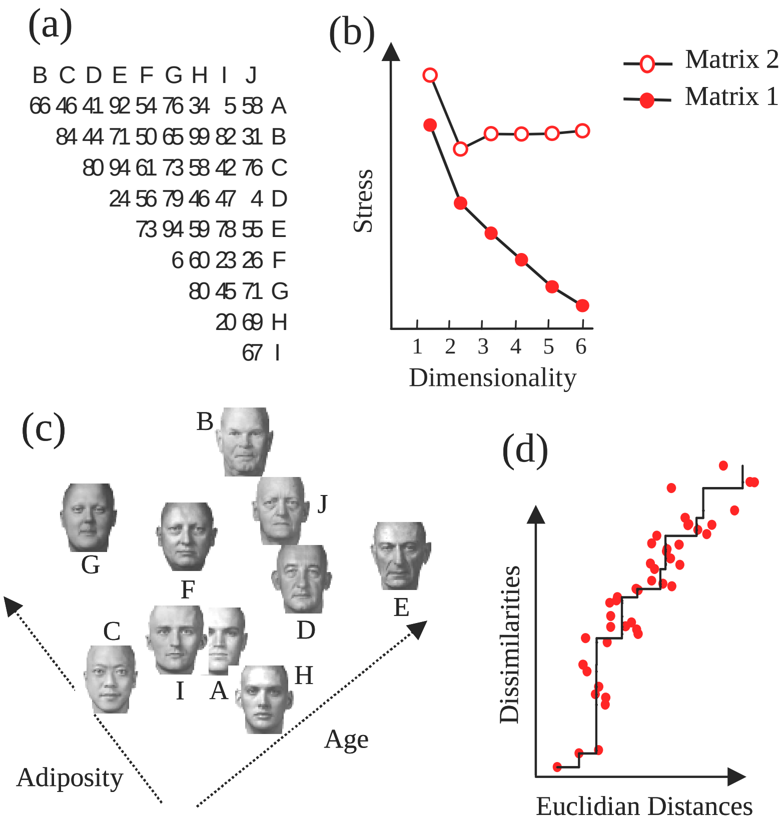
(a)
B C D E F G H I J
66 46 41 92 54 76 34 5 58 A
84 44 71 50 65 99 82 31 B
80 94 61 73 58 42 76 C
24 56 79 46 47 4 D
73 94 59 78 55 E
6 60 23 26 F
80 45 71 G
20 69 H
67 I
(b)
Stress
Dimensionality
1 2 3 4 5 6
Matrix 2
Matrix 1
(c)
B
J
G
F
E
D
C
I
A
H
Age
Adiposity
(d)
Dissimilarities
Euclidian Distances

Figure 1.