Overview of Multidimensional Scaling[1]

Anthony P.M. Coxon, Cardiff University

The information in large or complex sets of data is often difficult to conceptualise and describe. What is needed is a way of representing the relationships in the data in a way which can be easily assimilated and which at the same time accurately represents that information. One successful way of doing this is data visualization: representing the complexities of the data in a visual manner. Much has been claimed for visual representation of data in terms of the cognitive ease of taking in complex relationships and getting an overall view. Tukey originally commented that "a picture is worth a thousand words"; Simon went on to say "a diagram is (sometimes) worth ten thousand words" and Loftus recently claims "a picture is worth a thousand p values". But not any picture will do: one wishes to choose the "best" picture in the sense that it better or more accurately matches the data than any others do.

Multidimensional scaling (MDS) refers to a whole family of procedures for data analysis and data visualization. Each procedure is designed to take the structure in a set of data and graphically represent it as a "configuration" – a picture or "map" or pattern of points in three or two dimensions . (It can be one-dimensional – unidimensional scaling – or it can actually be more than three, but this gives rise to difficulties in representation). This picture is constructed in such a way that the geometric relationships in the picture mirror the data relationships as closely as possible.

MDS procedures can be used on a wide variety of data, using different models and allowing different assumptions about the level of measurement.

In the simplest case, a data matrix giving information about the similarity (or DISSIMILARITY) between a set of objects is represented by the proximity (or distance) between corresponding points in a low-dimensional space. Given a set of data, interpreted as "distances", it finds the map locations which generated them.

For example, in a study of how people categorize drugs, a list of drugs was elicited from a sample of users and non-users; 28 drugs were retained for a free-SORTING experiment and the co-occurrence frequency was used as the measure of similarity. The data were scaled in 2-dimensions, producing the following map:

[1] Early expanded version of entry "Multidimensional Scaling" by APMC in The SAGE Encyclopedia of Social Science Research Methods (2004)

[FIGURE about here]

In the case of perfect data, the correspondence between the dissimilarity data and the distances of the solution will be total, but with imperfect data, the degree of fit is given by size of the normalised Stress (residual sum of squares) value, which is a measure of badness of fit. In this case the fit is excellent ($stress_1$ =0.097) and it is three times smaller than random expectation. The solution configuration yields three highly distinct clusters (confirmed by independent hierarchical clustering) of Mild stimulant drugs (core prototypes: tobacco; caffeine), Hard recreational (core: Cocaine, Heroin) Household Prescription (core: Aspirin Penicillin)..

MDS can be either exploratory (simply providing a useful and easily-assimilable visualization of a data set) or it can be explanatory (giving a geometric representation of the structure in a data set, where the assumptions of the model are taken to represent the way in which the data were produced). Compared to other multivariate methods, MDS models are usually distribution-free, make conservative demands on the structure of the data, are unaffected by non-systematic missing data, can be used with a wide variety of measures, and the solutions are usually readily interpretable. The chief weaknesses are relative ignorance of the sampling properties of stress, prone-ness to local minima solutions and inability to represent the asymmetry of causal models.

THE BASIC MDS MODEL

The basic type of MDS is the analysis of 2-way, 1 mode data (e.g. a matrix of correlations or other dis/similarity measures), using the Euclidean distance model. The original *metric* version, "classical" MDS converted the "distances" into scalar-products, and then factored or decomposed them into a set of locations in a low-dimensional space (hence "smallest space analysis"). The first *non-metric* model was developed by Roger N. Shepard in 1962, who showed that the merely ordinal constraints of the data, if imposed in sufficient number, guarantee metric recovery, and he provided the first iterative computer program that implemented the claim. Kruskal (1964) gave OLS statistical substance to it, Young developed a frequently-used Alternating Least Squares program (ALSCAL) and probabilistic MAXIMUM LIKELIHOOD versions of MDS have also subsequently been developed (e.g. MULTISCALE, Ramsay ,1978).

VARIANTS OF MDS

The various types of MDS can be differentiated (Coxon 1982) in terms of:

- **The data** (whose "shape" is described in terms of its *way* (rows, columns, "third-way" replications) and its *mode* (the number of sets of distinct objects, such as variables, subjects)
- **The transformation (re-scaling) function** (or LEVEL OF MEASUREMENT, which specifies the extent to which the data properties are to be adequately represented in the solution). The primary distinction is *non-metric* (ordinal or lower) *vs metric* (interval and ratio) scaling
- **The model** (usually the Euclidean distance model, but also covering simpler Minkowski metrics such as City-Block, and different composition functions like the vector, scalar products or FACTOR model).

TWO-WAY, ONE-MODE DATA

Any non-negative, symmetric judgments or measures can provide input to the basic model: frequencies, counts; ratings/rankings of similarity; co-occurrence, co-location , confusion rates; measures of association, etc. and a two-dimensional solution should be stable with 12 or more points Non-metric analyses use distance models; metric models additionally may use vector/factor models. Stress values from simulation studies of random configurations provide an empirical yardstick for assessing obtained stress values .

2-WAY, 2-MODE DATA

Rectangular data matrices (usually with subjects as rows and variables as columns) consist of profile data or preference ratings/rankings. Distance (unfolding) or vector models are employed to produce a joint BIPLOT of the row and column elements as points. The vector model (MDPREF) is formally identical to simple CORRESPONDENCE ANALYSIS, and represents the row elements as unit vectors. Because such data are row-conditional, caution is needed in interpreting inter-set relations in the solution. If the sorting data of the example were entered directly as (individual x object matrix, with entries giving the category number) then a program such as MDSORT will represent both the objects and the categories.

3-WAY DATA

The "third way" consists of different individuals, time or spatial-points, sub-groups in the data, experimental conditions, different methods etc. The most common variant is 3-way, 2-mode data (a stack of 2-way, 1-mode matrices) -- which when represented by the metric weighted distance model, is termed INDSCAL (Individual Differences Scaling, Carroll 1970). In the INDSCAL solution, the objects are represented in a Group Space (whose dimensions are fixed, and should be readily interpretable), and each element of the third-way ("individual") has a set of non-negative dimensional weights, by which the individual differentially shrinks or expands each the dimensions of the Group Space to form a "Private Space". Individual differences are thus represented by a profile of salience weights applied to a common space. When plotted separately as the "Subject Space" , the angular separation between the points represents the closeness of their profiles, and the distance from the origin approximately represents the variance explained. Returning to the example, If the individuals were divided into sub-groups (such as Gender x Usage) and a co-occurrence matrix calculated within each group, then the set of matrices forms 3-way 2-mode data, and their scaling by INDSCAL would yield information about how far the groups differed in terms of the importance (weights) attributed to the dimensions underlying the drugs-configuration.

Other variants of MDS exist for other types of data (e.g. Tables, triads), other transformations (parametric mapping; power functions) and other models (non-Euclidean metrics, simple conjoint models, discrete models such as additive trees), and also for a wide range of three-way models and hybrid models such as CLASCAL (a development of INDSCAL which parameterises latent *classes* of individuals). Utilities exist for comparing configuration (Procrustean Rotation) and extensions for up to 7-way data are possible.

An extensive review of variants of MDS are contained in Carroll and Arabie (1981) and a wide-ranging bibliography of 3-way applications is accessible at: http://ruls01.fsw.leidenuniv.nl/~kroonenb/document/reflist.htm

APPLICATIONS

For the basic model, and for the input of aggregate measures, restrictions on the number of cases is normally irrelevant. But for 2-way,2-mode data and 3-way data, applications are rarely feasible (or interpretable) for more than 100 individual cases. In this case "pseudo-subjects" (subgroups of individuals chosen either on analytic grounds, or after CLUSTER ANALYSIS) are more appropriately used. Disciplines using MDS now extend well beyond the original focus in psychology and political science, and include other social sciences, such as economics, as well as biological sciences and chemistry.

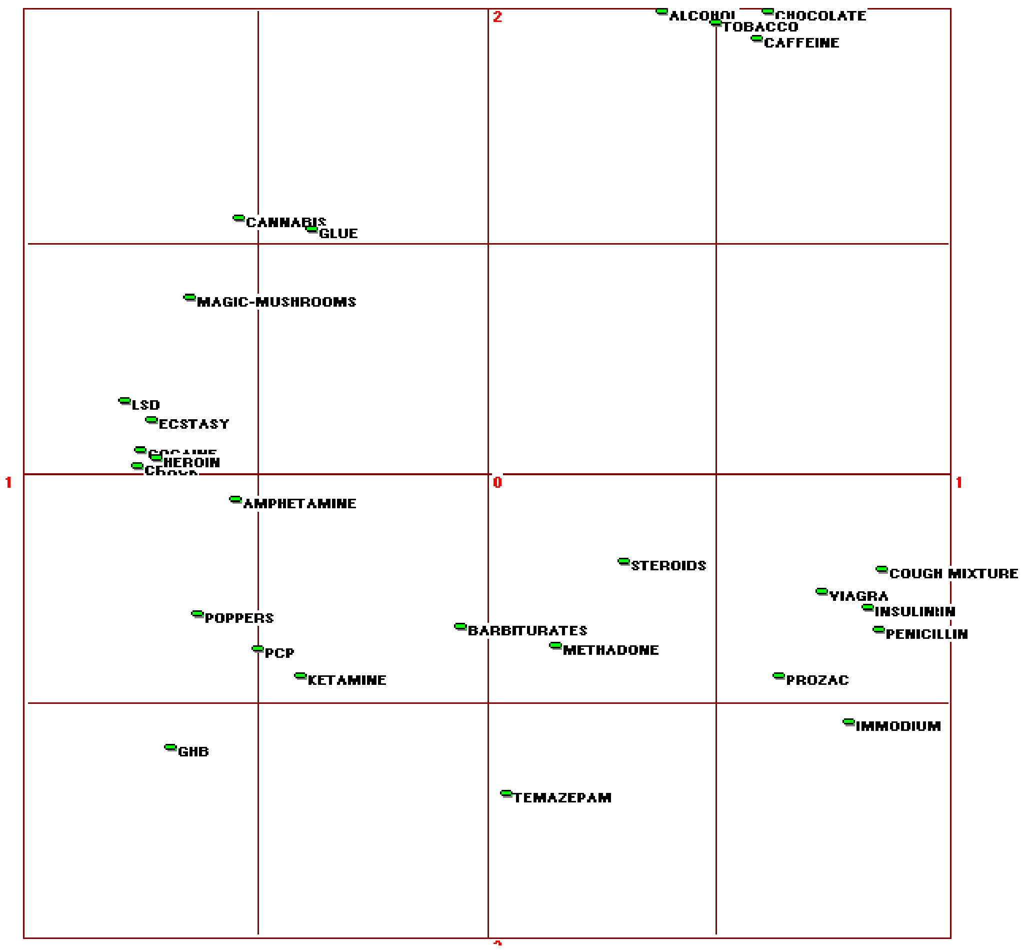
REFERENCES

Carroll, J.D. & J-J Chang (1970) Analysis of individual differences in multidimensional scaling via a N-way generalisation of Eckart-Young decomposition, *Psychometrika, 35,I 283-299 and 310-319*

Carroll, J.D & P. Arabie (1981) Multidimensional scaling, in M.H. Birnbaum, ed, *Measurement, Judgment and Decision Making*, San Diego: Academic Press

Coxon, A.P.M. (1982) *The User's Guide to Multidimensional Scaling*, London: Heinemann Educational

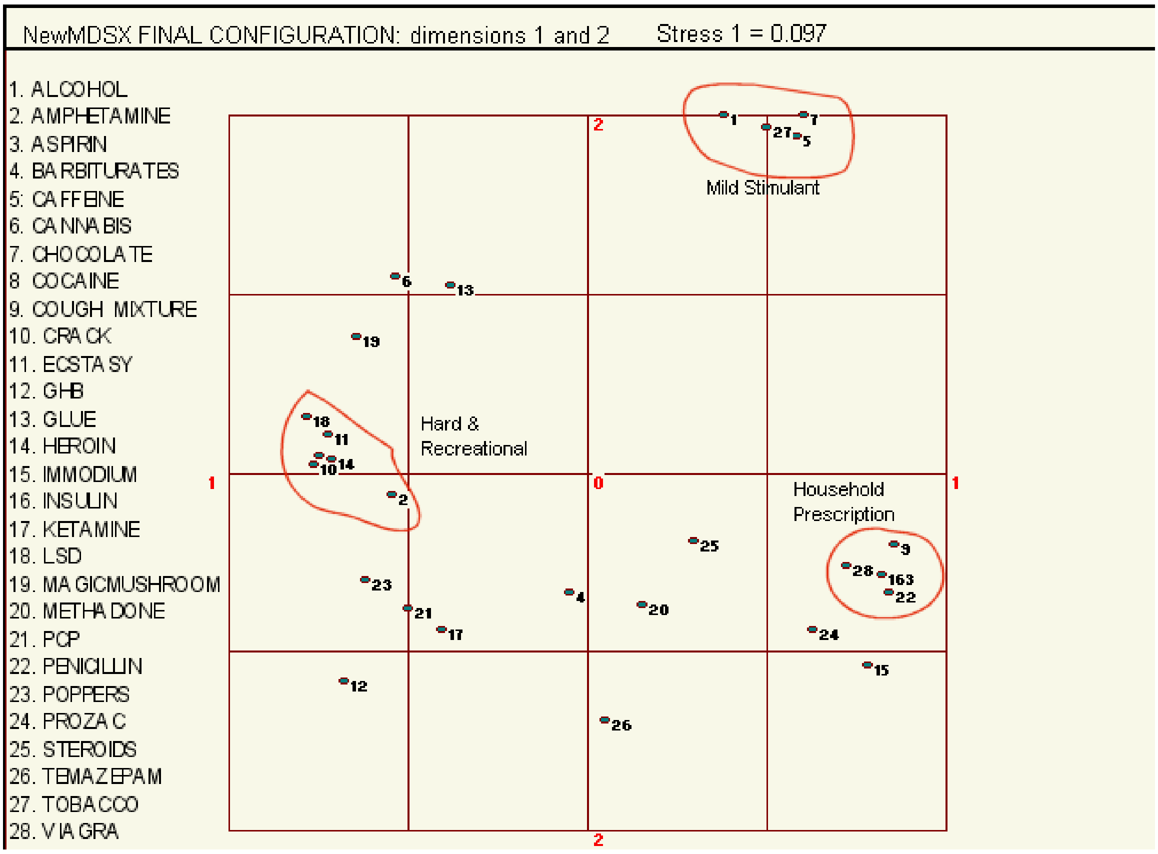
Kruskal. J B. (1964) Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis, *Psychometrika,29, 1-27 and 115-129*



<u>2D Solution (Basic Non-metric MDS Distance) Model</u>

Stress1 = 0.096768

STRESS1 BASED ON APPROXIMATION TO RANDOM DATA = 0.290413



FIGURE

REFERENCES

J.H. Larkin, H.A. Simon, Why a diagram is (sometimes) worth ten thousand words, Cognitive Science 11 (1987) 65 99.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley Publishing Company.

Loftus, G. R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. Behavior Research Methods, Instruments and Computers, 25, 250-256.