Multidimensional Scaling in R: SMACOF

Patrick Mair
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business

Jan de Leeuw
Department of Statistics
University of California, Los Angeles (UCLA)

Workshop Content

Throughout the workshop we use the R package smacof available on CRAN.

Materials at `http://statmath.wu.ac.at/~mair/CARME/`.

- Symmetric SMACOF
- Spherical SMACOF
- 3-Way SMACOF
- Constrained SMACOF
- Rectangular SMACOF

De Leeuw, J. & Mair, P. (2009). Multidimensional Scaling using Majorization: SMACOF in R. *Journal of Statistical Software, 31(3)*, p. 1-30.
URL: `http://www.jstatsoft.org`

Multidimensional Scaling (MDS)

- Family of data-analytic methods which represent distances between objects in a low-dimensional space.
- Input structure: Dissimilarity matrix.
- Computation: Optimize corresponding target function.
- Output: Configurations in low-dimensional space.
- Visualization: Configuration plot.

MDS Models

- Classical (metric) scaling (Torgerson, 1952): data on interval level; strain loss, stress loss.
- Nonmetric MDS (Shepard, 1962): data on ordinal level; monotonicity constraint on dissimilarities.
- Individual differences: K dissimilarity matrices.
- Unfolding: Preference data; n_1 raters, n_2 objects.
- Other related models: Procrustes, CA, Gifi.

R Implementation: SMACOF package

The R Project for Statistical Computing

- R is an open source software environment for statistical computing and graphics
- `http://www.R-project.org`
- $\sim$2800 packages available

The smacof package is available on

- CRAN: `http://CRAN.R-project.org`
- PsychoR: `http://r-forge.r-project.org/projects/psychor`.

Symmetric SMACOF

- Distance matrix Δ of dimension $n \times n$ with elements δ_{ij}.
- Problem to solve: Locate points (*configurations*) in a p-dimensional space such that the distances $d_{ij}(X)$ between the points approximate δ_{ij}.
- Configuration distances:

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}$$

- Minimize *stress* (*Majorization*; de Leeuw, 1977):

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \rightarrow \text{min!}$$

Example 1: French Map

Nonmetric symmetric SMACOF

- Dissimilarities δ_{ij} on ordinal scale.
- Monotonic transformation f: $\delta_{ij} < \delta_{i'j'} \Rightarrow f(\delta_{ij}) < f(\delta_{i'j'})$.
- Achieved by monotone regression fit in each iteration (PAVA).
- Nonmetric stress:

$$\sigma(X, \widehat{D}) = \sum_{i<j} w_{ij}(\hat{d}_{ij} - d_{ij}(X))^2$$

- Disparities $\hat{d}_{ij} = f(\delta_{ij})$

Example 2: Political statements

3-Way SMACOF

SMACOF for individual differences:

- $k = 1, \ldots, K$ separate symmetric distance matrices.
- Stress:

$$\sigma(X^*) = \sum_{k=1}^{K} \sum_{i<j} w_{ij,k}(\delta_{ij,k} - d_{ij}(X_k))^2.$$

- INDSCAL (unrestricted; Carrol & Chang, 1970).
- Restrictions on the configuration weight matrix $X_k = ZC_k$ (diagonal, identity).
- IDIOSCAL (Carrol & Wish, 1974).

Example 3: Wine tasting

Wine Tasting: Descriptives

	Price	Alcohol	Mean Rating
Jurtschitsch Chardonnay	14.99	13.00	2.00
Ziniel Chardonnay	7.00	12.00	2.60
Markowitsch Chardonnay	9.99	12.50	2.60
Ritinitis Noble Retsina	9.99	12.00	4.30
Retsina	2.99	11.50	4.60
Krems Chardonnay	5.99	12.50	2.70
Castel Nova Chardonnay	1.99	12.00	2.80

Spherical SMACOF

Restrictions on the configurations (*weakly constrained MDS*, Cox & Cox, 1991).

$$\mathbf{x}_i'\Lambda\mathbf{x}_i + 2\mathbf{x}_i'\beta + \gamma = 0$$

- $\mathbb{R}^2$: circle, ellipse, hyperbola, parabola.
- $\mathbb{R}^3$: sphere, ellipsoid, hyperboloid, paraboloid, cylinder.
- Optimization: Primal and dual methods available.

Example 4: Trading volume

Constrained SMACOF

Restrictions by means of external constraints (*confirmatory MDS*, de Leeuw & Heiser, 1980).

$$X = ZC$$

where Z is a known predictor matrix.

- Decomposition: $X = \begin{bmatrix} X_1 & ZC_1 & C_2 \end{bmatrix}$.
- X_1 is the unrestricted part and of dimension $n \times q_1$.
- ZC_1 is the linearly restricted part of dimension $n \times q_2$.
- C_2 is a diagonal matrix of order n.

Constrained SMACOF

Triplet notation (q_1, q_2, q_3):

- q_1 is the number of unrestricted dimensions.
- q_2 the number of linearly restricted dimensions.
- q_3 is either zero or one (absence/presence).
- Important special case: Bentler-Weeks uniqueness model $(q, 0, 1)$.
- Additional options: Simplex and circumplex fitting.

Example 5: Rectangles

Rectangular SMACOF (Unfolding)

Rectangular $n_1 \times n_2$ preference matrix Δ.

- Stress becomes

$$\sigma(X) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}(\delta_{ij} - d_{ij}(X_1, X_2))^2 \rightarrow \text{min!}$$

- Judge $n_1 \times p$ configuration matrix.
- Object $n_2 \times p$ configuration matrix.

Example 6: Company rating

References

- De Leeuw, J. & Mair, P. (2009). Multidimensional Scaling using Majorization: SMACOF in R. Journal of Statistical Software, 31(3), p. 1-30. URL: `http://www.jstatsoft.org`

- Borg, I., & Groenen, P. J. F. (2005). Modern Multidimensional Scaling. New York: Springer.
- Cox, T. F., & Cox, M. A. A. (2001). Multidimensional Scaling (2nd edition). Boca Raton, FL: Chapman & Hall/CRC.

Links and Contact

PsychoR project:

- Website: `http://r-forge.r-project.org/projects/psychor`
- Additional PsychoR topics: correspondence analysis (anacor), Gifi optimal scaling (homals), isotone optimization (isotone), aspect framework (aspect).

Patrick Mair
Institute for Statistics and Mathematics
WU Vienna University of Economics
Augasse 2-6
1090 Vienna

Email: `patrick.mair@wu.ac.at`
Website: `http://statmath.wu.ac.at/~mair`