R Packages smacofx, cops and stops

Flexible Multidimensional Scaling

Thomas Rusch

WU Vienna

Date: 2024-07-09

Outline

We introduce the R packages **smacofx**, **cops**, **stops** for flexible multidimensional scaling (fMDS). Together with **smacof** they comprise the smacofverse.

- General Idea of fMDS
- Methods in **smacofx**, **cops** for obtaining a configuration
- Example: Finding a configuration
- Hyperparameter selection via functionality in **stops**
- Example: Hyperparameter selection
- Post-Fit Infrastructure
- Conclusion

Joint work with Patrick Mair (Harvard), Jan De Leeuw (UCLA) and Kurt Hornik (WU Vienna)

Multidimensional Scaling - I

MDS methods represent objects as an $n \times p$ matrix X (the configuration) based on pairwise proximities in the original space. Usually p is small.

The proximities δ_{ij} (given or derived) are approximated by pairwise fitted distances $d_{ij}(X)$ in X based on the badness-of-fit criterion $L(\delta_{ij}, d_{ij}(X))$.

We look for

$$\arg\min_X L(\delta_{ij}, d_{ij}(X))$$

the X that minimizes the approximation error.

Unsupervised learning method popular for

- Similarity analysis
- Dimensionality reduction
- Graph drawing
- Data visualization and exploratory data analysis

Multidimensional Scaling - II

A popular badness-of-fit is STRESS:

$$\sigma_{stress}(X) = \sum_{i<j} \left(f(\delta_{ij}) - d_{ij}(X)\right)^2$$

to be minimized over X.

- $d_{ij}(X)$...fitted distances (usually Euclidean)
- δ_{ij}... dissimilarities
- f... implicit optimal scaling transformation, i.e., ratio, interval (metric), and ordinal (nonmetric)

We make this formulation more general by allowing for explicit parametric transformations of the δ_{ij} and/or $d_{ij}(X)$ and/or penalization → flexible MDS.

Flexible Multidimensional Scaling

Flexible MDS: Extend standard MDS methods by

- Introducing **parametric transformations** of proximities $T_\Delta(\delta_{ij}|\theta_\Delta)$ and fitted distances $T_D(d_{ij}(X)|\theta_D)$
- Residual weighting w_{ij}
- **Penalizing** the fit with index $I(X)$ to emphasize different aspects of the result

$$\sigma_{fMDS}(X|\theta) = \left(\sum_{i<j} w_{ij}\left(f(T_\Delta(\delta_{ij}|\theta_\Delta)) - T_D(d_{ij}(X)|\theta_D)\right)^2\right) + v_1 I(X)$$

- This is to be minimized over X. The weight v_1 governs the strength of the (optional) penality. We retain the implicit optimal scaling f for the transformed proximities.

fMDS Models - I

The fMDS formulation encompasses many MDS models for obtaining a configuration X including:

- Ratio, interval and nonmetric MDS: optimal scaling transformations f; $T_\Delta(\delta_{ij}|\theta_\Delta)$ and $T_D(d_{ij}(X)|\theta_D)$ the identity function.
- Sammon mapping, elastic scaling: Residual weighting with $w_{ij} = \delta_{ij}^{-1}$ and $w_{ij} = \delta_{ij}^{-2}$.
- ALSCAL: $T_\Delta(\delta_{ij}|\theta_\Delta)$ and $T_D(d_{ij}(X)|\theta_D)$ the square function (metric only).
- Multiscale MDS: $T_\Delta(\delta_{ij}|\theta_\Delta)$ and $T_D(d_{ij}(X)|\theta_D)$ are the logarithm (metric only).
- Power-Stress MDS, Box-Cox MDS: $T_\Delta(\delta_{ij}|\theta_\Delta)$ and $T_D(d_{ij}(X)|\theta_D)$ are power/Box-Cox transformations with θ_Δ, θ_D the powers (metric only).
- Curvilinear Component Analysis, Sparsified MDS: Residual weighting with $w_{ij}(X) = \mathbb{1}\{d_{ij}(X) \leq \tau\}$ (metric only).

Implemented in packages: **smacof** (metric and nonmetric MDS), **smacofx** (everything else).

fMDS Models - II

The fMDS formulation encompasses many MDS models for obtaining a configuration X:

- r-stress MDS: Optimal scaling f and $T_D(d_{ij}(X)|\theta_D)$ power transformation with θ_D the power (metric and nonmetric).
- Isomap, Local MDS: $T_\Delta(\delta_{ij}|\theta_\Delta)$ and/or $T_D(d_{ij}(X)|\theta_D)$ are neighbourhood-preserving transformations (e.g., geodesic distance transformation for k neighbours, metric only).
- Cluster Optimized Proximity Scaling (COPS-C): Optimal scaling, power transformations for $T_\Delta(\delta_{ij}|\theta_\Delta)$ and $T_D(d_{ij}(X)|\theta_D)$ and clusteredness penalty for $I(X)$ (metric and nonmetric).
- Flexible also means mix-and-match of optimal scaling, explicit transformations and penalties for new methods (e.g., creating Interval Curvilinear Component Sammon Mapping R-Stress COPS-C)

Available in packages **cops** (COPS-C) and **smacofx** (everything else).

Example Data

For illustration, we use the data of Koller et al. (2013):

- 1013 responses on 24 Likert items (1 through 5) from two scales
- Scale 1: Customer susceptibility to interpersonal influence (CSII)
- Scale 2: Self-concept clarity (SCC)
- Scales are related: Higher SCC associated with lower CSII

We use Gower dissimilarity for ordinal variables.

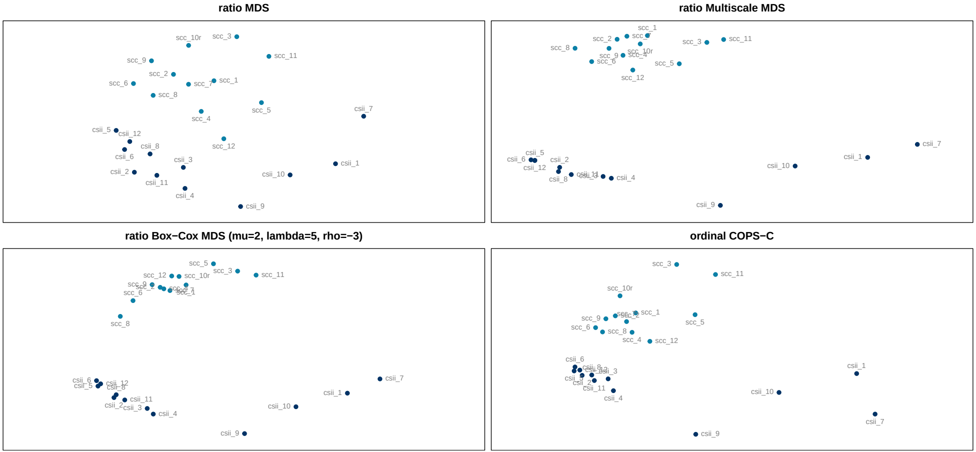
Example I: fMDS

Out of the many models available, we fit

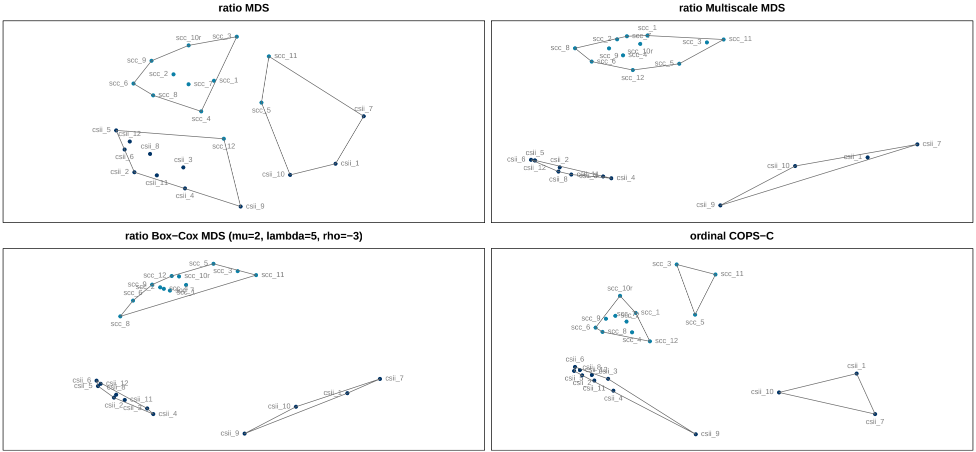
- Ratio MDS (`smacof::mds(delta)`)
- Ratio Multiscale MDS (`smacofx::multiscale(1+delta)`)
- Ratio Box-Cox MDS (`smacofx::bcmds(delta,mu=2,lambda=5,rho=-3)`)
- Ordinal COPS-C (`cops::copsc(delta,type="ordinal",cordweight=0.1,stressweight=0.9)`)

Configurations are Procrustes adjusted. We also draw the convex hull from the optimal k-means solution (with 3, 3, 3 and 4 centers).

Example I: fMDS



Example I: fMDS



Example I: fMDS

We see that

- ratio MDS shows less structure and also misassignment in clustering (e.g., SCC 5 and SCC 11)
- fMDS models can emphasize the structure better
- fMDS models reveals additional insight (CSII comprises two scales, perhaps SCC as well)

The extra flexibility in MDS models can bring out structure in the data more clearly which in turn may allow for more insightful visualization and exploration.

Hyperparameter Selection

In many fMDS models (e.g., the Box-Cox MDS from before), the explicit transformation hyperparameters, θ, needed to be supplied for finding an X from $\sigma_{fMDS}(X|\theta)$. How to choose them?

Hyperparameter Selection

In many fMDS models (e.g., the Box-Cox MDS from before), the explicit transformation hyperparameters, θ, needed to be supplied for finding an X from $\sigma_{fMDS}(X|\theta)$. How to choose them?

Hyperparameter selection via the STOPS framework:

- We want to have a certain structural quality in the configuration
- We measure that with with M univariate c-structuredness indices (e.g., "c-clusteredness", "c-functionality")
- Combine badness-of-fit measure with c-structuredness indices to an objective function for hyperparameter selection (stoploss)
- Optimize over the space of the hyperparameters (with Bayesian optimization, particle swarm, etc.)
- Select the transformation parameter θ^* that leads to the optimal stoploss.

We do this with a sort of profile method.

Hyperparameter Selection - Details

Let $X^*(\theta) := \arg\min_X \sigma'_{PS}(X|\theta)$. M different univariate indices with possible metaparameters γ, $I_m(X^*(\theta)|\gamma)$, measure c-structuredness. We look for optimal hyperparameters θ^* via

$$\theta^* = \arg\min_\theta \mathrm{aSTOPS}(\theta|v_0,\ldots,v_M,\Delta,\gamma) = v_0 \cdot \sigma'_{\mathrm{PS}}(X^*(\theta)|\theta) + \sum_{m=1}^{M} v_m I_m(X^*(\theta)|\gamma)$$

$$\theta^* = \arg\min_\theta \mathrm{mSTOPS}(\theta|v_0,\ldots,v_M,\Delta,\gamma) = \sigma'_{\mathrm{PS}}(X^*(\theta)|\theta)^{v_0} \cdot \prod_{m=1}^{M} I_m(X^*(\theta)|\gamma)^{v_m}$$

with $v_0 \in \mathbb{R}_+$ and $v_1,\ldots,v_M \in \mathbb{R}$ being the scalarization weights.

We optimize this with derivative-free heuristics, e.g., Bayesian optimization. The STOPS functionality is available in package **stops**.

Example II: Hyperparameter Selection

We now select hyperparameters for the Box-Cox MDS (we *ad hoc* used `mu=2`, `lambda=5`, `rho=-3` before). We want to have high c-clusteredness of the configuration.

- We measure c-clusteredness with the OPTICS Cordillera (*OC*, Rusch et al., 2018)
- Clusters must comprise at least 4 points, same weight for stress and *OC*
- Box constraints $\theta \in [(0.1, 3)^\top, (0.1, 3)^\top, (-3, 3)^\top]$

Example II: Hyperparameter Selection

We now select hyperparameters for the Box-Cox MDS (we *ad hoc* used `mu=2`, `lambda=5`, `rho=-3` before). We want to have high c-clusteredness of the configuration.

- We measure c-clusteredness with the OPTICS Cordillera (*OC*, Rusch et al., 2018)
- Clusters must comprise at least 4 points, same weight for stress and *OC*
- Box constraints $\theta \in [(0.1, 3)^\top, (0.1, 3)^\top, (-3, 3)^\top]$

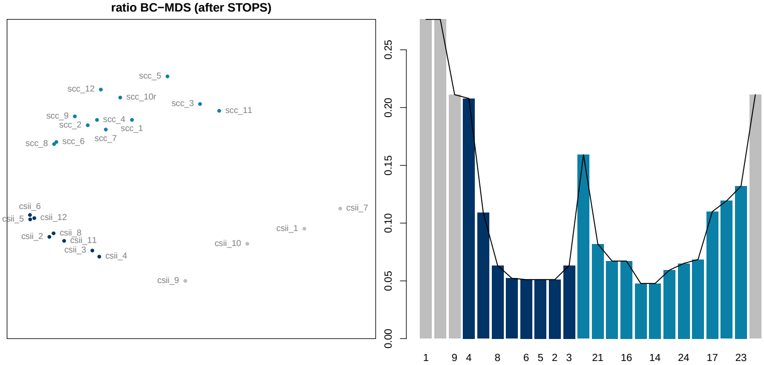
Call:

```
> st1 <- stops(delta,loss="bcmds",upper=c(3,3,3),lower=c(0.1,0.1,-3),
+               structures="cclusteredness",strucpars=list(minpts=4))
```

Selected hyperparameters were

```
> coef(st1)
       mu    lambda       rho
1.9473251 0.6326918 1.4585256
```

Example II: Hyperparameter Selection



Box-Cox MDS with these parameters gives a rather clustered result with clusters of at least 4 objects (4 objects as noise due to density)

Post-Fit Infrastructure

We feature an object-oriented approach for all objects returned by functions in **smacofx**, **cops** and **stops** to use/match the functionality in **smacof**.

- Standard S3 Generics: `print`, `summary`, `plot`, `coef` (gives the θ)
- Plotting: Configuration plot, Shepard plot, bubbleplot, stress-decomposition plot, residual plot, transformation plot (argument `plot.type`), Biplots (`biplotmds`)
- Uncertainty quantification: MDS jackknife (`jackmds`), MDS bootstrap (`bootmds`), permutation test (`permtest`)
- Local Minima Diagnostic: Effect of starting configuration (`icGenExplore`), different starting configuration (`multistart`)

Conclusions

We introduced functionality for fMDS into the smacofverse via the packages **smacofx**, **cops**, **stops**

- Many MDS models for metric and nonmetric MDS
- Flexibility is achieved via combinations of optimal scaling, explicit transformations, weighting and penalization.
- The packages currently feature 17 fMDS methods, with support for hyperparameter selection in the STOPS framework
- Object-oriented approach that mimics/is compatible with the approach in **smacof**
- Post-fit infrastructure that is streamlined between all the packages

Conclusions

We introduced functionality for fMDS into the smacofverse via the packages **smacofx**, **cops**, **stops**

- Many MDS models for metric and nonmetric MDS
- Flexibility is achieved via combinations of optimal scaling, explicit transformations, weighting and penalization.
- The packages currently feature 17 fMDS methods, with support for hyperparameter selection in the STOPS framework
- Object-oriented approach that mimics/is compatible with the approach in **smacof**
- Post-fit infrastructure that is streamlined between all the packages

The smacofverse (comprising smacof, smacofx, cops, stops) is currently the most versatile and complete software suite for MDS in existence.

References

- Koller, M, Floh, A., Zauner, A., Rusch, T. (2013). Persuasibility and the self: Investigating heterogeneity among consumers. *Australasian Marketing Journal, 21*, 94–104.
- Rusch, T., De Leeuw, J., Mair, P., and Hornik, K. (in preparation). smacofverse I: Flexible multidimensional scaling with the package smacofx.
- Rusch, T., Hornik, K., and Mair, P. (2018). Assessing and quantifying clusteredness: The OPTICS Cordillera. *Journal of Computational and Graphical Statistics, 27*, 220–233.
- Rusch, T., Mair, P., and Hornik, K. (2021). Cluster optimized proximity scaling. *Journal of Computational and Graphical Statistics, 30*,1156–1167.
- Rusch, T., Mair, P., and Hornik, K. (2023). Structure-based hyperparameter selection with Bayesian optimization in multidimensional scaling. *Statistics and Computing*, 33 (28).
- Rusch, T., Mair, P., and Hornik, K. (in preparation). smacofverse II: More flexible multidimensional scaling with the packages cops and stops.