FreeSortR : a R package for the analysis of free sorting data.

Philippe Courcoux

ONIRIS, Sensometrics and Chemiometrics Lab. BP43222 Nantes, France
E-mail : *philippe.courcoux@oniris-nantes.fr*

Abstract
FreeSortR is a R package dedicated to the analysis of data coming from a free sorting experiment: factorial representation with confidence regions, analysis of verbalization task and computing of consensus partition.
Keywords: free sorting task, Mds, consensus partition.

Résumé en Français
FreeSortR est un paquet R dédié à l'analyse de données issues d'une épreuve de tri libre : représentation factorielle avec régions de confiance, analyse des données de verbalisation et calcul de la partition consensus.
Mots-clés : tri libre, Mds, partition consensus

1. Introduction

Free sorting is becoming more and more popular as a tool to investigate the perception of a panel of subjects towards a set of stimuli.
In the course of a free sorting experiment, each subject is asked to give a partition of the whole set of objects. This task is very natural and simple and may be used with untrained assessors.
The package FreeSortR gives to the practitionner several tools for analysing data resulting from this task:

- Factorial representation of the distances between products using Mds techniques,
- Interpretation of the configuration of stimuli by representation of the words given during a verbalization task,
- Computation of the consensus partition among the subjects, giving a segmentation of the products.

The different techniques will be illustrated on the basis of a sensory experiment: stimuli were 16 samples of aromas used for training olfactory memory. A panel of thirty one subjects were presented with the whole set of aromas and were asked to group the samples according to the similarity of the odour of the products.

2. Representation of items using Mds

2.1 Sorting task

After a free sorting task, data for each subject are often expressed in a symmetric (products by products) matrix depicting the dissimilarity between stimuli (1 for two items in different groups and 0 if not). At the panel level, the individual dissimilarity matrices can be averaged across subjects yielding a global dissimilarity matrix between products.

The package FreeSortR gives several Multidimensional Scaling (MDS) solutions for representing the global dissimilarity matrix: metric and non-metric Mds based on the Smacof algorithm.

Confidence regions around products may be drawn using a bootstrapping on subjects.
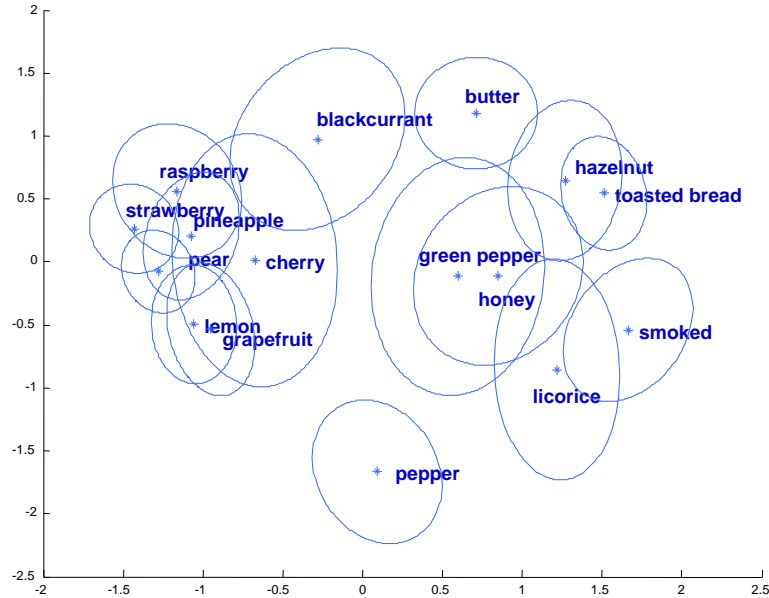
Mds and confidence regions around the aroma.

2.2   Verbalization task

Usually, a verbalization task is performed by the subjects after the sorting task : subjects are asked to freely describe with their own words and expressions the groups they formed. These verbatims are used for interpret the products configuration.
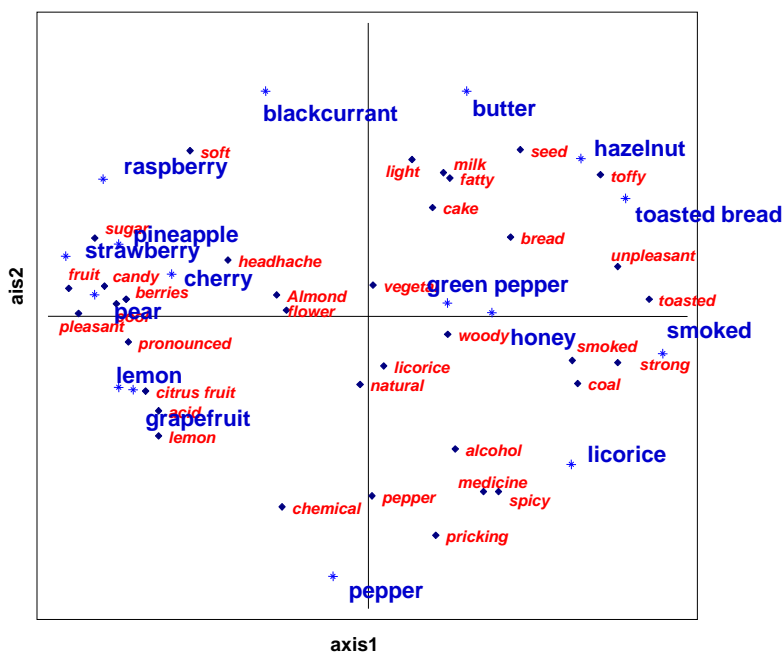
Correlations between verbatims and Mds dimensions.

3.   Clustering of subjects and consensus partition of stimuli

Very often, the stimuli are the center of interest of the study but there are several cases where the subjects themselves are the main target of the investigations. In any case, it is interest to assess the agreement among the subjects and segment them on the basis of how they have sorted the stimuli.

The similarity between two subjects can be assessed by computing the adjusted Rand Index between the two partitions given by these subjects. A strategy for segmenting the panel may be based on a clustering the subjects on the basis of the matrix of adjusted Rand Index between subjects.

It may also be of interest to find a central partition that stands as a group average partition to a group of subjects. The Rand Index is used to determine a central (or consensus) partition to the set of partitions given by a group of subjects. This partition is assumed to be the closest partition to the partitions under consideration (Courcoux et al., 2013).

In the case of the aroma, the consensus partition has seven classes, which are the following: {Lemon, Grapefruit, Pineapple, Pear, Strawberry, Raspberry} {Honey, Toasted bread, Hazelnut} {Cherry} {Blackcurrant, Butter} {Pepper} {Licorice, Smoked} {Green pepper}

The mean Adjusted Rand index between the consensus partition and the 31 original partitions is 0.3080. This means that this consensus partition may be considered as a good summary of the partitions given by the subjects.

References

Courcoux P., Faye P., E.M.Qannari E.M. (2013) Determination of the consensus partition and cluster analysis of subjects in a free sorting task experiment. Food Quality and Preference 32:107–112.

Faye P., Courcoux P., Qannari E.M. (2011) Méthodes de traitement statistique des données issues d'une épreuve de tri libre. Revue Modulad, 43:1-24.

Guénoche A. (2011) Consensus of partitions : a constructive approach. Advances in Data Analysis and Classification, 5, 215-229

Hubert L. J., Arabie P. (1985). Comparing partitions. Journal of Classification, 2, 193–218.

Krieger A.M., Green P.E. (1999) A generalized rand-index method for consensus clustering of separate partitions of the same data base. Journal of Classification 16, 63-69.