
Analyzing categorization data

Marine Cadoret, Sébastien Lê, and Jérôme Pagès

Agrocampus Ouest, Laboratoire de mathématiques appliquées, 65 rue de
Saint-Brieuc, CS 84215, 35042 RENNES Cedex

`marine.cadoret@agrocampus-ouest.fr`

Summary. Sorting task (or categorization) is a data collection which consists in asking J subjects to partition a set of I objects function of their similarities: a subject puts two objects in the same group if he considers them similar and in two different groups if not. This data collection was used for the first time in psychology and is becoming now more and more popular in sensory analysis.

The current practice consists in gathering the data issued from the J subjects in a cooccurrences data table C of dimension $I \times I$, where $C_{ii'}$ denotes the number of subjects who gathered the two objects i and i' in the same group. This data table is then analyzed by multidimensional scaling (MDS) [1].

An alternative consists in gathering the data in a table X of dimension $I \times J$ where x_{ij} denotes the index of the group in which the subject j puts the object i . Each column of X can then be considered as a qualitative variable (partition); therefore X can be analyzed by multiple correspondence analysis (MCA). This MCA is the core of a methodology which provides a representation of the objects (but compared with the one of MDS, this representation can be completed by confidence ellipses around objects) and a representation of the subjects linked to the one of the objects.

The aim of this communication is to describe this methodology, to compare it with the classical approach and to illustrate it with an example. In this example, the subjects used words to describe the groups they provided.

Key words: Categorization data, Multidimensional Scaling, Multiple Correspondence Analysis

References

1. H.T. Lawless. Exploration of fragrance categories and ambiguous odors using multidimensional scaling and cluster analysis. *Chemical Senses*, 14:349–360, 1989.