Hierarchical Multiple Factor Analysis: Application to the comparison of sensory profiles

S. Le Dien and J. Pagès

Laboratoire de mathématiques appliquées, ENSAR, F-35042 Rennes Cedex, France

Analysis of a set of variables structured according to a hierarchical tree

Partition structure. In the analysis of an individuals × variables table, variables are frequently structured into groups according to a partition. In sensory analysis, products are often described by both sensory descriptors and instrumental measures, for instance. Taking into account a partition structure on variables in a global analysis is a classic problem which can be handled by various statistical methods such as GPA, Multiple Factor Analysis (a brief introduction to MFA, as well as an application of the method, can be found in Pagès and Husson, 2001) and STATIS. In concrete terms, it implies balancing the groups in the analysis (weighting of the variables) and providing specific outputs (simultaneous display of the individuals according to each group, graphical display of the groups and numerical indicators). The benefits gained from these outputs are many folds, ranging from assessing relationships among the variables at hand to shedding more light on the similarities among individuals from different viewpoints.

Hierarchy structure. In many datasets, variables are structured according to a hierarchy leading to groups and subgroups of variables. In the example below, 5 dark chocolates were profiled by 6 trained sensory panels on the one hand and an untrained sensory panel on the other hand. Analysing such data implies balancing the part of the untrained panel as well as that of the trained panels all together on the one hand, but also that of the trained panels among them on the other hand. To do so, it seems necessary to consider the hierarchy represented in figure 1. The usual methods mentioned above do not suit this type of problem since they lead to outputs where a point of view of a group of variables may be preponderant in comparison to the point of view of other groups.

Taking a hierarchy into account. The approach to consider such a structure on the variables in a global analysis involves balancing the groups of variables within every node of the hierarchy. It is also necessary to have appropriate outputs as in particular a graphical display of each chocolate as profiled by the untrained panel, by the trained panels as a whole and by each trained panel in particular. Such graphical displays make it possible to investigate the performance of the assessors and to compare, product by product, the assessment of the untrained panel to the trained panels, which is one of the aim of the study.
Hierarchical multiple factor analysis

Balancing subgroups of variables within each node. Hierarchical multiple factor analysis (HMFA) is an extension of MFA to the case where variables are structured according to a hierarchy.

In MFA, we consider a partition on the variables: observations are described by variables partitioned into groups. A paramount feature of MFA is that it balances the importance of each group before performing principal component analysis on the merged groups. The weighting of the groups results in interesting properties that make it possible to depict several graphical displays showing the similarities among individuals from different point of view.

In HMFA, a succession of MFA is applied to each node of the hierarchy in order to balance the groups of variables within every node, by going through the hierarchical tree from the bottom up. In the example outlined above, we applied, in a first step, MFA to the chocolates profiled by the 6 trained panels in which each group of variables corresponds to a trained panel. In a second step, we applied MFA to the groups associated with the untrained panel on the one hand and with the trained panels considered as one group of variables on the other hand, the variables of this latter group being balanced according to the first step.

Superimposed representation. Not only HMFA provides a graphical display of the individuals according to the whole set of (weighted) variables, but it also displays the individuals as described by each group of variables (cf. figure 2): by definition, an individual which is described by just one group of variables is called a “partial individual”.

![Figure 2. Superimposed representation stemming from HMFA. a, ..., e denote the 5 chocolates; U (resp. T) denotes the partial individual associated with the untrained panel (resp. the trained panels all together); 1, ..., 6 denote the partial individuals associated respectively with one of the trained panel.](image)

It can be seen from figure 2 that:

- for each product, the representation of the partial individuals associated with the untrained panel (labelled $U$) and with the trained panels (labelled $T$) are to a certain extent close to each other;
- for each product, the representation of the partial individuals associated with the trained panels gives an idea of their variability. In a former study, Pagès and Husson (2000) established a satisfactory reproductibility of the same trained panels. HMFA allows to find the same result again: for each product, the partial individuals related to the trained panels are close to their barycenter;
- globally, the partial individuals associated with the untrained panel are positioned within the variability of the trained panels; let us notice that the untrained panel granted more importance to the second dimension than the trained panels.

Conclusion

In this example, where simple descriptors were considered, the untrained panel gave a profile of the products which was very close to that of the trained panels. To obtain a basic description of products, which is often sufficient in practice, a single session training can turn out satisfactory.
In practice, this structure of hierarchy on the variables is quite common: HMFA offers a straight and operational solution to this problem.

References