Analysis of sensory profiling data : from discrimination to description

M. Hanafi[1], E. M. Qannari[1], Ph. Courcoux[1], G. Organ[2]

[1]ENITIAA-INRA, Statistique appliquée, Rue de la Géraudière, 44072 Nantes, Cedex 03, France
[2]Lion Nathan, Locked bag 58, Silverwater, 1811, Australia

General introduction

Consider the situation where *m* data sets are obtained by means of sensory profiling of *n* products using a fixed vocabulary procedure with *p* attributes. Several procedures for analysing these data are proposed ranging from performing principal components analysis (*PCA*) on the average data set to carrying out methods akin to discriminant analysis by considering the products as categories. Another procedure which is intuitively appealing consists in merging vertically the m data sets thus forming a data table *X* with *nxm* rows and *p* columns, then *PCA* is performed on *X*. Thereafter, products are depicted using the average scores associated with the retained principal components.

It is clear that methods based on *PCA* tend to recover the total variability in the data whereas methods based on discriminant analysis aim at highlighting differences among products no matter whether these differences are reflected by variables with small variances or large variances.

We discuss a general approach of analysis which encompasses the three methods outlined above as particular cases and provide a range of intermediate situations which give insight into the profiling data from different perspectives. The approach bears some similarity to the concept of continuum regression introduced by Books and Stone (1990, 1994).

Method of analysis

Consider matrix *X* introduced above and suppose that its columns are centred. *PCA* performed on matrix *X* consists, in a first step, in seeking a vector of loadings, *a*, associated with the *p* variables such that $a^T Ta$ is maximized under the constraint $a^T a=1$; where $T=X^T X$. Similarly, *PCA* carried out on matrix *M* (average data set) consists in determining a vector *a* of loadings associated with the *p* attributes such that $a^T Ba$ is maximized under the constraint $a^T a=1$ where $B=M^T M$. Disciminant analysis consists in determining, in a firsts step, a vector of loading associated with the *p* variables which maximizes $a_1^T Ba_1 / a_1^T Ta_1$. Once the vectors of loading are computed, scores (principal components or discriminant functions) are computed allowing depiction/discrimination of the products. Subsequent scores may also be computed by considering the same optimisation criteria and imposing orthogonality constraints.

Let α be a positive scalar, we consider the following problem which consists in seeking a vector of loadings *a* such that $f_a(\boldsymbol{a})=(\boldsymbol{a}^T \boldsymbol{Ba})(\boldsymbol{a}^T \boldsymbol{Ta})^{a-1}$ is maximized under the constraint $a^T a=1$.

Each choice of a implies a particular strategy of analysis. For instance, if a is chosen between 0 and 1 then the stress is put on the discrimination of the products. From this perspective, it appears that the smaller a is, the heavier the stress is on the discrimination aspect. As an extreme case $a=0$ leads to discriminant analysis discussed above. When a is larger than 1 then obviously, the aim is to recover the total variability in the data. When a

tends to infinity, we retrieve *PCA* on data set *X*. The particular case $a=1$ leads to PCA on the average data set *M*.
We discuss a procedure of determining optimum vectors of loadings and we illustrate the method of analysis using real data sets.

References

- Brooks, R. J, Stone, M. (1994). Joint Continuum regression for multiple predictands. *JASA*, vol. 89, 428, 1374-1377
- Stone, M. Brooks, R. J. (1990). Continuum regression : Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *J. R. Statist. Soc,* B, 237-269.