Weighted PLS-Discriminant analysis with application to conventional sensory profiling

Stéphane Verdun Véronique Cariou El Mostafa Qannari

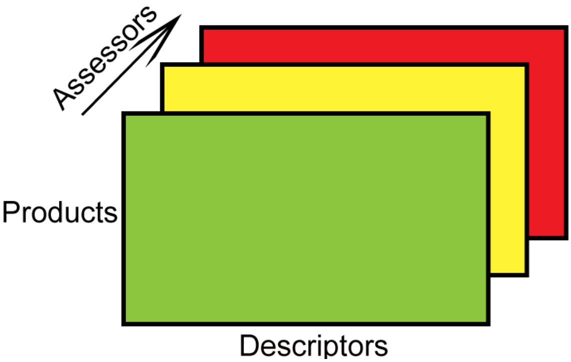
Sensometrics and Chemometrics Laboratory
Oniris, Nantes

10^{th} Sensometrics

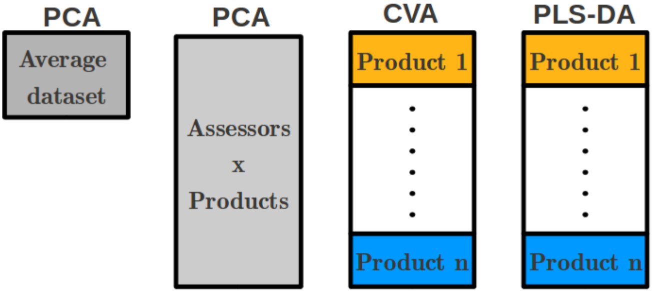
Overview

Conventional sensory profiling

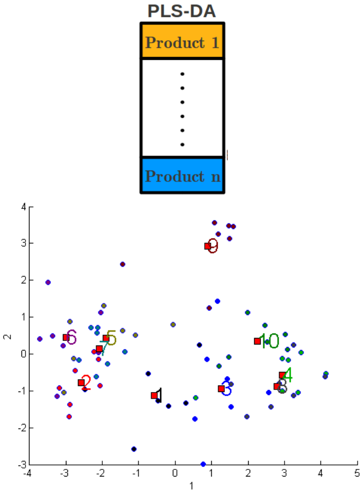
The assessors score the products for various descriptors, leading to the 3-way matrix X :



Popular methods



PLS-DA



- Focus on PLS-DA in this presentation
- Seek components which maximize the between product variation
- See poster by Rossini et al.

Assessors' performance

- A good performance ensures a good discrimination
- Which actions should be taken in case of bad performance ?
  - $\rightarrow$ Discard the assessors
  - $\rightarrow$ Downweight the assessors (Statis (Schlich, 1996), GPA (Qannari et al., 1999))

Overall downweighting of assessors

General strategy involves in GPA or Statis :

- Objective : find weights for the assessors, according to their agreement
- Compute the similarity matrix between the assessors
- Extract the first eigenvector
- Assign the components of the eigenvector as assessors' weights

Downweighting per case (assessor x product)

It may happen that an assessor has a good agreement with the panel except for one specific product
⟶ Downweight each case (assessor x product)

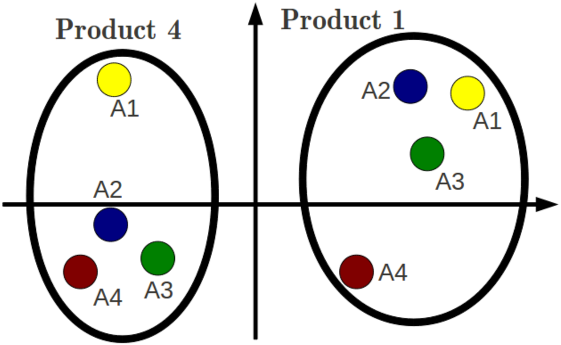


FIGURE: Product space

Within each product

- Compute a similarity matrix S
- Normalize to a stochastic matrix P
- Extract the dominant eigenvector
- Assign the components of the eigenvector as weights
- Justification : graph theory, Markov chains, *Reaching a consensus*, De Groot (1974)

Weighted PLS-DA

- The weights can be used to compute robust means, variances...
- The algorithm of weighted PLS-DA is the same as PLS-DA except that the means and the between products covariance matrix is computed using the weights

Example of similarities between two cases *i* and *j*

- Gaussian similarity :

$$s_{ij} = exp(-\frac{d_{ij}^2}{2\sigma^2}) \text{ where } \sigma \text{ is a tuning parameter}$$

- Proportion of common neighbours within the *k* nearest neighbours, *k* is a tuning parameter

How to tune the parameter σ or k (number of nearest neighbours) ?

Jacknife procedure (leave-one-out) on the assessors and choose a parameter σ or k that ensures the highest stability of a two (or three...) dimensional representation of the products (by means of PLS-DA).

Data

Data

- QDA experiment
- 10 varieties (ciders) evaluated according to 10 descriptors by 7 assessors

Factorial plane

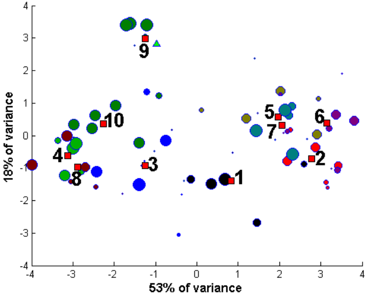


FIGURE: First factorial plane - Red squares are weighted means and the green triangle is a classic mean

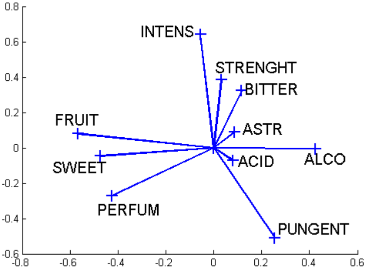


FIGURE: First factorial plane - Map of the variables

Example of weights

- Weights for some products and some assessors :

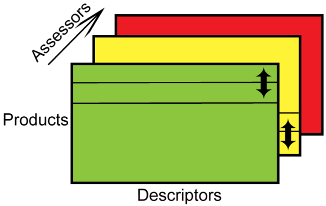
	P1	...	P3	...	P5	P7	P9	Mean weight
A1	0,063	...	0,086	...	0,183	0,056	0,143	0,132
A2	0,055	...	0,195	...	0,100	0,204	0,206	0,157
A3	0,162	...	0,116	...	0,100	0,167	0,206	0,165
...	...	...	...	...	...	...	...	...
A5	0,169	...	0,143	...	0,176	0,000	0,048	0,093

TABLE: Example of weights and mean weights for some assessors and some products

- The assessors' capability over all products can be measured by the mean of the weights, but there is a loss of information

Stability of the factorial plane

- A perturbation is introduced in the data by permutating the answers of an assessor



- We have instances of (local) disagreement involving product 3
- The same with the product 9

Confidence ellipses

Perturbated products : products 3 and 9

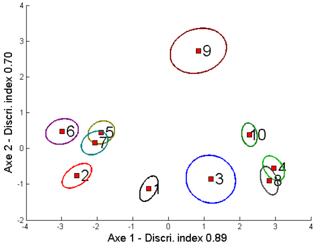


FIGURE: First factorial map, without weights

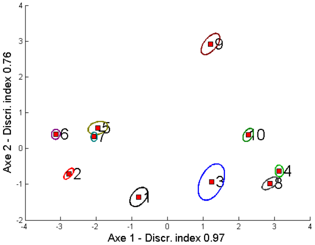


FIGURE: First factorial map, neighbourhood similarity

Conclusion

- Better insight into the assessors' performance
- The weighting strategy improves the stability of the factorial plane, leading to more robust representations of the products
- The weighting strategy is flexible (use of different similarities) and versatile (use within different factorial methods)
- The parameters of the similarities can be tuned according to different objectives : stability, discrimination...

De Groot, M.H. (1974). Reaching a consensus. *J. Amer. Statist. Assoc.*, 69. pp. 118 – 121

Schlich P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. In T. Naes & E. Risvik (Eds.) *Multivariate analysis of data in sensory science*. Elsevier Science Publishers

Qannari, E.M., MacFie, H.J.H, Courcoux, P. (1999). Performance indices and isotropic scaling factors in sensory profiling. *Food Quality and Preference*, 10. pp 17 – 21