

GPA-PLSR: An integrated technique for analysing sensory science data

Garnt Dijksterhuis, Harald Martens, Magni Martens

KVL, Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Sensory Science Group, Rolighedsvej 30, DK 1958 Frederiksberg C., Denmark

We aim to merge two different data analytical methods: Generalised Procrustes Analysis (GPA, Gower 1975) and Partial Least Squares Regression (PLSR, Wold et al. 1983). The merger is such that the criterion of inter-subjectivity, inherent in sensometric use of GPA, and the criterion of correspondence to physics and chemistry, inherent in sensometric use of PLSR, are combined.

GPA works on K data sets X_k , $k=1, \dots, K$. They correspond to one data table for each sensory assessor; rows corresponding to products, columns to sensory terms. In the first step of GPA, the input data are modified so that systematic between-assessors variation (level effects, range effects, rotation effects) are removed. The second step is to form a group average. Finally the main structure is extracted by a Principal Component Analysis (PCA). PLSR analyses the relation between two data sets, Y and X . In sensometrics one is free to choose what to put in X and in Y . But an explorative way to use PLSR, called DPLSR (Martens & Martens 2000), is to define X as sensory data (e.g. average over the K assessors) and let Y be design facts and physical/ chemical measurements. In the first step in PLS the X -variables are compressed into a few Y -relevant PCs. The second step is to model both X and Y in terms of these PCs. It can so happen that an individual data set X_k is off w.r.t. the other sets but in accordance with the external information in Y . In that case this X_k should get the opportunity to make itself heard. This will be accomplished by the PLSR of the X_k 's towards the external data Y . Analogously, when the majority of sets X_k , are in strong agreement with each other, they will together mainly shape the group average, and the effect of the PLSR of Y should be outweighed by them. Merging GPA and PLSR means bridging two statistical traditions. In GPA the main modelling takes place in the full dimensional space where no noise reduction takes place; only at the very end of the analysis a PCA rank reduction may be performed to inspect the results. In the PLSR tradition, the rank reduction is done as part of the main modelling. Our goal for the combined GPA-PLSR technique is to have a way to extract reliable and relevant information from sensory panel data, by obtaining a group average which strikes a balance between the two 'truth' criteria:

- Inter-subjective agreement among the K assessors indicates reliable sensory data. This "democratic, cultural process" may correspond to how social scientists think of valid structure in data
- Correspondence to external design or to physics and chemistry data indicates reliable sensory data. This may correspond to the more or less "causal" thinking of the natural scientist.

References

- Gower, J.C. (1975). Generalized Procrustes analysis, *Psychometrika*, 40, 33-51.
- Martens H. and Martens M (2000). *Multivariate Analysis of Quality. An Introduction*. J.Wiley & Sons, Ltd.
- Wold, S., Martens, H. and Wold, H. (1983). The Multivariate Calibration Problem in Chemistry solved by the PLS Method. *Proc. Conf. Matrix Pencils*, (A. Ruhe and B.

Kågström, eds.), March 1982, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, 286-293.