Food Quality and Preference 40 (2015) 326–333

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies

C. Peltier*, M. Visalli, P. Schlich

CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France
INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France
Université de Bourgogne, UMR Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form 5 May 2014
Available online 16 May 2014

Keywords:
CVA
PCA
Meta-analysis
SensoBase

ABSTRACT

Although Principal Component Analysis (PCA) of product mean scores is most often used to generate a product map from sensory profiling data, it does not take into account variance of product mean scores due to individual variability. Canonical Variate Analysis (CVA) of the product effect in the two-way (product and subject) multivariate ANOVA model is the natural extension of the classical univariate approach consisting of ANOVAs of every attribute. CVA generates successive components maximizing the ANOVA F-criterion. Thus, CVA is theoretically more adapted than PCA to represent sensory data. However, CVA requires a matrix inversion which can result in computing instability when the sensory attributes are highly correlated.

Based on the analysis of 422 descriptive sensory studies gathered in SensoBase (www.sensobase.fr), this paper compares the maps obtained by covariance PCA and CVA, both performed on significant ($p = 0.1$) attributes for the product effect. Results suggest that 9 times out of 10, PCA and CVA maps are very similar. However, differences between these maps increase with product space complexity. It is thus recommended to analyze and map each sensory modality (texture, aroma...) separately.

Introduction

Sensory profiling is the process by which a panel of trained panelists scores the perceived intensities of a number of sensory attributes on a number of products possibly with replications. The resulting data are usually analyzed with Analysis of Variance (ANOVA) to detect differences between products for each attribute taken separately and then with Principal Component Analysis (PCA) to get a product map based on all the attributes simultaneously. Whereas ANOVA takes into account the subject variability around the product means to assess product differences, PCA draws the product map based on product means with no consideration of individual variability. Multivariate Analysis of Variance (MANOVA) is the natural extension of ANOVA to several variables analyzed simultaneously. It allows us to assess statistical significance of product differences in the space generated by the set of attributes being scored. Canonical Variate Analysis (CVA) is a mapping method (cf. Appendix A) based on MANOVA which derives successive components on which the products are as much discriminated as possible. The meaning of discrimination is to separate product means as much as possible, while individual assessments of a given product are clustered as much as possible close to its product mean. CVA is thus theoretically more adapted than PCA for the multivariate analysis of profiling data. Indeed, Heymann and Noble (1989), Monrozier and Danzart (2001), Noble, Williams, and Langron (1984), Porcherot and Schlich (2000) and Schlich (2004) recommended CVA for the analysis of sensory descriptive data.

Only few comparisons of the two mapping methods have already been reported so far. Martinez and Kak (2001) proved that Linear Discriminant Analysis (a one-way model of CVA) was not always better than PCA to recognize faces. In sensory analysis, Heymann and Noble (1989) showed that PCA and CVA maps provided similar results using wine datasets. Brockhoff (2000) emphasizes the importance of taking into account the variability around the product means with 39 sensory datasets and consequently claimed that CVA was better than PCA. Finally, Monrozier and Danzart (2001) compared PCA, CVA and other different MANOVA tests with resampling and concluded that PCA was very robust whereas complex MANOVA models led to unstable results. These comparisons were carried out with only a few sensory datasets.

This paper presents a comparison of the maps obtained by PCA and CVA on 422 sensory profile datasets. The first objective of this

* Corresponding author Address: Centre des Sciences du Goût et de l'Alimentation, 9E boulevard Jeanne d'Arc, F-21000 Dijon, France. Tel.: +33 3 80681676.
E-mail address: caroline.peltier@dijon.inra.fr (C. Peltier).

http://dx.doi.org/10.1016/j.foodqual.2014.05.005
0950-3293/

paper is to quantify the differences between PCA and CVA observed on these numerous datasets. For this purpose, criteria for comparing PCA and CVA have been defined. The second objective of this paper is to explain potential differences between PCA and CVA maps in function of the parameters of the datasets (number of products, number of attributes…). Our hypothesis is that the more complex the product space, the more different the maps. The complexity of the product space corresponds to its dimensionality which depends on the number of attributes and their level of correlation (the larger the correlations, the smaller the dimensionality). Consequently, decomposing the space into subspaces by sensory modality (texture, smell, taste…) is not only intuitively appealing but could also lead to a better agreement between PCA and CVA maps. Such decomposition has been undertaken for datasets showing the largest difference between PCA and CVA maps.

Material and methods

Selection of datasets and attributes

SensoBase (www.sensobase.fr) was created (Pineau, 2006) and maintained by the laboratory of the authors of the present paper. It is a sensory profile database containing 1107 datasets. 422 datasets presenting the following characteristics were extracted from this database:

- Including between 3 and 20 products.
- Including more than 4 subjects.
- Having at least 2 significant ($p = 0.1$) attributes for the product effect.
- Being balanced (same number of evaluations of each product by subject).

These 422 datasets totalized 2130 subjects, 2605 products, 7217 attributes and 1,425,056 observations (scores). For each dataset, covariance PCA and CVA were computed with a R-package developed by the corresponding author. The section ‘Comparison with correlation PCA’ will address whether a correlation PCA would have given the same results or not. In order to reduce noise in the data, only significant ($p = 0.1$) attributes for the product effect in the two-way ANOVA were included in the analysis.

Number of dimensions used in PCA and CVA

In CVA, a statistical test (Mardia, Kent, & Bibbly, 1994) allows us to find the number D of significant dimensions of the product space (cf. Appendix B). Consequently, the first D axes of the CVA theoretically account for all relevant information, the other axes representing only noise. However, many users often prefer to limit their interpretation to the two first axes. Consequently, CVA and PCA plots were compared on the first two dimensions (1–2 comparison), but also in the product space generated by the first D dimensions (1-D comparison).

Map characteristics

CVA or PCA plots are composed of a product representation and an attribute representation. Fig. 1 is an example of CVA plots from a dataset of the SensoBase, having 6 products and 4 significant attributes: Sweet, Salty, Sour and Dry.

In order to be as exhaustive as possible in the comparison, the different elements composing CVA and PCA plots have been listed. They are the same in PCA and CVA plots.

The plot on the left is the representation of the **product configuration**, with points indicating product means. Distances between products represent differences between them. In the example, P_1 appears different from P_5.

The plot on the right is the representation of attribute configuration, with arrow coordinates being correlation coefficients between attributes and components. The joint interpretation of the two graphs provides a **sensory interpretation**. A product with a large coordinate on the first axis is likely to have large scores on the attributes strongly positively correlated with the first axis. For example, P_5 and P_6 should have larger Salty scores than other products.

The ellipses on the product plot represent the variability of the subject scores around the mean. They delimitate an area where the product means have a 90% probability to be truly located. The overlapping of ellipses gives indications on **pairwise product comparisons**. For example, since P_1 and P_6 ellipses are well separated, they are expected to be different at least on the subspace generated by the first two axes. Note that confidence ellipses can also be built in PCA. Although PCA was computed from the product mean scores, it is possible to project the individual scores as supplementary

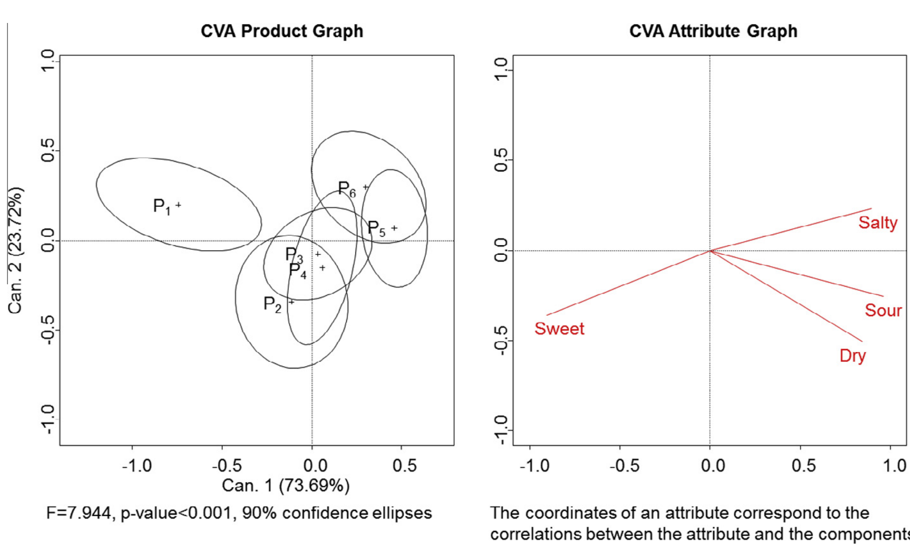

Fig. 1. Example of CVA plots (on the first two components).