Use of symbolic principal component analysis to account for imprecision within the sensory map of products

Cordier C.[1], Grandière C.[1], Coulon-Leroy C. [2], Cariou V. [1]

[1] *ONIRIS, Rue de la Géraudière, BP 82225 44322 Nantes Cedex 03, France*
@ cynthia.cordier@oniris-nantes.fr, claire.grandiere@oniris-nantes.fr, veronique.cariou@oniris-nantes.fr

[2] *Unité de Recherche GRAPPE, Univ Bretagne Loire, Ecole Supérieure d'Agricultures (ESA)-INRA, 55 rue Rabelais, BP 30748, 49007 Angers Cedex, France.*
E-mail : *c.coulon@groupe-esa.com*

Abstract
This study fits within the general framework of conventional sensory profiling where a panel is used to determine the characteristics of several products in terms of their sensory properties. We consider, herein, an experimental design where on the one hand the panel is an untrained panel consisting of students in food science and on the second hand the panel is trained (expert panel). To take account of the panelist's possible difficulties in the sensory rating task, we propose to introduce a fuzzy rating through a triangular measure. This latter one makes each panelist possible to define the standard value associated to each product for each sensory descriptor both with an interval reflecting his lack of precision. Herein, we study the impact of the training level on the imprecision both with the representation of the imprecision on the product map. In order to provide a sensory map of products on the basis of fuzzy data, a symbolic principal component analysis is considered. Symbolic principal component analysis provides both the sensory map of products and the representation of the imprecision expressed by the panelists through hyper-rectangles. This approach is illustrated with two datasets pertaining to the evaluation of compotes.

Keywords: Sensory analysis; Symbolic data**;** Fuzzy data; Principal Component Analysis

1. Introduction

This study fits within the general framework of sensory profiling. Classical sensory evaluation is generally performed using conventional sensory (AFNOR, 1995) profiling involving a panel of trained assessors. The establishment of the panel is binding: each assessor is enrolled after three selection stages and he has to master vocabulary of product space. Alternatively, several methods such as Free Choice Profiling and Flash Profiling involve the use of untrained consumers rather than a trained panel making these techniques faster and cheaper to conduct.
In the scope of conventional sensory profiling, we consider, herein, the use of an untrained panel and compare it with a trained one. The untrained panel is a "semi" naive one since it consists of students in food science. "Untrained" means herein that the assessors are not accommodated with the product space. Because of the lack of specific training, assessors may face difficulties with products rating involving some imprecision. In order to account for their imprecision, we propose to introduce a fuzzy rating. In such a rating, each consumer is asked to rate a product according to a descriptor with a standard value both with an interval surrounding this standard value. Several studies have already been made to deal with imprecision and variability, among which we can cite symbolic and fuzzy methods. Hence, the aim is to study the relevancy of fuzzy and symbolic approaches in the scope of

conventional sensory profiling task with an untrained panel. Moreover, we focus in the determination of a sensory map of products accounting for imprecision. This sensory map is obtained by the use of symbolic principal component analysis. In the next section, a description of two symbolic methods is proposed and a comparison is performed on the basis of a compotes dataset.

2. Materials and methods

2.1 Datasets

This study was led using two datasets: the first one consisted in a sensory evaluation with a trained panel, composed by fourteen assessors while the other one was based on the evaluation of a panel of thirty-four students in food science without previous common training. These datasets came from ESA (Ecole Supérieure d'Agriculture). Each assessor had to rate eight compotes having different acidity and sugar rates during two sessions with three repetitions. The evaluation has been performed on the basis of nine descriptors: fluidity, granularity, sugar, acidity, bitterness, global interaction between aromas, raw apple flavor, cooked apple flavor and finally oxidized apple flavor. In order to take account of the lack of training, each assessor was asked to give both a standard value and if necessary an interval (minimum and maximum value) reflecting his lack of precision. Thus, for each descriptor the judge give three notes, leading to a triangular observation (see figure 1).

Figure 1: Triangular rating

2.2 Fuzzy principal component analysis on triangular data

2.1.1 PCA-TF: Principal Components Analysis for Trapezoidal Fuzzy Numbers

This method is an extension, proposed by Pacheco (2010), which is based on Centered Principal Component Analysis (C-PCA). C-PCA is a principal component analysis on intervals data proposed by Cazes & al. (1997). Principal Components Analysis for Trapezoidal or Triangular Fuzzy Numbers (PCA-TF) is performed on numbers defined by three or four figures which define the support and the core of these fuzzy numbers.

For example, a trapezoidal fuzzy number Y is defined by $Y = (Y^{(1)}, Y^{(2)}, Y^{(3)}, Y^{(4)})$ with the following membership function (Pacheco, 2010):

$$\mu_Y(x) = \begin{cases} 0 & \text{if } x < Y^{(1)} \\ \frac{x - Y^{(1)}}{Y^{(2)} - Y^{(1)}} & \text{if } Y^{(1)} \leq x \leq Y^{(2)} \\ 1 & \text{if } Y^{(2)} \leq x \leq Y^{(2)} \\ \frac{Y^{(4)} - x}{Y^{(4)} - Y^{(3)}} & Y^{(3)} \leq x \leq Y^{(4)} \\ 0 & \text{if } Y^{(4)} < x \end{cases}$$

With $\sup(Y)=[Y^{(1)},Y^{(4)}]$ and $\text{core}Y=[Y^{(2)},Y^{(3)}]$

For triangular fuzzy numbers $Y^{(2)} = Y^{(3)}$.

The initial dataset is "defuzzified" such that each triangular observation is recoded by a single value: $X_E=\frac{Y^{(1)}+2Y^{(2)}+Y^{(4)}}{4}$.

Hence, a new matrix is created containing center of intervals (X_E) for each descriptor, each judge and each repetition. Then a classical PCA is performed on this matrix like in C-PCA (Cazes and al, 1997). A graphical representation of PCA-TF results is obtained depicting the standard value associated to each product both with a rectangle reflecting imprecision around this standard value. Rectangles correspond to the minimum envelop containing all the component scores for vertices associated with the triangular initial values.

2.1.2 LMR-PCA

Another method to analyze fuzzy data is LMR-PCA (Coppi & al., 2006) which doesn't involve the recoding of the triangular data into a single valued data like PCA-TF. LMR-PCA deals with a general class of fuzzy data represented by "LR" fuzzy data. Specifically, each observation is characterized by its center both with its left and right spreads. Then three matrices are obtained: M contains standard values, L contains left spreads and R contains right spreads.
This method aims at recovering the underlying structure of fuzzy data taking into account standard and interval values. It is based on the minimization of the criterion ä:

$$ä = (2^J+1)\|M-M*\|\ 2 + 2^J\text{-}1\|L-L*\|\ 2 + 2J\text{-}1\|R-R*\|\ 2 - 2J\mathrm{tr}[(M-M*)T\ (L-L*)]+2^J\ \mathrm{tr}[(M-M*)T\ (R-R*)]$$

Where J is the number of descriptors and M*, L* and R* are the estimated matrices of M, L and R. The estimation procedure is based on an iterative gradient descent procedure which produces scores and loadings as in PCA. The graphical representation of LMR-PCA is a factorial map obtained in a similar way than PCA-TF.

3. Results

PCA-TF and LMR-PCA have been performed on the two different datasets. In this section, we focus on the comparison of C-PCA with a classical PCA applied on standard values discarding imprecision.

3.1 PCA on standard values

Before considering imprecision, a first PCA has been performed on standard values discarding intervals.

3.1.1 Untrained panel.

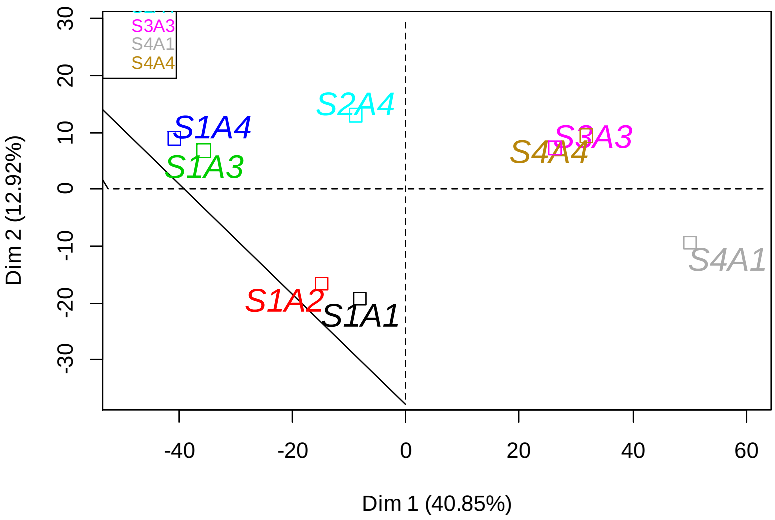

Figure 2: Product map

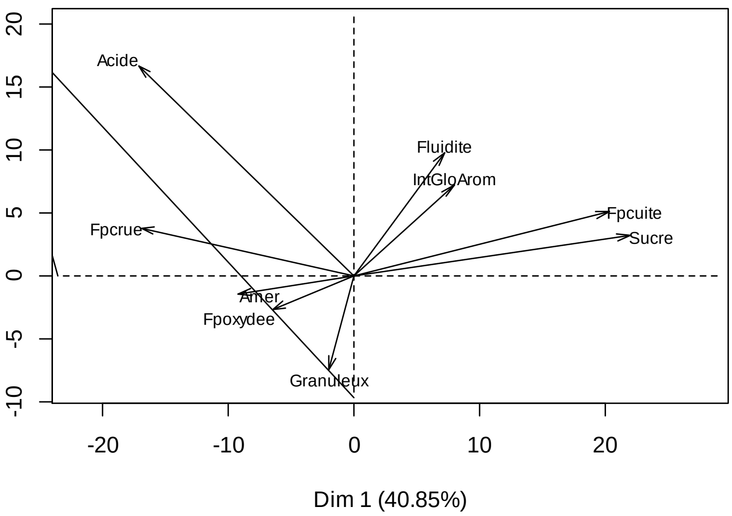

Figure 3: Loadings map

On the product map (see figure 2), we can see that the first dimension describes the intensity of sugar and the second dimension is most correlated with acid intensity.

3.1.2 Trained panel

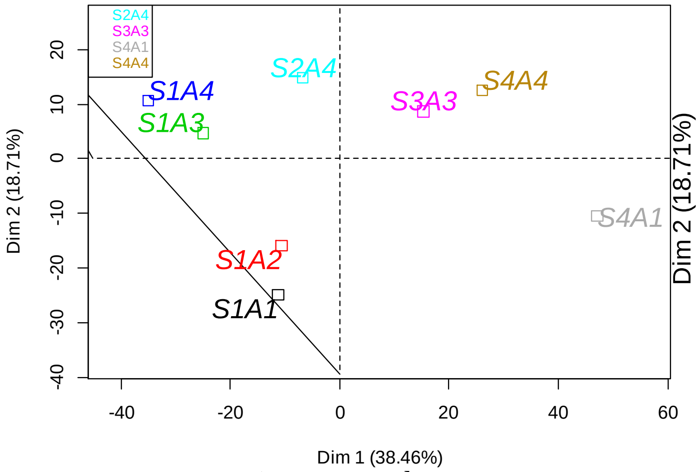

Figure 4: Product map

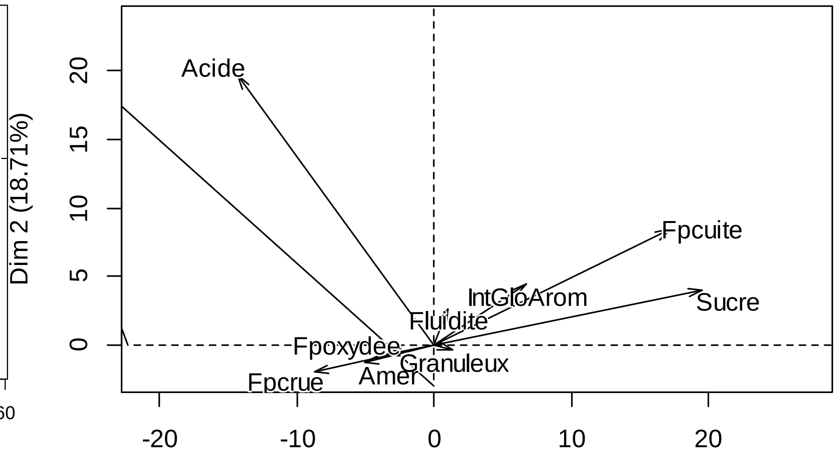

Figure 5: Loadings map

As for the untrained panel, the first dimension describes the intensity of sugar and the second dimension is most correlated with acid intensity (see figure 4). Acidity and sugar descriptors are uncorrelated.

3.2 C-PCA product maps

C-PCA results are depicted in figures 6 and 7.

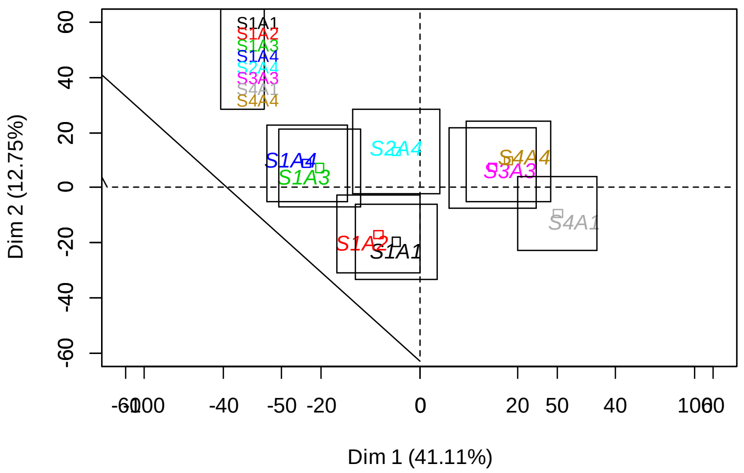

Figure 6: Product map: untrained panel

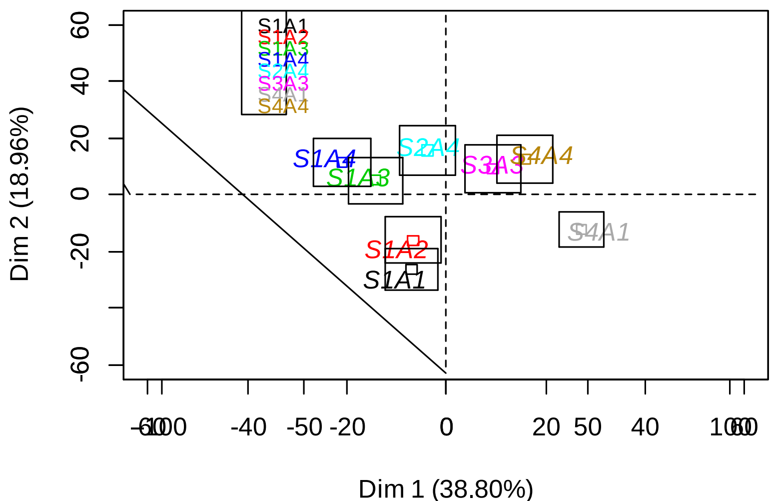

Figure 7: Product map: trained panel

Product maps (figs 6-7) show that trained-panel intervals are closer to centers than untrained-panel ones and groups are better defined. For the untrained panel, accounting imprecision in C-PCA provides a slightly better separation of the products on the second axis compared to PCA on standard values.

4. Conclusion

In this first part of the study, we focus on imprecision visualization thanks to a method for recovering the underlying structure of fuzzy interval data. Imprecision is provided through rectangles surrounding each standard value. More research is needed to further explore the accounting of imprecision in the rating task. For example, indicators need to be defined in order to assess whether those kinds of methods are useful to perform conventional sensory profiling with an uncommonly trained panel.

References

AFNOR, Association française de normalisation, 1995. NF ISO 11035: Analyse sensorielle - Recherche et sélection de descripteurs pour l'élaboration d'un profil sensoriel, par approche multidimensionnelle. Paris: AFNOR, 26 p.

Cazes P., Chouakria A., Diday E., Schektman Y. (1997). Extension de l'ACP à des données de type intervalle. *Revue de Statistique Appliquée*, XLV (3), pp. 5-24.

Coppi R., Giordani P., D'Urso P. (2006). Component Models for Fuzzy Data. *Psychometrika*, 71, pp. 733-761.

Coulon-Leroy C., Alonso Moral J.M., Symoneaux R., Sutton-Charani N., Montmain J. (2015). Consideration of uncertainty during data acquisition and processing chain on sensory data. 11th Pangborn Sensory Science Symposium. Gothenburg, Sweden.

D'Urso P., Giordani P. (2004). A least squares approach to principal component analysis for interval valued data. *Chemometrics and Intelligent Laboratory Systems*, 70, pp. 179-192.

Giordani P., Kiers H.A.L. (2006). A comparison of three methods for principal component analysis of fuzzy interval data. *Computational Statistics & Data Analysis*, 51, pp. 379 – 397.

Pacheco A., Rodríguez O. (2010). *Principal Components Analysis for Trapezoidal Fuzzy Numbers*. In: Locarek-Junge H., WEIHS C., Classification as a Tool for Research (pp. 351-360). Heidelberg: Springer-Verlag.