Food Quality and Preference 32 (2014) 56–64

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

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

Multiblock modeling for complex preference study. Application to European preferences for smoked salmon

Bougeard Stéphanie [a,*], Cardinal Mireille [b]

[a] Anses (French Agency for Food, Environmental and Occupational Health Safety), Department of Epidemiology, F-22440 Ploufragan, France

[b] Ifremer (French Research Institute for Exploration of the Sea), Laboratoire Science et Technologie de la Biomasse Marine, F-44311 Nantes 03, France

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form 30 January 2013
Accepted 7 June 2013
Available online 14 June 2013

Keywords:
Multiblock modeling
Multiblock redundancy analysis
External preference mapping
Smoked salmon

ABSTRACT

The aim of the paper is to propose an alternative method to external preference mapping for the case of complex data where explanatory variables are organized in meaningful blocks. We propose an innovative method in the multiblock modeling framework, called multiblock Redundancy Analysis. The interest and relevance of this method is illustrated on the basis of a European consumer preference study for cold-smoked salmon. The study aims at explaining six homogeneous clusters of preference with explanatory parameters organized in five thematic blocks related to physico-chemical measurements, microbiological characterization, appearance attributes, odor/flavor characterization and texture descriptors. Overall indexes and graphical displays associated with different interpretation levels are proposed to sort the key drivers of preference by order of priority at the variables and at the block level. On the basis of these data, multiblock Redundancy Analysis is also compared to standard preference mapping in terms of model quality; the best model is here associated with the multiblock method.

1. Context

Product development is often based on external preference mapping (prefmap) where sensory profiles are used to model consumer likings, all the variables being measured on the same products. This approach gives a reliable basis for creating products which correspond to consumer expectations. In this framework, we mainly focus on identifying the key drivers which impact the consumer preferences. External preference mapping assumes that consumers have a common perceptual space and that it can be modeled with sensory data (Jaeger, Wakeling, & Macfie, 2000). It is worth noting that other parameters are usually measured on products such as physical and chemical measurements, price and packaging descriptions. In order to improve the preference modeling and then get an overall vision of the preferred products or of the products to be developed, it is of paramount importance to explain consumer preferences not only with sensory attributes but also with these additional parameters. This could improve one of the main criticisms of prefmap, namely the poor model quality due to a product attribute space inadequate to the preference one. As a way to enhance modeling quality, we focus on external preference mapping applied not only to sensory but also to external attributes. This problematic is related to the explanation of a composite dataset, *i.e.*, the consumer preferences ($\mathbf{Y}$) with explanatory variables organized in several meaningful blocks, *e.g.*, sensory attributes ($\mathbf{X_1}$), physico-chemical parameters ($\mathbf{X_2}$) and packaging description ($\mathbf{X_3}$). All these variables are measured on the same observations, *i.e.*, the products under study, as illustrated in Fig. 1.

For the time being, three data processing solutions remain for the user to take account of the explanatory block structure. (i) The widely used solution consists in linking at first sensory attributes to preferences with a two-block method such as Partial Least Squares (PLS). In a further stage, other measurements are linked to preferences to get a more accurate characterization (Semenou, Courcoux, Cardinal, Nicod, & Ouisse, 2007). But this leads to a sequential resolution where preferences are actually only explained with sensory attributes. (ii) The second kinds of methods pertain to the field of Structural Equation Modeling–PLS Path Modeling (PLS-PM) being the well-known in sensometrics-initially developed for more complex data (Pagès & Tenenhaus, 2001). This method brings information on links between blocks achieved with the inner model but these coefficients can only be given separately for each dimension under study. Nevertheless, models are usually multidimensional, especially in biological fields. In addition, PLS-PM is based on a complicated iterative algorithm with any formal convergence proofs (Henseler, 2010). (iii) Finally, some multiblock methods, such as Parallel Orthogonalized PLS (Måge, Menichelli, & Naes, 2012) are proposed. The PO-PLS method is especially developed for external preference mapping and focuses on common and unique information in each

* Corresponding author. Tel.: +33 296 010 150.
E-mail address: stephanie.bougeard@anses.fr (B. Stéphanie).

0950-3293/$ - see front matter
http://dx.doi.org/10.1016/j.foodqual.2013.06.002

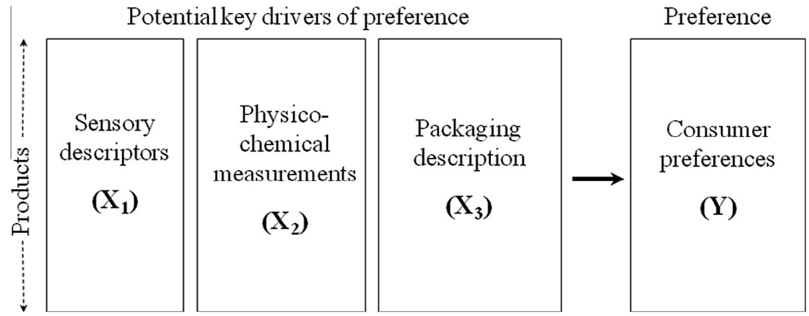

Fig. 1. Example of multiblock explanatory data which aims at explaining consumer preferences.

block. But the iterative algorithm and its complexity restrict the use of this method for more than two blocks (Måge et al., 2012). In this paper, we will stand in this interesting latter framework of multiblock methods while proposing an alternative approach with a direct eigensolution.

Among methods pertaining to the multiblock $(K + 1)$ setting, we single out those which are based on an optimization criterion that reflects the objectives to be addressed and leads to a direct eigensolution. Three methods which meet these constraints are available: Generalized Canonical Analysis with a Reference Table, GCA-RT (Kissita, 2003), multiblock Redundancy Analysis, mbRA (Bougeard, Qannari, & Rose, 2011) and multiblock Partial Least Squares, mbPLS (Wold, 1984). The method GCA-RT, is interesting from a theoretical point of view but may lead in practice to unstable model in case of quasi-collinear variables. The method mbPLS is a helpful and popular method in regards with its stability in case of multicollinearity but leads to solutions often not much linked to the dependent dataset. In addition, for our particular case of a single dataset to be explained, mbPLS leads to a simple PLS of **Y** and the merged dataset **X** (Westerhuis, Kourti, & MacGregor, 1998). We focus afterward on multiblock Redundancy Analysis which appears to take account of the multiblock structure of data and to lead to a model with a good fitting ability in spite of its lack of stability in case of high quasi-collinear variables (Bougeard & Qannari, 2011). Our purpose is to apply this original multiblock method to external preference mapping. This can be viewed as an extension of external preference mapping to the multiblock framework. Several interpretation tools pertaining to the field of factorial analysis and modeling are provided to further investigate the relationships among variables and datasets (Bougeard et al., 2011). All these methodological contributions are presented in Section 2. The interest of multiblock modeling analysis is illustrated on the basis of a European preference study of smoked salmons in Section 3. A discussion both on method and application is proposed in Section 4.

2. Material and method

2.1. Multiblock data and aims: European preferences for smoked salmon

The interest of multiblock modeling is illustrated on data from a European project (Adriant & IMR, 2004). A preference study is conducted on thirty smoked salmon, representatives of the market range (Cardinal et al., 2004) tested by 1063 consumers. As the individual preferences of the 1063 consumers are not uniform, homogeneous clusters of hedonic assessments are provided through a latent class vector model (Semenou et al., 2007). Six clusters, respectively containing 121, 74, 349, 78, 404 and 37 consumers, are highlighted. For simplicity sake and as often in external preference mapping, we take account of homogeneous clusters instead of individual likings. It follows that the quantitative dependent dataset **Y** involves thirty observations (salmons) described by six variables (clusters of preferences), each salmon being described by the associated cluster preference average. Physical, chemical, microbiological and sensorial measurements are carried out on the same salmons. We choose to organize these forty-four explanatory parameters in five thematic blocks related to the physico-chemical measurements ($\mathbf{X_1}$ dataset, 13 variables), the microbiological characterization ($\mathbf{X_2}$, 6 variables), the appearance attributes ($\mathbf{X_3}$, 6 variables), the odor and flavor characterization ($\mathbf{X_4}$, 14 variables) and the texture attributes ($\mathbf{X_5}$, 5 variables) (see description in appendix). As all variables are expressed in non-comparable range of measurements, they are column centered and scaled to unit variance. However, it is worth noting that as the variables have been standardized, the total variance in each block is equal to the number of variables in this block. This motivates the block weighting to put the blocks to the same footing (Westerhuis & Coenegracht, 1997). Each of the $(K = 5)$ explanatory block is accommodated with an isotropic scaling factor to set them to the same total variance, chosen equal to $1/K$. Therefore the merged explanatory dataset **X** (resp. **Y**) has a total variance equal to one.

These data have already been processed from many different ways (Cardinal et al., 2004; Courcoux, Qannari, & Schlich, 2006; Semenou et al., 2007). The latter authors propose internal and external preference mapping, the physico-chemical variables being related to preferences in a second step. We propose to consider the whole data, namely preferences, sensory analysis, physico-chemical and microbiological measurements in a single analysis. The first aim is descriptive and consists in explaining the consumer preferences in relation to all the explanatory variables organized in thematic blocks. This leads to two main questions:

Q1. Are there any relationships between the smoked salmon clusters of preferences $\mathbf{Y} = (y_1 \ldots y_6)$ and the external salmon attributes $\mathbf{X} = (x_1 \ldots x_{44})$?

Q2. Do the thirty smoked salmons have the same features in terms of their external description (**X**) in relation to preferences (**Y**)?

The second and pivotal aim is devoted to assess the key drivers of the salmon preferences from the forty-four external attributes. Three questions pertaining to the modeling framework can be asked:

Q3. Are there significant links between all the variables describing the external attributes $\mathbf{X} = (x_1 \ldots x_{44})$ and each clusters of preferences $\mathbf{Y} = (y_1 \ldots y_6)$?

Q4. Is it possible to sort by order of priority all the external variables describing the smoked salmons $\mathbf{X} = (x_1 \ldots x_{44})$ in relation to the overall preferences (**Y**)?

Q5. Is it possible to sort by order of priority the various external blocks $(\mathbf{X_1} \ldots \mathbf{X_5})$ in relation to the overall preferences (**Y**)?