Food Quality and Preference 32 (2014) 198–209

Contents lists available at ScienceDirect

Food Quality and Preference

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

Interpretation, validation and segmentation of preference mapping models

Isabella Endrizzi [a,*], Flavia Gasperi [a], Marit Rødbotten [b], Tormod Næs [b]

[a] Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund Mach (FEM), Via E. Mach, 1, 38010 San Michele all'Adige, Italy
[b] Nofima Mat AS, Osloveien 1, NO-1430 Ås, Norway and Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, 1958 Fredriksberg Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 21 February 2013
Received in revised form 3 September 2013
Accepted 3 October 2013
Available online 12 October 2013

Keywords:
Individual differences
Preference mapping
ANOVA
PCA
PLS-DA
Validation

ABSTRACT

In this paper we discuss an extension to preference mapping of the method proposed in [Endrizzi, I., Menichelli, E., Johansen, S. B., Olsen, N. V., & Næs, T. (2011). Handling of individual differences in rating-based conjoint analysis. Food Quality and Preference, 22, 241–254] for accommodating both population averages and individual differences in the same model. The method, based on average estimates and residuals, is a combination of ANOVA, PCA and PLS-DA, which are well-known techniques that can be run in almost all statistical software packages. Main attention is given to the relation between the double-centred residual matrix which highlights differences between consumers in their relative position as compared to the average consumer values and the standard centring in preference mapping. This approach has been found particularly useful for highlighting differences in preference pattern among the consumers. Furthermore, the interpretation and the segmentation, that is here taking place based on differences in acceptance pattern, are graphically oriented. In addition, some possible alternatives to the generally used validation method in PCA are suggested. The approach is then illustrated using two data-sets from consumer studies of apple and raspberry juice, showing that when individual differences are analysed by the present method, interesting results regarding individual differences in response pattern are detected.

1. Introduction

Preference mapping (Chang & Carroll, 1969; Carroll, 1972) is an important and much used methodology for modelling, analysing and understanding consumer preferences and their relation to product characteristics. Because of its attractive properties, preference mapping has been used for a number of different purposes, for instance to identify sensory drivers of liking (Michon, O'Sullivan, Sheehan, Delahunty, & Kerry, 2010; Sinesio et al., 2010), to find the best product composition (Endrizzi, Pirretti, Calò, & Gasperi, 2009; Felberg, Deliza, Farah, Calado, & Donangelo, 2010) and as a method for product optimisation (Busing, Heiser, & Cleaver, 2010; Ares, Varela, Rado, & Giménez, 2011). The most used methods in the area are the internal and external linear preference mapping methods (McEwan, 1996; van Kleef, van Trijp, & Luning, 2006). The main advantage of these techniques is that they are simple to use and interpret and therefore provide also non-statisticians with useful information. In some cases one needs to extend the methodology for handling non-linear preference tendencies. This can be done through the ideal point modelling strategies based on polynomial regression (McEwan, 1996). A practical limitation to this type of methods, however, is that they require a larger set of samples for giving precise model estimates. A strategy based on serving the consumers different samples with subsequent fuzzy clustering was proposed in Johansen, Herseleth, and Næs (2010) for solving this problem. For an alternative probabilistic approach to ideal points modelling we refer to MacKay (2001), Ennis and Rousseau (2004), and MacKay (2006). Comparisons between this approach and regression based methods can be found in Meullenet, Xiong, and Findlay (2007), Busing et al. (2010) and Rousseau, Ennis, and Rossi (2012). Alternative ways of solving similar problems can be found in Carroll (1972), Danzart, Sieffermann, and Delarue (2004), Busing, and Van Deun (2005), Meullenet, Lovely, Threlfall, Morris, and Striegel (2008), van de Velden, De Beuckelaer, Groenen, and Busing (2013). For an exhaustive comparison of methods and software programs to perform regression based preference mapping methods we refer to Yenket, Chambers, and Adhikari (2011).

Although the main purpose of preference mapping is to understand individual differences and their relation to the product characteristics, in many cases one is also interested in understanding how these differences relate to consumer characteristics such as age, gender, attitudes and habits. This type of information can

* Corresponding author. Tel.: +39 0461615388.
E-mail address: isabella.endrizzi@fmach.it (I. Endrizzi).

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.10.002

sometimes be essential for developing for instance successful marketing strategies. A number of different methods are developed for this purpose, both one step procedures (Martens et al., 2005; Vigneau, Endrizzi, & Qannari, 2011) and procedures based on first analysing the liking pattern and then relating this pattern to the external data (see e.g. Næs, Brockhoff, & Tomic, 2010).

The focus of the present paper is on regression based approaches to preference mapping with special emphasis on aspects that need a more detailed consideration. In particular, this paper is devoted to a discussion of validation, segmentation and different centring of the preference data. One of our aims is to generalise the method proposed in Endrizzi, Menichelli, Johansen, Olsen, and Næs (2011) for conjoint data to incorporate also preference mapping. This is a method that accommodates both population averages and individual differences in the same analysis of variance approach. It is of particular interest to discuss how this method differs from standard preference mapping when comes to segmentation. When concerns validation of preference mapping results, we will propose a new approach based on permutation testing which is particularly useful when the number of samples is very low, as is often the case in this type of studies. We will also highlight the importance of segmentation based on visual inspection of plots (Endrizzi et al., 2011) since it allows for segmentation according to interpretation and thus represents a more flexible approach than automatic procedures (Wajrock, Antille, Rytz, Pineau, & Hager, 2008). A major aspect here is also that one can seldom expect clearly separated consumer segments and different automatic clustering procedures can therefore easily end up with different proposals or proposals with little meaning. For illustrating the methodology we will use two data sets from consumer studies on apple juice and raspberry juice. In both cases, the general tendency of population liking has already been studied, but it will be shown that more information can be extracted handling individual differences as proposed here.

2. Methods

2.1. Linear and ideal point preference mapping

Linear preference mapping is based on relating sensory data to consumer liking data using a multivariate linear regression model, usually either PCR or PLS regression (McEwan, 1996; Martens, Esposito Vinzi, & Martens, 2007). For so-called internal preference mapping, the consumer data are treated as independent variables while for external mapping the sensory data are considered the independent ones. Both methods are used in practice and they have their own advantages and disadvantages as discussed in for instance by Næs et al. (2010). They both provide scores, sensory loadings and consumer loadings which are interpreted in the same way using scatter plots. Ideal point preference mapping (as discussed in McEwan, 1996) by the use of polynomial regression is strongly related to linear external mapping, the only differences being that quadratic and interactions terms between the principal components of the sensory data are added to the linear model. Both these methods will form the basis for the rest of the developments in this paper, with main focus on linear mapping.

2.2. An alternative approach

The single framework approach for conjoint analysis mentioned above (Endrizzi et al., 2011) is based on estimating the average population liking effects using a standard ANOVA model and considering the residuals for interpretation of the individual differences. In other words, the averages are calculated to represent the population structure and the residuals are calculated to represent the individual differences representing the deviations from the average effects. An advantage of this approach is that all results can be understood within the same framework model. For interpretation it is, however, important to note that the residuals obtained in this way are double-centred as will be discussed below.

For our purpose, the appropriate ANOVA model can be written as

$$y_{ij} = \mu + \alpha_i + C_j + \varepsilon_{ij}, \quad i = 1, \cdot, I, \quad j = 1, \cdot, J \tag{1}$$

where y_{ij} is the (ij)th observation, μ is the general mean, α_i is the main effect of the tested products, C_j is the random main effect of consumers and ε_{ij} is the random error. Since each consumer tests each sample once, no interaction between consumer and product is possible. The information about individual differences and how they interact with the product effect is therefore to be found in the residuals and **only** there. The average liking for the different products will be found in the estimates of the α's. With the product effect and the consumer main effect in the model, the residuals can be written as

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{C}_j. \tag{2}$$

These residuals are double centred, i.e. they are mean centred across products and across consumers for each combination of i and j. The same values can in fact be obtained by just double centring the original data matrix without going via the ANOVA model, but the link to the model clarifies their relation to the individual differences.

As for standard internal preference mapping these residuals can also be analysed by a PCA. In Endrizzi et al. (2011) it was illustrated how the scores and loadings plot from such a PCA can be used for visual interpretation and segmentation. If external consumer characteristics are available the consumer loadings for the PCA can be related to these values using for instance PLS regression. If useful segments are identified, one can use discriminant PLS instead. It is also possible to regress the sensory attributes onto the scores of this PCA in order to understand how the different dimensions relate to the sensory data.

2.3. Interpretation of individual differences

Although the residuals approach will here mainly be used for visual segmentation according to interpretation, it is worthwhile to discuss the relation between the PCA results of the raw data (identical to internal mapping) and the residuals.

Using only centring per consumer as done for standard internal mapping (like in MDPREF, see e.g. McEwan, 1996), the PCA results represent how the different consumers rate the relative differences between products. The concept "relative" here refers to the fact that the level of liking is removed due to the centring. For a meaningful interpretation of the PCA plot, one implicitly assumes that there is an underlying space of relative differences that all consumers have in common (scores), but one allows for the consumers to use this underlying space differently. This means that the consumers, which represent the loadings, may in principle be spread out all over the place in the plot. All this implies that the liking values of the different **products** play the most important role in the interpretation. Note that regular MDPREF implicitly, since only averages for each consumer are subtracted, contains information about both average product effects and individual differences in liking.

Due to double-centring, in the residual approach the subsequent PCA has another interpretation. For standard preference mapping with only centring of the consumers, the focus is on the relative differences between the products in the PCA scores space. For the residual approach, however, the sum of each of the rows is also equal to 0 which means that each of the values in the row refers to the average consumer value for that product. One can think