CLV3W: A clustering around latent variables approach to detect panel disagreement in three-way conventional sensory profiling data

Tom F. Wilderjans a,b, Véronique Cariou c,d,*

a Methodology of Educational Sciences Research Group, Faculty of Psychology and Educational Sciences, KU Leuven, Andreas Vesaliusstraat 2, Box 3762, 3000 Leuven, Belgium
b Research Group of Methodology and Statistics, Institute of Psychology, Faculty of Social and Behavioral Sciences, Leiden University, Pieter de la Court Building, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands
c LUNAM Université, ONIRIS, Unité de Sensométrie et Chimiométrie, Nantes F-44322, France
d INRA, Nantes F-44316, France

A R T I C L E   I N F O
Article history:
Received 15 September 2014
Received in revised form 10 March 2015
Accepted 19 March 2015
Available online 27 March 2015

Keywords:
Sensory profiling
Fixed vocabulary
Panel (dis)agreement
Assessor weight assignment
Clustering of variables
CLV3W
Clusterwise Parafac
Latent components

A B S T R A C T
To detect panel disagreement, we propose the clustering around latent variables for three-way data (CLV3W) approach which extends the clustering of variables around latent components (CLV) approach to three-way data typically obtained from a conventional sensory profiling procedure (i.e., assessors rating products on various descriptors). The CLV3W method groups the descriptors into Q clusters and estimates for each cluster an associated latent sensory component such that the attributes within each cluster are as much related (i.e., highest squared covariance) as possible with the latent component. Simultaneously, for each latent sensory component separately, a system of weights is estimated that yields information regarding the extent to which an assessor (dis)agrees with the rest of the panel according to the latent sensory component under study. Our new approach is illustrated with a dataset pertaining to Quantitative Descriptive Analysis applied to cider varieties. It is shown that CLV3W, as opposed to related approaches, is able to detect differential panel disagreement on various latent sensory components.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In conventional sensory profiling, a panel of assessors is used to determine the differences among a set of products in terms of their sensory properties. To this end, a fixed vocabulary of descriptive terms is chosen and each assessor belonging to the panel is asked to rate each product according to each term of this (fixed) set. Of paramount interest in conventional sensory profiling is the assessors’ reliability as it has a large impact on the interpretation of the sensory properties of the products, which is very important for product optimization purposes. To evaluate assessors’ reliability, researchers try to determine for each assessor their discriminability, repeatability and agreement with the (other) members of the panel (Latreille et al., 2006). Focusing on the latter criterion, the goal is to evaluate whether a consensus within the panel exists. This situation may, for example, occur when the attributes are difficult to define or when different assessors attach a different meaning to the same attributes (Dijksterhuis, 1995).

In the past, many approaches have been proposed to account for individual differences in panel performance. For example, Dijksterhuis (1995) introduced a consonance index that was based on Principal Component Analysis (PCA) to detect (individual) differences in the use of each attribute. Studying the individual differences on each attribute separately has also been done by researchers that examine the assessors × products interaction term in an analysis of variance (Couronne, 1997). In this regard, for example, the egg-shell plot has been proposed in order to highlight an assessor’s agreement with the panel’s ranking (Lea, Redbotten, & Næs, 1995). Another example is Peron (2000) in which an analysis of variance is performed to select discriminant attributes and the consonance index is used to reveal assessor’s agreement. More recently, some authors applied linear mixed models to measure the reliability of a panel (Brockhoff, 2003; Latreille et al., 2006). Schlich (1996) proposed a method called Control of Assessor Performances (CAP) to study the discrimination and the agreement of each judge, while Derks (2010) proposed Panel Concordance Analysis (PANCA) as a tool for panel leaders...
to identify disagreement between the panelists. A disadvantage of all these methods is that they study each attribute separately, herewith ignoring the (mutual) relationships among the attributes.

To take the mutual relationships between the attributes into account, researchers proposed to use a multivariate approach. Because conventional profiling data can be studied using different viewpoints, a wide range of statistical multivariate methods to analyze such data have been introduced in the past (Rossini, Verdun, Cariou, Qannari, & Fogliatto, 2012). Some methods that are associated with the RV coefficient (Schlich, 1996; Tomic, Forde, Delahuntyn, & Næs, 2013) or that rely on Generalized Procrustes Analysis (Collins, 1992; Gower, 1975; Qannari, MacPie, & Courcoux, 1999; ten Berge, 1977) determine a global configuration of the products which is based on the consensus between the assessors. To study panel disagreement, these methods calculate a coefficient measuring assessors’ performance and this coefficient is used to down-weight bad performers. The rationale behind such a strategy, which has also been adopted in Qannari and Meyners (2001) and in Ledauphin, Hanafi, and Qannari (2006), is to assign a unique weight to each assessor that reflects assessors’ overall agreement with the rest of the panel, herewith ignoring that assessors may agree with the panel on one latent dimension but that they may totally disagree on another latent dimension (Kunert & Qannari, 1999; Tomic et al., 2013). In order to refine the weighting scheme, Verdun, Cariou, and Qannari (2012) introduced a system of weighting that allows each assessor-product combination to be associated with a different weight.

Although many multivariate analysis methods have been proposed in the past for the analysis of fixed vocabulary profiling data, three-way methods have surprisingly been almost neglected. To the best of our knowledge, for sensory profiling, we are only aware of three-way methods that use a Tucker approach (Brockhoff, Hirst, & Næs, 1996; Dahl & Næs, 2009; Romano, Brockhoff, Hersleth, Tomic, & Næs, 2008). Three-way methods, however, may be very useful for analyzing sensory profiling data as they allow a more precise weighting scheme that is based on the analysis of assessor-sensory dimension combinations rather than on assessor-product combinations. The former weighting approach appears more suited to account for panel (dis)agreement since it aims at detecting the differences between assessors in their ability to perceive sensory dimensions. Moreover, this weighting strategy makes it possible to highlight assessors who need more training for only a subset of sensory descriptors (related to the problematic sensory dimensions), which is of paramount interest for a panel leader.

To this end, we introduce in this paper a clustering of variables (Qannari, Vigneau, Luscan, Lefebvre, & Vey, 1997) approach that yields the sensory dimensions (i.e., one dimension for each cluster of variables) underlying the data along with a weighting scheme that reveals each assessors’ degree of agreement with the panel on each sensory dimension (i.e., cluster of variables). This approach extends Clustering of variables around latent components (CLV) analysis (Vigneau & Qannari, 2003; Vigneau, Qannari, Sahmer, & Ladiray, 2006), a method that already has been successfully applied in the context of conventional sensory profiling; CLV aims at clustering sensory descriptors (i.e., variables) along with summarizing each descriptor cluster by a latent component that captures the underlying sensory dimension. In particular, the groups and latent variables are determined in such a way that the (observed) variables in each group are as much related (in terms of squared covariance) to their latent variable as possible. In the context of sensory profiling procedures, however, CLV was only proposed for analyzing two-way data, which are usually obtained by aggregating (e.g., taking the mean) the sensory data over the assessors, resulting in the loss of information regarding the differences between assessors which is of special interest to us. Therefore, we propose the clustering around latent variables for three-way data (CLV3W) approach which extends CLV to three-way profiling data such that assessor differences in the use of the underlying sensory dimensions can be disclosed.

The rest of the paper is organized as follows. In Section 2, we discuss how CLV can be extended to three-way data structures and we present an algorithm to estimate the clustering and the underlying sensory dimensions. In Section 3, the new method is applied to a sensory profiling data set that pertains to the sensory evaluation of ciders. We end the paper by drawing general conclusions and pointing to possible future developments.

2. The CLV3W model for three-way data (material and methods)

2.1. Conventional sensory profiling data structure

In conventional sensory profiling studies, K assessors score I products according to a set of J attributes, stored in the I \times J \times K data array X (see Fig. 1). To present the CLV3W model with a single component, we will start from the matrix X_j (I \times K), which is the jth lateral slice of X (Kiers, 2000) and which contains the scores from the I products on attribute j as given by the K assessors. Without loss of generality, we assume that all X_j (j = 1, ..., J) are column-wise centered (i.e., a mean product score of zero for each attribute-assessor combination). As such, known variations among the assessors are accounted for by removing the assessors’ main (or shift) effect (i.e., assessors using different levels of the rating scale).

2.2. The CLV3W model

The goal of the CLV3W analysis is to cluster the J attributes into Q clusters and to determine Q latent variables t_1, t_2, ..., t_Q along with cluster-specific assessor weights w_q such that the following function is maximized:

\[
g = \sum_{j=1}^{J} \sum_{q=1}^{Q} p_{jq} \text{cov}^2(X_{tq}, t_q)\]

with p_{jq} indicating whether attribute j belongs (p_{jq} = 1) or does not belong (p_{jq} = 0) to cluster G_q. With regard to cluster G_q, this function implies that the sum of squared covariances between t_q and the weighted average of the scores of the assessors for attributes j belonging to G_q is maximized. Note that the weights w_1, w_2, ..., w_Q differ depending on the cluster q to which the attribute in question belongs. As such, this is equivalent to minimize:
\[ f = \frac{1}{J} \sum_{j=1}^{J} \sum_{q=1}^{Q} p_{jq} |X_j - \mathbf{w}_{q} \mathbf{t}_q^T|^2 \]  

(2)

where \( \mathbf{w}_q \) corresponds to the loading of the attribute \( j \) in cluster \( \mathbf{G}_q \) (with this loading being zero when variable \( j \) does not belong to \( \mathbf{G}_q \)). This latter criterion boils down to a Clusterwise Parafac model with \( Q \) clusters and one component in each cluster (Wilderman & Ceulemans, 2013; see also the Parafac with Optimally Clustered Variables – PFOCV – model as presented in Krijnen, 1993).

It is worth noting that in the case of having only a single assessor (i.e., \( K = 1 \) and \( X_j \) being reduced to a vector \( x_j \) with scores from a single assessor) or having scores obtained by computing the (weighted) average over assessors, minimizing this criterion is equivalent to minimizing the CLV criterion (Vigneau et al., 2006) which is:

\[ E = \frac{1}{J} \sum_{j=1}^{J} \sum_{q=1}^{Q} p_{jq} |X_j - \mathbf{w}_{q} \mathbf{t}_q^T|^2. \]  

(3)

In this case, Vigneau et al. (2006) have shown that \( \mathbf{t}_q \) is given by the first standardized principal component of the variables belonging to group \( \mathbf{G}_q \).

Relation with a Consensus PCA model

The three-way data matrix \( X \) can be considered as a multitask data set consisting of multiple blocks \( X_j (j = 1, \ldots, J) \) of size \( (I \times K) \). Given a partition of the variables, CLV3W with a single component (i.e., a one-dimensional model) is closely related to Consensus PCA (Westerhuis, Kourtz, & MacGregor, 1998) in which the vector of loadings \( \mathbf{t}_q \) is restricted to be the same for all blocks \( X_j \) belonging to the same cluster \( \mathbf{G}_q \) (Cariou, Hanafi, & Qannari, 2010). More specifically, maximizing the criterion \( g \) in (1) is equivalent to maximizing the Consensus PCA criterion \( \sum_{q=1}^{Q} \sum_{j \in \mathbf{G}_q} \text{cov}^2(X_j, \mathbf{t}_q) \) with the constraints that \( \mathbf{t}_q \) equals \( \mathbf{w}_q \) for all those \( X_j \) belonging to cluster \( \mathbf{G}_q(q = 1, \ldots, Q); \mathbf{t}_q \) turns out to be the global scores over the blocks \( X_j \) belonging to cluster \( \mathbf{G}_q \) whereas the block scores for each \( X_j \) correspond to \( \mathbf{X}_j \).

2.3. Algorithm

To fit a \( Q \)-cluster CLV3W model to a three-way data set at hand, the following algorithm is used. First, an initial partition of the variables is obtained. Next, the CLV3W algorithm iterates the following two updating steps until convergence: (1) updating the cluster membership \( p_{jq} \) of each variable conditional on the cluster-specific parameters (i.e. \( t_q \), \( \mathbf{w}_q \) and \( \mathbf{G}_q \)) and (2) re-estimating the cluster-specific parameters conditional upon the cluster memberships. In order to minimize the risk of the algorithm getting stuck in a local optimal solution, a multi-start procedure is adopted. A schematic overview of the algorithm (in terms of pseudo-code) can be found in Appendix I. Software to perform a CLV3W analysis has been implemented in Matlab (version 2014b) and in \( R \) (version 3.1.0) and is available upon request from the authors. Moreover, \( R \) code to perform a CLV3W analysis will soon be added to the \( R \) package ClustVarLV. In the following, the initialization step, the multi-start procedure and the two updating steps will be discussed in detail.

2.3.1. Obtaining an initial variable partition

An initial variable partition with \( Q \) clusters can be determined in a random or in a more rational way. A random partition of the variables may be obtained by randomly assigning the \( J \) variables to \( Q \) clusters. As random partitions may differ quite a lot from the optimal partition, it may be better to look for an initial partition that is more close to the optimal one. One way to go, when the user has some previous knowledge (e.g., results from an earlier analysis) or has some expectations regarding the partition of the variables, is to incorporate this information into the algorithmic procedure by adding one (or more) user-specified starting partition(s). A second way out, as is often done in classical clustering, consists of applying an Agglomerative Hierarchical Clustering (AHC) based on criterion \( f \) in (2) using Ward’s aggregation criterion. At the first step of the hierarchical clustering, each variable (i.e. \( X_j \)) forms a group on its own. In this case, the \( f \) criterion in (2) is equal to:

\[ f_1 = \sum_{j=1}^{J} ||X_j - \mathbf{t}_q \mathbf{w}_q^T||^2, \]  

(4)

with \( \mathbf{t}_q \) being the first standardized principal component that is associated with the largest eigenvalue of \( X_j \) and \( \mathbf{w}_q \) the associated (standardized) vector of loadings. We recall that, in the CLV3W case, \( \mathbf{w}_q \) represents the assessors’ weighting scheme for \( X_j \).

This latter criterion boils down to a consensus criterion (Steinley, 2006) which will be discussed in detail. At each step, the hierarchical strategy consists of aggregating the two groups which lead to the smallest increase in \( \Delta f \). This aggregating of variable groups continues until \( Q \) clusters of variables are determined. Note that merging two clusters of variables \( \mathbf{G}_q \) with \( \mathbf{G}_q \) implies that simultaneously the restrictions (1) \( t_q \) equals \( t_{q+1} \) and (2) \( W_q \) is equal to \( W_{q+1} \) are imposed.

\( \Delta f = - \sum_{j \in \mathbf{G}_q} ||X_j - \mathbf{w}_q \mathbf{t}_q^T||^2 - \sum_{j \notin \mathbf{G}_q} ||X_j - \mathbf{w}_{q+1} \mathbf{t}_{q+1}^T||^2 + \sum_{j \notin \mathbf{G}_q} ||X_j - \mathbf{w}_q \mathbf{t}_q^T||^2. \]  

(5)

This criterion will always be greater than zero, and it will increase when going from one step to the subsequent step of the hierarchy. At each step, the hierarchical strategy consists of aggregating the two groups which lead to the smallest increase in \( \Delta f \). This aggregating of variable groups continues until \( Q \) clusters of variables are determined. Note that merging two clusters of variables \( \mathbf{G}_q \) and \( \mathbf{G}_q \) implies that simultaneously the restrictions (1) \( t_q \) equals \( t_{q+1} \) and (2) \( W_q \) is equal to \( W_{q+1} \) are imposed. As such, it is guaranteed that the loss function \( f \) in (2) will increase (i.e., \( \Delta f \) being positive) as a constrained model always will fit the data worse (or equal) than an unconstrained (or less constrained) model (when having for both models the global optimal least-square estimates of their parameters).

2.3.2. Multi-start procedure

As CLV3W involves a clustering of the variables, the CLV3W optimization problem is very hard to solve (Steinley, 2006a). As a consequence, as is true for classical K-means (Steinley, 2003, 2006b), the CLV3W algorithm is not guaranteed to converge to the global optimal solution and its performance strongly depends on the initialization used. Therefore, in order to minimize the risk of the CLV3W algorithm to get stuck in a local optimal solution, it is advised to use a multi-start procedure (i.e., a similar advise is given for K-means, see Steinley, 2003, 2006b). This procedure consists of generating an initial variable partition and performing the two updating steps until convergence multiple times, each time starting with a different initial partition of the variables, and retaining the solution that yields the lowest loss function value. Regarding initialization, we advise to always use the (rationally determined) variable partition obtained by applying the Agglomerative Hierarchical Clustering procedure as one of the initial partitions. Further, we strongly encourage researchers to complement this rational start with at least 50 random starts, and, when available, with one (or more) user-defined starting partition(s). However, when there are a large number of variables and/or when the user wants to fit a model with many clusters, a larger number of (random) starts is preferred. Note that, in general, determining a rational start by means of the Agglomerative Hierarchical Clustering procedure is
more time-consuming than generating a random (or adding a user-specified) start, but that, in general, the rational initial partition will be closer to the optimal one than the random initial partition. The goal of combining different types of starts (i.e., rational, random and user-specified) is to cover a larger part of the solution space, resulting in a lower risk of the algorithm retaining a local optimal solution.

2.3.3. Two updating steps

To update the cluster membership of variable \( j \), the optimal \( z_{jq} \) for each cluster \( C_q \) given \( t_{jq} \) and \( w_q \) is computed by means of linear regression (for more information, see Smilde, Bro, & Geladi, 2004). Next, the criterion \( f_{jq} = X_j - z_{jq}(t_{jq}w_qe_q^T) \) is computed for each cluster \( C_q \) (i.e., the extent to which variable \( j \) does not fit in cluster \( C_q \)), and variable \( j \) is assigned to the cluster \( C_q \) for which \( f_{jq} \) is minimal. When two (or more) clusters \( C_q \) exist that yield the same minimal \( f_{jq} \)-value, which will almost never be encountered in empirical datasets, variable \( j \) is assigned at random to one of these clusters. After updating the cluster membership of each variable, the cluster-specific parameters are re-estimated by fitting a Parafac model (Carroll & Chang, 1970; Harshman, 1970; Hitchcock, 1927) with one component to each \( X_{(i,j,q)} \) \((q = 1, \ldots, Q)\), which is a three-way array that only consists of the variables that belong to cluster \( C_q \). As no closed-form expression exists for the optimal parameters of a Parafac model, the cluster-specific parameters \( t_q, w_q \) and \( z_{jq} \) for each \( X_{(i,j,q)} \) are estimated by means of an alternating least squares algorithm (ten Berge, 1993). In this algorithm, each set of parameters (i.e., \( t_q, w_q \) or \( z_{jq} \)) is re-estimated alternatingly, herewith keeping the other parameters fixed; this alternating procedure boils down to solving a series of multivariate linear regression problems (for more information and a comparison of algorithms for Parafac, see Bro, 1997; Faber, Bro, & Hopke, 2003; Tomaso & Bro, 2006; for Matlab and R based software to fit Parafac models, see Andersson & Bro, 2000; Giordani, Kiers, & Del Ferraro, 2014). It should be noted that finding the optimal least-squares parameters for a Parafac model is not a trivial task as the iterative procedure described above may get stuck in a local optimal solution or may produce a degenerate solution in which some components are highly (negatively) correlated (for a discussion of these problems in the context of Parafac, see De Silva & Lim, 2008; Harshman, 1970; Krijnen, Dijkstra, & Stegeman, 2008; Kronenberg, 2008; Mitchell & Burdick, 1994; Smilde et al., 2004; Stegeman, 2006, 2007). It should further be noted that \( z_{jq} \) is computed in both updating steps and that, in general, the optimal value for \( z_{jq} \) after the first updating step will be different from the optimal value \( z_{jq} \) after the second step.

After each update of all cluster memberships and all cluster-specific parameters (i.e., after each iteration which consists of performing both updating steps one time), it is checked whether there are empty clusters (i.e., resulting in a model with less than \( Q \) clusters). When empty clusters are encountered, the variable fitting its cluster the worst is reassigned to the empty cluster and the cluster-specific parameters are re-calculated (see second updating step). This procedure is repeated until all clusters contain at least one variable. Next, it is determined whether or not the algorithm has been converged. This is the case when (1) after updating the cluster memberships of all variables, the same clustering and, consequently, the same cluster-specific parameters and loss function value is obtained or (2) the decrease in the loss function value is smaller than some pre-defined tolerance value (i.e., .0000001). When no convergence has been obtained, the algorithm returns to the first updating step (i.e., updating of cluster memberships; see pseudo-code in Appendix I).

2.4. Selecting the number of clusters

To determine the optimal number of clusters, one performs multiple CLV3W analyses with increasing numbers of clusters (e.g., one, two, three, etc.). Subsequently, one selects a solution that has the best balance between model fit (i.e., sum of squared difference between data and predicted data by the model) and model complexity (i.e., number of clusters). To this end, one may rely on the scree test of Cattell (1966) or a generalized version thereof (Ceulemans & Kiers, 2006, 2009; Wilderjans, Ceulemans, & Meers, 2013). Applied to CLV3W, a scree test consists of plotting the loss function value (2) against the number of clusters and retaining the solution that lies at the sharpest elbow in the plot. This solution nicely balances fit and model complexity as retaining a less complex model will result in a substantial drop in model fit, whereas adding an extra cluster only yields a small gain in the fit of the model. In order to determine the sharpest elbow in the plot in a more automated way (instead of visually eyeballing), one may make use of the CHull method which looks for the model that yields the largest scree-ratio among the models that are located on the convex hull of the plot of model complexity versus model (mis)fit (see Ceulemans & Kiers, 2006; Wilderjans et al., 2013). It should be noted that the optimal number of clusters can also be determined by detecting an elbow in the scree diagram which depicts the evolution of the aggregation ratio associated with the dendrogram of the Agglomerative Hierarchical Clustering (i.e., stepsze criterion) procedure. One should, however, acknowledge that the optimal Q-cluster partition can be quite different from the clustering obtained by cutting the dendrogram at Q clusters, and, as a consequence, that the evolution of the aggregation ratio and the evolution of the sum of squared prediction errors can give a different view on which number of clusters to retain. When selecting a good model, one, of course, should not blindly rely on some automated model selection procedure but always should also take the interpretability and stability of the solution into account.

3. Application

3.1. Ciders data set

In order to illustrate the use of the CLV3W method, we consider a case study pertaining to Quantitative Descriptive Analysis (QDA) applied to ten varieties of cider. The sensory panel consists of seven trained assessors who were asked to rate ten varieties of cider using a list of ten sensory attributes, namely: sweet, acid, bitter, astringency, odor strength, pungent, alcohol, perfume, intensity, and fruity (Verdun et al., 2012).

3.2. Pre-processing and analyzing the data

The data associated with each assessor can be presented as an \((I \times J)\) matrix, which will be denoted by \(X_k\) \((k = 1, \ldots, K)\). The rows of this matrix refer to the products and the columns to the attributes. Before analyzing, in order to deal with some known variations among the assessors, each matrix is column-wise centered to remove the assessors’ main (or shift) effect (i.e., assessors using different levels of the rating scale). Further, to control for assessors using different ranges of the scoring scales, isotopic scaling factors were applied to each \(X_k\) (Kunert & Qannari, 1999). In particular, to shrink the configuration of assessors using larger scale ranges and expand the configuration of assessors adopting relatively narrow ranges of the scales, we multiplied the data of each assessor \(X_k\) \((k = 1, \ldots, K)\) by a scaling factor \(S_k\) that is computed as follows: (1) calculate the total variance \(I(X_k)\) of dataset \(X_k\) as the sum of the variances of the columns of \(X_k\); (2) compute \(I_i\) as the average
of $l_2(X_k)$ ($k = 1, \ldots, K$); (3) calculate $\gamma_k = \frac{c_k}{\sum c_k}$. Note that using the isotropic scaling procedure results in each $X_k$ having the same total variance (i.e., $l_k$). Next, we analyzed the pre-processed data with CLV3W using one to seven clusters. We adopted a multi-start procedure consisting of one rational starting partition (i.e., the partition obtained with the Agglomerative Hierarchical Clustering procedure, see Section 2.3) and 50 random initial variable partitions and retained the solution that yielded the lowest loss function value $f$ in (2).

3.3. Results and discussion

3.3.1. Determining the number of clusters

In Fig. 2, which displays the scree plot (only considering the best run among the 50 multi-starts and the rational starting partition) for the cider data with the number of clusters ranging from one to seven, one can see that the solution with two clusters lies at the sharpest elbow, suggesting that the solution with two clusters should be retained. Note that this solution has been obtained both from several random initial partitions and from the rational Agglomerative Hierarchical one (see further). When inspecting the scree ratios (see Section 2.4), the solution with two clusters should be preferred as this solution has the largest ratio (i.e., the scree ratio is 2.98, 1.12, 1.26 and .82 for the solution with two, three, four and five clusters, respectively).

3.3.2. Results

For the CLV3W solution with two clusters, the obtained clustering of the descriptors along with their component loadings are plotted in Fig. 3a, whereas the product scores (resp. assessor weights) for each cluster are depicted in Fig. 4a (resp. Fig. 4b). Note that in Figs. 3a and 4a, the two axes D1 and D2 correspond to the two clusters (i.e., the component loadings and the product scores for the first and second cluster are presented on D1 and D2, respectively).

Looking at the solution with two clusters, it appears that the two main dimensions of this dataset are clearly disclosed: the first cluster mainly contains descriptors that refer to taste with fruity and sweet on the one hand (i.e., positive loading) and bitter and acid (i.e., negative loading) on the other hand. The second cluster contains the descriptors pungent (with a positive loading), odor strength and intensity (with a negative loading). When inspecting the product scores (see Fig. 4a), one can see that the latent variable associated with the first cluster (i.e., D1 axis) clearly separates ciders 2/5–7 (with a negative score) from ciders 3/4/8/10 (with a positive score), whereas the second latent variable (i.e., D2 axis) mainly distinguishes between cider 1–3 (positive score) on the one hand and cider 9 (negative score) on the other hand. The assessor weights (see Fig. 4b), which may differ across latent dimensions and assessors, indicate the importance of each latent dimension for each assessor when discriminating between the products. As a consequence, the assessor weights can be used to detect panel (dis)agreement by comparing for each dimension the weights across assessors. In Fig. 4b, one can see that there is a large consensus regarding the descriptors that are associated with the first latent dimension. For the second dimension, however, clear differences between assessors emerge. In particular, compared to the average panel member, assessor 1 and 5 attach...
a lower importance and assessor 3 and 6 a higher importance to this dimension. It can be concluded that the panel disagrees on the second dimension but not regarding the first dimension.

We also inspected CLV3W solutions with a larger number of clusters. It appears that the obtained partitions are nested and that they mostly correspond to the Agglomerative Hierarchical Clustering solutions. In the three cluster solution, for example, the largest cluster of the two cluster solution is split into two smaller ones. Note that CLV3W does not necessarily yield nested clusters (i.e., the observed nesting structure is a feature of the cider data, not of the CLV3W method). We also looked at the variation in two-cluster partitions that has been encountered across the 50 multi-starts. From Table 1, one can see that 13 different final partitions were obtained, with 12 of them being solutions that are only locally optimal. It further appears that only 5 out of the 50 random starts yielded the same optimal partition and that the partitions from locally optimal solutions are quite different from the optimal partition in terms of Adjusted Rand Index (Hubert & Arabie, 1985). However, when comparing the partitions for the three best solutions, which have low function values that are quite close to each other, it can be concluded that the final partition is stable as only for bitter (d4) and pungent (d7) some uncertainty exists regarding the cluster to which these attributes belong. Finally, each random run took, on average, 8.14 s to arrive at a final partition, while the run with the Agglomerative Hierarchical Clustering rational initialization, which directly converged to the optimal solution, only needed 6.83 s.

3.4. Comparison with other methods

In this section, we will compare the CLV3W results for the cider data with the results obtained with related methods applied to the same data set. In particular, we will compare the CLV3W results with the results of (1) a Weighted Partial Least Squares-Discriminant Analysis (PLS-DA), (2) a Parafac analysis with two components (without a clustering of the variables) and (3) a CLV on the data averaged over the assessors. For the PLS-DA we used Matlab (version 2014b), for Parafac the R package ThreeWay (Giordani et al., 2014) and for CLV the R package ClustVarLV.

3.4.1. Weighted PLS-DA

Weighted Partial Least Squares-Discriminant Analysis (PLS-DA) has been proposed as a robust analysis method for conventional sensory profiling data (Verdun et al., 2012). In the context of conventional sensory profiling, a weighted PLS-DA consists of two steps. First, for each combination of an assessor and a product, a weight is determined that reflects the (dis)agreement of the assessor with the rest of the panel with respect to the product in question. Second, the data are averaged over the assessors, using the weights determined in the first step, and the PLS discriminant components are computed (on the weighted data). When analyzing conventional sensory profiling data, PLS-DA is applied to the (product by attribute) datasets $X_k (k = 1, \ldots, K)$ concatenated vertically and the product memberships (i.e., to which product each row belongs) are used as the dependent variable (Rossini et al., 2012). When comparing the Weighted PLS-DA results for the cider data (as presented in Verdun et al., 2012) with our results, it appears that the obtained CLV3W partition shows a strong similarity with the first two Weighted PLS-DA components: the sweet-bitter cluster corresponds to the first component and the descriptors of the second cluster are closely related to the second Weighted PLS-DA component.
However, whereas Verdun et al. (2012) emphasize the possible disagreement between products (e.g., focusing on ciders 7 and 9), CLV3W clearly identifies the sensory descriptors, and naturally the associated underlying sensory dimension, that are problematic. In particular, with regard to the second cluster (including intensity, odor strength and pungent), assessors 1 and 5 are down-weighted, suggesting that they have difficulties with using the associated descriptors when judging the products. The down-weighting of assessors 1 and 5 was also suggested by Verdun et al. (2012), whereas Ledauphin et al. (2006) only identified assessor 5 as a bad performer. Moreover, besides detecting bad performers, CLV3W, which is a multivariate approach instead of a univariate one, makes it also possible to detect which (kind of) attributes are badly rated. As such, we can clearly distinguish the attributes which seem to be more difficult to evaluate (for all or some of the assessors) from those that yield a large consensus within the panel (i.e. sensory descriptors belonging to the first cluster).

### 3.4.2. Two component Parafac

In order to compare our approach with an alternative three-way weighting strategy, we also performed a two component Parafac on the pre-processed data (see Section 3.2). One can see in Fig. 3b that the first Parafac component is quite similar to the latent dimension of the first CLV3W cluster in that it mainly opposes sweet, perfume and fruity to alcohol attributes. The second Parafac component clearly distinguishes intensity and fruity from pungent, which differs from the second CLV3W dimension. When inspecting the product scores which are depicted in Fig. 5a for the Parafac solution, no substantial differences are encountered compared to the product scores of the CLV3W solution (in Fig. 4a). Fig. 5b displays the assessor weights from the two component Parafac solution. Looking at the first Parafac component, the weights associated to the assessors are nearly identical and vary from .30 to .43. With regard to the second component, as is the case for CLV3W (see Fig. 4b), assessor 5 is down-weighted compared to the others.

![Fig. 5. Configuration of the products (a) and assessor weights (b) for the Parafac solution with two components for the cider data.](image)

![Fig. 6. Dendrogram of variables obtained from (a) CLV on aggregated pre-processed data, (b) CLV3W with the Agglomerative Hierarchical Clustering procedure for the cider data.](image)
Unlike CLV3W, however, assessor 1 is not down-weighted in the second component of the Parafac solution.

We can point out that although Parafac also exhibits the group of sensory descriptors including sweet, perfume, fruity and alcohol, the descriptor loadings of the Parafac model (see Fig. 3b) are more difficult to interpret. This is especially the case for the second Parafac component which mainly highlights the attributes pungent and fruity, which are clustered together in one solution but not in the other one (and vice versa). To our point of view, these differences mainly arise because loadings associated with descriptors from a different cluster are set to zero. Note, however, that rotating a (optimal) Parafac solution, in general, results in a worse fitting model (i.e. Parafac has no rotational freedom).

3.4.3. CLV

Finally, we also compared CLV3W with a CLV analysis on the (pre-processed) data averaged across the assessors, which implies a weighting strategy in which each assessor is weighted equally (for all dimensions). To get more insight into the differences between both methods, we will look at the dendrogram that is obtained during the rational initialization phase of both algorithms. The resulting CLV dendrogram is displayed in Fig. 6 (left panel) along with the CLV3W dendrogram (right panel). As for CLV3W, a partition with two clusters is retained. Note that for both CLV and CLV3W this partition into two clusters is also the final partition optimizing the CLV and CLV3W loss functions. While some attributes are grouped together in both solutions (e.g., intensity and odor strength or fruity and sweet), other attributes are clustered together in one solution but not in the other one (and vice versa). To our point of view, these differences mainly arise because of the disagreement among assessors regarding these (dimensions of) attributes (e.g., pungent). Indeed, while CLV3W accounts for the differences between raters in the use/importance of the various latent dimensions (i.e., assessor weights), CLV does not because averaging the data across assessors removes important information regarding assessor differences.

4. Conclusion

In the context of a clustering around latent variables approach, we introduced CLV3W as a new method that extends the CLV procedure of Vigneau and Qannari (2003) and Vigneau et al. (2006) to three-way data. Dealing with conventional profiling data, this approach makes it possible to simultaneously (1) exhibit clusters sensory attributes such that the main sensory dimensions are exhibited. Finally, the CLV3W optimization criterion appears to be equivalent to a Clusterwise Parafac criterion (Wilderjans & Ceulemans, 2013) given a one component Parafac model within each cluster.

More research is needed to further explore the properties of this analysis method and to extend it to other three-way structures which are often encountered in consumer research. Indeed, such three-way structures are more and more collected by food companies who need detailed information about how consumers perceive their products according to several aspects (e.g., odor, taste, texture, global liking) or who want to find out which emotions are activated during the evaluation of their products. As demonstrated here in the special case of conventional sensory profiling data, a strategy which consists in first aggregating the data across assessors before applying a standard two-way or multi-block approach leads to a loss of information regarding the individual differences between assessors (or alternatively between consumers). In the same vein, more research is needed to extend this approach to L-shaped data in the context of consumer preference analysis.

Acknowledgements

Tom F. Wilderjans is a post-doctoral Fellow at the Fund of Scientific Research (FWO) Flanders (Belgium).

Appendix I. Schematic overview of the CLV3W algorithm

**Input:** a dataset \( X \), the number of cluster \( Q \)

**Initialization**

- Obtain \( W \) starting partitions \( P_{\text{start}}(w = 1, \ldots, W) \) by combining the following methods (and obtain the cluster-specific parameters - see second updating step - associated with each \( P_{\text{start}}(w) \))
  - Rational: perform the Agglomerative Hierarchical Clustering procedure on \( X \) and cut the dendrogram at \( Q \) clusters
  - Random: obtain a (multiple) random starting partition(s) with \( Q \) clusters
  - User-specified: specify a (multiple) partition(s) with \( Q \) clusters based on previous knowledge or expected group memberships

**Optimization/iteration**

- For each obtained (rational, random or user-specified) starting partition \( P_{\text{start}}(w = 1, \ldots, W) \)
  - Run until convergence the following steps
    - Update the clustering (see first updating step)
    - Update the cluster-specific parameters (see second updating step)
    - Check for empty clusters
    - Check for convergence

- Store the converged solution and associated loss value \( f_{\text{Q}}(w = 1, \ldots, W) \)
- Retain the solution associated with \( \min(f_1, f_2, \ldots, f_W) \)

**Output:** an optimal partition and associated cluster-specific parameters

References
