External preference segmentation with additional information on consumers.

Evelyne VIGNEAU
Mathilde CHARLES
Mingkun CHEN
External preference mapping and panel segmentation

Consumers

products

Sensory experts

liking scores

Internal preference segmentation

Sensory profiles

Regression model of the segment’s mean
Clustering around Latent Variables (CLV) with co-variables measured on samples

**Objective**: merge together consumers who have similar drivers of preference

Simultaneously, define:

- groups of consumers and
- in each group, the linear combination of the sensory attributes which explains at much as possible the liking scores.

CLV with co-variables measured on samples

Solution

K segments

in each segment \( k \),

\( t_k \) is the first PLS regression component of \( \overline{Y}_k \) on \( X \)

\[
S = \sum_{k=1}^{K} \sum_{j=1}^{p} \delta_{kj} \text{cov}( y_j, t_k )
\]

with

\[
t_k = X a_k \quad a_k^T a_k = 1
\]

\( y_j \): scores of likings for consumer \( j \) (\( j=1,\ldots,p \))

\( c_k \): latent variable in group \( k \) (\( k=1,\ldots,K \))

\( \delta_{kj}=1 \) if \( x_j \) belongs to \( G_k \), =0 otherwise

\( \overline{Y}_k \)
Specificity of « statistical » clustering method in hedonic studies (crisp algorithm)

Each consumer belongs to one, and only one, group

**BUT** not all consumers are well represented by their group’s mean (« non typical » or « spurious » likings)

**OR** some consumers are almost between two groups (degree of neighborhood between segments)

possible difficulties in the interpretation of each segment ... specifically by means of external consumers attributes
Taking account of additional information on consumers

**Working at the individual level:**

- and sum up the consumers in a segment.
- after discarding those consumers with low cluster contribution and/or high between-cluster position
  - $R_2$ with the own cluster, $R_2$ with the next nearest cluster (~silhouette indices)
  - cluster membership’s values from fuzzy clustering.

**working on the segment level**

- On the basis of latent components = central tendencies in the hedonic space associated with the consumers attributes and to the products attributes.
  - $L$-CLV approach
Clustering around Latent Variables (CLV) with co-variables measured on products and additional information on consumers

L-shaped data

Information collected by means of a questionnaire on consumers

\[ Z' \]

\((M \times p)\)

Liking scores

\[ Y \]

\((n \times p)\)

Sensory attributes

\[ X \]

\((n \times Q)\)
L-CLV

maximize

\[ \tilde{S}_X^Z = \sum_{k=1}^{K} \text{cov}(c_k, t_k) \]

with

\[ c_k = P_k u_k \]
\[ u_k' u_k = 1 \]
\[ P_k = Y_k Z_k \]

\[ P_k \leftrightarrow \text{interaction between } Y \text{ and } Z \]

\[ t_k = X a_k \]
\[ a_k' a_k = 1 \]

\[ \tilde{S}_X^Z = \frac{1}{n} \sum_{k=1}^{K} u_k' Z_k Y_k' X a_k \]
Apple Case study (COSI-VEG 2010-2013)
31 apple cultivars locally produced (Loire Valley, France)

Consumer questionnaire
- Frequency of Consumption,
- Apple cultivars known
- Important sensory attributes,
- Modalities of consumption (peeled/during meal/ ...)
- Purchase criteria
- Supply location
- 14 questions « eater style » (likert scale)
- 7 questions « opinion on apple » (likert scale)
- Age, gender, professional activity....

Hedonic test
224 regular apple consumers
- Liking score on a 9-points
- 5 sessions during 3 weeks

Sensory descriptive analysis
15 assessors, 15 attributes
Crunchy A_Pineapple/Banana
Juicy A_Sweet/Rose
Fondant A_Woody/Earthy
A_Rustic
Sweet A_Lemon
Acid A_White flowers
A_Ripe fruit
Odour intensity A_Green
Aroma intensity

11th Sensometrics meeting- 10-13 july 2012, Rennes, France
Apple Case study (COSI-VEG 2010-2013)
31 apple cultivars locally produced (Loire Valley, France)

Consumers attributes
- Categorical: dummy variables, globally scaled
- Numerical: centered and unit scaled

Liking scores
- Centered and unit scaled

Sensory attributes
- Centered and unit scaled

31 apple cultivars locally produced (Loire Valley, France)
L-CLV: choice of the number of segments

Evolution of the aggregation criterion

CLV Dendrogram

Segment L3-1: 82 consumers (37%)
Segment L3-2: 96 consumers (43%)
Segment L3-3: 46 consumers (20%)

11th Sensometrics meeting - 10-13 July 2012, Rennes, France
L-CLV:
Consumers segments represented on the internal preference mapping.
L-CLV:
Consumers segments represented on the internal preference mapping

![Diagram showing consumer segments on an internal preference mapping.](image)
**L-CLV**: Loadings associated to the sensory attributes \((a_k)\)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
<th></th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crunchy</td>
<td>0.308</td>
<td>0.202</td>
<td>-0.214</td>
<td>Crunchy _sq</td>
<td>-0.138</td>
<td>-0.098</td>
<td>-0.102</td>
</tr>
<tr>
<td>Juicy</td>
<td>0.307</td>
<td>0.154</td>
<td>-0.073</td>
<td>Juicy _sq</td>
<td>-0.150</td>
<td>-0.083</td>
<td>-0.047</td>
</tr>
<tr>
<td>Fondant</td>
<td>-0.222</td>
<td>-0.147</td>
<td>0.297</td>
<td>Fondant _sq</td>
<td>-0.122</td>
<td>-0.134</td>
<td>-0.020</td>
</tr>
<tr>
<td>Sweet</td>
<td>0.330</td>
<td>0.376</td>
<td>0.278</td>
<td>Sweet _sq</td>
<td>0.025</td>
<td>-0.033</td>
<td>-0.154</td>
</tr>
<tr>
<td>Acid</td>
<td>-0.068</td>
<td>0.060</td>
<td>-0.298</td>
<td>Acid _sq</td>
<td>-0.120</td>
<td>-0.077</td>
<td>-0.237</td>
</tr>
<tr>
<td>Odour Int</td>
<td>0.203</td>
<td>0.189</td>
<td>0.038</td>
<td>Odour Int _sq</td>
<td>-0.007</td>
<td>-0.022</td>
<td>-0.190</td>
</tr>
<tr>
<td>Aroma Int</td>
<td>0.282</td>
<td>0.344</td>
<td>0.066</td>
<td>Aroma Int _sq</td>
<td>0.091</td>
<td>0.040</td>
<td>0.018</td>
</tr>
<tr>
<td>A_Pineapple/Banana</td>
<td>0.320</td>
<td>0.390</td>
<td>0.067</td>
<td>A_Pineapple/Banana _sq</td>
<td>0.166</td>
<td>0.205</td>
<td>-0.179</td>
</tr>
<tr>
<td>A_Sweet/Rose</td>
<td>0.322</td>
<td>0.322</td>
<td>0.292</td>
<td>A_Sweet/Rose _sq</td>
<td>0.125</td>
<td>0.143</td>
<td>0.097</td>
</tr>
<tr>
<td>A_Woody/Earthy</td>
<td>-0.234</td>
<td>-0.270</td>
<td>-0.138</td>
<td>A_Woody/Earthy _sq</td>
<td>-0.079</td>
<td>-0.089</td>
<td>-0.078</td>
</tr>
<tr>
<td>A_Rustic</td>
<td>-0.232</td>
<td>-0.293</td>
<td>0.055</td>
<td>A_Rustic _sq</td>
<td>-0.005</td>
<td>-0.059</td>
<td>0.082</td>
</tr>
<tr>
<td>A_Lemon</td>
<td>-0.071</td>
<td>0.077</td>
<td>-0.242</td>
<td>A_Lemon _sq</td>
<td>-0.212</td>
<td>-0.116</td>
<td>-0.207</td>
</tr>
<tr>
<td>A_White flowers</td>
<td>0.049</td>
<td>-0.033</td>
<td>-0.095</td>
<td>A_White flowers _sq</td>
<td>0.038</td>
<td>0.009</td>
<td>-0.044</td>
</tr>
<tr>
<td>A_Ripe fruit</td>
<td>0.020</td>
<td>0.090</td>
<td>0.258</td>
<td>A_Ripe fruit _sq</td>
<td>0.004</td>
<td>-0.021</td>
<td>0.174</td>
</tr>
<tr>
<td>A_Green</td>
<td>-0.147</td>
<td>-0.202</td>
<td>-0.350</td>
<td>A_Green _sq</td>
<td>-0.089</td>
<td>-0.113</td>
<td>-0.248</td>
</tr>
</tbody>
</table>
**L-CLV** : Loadings associated to the sensory attributes \( (a_k) \)

- **Segments L3-1 and L3-2** : similar sensory keydrivers (texture crunchy and juicy, sweet flavor, « pineapple/banana » aroma) ... slight differences for acidity, « lemon » aroma.

- **Segment L3-3** : do not reject fondant texture, appreciate more « rustic » and « ripe-fruit » aroma than « pineapple/banana » aroma, clearly reject acidity, « green » aroma.
L-CLV: Loadings associated to the consumers attributes \((u_k)\)

The most discriminant modalities associated to the consumers attributes are highlighted.
(\text{chi-square test/v.test})
**L-CLV**: Loadings ($u_k$) associated to agreement measurements

(likert-type scale, centered data)

- **Clear opposition between segment 2 and segment 1**
  Segment 1: apple is not a « good fruit », not for every day, don’t eat at lot of fruits and vegetables

- **Segment 3**: don’t like acid fruits, prefer sweet fruits but don’t add sugar, are very not fond of apples

11th Sensometrics meeting- 10-13 July 2012, Rennes, France
Comparison with CLV on Y with external X-block ... without Z-block
Comparison regarding the sensory key-drivers

**L - CLV on Y with external X-block and Z-block**
Partition in 3 groups

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Main sensory drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3-1</td>
<td>82</td>
<td>+ Sweetness + Juicy + Crunchy + A_Pineapple/banana + A_Sweet/rose</td>
</tr>
<tr>
<td>L3-2</td>
<td>96</td>
<td>++ A_Pineapple/banana ++ Sweetness + Aroma intensity 0 A_Lemon</td>
</tr>
<tr>
<td>L3-3</td>
<td>46</td>
<td>-- A_green -- Acidity - A_lemon + Fondant 0 A_Pineapple/banana</td>
</tr>
</tbody>
</table>

**CLV on Y with external X-block**
(without Z-block)
Partition in 4 groups

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Main sensory drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>YX4-1</td>
<td>67</td>
<td>++ Juicy ++ Crunchy + Sweetness + A_Pineapple/banana</td>
</tr>
<tr>
<td>YX4-2</td>
<td>45</td>
<td>++A_Pineapple/banana - A_Rustic + Aroma intensity + A_Lemon</td>
</tr>
<tr>
<td>YX4-3</td>
<td>40</td>
<td>-- A_green -- Acidity - A_lemon + Fondant 0 A_Pineapple/banana</td>
</tr>
<tr>
<td>YX4-4</td>
<td>72</td>
<td>++ Sweetness ++ A_Sweet/rose + A_Pineapple/banana - A_green</td>
</tr>
</tbody>
</table>

Correlation coeff. between the latent variables $t_k$

11th Sensometrics meeting - 10-13 July 2012, Rennes, France
For each partition, the most discriminant consumers attributes are highlighted (chi-square/ANOVA).

Comparison regarding the consumers attributes:

- much more discriminant information for the partition from L-CLV (as expected).
- segmentation according to the age category, opinion and knowledge on apple, purchase criteria...

11th Sensometrics meeting- 10-13 july 2012, Rennes, France
Conclusion

In external preference mapping/segmentation, by taking into account only the external information on products, no relevant information is necessarily gained with the subsequent use of the consumers attributes.

Taking into account simultaneously external information on products attributes and consumers attributes makes it possible to reveal a segmentation of consumers interpretable in terms of sociological and behavioural parameters in relation with the sensory key-drivers.

L-CLV method is suitable for this purpose (marketing research).
Thank you for your attention!
Comment: Clustering in hedonic studies

STATISTICAL PACKAGE CLUSTERING MAY NOT BE BEST FOR GROUPING CONSUMERS TO UNDERSTAND THEIR MOST LIKED PRODUCTS

RENOO YENKET¹, EDGAR CHAMBERS IV¹² and DALLAS E. JOHNSON²

Each consumer belongs to one, and only one, group (crisp algorithm).

BUT not all consumers are well represented by their group’s mean (« non typical » or « spurious » likings) .... low « cluster contribution »

OR some consumers are almost between two groups (degree of neighborhood between segments) ....... high « between-cluster position »

Cluster contribution

\[ R_{\text{own},j}^2 = \max \left\{ 1, r(y_{j \in G_k}, c_k) \right\}^2 \]

\[ R_{\text{nearest},j}^2 = \max \left( 0, \max_{\ell \neq k} r(y_{j \in G_k}, c_{\ell}) \right)^2 \]

between-cluster position

\[ 1 - R^2 \text{ ratio } j = \frac{1 - R_{\text{own},j}^2}{1 - R_{\text{nearest},j}^2} \]
Alternative approaches

**CLV with Y** without external X-block and Z-block
Interpretation of the segments from CLV on Y with external X-block, without or with cleaning

**cleaning** = discard « spurious » and/ or « between-clusters » consumers, select the consumers near their group’s center.

Here selection criterion 1-$R^2$ ratio < 0.90

\[ R_{own,j}^2 = \max \{ 0, r(y_{j \in G_k}, c_k) \} \]
\[ R_{nearest,j}^2 = \max_{\ell \neq k} \{ 0, r(y_{j \in G_k}, c_\ell) \} \]

\[ 1 - R^2 \text{ratio}_j = \frac{1 - R_{own,j}^2}{1 - R_{nearest,j}^2} \]

\( \Rightarrow 115 \) selected / 109 excluded

No clear improvement by the cleaning process

The segments can’t be better been explained by the consumers attributes, except « Age category » and « Origin » for purchase criteria which become more discriminant.