A Variation on External Preference Mapping

Sensometrics 9
July 22, 2008
Dave Plaehn
Introduction

- Major goal of product developers: Create well-liked products for given target populations
- Some potential methods that could be used in this regard (assuming existence of “competitor” products/prototypes):
• Another approach:
  – Based on external preference mapping (EPM)
  – Tries to address concerns with/criticisms of traditional EPM
  – Gives optima in original sensory space
Data Requirements

• Product attribute data (sensory/physico-chemical descriptors):
  – $Z: I \times J$ products × attributes

• Consumer Preference Data:
  – $Y, I \times N$ consumers
EPM: Criticisms/Concerns

- Many poor quality models
  - Product attribute space inadequate to model preference space (Jaeger, Wakeling and MacFie, 2000; Muellenet, Xiong and Findlay, 2007)
  - Too few or wrong latent variables used (Greenhof and MacFie, 1994; Heyd and Danzart, 1998; Faber, Mojet and Poelman, 2003)
  - Overfitting/Multicollinearity (MacKay, 2006; Faber et al, 2003)
- Problems with model selection process
  - No use of validation methods
  - Application of “first F test” (Faber et al, 2003)
- Model culling cutoff ("level of selection") possibly “too liberal” ("second F test", Faber et al, 2003)
- Other
  - Preference directions hard to interpret (Guinard, Uotani and Schlich, 2001)
  - Traditional approach did not give precise optima
  - Problems with recent “barycenter” optimization approach (Danzart, Sieffermann and Delarue, 2004; Blumenthal, 2004)
“Although these methods [MDPREF and PREFMAP] provide information about the relationship between liking by groups of consumers and sensory attributes, the optimal level of a specific attribute is not necessarily identified. From the standpoint of product development, this causes considerable problems. It is one thing to find through MDPREF or PREFMAP that saltiness drives liking, it is yet another to determine how much salt is enough or how much is too much. This is where we see preference mapping in its original forms fail to provide enough information to the product developer to formulate an optimal product from the sensory standpoint.”

## Traditional vs. New Approach

<table>
<thead>
<tr>
<th>Traditional</th>
<th>New Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Create product latent variables (LV’s) using PCA or GPA</td>
<td>• Allow for LV’s from other sources such as PLS (more relevant to preference data?)</td>
</tr>
<tr>
<td>• Use first 2 or 3 LV’s for modeling</td>
<td>• Use “many” LV’s for modeling</td>
</tr>
<tr>
<td>• Use OLS for individual models</td>
<td>• Use PLS or similar approach to avoid multicollinearity problems</td>
</tr>
<tr>
<td>• Create up to 4 models per individual (vector, circular, elliptical and full quadratic) and select “best” using a nested F-test</td>
<td>• Create all possible or many 1st and 2nd degree models across multiple LV ranges (1:2, 1:3, 1:4, …) and select “best” using cross-validation or like approach</td>
</tr>
<tr>
<td>• Cull individual models using “2nd F-test”</td>
<td>• Cull individual models using $R^2$ of validation or similar</td>
</tr>
</tbody>
</table>
### Traditional vs. New Approach, (cont.)

<table>
<thead>
<tr>
<th>Traditional</th>
<th>New Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Estimate optimal properties from preference map or use more exact method (&quot;barycenter&quot;, …)</td>
<td>• Use more sophisticated approach (&quot;gradient descent&quot;, quadratic programming, &quot;simplex&quot; genetic algorithm, …) capable of giving optima in original sensory space</td>
</tr>
</tbody>
</table>
Outline

- Apply new approach to a real data set, getting specific where necessary
- Compare results using various numbers of latent variables
  - Fitting the preference data
  - Proportion retained models
  - Optimization
  - Sensitivities
- Recap/Conclusions
Example: Muscadine Grape Juices

• Downloaded from www.uark.edu/ua/multivariate/index.htm


• 61 consumers, 10 muscadine grape juices, 15 sensory attributes, 9 trained panelists. Consumers rated each juice on a 9 point hedonic scale
# Juicy Stats

<table>
<thead>
<tr>
<th>Product Name</th>
<th>ANOVA Mean OL ($p = .027$)</th>
<th>MSD Lettering ($p &lt; 0.05$, LSD)</th>
<th>Median OL</th>
<th>Freidman Scores ($p = .143$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Beauty</td>
<td>6.9</td>
<td>'a'</td>
<td>7</td>
<td>5.8</td>
</tr>
<tr>
<td>Carlos</td>
<td>6.8</td>
<td>'ab'</td>
<td>7</td>
<td>5.6</td>
</tr>
<tr>
<td>Commercial Red</td>
<td>6.3</td>
<td>'b'</td>
<td>7</td>
<td>4.9</td>
</tr>
<tr>
<td>Commercial White</td>
<td>6.3</td>
<td>'b'</td>
<td>7</td>
<td>4.6</td>
</tr>
<tr>
<td>Granny Val</td>
<td>7.0</td>
<td>'a'</td>
<td>7</td>
<td>5.7</td>
</tr>
<tr>
<td>Ison</td>
<td>6.9</td>
<td>'a'</td>
<td>7</td>
<td>5.9</td>
</tr>
<tr>
<td>Nesbitt</td>
<td>6.6</td>
<td>'ab'</td>
<td>7</td>
<td>5.4</td>
</tr>
<tr>
<td>Southern Home</td>
<td>6.9</td>
<td>'a'</td>
<td>7</td>
<td>5.9</td>
</tr>
<tr>
<td>Summit</td>
<td>7.0</td>
<td>'a'</td>
<td>7</td>
<td>5.8</td>
</tr>
<tr>
<td>Supreme</td>
<td>6.6</td>
<td>'ab'</td>
<td>7</td>
<td>5.4</td>
</tr>
</tbody>
</table>
Determine LV method and maximum number of LV’s to use

- We used PLS, i.e. LV’s were PLS scores ($T$) from linear model $Y = ZB + E = TQ' + E$
- Try to use as many LV’s as possible (9), given execution time constraints, to increase model variety and to reduce error when transforming between manifest and latent spaces
  - Latent $\rightarrow$ Raw: $Z = TP'$
  - Raw $\rightarrow$ Latent: $T = ZR$
    where $R = W(P'W)^{-1}$
For each individual try to find “best” model per LV range for a number of ranges

Latent variables (linear terms) + Consumer k ratings

Selection process: (1) calculated all possible or used GA* (2) linear terms always in model (3) select over square and interaction terms (4) find lowest RMSECV model

- Store best model($k;2$)
- Store best model($k;3$)
- Store best model($k;9$)

(*genetic algorithm, Holland, 1975; Michalewicz, 1999)
Highest $R^2$ of validation as maximum number of LV’s ($A_{max}$) changes

<table>
<thead>
<tr>
<th>LV Range/Resp #</th>
<th>1:2</th>
<th>1:3</th>
<th>1:4</th>
<th>1:5</th>
<th>1:6</th>
<th>1:7</th>
<th>1:8</th>
<th>1:9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.27</td>
<td>0.82</td>
<td><strong>0.84</strong></td>
<td>0.79</td>
<td>0.84</td>
<td>0.77</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>-0.40</td>
<td><strong>0.89</strong></td>
<td>0.82</td>
<td>0.81</td>
<td>0.58</td>
<td>0.63</td>
<td>0.65</td>
<td>0.68</td>
</tr>
<tr>
<td>3</td>
<td>-0.31</td>
<td>0.68</td>
<td>0.86</td>
<td>0.79</td>
<td><strong>0.99</strong></td>
<td>0.81</td>
<td>0.70</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>-0.02</td>
<td>0.09</td>
<td>0.20</td>
<td>0.43</td>
<td><strong>0.92</strong></td>
<td>0.61</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>5</td>
<td>0.47</td>
<td>0.72</td>
<td><strong>0.86</strong></td>
<td>0.79</td>
<td>0.58</td>
<td>0.62</td>
<td>0.63</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>-0.23</td>
<td>-0.03</td>
<td>-0.10</td>
<td>0.31</td>
<td>0.65</td>
<td><strong>0.78</strong></td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>7</td>
<td>0.35</td>
<td><strong>0.89</strong></td>
<td>0.75</td>
<td>0.75</td>
<td>0.74</td>
<td>0.77</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>8</td>
<td>0.33</td>
<td>0.45</td>
<td>0.59</td>
<td><strong>0.93</strong></td>
<td>0.85</td>
<td>0.70</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>9</td>
<td>-0.29</td>
<td>-0.25</td>
<td>0.54</td>
<td>0.71</td>
<td><strong>0.76</strong></td>
<td>0.38</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td>10</td>
<td>-0.24</td>
<td>-0.03</td>
<td>-0.22</td>
<td>0.59</td>
<td>0.83</td>
<td>0.90</td>
<td><strong>0.92</strong></td>
<td>0.72</td>
</tr>
</tbody>
</table>

** = maximum per individual
Maximum $R^2$ of Validation as a Function of Number of LV’s
(using a threshold of 0.7 for the LV range {1:9})
Cull models based on $R^2$ of validation threshold

Threshold ($\kappa$) picked “by inspection” with an aim of achieving both a high overall $R^2$ of validation and a high proportion of retained models.
Overall Model Stats
(based on retained models, $\kappa = 0.7$)
Optimization

• Goal: Find product profile in original space that maximizes (mean/median) predicted liking for a given population
• Lots of approaches (Fletcher, 2000)
• Used fmincon.m, constrained optimization, from Matlab® Optimization Toolbox (Mathworks, 2008) with multiple starting points
• Used various constraints to avoid model extrapolation
Optimization: Objective Function

- An optimization program tries to minimize or maximize a so-called *objective function*, $f$.
- For the case at hand:
  - **Inputs:**
    - Sensory profile, $z$
    - Individual models (regression coefficients, LV space for best model, transformation for raw to latent space) for group of interest
  - **Output:** *mean (or median …) predicted liking subject to constraints.*
Optimization: Constraint Options

• Apply constraints to prevent model extrapolation

• Possible (optional) constraints
  – Optima “candidates” (z) must be within Z data bounds.
  – t = zR must be in associated LV bounds
  – Individual model predictions must be within given range e.g. [1 9].

I want to extrapolate!!
Optimization: Consistency of Optima

• As one changes the maximum number of LV’s \((A_{\text{max}})\) how similar are the resulting optima?
• We ran the cases where \(A_{\text{max}} = 4, 5, \ldots, 9\)
Overall optima obtained using best models for first 4, 5, …, 9 LV’s above threshold (0.7)

<table>
<thead>
<tr>
<th>LV Range</th>
<th>{1:4}</th>
<th>{1:5}</th>
<th>{1:6}</th>
<th>{1:7}</th>
<th>{1:8}</th>
<th>{1:9}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Retained</td>
<td>0.41</td>
<td>0.74</td>
<td>0.84</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Ckd Musc aroma</td>
<td>6.87</td>
<td>4.42</td>
<td>6.20</td>
<td>6.36</td>
<td>5.02</td>
<td>6.84</td>
</tr>
<tr>
<td>Ckd Grape aroma</td>
<td>0.55</td>
<td>2.74</td>
<td>1.11</td>
<td>1.63</td>
<td>2.02</td>
<td>1.85</td>
</tr>
<tr>
<td>Musty aroma</td>
<td>1.48</td>
<td>0.77</td>
<td>1.43</td>
<td>1.50</td>
<td>1.40</td>
<td>1.48</td>
</tr>
<tr>
<td>Sweet</td>
<td>8.01</td>
<td>7.25</td>
<td>8.19</td>
<td>7.92</td>
<td>8.19</td>
<td>7.96</td>
</tr>
<tr>
<td>Sour</td>
<td>5.68</td>
<td>6.44</td>
<td>5.12</td>
<td>4.58</td>
<td>5.29</td>
<td>5.53</td>
</tr>
<tr>
<td>Cooked Muscadine</td>
<td>8.55</td>
<td>5.51</td>
<td>7.36</td>
<td>7.76</td>
<td>6.47</td>
<td>7.57</td>
</tr>
<tr>
<td>Cooked Grape</td>
<td>0.65</td>
<td>3.50</td>
<td>1.60</td>
<td>1.59</td>
<td>2.86</td>
<td>1.75</td>
</tr>
<tr>
<td>Musty</td>
<td>1.76</td>
<td>0.86</td>
<td>1.77</td>
<td>1.78</td>
<td>1.79</td>
<td>1.83</td>
</tr>
<tr>
<td>Green Unripe</td>
<td>1.81</td>
<td>1.49</td>
<td>0.50</td>
<td>0.52</td>
<td>0.45</td>
<td>1.19</td>
</tr>
<tr>
<td>Floral</td>
<td>0.32</td>
<td>0.43</td>
<td>0.78</td>
<td>0.69</td>
<td>0.87</td>
<td>0.65</td>
</tr>
<tr>
<td>Apple/Pear</td>
<td>0.98</td>
<td>0.96</td>
<td>1.12</td>
<td>1.24</td>
<td>1.06</td>
<td>1.14</td>
</tr>
<tr>
<td>Fermented</td>
<td>0.18</td>
<td>0.45</td>
<td>0.64</td>
<td>0.70</td>
<td>0.71</td>
<td>0.56</td>
</tr>
<tr>
<td>Metallic (af)</td>
<td>0.72</td>
<td>0.99</td>
<td>0.65</td>
<td>0.69</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td>Astringent</td>
<td>6.83</td>
<td>6.66</td>
<td>6.80</td>
<td>6.86</td>
<td>6.78</td>
<td>6.50</td>
</tr>
<tr>
<td>Metallic (ff)</td>
<td>0.51</td>
<td>1.20</td>
<td>0.57</td>
<td>0.44</td>
<td>0.74</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean predicted OL</td>
<td>7.53</td>
<td>7.12</td>
<td>7.20</td>
<td>7.21</td>
<td>7.21</td>
<td>7.22</td>
</tr>
</tbody>
</table>
MDS plot of products and optima
(with LSD multiple comparison lettering)

In most cases consumers equally likes [sic] products with really different sensory properties.

(Danzart, Sieffermann and Delarue, 2004)

These optima are all over the place! Wha’s up?
Euclidean distances of optima from given

Black Beauty
Carlos
Commercial Red
Commercial White
Granny Val
Ison
Nesbitt
Southern Home
Summit
Supreme

LV{1:4} LV{1:5} LV{1:6} LV{1:7} LV{1:8} LV{1:9}

Lighter → Closer
How ‘stable’ are the liking predictions for the various optima across the various LV model ranges?
Understanding product attributes: Sensitivities

- The objective function, $f$, can be used to create “sensitivity plots”
- Attribute “sensitivities” are used to try to estimate an attribute's importance
- They are conditioned on a given estimated optimal profile i.e. they are “local” and are thus possibly of limited value.
- With all other attributes fixed the attribute of interest is sampled across its range and predicted liking is calculated for each sample point.

\[
\text{sensitivity} = \frac{\text{sign( change in } f \text{ ) \times range of } f}{\text{range of } Z \text{ variable}}
\]
Sensitivities (cont.)

- Floral: 0.65
- Metallic (af): 0.51
- Musty: 0.37
- Musty aroma: 0.35
- Metallic (ff): 0.13
- Fermented: 0.11
- Sour: 0.11
- Cooked Muscadine: 0.04
- Green Unripe: 0.04
- Sweet: 0.03
- Ckd Grape aroma: 0.02
- Cooked Grape: 0.02
- Ckd Musc aroma: -0.03
- Apple/ Pear: -0.07
- Astringent: -0.58

Sensitivities (values range from -0.60 to 0.60)
Sensitivity Plots for LV 1:9 Optimum

Although it's possible for the optimization program to get "stuck" at a "local" optimum, at least one constraint violation occurred at the points in question here.

How come the optimum isn't at the highest liking? Check out **Sour**, for example.
Sensitivity Plots (cont.)

- **Cooked Grape**
  - ZOpt, YOpt
  - Plot shows a steady increase from 6.7 to 7.4

- **Musty**
  - ZOpt, YOpt
  - Plot shows a linear increase from 6.7 to 7.4

- **Green Unripe**
  - ZOpt, YOpt
  - Plot shows a steady increase from 6.7 to 7.4

- **Floral**
  - ZOpt, YOpt
  - Plot shows a peak at 6.7, with a decrease to 6.9

- **Apple/Pear**
  - ZOpt, YOpt
  - Plot shows a steady decrease from 7.4 to 6.7

- **Fermented**
  - ZOpt, YOpt
  - Plot shows a peak at 7.4, with a decrease to 7.1
Sensitivity Plots (cont.)

Metallic (af) (Z_{Opt}, Y_{Opt})

Astringent (Z_{Opt}, Y_{Opt})

Metallic (ff) (Z_{Opt}, Y_{Opt})
Conclusions

• A method of estimating optimal products given product attribute and consumer liking data was presented:
  – Roughly based on EPM
  – Non-parametric
  – Model selection and culling using validation results, hopefully providing a robust “overall model”
  – “Sophisticated” optimization approach used allowing for several types of constraints
  – Currently, very computationally intensive

• As applied to the Muscadine grape juice data, the method
  – Fit the liking data well in both a calibration and validation sense
  – A large proportion of the individual models were retained with the higher LV ranges.
  – Some apparent model stabilization beginning with the {1:6} models; despite differences between optima for {1:6}—{1:9} the predicted liking for these optima changed little across these same models.
  – Utility of optima need to be confirmed
  – Took about 20 hours on 1 processor, 5 hours with 4 processors in parallel.
Future Work

• Speed up model selection process
• Find more exact method of calculating $R^2$ of validation threshold e.g. maximize $(proportion\ retained) \times (R^2 \ of \ validation)$
• Validate method with follow-up studies
• Compare to alternatives
Acknowledgements

Thanks to Mary Browning and Michelle Funk and others at Insights Now, Inc. for their help in creating this presentation.
References

Optimization: Objective function, $f$:

Sensory profile

$Z$

Individual models

$R$

LV model ranges

$z$

Change to LV space: $t = zR$

For each model:

Truncate $t$ to appropriate length

Make appropriate squares & interactions

Make individual prediction

$y = \text{mean of predictions}$

Stored info

$y$

Achtung!

$z$'s can be rejected for violating various (optional) constraints.