

Meta-Attributes in Sensory Descriptive Analysis B.C. Franczak¹, P.D. McNicholas¹ and C. Findlay²

¹Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario, Canada ²Compusense Inc., Guelph, Ontario, Canada



Introduction

- Results
- The objective of this project is to determine if there are groups of homogeneous attributes, which we will call "meta-attributes" (MAs), present in two data sets.
- These data sets resulted from two descriptive analyses, one in 2010 and one in 2014, conducted by trained panels on potato varieties with 52 well-defined attributes. An example of the features of these data sets is given in Table 1.
- We fitted the CUU model for $G=1, \ldots, 10$ components, $q=1, \ldots, 20$ latent factors, and initialized each model using two different starting points.
- We choose the best fitting model using the Bayesian Information Criterion (BIC; Schwarz, 1978).

The 2010 data set



Table 1: A subset of the mean panelists' scores for each attribute of the potato varieties evaluated in 2014.

	C1	A1	A2	A3	A4	A5	A6
1	49.80	48.45	32.70	22.95	29.35	46.70	15.00
2	62.55	50.05	32.80	20.90	21.75	50.05	18.15
3	29.85	46.35	29.70	28.85	33.55	31.95	8.45
4	21.65	48.50	26.35	18.80	25.15	27.75	8.60
5	56.70	49.25	32.70	20.90	21.05	39.50	13.70
6	30.35	46.15	31.00	22.85	26.15	40.10	8.05
7	44.40	48.60	27.00	26.60	23.85	33.35	7.15
8	59.95	51.70	34.35	29.20	22.70	26.70	7.45
9	40.45	45.25	29.50	18.95	20.20	40.50	12.30
10	40.65	45.65	24.60	21.25	27.70	41.30	10.55

Methodology

• A model-based clustering approach is utilized to find the MAs; specifically, we define each MA using a constrained mixture of factor analyzers model called the CUU model (see Figure 1 for an example).



• For each starting point, the BIC selected CUU models with identical features (see Table 2) with each suggesting that there are two MAs.

Table 2: Features of the best fitting CUU models (A,B) for the 2010 potato varieties

	MAs	q	BIC	Magnitude of each MA
A	2	16	-140485.67	G1 = 22, G2 = 30
B	2	16	-140743.95	$ { m G1} =23$, $ { m G2} =29$

• Comparing each result shows that both models classified 45 of the 52 attributes the same way (Table 3)

Table 3: A cross tabulation of the MAs selected by the best fitting CUU models given in Table 2.

		E	3
		MA1	MA2
Δ	MA1	19	3
A	MA2	4	26

The 2014 data set

- For each initialization strategy, the BIC selected a different CUU model (see Table 4).
 - Table 4: Features of the best fitting CUU models (A,B) for the 2014 potato varieties

Figure 2: The direction of each attribute in the space of the first two principal components.

- The PCA does not give any evidence that the other attributes that make up MA3 are over-expressed, nor does it suggest that they are acting together.
- Refitting the CUU model, for $G = 1, \ldots, 3$ components and $q = 1, \ldots, 10$ latent factors, to a subset of the 2014 data set that includes only the attributes that make up MA3 provides no evidence that the considered attributes are not homogeneous with each other.
- We also refitted the CUU model, under the same conditions given in the previous point, to the entire 2014 data set with only *Flesh* Colour removed. This result is summarized in Table 6.

Table 6: Features of the best fitting CUU models for the 2014 potato varieties with *Flesh Colour* removed.



Figure 1: A bivariate example of a two-component CUU model. We use the characteristics of each component of the curve to define a group of observations.

• The CUU model has density

 $f\left(\mathbf{x}\mid \boldsymbol{artheta}
ight) = \sum_{g=1}^{G} \pi_{g} \phi_{p}\left(\mathbf{x}\mid \boldsymbol{\mu}_{g}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}_{g}
ight),$

where $\pi_g > 0$, s.t., $\sum_{g=1}^G \pi_g = 1$ are the mixing proportions and each component density function, $\phi_p\left(\mathbf{x} \mid \boldsymbol{\mu}_q, \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}_g
ight)$, is multivariate Gaussian.

• From a practical viewpoint, this model lets groups share covari-

	MAs	q	BIC	Magnitude of each MA
A	2	16	-129034.58	G1 = 18, $ G2 = 34$
В	3	4	-142497.05	$ { m G1} =7$, $ { m G2} =21$, $ { m G3} =24$

• This time, we notice that each model has identified a different number of MAs. However, like before, we notice a large amount of overlap between the each MA (see Table 5).

Table 5: A cross tabulation of the MAs selected by the best fitting CUU models given in Table 4.

			В	
		MA1	MA2	MA3
٨	MA1	4	14	0
A	MA2	3	7	24

• For the 2010 data set, the results of a PCA do not suggest that any individual attribute is acting independently of the rest. However, for the 2014 data set, one attribute, *Flesh Colour*, is strongly correlated with the first PC and is deviating from the other attributes (see Figure 2).

• This result raises the following set of questions:

1. For model 2, is MA3 composed of over-expressed attributes? 2. Should *Flesh Colour* be considered its own group?

	MAs	q	BIC	Magnitude of each MA
A	2	15	-125895.5	$ {\sf G1} =19$, $ {\sf G2} =32$

• This result provides more evidence that the attributes within MA3, aside from *Flesh Colour*, are homogeneous with one another.

Conclusions

• There is evidence that, within these data sets, there are at least two MAs.

• Comparing the best fitting CUU models between both data sets indicates that 38 attributes were classified in the same way (Table 7).

Table 7: A cross tabulation of the MAs selected by the first models listed in Tables 2 and 4.

		A4		
		MA1	MA2	
٨٥	MA1	13	9	
A2	MA2	5	25	

• These MAs could be used to construct a sensory informed design (SID, Franczak et al., 2015) or when performing quality control.

Acknowledgements

ances while allowing variances to differ from group-to-group.

• In addition to fitting the CUU model, we also use principal component analysis (PCA) to determine if any individual attributes are deviating from the other attributes.

3. If we remove *Flesh Colour* are the remaining attributes homogenous?

• To address each of these questions we use the results of the PCA and refitted the CUU model with the suspect attribute removed.

This work is supported by a grant-in-aid from Compusense Inc., as well as Collaborative Research and Development Grant from the Natural Sciences and Engineering Research Council of Canada.