A Preliminary Review of Multiple Group Principal Component Analysis for Descriptive Sensory Data

Moyi Li[1], Ryan P. Browne[1], Paul D. McNicholas[1], John C. Castura[2]
[1]University of Guelph, Guelph, Ontario, Canada
[2]Compusense, Guelph, Ontario, Canada

Principal component analysis (PCA) is frequently used to analyse sensory descriptive analysis data to better understand the multivariate sensory space. Consider that even well-trained descriptive sensory panelists might retain some distinctive characteristics, including a tendency to use somewhat different scale levels and ranges than other panelists. Panelists might also show other innate differences in sensitivity to particular attributes or differences in response patterns due to attribute understanding. Often these differences are averaged out prior to conducting PCA. We explored multiple group principal component analysis (MGPCA; Thorpe, 1988) as an alternative multivariate approach.

MGPCA is a relatively simple technique related to canonical variate analysis (CVA; Hotelling, 1936; Thorpe, 1988). Where PCA might perform singular value decomposition on the variance-covariance matrix obtained (conventionally) from panel averages, MGPCA can be performed by singular value decomposition of a pooled variance-covariance matrix derived from the weighted average of the panelists' variance-covariance matrices. MGPCA provides a within-class analysis that derives a consensus sensory space in which the individual panellist responses for products are also represented. Agreement amongst panelists is readily evaluated by inspection.

In this respect, MGPCA provides richer output than PCA. It derives a similar consensus space as generalized Procrustes analysis (GPA) without performing translation, rotation, isotropic scaling transformations. This preliminary investigation reveals some advantages to MGPCA for sensory data, and interpreted results from previous descriptive analysis studies were comparable to those obtained from other multivariate approaches, indicating that the MGPCA approach warrants further investigation.

A high level comparison of three multivariate analysis methods follows…

PCA

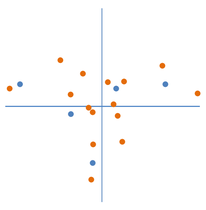
1. Obtain average (or median) data over all panelists and reps

Attributes / Products: $\bar{X}$

2. Obtain covariance matrix based on average data

$S_{\bar{X}}$

3. Decompose covariance matrix $S_{\bar{X}}$ (e.g. SVD decomposition)



MGPCA

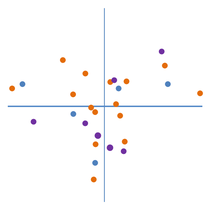
1. Obtain covariance matrix for each panelist (averaging over reps)

S_1 . . . S_N

2. Obtain average covariance matrix

$\bar{S}$

3. Decompose covariance matrix $\bar{S}$ (e.g. SVD decomposition)

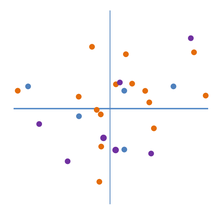


GPA

1. Average over reps

Panelist 1 Attributes / Products . . . **Panelist n** Attributes / Products

2. Translate data (centering)
3. Iterative rotation/reflection and scaling until convergence (using loss function and threshold)
4. PCA (for presentation of results)



5. Permutation test to evaluate solution

A (toy) example:

	X_1		X_2		$\bar{X}$		$S_{\bar{X}}$	
	Sweet	Sour	Sweet	Sour	Sweet	Sour	Sweet	Sour
1	10	10	30	10	20	10	0	0
2	20	20	20	20	20	20	0	100
3	30	30	10	30	20	30		

PCA

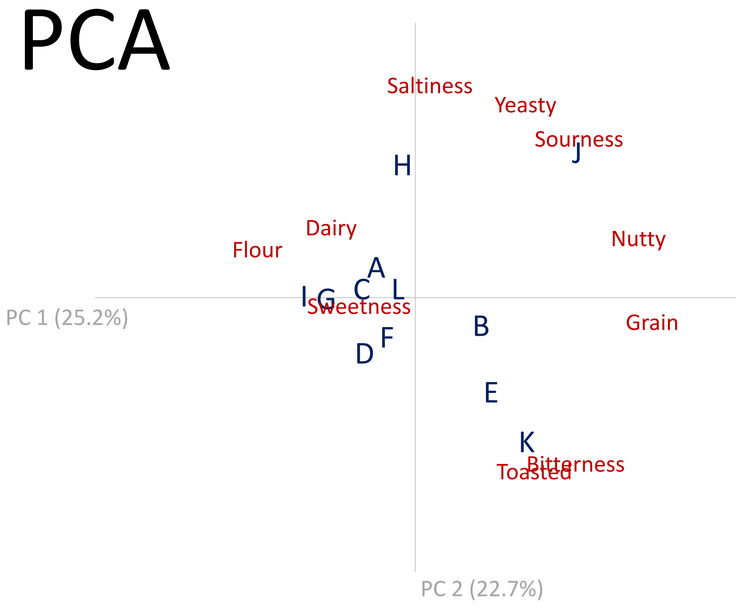
	S_1		S_2		$\bar{S}$	
	Sweet	Sour	Sweet	Sour	Sweet	Sour
	100	100	100	-100	100	0
	100	100	-100	100	0	100

MGPCA

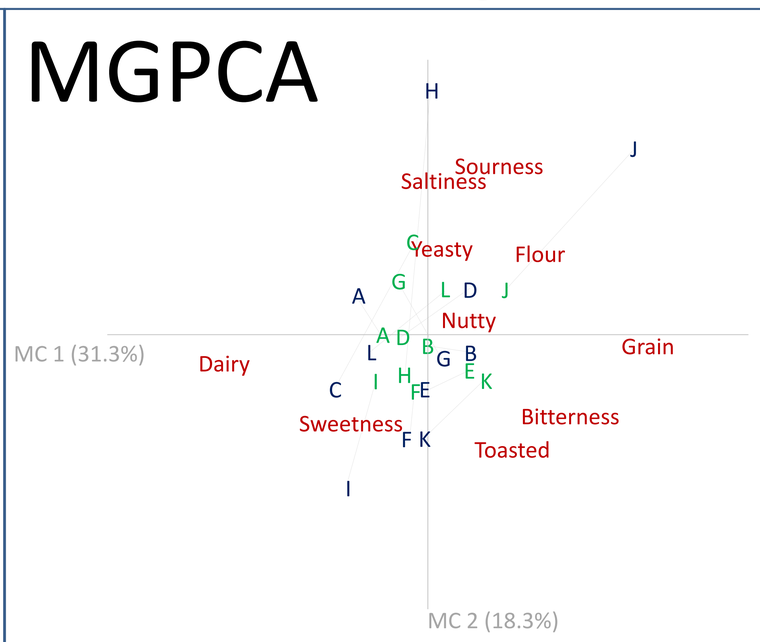
Is a more appropriate estimate of variance of "sweet" 0 (as in PCA) or 100 (as in MGPCA)?

Descriptive sensory data from panel trained to evaluate whole grain bread (flavour data, reduced to 2 dimensions)

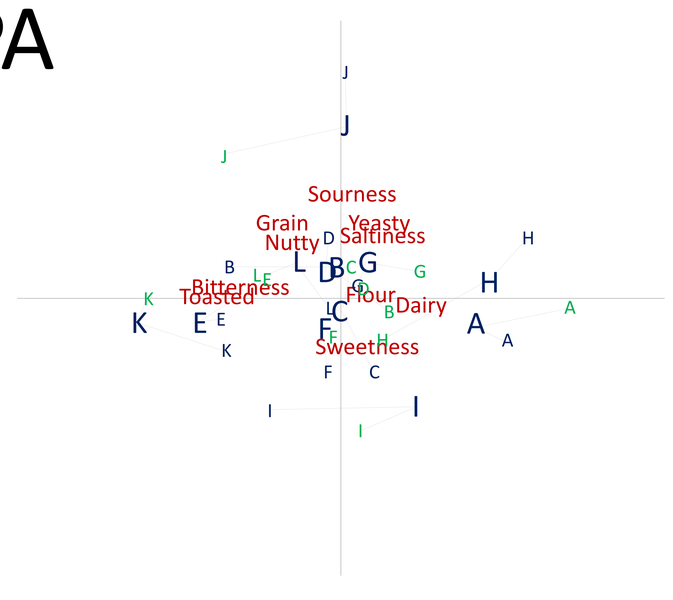
PCA



MGPCA



GPA



References:
Hotelling, H. (1936) Relations between two sets of variates. *Biometrika, 28,* 321-377
Thorpe, R.S. 1988. Multiple Group Principal Component Analysis and Population Differentiation. *Journal of Zoology, 216*, 37-40.

Pangborn 2013
10th Pangborn Sensory Science Symposium
Windsor Barra Hotel, Rio de Janeiro, Brazil 11 – 15 August 2013

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