

Stabiliser des cartes de préférences en débruitant les données d'appréciation

Rebhi Ibtihel¹²

¹Unité signaux et systèmes, Ecole Nationale d'Ingénieurs de Tunis, Université Tunis ElManar

²Ecole Supérieure de la Statistique et de l'Analyse de l'Information, Université Carthage

E-mail : ibtihelrebhi@yahoo.fr

Résumé :

Dans le domaine de l'analyse sensorielle, nous pouvons considérer les données d'appréciation des consommateurs comme une structure d'un signal corrompu par un bruit. Ce bruit s'explique par le rang de dégustation du produit, le moment de la dégustation, l'arrière-effet du produit précédent, etc.

Nous proposons dans ce travail un débruitage de ce type de données par la méthode d'Analyse en Composantes Principales régularisée permettant de mieux reconstituer et visualiser le signal sous-jacent en fournissant une meilleure estimation au sens de l'erreur quadratique moyenne du signal (Verbranck *et al.*, 2013).

Nous construisons alors une carte des préférences en utilisant les données d'appréciation débruitées et les données de caractéristiques sensorielles de ces produits. La carte est obtenue en réalisant une ACP sur les données sensorielles, en projetant les préférences en supplémentaire et en modélisant par régression quadratique pour chaque consommateur les notes données aux différents produits (Danzart, 2004).

Nous évaluons, sur un exemple de données de biscuits, que l'élimination du bruit par ACP régularisée améliore la qualité de cartographie comparée à celle sans débruitage. Nous évaluerons les qualités de cette méthode par des simulations.

Mots-clés : Débruitage, ACP régularisée, Données de consommateurs, Cartographie externe des préférences, Stabilité.

Abstract :

In sensory analysis, we can consider the data of consumer ratings as a structure of a signal corrupted by noise. This noise is due to the row, the moment, the effect of the previous tasting product, etc.

We propose in this paper a denoising of consumers data given by regularized Principal Component Analysis method to better reconstruct and visualize the underlying signal by providing a better estimate in the sense of the mean square error signal (Verbranck, 2013).

We carry after an external preference mapping constructed from two types of data: the assessments data already denoised and sensory characteristics data of these products. The map is obtained by performing a PCA on sensory data, projecting in supplementary the preferences and modeling by quadratic regression for each consumer the ratings given to different products (Danzart , 2004).

We show, on data cookies tasted by Pakistani, that the elimination of noise corrected by regularized PCA improves the quality of mapping compared to the noised one. We will assess the qualities of this method by simulations.

Keywords : Denoising, Regularised PCA, Consumers data, External preference mappin, stability.

1. Introduction

Recueillir des mesures d'appréciation est une tâche difficile. D'une part cette mesure est subjective et tous les consommateurs n'ont pas les mêmes préférences. La moyenne, qui est une mesure très stable, ne peut pas être utilisée sur des données de préférences car cette moyenne n'a pas nécessairement de sens. Imaginons le cas d'une dégustation de jus d'orange. Certains consommateurs apprécient fortement les jus d'orange très pulpeux et d'autres les jus d'orange sans pulpe. La moyenne des appréciations conduirait à dire que l'ensemble des individus apprécie des jus d'orange moyennement pulpeux, ce qui ne convient à personne. On est donc amené à conserver le caractère individuel de l'appréciation lorsqu'on construit des cartes de préférences. D'autre part, la stabilité des jugements hédoniques individuels est mauvaise. En effet, un même juge ne donnera pas les mêmes notes d'appréciation à un même ensemble de produits lors de deux séances de dégustations. Ceci peut s'expliquer par de multiples raisons : le rang de dégustation des produits, l'arrière-effet (l'effet du produit précédent : si le produit précédent était très peu apprécié, le produit présent risque d'être sur-noté), le moment de la dégustation, le fait que d'un jour à l'autre ou d'un moment à l'autre on n'apprécie pas les mêmes saveurs, etc. Ce manque de stabilité des jugements hédoniques influe aussi sur les cartes de préférences.

Les cartes de préférences sont construites de la façon suivante (Danzart, 2004): une carte des produits est obtenue en réalisant une ACP sur les données sensorielles; les préférences individuelles sont ensuite utilisées pour prédire en tout point de la carte la préférence d'un individu; la carte d'un individu est ensuite discrétisée en 2 parties (une zone d'appréciation et une zone de non-appréciation). Les cartes individuelles sont ensuite sommées pour obtenir la carte des préférences du jury de consommateurs.

L'idée de l'exposé est de réduire le bruit présent dans les données d'appréciation afin d'utiliser des mesures qui se rapprochent du signal avant de construire des cartes de préférences. Ce débruitage doit permettre de stabiliser les cartes de préférences.

2. Débruitage par ACP régularisée

Un modèle classique pour l'ACP est le modèle à effets fixes introduit par (Caussinus, 1986) qui est un modèle de type signal plus bruit.

Une matrice X_{np} de données d'appréciation peut se décomposer en un signal auquel s'ajoute un bruit gaussien. Sa structure sera donc :

$$X_{np} = \tilde{X}_{np} + \varepsilon_{np}$$
$$x_{ij} = \sum_{s=1}^S \sqrt{d_s} q_{is} r_{js} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- d_s est la $s^{\text{ième}}$ valeur propre de $\tilde{X}\tilde{X}$
- $\tilde{X}\tilde{X}$ n fois la matrice variance-covariance celle du signal
- $r_s = \{r_{1s} \dots r_{js} \dots r_{ps}\}$ le vecteur propre associé
- $q_s = \{q_{1s} \dots q_{is} \dots q_{ns}\}$ le $s^{\text{ième}}$ vecteur propre de $\tilde{X}\tilde{X}$

Plusieurs traitements ont été réalisés à partir de cette structure de données : effectuer une décomposition en valeurs singulières de X , minimiser l'erreur quadratique moyenne du signal et estimer le terme de seuillage. Ainsi l'ACP régularisée est définie par la multiplication de la solution de maximum de vraisemblance par ce terme de seuillage ce qui conduit à obtenir une matrice dont les données sont les plus proches du signal reconstitué et visualisé (Verbrancq *et al.*, 2013).

3. Application sur le jeu de données biscuits

Afin d'évaluer l'intérêt de débruiter les données de préférences avant de construire une carte de préférences, nous avons réaliser des simulations pour comparer les cartes obtenues avec et sans débruitage. Pour évaluer les cartes, nous avons utiliser un jeu de données réel que nous avons diviser en 2 sous-échantillons. Nous avons alors construit les cartes de préférences sur chaque sous-échantillon et comparons la stabilité des cartes d'un sous-échantillon à l'autre quand les données sont débruitées et quand elles ne le sont pas.

Nous disposons d'un jeu de données de 8 biscuits (4 biscuits fabriqués au Pakistan et 4 biscuits Français). Une évaluation sensorielle a été organisée au Pakistan pour caractériser ces produits. Un jury sensoriel constitué de 12 étudiants (entraînés) a évalué chacun des 8 biscuits lors de 2 séances de dégustation selon une liste de 23 descripteurs. Les descripteurs sont les suivants : couleur, épaisseur, odeur d'œuf, odeur de lait, odeur de beurre, odeur vanille, odeur caramel, odeur citron, dureté-croquant, friabilité, granuleux-sableux, farineux, craquant-croustillant, pâteux, fondant, goût salé, goût sucré, goût œuf, goût lait, goût beurre, goût vanille, goût caramel, goût citron. Les notes varient entre 0 et 10 ; une note de 0 signifie que l'odeur ou la saveur évaluée est extrêmement faible et une note de 10 signifie qu'elle est très importante.

Par ailleurs, un jury de 150 consommateurs Pakistanais a donné une note d'appréciation pour chaque biscuit allant de 0 (je n'aime pas du tout ce biscuit) à 10 (j'aime beaucoup ce biscuit) (Pagès, *et al.*, 2007).

Nous avons alors constitué deux sous-échantillons indépendants de 72 consommateurs choisis aléatoirement, dans laquelle chaque répondant évalue chaque produit.

Nous avons appliqué l'approche de débruitage sur ces deux échantillons. Les cartes obtenues après débruitage construisent un espace sensoriel stable où les zones de préférences paraissent plus claires et les lignes de niveaux sont moins erratiques que celles obtenues à partir des données bruitées [Fig. 1].

A travers cette étude, on montre qu'un débruitage des données hédoniques permet de stabiliser les cartes de préférences.

Bibliographie

[1] Marie Verbanck, Julie Josse, François Husson(2003), Regularised PCA to denoise and visualise data

[2] Caussinus H (1986) Models and uses of principal component analysis (with discussion), *DSWO Press*, 149–178

[3] Pagès, J., Bertrand, C., Ali, R., Husson, F. & Lê, S. (2007). Sensory analysis comparison of eight biscuits by French and Pakistani panels. *Journal of Sensory Studies*. 22 (6), 665-686.

[4] Daniel M. Ennis and John M. Ennis (2013), Mapping Hedonic Data: A Process Perspective, Published in *Journal of Sensory Studies*, 28, 324-334.

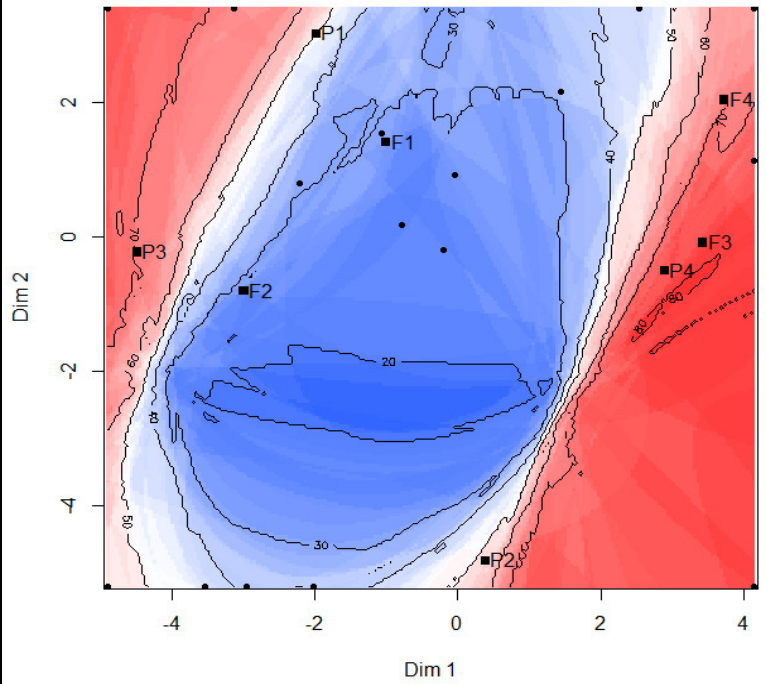
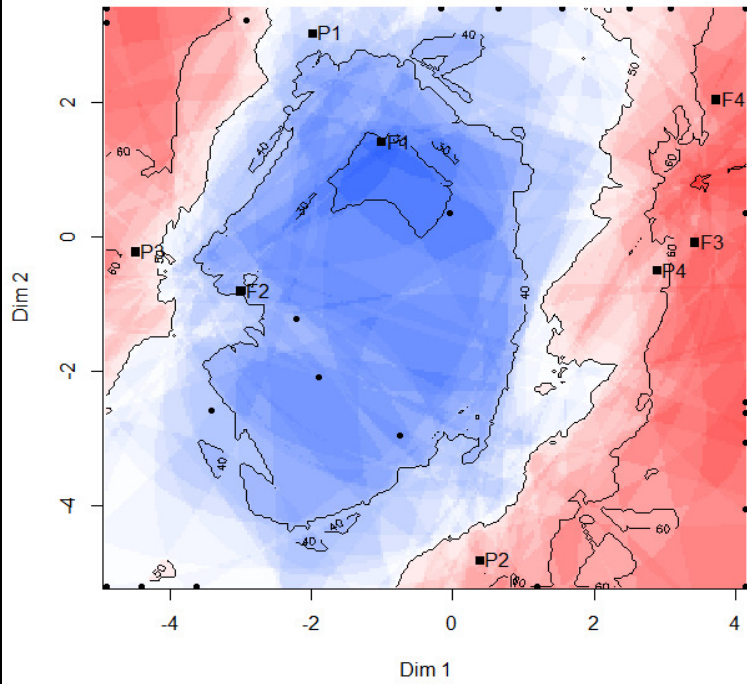
[5] Danzart M., Sie_ermann J.M., Delarue J. (2004). New developments in preference mapping techniques: finding out a consumer optimal product, its sensory profile and the key sensory attributes. 7th Sensometrics Conference, July 27-30, 2004, Davis, CA

Cartographie sur données bruitées

Cartographie sur données débruitées

Cartographie sur données bruitées (Echantillon1)

Cartographie sur données débruitées (Echantillon1)



Cartographie sur données débruitées (Echantillon2)

Cartographie sur données débruitées (Echantillon2)

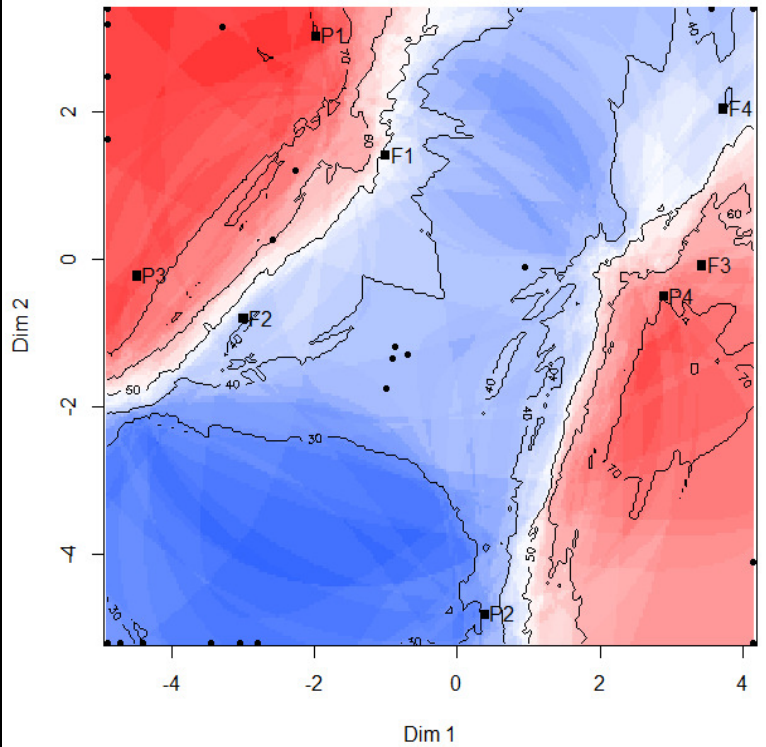
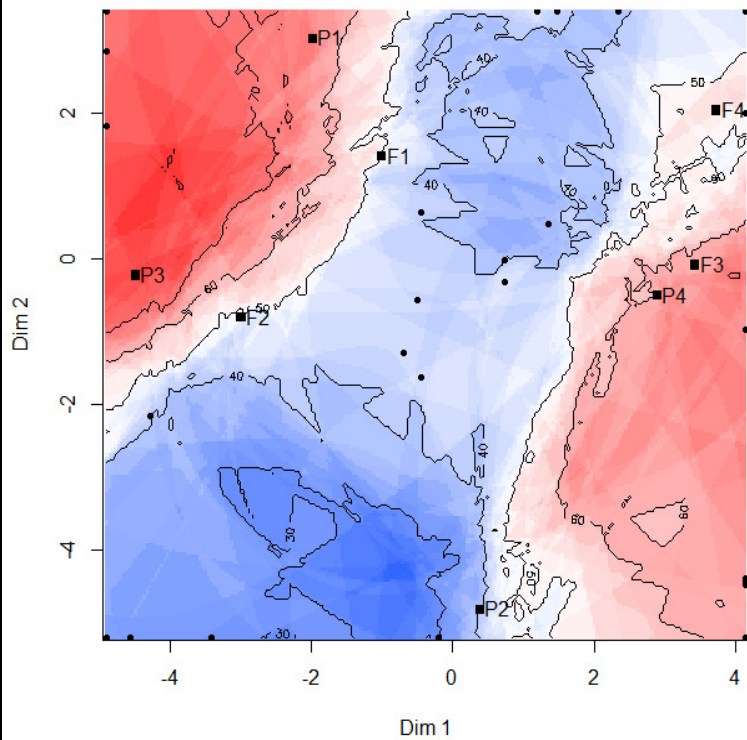


Fig. 1 : Cartographie sur données bruitées et débruitées

