

Clustering of categorical variables around latent components

Application to consumer survey data

Justine Chevalier⁽¹⁾, Marlène Herbreteau⁽¹⁾, Agnès Tariel⁽¹⁾

⁽¹⁾ ONIRIS - Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering

► Data set = Consumer survey data

- ✓ Focus on 8 categorical variables / 799 subjects
 - ✓ « If you were a brand what would you be? » (LU, Innocent,...)
 - ✓ « If innovation was a verb what would it be? » (to create, to discover, ...)
 - ✓ ...

► Objectives

- ✓ Compare HCA with different association criteria, and a latent component approach « ClustOfVar⁽¹⁾ »

(1) Vanessa Kuentz, PhD thesis, 2009

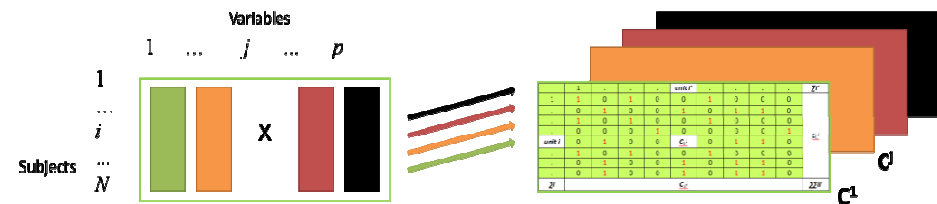
- ✓ Propose a clustering strategy around latent components based on « Statis » compromise both to summarize categorical variables and to display a factorial representation

Clustering of categorical variables

Clustering on the basis of latent variables

► Methods

- Encoding into a relational matrix (Condorcet)



- Dissimilarity matrix based on Rand, Adjusted Rand, Tschuprow, Janson Végélius ...

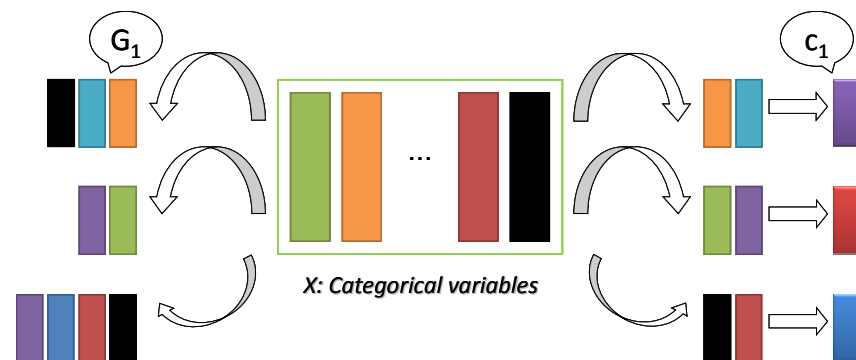


Fig.1 : Representation of the two approaches

Clustering of categorical variables around latent components

Application to consumer survey data

Justine Chevalier⁽¹⁾, Marlène Herbreteau⁽¹⁾, Agnès Tariel⁽¹⁾

⁽¹⁾ ONIRIS - Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering

► Clustering of variables based on « Statis » compromise

- ✓ From the matrices C^j , matrix W_k associated with group G_k is obtained by minimizing the criteria

$$\min \sum_k \sum_{j \in G_k} \|C^j - \alpha_j W_k\|^2$$

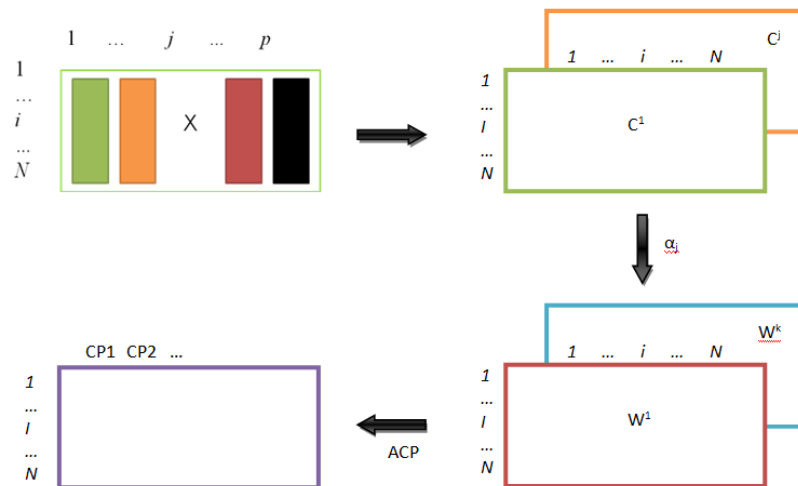


Fig.2 : Computation of the compromise

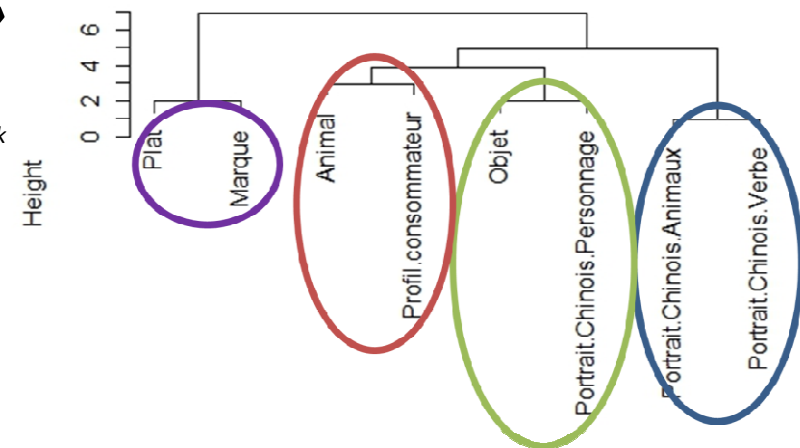


Fig.3 : Dendrogram based on Statis compromise

► Comparison of the results

- ✓ Dendrograms obtained by ClustOfVar, the Statis compromise and the Tschuprow criterion look quite similar
- ✓ More investigations should be undertaken in order to compare the different strategies
- ✓ Statis compromise could be extended to other association criteria