Three-way clustering around latent variables approach with constraints to improve configurations’ interpretability

Véronique Cariou\textsuperscript{a}, Tom F. Wilderjans\textsuperscript{b,c}

\textsuperscript{a}StatSC, ONIRIS, INRA, 44322 Nantes, France
\textsuperscript{b}Methodology and Statistics Research Unit, Leiden University, The Netherlands
\textsuperscript{c}Research Group of Quantitative Psychology and Individual Differences, Leuven, Belgium

XXVIe Rencontres de la Société Francophone de Classification
3 – 5 sept 2019, Nancy
Clustering of variables around latent components with CLV:

✓ Understand the underlying structures
✓ Detect redundancies between variables
✓ Reduce the number of variables

Maximize

\[ T^{(Q)} = \sum_{q=1}^{Q} \sum_{j=1}^{J} \delta_{qj} \text{cov}^2(x_j, c_q) \]

with \( \text{var}(c_q) = 1 \)

Introduction

Extension of CLV to three-way arrays with CLV3W:

Connexion with clusterwise models and other clustering techniques

Overview

Clustering around latent components for three-way data
Adding a non negativity constraint on loadings
Application on sensory data for consumers’ segmentation
Imposing a global weighting scheme
Conclusion and perspectives
Clustering around latent components for Three-way data

$x_{ijk}$: score associated with the $i^{th}$ product rated by the $j^{th}$ consumer along with the $k^{th}$ parameters

$X_j$ — $j^{th}$ lateral slice of $X$ — contains the scores of $I$ samples on the $j^{th}$ variable according to $K$ parameters.
Clustering around latent components for Three-way data

Three-way partitioning of the 2\textsuperscript{nd} mode:
✓ Maximise the adequation between each lateral slice and the estimated one associated with its group

Minimize the Loss function:

\[ f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| x_j - \alpha_{jq} (t_q w_q^\top) \right\|_F^2 \]

1\textsuperscript{st} mode latent component: \( t_q \)
2\textsuperscript{nd} mode loadings of variables: \( \alpha_q \)
3\textsuperscript{rd} mode attributes’ weights: \( w_q \)
cluster dummy: \( \delta_{jq} \)
Clustering around latent components for Three-way data

Minimize the Loss function: 

\[ f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| X_j - \alpha_{jq} (t_q w_q^T) \right\|_F^2 \]

Maximise the squared covariance function:

\[ g = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \text{cov}^2 (X_j w_q, t_q) \]

with \( \|w_q\| = 1 \) and \( \|t_q\| = 1 \)


Initialisation

• Using a multi-start procedure obtained by randomly assigning the variables to $Q$ clusters

• Using a rational initial partitioning obtained by applying a Hierarchical Ascendant Algorithm based on criterion $f$ and Ward’s aggregation criterion

  • The variation of the criterion from step $l$ to step $l+1$ corresponds to the aggregation of two clusters, say $G_A$ and $G_B$.
  • It can be written as:

$$
\Delta f = - \sum_{X_j \in G_A} \|X_j - \alpha_{jG_A}(t_{G_A} w_{G_A}^\top)\|_F^2 - \sum_{X_j \in G_B} \|X_j - \alpha_{jG_B}(t_{G_B} w_{G_B}^\top)\|_F^2 \\
+ \sum_{X_j \in (G_A \cup G_B)} \|X_j - \alpha_{j(G_A \cup G_B)}[t_{(G_A \cup G_B)} w_{(G_A \cup G_B)}^\top]\|_F^2
$$
Algorithm

Starting from an initial partitioning into $Q$ clusters

Iterate the following two steps until convergence:

1. computing the cluster-specific parameters conditional on the cluster memberships
2. updating the cluster membership $\delta_{jq}$ of each variable conditional on the cluster-specific parameters (i.e., $t_q$ and $w_q$)

For the assignment, variable $j$ is assigned to the cluster $G_q$ for which $f_{jq}$ is minimal: $f_{jq} = \|X_j - \alpha_{jq}(t_q w_q^\top)\|_F^2$ with $\|t_q\|=1$, $\|w_q\|=1$

Determination of $\alpha_{jq}$ by means of a linear regression
Adding a non negativity constraint

Minimize the Loss function:  
\[ f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \| X_j - \alpha_{jq} (t_q w_q^T) \|_F^2 \]

Maximise the covariance function:  
\[ g = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \text{cov}^2(X_j w_q, t_q) \]

with  \( \| w_q \| = 1, \alpha_{jq} \geq 0 \) and  \( \| t_q \| = 1 \)

Determination of  \( \alpha_{jq} \) by means of a non negative linear regression
Consumers’ segmentation

Emotion ratings associated to coffee aromas

- 84 undergraduate students from Oniris
- 12 coffee aromas
  - Earth, Hay, Cedar; Vanilla, Coriander seeds, Flower Coffee, Apricot, Lemon, Honey, Basmati Rice, Hazelnut, Medicine
- 15 emotions rated with a 5-point Likert scale
  - Happy, Well, Amused
  - Angry, Unpleasant, Irritated, Disgusted, Upset
  - Energic, Calm, Nostalgic, Free, Excited, Unique, Surprised

consumer agreement on the product ratings
Consumers’ segmentation

Evolution of the aggregation criterion

% of explained inertia / number of clusters

Sensitivity to initialization

22.8%

3-5 Sept. 2019

Rencontres de la SFC
Consumers’ segmentation

Functions in Package ClustVarLV

```r
res <- CLV3W_kmeans(coffee,K=2,mode.scale=2, NN=TRUE,init=50,cp.rand=5)
summary(res)
```

```
$size
clusters
1  2
42 42

$prop_explained_per_cluster

$prop_explained_total
 [1] 22.812

$comp
  Vanilla B.Rice Lemon Coffee.Flower
    Comp1    Comp2   Comp1    Comp2
[1,] -0.306  0.058  0.524  0.422
[2,]  0.061 -0.218  0.435  0.278

$weight
  Calm Nostalgic Angry Disgusted Unique Excited Unpleasant Disappointed Free Surprised
Comp1  0.215  0.188 -0.224  0.477  0.094  0.138 -0.337  0.072 -0.232  0.229
Comp2  0.179  0.227 -0.222  0.371  0.225  0.207  0.327  0.528 -0.281  0.188

$groups
$groups$`cluster 1`
   loading cor.in.group cor.next.group
  1  0.218  0.824  0.504
  2  0.078  0.605  0.343
  3  0.152  0.600  0.135
  6  0.134  0.623  0.408
  8  0.086  0.384  0.157
  9  0.072  0.528  0.055
 11  0.000 -0.434 -0.115
 12  0.128  0.694  0.220
 14  0.154  0.692  0.345
 15  0.090  0.366  0.160
 18  0.236  0.808  0.505
 19  0.091  0.376 -0.043
 20  0.129  0.536  0.402
 22  0.144  0.639  0.382

$scormatrix
  Comp.1 Comp.2
Comp.1  1.000  0.586
Comp.2  0.586  1.000
```
Consumers’ segmentation

plot_var.clv3w(res.clv3w,K=2,labels=TRUE,cex.lab=0.8,beside=FALSE,mode3=TRUE)
Consumers’ segmentation

Cluster 1 weights
Cluster 2 weights

emotion agreement among the 2 clusters
Imposing a global weighting scheme

Minimise the Loss function:  

\[ f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \| x_j - \alpha_{jq} (t_q w^\top) \|^2_F \]

Maximise the squared covariance function:

\[ g = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \text{cov}^2(x_j w, t_q) \]

with  \( \|w\| = 1 \) and  \( \|t_q\| = 1 \)
Iterate the following two steps until convergence:

1. computing parameters conditional on the cluster memberships:
   - initializing \( w \), the two steps are alternated until convergence:
   - Each \( t_q \) is updated as the left singular vector corresponding to the largest singular value of the \((I \times \#G_q)\) matrix of the block components associated with the variables of \( G_q : \left[ \cdots |X_j w| \cdots \right]_{x_j \in G_q} \)
   - Global \( w \) is updated as the left singular vector associated with the largest singular value from of the \((K \times J)\) matrix \[ \left[ \cdots |X_j^\top t_q| \cdots \right]_{j=1,\ldots,J} \]

2. updating the cluster membership \( \delta_{jq} \) of each variable conditional on the cluster-specific \( t_q \) and the global weighting scheme \( w \)
Consumers’ segmentation

Sensitivity to initialisation

Inertia explained with NN constraint: 22.8%

Inertia explained with NN and Global Weights: 22.5%

Size of cluster 1: 44; Size of cluster 2: 38
2 consumers set aside
Consumers’ segmentation
Conclusion

Cluster analysis of three-way data with CLV3W providing:

✓ Latent component associated with the 1st mode samples
✓ Loadings reflecting the adequation of each slice in its cluster
✓ Weights deriving a block component for each slice

Available in Cran R

Easier interpretation with constraints on loadings and weights
Choice of the number of clusters:

- Adaptation of Hartigan criterion

\[ H_Q = (J - Q - 1) \left( \frac{f_Q}{f_{Q+1}} - 1 \right) \]

Index corresponding to the largest difference between two consecutive ones

Choice between full model or common weights constraint

Further comparisons needed with other strategies such as simultaneous decomposition and clustering models

Minimize the Loss function:

\[ f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| X_j - t_q w^T \right\|_F^2 \quad \text{with} \quad \left\| w_q \right\| = 1 \]

Thank you for your attention