SMILES

Faicel Chamroukhi

Journée NormaSTIC et NorMAth
Rouen, 12 octobre 2018
**SMILES (nov 2018, for 42 months)**

Statistical Modeling and Inference for unsupervised Learning at LargE-Scale

**Partenaires :**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="UNCAEN_LMNO" /></td>
<td>F. Chamroukhi (coord.), A. Sesboüé, J. Fadili</td>
</tr>
<tr>
<td><img src="image" alt="UNIVERSITÉ DE ROUEN_LMRS" /></td>
<td>G. Chagny, Antoine Channarond, N. Vergne, C. Bérard, A. Roche</td>
</tr>
<tr>
<td><img src="image" alt="LIS" /></td>
<td>H. Glotin, S. Paris, J. Razik, M. Richard</td>
</tr>
<tr>
<td><img src="image" alt="UNIVERSITÉ DE TOULON" /></td>
<td>C. Biernacki, V. Vandewalle</td>
</tr>
</tbody>
</table>
Research Axes

1. Task 1 : Models and inference for unsupervised large-scale data classification.
   - Sub-task 1.1 : Large-scale model-based clustering. (LMNO, INRIA) :
   - Sub-task 1.2 : Large-scale LDM and inference for functional data. (LMNO, INRIA) :
   - Sub-task 1.3 : Large-Scale LDM and inference for discrete data. (LMRS, LMNO, INRIA) :

2. Task 2 : Models and inference for large-scale data representation.
   - Sub-task 2.1 : High-dimensional (non-)parametric sparse regression for large-scale representation (LMRS, LMNO) :
   - Sub-task 2.2 : Unsupervised large-scale multimodal data representation (LIS, LMNO) :

3. Task 3 : Validation and applications.
   - i) Large-scale functional data analysis (LMNO-INRIA-LMRS) of heterogeneous multivariate times series and fMRI images
   - ii) Large-scale Bioacoustical data analysis (LIS-LMNO) for environmental survey
   - iii) Large-scale biological data analysis (LMRS) by inferring large-scale biological sequences from high-throughput sequencing
Research Axes

1. Task 1: Models and inference for unsupervised large-scale data classification.
   - Sub-task 1.1: Large-scale model-based clustering. (LMNO, INRIA):
   - Sub-task 1.2: Large-scale LDM and inference for functional data. (LMNO, INRIA):
   - Sub-task 1.3: Large-Scale LDM and inference for discrete data. (LMRS, LMNO, INRIA):

2. Task 2: Models and inference for large-scale data representation.
   - Sub-task 2.1: High-dimensional (non-)parametric sparse regression for large-scale representation (LMRS, LMNO):
   - Sub-task 2.2: Unsupervised large-scale multimodal data representation (LIS, LMNO):
Research Axes

1. Task 1: Models and inference for unsupervised large-scale data classification.
   ▶ Sub-task 1.1: Large-scale model-based clustering. (LMNO, INRIA):
   ▶ Sub-task 1.2: Large-scale LDM and inference for functional data. (LMNO, INRIA):
   ▶ Sub-task 1.3: Large-Scale LDM and inference for discrete data. (LMRS, LMNO, INRIA):

2. Task 2: Models and inference for large-scale data representation.
   ▶ Sub-task 2.1: High-dimensional (non-)parametric sparse regression for large-scale representation (LMRS, LMNO):
   ▶ Sub-task 2.2: Unsupervised large-scale multimodal data representation (LIS, LMNO):

3. Task 3: Validation and applications.
   ▶ i) Large-scale functional data analysis (LMNO-INRIA-LMRS) of heterogeneous multivariate times series and fMRI images
   ▶ ii) Large-scale Bioacoustical data analysis (LIS-LMNO) for environmental survey
   ▶ iii) Large-scale biological data analysis (LMRS) by inferring large-scale biological sequences from high-throughput sequencing
Objectifs

- Données complexes ➔ hétérogènes, temporelles dynamiques, fonctionnelles, incomplètes, de grande dimension, et disponibles en masse

- **Objectif** : Transformation de telles données en connaissances :
  ➔ Reconstruction/révélation de structures cachées, i.e, (hiérarchie de) groupes ; sélection de variables et prédiction, etc

- SMILES vise à élaborer un cadre scientifique et technique pour traiter et analyser des données massives hétérogènes et peu ou non-annotées

- Avec une visibilité à l’international

Axes du projet

1. Modélisation non supervisée par des modèles à variables latentes (MVL)

2. Inférence efficace non supervisée à grande échelle des MVL
Modélisation statistique par des MVL à l’échelle

Cadre scientifique général

_modèles_ statistiques à _variables latente_ :
\[ f(x|\theta) = \int_Z f(x, z|\theta) \, dz \]

_inférence_ à grande échelle par régularisation et échantillonnage :
\[ \hat{\theta} \in \arg\max_{\theta} \ell(\theta) - \text{Pen}_\lambda(\theta) \]

Modélisation statistique non supervisée à grande échelle par des MVL

- Apprentissage génératif, via des modèles à variables latentes (_régression_ et _clustering_).
- représenter explicitement la structure des données brutes et la révéler
  \[ \exists \text{ fondement théorique solide} \]
  \[ \Rightarrow \text{ Outils afférents d’estimation et de choix de modèle} \]
- n’ont pas été considérés avec succès pour une analyse à grande échelle
Inférence non supervisée à grande échelle des MVL

Inférence en grande dimension

- L’inférence se ramène en général à l’optimisation de problèmes non linéaires complexes. À grande échelle :
  - suggère de nouvelles stratégies de régularisation pour pallier la grande dimension
  - Méthodes parcimonieuses pour une meilleure représentation

Données de gros volume

- la distribution des calculs est une façon naturelle de s’y prendre
- méthodo : échantillonnage et inférence des modèles agrégés à partir d’un gros volume de données
  - Nouvelles stratégies d’agrégation d’estimateurs et de sélection de modèle
Données longitudinales de plus en plus fréquentes

Railway switch curves

Yeast cell cycle curves

Phonemes curves

Satellite waveforms
Clustering/segmentation de données temporelles
Clustering/segmentation de données temporelles

![Graph showing power (Watt) over time (Second)](image-url)
Clustering de données représentées par des graphes
Décomposition parcimonieuse non-supervisée
Décomposition parcimonieuse non-supervisée
Model-Based Co-Clustering of Multivariate Functional Data
Joint work with Christophe Biernacki, INRIA-Lille
Outline

1 Model-Based Co-Clustering of Multivariate Functional Data
   - Motivation
   - Model-based co-clustering
   - Temporal curve segmentation (RHLP)
   - Model-based co-clustering embedding RHLP
   - Conclusion and perspectives

2 Regularized Mixture-of-Experts for high-dimensional data
Functional data are increasingly frequent

[James and Hastie, 2001; James and Sugar, 2003]
[Ramsay and Silverman, 2005]
[Chamroukhi et al., 2010]
[Bouveyron and Jacques, 2011]
[Samé et al., 2011]
[Jacques and Preda, 2014]
[Bouveyron et al., 2018]
[Chamroukhi and Nguyen, 2018]
Clustering of functional data

→ a growing investigation of Model-Based Clustering (MBC) for functional data

Some Reviews on MBC for functional data: [Jacques and Preda, 2014; Chamroukhi and Nguyen, 2018]
Clustering of functional data

→ a growing investigation of Model-Based Clustering (MBC) for functional data

Some Reviews on MBC for functional data: [Jacques and Preda, 2014; Chamroukhi and Nguyen, 2018]

Tecator data set\(^1\): \(n = 240\) spectra with \(m = 100\)

Figure – Original data and clustering results from Chamroukhi [2016b] for the data considered in the same setting as in Hébrail et al. [2010] (six clusters, each cluster is approximated by five linear segments \((R = 5, p = 1)\))
Clustering of functional data

Topex/Poseidon satellite data\(^2\) : \(n = 472\) waveforms of \(m = 70\) measured echoes

Figure – Original data and clustering results from Chamroukhi [2016b] with the same setting as in Hébrail et al. [2010] : twenty clusters and a piecewise linear approximation of four segments.

---

Clustering of functional data

Phonemes data set: $n = 1000$ log-periodograms for $m = 150$ frequencies

Figure – Original data and clustering results from Chamroukhi [2016b]

Clustering of functional data

Clustering real curves of high-speed railway-switch operations
Data: $n = 115$ curves of $m \approx 510$ observations
$K = 2$ clusters: operating state without/with possible defect
Clustering switch operations

Clustering real curves of high-speed railway-switch operations
Data: $n = 115$ curves of $m \approx 510$ observations
$K = 2$ clusters: operating state without/with possible defect
Outline

1 Model-Based Co-Clustering of Multivariate Functional Data
   - Motivation
   - Model-based co-clustering
   - Temporal curve segmentation (RHLP)
   - Model-based co-clustering embedding RHLP
   - Conclusion and perspectives

2 Regularized Mixture-of-Experts for high-dimensional data
This talk: **Multivariate functional data clustering**

- Multivariate functional data are increasingly present
- e.g. : Data continuously recorded for different subjects from multiple subject’s sensors

$\leftrightarrow$ Measurements collected from different network elements (transceivers, cells, sites...):

**Figure** – An example with $d = 30$ and $n = 20$ daily observations [Ben Slimen et al., 2016].
Questioning

Clustering of highly multivariate functional data with two guidelines:

- (1) Mathematical guideline: warranty for estimation and selection
- (2) User guideline: keep a user-friendly meaning of the process

Both are important because clustering is a highly risky task...

Proposed answering

(1) Model-based co-clustering with (2) temporal curve segmentation

Novelty corresponds to combining both (1) and (2)
Difference between clustering and co-clustering

- Simultaneous clustering of lines/indiv. ($Z$) and columns/var. ($W$)
- Can be used as a way to reduce dimensionality (var. $\rightarrow W$)

**Figure** – Binary data set with $n = 500$, $d = 300$, $K = M = 3$
The Latent Block Model [Govaert and Nadif, 2013]

\[
f(X; \Psi) = \sum_{(z,w) \in Z \times W} \mathbb{P}(Z, W; \pi, \rho) f(X|Z, W; \theta)
\]

Hypotheses

- The latent variables \( Z \) and \( W \) are independent: \( \mathbb{P}(Z, W) = \mathbb{P}(Z)\mathbb{P}(W) \) and iid:
  \[
  \mathbb{P}(Z) = \prod_i \mathbb{P}(z_i) \text{ with } z_i \sim \text{Multinomial}(\pi_1, \ldots, \pi_K) \text{ where } \pi_k = \mathbb{P}(z_k = k)
  \]
  \[
  \mathbb{P}(W) = \prod_j \mathbb{P}(w_j) \text{ with } w_j \sim \text{Multinomial}(\rho_1, \ldots, \rho_M) \text{ where } \rho_\ell = \mathbb{P}(w_j = \ell)
  \]
- Conditional independence: \( x_{ij}|(z_i, w_j) \perp x_{i'j'}|(z_i', w_j') \)
**Latent block model for co-clustering**

The Latent Block Model [Govaert and Nadif, 2013]

\[
f(X; \Psi) = \sum_{(z,w) \in Z \times W} P(Z, W; \pi, \rho) f(X|Z, W; \theta)
\]

data kind dependent

**Hypotheses**

- The latent variables \(Z\) and \(W\) are independent: \(P(Z, W) = P(Z)P(W)\) and iid:
  \[
P(Z) = \prod_i P(z_i) \text{ with } z_i \sim \text{Multinomial}(\pi_1, \ldots, \pi_K) \text{ where } \pi_k = P(z_k = k)
  
P(W) = \prod_j P(w_j) \text{ with } w_j \sim \text{Multinomial}(\rho_1, \ldots, \rho_M) \text{ where } \rho_\ell = P(w_j = \ell)
  
- Conditional independence: \(x_{ij}|(z_i, w_j) \perp x_{i'j'}|(z_{i'}, w_{j'})\)

← binary data: binary [Govaert and Nadif, 2003, 2008; Keribin et al., 2012],
← categorical data: multinomial [Keribin et al., 2014]
← continuous data: Gaussian [Lomet, 2012; Govaert and Nadif, 2013]
← functional data: functional PCA + Gaussian, see further [Ben Slimen et al., 2016]
Inference of the latent block model

- variational block EM (VBEM) for maximum likelihood estimation and fuzzy co-clustering [Govaert and Nadif, 2006, 2008].


- Number of blocks estimation : ICL criterion [Lomet, 2012; Keribin et al., 2014]
Package blockcluster on the cloud

massiccc.lille.inria.fr

Massive Clustering with Cloud Computing
Clustering of heterogeneous data with missing values.
Hosted in the cloud. No installation or configuration required.
Upload your data, and get results straight away.

How MASSICCC Platform works
MASSICCC Platform lets you upload your data and select from a range of analysis tools to extract meaningful information. No need to install any software or configure some tools. You’ll see results straight away.

- Upload your data securely
  Upload all or part of your data to MASSICCC Platform and run multiple algorithms on them. You will have full control on the data you upload and only you will be able to access them.

- Focus on the data
  Minimum configuration is required to use our algorithms since we’ll select the most sensible options by default based on the data you provide. No scientific background is required to start working and get results. Advanced configuration options are available if you need specific functions.
Functional data notation

- Data: (discretized) values of underlying smooth functions, not just vectors
- Data: A sample of \( n \) heterogeneous univariate curves \((x_1, y_1), \ldots, (x_n, y_n)\)
- \((x_i, y_i)\) consists of \( m_i \) observations \( y_i = (y_{i1}, \ldots, y_{imi})\) observed at the independent covariates, (e.g., time \( t \) in time series), \((x_{i1}, \ldots, x_{imi})\)
Functional data modeling: “classical” approach

[Ramsay and Silverman, 2005] and many others

- **Step 1**: \((x, y)\) decomposed into a finite basis of function (B-spline...) : 
  \[ Y_i(t) \approx \sum_{r=1}^{d} c_{ir} \phi_r(x_i(t)) \] with \(c\) estimated by OLS

- **Step 2**: functional principal components analysis (PCA) which is performed as a usual PCA of the basis expansion coefficients \(c\) using a metric defined by the inner products between the basis functions

- **Step 3**: set a probability distribution on \(c\), typically Gaussian

It defines a distribution on \(c\) instead of \(y\)...
Alternatively, use a segmentation via generative piecewise polynomial regression modeling of $f(y|x)$ [Chamroukhi et al.]

Regression with Hidden Logistic Process (RHLP)
See formula later

It gives a distribution on $y$ and also a meaningful segmentation of the curve
RHL for modeling different types of functions

Faicel Chamroukhi  
Projet ANR SMILES
Package mixtcomp on the cloud

massiccc.lille.inria.fr

Massive Clustering with Cloud Computing
Clustering of heterogeneous data with missing values.
Hosted in the cloud. No installation or configuration required.
Upload your data, and get results straight away.

Developed by Inria

How MASSICCC Platform works
MASSICCC Platform lets you upload your data and select from a range of analysis tools to extract meaningful information. No need to install any software or configure some tools. You’ll see results straight away.

- Upload your data securely
  Upload all or part of your data to MASSICCC Platform and run multiple algorithms on them. You will have full control of the data you upload and only you will be able to access them.
- Focus on the data
  Minimum configuration is required to use our algorithms since we’ll select the most sensible options by default based on the data you provide. No scientific background is required to start working and get results. Advanced configuration options are available if you need specific functions.
Multivariate functional data co-clustering

[Chamroukhi and Biernacki, 2017]

- Data: \( \mathbf{Y} = (\mathbf{y}_{ij}) \) a data sample matrix of \( n \) individuals defined on a set \( \mathcal{I} \) and \( d \) continuous functional variables defined on a set \( \mathcal{J} \).
- Each variable \( \mathbf{y}_{ij} \) is an univariate curve \( \mathbf{y}_{ij} = (y_{ij}(t_1), \ldots, y_{ij}(t_{T_{ij}})) \) of \( T_{ij} \) observations \( y(t) \in \mathbb{R} \) linked to covariates \( \mathbf{x}_{ij} = (x_{ij}(t_1), \ldots, x_{ij}(t_{T_{ij}})) \) at the points \( (t_1, \ldots, t_{T_{ij}}) \), typically a sampling time.
Embedding RHLP in co-clustering

[Chamroukhi and Biernacki, 2017]

Functional Latent Block Model for Co-clustering:

\[
\begin{align*}
    f(Y|X; \Psi) &= \sum_{(z,w) \in Z \times W} \mathbb{P}(Z; \pi)\mathbb{P}(W; \rho)f(Y|X, Z, W; \theta) \\
    &= \sum_{(z,w) \in Z \times W} \prod_{i,k} \pi^{z_{ik}} \prod_{j,\ell} \rho^{w_{j\ell}} \prod_{i,j,k,\ell} f(y_{ij}|x_{ij}; \theta^{z_{ik}w_{j\ell}}).
\end{align*}
\]

with parameter vector \( \Psi = (\pi^T, \rho^T, \theta^T)^T \), where \( \pi = (\pi_1, \ldots, \pi_K)^T \), \( \rho = (\rho_1, \ldots, \rho_M)^T \), and \( \theta = (\theta_{11}^T, \ldots, \theta_{k\ell}^T, \ldots, \theta_{KM}^T)^T \).
Embedding RHLP in co-clustering

- **RHLP** [Chamroukhi et al., 2009]: model the conditional data distribution for each block $kl$, assuming that each functional variable $y_{ij}$ is governed by an $S_{k\ell}$-state hidden process of $y_{ij}$:

$$f(y_{ij}|x_{ij}; \theta_{k\ell}) = \prod_{t=1}^{T_{ij}} \sum_{r=1}^{S_{k\ell}} \alpha_{k\ell r}(t; \xi_{k\ell}) N(y_{ij}(t); \beta_{k\ell r}^T x_{ij}(t), \sigma_{k\ell r}^2)$$

where the dynamical weights $\alpha$'s are given by the multinomial logistic:

$$\alpha_{k\ell r}(t; \xi_{k\ell}) = \frac{\exp (\xi_{k\ell r0} + \xi_{k\ell r1} t)}{1 + \sum_{r'=1}^{S_{k\ell}-1} \exp (\xi_{k\ell r'0} + \xi_{k\ell r'1} t)}.$$

← Can be seen as a generative piecewise polynomial regression model where the transition points are smoothly controlled by logistic weights

← a particular mixture-of-experts model [Jacobs et al., 1991; Jordan and Jacobs, 1994]/(parametric) mixture of regressions with predictor-dependent mixing proportions [Young and Hunter, 2010]
Block mean curve approximation and segmentation

- Approximation: a prototype mean curve

\[
y_t \mid (z_i, w_j) \approx \hat{y}_t = \mathbb{E}[Y(t) \mid z_i, w_j, x(t); \hat{\Psi}] = \sum_{s=1}^{S_{kl}} \alpha_{k\ell r}(t; \hat{\xi}_{k\ell}) \beta_{k\ell r}^T x_i(t)
\]

\[\hookrightarrow\] A smooth and flexible approximation thanks to the logistic weights

- Curve segmentation:

\[
\hat{h}_t \mid (z_i, w_j) = \arg \max_{1 \leq s \leq S_{kl}} \mathbb{E}[H_t \mid z_i, w_j, x_{ij}(t); \hat{\xi}] = \arg \max_{1 \leq k \leq K} \alpha_{k\ell r}(t; \hat{\xi}_{k\ell})
\]
Parameter estimation: EM not feasible

EM algorithm:

\[ \Phi^{(q+1)} \in \arg \max_{\Psi} \mathbb{E} \left[ \log L_c(\Phi) | D, \Phi^{(q)} \right] \]

- The complete-data log-likelihood:

\[
\log L_c(\Phi) = \log f(Y, Z, W, H | X; \Phi) \\
= \sum_{i,k} z_{ik} \log \pi_k + \sum_{j,\ell} w_{j\ell} \log \rho_{\ell} \\
+ \sum_{i,j,k,\ell,t,r} z_{ik} w_{j\ell} h_{tr} \log \left[ \alpha_{k\ell r}(t; \xi_{k\ell}) \mathcal{N}(y_{ij}(t); \beta_{k\ell r}^T x_{ij}(t), \sigma^2_{k\ell r}) \right]
\]

where \((h_{tr}; t = 1, \ldots, T_{ij}, r = 1, \ldots, S_{k\ell})\) is a binary variable indicating from which state the observation \(y_{ij}(t)\) within the block cluster \(k\ell\) is originated.
Parameter estimation: EM not feasible

- The E-Step computes the expected complete-data log-likelihood, given the observed curves \((X, Y)\), and the current parameter estimation \(\Psi^{(q)}\)

\[
Q(\Psi, \Psi^{(q)}) = \mathbb{E} \left[ \log L_c(\Psi) \mid X, Y; \Psi^{(q)} \right]
\]

\[
= \sum_{i,k} \mathbb{P}(z_{ik} = 1 \mid y_{ij}, x_{ij}) \log \pi_k + \sum_{j,\ell} \mathbb{P}(w_{j\ell} = 1 \mid y_{ij}, x_{ij}) \log \rho_{\ell}
\]

\[
+ \sum_{i,j,k,\ell,t,r} \mathbb{P}(z_{ik}w_{j\ell} = 1 \mid y_{ij}, x_{ij}) \mathbb{P}(h_{tr} = 1 \mid z_{ik}, w_{j\ell}, y_{ij}(t), x_{ij}(t)) \times
\]

\[
\log \left[ \alpha_{k\ell r}(t; \xi_{k\ell}) N \left( y_{ij}(t); \beta_{k\ell r}^T x_{ij}(t), \sigma_{k\ell r}^2 \right) \right]
\]

\[\hookrightarrow\] Requires the calculation of the posterior joint distribution \(\mathbb{P}(z_{ik}w_{j\ell} = 1 \mid y_{ij}, x_{ij})\)

\[\hookrightarrow\] does not factorize due to the conditional dependence on the observed curves of the row and the column labels


\[\hookrightarrow\] We adopt this variational approximation in our context
Variational block EM algorithm

\[ P(z_{ik}w_{j\ell} = 1 | y_{ij}, x_{ij}) \approx P(z_{ik} = 1 | y_{ij}, x_{ij}) \times P(w_{j\ell} = 1 | y_{ij}, x_{ij}) \]
**Variational block EM algorithm**

\[
\mathbb{P}(z_{ik} w_{j\ell} = 1 | y_{ij}, x_{ij}) \approx \mathbb{P}(z_{ik} = 1 | y_{ij}, x_{ij}) \times \mathbb{P}(w_{j\ell} = 1 | y_{ij}, x_{ij})
\]

**Initialization** : start from an initial solution at iteration \( q = 0 \), and then alternate at the \((q + 1)\)th iteration between the following variational E- and M- steps until convergence :

**VE Step** Estimate the variational approximated posterior memberships :

1. \( \tilde{z}^{(q+1)}_{ik} \propto \pi^{(q)}_{k} \exp \left( \sum_{j,l,t,r} \tilde{w}^{(q)}_{j\ell} \tilde{h}^{(q)}_{tr} \log \left[ \alpha_{k\ell r}(t; \xi^{(q)}_{k\ell}) \mathcal{N} \left( y_{ij}(t); \beta^{T(q)}_{k\ell r} x_{ij}(t), \sigma^{(q)2}_{k\ell r} \right) \right] \right) \)

2. \( \tilde{w}^{(q+1)}_{j\ell} \propto \rho^{(q)}_{\ell} \exp \left( \sum_{i,k,t,r} \tilde{z}^{(q)}_{ik} \tilde{h}^{(q)}_{tr} \log \left[ \alpha_{k\ell r}(t; \xi^{(q)}_{k\ell}) \mathcal{N} \left( y_{ij}(t); \beta^{T(q)}_{k\ell r} x_{ij}(t), \sigma^{(q)2}_{k\ell r} \right) \right] \right) \)

3. \( \tilde{h}^{(q+1)}_{tr} \propto \alpha^{(q)}_{k\ell r}(t; \xi^{(q)}_{k\ell}) \mathcal{N} \left( y_{ij}(t); \beta^{(q)T}_{k\ell r} x_{ij}(t), \sigma^{(q)2}_{k\ell r} \right) \)

where :

- \( \tilde{z}_{ik} = \mathbb{P}(z_{ik} = 1 | y_{ij}, x_{ij}) \),
- \( \tilde{w}_{j\ell} = \mathbb{P}(w_{j\ell} = 1 | y_{ij}, x_{ij}) \),
- \( \tilde{h}_{tr} = \mathbb{P}(h_{tr} = 1 | z_i, w_j, y_{ij}(t), x_{ij}(t)) \)
Variational block EM algorithm

**M Step** update the parameters estimates $\theta^{(q+1)}$ given the estimated posterior memberships at the current iteration $q + 1$:

1. $\pi_k^{(q+1)} = \frac{\sum_i \tilde{z}_{ik}^{(q+1)}}{n}$

2. $\rho_{\ell}^{(q+1)} = \frac{\sum_j \tilde{w}_{j\ell}^{(q+1)}}{d}$
Variational block EM algorithm

M Step update the parameters estimates $\theta^{(q+1)}$ given the estimated posterior memberships at the current iteration $q + 1$:

1. $\pi_k^{(q+1)} = \sum_i \frac{\tilde{z}_{ik}^{(q+1)}}{n}$

2. $\rho_{\ell}^{(q+1)} = \sum_j \frac{\tilde{w}_{j\ell}^{(q+1)}}{d}$

The update of each block parameters $\theta_{k\ell}$ consists in a weighted version of the RHLP updating rules:

3. $\xi_{k\ell}^{(new)} = \xi_{k\ell}^{(old)} - \left[ \frac{\partial^2 F(\xi_{k\ell})}{\partial \xi_{k\ell} \partial (\xi_{k\ell})^T} \right]^{-1} \frac{\partial F(\xi_{k\ell})}{\partial \xi_{k\ell}} \bigg|_{\xi_{k\ell} = \xi_{k\ell}^{(old)}}$ which is the IRLS maximisation of $F(\xi_{k\ell}) = \sum_{i,j,t} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} \tilde{h}_{tr}^{(q)} \log \alpha_{k\ell r}(t; \xi_{k\ell})$ w.r.t $\xi_{k\ell}$.
Variational block EM algorithm

**M Step** update the parameters estimates $\theta^{(q+1)}$ given the estimated posterior memberships at the current iteration $q + 1$:

1. $\pi_k^{(q+1)} = \frac{\sum_i \tilde{z}_{ik}^{(q+1)}}{n}$

2. $\rho_\ell^{(q+1)} = \frac{\sum_j \tilde{w}_{j\ell}^{(q+1)}}{d}$

The update of each block parameters $\theta_{k\ell}$ consists in a weighted version of the RHLP updating rules:

3. $\xi_{k\ell}^{(\text{new})} = \xi_{k\ell}^{(\text{old})} - \left[ \frac{\partial^2 F(\xi_{k\ell})}{\partial \xi_{k\ell} \partial \xi_{k\ell}^T} \right]^{-1} \left[ \frac{\partial F(\xi_{k\ell})}{\partial \xi_{k\ell}} \right] |_{\xi_{k\ell} = \xi_{k\ell}^{(\text{old})}}$ which is the IRLS maximisation of $F(\xi_{k\ell}) = \sum_{i,j,t} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} \tilde{h}_{tr}^{(q)} \log \alpha_{k\ell r}(t; \xi_{k\ell})$ w.r.t $\xi_{k\ell}$.

The regression parameters updates consist in analytic WLS problems:

4. $\beta_{k\ell r}^{(q+1)} = \left[ \sum_{i,j} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} X_{ij}^T \Lambda_{ijkr}^{(q)} X_{ij} \right]^{-1} \sum_{i,j} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} X_{ij}^T \Lambda_{ijkr}^{(q)} y_{ij}$

5. $\sigma_{k\ell r}^{2(q+1)} = \frac{\sum_{i,j} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} \| \sqrt{\Lambda_{ijkr}^{(q)}} (y_{ij} - X_{ij} \beta_{k\ell r}^{(q+1)}) \|^2}{\sum_{i,j} \tilde{z}_{ik}^{(q)} \tilde{w}_{j\ell}^{(q)} \text{trace}(\Lambda_{ijkr}^{(q)})} \text{ where } X_{ij} \text{ is the design matrix for the } i\text{th curve, } \Lambda_{ijkr}^{(q)} \text{ is the diagonal matrix whose diagonal elements are the posterior segment memberships } \{\tilde{h}_{i jtr}^{(q)}; t = 1, \ldots, T_{ij}\}.
It is also possible to use the Classification EM (CEM) approximation of EM [Celeux and Govaert, 1992].

Parameter estimation by an SEM algorithm : SEM-FLBM

- The SEM algorithm [Celeux and Diebolt, 1985] allows to overcome some drawbacks of the variational-EM algorithm, including its sensitivity to starting values; SEM does not use an approximation.
- Eg. SEM for latent block models for categorical data [Keribin et al., 2012, 2014]
- The formulas of VEM-FLBM and SEM-FLBM are essentially the same, except that we incorporate a stochastic step consisting of sampling binary indicator variables $z_{ik}$, $w_{j\ell}$ and $h_{tr}$ according to $\tilde{z}_{ik}$, $\tilde{w}_{j\ell}$ and $\tilde{h}_{tr}$.
Conclusion and perspectives

Conclusion

- A full generative framework for the cluster analysis and segmentation of high-dimensional non-stationary functional data
- The model inference can be performed by a variational EM algorithm or SEM

Perspectives

- Numerical experiments
- Package
Y. Ben Slimen, S. Allio, and J. Jacques. Model-Based Co-clustering for Functional Data. HAL preprint hal-01422756, December 2016. URL https://hal.inria.fr/hal-01422756.


References II


Thank you for your attention!