Learning probabilistic latent process models from temporal data

Faicel Chamroukhi

Information Sciences and Systems Laboratory (LSIS), UMR CNRS 7296
University of the South Toulon-Var (USTV), France

email: `chamroukhi@univ-tln.fr`
web: `chamroukhi.univ-tln.fr`

ERMITES 2012, La Seyne/Mer

01 september 2012

Outline

Contexts

1. Supervised learning : data (x,y)
2. Unsupervised learning : data x, y ?
3. semi-supervised, partially supervised, ...

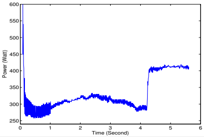
Generative/Discriminative

1. Discriminative approach : Directly learn $p(y|x)$
2. Generative approach : Learn $p(x,y) \Rightarrow p(y|x) \propto p(x|y)p(y)$
   - "understand" the process generating the data
   - easily adaptable to the unsupervised context

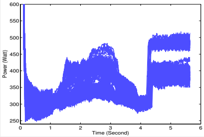
$\Rightarrow$ Latent data models

Context

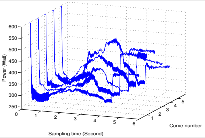
- Temporal data (signals, sequences, functions, time series, ..)



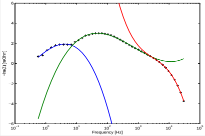
A curve



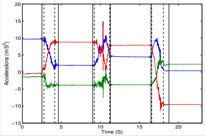
A set of curves



Sequence of curves



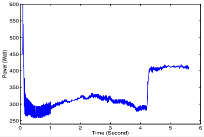
impedance signal of a fuel cell



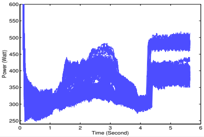
Acceleration measures

Context

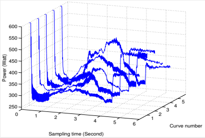
- Temporal data (signals, sequences, functions, time series, ..)



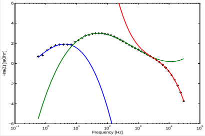
A curve



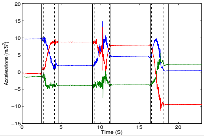
A set of curves



Sequence of curves



impedance signal of a fuel cell



Acceleration measures

1. Several regimes over time (temporal aspect) ⇒ Abrupt and/or regime changes
2. Many curves to analyse
3. Multidimensional temporal data

Objectives

- Take into account the class dispersion and the regime changes
- Explicitly integrate this complexity (class dispersion, regimes,..)

  ⇒ Learning probabilistic generative models

Objectives

- Take into account the class dispersion and the regime changes
- Explicitly integrate this complexity (class dispersion, regimes,..)

  ⇒ Learning probabilistic generative models

- The generative approach is interested in the process generating the data
- Better suited to unsupervised context that the discriminative approach

Curves modeling (functions)

- Approaches based on hidden process regression
- Probabilistic formalization of regime changes

Curves Classification (supervised and unsupervised)

- Mixture model-based approach for curve classification
  - model-based functional cluster analysis (in the unsupervised case)
  - model-based functional discriminant analysis (in the supervised case)

Regression-based models

Data : 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

Regression-based models

Data : 1 curves $\mathbf{y}=(y_1,\ldots,y_m)$ regularly observed at instants $(t_1,\ldots,t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

- simple polynomial regression : f est une fonction polynôme

Regression-based models

Data : 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

- simple polynomial regression : f est une fonction polynôme
- polynomial spline regression (Deboor, 1978)

Regression-based models

Data : 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

- simple polynomial regression : f est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- f is a piecewise polynomial function (McGee et Carleton, 1970)

  $\Rightarrow$ exact Optimization with dynamic prog. (Bellman, 1961)

Regression-based models

Data : 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

- simple polynomial regression : f est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- f is a piecewise polynomial function (McGee et Carleton, 1970)

  $\Rightarrow$ exact Optimization with dynamic prog. (Bellman, 1961)

  - dynamic programming can be time consuming
  - Adapted to abrupt regime changes but a problem of continuity

Regression-based models

Data : 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)$$

- simple polynomial regression : f est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- f is a piecewise polynomial function (McGee et Carleton, 1970)

  $\Rightarrow$ exact Optimization with dynamic prog. (Bellman, 1961)

  - dynamic programming can be time consuming
  - Adapted to abrupt regime changes but a problem of continuity
- Polynomial regression governed by a Markov chain (Fridman, 1993)

  The model for a curve : $y_j = \boldsymbol{\beta}_{z_j}^T \mathbf{t}_j + \sigma_{z_j}\epsilon_j \quad (j = 1, \ldots, m)$

  - Continuity not guaranteed
  - Not adapted to approximate a set of curves

The Regression model with a Hidden Logistic Process (RHLP)

The proposed Regression model with a Hidden Logistic Process (RHLP) : Neural Networks - Elsevier, 22(5-6) :593-602, 2009.

Temporal data $\mathbf{y} = (y_1, \ldots, y_m)$ observed at instants $\mathbf{t} = (t_1, \ldots, t_m)$

Model definition

$$y_j = \boldsymbol{\beta}_{z_j}^T \mathbf{t}_j + \sigma_{z_j} \epsilon_j \quad ; \quad \epsilon_j \sim \mathcal{N}(0,1), \quad (j = 1, \ldots, m)$$

- z_j hidden label of the regime of y_j
- $\mathbf{z} = (z_1, \ldots, z_n)$ is hidden logistic process

The proposed Regression model with a Hidden Logistic Process (RHLP) : Neural Networks - Elsevier, 22(5-6) :593-602, 2009.

Temporal data $\mathbf{y} = (y_1, \ldots, y_m)$ observed at instants $\mathbf{t} = (t_1, \ldots, t_m)$

Model definition

$$y_j = \boldsymbol{\beta}_{z_j}^T \mathbf{t}_j + \sigma_{z_j} \epsilon_j \quad ; \quad \epsilon_j \sim \mathcal{N}(0,1), \quad (j = 1, \ldots, m)$$

- z_j hidden label of the regime of y_j
- $\mathbf{z} = (z_1, \ldots, z_n)$ is hidden logistic process

$z_j | t_j \sim \mathcal{M}(1, \pi_1(t_j; \mathbf{w}), \ldots, \pi_R(t_j; \mathbf{w}))$; où

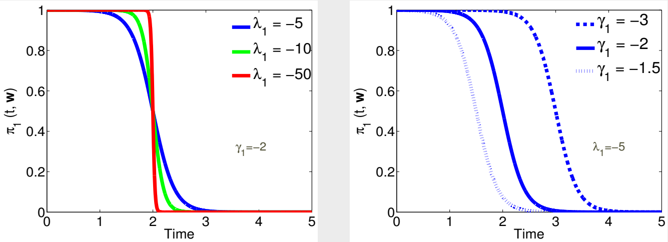
$$\pi_r(t_j; \mathbf{w}) = p(z_j = r | t_j; \mathbf{w}) = \frac{\exp(\lambda_r(t_j + \gamma_r))}{\sum_{\ell=1}^R \exp(\lambda_\ell(t_j + \gamma_\ell))}$$

- $\boldsymbol{w}_r = (\lambda_r, \gamma_r)^T$ parameter of the rth logistic function
- $\mathbf{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_R)$ parameter of the R logistic functions

Flexibility of the logistic transformation

Temporal variation of the logistic function w.r.t $\boldsymbol{w}$

- $\pi_r(t_i;\mathbf{w}) = \frac{\exp(\lambda_r(t_i+\gamma_r))}{\sum_{\ell=1}^{R}\exp(\lambda_\ell(t_i+\gamma_\ell))}$
- Example with $K = 2$ regimes



⇒ The parameter λ_r controls the quality of transitions between regimes

⇒ The parameter γ_r is related to the transition time

Illustration of the principle of the method

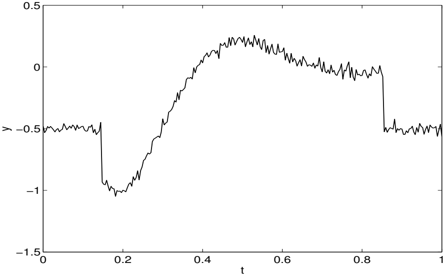
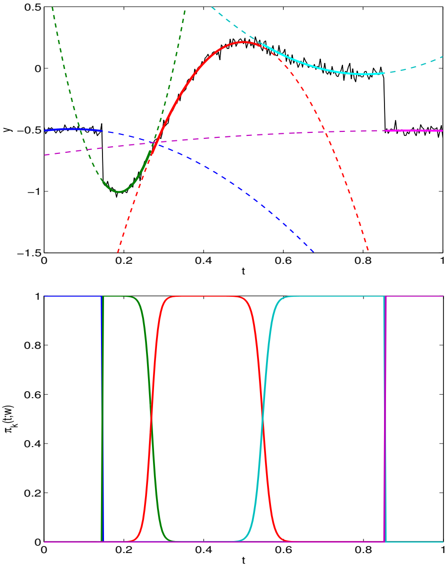


Illustration of the principle of the method



Parameter estimation with ML via EM

- parameter vector of the model : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-likelihood : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$

Parameter estimation with ML via EM

- parameter vector of the model : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-likelihood : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Completed Log-likelihood :
  $\mathscr{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log \left[\pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)\right]$

Parameter estimation with ML via EM

- parameter vector of the model : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-likelihood : $\mathcal{L}(\boldsymbol{\theta}) = \sum_{j=1}^m \log \sum_{r=1}^R \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Completed Log-likelihood :
  $\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^m \sum_{r=1}^R z_{jr} \log [\pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)]$

EM Algorithm for the proposed model

Initial Parameter $\boldsymbol{\theta}^{(0)}$:

1. E-Step : Conditional expectation of the complete log-likelihood $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) =$

$$\mathbb{E}\left[\mathcal{L}_c(\boldsymbol{\theta}) | \mathbf{y}, \mathbf{t}; \boldsymbol{\theta}^{(q)}\right] = \underbrace{\sum_{r=1}^R \sum_{j=1}^m \tau_{jr}^{(q)} \log \pi_r(t_j; \mathbf{w})}_{Q_{\mathbf{w}}} + \sum_{r=1}^R \underbrace{\sum_{j=1}^m \tau_{jr}^{(q)} \log \mathcal{N}\left(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2\right)}_{Q_{\boldsymbol{\theta}_r}}$$

$\Rightarrow$ calculating the posterior probabilities $\tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \boldsymbol{\theta}^{(q)})$

Parameter estimation with ML via EM

- parameter vector of the model : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-likelihood : $\mathcal{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Completed Log-likelihood :
$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log\left[\pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)\right]$

EM Algorithm for the proposed model

Initial Parameter $\boldsymbol{\theta}^{(0)}$:

1. E-Step : Conditional expectation of the complete log-likelihood $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) =$

$$\mathbb{E}\left[\mathcal{L}_c(\boldsymbol{\theta}) | \mathbf{y}, \mathbf{t}; \boldsymbol{\theta}^{(q)}\right] = \underbrace{\sum_{r=1}^{R} \sum_{j=1}^{m} \tau_{jr}^{(q)} \log \pi_r(t_j; \mathbf{w})}_{Q_{\mathbf{w}}} + \sum_{r=1}^{R} \underbrace{\sum_{j=1}^{m} \tau_{jr}^{(q)} \log \mathcal{N}\left(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2\right)}_{Q_{\boldsymbol{\theta}_r}}$$

$\Rightarrow$ calculating the posterior probabilities $\tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \boldsymbol{\theta}^{(q)})$

2. M-Step : $\boldsymbol{\theta}^{(q+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) = Q_{\mathbf{w}}(\mathbf{w}, \boldsymbol{\theta}^{(q)}) + \sum_{r=1}^{R} Q_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)})$
$\Rightarrow$ Separate Maximizations of $Q_{\boldsymbol{\theta}_r}$ and $Q_{\mathbf{w}}$

- Maximisation of $Q_{\boldsymbol{\theta}_r}$: exact solutions of weighted regressions (weights τ_{jr})

$$\boldsymbol{\beta}_r^{(q+1)} = \Big[\sum_{j=1}^{m} \tau_{jr}^{(q)} \mathbf{t}_j \mathbf{t}_j^T\Big]^{-1} \sum_{j=1}^{m} \tau_{jr}^{(q)} y_j \mathbf{t}_j$$

$$\sigma_r^{2(q+1)} = \frac{1}{\sum_{j=1}^{m} \tau_{jr}^{(q)}} \sum_{j=1}^{m} \tau_{jr}^{(q)} (x_{ij} - \boldsymbol{\beta}_{gkr}^{T(q+1)} \mathbf{t}_j)^2.$$

- Maximisation of $Q_{\mathbf{w}}$: a convex problem of multi-class logistic regression weighted by $\tau_{jr}^{(q)}$ $\Rightarrow$ iterative method : IRLS

$$\mathbf{w}^{(q,l+1)} = \mathbf{w}^{(l)} - \Big[\frac{\partial^2 Q_{\mathbf{w}}(\mathbf{w}, \boldsymbol{\theta}^{(q)})}{\partial \mathbf{w} \partial \mathbf{w}^T}\Big]^{-1}_{\mathbf{w}=\mathbf{w}^{(l)}} \frac{\partial Q_{\mathbf{w}}(\mathbf{w}, \boldsymbol{\theta}^{(q)})}{\partial \mathbf{w}}\Big|_{\mathbf{w}=\mathbf{w}^{(l)}}$$

Curve approximation and segmentation

Curve approximation

$$\mathbb{E}[y_j|t_j;\widehat{\boldsymbol{\theta}}] = \sum_{r=1}^{R} \pi_r(t_j;\widehat{\mathbf{w}})\widehat{\boldsymbol{\beta}}_r^T \mathbf{t}_j$$

Weighted sum of polynomials weighted by logistic functions

⇒ Adapted to abrupt and smooth regime change

⇒ Ensure the continuity and the regularity of the estimated curve

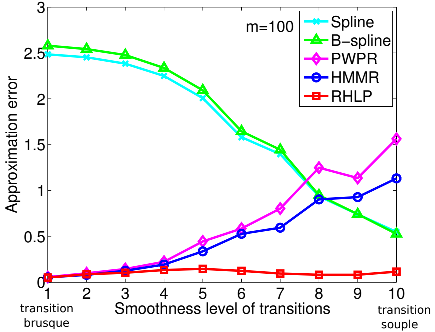
Curve Segmentation

$$\widehat{z}_j = \arg\max_r \pi_r(t_j;\widehat{\mathbf{w}}), \quad (j = 1,\ldots,m)$$

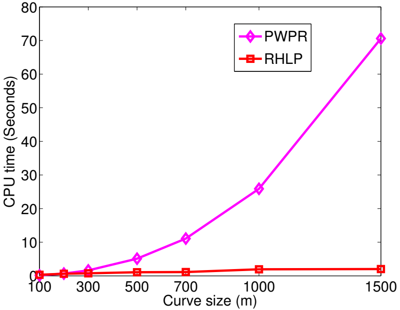
Choice of (R,p) ⇒ $\mathrm{BIC}(R,p) = \mathscr{L}(\widehat{\boldsymbol{\theta}}) - \frac{\nu_{\boldsymbol{\theta}} \log(m)}{2}$

Evaluation in terms of modeling and segmentation

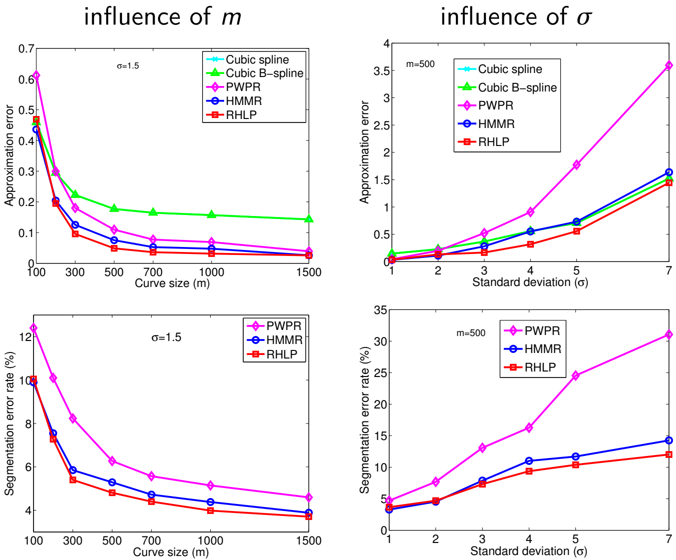
Approximation error as a function of the speed of transitions



Computing time



Evolution of the approximation and segmentation errors



Curves classification

Curves classification

Curves clustering

1. Model-based clustering :

   Regression mixtures, splines, B-splines (Gaffney, 2004 ; James and Sugar, 2003 ; Liu and Yang,2009),

   Mixture of HMMs (Smyth, 1996)

2. Distance-based approach (K-means like) (Hébrail et al., 2010)

Curves classification

Curves clustering

1. Model-based clustering :

   Regression mixtures, splines, B-splines (Gaffney, 2004 ; James and Sugar, 2003 ; Liu and Yang,2009),

   Mixture of HMMs (Smyth, 1996)

2. Distance-based approach (K-means like) (Hébrail et al., 2010)

Curves classification

1. Functional Linear Discriminant Analysis (James and Hastie, 2001)
2. Functional Mixture Discriminant Analysis (Gui and Li, 2003) (B-splines)

Curves clustering

Data : n independent curves $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ (observed)
hidden classes $(h_1, \ldots, h_n)$, hidden regimes $(z_{1k}, \ldots, z_{mk})$ of class k

Curves clustering

Data : n independent curves $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ (observed)
hidden classes $(h_1, \ldots, h_n)$, hidden regimes $(z_{1k}, \ldots, z_{mk})$ of class k

Unsupervised learning, clustering and segmentation

(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

$$p(\mathbf{y}_i|\mathbf{t};\boldsymbol{\Psi}) = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{cluster prob.}} \overbrace{\prod_{j=1}^{m} \sum_{r=1}^{R} \underbrace{\pi_{kr}(t_j;\mathbf{w}_k)}_{\text{regime prob.}} \underbrace{\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T \mathbf{t}_j, \sigma_{kr}^2)}_{\text{Noisy polynomial regime}}}^{\text{cluster : RHLP component density}}$$

Curves clustering

Data : n independent curves $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ (observed)
hidden classes $(h_1, \ldots, h_n)$, hidden regimes $(z_{1k}, \ldots, z_{mk})$ of class k

Unsupervised learning, clustering and segmentation

(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

$$p(\mathbf{y}_i|\mathbf{t};\mathbf{\Psi}) = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{cluster prob.}} \overbrace{\prod_{j=1}^{m}\sum_{r=1}^{R} \underbrace{\pi_{kr}(t_j;\mathbf{w}_k)}_{\text{regime prob.}} \underbrace{\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T\mathbf{t}_j,\sigma_{kr}^2)}_{\text{Noisy polynomial regime}}}^{\text{cluster : RHLP component density}}$$

- Log-likelihood :

$$\mathcal{L}(\mathbf{\Psi};\mathbf{Y},\mathbf{t}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m}\sum_{r=1}^{R} \pi_{kr}(t_j;\mathbf{w}_k)\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T\mathbf{t}_j,\sigma_{kr}^2)$$

Curves clustering

Data : n independent curves $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ (observed)
hidden classes $(h_1, \ldots, h_n)$, hidden regimes $(z_{1k}, \ldots, z_{mk})$ of class k

Unsupervised learning, clustering and segmentation

(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

$$p(\mathbf{y}_i|\mathbf{t};\mathbf{\Psi}) = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{cluster prob.}} \overbrace{\prod_{j=1}^{m} \sum_{r=1}^{R} \underbrace{\pi_{kr}(t_j;\mathbf{w}_k)}_{\text{regime prob.}} \underbrace{\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T\mathbf{t}_j,\sigma_{kr}^2)}_{\text{Noisy polynomial regime}}}^{\text{cluster : RHLP component density}}$$

- Log-likelihood :

$$\mathcal{L}(\mathbf{\Psi};\mathbf{Y},\mathbf{t}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j;\mathbf{w}_k)\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T\mathbf{t}_j,\sigma_{kr}^2)$$

- Maximisation of the log-likelihood by using the EM algorithm (ADAC, 2011)

EM Algorithm

- **Initialisation :** $\mathbf{\Psi}^{(0)}$, $q \longleftarrow 0$ (q itération)

EM Algorithm

- **Initialisation :** $\boldsymbol{\Psi}^{(0)}$, $q \longleftarrow 0$ (q itération)

1. **Étape E : Expectation**

$$
\begin{aligned}
Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(q)}) &= \mathbb{E}\Big[\mathscr{L}_c(\boldsymbol{\Psi};\mathbf{Y},\mathbf{t},\mathbf{h},\mathbf{z}_1,\ldots,\mathbf{z}_K)|\mathbf{Y},\mathbf{t};\boldsymbol{\Psi}^{(q)}\Big]\\
&= \sum_{k=1}^{K}\sum_{i=1}^{n}\tau_{ik}^{(q)}\log\alpha_g + \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m}\tau_{ik}^{(q)}\gamma_{ijkr}^{(q)}\log\pi_{kr}(t_j;\mathbf{w}_k)\\
&+ \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m}\tau_{ik}^{(q)}\gamma_{ijkr}^{(q)}\log\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^T\mathbf{r}_j,\sigma_{kr}^2)
\end{aligned}
$$

$$\tau_{ik}^{(q)} = p(h_i = k|\mathbf{y}_i;\boldsymbol{\Psi}^{(q)})$$: Posterior prob that $\mathbf{y}_i$ belongs to class k

$$\gamma_{ijkr}^{(q)} = p(z_{jk} = r|y_{ij};\boldsymbol{\Psi}^{(q)})$$: Posterior prob that y_{ij}belong to regime r of class k

EM Algorithm

- **Initialisation :** $\mathbf{\Psi}^{(0)}$, $q \longleftarrow 0$ (q itération)

1. **Étape E : Expectation**

$$
\begin{aligned}
Q(\mathbf{\Psi},\mathbf{\Psi}^{(q)}) &= \mathbb{E}\Big[\mathscr{L}_c(\mathbf{\Psi};\mathbf{Y},\mathbf{t},\mathbf{h},\mathbf{z}_1,\ldots,\mathbf{z}_K)|\mathbf{Y},\mathbf{t};\mathbf{\Psi}^{(q)}\Big]\\
&= \sum_{k=1}^{K}\sum_{i=1}^{n}\tau_{ik}^{(q)}\log\alpha_g + \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m}\tau_{ik}^{(q)}\gamma_{ijkr}^{(q)}\log\pi_{kr}(t_j;\mathbf{w}_k)\\
&+ \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m}\tau_{ik}^{(q)}\gamma_{ijkr}^{(q)}\log\mathcal{N}(y_{ij};\boldsymbol{\beta}_{kr}^{T}\boldsymbol{r}_j,\sigma_{kr}^2)
\end{aligned}
$$

$$
\begin{aligned}
\tau_{ik}^{(q)} &= p(h_i=k|\mathbf{y}_i;\mathbf{\Psi}^{(q)}) \text{ : Posterior prob that } \mathbf{y}_i \text{ belongs to class } k\\
\gamma_{ijkr}^{(q)} &= p(z_{jk}=r|y_{ij};\mathbf{\Psi}^{(q)}) \text{ : Posterior prob that } y_{ij} \text{belong to regime } r \text{ of class } k
\end{aligned}
$$

2. **Étape M : Maximisation** : $\mathbf{\Psi}^{(q+1)} = \arg\max_{\mathbf{\Psi}} Q(\mathbf{\Psi},\mathbf{\Psi}^{(q)})$

EM Algorithm

- **Initialisation :** $\mathbf{\Psi}^{(0)}$, $q \longleftarrow 0$ (q itération)

1. **Étape E : Expectation**

$$
\begin{aligned}
Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)}) &= \mathbb{E}\Big[\mathscr{L}_c(\mathbf{\Psi}; \mathbf{Y}, \mathbf{t}, \mathbf{h}, \mathbf{z}_1, \ldots, \mathbf{z}_K)|\mathbf{Y}, \mathbf{t}; \mathbf{\Psi}^{(q)}\Big] \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n} \tau_{ik}^{(q)} \log \alpha_g + \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \pi_{kr}(t_j; \mathbf{w}_k) \\
&+ \sum_{k=1}^{K}\sum_{r=1}^{R_k}\sum_{i=1}^{n}\sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{kr}^T \boldsymbol{r}_j, \sigma_{kr}^2)
\end{aligned}
$$

$$
\begin{aligned}
\tau_{ik}^{(q)} &= p(h_i = k|\mathbf{y}_i; \mathbf{\Psi}^{(q)}) : \text{Posterior prob that } \mathbf{y}_i \text{ belongs to class } k \\
\gamma_{ijkr}^{(q)} &= p(z_{jk} = r|y_{ij}; \mathbf{\Psi}^{(q)}) : \text{Posterior prob that } y_{ij} \text{belong to regime } r \text{ of class } k
\end{aligned}
$$

2. **Étape M : Maximisation** : $\mathbf{\Psi}^{(q+1)} = \arg\max_{\mathbf{\Psi}} Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})$

- $q \longleftarrow q + 1$

M-step :

1. separate maximizations w.r.t the mixing proportions $(\alpha_1,\ldots,\alpha_K)$, the regression parameters $\{\boldsymbol{\beta}_{kr},\sigma^2_{kr}\}$ and the processes' parameters $\{\mathbf{w}_k\}$.

M-step :

1. separate maximizations w.r.t the mixing proportions $(\alpha_1, \ldots, \alpha_K)$, the regression parameters $\{\boldsymbol{\beta}_{kr}, \sigma^2_{kr}\}$ and the processes' parameters $\{\mathbf{w}_k\}$.
2. Updating the mixing proportions : $\alpha_1^{(q+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}^{(q)} \quad (k = 1, \ldots, K)$,

M-step :

1. separate maximizations w.r.t the mixing proportions $(\alpha_1, \ldots, \alpha_K)$, the regression parameters $\{\boldsymbol{\beta}_{kr}, \sigma^2_{kr}\}$ and the processes' parameters $\{\mathbf{w}_k\}$.
2. Updating the mixing proportions : $\alpha_1^{(q+1)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_{ik}^{(q)} \quad (k = 1, \ldots, K)$,
3. for the regression parameters : separate analytic solutions of weighted least-squares problems

$$\boldsymbol{\beta}_{kr}^{(q+1)} = \Big[\sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}^{(q)} \mathbf{t}_j \mathbf{t}_j^T\Big]^{-1} \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}^{(q)} y_{ij} \mathbf{t}_j$$

$$\sigma_{kr}^{2(q+1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}^{(q)} (y_{ij} - \boldsymbol{\beta}_{kr}^{T(q+1)} \mathbf{t}_j)^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}^{(q)}}.$$

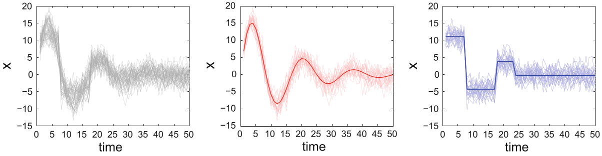
M-step :

1. separate maximizations w.r.t the mixing proportions $(\alpha_1,\ldots,\alpha_K)$, the regression parameters $\{\boldsymbol{\beta}_{kr},\sigma^2_{kr}\}$ and the processes' parameters $\{\mathbf{w}_k\}$.
2. Updating the mixing proportions : $\alpha_1^{(q+1)} = \frac{1}{n}\sum_{i=1}^{n}\gamma_{ik}^{(q)} \quad (k=1,\ldots,K)$,
3. for the regression parameters : separate analytic solutions of weighted least-squares problems

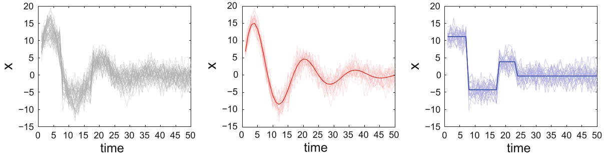
$$
\begin{aligned}
\boldsymbol{\beta}_{kr}^{(q+1)} &= \Big[\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}\mathbf{t}_j\mathbf{t}_j^T\Big]^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}y_{ij}\mathbf{t}_j\\
\sigma_{kr}^{2(q+1)} &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}(y_{ij}-\boldsymbol{\beta}_{kr}^{T(q+1)}\mathbf{t}_j)^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}}.
\end{aligned}
$$

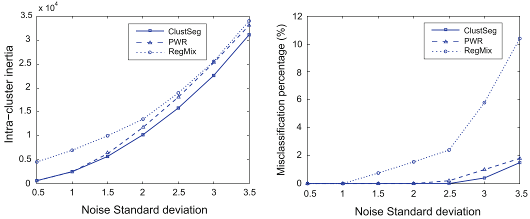
4. the maximization w.r.t the logistic processes parameters $\{\mathbf{w}_{gk}\}$ consists in solving multinomial logistic regression problems weighted by $\gamma_{igk}^{(q)}\tau_{ijgkr}^{(q)} \Rightarrow$ solved with a multi-class IRLS algorithm

Experiments on simulated data



Experiments on simulated data





Curve classification (Discrimination)

Data : n independent labeled curves $((\mathbf{y}_1, c_1), \ldots, (\mathbf{y}_n, c_n))$

- Generative Functional discriminant analysis
- Assign a (new) curve $\mathbf{y}_i$ to the class c_i using the MAP rule :

$$c_i = \arg \max_{1 \leq g \leq G} \frac{\overbrace{w_g}^{prior} \overbrace{p(\mathbf{y}_i | c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g)}^{conditional}}{\text{Cst}}$$

- Classification directly in the space of curves

Curve classification (Discrimination)

Data : n independent labeled curves $((\mathbf{y}_1, c_1), \ldots, (\mathbf{y}_n, c_n))$

- Generative Functional discriminant analysis
- Assign a (new) curve $\mathbf{y}_i$ to the class c_i using the MAP rule :

$$c_i = \arg \max_{1 \leq g \leq G} \frac{\overbrace{w_g}^{prior} \overbrace{p(\mathbf{y}_i | c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g)}^{conditional}}{\mathrm{Cst}}$$

- Classification directly in the space of curves

There are different ways to model the conditional density $p(\mathbf{y}_i | c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g)$:

1. Functional Linear (or Quadratic) Discriminant Analysis (FLDA) (James 2001))
2. Functional Mixture Discriminant Analysis (FMDA) (Gui 2003).

Curve classification (Discrimination)

Homogeneous classes : Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve “model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$

Curve classification (Discrimination)

Homogeneous classes : Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve "model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
$p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$
- Parameter Estimation with EM similar to the case of a single curve
(Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)

Curve classification (Discrimination)

Homogeneous classes : Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve “model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$
- Parameter Estimation with EM similar to the case of a single curve *(Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)*

Dispersed classes : Functional Mixture Discriminant Analysis

- Summarise the curves by several models, each model is associated with a subclass
- Mixture distribution for each class of curves (MixRHLP)
  $p(\mathbf{y}_i|c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^{m} \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; \mathbf{w}_{gk}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gkr}^T \mathbf{t}_j, \sigma_{gkr}^2)$

Curve classification (Discrimination)

Homogeneous classes : Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve “model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}\big(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2\big)$
- Parameter Estimation with EM similar to the case of a single curve
  (Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)

Dispersed classes : Functional Mixture Discriminant Analysis

- Summarise the curves by several models, each model is associated with a subclass
- Mixture distribution for each class of curves (MixRHLP)
  $p(\mathbf{y}_i|c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^{m} \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; \mathbf{w}_{gk}) \mathcal{N}\big(y_{ij}; \boldsymbol{\beta}_{gkr}^T \mathbf{t}_j, \sigma_{gkr}^2\big)$
- Parameter estimation via EM as in the clustering case
  (ESANN 2012, IJCNN 2012)

Experiments

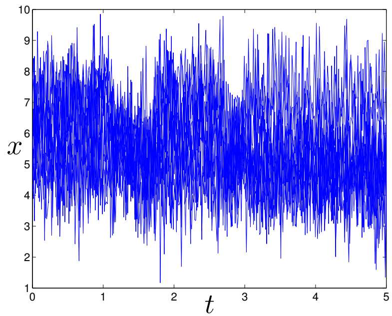
- evaluation of the proposed FMDA-MixRHLP approach on simulated data
- comparisons with FLDA approaches using a polynomial regression (FLDA-PR) or a spline regression (FLDA-SR) model (James 2001)), and the one that uses a single RHLP model (FLDA-RHLP) (Chamroukhi et al. 2010).
- comparisons with alternative FMDA approaches that use polynomial regression mixtures (FMDA-PRM) (Gaffney 2004), and spline regression mixtures (FMDA-SRM) (Gui 2003)

Evaluation criteria

- the misclassification error rate computed by a 5-fold cross-validation
- the intra-class inertia :
  - For FLDA : inertia $= \sum_g \sum_{i|y_i=g} \|\mathbf{x}_i - \mathbf{m}_g\|^2$
  - for FMDA : inertia $= \sum_g \sum_{i|y_i=g} \sum_{k=1}^{K_g} \|\mathbf{x}_i - \mathbf{m}_{gk}\|^2$

Experiments using simulated data

- simulated curves issued from two classes of piecewise noisy functions
- Including a complex shaped class composed of three sub-classes, and a homogeneous class
- Each curve consists of three piecewise regimes and is composed of 200 points



Simulation results

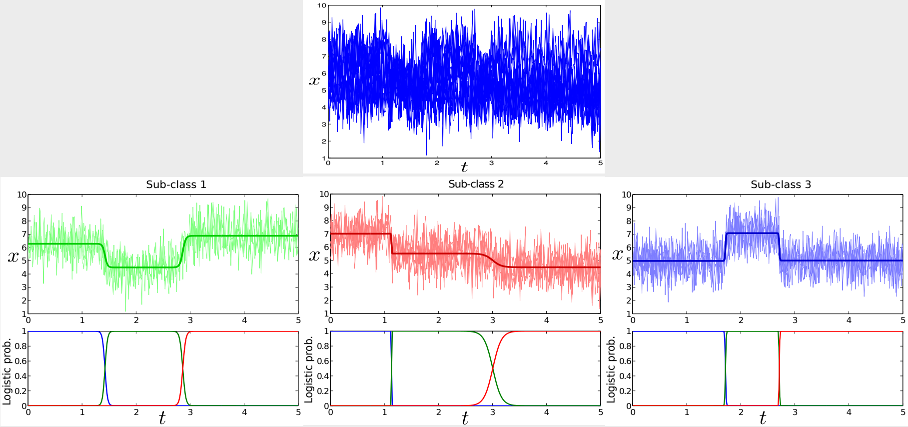


FIGURE: The estimated sub-classes colored according to the partition given by the EM algorithm for the proposed approach (top); Then are presented separately each sub-class of curves with the estimated mean curve in bold line (top sub-plot) and the corresponding logistic probabilities that govern the hidden regimes

Simulation Results

Discrimination Approach	Classif. error rate (%)	Intra-class inertia ($\times 10^3$)
FLDA-PR	21	7.1364
FLDA-SR	19.3	6.9640
FLDA-RHLP	18.5	6.4485
FMDA-PRM	11	6.1735
FMDA-SRM	9.5	5.3570
FMDA-MixRHLP	5.3	3.8095

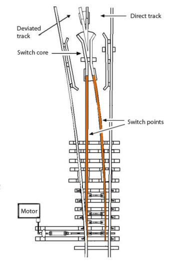
- FMDA approaches provide better results compared to FLDA approaches
- using a single model for complex-shaped classes (i.e., when using FLDA approaches) is not adapted for complex-shaped classes
- the proposed FMDA approach based on hidden logistic process regression (FMDA-MixRHLP) outperforms the alternative FMDA approaches thanks to the flexibility of the logistic process well adapted for curves with regime changes.

Application to the study of a railway system

Context

- Collaboration with la SNCF
- Diagnosis and monitoring of a component of the railway infrastructure

Switch mechanism

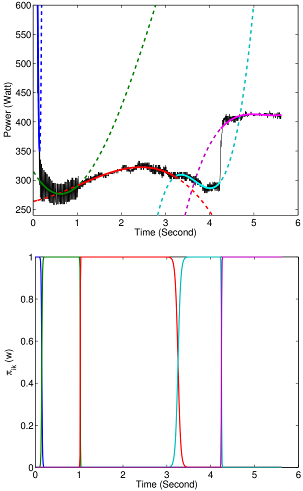


Objectives

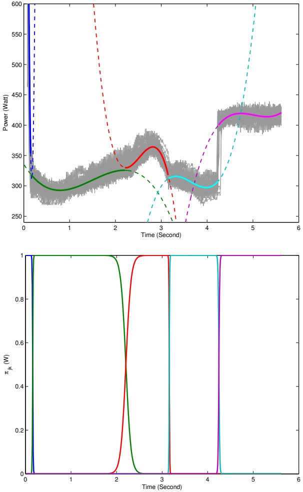
- Estimating the state of component (diagnosis)
- Monitor its status over time

Visualization of modeling results on real data

One curves

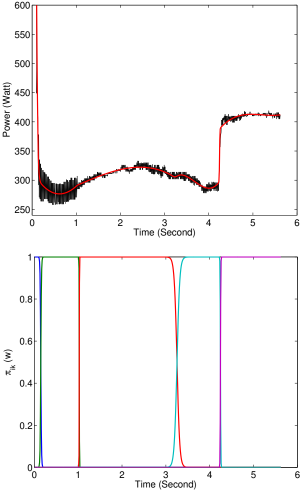


A class of curves

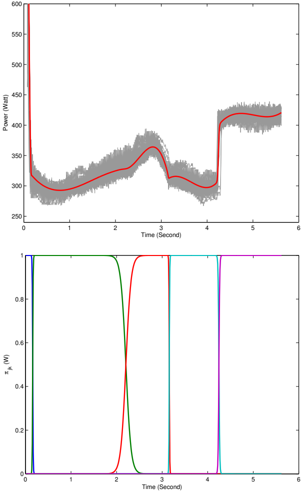


Visualization of modeling results on real data

One curves



A class of curves



Classification of switch operation signals

- Diagnosis results

Approaches	Classification error rate (%)
PSR-MDA	13 $\pm$(4.5)
PWPR-MDA	12 $\pm$(1.7)
HMMR-MDA	9 $\pm$(2.25)
RHLP-MDA	**4** $\pm$**(1.33)**
Functional LDA	
PSR-MAP	7.3 $\pm$ (4.36)
PWPR-MAP	1.82 $\pm$ (5.74)
RHLP-MAP	**1.67** $\pm$ **(2.28)**

Classification of real data by FMDA

- Database issued from a real french railway diagnosis application
- two classes :
  - Class 1 : curves with no defect or with a minor defect
  - Class 2 : curves with a critical defect

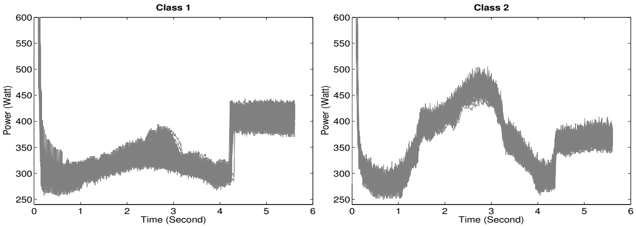
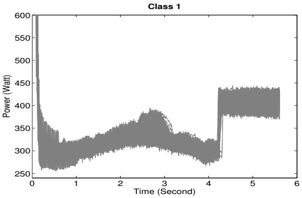
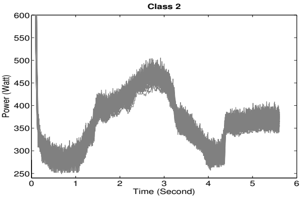


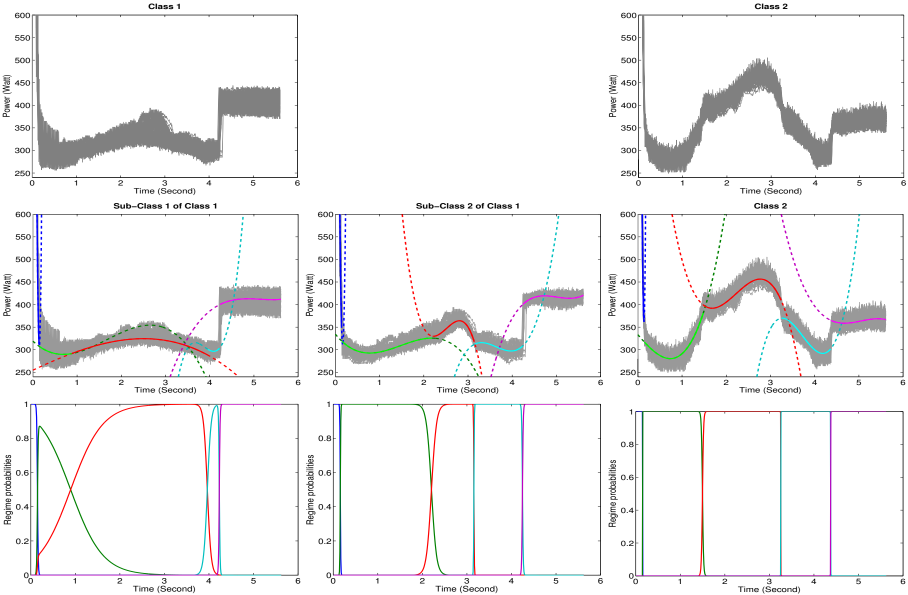
FIGURE: 75 switch operation curves from the first class (left) and 45 curves from the second class (right).

Results

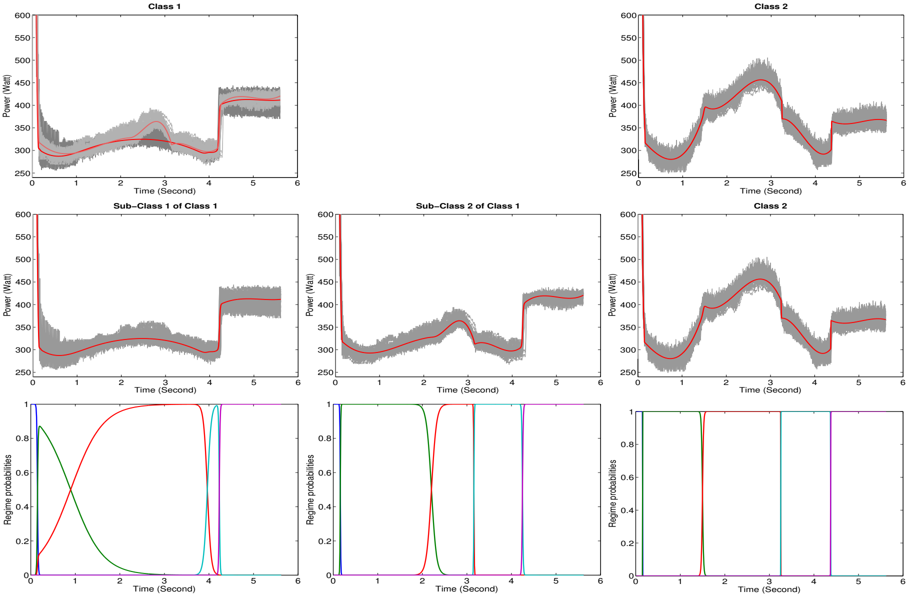

Class 1
Power (Watt)
Time (Second)
600
550
500
450
400
350
300
250
0
1
2
3
4
5
6


Class 2
Power (Watt)
Time (Second)
600
550
500
450
400
350
300
250
0
1
2
3
4
5
6

Results


Class 1
Class 2
Sub−Class 1 of Class 1
Sub−Class 2 of Class 1
Class 2
Power (Watt)
Time (Second)
Regime probabilities

Results


Class 1
Sub−Class 1 of Class 1
Sub−Class 2 of Class 1
Class 2
Class 2
Power (Watt)
Time (Second)
Regime probabilities

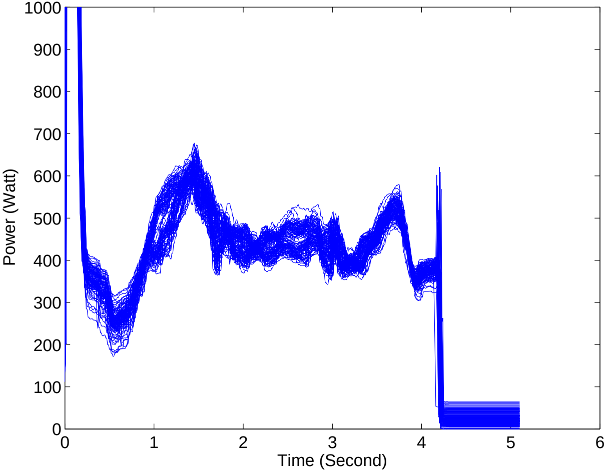
Results

Discrimination approach	Classif. error rate (%)	Intra-class inertia ($\times 10^9$)
FLDA-PR	11.5	10.7350
FLDA-SR	9.53	9.4503
FLDA-RHLP	8.62	8.7633
FMDA-PRM	9.02	7.9450
FMDA-SRM	8.50	5.8312
FMDA-MixRHLP	6.25	3.2012

TABLE: Obtained results for the real curves.

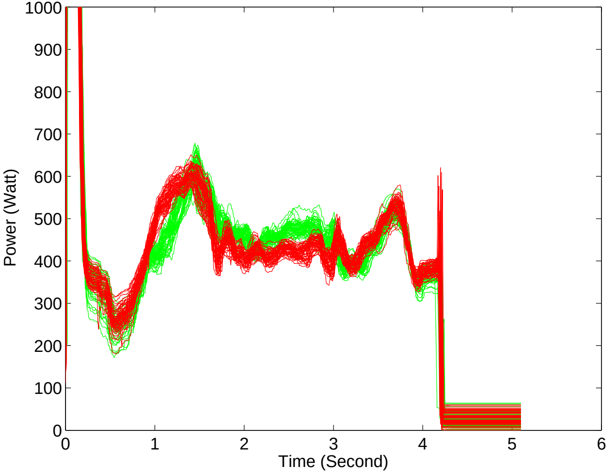
Curves clustering

Real Data : 115 signals



Curves clustering

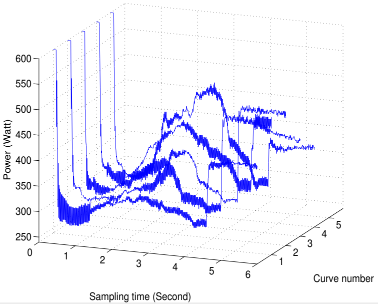
Real data : graphical results



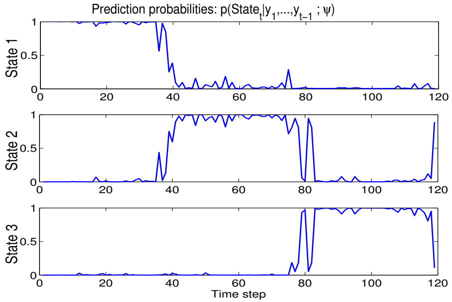
Temporal monitoring of the curves

non-homogeneous autoregressive HMM : prediction results

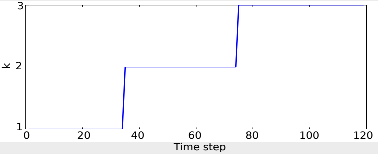
Sequence of curves



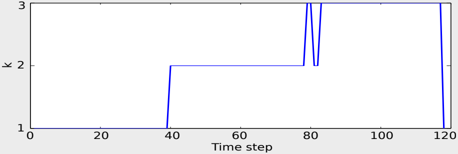
prediction prof the system state



True sequence



Estimated sequence



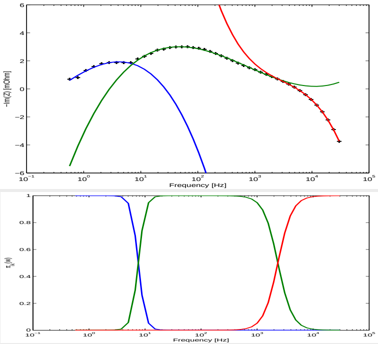
Modeling of impedance spectra of Fuel Cells

- Context : energy of transportation
- *Objective* : Estimation of the fuel cell lifetime.
- Representation of impedance spectrum data

Modeling of impedance spectra of Fuel Cells

- Context : energy of transportation
- *Objective* : Estimation of the fuel cell lifetime.
- Representation of impedance spectrum data

  → a probabilistic approach : the RHLP model *(IEEE ICMLA 2009)*



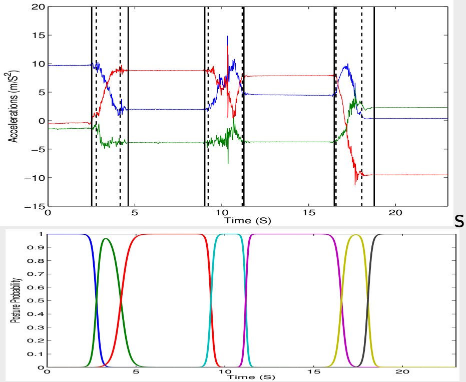
Activity recognition from acceleration data

- Context : Assistive Robotics
- *Objective : Gesture recognition for helping ageing people*
- Data : accelerations, ...

(a) (b) (c) (d) (e) (f)

Unsupervised classification of acceleration data

- ⇒ Joint segmentation of temporal multidimensional data
- Multidimensional regression with a hidden process process
    - *IEEE Transactions on Automation Science and Engineering (Accepted)*
    - *Neurocomputing (in revision)*



Conclusions and perspectives

Conclusions

1. Curves modeling :
   - A Hidden process regression model for curves with regime changes
     - Suited to abrupt and/or smooth regime changes
     - Unsupervised Learning based on the EM algorithm

Conclusions

1. Curves modeling :
    - A Hidden process regression model for curves with regime changes
        - Suited to abrupt and/or smooth regime changes
        - Unsupervised Learning based on the EM algorithm
2. Curves classification (Discrimination and clustering)
    - Direct Classification of curves without preliminary representation
    - Adapted to heterogeneous classes

Conclusions

1. Curves modeling :
   - A Hidden process regression model for curves with regime changes
     - Suited to abrupt and/or smooth regime changes
     - Unsupervised Learning based on the EM algorithm
2. Curves classification (Discrimination and clustering)
   - Direct Classification of curves without preliminary representation
   - Adapted to heterogeneous classes
3. Application to real problems

Merci de votre attention !