Apprentissage de modèles probabilistes à processus latent à partir de données temporelles

Faicel Chamroukhi
Maître de Conférences
USTV, LSIS UMR CNRS 6168

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

Séminaire GFD-LIPADE Univ. Paris 5

05 juillet 2012

Plan

Contextes

1. Apprentissage supervisé : données (x,y)
2. Apprentissage non supervisé : données x, y ?
3. semi supervisé, partiellement supervisé ..

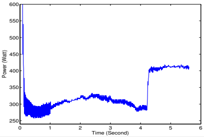
Génératif/Discriminatif

1. Approche discriminative : apprendre directement $p(y|x)$
2. Approche générative : apprendre $p(x,y) \Rightarrow p(y|x) \propto p(x|y)p(y)$
   - S'intéresse au processus de génération des données
   - S'adapte "facilement" au contexte non supervisé

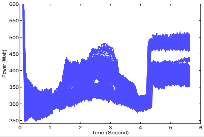
$\Rightarrow$ Modèles à variable latente

Contexte

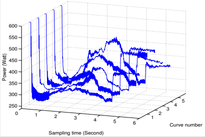
- Données temporelles (signaux, séquences, fonctions, séries temporelles, ..)



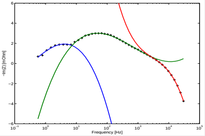
Courbe



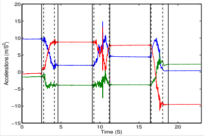
Ensemble de courbes



Séquence de courbes



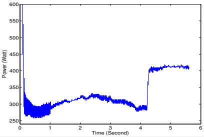
Signal d'impédance de PAC



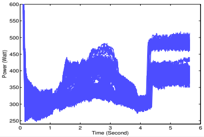
Mesures d'accélérations

Contexte

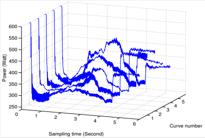
- Données temporelles (signaux, séquences, fonctions, séries temporelles, ..)



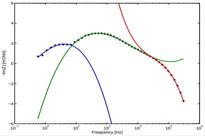
Courbe



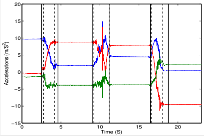
Ensemble de courbes



Séquence de courbes



Signal d'impédance de PAC



Mesures d'accélérations

1. Plusieurs régimes se succédant au cours du temps (aspect temporel) ⇒ Changements brusques ou/et lents de régimes

2. Plusieurs courbes à analyser

3. Données temporelles multidimensionnelles

Objectifs

- Prendre en compte la dispersion des classes et les changements de regime
- Intégrer explicitement cette complexité (class dispersion, régimes,..)

  ⇒ Apprendre des modèles probabilistes génératifs

Objectifs

- Prendre en compte la dispersion des classes et les changements de regime
- Intégrer explicitement cette complexité (class dispersion, régimes,..)

  ⇒ Apprendre des modèles probabilistes génératifs

- L'approche générative s'intéresse au processus générant les données
- Mieux adaptée au contexte non-supervisé que la discriminative

Modélisation de courbes (fonctions)

- Approches à base de régression dynamique à processus latent
- Formalisation probabiliste des changements de régimes

Classification de courbes (supervisée et non supervisée)

- Approche à base de modèle de mélange pour la classification
  - model-based functional cluster analysis (en non supervisé)
  - model-based functional discriminant analysis (en supervisé)

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

- Régression polynomiale simple : f est une fonction polynôme

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

- Régression polynomiale simple : f est une fonction polynôme
- Régression polynomiale par splines (Deboor, 1978)

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

- Régression polynomiale simple : f est une fonction polynôme
- Régression polynomiale par splines (Deboor, 1978)
- f est un polynôme par morceaux (McGee et Carleton, 1970)

  $\Rightarrow$ Optimisation exacte par la prog. dynamique (Bellman, 1961)

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

- Régression polynomiale simple : f est une fonction polynôme
- Régression polynomiale par splines (Deboor, 1978)
- f est un polynôme par morceaux (McGee et Carleton, 1970)

$\Rightarrow$ Optimisation exacte par la prog. dynamique (Bellman, 1961)

- La prog. dynamique peut s'avérer coûteuse en temps de calcul
- Adaptée aux changements brusques mais problème de continuité

Modélisation par régression

Données : 1 courbe $\mathbf{y} = (y_1, \ldots, y_m)$ observée régulièrement aux instants $(t_1, \ldots, t_m)$

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

- Régression polynomiale simple : f est une fonction polynôme
- Régression polynomiale par splines (Deboor, 1978)
- f est un polynôme par morceaux (McGee et Carleton, 1970)

  ⇒ Optimisation exacte par la prog. dynamique (Bellman, 1961)

  - La prog. dynamique peut s'avérer coûteuse en temps de calcul
  - Adaptée aux changements brusques mais problème de continuité
- Régression polynomiale régi par une chaîne de Markov (Fridman, 1993)

  Modèle pour une courbe : $y_j = \boldsymbol{\beta}_{z_j}^T \mathbf{t}_j + \sigma_{z_j}\epsilon_j \quad (j = 1, \ldots, m)$

  - Continuité pas certaine
  - N'est pas adapté pour approximer un ensemble de courbes

Modèle de régression à processus latent (RHLP) proposé

Régression à processus logistique latent (RHLP) proposé :

Neural Networks - Elsevier, 22(5-6) :593-602, 2009.

Données temporelles $\mathbf{y} = (y_1, \ldots, y_m)$ observées aux instants $\mathbf{t} = (t_1, \ldots, t_m)$

Définition du modèle

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

- z_j label caché du modèle de y_j
- $\mathbf{z} = (z_1, \ldots, z_n)$ est un processus logistique caché

Régression à processus logistique latent (RHLP) proposé :

Neural Networks - Elsevier, 22(5-6) :593-602, 2009.

Données temporelles $\mathbf{y} = (y_1, \ldots, y_m)$ observées aux instants $\mathbf{t} = (t_1, \ldots, t_m)$

Définition du modèle

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

- z_j label caché du modèle de y_j
- $\mathbf{z} = (z_1, \ldots, z_n)$ est un processus logistique caché

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

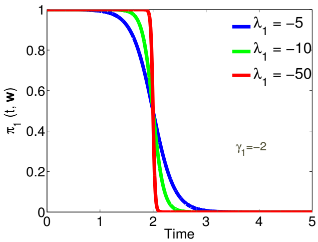
$$\pi_r(t_j; \mathbf{w}) = p(z_j = r | t_j; \mathbf{w}) = \frac{\exp(\lambda_r(t_j + \gamma_r))}{\sum_{\ell=1}^R \exp(\lambda_\ell(t_j + \gamma_\ell))}$$

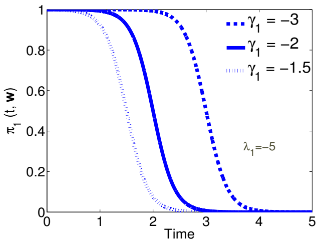
- $\boldsymbol{w}_r = (\lambda_r, \gamma_r)^T$ paramètre de la *rieme* fonction logistique
- $\mathbf{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_R)$ paramètre pour les R fonctions logistiques

Flexibilité de la transformation logistique

Variation temporelle de la fonction logistique en fonction de $\boldsymbol{w}$

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





⇒ Le paramètre λ_r contrôle la vitesse des transitions entre les régimes

⇒ Le paramètre γ_r contrôle l'instant de transition

Illustration du principe de la méthode

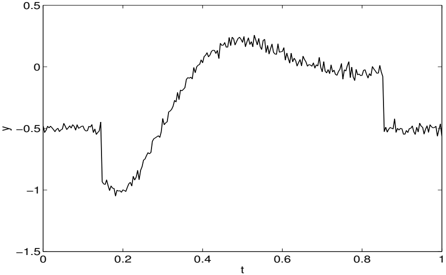
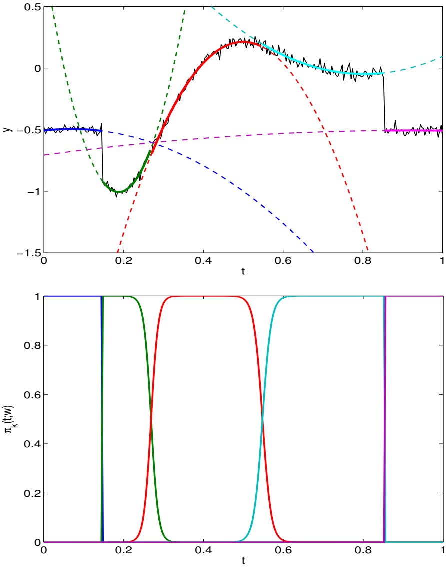


Illustration du principe de la méthode



Estimation des paramètres par MV via EM

- Vecteur paramètre du modèle : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-vraisemblance : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$

Estimation des paramètres par MV via EM

- Vecteur paramètre du modèle : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-vraisemblance : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Log-vraisem. complétée : $\mathscr{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log \left[\pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)\right]$

Estimation des paramètres par MV via EM

- Vecteur paramètre du modèle : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-vraisemblance : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Log-vraisem. complétée : $\mathscr{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log \left[\pi_r(t_j; \mathbf{w}) \mathcal{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)\right]$

Algorithme EM pour le modèle proposé

Paramètre initial $\boldsymbol{\theta}^{(0)}$:

1. E-Step : Espérance conditionnelle de la log-vraisemblance complétée $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) =$

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

$\Rightarrow$ Calcul des probabilités a posteriori $\tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \boldsymbol{\theta}^{(q)})$

Estimation des paramètres par MV via EM

- Vecteur paramètre du modèle : $\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Log-vraisemblance : $\mathscr{L}(\boldsymbol{\theta}) = \sum_{j=1}^m \log \sum_{r=1}^R \pi_r(t_j; \mathbf{w}) \mathscr{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)$
- Log-vraisem. complétée : $\mathscr{L}_c(\boldsymbol{\theta}) = \sum_{j=1}^m \sum_{r=1}^R z_{jr} \log [\pi_r(t_j; \mathbf{w}) \mathscr{N}(y_j; \boldsymbol{\beta}_r^T \mathbf{t}_j, \sigma_r^2)]$

Algorithme EM pour le modèle proposé

Paramètre initial $\boldsymbol{\theta}^{(0)}$:

1. E-Step : Espérance conditionnelle de la log-vraisemblance complétée $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) =$

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

$\Rightarrow$ Calcul des probabilités a posteriori $\tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \boldsymbol{\theta}^{(q)})$

2. M-Step : $\boldsymbol{\theta}^{(q+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) = Q_{\mathbf{w}}(\mathbf{w}, \boldsymbol{\theta}^{(q)}) + \sum_{r=1}^R Q_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)})$
   $\Rightarrow$ Maximisations séparées de $Q_{\boldsymbol{\theta}_r}$ et de $Q_{\mathbf{w}}$

- Maximisation de $Q_{\boldsymbol{\theta}_r}$: solutions exactes de régressions pondérées par les 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 de $Q_{\mathbf{w}}$: un problème convexe de régression logistique multi-classes pondéré par les $\tau_{jr}^{(q)}$ $\Rightarrow$ méthode itérative IRLS

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

Approximation et segmentation d'une courbe

Approximation d'une courbe

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

Somme de polynômes pondérés par des fonctions logistiques

⇒ Adaptée pour les changement lents ou rapides

⇒ Assure la continuité et la régularité de la courbe estimée

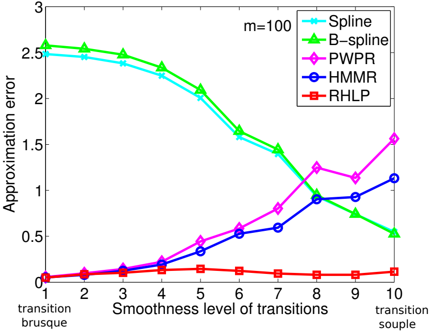
Segmentation d'une courbe

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

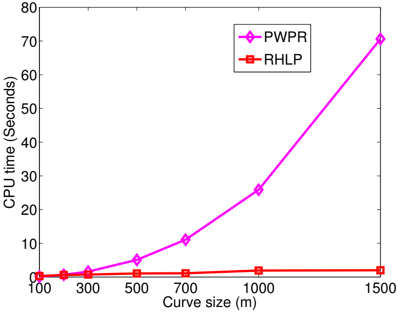
Choix de (R,p) ⇒ $\mathrm{BIC}(R,p) = \mathscr{L}(\widehat{\boldsymbol{\theta}}) - \frac{\nu_{\boldsymbol{\theta}} \log(m)}{2}$

Évaluation en terme de modélisation et de segmentation

Évolution de l'erreur d'approximation en fonction de la vitesse des transitions



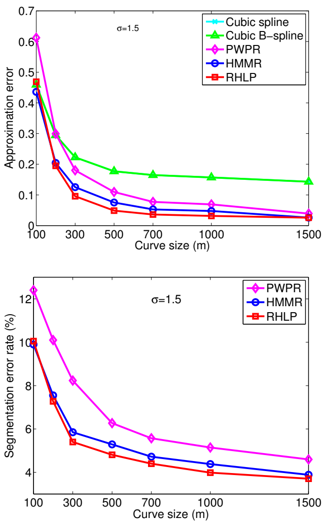
Temps de calcul en fonction de la taille m d'une courbe



Évolution de l'erreur d'approximation et de segmentation

influence de m

influence de σ



Classification de courbes indépendantes

Classification de courbes

Classification non supervisée

1. Clustering à base de modèle :

   Mélange de régressions polynomiales, splines, B-splines (Gaffney, 2004 ; James and Sugar, 2003 ; Liu and Yang,2009),

   Mélange de HMMs (Smyth, 1996)

2. Approche à base de critère de distance : Régression par morceaux (Hébrail et al., 2010)

Classification de courbes

Classification non supervisée

1. Clustering à base de modèle :

   Mélange de régressions polynomiales, splines, B-splines (Gaffney, 2004 ; James and Sugar, 2003 ; Liu and Yang,2009),

   Mélange de HMMs (Smyth, 1996)

2. Approche à base de critère de distance : Régression par morceaux (Hébrail et al., 2010)

Classification supervisée

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

Classification non supervisée de courbes

Données : n courbes indépendantes $(\mathbf{y}_1,\ldots,\mathbf{y}_n)$ observées
classes $(h_1,\ldots,h_n)$ cachées, régimes $(z_{1k},\ldots,z_{mk})$ cachés de la classe k

Classification non supervisée de courbes

Données : n courbes indépendantes $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ observées
classes $(h_1, \ldots, h_n)$ cachées, régimes $(z_{1k}, \ldots, z_{mk})$ cachés de la classe k

Apprentissage non supervisée pour la classification et la segmentation

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

- Modèle de mélange régressions à processus latent RHLP (MixRHLP)

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

Classification non supervisée de courbes

Données : n courbes indépendantes $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ observées
classes $(h_1, \ldots, h_n)$ cachées, régimes $(z_{1k}, \ldots, z_{mk})$ cachés de la classe k

Apprentissage non supervisée pour la classification et la segmentation

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

- Modèle de mélange régressions à processus latent RHLP (MixRHLP)

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

- Log-vraisemblance :

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

Classification non supervisée de courbes

Données : n courbes indépendantes $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ observées
classes $(h_1, \ldots, h_n)$ cachées, régimes $(z_{1k}, \ldots, z_{mk})$ cachés de la classe k

Apprentissage non supervisée pour la classification et la segmentation

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

- Modèle de mélange régressions à processus latent RHLP (MixRHLP)

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

- Log-vraisemblance :

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

- Maximisation de la log-vraisemblance l'algorithme EM (ADAC, 2011)

Algorithme EM

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

Algorithme EM

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

1. **Étape E : Espérance**

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

$\tau_{ik}^{(q)} = p(h_i = k|\mathbf{y}_i;\mathbf{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de $\mathbf{y}_i$ à la classe k

$\gamma_{ijkr}^{(q)} = p(z_{jk} = r|y_{ij};\mathbf{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de y_{ij} au régime r de la classe k

Algorithme EM

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

1. **Étape E : Espérance**

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

$\tau_{ik}^{(q)} \;=\; p(h_i = k | \mathbf{y}_i; \boldsymbol{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de $\mathbf{y}_i$ à la classe k

$\gamma_{ijkr}^{(q)} \;=\; p(z_{jk} = r | y_{ij}; \boldsymbol{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de y_{ij} au régime r de la classe k

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

Algorithme EM

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

1. **Étape E : Espérance**

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

$\tau_{ik}^{(q)} = p(h_i = k|\mathbf{y}_i;\mathbf{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de $\mathbf{y}_i$ à la classe k

$\gamma_{ijkr}^{(q)} = p(z_{jk} = r|y_{ij};\mathbf{\Psi}^{(q)})$: probabilité a posteriori d'appartenance de y_{ij} au régime r de la classe k

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 :

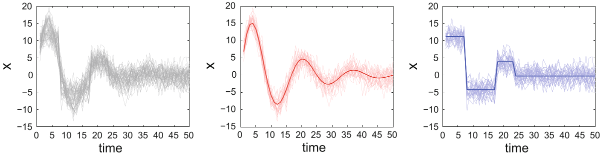
1. separate maximizations w.r.t the mixing proportions $(\alpha_1,\ldots,\alpha_K)$, the regression parameters $\{\boldsymbol{\beta}_{kr},\sigma^2_{kr}\}$ and the processes' parameters $\{\mathbf{w}_k\}$.
2. Updating the mixing proportions : $\alpha_1^{(q+1)} = \frac{1}{n}\sum_{i=1}^n \gamma_{ik}^{(q)}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

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

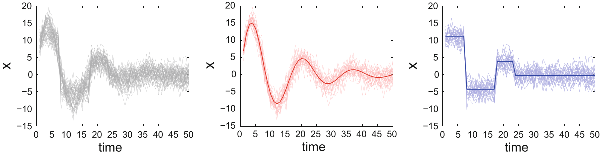
$$\sigma_{kr}^{2(q+1)} = \frac{\sum_{i=1}^n\sum_{j=1}^m \gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}(y_{ij}-\boldsymbol{\beta}_{kr}^{T(q+1)}\mathbf{t}_j)^2}{\sum_{i=1}^n\sum_{j=1}^m \gamma_{ik}^{(q)}\tau_{ijkr}^{(q)}}.$$

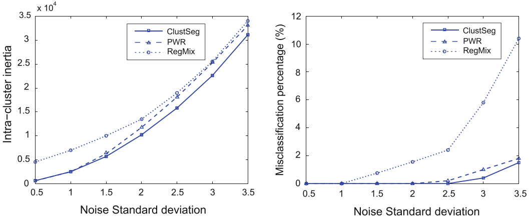
4. the maximization w.r.t the logistic processes parameters $\{\mathbf{w}_{gk}\}$ consists in solving multinomial logistic regression problems weighted by $\gamma_{igk}^{(q)}\tau_{ijgkr}^{(q)} \Rightarrow$ solved with a multi-class IRLS algorithm

Expérimentations sur données simulées



Expérimentations sur données simulées





Classification supervisée de courbes

Données : n courbes indépendantes étiquetées $((\mathbf{y}_1, c_1), \ldots, (\mathbf{y}_n, c_n))$

- 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 directement dans l'espace des courbes

Classification supervisée de courbes

Données : n courbes indépendantes étiquetées $((\mathbf{y}_1, c_1), \ldots, (\mathbf{y}_n, c_n))$

- 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 directement dans l'espace des courbes

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 [?])
2. Functional Mixture Discriminant Analysis (FMDA) (Gui [?]).

Classification supervisée de courbes

Classes homogènes : Functional Linear Discriminant Analysis

- Résumer une classe de courbes en une courbe “modèle" (l'espérance)
- Distribution d'une classe homogène de courbes (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$

Classification supervisée de courbes

Classes homogènes : Functional Linear Discriminant Analysis

- Résumer une classe de courbes en une courbe “modèle" (l'espérance)
- Distribution d'une classe homogène de courbes (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$
- Estimation des paramètres par EM similaire au cas d'une courbe *(Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)*

Classification supervisée de courbes

Classes homogènes : Functional Linear Discriminant Analysis

- Résumer une classe de courbes en une courbe “modèle" (l'espérance)
- Distribution d'une classe homogène de courbes (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$
- Estimation des paramètres par EM similaire au cas d'une courbe *(Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)*

Classes dispersées : Functional Mixture Discriminant Analysis

- Résumé en des courbes modèles, chacune associée à une sous classe
- Distribution mélange d'une classe de courbes (MixRHLP)
  $p(\mathbf{y}_i|c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^{m} \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; \mathbf{w}_{gk}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gkr}^T \mathbf{t}_j, \sigma_{gkr}^2)$

Classification supervisée de courbes

Classes homogènes : Functional Linear Discriminant Analysis

- Résumer une classe de courbes en une courbe “modèle" (l'espérance)
- Distribution d'une classe homogène de courbes (RHLP)
  $p(\{\mathbf{y}_i\}|c_i = g, \mathbf{t}; \boldsymbol{\theta}_g) = \prod_i \prod_{j=1}^m \sum_{r=1}^R \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gr}^T \mathbf{t}_j, \sigma_{gr}^2)$
- Estimation des paramètres par EM similaire au cas d'une courbe
  (Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)

Classes dispersées : Functional Mixture Discriminant Analysis

- Résumé en des courbes modèles, chacune associée à une sous classe
- Distribution mélange d'une classe de courbes (MixRHLP)
  $p(\mathbf{y}_i|c_i = g, \mathbf{t}; \boldsymbol{\Psi}_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^m \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; \mathbf{w}_{gk}) \mathcal{N}(y_{ij}; \boldsymbol{\beta}_{gkr}^T \mathbf{t}_j, \sigma_{gkr}^2)$
- Estimation des paramètres par EM comme dans le cas du clustering
  (ESANN 2012, IJCNN 2012)

Application à l'étude de système ferroviaire

1 Contexte et objectifs

2 Modélisation probabiliste à processus latent de courbes

3 Classification de courbes

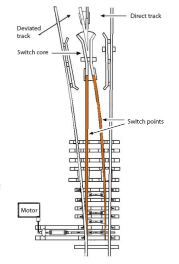
4 Applications
- Diagnostic et télésurveillance ferroviaire
- Energie des Transports
- Robotique Assistive

5 Conclusions

Contexte

- Collaboration avec la SNCF
- Diagnostic et télésurveillance d'un composant de l'infrastructure ferroviaire

Mécanisme d'aiguillage

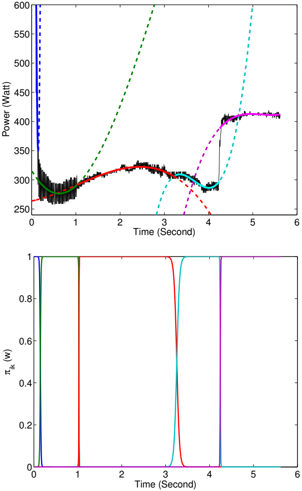


Objectifs visés

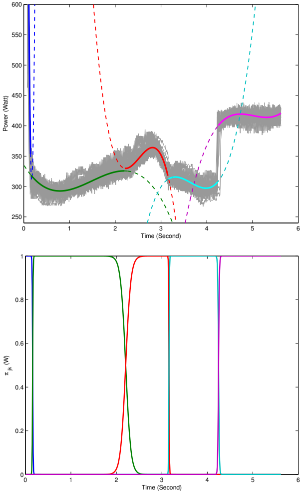
- Estimer l'état de fonctionnement du composant (diagnostic)
- Surveiller son état au cours du temps (suivi temporel)

Visualisation de résultats de modélisations pour données réelles

Une courbe

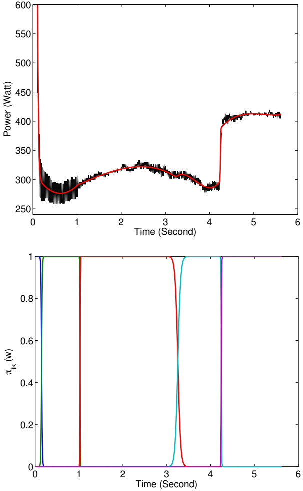


Une classe de courbes

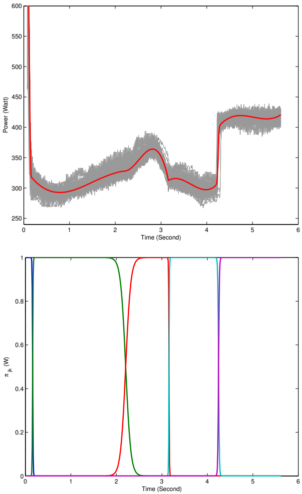


Visualisation de résultats de modélisations pour données réelles

Une courbe



Une classe de courbes



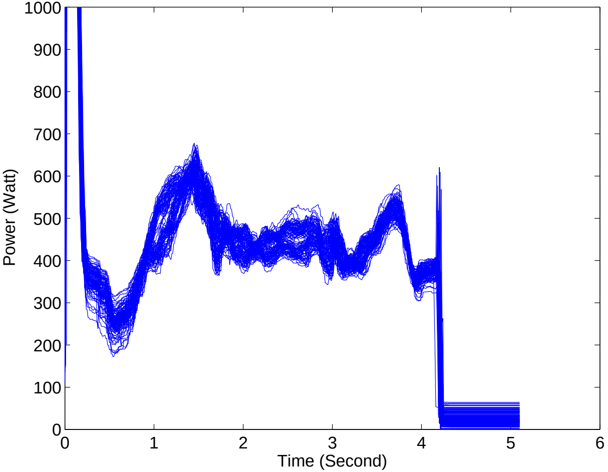
Classification des signaux de manœuvres d'aiguillages

- Résultats de diagnostic

Approches	Taux d'erreur de classification (%)
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)**

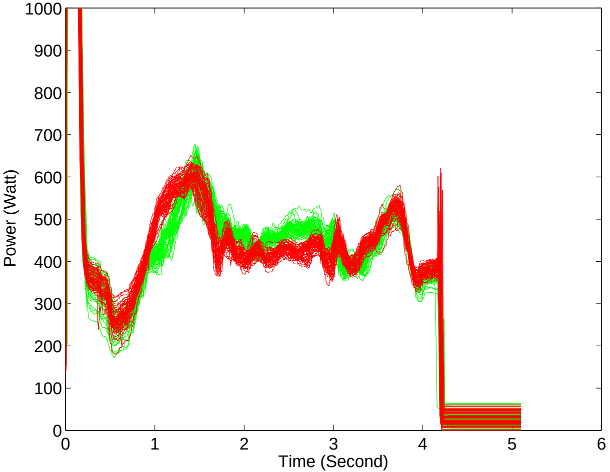
Clustering de courbes : Aide à la décision

Données réelles : 115 signaux de 511 points



Clustering de courbes : Aide à la décision

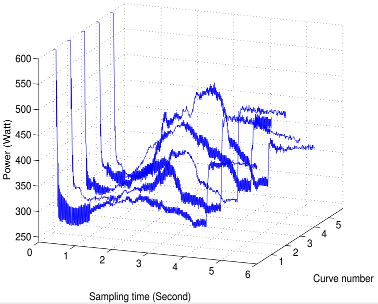
Données réelles : Résultats graphiques



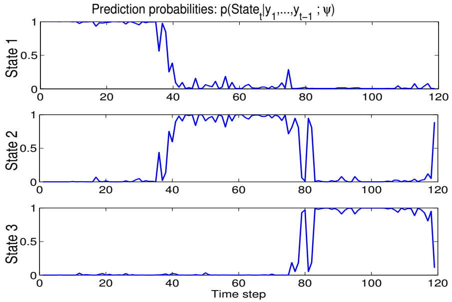
Suivi temporel de courbes

HMM autorégressif non-homogène : Résultats de prédiction

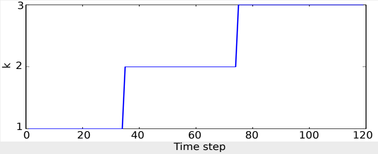
Séquence de courbes



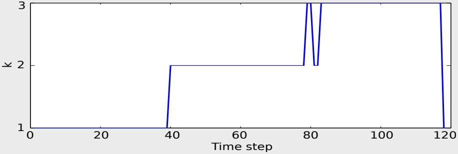
Probas de prédiction de l'état du système



Vrai séquence



Séquence estimée



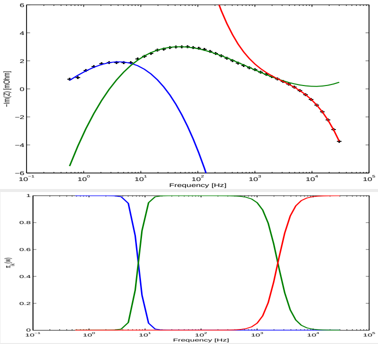
Modélisation de spectres d'impédances de PAC

- Contexte : énergie de transport
- *Objectif* : Estimation de la durée de vie des piles à combustible (PAC).
- Représentation des données de spectre d'impédance

Modélisation de spectres d'impédances de PAC

- Contexte : énergie de transport
- *Objectif* : Estimation de la durée de vie des piles à combustible (PAC).
- Représentation des données de spectre d'impédance

→ Approche probabiliste à base de modèle RHLP *(IEEE ICMLA 2009)*



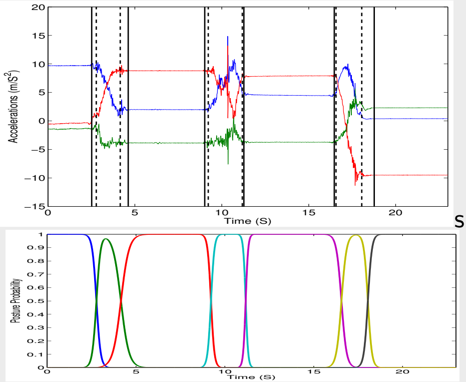
Reconnaissance de postures à partir de mesures d'accélération

- Contexte : Robotique de service
- *Objectif : Reconnaissance de postures pour l'aide à la personne*
- Données : accélérations, ...

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

Classification non supervisé de mesures d'accélération

- ⇒ Segmentation jointe de données temporelles multidimensionnelles *(Neurocomputing, en révision)*
- Régression multidimensionnelle à processus latent



Conclusions et perspectives

Conclusions

1. Modélisation de courbes :
   - Modèle de régression à processus latent proposé pour des courbes à changements de régimes
     - Adaptation aux changements doux ou rapides des régimes
     - Cadre élégant pour l'optimisation des paramètres : EM

Conclusions

1. Modélisation de courbes :
    - Modèle de régression à processus latent proposé pour des courbes à changements de régimes
        - Adaptation aux changements doux ou rapides des régimes
        - Cadre élégant pour l'optimisation des paramètres : EM
2. Classification de courbes (cadre supervisé et non supervisé)
    - Classification directe dans l'espace des courbes
    - Adaptée pour les classes hétérogènes

Conclusions

1. Modélisation de courbes :
    - Modèle de régression à processus latent proposé pour des courbes à changements de régimes
        - Adaptation aux changements doux ou rapides des régimes
        - Cadre élégant pour l'optimisation des paramètres : EM
2. Classification de courbes (cadre supervisé et non supervisé)
    - Classification directe dans l'espace des courbes
    - Adaptée pour les classes hétérogènes

Conclusions

1. Modélisation de courbes :
    - Modèle de régression à processus latent proposé pour des courbes à changements de régimes
        - Adaptation aux changements doux ou rapides des régimes
        - Cadre élégant pour l'optimisation des paramètres : EM
2. Classification de courbes (cadre supervisé et non supervisé)
    - Classification directe dans l'espace des courbes
    - Adaptée pour les classes hétérogènes
3. Application des modèles développés sur des problèmes réels

Perspectives

- Maximum classification likelihood for curve clustering and segmentation (en cours de soumission)
- CEM(-like) algorithm for the mixture of hidden process regression models

  Dedicated to classification rather than estimation
- Apprentissage en ligne à partir de courbes
- Approche Bayésienne pour le controle de la comlexité du modèle

plus générales

- Apprentissage à partir de données fonctionnelles (se poursuit)
- Generative Topographic Mapping (en cours avec le LIPN)
- Apprentissage non supervisé de modèles probabilistes Bayesiens pour la description parcimonieuse et l'analyse de scènes complexes

Merci de votre attention !