

Contributions à l'analyse de données temporelles

Ahlame Douzal

AMA
Laboratoire d'Informatique de Grenoble
Université Joseph Fourier

29 Novembre 2012

Introduction

Définition

- Un cas particulier de données de séquences :
 - une suite ordonnée d'observations
 - critère d'ordre temporel

Omniprésence des données temporelles

- Analyse de comportements d'utilisateurs du web
- Analyse de l'évolution structurelle ou informationnelle au sein de réseaux sociaux
- Analyse de courbes de charge de consommation d'énergie
- Analyse de données issues de réseaux de capteurs

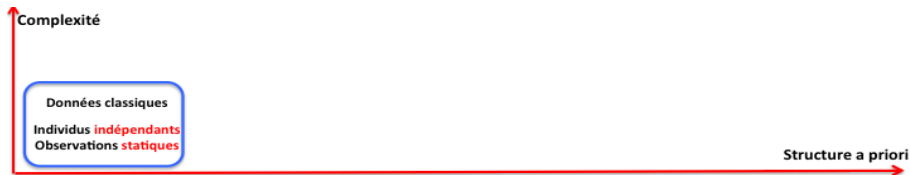
Définition

- Un cas particulier de données de séquences :
 - une suite ordonnée d'observations
 - critère d'ordre temporel

Omniprésence des données temporelles

- Analyse de comportements d'utilisateurs du web
- Analyse de l'évolution structurelle ou informationnelle au sein de réseaux sociaux
- Analyse de courbes de charge de consommation d'énergie
- Analyse de données issues de réseaux de capteurs

Positionnement : analyse de données et structures *a priori*



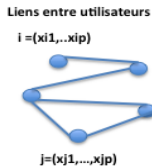
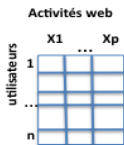
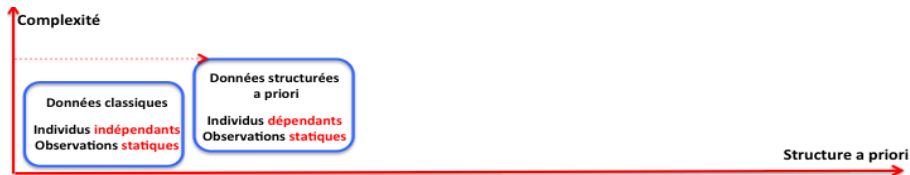
Activités web

	X1	...	Xp
1			
...			
n			

utilisateurs

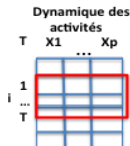
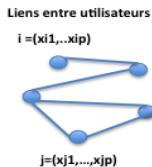
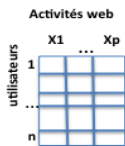
- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités

Positionnement : analyse de données et structures *a priori*



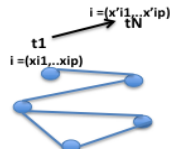
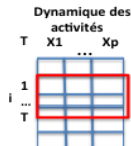
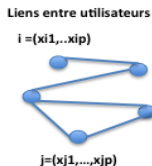
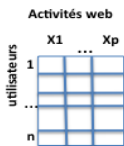
- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités
- Mesurer la dépendance entre l'information structurelle et informationnelle
- Identifier les activités homogènes dans les voisinage (hot spot)

Positionnement : analyse de données et structures *a priori*



- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités
- Mesurer la dépendance entre l'information structurelle et informationnelle
- Identifier les activités homogènes dans les voisinage (hot spot)
- Prédire un comportement type d'usage
- Identifier les caractéristiques dynamiques d'une classe d'usage

Positionnement : analyse de données et structures *a priori*



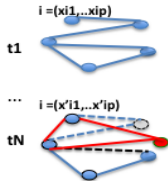
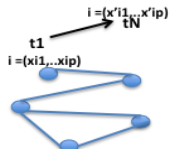
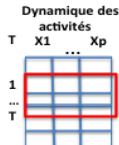
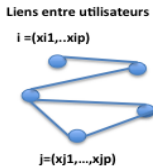
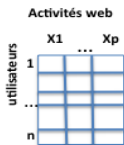
- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités

- Mesurer la dépendance entre l'information structurelle et informationnelle
- Identifier les activités homogènes dans les voisinage (hot spot)

- Prédire un comportement type d'usage
- Identifier les caractéristiques dynamiques d'une classe d'usage

- Mesurer la dépendance entre les dynamiques observées et la structure
- Définir une cartographie des dynamiques

Positionnement : analyse de données et structures *a priori*



- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités

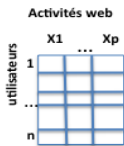
- Mesurer la dépendance entre l'information structurelle et informationnelle
- Identifier les activités homogènes dans les voisinage (hot spot)

- Prédire un comportement type d'usage
- Identifier les caractéristiques dynamiques d'une classe d'usage

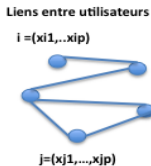
- Mesurer la dépendance entre les dynamiques observées et la structure
- Définir une cartographie des dynamiques

- Identifier les relations ou structures stables
- Mesurer l'effet de la structure évolutive sur les observations
- Prédire l'apparition de liens ou nœuds

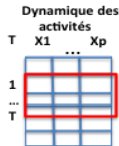
Positionnement : analyse de données et structures *a priori*



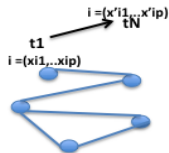
- Identifier des classes d'usage du web
- Mesurer des dépendances entre type d'activités



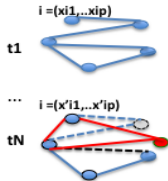
- Mesurer la dépendance entre l'information structurelle et informationnelle
- Identifier les activités homogènes dans les voisinage (hot spot)



- Prédire un comportement type d'usage
- Identifier les caractéristiques dynamiques d'une classe d'usage



- Mesurer la dépendance entre les dynamiques observées et la structure
- Définir une cartographie des dynamiques



- Identifier les relations ou structures stables
- Mesurer l'effet de la structure évolutive sur les observations
- Prédire l'apparition de liens ou nœuds

Analyse de données temporelles

- **Représentation des données temporelles**
- **Définition et apprentissage de métriques**
- **Classification et prédiction de données temporelles**

Analyse de données temporelles

Représentation des données temporelles

- **Segmentation** : représentation compacte de séries multivariées
 - Collaborations avec G. Benchetrit PRETA (TIMC-IMAG) et C. Garbay, encadrements de masters 2R

Métriques associées aux données temporelles

- **Appariements discriminants** : discrimination de classes de séries
 - Thèse C. Frambourg (MENRT, soutenance janvier 2013)
 - Co-direction avec J. Demongeot (TIMC-IMAG)
 - Collaboration avec E. Gaussier
- **"Forme" des séries** : métriques intégrant la forme des séries
 - Thèse A. Diallo (soutenue en 2010)
 - Co-direction avec F. Giroud (TIMC-IMAG)
 - Collaboration avec P. Nagabhushan (Université de Mysore), encadrement de masters 2R

Classification et prédiction de données temporelles

- **Classification par arbre** : extension des arbres de classification à des séries
 - Collaboration avec C. Amblard (AMA), encadrement de masters 2R

Analyse de données temporelles

Représentation des données temporelles

- **Segmentation** : représentation compacte de séries multivariées
 - Collaborations avec G. Benchetrit PRETA (TIMC-IMAG) et C. Garbay, encadrements de masters 2R

Métriques associées aux données temporelles

- **Appariements discriminants** : discrimination de classes de séries
 - Thèse C. Frambourg (MENRT, soutenance janvier 2013)
 - Co-direction avec J. Demongeot (TIMC-IMAG)
 - Collaboration avec E. Gaussier
- **"Forme" des séries** : métriques intégrant la forme des séries
 - Thèse A. Diallo (soutenue en 2010)
 - Co-direction avec F. Giroud (TIMC-IMAG)
 - Collaboration avec P. Nagabhushan (Université de Mysore), encadrement de masters 2R

Classification et prédiction de données temporelles

- **Classification par arbre** : extension des arbres de classification à des séries
 - Collaboration avec C. Amblard (AMA), encadrement de masters 2R

Analyse de données temporelles

Représentation des données temporelles

- **Segmentation** : représentation compacte de séries multivariées
 - Collaborations avec G. Benchetrit PRETA (TIMC-IMAG) et C. Garbay, encadrements de masters 2R

Métriques associées aux données temporelles

- **Appariements discriminants** : discrimination de classes de séries
 - Thèse C. Frambourg (MENRT, soutenance janvier 2013)
 - Co-direction avec J. Demongeot (TIMC-IMAG)
 - Collaboration avec E. Gaussier
- **"Forme" des séries** : métriques intégrant la forme des séries
 - Thèse A. Diallo (soutenue en 2010)
 - Co-direction avec F. Giroud (TIMC-IMAG)
 - Collaboration avec P. Nagabhushan (Université de Mysore), encadrement de masters 2R

Classification et prédiction de données temporelles

- **Classification par arbre** : extension des arbres de classification à des séries
 - Collaboration avec C. Amblard (AMA), encadrement de masters 2R

Analyse de données temporelles

Représentation des données temporelles

- **Segmentation** : représentation compacte de séries multivariées
 - Collaborations avec G. Benchetrit PRETA (TIMC-IMAG) et C. Garbay, encadrements de masters 2R

Métriques associées aux données temporelles

- **Appariements discriminants** : discrimination de classes de séries
 - Thèse C. Frambourg (MENRT, soutenance janvier 2013)
 - Co-direction avec J. Demongeot (TIMC-IMAG)
 - Collaboration avec E. Gaussier
- **"Forme" des séries** : métriques intégrant la forme des séries
 - Thèse A. Diallo (soutenue en 2010)
 - Co-direction avec F. Giroud (TIMC-IMAG)
 - Collaboration avec P. Nagabhushan (Université de Mysore), encadrement de masters 2R

Classification et prédiction de données temporelles

- **Classification par arbre** : extension des arbres de classification à des séries
 - Collaboration avec C. Amblard (AMA), encadrement de masters 2R

Introduction : analyse de la dépendance (1)

• Données indépendantes

$X_{(n \times p)}$: la matrice de description de n individus par p variables, $V(X)$ la matrice de variance covariance associée :

$$V(X) = X^t(I - UP)^t P(I - UP)X$$

$I_{(n \times n)}$ la matrice diagonale identité, $U_{(n \times n)}$ la matrice unité, et P la matrice diagonale des poids des individus $p_i = \frac{1}{n}$ (cas équipondéré).

- $U_{(n \times n)}$: une matrice d'adjacence où toutes les observations sont voisines (coïncidence entre indépendance totale et dépendance totale des observations)
- $U_{(n \times n)}$ introduit un centrage par rapport à la moyenne globale

Introduction : analyse de la dépendance (2)

- **Données dépendantes**

$M_{(n,n)}(m_{ii'})$ une matrice d'adjacence avec $m_{ii'} = 1$ si i et i' contigus, 0 sinon.

La matrice de variance/covariance locale $V_M(X)$:

(Moran 1948, Geary 1954, Cliff et Ord 1973, Lebart 1969)

$$V_M(X) = X^t(I - M')^t P(I - M')X$$

- M' (M normée en ligne) introduit un centrage par rapport à la **moyenne de voisinage**

$$v_{jl} = \frac{1}{2m} \sum_{i,i'} m_{ii'} (x_{ij} - x_{i'j})(x_{il} - x_{i'l})$$

Introduction : analyse de la dépendance (3)

- **Approches classiques d'analyse de la dépendance (analyse globale/locale)**

(Banet & Lebart 1984, Wartenberg 1985, Mom 1988, Thioulouse et al. 1995)

- Structure de contiguïté d'origine spatiale, géographique..., rarement temporelle
- Structure connue *a priori*

- **Nos travaux**

- Étudier les concepts classiques dans un contexte de dépendance temporelle et mise en évidence de propriétés utiles à l'analyse des séries
- Explorer le cadre de structures de dépendance non connues mais apprises

Introduction : analyse de la dépendance (3)

- **Approches classiques d'analyse de la dépendance (analyse globale/locale)**

(Banet & Lebart 1984, Wartenberg 1985, Mom 1988, Thioulouse et al. 1995)

- Structure de contiguïté d'origine spatiale, géographique..., rarement temporelle
- Structure connue *a priori*

- **Nos travaux**

- Étudier les concepts classiques dans un contexte de dépendance temporelle et mise en évidence de propriétés utiles à l'analyse des séries
- Explorer le cadre de structures de dépendance non connues mais apprises

Introduction : analyse de la dépendance (3)

- **Approches classiques d'analyse de la dépendance (analyse globale/locale)**

(Banet & Lebart 1984, Wartenberg 1985, Mom 1988, Thioulouse et al. 1995)

- Structure de contiguïté d'origine spatiale, géographique..., rarement temporelle
- Structure connue *a priori*

- **Nos travaux**

- Étudier les concepts classiques dans un contexte de dépendance temporelle et mise en évidence de propriétés utiles à l'analyse des séries
- Explorer le cadre de structures de dépendance non connues mais apprises

Introduction : analyse de la dépendance (3)

- **Approches classiques d'analyse de la dépendance (analyse globale/locale)**

(Banet & Lebart 1984, Wartenberg 1985, Mom 1988, Thioulouse et al. 1995)

- Structure de contiguïté d'origine spatiale, géographique..., rarement temporelle
- Structure connue *a priori*

- **Nos travaux**

- Étudier les concepts classiques dans un contexte de dépendance temporelle et mise en évidence de propriétés utiles à l'analyse des séries
- Explorer le cadre de structures de dépendance non connues mais apprises

Première partie

Représentation compacte de séries temporelles
par **segmentation**

Problème et objectif

- **Problème & objectif**

- Analyser de longues séries multivariées (Nb. d'obs./séries $\approx 10^3$)
- Nombre d'observations constitue une limite à l'analyse
- **Objectif** : Réduction du nombre d'observations des séries multivariées par segmentation

- **Limites des approches existantes**

- Dédiées à la segmentation de séries univariées
- Critères de segmentation non fondés sur la dépendances des variables

- **Contribution**

- Segmentation de séries multivariées préservant les corrélations des variables, information fondamentale en analyse des données

Problème et objectif

- **Problème & objectif**
 - Analyser de longues séries multivariées (Nb. d'obs./séries $\approx 10^3$)
 - Nombre d'observations constitue une limite à l'analyse
 - **Objectif** : Réduction du nombre d'observations des séries multivariées par **segmentation**
- **Limites des approches existantes**
 - Dédiées à la segmentation de séries univariées
 - Critères de segmentation non fondés sur la dépendances des variables
- **Contribution**
 - Segmentation de séries multivariées préservant les corrélations des variables, information fondamentale en analyse des données

Problème et objectif

- **Problème & objectif**

- Analyser de longues séries multivariées (Nb. d'obs./séries $\approx 10^3$)
- Nombre d'observations constitue une limite à l'analyse
- **Objectif** : Réduction du nombre d'observations des séries multivariées par segmentation

- **Limites des approches existantes**

- Dédiées à la segmentation de séries univariées
- Critères de segmentation non fondés sur la dépendances des variables

- **Contribution**

- Segmentation de séries multivariées préservant les corrélations des variables, information fondamentale en analyse des données

Problème et objectif

- **Problème & objectif**

- Analyser de longues séries multivariées (Nb. d'obs./séries $\approx 10^3$)
- Nombre d'observations constitue une limite à l'analyse
- **Objectif** : Réduction du nombre d'observations des séries multivariées par segmentation

- **Limites des approches existantes**

- Dédiées à la segmentation de séries univariées
- Critères de segmentation non fondés sur la dépendances des variables

- **Contribution**

- Segmentation de séries multivariées préservant les corrélations des variables, information fondamentale en analyse des données

La segmentation : état de l'art

Problème de segmentation

- Recherche de u_1, \dots, u_k instants (k connu ou estimé) visant à optimiser un des deux critères :

Des points de changements significatifs des caractéristiques de distributions (gaussiennes)

- Estimation du rapport de vraisemblance (Basevilles 1993, Gustafsson 2000, Guralnik 1999), comparaison des densités de probabilités estimées (Brodsky 1993, Csörgö 1988), mesure Kulback empirique (Kawahara et al. 2009)

Des points minimisant les résidus de modèles a priori estimés sur les segments (linéaires)

- Localisation spatiale de *hot spot* Fused-Lasso (Tibshirani 2008), détection de multiples points de changement d'une ou de plusieurs séries (Harchaoui et al. 2010, Vert et al. 2010)

La segmentation : état de l'art

Problème de segmentation

- Recherche de u_1, \dots, u_k instants (k connu ou estimé) visant à optimiser un des deux critères :

Des points de changements significatifs des caractéristiques de distributions (gaussiennes)

- Estimation du rapport de vraisemblance (Basevilles 1993, Gustafsson 2000, Guralnik 1999), comparaison des densités de probabilités estimées (Brodsky 1993, Csörgö 1988), mesure Kulback empirique (Kawahara et al. 2009)

Des points minimisant les résidus de modèles a priori estimés sur les segments (linéaires)

- Localisation spatiale de *hot spot* Fused-Lasso (Tibshirani 2008), détection de multiples points de changement d'une ou de plusieurs séries (Harchaoui et al. 2010, Vert et al. 2010)

La segmentation : état de l'art

Problème de segmentation

- Recherche de u_1, \dots, u_k instants (k connu ou estimé) visant à optimiser un des deux critères :

Des points de changements significatifs des caractéristiques de distributions (gaussiennes)

- Estimation du rapport de vraisemblance (Basevilles 1993, Gustafsson 2000, Guralnik 1999), comparaison des densités de probabilités estimées (Brodsky 1993, Csörgö 1988), mesure Kulback empirique (Kawahara et al. 2009)

Des points minimisant les résidus de modèles a priori estimés sur les segments (linéaires)

- Localisation spatiale de *hot spot* Fused-Lasso (Tibshirani 2008), détection de multiples points de changement d'une ou de plusieurs séries (Harchaoui et al. 2010, Vert et al. 2010)

Caractéristiques des critères classiques

- Recherche de points changements de caractéristiques (moyenne, variance), ne rend pas compte de la dépendance temporelle des observations,
- Prise en compte de la dépendance des observations fondée sur des modèles a priori,
- Approches souvent limitées à des séries univariées, nombre de segments connu (estimé)

Contribution

- Un nouveau critère de segmentation de séries multivariées tenant compte de la dépendance temporelle des observations (variance temporelle)
- Approche locale (fenêtre glissante) pour l'extraction de segments, de nombre inconnu, de variances temporelles équivalentes
- Une segmentation guidée par la préservation des corrélations entre les variables

Caractéristiques des critères classiques

- Recherche de points changements de caractéristiques (moyenne, variance), ne rend pas compte de la dépendance temporelle des observations,
- Prise en compte de la dépendance des observations fondée sur des modèles a priori,
- Approches souvent limitées à des séries univariées, nombre de segments connu (estimé)

Contribution

- Un nouveau critère de segmentation de séries multivariées tenant compte de la **dépendance temporelle** des observations (variance temporelle)
- Approche locale (fenêtre glissante) pour l'extraction de segments, **de nombre inconnu**, de **variances temporelles** équivalentes
- Une segmentation guidée par **la préservation des corrélations** entre les variables

Variance/Covariance : dépendance temporelle

S une série multivariée portant sur n observations

La variance/covariance temporelle $V_T(S)$ (ordre r) : $V_T(S) = V_M(S)$

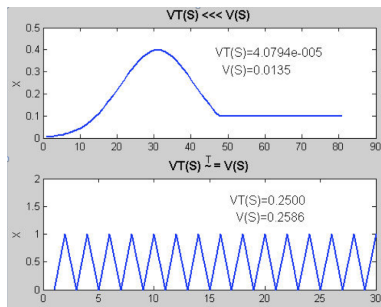
$M_{(n,n)}(m_{ii'})$: $m_{ii'} = 1$ si $|i - i'| \leq r$, 0 sinon

Variance/Covariance : dépendance temporelle

S une série multivariée portant sur n observations

La variance/covariance temporelle $V_T(S)$ (ordre r) : $V_T(S) = V_M(S)$

$M_{(n,n)}(m_{ii'})$: $m_{ii'} = 1$ si $|i - i'| \leq r$, 0 sinon



Définitions et propriétés

S : une série multivariée à T instants, s_k un segment de S à n_k obs,
 s_i^j : le segment couvrant la période $[i, j]$, $1 \leq i < j \leq T$

Inertie de S

$$I(S) = \text{Tr}(V_T(S)) ; (\|V_T(S)\|_F)$$

Contribution de s_k à l'inertie de S

$$C_S(s_k) = \frac{n_k - 1}{T - 1} \text{Tr}(V_T(s_k)) = \frac{n_k - 1}{T - 1} I(s_k)$$

C_S : additive, monotone croissante

Pour toute segmentation de S en K segments s_1, \dots, s_K

$$I(S) = C_S(S) = \sum_{k=1}^K C_S(s_k)$$

$$\forall j < l \quad C_S(s_i^j) \leq C_S(s_i^l)$$

Définitions et propriétés

S : une série multivariée à T instants, s_k un segment de S à n_k obs,
 s_i^j : le segment couvrant la période $[i, j]$, $1 \leq i < j \leq T$

Inertie de S

$$I(S) = \text{Tr}(V_T(S)) ; (\|V_T(S)\|_F)$$

Contribution de s_k à l'inertie de S

$$C_S(s_k) = \frac{n_k - 1}{T - 1} \text{Tr}(V_T(s_k)) = \frac{n_k - 1}{T - 1} I(s_k)$$

C_S : additive, monotone croissante

Pour toute segmentation de S en K segments s_1, \dots, s_K

$$I(S) = C_S(S) = \sum_{k=1}^K C_S(s_k)$$
$$\forall j < l \quad C_S(s_i^j) \leq C_S(s_i^l)$$

Définitions et propriétés

S : une série multivariée à T instants, s_k un segment de S à n_k obs,
 s_i^j : le segment couvrant la période $[i, j]$, $1 \leq i < j \leq T$

Inertie de S

$$I(S) = \text{Tr}(V_T(S)) ; (\|V_T(S)\|_F)$$

Contribution de s_k à l'inertie de S

$$C_S(s_k) = \frac{n_k - 1}{T - 1} \text{Tr}(V_T(s_k)) = \frac{n_k - 1}{T - 1} I(s_k)$$

C_S : additive, monotone croissante

Pour toute segmentation de S en K segments s_1, \dots, s_K

$$I(S) = C_S(S) = \sum_{k=1}^K C_S(s_k)$$
$$\forall j < l \quad C_S(s_i^j) \leq C_S(s_i^l)$$

Définitions et propriétés

S : une série multivariée à T instants, s_k un segment de S à n_k obs,
 s_i^j : le segment couvrant la période $[i, j]$, $1 \leq i < j \leq T$

Inertie de S

$$I(S) = \text{Tr}(V_T(S)) ; (\|V_T(S)\|_F)$$

Contribution de s_k à l'inertie de S

$$C_S(s_k) = \frac{n_k - 1}{T - 1} \text{Tr}(V_T(s_k)) = \frac{n_k - 1}{T - 1} I(s_k)$$

C_S : additive, monotone croissante

Pour toute segmentation de S en K segments s_1, \dots, s_K

$$I(S) = C_S(S) = \sum_{k=1}^K C_S(s_k)$$
$$\forall j < l \quad C_S(s_i^j) \leq C_S(s_i^l)$$

Définitions et propriétés

S : une série multivariée à T instants, s_k un segment de S à n_k obs,
 s_i^j : le segment couvrant la période $[i, j]$, $1 \leq i < j \leq T$

Inertie de S

$$I(S) = \text{Tr}(V_T(S)) ; (\|V_T(S)\|_F)$$

Contribution de s_k à l'inertie de S

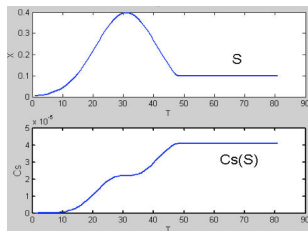
$$C_S(s_k) = \frac{n_k - 1}{T - 1} \text{Tr}(V_T(s_k)) = \frac{n_k - 1}{T - 1} I(s_k)$$

C_S : additive, monotone croissante

Pour toute segmentation de S en K segments s_1, \dots, s_K

$$I(S) = C_S(S) = \sum_{k=1}^K C_S(s_k)$$

$$\forall j < l \quad C_S(s_i^j) \leq C_S(s_i^l)$$



Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{C_S} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{C_S} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

$O(np^2)$

4. **Estimation de la perte de corrélation**

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{C_S}**

- Augmenter α_{C_S} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{CS} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{CS} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

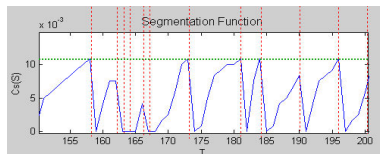
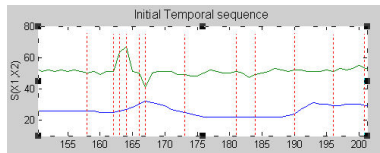
4. **Estimation de la perte de corrélation**

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{CS}**

- Augmenter α_{CS} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

$O(np^2)$



Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{CS} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{CS} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

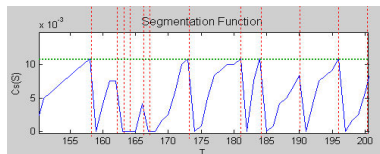
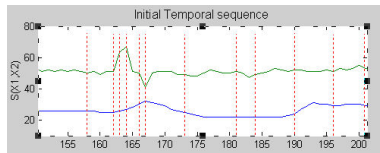
4. Estimation de la perte de corrélation

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{CS}**

- Augmenter α_{CS} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

$O(np^2)$



Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{CS} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{CS} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

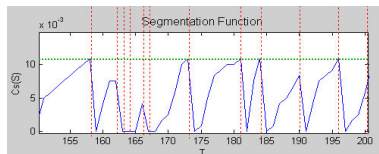
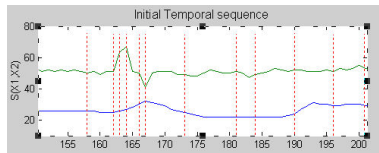
4. **Estimation de la perte de corrélation**

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{CS}**

- Augmenter α_{CS} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

$O(np^2)$



Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{CS} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{CS} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

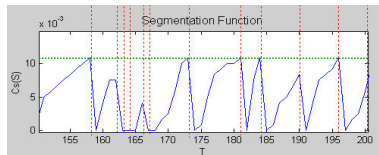
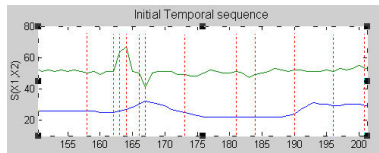
4. **Estimation de la perte de corrélation**

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{CS}**

- Augmenter α_{CS} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

$O(np^2)$



Segmentation pour la réduction du nombre d'observations

1. **Initialisation** : S , $\alpha \in [0, 1]$ perte de corrélations,
 $\alpha_{CS} \in [0, 1]$ seuil de contribution à $I(S)$

2. **Segmentation** : Extraction de $s_1^{j_1}, s_{j_1+1}^{j_2} \dots$

$$C_S(s_{j_{k-1}+1}^{j_k}) \leq \alpha_{CS} I(S) \leq C_S(s_{j_{k-1}+1}^{j_{k+1}})$$

3. **Regression linéaire multiple** :

- Régression sur chaque segment extrait
- Décrire chaque segment par deux valeurs estimées aux bornes (S' la série réduite)

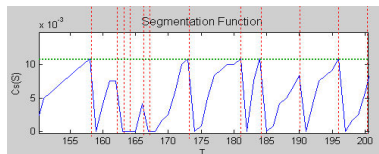
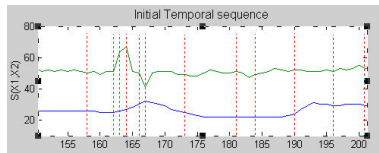
4. **Estimation de la perte de corrélation**

$$e(S, S') = \frac{\|Cor(S) - Cor(S')\|_F}{\|Cor(S)\|_F}$$

5. **Ajustement de α_{CS}**

- Augmenter α_{CS} si $e(S, S') < \alpha$, le diminuer sinon
- Répéter les étapes 2 à 4 jusqu'à la réduction maximale de S préservant les corrélations au seuil $1 - \alpha$

$O(np^2)$

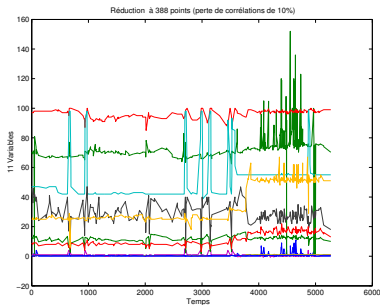
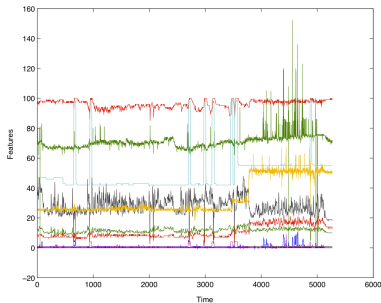


Resultats : données de monitoring en anesthésie-réanimation

PRETA (TIMC-IMAG)

$p = 11, n \approx 5000,$
Nb. itérations ≈ 5

- Réduction de la série de 5000 à 388 obs. (taux réduction 92%, perte des corrélations 10%)



Deuxième partie

Apprentissage d'appariements discriminants

Problème et objectif

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes
- **Objectif** : Classification fondée sur les caractéristiques dynamiques discriminantes

- **Limites des approches existantes**

- Hypothèse de dynamiques globales similaires dans les classes

- **Contribution**

- Apprentissage d'appariements discriminants (révélant les caractéristiques communes dans les classes et différentielles entre les classes)
- Dériver une métrique pour la classification de séries

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes
- **Objectif** : Classification fondée sur les caractéristiques dynamiques discriminantes

- **Limites des approches existantes**

- Hypothèse de dynamiques globales similaires dans les classes

- **Contribution**

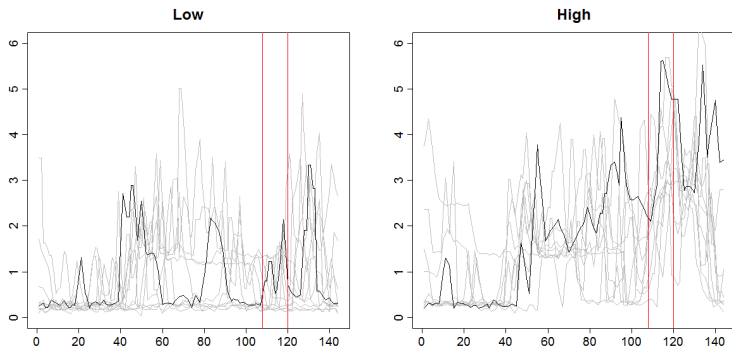
- Apprentissage d'appariements discriminants (révélant les caractéristiques communes dans les classes et différentielles entre les classes)
- Dériver une métrique pour la classification de séries

Problème et objectif

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes

Courbes de consommations électriques journalières



- **Objectif** : Classification fondée sur les caractéristiques dynamiques discriminantes

Problème et objectif

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes
- **Objectif** : Classification fondée sur les caractéristiques dynamiques **discriminantes**

- **Limites des approches existantes**

- Hypothèse de dynamiques globales similaires dans les classes

- **Contribution**

- Apprentissage d'appariements discriminants (révélant les caractéristiques communes dans les classes et différentielles entre les classes)
- Dériver une métrique pour la classification de séries

Problème et objectif

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes
- **Objectif** : Classification fondée sur les caractéristiques dynamiques **discriminantes**

- **Limites des approches existantes**

- Hypothèse de dynamiques globales similaires dans les classes

- **Contribution**

- Apprentissage d'appariements discriminants (révélant les caractéristiques communes dans les classes et différentielles entre les classes)
- Dériver une métrique pour la classification de séries

Problème et objectif

- **Problème & objectif**

- Classification de séries multivariées complexes
- Différentes dynamiques au sein des classes, faibles différences entre les classes
- **Objectif** : Classification fondée sur les caractéristiques dynamiques **discriminantes**

- **Limites des approches existantes**

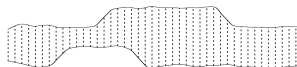
- Hypothèse de dynamiques globales similaires dans les classes

- **Contribution**

- Apprentissage **d'appariements discriminants** (révélant les caractéristiques communes dans les classes et différentielles entre les classes)
- Dériver une métrique pour la classification de séries

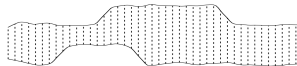
Stratégies classiques d'alignement des séries

- Alignement des observations de mêmes instants (alignement euclidien)
- Alignement tenant compte de délais, Dynamic Time Warping DTW (Sankoff & Kruskal 1983, Kruskal & Liberman 1983).
- Alignement sous contraintes globales/locales (Sakoe-Chiba 1975, Itakura 1971, Rabiner 1978)



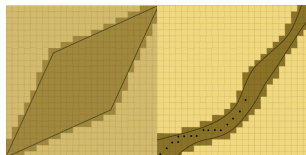
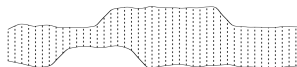
Stratégies classiques d'alignement des séries

- Alignement des observations de mêmes instants (alignement euclidien)
- Alignement tenant compte de délais, Dynamic Time Warping DTW (Sankoff & Kruskal 1983, Kruskal & Liberman 1983).
- Alignement sous contraintes globales/locales (Sakoe-Chiba 1975, Itakura 1971, Rabiner 1978)



Stratégies classiques d'alignement des séries

- Alignement des observations de mêmes instants (alignement euclidien)
- Alignement tenant compte de délais, Dynamic Time Warping DTW (Sankoff & Kruskal 1983, Kruskal & Liberman 1983).
- Alignement sous contraintes globales/locales (Sakoe-Chiba 1975, Itakura 1971, Rabiner 1978)



Alignements classiques

Limite : hypothèse de dynamiques globales similaires dans les classes

- Approches pair-à-pair
 - Limitées au couple de séries à aligner
 - Ne rend pas compte des dynamiques de l'ensemble des séries
 - Dissociées des processus d'analyse des séries
- Contraintes temporelles rigides
 - Croisements non permis

Propositions plus récentes

- Apprentissage de contraintes d'alignement de la DTW en vue d'un k-NN à vase marge
[Yu et al. 2011](#), [Jeong et al. 2011](#)
- Modèle probabiliste assurant conjointement les processus de classification non supervisée des séries et leur alignements [Ramsay et al. 1998](#), [Gaffney et al. 2005](#)
- Modèle bayésien hiérarchique pour l'identification de faibles différences entre classes
[Listgarten et al. 2007](#)

Alignements classiques

Limite : hypothèse de dynamiques globales similaires dans les classes

- Approches pair-à-pair
 - Limitées au couple de séries à aligner
 - Ne rend pas compte des dynamiques de l'ensemble des séries
 - Dissociées des processus d'analyse des séries
- Contraintes temporelles rigides
 - Croisements non permis

Propositions plus récentes

- Apprentissage de contraintes d'alignement de la DTW en vue d'un k-NN à vase marge
[Yu et al. 2011](#), [Jeong et al. 2011](#)
- Modèle probabiliste assurant conjointement les processus de classification non supervisée des séries et leur alignements [Ramsay et al. 1998](#), [Gaffney et al. 2005](#)
- Modèle bayésien hiérarchique pour l'identification de faibles différences entre classes
[Listgarten et al. 2007](#)

Alignements classiques

Limite : hypothèse de dynamiques globales similaires dans les classes

- Approches pair-à-pair
 - Limitées au couple de séries à aligner
 - Ne rend pas compte des dynamiques de l'ensemble des séries
 - Dissociées des processus d'analyse des séries
- Contraintes temporelles rigides
 - Croisements non permis

Propositions plus récentes

- Apprentissage de contraintes d'alignement de la DTW en vue d'un k-NN à vase marge
[Yu et al. 2011](#), [Jeong et al. 2011](#)
- Modèle probabiliste assurant conjointement les processus de classification non supervisée des séries et leur alignements [Ramsay et al. 1998](#), [Gaffney et al. 2005](#)
- Modèle bayésien hiérarchique pour l'identification de faibles différences entre classes
[Listgarten et al. 2007](#)

Discrimination de séries temporelles

Objectif

- Discrimination de classes de séries complexes (dynamiques différentes au sein des classes, faibles différences entre les classes)

Approche

- Élargir la stratégie d'alignement à un appariement moins contraints temporellement
- Appariement multiples : mettre en regard l'ensemble des dynamiques des séries
- Révéler des caractéristiques communes dans les classes et différentielles entre les classes

Discrimination de séries temporelles

Objectif

- Discrimination de classes de séries complexes (dynamiques différentes au sein des classes, faibles différences entre les classes)

Approche

- Élargir la stratégie d'alignement à un appariement moins contraints temporellement
- Appariement multiples : mettre en regard l'ensemble des dynamiques des séries
- Révéler des caractéristiques communes dans les classes et différentielles entre les classes

Discrimination de séries temporelles

Objectif

- Discrimination de classes de séries complexes (dynamiques différentes au sein des classes, faibles différences entre les classes)

Approche

- Élargir la stratégie d'alignement à un appariement moins contraints temporellement
- Appariement multiples : mettre en regard l'ensemble des dynamiques des séries
- Révéler des caractéristiques communes dans les classes et différentielles entre les classes

Discrimination de séries temporelles

Objectif

- Discrimination de classes de séries complexes (dynamiques différentes au sein des classes, faibles différences entre les classes)

Approche

- Élargir la stratégie d'alignement à un appariement moins contraints temporellement
- Appariement multiples : mettre en regard l'ensemble des dynamiques des séries
- Révéler des caractéristiques communes dans les classes et différentielles entre les classes

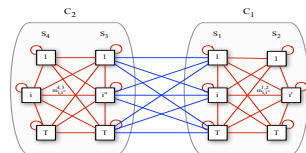
Contribution

Idée

- A tout appariement entre des séries correspond une variabilité intra et inter classe définie par les observations liées à l'intérieur des classes et entre les classes
- Révéler les caractéristiques communes au sein des classes par la minimisation de la variance intra classe
- Révéler les attributs différentiels par la maximisation de la variance inter classe

Comment ?

- Étendre les concepts de variance/covariance à un ensemble, ainsi qu'à une partition de séries
- Pénaliser (renforcer) les liens en fonction de leur contribution à la variance intra et inter classes



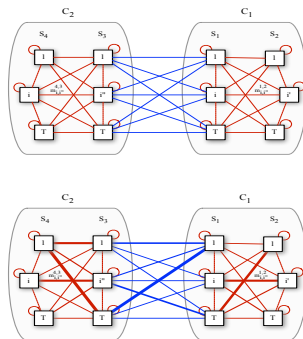
Contribution

Idée

- A tout appariement entre des séries correspond une variabilité intra et inter classe définie par les observations liées à l'intérieur des classes et entre les classes
- Révéler les caractéristiques communes au sein des classes par la minimisation de la variance intra classe
- Révéler les attributs différentiels par la maximisation de la variance inter classe

Comment ?

- Étendre les concepts de variance/covariance à un ensemble, ainsi qu'à une partition de séries
- Pénaliser (renforcer) les liens en fonction de leur contribution à la variance intra et inter classes



Variance/Covariance induite par des séries temporelles

S_1, \dots, S_n séries multivariées, S_l décrit l'observation de p variables à T instants

$X_{n, T \times p}(x_{ij}^l)$: matrice de description des séries S_1, \dots, S_n par p variables

x_{ij}^l : la valeur de la variable j prise à l'instant i de S_l

Appariement a priori des séries

- Pour tout appariement a priori des séries S_1, \dots, S_n (exp. DTW)
- On associe $M_{(n,n)}(M_{ll'})$ une matrice d'adjacence par blocs $M_{ll'}(m_{ii}^{ll'})$ liant les observations i de S_l et i' de $S_{l'}$
- $m_{ii}^{ll'} = 1$ si i et i' entre S_l et $S_{l'}$ sont alignées, 0 sinon

Variance/covariance totale associée à l'ensemble des séries

$$V_M(X) = X^t(I - M')^t P(I - M')X$$

Variance/covariance intra et inter classes de séries

$$V_{M_{Intra}}(X) = X^t(I - M_{Intra}')^t P(I - M_{Intra}')X$$

$$V_{M_{Inter}}(X) = X^t(I - M_{Inter}')^t P(I - M_{Inter}')X$$

- M_{Intra} : $m_{ii}^{ll'} = 1$ si i et i' sont des observations de séries de même classe, 0 sinon.
- M_{Inter} : $m_{ii}^{ll'} = 1$ si i et i' sont des observations de séries de classes différentes, 0 sinon.

Variance/Covariance induite par des séries temporelles

S_1, \dots, S_n séries multivariées, S_I décrit l'observation de p variables à T instants

$X_{n, T \times p}(x_{ij}^I)$: matrice de description des séries S_1, \dots, S_n par p variables

x_{ij}^I : la valeur de la variable j prise à l'instant i de S_I

Appariement a priori des séries

- Pour tout appariement a priori des séries S_1, \dots, S_n (exp. DTW)
- On associe $M_{(n,n)}(M_{II'})$ une matrice d'adjacence par blocs $M_{II'}(m_{ii'}^{II'})$ liant les observations i de S_I et i' de $S_{I'}$
- $m_{ii'}^{II'} = 1$ si i et i' entre S_I et $S_{I'}$ sont alignées, 0 sinon

Variance/covariance totale associée à l'ensemble des séries

$$V_M(X) = X^t(I - M')^t P(I - M')X$$

Variance/covariance intra et inter classes de séries

$$V_{M_{Intra}}(X) = X^t(I - M_{Intra})^t P(I - M_{Intra})X$$

$$V_{M_{Inter}}(X) = X^t(I - M_{Inter})^t P(I - M_{Inter})X$$

- M_{Intra} : $m_{ii'}^{II'} = 1$ si i et i' sont des observations de séries de même classe, 0 sinon.
- M_{Inter} : $m_{ii'}^{II'} = 1$ si i et i' sont des observations de séries de classes différentes, 0 sinon.

Variance/Covariance induite par des séries temporelles

S_1, \dots, S_n séries multivariées, S_I décrit l'observation de p variables à T instants

$X_{n, T \times p}(x_{ij}^I)$: matrice de description des séries S_1, \dots, S_n par p variables

x_{ij}^I : la valeur de la variable j prise à l'instant i de S_I

Appariement a priori des séries

- Pour tout appariement a priori des séries S_1, \dots, S_n (exp. DTW)
- On associe $M_{(n,n)}(M_{II'})$ une matrice d'adjacence par blocs $M_{II'}(m_{ii'}^{II'})$ liant les observations i de S_I et i' de $S_{I'}$
- $m_{ii'}^{II'} = 1$ si i et i' entre S_I et $S_{I'}$ sont alignées, 0 sinon

Variance/covariance totale associée à l'ensemble des séries

$$V_M(X) = X^t(I - M')^t P(I - M')X$$

Variance/covariance intra et inter classes de séries

$$V_{M_{Intra}}(X) = X^t(I - M_{Intra}')^t P(I - M_{Intra}')X$$

$$V_{M_{Inter}}(X) = X^t(I - M_{Inter}')^t P(I - M_{Inter}')X$$

- M_{Intra} : $m_{ii'}^{II'} = 1$ si i et i' sont des observations de séries de même classe, 0 sinon.
- M_{Inter} : $m_{ii'}^{II'} = 1$ si i et i' sont des observations de séries de classes différentes, 0 sinon.

Variance/Covariance induite par des séries temporelles

S_1, \dots, S_n séries multivariées, S_l décrit l'observation de p variables à T instants
 $X_{n, T \times p}(x_{ij}^l)$: matrice de description des séries S_1, \dots, S_n par p variables
 x_{ij}^l : la valeur de la variable j prise à l'instant i de S_l

Appariement a priori des séries

- Pour tout appariement a priori des séries S_1, \dots, S_n (exp. DTW)
- On associe $M_{(n,n)}(M_{ll'})$ une matrice d'adjacence par blocs $M_{ll'}(m_{ii}^{ll'})$ liant les observations i de S_l et i' de $S_{l'}$
- $m_{ii}^{ll'} = 1$ si i et i' entre S_l et $S_{l'}$ sont alignées, 0 sinon

Variance/covariance totale associée à l'ensemble des séries

$$V_M(X) = X^t(I - M')^t P(I - M')X$$

Variance/covariance intra et inter classes de séries

$$V_{M_{Intra}}(X) = X^t(I - M_{Intra})^t P(I - M_{Intra})X$$

$$V_{M_{Inter}}(X) = X^t(I - M_{Inter})^t P(I - M_{Inter})X$$

- M_{Intra} : $m_{ii}^{ll'} = 1$ si i et i' sont des observations de séries de même classe, 0 sinon.
- M_{Inter} : $m_{ii}^{ll'} = 1$ si i et i' sont des observations de séries de classes différentes, 0 sinon.

Objectif

Rechercher les appariements M_{Intra} et M_{Inter} par :
minimisation de $V_{M_{Intra}}$ et maximisation $V_{M_{Inter}}$

$$M_{Intra}^{l,l'} = \begin{cases} \mathbf{I} & \text{si } l = l' \\ \neq \mathbf{0} & \text{si } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{si } y_l \neq y_{l'} \end{cases} \quad M_{Inter}^{l,l'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases}$$

\mathbf{I} et $\mathbf{0}$ les matrices identités et nulles
 y_l la classe de S_l

Deux phases consécutives :

1. Apprentissage des appariements intra classes (M_{Intra}^*)
 - Initialisation M_{Intra}^0 non contrainte
2. Apprentissage des appariements inter classes (M^*)
 - Initialisation de M_{Inter}^0 à partir de M_{Intra}^*
(affiner M_{Intra}^* pour révéler les liens à la fois communs et différentiels)

Objectif

Rechercher les appariements M_{Intra} et M_{Inter} par :
minimisation de $V_{M_{Intra}}$ et maximisation $V_{M_{Inter}}$

$$M_{Intra}^{l,l'} = \begin{cases} \mathbf{I} & \text{si } l = l' \\ \neq \mathbf{0} & \text{si } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{si } y_l \neq y_{l'} \end{cases} \quad M_{Inter}^{l,l'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases}$$

\mathbf{I} et $\mathbf{0}$ les matrices identités et nulles
 y_l la classe de S_l

Deux phases consécutives :

1. Apprentissage des appariements intra classes (M_{Intra}^*)
 - Initialisation M_{Intra}^0 non contrainte
2. Apprentissage des appariements inter classes (M^*)
 - Initialisation de M_{Inter}^0 à partir de M_{Intra}^*
(affiner M_{Intra}^* pour révéler les liens à la fois communs et différentiels)

Objectif

Rechercher les appariements M_{Intra} et M_{Inter} par :
minimisation de $V_{M_{Intra}}$ et maximisation $V_{M_{Inter}}$

$$M_{Intra}^{l,l'} = \begin{cases} \mathbf{I} & \text{si } l = l' \\ \neq \mathbf{0} & \text{si } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{si } y_l \neq y_{l'} \end{cases} \quad M_{Inter}^{l,l'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases}$$

\mathbf{I} et $\mathbf{0}$ les matrices identités et nulles
 y_l la classe de S_l

Deux phases consécutives :

1. Apprentissage des appariements intra classes (M_{Intra}^*)
 - Initialisation M_{Intra}^0 non contrainte
2. Apprentissage des appariements inter classes (M^*)
 - Initialisation de M_{Inter}^0 à partir de M_{Intra}^*
(affiner M_{Intra}^* pour révéler les liens à la fois communs et différentiels)

Objectif

Rechercher les appariements M_{Intra} et M_{Inter} par :
minimisation de $V_{M_{Intra}}$ et maximisation $V_{M_{Inter}}$

$$M_{Intra}^{l,l'} = \begin{cases} \mathbf{I} & \text{si } l = l' \\ \neq \mathbf{0} & \text{si } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{si } y_l \neq y_{l'} \end{cases} \quad M_{Inter}^{l,l'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases}$$

\mathbf{I} et $\mathbf{0}$ les matrices identités et nulles
 y_l la classe de S_l

Deux phases consécutives :

1. Apprentissage des appariements intra classes (M_{Intra}^*)
 - Initialisation M_{Intra}^0 non contrainte
2. Apprentissage des appariements inter classes (M^*)
 - Initialisation de M_{Inter}^0 à partir de M_{Intra}^*
(affiner M_{Intra}^* pour révéler les liens à la fois communs et différentiels)

Objectif

Rechercher les appariements M_{Intra} et M_{Inter} par :
minimisation de $V_{M_{Intra}}$ et maximisation $V_{M_{Inter}}$

$$M_{Intra}^{l,l'} = \begin{cases} \mathbf{I} & \text{si } l = l' \\ \neq \mathbf{0} & \text{si } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{si } y_l \neq y_{l'} \end{cases} \quad M_{Inter}^{l,l'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases}$$

\mathbf{I} et $\mathbf{0}$ les matrices identités et nulles
 y_l la classe de S_l

Deux phases consécutives :

1. Apprentissage des appariements intra classes (M_{Intra}^*)
 - Initialisation M_{Intra}^0 non contrainte
2. Apprentissage des appariements inter classes (M^*)
 - Initialisation de M_{Inter}^0 à partir de M_{Intra}^*
(affiner M_{Intra}^* pour révéler les liens à la fois communs et différentiels)

Apprentissage des appariements intra classes

$S_{I_1} = (x_1^{I_1}, \dots, x_T^{I_1})$, $S_{I_2} = (x_1^{I_2}, \dots, x_T^{I_2})$, de même classe C_k ($|C_k| = n_k$)

$M \setminus (i, i', l, l')$: M après suppression du lien (i, i') entre S_l et $S_{l'}$ ($m_{ii'}^{ll'} = 0$)

Principe de l'algorithme

- 1 Initialiser M_{Intra} par un couplage complet :

$$\forall i, i' \in \{1, \dots, T\} \text{ et } S_l, S_{l'} \text{ de même classe } m_{ii'}^{ll'} = 1$$

- 2 Mesurer la contribution $C_{i_1 i_2}^{I_1 I_2}$ de chaque lien i_1, i_2 entre S_{I_1} et S_{I_2} à la variance $V_{M_{Intra}}$

$$C_{i_1 i_2}^{I_1 I_2} = V_{M_{Intra}} - V_{M_{Intra} \setminus (i_1, i_2, I_1, I_2)}$$

- 3 Pénaliser les liens (i_1, i_2) ($m_{i_1 i_2}^{I_1 I_2} = 0$) de contributions $C_{i_1 i_2}^{I_1 I_2} > 0$

- 4 Répéter les étapes 2 et 3 jusqu'à stabilisation de la variance intra

Apprentissage des appariements intra classes

$S_{I_1} = (x_1^{I_1}, \dots, x_T^{I_1})$, $S_{I_2} = (x_1^{I_2}, \dots, x_T^{I_2})$, de même classe C_k ($|C_k| = n_k$)

$M \setminus (i, i', l, l')$: M après suppression du lien (i, i') entre S_l et $S_{l'}$ ($m_{ii'}^{ll'} = 0$)

Principe de l'algorithme

- 1 Initialiser M_{Intra} par un couplage complet :

$$\forall i, i' \in \{1, \dots, T\} \text{ et } S_l, S_{l'} \text{ de même classe } m_{ii'}^{ll'} = 1$$

- 2 Mesurer la contribution $C_{i_1 i_2}^{I_1 I_2}$ de chaque lien i_1, i_2 entre S_{I_1} et S_{I_2} à la variance $V_{M_{Intra}}$

$$C_{i_1 i_2}^{I_1 I_2} = V_{M_{Intra}} - V_{M_{Intra} \setminus (i_1, i_2, I_1, I_2)}$$

- 3 Pénaliser les liens (i_1, i_2) ($m_{i_1 i_2}^{I_1 I_2} = 0$) de contributions $C_{i_1 i_2}^{I_1 I_2} > 0$

- 4 Répéter les étapes 2 et 3 jusqu'à stabilisation de la variance intra

Apprentissage des appariements intra classes

$S_{I_1} = (x_1^{I_1}, \dots, x_T^{I_1})$, $S_{I_2} = (x_1^{I_2}, \dots, x_T^{I_2})$, de même classe C_k ($|C_k| = n_k$)

$M \setminus (i, i', l, l')$: M après suppression du lien (i, i') entre S_l et $S_{l'}$ ($m_{ii'}^{ll'} = 0$)

Principe de l'algorithme

- 1 Initialiser M_{Intra} par un couplage complet :

$$\forall i, i' \in \{1, \dots, T\} \text{ et } S_l, S_{l'} \text{ de même classe } m_{ii'}^{ll'} = 1$$

- 2 Mesurer la contribution $C_{i_1 i_2}^{I_1 I_2}$ de chaque lien i_1, i_2 entre S_{I_1} et S_{I_2} à la variance $V_{M_{Intra}}$

$$C_{i_1 i_2}^{I_1 I_2} = V_{M_{Intra}} - V_{M_{Intra} \setminus (i_1, i_2, I_1, I_2)}$$

- 3 Pénaliser les liens (i_1, i_2) ($m_{i_1 i_2}^{I_1 I_2} = 0$) de contributions $C_{i_1 i_2}^{I_1 I_2} > 0$

- 4 Répéter les étapes 2 et 3 jusqu'à stabilisation de la variance intra

Apprentissage des appariements intra classes

$S_{I_1} = (x_1^{I_1}, \dots, x_T^{I_1})$, $S_{I_2} = (x_1^{I_2}, \dots, x_T^{I_2})$, de même classe C_k ($|C_k| = n_k$)

$M \setminus (i, i', l, l')$: M après suppression du lien (i, i') entre S_l et $S_{l'}$ ($m_{ii'}^{ll'} = 0$)

Principe de l'algorithme

- 1 Initialiser M_{Intra} par un couplage complet :

$$\forall i, i' \in \{1, \dots, T\} \text{ et } S_l, S_{l'} \text{ de même classe } m_{ii'}^{ll'} = 1$$

- 2 Mesurer la contribution $C_{i_1 i_2}^{I_1 I_2}$ de chaque lien i_1, i_2 entre S_{I_1} et S_{I_2} à la variance $V_{M_{Intra}}$

$$C_{i_1 i_2}^{I_1 I_2} = V_{M_{Intra}} - V_{M_{Intra} \setminus (i_1, i_2, I_1, I_2)}$$

- 3 Pénaliser les liens (i_1, i_2) ($m_{i_1 i_2}^{I_1 I_2} = 0$) de contributions $C_{i_1 i_2}^{I_1 I_2} > 0$

- 4 Répéter les étapes 2 et 3 jusqu'à stabilisation de la variance intra

Conditions de non dégénérescence & convergence

$$\forall k \in \{1, \dots, K\}, \forall (l, l') \in C_k, \forall (i, i') \in [1, T]^2$$

Contraintes liées à la définition de la variance

- 1- $m_{ii}^{ll} > 0$
- 2- M_{Intra} normée en ligne : $\sum_{i'=1}^{n_k} \sum_{i'=1}^T m_{ii'}^{ll'} = 1$

Contrainte de non dégénérescence de la variance

- 3- Toute obs. i de S_l doit être connectée à au moins une obs. de $S_{l'}$: $\sum_{i'=1}^T m_{ii'}^{ll'} > 0$

Contrainte de convergence de la minimisation de la variance

- 4- La suppression d'un lien (i, i') influence le poids des voisins de i et i' (ligne i et i') :
suppression par itération d'au plus un lien par ligne (max des contributions $C_{i_1 i_2}^{l_1 l_2} > 0$)

Conditions de non dégénérescence & convergence

$$\forall k \in \{1, \dots, K\}, \forall (l, l') \in C_k, \forall (i, i') \in [1, T]^2$$

Contraintes liées à la définition de la variance

- 1- $m_{ii}^{ll} > 0$
- 2- M_{Intra} normée en ligne : $\sum_{i'=1}^{n_k} \sum_{i'=1}^T m_{ii'}^{ll'} = 1$

Contrainte de non dégénérescence de la variance

- 3- Toute obs. i de S_l doit être connectée à au moins une obs. de $S_{l'}$: $\sum_{i'=1}^T m_{ii'}^{ll'} > 0$

Contrainte de convergence de la minimisation de la variance

- 4- La suppression d'un lien (i, i') influence le poids des voisins de i et i' (ligne i et i') :
suppression par itération d'au plus un lien par ligne (max des contributions $C_{i_1 i_2}^{l_1 l_2} > 0$)

Conditions de non dégénérescence & convergence

$$\forall k \in \{1, \dots, K\}, \forall (l, l') \in C_k, \forall (i, i') \in [1, T]^2$$

Contraintes liées à la définition de la variance

- 1- $m_{ii}^{ll} > 0$
- 2- M_{Intra} normée en ligne : $\sum_{i'=1}^{n_k} \sum_{i''=1}^T m_{i''}^{ll'} = 1$

Contrainte de non dégénérescence de la variance

- 3- Toute obs. i de S_l doit être connectée à au moins une obs. de $S_{l'}$: $\sum_{i''=1}^T m_{i''}^{ll'} > 0$

Contrainte de convergence de la minimisation de la variance

- 4- La suppression d'un lien (i, i') influence le poids des voisins de i et i' (ligne i et i') : suppression par itération d'au plus un lien par ligne (max des contributions $C_{i_1 i_2}^{l_1 l_2} > 0$)

- Par classe et par itération :

$$O\left(\sum_{k=1}^K n_k^3 T^3\right)$$

- Parallélisation de l'apprentissage de M_{Intra} et M_{Inter} par classe et par ligne

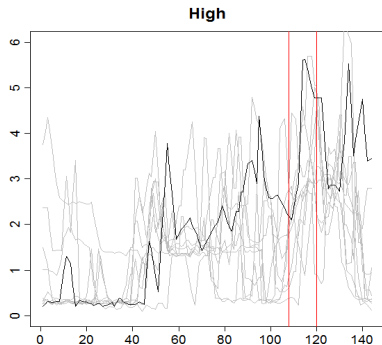
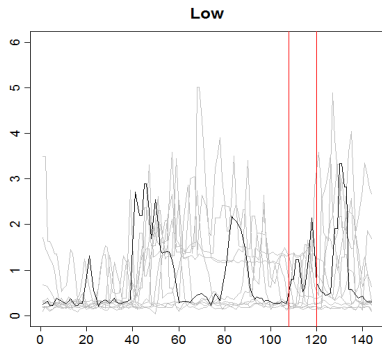
Algorithme: apprentissage de M_{Intra}

```
M = complete intra-class matching 1
repeat
  LinkRemoved = false
  for all (l, l') with y_l = y_l' and l ≠ l' do
    for all (i, i') ∈ [1, T] × [1, T] do
      C_{ii'}^{ll'} evaluation 2
    end for
  end for
  for all (i, l) ∈ [1, T] × [1, n] do
    Link = argmax_{i', l'} (C_{ii'}^{ll'}) (C.3-4) 3
    if Link ≠ ∅ then
      Remove Link (m_{i,i'}^{l,l'} = 0) and
      Update weights (C.1-2)
      LinkRemoved = true
    end if
  end for
until ¬LinkRemoved 4
return(M_{Intra} = M)
```

Applications : courbes de consommations électriques

Georges Hébrail (EDF Clamart)

- Mesures annuelles de la consommation électrique au sein d'un foyer individuel
- Courbes de consommations électriques journalières (échantillonnage / 10mn)



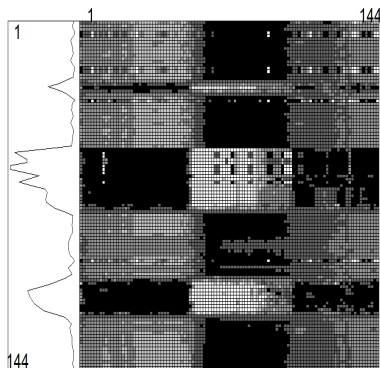
Objectif : prédiction précoces de la consommation électrique entre 18h-20h

Applications : courbes de consommations électriques

Georges Hébrail (EDF Clamart)

Appariements discriminant appris (CONSLEVEL)

M_{Intra}^* (Low)



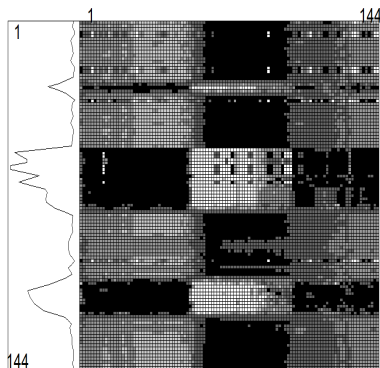
M_{Inter}^* (Low vs. High)

Applications : courbes de consommations électriques

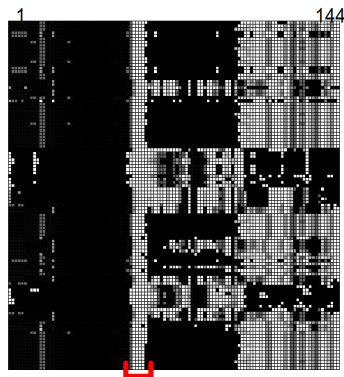
Georges Hébrail (EDF Clamart)

Appariements discriminant appris (CONSLEVEL)

M_{Intra}^* (Low)



M_{Inter}^* (Low vs. High)



Applications : courbes de consommations électriques

- D fondée sur les alignements discriminants

Table 2: k -Nearest Neighbor classification error rates

	k	D	DE	DTW
BME	1	0.032	0.165	0.130
	3	0.034	0.208	0.132
	5	0.062	0.234	0.136
	7	0.079	0.297	0.191
UMD	1	0.055	0.173	0.121
	3	0.111	0.333	0.177
	5	0.173	0.343	0.225
	7	0.222	0.378	0.274
CONSLEVEL	1	0.056	0.306	0.289
	3	0.044	0.267	0.261
	5	0.028	0.233	0.239
	7	0.017	0.233	0.233
CONSSEASON	1	0.094	0.239	0.283
	3	0.128	0.228	0.311
	5	0.205	0.200	0.300
	7	0.111	0.222	0.306
TRAJ	1	0.014	0.012	0.019
	3	0.018	0.017	0.022
	5	0.022	0.021	0.028
	7	0.019	0.021	0.026

- Distribution des courbes de consommations (CONSLEVEL)

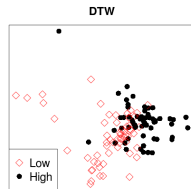
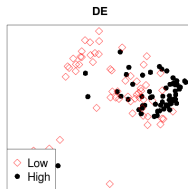
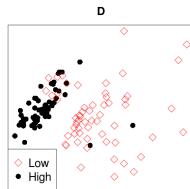
Applications : courbes de consommations électriques

- Distribution des courbes de consommations (CONSLEVEL)

- D fondée sur les alignements discriminants

Table 2: k -Nearest Neighbor classification error rates

	k	D	DE	DTW
BME	1	0.032	0.165	0.130
	3	0.034	0.208	0.132
	5	0.062	0.234	0.136
	7	0.079	0.297	0.191
UMD	1	0.055	0.173	0.121
	3	0.111	0.333	0.177
	5	0.173	0.343	0.225
	7	0.222	0.378	0.274
CONSLEVEL	1	0.056	0.306	0.289
	3	0.044	0.267	0.261
	5	0.028	0.233	0.239
	7	0.017	0.233	0.233
CONSSEASON	1	0.094	0.239	0.283
	3	0.128	0.228	0.311
	5	0.205	0.200	0.300
	7	0.111	0.222	0.306
TRAJ	1	0.014	0.012	0.019
	3	0.018	0.017	0.022
	5	0.022	0.021	0.028
	7	0.019	0.021	0.026



Troisième partie

Intégration de la "forme" des séries

Problème et objectif

- **Problème & objectif**

- Classification des séries sur la base de la "forme"
- **Objectif** : Métrique intégrant la composante "forme" des séries

- **Limites des approches existantes**

- Comparaison des valeurs des séries (ignore le voisinage temporel)

- **Contributions**

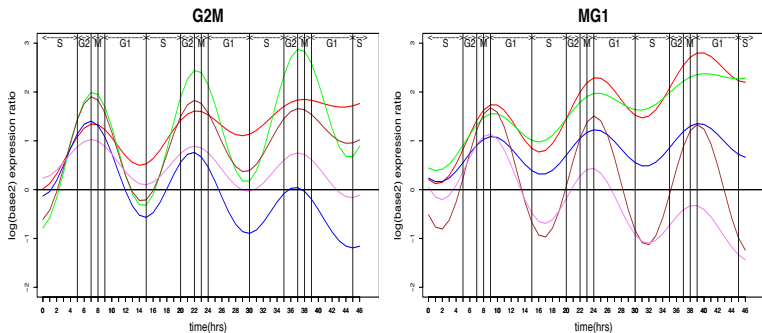
- Définition d'une métrique capturant la forme des séries
- Métriques adaptatives alliant les composantes formes et valeurs.
- Extension des arbres de classification à des séries

Problème et objectif

- **Problème & objectif**

- Classification des séries sur la base de la "forme"

Profils d'expression de gènes



- Objectif : Métrique intégrant la composante "forme" des séries

- Limites des approches existantes

Problème et objectif

- **Problème & objectif**
 - Classification des séries sur la base de la "forme"
 - **Objectif** : Métrique intégrant la composante "forme" des séries
- Limites des approches existantes
 - Comparaison des valeurs des séries (ignore le voisinage temporel)
- Contributions
 - Définition d'une métrique capturant la forme des séries
 - Métriques adaptatives alliant les composantes formes et valeurs.
 - Extension des arbres de classification à des séries

Problème et objectif

- **Problème & objectif**
 - Classification des séries sur la base de la "forme"
 - **Objectif** : Métrique intégrant la composante "forme" des séries
- **Limites des approches existantes**
 - Comparaison des valeurs des séries (ignore le voisinage temporel)
- **Contributions**
 - Définition d'une métrique capturant la forme des séries
 - Métriques adaptatives alliant les composantes formes et valeurs.
 - Extension des arbres de classification à des séries

Problème et objectif

- **Problème & objectif**
 - Classification des séries sur la base de la "forme"
 - **Objectif** : Métrique intégrant la composante "forme" des séries
- **Limites des approches existantes**
 - Comparaison des valeurs des séries (ignore le voisinage temporel)
- **Contributions**
 - Définition d'une métrique capturant la forme des séries
 - Métriques **adaptatives** alliant les composantes formes et valeurs.
 - Extension des arbres de classification à des séries

Comparaison de la forme des séries

Quelques travaux

- Coefficient de Kendall, distance qualitative pour la comparaison des accroissements
Todorovski et al. 2002, Son et Baek 2008
- Coefficient de Spearman pour la comparaison des rangs des observations Balasubramanian et al 2004
- Coefficient de corrélation de Pearson, autocorrélations Abraham et al. 2010, Cabestaing et al. 2007, Rydell et al. 2008, Gaidon et al. 2011

Limites principales

- Coefficient de corrélation de Pearson
 - Implique tous les couples i, i' (présume de l'indépendance des observations)
 - Surestime la similarité entre formes
- Autocorrélation
 - Invariant pour des dynamiques opposées
 - Autocorrélations proches \nRightarrow dynamiques similaires !

Comparaison de la forme des séries

Quelques travaux

- Coefficient de Kendall, distance qualitative pour la comparaison des accroissements
[Todorovski et al. 2002](#), [Son et Baek 2008](#)
- Coefficient de Spearman pour la comparaison des rangs des observations [Balasubramanian et al 2004](#)
- Coefficient de corrélation de Pearson, autocorrélations [Abraham et al. 2010](#), [Cabestaing et al. 2007](#), [Rydell et al. 2008](#), [Gaidon et al. 2011](#)

Limites principales

- Coefficient de corrélation de Pearson
 - Implique tous les couples i, i' (présume de l'indépendance des observations)
 - Surestime la similarité entre formes
- Autocorrélation
 - Invariant pour des dynamiques opposées
 - Autocorrélations proches \nRightarrow dynamiques similaires !

Comparaison de la forme des séries

Quelques travaux

- Coefficient de Kendall, distance qualitative pour la comparaison des accroissements
[Todorovski et al. 2002](#), [Son et Baek 2008](#)
- Coefficient de Spearman pour la comparaison des rangs des observations [Balasubramanian et al 2004](#)
- Coefficient de corrélation de Pearson, autocorrélations [Abraham et al. 2010](#), [Cabestaing et al. 2007](#), [Rydell et al. 2008](#), [Gaidon et al. 2011](#)

Limites principales

- Coefficient de corrélation de Pearson
 - Implique tous les couples i, i' (présume de l'indépendance des observations)
 - Surestime la similarité entre formes
- Autocorrélation
 - Invariant pour des dynamiques opposées
 - Autocorrélations proches \nRightarrow dynamiques similaires !

Comparaison de la forme des séries

Quelques travaux

- Coefficient de Kendall, distance qualitative pour la comparaison des accroissements
[Todorovski et al. 2002](#), [Son et Baek 2008](#)
- Coefficient de Spearman pour la comparaison des rangs des observations [Balasubramanian et al 2004](#)
- Coefficient de corrélation de Pearson, autocorrélations [Abraham et al. 2010](#), [Cabestaing et al. 2007](#), [Rydell et al. 2008](#), [Gaidon et al. 2011](#)

Limites principales

- Coefficient de corrélation de Pearson
 - Implique tous les couples i, i' (présume de l'indépendance des observations)
 - Surestime la similarité entre formes
- Autocorrélation
 - Invariant pour des dynamiques opposées
 - Autocorrélations proches \nRightarrow dynamiques similaires !

Comparaison de la forme des séries

Quelques travaux

- Coefficient de Kendall, distance qualitative pour la comparaison des accroissements
[Todorovski et al. 2002](#), [Son et Baek 2008](#)
- Coefficient de Spearman pour la comparaison des rangs des observations [Balasubramanian et al 2004](#)
- Coefficient de corrélation de Pearson, autocorrélations [Abraham et al. 2010](#), [Cabestaing et al. 2007](#), [Rydell et al. 2008](#), [Gaidon et al. 2011](#)

Limites principales

- Coefficient de corrélation de Pearson
 - Implique tous les couples i, i' (présume de l'indépendance des observations)
 - Surestime la similarité entre formes
- Autocorrélation
 - Invariant pour des dynamiques opposées
 - Autocorrélations proches \nrightarrow dynamiques similaires !

Contribution

Coefficient de corrélation temporelle $Cort(x, y)$ d'ordre r

$$Cort(x, y) = \frac{\sum_{i,i'} m_{ii'} (x_i - x_{i'}) (y_i - y_{i'})}{\sqrt{\sum_{i,i'} m_{ii'} (x_i - x_{i'})^2} \sqrt{\sum_{i,i'} m_{ii'} (y_i - y_{i'})^2}} = \frac{Cov_M}{\sqrt{Var_M}}$$

$M(m_{ii'})$: $m_{ii'} = 1$ si $|i' - i| \leq r$, 0 sinon (dépendance temporelle de voisinage r)

Force et faiblesse

- + Dynamiques similaires \Leftrightarrow Coefficient de corrélation temporelle fort
- Sensible aux bruits, pris en compte pour fixer r

Analyses fondées sur $Cort$

- Analyse de données d'expression de gènes (Thèse : A. Diallo, F. Giroud)
- Définition de métriques adaptatives alliant formes et valeurs des séries (P. Nagabhushan)
- Extension des arbres de classification à des séries (C. Amblard)

Contribution

Coefficient de corrélation temporelle $Cort(x, y)$ d'ordre r

$$Cort(x, y) = \frac{\sum_{i, i'} m_{ii'} (x_i - x_{i'}) (y_i - y_{i'})}{\sqrt{\sum_{i, i'} m_{ii'} (x_i - x_{i'})^2} \sqrt{\sum_{i, i'} m_{ii'} (y_i - y_{i'})^2}} = \frac{Cov_M}{\sqrt{Var_M}}$$

$M(m_{ii'})$: $m_{ii'} = 1$ si $|i' - i| \leq r$, 0 sinon (dépendance temporelle de voisinage r)

Force et faiblesse

- + Dynamiques similaires \Leftrightarrow Coefficient de corrélation temporelle fort
- Sensible aux bruits, pris en compte pour fixer r

Analyses fondées sur $Cort$

- Analyse de données d'expression de gènes (Thèse : A. Diallo, F. Giroud)
- Définition de métriques adaptatives alliant formes et valeurs des séries (P. Nagabhushan)
- Extension des arbres de classification à des séries (C. Amblard)

Contribution

Coefficient de corrélation temporelle $Cort(x, y)$ d'ordre r

$$Cort(x, y) = \frac{\sum_{i, i'} m_{ii'} (x_i - x_{i'}) (y_i - y_{i'})}{\sqrt{\sum_{i, i'} m_{ii'} (x_i - x_{i'})^2} \sqrt{\sum_{i, i'} m_{ii'} (y_i - y_{i'})^2}} = \frac{Cov_M}{\sqrt{Var_M}}$$

$M(m_{ii'})$: $m_{ii'} = 1$ si $|i' - i| \leq r$, 0 sinon (dépendance temporelle de voisinage r)

Force et faiblesse

- + Dynamiques similaires \Leftrightarrow Coefficient de corrélation temporelle fort
- Sensible aux bruits, pris en compte pour fixer r

Analyses fondées sur $Cort$

- Analyse de données d'expression de gènes (Thèse : A. Diallo, F. Giroud)
- Définition de métriques adaptatives alliant formes et valeurs des séries (P. Nagabhushan)
- Extension des arbres de classification à des séries (C. Amblard)

Classification par arbre de séries ("Digits")

Nouveau critère de coupure

- Métrique alliant les composantes **forme et valeurs** des séries
- **Métrique adaptative** adapter le compromis forme/valeurs aux propriétés dynamiques des séries
- Localisation de **sous-séquences discriminantes**
- **Interprétabilité** de l'arbre induit

Classification par arbre de séries ("Digits")

Nouveau critère de coupure

- Métrique alliant les composantes **forme et valeurs** des séries
- **Métrique adaptative** adapter le compromis forme/valeurs aux propriétés dynamiques des séries
- Localisation de **sous-séquences discriminantes**
- **Interprétabilité** de l'arbre induit

Classification par arbre de séries ("Digits")

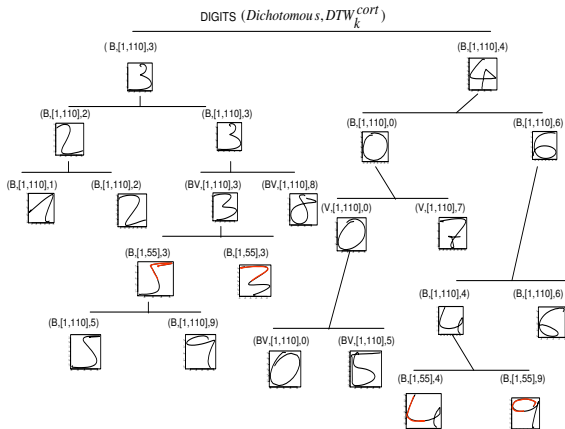
Nouveau critère de coupure

- Métrique alliant les composantes **forme et valeurs** des séries
- **Métrique adaptative** adapter le compromis forme/valeurs aux propriétés dynamiques des séries
- Localisation de **sous-séquences discriminantes**
- **Interprétabilité** de l'arbre induit

Classification par arbre de séries ("Digits")

Nouveau critère de coupure

- Métrique alliant les composantes **forme** et **valeurs** des séries
- **Métrique adaptative** adapter le compromis forme/valeurs aux propriétés dynamiques des séries
- Localisation de **sous-séquences discriminantes**
- **Interprétabilité** de l'arbre induit



Conclusion

- **Analyse de longues séries multivariées**
 - **Segmentation** : nouvelles approches de segmentation de séries multivariées préservant les corrélations
- **Classification de séries de structures complexes**
 - **Appariements discriminants** : nouvelle stratégie d'appariement révélant les caractéristiques discriminants des classes de séries
- **Classification des séries intégrant la composante "Forme"**
 - **Forme des séries** : a) une mesure de similarité fondée sur la forme des séries, b) métriques adaptatives alliant les composantes forme et valeurs et c) nouveau critère de coupure étendant les arbres de classification à des séries temporelles

Conclusion

- **Analyse de longues séries multivariées**
 - **Segmentation** : nouvelles approches de segmentation de séries multivariées préservant les corrélations
- **Classification de séries de structures complexes**
 - **Appariements discriminants** : nouvelle stratégie d'appariement révélant les caractéristiques discriminants des classes de séries
- **Classification des séries intégrant la composante "Forme"**
 - **Forme des séries** : a) une mesure de similarité fondée sur la forme des séries, b) métriques adaptatives alliant les composantes forme et valeurs et c) nouveau critère de coupure étendant les arbres de classification à des séries temporelles

Conclusion

- **Analyse de longues séries multivariées**
 - **Segmentation** : nouvelles approches de segmentation de séries multivariées préservant les corrélations
- **Classification de séries de structures complexes**
 - **Appariements discriminants** : nouvelle stratégie d'appariement révélant les caractéristiques discriminants des classes de séries
- **Classification des séries intégrant la composante "Forme"**
 - **Forme des séries** : a) une mesure de similarité fondée sur la forme des séries, b) métriques adaptatives alliant les composantes forme et valeurs et c) nouveau critère de coupure étendant les arbres de classification à des séries temporelles

Cinq revues majeures

- Advanced Data Analysis and Classification (2007)
- Computational Statistics and Data Analysis (2009)
- Pattern Recognition Letters (2010)
- Statistical Analysis and Data Mining (2011)
- Pattern Recognition (2012)

Publications (36 publications)

- 5 revues internationales, 2 revues nationales avec comité de lecture, 4 chapitres de livres, 9 conférences internationales avec publication des actes, 6 conférences internationales à publications courtes, 10 conférences nationales avec publication des actes, 3 thèses co-encadrées

À très brèves échéances

- Pallier la surparamétrisation des appariements
 - Régularisation des blocs $M_{ll'}$
 - Initialisation par bloc diagonaux entre segments saillants ([DIALIGN](#), [Brudno et al. 2003](#))

Analyse de la dépendance temporelle

- Classification non supervisée par apprentissage d'appariements (définition de prototypes)
- Étude de noyaux dédiés aux séries (analyse des séries par des méthodes à noyaux)

Analyse de la dépendance liée à des structures plus générales

- Analyse de structures impliquant différents types de dépendances (par décomposition de matrices non négatives)
- Analyse de graphes évolutifs, de structures spatio-temporelles

À très brèves échéances

- Pallier la surparamétrisation des appariements
 - Régularisation des blocs $M_{ll'}$
 - Initialisation par bloc diagonaux entre segments saillants ([DIALIGN](#), [Brudno et al. 2003](#))

Analyse de la dépendance temporelle

- Classification non supervisée par apprentissage d'appariements (définition de prototypes)
- Étude de noyaux dédiés aux séries (analyse des séries par des méthodes à noyaux)

Analyse de la dépendance liée à des structures plus générales

- Analyse de structures impliquant différents types de dépendances (par décomposition de matrices non négatives)
- Analyse de graphes évolutifs, de structures spatio-temporelles

À très brèves échéances

- Pallier la surparamétrisation des appariements
 - Régularisation des blocs $M_{ll'}$
 - Initialisation par bloc diagonaux entre segments saillants ([DIALIGN](#), [Brudno et al. 2003](#))

Analyse de la dépendance temporelle

- Classification non supervisée par apprentissage d'appariements (définition de prototypes)
- Étude de noyaux dédiés aux séries (analyse des séries par des méthodes à noyaux)

Analyse de la dépendance liée à des structures plus générales

- Analyse de structures impliquant différents types de dépendances (par décomposition de matrices non négatives)
- Analyse de graphes évolutifs, de structures spatio-temporelles

Projets en cours

Analyse de la dépendance temporelle

● Biologie

- ◇ ANR DYNOTEP (SVSE6) avec le CEA (début 2013, 25%)
 - Exploration de profils d'abondance de peptides pour l'analyse de la dynamique de cellules endocytaires

● Réseaux de capteurs

- ◇ Schneider : Innovation pour l'Efficacité Energétique
 - Étude de méthodes d'apprentissage adaptées aux réseaux de capteurs
 - Thèse : Cao-Tri Do (Cifre, début : janvier 2013), Co-Dir M. Rambaut

Analyse de la dépendance liée à des structures plus générales

● Réseaux sociaux

- ◇ BestOfMedia
 - Détection et prédiction de thèmes émergents dans les médias sociaux
 - Thèse : F. Kawala (Cifre, début : sep 2011), Dir. E. Gaussier
- ◇ Namayan Farda Company
 - Prédiction de liens dans des réseaux évolutifs
 - Thèse : S. Souheily (début : mars 2013), Co-Dir. E. Gaussier

● Spatio/temporel

- ◇ Apprentissage et recommandation d'itinéraires partagés à partir de trajectoires GPS (spatio-temporel), H. Martin

Projets en cours

Analyse de la dépendance temporelle

● Biologie

- ◇ ANR DYNOTEP (SVSE6) avec le CEA (début 2013, 25%)
 - Exploration de profils d'abondance de peptides pour l'analyse de la dynamique de cellules endocytaires

● Réseaux de capteurs

- ◇ Schneider : Innovation pour l'Efficacité Energétique
 - Étude de méthodes d'apprentissage adaptées aux réseaux de capteurs
 - Thèse : Cao-Tri Do (Cifre, début : janvier 2013), Co-Dir M. Rambaut

Analyse de la dépendance liée à des structures plus générales

● Réseaux sociaux

- ◇ BestOfMedia
 - Détection et prédiction de thèmes émergents dans les médias sociaux
 - Thèse : F. Kawala (Cifre, début : sep 2011), Dir. E. Gaussier
- ◇ Namayan Farda Company
 - Prédiction de liens dans des réseaux évolutifs
 - Thèse : S. Souheily (début : mars 2013), Co-Dir. E. Gaussier

● Spatio/temporel

- ◇ Apprentissage et recommandation d'itinéraires partagés à partir de trajectoires GPS (spatio-temporel), H. Martin

Projets en cours

Analyse de la dépendance temporelle

● Biologie

- ◇ ANR DYNOTEP (SVSE6) avec le CEA (début 2013, 25%)
 - Exploration de profils d'abondance de peptides pour l'analyse de la dynamique de cellules endocytaires

● Réseaux de capteurs

- ◇ Schneider : Innovation pour l'Efficacité Energétique
 - Étude de méthodes d'apprentissage adaptées aux réseaux de capteurs
 - Thèse : Cao-Tri Do (Cifre, début : janvier 2013), Co-Dir M. Rambaut

Analyse de la dépendance liée à des structures plus générales

● Réseaux sociaux

- ◇ BestOfMedia
 - Détection et prédiction de thèmes émergents dans les médias sociaux
 - Thèse : F. Kawala (Cifre, début : sep 2011), Dir. E. Gaussier
- ◇ Namayan Farda Company
 - Prédiction de liens dans des réseaux évolutifs
 - Thèse : S. Souheily (début : mars 2013), Co-Dir. E. Gaussier

● Spatio/temporel

- ◇ Apprentissage et recommandation d'itinéraires partagés à partir de trajectoires GPS (spatio-temporel), H. Martin

Projets en cours

Analyse de la dépendance temporelle

● Biologie

- ◇ ANR DYNOTEP (SVSE6) avec le CEA (début 2013, 25%)
 - Exploration de profils d'abondance de peptides pour l'analyse de la dynamique de cellules endocytaires

● Réseaux de capteurs

- ◇ Schneider : Innovation pour l'Efficacité Energétique
 - Étude de méthodes d'apprentissage adaptées aux réseaux de capteurs
 - Thèse : Cao-Tri Do (Cifre, début : janvier 2013), Co-Dir M. Rambaut

Analyse de la dépendance liée à des structures plus générales

● Réseaux sociaux

- ◇ BestOfMedia
 - Détection et prédiction de thèmes émergents dans les médias sociaux
 - Thèse : F. Kawala (Cifre, début : sep 2011), Dir. E. Gaussier
- ◇ Namayan Farda Company
 - Prédiction de liens dans des réseaux évolutifs
 - Thèse : S. Souheily (début : mars 2013), Co-Dir. E. Gaussier

● Spatio/temporel

- ◇ Apprentissage et recommandation d'itinéraires partagés à partir de trajectoires GPS (spatio-temporel), H. Martin

Projets en cours

Analyse de la dépendance temporelle

● Biologie

- ◇ ANR DYNOTEP (SVSE6) avec le CEA (début 2013, 25%)
 - Exploration de profils d'abondance de peptides pour l'analyse de la dynamique de cellules endocytaires

● Réseaux de capteurs

- ◇ Schneider : Innovation pour l'Efficacité Energétique
 - Étude de méthodes d'apprentissage adaptées au réseaux de capteurs
 - Thèse : Cao-Tri Do (Cifre, début : janvier 2013), Co-Dir M. Rambaut

Analyse de la dépendance liée à des structures plus générales

● Réseaux sociaux

- ◇ BestOfMedia
 - Détection et prédiction de thèmes émergents dans les médias sociaux
 - Thèse : F. Kawala (Cifre, début : sep 2011), Dir. E. Gaussier
- ◇ Namayan Farda Company
 - Prédiction de liens dans des réseaux évolutifs
 - Thèse : S. Souheily (début : mars 2013), Co-Dir. E. Gaussier

● Spatio/temporel

- ◇ Apprentissage et recommandation d'itinéraires partagés à partir de trajectoires GPS (spatio-temporel), H. Martin

Merci !

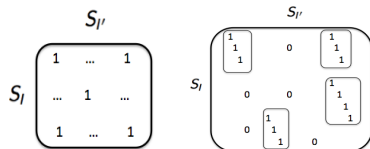
Considérations calculatoires

- Le calcul des contributions $C_{i_1 i_2}^{l_1 l_2}$

$$\begin{aligned}
 C_{i_1 i_2}^{l_1 l_2} &= V_{M^{Intra}} - V_{M^{Intra} \setminus (i_1, i_2, l_1, l_2)} \\
 &= \frac{1}{nT} (x_{i_1}^{l_1} - \sum_{l'=1}^{n_k} \sum_{i'=1}^T m_{i_1 i'}^{l_1 l'} x_{i'}^{l'})^2 - \frac{1}{nT} (x_{i_1}^{l_1} - \sum_{l'=1}^{n_k} \sum_{i'=1}^T \frac{m_{i_1 i'}^{l_1 l'}}{1 - m_{i_1 i_2}^{l_1 l_2}} x_{i'}^{l'})^2 \quad (1)
 \end{aligned}$$

$C_{i_1 i_2}^{l_1 l_2}$ se réduit à la différence entre deux valeurs centrées, impliquant les seules lignes i_1 , et i_2 de M^{Intra}

- Couplage initial par blocs diagonaux entre segments saillants (DIALIGN, Brudno et al. 2003)



Complexité

Algorithme: apprentissage de M_{Intra}

- Nb de liens évalués :

$$\approx O\left(\sum_{k=1}^K l_k n_k^2 T^2\right)$$

Au pire, $l_k = (n_k - 1)(T - 1)$

$$\approx O\left(\sum_{k=1}^K n_k^3 T^3\right)$$

- Parallélisation de l'apprentissage de M_{Intra} et M_{Inter} par classe et par ligne

$M =$ complete intra-class matching

repeat

$LinkRemoved = false$

for all (l, l') with $y_l = y_{l'}$ and $l \neq l'$ do

for all $(i, i') \in [1, T] \times [1, T]$ do

$C_{ii'}^{ll'}$ evaluation

end for

end for

for all $(i, l) \in [1, T] \times [1, n]$ do

$Link = \operatorname{argmax}_{i', l'} (C_{ii'}^{ll'})$

if $Link \neq \emptyset$ then

Remove $Link$ ($m_{i', i'}^{l, l'} = 0$) and

Update weights

$LinkRemoved = true$

end if

end for

until $\neg LinkRemoved$

return($M_{Intra} = M$)

1

2

3

4

2. Comparaison classique des séries

$\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_m)$ éléments de $X = \mathfrak{R}^d$

$$DTS(\mathbf{x}, \mathbf{y}) = \min_{\pi \in A(n, m)} D_{\mathbf{x}, \mathbf{y}}(\pi)$$

$$D_{\mathbf{x}, \mathbf{y}}(\pi) = \sum_{i=1}^{|\pi|} \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})$$

$A(n, m)$: l'ensemble des alignements $\pi(\mathbf{x}, \mathbf{y})$

$(\pi_1(i), \pi_2(i))$: i^{eme} couple d'observations alignées

$\varphi(x_{\pi_1(i)}, y_{\pi_2(i)}) = \|x_{\pi_1(i)} - y_{\pi_2(i)}\|^2$, ignore le voisinage temporel des valeurs

Comparaison de la forme des séries : autocorrélation

Mesures fondées sur les autocorrélations (Chatfield 2009)

$$\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$$

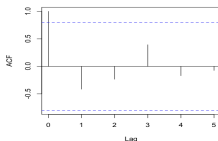
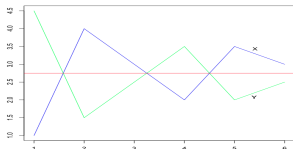
$$\rho_\tau(x) = \frac{\sum_{i=1}^{T-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x})}{\sum_{i=1}^T (x_i - \bar{x})^2}$$

$$\tilde{x} = (\rho_1(x), \dots, \rho_K(x)) \quad \tilde{y} = (\rho_1(y), \dots, \rho_K(y))$$

$$D(x, y) = \varphi(\tilde{x}, \tilde{y})$$

Faiblesse

- Coefficient d'autocorrélations ρ_τ proches \nRightarrow dynamiques similaires !



$$\tilde{x} = \tilde{y} = (1.000, -0.415, -0.234, 0.394, -0.170, -0.074)$$

Comparaison de la forme des séries : corrélations & autocorrélation

Coefficient de corrélations de Pearson

$$\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$$

$$\text{Cor}(x, y) = \frac{\sum_{i,i'} (x_i - x_{i'})(y_i - y_{i'})}{\sqrt{\sum_{i,i'} (x_i - x_{i'})^2} \sqrt{\sum_{i,i'} (y_i - y_{i'})^2}}$$

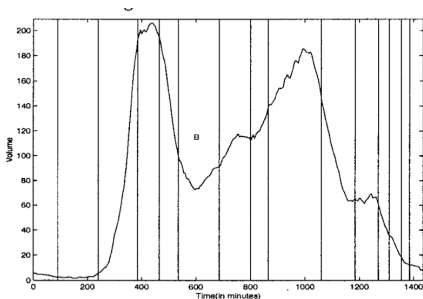
Faiblesses

- Implique tous les couples i, i' (présume les obs. indépendantes)
- Surestime la similarité entre formes

Segmentation

Caractéristiques des approches standards

- Recherche de points changements de caractéristiques (moyenne, variance), ne rend pas compte de la dépendance temporelle des observations,



Change points detection by likelihood estimate [Guralnik et al. 99](#)