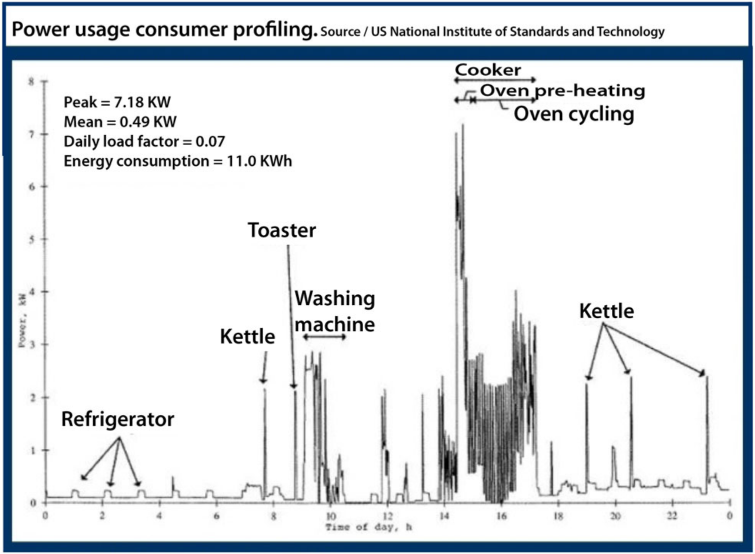
Intelligence Artificielle & Machine Learning pour la modélisation de séries temporelles et de signaux

Séance 7 : Détection et extraction de motifs

Laurent Oudre
laurent.oudre@ens-paris-saclay.fr

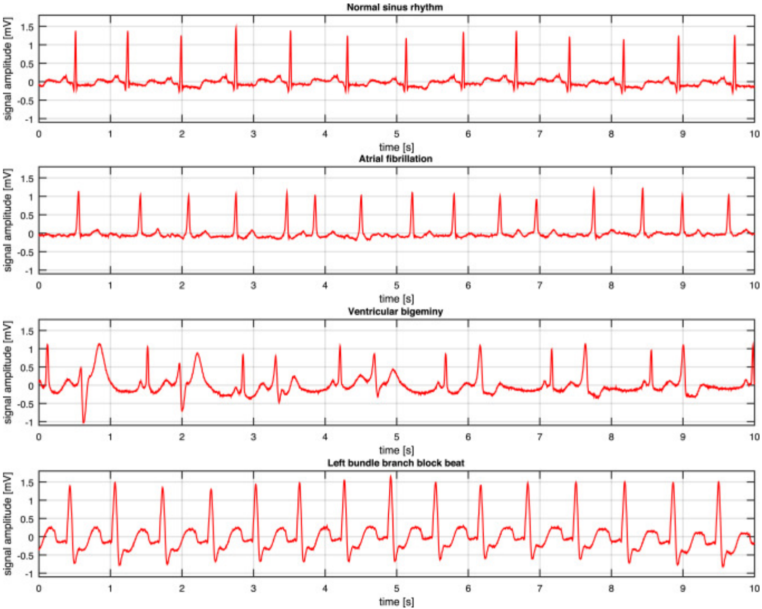
Diplôme ARIA
ENS Paris Saclay
2025-2026

Motivation



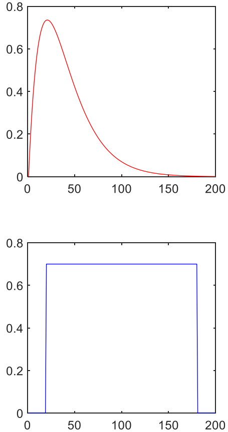
Power usage consumer profiling. Source / US National Institute of Standards and Technology
Peak = 7.18 KW
Mean = 0.49 KW
Daily load factor = 0.07
Energy consumption = 11.0 KWh
Cooker
Oven pre-heating
Oven cycling
Toaster
Washing machine
Kettle
Kettle
Refrigerator
Power, kW
Time of day, h

Motivation

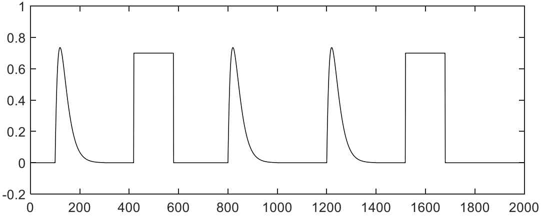


Normal sinus rhythm
signal amplitude [mV]
time [s]
Atrial fibrillation
signal amplitude [mV]
time [s]
Ventricular bigeminy
signal amplitude [mV]
time [s]
Left bundle branch block beat
signal amplitude [mV]
time [s]

Problème 1 : Détection de motifs

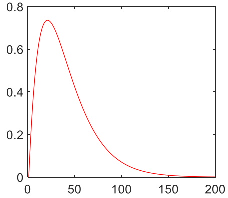


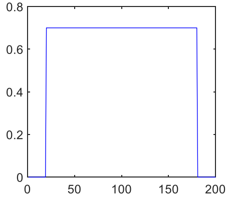
Dictionary of patterns



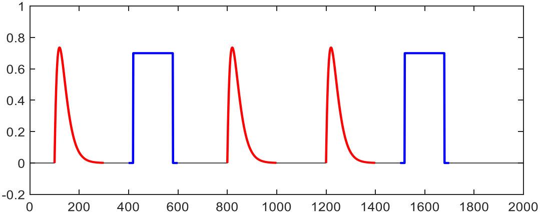
Input time series

Problème 1 : Détection de motifs





Dictionary of patterns

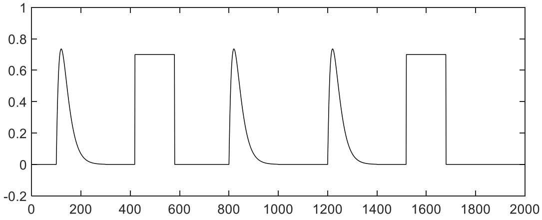


Annotated time series

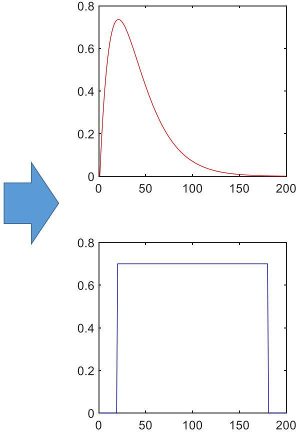
Problème 1 : Détection de motifs

- A partir d'un dictionnaire de motifs $\mathcal{P}$, retrouver ces motifs dans une série temporelle $\mathbf{x}$
- Les motifs de $\mathcal{P}$ peuvent avoir des longueurs différentes
- Les motifs peuvent être annotés, c'est à dire correspondant chacun à un phénomène d'intérêt. Dans ce cas, la détection de motifs fournira une annotation automatique de la série temporelle

Problème 2 : Extraction de motifs



Input time series



Extracted patterns

Problème 2 : Extraction de motifs

- Etant donnée une série temporelle $\mathbf{x}$ (ou plusieurs), apprendre un dictionnaire de motifs $\mathcal{P}$
- Un motif correspond à une *forme* qui apparaît de façon répétitive dans la série temporelle : il s'agit d'une notion assez floue
- On suppose dans ce cas que tous les motifs ont la même taille
- Les motifs appris peuvent servir à caractériser la série temporelle ou étudiés individuellement

Programme de la séance

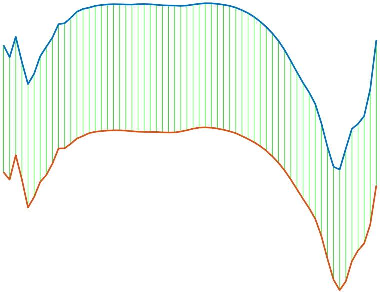
- Introduction de différentes *distances* permettant de comparer des séries temporelles
- Algorithmes pour la détection et l'extraction de motifs dans les séries temporelles

Session 7 : Pattern detection and extraction

Plan du cours

Plan du cours

Distance euclidienne

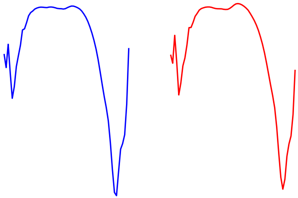


$$d_{EUC}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}$$

- Sensible aux décalages temporels, aux changements d'amplitude et d'ordonnées et aux contractions/dilatations
- Nécessite un alignement parfait des lignes temporelles
- Sensible au bruit impulsionnel mais relativement résistante au bruit blanc additif gaussien

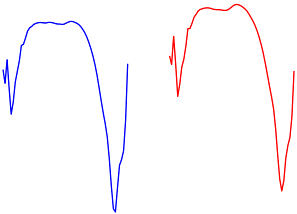
Robustesse de la distance euclidienne

Cas standard



$$d_{EUC} = 3.5$$

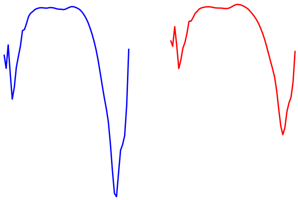
Changement d'ordonnée à l'origine



$$d_{EUC} = 9.6$$

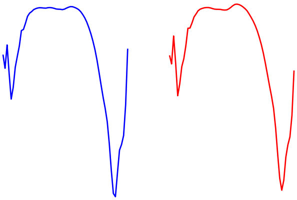
Robustesse de la distance euclidienne

Changement d'amplitude (70 %)



$$d_{EUC} = 12.8$$

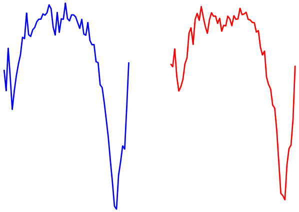
Décalage temporel (1.6 %)



$$d_{EUC} = 5.3$$

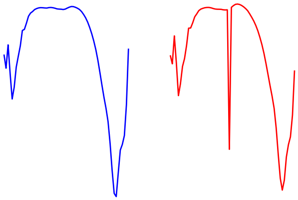
Robustesse de la distance euclidienne

Bruit blanc additif gaussien



$$d_{EUC} = 6.9$$

Outliers



$$d_{EUC} = 10.3$$

Distance euclidienne normalisée

Afin d'améliorer la robustesse par rapport aux changements d'amplitude et d'ordonnées, on peut utiliser une distance euclidienne normalisée :

$$\boxed{d_{nEUC}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{N} (\tilde{x}_i - \tilde{y}_i)^2}}$$

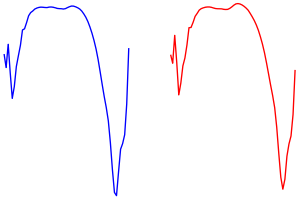
où

$$\tilde{x}_i = \frac{x_i - \mu_x}{\sigma_x} \quad \text{normalisation z-score}$$

- Toujours sensible aux décalages temporels, contraction/dilatation, outliers et lignes temporelles
- Attention : la normalisation peut accroître la sensibilité au bruit blanc additif gaussien

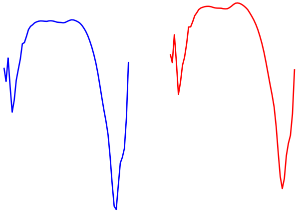
Robustesse de la distance euclidienne normalisée

Cas standard



$d_{nEUC} = 0.8$

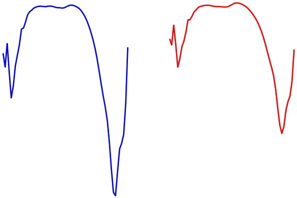
Changement d'ordonnée à l'origine



$d_{nEUC} = 0.8$

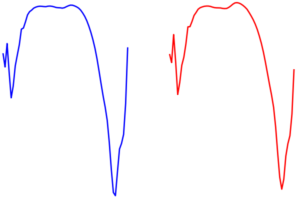
Robustesse de la distance euclidienne normalisée

Changement d'amplitude (70 %)



$d_{nEUC} = 0.8$

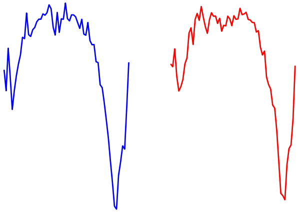
Décalage temporel (1.6 %)



$d_{nEUC} = 1.9$

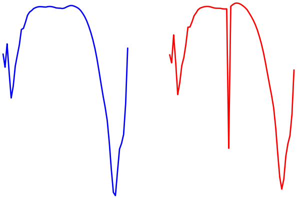
Robustesse de la distance euclidienne normalisée

Bruit blanc additif gaussien



$$d_{nEUC} = 1.9$$

Outliers



$$d_{nEUC} = 2.7$$

Questions en suspens

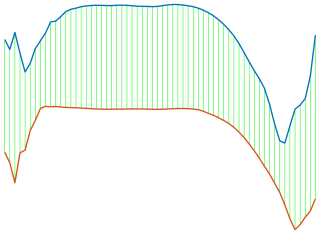
Questions en suspens :

- Comment diminuer la sensibilité par rapport aux lignes temporelles (contraction/dilatation, décalages temporels...) ?
- Comment comparer deux séries temporelles de tailles différentes ?
- Comment étendre la notion de distance euclidienne à des modifications non linéaires de lignes temporelles ?

Nouvelle notion de distance : Dynamic Time Warping (DTW)

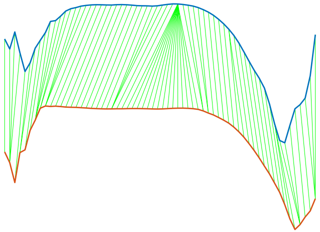
Différences entre la distance euclidienne et la DTW

Distance euclidienne



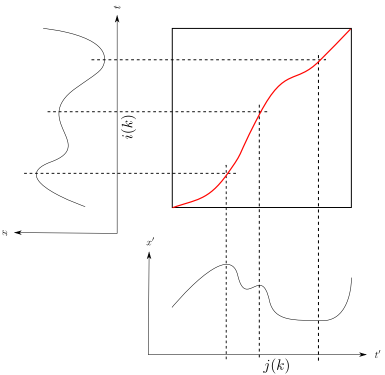
L'échantillon x_i est associé à l'échantillon y_i

DTW



L'échantillon $x_{i(k)}$ est associé à l'échantillon $y_{j(k)}$

Notion de chemin



- La DTW calcule une correspondance entre les éléments de **x** et ceux de **y**.
- Fonction de correspondance appelée **chemin** :

$$P = ((i(1), j(1)), \cdots, (i(K_P), j(K_P))) \in (\mathbb{N} \times \mathbb{N})^{K_P}, K_P \in \mathbb{N}$$

$$(i(k), j(k)) \in P \Leftrightarrow y_{j(k)} \text{ est associé à } x_{i(k)}$$

Choix du chemin optimal

- Le chemin P est évalué grâce à une fonction de coût

$$w(P) = \sum_{k=1}^{K_P} (x_{i(k)} - y_{j(k)})^2$$

- La valeur de la DTW entre deux séries temporelles est la valeur minimale de la fonction de coût

$$\boxed{d_{DTW}(\mathbf{x}, \mathbf{y}) = \sqrt{\min_{P \in \mathcal{P}} w(P)}}$$

- Si $K_P = N$ et $i(k) = j(k) = k$, la DTW est exactement égale à la distance euclidienne

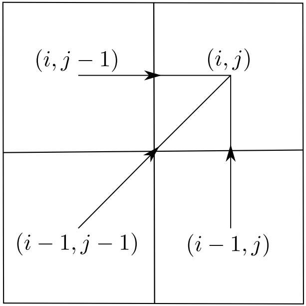
Ensemble des chemins acceptables $\mathcal{P}$

Il existe une infinité de chemins possibles : comment choisir l'ensemble des chemins acceptables $\mathcal{P}$? Etant données deux séries temporelles $\mathbf{x}$ et $\mathbf{y}$ de longueurs respectives M et N, le chemin P doit être :

- **Continu** $i(k) - i(k-1) \leq 1$ et $j(k) - j(k-1) \leq 1$
  Si l'on imagine le plan comme un plateau de jeu d'échecs, le chemin ne peut suivre que les positions d'un roi ;
- **Monotone** $i(k-1) \leq i(k)$ et $j(k-1) \leq j(k)$
  Le chemin peut aller vers le haut, vers la droite ou selon la diagonale haut/droite ;
- **Borné** $(i(1), j(1)) = (1, 1)$ et $(i(K_P), j(K_P)) = (M, N)$
  Le chemin commence et termine respectivement par les premiers/derniers échantillons des deux signaux

$\mathcal{P}$: ensemble des chemins acceptables selon ces contraintes

Ensemble des chemins acceptables $\mathcal{P}$



A partir de $(i(k), j(k))$, il reste trois possibilités de déplacement

$$(i(k-1), j(k-1)) = \begin{cases} & (i(k)-1, j(k)) \\ \text{or} & (i(k), j(k)-1) \\ \text{or} & (i(k)-1, j(k)-1). \end{cases}$$

Algorithme de DTW

1. Calculer les distances deux à deux des échantillons de $\mathbf{x}$ et $\mathbf{y}$, et les stocker dans une matrice $\mathbf{D} \in \mathbb{R}^{M \times N}$:

$$D(i,j) = (x_i - y_j)^2$$

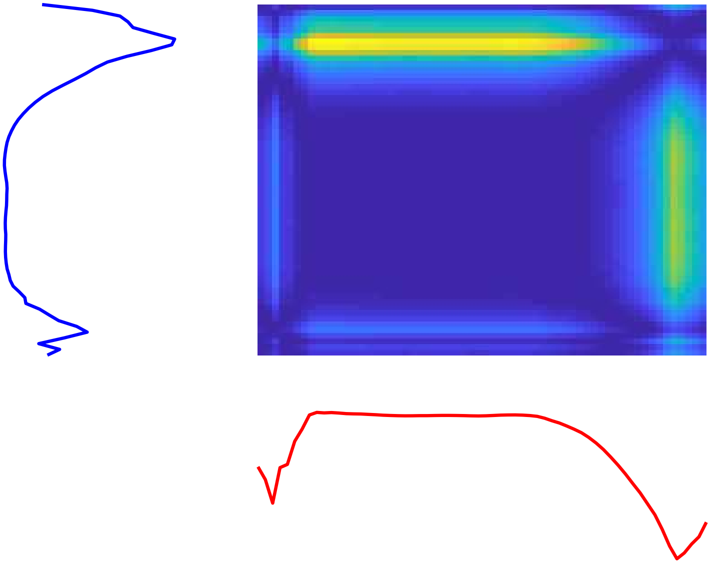
2. Calculer la matrice des distances cumulées $\mathbf{C} \in \mathbb{R}^{M \times N}$ grâce à de la programmation dynamique

$$\begin{aligned} C(i,1) &= D(i,1), \\ C(1,j) &= D(1,j), \\ C(i,j) &= D(i,j) + \min \begin{cases} C(i,j-1) \\ C(i-1,j-1) \\ C(i-1,j) \end{cases} \end{aligned}$$

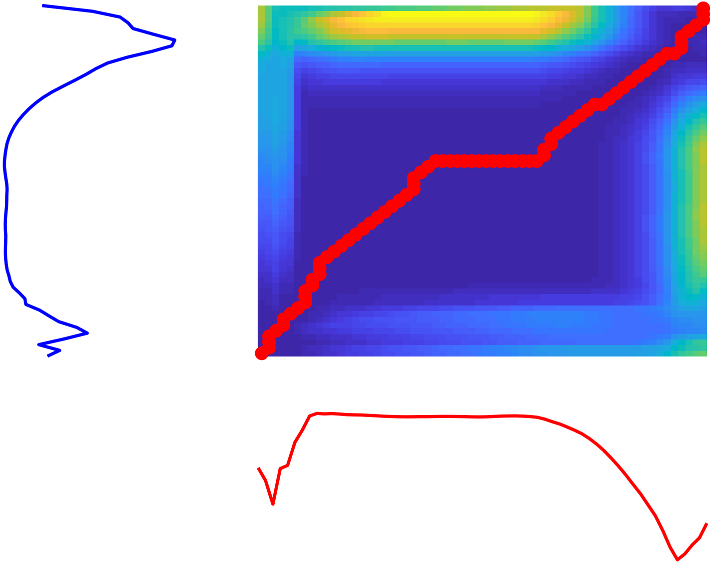
3. La DTW est calculée comme

$$d_{DTW}(\mathbf{x}, \mathbf{y}) = \sqrt{C(M,N)}$$

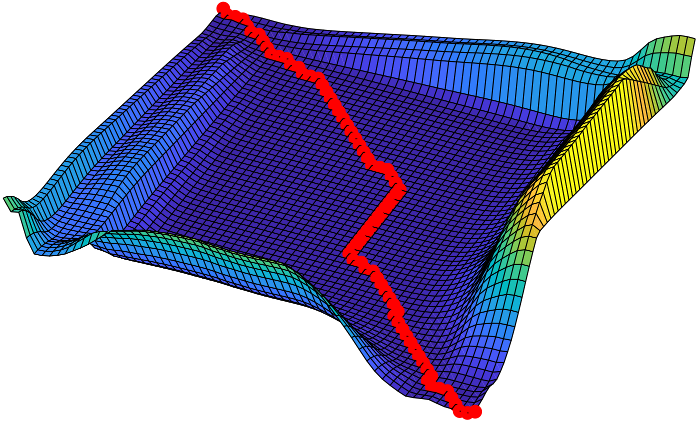
Etape 1 : Calcul de la matrice des distances



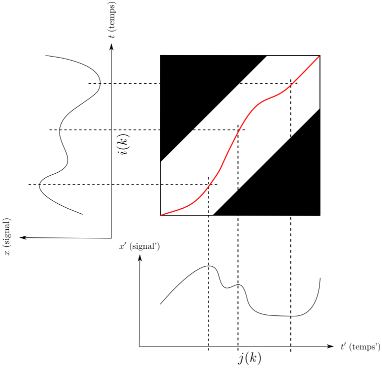
Etape 2 : Calcul de la matrice des distances cumulées



Etape 2 : Calcul de la matrice des distances cumulées



Variantes



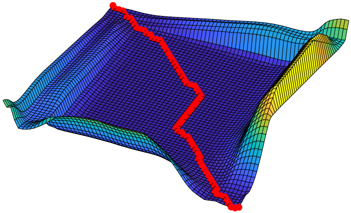
- Si les lignes temporelles de **x** et **y** sont relativement similaires, il n'est pas nécessaire (ni recommandé) de calculer toutes les valeurs de $C(i, j)$
- Des variantes contraintes plus rapides peuvent être implémentées, par exemple en calculant $C(i, j)$ que si

$$|i(k) - j(k)| < \lambda \text{ (Sakoe-Chiba band)}$$

- Ceci permet de considérablement diminuer le temps de calcul, tout en améliorant parfois les performances

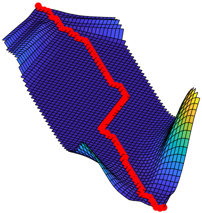
Variantes

DTW classique



9 ms

DTW contrainte



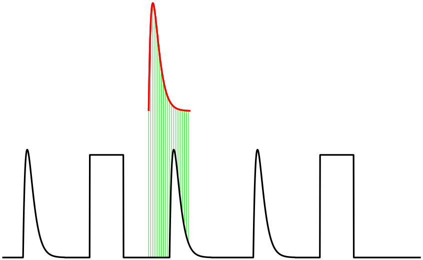
2 ms

Robustesse de la DTW

- Par construction, la DTW est robuste aux changements de ligne temporelle
- En revanche, elle suppose que les premiers et derniers échantillons sont alignés, ce qui peut parfois se révéler problématique
- La DTW est sensible à l'échelle (ce qui peut se contourner en normalisant les signaux préalablement), au bruit et aux outliers

Plan du cours

Problème posé



- Afin de retrouver le motif $\mathbf{p} \in \mathcal{P}$ de longueur N_p dans la séries temporelle $\mathbf{x} = x_{1:N}$ de longueur N, il faut calculer toutes les distances

$$d_i = d(\mathbf{p}, x_{i:i+N_p-1})$$

pour $1 \leq i \leq N - N_d + 1$ et où $x_{i:i+N_d-1}$ est la séquence de $\mathbf{x}$ commençant à l'échantillon i et de longueur N_p

- Intuitivement, il s'agit d'une opération coûteuse en temps de calcul puisque les distances doivent être calculées $N - N_d + 1$ fois

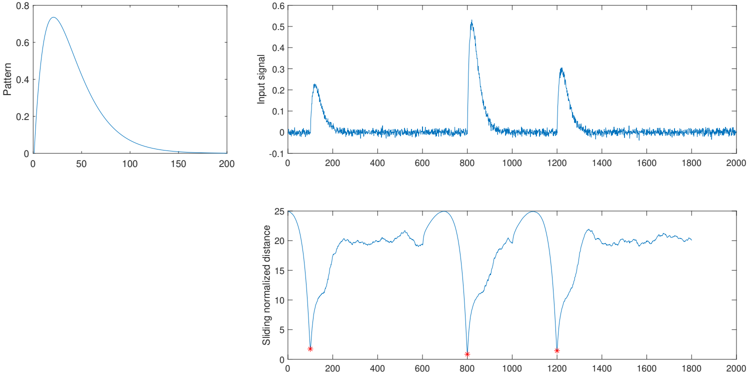
Calculs rapides

- Le vecteur $\mathbf{d}$ qui stocke la distance entre le motif et toutes les séquences de tailles N_p dans la série temporelle est appelé **profil de distance** (ou distance profile en anglais)
- Pour la distance euclidienne (normalisée ou pas), il existe des algorithmes extrêmement rapides pour calculer cette quantité, basés sur le calcul de la Fast Fourier Transform (FFT) qui est un algorithme optimisé pour calculer la transformée de Fourier discrète
- Pour la DTW, c'est plus compliqué mais on peut facilement retrouver

$$d^* = \min_{1 \leq i \leq N - N_d + 1} d_{DTW}(\mathbf{p}, x_{i:i+N_p-1})$$

c'est à dire la position et la valeur de la DTW minimale entre le motif $\mathbf{p}$ et toutes les séquences de $\mathbf{x}$ de taille N_p

Résultats avec la nEUC



Les motifs peuvent facilement être retrouvés dans le signal en cherchant les pics correspondant à une petite distance

Plan du cours

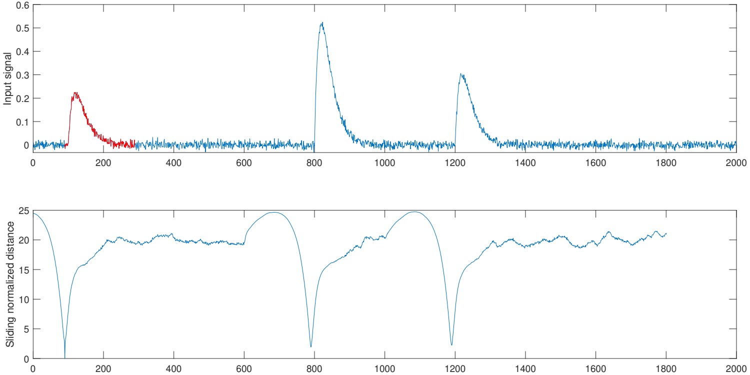
Extraction de motifs

- Nous avons vu comment rechercher des motifs dans des séries temporelles en partant d'un dictionnaire pré-défini
- On va s'intéresser ici à la question inverse : comment extraire des motifs à partir d'une ou plusieurs séries temporelles ? C'est à dire comment apprendre le dictionnaire de motifs ?
- Tâche non supervisée : aucune connaissance a priori à part la durée moyenne des motifs que l'on recherche (notée L)
- **Qu'est-ce qu'un motif ?** Question difficile : notion de formes répétitives, éventuellement de périodicité...

Extraction de motifs basée sur la notion de distance

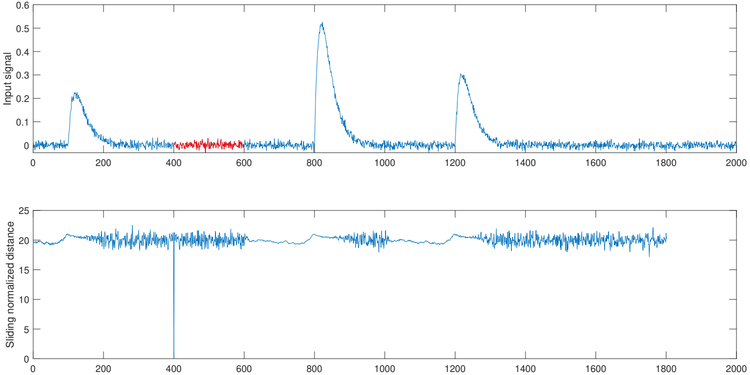
- Intuitivement, on peut ré-utiliser les résultats de la partie précédente pour détecter automatiquement des motifs
- Un motif est une séquence de longueur L qui est susceptible d'apparaître plusieurs fois dans la série temporelle
- En calculant une distance glissante entre cette séquence et toutes les autres séquences de longueur L, on devrait voir apparaître de façon claire que plusieurs d'entre elles sont *proches* d'elle
- **Solution :** Utiliser un algorithme de force brute pour calculer efficacement toutes les distances

Exemple



Lorsque l'on observe le profil de distance associé à cette séquence, on voit apparaître plusieurs pics vers le bas, ce qui montre que le motif est présent ailleurs : il s'agit donc d'une séquence susceptible de correspondre à un motif

Exemple



A l'inverse, pour cette séquence, toutes les distances sont élevées : ce n'est a priori pas un motif.

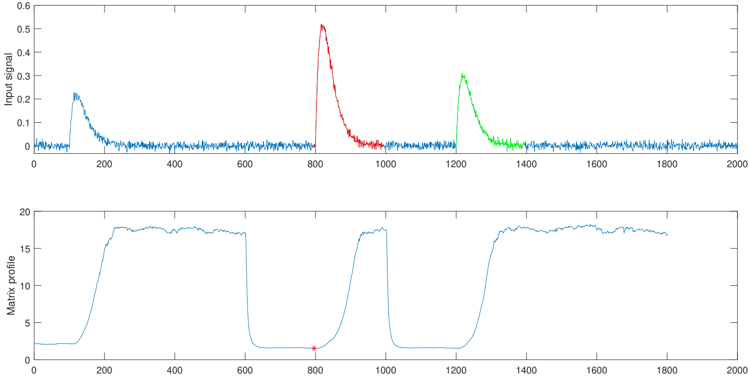
Matrix profile

- Afin de détecter un motif potentiel, il suffit de regarder sa distance minimale à toutes les autres séquences de taille L
- Attention aux correspondances triviales ! Il ne faut comparer que des séquences qui ne se chevauchent pas (sinon la petite distance peut être liée au fait qu'ils ont beaucoup de points en communs)
- **Matrix profile** : Etant donnée une longueur L, calcul de

$$m_i = \min_{j>i+L \text{ or } j<i-L} d(x_{i:i+L-1}, x_{j:j+L-1})$$

- Une petite valeur de matrix profile à l'indice i indique qu'il existe au moins une séquence très proche de $x_{i:i+L-1}$, donc que $x_{i:i+L-1}$ est un bon candidat pour être un motif
- Le calcul de cette quantité peut être rendu très rapide grâce aux techniques déjà mentionnées

Exemple



En rouge : séquence qui a un matrix profil minimal. En vert : la séquence la plus proche de celle en rouge (meilleure correspondance entre deux séquences)

Références

- Berndt, D. J., & Clifford, J. (1994, July). Using dynamic time warping to find patterns in time series. In KDD workshop (Vol. 10, No. 16, pp. 359-370).
- Mueen, A., Keogh, E., Zhu, Q., Cash, S., & Westover, B. (2009, April). Exact discovery of time series motifs. In Proceedings of the 2009 SIAM international conference on data mining (pp. 473-484). Society for Industrial and Applied Mathematics.
- Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016, December). Matrix profile I : all pairs similarity joins for time series : a unifying view that includes motifs, discords and shapelets. In 2016 IEEE 16th international conference on data mining (ICDM) (pp. 1317-1322). Ieee.
- Geurts, P. (2001, September). Pattern extraction for time series classification. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 115-127). Springer, Berlin, Heidelberg.