

# Apprentissage de dictionnaire invariant aux déformations temporelles pour le clustering de séries temporelles

Soutenance de projet — Séries Temporelles

Colin COËRCHON & Chloé HABASQUE

ENS Paris-Saclay

9 janvier 2025

# Informations clefs

- Professeurs référents : **Laurent Oudre & Charles Truong**

école \_\_\_\_\_  
normale \_\_\_\_\_  
supérieure \_\_\_\_\_  
paris—saclay \_\_\_\_\_



# Sommaire

- 1 Introduction
- 2 Méthode
- 3 Jeux de données
- 4 Résultats
- 5 Conclusion

- 1 Introduction
- 2 Méthode
- 3 Jeux de données
- 4 Résultats
- 5 Conclusion

# Présentation de l'article

## Les points importants

Nous avons décidé de travailler sur l'article intitulé "Time warp invariant dictionary learning for time series clustering : application to music data stream analysis". L'article est plutôt récent (date de parution en 2018).

# Présentation de l'article

## Les points importants

Nous avons décidé de travailler sur l'article intitulé "Time warp invariant dictionary learning for time series clustering : application to music data stream analysis". L'article est plutôt récent (date de parution en 2018).

Différentes notions phares sont abordées dans l'article :

- Codage parcimonieux
- Apprentissage de dictionnaire
- Clustering robuste aux décalages temporels

# Motivations

Les motivations présentées dans l'article sont claires :

- ➊ Développer un algorithme de codage parcimonieux invariant au décalage temporel.  
↪ TWI-OMP
- ➋ Proposer une méthode d'apprentissage de dictionnaires discriminants par cluster.

# Motivations

Les motivations présentées dans l'article sont claires :

- ➊ Développer un algorithme de codage parcimonieux invariant au décalage temporel.  
↪ TWI-OMP
- ➋ Proposer une méthode d'apprentissage de dictionnaires discriminants par cluster.
- ➌ Développer une approche de clustering utilisant ces deux premières idées.  
↪ TWI-DLCLUST

# Motivations

Les motivations présentées dans l'article sont claires :

- ➊ Développer un algorithme de codage parcimonieux invariant au décalage temporel.  
↪ TWI-OMP
- ➋ Proposer une méthode d'apprentissage de dictionnaires discriminants par cluster.
- ➌ Développer une approche de clustering utilisant ces deux premières idées.  
↪ TWI-DLCLUST
- ➍ Valider ces méthodes sur des données variées, notamment sur celles de Deezer.

## 1 Introduction

## 2 Méthode

- Formulation mathématique
- Une approche en 3 étapes

## 3 Jeux de données

## 4 Résultats

## 5 Conclusion

# Formalisation mathématique

- On dispose d'un ensemble de séries temporelles  $X = \{x_i\}_{i=1}^N$ .
- On souhaite partitionner cet ensemble en  $K$  clusters  $\{C_\ell\}_{\ell=1}^K$ .

# Formalisation mathématique

- On dispose d'un ensemble de séries temporelles  $X = \{x_i\}_{i=1}^N$ .
- On souhaite partitionner cet ensemble en  $K$  clusters  $\{C_\ell\}_{\ell=1}^K$ .

## Fonction objectif

$$\min_{C, D} \sum_{\ell=1}^K \sum_{x_i \in C_\ell} E(x_i, D_\ell)$$

$$E(x_i, D_\ell) = \min_{\alpha_i^\ell} \|x_i - \mathcal{F}_i(D_\ell) \alpha_i^\ell\|_2^2 \quad \text{avec} \quad \|\alpha_i^\ell\|_0 \leq \tau$$

Avec  $\mathcal{F}_i(D_\ell) = [f_i(d_1^\ell), \dots, f_i(d_{K_\ell}^\ell)]$  où  $f_i(d_j^\ell) = \Delta_{ij}^\ell d_j^\ell$ , et  $\Delta_{ij}$  est une matrice de projection obtenue par warping temporel.

## 1 Introduction

## 2 Méthode

- Formulation mathématique
- Une approche en 3 étapes

## 3 Jeux de données

## 4 Résultats

## 5 Conclusion

# Étape 1 : TWI-OMP

## Idée principale

- **Aligner dynamiquement** les atomes d'un dictionnaire  $D_\ell$  grâce à la transformation  $\mathcal{F}_i(D_\ell)$  qui applique un *warping* temporel  $\Delta_{ij}^\ell$ .  
 $\hookrightarrow$  COSTW
- **Sélectionner itérativement** les atomes alignés qui minimisent l'erreur  $\|x_i - \mathcal{F}_i(D_\ell) \alpha_i^\ell\|_2^2$  (*Orthogonal Matching Pursuit* modifié).
- **Estimer les coefficients**  $\alpha_i^\ell$ .  
 $\hookrightarrow$  À chaque itération :  $\alpha_\Omega = (S_\Omega^T S_\Omega)^{-1} S_\Omega^T x$   
 avec  $S_\Omega = [\Delta_{ij}^\ell d_j^\ell]_{j \in \Omega}$

# Mesure COSTW et lien avec DTW

## Définition du **COSTW**

$$\text{COSTW}(x_i, d_j^\ell) = \cos(\theta(x_i, \Delta_{ij}^\ell d_j^\ell)) = \frac{x_i^T (\Delta_{ij}^\ell d_j^\ell)}{\|x_i\|_2 \|\Delta_{ij}^\ell d_j^\ell\|_2}$$

- Cherche l'*alignement optimal* (via  $\Delta_{ij}^\ell$ ) qui **maximise** la similarité cosinus.
- COSTW se rapproche de la logique DTW (Dynamic Time Warping) pour gérer les variations temporelles, **mais** il s'intègre directement dans le codage parcimonieux (TWI-OMP).

## Étape 2 : Apprentissage de dictionnaires

### Objectif

Minimiser l'erreur de reconstruction  $\mathcal{J}_\ell$  de chaque cluster  $C_\ell$  :

$$\mathcal{J}_\ell = \sum_{x_i \in C_\ell} \left\| x_i - \mathcal{F}_i(D_\ell) \alpha_i^\ell \right\|_2^2$$

## Étape 2 : Apprentissage de dictionnaires

### Objectif

Minimiser l'erreur de reconstruction  $\mathcal{J}_\ell$  de chaque cluster  $C_\ell$  :

$$\mathcal{J}_\ell = \sum_{x_i \in C_\ell} \left\| x_i - \mathcal{F}_i(D_\ell) \alpha_i^\ell \right\|_2^2$$

### Descente de gradient

$$d_j^{\ell(m+1)} = d_j^{\ell(m)} - \eta \frac{\partial \mathcal{J}_\ell}{\partial d_j^{\ell(m)}}$$

$$\text{où } \frac{\partial \mathcal{J}_\ell}{\partial d_j^\ell} = -2 \sum_{x_i \in C_\ell} \alpha_{ij}^\ell \Delta_i^\ell \left( x_i - \mathcal{F}_i(D_\ell) \alpha_i^\ell \right)$$

# Étape 3 : TWI-DLCLUST

## Clustering final

Une unique boucle composée de 2 étapes :

### ① Attribution des séries temporelles :

- ▶ Codage parcimonieux via TWI-OMP pour chaque  $D_\ell$ .
- ▶ Associer chaque série  $x_i$  au cluster minimisant  $E(x_i, D_\ell)$ .

### ② Mise à jour des dictionnaires :

- ▶ Mise à jour des atomes  $d_j^\ell$  par descente de gradient.

## Étape 3 : TWI-DLCLUST

### Clustering final

Une unique boucle composée de 2 étapes :

**① Attribution des séries temporelles :**

- ▶ Codage parcimonieux via TWI-OMP pour chaque  $D_\ell$ .
- ▶ Associer chaque série  $x_i$  au cluster minimisant  $E(x_i, D_\ell)$ .

**② Mise à jour des dictionnaires :**

- ▶ Mise à jour des atomes  $d_j^\ell$  par descente de gradient.

### Critère d'arrêt

L'itération s'arrête lorsque les clusters ne changent plus entre deux itérations consécutives.

- 1 Introduction
- 2 Méthode
- 3 Jeux de données
  - ECG200
  - Trace
- 4 Résultats
- 5 Conclusion

# ECG Dataset

## Caractéristiques :

- Signaux d'électrocardiogrammes.
- 100 séries temporelles de taille 96.
- 2 classes : **battement de cœur normal** ou **infarctus du myocarde**.

# Dataset ECG200

## Pré-traitement :

- Normalisation
- Smoothing

# Dataset ECG200

## Pré-traitement :

- Normalisation
- Smoothing

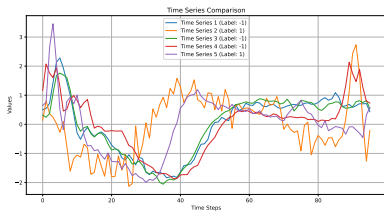


Figure – Avant

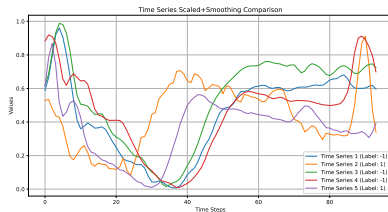


Figure – Après

- 1 Introduction
- 2 Méthode
- 3 Jeux de données
  - ECG200
  - Trace
- 4 Résultats
- 5 Conclusion

# Dataset Trace

## Caractéristiques :

- Jeu de données synthétique conçu pour simuler des défaillances d'instrumentation dans une centrale nucléaire.
- 100 séries temporelles de taille 275.
- 4 classes numérotées de 1 à 4.

# Dataset Trace

## Caractéristiques :

- Jeu de données synthétique conçu pour simuler des défaillances d'instrumentation dans une centrale nucléaire.
- 100 séries temporelles de taille 275.
- 4 classes numérotées de 1 à 4.

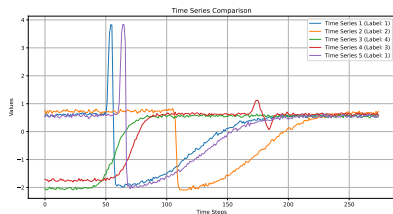


Figure – Avant

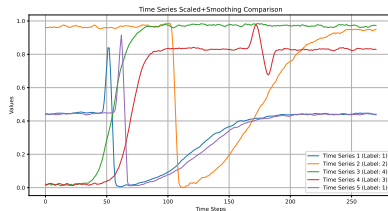


Figure – Après

## 1 Introduction

## 2 Méthode

## 3 Jeux de données

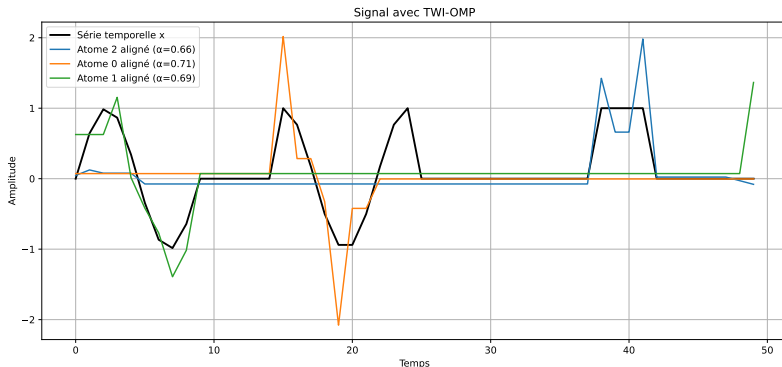
## 4 Résultats

- Résultats pour TWI-OMP
- Résultats pour l'apprentissage de dictionnaire
- Résultats pour TWI-DLCLUST

## 5 Conclusion

- La série temporelle est une suite de 3 motifs précis : 2 sinusoides et une impulsion.
- 3 atomes correspondent aux 3 motifs (avec un peu de bruit).

**Résultat de TWI-OMP :** Les 3 atomes ( $d_j$ ) s'alignent étroitement avec leur motif respectif ( $\Delta_{ij}d_j\alpha_{ij}$ ).



## 1 Introduction

## 2 Méthode

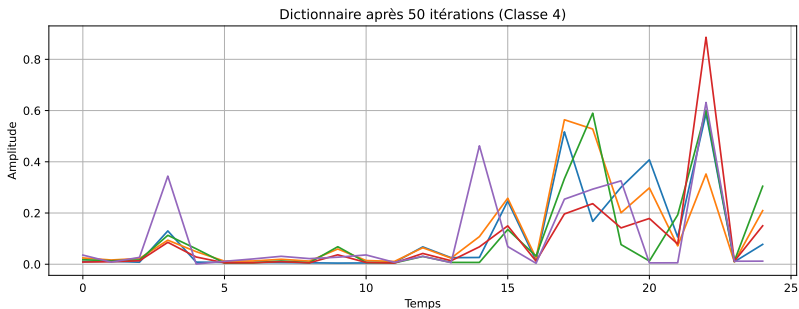
## 3 Jeux de données

## 4 Résultats

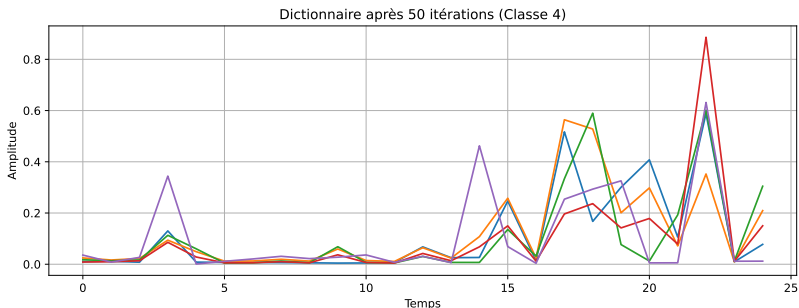
- Résultats pour TWI-OMP
- Résultats pour l'apprentissage de dictionnaire
- Résultats pour TWI-DLCLUST

## 5 Conclusion

Résultat de l'apprentissage de dictionnaire par descente de gradient.  $\eta$  est choisi à 0,01.



Résultat de l'apprentissage de dictionnaire par descente de gradient.  $\eta$  est choisi à 0,01.



**Sûrement une erreur**

C'est très certainement cette partie de l'implémentation qui a provoqué nos mauvais résultats...

## 1 Introduction

## 2 Méthode

## 3 Jeux de données

## 4 Résultats

- Résultats pour TWI-OMP
- Résultats pour l'apprentissage de dictionnaire
- Résultats pour TWI-DLCLUST

## 5 Conclusion

## ECG200

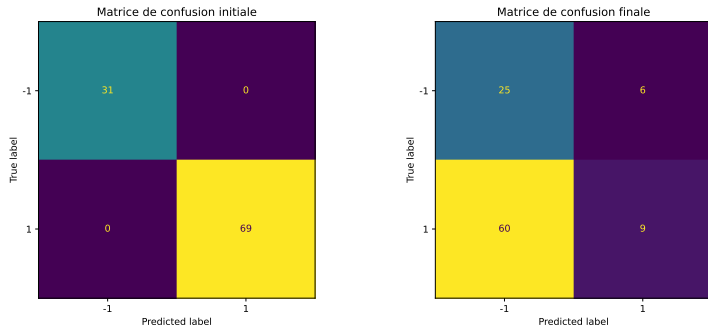


Figure – Matrice de confusion du clustering de ECG200

## ECG200

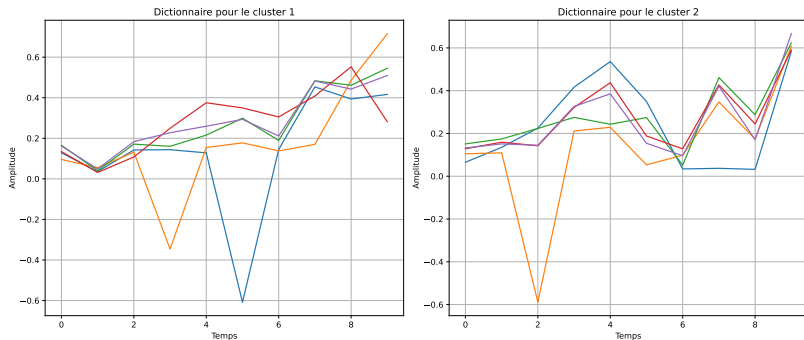


Figure – Dernière mise à jour des dictionnaires pour ECG200

# ECG200

Pour le dataset *ECG200*, on n'observe pas de patterns qui permettrait de différencier facilement chacune des classes, contrairement à *Trace*.

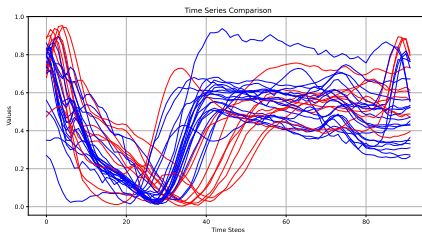


Figure – ECG200

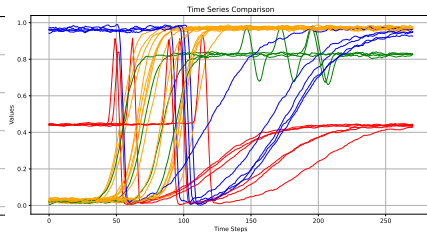


Figure – Trace

# Trace

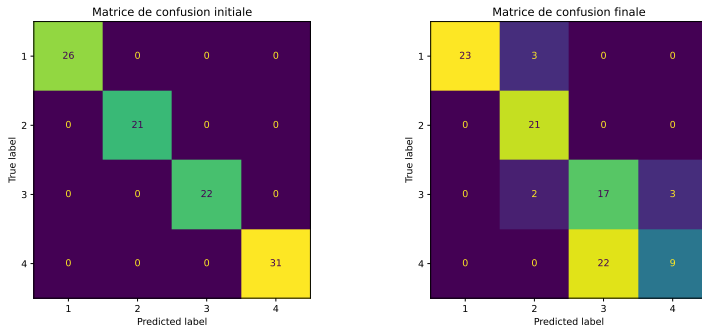


Figure – Matrice de confusion du clustering de Trace

# Trace

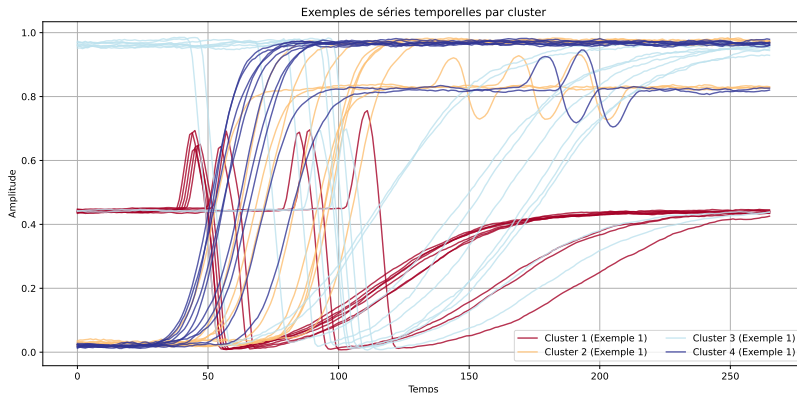


Figure – Clustering effectué sur Trace par TWI-DLCLUST

- 1 Introduction
- 2 Méthode
- 3 Jeux de données
- 4 Résultats
- 5 Conclusion**

# Conclusion

Les résultats principaux :

| Dataset | K (Clusters) | Taille des atomes | KI (Atomes par dic.) | $\tau$ | ARI    |
|---------|--------------|-------------------|----------------------|--------|--------|
| ECG     | 2            | 10                | 5                    | 1      | 0.0294 |
| Trace   | 4            | 25                | 5                    | 3      | 0.5486 |

## Les points à retenir

- Une approche innovante et intéressante.
- Des résultats éloignés des résultats de l'article (entre 0.77 et 0.97, sauf pour le dataset ECG).
- Le problème vient très certainement de l'apprentissage de dictionnaire.