

Apprentissage automatique pour les séries temporelles

Partie II : Représentation par sacs de mots et classification

Yann Cabanes, Johann Faouzi, Romain Tavenard

Tutoriel présenté à PFIA'23 : tslearn-team.github.io/tutoriel-cap2023/

4 juillet 2023

Introduction

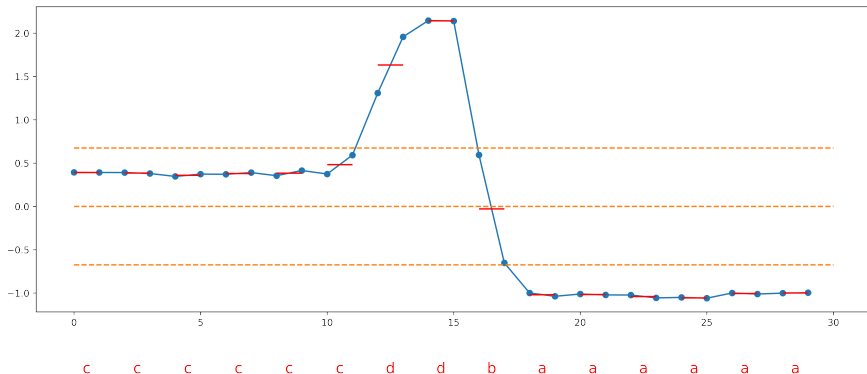
- Approximation : concept courant en apprentissage automatique (sur-apprentissage, généralisation)
- Exemples : arbres de décision et les algorithmes basés dessus (notamment LightGBM).
- Pour les séries temporelles : identifier des motifs et calculer leurs fréquences

Principe général des algorithmes

- 1 Extraction de sous séries temporelles avec une fenêtre glissante.
- 2 Approximation de chaque sous série temporelle en un mot.**
- 3 Calcul de la fréquence de chaque mot pour la série temporelle.
- 4 Classification basée sur cette représentation.**

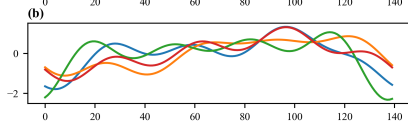
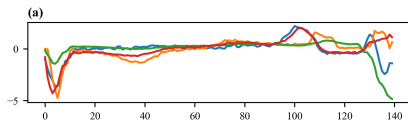
Approximation dans le domaine temporel

Symbolic Aggregate approXimation (SAX)

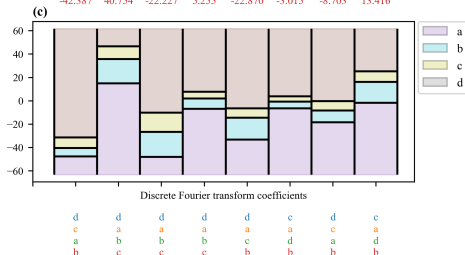


Approximation dans le domaine fréquentiel

Symbolic Fourier Approximation (SFA)

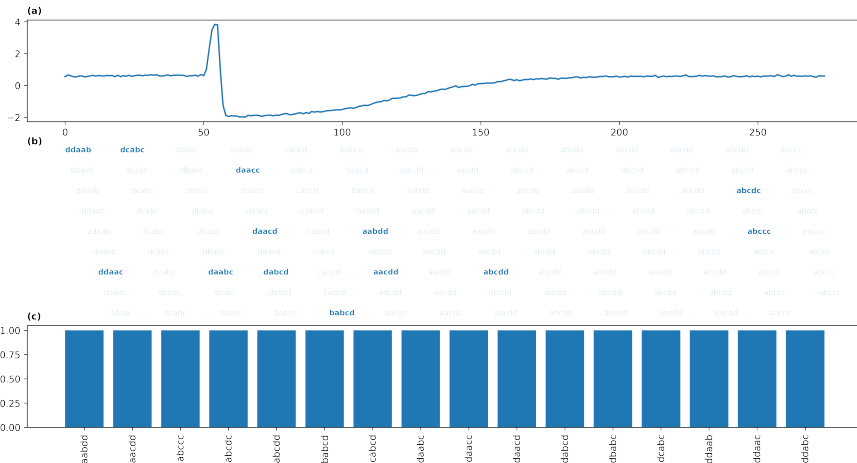


-2.743	64.421	-0.365	18.673	-3.777	0.850	12.030	19.556
-31.732	-53.114	-48.736	-26.040	-35.874	-14.202	-5.741	-6.000
-51.738	32.518	-39.331	-1.874	-8.266	6.170	-19.610	29.268
-42.387	40.754	-22.227	5.255	-22.870	-3.015	-8.703	13.416



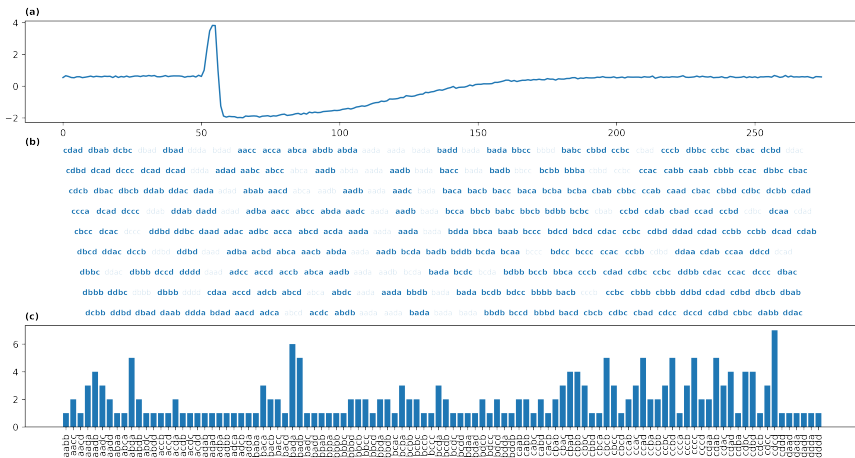
Histogramme des fréquences de mots

Bag of Patterns (BOP)



Histogramme des fréquences de mots

Bag of Symbolic Fourier Approximation Symbols (BOSS)

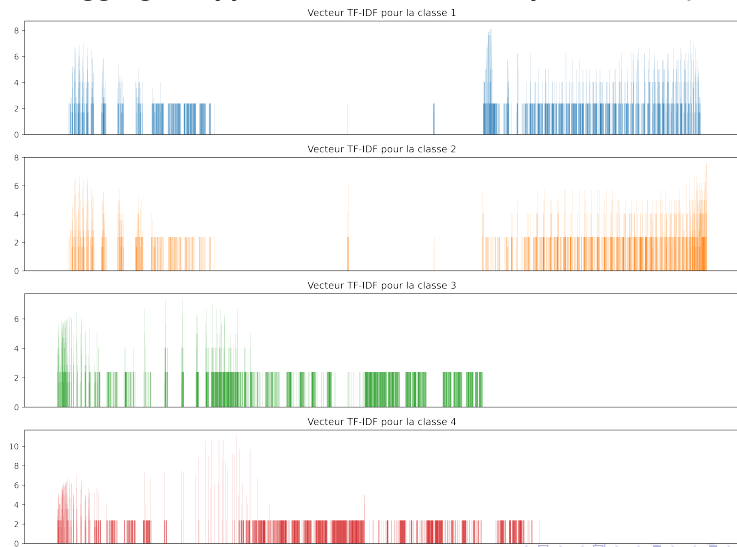


Modèle vectoriel (*vector space model*)

- Idée : représenter chaque classe comme un vecteur de fréquences de mots
- Calcul (entraînement) : *term frequency-inverse document frequency* (TF-IDF)
- Prédiction (inférence) : classe ayant la similarité cosinus la plus élevée entre le vecteur de la série temporelle et le vecteur TF-IDF de chaque classe

Modèle vectoriel

Symbolic Aggregate approXimation in Vector Space Model (SAXVSM)



Modèle vectoriel

Bag of SFA Symbols in Vector Space (BOSSVS)

