

# Apprentissage non supervisé de séries temporelles à l'aide des $k$ -Means et d'une nouvelle méthode d'agrégation de séries

Rémi Gaudin, Nicolas Nicoloyannis

LABORATOIRE ERIC 3038 Université Lumière - Lyon2  
Batiment L 5 av. Pierre Mendès-France 69676 BRON cedex FRANCE  
Remi.Gaudin@univ-lyon2.fr, Nicolas.Nicoloyannis@univ-lyon2.fr

**Résumé.** L'utilisation d'un algorithme d'apprentissage non supervisé de type  $k$ -Means sur un jeu de séries temporelles amène à se poser deux questions : Celle du choix d'une mesure de similarité et celle du choix d'une méthode effectuant l'agrégation de plusieurs séries afin d'en estimer le centre (*i.e.* calculer les  $k$  moyennes). Afin de répondre à la première question, nous présentons dans cet article les principales mesures de similarité existantes puis nous expliquons pourquoi l'une d'entre elles (appelée *Dynamic Time Warping*) nous paraît la plus adaptée à l'apprentissage non supervisé. La deuxième question pose alors problème car nous avons besoin d'une méthode d'agrégation respectant les caractéristiques bien particulières du *Dynamic Time Warping*. Nous pensons que l'association de cette mesure de similarité avec l'agrégation Euclidienne peut générer une perte d'informations importante dans le cadre d'un apprentissage sur la "forme" des séries. Nous proposons donc une méthode originale d'agrégation de séries temporelles, compatible avec le *Dynamic Time Warping*, qui améliore ainsi les résultats obtenus à l'aide de l'algorithme des  $k$ -Means.

**Mots-clés :** Fouille de données et Apprentissage non supervisé, Séries temporelles, K-Means, Dynamic Time Warping

## 1 Introduction

Les séries temporelles sont des données ordonnées dans le temps et cet ordonnancement a une signification que l'on ne peut ignorer. Ainsi, on ne peut pas leur appliquer des méthodes de fouille de données classiques mais bien des méthodes spécialement adaptées, qui respectent la temporalité de ce type de donnée. Nous nous intéresserons ici uniquement à l'apprentissage non supervisé à partir des séries temporelles.

L'utilisation d'un algorithme d'apprentissage non supervisé de type "moyenne mobile" (le plus connu étant les  $k$ -Means) sur un jeu de séries temporelles amène à se poser les questions du choix d'une mesure de distance entre deux séries temporelles et celle du choix d'une méthode effectuant l'agrégation de plusieurs séries temporelles afin d'en estimer le centre (*i.e.* calculer les  $k$  moyennes). Afin de répondre à la première question, nous allons dresser l'état des lieux des principales méthodes de comparaison de séries temporelles déjà existantes (paragraphe 2), puis nous allons discuter l'intérêt de chacune d'entre elles dans le cadre d'un apprentissage non supervisé (paragraphe 2.4).