

# 1d-SAX : une nouvelle représentation symbolique pour les séries temporelles

Simon Malinowski<sup>\*,\*\*\*</sup>, Thomas Guyet<sup>\*,\*\*\*</sup>, René Quiniou<sup>\*\*,\*\*</sup>, Romain Tavenard<sup>\*\*\*\*</sup>

\*AGROCAMPUS-OUEST

\*\*INRIA, Centre de Rennes - Bretagne Atlantique

\*\*\*IRISA - UMR 6074 Campus de Beaulieu, F - 35 042 Rennes Cedex

\*\*\*\*IDIAP Research Institute, Martigny, Switzerland

simon.malinowski@agrocampus-ouest.fr, thomas.guyet@agrocampus-ouest.fr,  
rene.quiniou@inria.fr, romain.tavenard@idiap.ch

**Résumé.** SAX (Symbolic Aggregate approXimation) est une des techniques majeures de symbolisation des séries temporelles. La non prise en compte des tendances dans la symbolisation est une limitation bien connue de SAX. Cet article présente 1d-SAX, une méthode pour représenter une série temporelle par une séquence de symboles contenant des informations sur la moyenne et la tendance des fenêtres successives de la série segmentée. Nous comparons l'efficacité de 1d-SAX vs SAX dans une tâche de classification de séries temporelles d'images satellites. Les résultats montrent que 1d-SAX améliore les taux de classification pour une quantité d'information identique utilisée.

## 1 Introduction

La fouille de séries temporelles (FST) a récemment focalisé l'attention des chercheurs en fouille de données en raison de l'augmentation de la disponibilité de données comportant une dimension temporelle. Les algorithmes de FST tels que la classification/regroupement des séries temporelles, l'extraction de motifs ou la recherche de similarités nécessitent une mesure de distance entre séries temporelles. Le calcul de ces distances repose principalement sur la classique distance euclidienne ou l'alignement temporel dynamique (DTW - Dynamic Time Warping) qui peuvent conduire à des temps de calcul trop importants pour s'attaquer à de longues séries ou à des bases de séries volumineuses. Aussi, de nombreuses représentations approchées des séries temporelles ont émergé au cours de la dernière décennie. La représentation symbolique est une technique pour approximer les séries temporelles. L'algorithme SAX proposé par Lin et al. (2003) est un des plus utilisés pour la symbolisation. C'est une technique très simple, permettant de symboliser par segments les séries temporelles sans nécessiter d'information a priori. Lin et al. (2003) ont montré que SAX possède de bonnes performances pour la FST. Elle ne permet néanmoins pas de prendre en compte les informations de tendance dans les segments de séries temporelles. Plusieurs extensions de la représentation SAX ont été proposées pour pallier ce manque (Esmael et al. (2012); Lkhagva et al. (2006); Zalewski et al. (2012)). Cependant l'information sur la pente dans ces travaux est soit très simplifiée,