

Sélections simultanées de variables et de représentations pour la classification de séries temporelles

Alexis Bondu*, Dominique Gay**, Vincent Lemaire*, Marc Boullé*, Eole Cervenka***

* Orange Labs, Paris, France

alexis.bondu, vincent.lemaire, marc.boulle@orange.com

** LIM-EA2525, Université de La Réunion, France

dominique.gay@univ-reunion.fr

*** Boston Consulting Group, San Francisco, United States

eolecvk@gmail.com

Résumé. Cet article présente une méthode de classification de séries temporelle qui sélectionne des représentations alternatives (telles que les dérivées, les intégrales cumulatives, le spectre de puissance) et en extrait des descripteurs informatifs. L’approche proposée est décomposée en trois étapes : i) les séries temporelles originales sont transformées en plusieurs représentations stockées sous forme de données relationnelles ; ii) ensuite, une méthode de propositionnalisation est appliquée pour “aplatir” les données relationnelles et en extraire des descripteurs ; (iii) enfin, un classificateur Bayésien est appris à partir des descripteurs résultants. Les étapes précédentes sont répétées par un algorithme de sélection de type “forward / backward” pour trouver le sous-ensemble de représentations le plus informatif. L’approche proposée s’avère très compétitive par rapport aux méthodes de l’état de l’art, et extrait des descripteurs interprétables.

1 Introduction

La littérature sur les séries temporelles traite diverses tâches d’apprentissage telles que la prévision, le clustering, etc. Nous abordons le problème de la Classification de Séries Temporelles (CST). Pour une série temporelle univariée notée $\tau_i = \langle (t_1, x_1), (t_2, x_2), \dots, (t_m, x_m) \rangle$ où x_k est la valeur de la série à l’instant t_k , l’objectif est de prédire la valeur l’appartenance de τ_i à l’une des valeurs d’une variable cible catégorielle à partir d’un ensemble de séries temporelles préalablement étiquetées. Un consensus s’est dégagé au sein de la communauté sur le fait que la transformation des séries temporelles du domaine temporel vers un espace de représentation alternatif est l’un des meilleurs moyens pour améliorer la précision des modèles. Selon les auteurs, HIVE-COTE Lines et al. (2018) fournit des résultats quasiment optimaux en termes de précision, et désormais, l’objectif des chercheurs est de développer des approches dont la précision serait comparable, mais qui améliorent d’autres critères comme l’évolutivité, l’interprétabilité, l’automatisation etc. Cela nous a motivés à mettre au point **FEARS** : une nouvelle méthode CST rapide et performante. Cette méthode extrait des descripteurs informatifs tout en sélectionnant simultanément des représentations utiles pour la CST.