

Régularisation de noyaux temporellement élastiques et analyse en composantes principales non-linéaire pour la fouille de séries temporelles

Pierre-François Marteau*

*UMR 6074 IRISA, Université de Bretagne Sud, 56000 Vannes, France
pierre-francois DOT marteau AT irisa DOT fr,
<http://people.irisa.fr/Pierre-Francois.Marteau/>

Résumé. Dans le domaine de la fouille de séries temporelles, plusieurs travaux récents exploitent des noyaux construits à partir de distances élastiques de type Dynamic Time Warping (DTW) au sein d’approches à base de noyaux. Pourtant les matrices, apparentées aux matrices de Gram, construites à partir de ces noyaux n’ont pas toujours les propriétés requises ce qui peut les rendre *in fine* impropres à une telle exploitation. Des approches émergentes de régularisation de noyaux élastiques peuvent être mises à profit pour répondre à cette insuffisance. Nous présentons l’une de ces méthodes, K_{DTW} , pour le noyau DTW, puis, autour d’une analyse en composantes principales non-linéaire (K-PCA), nous évaluons la capacité de quelques noyaux concurrents (élastiques v.s non élastiques, définis v.s. non définis) à séparer les catégories des données analysées tout en proposant une réduction dimensionnelle importante. Cette étude montre expérimentalement l’intérêt d’une régularisation de type K_{DTW} .

1 introduction

Les méthodes à noyau utilisées en analyse exploratoire des données (K-PCA, K-LDA, K-CCA, etc.) ou pour traiter des tâches de classification ou de régression (machines à support vectoriel, SVM) nécessitent, dans leurs fondements, l’usage de noyaux définis (positifs ou négatifs). Pourtant, bon nombre d’études relativement récentes en fouille de données temporelles présentent des résultats produits par de telles méthodes exploitant des noyaux temporellement *élastiques* (NTE) non définis Haasdonk (2005) Zhang et al. (2010) ou régularisés par des méthodes spectrales Narita et al. (2007). L’émergence de nouvelles méthodes de régularisation pour NTE offre aujourd’hui des alternatives à l’exploitation des noyaux élastiques non définis que nous nous proposons d’évaluer de manière comparative sur des jeux de données simples mais potentiellement explicites. D’une manière générale, les procédures de régulation ont été développées pour approximer des noyaux non définis par des noyaux définis (ou semi-définis). Les premières approches appliquent directement des transformations spectrales aux matrices de Gram issues des noyaux non définis. Ces méthodes Wu et al. (2005) Chen et al. (2009) consistent à i) changer le signe des valeurs propres négatives ou décaler ces valeurs propres en utilisant la valeur de décalage minimal nécessaire pour rendre le spectre des valeurs propres