NEW EFFICIENT ALGORITHMS FOR MULTIPLE CHANGE-POINT DETECTION WITH KERNELS

ALAIN CELISSE
Inria Lille Nord Europe – Équipe-projet Inria MODAL & Univ. Lille Sciences et Technologies, CNRS, UMR 8524 – Laboratoire Paul Painlevé
e-mail: alain.celisse@math.univ-lille1.fr

GUILLEMETTE MAROT
Inria Lille Nord Europe – Équipe-projet Inria MODAL & Univ. Lille Droit et Santé – EA 2694 – CERIM
e-mail: guillemette.marot@inria.fr

MORGANE PIERRE-JEAN
LaMME – Univ. d’Evry, UMR CNRS 8071, ENSIIE, USC INRA & Univ. Lille Droit et Santé – EA 2694 – CERIM
e-mail: morgane.pierrejean@genopole.cnrs.fr

GUILLEM RIGAILL
IPS2 – UMR 9213/UMR1403, CNRS, INRA, Univ. P-Sud, Univ. d’Evry & LaMME, Univ. d’Evry, UMR CNRS 8071, ENSIIE, USC INRA
e-mail: guillem.rigaill@inra.fr

Several statistical approaches based on reproducing kernels have been proposed to detect abrupt changes arising in the full distribution of the observations and not only in the mean or variance. Some of these approaches enjoy good statistical properties (oracle inequality, consistent). Nonetheless, they have a high computational cost both in terms of time and memory. This makes their application difficult even for small and medium sample sizes ($n < 10^4$). This computational issue is addressed by first describing a new efficient and exact algorithm for kernel multiple change-point detection with an improved worst-case complexity that is quadratic in time and linear in space. It allows dealing with medium size signals (up to $n \approx 10^5$). Second, a faster but approximate algorithm is described. It is based on a low-rank approximation to the Gram matrix. It is linear in time and space. This approximate algorithm can be applied to large-scale signals ($n \geq 10^6$). These exact and approximate algorithms have been implemented in R and C for various kernels. The computational and statistical performances of these new algorithms have been assessed through empirical experiments. The runtime of the new algorithms is observed to be faster than that of other considered procedures.