

Multi-task learning models for functional data and application to the prediction of sports performances

Arthur Leroy

Abstract:

The present document is dedicated to the analysis of functional data and the definition of multi-task models for regression and clustering. The purpose of this work is twofold and finds its origins in the problem of talent identification in elite sports. This context provides a leading thread illustrative example for the methods and algorithms introduced subsequently while also raising the problem of studying multiple time series, assumed to share information and generally observed on irregular grids. The central method and the associated algorithm developed in this thesis focus on the aspects of functional regression by using multi-task Gaussian processes (GPs) models. This non-parametric probabilistic framework proposes to define a prior distribution on functions, generating data associated with several individuals. Sharing information across those different individuals, through a mean process, offers enhanced modelling compared to a single-task GP, along with a thorough quantification of uncertainty. An extension of this model is then proposed from the definition of a multi-task GPs mixture. Such an approach allows us to extend the assumption of a unique underlying mean process to multiple ones, each being associated with a cluster of individuals. These two methods, respectively called MAGMA_{CLUST} and MAGMA, provide new insights on GP modelling as well as state-of-the-art performances both on prediction and clustering aspects. From the applicative point of view, the analyses focus on the study of performance curves of young swimmers, and preliminary exploration of the real datasets highlights the existence of different progression patterns during the career. Besides, the algorithm MAGMA provides, after training on a dataset, a probabilistic prediction of the future performances for each young swimmer, thus offering a valuable forecasting tool for talent identification. Finally, the extension proposed by MAGMA_{CLUST} allows the automatic construction of clusters of swimmers, according to their similarities in terms of progression patterns, leading once more to enhanced predictions. The methods proposed in this thesis have been entirely implemented and are freely available.

Mots-Clefs : Gaussian Processes, multi-task learning, functional data, curve clustering, EM algorithms, variational inference