Bayesian repulsive mixture model for functional data: contaminating the normal distribution to repel similarly shaped curves

Rosangela Loschi
Universidade Federal de Minas Gerais, Brazil

We introduce a repulsive mixture model for multivariate functional data having as one of the main goal to clustering curves with similar shapes. We assume that more than one functional curve is observed for each individual. The clustering process takes into account both individual-specific information, as we assume covariate-dependent mixture weights, and the similarity of the curve shape. To facilitate the identification of well-differentiated clusters while avoiding redundant clusters in relation to their curve shapes, we propose a repulsive prior distribution for the component-specific location vector of the B-spline coefficients. The proposed prior distribution is a distortion of the multivariate normal distribution, which is multiplied by a functional repulsive factor which depends on a B-spline curve-tailored distance extending existing repulsive priors to the context of functional data.To sample from the posterior distribution, we propose a MCMC algorithm with a split-merge step. The split-merge step significantly improves the mixing of the chain. Different features of the proposed model, including the effects of repulsion and covariates in the clustering, are investigated through simulation. The model is used to identify different movement characteristics in individuals with Chronic Ankle Instability (CAI data). We also evaluate the effect of repulsion on clustering of functional NBA data which has a high noise level.

Joint work with Ricardo Cunha Pedroso (INSPER and Universidade Federal de Minas Gerais-Brazil) and Fernando A. Quintana (Pontificia Universidad Catolica de Chile)