funFEM: an R package for functional data clustering

Charles Bouveyron, Julien Jacques

funFEM: an R package for functional data clustering

C. Bouveyron[a] and **J. Jacques**[b]

[a]Laboratoire MAP5, UMR CNRS 8145
Université Paris Descartes
45 rue des Saints Pères, 75006 Paris, France
charles.bouveyron@parisdescartes.fr

[b]Laboratoire ERIC
Université de Lyon - Lumière
5 av. Pierre Mendès France, 69755 Bron, France
julien.jacques@univ-lyon2.fr

Mots clefs: clustering, functional data.

1 Introduction

A new model-based clustering algorithm for times series (or more generally functional data), called FunFEM, has been proposed in [1]. It is based on a functional mixture model which allows the clustering of the data in a discriminative functional subspace. This model presents the advantage to be parsimonious and to allow the visualization of the clustered systems. This paper presents the funFEM package for R which implements this new clustering algorithm.

2 The discriminative functional mixture model

Let $\{x_1, ..., x_n\}$ be the observed curves we want to cluster. In practice, the functional expressions of these curves are not known and we only have access to the discrete observations $x_i(t_{is})$ at a finite set of ordered times $\{t_{is} : s = 1, \ldots, m_i\}$. The first step in functional data analysis usually consists in recovering the functional nature of data, and for this we assume that the observed curves can be decomposed in a finite basis of function

$$x_i(t) = \sum_{j=1}^{p} \gamma_{ij}\psi_j(t).$$

The functional mixture model proposed in [1] assumes that the basis expansion coefficient $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{ip})^t$ of curve x_i follows a mixture of Gaussians:

$$p(\gamma) = \sum_{k=1}^{K} \pi_k \phi(\gamma; U\mu_k, U^t\Sigma_k U + \Xi), \tag{1}$$

where π_k is the prior probability of the kth group, ϕ is the standard Gaussian density function, U is a $p \times d$ matrix mapping the coefficient γ into the discriminative subspace (of dimension $d < p$), μ_k and Σ_k are the mean vector and covariance matrice (for cluster k) of the mapping of γ into the discriminative subspace, and Ξ the noise covariance matrix.

3 The funFEM package

Model inference is based on an EM-like algorithm, including an additional step between the traditional E and M steps in which the orientation matrix U is updated. This algorithm is implemented in the funFEM package for R, available on the CRAN.
The main function, `funFEM` has only two mandatory arguments: the functional data, defined as a functional object of the fda package, and the number of clusters (or a vector of). The outputs of `funFEM` are: the posterior probabilities and the estimated clusters, the model parameter estimation and several model selection criteria. These latter can be used in order to choose the optimal number of clusters.
The use of `funFEM` is now illustrated on the Velov dataset (available in the funFEM package), which contains one week of loading curves (i.e. the proportion of available bikes) of the bike sharing system stations of Lyon (called Velov). The number of Velov stations in Lyon is 345, and the curves are sampled approximatively each hour. Due to the periodic nature of the curves, a Fourier basis is considered. Below is given the R script used to launch the clustering analysis with `funFEM`.

```
# Load the velov data and smoothing
R> library(funFEM)
R> data(velov)
R> basis <- create.fourier.basis(c(0, 181), nbasis=25)
R> fdobj <- smooth.basis(1:181,t(velov$data),basis)$fd
# Clustrering with FunFEM
R> res = funFEM(fdobj,K=6,model="AkjBk",init="kmeans",lambda=0,disp=TRUE)
```

Figure 1 presents the resulting mean curves of each cluster as well as the geographical repartition of clusters. Such results typically allow us to analyse the bike sharing system of Lyon.

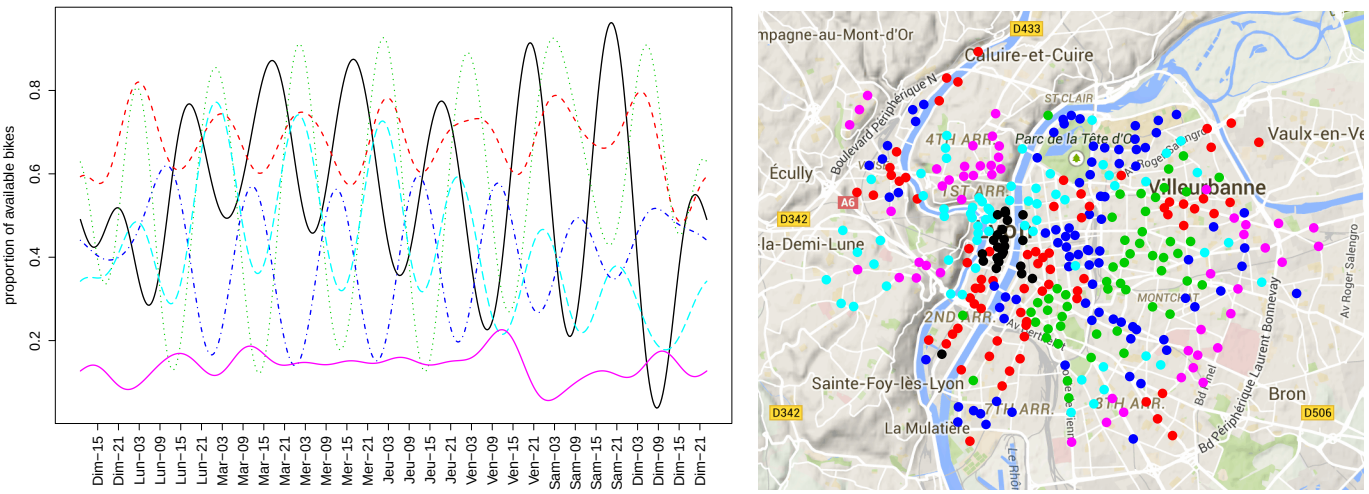

Figure 1: Mean profiles of the 6 clusters (left) and position of clusters in Lyon (right)

References

[1] Bouveyron, C., Côme E., and Jacques, J. (2014). The discriminative functional mixture model for the analysis of bike sharing systems. *Preprint HAL n◦01024186*