Coupled Gaussian Processes for Functional Data Analysis

Processi gaussiani per l'analisi dei dati funzionali

L. Fontanella, S. Fontanella, R. Ignaccolo, L. Ippoliti, P. Valentini

Abstract We present an approach for modelling multivariate dependent functional data. To account for the dominant structural features of the data, we rely on the theory of Gaussian Processes and extend hierarchical dynamic linear models for multivariate time series to the functional data setting. We illustrate the proposed methodology within the framework of bivariate functional data and discuss problems referring to detection of spatial patterns and curve prediction.

Abstract *In questo lavoro, viene presentato un approccio idoneo alla modellazione di dati funzionali multivariati che presentano dipendenza. Per considerare le caratteristiche strutturali dominanti dei dati, ci si avvale della teoria dei processi gaussiani e si considera l'estensione dei modelli lineari dinamici gerarchici per serie storiche nell'ambito dei dati funzionali. La metodologia proposta viene illustrata con riferimento a dati funzionali bivariati. Inoltre, vengono esaminati i problemi connessi all'individuazione di patterns spaziali e previsioni funzionali.*

Key words: Gaussian Processes, Functional data, Simultaneous diagonalization, Derivative process

L. Fontanella
University G. d'Annunzio, Chieti-Pescara, Italy, e-mail: lfontan@unich.it

S. Fontanella
University of Torino, Italy e-mail: sara.fontanella@unito.it

R. Ignaccolo
University of Torino, Italy e-mail: rosaria.ignaccolo@unito.it

L. Ippoliti
University G. d'Annunzio, Chieti-Pescara, Italy e-mail: ippoliti@unich.it

P. Valentini
University G. d'Annunzio, Chieti-Pescara, Italy e-mail: pvalent@unich.it

1 Introduction

Effective statistical modelling under complex designs for functional data is still under development and requires innovative theories. In the following we focus on multivariate *dependent* functional data where the dependence can arise via multiple responses, temporal or spatial effects. In particular, we consider two different problems for bivariate functional data and illustrate the proposed methodology in the frameworks of detection of spatial patterns and curve prediction. Specific applications within these frameworks will be discussed in an extended version of the present article.

2 Detection of spatial patterns using coupled GPs

We focus on the identification of patterns of oscillations considering the *simultaneous orthogonal* expansions of Gaussian Processes (GPs). Orthogonal expansion of GPs has been extensively used for both theoretically investigation and applications. In the case of univariate processes, the theory is based on the probabilistic corollary of Mercer's theorem which is known as Karhunen-Loève expansion. Following Root [2], we consider an extension of this expansion to the case of two kernels that allows the simultaneous orthogonal expansion of two Gaussian processes.
Specifically, we consider the functional data $Y(\mathbf{s},\tau)$, where $\mathbf{s} \in \mathscr{S} \subseteq \mathscr{R}^2$ is a continuous spatial index and $\tau \in \mathscr{T} \subseteq \mathscr{Z}$ is an index of time, as a sample function of either one of two zero-mean Gaussian processes, with spatial covariances $Q_1(\mathbf{s},\mathbf{s}')$ and $Q_2(\mathbf{s},\mathbf{s}')$. Then, it can be shown [2] that the observed process can be expanded in terms of a series of *spatial patterns*, $w_k(\mathbf{s}) = (Q_1^{1/2}u_k)(\mathbf{s})$ as

$$Y(\mathbf{s},\tau) = \sum_{k=1}^{\infty} \alpha_{y,k}(\tau)\, w_k(\mathbf{s}) \tag{1}$$

where $\alpha_{y,k}(\tau)$ is a sequence of independent Gaussian variables and u_k are orthonormalized eigenfunctions in the domain of $Q_1^{-1/2}$.

If only a limited number, K, of patterns are considered, we have a truncated expansion of equation (1), which leads to the following measurement equation

$$\begin{aligned} Y(\mathbf{s},\tau) &= \sum_{k=1}^{K} \alpha_{y,k}(\tau)\, w_k(\mathbf{s}) + \sum_{k=K+1}^{\infty} \alpha_{y,k}(\tau)\, w_k(\mathbf{s}) \\ &= Y^{(K)}(\mathbf{s},\tau) + \mathbf{e}(\mathbf{s},\tau). \end{aligned} \tag{2}$$

Equation (2) allows for extending hierarchical dynamic linear models for multivariate time series to the functional data setting. Moreover, it can be shown that the proposed framework also allows to leverage knowledge from one process when solving an inferential task for another and, accordingly, it forms the motivation for *transfer learning* and prediction of functional data at new sites.

3 Interpolation using derivatives

For the sake of simplicity, consider the problem of representing a one-dimensional real function $Y(\tau)$, $\tau \in \mathscr{T} \subseteq \mathscr{R}$, or estimating its derivative, $X(\tau)$, using only a limited amount of data at points, $\tau_1, \tau_2, \ldots, \tau_n$. This is useful in many applications, including shape analysis, Bayesian optimization and reconstruction of surfaces and signals.

Suppose that all the known values **y** and derivatives **x** are collected into a vector **z**. Let κ be a vector of corresponding indices to show the order of the derivative. For each site τ_j there may be several choices of κ_i if the value of the function and of some of its derivatives are all known at that site. For a bivariate process, $\kappa_i = 0$ if $z(\tau_i) = y(\tau_i)$ is a data value, and $\kappa_i = 1$ if $z(\tau_i) = x(\tau_i)$ is a first derivative.

Denote with $\mathbf{U} = \{u_{im}\}$ the (n, r) matrix of drift terms

$$u_{im} = \frac{\partial^{|\kappa_i|}}{\partial \tau_i^{\kappa_i}}(\tau_i^m), \quad 1 \leq i \leq n, \;\; |m| \leq r,$$

where r is the order of the polynomial drift, and with u_0 the vector of drift terms at a new site τ_0. Assume also that $\Sigma = \{\sigma_{ij}\}$ is a non-singular block covariance matrix where each block has entries given by

$$\sigma_{ij} = (-1)^{|\kappa_j|}\sigma^{(\kappa_i+\kappa_j)}(\tau_i - \tau_j), \quad 1 \leq i, j \leq n$$

and where $\sigma^{(\kappa_i)}$ denotes the partial derivative of $\sigma(\tau_i - \tau_j)$ of order κ_i.

It can be shown that the problem of predicting $Y(t_0)$ at some new points $\tau_0 \in \mathscr{R}$, reduces to find a predictor of the form

$$\hat{Y}(\tau_0) = \sum_{k=1}^{M} a_j u_{j0} + \sum_{i=1}^{n} b_j \sigma(\tau_0, \tau_i). \tag{3}$$

where a_k and b_i are appropriate parameters to be estimated.

Note that this framework allows to use derivative data to predict derivatives [1]. Estimating derivatives and representing the dynamics for functional data is often crucial as they can reveal patterns in a (functional) dataset that address important research questions. The proposed framework also leads to the notion of *derivative principal component analysis*, which complements functional principal component analysis, one of the most popular tools of functional data analysis.

References

1. Mardia K.V., Kent, J. T., Goodall C.R.: Kriging and Splines with Derivative Information. Biometrika **83**, 217–285 (1996)
2. Root W. L.: Singular Gaussian measures in detection theory. In: Proc. Symposium on Time Series Analysis, pp. 292–315. John Wiley, New York (1963)