Journal of Multivariate Analysis 163 (2018) 51–66

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Function-on-function regression with thousands of predictive curves

Xin Qi [a,*], Ruiyan Luo [b]

[a] Department of Mathematics and Statistics, Georgia State University, 30 Pryor Street, Atlanta, GA 30303, United States
[b] Division of Epidemiology and Biostatistics, Georgia State University School of Public Health, 140 Decatur Street SE, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 20 December 2016
Available online 25 October 2017

Keywords:
Function-on-function regression
High-dimension
Simultaneous sparse-smooth penalty

ABSTRACT

With the advance of technology, thousands of curves can be simultaneously recorded by electronic devices, such as simultaneous EEG and fMRI data. To study the relationship between these curves, we consider a functional linear regression model with functional response and functional predictors, where the number of predictive curves is much larger than the sample size. The high dimensionality of this problem poses theoretical and practical difficulties for the existing methods, including estimation inconsistency and prediction inaccuracy. Motivated by the simultaneous EEG and fMRI data, we focus on models with sparsity structures where most of the coefficient functions of the predictive curves have small norms. To take advantage of this sparsity structure and the smoothness of coefficient functions, we propose a simultaneous sparse-smooth penalty which is incorporated into a generalized functional eigenvalue problem to obtain estimates of the model. We establish the asymptotic upper bounds for the prediction and estimation errors as both the sample size and the number of predictive curves go to infinity. We implement the proposed method in the `R` package `FRegSigComp`. Simulation studies show that the proposed method has good predictive performance for models with sparsity structures. The proposed method is applied to a simultaneous EEG and fMRI dataset.

1. Introduction

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) have been widely used to study brain activities in neuroscience. EEG captures the electrical signals generated by nerve cells in the brain through dozens of electrodes placed around the head. An fMRI scan produces a three-dimensional image which can simultaneously record the activation levels of thousands of voxels, i.e., small three-dimensional cuboids in a brain image. In recent years, the simultaneous recording and analysis of EEG and fMRI has received substantial attention [14]. This technology records the EEG and fMRI time series simultaneously while subjects perform tasks.

Motivated by the study of the association between EEG and fMRI data, we consider a functional linear regression model with functional response, e.g., the time series of a particular EEG feature and a high dimensional number of predictive curves, e.g., the time series curves of thousands of voxels in fMRI data. Specifically, we consider the model

$$\forall_{t\in[a,b]} \quad y(t) = \mu(t) + \sum_{j=1}^{p} \int_0^1 x_j(s)\beta_j(s,t)ds + \varepsilon(t), \tag{1}$$

* Corresponding author.
E-mail addresses: xqi3@gsu.edu (X. Qi), rluo@gsu.edu (R. Luo).

https://doi.org/10.1016/j.jmva.2017.10.002
0047-259X/

where $y(t)$ is the response function defined in an interval $[a, b]$ and the predictive curves, $x_1(s), \ldots, x_p(s)$, without loss of generality, are random functions defined in $[0, 1]$. The number p of predictive curves can be much larger than the sample size in practice, and correlation usually exists among predictive functions. The functions $\beta_1(s, t), \ldots, \beta_p(s, t)$ are the p coefficient functions, while $\mu(t)$ and $\varepsilon(t)$ are the intercept function and the noise function, respectively. Within-function correlation can exist in $\varepsilon(t)$.

For models with functional predictors, many studies have been conducted on scalar-on-function regression where the response is scalar, including but not limited to [4,5,7,12,16,24,28,30,31,33,34]. Some works such as [17,20] impose sparsity penalties to select functional predictors. Relatively fewer works have been conducted on function-on-function regression with functional response and functional predictors. The existing methods [6,15,21,22,25,27,31,35,37,39,40] focus either on one or a small number of predictive curves and establish asymptotic results with p fixed. These are not appropriate for Model (1), which has a large number of predictive curves. In theory, since we assume that $p \gg n$ in Model (1), it is more natural to consider what happens when both p and n go to infinity. In practice, the prediction errors of the existing methods can increase fast as p increases. For example, in Simulation 4 in Luo and Qi [21], the prediction errors when $p = 30$ are much larger than those with $p < 10$. Furthermore, heavy computational load prevents the application of these methods to models with a large number of predictive curves.

In the simultaneous EEG-fMRI data, it is believed that a specific brain activity is only related to local regions in the brain, which implies that in Model (1), only a few coefficient functions are nonzero among $\beta_1(s, t), \ldots, \beta_p(s, t)$. Thus, there is a sparsity structure in the model for the simultaneous EEG-fMRI data. In this paper, we consider Model (1) with a more general sparsity structure and assume that most coefficient functions have relatively small L^2 norms.

Specifically, using $\|\boldsymbol{\beta}\|_{L^2,1} = \|\beta_1\|_{L^2} + \cdots + \|\beta_p\|_{L^2}$ to measure the sparsity of the vector $\boldsymbol{\beta}(s, t) = (\beta_1(s, t), \ldots, \beta_p(s, t))^\top$, we assume that $\sqrt{\ln(p)/n} \to 0$ and the sparsity $\|\boldsymbol{\beta}\|_{L^2,1}$ increases more slowly than $\sqrt{n/\ln(p)}$ as $n \to \infty$. Under these assumptions, we aim to estimate the expansion of $\boldsymbol{\beta}(s, t)$ introduced in Luo and Qi [21]. We propose a simultaneous sparse-smooth penalty which is incorporated in a generalized functional eigenvalue problem to obtain the estimates. We propose an efficient algorithm to solve the penalized generalized functional eigenvalue problem with thousands of predictive curves and the method for choosing the number of components and tuning parameters. We establish asymptotic upper bounds for the estimation error and the prediction error. Simulation studies are performed to evaluate the estimation and prediction performance and the effects of penalties. The proposed method is applied to an EEG-fMRI data set.

The rest of the paper is organized as follows. We introduce our method in Section 2 and provide asymptotic results in Section 3. The choice of tuning parameters and the number of components are discussed in Section 4. Simulation studies and an application study are provided in Sections 5 and 6, respectively. We summarize this paper with discussion in Section 7 and provide all the proofs, the details of computations and algorithms, and additional figures and tables in the Online Supplement.

2. Methodology

We first introduce some notations. We use vectors and matrices of functions, and the transpose and multiplication operations on them are the same as those on the usual vectors and matrices. For example, let $\mathbf{A}(t)$ and $\mathbf{C}(s, t)$ denote $q \times p$ and $p \times r$ matrices of functions with (i, j)th entry $a_{ij}(t)$ and (j, k)th entry $c_{jk}(s, t)$, respectively. The multiplication $\mathbf{A}(t)\mathbf{C}(s, t)$ is a $q \times r$ matrix of functions whose (i, k)th entry is the function $a_{i1}(t)c_{1k}(s, t) + \cdots + a_{ip}(t)c_{pk}(s, t)$.

We define the element-wise integrals of vectors and matrices of functions. For example, $\int \mathbf{C}(s, t)ds$ is a $p \times r$ matrix of functions with (j, k)th entry equal to $\int c_{jk}(s, t)ds$ which is a function of t. We use $\|\cdot\|_2$ to denote the usual Euclidean norm of a vector and $\|\cdot\|_{L^2}$ to denote the L^2 norm of a square integrable function. For any vector $\mathbf{d}(s)$ (or $\mathbf{d}(s, t)$) of functions, we define the following L^1-type and L^2-type norms

$$\|\mathbf{d}\|_{L^2,1} = \sum_{j=1}^{p} \|d_j\|_{L^2}, \quad \|\mathbf{d}\|_{L^2,2} = \sqrt{\sum_{j=1}^{p} \|d_j\|_{L^2}^2},$$

respectively. Under these notations, Model (1) can be written as

$$y(t) = \mu(t) + \int_0^1 \mathbf{x}(s)^\top \boldsymbol{\beta}(s, t)ds + \varepsilon(t), \tag{2}$$

where $\mathbf{x}(s) = (x_1(s), \ldots, x_p(s))^\top$ is the p-dimensional vector of predictive functions, $\boldsymbol{\beta}(s, t) = (\beta_1(s, t), \ldots, \beta_p(s, t))^\top$ is the p-dimensional vector of coefficient functions with sparsity structure in that most of its coordinates have small norms. Throughout this paper, we assume that each $\beta_j(s, t)$ is a nonrandom function in $L^2([0, 1] \times [a, b])$ and each $x_j(s)$ is a random function taking values in $L^2[0, 1]$ with $\mathrm{E}\{x_j(s)\} = \mathbf{0}$ for all $j \in \{1, \ldots, p\}$. Furthermore, $\varepsilon(t)$ is a random function taking values in $L^2[a, b]$ with $\mathrm{E}\{\varepsilon(t)\} = \mathbf{0}$, and is independent of $\mathbf{x}(s)$.

We will generalize the method in Luo and Qi [21] to Model (1) with $p \gg n$. We first briefly review the decomposition of $\boldsymbol{\beta}(s, t)$ introduced in Luo and Qi [21] under the notations of this paper in Section 2.1, and then introduce our two-step procedure to estimate the decomposition and hence Model (1) with $p \gg n$ in Section 2.2.