Computational Statistics and Data Analysis 93 (2016) 86–96

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Wavelet-based scalar-on-function finite mixture regression models[☆]

Adam Ciarleglio[a,*], R. Todd Ogden[b]

[a] Department of Child and Adolescent Psychiatry, Division of Biostatistics, New York University, United States

[b] Department of Biostatistics, Columbia University Mailman School of Public Health, United States

ARTICLE INFO

Article history:
Received 15 August 2014
Received in revised form 25 November 2014
Accepted 26 November 2014
Available online 17 December 2014

Keywords:
EM algorithm
Functional data analysis
Lasso
Wavelets

ABSTRACT

Classical finite mixture regression is useful for modeling the relationship between scalar predictors and scalar responses arising from subpopulations defined by the differing associations between those predictors and responses. The classical finite mixture regression model is extended to incorporate functional predictors by taking a wavelet-based approach in which both the functional predictors and the component-specific coefficient functions are represented in terms of an appropriate wavelet basis. By using the wavelet representation of the model, the coefficients corresponding to the functional covariates become the predictors. In this setting, there are typically many more predictors than observations. Hence a lasso-type penalization is employed to simultaneously perform feature selection and estimation. Specification of the model is discussed and a fitting algorithm is provided. The wavelet-based approach is evaluated on synthetic data as well as applied to a real data set from a study of the relationship between cognitive ability and diffusion tensor imaging measures in subjects with multiple sclerosis.

1. Introduction

Let $Y_i \in \mathbb{R}$ be the scalar response of interest for observation i, $i = 1, \ldots, n$ and let X_i be a random predictor process that is square integrable on a compact support $I \subset \mathbb{R}$ (i.e., $\int_I X_i^2(t)dt < \infty$). It is increasingly common to model the relationship between Y_i and X_i via a functional linear model (FLM) given by:

$$Y_i = \alpha + \int_I X_i(t)\omega(t)dt + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where α is the scalar intercept and ε_i is the error term such that $\varepsilon_i \sim N(0, \sigma^2)$. ω is a square integrable coefficient function that relates the predictor process to the response. The magnitude of $\omega(t)$ indicates the relative importance of the predictor X_i at a given value of t. If $|\omega(t_0)|$ is large, this means that changes in the predictor process at t_0 are important in predicting the response. A variety of approaches have been developed for estimating the coefficient function in (1.1) (James, 2002; Cardot et al., 2003; Cardot and Sarda, 2005; Ramsay and Silverman, 2005; James and Silverman, 2005; Cai and Hall, 2006; Reiss and Ogden, 2007; Müller and Yao, 2008; Zhao et al., 2012).

☆ Supplementary materials, including R code for implementing the proposed method and conducting analyses, as well as additional simulation results can be found in the electronic version of this paper (see Appendix A).

* Corresponding author. Tel.: +1 646 754 5463.
E-mail address: Adam.Ciarleglio@nyumc.org (A. Ciarleglio).

http://dx.doi.org/10.1016/j.csda.2014.11.017
0167-9473/

Although (1.1) is adequate for modeling the relationship between a scalar response and a functional predictor when the association between the response and predictor is the same for all observations, it is not appropriate for settings in which the coefficient function differs across subgroups of the observations. If there are C different associations corresponding to C different coefficient functions, then we can think of each observation as coming from one of C distinct subpopulations/components and would need C distinct FLMs to adequately describe the relationship between the response and the predictor. We are concerned with settings in which subpopulation membership is not observed and will need to be estimated along with the component-specific coefficient functions. In order to appropriately model the relationship between X_i and Y_i in this context, we combine finite mixture and functional linear modeling strategies.

Although the underlying theory of finite mixture regression models and methods for estimating those models have been well-studied when the predictors are scalars (McLachlan and Peel, 2000; Schlattmann, 2009), methods for finite mixture regression remain relatively undeveloped when the predictors are functions. To our knowledge, Yao et al. (2011) are the only ones to investigate such an extension. In their approach to functional mixture regression (FMR) models, they first represent each functional predictor in terms of some suitably chosen number of functional principal components and apply standard mixture regression techniques in the new coordinate space.

The FMR model is given by

$$Y_i = \alpha_r + \int_I X_i(t)\omega_r(t)dt + \varepsilon_i \quad \text{if subject } i \text{ belongs to the } r\text{th group,} \tag{1.2}$$

where C is the number of components or distinct subpopulations, α_r is the rth component-specific intercept, and ω_r is the regression function for the rth group, $r = 1, \dots, C$.

In contrast to Yao et al. (2011), we propose to take a wavelet-based approach to FMR models. Our approach is distinct from that of Yao et al. (2011) in several ways. First, just as with using functional principal components, using a wavelet basis initially provides no dimension reduction. However, wavelets are useful for providing sparse representations of functions so that most of the information about the function is contained in relatively few wavelet coefficients. The method that we use to achieve dimension reduction relies on this sparsity property and is very different from simply choosing a small number of principal components that explain some specified proportion of the variance in the predictors. We take a fully functional approach which performs dimension reduction and model estimation simultaneously so that dimension reduction is driven by the relationship between the functional predictor and the response of interest. This is not the case with the functional principal components-based approach. Furthermore, the approach that we present here is flexible in that we can choose from a host of wavelet families to represent the functional predictors and component coefficient functions. Our approach comes with the added cost of needing to select more tuning parameters, but we propose methods for tuning parameter selection that perform well in simulations and we show that we can make substantial gains in estimation accuracy over using functional principal components when the true component coefficient functions are characterized by local features. Finally, it is computationally trivial to extend our approach to higher-dimensional predictors such as 2- or 3-dimensional images. This is due to the fact that there are several software packages available for performing wavelet decompositions for 2- or 3-dimensional data. To our knowledge, there is no readily available software for performing FPCA for such data.

The rest of the paper is organized as follows. In Section 2, we provide a brief discussion of wavelets and the wavelet-based functional linear model followed by specification of the wavelet-based (WB) functional finite mixture regression model. In Section 3, we outline an algorithm for fitting the WB model. Section 4 discusses the various tuning parameters in the WB model. Section 5 presents simulation results showing the performance of the WB method and an application of our method to a real data set. We conclude with a brief discussion in Section 6.

2. Methodology

2.1. Wavelets, wavelet decomposition, and wavelet representation of the FLM

We focus here on the wavelet basis for several reasons. Wavelets are particularly well suited to handle many types of functional data, especially functional data that contain features on multiple scales. They have the ability to adequately represent global and local attributes of functions and can handle discontinuities and rapid changes. Furthermore a large class of functions can be well represented by a wavelet expansion with relatively few non-zero coefficients. This is a desirable property from a computational point of view as it aids in achieving the goal of dimension reduction.

In $L^2(\mathbb{R})$, a wavelet basis is generated by two kinds of functions: a father wavelet, $\phi(t)$, and a mother wavelet, $\psi(t)$, satisfying $\int \phi(t)dt = 1$ and $\int \psi(t)dt = 0$. Here we restrict ourselves to orthonormal wavelet basis families (Daubechies, 1988).

Any particular wavelet basis consists of translated and dilated versions of its father and mother wavelets given by $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ where the integer j is the dilation index referring to the scale and k is an integer that serves as a translation index. These functions can be adapted via implementation of one of several boundary handling schemes that represent a given function on a specified interval. Without loss of generality, we take that interval to be [0, 1]. Hence if we assume that $\omega \in L^2([0, 1])$ then we can represent ω in the wavelet domain by

$$\omega(t) = \sum_{k=0}^{2^{j_0}-1} \beta'_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(t), \tag{2.1}$$