Accepted Manuscript

Sparse estimation for functional semiparametric additive models

Peijun Sang, Richard A. Lockhart, Jiguo Cao

PII: S0047-259X(18)30338-5
DOI: https://doi.org/10.1016/j.jmva.2018.06.010
Reference: YJMVA 4378

To appear in: *Journal of Multivariate Analysis*

Received date : 5 January 2018

Please cite this article as: P. Sang, R.A. Lockhart, J. Cao, Sparse estimation for functional semiparametric additive models, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.06.010

Sparse estimation for functional semiparametric additive models

Peijun Sang, Richard A. Lockhart, Jiguo Cao*

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada V5A1S6

Abstract

We propose a functional semiparametric additive model for the effects of a functional covariate and several scalar covariates and a scalar response. The effect of the functional covariate is modeled nonparametrically, while a linear form is adopted to model the effects of the scalar covariates. This strategy can enhance flexibility in modeling the effect of the functional covariate and maintain interpretability for the effects of scalar covariates simultaneously. We develop the method for estimating the functional semiparametric additive model by smoothing and selecting non-vanishing components for the functional covariate. Asymptotic properties of our method are also established. Two simulation studies are implemented to compare our method with various conventional methods. We demonstrate our method with two real applications.

Keywords: Functional data analysis, Functional linear model, Functional principal component analysis

1. Introduction

High-dimensional data sets of large volume and complex structure are rapidly emerging in various fields. Functional data analysis, due to its great flexibility and wide applications in dealing with high-dimensional data, has received considerable attention. One important problem in functional data analysis is functional linear regression (FLR). One type of FLR models the relationship between a functional covariate and a univariate scalar response of interest. Due to potential lack of fit with FLR models, [24] proposed functional additive models (FAM), in which a scalar response depends on an additive form of the functional principal component (FPC) scores of a functional covariate. A local linear smoother was employed to estimate each component in the additive form and consistency was established for this estimator.

However, in many cases, not only functional covariates but also some scalar covariates may play a role in explaining variation of response. For instance, the Tecator dataset (see Section 4.1 for a more detailed description), which consists of three contents (fat, water, and protein) and 100-channel spectral trajectories of absorbance, has been analyzed with various models, where the response of interest is one of the three contents. Previous studies have focused on regressing the response on the spectral trajectories, which can be viewed as a functional covariate. Zhu et al. [36], for example, employed a regularized functional additive model, where scaled FPC scores are treated as covariates to predict the protein content. However, pairwise scatter plots of the three contents suggest that the other two contents are highly correlated with the protein content as well; thus it may be beneficial to add them into the regression model. In light of this fact, we aim to build a model which can incorporate the effects of both the spectral trajectories and the fat and water contents on the prediction of the protein content.

Motivated by the above example, we propose a functional semiparametric additive model (FSAM) to describe the relationship between a functional covariate, a finite number of scalar covariates and a response variable of interest. In this model, the effect of a functional covariate is represented by its scaled leading FPC scores while scalar covariates are modeled linearly. As a result, this model enables us to acquire flexibility in calibrating the effect of the functional covariate while retaining easy interpretation of the effects of the scalar covariates. There are two main difficulties associated with this new model: the first one is the model estimation and the second concerns theoretical properties.

*Corresponding author. Email: jiguo_cao@sfu.ca