Statistica Sinica **33** (2023), 1-22
doi:https://doi.org/10.5705/ss.202020.0480

OPTIMAL FUNCTION-ON-FUNCTION REGRESSION WITH INTERACTION BETWEEN FUNCTIONAL PREDICTORS

Honghe Jin, Xiaoxiao Sun and Pang Du

University of Georgia, University of Arizona and Virginia Polytechnic Institute and State University

Abstract: We consider a functional regression model in the framework of reproducing kernel Hilbert spaces, where the interaction effect of two functional predictors, as well as their main effects, over the functional response is of interest. The regression component of our model is expressed by one trivariate coefficient function, the functional ANOVA decomposition of which yields the main and interaction effects. The trivariate coefficient function is estimated by optimizing a penalized least squares objective with a roughness penalty on the function estimate. The estimation procedure can be implemented easily using standard numerical tools. Asymptotic results for the proposed model, with or without functional measurement errors, are established under the reproducing kernel Hilbert space (RKHS) framework. Extensive numerical studies show the advantages of the proposed method over existing methods in terms of the prediction and estimation of the coefficient functions. An application to the histone modifications and gene expressions of a liver cancer cell line further demonstrates the better prediction accuracy of the proposed method over that of its competitors.

Key words and phrases: Functional ANOVA, functional interaction, function-on-Function regression, measurement errors, minimax convergence rate, penalized least squares, tensor product.

1. Introduction

Functional regression models, such as scalar-on-function, function-on-scalar, and function-on-function regression models, have attracted much attention (Ramsay and Silverman (2005); Ferraty and Vieu (2006)). In this article, we consider a second-order function-on-function regression model. For $1 \leq i \leq n$, the ith response function $Y_i(\cdot)$ is related to two independent functional predictors $X_i(\cdot)$ and $Z_i(\cdot)$ through

$$Y_i(t) = \int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta(t,r,s)\,dsdr + \epsilon_i(t), \quad t \in I_y. \tag{1.1}$$

Corresponding author: Xiaoxiao Sun, Department of Epidemiology and Biostatistics, The University of Arizona, Tucson, AZ 85724, USA. E-mail: xiaosun@arizona.edu.