ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian functional joint models for multivariate longitudinal and time-to-event data

Kan Li [a], Sheng Luo [b,*]

[a] Merck Research Lab, Merck & Co, 351 North Sumneytown Pike, North Wales, PA 19454, USA
[b] Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2400 Pratt St, 7040 North Pavilion, Durham, NC 27705, USA

ARTICLE INFO

Article history:
Received 30 January 2018
Received in revised form 30 June 2018
Accepted 26 July 2018
Available online xxxx

Keywords:
Longitudinal functional data
Joint modeling
Dynamic prediction
Alzheimer's disease

ABSTRACT

A multivariate functional joint model framework is proposed which enables the repeatedly measured functional outcomes, scalar outcomes, and survival process to be modeled simultaneously while accounting for association among the multiple (functional and scalar) longitudinal and survival processes. This data structure is increasingly common across medical studies of neurodegenerative diseases and is exemplified by the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study, in which serial brain imaging, clinical and neuropsychological assessments are collected to measure the progression of Alzheimer's disease (AD). The proposed functional joint model consists of a longitudinal function-on-scalar submodel, a regular longitudinal submodel, and a survival submodel which allows time-dependent functional and scalar covariates. A Bayesian approach is adopted for parameter estimation and a dynamic prediction framework is introduced for predicting the subjects' future health outcomes and risk of AD conversion. The proposed model is evaluated by a simulation study and is applied to the motivating ADNI study.

1. Introduction

The growing public health threat posed by Alzheimer's disease (AD) has raised the urgency to discover and assess markers for the early detection of the disease. In this regard, a great deal of effort has been dedicated to building models for predicting AD based on a single marker, or a combination of multiple markers, which captures the heterogeneity among subjects and detects the disease progression of subjects at risk (Weiner et al., 2013). Since mild cognitive impairment (MCI) is often considered as a transitional stage to AD, MCI patients are usually enrolled as the target population for early prognosis and evaluating interventions (Petersen et al., 1999). Existing research has identified a number of biomarkers in predicting an individual's likelihood of converting to AD, as well as differences in biomarker values among MCI and AD individuals (Perrin et al., 2009; Schmand et al., 2010). It is widely acknowledged that magnetic resonance imaging (MRI) based measures of atrophy in key brain regions, such as the hippocampus, are predictive of progression from MCI to AD (Du, 2001; Frisoni et al., 2010). Although most of the current studies measure regional atrophy using a single volume-based value, some researchers (Apostolova et al., 2010; Qiu et al., 2009) demonstrated that the surface-based morphology analysis offers more advantages because this method studies patterns of subregion atrophy and produces detailed pointwise correlation between atrophy and cognitive decline. Li and Luo (2017b) proposed a functional joint model (FJM) that incorporates surface-based hippocampus measure as a functional predictor in the joint model of longitudinal and survival framework. They developed a dynamic prediction method and demonstrated that using such a functional predictor, in addition to other scalar markers, improves

* Corresponding author.
E-mail address: sheng.luo@duke.edu (S. Luo).

https://doi.org/10.1016/j.csda.2018.07.015
0167-9473/

predictive performance of the progression of MCI to AD (Li and Luo, 2017a). However, the proposed FJM only accommodates baseline imaging marker as a time-invariant function predictor. Since the imaging markers (e.g., hippocampus) from MRI, along with other neurocognitive markers, are often collected repeatedly in the studies of AD, it is of scientific interest to investigate the combined predictive performance of these repeatedly measured functional and scalar outcomes.

Several methods for the analysis of repeatedly measured functional outcome exist in the literature. One category of the methods is based on functional principal component analysis (FPCA), as well as its extension for multilevel FPCA by Di et al. (2009), longitudinal FPCA by Greven et al. (2010) and by Park and Staicu (2015). These methods modeled subject-specific deviations from a population mean by using low dimensional basis functions estimated from the empirical covariance matrix. However, they were inflexible to estimate the effect of covariates (e.g., age) on the functional outcome. Brumback and Rice (1998) and Guo (2002) proposed a function-on-scalar mixed effect model in which population level effects and individual level deviations were modeled by using penalized splines. Wavelet-based Bayesian functional mixed models were presented in Morris and Carroll (2006), which used a discrete wavelet transform of the observed functional data and modeled coefficients in the wavelet domain. Goldsmith and Kitago (2016) developed a Bayesian framework for penalized spline function-on-scalar regression, allowing the joint modeling of population level fixed effects, individual level random effects, and residual functions. However, these works focused on the statistical inference on longitudinal functional data without considering the survival process and not for prediction purpose.

Joint model is an appropriate framework to modeling longitudinal data and time-to-event data since it has potential to reduce parameter estimate bias, account for dropout in longitudinal studies, and enable the inclusion of longitudinal covariates (both scalar and functional) measured with error in time-to-event models (Tsiatis and Davidian, 2004; Henderson et al., 2000). Multivariate joint models have been well studied by considering multivariate continuous, binary, ordinal, or a mixture of different outcome types. Hickey et al. (2016) gave an excellent review of multivariate joint modeling research. However, no previous study investigates how to incorporate the longitudinal functional (high-dimensional) outcome in a multivariate joint model framework. To this end, we propose a novel joint model that incorporates the growing volume of repeatedly measured functional outcomes in the longitudinal–survival setting. Specifically, we develop a multivariate functional joint model (MFJM) that could simultaneously analyze a longitudinal functional outcome, a longitudinal scalar outcome, and a survival outcome. The principle of the MFJM is to define three type of submodels: (1) a functional mixed effect submodel for the longitudinal functional outcome, (2) a regular mixed effects submodel, or multiple regular mixed effects submodels, to describe the evolution of the longitudinal scalar outcome(s), and (3) a Cox submodel for the survival outcome which is linked with (1) and (2) using a common latent structure. The MFJM is flexible to account for the correlation between repeated measures and correlation among multiple outcomes. We estimate the coefficient functions in the functional regression using penalized spline approach and parameters are jointly estimated in a Bayesian framework.

Compared with the existing literature, we make two major contributions to both multivariate joint modeling and functional data analysis: (1) We propose a multivariate joint model considering both longitudinal functional and scalar outcomes. To the best of our knowledge, this paper is the first to model the repeatedly measured functional outcomes, scalar outcomes, survival process simultaneously while accounting for the associations among the processes. (2) We propose a dynamic prediction framework that provides accurate personalized predictions of disease risk and progression. We investigate the potential capability of the longitudinal functional outcome in improving the prediction of AD progression. Previous studies involving functional data mainly focused on model inference rather than prediction of risk and longitudinal outcome trajectories. These important predictive tools can provide valuable information to monitor each patient's disease progression and to make early decisions about targeted prevention and treatment selection.

The remainder of the article is organized as follows. In Section 2, we describe the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study and the data structure. In Section 3, we discuss the multivariate functional joint model, Bayesian inference procedure, and dynamic prediction framework. In Section 4, we apply the proposed method to the motivating ADNI study. In Section 5, we conduct a simulation study to assess the performance of the method. Concluding remarks and discussion are presented in Section 6.

2. A motivating clinical study

The methodology development is motivated by Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The primary goal of the study is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF) markers, and neuropsychological assessments can be combined to measure the progression of AD. The phase one of the ADNI study (ADNI-1) recruited more than 800 adults, of which about 200 cognitively normal individuals, 400 mild cognitive impairment (MCI) patients, and 200 early AD patients. Participants were reassessed at 6, 12, 18, 24 and 36 months, and additional follow-ups were conducted annually as part of ADNI-2. At each visit, various neuropsychological assessments, brain image, and clinical measures were collected. Detailed information about the ADNI study procedures, including participant inclusion and exclusion criteria and complete study protocol can be found at http://www.adni-info.org.

MCI is commonly considered as a transitional stage between normal cognition and Alzheimer's disease and used as the target population for evaluating prognosis and early treatment. To this end, our analysis focuses on 355 MCI patients in the ADNI-1 study without missing data in covariates of interests, and we consider time from baseline to AD diagnosis among MCI patients to be the survival event of interest. In the ADNI-1 study, the 355 MCI patients were followed up for a mean

Please cite this article in press as: Li K., Luo S., Bayesian functional joint models for multivariate longitudinal and time-to-event data. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.07.015.