Neurocomputing 000 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Kernel-based random effect time-varying coefficient model for longitudinal data

Jooyong Shim [a], Insuk Sohn [b], Changha Hwang [c,*]

[a] Institute of Statistical Information, Department of Statistics, Inje University, Gyeongnam 621-749, South Korea
[b] Biostatistics and Clinical Epidemiology Center, Samsung Medical Center, Seoul 06351, South Korea
[c] Department of Applied Statistics, Dankook University, Gyeonggido 448-160, South Korea

ARTICLE INFO

Article history:
Received 4 October 2016
Revised 5 June 2017
Accepted 9 June 2017
Available online xxx

Communicated by Shiliang Sun

Keywords:
Generalized cross validation
Kernel technique
Least squares support vector machine
Longitudinal data
Model selection
Random effect
Time-varying coefficient model

ABSTRACT

Lots of efforts have been devoted to develop effective estimation methods for parametric and nonparametric longitudinal data models. Varying coefficient regression model has received a great deal of attention as an important tool for modeling the relation between a response and a group of predictor variables. The varying coefficient model is particularly useful in longitudinal data analysis. A random effect time-varying coefficient model is proposed for analyzing longitudinal data, which is based on the basic principle of least squares support vector machine along with the kernel technique. A generalized cross validation method is also considered for choosing the tolerance level and the hyperparameters which affect the performance of the proposed model. The proposed model is evaluated through numerical studies.

1. Introduction

Longitudinal data analysis has received a lot of attention during the last twenty years due to applications in various fields, such as economics, finance, biology, sociology, psychology, medicine and so on. Parametric regression models for longitudinal data have been well developed in the last twenty years [6,9,28]. Nonparametric and semiparametric regression models for longitudinal data using kernel and spline methods have enjoyed substantial developments in the last ten years [7,20]. On the other hand, mixed effects models are becoming increasingly popular for longitudinal data. Mixed effects models constitute both fixed and random effects. Both fixed and random effects may be included for vector of longitudinal covariates. In particular, linear mixed effects models have been widely used [9,27]. The most commonly employed approaches for nonparametric longitudinal data analysis are nonparametric mixed effects models and functional regression models based on principal component analysis through conditional expectation (PACE) [2,18,19,31,33].

Existing nonparametric mixed effects models suffer the drawbacks of not including the effects of covariates with time into the model. In fact, it is not unusual that covariates may depend on time progresses. A promising alternative to overcome this problem is to consider both random effect and time-varying coefficient (TVC) simultaneously as in [23]. The varying coefficient model (VCM) was firstly proposed by [8] and further studied by [10]. Some more VCMs for longitudinal data analysis have been developed and discussed. See, for example, [3,11,21,29,30,32] and [17].

For longitudinal data [23] considered the following random effect TVCM, and presented an estimation procedure for the within subject or cluster correlation structure of the proposed model.

$$y_{ij} = \boldsymbol{\beta}(t_{ij})^t \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_i^t \boldsymbol{z}_{ij} + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n_i, \tag{1}$$

where y_{ij} is the response, $\boldsymbol{x}_{ij} \in R^p$, $\boldsymbol{\beta}(t_{ij}) = (\beta_1(t_{ij}), \ldots, \beta_p(t_{ij}))^t$ is a vector of unknown smooth functions of $t_{ij} \in R$, $\boldsymbol{z}_{ij} \in R^q$ is a vector of covariates associated with random effects, and ϵ_{ij}'s are measurement errors, which are assumed to be i.i.d. with $E(\epsilon_{ij}) = 0$ and $Var(\epsilon_{ij}) = \sigma^2$. Here, $\boldsymbol{\gamma}_i$'s are random effects across the subjects or clusters, which are assumed to be i.i.d. with $E(\boldsymbol{\gamma}_i) = 0$ and $Cov(\boldsymbol{\gamma}_i) = \boldsymbol{\Sigma}$. Furthermore, $\boldsymbol{\gamma}_i$ is independent of ϵ_{ij}.

For longitudinal data [14] considered the following TVCM with fixed effects, and developed a procedure of selecting the significant

* Corresponding author.
E-mail addresses: ds1631@hanmail.net, insuks@gmail.com, chwang@dankook.ac.kr (J. Shim).

http://dx.doi.org/10.1016/j.neucom.2017.06.039
0925-2312/© 2017 Published by Elsevier B.V.

variables.

$$y_{ij} = \boldsymbol{\beta}(t_{ij})^t \boldsymbol{x}_{ij} + \mu_i + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n_i, \tag{2}$$

where μ_i reflects the unobserved individual effect, and ϵ_{ij} is independent of μ_i and $\boldsymbol{x}_{ij} \in R^p$. Here, μ_i is time-invariant and it accounts for the individual's unobserved ability. Model (2) is called as a fixed effect TVCM when μ_i is allowed to be correlated with $\boldsymbol{x}_{ij}$. Model (2) reduces to a random effect TVCM when μ_i is uncorrelated with $\boldsymbol{x}_{ij}$. For identification purpose, we impose the restriction for the fixed effect model that $\sum_{i=1}^m \mu_i = 0$. The usual way to incorporate unobservable variables in a statistical model is via random effects.

For longitudinal data we are going to propose the kernel-based random effect TVCM (KRETVCM) by applying the basic principle of least squares support vector machine (LS-SVM) to the random effect TVCM after preprocessing the time variable t_{ij} by a nonlinear feature mapping function into some feature space. LS-SVM is a least squares version of support vector machine (SVM) and was initially introduced by [24]. The SVM, first developed by [26] and his group at AT&T Bell Laboratories, has been successfully applied to a number of real world problems related to classification and regression problems. LS-SVM has been proved to be a very appealing and promising method [24,25]. There are some strong points of LS-SVM. Here we will consider four of them. The first is that LS-SVM can theoretically achieve a global optimum solution likewise SVM. Thus, it obtains the unique solution. The second is that LS-SVM uses the linear equation which is simple to solve and good for computational time saving. The third is that LS-SVM has good generalization performance likewise SVM. Thus, LS-SVM generally has excellent performance on unseen test data. The fourth is that LS-SVM makes it possible to derive the generalized cross validation (GCV) function which can be utilized efficiently for the model selection. Thus, it is meaningful to compare the proposed KRETVCM with some existing models in terms of fitting and generalization performances.

For the purpose of developing the KRETVCM, we consider the random effect TVCM (1) with $\boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^t$ and $\boldsymbol{\beta}(t_{ij}) = (\beta_0(t_{ij}), \beta_1(t_{ij}), \ldots,\ \beta_p(t_{ij}))^t$. This means that we basically consider the TVCM with the intercept term depending on time as in [10]. The rest of this paper is organized as follows. Section 2 briefly describes the basic principle of LS-SVM and proposes the KRETVCM along with its model selection procedure.Section 3 and Section 4 present numerical studies and conclusion, respectively.

2. KRETVCM for longitudinal data

In this section we review LS-SVM regression and illustrate KRETVCM with estimation and model selection procedures. The underlying idea of KRETVCM is that the true mean specification is approximated by a combination of linear LS-SVM regression and the random effect TVCM based on the time variable t_{ij} preprocessed by a nonlinear feature mapping function.

2.1. LS-SVM regression

The foundations of LS-SVM have been originally proposed by [24]. We now briefly review the standard LS-SVM regression. We basically illustrate the case of the linear LS-SVM regression. Given the training data set $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ with covariate vector $\boldsymbol{x}_i \in \mathbb{R}^p$ and the response $y_i \in \mathbb{R}$, the optimization problem of the linear LS-SVM regression in primal weight space is given as follows:

$$\min_{\boldsymbol{w}, b, \epsilon_i} \mathcal{J} = \frac{1}{2}\boldsymbol{w}^t\boldsymbol{w} + \frac{\lambda}{2}\sum_{i=1}^n \epsilon_i^2 \tag{3}$$

over ($\boldsymbol{w}$, b, ϵ_i) subject to equality constraints

$$y_i = \boldsymbol{w}^t\boldsymbol{x}_i + b + \epsilon_i,\ \ i = 1, \ldots, n \tag{4}$$

with weight vector $\boldsymbol{w} \in R^p$ in primal weight space, bias term b and error variables $\epsilon_i \in R$. The regularization parameter λ is a positive real constant and is considered as a tuning parameter in the algorithm.

The key idea is to construct the primal Lagrange function

$$\mathcal{L} = \mathcal{J} - \sum_{i=1}^n \alpha_i \big(\boldsymbol{w}^t\boldsymbol{x}_i + b + \epsilon_i - y_i\big), \tag{5}$$

where $\alpha_i \geq 0$ are Lagrange multipliers. The conditions for optimality are given by

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{0} \to \boldsymbol{w} = \sum_{i=1}^n \alpha_i \boldsymbol{x}_i \\ &\frac{\partial \mathcal{L}}{\partial b} = 0 \to \sum_{i=1}^n \alpha_i = 0 \\ &\frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \to \boldsymbol{w}^t\boldsymbol{x}_i + b + \epsilon_i - y_i = 0,\ i = 1, \ldots, n \\ &\frac{\partial \mathcal{L}}{\partial \epsilon_i} = 0 \to \alpha_i = \lambda\epsilon_i,\ i = 1, \ldots, n. \end{aligned} \tag{6}$$

After eliminating ϵ_i and $\boldsymbol{w}$, we have the optimal values of α_i and b which are obtained from the linear equation as follows:

$$\begin{pmatrix} 0 & \boldsymbol{1}_n^t \\ \boldsymbol{1}_n & \boldsymbol{\Omega}\lambda^{-1}\boldsymbol{I}_n \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{y} \end{pmatrix}, \tag{7}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^t$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^t$, $\boldsymbol{1}_n$ is the vector of ones of dimension n, $\boldsymbol{I}_n$ is the identity matrix of dimension n and $\boldsymbol{\Omega}$ is the $n \times n$ matrix with elements $\boldsymbol{x}_i^t\boldsymbol{x}_j, i, j = 1, \ldots, n$.

Then, the regression function given $\boldsymbol{x}$ is estimated as follows:

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^n \hat{\alpha}_i \boldsymbol{x}_i^t \boldsymbol{x} + \hat{b}, \tag{8}$$

where $\hat{\alpha}_i$ and $\hat{b}$ are the solutions of the linear Eq. (7).

We now consider how to make the linear LS-SVM regression algorithm nonlinear. This could be achieved by simply preprocessing the input vectors $\boldsymbol{x}_i$ by a nonlinear feature mapping function $\boldsymbol{\phi} : \mathbb{R}^p \to \mathcal{F}$ into some feature space $\mathcal{F}$, and then applying the linear LS-SVM regression algorithm. Thus, we only need to use the kernel trick $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\phi}(\boldsymbol{x}_i)^t\boldsymbol{\phi}(\boldsymbol{x}_j)$ in the Eqs. (7) and (8) associated with the linear LS-SVM regression algorithm [16]. We never need to know explicitly what $\boldsymbol{\phi}$ is. The most popular kernel is Gaussian kernel defined by

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/2\kappa\right),\ i, j = 1, \ldots, n, \tag{9}$$

where $\kappa > 0$ is the kernel parameter.

2.2. Estimation procedure of KRETVCM

Given the training data set $\mathcal{D} = \{(t_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, y_{ij})\}_{i,j=1}^{m,n_i}$, we first consider a random effect TVCM of the form

$$\begin{aligned} y_{ij} &= f(t_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) + \epsilon_{ij} \\ &= \boldsymbol{\beta}(t_{ij})^t\boldsymbol{x}_{ij} + \boldsymbol{\gamma}_i^t\boldsymbol{z}_{ij} + \epsilon_{ij}, i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n_i, \end{aligned} \tag{10}$$

where $\boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^t$, $\boldsymbol{\beta}(t_{ij}) = (\beta_0(t_{ij}), \beta_1(t_{ij}), \ldots, \beta_p(t_{ij}))^t$, $y_{ij} \in R$ is the jth response variable of the ith subject corresponding to $(p+1) \times 1$ fixed effect covariate vector $\boldsymbol{x}_{ij}$ and $q \times 1$ random effect covariate vector $\boldsymbol{z}_{ij}$, with the $q \times 1$ random effect parameter vector $\boldsymbol{\gamma}_i \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $n_i \times 1$ error vector $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^t \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i)$, and $\boldsymbol{\gamma}_i$ is independent of ϵ_{ij}. Here p and q are the numbers of fixed