IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 April 2022, accepted 31 May 2022, date of publication 1 July 2022, date of current version 18 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187767

An Advanced Hybrid Logistic Regression Model for Static and Dynamic Mixed Data Classification

MINGXUE QUAN
School of Mathematics, Renmin University of China, Beijing 100872, China
e-mail: quanmingxue@ruc.edu.cn

This work was supported by the Educational Research Projects of Young and Middle-Aged Teachers in Fujian Education Department under Grant JAT200297.

ABSTRACT We consider the binary classification problem of static and dynamic mixed data in this paper. Different from mixed categorical and numerical data, the dynamic variables in the new type of data vary with time and are recorded at discrete time points. This discrete form results in the high correlation within each variable, at the same time, more shape and dynamic information need to be explored, then an efficient fusion model is urgently needed. To tackle the challenge, we propose a novel fusion method, where the discrete observations from dynamic variables are transformed to continuous functions via basis expansion, and then are combined with static variables via a hybrid logistic regression model, with a group lasso penalty term to select the important features. Consequently, the proposed method makes full use of the correlation and dynamic information, then discards the useless information. It can be regarded as an efficient tool to do the classification. In addition, two numerical examples and a real dataset are utilized to further illustrate the effectiveness of the proposed method.

INDEX TERMS Mixed data, binary classification, variable selection, functional data, functional principal component analysis.

I. INTRODUCTION

The mixed data usually refer to the data including categorical and numerical variables, and they are ubiquitous in many practical applications. There are extensive relevant studies in the literature, among many others, by [13], [23] on classification, by [8], [27] on clustering, and by [12] on regression. Recently, owing to the advancement of technology, some new kinds of mixed data emerge and have received considerable attention. For example, [34] handled the classification of mixed data stream, where the data are not completely collected but come sequentially; [17] considered clustering the data with a mixture of categorical, numerical, sequential, network and other kinds of variables. In this paper, we consider the data with mixture of static and dynamic variables, where the static variables refer to the ones remaining unchanged for a long time or even forever, while dynamic variables vary over time continuously, to some extent, which can be regarded as time series. This type of data is denoted as static and dynamic mixed data, and it can be regarded as an extension of classic mixed data. More details and illustration can be seen in Figure 1.

The static and dynamic mixed data exist in quite a few fields, for a better understand, two examples are displayed as follows. The first example is physical examination data, where height, weight and the indicators in routine blood tests are static, while the data from electrocardiogram (ECG) or electroencephalogram (EEG) are constantly recorded and presented in the curve form. In addition, the data for evaluation of urban comprehensive strength are also representative, where the indicators from economy, politics, technology, ecology and education fields all need to be considered to do the evaluation. Among these indicators, some remain unchanged for a long time such as the greening rate and number of cultural relics, while others like Gross Domestic Product (GDP), personal income and urban population are always changeable.

We consider the binary classification for static and dynamic mixed data in this paper. Since the classification of

The associate editor coordinating the review of this manuscript and approving it for publication was Gianmaria Silvello.

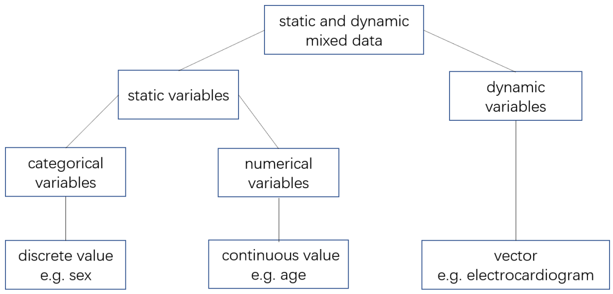


FIGURE 1. The illustration of static and dynamic mixed data. Dynamic variables vary with time while are recorded in discrete time points. Therefore, different from numerical variables, dynamic variables are represented with the form of vector, and the length is determined by sampling frequency and duration.

mixed categorical and numerical data has been well studied, naturally, an extension of the existing methods to the new setting is considered, and the methods for mixed categorical and numerical data are summarized as follows. As we expect, one popular approach is to handle categorical and numerical variables separately, then combine the results to do a comprehensive analysis [32]. In another framework, all variables are treated as one kind. For example, categorical variables are transformed into numerical ones, or numerical variables are treated as categorical by discretization. More statistically significant, some research assume a joint distribution with prior information, then figure out the probability of different classes. Classic models contain conditional Gaussian regression model [36], general mixed-data model [13], [14], and generalized linear mixed model [6], [11], [21].

However, the above approaches cannot achieve a satisfying performance or even cannot work for static and dynamic mixed data classification, since there are two challenges simultaneously. First, dynamic variables continuously vary with time, however in practice, the data are recorded in discrete time points with the measurement error, then two problems appears. On the one hand, as mentioned above, the discrete record makes dynamic variables exist in the form of vector. Consequently, different frequency and duration generate different length of vector, even including high dimensional or high frequency case. On the other hand, the data from the same dynamic variable are highly correlated, which can be called as within-variable correlation, and is similar to within-subject correlation in [45] and [47]. The within-variable correlation contains more shape, dynamic and changing information, which can't be ignored in the classification.

Another main challenge is related with the fusion method of static and dynamic information. If discrete form of dynamic variables are combined with static ones directly, the correlation and dynamic information will be lost. Additionally, high-dimensional discrete dynamic variables will reduce the predictive power of static variables.

A satisfactory recovery and fusion method is urgently required while there are relatively scarce studies, and we notice that most existing methods base on deep learning framework. For example, [33] adopted Hidden Markov Models (HMM) and Long Short-Term Memory (LSTM) while [24] utilized Recurrent Neural Network (ANN) to extract and combine information, then do the classification. The deep learning framework is merely a black box, and overly depends on the data. In this paper, we proposed a fusion model via an efficient tool named functional data.

Functional data [3], [40], [44] refer to the data collected from dynamic and continuous process while only discrete forms are attained owing to the limitations in reality. The same essence makes the functional data become an efficient tool to handle dynamic variables. Furthermore, it is worth mentioning that there are plenty of research focusing on the functional data classification problem [1], [2], [29], [42], to a certain degree, which can be regarded as the classification for dynamic variable. The majority of existing research work focus on the univariate functional data classification, and until recently the multivariate functional data classification begins to receive limited attention. However, in some scenes, it is also not comprehensive because the static information is lost.

In this paper, we propose a complete and systemic framework to classify static and dynamic mixed data. In the proposed method, the dynamic variables are regarded as functional predictors, and static variables are regarded as scalar predictors. We adopt a sequence of data-driven basis functions to recover the functional predictors, then combine the comprehensive information by a hybrid logistic regression model. In addition, a grouping rule is proposed and the corresponding penalty term are used to select important features. Consequently, the useful information is sufficiently captured and extracted, while redundant information is discarded. Two numerical examples and a real dataset in Section III and IV illustrate that the proposed method has a higher classification accuracy and more robust performance.

The paper is structured as follows. In Section II, we formally introduce an advanced hybrid logistic model to classify static and dynamic mixed data. In Section III, the numerical performance of the proposed method is evaluated by two numerical experiments. Besides, a real dataset is adopted to further demonstrate the efficiency in Section IV. Finally, Section V concludes the paper and provides directions for future work.

II. HYBRID LOGISTIC REGRESSION WITH PENALTY TERM

A. PRELIMINARIES

Define static and dynamic mixed data:

$$\{X_1(t), \ldots, X_p(t), Z_1, \ldots, Z_m, Y\},$$

and $X(t) = (X_1(t), \ldots, X_p(t)) \in L^2(\mathcal{T})$ are termed as dynamic variables or functional predictors, where $X_j(t) = (X_j(t_{i1}), \ldots, X_j(t_{il_i}))$ for $j = 1, 2, \ldots, p$. It's because $X_j(t)$ are recorded in discrete time points and are represented in the form of discrete vectors, and it's remarkable that the time points can be regular or irregular. $Z = (Z_1, \ldots, Z_m)$ are termed as static variables or scalar predictors, and Y is a binary response. Our goal is to propose a model to combine mixed predictors and then finish the classification. Note that the domains of $X_j(t), j = 1, 2, \ldots, p$ can be different, while we assume the same domain in this section for simplicity, then we consider the different domains with a real dataset in Section IV.

B. HYBRID LOGISTIC REGRESSION MODEL

Assume the static and dynamic mixed data come from a hybrid logistic regression model:

$$logit(Y) = \sum_{j=1}^{p} \int_T \beta_j(t) X_j(t)\, dt + Z^\top \gamma + \varepsilon, \tag{1}$$

where $\beta(t) = (\beta_1(t), \ldots, \beta_p(t))^\top$ and $\gamma = (\gamma_1, \ldots, \gamma_m)^\top$ are the coefficients of functional predictors and scalar predictors respectively. The intercept term is not contained since the data are zero-centered. ε is the noise term with $\mathbb{E}\varepsilon = 0$ and $\mathbb{E}\varepsilon^2 = \sigma^2 < \infty$. Once the coefficients $\beta(t)$ and γ are estimated, we can use the model to conduct the classification.

C. MODEL TRANSFORMATION

The estimation of $\beta(t)$ is not trivial due to the form of function, for that we adopt the basis expansion method to tackle the challenge. The idea is to expand $X_{i,j}(t)$ and $\beta_j(t)$ with a sequence of basis functions $\phi_1(t), \phi_2(t), \ldots$, specifically,

$$X_{i,j}(t) = \sum_{k=1}^{L_j} \xi_{ijk} \phi_k(t),$$

and

$$\beta_j(t) = \sum_{k=1}^{L_j} b_{jk} \phi_k(t),$$

where $\phi(t)$ can be pre-determined or data-driven, and $L_j, j = 1, 2, \ldots, p$ are the number of bases, which can be determined by 90% cumulative variance contribution rate. In order to make the method more adaptive, we utilize the principle component basis functions, which come from original data. Specifically, for a functional data object $X(t), t \in \mathcal{T}$ with mean function $\mu(t)$ and covariance function $\Sigma_X(s,t) = Cov(X(s), X(t))$, we do the eigen-decomposition to the covariance function:

$$\Sigma_X(s,t) = \mu(t) + \sum_{k=1}^{+\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where $\lambda_k, k = 1, 2, \ldots,$ are ordered eigenvalues, and $\phi_k(t), k = 1, 2, \ldots$ are orthogonal eigenfunctions, also named functional principal component basis.

With the basis expansion method, the discrete record of dynamic variables are transformed into continuous terms, and the dynamic and correlation information are fully utilized. More importantly, we just need to estimate expansion coefficients b_{jk} for $k = 1, 2, \ldots, L_j$ and $j = 1, 2, \ldots, p$ but not the functions $\beta_j(t)$. Consequently, the model (1) turns into:

$$logit(Y_i) = \sum_{j=1}^{p} \sum_{k=1}^{L_j} b_{jk} \xi_{ijk} + Z^\top \gamma. \tag{2}$$

Denote $(b_{11}, \ldots, b_{1L_1}, b_{21}, \ldots, b_{2L_2}, \ldots, b_{p1}, \ldots, b_{pL_p})^\top$ as b, which is a vector of length $L_1 + L_2 + \ldots + L_p$. Besides, define the parameter vector $a = (b, \gamma)^\top$, if a is estimated, the model can be adopted to do the classification.

D. ESTIMATION WITH PENALTY TERM

As mentioned above, we truncate the basis series with first L_j basis functions according to cumulative variance contribution rate, and some redundant information are discarded. However, When the model (1) are adopted to do classification, some functional and scalar predictors may have little contribution to classification. Therefore, we introduce a penalty term to select the important predictors when do the estimation.

We notice that some connections exist in the parameter vector $a = (b, \gamma)^\top$: $(b_{j1}, b_{j2}, \ldots, b_{jL_j})^\top$ is the expansion coefficient of $\beta_j(t)$ for $j = 1, 2, \ldots, p$, then there is a natural group structure existing in the b. Naturally, we put forward a grouping principle: divide $b_{j1}, b_{j2}, \ldots, b_{jL_j}$ into group j, and divide each element in $\gamma = (\gamma_1, \gamma_1, \ldots, \gamma_m)^\top$ into one group, that is, γ_j is in group $p + j$ for $j = 1, 2, \ldots, m$. Finally, the total number of group is $p + m$. With the log likelihood loss function and penalty term, the coefficients a can be estimated by solving the minimization problem,

$$\hat{a} = \min_a \{-\ln L(a) + \lambda \sum_{j=1}^{J} \|a_j\|_2^2\}, \tag{3}$$

where a_j for $j = 1, 2, \ldots, (p + m)$ are the elements in the jth group, and $\ln L(a)$ is defined as $\ln a = \ln L(b, \gamma) = \sum_{i=1}^{n} [Y_i(\sum_{j=1}^{p} \sum_{k=1}^{L_j} \xi_{ijk} b_{jk} + \sum_{j=1}^{m} Z_{ij}\gamma_j) - ln(1 + e^{\sum_{j=1}^{p} \sum_{k=1}^{L_j} \xi_{ijk} b_{jk} + \sum_{j=1}^{m} Z_{ij}\gamma_j})]$. Additionally, λ is a turning parameter to control the penalty term, and it can be selected by K-fold cross validation method. We can adopt gradient descent or expectation maximum methods to do the estimation for the objective function (3).

After attaining the estimator $\hat{a}$, the coefficient functions $\beta_j(t)$ can be estimated by $\hat{\beta}_j(t) = \sum_{k=1}^{L_j} \hat{b}_{jk} \phi_k(t)$, and the model (1) can be used for binary classification. Combining the model 1 and the penalty term, we call the proposed method hybrid logistic model with group penalty (HLR_{group}) method.

III. SIMULATION

A. ILLUSTRATION

To evaluate the performance of the proposed method, two numerical experiments are displayed in this section. Since we are interested in binary classification problem, without loss of generality, two classes are denoted as 0 and 1 with equal amount of observations. Both functional and scalar predictors are contained to predict the class label, and we mainly consider the period and monotonous cases for functional predictors, then normal, uniform and two-point cases for scalar predictors.

Different noise levels and train-test split ratio (abbreviated as 'split ratio' in the following text) are considered to further illustrate the behavior of the proposed method. Specifically, the noise of data is generated with two levels of SNR, namely, SNR = 2 and SNR = 4, where SNR is the signal-to-noise ratio, which is defined by $\mathbb{E}||x - \mu||^2/\mathrm{Var}(\varepsilon)$. Additionally, we divide the data into train and test set under the split ratio of 2:8, 3:7, 5:5, 7:3 and 8:2 to do the stability and robustness analysis.

For the purpose of comparison, we also adopt other approaches to conduct the classification. Two groups of approaches including discrete data-based and functional data-based methods are considered. From the discrete data perspective, we focus on the logistic regression (LR) and related methods in this paper. It's because that the hybrid logistic model (1) can be regarded as the extension of logistic regression, and the comparison is more intuitive. Although the data are discrete, they are highly correlated within the same dynamic variable. We additionally add L_2 and group lasso penalty term but not L_1 penalty to LR model. As a result, Group 1 includes three approaches, namely, logistic regression (LR), logistic regression with L_2 penalty (LR_{L^2}) and logistic regression with group penalty (LR_{group}). Similarly, Group 2 consists of the hybrid logistic regression model (HLR), hybrid logistic regression model with L_2 penalty (HLR_{L^2}) and group lasso penalty (HLR_{group}).

The performance of the method is evaluated by classification accuracy, which is defined as

$$\text{accuracy} = \frac{\sum_{i=1}^{n_{te}} I(\hat{Y}_i == Y_i)}{n_{te}},$$

where Y_i is the true class label while $\hat{Y}_i$ is the estimated one, and n_{te} is the sample size of the test set. All methods run twenty times in R to reduce error.

B. NUMERICAL EXPERIMENT 1

1) DATA DESCRIPTION

The numerical example 1 contains 200 observations. Six functional predictors $X_j(t), j = 1, 2, \ldots, 6$ and three scalar predictors $Z_k, k = 1, 2, 3$ exist in this data set, and six functional predictors are generated from:

$$\begin{aligned}
X_1(t) &= at, \\
X_2(t) &= sin(b_1\pi t) + b_2 sin(b_3\pi t + \frac{b_4}{2}\pi), \\
X_3(t) &= c_1 sin(c_2\pi t), \\
X_4(t) &= |t - 0.5|, \\
X_5(t): dX_t &= (\mu_1 - x_t)dt + \sigma_1 dW_t, \\
X_6(t): dX_t &= (\mu_2 - x_t)dt + \sigma_2 dW_t,
\end{aligned}$$

where W_t is a wiener process. The domains of these six functional predictors are all set to [0, 1], which is then divided equally into 100 discrete time points. The parameters $a \sim N(1, 2.5)$, $b_2 = 0.5$, $c_1 = 0.8$, $\mu_1 = 2$ and $\mu_2 = 1$ are identical for $Y = 0$ and $Y = 1$, and other parameters are set as follows: $b_1 = 2$, $b_3 = 2$, $b_4 = 0$, $c_2 = 6$, $\sigma_1 = 1$, $\sigma_2 = 1$ in $Y = 0$ case; $b_1 = 4$, $b_3 = 4$, $b_4 = 1$, $c_2 = 5$, $\sigma_1 = 3$, $\sigma_2 = 3$ in $Y = 1$ case. Some functional subjects with noise are shown in Figure 2.

Z_1, Z_2, Z_3 are scalar predictors, where Z_1, Z_2 are numerical variables following normal distribution and uniform distribution respectively, and the specific form is presented in (4). Z_3 is a category variable randomly sampled from {0, 1}. The boxplot of the scalar predictors is displayed in Figure 3.

$$Z_1 = \begin{cases} N(40, 1) & Y = 0 \\ N(39, 1) & Y = 1, \end{cases} \quad Z_2 = \begin{cases} U(5, 10) & Y = 0 \\ U(0, 5) & Y = 1. \end{cases} \tag{4}$$

2) RESULTS AND ANALYSIS

The classification results are summarized in table 1, and we notice that:

First, as expected, almost in all settings, the performance of the approaches improve as the noise becomes smaller or the sample size of training set increases.

The approaches in Group 2 perform significantly better than those in Group 1 with different noise levels and split ratios. This is because in the methods of Group 2, the correlation within the functional predictors, as important information, are extracted and combined to do the classification, and it makes classification more accurate and robust.

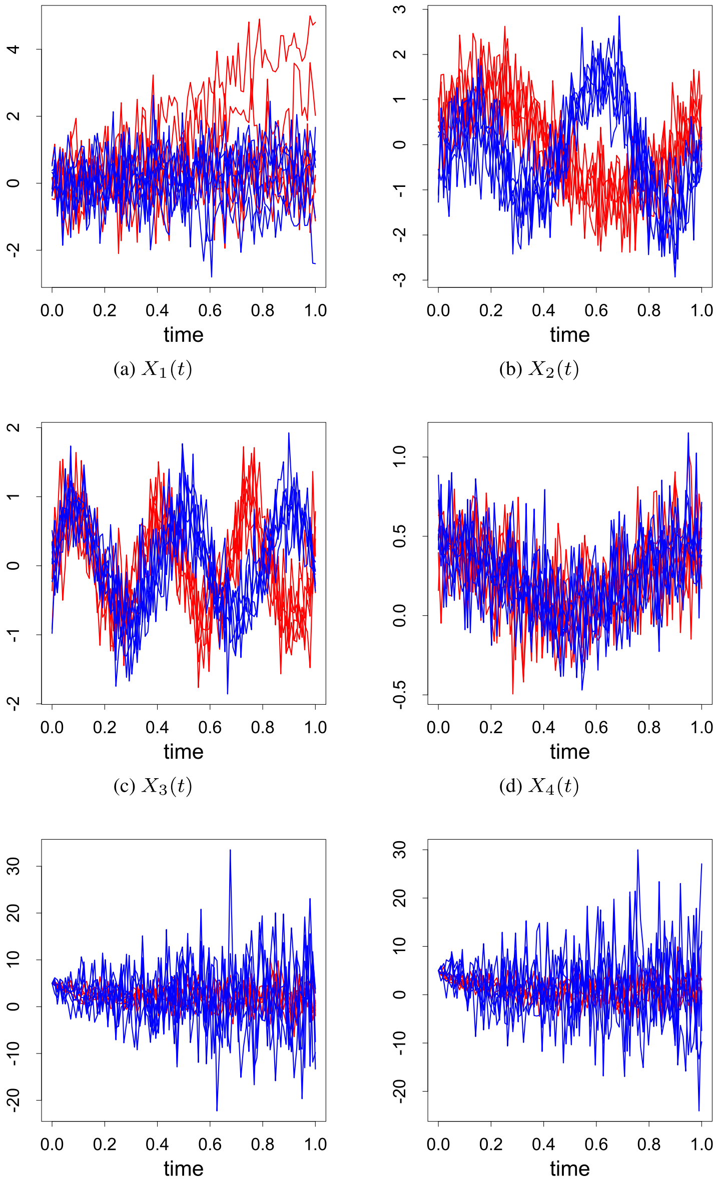
Now, we compare the results of different approaches in fixed noise level and split ratio. In most cases, the proposed method can achieve the highest classification accuracy, while LR_{group} approach outperforms the others in some scenarios, fortunately, where the proposed method can achieve almost the same accuracy. For example, when SNR is fixed to 4 and the split ratio is set to 5:5, 3:7 and 2:8 respectively, the highest accuracy is 98.90%, 98.43% and 97.44%, and they are all achieved by LR_{group} model. While the corresponding accuracy of the proposed method are 98.65%, 98.25% and 97.00%, which are pretty close to the best results.

In order to make the results more intuitive, we transform the contents of table 1 into Figure 4, then the curves of classification results with different split ratios and methods are depicted. According to the trend and comparison of curves, the proposed method almost achieves the highest classification accuracy in different noise levels and split ratios. Additionally, it's noticed that both the penalty term and the tool of functional data can improve the classification accuracy.

C. NUMERICAL EXPERIMENT 2

1) DATA DESCRIPTION

Different from numerical example 1, we extend the domains of functional predictors to [0, 10], and narrow the difference among the parameters in different classes, then classification task becomes more difficult in this numerical experiment.

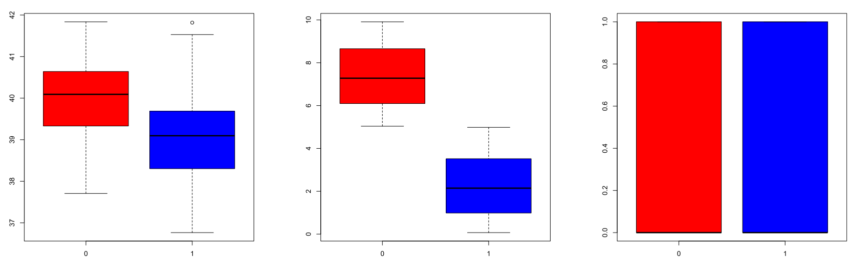


(a) $X_1(t)$ (b) $X_2(t)$

(c) $X_3(t)$ (d) $X_4(t)$

(e) $X_5(t)$ (f) $X_6(t)$

FIGURE 2. Some functional subjects with SNR = 2 noise in the numerical example 1, where color red represents $Y = 0$, while blue represents $Y = 1$.



(a) Z_1 (b) Z_2 (c) Z_3

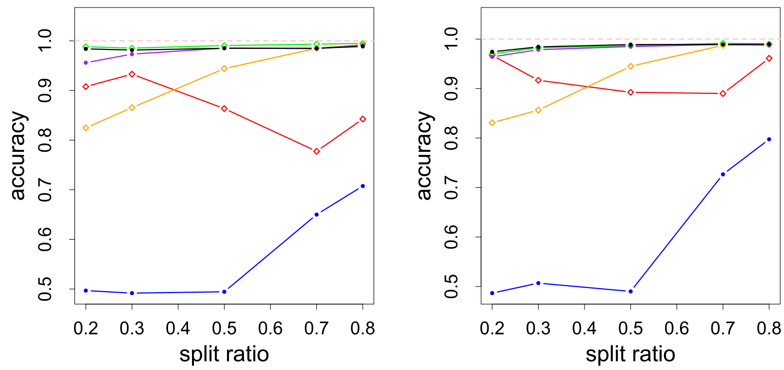
FIGURE 3. Scalar predictors in the numerical example 1, where color red represents $Y = 0$, while blue represents $Y = 1$.

The dataset, where exist 200 observations, includes four functional predictors $X_j(t), j = 1, 2, 3, 4$, and three scalar predictors $Z_k, k = 1, 2, 3$. The functional predictors are generated by

$$X_1(t) = 10 - a_1 t,$$
$$X_2(t) = b_1 sin(b_2 \pi t),$$

TABLE 1. Comparison of average classification accuracy (%) of numerical experiment 1.

	Method	SNR	8:2	7:3	5:5	3:7	2:8
Group 1	LR	SNR=2	70.75	65.00	49.45	49.18	49.69
	LR_{L^2}	SNR=2	99.13	98.42	98.55	97.32	95.59
	LR_{group}	SNR=2	98.88	98.50	98.50	98.14	98.38
Group 2	HLR	SNR=2	84.25	77.75	86.35	93.29	90.78
	HLR_{L^2}	SNR=2	99.25	98.50	94.40	86.54	82.47
	HLR_{group}	SNR=2	**99.50**	**99.33**	**99.05**	**98.54**	**98.81**
Group 1	LR	SNR=4	79.75	72.67	49.00	50.68	48.66
	LR_{L^2}	SNR=4	98.75	98.92	98.50	97.86	96.41
	LR_{group}	SNR=4	**99.00**	98.92	**98.90**	**98.43**	**97.44**
Group 2	HLR	SNR=4	96.13	89.00	89.25	91.68	96.72
	HLR_{L^2}	SNR=4	98.88	98.75	94.50	85.64	83.09
	HLR_{group}	SNR=4	**99.00**	**99.17**	98.65	98.25	97.00



(a) SNR=2 (b) SNR=4

FIGURE 4. The classification accuracy comparison in numerical experiment 1 with the methods of LR (blue), LR_{L^2} (purple) and LR_{group} (black), HLR (red), HLR_{L^2} (orange), HLR_{group} (green) respectively.

$$X_3(t) = sin(c_1\pi t) + cos(c_2\pi t),$$
$$X_4(t) = d_1 t^2,$$

where the domains of functional predictors are all divided into 1000 discrete time points. The parameters $a_1 \sim N(1, 2.5)$, $d_1 \sim N(1, 0.5)$ and $b_1 = 2$ remain unchanged for different classes, and other parameters are set as follows: $b_2 = 2$, $c_1 = 4$, $c_2 = 4$ for $Y = 0$, while $b_2 = 1.5$, $c_1 = 3$, $c_2 = 3$ for $Y = 1$. Some functional subjects with SNR $= 2$ noise are presented in Figure 5. Moreover, the scalar predictors Z_1 and Z_2 follow the distributions described in (5), and Z_3 is randomly sampled from $\{0, 1\}$. The boxplot of scalar predictors is displayed in Figure 6.

$$Z_1 = \begin{cases} N(3,3) & Y = 0 \\ N(2,2) & Y = 1, \end{cases} \quad Z_2 = \begin{cases} U(0,1) & Y = 0 \\ U(0,1) & Y = 1. \end{cases} \tag{5}$$

2) RESULTS AND ANALYSIS

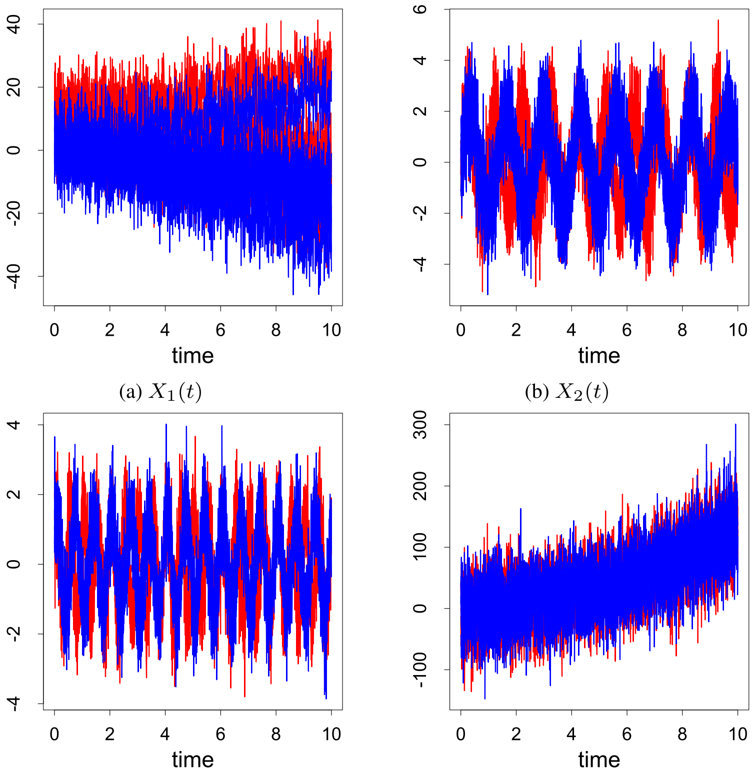
The classification result is summarized in table 2. It's not surprising that there are many similar results with numerical experiment 1. For example, the performance of almost all methods improve with the decrease of noise and increase of split ratio, and the methods in Group 2 can achieve higher classification accuracy. Besides, LR_{L^2} method seems preferred with a high split ratio. The method has a poor performance when there are limited training samples, but as training samples increase, the classification accuracy quickly improves.

Compared with numerical example 1, the observations are recorded for a longer time in this numerical example, and more discrete values of functional predictors are represented. Furthermore, we narrow the difference of parameters, all of these lead to a more difficult classification task. We notice that the proposed method achieves the highest classification accuracy in all cases and has an absolute advantage in this dataset.

D. SUMMARY

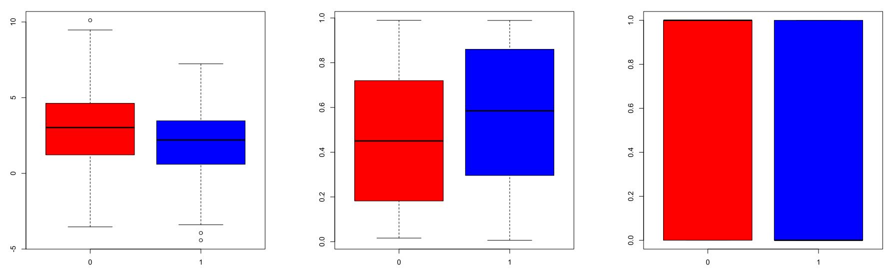
From the results of two numerical examples, the proposed method has a favorable and robust performance, even with large noise and a small split ratio. Although LR_{group} and HLR_{L^2} also have competitive performance in some cases, when we are not familiar with data, the proposed method is a safer option.

The data in these two numerical examples are balanced and the domains of all functional predictors are identical.



(a) $X_1(t)$ (b) $X_2(t)$ (c) $X_3(t)$ (d) $X_4(t)$

FIGURE 5. Functional predictors with SNR = 2 noise in the numerical example 2, where color red represents $Y = 0$, while blue represents $Y = 1$.



(a) $Z_1(t)$ (b) $Z_2(t)$ (c) $Z_3(t)$

FIGURE 6. Scalar predictors in the numerical example 2, where color red represents $Y = 0$, while blue represents $Y = 1$.

However, the datasets in practice present more complex features. As a result, we additionally utilize a real dataset to further illustrate the efficiency of the proposed method in the following section.

IV. APPLICATIONS

A. DATA DESCRIPTION

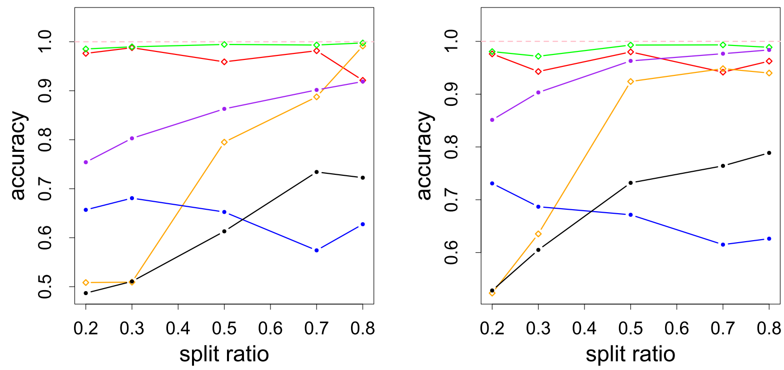
In this section, we applied the methods in Section III to analyze a Diffusion Tensor Imaging (DTI) data set, which was collected at Johns Hopkins University and the Kennedy-Krieger Institute. To better understand the data, we briefly introduce the DTI technology.

Diffusion Tensor Imaging (DTI) is a novel method to describe brain structure, and it is a special form of Magnetic Resonance Imaging (MRI). Similar to tracking the Hydrogen atoms in water molecules at MRI, DTI is based on the direction of water molecules moving, and the effects of brain tumors on nerve cell connections can be revealed. Consequently, it can represent the abnormal changes associated with stroke, multiple sclerosis, schizophrenia, and other brain diseases. It is a useful tool for diseases diagnosis.

The DTI data set in this section records the detailed information of multiple sclerosis patients and persons without multiple sclerosis, in which 255 observations with 26 controls and 229 cases are contained. Every observation contains two functional predictors and five scalar predictors. Two functional predictors are denoted as 'cca' and 'rcst', describing the direction of water molecules moving in the corpus callosum (cca) and right corticospinal tract (rcst) region. Different from the data in Section III, the time domain of these two

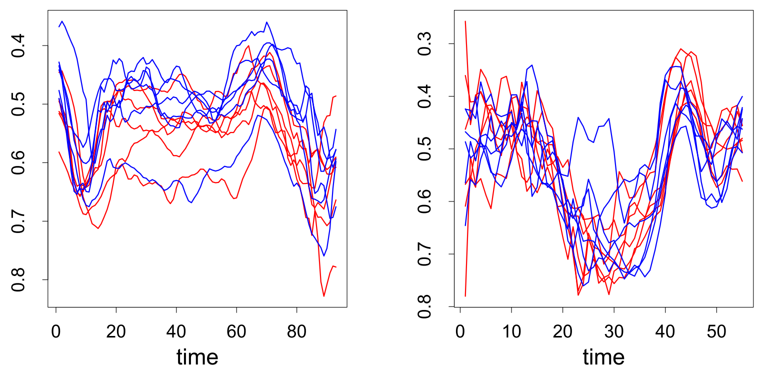
TABLE 2. Comparison of average classification accuracy (%) of numerical experiment 2.

	Method	SNR	8:2	7:3	5:5	3:7	2:8
Group 1	LR	SNR=2	62.75	57.42	65.25	68.07	65.69
	$\mathrm{LR}_{\mathrm{L}^2}$	SNR=2	91.88	90.17	86.30	80.29	75.41
	$\mathrm{LR}_{\mathrm{group}}$	SNR=2	72.25	73.42	61.30	51.07	48.69
Group 2	HLR	SNR=2	92.13	98.17	95.90	98.78	97.63
	$\mathrm{HLR}_{\mathrm{L}^2}$	SNR=2	99.13	88.75	79.50	50.93	50.84
	$\mathrm{HLR}_{\mathrm{group}}$	SNR=2	**99.75**	**99.33**	**99.45**	**98.96**	**98.53**
Group 1	LR	SNR=4	62.63	61.50	67.15	68.68	73.09
	$\mathrm{LR}_{\mathrm{L}^2}$	SNR=4	98.38	97.67	96.30	90.32	85.13
	$\mathrm{LR}_{\mathrm{group}}$	SNR=4	76.88	76.42	73.20	60.50	52.81
Group 2	HLR	SNR=4	96.25	94.17	98.00	94.29	97.63
	$\mathrm{HLR}_{\mathrm{L}^2}$	SNR=4	94.00	94.83	92.40	63.54	52.31
	$\mathrm{HLR}_{\mathrm{group}}$	SNR=4	**98.88**	**99.33**	**99.30**	**97.18**	**98.06**



(a) SNR=2 (b) SNR=4

FIGURE 7. The classification accuracy comparison in numerical experiment 2 with the methods of LR (blue), $\mathrm{LR}_{\mathrm{L}^2}$ (purple) and $\mathrm{LR}_{\mathrm{group}}$ (black), HLR (red), $\mathrm{HLR}_{\mathrm{L}^2}$ (orange), $\mathrm{HLR}_{\mathrm{group}}$ (green) respectively.



(a) cca region (b) rcst region

FIGURE 8. The functional predictors in DTI data, where color red represents the persons without multiple sclerosis, while blue represents patients.

predictors are different, and the scan time is 93s in the cca region while 55s in the rcst region. Five scalar predictors are subject-specific visit numbers (visit), the subject-specific visit time measured in days since the first visit (visit.time), the total number of visits for each subject (Nscans), subject's sex (sex), and the PASAT score at each visit (pasat).

The functional predictors are depicted in Figure 8, and the boxplot of five scalar predictors is displayed in Figure 9. From the above figures, the scalar predictors belonging to different classes are easy to distinguish, however, two functional predictors are hard to distinguish. Unfortunately, the information of scalar predictors takes up a very little proportion, and

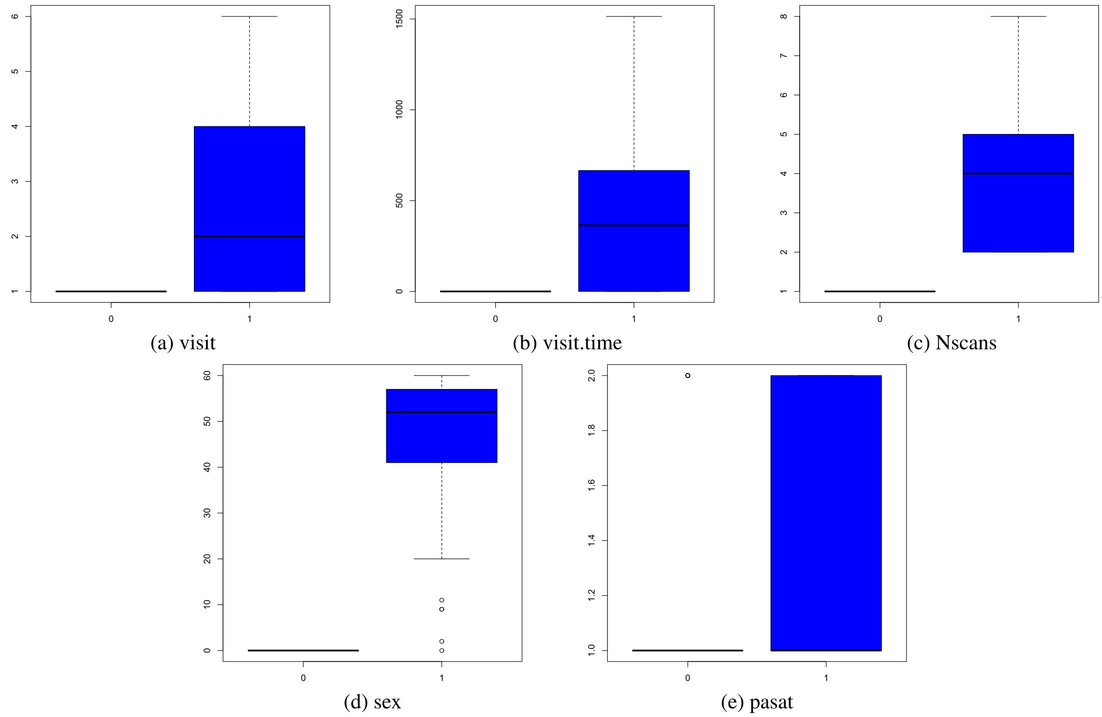


FIGURE 9. **The scalar predictors in DTI data, where color red represents the persons without multiple sclerosis, while blue represents patients.**

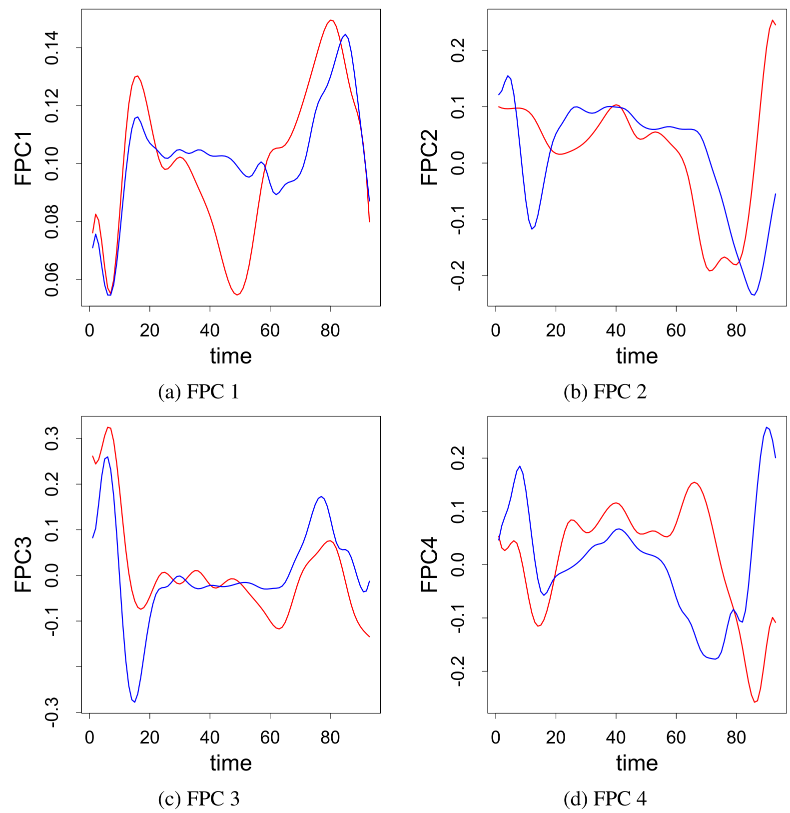
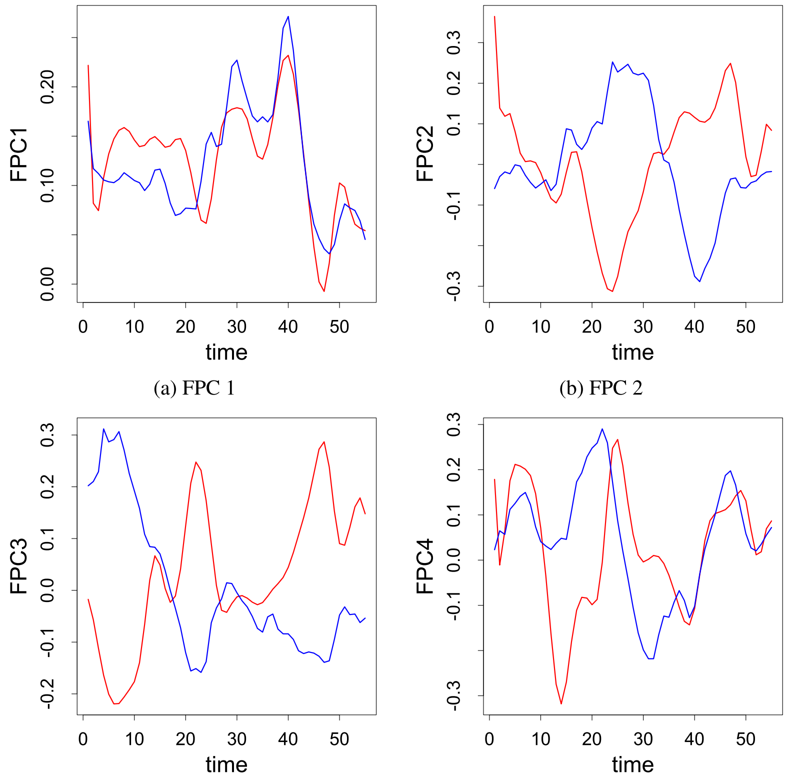


FIGURE 10. **The first four FPCs of the data in cca region, where color red represents the persons without multiple sclerosis, while blue represents patients.**

the discrete date from functional predictors are the main influence factors, the classification thus becomes challenging. Besides, the imbalanced class labels and different domains increase the difficulty of classification.



(a) FPC 1 (b) FPC 2

(c) FPC 3 (d) FPC 4

FIGURE 11. The first four FPCs of the data in rcst region, where color red represents the persons without multiple sclerosis, while blue represents patients.

TABLE 3. Comparison of average classification accuracy of real data.

	Method	8:2	7:3	5:5	3:7	2:8
Group 1	LR	87.94	83.82	59.49	59.78	60.32
	LR_{L^2}	89.41	88.95	89.84	89.80	89.98
	LR_{group}	97.25	97.83	95.59	93.68	92.70
Group 2	HLR	97.16	97.50	94.72	91.15	65.02
	HLR_{L^2}	90.84	91.58	91.11	90.57	90.40
	HLR_{group}	**99.31**	**98.75**	**97.48**	**98.57**	**95.54**

B. RESULTS AND ANALYSIS

Since we adopt the data-driven basis functions in this paper, the functional principal components(FPCs) represent important information in real date analysis. For DTI data, the first four FPCs are extract and depicted in Figure 10 and 11, and their aggregated fractions of variance explained (FVE) exceed 90% and 80% respectively. From the figures, we observe that some FPCs show significant differences between patients and persons without multiple sclerosis, specifically, for FPCs in the cca region, the first FPC of persons without multiple sclerosis presents fluctuation while that of patients maintains relatively stable from 20s to 60s, in contrast, the third FPC of persons without multiple sclerosis maintain stable while that of patients show drastic fluctuations from 10s to 20s. For FPCs in the rcst region, the FPC pairs present an opposite trend from 15s to 20s in the second FPC and from 10s to 20s in the fourth FPC. After making the comparison, we find that the differences between FPC pairs in rcst region are more significant than cca region, and we are supposed to pay more attention to the rcst data, which can help us to explore more information about multiple sclerosis.

Now we consider the classification accuracy, and the results are shown in table 3 and Figure 12. For this imbalanced dataset, the proposed method keeps the best performance in all settings, and the classification accuracy exceeds that of LR_{group} about 2%-3%, and exceeds that of HLR_{L^2} about 6%-8%. These results demonstrate the effectiveness of the proposed method. It's also noticed that the approaches in Group 2 still perform significantly better than those in Group 1, and the models with penalty term surpass the models without penalty, especially with small split ratio.

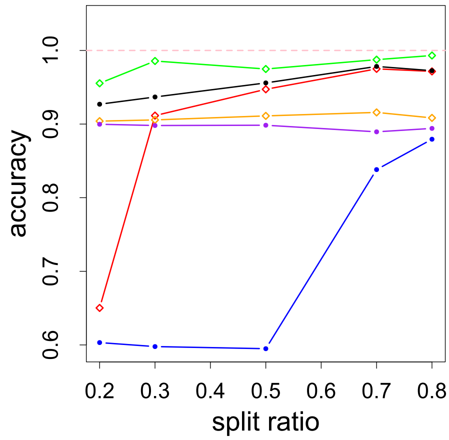


FIGURE 12. The classification accuracy comparison of DTI data with the methods of LR (blue), LR_{L_2} (purple) and LR_{group} (black), HLR (red), HLR_{L_2} (orange), HLR_{group} (green) respectively.

V. CONCLUSION REMARKS

In this paper, we consider static and dynamic mixed data classification problem, and the main challenges are the discrete form and high correlation of dynamic variables. To tackle the challenge, we proposed the hybrid logistic with group penalty (HLR_{group} method), where HLR model is adopted to fully combine static and dynamic predictors, then the group penalty term is used to select important group variables. Consequently, the proposed method utilizes more useful information like the correlation and shape of curve, and discards redundant information, then it has a favorable performance in the classification. The numerical experiments and the comparison with other approaches also demonstrate that the proposed method has a higher classification accuracy and more robust performance, and it seems preferred when the data are imbalanced and the split ratio is small. The classification of static and dynamic data exists in plenty of applications, and we can consider multi-class classification problem and the extension to static and dynamic mixed data stream in the future.

REFERENCES

[1] A. M. Aguilera, M. Escabias, and M. J. Valderrama, "Discussion of different logistic models with functional data. Application to systemic lupus erythematosus," *Comput. Statist. Data Anal.*, vol. 53, no. 1, pp. 151–163, Sep. 2008.

[2] A. Ahmedou, J.-M. Marion, and B. Pumo, "Generalized linear model with functional predictors and their derivatives," *J. Multivariate Anal.*, vol. 146, pp. 313–324, Apr. 2016.

[3] G. Aneiros, R. Cao, R. Fraiman, C. Genest, and P. Vieu, "Recent advances in functional data analysis and high-dimensional statistics," *J. Multivariate Anal.*, vol. 170, pp. 3–9, Mar. 2019.

[4] V. Audigier, F. Husson, and J. Josse, "A principal component method to impute missing values for mixed data," *Adv. Data Anal. Classification*, vol. 10, no. 1, pp. 5–26, 2016.

[5] R. Blanquero, E. Carrizosa, A. Jiménez-Cordero, and B. Martín-Barragán, "Variable selection in classification for multivariate functional data," *Inf. Sci.*, vol. 481, pp. 445–462, May 2019.

[6] B. M. Bolker, M. E. Brooks, C. J. Clark, and S. W. Geange, "Generalized linear mixed models: A practical guide for ecology and evolution," *Trends Ecol. Evol.*, vol. 24, no. 3, pp. 127–135, 2009.

[7] C. Bouveyron and G. Celeux, *Model-Based Clustering and Classification for Data Science: With Applications in R*, vol. 50. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[8] G. Caruso, S. A. Gattone, F. Fortuna, and T. Di Battista, "Cluster analysis for mixed data: An application to credit risk evaluation," *Socio-Econ. Planning Sci.*, vol. 73, Feb. 2021, Art. no. 100850.

[9] J. Chen, X. Lin, Q. Xuan, and Y. Xiang, "FGCH: A fast and grid based clustering algorithm for hybrid data stream," *Int. J. Speech Technol.*, vol. 49, no. 4, pp. 1228–1244, Apr. 2019.

[10] J.-Y. Chen and H.-H. He, "A fast density-based data stream clustering algorithm with cluster centers self-determined for mixed data," *Inf. Sci.*, vol. 345, pp. 271–293, Jun. 2017.

[11] D. G. Clayton, "Generalized linear mixed models," in *Markov Chain Monte Carlo in Practice*, vol. 1. Boca Raton, FL, USA: Chapman & Hall, 1996, pp. 275–302.

[12] C. M. Cuadras and C. Arenas, "A distance based regression model for prediction with mixed data," *Commun. Statist.-Theory Methods*, vol. 19, no. 6, pp. 2261–2279, Jan. 1990.

[13] A. R. de Leon, A. Soo, and T. Williamson, "Classification with discrete and continuous variables via general mixed-data models," *J. Appl. Statist.*, vol. 38, no. 5, pp. 1021–1032, May 2011.

[14] A. R. de Leon and K. C. Carriègre, "General mixed-data model: Extension of general location and grouped continuous models," *Can. J. Statist.*, vol. 35, no. 4, pp. 533–548, Dec. 2007.

[15] A. R. De Leon and K. C. Chough, *Analysis of Mixed Data: Methods & Applications*. Boca Raton, FL, USA: CRC Press, 2013.

[16] M. Denhere and N. Billor, "Robust principal component functional logistic regression," *Commun. Statist.-Simul. Comput.*, vol. 45, no. 1, pp. 264–281, Jan. 2016.

[17] P. D'Urso and R. Massari, "Fuzzy clustering of mixed data," *Inf. Sci.*, vol. 505, pp. 513–534, Dec. 2019.

[18] F. Jianqing, H. Liu, Y. Ning, and H. Zou, "High dimensional semiparametric latent graphical model for mixed data," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 79, no. 2, pp. 405–421, Sep. 2017.

[19] R.-E. Fan and K.-W. Chang, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, 1871–1874, Aug. 2008.

[20] T. Hastie, *Statistical Learning With Sparsity: The Lasso and Generalizations*. Boca Raton, FL, USA: CRC Press, 2015.

[21] D. Hedeker, "Generalized linear mixed models," in *Encyclopedia of Statistics in Behavioral Science*. Wiley, 2005, pp. 1–10.

[22] C.-C. Hsu and Y.-C. Chen, "Mining of mixed data with application to catalog marketing," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 12–23, 2007.

[23] C.-C. Hsu, Y.-P. Huang, and K.-W. Chang, "Extended naive Bayes classifier for mixed data," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1080–1083, Oct. 2008.

[24] T.-C. Hsu, S.-T. Liou, Y.-P. Wang, and Y.-S. Huang, "Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1572–1576.

[25] L. Huang, J. Goldsmith, P. T. Reiss, D. S. Reich, and C. M. Crainiceanu, "Bayesian scalar-on-image regression with application to association between intracranial DTI and cognitive outcomes," *NeuroImage*, vol. 83, pp. 210–223, Dec. 2013.

[26] L. Huang and J. Zhao, "Robust shrinkage estimation and selection for functional multiple linear model through LAD loss," *Comput. Statist. Data Anal.*, vol. 103, pp. 384–400, Nov. 2016.

[27] L. Hunt and M. Jorgensen, "Mixture model clustering for mixed data with missing information," *Comput. Statist. Data Anal.*, vol. 41, nos. 3–4, pp. 429–440, Jan. 2003.

[28] G. M. James, "Generalized linear models with functional predictors," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 64, no. 3, pp. 411–432, Aug. 2002.

[29] C.-R. Jiang and L.-H. Chen, "Filtering-based approaches for functional data classification," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 12, no. 4, p. e1490, 2020.

[30] A. Jiménez-Cordero and S. Maldonado, "Automatic feature scaling and selection for support vector machine classification with functional data," *Appl. Intell.*, vol. 51, pp. 161–184, Aug. 2020.

[31] H. Kim and H. Kim, "Functional logistic regression with fused lasso penalty," *J. Stat. Comput. Simul.*, vol. 88, no. 15, pp. 2982–2999, Oct. 2018.

[32] W. J. Krzanowski, "Distance between populations using mixed continuous and categorical variables," *Biometrika*, vol. 70, no. 1, pp. 235–243, 1983.

[33] A. Leontjeva and I. Kuzovkin, "Combining static and dynamic features for multivariate sequence classification," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2016, pp. 21–30.

[34] Q. Li, Q. Xiong, S. Ji, Y. Yu, C. Wu, and M. Gao, "Incremental semi-supervised extreme learning machine for mixed data stream classification," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115591.

[35] H. Matsui, "Sparse group lasso for multiclass functional logistic regression models," *Commun. Statist.-Simul. Comput.*, vol. 48, no. 6, pp. 1784–1797, Jul. 2019.

[36] M. J. McGeachie, H.-H. Chang, and S. T. Weiss, "CGBayesNets: Conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data," *PLoS Comput. Biol.*, vol. 10, no. 6, Jun. 2014, Art. no. e1003676.

[37] S. N. Mousavi and H. Sørensen, "Multinomial functional regression with wavelets and LASSO penalization," *Econometrics Statist.*, vol. 1, pp. 150–166, Jan. 2017.

[38] S. N. Mousavi and H. Sørensen, "Functional logistic regression: A comparison of three methods," *J. Stat. Comput. Simul.*, vol. 88, no. 2, pp. 250–268, Jan. 2018.

[39] A. J. Onwuegbuzie, J. R. Slate, N. L. Leech, and K. M. Collins, "Mixed data analysis: Advanced integration techniques," *Int. J. Multiple Res. Approaches*, vol. 3, no. 1, pp. 13–33, 2009.

[40] J. O. Ramsay, "Functional data analysis," in *Encyclopedia of Statistical Sciences*. New York, NY, USA: Springer, 2004.

[41] S. J. Ratcliffe, G. Z. Heller, and L. R. Leader, "Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression," *Statist. Med.*, vol. 21, no. 8, pp. 1115–1127, Apr. 2002.

[42] F. Rossi and N. Villa, "Support vector machine for functional data classification," *Neurocomputing*, vol. 69, nos. 7–9, pp. 730–742, Mar. 2006.

[43] E. Tavazzi, S. Daberdaku, R. Vasta, A. Calvo, A. Chiò, and B. Di Camillo, "Exploiting mutual information for the imputation of static and dynamic mixed-type clinical data with an adaptive k-nearest neighbours approach," *BMC Med. Informat. Decis. Making*, vol. 20, no. S5, pp. 1–23, Aug. 2020.

[44] J.-L. Wang, J.-M. Chiou, and H.-G. Müller, "Functional data analysis," *Annual Rev. Stat. Appl.*, vol. 3, pp. 257–295, Jun. 2016.

[45] N. Wang, "Marginal nonparametric kernel regression accounting for within-subject correlation," *Biometrika*, vol. 90, no. 1, pp. 43–52, Mar. 2003.

[46] Y. Yang and H. Zou, "A fast unified algorithm for solving group-lasso penalize learning problems," *Statist. Comput.*, vol. 25, no. 6, pp. 1129–1141, Nov. 2015.

[47] B. Yao, L. Wang, and X. He, "Semiparametric regression analysis of panel count data allowing for within-subject correlation," *Comput. Statist. Data Anal.*, vol. 97, pp. 47–59, May 2016.

[48] M. Yinfeng and L. Jiye, "Linear regularized functional logistic model," *J. Comput. Res. Develop.*, vol. 57, no. 8, p. 1617, 2020.

[49] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[50] Y.-C. Zhang and L. Sakhanenko, "The naive Bayes classifier for functional data," *Statist. Probab. Lett.*, vol. 152, pp. 137–146, Sep. 2019.

MINGXUE QUAN was born in November, in 1995, Hebei, China. She received the B.S. degree in applied statistics from the Hebei University of Technology, in 2018. She is currently pursuing the Ph.D. degree in probability and statistics with the Renmin University of China.

She has published two papers, which are Some Strong Deviation Theorems of k-ordered Non-homogeneous Markov Chains on a Non-Homogenous Tree and Many Strong Limit Theorems for Tree Indexed m-ordered Non-Homogenous Markov Chains. Her research interests include statistical modeling and machine learning algorithm.