AN OUTLYINGNESS MATRIX FOR MULTIVARIATE FUNCTIONAL DATA CLASSIFICATION

Wenlin Dai and Marc G. Genton

King Abdullah University of Science and Technology

Abstract: The classification of multivariate functional data is an important task in scientific research. Unlike point-wise data, functional data are usually classified by their shapes rather than by their scales. We define an outlyingness matrix by extending directional outlyingness, an effective measure of the shape variation of curves that combines the direction of outlyingness with conventional statistical depth. We propose classifiers based on directional outlyingness and the outlyingness matrix. Our classifiers provide better performance compared with existing depth-based classifiers when applied on both univariate and multivariate functional data from simulation studies. We also test our methods on two data problems: speech recognition and gesture classification, and obtain results that are consistent with the findings from the simulated data.

Key words and phrases: Directional outlyingness, functional data classification, multivariate functional data, outlyingness matrix, statistical depth.

1. Introduction

Functional data are frequently collected by researchers in such fields as biology, finance, geology, medicine, and meteorology. As with other types of data, problems such as ranking, registration, outlier detection, classification, and modeling also arise with functional data. Many methods have been proposed to extract useful information from functional data (Ramsay and Silverman (2006), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012)). Functional classification is an essential task in many applications, e.g., diagnosing diseases based on curves or images from medical test results, recognizing handwriting or speech patterns, and classifying products (Epifanio (2008), Delaigle and Hall (2012), Alonso, Casado and Romo (2012), Sguera, Galeano and Lillo (2014), and Galeano, Joseph and Lillo (2015)).

Statistical depth was initially defined to rank multivariate data, mimicking the natural order of univariate data. Zuo and Serfling (2000) presented details on statistical depth. Recently, the concept has been generalized to functional

depth to rank functional data from the center outward (Fraiman and Muniz (2001), Cuevas, Febrero and Fraiman (2007), López-Pintado and Romo (2009), López-Pintado et al. (2014), and Claeskens et al. (2014)). An alternative way to rank functional data is the tilting approach proposed by Genton and Hall (2016). Functional depth, as a measure of the centrality of curves, has been used extensively to classify functional data, especially if the dataset is possibly contaminated (Sguera, Galeano and Lillo (2014)). López-Pintado and Romo (2006) defined (modified) band depth for functional data, based on which they proposed two methods for classification research of functional data: "distance to the trimmed mean" and "weighted averaged distance". Cuevas, Febrero and Fraiman (2007) introduced random projection depth and the "within maximum depth" criterion. Sguera, Galeano and Lillo (2014) defined kernelized functional spatial depth and comprehensively investigated the performance of depth-based classifiers. Cuesta-Albertos, Febrero-Bande and de la Fuente (2015) and Mosler and Mozharovskyi (2015) discussed functional versions of the depth-depth (DD) classifier (Li, Cuesta-Albertos and Liu (2012) and Liu, Parelius and Singh (1999)). Hubert, Rousseeuw and Segaert (2016) investigated functional bag distance and a distance-distance plot to classify functional data. Kuhnt and Rehage (2016) proposed a graphical approach using the angles in the intersections of one observation with the others.

There have been many other attempts to tackle the challenge of functional data classification, a great number of which sought to generalize finite-dimensional methods to functional settings. These approaches map functional data to finite-dimensional data via dimension reduction and then apply conventional classification methods, e.g., linear discriminant analysis (LDA) or support vector machines (SVM) (Boser, Guyon and Vapnik (1992) and Cortes and Vapnik (1995)), to the finite-dimensional data. Dimension reduction techniques mainly fall into two categories: *regularization* and *filtering*. The regularization approach treats functional data as multivariate data observed at discrete time points or intervals (Li and Yu (2008) and Delaigle, Hall and Bathia (2012)), and the filtering approach approximates each curve by a linear combination of a finite number of basis functions, representing the data by their corresponding basis coefficients (James and Hastie (2001), Rossi and Villa (2006), Epifanio (2008), Galeano, Joseph and Lillo (2015), Li, Chiou and Shyr (2017), and Yao, Wu and Zou (2016)).

Most of the aforementioned methods focus on univariate functional data. Very little attention has been paid to multivariate functional cases, now frequently observed in scientific research. Examples of multivariate functional cases

are gait data and handwriting data (Ramsay and Silverman (2006)), height and weight of children by age (López-Pintado et al. (2014)) and various records from weather stations (Claeskens et al. (2014)). Classifying such multivariate functional data jointly rather than marginally is necessary because a joint method takes into consideration the interaction between components and one observation may be marginally assigned to different classes by different components.

Locations/coordinates are used to classify point-wise data; however, the variation between different groups of curves in functional data classification usually results from the data's different patterns/shapes rather than their scales. We refer the readers to the simulation settings and applications in a number of references (Cuevas, Febrero and Fraiman (2007), Alonso, Casado and Romo (2012), Epifanio (2008), Galeano, Joseph and Lillo (2015), and Sguera, Galeano and Lillo (2014)). This important feature of functional data classification cannot be handled by conventional functional depths which do not effectively describe the differences in shapes of curves. A recently proposed notion of directional outlyingness (Dai and Genton (2018)) overcomes these drawbacks. The authors pointed out that the direction of outlyingness is crucial to describing the centrality of multivariate functional data. By combining the direction of outlyingness with the conventional point-wise outlyingness, they established a framework that can decompose total functional outlyingness into shape outlyingness and scale outlyingness. The shape outlyingness measures the change of point-wise outlyingness in view of both level and direction. It thus effectively describes the shape variation between curves. We extend the scalar outlyingness to an outlyingness matrix, which contains pure information of shape variation of a curve. Based on directional outlyingness and the outlyingness matrix, we propose two classification methods for multivariate functional data.

The remainder of the paper is organized as follows. In Section 2, we briefly review the framework of directional outlyingness, define the outlyingness matrix and propose two classification methods for multivariate functional data using this framework. In Section 3, we evaluate our proposed classifiers on both univariate and multivariate functional data via simulation studies. In Section 4, we use two datasets to illustrate the performance of the proposed methods in practice. We end the paper with a short discussion in Section 5. Two illustrative figures of multivariate functional data and proofs for the theoretical results are provided in an online supplement.

2. Directional Outlyingness and Classification Procedure

With $K \geq 2$ groups of data as training sets, to classify a new observation from the test set, $\mathbf{X}_0$, into one of the groups, one needs to find an effective measure of distance between $\mathbf{X}_0$ and each group. Such a measure is the Bayesian probability for the naive Bayes classifier, the Euclidean distance for the k-nearest neighbors classifier, or the functional outlyingness/depth for the depth-based classifier. Our classification methods fall into the latter category. In what follows, we first review the framework of directional outlyingness as our measure for the distance between a new curve and a labeled group of curves, and then propose two classification methods based on this framework.

2.1. Directional outlyingness

Consider a p-variate stochastic process of continuous functions, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$, with each X_k $(1 \leq k \leq p)$: $\mathcal{I} \to \mathbb{R}$, $t \mapsto X_k(t)$ from the space $\mathcal{C}(\mathcal{I}, \mathbb{R})$ of real continuous functions on $\mathcal{I}$. At each fixed time point, t, $\mathbf{X}(t)$ is a p-variate random variable. Here, p is a finite positive integer that indicates the dimension of the functional data and $\mathcal{I}$ is a compact time interval. We get univariate functional data when $p = 1$ and multivariate functional data when $p \geq 2$. Denote the distribution of $\mathbf{X}$ as $F_{\mathbf{X}}$ and the distribution of $\mathbf{X}(t)$, which is the function value of $\mathbf{X}$ at time point t, as $F_{\mathbf{X}(t)}$. For a sample of curves from $F_{\mathbf{X}}$, $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the empirical distribution is denoted as $F_{\mathbf{X},n}$; correspondingly, the empirical distribution of $\mathbf{X}_1(t), \ldots, \mathbf{X}_n(t)$ is denoted as $F_{\mathbf{X}(t),n}$. Let $d(\mathbf{X}(t), F_{\mathbf{X}(t)})$: $\mathbb{R}^p \longrightarrow [0, 1]$ be a statistical depth function for $\mathbf{X}(t)$ with respect to $F_{\mathbf{X}(t)}$. The finite sample depth function is then denoted as $d_n(\mathbf{X}(t), F_{\mathbf{X}(t),n})$.

Directional outlyingness (Dai and Genton (2018)) is defined by combining conventional statistical depth with the direction of outlyingness. For multivariate point-wise data, assuming $d(\mathbf{X}(t), F_{\mathbf{X}(t)}) > 0$, the directional outlyingness is defined as

$$\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) = \left\{ \frac{1}{d(\mathbf{X}(t), F_{\mathbf{X}(t)})} - 1 \right\} \cdot \mathbf{v}(t),$$

where $\mathbf{v}(t)$ is the unit vector pointing from the median of $F_{\mathbf{X}(t)}$ to $\mathbf{X}(t)$. Specifically, assuming that $\mathbf{Z}(t)$ is the unique median of $F_{\mathbf{X}(t)}$, $\mathbf{v}(t)$ can be expressed as $\mathbf{v}(t) = \{\mathbf{X}(t) - \mathbf{Z}(t)\} / \|\mathbf{X}(t) - \mathbf{Z}(t)\|$, where $\|\cdot\|$ denotes the L_2 norm. Then, Dai and Genton (2018) defined three measures of directional outlyingness for functional data. The functional directional outlyingness (FO) is

$$\text{FO}(\mathbf{X}, F_{\mathbf{X}}) = \int_{\mathcal{I}} \|\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})\|^2 w(t)\mathrm{d}t;$$

the mean directional outlyingness (**MO**) is

$$\mathbf{MO}(\mathbf{X}, F_{\mathbf{X}}) = \int_{\mathcal{I}} \mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) w(t)\mathrm{d}t;$$

the variation of directional outlyingness (VO) is

$$\text{VO}(\mathbf{X}, F_{\mathbf{X}}) = \int_{\mathcal{I}} \|\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) - \mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\|^2 w(t)\mathrm{d}t,$$

where $w(t)$ is a weight function defined on $\mathcal{I}$, which can be constant or pro-
portional to the local variation at each time point (Claeskens et al. (2014)).
Throughout, we use a constant weight function, $w(t) = \{\lambda(\mathcal{I})\}^{-1}$, where $\lambda(\cdot)$ is
Lebesgue measure. **MO** indicates the position of a curve relative to the center
on average, which measures the scale outlyingness of this curve; VO represents
the variation in the quantitative and directional aspects of the directional out-
lyingness of a curve and measures the shape outlyingness of that curve. We can
link the three measures of directional outlyingness by

$$\text{FO}(\mathbf{X}, F_{\mathbf{X}}) = \|\mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\|^2 + \text{VO}(\mathbf{X}, F_{\mathbf{X}}). \tag{2.1}$$

Then, FO can be regarded as the overall outlyingness and is equivalent to the
conventional functional outlyingness. When the curves are parallel to each other,
VO is zero and a quadratic relationship exists between FO and **MO**. Many
existing statistical depths can be used to construct their corresponding directional
outlyingness, among which we suggest the distance-based depths, e.g., random
projection depth (Zuo (2003)) and the Mahalanobis depth (Zuo and Serfling
(2000)). In the current paper, we choose the Mahalanobis depth to construct
directional outlyingness for all the numerical studies. As an intuitive illustration
of this framework, an example is provided in the supplement.

Compared with conventional functional depths, directional outlyingness more
effectively describes the centrality of functional data, especially the shape varia-
tion, because VO accounts for not only variation of absolute values of point-wise
outlyingness but also for the change in their directions. This advantage coincides
with the functional data classification task, which is essentially to distinguish
curves by their differences in shapes rather than scales. With the above advan-
tages, we adopt the functional directional outlyingness to measure the distance
between the curve to be classified and the labeled groups of curves. In the next
two subsections, we propose two classification methods for multivariate functional
data. Both are based on a similar idea used by the maximum depth classifier:

a new curve should be assigned to the class leading to the smallest outlyingness value.

2.2. Two-step outlyingness

In the first step, directional outlyingness maps one p-variate curve to a $(p+1)$-dimensional vector, $\mathbf{Y} = (\mathbf{MO}^{\mathrm{T}}, \mathrm{VO})^{\mathrm{T}}$, that involves both magnitude outlyingness and shape outlyingness of this curve. As shown in Figure S1 of the supplement, the $\mathbf{Y}_i$'s that correspond to the outlying curves are also isolated from the cluster of points corresponding to non-outlying curves. In the second step, we can simply measure the outlyingness of the point, $\mathbf{Y}_i$, to assess the outlyingness of its respective curve, $\mathbf{X}_i$. Specifically, we calculate the Mahalanobis distance (Mahalanobis (1936)) of $\mathbf{Y}_i$ and employ this distance as a two-step outlyingness of the raw curve.

For a set of n observations, $\mathbf{Y}_i$ $(i = 1, \ldots, n)$, a general form of the Mahalanobis distance is

$$\mathrm{D}(\mathbf{Y}, \boldsymbol{\mu}) = \sqrt{(\mathbf{Y} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{Y} - \boldsymbol{\mu})},$$

where $\boldsymbol{\mu}$ is the mean vector of the $\mathbf{Y}_i$'s and $\mathbf{S}$ is the covariance matrix. Various estimators of $\mathbf{S}$ exist in the literature, among which the minimum covariance determinant (MCD) estimator (Rousseeuw (1985)) is quite popular due to its robustness. To subtract the influence of potential outliers, we utilize this estimator to calculate the distance for our method.

In particular, the robust Mahalanobis distance based on MCD and a sample of size $h \leq n$ can be expressed as

$$\mathrm{RMD}(\mathbf{Y}) = \sqrt{(\mathbf{Y} - \bar{\mathbf{Y}}_{\mathrm{J}}^*)^{\mathrm{T}} \mathbf{S}_{\mathrm{J}}^{*-1} (\mathbf{Y} - \bar{\mathbf{Y}}_{\mathrm{J}}^*)},$$

where J denotes the set of h points that minimizes the determinant of the corresponding covariance matrix, $\bar{\mathbf{Y}}_{\mathrm{J}}^* = h^{-1} \sum_{i \in \mathrm{J}} \mathbf{Y}_i$ and $\mathbf{S}_{\mathrm{J}}^* = h^{-1} \sum_{i \in \mathrm{J}} (\mathbf{Y}_i - \bar{\mathbf{Y}}_{\mathrm{J}}^*)(\mathbf{Y}_i - \bar{\mathbf{Y}}_{\mathrm{J}}^*)^{\mathrm{T}}$. The sub-sample size, h, controls the robustness of the method. For a $(p+1)$-dimensional distribution, the maximum finite sample breakdown point is $[(n-p)/2]/n$, where $[a]$ denotes the integer part of $a \in \mathbb{R}$. Assume that we have $K \geq 2$ groups of functional observations, G_i $(i = 1, \ldots, K)$. To classify a new curve, $\mathbf{X}_0$, into one of the groups, we use the classifier

$$C_1 = \arg \min_{1 \leq i \leq K} \left(\mathrm{RMD}_{G_i}(\mathbf{X}_0) \right),$$

where C_1 is the group label, to which we assign $\mathbf{X}_0$, and $\mathrm{RMD}_{G_i}(\mathbf{X}_0)$ is the robust Mahalanobis distance of $\mathbf{X}_0$ to G_i. This classifier is based on an idea similar to the "within maximum depth" criterion (Cuevas, Febrero and Fraiman (2007))

that assigns a new observation to the group that leads to a larger depth. The difference is that we use a two-step outlyingness, which can better distinguish shape variation between curves compared with conventional functional depths utilized in existing methods.

2.3. Outlyingness matrix

Unlike conventional statistical depth, point-wise directional outlyingness of multivariate functional data, $\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})$, is a vector that allows us to define two additional statistics to describe the centrality of multivariate functional data.

Definition 1. (**Outlyingness Matrix of Multivariate Functional Data**): *Consider a stochastic process,* $\mathbf{X} : \mathcal{I} \longrightarrow \mathbb{R}^p$, *that takes values in the space* $\mathcal{C}(\mathcal{I}, \mathbb{R}^p)$ *of real continuous functions defined from a compact interval,* $\mathcal{I}$, *to* $\mathbb{R}^p$ *with probability distribution* $F_{\mathbf{X}}$. *The functional directional outlyingness matrix* (**FOM**) *is*

$$\mathbf{FOM}(\mathbf{X}, F_{\mathbf{X}}) = \int_{\mathcal{I}} \mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})\mathbf{O}^{\mathrm{T}}(\mathbf{X}(t), F_{\mathbf{X}(t)})w(t)\mathrm{d}t;$$

and the variation of directional outlyingness matrix (**VOM**) *is*

$$\mathbf{VOM}(\mathbf{X}, F_{\mathbf{X}}) = \int_{\mathcal{I}} \left\{\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) - \mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\right\} \left\{\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})\right.$$
$$\left. - \mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\right\}^{\mathrm{T}}w(t)\mathrm{d}t.$$

FOM can be regarded as a matrix version of the total outlyingness, FO, and **VOM** corresponds to the shape outlyingness, VO. A decomposition of **FOM** and its connection with the scalar statistics are on exhibit in the following.

Theorem 1. (**Outlyingness Decomposition**): *For the statistics in Definition 1, we have*

(i) $\mathbf{FOM}(\mathbf{X}, F_{\mathbf{X}}) = \mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\mathbf{MO}^{\mathrm{T}}(\mathbf{X}, F_{\mathbf{X}}) + \mathbf{VOM}(\mathbf{X}, F_{\mathbf{X}});$

(ii) $\mathrm{FO}(\mathbf{X}, F_{\mathbf{X}}) = \mathrm{tr}\left\{\mathbf{FOM}(\mathbf{X}, F_{\mathbf{X}})\right\}$ *and* $\mathrm{VO}(\mathbf{X}, F_{\mathbf{X}}) = \mathrm{tr}\left\{\mathbf{VOM}(\mathbf{X}, F_{\mathbf{X}})\right\}$, *where* $\mathrm{tr}(\cdot)$ *denotes the trace of a matrix.*

Theorem 2. (**Properties of the Outlyingness Matrix**): *Assume that* $\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})$ *is a valid directional outlyingness for point-wise data from* Dai and Genton (2018). *Then, for a constant weight function, we have*

$$\mathbf{VOM}\left(\mathbf{T}(\mathbf{X}_g), F_{\mathbf{T}(\mathbf{X}_g)}\right) = \mathbf{A}_0\mathbf{VOM}\left(\mathbf{X}, F_{\mathbf{X}}\right)\mathbf{A}_0^{\mathrm{T}},$$

where $\mathbf{T}(\mathbf{X}_g(t)) = \mathbf{A}\left\{g(t)\right\}\mathbf{X}\left\{g(t)\right\} + \mathbf{b}\left\{g(t)\right\}$ *is a transformation of* $\mathbf{X}$ *in both the response and support domains,* $\mathbf{A}(t) = f(t)\mathbf{A}_0$ *with* $f(t) > 0$ *for* $t \in \mathcal{I}$ *and*

$\mathbf{A}_0$ *an orthogonal matrix,* $\mathbf{b}(t)$ *is an p-vector at each time t, and g is a bijection on the interval* $\mathcal{I}$.

We focus on the cases when the distinction between different groups of functional data depends on their patterns/shapes. **VOM** effectively measures the level of shape variation between one curve and a group of curves. Our second classifier is

$$C_2 = \arg\min_{1 \le i \le K} \left(\|\mathbf{VOM}(\mathbf{X}_0, F_{G_i})\|_F \right),$$

where $\| \cdot \|_F$ denotes the Frobenius norm of a matrix and C_2 is the group label, to which we assign $\mathbf{X}_0$. Compared with our first classifier, this classifier is based purely on the shape information. We choose the Frobenius norm to get a scalar to take into consideration the interaction between outlyingness in different directions (the off-diagonal elements of **VOM** for group G_i).

3. Simulation Studies

In this section, we report on some simulation studies to assess finite-sample performances of the proposed classification methods and to compare them with those of some existing methods based on conventional statistical depth. We investigate both univariate and multivariate functional data cases.

3.1. Classification methods

We calculated the point-wise directional outlyingness with the Mahalanobis depth (MD) (Zuo and Serfling (2000)) for our proposed methods, two-step outlyingness, denoted by RMD, and outlyingness matrix, denoted by VOM. We considered the "within maximum depth" criterion (Cuevas, Febrero and Fraiman (2007)) for existing methods, using four conventional functional depths that can handle both univariate and multivariate functional data.

Method FM1. Integrated depth defined by Fraiman and Muniz (2001), which calculates functional depth as the integral of point-wise depth across the whole support interval of a curve. We used random Tukey depth (TD) (Tukey (1975)) as the point-wise depth for this method.

Method FM2. Integrated depth with MD as the point-wise depth. The R functions *depth.FM* and *depth.FMp* in the package *fda.usc* were used to calculate FM1 and FM2 for univariate and multivariate cases, respectively.

Method RP1. Random projection depth defined by Cuevas, Febrero and Fraiman (2007). In this method, we randomly chose NR directions, projected

the curves onto each direction, calculated the statistical depth based on the projections for each direction and took the average of the direction-wise depth. Here we set the number of random directions, NR = 50. The direction in this method refers to a random function, a, in the Hilbert space $L^2[0,1]$ so that the projection of a datum, X, is given by the standard inner product $\langle a, X \rangle = \int_0^1 a(t)X(t)\mathrm{d}t$. We used TD as the direction-wise depth for this method.

Method RP2. Random projection depth with MD as the direction-wise depth. The R functions *depth.RP* and *depth.RPp* in the package *fda.usc* were used to calculate RP1 and RP2 for univariate and multivariate cases, respectively.

TD and MD were selected as representatives of rank-based and distance-based depths, respectively. Aside from them, many other notions have been proposed in the literature. Some methods can be regarded as special cases of FM1 (with different point-wise depths), including modified band depth (López-Pintado and Romo (2009)), half-region depth (López-Pintado and Romo (2011)), simplicial band depth (López-Pintado et al. (2014)), multivariate functional half-space depth (Claeskens et al. (2014)), and multivariate functional skew-adjusted projection depth (Hubert, Rousseeuw and Segaert (2015)). Some methods have been specifically designed for univariate functional data, including kernelized functional spatial depth (Sguera, Galeano and Lillo (2014)) and extremal depth (Narisetty and Nair (2016)).

3.2. Univariate functional data

We considered three univariate settings. Different groups of curves vary in terms of patterns or shapes rather than scales. Each pair of curves thus oscillates within a similar range in different fashions in our settings.

Data 1. Class 0: $X_0(t) = A_{01}\sin(2\pi t) + A_{02}\cos(2\pi t) + \varepsilon(t)$ and class 1: $X_1(t) = A_{11}\sin(2\pi t) + A_{12}\cos(2\pi t) + \varepsilon(t)$, where A_{01} and A_{02} were generated independently from a uniform distribution $U(0.5, 1)$, A_{11} and A_{12} were i.i.d. observations from $U(1, 1.2)$ and $\varepsilon(t)$ was a Gaussian process with covariance function

$$\mathrm{cov}\{\varepsilon(t), \varepsilon(s)\} = 0.25\exp(-(t-s)^2), \quad t, s \in [0, 1].$$

This setting has been considered by Sguera, Galeano and Lillo (2014).

Data 2. Class 0: $X_0(t) = 10\sin(2\pi t) + \varepsilon(t)$ and class 1: $X_1(t) = 10\sin(2\pi t) + \sin(20\pi t) + \varepsilon(t)$. A similar setting has been considered by Cuevas, Febrero and Fraiman (2007).

Data 3. Class 0: $X_0(t) = A_0\sin(2\pi t) + \varepsilon(t)$ and class 1: $X_1(t) = A_1 + \varepsilon(t)$, where
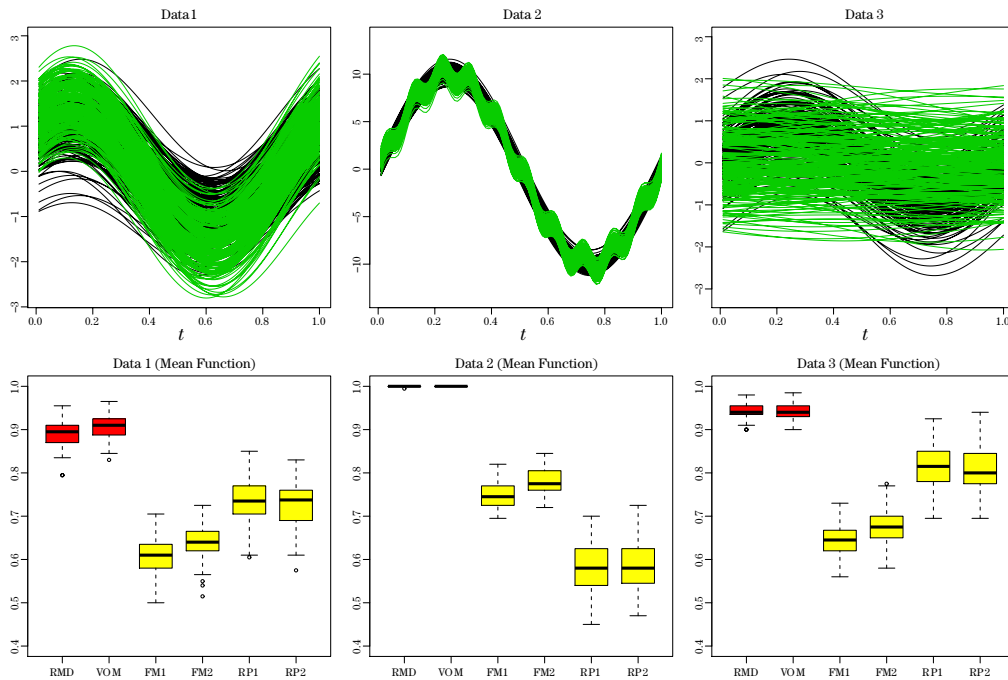
Figure 1. Top panel: Realizations of three univariate functional data settings (Data 1, 2, 3) with two classes. Bottom panel: correct classification rates of our two proposed methods, RMD and VOM, and four existing methods, FM1, FM2, RP1, and RP2, for three settings based on 100 simulations.

A_0 was generated from $U(0.5, 1)$ and A_1 was generated from $U(-1, 1)$. López-Pintado and Romo (2009) considered a similar setting for outlier detection.

In the top panel of Figure 1, we provide one realization of two classes of curves for each setting. The functions were evaluated at 50 equidistant points on $[0, 1]$. We independently generated 200 samples from both classes of each data setting, randomly chose 100 of them as the training set, and treated the remaining 100 samples as the testing set. We applied the six methods to the generated data and calculated the correct classification rate, p_c, for each method. We repeated the above procedure 100 times. The results are presented in the bottom panel of Figure 1. Under all three settings, our proposed methods performed significantly better than the four existing classification methods. For example, the classification result from our methods are almost perfect, whereas the other four methods achieve p_c less than 80% in the second setting, because our methods describe the shape variation of a curve more effectively than does conventional functional depth.
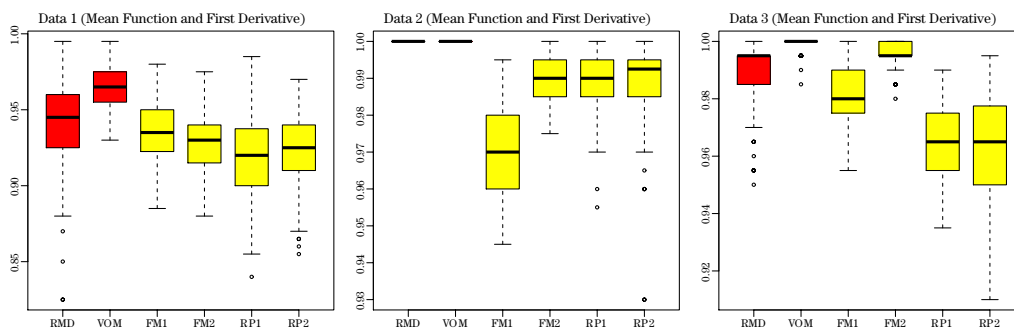
Figure 2. Correct classification rates of our two proposed methods, RMD and VOM, and four existing methods, FM1, FM2, RP1, and RP2, for three settings (Data 1, 2, 3) based on 100 simulations, using both mean functions and first-order derivatives.

3.3. Multivariate functional data

Typically, multivariate functional data are obtained from two sources: combining raw univariate curves and their derivatives (Cuevas, Febrero and Fraiman (2007) and Claeskens et al. (2014)) or functional data with multiple responses (Hubert, Rousseeuw and Segaert (2016) and Hubert, Rousseeuw and Segaert (2015)). We conducted simulation studies on both sources.

In the first scenario, we combined mean functions and the first-order derivatives of Data 1, 2, and 3 to get bivariate functional data. Under the same setting for sample sizes, design points, and repeated times, we applied the six methods to the resulting data and present the classification results in Figure 2. On the three datasets, RMD and VOM perform better than the existing methods and VOM always performs the best. The performance of the existing methods improves by combining the first-order derivatives with the mean function for classification; the derivatives are no longer of the same scale for different groups, which makes classifying by conventional functional depths easier.

In the second scenario, we considered three settings: two bivariate cases and one three-variate case. Again, the two classes of simulated data possess the same range but different patterns.

Data 4. Class 0: $\mathbf{X}_0 = (X_{01}, X_{02})^{\mathrm{T}}$ with $X_{01}(t) = \sin(4\pi t) + e_1(t)$ and $X_{02}(t) = \cos(4\pi t) + e_2(t)$ and class 1: $\mathbf{X}_1 = (X_{11}, X_{12})^{\mathrm{T}}$ with $X_{11}(t) = \sin(4\pi t) + \sin(20\pi t)/10 + e_1(t)$ and $X_{12}(t) = \cos(4\pi t) + \cos(20\pi t)/10 + e_2(t)$, where $\mathbf{e}(t) = \{e_1(t), e_2(t)\}^{\mathrm{T}}$ was a bivariate Gaussian process with zero mean and covariance function (Gneiting, Kleiber and Schlather (2010) and Apanasovich, Genton and Sun (2012)):

$$\mathrm{cov}\{e_i(s), e_j(t)\} = \rho_{ij}\sigma_i\sigma_j \mathcal{M}(|s - t|; \nu_{ij}, \alpha_{ij}), \quad i, j = 1, 2,$$

where ρ_{12} is the correlation between $X_{i1}(t)$ and $X_{i2}(t)$ ($i = 0, 1$), $\rho_{11} = \rho_{22} = 1$, σ_i^2 is the marginal variance, and $\mathcal{M}(h; \nu, \alpha) = 2^{1-\nu}\Gamma(\nu)^{-1}(\alpha|h|)^\nu \mathcal{K}_\nu(\alpha|h|)$ with $|h| = |s - t|$ is the Matérn class (Matérn (1960)) where $\mathcal{K}_\nu$ is a modified Bessel function of the second kind of order ν, $\nu > 0$ is a smoothness parameter, and $\alpha > 0$ is a range parameter. Here, we set $\sigma_1 = \sigma_2 = 0.01$, $\nu_{11} = \nu_{22} = \nu_{12} = 2$, $\alpha_{11} = 0.2$, $\alpha_{22} = 0.1$, $\alpha_{12} = 0.16$, and $\rho_{12} = 0.6$.

Data 5. Class 0: $\mathbf{X}_0 = (X_{01}, X_{02})^{\mathrm{T}}$ with $X_{01}(t) = U_{01} + e_1(t)$ and $X_{02}(t) = U_{02} + e_2(t)$ and class 1: $\mathbf{X}_1 = (X_{11}, X_{12})^{\mathrm{T}}$ with $X_{11}(t) = U_{11} + \sin(4\pi t) + e_1(t)$ and $X_{12}(t) = U_{12} + \cos(4\pi t) + e_2(t)$, where U_{01} were generated independently from $U(-1.5, 1.5)$, U_{01} and U_{02} were generated independently from $U(-2, 2)$; U_{11} and U_{12} were generated independently from $U(-0.5, 0.5)$.

Data 6. Class 0: $\mathbf{X}_0 = (X_{01}, X_{02}, X_{03})^{\mathrm{T}}$ with three components generated from class 0 of Data 1, 2, and 3. Class 1: $\mathbf{X}_1 = (X_{11}, X_{12}, X_{13})^{\mathrm{T}}$ with three components generated from class 1 of Data 1, 2, and 3. Data 6 is a three-variate setting.

Realizations of two classes of curves for each setting are illustrated in the top panel of Figure 3. The functions were evaluated at 50 equidistant points from $[0, 1]$, i.e., $t_i = i/50$. We independently generated 200 samples from both classes of each data setting, randomly chose 100 of them as the training set, and treated the remaining 100 samples as the testing set. We applied the six methods to the simulated data and calculated the correct classification rate for each method. We repeated the above procedure 100 times and present the results in the bottom panel of Figure 3. As illustrated, our proposed methods attain much higher p_c than do the existing methods. In particular, VOM has almost perfect classification results for the three settings. Sometimes the four existing methods provide results that are only slightly better than results from completely random classification. Data 5 is an example. These simulation results again validate our claim that the proposed methods based on directional outlyingness are much more effective in distinguishing curve groups that vary by shape.

Besides the non-contaminated settings, we also considered a contaminated setting. **Data 1C**. Class 0: $X_0(t) = \{I_{(V \geq 0.1)}A_{01} + (1 - I_{(V \geq 0.1)})A_{11}\}\sin(2\pi t) + A_{02}\cos(2\pi t) + \varepsilon(t)$ and class 1: $X_1(t) = A_{11}\sin(2\pi t) + A_{12}\cos(2\pi t) + \varepsilon(t)$, where I_A is an indicator function: I_x equals to 1 if $x \in A$ and 0 otherwise; V was generated from $U(0, 1)$. Class 0 was contaminated by outliers with a probability of 0.1. Sguera, Galeano and Lillo (2014) considered a similar setting. The functions were evaluated at 50 equidistant points on $[0, 1]$, i.e., $t_i = i/50$. We
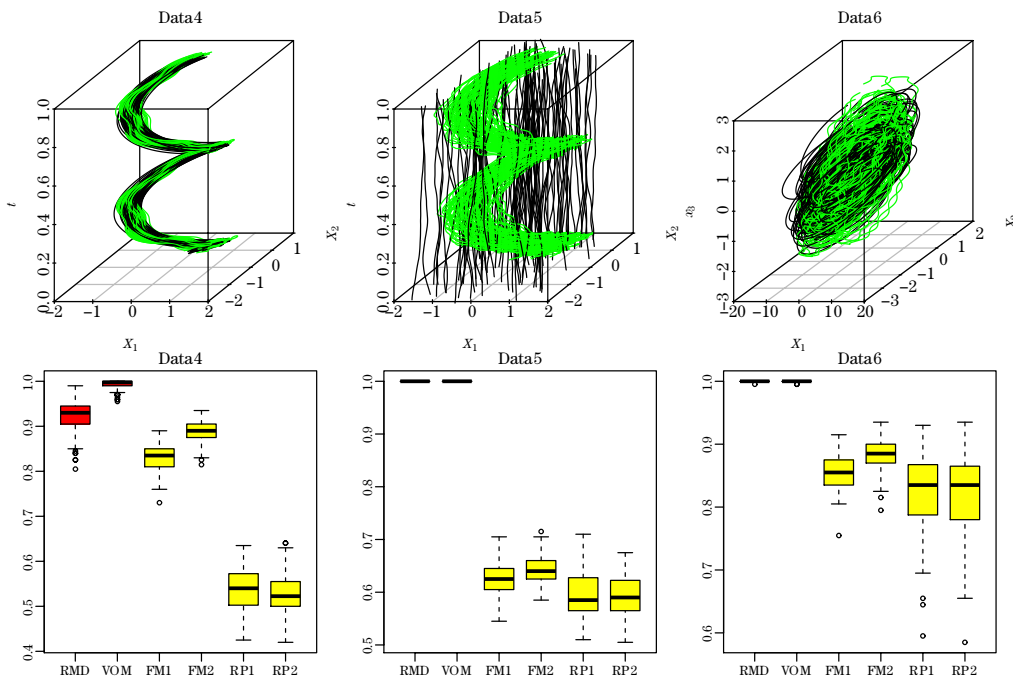
Figure 3. Top panel: realizations of three multivariate functional data settings (Data 4, 5, 6). Bottom panel: correct classification rates of our two proposed methods, RMD and VOM, and four existing methods, FM1, FM2, RP1, and RP2, for three settings based on 100 simulations.
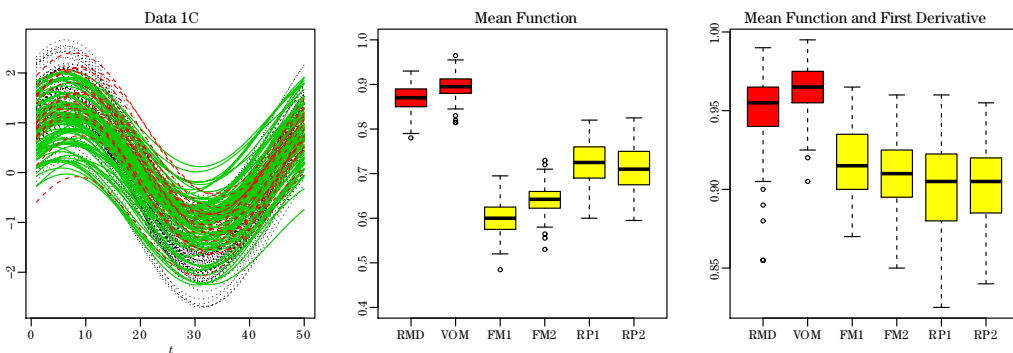

Figure 4. Left plot: realizations of the setting of Data 1C with two classes. The long-dashed curves are the outliers contaminating Class 0. Middle plot: correct classification rates of the six methods using the mean curves. Right plot: correct classification rates of the six methods using the combination of the mean curves and their first-order derivatives.

independently generated 200 samples from both classes, randomly chose 100 of them as the training set, and treated the remaining 100 samples as the testing set. We calculated the correct detection rates of the six methods based on the mean curves and the combination of the mean curves and their first-order derivatives, respectively. The results as illustrated in Figure 4, are quite similar to the results from Data 1, suggesting that our proposed methods are robust to the presence of outliers.

4. Data Applications

We evaluated our methods on two datasets: the first univariate and the second multivariate. Comparisons with existing methods are provided as well.

4.1. Phoneme data

We applied our methods to the benchmark phoneme dataset. Phoneme is a speech-recognition problem introduced by Hastie, Buja and Tibshirani (1995). We obtained the data from the R package *fds*. The dataset comprises five phonemes extracted from the TIMIT database (TIMIT Acoustic-Phonetic Continuous Speech Corpus, NTIS, U.S. Department of Commerce). The phonemes are transcribed as follows: "sh" as in "she", "dcl" as in "dark", "iy" as the vowel in "she", "aa" as the vowel in "dark", and "ao" as the first vowel in "water". A log-periodogram was computed from each speech frame; this is one of several widely used methods for translating speech data into a form suitable for speech recognition. For each log-periodogram, we considered the first 150 frequencies. In our study, we randomly selected 400 samples for each class and consequently, 2,000 samples were considered in total. Ten samples from each class are illustrated in Figure S2 of the supplement. As shown, the five types of curves vary within the same range with different shapes.

We randomly selected 1,500 samples as the training set (300 for each class) and treated the remaining 500 samples as the testing set (100 for each class). We applied the six aforementioned methods in two ways: using only the raw data (univariate); using both raw data and their first-order derivatives (bivariate). For each method, we calculated the correct classification rate and repeated this procedure 50 times. The simulation results are presented in Figure 5. Based on the raw data, our methods perform better than the existing methods. After taking their first derivatives into consideration, the performance of all methods except for RMD is improved significantly and VOM achieves the highest correct classification rate.
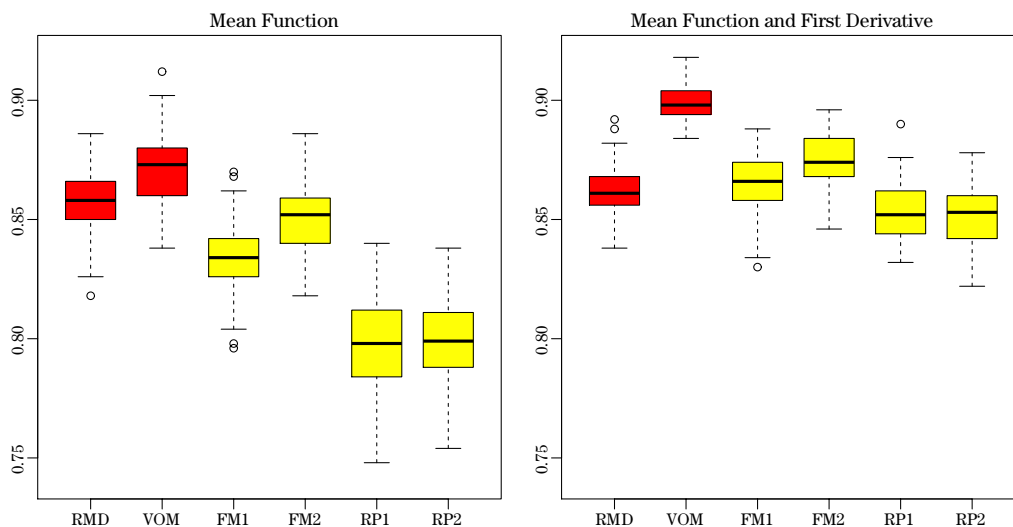
Figure 5. Correct classification rates of our two proposed methods RMD and VOM, and four existing methods, FM1, FM2, RP1, and RP2, of the phoneme data. Left: results based on only raw data; right: results based on both raw data and their first-order derivatives.

4.2. Gesture data

Gesture commands are widely used to interact with or control external devices, e.g., playing gesture-based games and controlling interactive screens. The problem is how to recognize one observation accurately as a particular gesture. Our second dataset includes gesture data comprising the eight simple gestures shown in Figure S3 of the supplement. These gestures have been identified by a Nokia research study as preferred by users for interaction with home appliances.

We downloaded this dataset from Chen et al. (2015). This dataset has been analyzed by Shokoohi-Yekta et al. (2015) with the dynamic time warping algorithm in a time series context. We used it to illustrate our functional data analysis approach. It includes 4,480 gestures: 560 for each type of action made by eight participants ten times per day during one week. Each record contains accelerations on three orthogonal directions (X, Y and Z), which means we need to classify three-dimensional curves. We found the median curve of acceleration for three directions of each gesture with the functional boxplot (Sun and Genton (2011)) as shown in Figure 6. Generally, most of the acceleration curves oscillate between -3 and 3. We applied the six methods to the gesture data in four ways: combining all three components together, (X, Y, Z), and selecting two components out of three, (X, Y), (X, Z), and (Y, Z). For each numerical study,
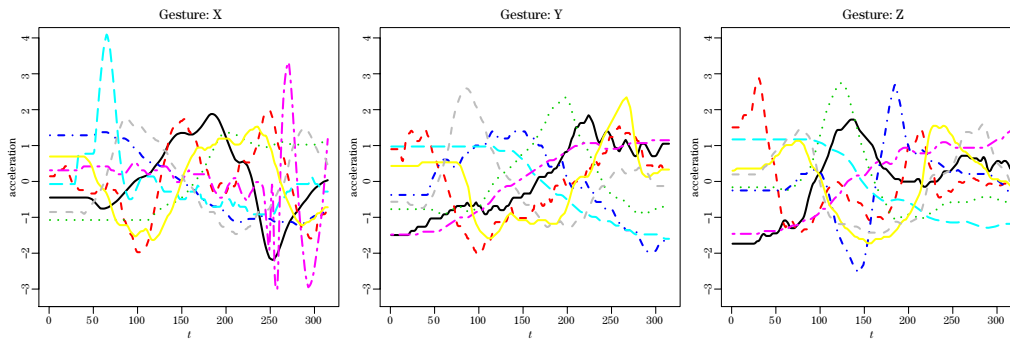
Figure 6. Left column: eight median curves of X-accelerations of eight gestures; middle column: eight median curves of Y-accelerations; right column: eight median curves of Z-accelerations.

we randomly selected 3,200 samples as the training set (400 for each class) and treated the remaining 1,280 samples as the testing set (160 for each class). We repeated this procedure for 50 times and report the correct classification rates of each method in Figure 7.

In the four combinations, our proposed methods are always better than the four existing methods except for RMD of (X, Z). For three cases, VOM achieves the best performance among the six methods. Overall, the correct classification rates improve as we raise the dimensions of the curves. We define the marginal effect of component X as the averaged difference between p_c for (X, Y, Z) and (Y, Z). This quantity measures how informative a component is for a classification task. By comparing the plot of (X, Y, Z) with the other three cases, we find that the marginal effect of Y is the smallest. This finding is consistent with the fact that the acceleration curves in direction Y are more alike with each other. For example, the black and yellow curves in the middle graph of Figure 6 are quite similar to the purple and red curves, respectively. In contrast, the shapes of the acceleration curves in the other two directions differ, which leads to their higher marginal effects. The gestures included in the dataset were mainly collected from the screens of smart phones, which means that the direction orthogonal to the screen is not as informative as the other two directions.

5. Discussion

The proposed methods can be simply generalized to image or video data (Genton et al. (2014)), where the support of functional data is two-dimensional. We plan to investigate more general settings for both classifiers and data struc-
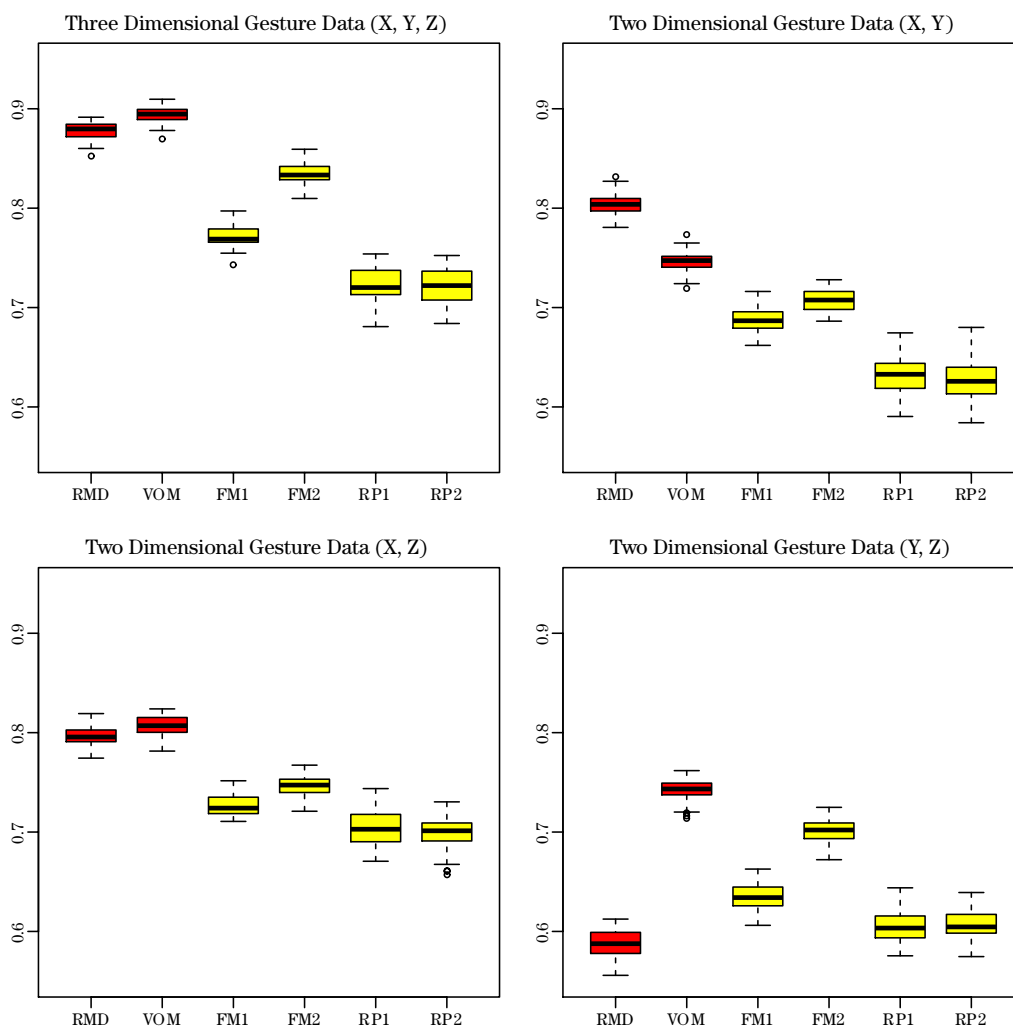
Figure 7. Correct classification rates of our two proposed methods, RMD and VOM, and four existing methods, FM1, FM2, RP1, and RP2, of the gesture data. Top left: gesture data (X, Y, Z); top right: gesture data (X, Y); bottom left: gesture data (X, Z); bottom right: gesture data (Y, Z).

tures. Rather than the constant weight function considered in the current paper, we believe that a weight function proportional to local variation could further improve our methods. It is reasonable to put more weight on the time points where the curves differ a lot and less weight on those where the curves are quite alike. For functional data observed at irregular or sparse time points (López-Pintado and Wei (2011)), we may fit the trajectories with a set of basis functions and

then estimate depth of the discrete curves based on their continuous estimates. The functional data within each group could be correlated in general data structures. An example is spatio-temporal precipitation (Sun and Genton (2012)). Our methods need further modifications to account for the correlations between functional observations as well.

Supplementary Materials

Supplementary material includes an illustrative example of functional directional outlyingness framework, two figures of real data, and technical proofs for the theoretical results.

Acknowledgment

References

Alonso, A. M., Casado, D. and Romo, J. (2012). Supervised classification for functional data: A weighted distance approach. *Computational Statistics & Data Analysis* **56**, 2334–2346.

Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association* **107**, 180–193.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. 144–152. ACM.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A. and Batista, G. (2015). The ucr time series classification archive. `www.cs.ucr.edu/eamonn/time_series_data/`.

Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association* **109**, 411–423.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20**, 273–297.

Cuesta-Albertos, J. A., Febrero-Bande, M. and de la Fuente, M. O. (2015). The DD^G-classifier in the functional setting. *arXiv preprint arXiv:1501.00372*.

Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics* **22**, 481–496.

Dai, W. and Genton, M. G. (2018). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, to appear. DOI: 10.1016/j.csda.2018.03.017.

Delaigle, A. and Hall, P. (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **74**, 267–286.

Delaigle, A., Hall, P. and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika* **99**, 299–313.

Epifanio, I. (2008). Shape descriptors for classification of functional data. *Technometrics* **50**, 284–294.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *TEST* **10**, 419–440.

Galeano, P., Joseph, E. and Lillo, R. E. (2015). The mahalanobis distance for functional data with applications to classification. *Technometrics* **57**, 281–291.

Genton, M. G. and Hall, P. (2016). A tilting approach to ranking influence. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **78**, 77–97.

Genton, M. G., Johnson, C. Potter, K., Stenchikov, G. and Sun, Y. (2014). Surface boxplots. *Stat* **3**, 1–11.

Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**, 1167–1177.

Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *The Annals of Statistics* **23**, 73–102.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications.* Springer.

Hubert, M., Rousseeuw, P. J. and Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications* **24**, 177–202.

Hubert, M., Rousseeuw, P. J. and Segaert, P. (2016). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification* **11**, 445–466.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **63**, 533–550.

Kuhnt, S. and Rehage, A. (2016). An angle-based multivariate functional pseudo-depth for shape outlier detection. *Journal of Multivariate Analysis* **146**, 325–340.

Li, B. and Yu, Q. (2008). Classification of functional data: A segmentation approach. *Computational Statistics & Data Analysis* **52**, 4790–4800.

Li, J., Cuesta-Albertos, J. A. and Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association* **107**, 737–753.

Li, P.-L., Chiou, J.-M. and Shyr, Y. (2017). Functional data classification using covariate-adjusted subspace projection. *Computational Statistics & Data Analysis* **115**, 21–34.

Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics* **27**, 783–858.

López-Pintado, S. and Romo, J. (2006). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* **72**, 103–120.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association* **104**, 718–734.

López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Computational Statistics & Data Analysis* **55**, 1679–1695.

López-Pintado, S., Sun, Y., Lin, J. K. and Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification* **8**, 321–338.

López-Pintado, S. and Wei, Y. (2011). Depth for sparse functional data. In *Recent Advances in Functional Data Analysis and Related Topics* (Edited by F. Ferraty) pp. 209–212. Springer.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* **2**, 49–55.

Matérn, B. (1960). *Spatial Variation.* Springer.

Mosler, K. and Mozharovskyi, P. (2015). Fast DD-classification of functional data. *Statistical Papers* **58**, 1055–1089.

Narisetty, N. N. and Nair, V. N. (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association* **111**, 1705–1714.

Ramsay, J. O. and Silverman, B. W. (2006). *Functional Data Analysis.* Springer.

Rossi, F. and Villa, N. (2006). Support vector machine for functional data classification. *Neurocomputing* **69**, 730–742.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Volume B* (Edited by W. Grossmann, G. Pflug, I. Vincze and W. Wert), 283–297. Reidel, Dordrecht.

Sguera, C., Galeano, P. and Lillo, R. (2014). Spatial depth-based classification for functional data. *TEST* **23**, 725–750.

Shokoohi-Yekta, M., Hu, B., Jin, H., Wang, J. and Keogh, E. (2015). Generalizing dynamic time warping to the multi-dimensional case requires an adaptive approach. SDM.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20**, 316–334.

Sun, Y. and Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics* **23**, 54–64.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians.* 523–531.

Yao, F., Wu, Y. and Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *TEST* **25**, 1–22.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics* **31**, 1460–1490.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics* **28**, 461–482.

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

E-mail: wenlin.dai@kaust.edu.sa

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

E-mail: marc.genton@kaust.edu.sa