Learning Shape Sensitive Descriptors for Classifying Functional Data

Wojciech Rafajłowicz[(✉)] and Ewaryst Rafajłowicz

Department of Control Systems and Mechatronics, Wroclaw University of Science and Technology, Wyb. Wyspianskiego 27, 50 370 Wrocław, Poland
{wojciech.rafajlowicz,ewaryst.rafajlowicz}@pwr.edu.pl

Abstract. We propose a new method of learning descriptors for constructing classifiers of functional data. These descriptors are moments of a curve derivative, but their learning is based solely on samples of the curve itself. Furthermore, the derivative itself is not directly estimated. This is possible due to the trick of using simultaneously two different bases of a functional space.

The advantage of extracting features from the derivative instead of from a curve itself is in raising their sensitivities to a shape of a curve. As expected, this may result in better classification accuracy. The simulation experiments that are based on an augmented real data support this claim, but it is not unconditional. Namely, noticeable improvements can be obtained when an appropriate classifier is selected.

Keywords: Functional data classification · Shape sensitive descriptors · Learning derivatives · Orthogonal expansion

1 Introduction

In recent several years, one can observe a growing interest of researchers to analyze and classify functional data (see monographs and survey papers cited in the next section). Clearly, important contributions in this direction are much earlier (see, e.g., [1,10] for reviews on classifying electrocardiogram (ECG) signals and [3,7] for electroencephalogram (EEG) signals classification and features selection [4,5] as well as [2] for the survey on the analysis of electromyography signals).

The renewed interest has its origin in growing possibilities of acquiring large number of samples from new sensors and storing them on cloud databases. An example of this kind is provided at the end of the paper, where curves from accelerometers are classified. Their distinctive feature is that they are repetitive, but in a stochastic sense, i.e., their underlying probability distributions remain the same for each class, although they are unknown. As the result, curves differ more by shape than by amplitudes. Therefore, our aim is to derive descriptors of curves that are based on curves' derivatives, without having an access to them directly.

The paper is organized as follows. In the next section, we justify using curve descriptors that are based on moments of its derivative instead of the curve itself. The main result of this section is the derivation of a relationship between these two kinds of moments. This relationship is crucial for Sect. 3 to construct an algorithm for learning the derivative descriptor, without having access to samples of the derivative curve. In Sect. 3 also an interplay between learning these descriptors and learning a classifier is described. Finally, in Sect. 4, the extensive results of testing the proposed approach on an augmented real data are summarized. They aimed to investigate an influence of a classifier on possible improvements of the classification accuracy.

2   Descriptors Based on the Derivative Moments

Our derivations are based on the notion of square root velocity (SRV) of differentiable functions $\mathbf{X}(t)$, $\mathbf{Y}(t)$, $t \in [0, T]$ that can be interpreted as signals, curves etc., defined on a finite time interval of the length $T > 0$. For those $t \in [0, T]$ for which the derivative $\mathbf{X}'(t)$ is not zero, the SRV of $\mathbf{X}$, denoted further as $q(\mathbf{X}, t)$ is defined as follows

$$q(\mathbf{X}, t) = \frac{\mathbf{X}'(t)}{\sqrt{|\mathbf{X}'(t)|}}, \quad t \in [0, T] \tag{1}$$

or, equivalently,

$$q(\mathbf{X}, t) = sgn(\mathbf{X}'(t)) \sqrt{|\mathbf{X}'(t)|}. \tag{2}$$

From (1) it is clear that the SRV description of $\mathbf{X}$ is invariant in a scale and a vertical position, i.e., for any $c > 0$ and any $\beta \in \mathbb{R}$:

$$q(c\,\mathbf{X}, t) = q(\mathbf{X}, t), \quad q(\beta + \mathbf{X}, t) = q(\mathbf{X}, t), \quad t \in [0, T]. \tag{3}$$

Let $\mathbf{X}'$ and $\mathbf{Y}'$ be square-integrable on $[0, T]$, $\mathbf{X}', \mathbf{Y}' \in L_2(0, T)$. Then, from (2) we immediately obtain:

$$\int_0^T q^4(\mathbf{X} - \mathbf{Y}, t)\, dt = \int_0^T [\mathbf{X}'(t) - \mathbf{Y}'(t)]^2\, dt. \tag{4}$$

Strictly speaking, the squared right hand side of (4) is not a distance measure, since $\mathbf{X} - \mathbf{Y}$ that differ by a constant yield zero in (4).

However, this expression suggests that the derivatives of $\mathbf{X}$, $\mathbf{Y}$, ... can be useful in classifying curves in a shape-sensitive way. In particular, moments of $\mathbf{X}'$, $\mathbf{Y}'$, ... with respect to a selected basis in $L_2(0, T)$ are worthwhile candidates for descriptors of $\mathbf{X}$, $\mathbf{Y}$, ... when one attempts to classify them. We shall follow this line of reasoning.

2.1   Modeling Random Curves

The main difficulty is in learning such descriptors from samples of $\mathbf{X}$, $\mathbf{Y}$, ... instead of $\mathbf{X}'$, $\mathbf{Y}'$, ... that are frequently not directly available. In this respect we shall follow [13,14], where the approach to nonparametric estimation of derivatives from noisy observations of $\mathbf{X}(t)$ can be found. However, we emphasise that in our case $\mathbf{X}(t)$ is a random element of $L_2(0, T)$, which implies a different model of random errors than the one used in [13]. Furthermore, the estimation of $\mathbf{X}'$ is only an intermediate step, since our goal is to learn the moments of $\mathbf{X}'$ with respect to a selected orthonormal basis.

We refer the reader to [6,8,9,16,19] for more details on shape-sensitive description of random curves.

Let $\mathbf{v}_k(t)$, $t \in [0, T]$, $k = 1, 2, \ldots$ be a selected orthogonal and complete sequence in $L_2(0, T)$ with elements that are also normalized to 1 with respect to the standard norm $||\mathbf{v}_k||^2 = < \mathbf{v}_k, \mathbf{v}_k >$, where $< \mathbf{X}, \mathbf{Y} > = \int_0^T \mathbf{X}(t)\,\mathbf{Y}(t)\,dt$. Then, $\mathbf{X} \in L_2(0, T)$ has the representation:

$$\mathbf{X}(t) = \sum_{k=1}^{K} a_k\,\mathbf{v}_k(t) + \mathbf{R}_K(t), \quad t \in [0, T], \tag{5}$$

where

$$\mathbf{R}_K(t) \overset{def}{=} \sum_{k=(K+1)}^{\infty} \beta_k\,\mathbf{v}_k(t) \tag{6}$$

and the coefficients are given by $a_k = < \mathbf{X}, \mathbf{v}_k >$, $k = 1, 2, \ldots, K$, $\beta_k = < \mathbf{X}, \mathbf{v}_k >$, $k = (K+1), (K+2), \ldots$, while the convergence is understood the L_2 norm sense.

Collections of coefficients a_k's and β_k's are both random, but they play different roles in our derivations. Namely, a_k's are regarded as descriptors that are informative for curves classification, while β_k's are interpreted as coefficients of non-informative error $\mathbf{R}_K$.

Denote by $\mathbb{E}$ the expectation with respect to a_k's and β_k's. Although their distributions are not known, the following assumptions are made:

$$\mathbb{E}[\beta_k] = 0, \quad \mathbb{E}[\beta_k\,\beta_j] = 0, \ \ k \neq j \ k, j = (K+1), (K+2), \ldots, \tag{7}$$

$$\gamma(K) \overset{def}{=} \mathbb{E}||\mathbf{R}_K||^2 \ \to 0, \ \text{as } K \to \infty, \tag{8}$$

$$\mathbb{E}(a_k^2) < \infty, \ \mathbb{E}(a_k\,\beta_l) = 0, \ k = 1, 2 \ldots, K, \ l = (K+1), (K+2), \ldots \tag{9}$$

Assumption (8) implicitly imposes constraints on the variability of the residual curve $\mathbf{R}_K(t)$, $t \in [0, T]$ for large K.

For simplicity, $1 \leq K \leq \infty$ is assumed to be fixed and known. In practice, one should select K so as to minimize an estimate of the classification error plus a penalty term for too complicated model, such as in the AIC, BIC etc. criterions.

2.2   The Relationship Between Descriptors of a Curve and Its Derivative

In the next section, we provide details of learning moments of $\mathbf{X}'$ from equidistant observations of $\mathbf{X}$ only. Here, we outline a general idea. If $\mathbf{v}_k$'s are differentiable and the series (5) and (6) is term-by-term differentiable, then,

$$\mathbf{X}'(t) = \sum_{k=1}^{K} a_k \, \mathbf{v}'_k(t) + \mathbf{R}'_K(t), \quad t \in [0, T], \tag{10}$$

On the other hand, for $\mathbf{w}_k$, $k = 1, 2, \ldots$ being an orthonormal and complete sequence in $L_2(0, T)$, $\mathbf{X}' \in L_2(0, T)$ has the representation

$$\mathbf{X}'(t) = \sum_{k=1}^{K} b_k \, \mathbf{w}_k(t) + \mathbf{r}_K(t), \quad b_k = <\mathbf{X}', \mathbf{w}_k> \tag{11}$$

$$\mathbf{r}_K(t) = \sum_{k=(K+1}^{\infty} \eta_k \, \mathbf{w}_k(t), \quad \eta_k = <\mathbf{X}', \mathbf{w}_k>, \; k = (K+1). \ldots \tag{12}$$

For sufficiently large K, according to (8), we approximate $\mathbf{X}'$ in (10) by the first summand, which yields, after substituting it into (11),

$$b_k = \sum_{j=1}^{K} a_j <\mathbf{w}_k, \mathbf{v}'_j>, \quad k = 1, 2, \ldots, K. \tag{13}$$

Observe that these formulas are exact, if $<\mathbf{w}_k, \mathbf{R}'_K> = 0$, $k = 1, 2, \ldots$, but this is not postulated here.

Summarizing, moments b_k's of $\mathbf{X}'$ with respect to basis $\mathbf{w}_k$'s can be expressed as linear combinations of moments a_k's that are estimable from observations of $\mathbf{X}$ itself, assuming that for each $k = 1, 2, \ldots, K$

$$\text{at least one } <\mathbf{w}_k, \mathbf{v}'_j> \neq 0, \quad j = 1, 2, \ldots, K. \tag{14}$$

Additionally, elements $<\mathbf{w}_k, \mathbf{v}'_j>$ of $K \times K$ transformation matrix, say B, are either known or they can be approximated to any desired accuracy by quadrature formulas.

3 Learning Classifiers Based on Curves' Derivative Descriptors

Suppose, for simplicity of formulas only, that random curves like $\mathbf{X}$ are drawn from two classes, labelled by I and II, that are formed as follows: firstly, vector $\bar{a} \overset{def}{=} [a_1, a_2, \ldots, a_K]^{tr}$ is drawn from a cumulative distribution function (c.d.f.), which is either F_I or F_{II}. These c.d.f.'s are not known. We do not impose any special restrictions on them, except for the existence of the second moments of a_k's and (9). In this way, a large class of classification problems for informative part of $\mathbf{X}$ can be stated. The second step in modeling $\mathbf{X}$ is to draw β_k's. Their distributions are also unknown and only conditions (7), (8) and (9) are assumed to hold. Finally, $\mathbf{X}$ is formed according to (5) and (6). Thus, $\mathbf{X}$ may come from class I or II, depending whether $\bar{a}$ was according to c.d.f. F_I or F_{II}. The existence of a priori probabilities $0 < p_I < 1$, $0 < p_{II} < 1$, $p_I + p_{II} = 1$ that $\mathbf{X}$ is from class I or II is postulated, but they are unknown. Their estimation by fractions in the learning sequence is a simple task, unless an essential class imbalance does not appear, which is excluded in this paper.

3.1 Learning Sequence

A learning sequence that we have at our disposal is of the form:

$$\mathcal{L}_N \overset{def}{=} \{(\bar{x}^{(1)}, j_1), (\bar{x}^{(2)}, j_2), \ldots, (\bar{x}^{(N)}, j_N)\}, \tag{15}$$

where $j_n \in \{I, II\}$ are correct class labels (provided by an expert), while $\bar{x}^{(n)}$ are equidistant, in $[0, T]$, samples from curves $\mathbf{X}^{(n)}$, taken at time instants t_i, $i = 1, 2, \ldots, m$, $n = 1, 2, \ldots, N$. Samples forming $\bar{x}^{(n)}$'s have the following form:

$$x_i^{(n)} = \mathbf{X}^{(n)}(t_i) = \bar{\mathbf{v}}^{tr}(t_i)\,\bar{a}^{(n)} + \mathbf{R}_K(t_i), \quad i = 1, 2, \ldots, m, \tag{16}$$

where $\bar{a}^{(n)}$ are drawn either according to F_J or F_{II}, while

$$\bar{\mathbf{v}}^{tr}(t) \overset{def}{=} [\mathbf{v}_1(t), \mathbf{v}_2(t), \ldots \mathbf{v}_K(t)]. \tag{17}$$

Analogously, new $\mathbf{X}$ to be classified is represented only by $\bar{x}$ with elements

$$x_i = \mathbf{X}(t_i) = \bar{\mathbf{v}}^{tr}(t_i)\,\bar{a} + \mathbf{R}_K(t_i), \quad i = 1, 2, \ldots, m, \tag{18}$$

Problem Formulation. Using learning sequence $\mathcal{L}_N$, derive a classifier that classifies $\mathbf{X}$, represented only by $\bar{x}$, to class I or II. This classifier should be shape sensitive in the sense that, for a preselected orthonormal and complete sequence $\mathbf{w}_k$'s, the classifier decision is based on learning descriptors $b_k = <\mathbf{X}', \mathbf{w}_k>$, $k = 1, 2, \ldots, K$, which are directly not available.
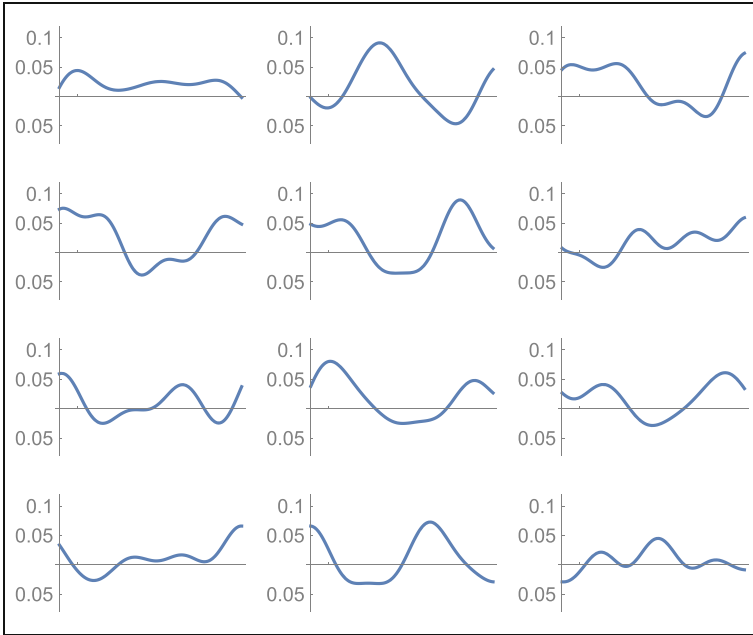
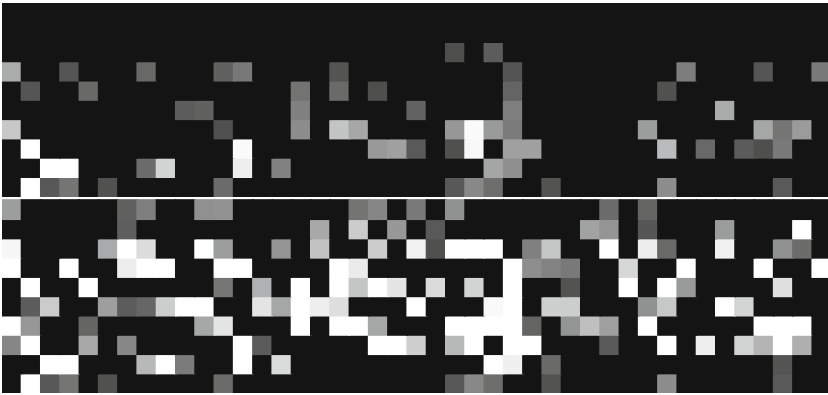**Fig. 1.** Examples of curves to be classified

Fig. 2. Descriptors of curves to be classified, stacked together, and displayed as images. Upper panel – classic DCT descriptors, lower panel – descriptors based on learning derivatives.

3.2   Learning Descriptors

Model of observations (18) and (16) suggest that for estimating primary descriptors $\bar{a}$ and $\bar{a}^{(n)}$'s one may use the method of minimizing the least squares error (LSE) in the nonparametric setting with deterministic regressors t_i's (see [12]).

However, in this case the ordinary (unweighted) LSE approach is not recommended, since $\mathbf{R}_K(t_i)$'s are correlated for moderate K.

Thus, $\bar{a}$ is estimated in more classic way as

$$\hat{\bar{a}} \,=\, \Delta_m \sum_{i=1}^{m} x_i\, \bar{\mathbf{v}}(t_i) \,=\, \Delta_m\, \bar{\mathbf{V}}\, \bar{x}, \quad \Delta_m \stackrel{def}{=} T/m, \tag{19}$$

where $\bar{\mathbf{V}}$ is $K \times m$ matrix composed of the columns: $\bar{\mathbf{v}}(t_i)$'s. It is not difficult to show that $\hat{\bar{a}}$ is asymptotically (as $m \to \infty$) unbised for $\bar{a}$. It is more tedious to bound the variances of $\hat{\bar{a}}_k$'s by $\zeta(K)/m^2$, where $\zeta(K) > 0$ depends on K in a polynomial way.

Transforming samples of the learning curves in the same way as in (19), we obtain the learning sequence, denoted as $\mathcal{A}_N$, composed of the classic descriptors:

$$\mathcal{A}_N \,=\, \{(\hat{\bar{a}}^{(n)},\, j_n),\; n = 1, 2, \ldots, N\},\; \hat{\bar{a}}^{(n)} = \Delta_m\, \bar{\mathbf{V}}\, \bar{x}^{(n)}, \tag{20}$$

where j_n's are joined at the original order.

Using the plug-in idea and (13), we can learn derivative-sensitive descriptors as follows:

$$\hat{b}_k \,=\, \sum_{j=1}^{K} \hat{a}_j <\mathbf{w}_k,\, \mathbf{v}'_j>, \quad k = 1, 2, \ldots, K \tag{21}$$

and they are also asymptotically unbised and with finite variances that can be reduced faster sampling (m larger).

Transforming the Learning Sequence. Elements of $\mathcal{L}_N$ are transformed into descriptors in the same way as in (21), providing learning sequence $\mathcal{B}_N$, say, of the form:

$$\mathcal{B}_N \,=\, \{(\hat{\bar{b}}^{(n)},\, j_n),\; n = 1, 2, \ldots, N\},\; \hat{\bar{b}}^{(n)} = \Delta_m\, B\, \bar{\mathbf{V}}\, \bar{x}^{(n)}, \tag{22}$$

while labels j_n's are rewritten from $\mathcal{L}_N$, accordingly.

Summarizing, original learning sequence $\mathcal{L}_N$, with usually long sequences of samples $\bar{x}^{(n)}$, was transformed into learning sequence $\mathcal{B}_N$ with descriptors for derivatives. Furthermore, this transformation is linear in $\bar{x}^{(n)}$, which allows for speeding up computations.

At this stage, it suffices to select a proper classifier, to learn and test it using $\mathcal{B}_N$ and to apply it for newly coming sample $\bar{x}$, after transforming it to $\hat{\bar{b}} = \Delta_m\, B\, \bar{\mathbf{V}}\, \bar{x}$. For brevity, the obtained classifier will be denoted as $CLname[\mathcal{B}_N;\, \bar{x}]$ or $CLname[\mathcal{A}_N;\, \bar{x}]$, when the learning is based on standard descriptors, for the sake of comparisons. For example, the support vector machine (SVM) classifier that was trained on $\mathcal{B}_N$ is denoted as $SVM[\mathcal{B}_N;\, \bar{x}]$ and its output is I or II class label.

As we shall see in the next section, this obvious route of building a classifier may lead to moderate or essential improvements of the classification accuracy, depending on the choice of a classifier.

4   Testing and Comparisons on Augmented Acceleration Data

Operators' cabins of large working machines are frequently subject to relatively high, repetitive accelerations. Benchmark data of this kind are freely available from [17], while in [18] their detailed description is provided.

The benchmark consists from $N = 43$ learning curves, each containing $m = 1K$ samples (see Fig. 1, where examples of curves are shown, after a low-pass filtering). Labels, either I or II, were attached to each curve, corresponding to lighter or heavier working conditions. Notice that curves in Fig. 1 differ mainly by shape rather than in amplitude.

As orthogonal systems in $L_2(0, T)$, we have selected the cosine series as $\mathbf{v}_k$'s and the sine series as $\mathbf{w}_k$'s. Descriptors $\hat{\bar{a}}^{(n)}$'s were computed according to (20) for $K = 16$. For illustration purposes, these $N = 43$ vectors were stacked into 43×16 matrix that is displayed in Fig. 2 – upper panel (dark places correspond to lower values of the descriptors).

Descriptors of derivatives $\hat{\bar{b}}^{(n)}$'s were computed according to (22). They are analogously visualized in Fig. 2 – lower panel. By a visual inspection of these two panels, we conclude that the variability of $\hat{\bar{b}}^{(n)}$'s is much larger than $\hat{\bar{a}}^{(n)}$'s. Thus, one may hope that the classification accuracy will also be larger.

Data Augmentation. Unfortunately, the learning sequence of the length $N = 43$ is far too short for learning and comparisons. Therefore, we augmented the original data as follows: each estimated $\hat{\bar{a}}^{(n)}$'s was repeated 1000 by adding to it the Gaussian perturbations with zero mean and dispersions 0.02 and keeping the same label. This augmentation corresponds to about 11% perturbations amplitudes. Notice that it suffices to add perturbations to $\hat{\bar{a}}^{(n)}$'s, instead adding them to original samples, since $\hat{\bar{a}}^{(n)}$'s depend on them linearly. In this way, we have obtained the augmented learning sequence $\mathcal{A}_{N_e}$ of the length $N_e = 43000$. Each descriptor from this sequence was transformed by (22), which led to the augmented learning sequence of the derivatives descriptors $\mathcal{B}_{N_e}$.

The next step was to learn the logistic regression classifier twice in order to obtain: $LogR(\mathcal{A}_{N_e}; \bar{x})$ and $LogR(\mathcal{B}_{N_e}; \bar{x})$ and their characteristics, such as accuracy, precision, recall etc. These characteristics are collected as pairs, separated by |, in the $LogR$ column in Table 1.

In the same way, the following classifiers were learned and validated:

LogR – the logistic regression classifier,
SVM – the support vector machine,
DecT – the decision tree classifier,
gbTr – the gradient boosted trees,
RFor – the random forests classifier,
5NN – the 5 nearest neighbors classifier.

The results are summarized in Table 1. Notice that the result on the left hand side in each cell of this table is intentionally the same as in [15], for the sake of comparisons.

Table 1. An account on testing the popular classifiers when the cosine moments (the left result) and the shape sensitive descriptors (the right result) are used for learning them. Abbreviations: Cohen – the Cohen κ coefficient, MCC – the Matthews Correlation Coefficient. For the abbreviations of the classifiers' names – see the text.

Classifier	LogR	SVM	DecT	gbTr	RFor	5 NN
Accuracy	0.91|0.97	0.94|0.96	0.84|0.86	0.92|0.90	0.91|0.92	0.90|0.92
Cohen	0.76|0.91	0.82|0.88	0.60|0.65	0.77|0.72	0.72|0.77	0.70|0.73
MCC	0.76|0.91	0.82|0.88	0.61|0.66	0.77|0.72	0.72|0.78	0.71|0.75
Precision	0.96|0.99	0.94|0.95	0.93|0.95	0.94|0.93	0.92|0.93	0.90|0.90
Recall	0.92|0.97	0.98|0.99	0.86|0.87	0.96|0.94	0.96|0.98	0.98|0.99
Specificity	0.88|0.95	0.80|0.83	0.60|0.85	0.79|0.77	0.72|0.75	0.65|0.65
F1 score	0.94|0.98	0.96|0.97	0.89|0.91	0.95|0.94	0.94|0.95	0.94|0.95

The analysis of this table leads to the following conclusions:

- when the *LogR* classifier is used together with descriptors based on derivatives b_k's, it provides a noticeable increase of the accuracy and other indicators (the Cohen and MCC) in comparison to applying the *LogR* classifier to classic descriptors a_k's,
- also the *SVM* classifier performs better on b_k's than on a_k's, but the improvements are less spectacular,
- only slight improvements, but pertaining all of the indicators, are visible when the decision trees, random forests and 5 NN classifiers are applied,
- somewhat unexpectedly to the authors, the gradient boosted trees classifier provided a slightly worse results when applied to b_k's descriptors, in other words, the *gbTr* was not able to take advantages from the derivative based descriptors.

5   Conclusions

The new way of learning descriptors of functional data is proposed and investigated from the view-point of the classification accuracy. Its essence is in learning descriptors of a curve derivative, without estimating it directly. Extensive simulations indicate that using these descriptors one may expect a better classification accuracy, but the improvement is essential when simultaneously an appropriate classifier of these descriptors is used. In the case study of accelerometer data, the proper choice was the logistic regression classifier, followed by the SVM.

The results are promising, but further efforts are necessary to reveal an influence of a kind of functional data on the choice of the classifier.

One of possible directions of generalizations of the proposed approach is to allow curves having derivatives with a finite number of jumps. Before learning their descriptors, it would be necessary to smooth samples in a jump-preserving way, as it was proposed in [11].

References

1. Abdulla, L., Al-Ani, M.: A review study for electrocardiogram signal classification. UHD J. Sci. Technol. **4**(1), 103–117 (2020). https://doi.org/10.21928/uhdjst.v4n1y2020.pp103-117. http://journals.uhd.edu.iq/index.php/uhdjst/article/view/711

2. Ahsan, M.R., Ibrahimy, M.I., Khalifa, O.O., et al.: EMG signal classification for human computer interaction: a review. Eur. J. Sci. Res. **33**(3), 480–501 (2009)

3. Azlan, W.A., Low, Y.F.: Feature extraction of electroencephalogram (EEG) signal - a review. In: 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), pp. 801–806 (2014). https://doi.org/10.1109/IECBES.2014.7047620

4. Gandhi, T., Panigrahi, B.K., Anand, S.: A comparative study of wavelet families for EEG signal classification. Neurocomputing **74**(17), 3051–3057 (2011)

5. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans. Neural Syst. Rehabil. Eng. **11**(2), 141–144 (2003). https://doi.org/10.1109/TNSRE.2003.814441

6. Harris, T., Tucker, J.D., Li, B., Shand, L.: Elastic depths for detecting shape anomalies in functional data. Technometrics 1–11 (2020)

7. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. **4**(2), R1–R13 (2007). https://doi.org/10.1088/1741-2560/4/2/r01

8. Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A.: Functional data analysis of amplitude and phase variation. Stat. Sci. **30**(4), 468–484 (2015). https://doi.org/10.1214/15-STS524

9. Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A.: Functional data analysis of amplitude and phase variation. Stat. Sci. **30**, 468–484 (2015)

10. Mironovova, M., Bíla, J.: Fast Fourier transform for feature extraction and neural network for classification of electrocardiogram signals. In: 2015 Fourth International Conference on Future Generation Communication Technology (FGCT), pp. 1–6 (2015). https://doi.org/10.1109/FGCT.2015.7300244

11. Pawlak, M., Rafajłowicz, E.: Jump preserving signal reconstruction using vertical weighting. Nonlinear Anal. Theory, Methods Appl. **47**(1), 327–338 (2001)

12. Rafajlowicz, E.: Nonparametric least squares estimation of a regression function. Statistics **19**(3), 349–358 (1988)

13. Rutkowski, L.: A general approach for nonparametric fitting of functions and their derivatives with applications to linear circuits identification. IEEE Trans. Circu. Syst. **33**(8), 812–818 (1986). https://doi.org/10.1109/TCS.1986.1086001

14. Rutkowski, L., Rafajłowicz, E.: On optimal global rate of convergence of some nonparametric identification procedures. IEEE Trans. Automatic Control **AC-34**, 1089–1091 (1989)

15. Skubalska-Rafajłowicz, E., Rafajłowicz, E.: Classifying functional data from orthogonal projections - model, properties and fast implementation. In: International Conference on Computational Science, Krakow, Poland (2021, accepted)

16. Srivastava, A., Klassen, E., Joshi, S.H., Jermyn, I.H.: Shape analysis of elastic curves in Euclidean spaces. IEEE J. Sel. Areas Commun. **10**(2), 391–400 (1992). https://doi.org/10.1109/49.126990

17. Więckowski, J.: Data from vibration in SchRs1200, Mendeley Data, V1. http://dx.doi.org/10.17632/htddgv2p3b.1. Accessed Jan 2021

18. Więckowski, J., Rafajlowicz, W., Moczko, P., Rafajlowicz, E.: Data from vibration measurement in a bucket wheel excavator operator's cabin with the aim of vibrations damping. Data Brief, 106836 (2021)

19. Xie, W., Chkrebtii, O., Kurtek, S.: Visualization and outlier detection for multivariate elastic curve data. IEEE Trans. Visual. Comput. Graph. **26**(11), 3353–3364 (2020). https://doi.org/10.1109/TVCG.2019.2921541