Learning Functional Descriptors Based on the Bernstein Polynomials – Preliminary Studies

Wojciech Rafajłowicz[1(✉)], Ewaryst Rafajłowicz[1],
and Jędrzej Więckowski[2]

[1] Faculty of Information and Comunication Technology, Wroclaw University
of Science and Technology, Wyb. Wyspianskiego 27, 50 370 Wrocław, Poland
{wojciech.rafajlowicz,jedrzej.wieckowski}@pwr.edu.pl

[2] Faculty of Mechanical Engineering, Wroclaw University of Science and Technology,
Wyb. Wyspianskiego 27, 50 370 Wrocław, Poland
ewaryst.rafajlowicz@pwr.edu.pl

Abstract. We propose a new method of constructing and estimating descriptors for classifying functional data. These descriptors are based on Bernstein polynomials and their estimation is based on noisy samples of a function (signal) to be classified.

The next step is to select an appropriate classifier, well suited to these descriptors. Although the result can be case dependent, we provide the methodology of running comparisons. As a vehicle for presenting the results, we choose benchmark data published in [32]. They represent shocks and vibrations of the operator's cabin in a large mechanical structure.

Keywords: Functional data classification · Machine learning · Estimating descriptors · Bernstein polynomials

1 Introduction

Descriptors of functional data, for example signals and curves, are created to extract features from the data to provide high-quality classification. At the same time, the descriptors should provide a significant degree of compression of the functional data, allowing it to be stored in computer memory in a cost-effective manner.

Approaches to creating descriptors can be divided into two large groups. The first includes methods tailored to a specific class of signals. These methods make significant use of specialized knowledge about a particular class of signals and their specific characteristics. A classic example of this class of methods is the recognition of electrocardiogram (ECG) signals based on so-called Q, R, S waveforms. We refer the reader to the following recent papers [1,17] [3] on classifying ECG signals. Specialized methods, dedicated to feature selection from electroencephalogram (EEG) signals, are developed and surveyed in [10,11], while in [2]

L. Rutkowski et al. (Eds.): ICAISC 2022, LNAI 13588, pp. 310–321, 2023.
https://doi.org/10.1007/978-3-031-23492-7_27

one can find the survey on electromyography signals. In [33] a representative artificial intelligence (AI) method applied to acoustic signals is described. A common feature of application-specific feature extraction methods is that they are highly labour intensive, which is justified mainly by applications in sensitive fields such as medicine. The second group of descriptor generation methods aims to significantly automate the feature extraction process for pattern, signal and image recognition. The expected result is a significant reduction in labor intensity, while maintaining satisfactorily high classification quality that is sufficient for applications in less demanding areas, for example, in technology and manufacturing processes.

Descriptors for Functional Data. The first examples of applications of methods from this group date back to the 1960 s s and are related to the development of algorithms known collectively as Fast Fourier Transform (FFT). In recent years, there has been renewed interest in this class of feature extraction methods due to the emergence of functional data classification methods. A special subclass within this group of methods are approaches that require the classifier to be sensitive to the shapes of the functions (signals) being classified. We refer the reader to [12,16,29,34] for more details on such approaches and to [23] for the latest contribution.

In these papers, the primary tool for obtaining the sensitivity of algorithms to the shape of signals is to consider the waveforms of their derivatives.

Advantages of Applying Bernstein Polynomials. In contrast, the approach proposed in this work is based on obtaining shape-sensitive descriptors of signals by comparing them with elements of the function space basis that have shape-preserving features. The best-known basis with these properties is that spanned by Bernstein polynomials. In the theoretical version of the proposed method, this comparison is implemented by computing scalar products between the signal to be classified and successive Bernstein polynomials. These products attain high values when individual signal (function) fragments are well matched to successive Bernstein polynomials and, conversely, the values are small when a given signal fragment is orthogonal to successive polynomial Bernsteins. For this reason, we choose these products (after possible normalization) as descriptors sensitive to signal shapes.

The question of whether to normalize descriptors or to use only non-normalized scalar products has no clear answer. In situations where the signal amplitudes vary considerably between classes, normalization is not advisable. On the other hand, when membership of a signal to a given class is determined only by its shape, the use of normalization will be useful.

Why is Learning Needed? In practice, we usually do not have a signal at all points of the observation interval, but only its samples, taken most often at equidistant moments of time, and observed with random disturbances. For these reasons, the process of learning the features of this signal is needed. In fact, we need to apply descriptor learning in two situations. The first one appears when we extract signal descriptors contained in the training sequence. The second one is needed when – after learning the classifier – we acquire new signals to be classified. In the first case, a learning process can be more accurate, since it is usually performed off-line. In the second one, it can be desirable (or necessary) to learn descriptors on-line.

Assumptions. A common feature of all approaches to the construction of classifiers for functional data is the assumption of statistical repeatability of signals and their (dis-)similarities when they come from the same or different classes. Since the description of probability distributions in function spaces is complex, in this paper we will make the simplifying assumption that we describe the probability distributions of signals of particular classes as finite-dimensional distributions of the coefficients of the expansion of that signal into a series of Bernstein polynomials of given degree $N > 1$. We refer the reader to [23,24] and [28] for a more advanced model of imposing a probability structure on random functions.

The well-known Weierstrass theorem on the approximation of a continuous function on a finite interval by a polynomial of a sufficiently high degree can serve as a justification of this assumption. Bernstein polynomials form the basis of a constructive proof of this theorem.

We emphasize that knowledge of these probability distributions is not assumed in this paper. On the contrary, we only assume their existence and the complete lack of knowledge about them. Thus, the proposed approach is non-parametric, even though it deals with a finite number $(N + 1)$ of parameters, since this number can be chosen depending on the number of observations n and can grow with it.

Our Approach. In summary, the method proposed in this work to construct classifiers for functional data consists of two steps. In the first one, we learn vectors of Bernstein descriptors for each class, based on the learning sequence. In the second stage, we select a descriptor classification method from among known algorithms in such a way that it gives a satisfactory classification quality for a given application.

Other Approaches Based on Bernstein Polynomials. Another approach to constructing classifiers based on Bernstein polynomials was proposed in [21]. The difference is that in [21] Bernstein polynomials were used to estimate the probability densities of the classes. Classifiers or predictors acting as neural networks based on Bernstein polynomials are constructed in a similar way (cf. [18,30]). Advantages of using Bernstein polynomials occurred to be useful in estimating quantile functions [19]. Recently, an interesting application of Bernstein polynomials to modeling Covid-19 growth was proposed in [25].

Paper Organization. The paper is organized as follows. The next section presents the basic properties of Bernstein polynomials that are needed later in the paper. In Sect. 3, we formulate the assumptions and pose the descriptor learning problem. We present the learning algorithm itself and its elementary properties in Sect. 4. In that section we also describe the interaction of this algorithm with the decision function of the classifier. We then illustrate the selection of the decision function using the example of classification of the acceleration signals of the excavator operator's cab as a function of ground hardness.

2 Descriptors Based on the Bernstein Polynomials

We refer the reader to [5,7,14] for classic and more recent results on Bernstein polynomials (BP) and to [6] for their application to nonparametric estimation of probability density functions (p.d.f.).

It should be emphasized that Bernstein polynomials do not form an orthogonal basis, but many formulas are similar to those typical for nonparametric estimation methods based on orthogonal expanssions (see, e.g., [20,26,27] and the bibliography cited therein).

Definition and Elementary Properties of Bernstein Polynomials
Bernstein polynomials are usually defined on the interval $X = [0, 1]$. Further in this paper we will assume that also all other functions considered here are defined on X.

For $x \in X$ k-th of order $N \geq k$ the Bernstein polynomial, denoted as $B_k^{(N)}(x)$, is defined as follows

$$B_k^{(N)}(x) = \binom{N}{k} x^k (1-x)^{N-k}, \quad k = 0, 1, \ldots, N.$$

We extend this definition by setting $B_k^{(N)}(x) \equiv 0$, if $k < 0$ or $k > N$.

We summarize and comment on the following, well-known, properties of the BPs.

$$\forall x \in X \quad \sum_{k=0}^{N} B_k^{(N)}(x) = 1, \quad 0 \leq B_k^{(N)}(x) \leq 1. \tag{BP 1}$$

Observe that (BP 1), being a partition of the unity, implies the ability of the BPs to restore constants exactly. Indeed, it suffices to set $a_k = 1$ for all k in formula (1) below.

For each sequence $a_k \in R, \quad k = 0, 1, \ldots, N$ the following function

$$w_N(x) = \sum_{k=0}^{N} a_k \cdot B_k^{(N)}(x) \tag{1}$$

is an N - th order polynomial in x. Let f be a continuous function on X. Then, it is well known that selecting $a_k = f(k/(N+1))$, $k = 0, 1, \ldots, N$ in (1) we obtain $w_N(x) \to f(x)$, uniformly in X, as $N \to \infty$.

The following expression is of importance for a proper scaling of integrals containing the BPs

$$\int_X B_k^{(N)}(x) dx = (N+1)^{-1}, \quad k = 0, 1, \ldots, N. \tag{BP 2}$$

Proposed Descriptors
Let $C^p(X)$, $p = 0,\ 1,\ 2\ldots$ denote the space of p-times differentiable functions in X with the convention that $C(X) = C^0(X)$ is the space of all functions that are continuous in X. Define the inner product

$$\forall f,\, g \in C(X) \qquad < f, g > = \int_X f(x)\, g(x)\, dx. \tag{2}$$

As descriptors of function (signal) $f \in C(X)$, denoted further as $d_k(f)$ (or d_k for brevity), we propose to take

$$d_k(f) \,=\, (N+1) < f, B_k^{(N)} >= \tag{3}$$

$$= \,(N+1) \int_X f(x)\, B_k^{(N)}(x)\, dx, \quad k = 0,\, 1,\, \ldots\,, N.$$

Note that $d_k(f)$'s depend also on N, but this dependence is not displayed, unless necessary.

It is worth mentioning also the normalized version of these descriptors, denoted further as $\breve{d}_k(f)$, that for $f \in C(X)$ is defined as follows

$$\breve{d}_k(f) \,=\, \frac{(N+1)\ < f, B_k^{(N)} >}{\max_{x\in X} |f(x)|}, \quad k = 0,\, 1,\, \ldots\,, N. \tag{4}$$

Note that $\breve{d}_k(f)$ is well defined, since for $f \in C(X)$ the maximum in the compact set X is attained. Furthermore,

$$\forall\, f \in C(X) \quad -1 \le \breve{d}_k(f) \le 1 \tag{5}$$

and $\breve{d}_k(f) = \pm 1$ for $f(x) = \pm 1$, $x \in X$. To prove this fact, it suffices to observe that

$$|< f, B_k^{(N)} >| \,=\, |\int_X f(x)\, B_k^{(N)}(x)\, dx| \le \int_X |f(x)|\, B_k^{(N)}(x)\, dx \le \tag{6}$$

$$\le \max_{x\in X} |f(x)| \int_X \, B_k^{(N)}(x)\, dx = \max_{x\in X} |f(x)|/(N+1),$$

due to (BP1) and (BP2).

Additionally, $\breve{d}_k(f) = 0$, if f is orthogonal to $B_k^{(N)}$. Thus, $\breve{d}_k(f)$'s are descriptors that are well suited for classification problems. One can interpret descriptors $d_k(f)$ and $\breve{d}_k(f)$ as indicators to what extent f is close to (or fits) $B_k^{(N)}$. Note that $d_k(f)$ and $\breve{d}_k(f)$ depend also on N, but this dependence is not displayed, unless necessary.

Sensitivity of Descriptors to Function Shapes

These descriptors are – to some extent – shape sensitive in the sense that is explained below. Our starting point is the following well-known – formula for iterative calculations of the derivative, denoted as D_x, of $B_k^{(N)}(x)$

$$D_x B_k^{(N)}(x) = N \cdot [B_{k-1}^{(N-1)}(x) - B_k^{(N-1)}(x)], \;\; k = 0, 1, .., N. \tag{BP 3}$$

Then, multiplying both sides of (BP 3) by $f \in C^1(X)$, integrating over X with the aid of the integration by parts (for $1 \le k \le (N-1)$ we have $B_k^{(N)}(0) = B_k^{(N)}(1) = 0$) and shifting index k we immediately obtain

$$< D_x f, B_{k+1}^{(N+1)} > \,=\, d_{k+1}(f) - d_k(f) \quad k = 0, 1, \ldots, (N-1). \tag{7}$$

These relationships can be interpreted as follows: if f is strictly increasing (decreasing) in X, then the left-hand side of (7) is positive (negative). Thus also the difference $d_{k+1}(f) - d_k(f)$ retains this property. In other words, if f is strictly increasing (decreasing) in X then also the sequence of $d_k(f)$ is, and this statement holds in a natural way, i.e., without having a priori knowledge or our intervention by imposing constraints. Dividing both sides of (7) by $\max_{x\in X}|f(x)|$ we conclude that this monotonicity preserving property holds also for the normalized descriptors $\breve{d}_k(f)$'s.

Assuming that $f \in C^2(X)$ and repeating the similar reasoning for $D_x^2 f(x)$, we come to the conclusion that if $D_x^2 f(x) > 0$, $x \in X$, which implies the convexity of f, then also sequences $d_k(f)$'s and $\breve{d}_k(f)$'s are also convex in the sense that their second order differences are positive.

These properties, important for classification of the descriptor sequence, are illustrated in Fig. 1.

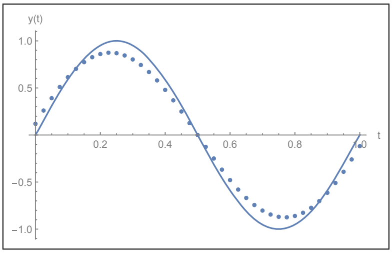

Fig. 1. Descriptors $d_k(f)$'s (dots) for $N = 50$ of function $f(x) = \sin(2\,\pi\,x)$, $x \in X$.

3 Learning Descriptors from Noisy Samples of Functional Data

In practice, the data is not available in functional form $f \in C(X)$, which means that the proposed descriptors cannot be computed directly. Most often we only have samples of f values, observed with noise. We adopt a standard description of this type of sampling, assuming that the samples are taken at equidistant points x_i (e.g., instants of time or spatial variables), with random additive perturbations ϵ_i, $i = 1, 2, \ldots, n$. We assume that these disturbances have zero expected values and finite variances. For simplicity, we assume that these variances are equal, and denote them by $0 < \sigma^2 < \infty$. In summary, the functional data samples y_i, $i = 1, 2, \ldots, n$ are of the form

$$y_i \;=\; f(x_i) \,+\, \epsilon_i,\; \mathbb{E}\epsilon_i = 0,\; \mathbb{E}\epsilon_i^2 = \sigma^2 < \infty, \qquad i \;=\; 1,\, 2, \ldots,\, n, \tag{8}$$

$\mathbb{E}[\epsilon_i\,\epsilon_j] = 0$ for $i \neq j$, where $\mathbb{E}$ is the expectation of a random variable.

Problem statement: having observations $(x_i,\, y_i)$, $i\,=\,1,\,2,\ldots,\,n$ at our disposal, our aim is to propose a learning algorithm for estimating descriptors $d_k(f)$, $k=0,1,\ldots,N$. For the sake of simplicity we assume that the original sample points are already transformed to $x_i \in [0,\,1]$ and $\Delta_n \overset{def}{=} x_{i+1}-x_i = 1/n$.

In the remainder of this paper, we will denote the descriptor estimates as $\hat{d}_k^{(n)}(\bar{y})$, $k=0,1,\ldots,N$, where $\bar{y}$ is a column vector of ordered observations y_i, $i\,=\,1,\,2,\ldots,\,n$ with possible upper indices when several functional elements f are considered.

According to (3), a natural and simple algorithm for $\hat{d}_k^{(n)}(\bar{y})$ is of the form

$$\hat{d}_k^{(n)}(\bar{y}) \;=\; \frac{N+1}{n}\sum_{i=1}^{n} y_i\, B_k^{(N)}(x_i), \quad k=0,1,\ldots,N. \tag{9}$$

Notice that noisy observations y_i's are directly inserted into (9) without any prefiltering (see [22] for a discussion on the advantages of using pre- or post-filtering). Nevertheless, $\hat{d}_k^{(n)}(\bar{y})$ still have satisfactory statistical properties, as stated below. One can consider more robust estimators of the expectation, e.g., the median or the trimmed mean in (9), but here we confine our attention to the classic mean, since Bernstein polynomials have a hidden ability to mitigate large errors.

Notice that for the bias $\delta_{kn} \overset{def}{=} d_k(f) - \mathbb{E}[\hat{d}_k^{(n)}(\bar{y})]$ we have

$$\delta_{kn} \;=\; (N+1)\,\Delta_n \sum_{i=1}^{n}[f(\tilde{x}_{ki})\,B_k^{(N)}(\tilde{x}_{ki}) - f(x_i)\,B_k^{(N)}(x_i)], \tag{10}$$

where $\tilde{x}_{ki}$'s are intermediate points in $I_i \overset{def}{=} [x_i-\Delta_n/2,\, x_i+\Delta_n/2]$ when the mean value theorem is applied to the integrals

$$\int_{I_i} f(x)\,B_k^{(N)}(x)\,dx = \Delta_n\, f(\tilde{x}_{ki})\,B_k^{(N)}(\tilde{x}_{ki}).$$

Lemma 1. *If f has a continuous derivative in $[0,\,1]$, then $|\delta_{kn}| = O(N/n)$ and the learning algorithm $\hat{d}_k^{(n)}(\bar{y})$ is asymptotically unbiased, as $n\to\infty$, for each finite and fixed N, $k=0,\,1,\ldots,\,N$.*

Indeed, the modulus of each summand in (1) is bounded by Δ_n multiplied by by the maximum over $[0,\,1]$ of the modulus of the derivative of $f(x)\,B_k^{(N)}(x)$, which – in turn – is bounded by

$$\max_x |f(x)| + N\,\max_x |f'(x)|.$$

due to BP3). This finishes the proof, since this bound is uniform in k.

For the variance of $\hat{d}_k^{(n)}(\bar{y})$ we have for $k=0,1,\ldots,N$

$$\mathbb{VAR}[\hat{d}_k^{(n)}(\bar{y})] \;=\; \left(\frac{N+1}{n}\right)^2\,\mathbb{E}\left[\sum_{i=1}^{n}\epsilon_i\,B_k^{(N)}(x_i)\right]^2 \le \sigma^2\,\frac{(N+1)^2}{n}, \tag{11}$$

due to the uncorrelatedness of ϵ_i's and the fact that $0 \leq B_k^{(N)}(x) \leq 1$.

Lemma 2. *Under the assumptions of Lemma 1, $\hat{d}_k^{(n)}(\bar{y})$'s are consistent in the mean squared error (MSE) sense as $n \to \infty$, for each finite and fixed N, $k = 0, 1, \ldots, N$.*

Indeed, it is well known that the MSE can be expressed as the sum of the variance and the squared bias. Thus, the result follows directly for Lemma 1 and (11).

Notice that the above results hold also in the case when f is a random element and descriptors $d_k(f)$'s are random variables. To this end, it suffices to consider the expectations as conditional ones, given $d_k(f)$'s.

4 Learning Classifiers Based on Bernstein Descriptors

We assume that random element f is drawn from a (sub-)class of continuously differentiable functions $\mathcal{F} : X \to R$. Two nonempty subsets $\mathcal{F}_I$ and $\mathcal{F}_{II}$ are distinguished in $\mathcal{F}$ and f is drawn from one of them with a priori probabilities $p_I > 0$, $p_{II} > 0$, respectively, and $p_I + p_{II} = 1$. These probabilities are unknown, but their estimation by fractions in the learning sequence is a simple task, unless there is no large imbalance between samples from class I and II in a learning sequence.

f is represented by random vector vector $\bar{d}(f)$ of its descriptors $d_k(f)$, $k = 0, 1, \ldots, N$, assuming fixed $N > 1$. Its choice is discussed later on. Probability distributions of $\bar{d}(f)$ depend on whether f is from class I or II, but they are unknown. Also $\bar{d}(f)$ is not directly available.

The only information that we have at our disposal is contained in a learning sequence, which is of the form:

$$\mathcal{L}_L \overset{def}{=} \{(\bar{y}^{(1)}, j_1), (\bar{y}^{(2)}, j_2), \ldots, (\bar{y}^{(L)}, j_L)\}, \tag{12}$$

where $j_k \in \{I, II\}$ are class labels, assumed to be correct, while $\bar{y}^{(k)}$ are vectors of equidistant samples from random elements $f^{(k)}$, drawn either from $\mathcal{F}_I$ or $\mathcal{F}_{II}$. These samples are taken at x_i, $i = 1, 2, \ldots, n$, according to (8), $k = 1, 2, \ldots, L$.

Now, our aim is to present an algorithm of learning, tuning, testing and selecting a classifier that classifies a random element f to classes I or II, based on its random samples $\bar{y}$ and using the estimates of the Bernstein descriptors.

To this end, let us denote by

$$cl.\,parameters = \mathbb{LEARN}[cl.\,name,\; learning\,seq.]$$

a generic learning procedure that takes a classifier name and a learning sequence as its inputs and provides tuning parameters of the classifier after learning as its outputs.

As *cl. name* one may select, e.g., one of the frequently used classifiers listed in Table 1 or even an ensemble of classifiers. We denote such a class of considered classifiers as $\mathcal{CL}$. The second tool that we need is a testing procedure:

Table 1. Examples of frequently used classifiers.

Acronym	Classifier
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
k-NN	The k nearest neighbors classifier

$$\{accuracy,\, precision, \ldots\} = \mathbb{TEST}[cl.\ name,\, parameters,\, testing\ seq.]$$

that takes the classifier name, its parameters and a testing sequence as inputs. Its output is a list of commonly used indicators of classifiers' quality, e.g., the accuracy, precision, recall, specificity, F1 and possibly many others. The $\mathbb{TEST}$ runs in a standard way, namely, it the selected classifier (with *parameters* from the learning procedure) on a supplied *testing sequence* and calculates the quality indicators. In a more advanced version, the testing inside $\mathbb{TEST}$ is performed many times on randomly selected subsequences of the *testing sequence* and the resulting indicators are averaged. It is further assumed that the $\mathbb{TEST}$ procedure is used in this more advanced version.

Selection and Learning a Classifier Based on Bernstein Descriptors

Data acquisition Collect samples of random elements, ask an expert to classify them and form learning sequence $\mathcal{L}_L$

Learning descriptors Select the order N of Bernstein descriptors. For the vector of samples $\bar{y}^{(l)}$ in $\mathcal{L}_L$ estimate the elements of the following list:

$$\bar{d}(\bar{y}^{(l)}) \overset{def}{=} \{\hat{d}_k^{(n)}(\bar{y}^{(l)}),\ k = 0, 1, \ldots, N\}, \tag{13}$$

using (9). To each $\bar{d}(\bar{y}^{(l)})$ attach label j_l that corresponds to $\bar{y}^{(l)}$ in $\mathcal{L}_L$ and form a transformed learning sequence $\mathcal{D}_L$ from pairs $(\bar{d}(\bar{y}^{(l)}), j_l)$, $l = 1, 2, \ldots, , L$.

Optional step $\mathcal{D}_L$ augmentation. Extend $\mathcal{D}_L$ by copying each of its elements $\eta > 1$ times and replacing $\bar{d}(\bar{y}^{(l)})$ vectors by their randomly perturbed copies with zero mean, but keeping the same class label. Perturbations by additive Gaussian random vectors are the first choice. Slightly abusing the notation, we shall further denote this extended learning sequence again by $\mathcal{D}_L$.

Preparations Select classifier CL_{cur} from $\mathcal{CL}$ and split at random $\mathcal{D}_L$ into two disjoint and covering $\mathcal{D}_L$ sets: tuning set $\mathcal{DL}_{L1}$ and testing set $\mathcal{DT}_{L2}$, $L1 + L2 = L$.

Learning Run $CL_{cur}\ parameters = \mathbb{LEARN}[CL_{cur}, \mathcal{DL}_{L1}]$.

Testing and Validation. Run the testing procedure:

$$\{accuracy,\, precision, \ldots\} = \mathbb{TEST}[CL_{cur},\, CL_{cur}\ parameters,\, \mathcal{DT}_{L2}]$$

```
and decide whether the quality indicators are satisfactory.
IF YES – STOP and provide $CL_{cur}$, $CL_{cur}$ parameters as the final result.
OTHERWISE
IF the admissible number of trials to select a proper classifier is not reached, then GO TO the Preparations step.
OTHERWISE
IF $N < n$ increase $N$ and GO TO the Learning descriptors step.
OTHERWISE
Declare the failure of the learning process and STOP.
```

If failure occurred, one may consider enlarging the number of observations n and/or extending the set of considered classifiers.

Testing on Samples from Shocks and Vibrations

Operators' cabins of large working machines repetitively undergo shocks and vibrations (see [31] for examples of signals of this kind and [32] for their interpretation). The data in [31] consists of $N = 43$ curves, sampled at $n = 1024$ equidistant points each. An expert assigned label I (light working conditions) or II (heavy working conditions) to each series of signal samples.

An optional step – data augmentation was applied, providing $\mathcal{D}_L$ with $L = 43000$. This was done by adding the Gaussian noise with zero mean and the disperssion 0.05 to the estimates obtained in the learning descriptors step of the algorithm.

The algorithm of learning and selecting good classifiers was run on the augmented data. The list of tested classifiers is presented in Table 1. Only two of them, namely the logistic regression and the SVM provided accuracy larger than 0.95 (for the LogR – 0.98 and for the SVM –0.951 were obtained). Other quality indicators of these classieifers were high: the recall was larger than 0.98 in both cases, the precision attained by the LogR was 0.98 and 0.93, respectively, by the SVM. The Cohen kappa coefficient was equal to 0.96 for the LogR and 0.9 for the SVM.

Conclusions and Possible Extensions. Summarizing, the proposed approach of selecting the descriptors based on Bernstein polynomials and testing an adequate classifier occurred to be successful in classifying functional data from their noisy samples.

These descriptors can also be used for estimating an observed signal by applying the following kernel $\mathcal{K}(x,\, x') \stackrel{def}{=} (N+1)\sum_{k=0}^{N} B_k^{(N)}(x)\, B_k^{(N)}(x')$, $x,\, x' \in X$. Although kernel $\mathcal{K}$ has different properties than those typically used in nonparametric regression estimation by Parzen kernel methods, it can be applied for change detection problems in a similar way as it was recently proposed in [8,9]. Our descriptors can also be used as a part of generating signature hybrid descriptors in a way similar to the one proposed recently in [35]. Another way of their applications include novelty detection in ways found fruitful in [13] and [24].

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.v4n1y
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. Augustyniak, P., Tadeusiewicz, R.: Optimization of ECG procedures chain for reliability and data reduction. In: Ubiquitous Cardiology: Emerging Wireless Telemedical Appl. pp. 202–227. IGI Global (2009)
4. 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
5. Chen, W.. Ditzian, Z.: Best polynomial and Durrmeyer approximation in $L_p(S)$ Indagationes Mathematicae **2**, 437–452 (1991)
6. Ciesielski, Z.: Nonparametric polynomial density estimation. Probab. Math. Stat. **9**(1), 1–10 (1988)
7. Derrienic, M.M.: On multivariate approximation by Bernstein-type polynomials. J. Approxim. Theory **45**, 155–166 (1985)
8. Galkowski, T., Krzyzak, A., Filutowicz, Z.: A new approach to detection of changes in multidimensional patterns. J. Artifi. Intell. Soft Comput. Res. **10**(2), 125–136 (2020). https://doi.org/10.2478/jaiscr-2020-0009
9. Galkowski, T., Krzyzak, A., Patora-Wysocka, Z., Filutowicz, Z., Wang, L.: A new approach to detection of changes in multidimensional patterns - Part II. J. Artifi. Intell. Soft Comput. Res. **11**(3), 217–227 (2021). https://doi.org/10.2478/jaiscr-2021-0013
10. Gandhi, T., Panigrahi, B.K., Anand, S.: A comparative study of wavelet families for EEG signal classification. Neurocomputing **74**(17), 3051–3057 (2011)
11. 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
12. Harris, T., Tucker, J.D., Li, B., Shand, L.: Elastic depths for detecting shape anomalies in functional data. Technometrics **63**(4), 466–476 (2020)
13. Homenda, W., Jastrzebska, A., Pedrycz, W., Yu, F.: Combining classifiers for foreign pattern rejection. J. Artifi. Intell. Soft Comput. Res. **10**(2), 75–94 (2020). https://doi.org/10.2478/jaiscr-2020-0006
14. Lorentz, G.G.: Bernstein Polynomials. American Mathematical Soc. (2013)
15. 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
16. 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
17. 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
18. Mohammad, A.J., Mohammad, I.J.: Summation-Integral Bernstein Type Of Neural Network Operators. Asian J. Math. Comput. Res. 74–86 (2017)

19. Pepelyshev, A., Rafajłłowicz, E., Steland, A.: Estimation of the quantile function using Bernstein-Durrmeyer polynomials. J. Nonpara. Stat. **26**(1), 1–20 (2014)
20. Rafajlowicz, E.: Nonparametric least squares estimation of a regression function. Statistics **19**(3), 349–358 (1988)
21. Rafajłowicz, E., Skubalska-Rafajłowicz, E.: Nonparametric regression estimation by Bernstein-Durrmeyer polynomials. Tatra Mt. Math. Publ. **17**, 227–239 (1999)
22. Pawlak, M., Rafajlowicz, E., Krzyzak, A.: Postfiltering versus prefiltering for signal recovery from noisy samples. IEEE Trans. Inf. Theory **49**(12), 3195–3212 (2003)
23. Rafajłowicz, W., Rafajłowicz, E.: Learning shape sensitive descriptors for classifying functional data. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) ICAISC 2021. LNCS (LNAI), vol. 12854, pp. 485–495. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87986-0_43
24. Rafajłowicz, W.: Learning novelty detection outside a class of random curves with application to covid-19 growth. J. Artifi. Intell. Soft Comput. Res. **11**(3), 195–215 (2021)
25. Rafajłowicz, W.: Learning Decision Sequences For Repetitive Processes—Selected Algorithms. SSDC, vol. 401. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-88396-6
26. Rutkowski, L.: A general approach for nonparametric fitting of functions and their derivatives with applications to linear circuits identification. IEEE Trans. Circ. Syst. **33**(8), 812–818 (1986). https://doi.org/10.1109/TCS.1986.1086001
27. Rutkowski, L., Rafajłowicz, E.: On optimal global rate of convergence of some nonparametric identification procedures. IEEE Trans. Autom. Control **AC-34**, 1089–1091 (1989)
28. Skubalska-Rafajłowicz, E., Rafajłowicz, E.: Classifying functional data from orthogonal projections – model, properties and fast implementation. In: Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) ICCS 2021. LNCS, vol. 12744, pp. 26–39. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77967-2_3
29. 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
30. Wang, C., Zhang, H., Fan, W., Fan, X.: A new wind power prediction method based on chaotic theory and Bernstein Neural Network. Energy **117**, 259–271 (2016)
31. Więckowski, J.: Data from vibration in SchRs1200, Mendeley Data, V1 (2021). http://dx.doi.org/10.17632/htddgv2p3b.1
32. 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. In: Data in Brief, p. 106836 (2021)
33. Wszolek, W., Tadeusiewicz, R., Chyla, A.: Recognition of selected helicopter types based on the generated acoustic signal with application of artificial intelligence methods. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 2001(4), pp. 1734–1737. Institute of Noise Control Engineering (2001)
34. Xie, W., Chkrebtii, O., Kurtek, S.: Visualization and outlier detection for multivariate elastic curve data. IEEE Trans. Visual Comput. Graphics **26**(11), 3353–3364 (2020). https://doi.org/10.1109/TVCG.2019.2921541
35. Zalasinski, M., Cpalka, K., Laskowski, L., Wunsch, D., Przybyszewski, K.: An algorithm for the evolutionary-fuzzy generation of on-line signature hybrid descriptors. J. Arti. Intell. Soft Comput. Res. **10**(3), 173–187 (2020). https://doi.org/10.2478/jaiscr-2020-0012