Applying classification methods for multivariate functional data

Tomasz Górecki, Mirosław Krzyśko, Waldemar Wołyński

Department of Mathematics and Computer Science
Adam Mickiewicz University, Poznań, Poland

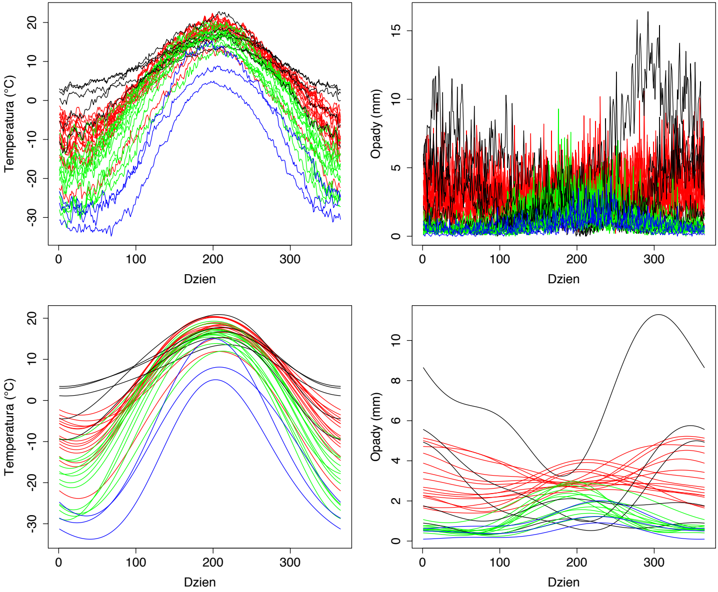
Conference of the International Federation of Classification Societies
San José 15-19.07.2024

Outline

Recently, methods for data representing functions or curves have received much attention. Such data are known in the literature as the functional data (Ramsay & Silverman, 2005). Examples of functional data can be found in several application domains, such as medicine, economics, meteorology, and many others. In many applications, statistical methods are needed for objects characterized by features observed at many time points (double multivariate data). Such data are called the multivariate functional data.

For multivariate functional data, various methods of classification are very often used. We have L different types of curves, and the aim is to classify a new function as one of the L types. Curve discrimination arises in many contexts and is a significant problem. A clear example is signal discrimination, which has been considered in several papers involving high-resolution radar returns for target detection or the recognition of speech signals.

Introduction



We recommend not to use classification methods in the original functional data space. For multivariate functional data, we construct the first discriminant coordinates (Górecki et al. (2018)). These coordinates are uncorrelated and have unit variances. This new space of functional discriminant coordinates is a very convenient space in which we can apply various classification methods.

Our second recommendation is to take into account the shape of functional data. Functions have shapes, and shapes are represented by functions. The curvature of a plane curve at point $P(x_0, y_0)$ defined by the function $y = f(x)$ in the Cartesian system is equal to

$$\kappa = |y_0''|/(1 + y_0'^2)^{3/2}.$$

Intuitively, the curvature is the amount by which a curve deviates from being a straight line. We see that the definition of the curvature of the plane curve is based on the first and second derivatives of the function f. Hence, we recommend extending the functional data space to include the first and second derivatives of functions representing this data.

Assume that our data is divided into L groups of objects and that each object is characterized by the values of the pair $(Y, \boldsymbol{X})$, where Y is a discrete random variable called a label with values from the set $\{1, 2, \ldots, L\}$ and $\boldsymbol{X} \in L_2^p(I)$ is p-dimensional Hilbert space of square-integrable functions on the time interval $I = [a, b]$.

Representation of multivariate functional data

We take into account the case when the dth component $X_d \colon I \longrightarrow \mathbb{R}$ of the process $\boldsymbol{X}$ belongs to the class twice, continuously differentiable functions on the time interval I and is represented by a finite number of orthonormal basis functions $\{\varphi_b\}$:

$$X_d(t) = \sum_{b=0}^{B_d} c_{db}\varphi_b(t), \tag{1}$$

where c_{db} are random variables such that $\mathrm{E}(c_{db}) = 0$, $t \in I$, $d = 1, 2, \dots, p$.

Using formula (1), the process $\boldsymbol{X}$ can be written as:

$$\boldsymbol{X}(t) = \boldsymbol{\Phi}(t)\boldsymbol{c}, \;\; t \in I, \tag{2}$$

where $\boldsymbol{c} = (c_{10}, \ldots, c_{1B_1}, \ldots, c_{p0}, \ldots, c_{pB_p})^\top$,
$\boldsymbol{\Phi}(t) = \mathrm{diag}(\boldsymbol{\varphi}_{B_1}^\top(t), \ldots, \boldsymbol{\varphi}_{B_p}^\top(t))$, $\boldsymbol{\varphi}_{B_d}(t) = (\varphi_0(t), \varphi_1(t), \ldots, \varphi_{B_d}(t))^\top$,
$d = 1, 2, \ldots, p$.

We can estimate the vector $\boldsymbol{c}$ on the basis of n independent realisations $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in_i}$ from the ith class, $i = 1, 2, \ldots, L$, of the random process $\boldsymbol{X}$ (functional data). Details of the least squares method estimation of the random coefficients $\boldsymbol{c}_{db}$ can be found in, e.g., in Górecki et al. (2018).

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$, where

$$X_d(t) = \sum_{b=0}^{B_d} c_{db}\varphi_b(t),\ \ t \in I,\ \ d = 1, 2, \ldots, p.$$

We compute the first derivative of the process X_d:

$$X'_d(t) = \sum_{b=0}^{B_d} c_{db}\varphi'_b(t),\ \ t \in I,\ \ d = 1, 2, \ldots, p.$$

Dataset extension

Let x'_{dj} denote the value of the process X'_d at time t_j, where $t_j \in I$, $j = 1, 2, \ldots, J$. Then our data consist of J pairs (t_j, x'_{dj}), $j = 1, 2, \ldots, J$, $d = 1, 2, \ldots, p$. This discrete data can be smoothed using a function:

$$\hat{X}'_d(t) = \sum_{b=0}^{B_d} e_{db}\varphi_b(t), \;\; t \in I, \;\; d = 1, 2, \ldots, p.$$

Then, we compute the second derivative of the process X_d:

$$X''_d(t) = \sum_{b=0}^{B_d} c_{db}\varphi''_b(t), \;\; t \in I, \;\; d = 1, 2, \ldots, p.$$

Let x''_{dj} be the value of the process X''_d at time t_j, where $t_j \in I$, $j = 1, 2, \ldots, J$.

Dataset extension

Now our data includes J pairs (t_j, x''_{dj}), $j = 1, 2, \ldots, J$, $d = 1, 2, \ldots, p$. This discrete data can be smoothed using a function:

$$\hat{X}''_d(t) = \sum_{b=0}^{B_d} h_{db}\varphi_b(t), \ \ t \in I, \ \ d = 1, 2, \ldots, p.$$

Finally, we add the information provided by derivatives to a pure p-multivariate process $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$, obtaining the extended process:

$$\boldsymbol{Z} = (X_1, X_2, \ldots, X_p, X'_1, X'_2, \ldots, X'_p, X''_1, X''_2, \ldots, X''_p)^\top.$$

From the formula (2), the estimates of independent realisations $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in_i}$ of the process $\boldsymbol{X}$ in the ith group have the form:

$$\hat{\boldsymbol{x}}_{ij}(t) = \boldsymbol{\Phi}(t)\hat{\boldsymbol{c}}_{ij}, \;\; t \in I, \;\; j = 1, 2, \ldots, n_i, \;\; i = 1, 2, \ldots, L.$$

For this functional data, we construct functional discriminant coordinates. We get $s = \min(B_1 + \cdots + B_p + p, L - 1)$ uncorrelated functional coordinates with unitary variances. This new s-dimensional space of functional discriminant coordinates is a very convenient classification space using a variety of classifiers. We can replace the p-dimensional process $\boldsymbol{X}$

with $3p$-dimensional extended process $\boldsymbol{Z}$ and proceed analogously.

Classifiers

In the s-dimensional vector space of functional discriminant coordinates, we take into account the following classifiers:

1. a classifier of k-nearest neighbors (kNN),
2. Naive Bayes classifier (NB),
3. decision trees (DT),
4. the support vector machine (SVM),
5. random forest (RF),
6. and XGBoost.

Classifiers

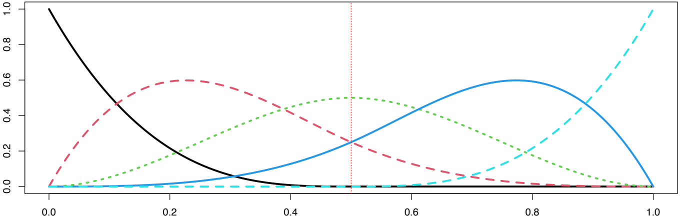
The percentage of correct classifications can be calculated for each of the six classifiers. The classification can be performed on the functional data related to the process $\boldsymbol{X}$ or the functional data associated with the extended process $\boldsymbol{Z}$. Since the data related to the extended $\boldsymbol{Z}$ process additionally contains information about the shape of the function, it should be expected that the classification performed on these data will contain fewer errors.

Datasets

Our experiments used time series data from the UEA MTSC archive (Bagnall et al. (2018)). Each dataset was divided into training and test sets. For this reason, we adopted the classification error rate on the test set as a quality measure.

Name	Train size	Test size	Dims	Length	Classes
AtrialFibrillation	15	15	2	640	3
BasicMotions	40	40	6	100	4
Epilepsy	137	138	3	206	4
EthanolConcentration	261	263	3	1751	4
ERing	30	270	4	65	6
FingerMovements	316	100	28	50	2
HandMovementDirection	160	74	10	400	4
JapaneseVowels	270	370	12	29	9
Libras	180	180	2	45	15
NATOPS	180	180	24	51	6
RacketSports	151	152	6	30	4
SelfRegulationSCP1	268	293	6	896	2
SelfRegulationSCP2	200	180	7	1152	2
StandWalkJump	12	15	4	2500	3
UWaveGestureLibrary	120	320	3	315	8

All calculations were performed in the R environment using the `fda` and `caret` packages. All classifier parameters were tuned automatically with the default settings of `caret` library. All hyperparameters were found using 10-fold cross-validation (10CV). During calculations, we used B-spline basis functions. B-spline basis functions have the advantages of speedy computation and great flexibility.



The first five B-spline basis functions on the interval [0, 1].

Mean classification accuracies (over 15 datasets) for selected classifiers. D states for derivative, and 0, 1, and 2 are raw data, first derivative and second derivative, respectively. The best method is bolded, and the worst is italicized.

Classifier	D0	D1	D2	D01	D02	D12	D012
kNN	0.60	0.53	*0.50*	0.61	0.60	0.51	**0.62**
NB	0.54	*0.46*	0.50	0.55	0.54	0.49	**0.57**
DT	0.48	0.47	*0.40*	0.49	0.48	0.45	**0.50**
SVM	0.50	*0.44*	0.46	0.50	0.51	0.48	**0.53**
RF	0.56	*0.45*	0.48	0.57	0.55	0.53	**0.59**
XGBoost	0.62	0.54	*0.49*	0.64	0.62	0.55	**0.67**

Future

1. Testing the proposed techniques using other bases (e.g., the Fourier base) would be appropriate.
2. Moreover, testing the methodology on a larger amount of data and larger data sets is reasonable.
3. Additionally, a dimension reduction method different from the one proposed, for example, PCA, is worth investigating.

Bibliography

1. A. Bagnall et al. (2018). The UEA multivariate time series classification archive. arXiv:1811.00075.
2. T. Górecki, M. Krzyśko, Ł. Waszak, W. Wołyński (2018). Selected statistical methods of data analysis for multivariate functional data. *Statistical Papers* 59:153–182.
3. T. Górecki, M. Łuczak (2014). First and Second Derivatives in Time Series Classification Using DTW. *Communications in Statistics - Simulation and Computation*, 43(9):2081–2092.
4. L. Horvath, P. Kokoszka (2012). Inference for Functional Data with Applications, Springer, New York.
5. J.O. Ramsay & B.W. Silverman (2005). *Functional data analysis*. Springer, New York.

Except Sir,
Of course!