New proposal of robust classifier for functional data

Daniel Kosiorowski[1], Dominik Mielczarek[2], Jerzy P. Rydlewski[3]

Abstract

A variety of economic research hypotheses may be translated into language of statistical discrimination analysis. The company's ability to adapt to changing environmental circumstances may be expressed in terms of a quality of a classifier used in a decision process by the company's management. The specific classifier is developed basing on company's experience expressed in the, so called, training sample. In practice, however, training samples contain outliers of various kinds, which influence the classifier quality. For this reason, robust classifiers, which are able to cope with various data imperfections, are especially desired. This paper focuses on robust classification issues for functional data. We present the state of art and indicate its consequences for the robust economic analysis. We propose an original classification rule appealing to the support vector machines methodology. We show its selected properties and apply it to an empirical issue related to monitoring of electricity market in Denmark.

Keywords: *classifier for functional objects, electricity market, robustness, support vector machines classifier*

JEL Classification: C14, C38, C44

DOI: 10.14659/SEMF.2018.01.20

1 Introduction

A variety of economic research hypotheses may be translated into a language of statistical discrimination analysis. Consider, for example, the company's ability to adapt to changing environmental circumstances, which may be expressed in terms of a quality of a classifier used in a decision process by the company's management. The specific classifier is developed basing on company's experience expressed in the, so called, training sample. In practice, however, training samples contain outliers of various kinds, which may adverse influence on the classifier quality. This fact motivates our studies. The paper focuses on robust classification issues for functional data related to the robust economic analysis of electricity market in Denmark in 2016. The recently developed statistical methodology named functional data analysis (FDA) enables for functional generalizations of well-known one and multivariate statistical techniques like analysis of variance, kernel regression or classification

[1] Corresponding author: Cracow University of Economics, Department of Statistics, ul. Rakowicka 27, 31-510 Kraków, e-mail: daniel.kosiorowski@uek.krakow.pl.
[2] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: dmielcza@wms.mat.agh.edu.pl.
[3] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: ry@agh.edu.pl.

techniques (see Horvath and Kokoszka, 2012; Górecki et al., 2014). The FDA enables us for effective analysis of economic data streams i.e., e.g., analysis of non-equally spaced observed time series, prediction of a whole future trajectory rather than single observations (see Kosiorowski, 2016). In the recent years several interesting for economic application procedures for functional data have been proposed (see Horvath and Kokoszka, 2012). The proposed techniques are not robust however.

In a context of functional data classification analysis, several issues are still unsolved. It should be stressed that a commonly acceptable definition of robustness for a classification procedure even in a multivariate case does not exist up to now (see Hubert et al., 2016). The robustness concept in this case should take into account a local nature of classification procedure (maybe robustness should be defined with respect to specified class rather than regarding the whole data set). One can propose a useful variant of qualitative robustness however: small changes of input should lead to small changes of output or a measure of quality of output. Therefore, we can adapt qualitative definition of robustness. It is possible to adapt Hampel's influence function as well. More recent and formal approaches may be found in Christmann et al. (2013). The main aim of the paper is to propose a new nonparametric statistical methodology, appealing to the Support Vector Machines method (SVM) (Schoelkopf and Smola, 2002), for classifying functional objects. We show its selected properties and apply our classifier to an empirical issue related to monitoring of electricity market in Denmark. Certain aspects of electricity prices modeling and forecasting have been described by Weron in his review paper (2014).

The rest of the paper is organized as follows. Section 2 sketches the basic concepts of classification in a functional setup. Section 3 introduces our procedure for classifying a functional data, and discusses the properties of the procedure through numerical simulations and tests the applicability of the proposed methodology on empirical examples. Section 4 describes the analysed electricity dataset. Section 5 conducts a short robustness analysis and a conclusion is provided.

2 A classifier for functional data

We are given a training sample, that is, n observations: U_1, U_2,…U_n and each observation can be described as U_i=(X_i, Y_i), where Y_i=-1 or Y_i=1. The X_i are called patterns, cases, inputs or instances, and Y_i are called labels, outputs or targets. We would like to classify a new object X into one of the two labelled groups, basing on knowledge included in the training sample. A classification rule defines a certain partition of the feature space into nonempty disjoint subsets, which are summed up to a whole feature space.

Classification methods for functional data include k-nearest neighbors (kNN) methods, reproducing kernel of a Hilbert space (RKHS) methods (see Schoelkopf and Smola, 2002), methods based on depth measures (see Cuevas and Fraiman, 2009; Kosiorowski et al., 2018), and neural networks methods. Functional outliers detection on the example of air quality monitoring has been recently described by Kosiorowski et al. (2018).

3 Our proposals

Let $X_1, X_{2,\dots}, X_m$ be any functional data from Hilbert space $L^2(\Omega)$ with the usual inner product defined by $< f, g >= \int_\Omega f(w)g(w)dw$, where Ω is a bounded subset of Euclidean space, and let numbers $Y_1, Y_{2,\dots}, Y_m$ be labels, i.e., $Y_i=1$ or $Y_i=-1$ for i=1,2,…,m. Patterns X_i are functions mapping set Ω into real numbers. We assume in the whole paper, that the set Ω is bounded, and then the space $L^2(\Omega)$ is separable, what is important in our considerations. Hence, there exists an orthonormal basis $\{Z_1, Z_2, \dots\}$and every function X from the space $L^2(\Omega)$ can be described as the following series $X = \sum_{n=1}^{\infty} \langle X, Z_n \rangle \cdot Z_n$, where the series convergence is a convergence in the sense of norm of the space $L^2(\Omega)$.

In our further considerations we set Ω=[0,T]. Patterns X_i belong to $L^2([0,T])$ space with the usual inner product defined by $< f, g >= \int_0^T f(t)g(t)dt$, and the orthonormal basis is either a standard Fourier basis or a spline basis.

In practice, we fix a natural number K and we determine a vector $c = (c_1, c_2, ..., c_K)$ such that $\hat{X} = \sum_{n=1}^{K} c_n \cdot Z_n$, so that they minimize a real function ϕ given by the following formula $\phi(c) = \left(X - Zc \right)^T \cdot \left(X - Zc \right)$, where $X^T = (X(t_1),\dots,X(t_M))$ and Z is a matrix of the form $\left[Z_j(t_i) \right]_{i=1,\dots,M}^{j=1,\dots,K}$ and $t_i \in [0,T]$ are knots. We propose a classifier for functional data of the form

$$f(X) = \int_\Omega X(\omega) W(\omega) d\omega + b$$

where *b* is any real number and weight function W is essentially bounded, i.e. $W \in L^2(\Omega)$ and chosen so that affine functional *f* be data-consistent i.e. $Y_i f(X_i) = 1$, for any i=1,2,…,m.

In other words, we are given empirical data (X_1, Y_1), (X_2, Y_2),…,(X_m, Y_m). Basing on the data, we classify a new functional observation X into one of the groups looking only on *sgn* (f(X)). The classifier doesn't work, if f(X) = 0 (for practical purposes one may assume zero probability for such an event). Existence of the weight function W, as can be shown, is guaranteed with linear independence of the random functions $X_1, X_2, \ldots, X_m$. Using functional analysis apparatus, i.a. Hahn-Banach Theorem, we have proved the following theorem.

Theorem. For any real number *b* there exists a function $W \in L^2(\Omega)$ such that

$$Y_i\left(\int_{\Omega} X_i(\omega)W(\omega)d\omega + b\right) = 1, \text{ for } i = 1, 2, \ldots, m$$

has a solution.

Note that, the weight function *W* satisfies

$$\int_{\Omega} X_i(\omega)W(\omega)d\omega + b = Y_i, \text{ for } i = 1, 2, \ldots, m.$$

It is now obvious, that in order to solve the classification problem it suffices to determine a weight function *W*, or equivalently to find a functional g such, that

$$g(X_i) = \int_{\Omega} X_i(\omega)W(\omega)d\omega = Y_i(1 - Y_i b) \text{ for } i = 1, 2, \ldots, m.$$

We show now, how to determine the hyperplane separating for functional data. Let fix an index i=1,2,…,m. We construct a bounded linear functional $g_i : L^2(\Omega) \to R$ such that $g_i(X_j) = \delta_{ij}$, where δ_{ij} is a Kronecker delta. Then, a functional *g* given by the formula $g = \sum_{j=1}^{m} Y_j(1 - Y_j b)g_j$ satisfies the equations (1).

In order to solve the set of equations (1) it suffices to indicate functionals g_i. Let us denote for any set of vectors $W_1, W_2, \ldots, W_m$ from Hilbert space with inner product $< \cdot, \cdot >$

$$M(W_1, \ldots, W_m) = \det\left[< W_i, W_j >\right]_{i=1,\ldots,m}^{j=1,\ldots,m}$$

The number $M(W_1, \ldots, W_m)$ is a Gram matrix determinant for a set of vectors $W_1, W_2, \ldots, W_m$. Recall, that a set of vectors is linearly independent if and only if $M(W_1, \ldots, W_m) > 0$. The functionals g_i we are looking, for any $Y \in L^2(\Omega)$ are given by the formula:

$$g_i(Y) = \frac{M(X_1, \ldots, X_{i-1}, Y, X_{i+1}, \ldots, X_m)}{M(X_1, \ldots, X_m)}$$

We now give a formula for a separating hyperplane. After conducting some simple computations, the weight function for the functional g is given by the following formula:

$$W = \sum_{i=1}^{m} \frac{X_i - P_i(X_i)}{\|X_i - P_i(X_i)\|^2}$$

where $P_i : L^2(\Omega) \to V$ is an orthogonal projection on $V = span\{X_1, ..., X_{i-1}, X_{i+1}, ..., X_m\}$.

It is the most right place here to remark, that as $L^2(\Omega)$ space is not a RKHS (Reproducing Kernel Hilbert Space), then the proposed classifier is not of that kind as well. Moreover, it can be easily seen that the classifier is affine invariant, i.e. invariant with respect to the mapping $A: X \to L(X) + s$, where L is a linear mapping and s is a translation.

4 Robustness of a classification rule for functional data

The robustness of the classifying rule toward outliers depends on the functional outliers type. It should be different for the functional shape outliers, functional amplitude outliers and for functional outliers with respect to (w.r.t.) the covariance structure. That's why it is not easy to approximate breakdown point or influence function. It should be stressed, that there is no agreement as to the breakdown point or influence function concepts in the functional classification case (see Hubert et al., 2015 and 2016, and references therein). Many of the classical robust classification methods assumes multivariate normality or elliptical symmetry, which is a simplification of the more complex problem. Hubert et al. (2016) discusses some robust approaches that can deal with functional data. They make use of the concept of depth and present in the article a new technique - classification in distance space. They carry out a distance transformation and use a bagdistance to obtain a robust classification rule.

Cuevas and Romo (1993) studied qualitative robustness of bootstrap approximations when the estimators are generated by a statistical functional *T*. They showed that the uniform continuity of statistical functional *T* is a sufficient condition for a qualitative robustness of the bootstrap estimator. Denote now $L_n(F)=L_n(T;F)$ is the sampling distribution of the statistic $T_n(X_1,\dots,X_n)$ where the sample comes from *F*, and $L[L_n(F_n)]$ is the sampling distribution of the generalized statistic $L_n(F_n)$ in the space of relevant probability measures. Cuevas and Romo (1993) definition states that „given a sequence $\{T_n\}$ of statistics generated by a statistical functional *T*, the sequence of bootstrap approximations $\{L_n(F_n)\}$ is said to be qualitatively robust at *F* when the sequence of transformations $\{G \to L[L_n(G_n)]\}$ is asymptotically equicontinuous at *F*.” Their concept is followed by i.e. Christmann et al. (2013) and by us. *Qualitative robustness* is thus defined as equicontinuity of the distribution of the considered statistic as the sample size is growing, and hence the concept of qualitative robustness is related to continuity of the statistic in the relevant space, which is now considered as a function in the weak* topology (see Rudin, 1991).

Remark. It can be proved, following lines of Christmann et al. (2013) proof, that our classifier is qualitatively robust as well. In the proof we exploit the fact, that Gram matrix used in classifier construction, as a matrix, is a continuous mapping. Notice, however, that our classification rule is not a kernel classification rule, so their proof cannot be directly applied.

Tarabelloni (2017) defined Max-Swap Algorithm, that can be used to obtain more robust classification rule. He suggests to compute the covariances for two groups obtained by including the unit either in the first or second group. The distances between the two groups are computed then. The new observation is attributed to the group for which the considered distance *is* the least. Finally, he suggests to choose the swapping units in such a way that the distance between the estimated covariance operators at the next step is higher than the distance between the estimated covariance operators at the preceding step. Controlling numbers of units in each groups affects the robustness of his algorithm (for details see Tarabelloni, 2017).

5 Empirical analysis of the electricity dataset

We consider data from an electricity market in Denmark, where each day is represented as a function. We use an electricity consumption data retrieved from www.nordpoolgroup.com. Using classifying methods, we would like to classify a new functional object into one of the considered groups: working day or weekend. Fig. 1 presents functional observations of Danish electricity consumption in working days in 2016, while Fig. 2 presents functional observations of Danish electricity consumption in weekend in 2016.

Training sample consists of the functional data of electricity consumption representing first 20 days of each month in 2015. Subsequently, separating hyperplanes have been computed for each month and a classification has been conducted for functional observations of electricity consumption representing each day of 2016.

Let now g denote our classification rule. The distribution of (X, Y) is unknown, so we estimate the empirical risk of misclassification for our rule

$$\hat{L}(g) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{g(X_i) \neq Y_i\}},$$

where 1_S denotes the indicator function of the set S. We have compared our classification rule for the analyzed empirical data with: RKHS kernel methods with Gaussian kernel and polynomial kernel (see Febrero-Bande and de la Fuente (2012) R package fda.usc). Empirical risk for our rule g is 26%, while empirical risk for Gaussian kernel RKHS method is 66%, for polynomial kernel RKHS method is 58%. Our method turned out to be computationally as intensive as the compared methods.

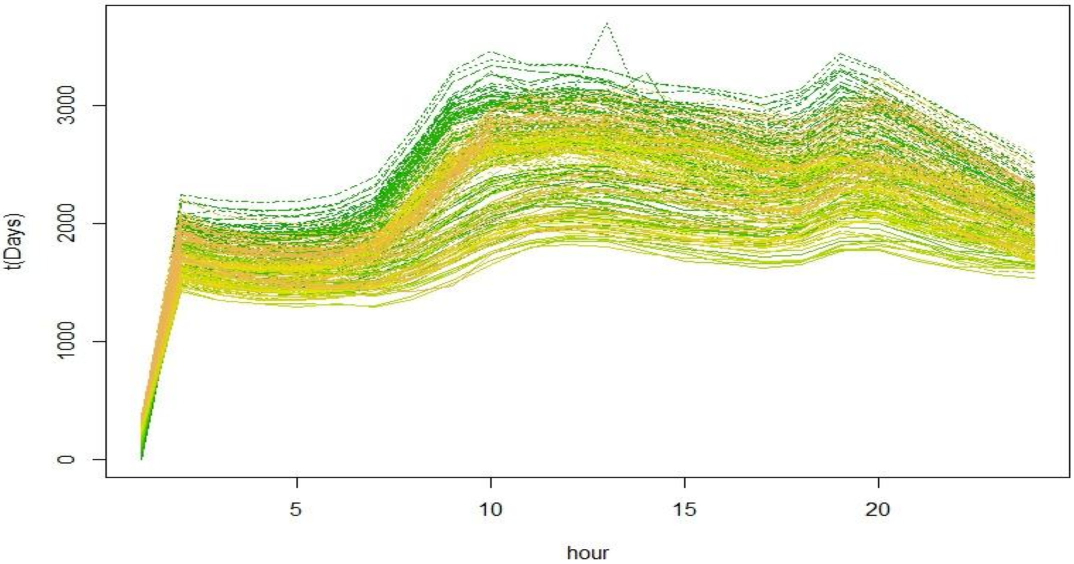

Fig.1. Functional observations electricity consumption in working days in 2016 for Denmark.

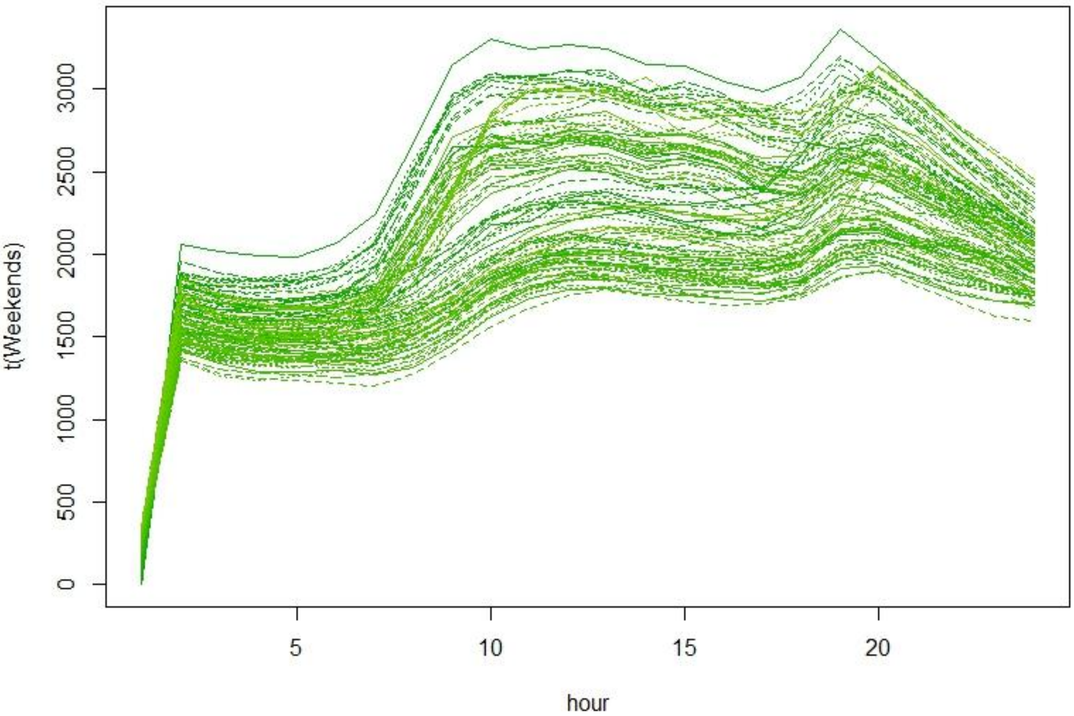

Fig.2. Functional observations electricity consumption in weekends in 2016 for Denmark.

Conclusions

This paper proposes new classification method for functional data. The presented method allows for monitoring phenomena appearing within the new economy, which are described by means of functions of a certain continuum. We show on a real data set, related to electricity consumption, some advantages and disadvantages of our proposal. In the future, we plan some further studies of the proposal involving gamma-regression or beta-regression.

Our method is significantly better w.r.t. empirical risk than the considered kernel RKHS methods. Nonetheless, note that our method is computationally as intensive as other methods.

Our proposals can be applied to different fields of an e-economy, i.e., (Web site management, protection of computer systems against hacking, spam filtering, etc. - as an e-economy provides a great deal of functional data) or to optimization of electricity production.

Acknowledgements

JPR and DM work was partially supported by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Ministry of Science and Higher Education, grant no. 11.11.420.004 and DK's research by the grant awarded to the Faculty of Management of CUE for preserving scientific resources for 2017 and 2018.

References

Christmann, A., Salibián-Barrera, M. & Van Aelst, S. (2013). Qualitative Robustness of Bootstrap Approximations for Kernel Based Methods. In: Becker, C., et al. (eds.), *Robustness and Complex Data Structures*, Springer, Berlin Heidelberg, 263-278.

Cuevas, A. & Romo, R. (1993). On robustness properties of bootstrap approximations. *Journal of Statistical Planning and Inference, 37*(2), 181–191.

Cuevas, A. & Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *Journal of Multivariate Analysis*, *100*(4), 753-766.

Febrero-Bande, M. & de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, *51*(4), 1-28.

Górecki, T., Krzyśko, M., Waszak, Ł. & Wołyński W. (2014). Methods of Reducing Dimension for Functional Data. *Statistics in Transition*, *15*(2), 231-242.

Horvath, L. & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Hubert, M., Rousseeuw, P. & Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications*, *24*(2), 177–202.

Hubert, M., Rousseeuw, P. & Segaert, P. (2016). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification*. Doi: 10.1007/s11634-016-0269-3

Kosiorowski, D. & Zawadzki, Z. (2017). DepthProc: An R Package for Robust Exploration of Multidimensional Economic Phenomena. Retrieved from https://arxiv.org/abs/1408.4542v9

Kosiorowski, D. (2016). Dilemmas of robust analysis of economic data streams. *Journal of Mathematical Sciences, 1*(2), 59-72.

Kosiorowski, D., Rydlewski, J. P. & Zawadzki, Z. (2018). Functional outliers detection by the example of air quality monitoring. *Statistical Review* (in Polish, to appear).

Rudin, W., (1991). *Functional Analysis*. McGraw-Hill.

Schoelkopf, B. & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press.

Tarabelloni, N., (2017), *Robust Statistical Methods in Functional Data Analysis,* PhD thesis and roahd R package, Politecnico di Milano.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting, 30*, 1030–1081.