Alternative Representations of Spectral Data for Classification

Diana Porro[1], Robert P. W. Duin[2], Isneri Talavera[1], Noslen Hernández[1],

[1]Advanced Technologies Application Centre, Cuba.
[2]Pattern Recognition Group, TU Delft, The Netherlands.
[1]{dporro,italavera,nhernandez}@cenatav.co.cu,[2] r.duin@ieee.org

Keywords: Dissimilarity Representation, Functional Data Analysis, Pattern Recognition, Chemometrics.

Abstract. Classification problems can be found in any research area, and one of its most essential facts is trying to have a representation of the data where can be resumed as much useful information as possible. In the specific case of chemical spectral data, although they are typically plotted as functions of wavelengths, product concentration, etc. they traditional representation is through a set of different features, ignoring important aspects as their functional nature i.e. connectivity between the measured points, shape, etc. Two approaches have been proposed to represent the spectral data: Functional Data Analysis and Dissimilarity Representation. These take into account the functional information and intrinsically constitute a reduction of the high-dimensionality of this data. In this paper, in order to show how the accuracy of the classifiers may be better when using these two approaches, we made a comparison of the realization of four classifiers on the two different representations and the traditional feature representation of four spectral chemical data.

1. Introduction

With the development of instrumental analysis equipments in the last years, the chemometricians have been provided of a lot of data. These have been very helpful to solve specially classification problems that before were very difficult to confront, e.g. recognition of persons by their DNA, types of drugs, types of soils and others. For the purpose of classifying these data many algorithms have been proposed: k-Nearest Neighbor (k-NN) [8], Normal density based linear/quadratic classifiers (NLC/NQC) [9], [24] and Support Vector Machine (SVM) [7], etc. But most of them are traditionally applied to a feature representation of the spectra, by which they are expressed as a set of different features (variables), disrespecting their functional nature, the continuity that exists between the values in the real curves.

With the aim of taking into account this important information about the spectral data, and therefore improving the results of the analysis made on it, the Functional Data Analysis (FDA) [23] has been proposed. FDA is an extension of the traditional multivariate analysis for data with a functional nature (functional data (FD)), and is based on the idea of considering the observed spectra as a continuous real-valued function instead of an array of individual observations. Among the goals of FDA mentioned in reference [23] are: to represent and transform the data in ways that aid further analysis, to display the data so as to highlight various characteristics and to study important sources of patterns and variations among the data. This is why several classical multivariate statistical methods have been extended to FD.

Recently, it has been introduced by Pekalska et al. [19], what is known as Dissimilarity Representation (DR). Although it is based in a different philosophy to FDA's, it also incorporates important information for the description of the data as connectivity, shape of the spectra, etc. This representation was mainly thought for classification; it is based on the idea that as classes are conformed by a set of similar objects, the proximity between them plays an important role in this type of problems. The DR consists basically in the representation of the objects by its dissimilarities with respect to the other objects, and the classifiers may be built on the dissimilarity space generated by a representation set, or in a feature space where the dissimilarity data is isometrically embedded. In this paper we will work with the first technique mentioned. It is worth to tell that any traditional classifier that works on feature spaces can be also used on the dissimilarity space.

As there is not a general dissimilarity measure for all types of data, the dissimilarity measure selected allows emphasizing the information that would be very useful to classify a particular type of spectra (in this case), e.g. connectivity, peak positions, shape, etc., which is one of the principal advantages of the DR. The use of DR is especially advantageous with: 1) spectral data, 2) when vector representations of objects live in a high dimensional space and 3) when the number of examples is very small [19], and these last two elements turn out to be the major problems when working with spectroscopic chemical data.

The goal of this paper is to show some of the advantages of the FD and DR representation for spectral data and how they may improve (sometimes substantially) the classification results obtained with the traditional feature representation. With this purpose, we will make a comparison of the performance of 1-NN, RNLC, RNQC (regularized versions of NLC and NQC) and SVM classifiers on the feature, FD and DR representations of four chemical spectral datasets.

The paper is organized in the following way. In Chapter 2, it has been made sort of a state of the art of what has been done on these approaches. Chapter 3 is dedicated to the essential things of the theory of Functional Data Analysis and Dissimilarity Representations. In Chapter 4, the description of the analyzed datasets can be found, as well as a relation of the methods and tools used. The results of the comparison between the different approaches and the discussion made around it are presented in Chapter 5. Finally, in Chapter 6, there are the conclusions to which we have arrived in the confection of the article.

2. Related work

Although DR and FDA are young techniques, they are having a very good acceptance in Pattern Recognition and Chemometrics applications respectively. FDA started to deploy with Ramsay and Dalzell [22] and coined by Ramsay and Silverman [23] where they expose many examples and methods for this analysis. Most of them are extensions of classical multivariate methods as for example: Principal Component Analysis (FPCA) [26], linear discriminant analysis [4], [10] and canonical correlation analysis [12]. The Functional PCA was the first method extended to FDA and it demonstrates the way in which a set of functional data varies from its mean, and, in terms of these modes of variability, quantifies the discrepancy from the mean of each individual functional datum. In the case of linear modeling there has been a large application in regression, generalizing to FD some dimensionality reduction approaches for linear regression e.g. Principal Component Regression (PCR) and Partial Least Squares (PLS) [2],[20], and also for classification as Linear Discriminant Analysis (LDA) [4], etc. More recently, a number of estimation methods for functional nonparametric classification and regression models have been also introduced. Namely, estimators based on functional data adaptations of classical neural networks [25], K-Nearest Neighbor classifier (k-NN) [5], kernel classifiers [1], [3], [8], and in Reproducing Kernel Hilbert Spaces (RKHS) [21] - the latter including Radial Basis Function (RBF) methods. Based on the RKHS have been proposed the extension of SV methods for classification [29] and regression [11] in functional spaces.

DR was introduced by Pekalska et al. [19] as a new approach that links the structural and statistical approaches. It is also inspired in the concept of kernels, as a way of handling the problem with indefinite kernels that can appear when working with some dissimilarity measures [16]. This type of representation can have several applications but the main researches have been done on classification [15], [17]. As was said before, the main advantages of this approach have been seen on sensory data, such as spectra and this has also been argued in [13], [14], where some dissimilarity measures for this specific type of data are also proposed.

3. Theory

3.1 Functional Data Analysis

In spectral chemical data as Near-Infrared, Ultra-Violet, etc., each spectrum is a function that maps the wavelengths of the illuminating light to the corresponding absorbencies of the studied samples; however, these spectra are usually observed and recorded discretely and so analyzed with multivariate data analysis techniques which consider the spectrum as high-dimensional vectors of different but high-correlated variables. So, when working with this type of representation many practical problems can be encountered as the characteristics of the real nature of the data are not taken into account.

On the other hand, if the algorithms work on the functional spaces, it can also lead to theoretical and practical difficulties as these have infinite dimensions. The idea of FDA is to retrieve the intrinsic characteristics of the underlying function from the discrete functional data, so the observations

can be seen as continuous single entities, instead of sets of different variables.

To deal with the infinite dimensional problem, most of FDA methods have been constructed on two general principles: filtering and regularization. In the filtering approach, the idea is to use representation methods that allow working in finite dimension. This is the way of approximation that is used here.

The first thing to do is to choose a proper family of basis functions matching best the underlying function (s) to be estimated. Of an existing variety of bases: Fourier series, polynomial, wavelet and splines; the basis of B-splines [30] seems to be most appropriate for spectrometric data, as they accurately represent one of the main characteristics of this type of data, smoothness. They are also very easy to compute. To make this basis of B-splines $\{\phi_k\}_{k=1}^{K}$ with K the number of basis functions, a number of knots between the start and end wavelengths are defined, and a B-spline is run from one knot to another; the different splines overlap.

So, the spectral function $x_i = x_i(\lambda)$ for sample i and wavelengths λ can be described by the linear combination of the basis functions:

$$x_i = \sum_{k=1}^{K} c_{ik}\phi_k \tag{1}$$

where c_{jk} , the B-spline weights (coefficients) are computed by the following expression:

$$\min \sum_{j=1}^{m} (x_{ij} - \sum_{k=1}^{K} c_{ik}\phi_k(\lambda_j))^2 \tag{2}$$

x_{ij} is an element of the matrix conformed by a set of spectra relating the *i*th sample with the *j*th wavelength. The expression in Eq. 2 corresponds then to minimizing the vertical distance between the observed spectral information and the fitted curve.

The function will be explained then by the coefficients and the methods will take these as the new representation of the data instead of the original data points.

3.2 Dissimilarity Representation

The Dissimilarity Representation (DR), based on the fact that the classes are conformed by objects that have similar characteristics, proposes to work on the space of the proximities between those objects, instead of the space defined by their characteristics (features), as it is usually done.

The first thing here is to select a suitable dissimilarity measure for the problem at hand, a measure that is appropriate for comparing objects given the known data characteristics. Next, a representation set $R(p_1, p_2, ..., p_n)$ has to be selected, which is a set of representative objects from the classes, called prototypes.

The DR for a set of objects X is the matrix $D(X,R)$ formed by the dissimilarities between each object $x \in X$ and the objects of R. Each element of the matrix would be then a dissimilarity value between two objects, $d(x_m, p_n)$. R can be a subset of X, $(R \subseteq X)$ or X itself, being then $D(X,X)$ a square dissimilarity matrix, or X and R can be completely different sets. There are many approaches to select the prototypes of the representation set [18].

In DR there are three main approaches. The first one addresses the given dissimilarities directly. The second is based on an approximate embedding of the dissimilarities into a (pseudo)Euclidean space and at last, the so-called dissimilarity space approach which is the one proposed to be used here. The dissimilarity space $\mathfrak{D} \subseteq \mathbb{R}^n$ is generated by the column vectors of the dissimilarity matrix, where each dimension corresponds to the dissimilarity between the objects and a prototype $d(\cdot, p_n)$. As the dissimilarities are computed to the representation set, even when this set is conformed by all the objects, it constitutes already a dimensionality reduction especially in spectrometric data, as one of the main problems of this type of data is the large number of variables and small number of samples (objects). Therefore it can be less computationally expensive. Besides, any traditional classifier that works on feature spaces can be also used on the dissimilarity space.

4 Material and methods

In order to evaluate the performance of different classifiers using the three different representations of the data, a comparative study with four of them: 1-NN, RNLC, RNQC and SVM will be made for 3 chemical spectral datasets.

The experiments were all performed in Matlab. For the case of FDA was used the FDAFuns toolbox. For the DR and classification of the data was used the PRTools toolbox.

For FDA representation each spectra was represented by an *n*th order B-spline approximation with *p* basis functions. The optimal values for the number of B-spline coefficients and the degree of the spline was chosen using leave-one-out cross validation on the training set.

For the classification of the data we used four classifiers: 1-NN, RNLC, RNQC and SVM. For the four datasets it was ran a 10 times k-fold cross validation using the dissimilarity matrix of the training set as input data. For this cross validation procedure were used 4-folds, 6-folds and 10-folds for Tecator, Wine, Oil and GadoiFish data sets respectively. The dissimilarity measures used are City Block and Euclidean distances. For the SVM classifier was used the linear kernel. To find the regularization parameters of RLNC and RQNC an automatic regularization process was done.

In order to obtain the final prediction error, an independent test set was used. The comparison of the accuracy among the models obtained for the FDA, DR (two dissimilarity measures) and feature representation with the four classifiers was done by using the classification error.

4.1 DataSets

We used three different datasets. The first one, named Tecator, comes from the food industry [28] (shown in Fig. 1). It consists of 215 near infrared absorbance spectra of meat samples (objects), recorded on a Tecator Infratec Food and Feed Analyzer. Each observation consists of a 100 channel absorbance spectrum (variables) in the 850-1050 nm wavelength range, and is associated to a content description of meat sample, obtained by analytic chemistry. The classification problem consists of separating meat samples into two classes. To the first class belong the samples with a high fat content (more than 20%), and the samples with a low fat content (less than 20%) conform the second class. From the 215 spectra, 43 are kept aside as testing set and the 172 remaining samples are used for training. Original spectra are preprocessed; each spectrum is reduced to zero mean and unit variance on the set of spectra.

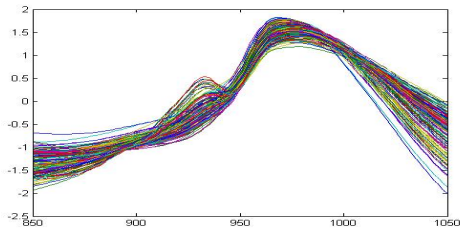

Fig. 1. Near-Infrared absorbance of meat samples. Tecator dataset.

The second dataset shown in Fig. 2, is named Wines [31] and is a set of 44 Red wines (objects), produced from the same grape (100% Cabernet Sauvignon), harvested in 4 geographical areas (Argentina, Chile, Australia, South Africa) which would be the classes conformed in this problem. The wine samples have been analyzed using a FT-IR commercial WineScan instrument provided by FOSS Analytical A/S (932-4999 nm), for a total of 842 (variables) wavelengths analyzed. The classification problem consists in defining to which of the four geographical areas each sample belongs. For the training set will be used 30 samples and 14 for testing.

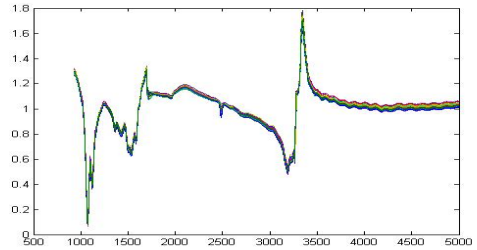

Fig. 2. FT-IR spectra of red wine samples. Wines dataset.

The third data consists of oil samples that were analyzed using three analytical techniques: Gas chromatography (GC), Mid-Infrared (MIR) and Near Infrared (NIR) [27]. We will only use here the MIR one shown in Fig. 3. The samples have three different origins (classes): A, B and D. There are 44 samples available for training and 27 for testing, with 571 variables measured.

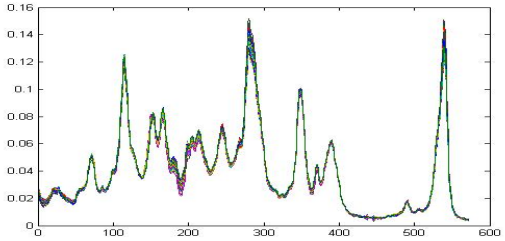

Fig. 3. Mid-Infrared spectra of oil samples. Oil dataset.

The fourth dataset shown in Fig. 4 is named GadoiFish [31]. It consists of 694 near magnetic resonance (NMR) spectra of fish samples (objects) of four species: whiting, haddock, saithe and cod (classes). The NMR relaxations were measured at 18 points spread throughout the whole fish with a Maran Benchtop Pulsed NMR analyzer operating at 23.2 MHz and equipped with an 18 mm variable temperature probe head. The receiver delay was set to 6 s. Transverse relaxations were measured using the CPMG sequence. For each measurement eight scans were performed with 1024 echoes and tau at 500 ms. Only even echoes were recorded of which only every second were used. This gives 256 echoes (variables) for each sample. All measurements were performed at 4°C. The fish were introduced into the NMR probe by placing samples of 2-4 g into glass tubes that matched the inner diameter of the 18 mm NMR sample tubes. The classification problem consists in defining to which of the four species (classes) each sample belongs. For the training set will be used 400 samples and 294 for testing. Original spectra are preprocessed; each spectrum is reduced to zero mean and unit variance on the set of spectra, as is suggested in literature for NMR spectra.

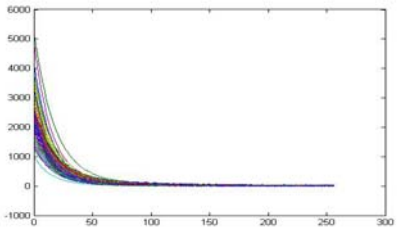

Fig. 4. Near Magnetic Resonance of fish samples. GadoiFish dataset.

5 Results and discussion

In this section, the classification results for the experiments on all datasets by using the three representations, for the specified classifiers: 1-NN, RNLC, RQDC and SVM will be shown.

For the functional approach, the leave-one-out error calculation leads to the selection of an optimal basis of 48-splines of order 4 for Tecator. The cross-validation error for the rest of the datasets also leaded to the selection of 100-splines of order 4 for Wines, 100-splines of order 6 for Oil, and 16-splines of order 6 for GadoiFish dataset. Since for Tecator dataset the shape of the spectrum appears to be so relevant, all the calculations for the FD were made on its second derivative. For the other three datasets, the calculations were made on the raw data and also on its second derivative. For the DR, the representation set $R(p_1, p_2,..., p_n)$ will be equal to the training set.

Table 1. Classification error of the test set of Tecator dataset for the three representations of the data with the different classifiers.

Representations	1-NN	RNLC	RNQC	SVM
Feature	0.023	0.023	0.070	0.069
FD	0.01	0.023	0.047	0.023
DR(CB)	0.046	0	0	0
DR(E)	0.023	0	0	0

Table 2. Classification error of the test set of Wines dataset for the three representations of the data with the different classifiers.

Representations	1-NN	RNLC	RNQC	SVM
Feature	0.214	0.543	0.440	0.5
FD	0.143	0.265	0.296	0.5
FD(2^{nd} D)	0.357	0.467	0.353	0.486
DR(CB)	0.357	0.167	0.195	0.429
DR(E)	0.429	0.115	0.186	0.357

Table 3. Classification error of the test set of Oil dataset for the three representations of the data with the different classifiers.

Representations	1-NN	RNLC	RNQC	SVM
Feature	0.074	0.77	0.037	0
FD	0.074	0	0	0
FD(2^{nd} D)	0.33	0.074	0.148	0.037
DR(CB)	0.148	0	0.037	0
DR(E)	0.22	0.074	0.11	0

Table 4. Classification error of the test set of GadoiFish dataset for the three representations of the data with the different classifiers.

Representations	1-NN	RNLC	RNQC	SVM
Feature	0.548	0.534	0.710	0.575
FD	0.487	0.422	0.435	0.418
FD(2^{nd} D)	0.550	0.512	0.621	0.479
DR(CB)	0.527	0.418	0.599	0.415
DR(E)	0.548	0.558	0.537	0.450

As can be seen in the tables above, most of the time, for most of the classifiers, the classification results improve by using the DR and FD representation of the data. Besides, both of them intrinsically make a dimensionality reduction. This is an important issue for spectral data, as they usually present the problem of ill-conditioned matrix, too much variables and a few samples. Nevertheless, for the fourth dataset there is no dimensionality reduction in the DR, as we are using the whole set of samples as representation set and this data matrix is unlike the others because it has more objects than variables.

It also shows that, the advantages of utilizing these representations depend on how the characteristics that contribute more to the classification of the spectra are measured. For example, in FDA, the use of the second derivative is suggested when the shape of the spectra is what matters, such is the case of Tecator (Table 1). But in spectra where the changes are mainly in the peak´s intensity, and there is merely a variation in the shape, as is the case of Wines, Oil and GadoiFish data sets (see Fig. 2, Fig.3 and Fig.4), the effect of the derivative application can be the opposite, the classification results can be worst due to a loss of information, or in the best case there might be no change (see Table 2 and Table 3).

In the case of DR, the results also depend on whether a suitable dissimilarity measure is used for the specific problem. In this paper we used two of the most common dissimilarity measures and also proposed for spectral data in [13], [14]. The results with the DR have outperformed the obtained with the feature representation in most cases, although the used measures might not be the most proper for these datasets; further researches must be done to select them. In general, for the four datasets very good results are obtained with the RNLC classifier on the DR. It can be explained by the fact that this classifier is optimal for data normally distributed and, according to the Central Limit Theorem applied to dissimilarities, when the dissimilarity measures are based on sum of differences (as is the case of Euclidean and City Block); the data tends to be approximately normal distributed. Nevertheless, the performance of all RLNC classifier in the feature representation of Oil dataset is not good, it can be due to that classes are not linearly separable; there is an overlapping between the samples of class A and B. The RNQC's performance is also good, and should be even better than the RNLC's, but when the representation set is the same training set, its accuracy may be worst. The exposed before about the distribution of the dissimilarity data for the used measures, might be the same reason of why sometimes the classifiers accuracy based on the DR, outperform the results obtained on the FD representation.

The results of 1-NN classifier on the DR are not as expected in most cases, but it can be due to the use of a bad dissimilarity measure. However, as usually happens, the RLNC/RQNC and SVC constructed on the DR outperform the k-NN based on the same representation set. Nevertheless, in GadoiFish dataset the behavior of all classifiers is almost the same and the error rates are high in general, which gives the idea that there could be some problem with the data. All of these results point out the importance of choosing the more suited dissimilarity measure for the problem at hand. It is of a high significance, how well i get to explain the influential characteristics of the curve by the measure, in order to obtain a better and more reliable classification of the data. For the spectral data, the characteristics can depend on the instrumental source, etc., a deeper investigation most be done on this issue. The SVM classifier also performs very well for all representations, but slightly more accurate for the DR than for the other two.

6 Conclusions

We presented two alternative representations for spectral chemical datasets, in order to improve the classification results of this type of data. The functional representation takes into account the functional nature of the data, as it considers the observed spectra as continuous real-valued functions by approximating it with splines. The DR makes use of the physical knowledge of the spectral background of the data by modeling their relations in a dissimilarity representation. Comparisons were made by classifying four chemical spectral datasets, expressed by their feature representation and the two alternatives mentioned. It was shown by the experiments that, with the studied representations the classifiers accuracy can be improved, sometimes substantially in dependence of how well the spectrum gets to be described by them. But it also shows that the use of either one or the other, as well as the classification results, depends on the spectra´s characteristics and the information that influences more the discrimination between the classes. For the case of functional representation, it is necessary to be careful with the use of derivatives; in some cases it can cause the loss of information. These two alternative approaches might be further improved by studying the modeling of the spectra and their dissimilarities.

References

[1]. Abraham, C., Biau, G., Cadre, B.: On the kernel rule for function classification. Annals of the Institute of Statistical Mathematics 58, 619--633 (2006).

[2]. Aguilera, A. M., Ocana, F., Valduama, M. J.: An approximated principal component prediction model for continuous-time stochastic processes. Appl. Stochastic Models Data Anal. 13, 61--72 (1999).

[3]. Biau, G., Bunea, F., Wegkamp, M.: Functional classification in Hilbert spaces.IEEE Transactions on Information Theory 51, 2163--2172 (2005).

[4]. Cardot, H., Ferraty, F., Sarda, P.: Functional linear model Statist. Probab. Lett. 45, 11--22 (1999).

[5]. Cérou, F., Guyader, A.: Nearest neighbor classification in infinite dimension. ESAIM: Probability and Statistics 10, 340--35 (2006).

[6]. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, UK (2000).

[7]. Duda, R., Hart, P., and Stork, D.: Pattern Classification. John Wiley & Sons, Inc., 2nd edition (2001).

[8]. Ferraty, F. and Vieu, P.: Nonparametric Functional Data Analysis. Springer Series in Statistics (2006).

[9]. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press. (1990).

[10]. Hastie, T., Buja, A., Tibshirani, R.: Penalized Discriminant Analysis. The Annals of Statistics 23, 73--102 (1995).

[11]. Hernández N., Biscay R. J., Talavera I.: Support Vector Regression Methods for Functional Data, LNCS, vol. 4756, pp. 564--573 (2007).

[12]. Leurgans, S. E., Moyeed, R. A., and Silverman, B.W.: Canonical Corrlation Analysis When the Data Are Curves. Journal of the Royal Statistical Society, Ser. B 55, 725--740 (1993).

[13]. Paclik, P.,Duin, R.P.W.: Classifying spectral data using relational representation. In: Spectral Imaging Workshop, Graz, Austria (2003).

[14]. Paclik, P., Duin, R.P.W.: Dissimilarity-based classification of spectra: computational issues. Real Time Imaging 9(4), 237--244 (2003).

[15]. Pekalska, E., Duin, R.P.W.: Classifiers for dissimilarity-based pattern recognition. In: International Conference on Pattern Recognition, vol. 2, pp. 12--16, Barcelona, Spain (2000).

[16]. Pekalska, E., Paclik, P., Duin, R.P.W.: A Generalized Kernel Approach to Dissimilarity Based Classification. Journal of Machine Learning Research 2 (2), 175--211 (2002).

[17]. Pekalska, E., Duin, R.P.W.: Dissimilarity representations allow for building good classifiers. Pattern Recognition Letters 23(8), 943--956 (2002).

[18]. Pekalska, E., Duin, R.P.W.: Prototype selection for finding efficient representations of dissimilarity data. In: R. Kasturi, D. Laurendeau, and C. Suen (eds). International Conference on Pattern Recognition, vol. 3, pp. 37--40, Quebec City, Canada (2002).

[19]. Pekalska E., Duin, R.P.W.: The Dissimilarity Representation For Pattern Recognition. Foundations and Applications. Machine Perception Artificial Intelligence, vol. 64 (2005).

[20]. Preda, C., Saporta, G.: PLS regression on stochastic processes. Comput. Statist. Data Anal. 48, 149--158 (2005).

[21]. Preda, C.: Regression models for functional data by reproducing kernel Hilbert space methods. J. Statistical Planning and Inference 137, 829--840 (2007).

[22]. Ramsay, J. O., Dalzell, C. J.: Some Tools for Functional Data Analysis (with Discussion). Journal of the Royal Statistical Society, Ser. B 53, 539--572 (1991).

[23]. Ramsay, J. O., Silverman, B. W.: Functional Data Analysis, New York. Springer (1997).

[24]. Ripley, B.: Pattern Recognition and Neural Networks. Cambridge University Press. Cambridge (1996).

[25]. Rossi, F., Conan-Gues, B.: Functional Multilayer perceptron: a nonlinear tool for functional data analysis. Neural Networks 18, 45--60 (2005).

[26]. Silverman, B.W.: Smoothed Functional Principal Components Analysis by Choice of Norm. The Annals of Statistics 24, 1--24 (1996).

[27]. ShootOut Oil dataset, http://cac2008.teledetection.fr/shootout.htm.

[28]. Tecator dataset, http:lib.stat.cmu.edu/datasets/tecator.

[29]. Villa, N., Rossi, F.: Support Vector Machine For Functional Data Classification. In: ESANN'05 Proceedings, http://apiacoa.org/publications/2005/svm-sann05.pdf

[30]. Wahba, G.: Spline Models for Observational Data. In: SIAM, Philadelphia (1990).

[31]. Wine and GadoiFish datasets, http://www.models.kvl.dk/research/data.