



ELSEVIER

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

The DTW-based representation space for seismic pattern classification

Mauricio Orozco-Alzate^{a,*}, Paola Alexandra Castro-Cabrera^a, Manuele Bicego^b,
John Makario Londoño-Bonilla^c^a Universidad Nacional de Colombia, Sede Manizales, Departamento de Informática y Computación, km 7 vía al Magdalena, Manizales 170003, Colombia^b Università degli Studi di Verona, Dipartimento di Informatica, Ca' Vignal 2, Strada le Grazie 15, Verona 37134, Italy^c Servicio Geológico Colombiano, Observatorio Vulcanológico y Sismológico de Manizales, Av. 12 de Octubre No. 15 – 47, Manizales 170001, Colombia

ARTICLE INFO

Article history:

Received 9 September 2014

Received in revised form

14 April 2015

Accepted 14 June 2015

Available online 17 June 2015

Keywords:

Classification

Dissimilarity space

Dynamic time warping

Seismic patterns

Volcano monitoring

ABSTRACT

Distinguishing among the different seismic volcanic patterns is still one of the most important and labor-intensive tasks for volcano monitoring. This task could be lightened and made free from subjective bias by using automatic classification techniques. In this context, a core but often overlooked issue is the choice of an appropriate representation of the data to be classified. Recently, it has been suggested that using a relative representation (i.e. proximities, namely dissimilarities on pairs of objects) instead of an absolute one (i.e. features, namely measurements on single objects) is advantageous to exploit the relational information contained in the dissimilarities to derive highly discriminant vector spaces, where any classifier can be used. According to that motivation, this paper investigates the suitability of a dynamic time warping (DTW) dissimilarity-based vector representation for the classification of seismic patterns. Results show the usefulness of such a representation in the seismic pattern classification scenario, including analyses of potential benefits from recent advances in the dissimilarity-based paradigm such as the proper selection of representation sets and the combination of different dissimilarity representations that might be available for the same data.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Analyzing the type and intensity of seismic activity is one of the most important tasks in volcano monitoring. Even though sophisticated monitoring techniques – mainly based on remote sensing equipment such as satellites and radars – are currently widely used, seismicity is still considered as the key *vital sign* that reveals changes in geophysical processes beneath the volcanic edifice. The analysis of such seismicity requires the exhausting and repetitive classification of seismic patterns into several pre-defined classes: a task that can be dramatically improved by applying novel statistical learning approaches for automated pattern classification. Personnel relieved of that duty may concentrate on interpretation, characterization and understanding of the volcanic phenomena; moreover, the classification task itself turns faster and, potentially, more accurate without the subjective judgments of experts or, in many cases, of temporary staff members.

Automatic pattern recognition systems are composed by two main parts: *representation* and *classification*. The first part is aimed

at characterizing the objects (also known as *patterns*) with a collection of descriptors, possibly resulting in a so-called feature space where classification can be easily and accurately carried out; the second one consists in learning, typically from a labelled training set, a classifier which is able to assign to new unlabeled patterns the corresponding class labels.

In the past, the statistical/machine learning community has been particularly focused on the design and application of classifiers, typically paying less attention to the step of representation – reasonably assuming that descriptors are already provided by some experts, being application or domain dependent (Duin et al., 2007). However, studying the representation issue is of paramount importance, mainly because any deficiencies in this step cannot be recovered later in the process of learning. Two types of representations are basically distinguished in the literature (Pekalska and Duin, 2005): *absolute representations* and *relative representations*. In the former case, which represents the classical option, patterns are represented as points in a feature vector space; in the latter case, only a set of (dis)similarity values derived from pairwise comparisons between objects is available – being also known as *(dis)similarity representations*. This second representation naturally arises in many different applications, particularly in those fields which involve human judgments – i.e. situations where it is easier to derive a relational measure rather

* Corresponding author.

E-mail addresses: morozcoa@unal.edu.co (M. Orozco-Alzate), pacastroc@unal.edu.co (P.A. Castro-Cabrera), manuele.bicego@univr.it (M. Bicego), jmakario@sgc.gov.co (J.M. Londoño-Bonilla).

than extracting descriptors, like when comparing medical images. More specifically, within the EU-FP7 SIMBAD¹ project, a number of advantages of the dissimilarity representation as alternative to the classical feature-based approach have been shown; results are reported in the recent book by Pelillo (2013).

Given the dissimilarity representations, the most natural choice for the classification step is to resort to the nearest neighbor rule: a simple yet effective classifier that labels an unseen pattern with the class label of its most similar (closest) pattern within a collection of representative examples; see (Theodoridis and Koutroubas, 2009, Section 2.6) for further details on this classical rule. Actually, it is only in recent years that dissimilarity-based classification techniques which go beyond the nearest neighbor rule have been investigated; among them, for instance, the so-called *dissimilarity space* (Duin and Pekalska, 2012), which is straightforward in its formulation and of high interest for practical applications. Such approach proposes to exploit the dissimilarity representations to derive a novel vector space, where dimensions are associated to a set of representative patterns. The coordinate value in a given dimension, for points in that space, is given by the dissimilarity between the corresponding representative patterns and individual patterns from another set. Given this novel vector space, built directly from the dissimilarity representations, any advanced classifier can be used, thus going well beyond the simple nearest neighbor rule. This paradigm, which has shown to be very effective in many different practical applications (Duin and Pekalska, 2012), has been hardly applied in the seismic scenario, with an early example appeared in Orozco-Alzate et al. (2006). However, such preliminary study was not conclusive enough for different reasons. First, only two metrics – the Euclidean distance and the L_1 distance – were applied for comparing pairs of equal-length one-dimensional power spectral densities of seismic signals. Therefore, with that choice, one of the most attractive, recently studied and promising advantages of the dissimilarity representation was not exploited: the potential informativeness of non-Euclidean (Duin et al., 2013) or, even, non-metric dissimilarity measures (Pekalska et al., 2006b; Plasencia-Calaña et al., 2013). Second, other recent developments were not considered in such study, such as those derived from the combination of different sets of differently measured dissimilarities (Ibba et al., 2010; Porro-Muñoz et al., 2011), or from the proper selection of representative objects employed to build the dissimilarity space (Pekalska et al., 2006a). From the seismic point of view, in addition, more advanced signal characterizations can be used, which are richer than one-dimensional spectra. One example is given by spectrograms, typically using the short-time Fourier transform, which map waveforms to the time–frequency domain and serve to analyze statistically nonstationary signals (Theodoridis and Koutroubas, 2009, Section 7.5.1) such the seismic ones; another example corresponds to a two-dimensional characterization using the so-called Fisher–Shannon method: an information-theory based feature representation that has been successfully applied by Telesca et al. (2013) for discriminating between tsunamigenic and non-tsunamigenic earthquakes as well as – among other analysis-oriented applications – to distinguish quarry blasts from earthquakes (Telesca et al., 2011). Finally, in Orozco-Alzate et al. (2006) only two not very challenging classification problems were tried: a two-class one and a three-class one. A much more realistic and challenging case study must include more class labels that are routinely assigned in volcano monitoring.

This paper makes one step further along this direction, exploiting recent advances of the dissimilarity-based representation

paradigm to derive a seismic signal classification procedure, which is tested on a challenging multi-class classification problem exhibiting realistic conditions. The proposed approach stems from the dynamic time warping (DTW) distance (Lemire, 2009; Lin et al., 2012), a widely used non-metric (dis)similarity measure able to take into account the intrinsic temporal and sequential nature of the seismic signals – thus being more accurate than lock-step measures. Moreover, the DTW measure does not require that both input signals are of the same length and is able to deal with distortions, e.g. shifting along the time axis (Lin et al., 2012) which, in practice, makes it robust to inaccuracies in the segmentation of individual isolated signals. It is important to note also that, despite the fact that using the DTW measure for classifying signals is a well-known and established practice, this distance has been almost always applied in a *template matching* framework, i.e. by applying the nearest neighbor rule for classification. The applicability of this measure in order to build dissimilarity spaces has hardly been investigated, for instance in a case of sign language recognition (Duin et al., 2013). In the proposed approach, the DTW dissimilarities have been computed starting both from signals, also known as waveforms, and spectrograms derived from the seismic events (see Fig. 1). Different classifiers have been tested in the resulting vector space, and compared with the traditional nearest neighbor rule (i.e. template matching). Moreover, some recent advanced versions of the dissimilarity based representation paradigm have also been tested, in particular the ones derived by combining dissimilarities or by properly selecting the representative sets. Experimental results, including learning curves, confirm that this approach represents a valid alternative to classic seismic classification techniques.

The remaining part of the paper is organized as follows. Basic concepts behind the dissimilarity representations and the dissimilarity space are briefly summarized in Section 2, together with some recent advances that we exploit in our proposed approach, which is described in detail in Section 3. Experimental results and their corresponding discussions are presented in Section 4. Finally, a number of concluding remarks are given in Section 5.

2. Background and framework

2.1. Dissimilarity representations and the dissimilarity space

Let \mathcal{T} be a set of labeled training patterns $\mathcal{T} = \{x_i^{(\mathcal{T})}; i = 1, \dots, |\mathcal{T}|\}$, where $|\mathcal{T}|$ is the cardinality of \mathcal{T} , and let x be an arbitrary pattern inside or outside \mathcal{T} . A *dissimilarity representation* (Pekalska and Duin, 2005) of x consists in a row vector $\mathbf{d}(x) = [d(x, x_i^{(\mathcal{T})}); i = 1, \dots, |\mathcal{T}|]$ where d is an appropriate dissimilarity measure for the nature of x , for instance the DTW dissimilarity measure in the case of our proposed approach. The DTW dissimilarity measure does not have a definition in the form of a mathematical formula but is computed via an algorithm with two nested loops (see Algorithm 1); a more detailed description of it is provided in Section 3.2.

In order to reduce the dimensionality of $\mathbf{d}(x)$ and, thereby, the amount of dissimilarity computations, a subset of \mathcal{T} is usually used as the so-called *representation set* \mathcal{R} : a collection of representative patterns also known as *prototypes*; in such a way, the dissimilarity representation for a single pattern x corresponds to $\mathbf{d}(x) = [d(x, x_i^{(\mathcal{R})}); i = 1, \dots, |\mathcal{R}|]$. New incoming patterns to be classified compose the so-called test set \mathcal{S} .

According to the above-indicated notation, training and test sets in the dissimilarity representation approach are contained in the following two matrices of pairwise dissimilarities:

¹ Similarity-based Pattern Analysis and Recognition (SIMBAD) project: <http://simbad-fp7.eu>.

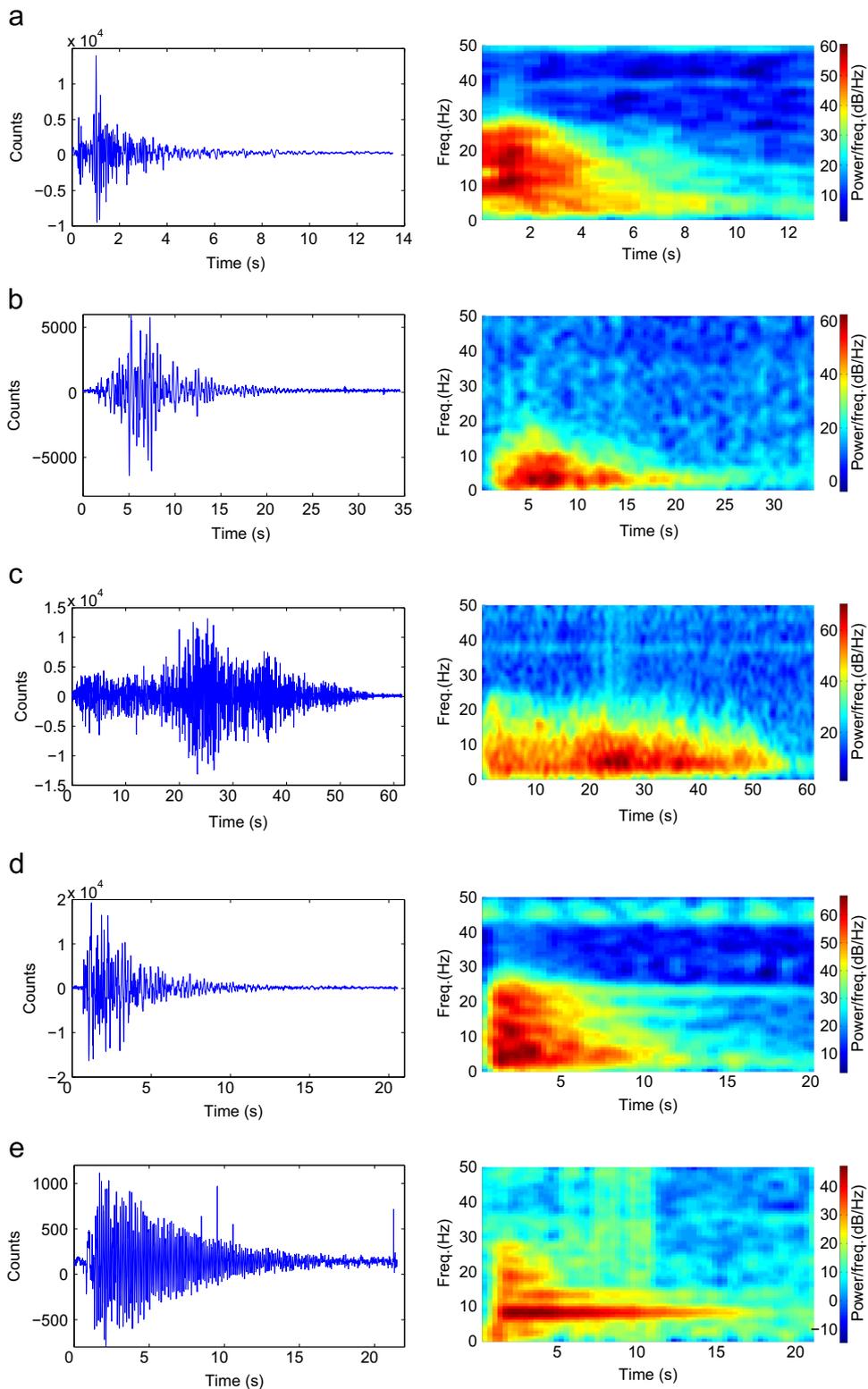


Fig. 1. An example of different classes of seismic-volcanic patterns (left) together with their corresponding spectrograms (right). (a) Volcano-tectonic (VT) event. (b) Long period (LP) event. (c) Tremor (TR). (d) Hybrid (HB) event. (e) screw-like (TO) event.

$$\mathbf{D}_{\mathcal{T}} = \mathbf{D}(\mathcal{T}, \mathcal{R}) = \left[\mathbf{d}(x_1^{(\mathcal{T})}); \mathbf{d}(x_2^{(\mathcal{T})}); \dots; \mathbf{d}(x_{|\mathcal{T}|}^{(\mathcal{T})}) \right]$$

$$\mathbf{D}_{\mathcal{S}} = \mathbf{D}(\mathcal{S}, \mathcal{R}) = \left[\mathbf{d}(x_1^{(\mathcal{S})}); \mathbf{d}(x_2^{(\mathcal{S})}); \dots; \mathbf{d}(x_{|\mathcal{S}|}^{(\mathcal{S})}) \right]$$

where semicolon denotes vertical concatenation. Notice that, consequently, sizes of matrices $\mathbf{D}_{\mathcal{T}}$ and $\mathbf{D}_{\mathcal{S}}$ are $|\mathcal{T}| \times |\mathcal{R}|$ and $|\mathcal{S}| \times |\mathcal{R}|$, respectively.

A straightforward way to classify objects from \mathcal{S} is to use the template matching approach (denoted hereafter as 1-NND), that is, directly applying the nearest neighbor rule to the given dissimilarities which, in practice, consists in finding the minima per row in $\mathbf{D}_{\mathcal{S}}$ and, afterwards, assigning to each test object the class label associated to the column (representation object) where the minimum dissimilarity was found. Clearly, this classification

strategy does not use all the information contained in \mathbf{D}_S , and completely disregards the information contained in \mathbf{D}_T . The alternative, proposed by Pekalska and Duin (2005), is to consider that entries of $\mathbf{d}(x)$ span a vector space equipped with the conventional inner product and Euclidean metric; such a space is called the *dissimilarity space* (Duin and Pekalska, 2012), has $|\mathcal{R}|$ dimensions and is populated with row vectors from \mathbf{D}_T and \mathbf{D}_S . In this way, all the information contained in the dissimilarity representations are used and maintained, but a vector space is now available: any traditional classifier can be built in the dissimilarity space, thus going far beyond the classical nearest neighbor approach. Please note that the dissimilarity space is now a proper vector space, equipped with the Euclidean distance, which can be computed between pairs of vectors. In this sense we will employ in our experiments also the standard nearest neighbor rule (denoted hereafter as 1-NN), which will permit to show the improvement in the discriminative capabilities gained by exploiting the dissimilarity space approach.

2.2. Advanced issues

The dissimilarity space has been largely investigated in recent years, from both a theoretical and a practical perspective, leading to many interesting variants and declinations (Duin and Pekalska, 2012). Among others, here we investigated two issues; the first one – known as *prototype selection* – involves techniques usable to prune the representation set \mathcal{R} (Pekalska et al., 2006a); the second one – known as *combining dissimilarity representations* – exploits the possibility of combining sets of differently measured dissimilarities (Ibba et al., 2010; Porro-Muñoz et al., 2011).

The first issue starts from the following observation: dissimilarity representations based on a large training set can be computationally demanding (Duin and Pekalska, 2012). Therefore, it is important to prune the representation set to a size significantly smaller than that of the training set. More formally, the prototype selection issue can be formulated as the task of finding a reduced representation set $\mathcal{R}^* \subset \mathcal{R}$ such that the classification performance does not deteriorate or, even, improves. The cardinality reduction of \mathcal{R} is achieved by applying prototype selection procedures that are analogous to the well-known ones that are used for feature selection in the traditional pattern recognition framework (Theodoridis and Koutroumbas, 2009, Chapter 5). An exhaustive study presented by Pekalska et al. (2006a) showed that prototype selection is a crucial aspect in dissimilarity-based classification. In our study we examine three different selection procedures: random selection (as a baseline approach), forward selection using the leave-one-out 1NN error estimation as criterion (a traditional feature selection procedure) and k -centres (an strategy from cluster analysis). Pekalska et al. (2006a) showed that the latter, in general, works well.

The second issue exploits the intuition that κ multiple dissimilarity matrices $\mathbf{D}_T^{(1)} \dots \mathbf{D}_T^{(\kappa)}$, often available for the same data, may be combined in such a way that the resulting performance is hopefully increased. Several options can be applied to perform the combination such as taking the entry-wise average, product, minimum and maximum operations (Pekalska and Duin, 2005, p. 457) as well as computing an optimized weighted average (Ibba et al., 2010) or, even, concatenating (extending) the available different representations (Plasencia-Calaña et al., 2013). In spite of the diversity of options, averaging the κ available matrices seems to be the simplest and cost-effective scheme, as shown in Duin and Pekalska (2012) and Ibba et al. (2010). Since different dissimilarity matrices may differ in their scales, it is important to normalize them according to a given criterion in order to prevent that one of them dominates in the combination. A global rescaling is the typical option for normalization by, for example, setting

either the maximum or the average dissimilarity to 1 and scaling all the entries of the matrix accordingly.

3. The proposed classification system

In this section the proposed system for classification of seismic events is described: the representation of the signals and the essential definitions and concepts of the DTW dissimilarity measure are presented, followed by the postulation of the DTW-based representation space for seismic pattern classification.

3.1. Representation

The input signals correspond to raw sequential data as provided by the acquisition system. They are typically *digital*; that is, both discrete in time at a given sampling rate (f_s) and quantized at a finite precision. Here we assume, as in many other studies, that automatic segmentation of isolated seismic events has been carried out, for example using the so-called STA/LTA algorithm or, otherwise, by manually stamping the start of the primary wave and the end of the so-called coda. Since seismic events naturally vary in duration, signals are sequences of different lengths. We call this basic representation raw time signals (i.e. waveforms).

Another widely used representation is based on the frequency content of the waveforms. The easiest way is to consider the *spectrum* of the signal: since this method was used in Orozco-Alzate et al. (2006), let us briefly summarize it here – it will also be considered in the experimental evaluation as a baseline comparison; see Section 4.5. The spectrum of a signal consists in the average power of it across the interval of frequencies $[-f_s/2, f_s/2]$ (Schilling and Harris, 2012). The spectral information of the first half of the spectrum is a mirrored version of that contained in the second half; in consequence, only the latter, corresponding to the interval $[0, f_s/2]$, is needed. The spectrum is typically computed by using the well-known Fast Fourier Transform (FFT), which is computationally efficient and able to produce outputs of a fixed length for inputs having arbitrary durations.

A richer representation is obtained by building a spectrogram (Fig. 1), which consists in computing the FFT for successive short overlapping frames. Such a computation produces, thereby, a sequence of spectra that reveals how the spectral content of the signal changes with the time. By fixing the length of the FFT output, spectrograms for signals with arbitrary different lengths coincide at least in one dimension: the axis corresponding to the frequency interval. Since the FFT gives complex numbers, only the magnitude (in dB) of the spectrograms was taken into account.

3.2. The dynamic time warping (DTW) dissimilarity measure

Dissimilarity measures are more general than metrics, in the sense that they are just required to obey conditions of being positive ($d(x_i, x_j) \geq 0$), reflexive ($d(x_i, x_i) = 0$) and definite or constant ($d(x_i, x_j) = 0$ if and only if x_i and x_j are identical). In contrast, metrics must also be symmetric ($d(x_i, x_j) = d(x_j, x_i)$) and satisfy the triangle inequality ($d(x_i, x_j) \leq d(x_i, x_k) + d(x_j, x_k)$). Advantages of the DTW dissimilarity measure over lock-step measures – e.g. the Euclidean distance and other classical metrics – were already mentioned in Section 1. Algorithm 1 shows the procedure to compute the DTW dissimilarity measure between two arbitrary sequences x and y having lengths m and n , respectively. The algorithm, $DTW_w(x, y)$, is based on the code by Wang (2013) and corresponds to an implementation using dynamic programming with a warping window of size w . The purpose of the warping window is to fix a global constraint in order to narrow the search

space of the best alignment between the two sequences and, therefore, speed up the computation (Lin et al., 2012).

Algorithm 1. DTW algorithm. Brackets denote positions of the arrays where sequences are stored.

Require: x, y : the two signals, of length m and n , respectively; w : the warping window size; $d(x[i], y[j])$: a distance between single elements of the signals
Ensure: $DTW_w(x, y)$: the DTW distance between x and y , given the warping window w .

```

// Adapt w
1:  $w = \max(w, \lfloor m - n \rfloor)$ 
// Initialize the array  $D$ , of size  $(m + 1) \times (n + 1)$ 
2: for  $i=1$  to  $n+1$  do
3:   for  $j=1$  to  $m+1$  do
4:      $D[i, j]=+\infty$ 
5:   end for
6: end for
7:  $D[1, 1] = 0$ 
// compute entries of  $D$  by dynamic programming
8: for  $i=1$  to  $m$  do
9:   for  $j = \max\{i - w, 1\}$  to  $\min\{i + w, n\}$  do
10:     $dist = d(x[i], y[j])$ 
11:     $D[i + 1, j + 1] = dist + \min\{D[i, j + 1], D[i + 1, j], D[i, j]\}$ 
12:   end for
13: end for
14: return  $D[m + 1, n + 1]$ 

```

Notice that there are two nested loops in Algorithm 1; therefore, even with the global constraint w , the DTW is computationally expensive: $O(wm)$ (Lemire, 2009). A popular rule-of-thumb to initialize w , before calling the algorithm, is setting it to the 10% of the longest sequence. When comparing one-dimensional signals, such as the considered input waveforms, the typical choice for the dissimilarity measure between single elements of the signals $d(x[i], y[j])$ is to use the Euclidean distance (see line 10 in Algorithm 1):

$$d(x[i], y[j]) = |x[i] - y[j]|$$

The algorithm can easily be generalized to a multi-dimensional case, provided that sizes of the multi-dimensional sequences x and y coincide in one of the dimensions; that is, if sizes of their associated arrays are $m \times p$ and $n \times p$, respectively. In such a case, $d(x[i], y[j])$ can be any norm between two vectors, i.e.

$$d(x[i], y[j]) = \|x[i, :] - y[j, :]\|$$

where colon denotes all entries in the second dimension of the array, i.e. slices of length p from x and y are compared. This multi-dimensional version of the DTW dissimilarity measure allows us to compare pairs of spectrograms.

3.3. The DTW-based dissimilarity space

Given a training set $\mathcal{T} = \{x_i^{(\mathcal{T})}; i = 1, \dots, |\mathcal{T}|\}$, a representation set $\mathcal{R} = \{x_i^{(\mathcal{R})}; i = 1, \dots, |\mathcal{R}|\}$ (which can be inside or outside \mathcal{T}), and a testing set $\mathcal{S} = \{x_i^{(\mathcal{S})}; i = 1, \dots, |\mathcal{S}|\}$, the problem is mapped to the dissimilarity space via the DTW dissimilarity by following the scheme proposed in Section 2.1. In particular, every seismic event $x \in \mathcal{T}, \mathcal{S}$, in its original time-domain form or considering the corresponding spectrogram, is described with the row vector

$$x \rightarrow [DTW_w(x, x_1^{(\mathcal{R})}), DTW_w(x, x_2^{(\mathcal{R})}), \dots, DTW_w(x, x_{|\mathcal{R}|}^{(\mathcal{R})})]$$

It is important to note that \mathcal{R} can be, in the basic case, the whole training set; better choices can also be performed by using the selection procedures described in Section 2.2.

3.4. Classifiers in the dissimilarity space

A plethora of classification rules can be applied in the dissimilarity space. However, a number of them are commonly preferred to be used; for example, the linear support vector machine (SVM) classifier, because of its ability in dealing with high dimensional spaces without suffering from the curse of the dimensionality problem; more than this, it has been shown that such linear classifier becomes non-linear when applied to the dissimilarity space (see Duin et al., 2013, Section 2.4 for more details). In our approach we considered three classic classifiers: the linear SVM, the 1-NN rule (as described in the previous section) and the Fisher's least square linear discriminant (Fisher).

4. Experimental evaluation

In this section all the details related to the experimental evaluation are given: data set, representations and classification experiments. The latter are discussed as figures and tables are presented.

4.1. Data set and preprocessing

The Volcanological and Seismological Observatory of Manizales (OVSM by its acronym in Spanish) from Servicio Geológico Colombiano (SGC) has deployed a seismic network to monitor volcanoes from the northern volcanic segment of Colombia, particularly the activity from Nevado del Ruiz volcano which is the most active one in that segment and have erupted a number of times during the last three decades. Experiments in this paper are based on seismic data from such volcano, which has been collected between January 2010 and September 2013. Signals, gathered at the BIS station, are relative to five classes, corresponding to the most common volcano-related events: volcano tectonic (VT) events, long period (LP) events, tremors (TR), hybrid (HB) events, and screw-like (TO) earthquakes; see Fig. 1 again. For every class, the following number of events is available: 153 (VT), 333 (LP), 242 (TR), 393 (HB) and 104 (TO). Even though seismic sensors deliver three components, only registers from the vertical one are considered. Signals are sampled at $f_s = 100$ Hz, quantized with a 16-bit analog-to-digital converter and segmented into isolated events.

Concerning the representations, spectrograms were built – as done in Castro-Cabrera et al. (2014) – with a 128-point FFT, a 64-point Hamming window to smooth each frame and an overlap of 50% between frames. As a baseline representation, a one dimensional spectrum was used, computed using the FFT with 512 bands; therefore, due to the mirrored property, the length of the one-dimensional spectra was 257. In the experiments we used the classifiers mentioned in Section 3.4 – 1-NN, Fisher and SVM. When needed, a priori class probabilities are estimated by class frequencies.

We designed three experiments, detailed in the following section, trying to highlight different aspects of the proposed approach. In all cases, averaged classification errors were presented (either 25 or 50 repetitions).

4.2. Experiment 1

In the first experiment we evaluated the behavior of the dissimilarity spaces based on two representations (waveforms and spectrograms) with learning curves for growing training sets,

randomly drawn from the data set; fifty patterns per class were chosen as the maximum size for the training set in order to guarantee that at least half of the patterns from the minority class (TO) remain in that case for testing. Classifiers are subsequently tested on the left-over patterns.

The main goal of this experiment is to illustrate how classifiers in the dissimilarity space are able to profit from the information in \mathcal{T} while keeping the computational cost, in the evaluation stage, fixed. Dissimilarities from training patterns to representation ones must, of course, also be computed; however, these computations are performed off-line; that is, in the training stage. As new incoming and unlabeled patterns arrive to the deployed classification system, only dissimilarities from them to the representation patterns remain to be calculated.

In these experiments, representation sets of various sizes are considered, particularly sets containing 1, 4, 7 and 10 patterns per class (randomly selected); therefore, the dimensions of the corresponding dissimilarity spaces are 5, 20, 35 and 50. Small representation set sizes are of special interest because the number of dissimilarity computations to classify a new incoming pattern, in a deployed automated system, is equal to $|\mathcal{R}|$. However, the extreme case when $\mathcal{R} = \mathcal{T}$ is also examined.

Results are reported in Figs. 2 and 3, for waveforms and spectrograms representations, respectively. 1-NN, Fisher and SVM represent the classifiers in the dissimilarity space built from the DTW (i.e. the proposed approach), while 1-NND is the reference approach (i.e. the Nearest Neighbor rule with the corresponding original dissimilarity matrix).

From a general point of view, these plots clearly show the advantage of learning from a training set. Notice, in particular, that template matching (1-NND) only yields better results when very small training sets are used. Such a behavior is most notorious for waveforms (Fig. 2).

Two general observations from Figs. 2 and 3 can be highlighted: (i) in most cases, Fisher is the best performing classifier; (ii) overall, the best dissimilarity-based representation is the one derived from the spectrograms. The first observation is – at first

glance – surprising, because Fisher is a linear and simple classifier. However, this simplicity and linearity is *apparent*, since it has been already shown by Duin et al. (2010) that linear classifiers in the dissimilarity space correspond to nonlinear ones in the original space. The second observation confirms that the spectrogram representation is richer than the other one; in addition, it is cheaper than the waveforms representation, because the width of the frames in the spectrograms allows to shorten the length of the sequences to be compared and, therefore, lighten the cost associated to DTW computations. Another general aspect to be noticed is that all learning curves are not saturated and, therefore, further improvements might be achieved by enlarging the training sets.

Additional particular observations are:

- Among the three dissimilarity-based classifiers, 1-NN is the worse one. In spite of that, it is also able to outperform 1-NND provided that a large enough training set is given.
- In spite that representation patterns were selected at random, it seems that a representation set with 7 or more patterns per class is enough to reach a good performance for large training sets. This is consistent with the known fact that random selection may work well (Pekalska et al., 2006a), unless a very small number of representation patterns is required. In such a case, systematic selection procedures must be applied.

Regarding the case $\mathcal{R} = \mathcal{T}$, it is interesting that 1-NND is better than the dissimilarity-based classifiers when applied on the waveform representation. In contrast, as shown in Fig. 4b, SVM is the best classifier when using spectrograms. These observations can be rephrased as follows: template matching is the best option to compare signals in the time domain but using the dissimilarity-based information contained in the training set really pays off when using spectrograms. These two claims, of course, are only valid for the case when $\mathcal{R} = \mathcal{T}$. Notice that, for all curves in Fig. 4, the dimensionality of the associated dissimilarity spaces grows from 5 (left part of the curve) to 250 (right part); consequently, dissimilarity spaces on the right part of the curve are very high-

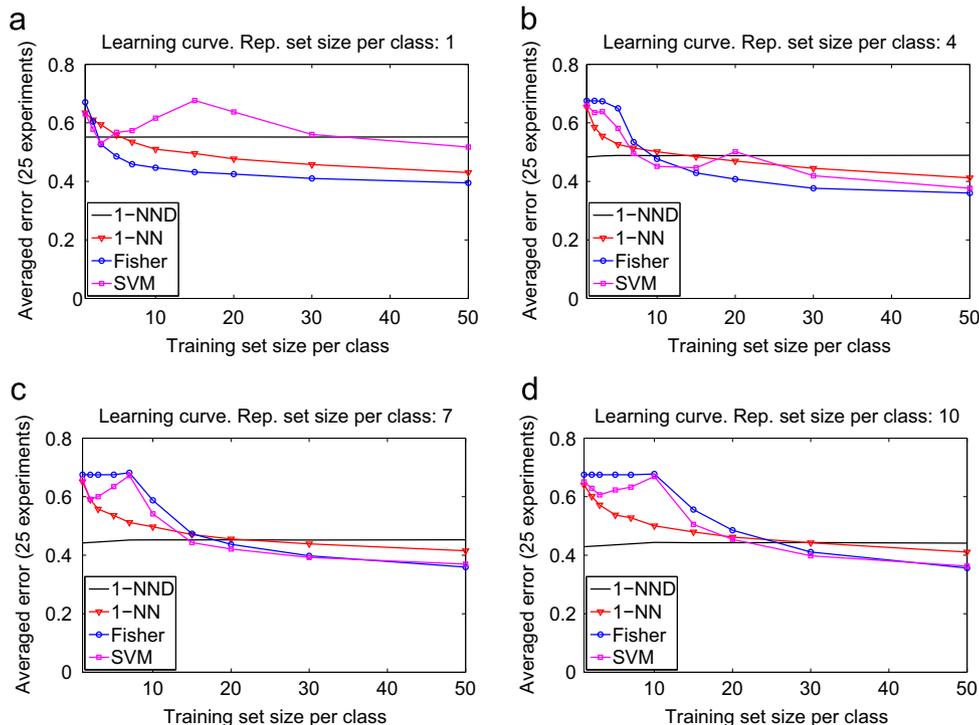


Fig. 2. Learning curves for the DTW-based dissimilarity matrix computed for the signals in the time-domain. (a) $|\mathcal{R}| = 5$. (b) $|\mathcal{R}| = 20$. (c) $|\mathcal{R}| = 35$. (d) $|\mathcal{R}| = 50$.

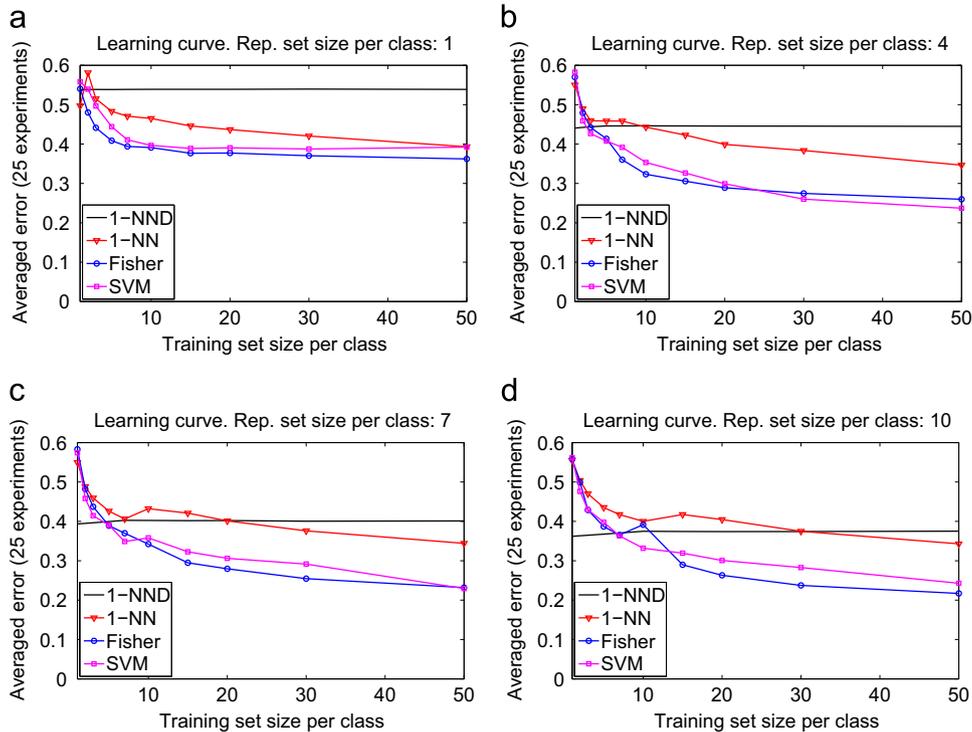


Fig. 3. Learning curves for the DTW-based dissimilarity matrix computed for the signal spectrograms (time/frequency-domain). (a) $R = 5$. (b) $R = 20$. (c) $R = 35$. (d) $R = 50$.

dimensional and demand, to a deployed system, as many dissimilarity computations as dimensions.

4.3. Experiment 2

As explained in the previous sections, dissimilarity representations based on a large training set can be computationally demanding (Duin and Pekalska, 2012). Therefore, it is important to prune the representation set to a size significantly smaller than that of the training set. This second experiment consists in examining curves for growing representation sets drawn from the training set that, in turn, corresponds to half of the data set; in other words, 50% of the data is used for training and the remaining 50% for testing. Two systematic selection procedures – (i) forward selection using the leave-one-out (LOO) 1-NN error as criterion and (ii) k -centres – as well as random selection are used; details of their implementations are not given here but can be found in Pekalska et al. (2006a). Experiments, in this case, were repeated more times (50) in order to better capture the average behavior when two random procedures are involved: partition into training-test sets and the random selection itself. Results are displayed

in Fig. 5, for both the waveform and the spectrogram representations.

The first evident observation is that dissimilarity-based classifiers are much better than template matching when using random selection and k -centres; see Fig. 5a and b and Fig. 5e and f, respectively. Forward selection finds good representation sets for 1-NND that are difficult to be enhanced by providing the additional information contained in the training sets; in fact, the only case – for forward selection – in which a dissimilarity-based classifier is slightly better than 1-NND is the one in Fig. 5d. This confirms that 1-NND is highly dependent on a carefully selected set of prototypes while classifiers in the dissimilarity space are not.

In spite of the superiority of forward selection, one must also consider that its computational cost is larger due to its supervised selection criterion: the minimization of the LOO 1-NN classification error. The other two procedures, in contrast, are unsupervised but still carried out in a class-wise way. The cheapest option is, of course, random selection since just implies the generation of a random permutation of indexes. Notice also that curves in Fig. 5 turn flat after having included a certain number of prototypes in R ; so, the computational demands associated to large representation

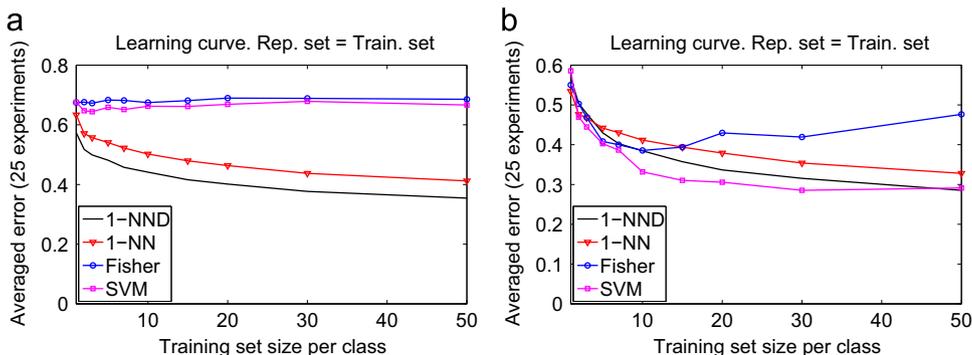


Fig. 4. Learning curves for $R = T$. (a) Waveforms. (b) Spectrograms.

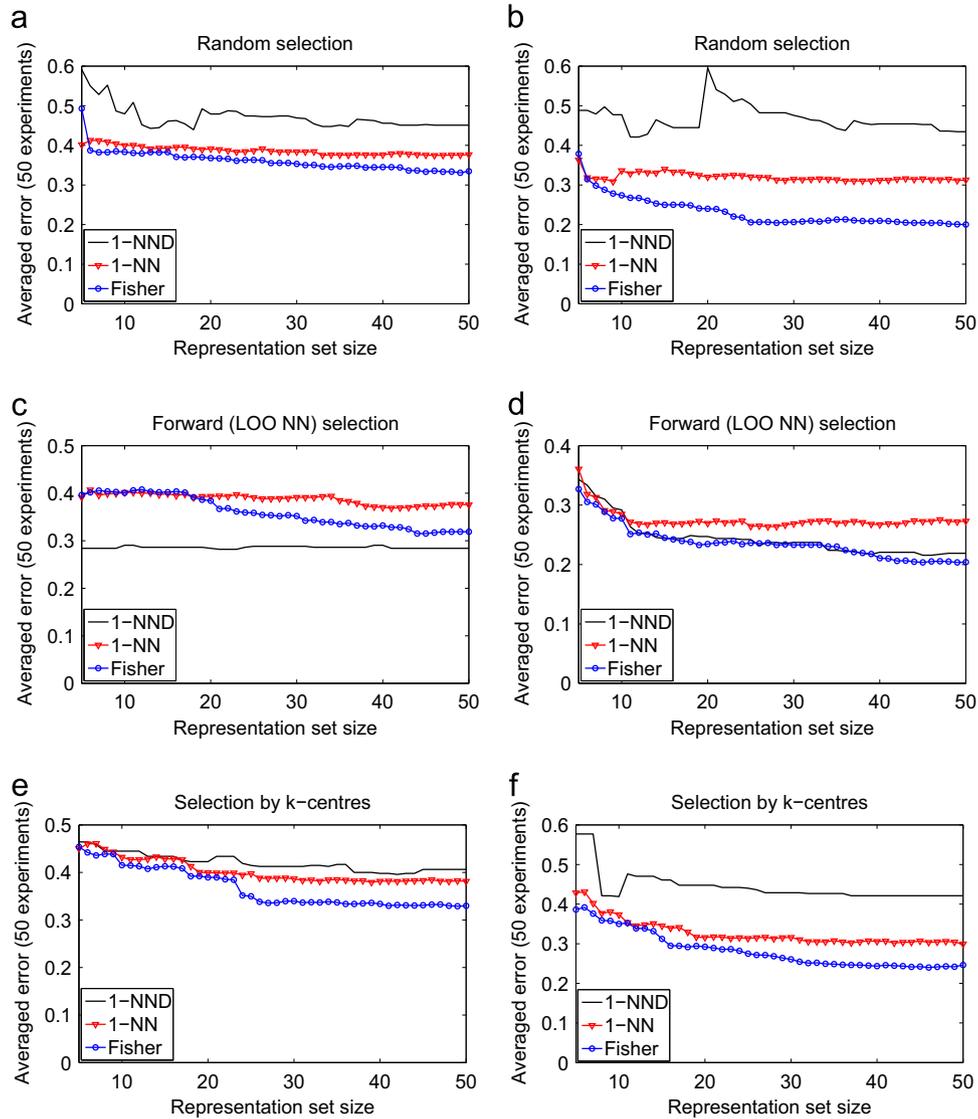


Fig. 5. Curves for growing representation set sizes. The entire training set, corresponding to half of the data set, is used for training the classifiers. (a) Random selection: waveforms. (b) Random selection: spectrograms. (c) Forward selection: waveforms. (d) Forward selection: spectrograms. (e) Selection by *k*-centres: waveforms. (f) Selection by *k*-centres: spectrograms.

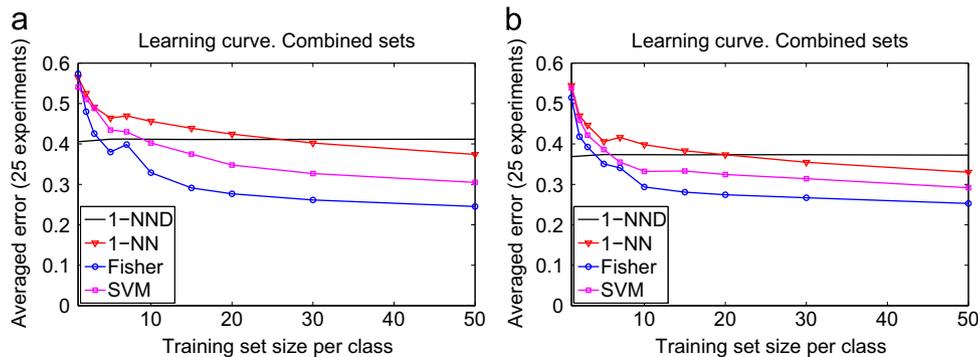


Fig. 6. Learning curves for the combined dissimilarity matrices. (a) Average combination of "waveforms+spectrograms" dissimilarities and (b) "waveforms+spectrograms+spectra" dissimilarities.

set sizes are not compensated by a significant increase in classification performance.

4.4. Experiment 3

This experiment is aimed at showing the effect on the

classification accuracies of the combination of different dissimilarity matrices, see Fig. 6. In particular, we considered the two dissimilarity matrices derived from the two representations (waveforms and spectrograms). Furthermore, we also considered another dissimilarity matrix, employing Euclidean distances between one-dimensional spectra, computed as described above. In both

cases (“waveforms+spectrograms” dissimilarities and “waveforms+spectrograms+spectra” dissimilarities), the combined dissimilarity matrix is obtained by simply averaging them after a normalization step such that the mean dissimilarity is 1.

A significant improvement due to the combination is not observed; in fact, performances are worse than the performance obtained for spectrograms alone (the best representation); cf. Fig. 3d. We are currently investigating this behavior, trying to understand whether this depends on the specific seismic scenario, or on more general factors, like the kind and the number of dissimilarities, the combination scheme, the normalization and so on.

4.5. Discussion

The three experiments in the previous sections were aimed at examining three different aspects of the proposed framework: the impact of the size of \mathcal{T} for different sizes of \mathcal{R} (Experiment 1), the impact of the different prototype selection methods (Experiment 2), and the impact of the dissimilarity combination schemes (Experiment 3): in all cases, learning curves are displayed, in order to understand at a fine-grain level the behavior of the proposed framework. To conclude the experimental evaluation, in this section we provide further discussions starting from some summarized results, which permit to have a better numerical and concise insight into the different aspects. In particular, for the three experiments, we extracted the best possible situation for every configuration: in other words, for a given analyzed configuration, we selected the best result over the different classifiers/training set sizes. Moreover, as previously announced in Section 3.1, results for representations built from pairwise Euclidean comparisons of signal spectra are also included as baseline comparisons.

The results are summarized in Tables 1–3 – since the reported values are classification errors, the lower the better. Different observations can be derived from these summarized results. In general, we can observe that the classification error in the DTW space is almost always better than the nearest neighbor error using the corresponding dissimilarity: only in 4 cases over 26, 1-NND does a better job. Such cases, however, are confined to two precise situations: (i) $\mathcal{R} = \mathcal{T}$ (for the first experiment), and (ii) the use of the forward selection method for selecting the prototypes (for the second experiment): interestingly, these two configurations represent two very computationally demanding versions of the proposed framework, not to be used in practical scenarios. Another general comment can be obtained by looking at the best

Table 1
Best (lowest) classification errors obtained in the first experiment.

Classifier	Representation set sizes				
	$ \mathcal{R} = 5$	$ \mathcal{R} = 20$	$ \mathcal{R} = 35$	$ \mathcal{R} = 50$	$ \mathcal{R} = \mathcal{T} $
	Representation: Waveform + DTW				
Best result for 1-NND	55.18	48.92	45.27	44.11	35.43
Best Diss.-based classifier	39.48 (Fisher)	36.00 (Fisher)	35.93 (Fisher)	35.61 (Fisher)	41.20 (1-NN)
	Representation: FFT + Euclidean				
Best result for 1-NND	55.60	46.73	40.88	40.51	31.33
Best Diss.-based classifier	42.61 (Fisher)	36.26 (Fisher)	33.72 (Fisher)	33.09 (Fisher)	29.58 (SVM)
	Representation: Spectrograms + DTW				
Best result for 1-NND	53.87	44.50	40.11	37.52	28.53
Best Diss.-based classifier	36.22 (Fisher)	23.71 (SVM)	22.96 (SVM)	21.70 (Fisher)	29.13 (SVM)

Table 2
Best (lowest) classification errors obtained in the second experiment.

Classifier	Selection method		
	Random selection	Forward selection	k -centres
	Representation: Waveform + DTW		
Best result for 1-NND	45.08	28.41	40.64
Best Diss.-based classifier	33.46 (Fisher)	31.89 (Fisher)	32.98 (Fisher)
	Representation: FFT + Euclidean		
Best result for 1-NND	41.63	26.36	57.36
Best Diss.-based classifier	30.94 (Fisher)	28.14 (Fisher)	32.26 (1-NN)
	Representation: Spectrograms + DTW		
Best result for 1-NND	43.43	21.89	42.11
Best Diss.-based classifier	20.01 (Fisher)	20.41 (Fisher)	24.63 (Fisher)

performance – obtained with the Spectrograms + DTW space with random selection of prototypes – which is an error of 20.01%. This error is rather high, this confirming the challenging nature of the seismic event classification problem when tested with realistically large data sets (as the one used in our study, involving 5 classes and more than 1200 events). Finally, it is interesting to observe that the best classifier in the dissimilarity space is in most cases a linear classifier (either Fisher or SVM), this confirming the goodness of the proposed representation space, in which classes are more linearly separable.

More specifically related to Experiment 1, we can observe from Table 1 the different behaviors of the three representation spaces: FFT + Euclidean (the baseline), Waveform + DTW and Spectrograms + DTW; in the results, the baseline (FFT + Euclidean) representation most of the times outperforms the Waveform + DTW one; however, the Spectrogram + DTW representation is consistently better than the other two. It makes a lot of sense since the spectrogram encodes information from both time- and frequency-domain. For what concerns Experiment 2 (Table 2), we can observe that the summarized results confirm that random selection is always a very good choice for selecting the prototypes, this being very interesting also from a practical point of view. Finally, looking at Table 3, an interesting fact emerges, related to the combination of different dissimilarities: the best performance for template matching (1-NND) is obtained when combining the three different representations while, in contrast, the best performance in the dissimilarity space was obtained with information from spectrograms and waveforms. As we pointed out at the end of Section 4.4, the behavior of combined dissimilarity representations requires further studies; nonetheless, results from Table 3 seem to indicate that dissimilarities from spectra add useful local information, which is beneficial for 1-NND while, in contrast, its inclusion slightly deteriorates the performance of dissimilarity-based classifiers which rely on more global information.

5. Conclusion

A dissimilarity space, based on the DTW distance measure, has been proposed to classify seismic volcanic patterns – either waveforms or spectrograms – as alternative to both the nearest neighbor rule directly applied to the dissimilarities and classification in a dissimilarity space based on the Euclidean distance between pairs of seismic spectra.

The proposal has been thoroughly studied by a set of experiments including learning curves to observe the behavior of the

Table 3
Best (lowest) classification errors obtained in the third experiment.

Classifier	Combined dissimilarity-based representations	
	“Waveforms+ spectrograms”	“Waveforms+ spectrograms+ spectra”
Best result for 1-NN	41.17	37.24
Best Diss.-based classifier	24.53 (Fisher)	25.27 (Fisher)

classification system over different training set sizes, curves to study the behavior for different dimensionalities of the DTW-based dissimilarity space and analyses on the possibilities of enhancing the classification performance by combining dissimilarity representations computed for different modalities of the seismic data (waveforms, spectra, spectrograms) or dissimilarity measures (DTW, Euclidean).

Results clearly showed the advantage of learning from a training set by building the DTW-based representation space, particularly when starting its construction from spectrograms. Low-dimensional versions of the proposed representation space allow to reduce the computational demands of a deployed system while preserving good enough classification performances; moreover, linear classifiers in the DTW-based dissimilarity space yield satisfactory classification accuracies as expected. Even though attempts to profit from the averaged combination of different dissimilarity representations were made, enhancements in the performances by applying this strategy were not observed. Further studies on this topic may be worthwhile.

Acknowledgements

Support from Dirección de Investigación - Sede Manizales (DIMA), Universidad Nacional de Colombia, is acknowledged as well as the Cooperint program from University of Verona. In addition, the authors would like to thank Programa de Formación Doctoral en Colombia - Becas COLCIENCIAS (Convocatoria 617 de 2013 - Convenio Especial No. 0656-2013, entre COLCIENCIAS y la Universidad Nacional de Colombia) for supporting the second author, and also the staff members from OVSM for providing the data set.

References

Castro-Cabrera, P.A., Orozco-Alzate, M., Adami, A., Bicego, M., Londoño-Bonilla, J.M., Castellanos-Domínguez, C.G., November 2014. A comparison between time-frequency and cepstral feature representations for seismic-volcanic pattern classification. In: Bayro-Corrochano, E., Hancock, E. (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Proceedings of the 19th Iberoamerican Congress on Pattern Recognition CIARP 2014*, Lecture Notes in Computer Science, IAPR, vol. 8827. Springer, Heidelberg, pp. 440–447.

Duin, R.P.W., Loog, M., Pekalska, E., Tax, D.M.J., 2010. Feature-based dissimilarity space classification. In: Únay, D., Çataltepe, Z., Aksoy, S. (Eds.), *Recognizing Patterns in Signals, Speech, Images and Videos: ICPR 2010 Contests*, Lecture Notes in Computer Science, vol. 6388. Springer, Berlin, Heidelberg, pp. 46–55.

Duin, R.P.W., Pekalska, E., 2007. The science of pattern recognition, achievements and perspectives. In: Duch, W., Mańdziuk, J. (Eds.), *Challenges for Computational Intelligence. Studies in Computational Intelligence*, vol. 63. Springer, Berlin, pp. 221–259.

Duin, R.P.W., Pekalska, E., 2012. The dissimilarity space: bridging structural and statistical pattern recognition. *Pattern Recognit. Lett.* 33 (7), 826–832.

Duin, R.P.W., Pekalska, E., Loog, M., 2013. Non-Euclidean dissimilarities: causes, embedding and informativeness. In: Pelillo, M. (Ed.), *Similarity-Based Pattern Analysis and Recognition. Advances in Computer Vision and Pattern Recognition*. Springer, London, pp. 13–44.

Ibba, A., Duin, R.P.W., Lee, W.-J., 2010. A study on combining sets of differently measured dissimilarities. In: 20th International Conference on Pattern Recognition (ICPR 2010), pp. 3360–3363.

Lemire, D., 2009. Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recognit.* 42 (9), 2169–2180.

Lin, J., Williamson, S., Borne, K.D., DeBarr, D., 2012. Pattern recognition in time series. In: Way, M.J., Scargle, J.D., Ali, K.M., Srivastava, A.N. (Eds.), *Advances in Machine Learning and Data Mining for Astronomy. Data Mining and Knowledge Discovery Series*. CRC Press, Boca Raton, FL, pp. 617–646 (Chapter 28).

Orozco-Alzate, M., García-Ocampo, M.E., Duin, R.P.W., Castellanos-Domínguez, C.G., 2006. Dissimilarity-based classification of seismic volcanic signals at Nevado del Ruiz volcano. *Earth Sci. Res. J.* 10 (December (2)), 57–65.

Pekalska, E., Duin, R.P.W., 2005. The dissimilarity representation for pattern recognition: foundations and applications. In: *Machine Perception and Artificial Intelligence*, vol. 64. World Scientific, Singapore.

Pekalska, E., Duin, R.P.W., Paclík, P., 2006a. Prototype selection for dissimilarity-based classifiers. *Pattern Recognit.* 39 (2), 189–208 (Special issue: Complexity Reduction).

Pekalska, E., Harol, A., Duin, R.P.W., Spillmann, B., Bunke, H., 2006b. Non-Euclidean or non-metric measures can be informative. In: Yeung, D.-Y., Kwok, J., Fred, A., Roli, F., de Ridder, D. (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition Joint IAPR International Workshops, SSPR 2006 and SPR 2006, Lecture Notes in Computer Science*, vol. 4109. Springer, Berlin, Heidelberg, pp. 871–880.

Pelillo, M. (Ed.), 2013. *Similarity-Based Pattern Analysis and Recognition. Advances in Computer Vision and Pattern Recognition*. Springer, London, UK.

Plasencia-Calaña, Y., Cheplygina, V., Duin, R.P.W., García-Reyes, E.B., Orozco-Alzate, M., Tax, D.M.J., Loog, M., July 2013. On the informativeness of asymmetric dissimilarities. In: Hancock, E.R., Pelillo, M. (Eds.), *Similarity-Based Pattern and Recognition: Proceedings of the Second International Workshop, SIMBAD 2013*, Lecture Notes in Computer Science, vol. 7953. Springer, Berlin, Germany, pp. 75–89.

Porro-Muñoz, D., Talavera, I., Duin, R.P.W., Orozco-Alzate, M., 2011. Combining dissimilarities for three-way data classification. *Comput. Syst.* 15 (1), 117–127.

Schilling, R.J., Harris, S.L., 2012. *Fundamentals of Digital Signal Processing using MATLAB*, 2nd edition. Cengage Learning, Stamford, CT.

Telesca, L., Lovallo, M., Chamoli, A., Dimri, V., Srivastava, K., 2013. Fisher-Shannon analysis of seismograms of tsunamigenic and non-tsunamigenic earthquakes. *Physica A: Stat. Mech. Appl.* 392 (16), 3424–3429.

Telesca, L., Lovallo, M., Kiszely, M.M., Toth, L., 2011. Discriminating quarry blasts from earthquakes in Vértes Hills (Hungary) by using the Fisher-Shannon method. *Acta Geophys.* 59 (5), 858–871.

Theodoridis, S., Koutroumbas, K., 2009. *Pattern Recognition*, 4th edition. Academic Press, Burlington, MA.

Wang, Q., August 2013. *Dynamic Time Warping (DTW) Package*. Available at (<http://www.mathworks.com/matlabcentral/fileexchange/>).