Classification of three-way data by the dissimilarity representation

Diana Porro-Muñoz [a,b,*], Robert P.W. Duin [b], Isneri Talavera [a], Mauricio Orozco-Alzate [c]

[a] *Advanced Technologies Application Center (CENATAV), Cuba*
[b] *Pattern Recognition Lab, TU Delft, The Netherlands*
[c] *Departamento de Informática y Computación, Universidad Nacional de Colombia, Sede Manizales, Colombia*

ARTICLE INFO

ABSTRACT

Representation of objects by multi-dimensional data arrays has become very common for many research areas e.g. image analysis, signal processing and chemometrics. In most cases, it is the straightforward representation obtained from sophisticated measurement equipments e.g. radar signal processing. Although the use of this complex data structure could be advantageous for a better discrimination between different classes of objects, it is usually ignored. Classification tools that take this structure into account have hardly been developed yet. Meanwhile, the dissimilarity representation has demonstrated advantages in the solution of classification problems e.g. spectral data. Dissimilarities also allow the representation of multi-dimensional objects in a way that the data structure can be used. This paper introduces their use as a tool for classifying objects originally represented by two-dimensional (2D) arrays. 2D measures can be useful to achieve this representation. A 2D measure to compute the dissimilarity representation from spectral data with this kind of structure is proposed. It is compared to existent 2D measures, in terms of the information that is taken into account and computational complexity.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The standard way of representing objects for classification is in a two-way structure (matrix), where a number of objects (rows) are simply characterized by feature vectors (1D representation). Nevertheless, in many research areas such as signal analysis, chemometrics and image analysis, objects observed by sensors are represented by higher-order generalizations of vectors and matrices i.e. several sets of features measured on objects, as for example, data collected at different times or conditions. The structure in which a set of objects with this representation is organized is called multi-way data.

Multi-way data analysis [1] is the extension of multivariate analysis when data is arranged in this multi-way structure. The most common is the three-way (3D) array, where objects are represented by matrices (2D representation) e.g. signals (objects, in the first direction of the array) represented by time points in the second direction and frequency in the third direction of the three-way data. The information obtained from such structures, e.g. interrelations between the different sets of variables, can be advantageous for many purposes as regression and/or classification.

Data will not be analyzed optimally by the traditional multivariate methods, which do not take into account the multi-way structure. If objects are analyzed in a 1D representation derived (unfolding for example) from the original higher-order one, the information of objects properties in one of the directions will be ignored. This may deteriorate the results. Moreover, fictitious relationships between the variables of the different directions may be created in this process. A number of methods for multi-way analysis have been proposed [1,2]. Most of

* Corresponding author at. Advanced Technologies Application Center (CENATAV), 7a # 21812 e/ 218 y 222, Rpto. Siboney, Playa, C.P. 12200, Ciudad de La Habana, Cuba. Tel.: +53 7 272 0422, 272 1670, 272 1676; fax: 53 7 273 0045.

E-mail addresses: dporro@cenatav.co.cu (D. Porro-Muñoz), r.duin@ieee.org (R.P. Duin), italavera@cenatav.co.cu (I. Talavera), morozcoa@bt.unal.edu.co (M. Orozco-Alzate).

these methods are for exploratory and regression purposes. Classification has been studied much less. This might be caused by the lack of classification tools able to operate on objects represented by multi-way arrays, that use all the available information.

Traditionally, pattern recognition systems are based on feature representations. Every object is represented by a feature vector that is constituted of object attributes that are characteristic for the differences between the pattern classes. The representation itself neglects possible dependencies between these attributes. They have to be found from the statistics in a training set. This holds in particular when as features samples of an object image a time signal or a spectrum is taken. The connectivity between pixels, time samples or frequencies is lost in the representation. By operations like a PCA they may be refound, but at the cost of a sufficient size of the training set.

To overcome the above problem of the feature representation, the Dissimilarity Representation (DR) has been developed [3–5]. It is based on a direct comparison of the total objects based on a dissimilarity measure. In this way, the geometry and the structure of a class are determined by a user defined dissimilarity measure in which application background information may be expressed. As this measure may respect the fact that the object has some shape in an image, as a function of time or in its spectrum, the above connectivity problem can be avoided.

The dissimilarity representation may generate a dissimilarity space, which is a general vector space. It can be used to train any of the traditional classifiers by a proper training set, represented by its dissimilarities to a selected set of prototype objects. This representation set may also be randomly selected or even be the entire training set.

In this paper, we introduce the use of the DR as a tool for classifying three-way data in such a way that objects are analyzed in their 2D structure. Consequently, the relationship between the object properties in the different variable directions of the three-way array can be included if a suitable dissimilarity measure is selected. Moreover, the relations between the objects are analyzed in the dissimilarity space. Thus, the key issue in this process is to find a dissimilarity measure that takes into account all relevant object differences for their classification. Information about the data that is not taken into account in the traditional representation or approaches can be included in the dissimilarity measure.

Although the introduced approach can be theoretically applied to any type of three-way data, we will focus in this paper on three-way spectral data e.g. signals represented by time points in one direction and frequency components in the other direction. With this purpose, we also try to construct a 2D measure that makes use of the 2D nature of the objects, and extracts the functional information e.g. shape and connectivity from this type of data. We will show how, by making use of the 2D structure, the discrimination of objects improves with respect to that obtained just using the traditional vectorial (1D) representation.

The paper is organized as follows. In Section 2, related works about DR e.g. on spectral data and other 2D measures proposed in the literature are analyzed. In Section 3, a brief introduction to the DR theory is given. In addition, the proposed generalization of the DR for three-way data is presented, together with the new 2D dissimilarity measure. This measure is compared to the other related measures in Section 2. The comparison is done in terms of the analyzed data and the computational complexity. In Section 4, the materials and methods applied in the experimental section are detailed. Following, the experimental results are presented and discussed in Section 5. In this section, in order to show the advantage of using the 2D representation of objects, the proposed approach is compared to the 1D representation of the analyzed data sets. Moreover, the new measure is compared to other ones in the literature for the two three-way spectral data sets analyzed in the paper. Some characteristics of the proposed measure are also analyzed. Finally, the drawn conclusions are presented in Section 6.

2. Related studies

Several studies to classify spectral data starting from their vectorial feature e.g. wavelengths representation have been done [6–12]. Nevertheless, for a wide variety of problems, the structure of the data can often be more complex than this; one can have several sets of variables measured on different samples, as for example, data collected at different times or conditions. There have not been many attempts to use this multi-dimensional representation as it is, and take advantage of it for a better discrimination of objects.

In signal processing, multi-dimensional representation of objects (mainly 2D) has been used e.g. video signal, space-time-wave analysis for source localization and detection [13], blind multiuser detection–estimation in direct-sequence code-division multiple-access (DS-CDMA) communication [14], 3D radar clutter modeling and mitigation [15–17], classification of sonar contacts [18], blind spatial signature estimation [19], detection of epilepsy [20] and alzheimer [21] from EEG signals, analysis of seismic signals [22–24], and sound and speech recognition [25,26], among others. However, what it is mainly done in these researches is to reduce the multi-way data to a vector for each object, by unfolding it in one of the variables direction e.g. from a spectrogram or scalogram, averaging in time to get the frequency spectrum, or vice versa. Afterwards, the authors do the processing of the data. For the classification task, a traditional classifier is applied after the data is transformed e.g. linear discriminant classifier and neural networks mainly [16,23–26].

In other research fields like chemometrics [27], the multi-way data analysis [2,28,1] has been one of the main topics for the last years, but most efforts are directed to solve regression and exploratory analysis problems [27,28]. These methods for multi-way data analysis have also been used in signal processing, but mainly for exploratory analysis (looking for the interrelation between the variables in the different directions of the multi-dimensional array) or for dimensionality reduction purposes [15,18,14]; applying a traditional classifier afterwards. So, classification has not been that explored in

multi-way analysis [1,29–31]. However, this approach is just based on a numerical analysis, where other aspects like the shape of the signal, which can also have valuable information for the discrimination between different patterns of signals (classification), is not taken into account.

On the other hand, although it is a rather new technique, the DR has been applied in many fields e.g. image analysis [32,33] and spectral data analysis [5,4]. In the latter, the DR has shown to be advantageous as it addresses the dimensionality ill-posed problem of most spectral data sets. Normally, this kind of data is characterized by having a few objects but represented with very high-dimensional vectors. Consequently, by using the DR, a dimensionality reduction is achieved, even if the entire training set is used as representation set. Moreover, background knowledge on the data can be expressed in the dissimilarity measure. Some studies have been done to find suitable dissimilarity measures for spectral data [5] to apply the DR approach, and more specifically for seismic volcanic signals [4,34] and chemical data [35].

The DR can be generated from any other representation of the objects e.g. vectors of numbers, graphs, or multi-way data, as long as the suitable dissimilarity measure is found. However, it has not been done yet for the last one. Nonetheless, some 2D measures (mainly for image processing) have been used and developed for the dissimilarity-based k-Nearest Neighbor (k-NN) classifier, with the purpose of making use of the two-dimensional (2D) data array representation e.g. Frobenius [36], Yang [37] and Volume [38] distances, and the Assembled Matrix Distance (AMD) [39]. But none of these measures takes the spectral information, e.g. continuity, shape, into account. Some measures are analyzed in the subsequent sections together with the one proposed in this paper.

3. Dissimilarity representation from three-way data

The Dissimilarity Representation [3] was proposed as a more flexible representation than the feature representation, with the purpose of having more information about the structure of the objects. It is seen as a link between the statistical and structural approaches, as both types of patterns can be described by the (dis)similarity measure. More possible prior knowledges about the data, e.g. rotation and size independence of images, shift independence of spectra, can be taken into account. The DR is also based on the role that (dis)similarities play in a class composition, where objects from the same class should be similar and objects from different classes should be different (compactness property). Hence, it should be easier for the classifiers to discriminate between them.

Using the DR, classifiers are trained in the space of the proximities between objects, instead of the traditional feature space. Thus, in place of the feature matrix $\mathbf{X} \in \mathbb{R}^{n \times q}$, where n runs over the objects and q over the variables, the set of objects is represented by the matrix $\mathbf{D}(\mathbf{X}, \mathbf{R})$. This matrix contains the dissimilarity values $d(x_i, r_j)$ between each object x_i of $\mathbf{X}$ and the objects r_j of the representation set $\mathbf{R}(r_1, \ldots, r_h)$, where h is the number of prototypes. We build from this matrix a dissimilarity space. Objects are represented in this space by the row
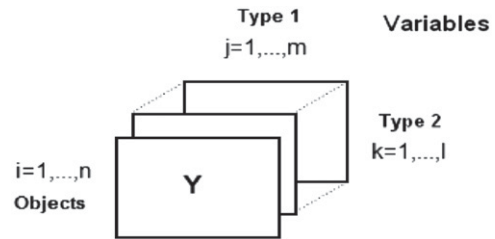

Fig. 1. Design of a three-way array.

vectors of the dissimilarity matrix. Each dimension corresponds to the dissimilarities with one of the representation objects.

The elements of $\mathbf{R}$ are called prototypes, and have preferably to be selected by a prototype selection method [3]. These prototypes are usually the most representative objects of each class, $\mathbf{R} \subseteq \mathbf{X}$ or $\mathbf{X}$ itself, resulting in a square dissimilarity matrix $\mathbf{D}(\mathbf{X}, \mathbf{X})$. $\mathbf{R}$ and $\mathbf{X}$ can also be chosen as different sets. As dissimilarities are computed to $\mathbf{R}$, a dimensionality reduction is reached if a good, small set can be found, resulting in less computationally expensive representation and classifiers.

For a t-dimensional array $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_t}$ (see Fig. 1), the theory of the DR is the same. In fact, one of the advantages of the DR is that it can be generated from any representation of the objects e.g. vectors of numbers or graphs, as long as we have a proper dissimilarity measure. This applies also to the multi-way data. Originally, each object is represented by a $(t-1)$-dimensional array of numerical values and all the objects together compose the t-dimensional array. Hence, to obtain the dissimilarity space, a mapping $\phi(\cdot, \mathbf{R}) : \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_{t-1}} \to \mathbb{R}^h$ is defined, such that for every object y, $\phi(y, \mathbf{R}) = [d(y, r_1), d(y, r_2), \ldots, d(y, r_h)]$, where h is the number of prototypes. Classifiers are then built in this space, as in any feature space.

Thus, to apply this approach to any classification problem the following steps should be followed:

1. Design the three-way (multi-way) data from the problem at hand, like $objects \times variable1 \times variable2$.
2. Define the dissimilarity measure according to the characteristics of the data. This measure should include all possible relevant knowledge, e.g. on the (in)significance of the tails of a spectrum.
3. Compute the dissimilarity matrix between the new objects and the representative objects (prototypes).
4. Prototype selection (if needed, depending on the classifier and the computational demands), with one of the methods reported in the references.
5. Build a classifier, using the dissimilarity matrix as input data, as the new description of the objects will be based on their dissimilarities to the prototypes from each class.

The issue to be addressed here is how to obtain the dissimilarities from the multi-way representation. Many ideas can arise to do this transformation. We propose as a first approach, focusing in three-way data, to take each object (matrix) (matrix) y of $\mathbf{Y}$ and compute the

dissimilarities between them by a 2D dissimilarity measure. Some 2D measures have been proposed in [39] for face and palm-print recognition. However, the selection of the suitable measure for the problem at hand is a very important aspect in the DR approach. To deepen in this task we will focus in our case of study on three-way spectral data i.e. objects represented by spectra and/or time signals in the two variable directions. Thus, each object is represented by a matrix (2D). A comparative study is made about the characteristics of each data set and the dissimilarity measure to be used. A 2D dissimilarity measure is proposed.

3.1. 2Dshape measure

In many types of data, e.g. spectral data, it is necessary to take into account the shape information and connectivity between the measured points. In this way, the observations in the spectrum can be seen as continuous single entities, instead of sets of different variables.

As mentioned in Section 2, some measures for 2D representation of samples have been proposed. Assume that the three-way array $\mathbf{Y} \in \mathbb{R}^{n \times m \times l}$, where n is the number of objects, and m and l the number of variables (related to a single object) in each of the other directions, respectively; $\forall j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, l$. Then, the AMD measure [39] for two objects y_a and y_b of $\mathbf{Y}$ is defined as

$$d_{AMD}(y_a, y_b) = \left(\sum_{k=1}^{l} \left(\sum_{j=1}^{m} (y_{a,j,k} - y_{b,j,k})^2 \right)^{p/2} \right)^{1/p} \quad (1)$$

The power p is used to emphasize either small or large differences between the elements, in dependence of the problem at hand. If $p < 1$, all the differences are reduced, thus the larger ones do not interfere much in the measure. On the other hand, if $p > 1$, the larger differences will be more pronounced, resulting in a heavy influence on the measure. This measure is a generalization of the Frobenius and Yang distance measures referenced in Section 2. When $p = 1$ in AMD, it is the same as the Yang distance, and for $p = 2$ is then the Frobenius distance.

These measures could be a good option when the spectral (functional) information can be assumed to be present in the data representation. However, this is not usually the case, as spectra are observed and recorded discretely. Consequently, they are analyzed with multivariate data analysis techniques which consider the spectrum as high-dimensional vectors of different but high-correlated variables, instead of a continuous single entity. Therefore, when the information is not taken into account in the representation of the data, the dissimilarity measure has to take care of it.

Hence, considering the results obtained with the Shape measure (Manhattan distance on the first Gaussian derivatives) for simple spectra [5], we propose to make use of the derivatives into the AMD measure. In such a way, we can take the ordering information into account as well as the shape of the spectra. A principle of the DR approach is that, instead of a single representation of a problem, one may also consider either a complex representation, built from many dissimilarity representations, where different aspects of the data are described in various ways [3]. Based on this and the previously stated, and in a way that the information available in both directions of the 2D data can be taken into account, we define the 2Dshape dissimilarity measure as follows:

1. Compute the matrix D^1

$$D^1_{a,b} = \left(\sum_{k=1}^{l} \left(\sum_{j=1}^{m} (y^\sigma_{a,j,k} - y^\sigma_{b,j,k})^2 \right)^{p_1/2} \right)^{1/p_1}$$

$$y^\sigma_{i,j,\cdot} = \frac{\mathrm{d}}{\mathrm{d}_j} G(j, \sigma) * y_{i,j,\cdot}$$

2. Compute the matrix D^2

$$D^2_{a,b} = \left(\sum_{j=1}^{m} \left(\sum_{k=1}^{l} (y^\sigma_{a,j,k} - y^\sigma_{b,j,k})^2 \right)^{p_2/2} \right)^{1/p_2}$$

$$y^\sigma_{i,\cdot,k} = \frac{\mathrm{d}}{\mathrm{d}_k} G(k, \sigma) * y_{i,\cdot,k}$$

3. Combine both dissimilarities matrices $D = \alpha_1 D^1 + \alpha_2 D^2$

The variables $y_{i,j,\cdot}$ and $y_{i,\cdot,k}$ stand for the kth columns and the jth rows of the ith matrix (object); $\forall i = 1, 2, \ldots, n$. Their expressions correspond to the computation of the first Gaussian (that is what G stands for) derivatives of spectra. Thus, a smoothing (blurring) is done by a convolution process ($*$) with a Gaussian filter and σ stands for a smoothing parameter [5]. The dissimilarities in step 1 and step 2 correspond to the first and second directions, respectively, as indicated by the notation e.g. spectra and time. This measure can also be used in three-way data where there are no variations in shape in one of the directions. In this case, it is enough to use the AMD measure in step 1 or step 2 only, such that only the differences in area are compared. Moreover, different p-values can be applied in the different directions. In the combination step, we included a weight for scaling. In this case, we defined $\alpha_c = D^c / max(D^c)$, to normalize the dissimilarity matrices.

If we analyze the computational complexity of all these measures, they are in the order of $\mathcal{O}(m \cdot l)$, as our samples have two-dimensions (two types of variables). In general, to compute the whole dissimilarity matrices, the computational complexity of all of them is $\mathcal{O}(n^2 \cdot m \cdot l)$. We are assuming here the worst case, in which all the training samples are used as prototypes. Nevertheless, if we really take into account the number of operations required for each measure, they are different. Between the Frobenius, Yang and AMD measures there are some slight differences in the number of operations required. Nevertheless, when we analyze the 2Dshape measure, it requires twice the number of operations needed for the other measures, plus the last sum operation. Moreover, in case of applying the derivatives, these operations are also added, and it depends on the selected smoothing parameter σ. So, from this point

of view, depending on the problem at hand, it is needed a trade-off between computational complexity and classification accuracy.

4. Materials and methods

Two three-way spectral data sets will be studied in this paper. The first is a public domain data set and the description has been taken from the website [40,41] for a better understanding of this paper. It consists of samples of red wine, produced from the same grape (*Cabernet Sauvignon*) and belonging to different geographical areas and producers. They were collected from local supermarkets and analyzed by means of HS-GC–MS (headspace gas chromatography/mass spectrometry). Separation of aroma compounds was carried out on a gas chromatography system (2700 columns from the scans of chromatographic profile). For each sample, a mass spectrum scan (m/z: 5–204) measured at the 2700 elution time points was obtained, providing a data cube of size $44 \times 2700 \times 200$ i.e. samples (objects) in first direction, elution time points in second direction and mass spectrum in third direction. The data set is composed of samples from three different geographical areas: South America (21 samples), Australia (12 samples) and South Africa (11 samples). For the two-way representation (1D representation of objects) of the data, the three-way data was unfolded in its second direction, obtaining a matrix of size $44 \times 540\,000$. All-zero columns were deleted in this representation (none of the samples have information in these columns), so the final data set has a size of $44 \times 117\,060$. The dissimilarity representation has a size of 44×44.

The second data set corresponds to seismic signals from the ice-capped Nevado del Ruiz volcano in the Colombian Andes, currently studied by the Volcanological and seismological observatory at Manizales. Signals were digitized at 100.16 Hz sampling frequency by using a 12 bit analog-to-digital converter. The data set for the experiments is composed of 12 032-point signals of two classes of volcanic events: 235 of Long-Period (LP) earthquakes, and 235 of Volcano-Tectonic (VT) earthquakes. A 2D time–frequency representation was computed with short-time Fourier transform (STFT), obtaining a spectrogram from each signal [10]. To compute these spectrograms, trying to achieve a trade-off between time and frequency resolution, a 256-point (window size) STFT was calculated with 50% overlap. With this technique, it can be known what frequency intervals are present in a time interval of the signal and use it for the discrimination between classes. The concatenation of the spectrograms of the different signals (objects) will result in a $470 \times 93 \times 129$ three-way data i.e. events (objects) in first direction, time points in second direction and frequency components in the third direction. For the 1D (spectral) representation of each object, we have computed the spectrum by using a 12 032-point fast Fourier transform (FFT), leading to a $470 \times 12\,032$ data. Consequently, the information in the whole signal is analyzed in both its 1D and 2D representation. The dissimilarity representation has a size of 470×100.

For both data sets, the related measures in Section 3.1 are used to obtain the DR from their 2D representation.

This way, we can analyze which should be a suitable measure for each problem, and compare how the results behave in each case. The experiments are also made for several p-values. Hence, we can also compare all the measures related before. For the wine data set, $p = [0.1, 0.5,1,2,3]$ and also for both directions of the volcano data set. In the case of the 2Dshape measure, experiments were also ran by exchanging the p-values of both directions. The σ parameter was optimized in a grid-search procedure with 10-fold cross-validation, for the different p-values in both data sets. In the case of wine data, as the number of objects is so small, the optimization procedure was done with the whole data set. The best results were obtained for $\sigma = 5$. As we are measuring shape in just one of the two directions, only one σ parameter had to be optimized. In the case of the seismic volcanic data, 170 samples (85 of each class) were used to optimize the σ parameter for each direction. The rest of the data was then used to evaluate the classification performances, by using the best σ values ($\sigma = 2$ for the time direction and $\sigma = 3$ for the frequency direction) overall p.

As mentioned previously, we will make a comparison between the classification accuracy by making use of the 2D structure or just the 1D. In the case of the 2D representation, the 2D measures explained in Section 3.1 will be used and compared. For 1D representation, we will use the shape measure introduced in [5]. It is defined as the Manhattan distance on the Gaussian derivatives of the spectra. As this measure is for vectors, only one σ has to be used. The σ parameter in this measure was also optimized in a cross-validation procedure and the best results were achieved with $\sigma = 15$ for the volcanic data set and $\sigma = 20$ for the wine data set.

The regularized linear discriminant classifier (RLDC) [42,43] was built on the DR obtained from the different representations of the two data sets. In order to find the regularization parameters, an automatic regularization (optimization over training set by cross-validation) process was done. Experiments were repeated 10 times. Training and test objects were randomly chosen from the total data sets, in a 10-fold cross-validation process. For the wine data set, as the size of the training set is so small, we decided to use all the objects as prototypes. A random prototype selection was performed with several numbers of prototypes for the volcanic data set. The best results were obtained for 100 prototypes, which are the ones shown here. In both cases, the same training and test sets were used for all the representations, so the results can be comparable. Classifiers performances are evaluated in terms of the average classification error (ACE), and the standard deviation from the different repetitions is taken into account.

The experiments were all performed in Matlab. For the computation of the spectrograms, we used the Signal Processing Toolbox from Matlab, and PRTools toolbox [44] for the computation of DR and classification of the data.

5. Experimental results and discussion

In this section we present several analyses. First, we want to compare the spectral data sets by applying the proposed measure and the other 2D measures of the

literature. This is in order to demonstrate the importance of selecting a suitable measure for the data at hand. An analysis of the influence of the parameter p in the results will also be performed. Afterwards, a comparison between the 2D and 1D representations of the objects will be done, to show the advantage of making use of the 2D structure data for a better discrimination of samples.

5.1. 2Dshape measure vs others measures

In chemometrics, in the case of the techniques combined with chromatography, in the chromatography direction (2nd) we will have the eluded peaks for all the components present in the substances. But in the spectral mode (third direction, mass spectra for GC–MS), each eluded component will only have one mass spectrum (mass fragments in which the molecule decomposes) independently of the class. This means that, with these techniques, if the classes differ because they have different components, the only thing we will see is the absence/presence of the peak or some differences in the concentration of the mass fragments.

Nevertheless, the other difference between classes that we can find is the relation between the eluded components in the chromatogram i.e. how the concentration of one of the peaks varies with respect to the others, for the different classes. In this case, it is important to take shape into account, because there is information in the ordering of the components (peaks) with different concentrations and also continuity.

Thus, for the wine data set we propose to adapt the 2Dshape measure defined in Section 3.1, which takes the information of both directions into account, to the specificities of the data. When computing the D^1 matrix for the chromatography direction, we will use the Gaussian derivatives to take into account the shape in the changes of concentration in the neighboring components. However, for the D^2 matrix from the mass spectra mode, the use of derivatives is meaningless, because there is no continuity between the mass fragments; just the differences between the concentration of the mass fragments will be computed.

In Figs. 2 and 3, the results for AMD and 2Dshape measures are shown. Notice again that when $p=1$, we are using the Yang distance and when $p=2$, the Frobenius. This way, we are comparing the three measures related in Section 3.1. If we analyze Fig. 2, we can see that the errors are in a range of 20–50% for the three measures (no functional information about the data is taken into account). Nevertheless, when we look at Fig. 3, the highest
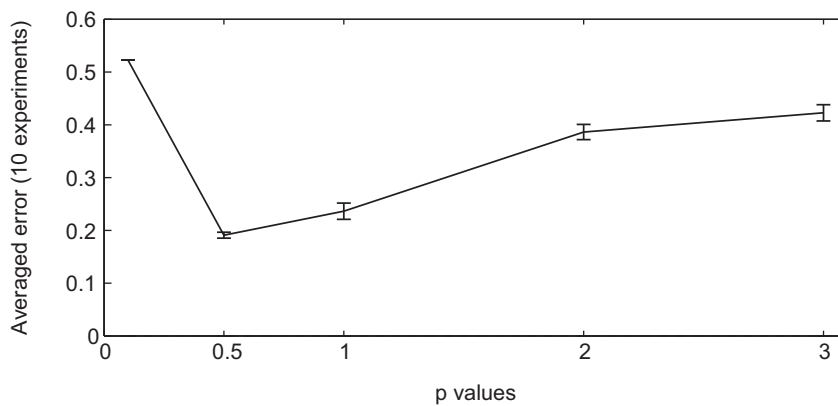

Fig. 2. Average cross-validation error (with standard deviation) for the AMD measure for different p-values for wine data set.
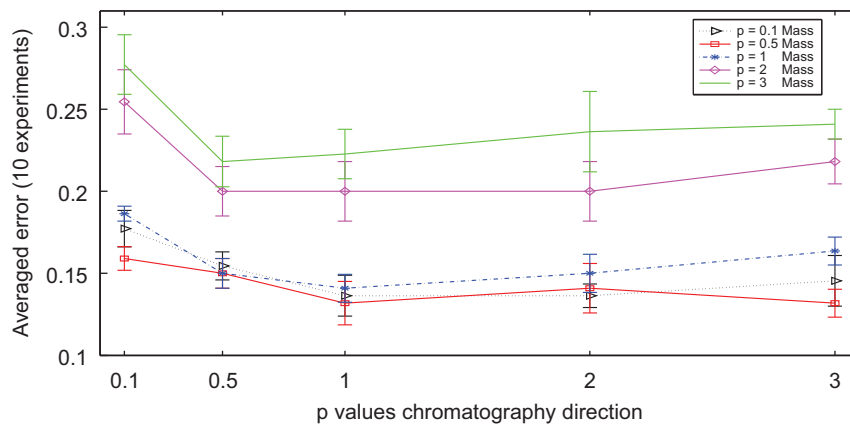

Fig. 3. Average cross-validation error (with standard deviation) for the 2Dshape measure for different p-values on both directions for wine data set.

error is around 28%, and coincides with the *p*-values for which the worst error was obtained with the AMD measure. It can be noticed that, by measuring the information in both directions and taking the functional, e.g. shape, information into account, we have decreased the errors to around 12–28%. Hence, from this data set we might conclude that, even when the *p* parameter is not optimized, by taking the spectral information into account better results can be achieved.

In Fig. 4, the classification results for the DR from the different directions separately are shown.

We can see from this figure that there is one direction for which better results are obtained. In this case it is the second direction (chromatography), which makes sense because we are analyzing the relation of concentration of components (shape) for the different classes. However, the third direction (mass spectrum), although is not that informative, can also help to discriminate. From this figure and Fig. 3, we can notice that if we combine the dissimilarity matrices from both directions (using the *p*-values for which the best results are obtained in each direction separately) the independent results for each direction are outperformed. This corroborates our initial hypothesis of getting a better discrimination between classes, if we take the information of both directions into account. In this case, we can also

see how the lowest error achieved in the combination (around 12%) was obtained by using $p=1$ in the second direction and $p=0.5$ in the third direction. These are the *p*-values for which the best results were obtained in the independent analysis.

A good example where the proposed measure in Section 3.1 can be useful is in this time–frequency representation of the second data set obtained by spectrograms. In this case, shape changes are present in the spectral (frequency) direction and connectivity in the time direction (due to the windows overlapping). In the next figures, the ACE are shown for the AMD (different *p*-values) measure in Fig. 5 and 2Dshape measure in Fig. 6.

From the figures we can observe a similar behavior to the ones of the Wine data set. With the proposed measure, where the shape and continuity information is taken into account, we got to outperform the AMD from a lowest error of around 27% of ACE, to 21%. This suggests that the proposed 2D measure is capable of capturing the information needed. Also, we should take a look at the results (ACE) when analyzing the directions separately (see Fig. 7). The results for the time direction (second) are not very good. The ones for the third direction (frequency) on the other hand are much better. This also makes sense, and is corroborated by the fact that spectral-based classification is often used for this type of data, as spectral content of
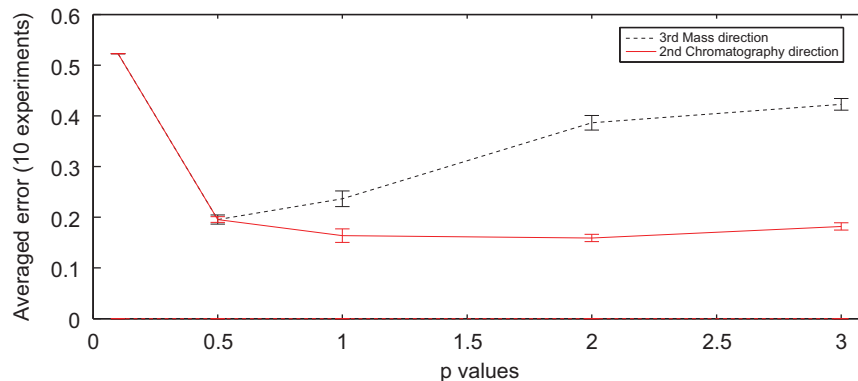

Fig. 4. Average cross-validation error (with standard deviation) for the two directions separately in the 2Dshape measure for wine data set. Different *p*-values are analyzed.
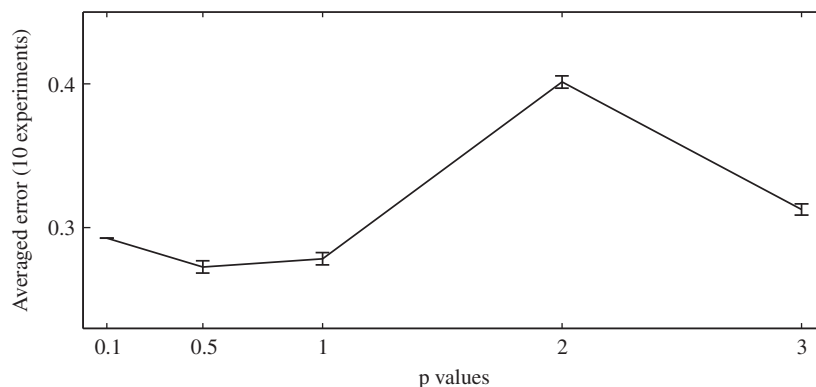

Fig. 5. Average cross-validation error (with standard deviation) for the AMD measure for different *p*-values for seismic volcanic data set.
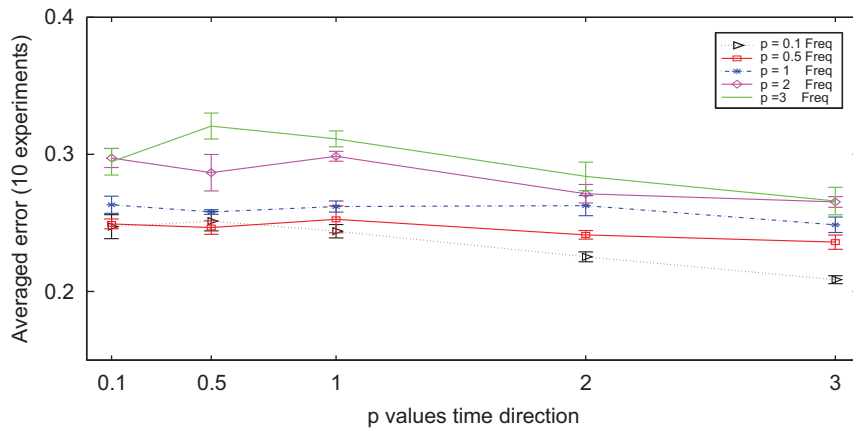
Fig. 6. Average cross-validation error (with standard deviation) for the 2Dshape measure for different *p*-values on both directions for seismic volcanic data set.
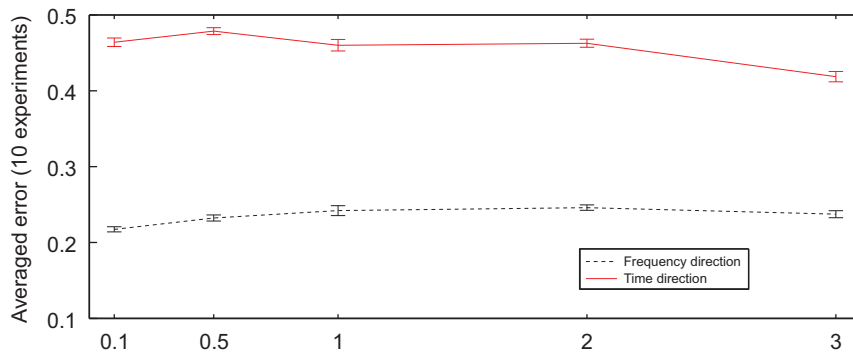

Fig. 7. Average cross-validation error (with standard deviation) for the two directions separately in the 2Dshape measure for seismic volcanic data set. Different *p*-values are analyzed.

these signals allows the discrimination between the events (events do not change in time heavily). Nevertheless, when the information from both directions is taken into account, better results are achieved, as shown in the results in Fig. 6. Moreover, we observe here that, in the case where the information from one of the directions is not sufficiently discriminative, the least we will get are the results similar to the ones for the discriminative direction.

We can also see from Figs. 6 and 7 that the best performances with the 2Dshape measure are achieved for the *p*-values for which the best results are obtained in each direction independently.

5.2. 1D vs 2D object representation

In this section we make a comparison between the classification results when using the vectorial (1D) of objects and when making use of the 2D structure. In Table 1, the ACE for the DR from the 1D and 2D representations of both data sets can be observed. In Section 5.1, an analysis of different 2D measures was done, and in both data sets the best results were obtained when applying a version of the proposed 2Dshape measure. Thus, in this section, the ACE values shown in

Table 1
Average cross-validation error in % (with standard deviation) for wine and seismic volcanic data sets.

Representations	Wine	Volcano
1D	36.1(1.4)	30.2(0.4)
2D	12.1(0.5)	20.9(0.2)

Table 1 for the 2D representation of both data sets are those best values obtained in the previous section. For the Wine data set, the results achieved with the unfolding procedure are not very good. By applying the DR on this representation, the high-dimensionality of the obtained data (which is one of the main problems in this procedure) is reduced. This is because now in the new dissimilarity space, the dimensionality is given by the number of objects in the representation set (training set in this data set). Still, fictitious connections between the end point of the variables in one direction and the start of the others are inserted. In any direction that we do the unfolding, the same phenomenon will happen; some relation will be lost or its benefit will not be used. However, these results are notably outperformed when using the 2D representation.

In the case of the seismic volcanic data set, the differences in 1D spectral content of a signal allow for the discrimination between the events. Nevertheless, with this representation we are not able to use the changes of frequency content in time to separate classes. It can also be observed in Table 1 that the ACE on the dissimilarity space generated from just the spectral data (1D) is around 30%. Nevertheless, when we analyze the error of the DR from the 2D representation we see a significant improvement. This ratifies the fact that the time–frequency relation is more discriminative than just the spectral information.

Thus, it can be observed that it is usually advantageous to make use of the multi-dimensional structure of the data for the classification process. The information and relationship between the variables of the different directions can be more discriminative, than if we just obtain a vector (1D) of variables from its 2D representation, ignoring its original structure.

6. Conclusions

We introduced the use of the dissimilarity representation as a tool for classifying three-way data. In this approach, objects are analyzed in their matrix (2D structure) representation by using 2D measures. Moreover, information about the data that is missing in the original representation, e.g. shape, can be considered in it. We developed a new 2D dissimilarity measure that allows taking into account the shape and continuity information in the directions of three-way spectral data. Furthermore, the relationship between the different dimensions is taken into account in this measure. It was compared to other three measures of the literature, in order to proof the importance of the selection of a suitable dissimilarity measure for the problem at hand. We also showed that, even when there is more discriminative information in one direction of the three-way data than in other, the results usually improve by combining.

The good performance of classifiers on the DR from the 2D representation of the objects, compared with the one from the traditional 1D, shows that this approach can be a good solution for the classification of objects with a 2D structure. Although this paper was focused on the solution for three-way data, it can be extended to multi-way e.g. in chemometrics, molecular entities of a substance can be separated by size on a chromatographic system and then detected by fluorescence, leading to a four-way data of (objects)×(fractions)×(excitation)×(emission). The proposed measure, as the dissimilarity representation, could be extended to multi-way data; it is part of the future work. Moreover, this procedure could be used in other types of problems where shape changes help for discrimination, like in image processing e.g. for the classification of faces.

Acknowledgments

References

[1] D. Porro-Muñoz, I. Talavera, R.P.W. Duin, Multi-way data analysis, Technical Report RNPS No. 2142, CENATAV, 2009.
[2] P.M. Kroonenberg, Applied Multiway Data Analysis, Wiley, Hoboken, NJ, 2008.
[3] E. Pekalska, R.P.W. Duin, The Dissimilarity Representation for Pattern Recognition. Foundations and Applications, World Scientific, 2005.
[4] M. Orozco-Alzate, M.E. García, R.P.W. Duin, C.G. Castellanos, Dissimilarity-based classification of seismic signals at Nevado del Ruiz Volcano, Earth Sciences Research Journal 10 (2) (2006) 57–65.
[5] P. Paclik, R.P.W. Duin, Dissimilarity-based classification of spectra: computational issues, Real Time Imaging 9 (4) (2003) 237–244.
[6] C.P. Grill, V.N. Rush, Analysing spectral data: comparison and application of two techniques, Biological Journal of the Linnean Society 69 (2000) 121–138.
[7] K. Varmuza, P. He, F. Kai-Tai, Boosting applied to classification of mass spectral data, Journal of Data Science 1 (2003) 391–404.
[8] F. Rossi, N. Villa, Support vector machine for functional data classification, Neurocomputing 69 (7–9) (2006) 730–742 (New Issues in Neurocomputing: 13th European Symposium on Artificial Neural Networks).
[9] H. Shinzawa, S. Morita, Y. Ozaki, R. Tsenkova, New method for spectral data classification: two-way moving window principal component analysis, Applied Spectroscopy 60 (8) (2006) 884–891.
[10] M. Benbrahim, A. Daoudi, K. Benjelloun, A. Ibenbrahim, Discrimination of seismic signals using artificial neural networks, in: C. Ardil (Ed.), WEC (2), Enformatika, Çanakkale, Turkey, 2005, pp. 4–7.
[11] M.C. Benítez, J. Ramírez, J.C. Segura, J.M. Ibáñez, J. Almendros, A. García-Yeguas, G. Cortés, Continuous HMM-based seismic-event classification at Deception Island, Antarctica, IEEE Transactions on Geoscience and Remote Sensing 45 (1) (2007) 138–146.
[12] G. Curilem, J. Vergara, G. Fuentealba, G. Acu na, M. Chacón, Classification of seismic signals at Villarrica Volcano (Chile) using neural networks and genetic algorithms, Journal of Volcanology and Geothermal Research 180 (1) (2009) 1–8.
[13] H. Becker, P. Comon, L. Albera, M. Haardt, I. Merlet, Multiway space-time-wave-vector analysis for source localization and extraction, in: EUSIPCO 10, XVIII European Signal Processing Conference, Aalborg, Denmark, 2010, pp. 23–27.
[14] N.D. Sidiropoulos, G.Z. Dimic, Blind multiuser detection in W-CDMA systems with large delay spread, IEEE Signal Processing Letters 8 (9) (2001) 87–89.
[15] T. Li, N.D. Sidiropoulos, G.B. Giannakis, PARAFAC STAP for the UESA Radar, in: Proceedings of ASAP2000, MIT Lincoln Laboratory, Lexington, MA, 2000.
[16] M.C. Wicks, M. Rangaswamy, R. Adve, T.B. Hale, Space-time adaptive processing: a knowledge-based perspective for airborne radar, Signal Processing Magazine, IEEE 23 (1) (2006) 51–65.
[17] V.F. Mecca, D. Ramakrishnan, J.L. Krolik, MIMO Radar space-time adaptive processing for multipath clutter mitigation, in: Proceedings Fourth IEEE Workshop on Sensor Array and Multichannel Processing, Waltham, MA, 2006, pp. 249–253.
[18] R.L. Oliveira, B.S.L.P. de Lima, N.F.F. Ebeckena, The use of multi-way analysis in the classification task of passive sonar contacts, Mecánica Computacional 29 (2010) 9389–9405.
[19] Y. Rong, S.A. Vorobyov, A.B. Gershman, N.D. Sidiropoulos, Blind spatial signature estimation via time-varying user power loading and parallel factor analysis, IEEE Transactions on Signal Processing 53 (5) (2005) 1697–1710.
[20] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, B. Yener, Multiway analysis of epilepsy tensors, Bioinformatics 23 (2007) i10–i18.
[21] D. Latchoumane, V. Charles Francois, A. Vialatte, F. Cichocki, J. Jeong, Multiway analysis of Alzheimer's disease: classification based on space-frequency characteristics of EEG time series, in: Proceedings of the World Congress on Engineering WCE 2008, vol. II, 2008.
[22] T. Bartosch, D. Seidl, Spectrogram analysis of selected tremorsignal using short-time Fourier transform and continuous wavelet transform, Annali di Geofisica 42 (3) (1999) 497–506.
[23] P. Lesage, F. Glangeau, J. Mars, Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events, Journal of Volcanology and Geothermal Research 114 (2002) 391–417.

[24] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, R. Campanini, Automatic Classification of Volcanic Tremor using Support Vector Machine Conception, Verification and Application of Innovative Techniques to Study Active Volcanoes, Istituto Nazionale di Geofisica e Vulcanologia Press, 2008.

[25] R. Rifkin, J. Bouvrie, K. Schutte, S. Chikkerur, M. Kouh, T. Ezzat, T. Poggio, Phonetic classification using hierarchical, feed-forward, spectro-temporal patch-based architectures, Technical Report MIT-CSAIL-TR-2007-019 CBCL-267, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, 2007.

[26] J. Dennis, H.D. Tran, H. Li, Spectrogram image feature for sound event classification in mismatched conditions, IEEE Signal Processing Letters 18 (2) (2011) 130–133.

[27] A.K. Smilde, R. Bro, P. Geladi, Multi-way analysis, in: Applications in the Chemical Sciences, Wiley, England, 2004.

[28] E. Acar, B. Yener, Unsupervised multiway data analysis: a literature survey, IEEE Transactions on Knowledge and Data Engineering 21 (2009) 6–20.

[29] J. Sádecká, J. Tóthová, Fluorescence spectroscopy and chemometrics in the food classification—a review, Czech Journal of Food Sciences 25 (2007) 159–173.

[30] D. Ebrahimi, J. Li, D.B. Hibbert, Classification of weathered petroleum oils by multi-way analysis of gas chromatography-mass spectrometry data using PARAFAC2 parallel factor analysis, Journal of Chromatography A 1166 (1–2) (2007) 163–170.

[31] G.J. Hall, J.E. Kenny, Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA, Analytica Chimica Acta 581 (2007) 118–124.

[32] E. Pekalska, R.P. W. Duin, On combining dissimilarity representations, in: MCS '01: Proceedings of the Second International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 2096, London, UK, 2001, pp. 359–368.

[33] S.W. Kim, R.P.W. Duin, On combining dissimilarity-based classifiers to solve the small sample size problem for appearance-based face recognition, in: CAI '07: Proceedings of the 20th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence, Lecture Notes in Computer Science, vol. 4509, 2007, pp. 110–121.

[34] M. Orozco-Alzate, M. Skurichina, R.P.W. Duin, Spectral characterization of volcanic earthquakes at Nevado del Ruiz Volcano using spectral band selection/extraction techniques, in: Progress in Pattern Recognition, Image Analysis and Applications. Proceedings of the 13th Iberoamerican Congress on Pattern Recognition CIARP 2008, Lecture Notes in Computer Science, vol. 5197, Springer, 2008, pp. 708–715.

[35] D. Porro-Muñoz, I. Talavera, R.P.W. Duin, N. Hernández, The representation of chemical spectral data for classification, in: Proceedings of the 14th Iberoamerican Congress on Pattern Recognition CIARP 2009, Lecture Notes in Computer Science, vol. 5856, Springer, 2009, pp. 513–520.

[36] J. Yang, J.Y. Yang, From image vector to matrix: a straightforward image projection technique-IMPCA vs. PCA, Pattern Recognition 35 (2002) 1997–1999.

[37] J. Yang, D. Zhang, A. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.

[38] J. Meng, W. Zhang, Volume measure in 2DPCA-based face recognition, Pattern Recognition Letters 28 (2007) 1203–1208.

[39] W. Zuo, D. Zhang, K. Wang, An assembled matrix distance metric for 2DPCA-based image recognition, Pattern Recognition Letters 27 (2006) 210–216.

[40] Wine dataset, 〈http://www.models.kvl.dk/datasets.html〉, 2008.

[41] T. Skov, D. Ballabio, R. Bro, Multiblock variance partitioning. A new approach for comparing variation in multiple data blocks, Analytica Chimica Acta 615 (1) (2008) 18–29.

[42] F.J. H, Regularized discriminant analysis, Journal of the American Statistical Association 84 (405) (1989) 165–175.

[43] R.O. Duda, P.E. Hart, D.G. Stork, in: Pattern Classification, John Wiley and Sons, New York, 2001.

[44] R.P.W. Duin, P. Juszczak, D. de Ridder, P. Paclik, E. Pekalska, D.M.J. Tax, Prtools, a matlab toolbox for pattern recognition, 〈http://www.prtools.org/download.html〉, 2004.