A study on the influence of shape in classifying small spectral data sets

Diana Porro-Muñoz[1,2], Robert P. W. Duin[2], Isneri Talavera[1] and Mauricio Orozco-Alzate[3]

[1] Advanced Technologies Application Center (CENATAV), Cuba, and Pattern Recognition Lab, TU Delft, The Netherlands,
[2] Universidad Nacional de Colombia Sede Manizales, Colombia
[1]{dporro,italavera}@cenatav.co.cu,[2]r.duin@ieee.org,
[3]morozcoa@bt.unal.edu.co

Abstract. Classification of spectral data has raised a growing interest in may research areas. However, this type of data usually suffers from the curse of dimensionality. This causes most statistical methods and/or classifiers to not perform well. A recently proposed alternative which can help avoiding this problem is the Dissimilarity Representation, in which objects are represented by their dissimilarities to representative objects of each class. However, this approach depends on the selection of a suitable dissimilarity measure. For spectra, the incorporation of information on their shape, can be significant for a good discrimination. In this paper, we make a study on the benefit of using a measure which takes shape of spectra into account. We show that the shape-based measure not only leads to better classification results, but that a certain number of objects is enough to achieve it. The experiments are conducted on three one-dimensional data sets and a two-dimensional one.

Key words: Object representation, classification,small sample size, dissimilarity representation, spectral data

1 Introduction

Classification of unknown objects is one of the main problems in many research areas. Object representation plays an important role in this task. In practical classification problems, the number of training samples is usually very small, represented by a very large number of features which are not always the best to describe them. Many studies have been done on this issue; when only a certain number of objects is available, a peaking phenomenon occurs in the classification accuracy as the number of features is increased. This is known as the curse of dimensionality [1, 2, 3]. Hence, the ideal situation in order to obtain a good classifier would be to have at least as many samples as features. It appears to be difficult to achieve this in a number of real-world problems.

A type of data which has raised a growing interest in advanced approaches to its automatic analysis is the spectral data. It is due to the increasing possibilities

of the different research fields e.g. chemometrics and signal processing, to obtain it, and the usefulness of the spectral information to describe and differentiate samples of different classes. This is the type of application where data sets are small because the cost to obtain them is very high, and they are usually much smaller than the dimensionality of the space. The traditional way of representing spectra is by sampling, as a sequence of individual observations made on the objects. The higher the sampling resolution, the more accurate the spectrum is described, which implies a representation in a high-dimensional space. However, this way of representation is not good for traditional procedures. It makes them suffer the curse of dimensionality. Furthermore, discriminative knowledge about spectra e.g. the continuity between the measured values, shape, is not taken into account in the traditional high-dimensional feature-based representation. Thus, it does not help avoiding the problem.

Recent works have studied alternative object representations instead of features, demonstrating that the curse of dimensionality can be avoided [3]. A recently developed alternative in the field of pattern recognition is the Dissimilarity Representation (DR) [4]. It is based on the important role that pairwise dissimilarities between objects play. Classifiers may be built in the dissimilarity space generated by a representation set. 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. It is important to remark that, any traditional classifier that is defined on feature spaces can also be used in the dissimilarity space.

With the DR, the problem of building classifiers in high-dimensional spaces can be tackled, as the dimensionality will depend now on the size of the representation set (usually smaller or equal to the size of the training set). However, the main issue in this approach is the selection of a suitable measure for the problem at hand. The more discriminative information we take into account when designing the dissimilarity measure, the more compact the classes are. The centroid of the data should remain approximately the same and the average distance to this mean should decrease or be constant [3], requiring less samples for its description and a good classification accuracy.

Due to benefits that the DR has shown, it has been explored in several applications like the discrimination of spectral data [4, 5, 6, 7]. In this paper, we will make an exhaustive experimental study on the DR for spectral data. We will focus on the usefulness of taking the shape of the curve into account in the dissimilarity measure. It will be shown that this can help achieving good classification results in small sample size problems. Recently, the use of the DR was also extended to 2D spectral data i.e. objects represented by matrices, where two types of spectral features are described [8]. Thus, the study will be generalized to this type of data. We will use three one-dimensional spectral data set and a 2D spectral one. In the experiments we compare the classification accuracy in measures which do not take shape into account with a measure which does. This analysis is done for several training set sizes and representation set sizes, to see how the measures influence the results. Moreover, for the measure which takes

shape into account, we study the sensitivity of the results to the optimization of the parameters (Gaussian filter parameter). The paper is structured as follows. In Section 2, a brief introduction to the DR will be done. Also, the 1D and 2D measures to be used in the experiments are referenced. Following, the data sets and experiments will be described in Section 3. Finally, a discussion and the drawn conclusions will be presented in Section 4.

2 Introduction to Dissimilarity Representation Approach

The Dissimilarity Representation (DR) [4] was proposed as a more flexible representation of the objects 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. The DR is also based on the role that (dis)similarities play in a class composition. 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 $\mathbf{n}$ runs over the objects and $\mathbf{q}$ 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 $\mathbf{R}(r_1, ..., r_h)$. We build from this matrix a dissimilarity space. Objects are represented in this space by the column vectors of the dissimilarity matrix. Each dimension corresponds to the dissimilarities with one of the representation objects.

When an object is represented by a matrix $\mathbf{Y} \in \mathbb{R}^{m \times l}$, the theory of the DR is the same [8]. 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, graphs, as long as we have a proper dissimilarity measure. Hence, to obtain the dissimilarity space, a mapping $\phi(\cdot, R) : \mathbb{R}^{m \times l} \rightarrow \mathbb{R}^h$ is defined, such that for every object $\mathbf{Y}$, $\phi(\mathbf{Y}, R) = [d(\mathbf{Y}, r_1), d(\mathbf{Y}, r_2), ..., d(\mathbf{Y}, r_h)]$. Classifiers are then built in this space, as in any feature space.

The elements of $\mathbf{R}$ are called prototypes, and have preferably to be selected by a prototype selection method [4]. 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 $\mathbf{D}(\mathbf{X}, \mathbf{X})$. $\mathbf{R}$ 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 classifiers.

2.1 1D and 2D dissimilarity measures for spectral data

A general dissimilarity measure for all types of data does not exist. Thus, the selection of the suitable measure for the problem at hand is the key issue in

the DR approach. In recent studies, some 1D [6, 9] and 2D [8] measures have been studied and proposed for spectral data. Such is the case of the very well known Manhattan (L1-norm) and Euclidean distances. However, although the previous dissimilarities are of the most used measures in the comparisons of chemical spectral data, the connectivity between the measured variables and/or shape, is not taken into account in neither of them. The variables could be easily reordered and the same dissimilarity value is obtained.

In [6], the authors propose to compute the Manhattan measure on the first Gaussian derivatives (See Eq. 1) of the curves (Shape measure). Thereby, the shape information that can be obtained from the derivatives is taken into account:

$$d(x_1, x_2) = \sum_{j=1}^{m} |x_{1j}^{\sigma} - x_{2j}^{\sigma}|, \quad x^{\sigma} = \frac{\mathrm{d}}{\mathrm{d}_j} G(j, \sigma) * x \tag{1}$$

The expression of x^{σ} corresponds to the computation of the first Gaussian (that is what G stands for) derivatives of spectra. A smoothing (blurring) is done by a convolution process ($*$) with a gaussian filter and σ stands for the smoothing parameter. Good performances have been obtained for chemical spectral data with this measure [6, 9].

For the 2D representation of objects, generalizations of the Manhattan and Euclidean distances have also been proposed. Assume that two objects y_a and $y_b \in \mathbb{R}^{m \times l}$, where m and l are the number of variables in each of the two directions respectively; $\forall j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, l$. Then, the AMD measure [10] is defined as:

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

The power p is used to emphasize either small or large differences between the elements, depending on 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 [11] and Yang [12] distance measures. 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 the case. Recently, considering the results obtained with the Shape measure for simple spectra, a new version for 2D spectral data (2Dshape measure) was introduced [8]:

1. Compute the matrix D^1

$$D^1_{a,b} = \left(\sum_{k=1}^{l} \left(\sum_{j=1}^{m} (y^{\sigma}_{a,j,k} - y^{\sigma}_{b,j,k})^2 \right)^{p_1/2} \right)^{1/p_1} \quad , \quad y^{\sigma}_{i,j,\cdot} = \frac{\mathrm{d}}{\mathrm{d}_j} G(j,\sigma) * y_{i,j,\cdot}$$

2. Compute the matrix D^2

$$D^2_{a,b} = \left(\sum_{j=1}^{m} \left(\sum_{k=1}^{l} (y^{\sigma}_{a,j,k} - y^{\sigma}_{b,j,k})^2 \right)^{p_2/2} \right)^{1/p_2} \quad , \quad y^{\sigma}_{i,\cdot,k} = \frac{\mathrm{d}}{\mathrm{d}_k} G(k,\sigma) * y_{i,\cdot,k}$$

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 k-th columns and the j-th rows of the i-th 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, as in the 1D measure. 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. With this measure, the connectivity between the measured points can be taken into account as well as the shape of the spectra.

The previously mentioned measures, which have been used for spectral data, will be used for the purpose of this paper.

3 Experimental Section

For the purpose of this paper, a set of experiments were conducted on small sample size data sets in high-dimensional spaces. Only one of them does not suffer from this problem, but still we want to show how also in this case, with the selection of a suitable dissimilarity measure, a reduced number of training samples can be enough to obtain good classification results with the DR. All of them consist of two-class classification problems. The data sets are described in the following subsection.

3.1 Data sets

The first data set, named Tecator, originates from the food industry [13]. It consists of 215 near infrared absorbance spectra of meat samples, recorded on a Tecator Infratec Food and Feed Analyzer. Each observation consists in a 100 channel absorbance spectrum in the 850-1050 nm wavelength range. It is associated to a content description of meat sample, obtained by analytic chemistry. The classification problem consists in separating 77 meat samples with a high

fat content (more than 20%), from 138 samples with a low fat content (less than 20%). Original spectra are preprocessed, each spectrum is reduced to zero mean and unit variance.

The second data set is a real-world data set, which was obtained from a cooperation with the Oil Industry in Cuba. It consists of 31 fuel samples of Fourier Transform Infrared (FT-IR) transmittance spectra in a wavelength range of 600-4000 cm^{-1}. A base line correction and smoothing were performed on the data. The classification problem consists in determining the fuel type of the samples: regular gasoline (16 samples) and especial gasoline (15 samples).

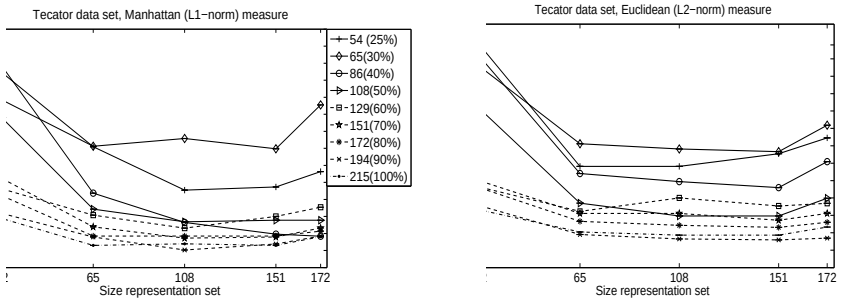
The third data set is another fuel real-world data set of 44 samples measured at 127 wavelengths in a range of 275-220 nm, but this time measures have been taken by a Ultra-Violet Visible (UV) spectrofotometer. The classification problem consists also in determining the fuel type of the samples: regular gasoline (23 samples) and especial gasoline (21 samples).

The fourth and last data set is a three-dimensional array, composed of objects naturally represented by 2D arrays. It is a public domain data set and the description has been taken from the website [14, 15] for a better understanding of this paper. It consists of samples of red wine belonging to different geographical areas and producers. They were 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 2 different geographical areas: South America (21 samples) and Australia (12 samples).

3.2 Experiments and Discussion

A set of experiments are conducted on the four data sets. A Fisher classifier is built on the dissimilarity space obtained for the two dissimilarity measures which do no take shape information into account (Manhattan and Euclidean) and also for the Shape measure. For the later, several experiments are shown, with different values for parameter σ. In the figures, learning curves are shown for various sizes of training and representation sets. The main idea of this experimental set up is to show how the use of a suitable measure e.g. measures shape in spectral data, can influence not only in the classifiers accuracy, but on the sample size problem.

Training and test objects were randomly chosen from the total data sets in a 10-fold cross-validation process, when the size of the training set allowed it. When the sizes of the training set was too small, a leave-one-out cross-validation was done. Experiments were repeated 10 times. For the training set of different sizes, a random selection of $[25, 30, 40, 50, 60, 70, 80$ and $90\%]$ is done from the total dissimilarity matrix. Different sizes for the representation set were also randomly selected $[10, 30, 50, 70$ and $80\%]$ of the total data set. When using the



(a) (b)

Fig. 1. Average cross-validation error (in %) for Tecator data set with (a) Manhattan and (b) Euclidean measures. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

Shape measure on the one-dimensional spectral data, the following values of σ were applied $[0.5, 1, 2, 3, 5, 7]$.

In the case of the Wine 2D spectral data, in the mass direction the classes only differ because they have different components. Therefore, the only thing we will see is the absence/presence of the peak or some differences in the concentration of the mass fragments. The other difference that we can find between these classes is related to the shape changes 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 several classes. Thus, for the Wine data set we will use the 2Dshape measure. The D1 matrix will be computed for the chromatography direction. The Gaussian derivatives are applied to take into account the shape in the changes of concentration in the neighboring components. In this case, the following values of σ were applied $[1, 2, 5$ and $8]$. However, for the D2 matrix from the mass spectra mode, we will only compute the overall sum of the differences between the concentration of the mass fragments. The use of derivatives is meaningless, because there is no continuity between the mass fragments.

In Figure 1 we can observe the same behavior for both measures: Manhattan(1(a)) and Euclidean (1(b)). Classifiers may perform better sometimes in one or the other. However, for both of them the classification error decreases as the training set and representation sets increase. When the training sets are too small, the errors are far higher that for larger training sets. On the other side, for larger training sets, the classification accuracy does not differ that much for different sizes (taking the standard deviation into account). There is even a point, where results are better or the same with 90% of the data, that with the full data set. Nevertheless, due to the so-called peaking phenomenon, when the number of prototypes starts reaching the size of the training set, the errors will increase.

Let us take a look at the results with the Shape measure (See Figure 2). Results have improved with respect to the dissimilarity measures which do not take the shape information into account. Of course, this is not for all values of σ.

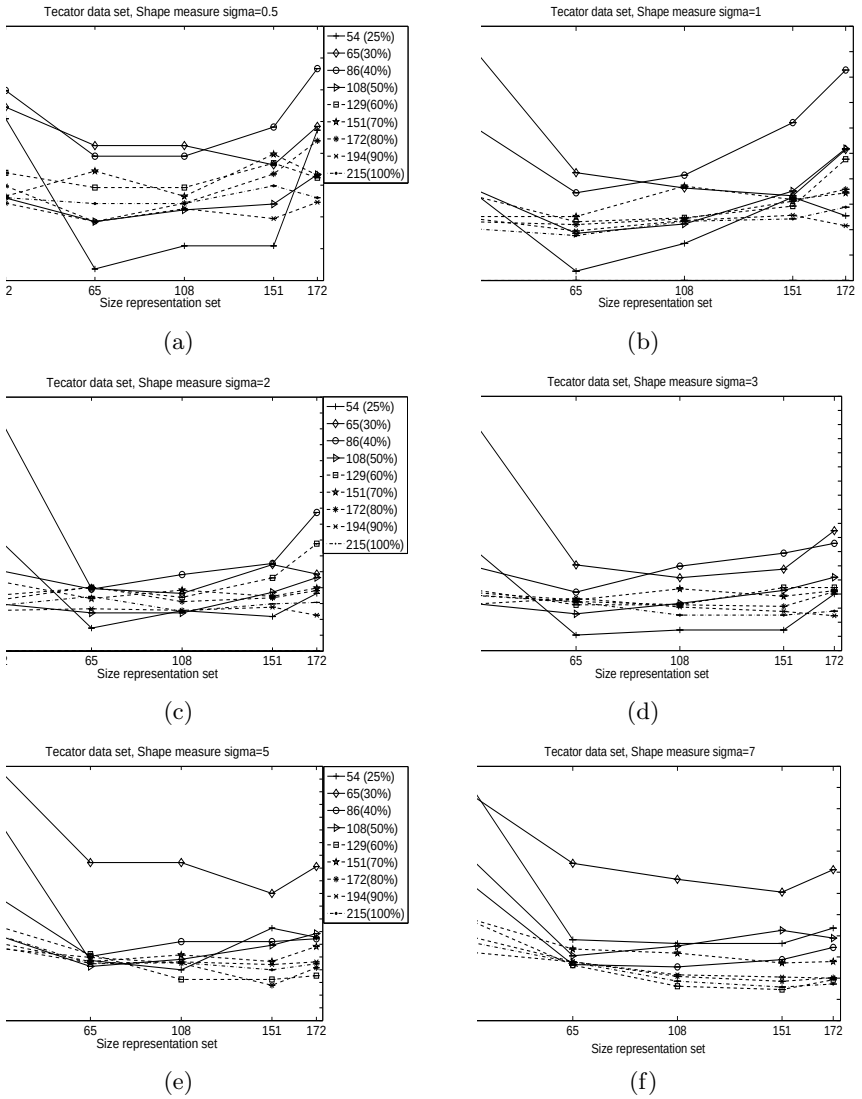
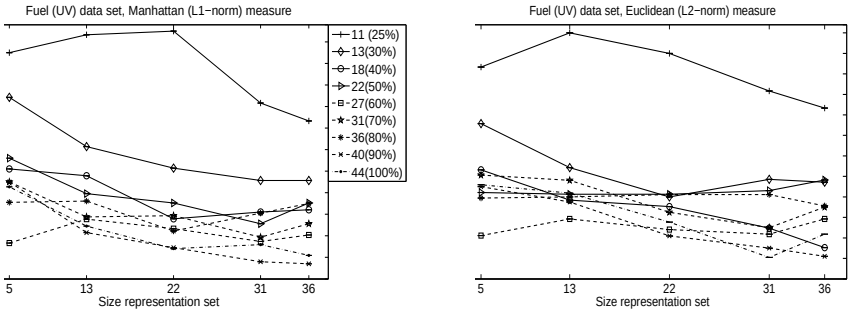


Fig. 2. Average cross-validation error (in %) for Tecator data set with Shape distance and different values of sigma for (a) sigma=0.5, (b) sigma=1, (c) sigma=2, (d) sigma=3, (e) sigma=5 and (f) sigma=7. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

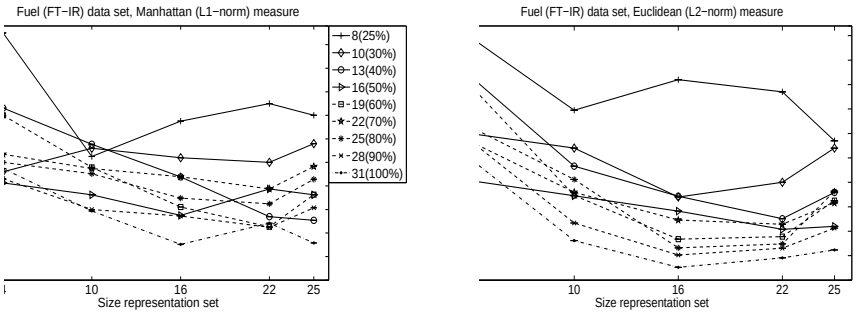
the optimization of the parameter does influence the results. From Figures 2(a) to 2(d), we can see that results are pretty much stable for all sizes of the training set. Although for $\sigma = 2$ and $\sigma = 3$, the classification errors start increasing. It seems that the parameter σ is better fixed to the data in the first two. Here, if

we take the standard deviation of the ten repetitions (around 0.5 the highest) into account, there is not much difference between using the smallest and the largest sample size. Thus, it seems that with this measure, a small training set is enough to reach even better results than with the other measures. In fact, in this case, the best results are always achieved with the smallest size of the training set. However, the results with 30% of the data are the highest, which could be influenced by the random selection of the samples. For the last two values of σ, the results keep increasing a lot, it can be due to the data is so smooth, that the measure starts failing. Thus, the importance of the optimization of the parameter.



(a) (b)

Fig. 3. Average cross-validation error (in %) for Fuel (UV) data set with (a) Manhattan and (b) Euclidean measures. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes



(a) (b)

Fig. 4. Average cross-validation error (in %) for Fuel (FT-IR) data set with (a) Manhattan and (b) Euclidean measures. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

With respect to the representation set, for all values of σ, the error always increases while the training set decreases, with the smallest representation set. It seems that the representation set is not representative enough. However, from that point on, the errors always start decreasing, until reaching the size of the training set, where they start increasing again due to the peaking phenomena.

With the two fuel data sets, we are facing a very complicated classification problem: discrimination of special and regular fuel, thereby the classification accuracy is not very good. Moreover, these are both affected by the small sample size problem. In Figures 3 and 4, we can observe the same phenomena as in the first figure. The errors decreasing while the size of the training set increases. In both cases, the results with the smallest sample size are far much higher; in this case the number of samples is really small.

However, if we take a look at Figure 5, we can see that with the Shape measure, for some values of σ parameter, the classifiers perform a lot better than for the Manhattan and Euclidean measures. With respect to the size of the training set, the behavior is a bit different. For the smaller sample sizes, the errors are very high, as in Figure 3. In this case, it can be explained by the fact that the original data suffers already from the small sample size problem. Thus, it would be too much to expect an improvement with so little samples. Nevertheless, from 60% of the data on, if we take the standard deviation into account, the results are very similar (for the σ with which the better results are obtained).

For the Fuel data set from FT-IR, we can see a bigger improvement by using the Shape measure than with the UV. This could be determined by the characteristics of the instrumental technique. It seems that the information obtained from the FT-IR spectra is more discriminative than that of UV-VIS. However, due to the small sample size problem, with the smallest training sets the results are still high. There is no sufficient data. Again, for the best σ values, the errors for the larger training sets, start behave very similar (taking standard deviation). It can also be noticed that for 40% of the whole data set, the best results are always achieved. It could be due to in the selection process, noisy data that could be affecting the results are no included.

The next data set is the three-way Wine data, where samples are represented by high-dimensional 2D matrices. In this case, we also compared the measures which take the shape information into account with does which do not. This is also a small sample size problem, in a very high-dimensional space. When analyzing the AMD measure with different values of p, which are the homologous for the Manhattan and Euclidean measures for one-dimensional data, the behavior is similar(See Figure 7). Although in this case the learning curve are a bit rough, we can see how the error decreases meanwhile the size of the training set increases.

In this case, although there is no shape information in both directions, it can be seen that the results also improve (See Figure 8) when taking this information into account (in the needed direction). If we take a look at the learning curve for all training set sizes, the best results are for $\sigma = 5$, so it seems to be the

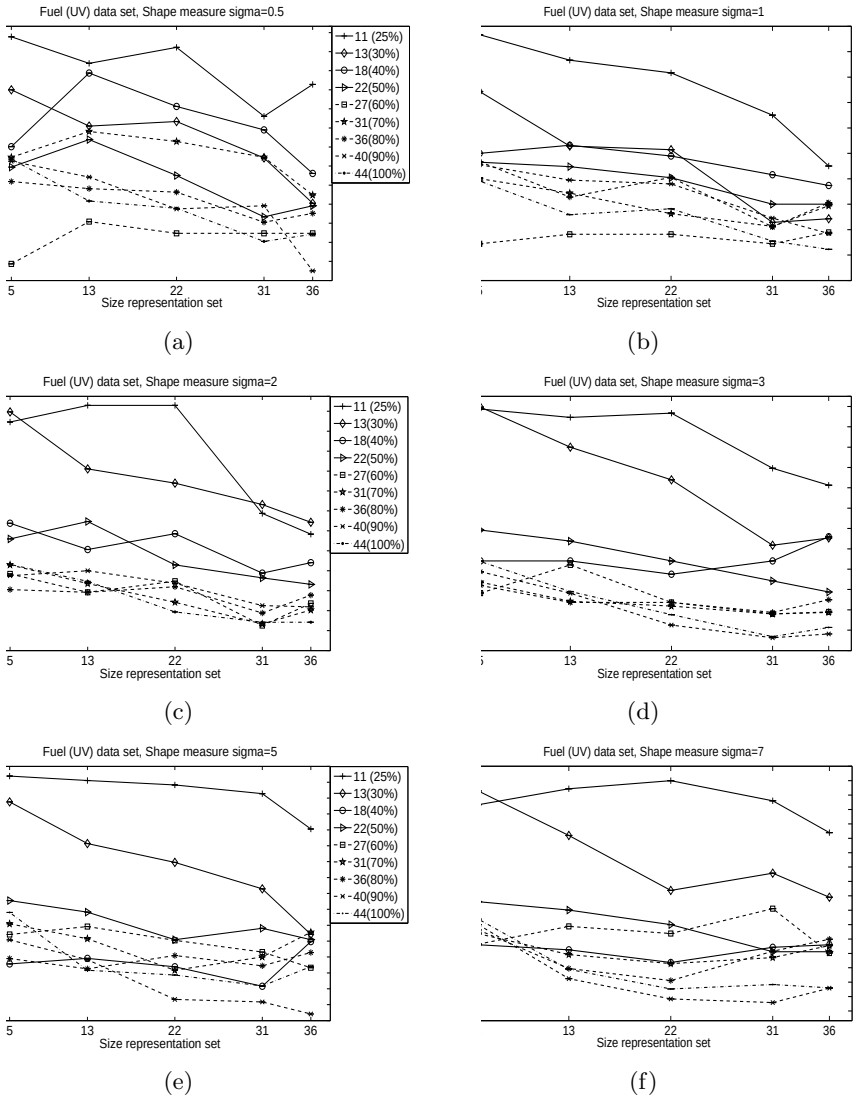


Fig. 5. Average cross-validation error (in %) for Fuel (UV) data set with Shape distance and different values of sigma for (a) sigma=0.5, (b) sigma=1, (c) sigma=2, (d) sigma=3, (e) sigma=5 and (f) sigma=7. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

best value for this parameter, in the range experimented. For this type of data, we can also observe, how by including certain discriminative knowledge in the measure i.e. shape, the results improve. With very small sample sizes the errors are still high (the available data is not enough to learn well). But, when the size

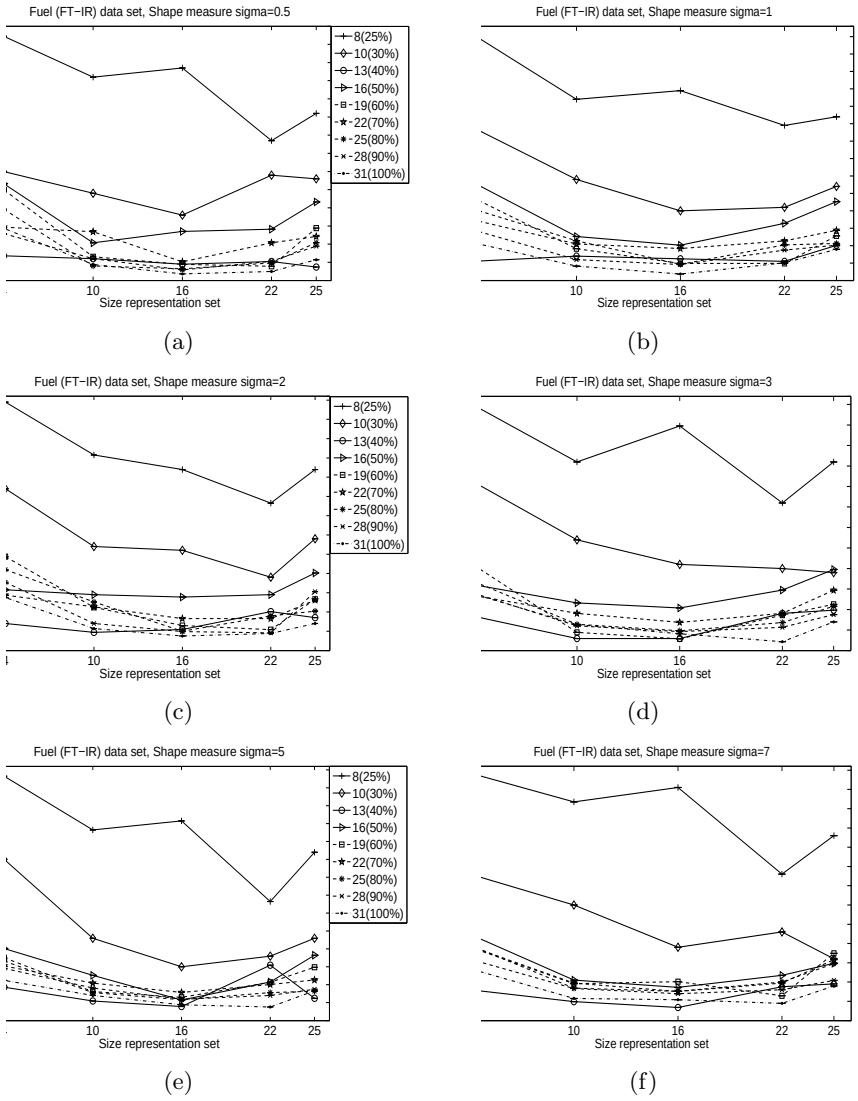


Fig. 6. Average cross-validation error (in %) for Fuel (FT-IR) data set with Shape distance and different values of sigma for (a) sigma=0.5, (b) sigma=1, (c) sigma=2, (d) sigma=3, (e) sigma=5 and (f) sigma=7. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

of the training sets start increasing, the errors are similar for most sizes (taking standard deviation into account). In which seems to be the best value for σ, the best results are achieved again, with only 50% of the total data set. Maybe, some noisy data which are influencing the results, are removed.

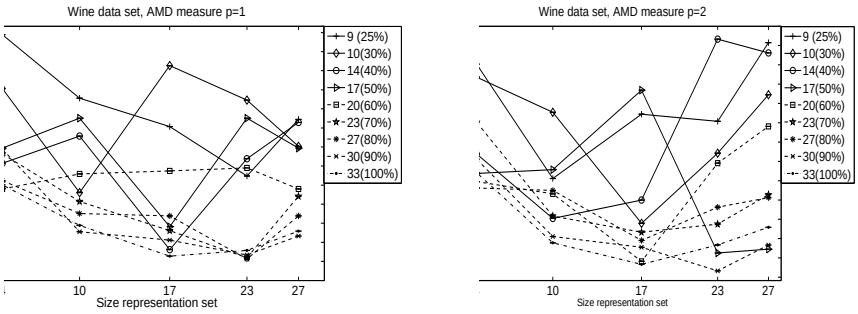


Fig. 7. Average cross-validation error (in %) for Wine three-way data set with (a) Yang (AMD p=1) and (b) Frobenius (AMD p=2) measures. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

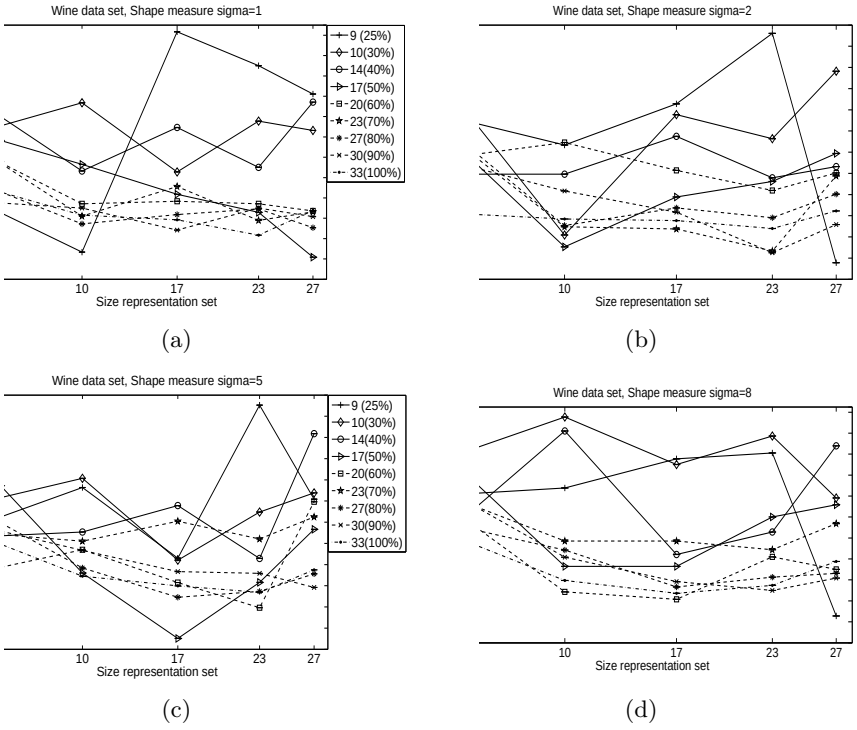


Fig. 8. Average cross-validation error (in %) for Wine three-way data set with Shape distance and different values of sigma for (a) sigma=1, (b) sigma=2, (c) sigma=5, (d) sigma=8, (e) sigma=10 and (f) sigma=15. The classifiers accuracy is analyzed for nine training set sizes and five representation set sizes

4 Discussion and Conclusions

The small sample size problem in high-dimensional spaces is very common in spectral data. Many statistical methods and classifiers fail with this type of data. Alternative representations for such data, to improve classification accuracy, have been explored. Such is the case of the Dissimilarity Representation. However, the key issue of this approach relies on the selection of a suitable dissimilarity measure for the problem at hand. In the case of spectral data, a discriminative feature is the knowledge about the connection between the neighboring points and shape.

In our experimental study, we showed the importance of taking the shape of the curve into account for the success of the DR. Even when we are facing small sample size problems, if we use the shape information, a few samples are enough for classifiers to learn better. For all data sets, there is some size for the training set (usually smaller than the original data size), from which adding new objects will not make much of a difference. This was also experimented, in a not so small data set, and we reached the same conclusions. In this case, we also benefit from lowering the computational complexity of the classifier. This behavior is not the same for measures which do not take discriminative information into account i.e. Manhattan, Euclidean or AMD for 2D data. In this case, the errors are smaller the larger the training set, so we are not solving a small sample size problem, as we would need more samples for the classifiers to learn better.

From the experiments with the measures which take shape into account i.e. Shape and 2Dshape, we can also observe the influence of the optimization of the Gaussian filter parameter. There is always a value of σ for which the classification results are better than without measuring shape. It also stabilizes the learning curves of the different sizes of training set, which are around the same performance.

The representation set is also very important. In all experiments we can observe that even with the large data set (Tecator) the error always increases while the training set decreases, with the smallest representation set. It seems that the representation set is not representative enough. However, from that point on, the errors always start decreasing, until reaching the size of the training set, where they start increasing again due to the peaking phenomena. However, these experiments are all based on two-class classification problems; for multi-class problems, further studies should be done.

In conclusion, the incorporation of shape information in the dissimilarity representation is important for the discrimination of spectral data. It helps avoiding the curse of dimensionality problem, allowing classifiers to perform well in small sample size situations.

5 Acknowledgment

We acknowledge financial support from the FET programme within the EU FP7, under the project "Similarity-based Pattern Analysis and Recognition -

SIMBAD" (contract 213250). We would also like to thank to the project Cálculo científico para caracterización e identificación en problemas dinámicos (code Hermes 10722) granted by Universidad Nacional de Colombia.

References

[1] Fukunaga, K., Hayes, R.: Effects of sample size in classifier design. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(8) (1991) 873–885
[2] Raudys, S.J., Jain, A.K.: Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Transactions on Pattern Analysis and Machine Intelligence **3**(3) (1991) 252–264
[3] Classifiers in almost empty spaces. In: 15th International Conference on Pattern Recognition. Volume 2., Barcelona , Spain, IEEE Computer Society (2000)
[4] Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation For Pattern Recognition. Foundations and Applications. World Scientific. (2005)
[5] Orozco-Alzate, M., Garca, M.E., Duin, R.P.W., Castellanos, C.G.: Dissimilarity-based classification of seismic signals at Nevado del Ruiz Volcano. Earth Sci. Res. J. **10**(2) (2006) 57–65
[6] Paclik, P., Duin, R.P.W.: Dissimilarity-based classification of spectra: computational issues. Real Time Imaging **9**(4) (2003) 237–244
[7] Porro-Muñoz, D., Talavera, I., Duin, R.P.W., Hernández, N., Orozco-Alzate, M.: Dissimilarity representation on functional spectral data for classification. Journal of Chemometrics **Early View** (2011)
[8] Porro-Muñoz, D., Duin, R.P.W., Orozco-alzate, M., Talavera, I., Londoño Bonilla, J.M.: The dissimilarity representation as a tool for three-way data classification: a 2D measure. In: S+SSPR2010. LNCS, Springer (August 2010)
[9] Porro-Muñoz, D., Talavera, I., Duin, R.P.W., Hernández, N.: The representation of chemical spectral data for classification. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Proceedings of the 14th Iberoamerican Congress on Pattern Recognition CIARP 2009. Volume 5856 of LNCS., Springer (November 2009) 513–520
[10] Zuo, W., Zhang, D., Wang, K.: An assembled matrix distance metric for 2DPCA-based image recognition. Pattern Recognition Letters **27** (2006) 210–216
[11] Yang, J., Yang, J.Y.: From image vector to matrix: A straightforward image projection technique-IMPCA vs. PCA. Pattern Recognition **35** (2002) 1997–1999
[12] Yang, J., Zhang, D., Frangi, A., Yang, J.Y.: wo-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Machine Intell **26**(1) (2004) 131–137
[13] Thodberg, H.H.: Tecator dataset. Available at: http:lib.stat.cmu.edu/datasets/tecator (1995) Danish Meat Research Institute.
[14] Skov, T.: Wine dataset. Available at: http:www.models.kvl.dk/datasets.html (2008)
[15] Skov, T., Ballabio, D., Bro, R.: Multiblock variance partitioning. A new approach for comparing variation in multiple data blocks. Analytica Chimica Acta **615**(1) (2008) 18–29