Classifying Spectral Data using Relational Representation

P.Paclík, R.P.W. Duin
Pattern Recognition Group, Faculty of Applied Science,
Delft University of Technology, The Netherlands

Abstract

Appropriate representation of spectral data is crucial for successful classification. Traditionally, spectra are represented in feature spaces, spanned by individual spectral bands. This approach neglects the existing connectivity in the spectra. In the paper, we study an alternative representation where spectra are represented by their relations to other spectra. An advantage of this relative representation is that spectrum may be treated as a connected entity. A dissimilarity measure may be designed to reflect more naturally the important clues such as shape. Using a real dataset, we study the performance of several classifiers built directly on dissimilarities and in dissimilarity spaces.

keywords: relative representation, dissimilarity, spectral data, classification

1 Introduction

Traditionally, spectra are represented in a feature space, spanned by the spectral bands (see Fig.1). This approach brings a number of difficulties. The high resolution of spectral data translates into excessive feature space dimensionality. As a consequence, unrealistic amount of labeled training data is required to design classifiers. This problem is tackled by different dimensionality reduction techniques [3].

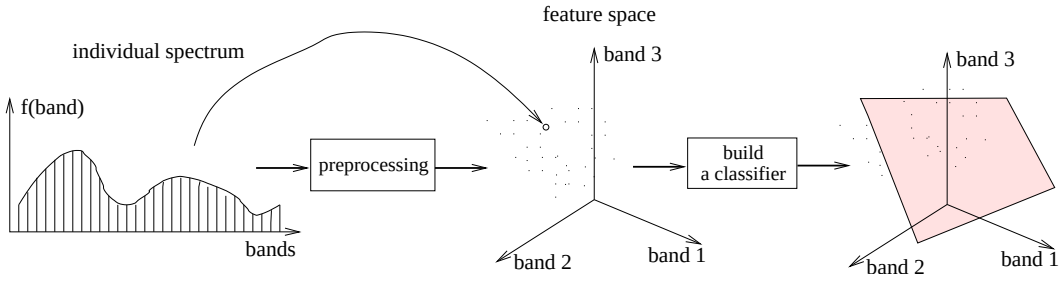

Figure 1: Feature-based absolute representation of spectral data.

We propose to represent spectra by dissimilarities to other spectra. These dissimilarities are expressed by the measure which may emphasize clues, important for particular type of spectra and application (e.g. shape, amplitude changes, etc.) The dissimilarity measure works on raw spectral data and may therefore take advantage of a number of existing signal processing procedures. The basic question we ask when designing a dissimilarity measure is: what makes two spectra similar/dissimilar. Is it their alignment or agreement in signal amplitudes? Or is it the shape what matters? How much is a local change of spectral shape affecting our judgment? Is it possible, that the material is manifested by a unique configuration of peaks which may shift slightly? We think that this approach for designing a pattern recognition system is closer to the actual expert practice than the traditional feature-based procedure. When raw spectral bands are taken as features, the above mentioned clues regarding connectivity, shape or peak positions must be re-invented in the high-dimensional feature space in terms of clusters or manifolds. Classification of spectral data, based on relative representation, is depicted in the Figure 3. An important difference between both approaches is that the feature-based method represents each spectrum in an absolute way while the proposed method relates the observations to each other.

The spectral data are represented by a matrix of dissimilarities. Classification may be performed in a several different ways. Traditionally, a nearest neighbor classifier is used on dissimilarity data [7]. We have shown in [4] that other classification techniques with better generalization capabilities and often lower computational requirements may be used. The dissimilarities may be either isometrically embedded in a feature space or classified in a dissimilarity space, generated by the representation set. In this study, we use the later approach to classification of relational data – the dissimilarity space.

In the following section, three dissimilarity measures between spectral data are presented. Section 3 describes experiments conducted on the real spectral dataset. Finally, we add some concluding remarks.

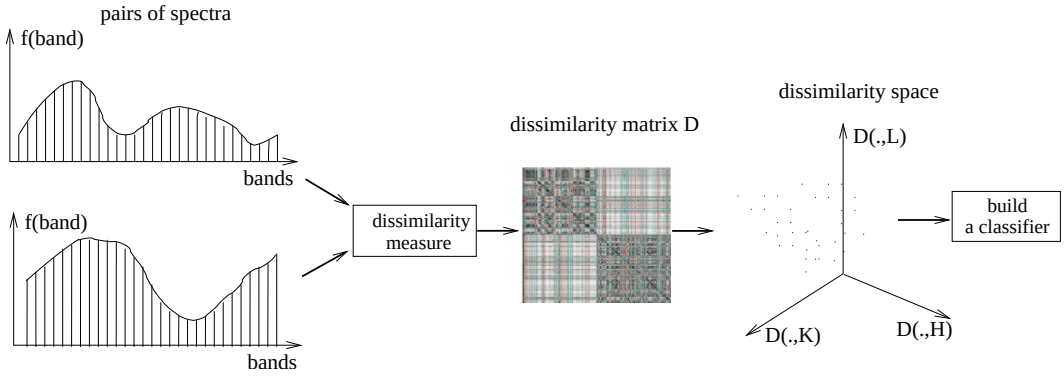

Figure 2: Relative representation of spectral data using dissimilarities.

2 Dissimilarity measures for spectral data

In this section, we introduce three dissimilarity measures between spectra H and K, each with N spectral bands, $\omega = 1, ..., N$. The *Minkowski distance* is defined as:

$$D_M(H,K) = \left(\sum_{\omega} |h_\omega - k_\omega|^r \right)^{1/r}. \tag{1}$$

It is a band-to-band measure neglecting the connectivity in the spectral domain [5]. Minkowski distance between two spectra yields the same value even if all the bands are randomly permuted.

Spectra, normalized to a unit area, may be compared using dissimilarity measures, developed for probability distributions. An example is a *Kolmogorov distance*:

$$D_K(H,K) = \max_{\omega}(|\hat{h}_\omega - \hat{k}_\omega|), \tag{2}$$

where $\hat{h}$ and $\hat{k}$ are cumulative distribution functions, $\hat{h}_i = \sum_{j \le i} h_j$, similarly for $\hat{k}$. This measure compares areas under original spectra and therefore reflects the distribution shape [5].

We propose to compute a dissimilarity between spectra by integrating the differences between shapes of both curves. Shape information is obtained by computing the first Gaussian derivative:

$$H^\sigma = \frac{d}{d\omega} G(\omega, \sigma) * H, \tag{3}$$

where $*$ denotes convolution and σ stands for a smoothing parameter. The proposed *shape dissimilarity measure* D_s is then a sum of absolute differences between spectra derivatives H^σ and K^σ.

$$D_s(H,K) = \sum_{\omega} (|h^\sigma_\omega - k^\sigma_\omega|). \tag{4}$$

3 Experimental results

In order to illustrate practical application of the relational representation, we use a set of autofluorescence spectra acquired in oral cavity [6]. The dataset contains 111 spectra from two classes: healthy and diseased mucosa with 78 and 33 samples, respectively. Each spectrum consists of 199 bins (spectral wavelengths). Details about data acquisition and preprocessing may be found in [1].

In all experiments, we used the leave-one-out procedure to estimate the error rates. For each type of distance (Minkowski, Kolmogorov and shape), the following setup was used: The full distance matrix between all data samples was computed. Directly on this set of dissimilarities, an error of nearest neighbor classifier err_{KNN} was estimated. Then, we used the dissimilarity matrix as a new feature space (also called dissimilarity space [2]) and estimated performance of several classifiers built in such a space. An example scatter plot of a dissimilarity space is shown in Figure 4.

Dimensionality of a dissimilarity space may be reduced in a number of ways. We have used the simplest procedure - the random selection. For each of two classes, 5 samples were chosen. Distances to these 10 samples generate a 10 dimensional dissimilarity space.

Following classifiers were used:

- nearest neighbor classifier with k=1 (KNN)
- Fisher linear discriminant
- Linear discriminant assuming normal densities (LD)

The experimental results are given in the Table 1. It can be seen that the nearest neighbor error err_{NN}, directly computed on Minkowski and Kolmogorov distances, is lower then for the proposed shape dissimilarity D_s. We can also observe, that the performance on D_s may be further considerably improved by building classifiers in dissimilarity spaces. No such improvement is, however, observable for the Minkowski and Kolmogorov distances. This result is conformable with our previous findings [4].

	D_M $err_{NN} = 0.144$		D_K $err_{NN} = 0.108$		D_s ($\sigma = 1.0$) $err_{NN} = 0.171$	
	all	$d_R = 10$	all	$d_R = 10$	all	$d_R = 10$
nearest neighbor $k = 1$	0.198	0.234	0.189	0.216	0.216	0.225
Fisher discriminant	0.144	0.198	0.459	0.153	0.288	0.117
Linear discriminant	0.297	0.198	0.297	0.153	**0.099**	0.117

Table 1: Classification results (Leave-one-out) using three dissimilarity measures and three classifiers. For each measure, directly computed nearest neighbor error is denoted err_{NN}. All classifiers were computed in a dissimilarity space with either all 110 dimensions or randomly selected 10 dimensions.

The proposed shape dissimilarity D_s requires one parameter to be set – the smoothing σ of a Gaussian derivative filter. In the following experiment, we investigated the effect of different smoothing on the classification performance. For each value of σ a lave-one-out procedure was run to estimate the classification error in the full dissimilarity space with 110 dimensions. The results are presented in the Figure 3. Surprisingly, the performance of the Fisher linear discriminant still improves reaching the lowest overall error 0.072 for $\sigma = 2.0$. The performance of both remaining classifiers (nearest neighbor and linear discriminant) is also improved with regards to the default unit smoothing.

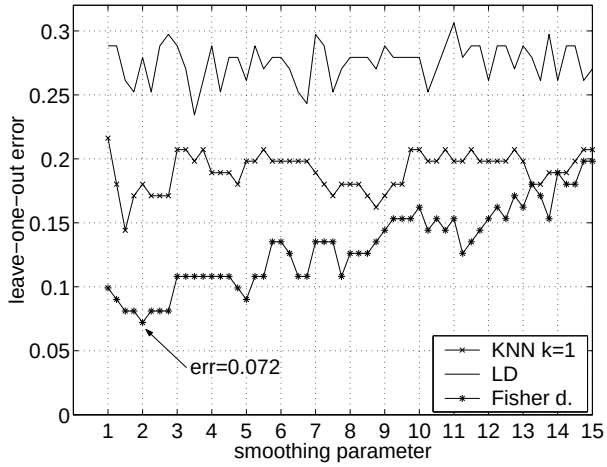

Figure 3: Leave-one-out error for different settings of smoothing parameter σ.

4 Conclusions

In this short study, we shown, how the spectral data may be represented by their relations to other spectra. This is an alternative to a commonly used absolute representation, where individual spectra are represented as points in high-dimensional feature-spaces. An advantage of the relative representation is that spectrum may be

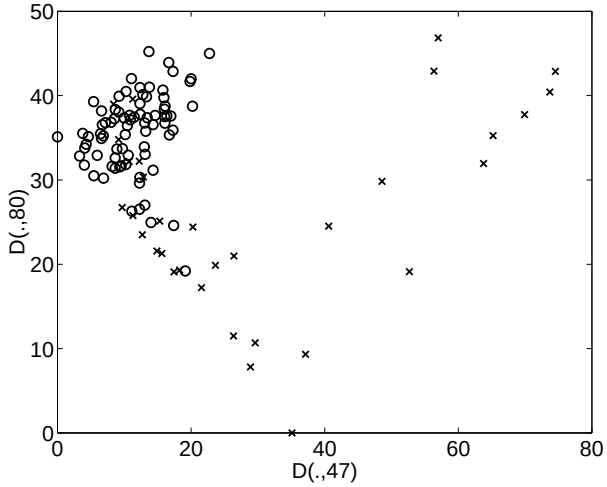

Figure 4: A scatter plot of a dissimilarity space. Each axis measures the dissimilarity to a data sample from a representation set (here samples number 47 and 80). These points are also visible on both axes, because their self-dissimilarity is zero.

treated as a connected entity. We think, that some characteristics, such as the shape, could be refined more easily when comparing two spectra than estimated from scattered points in a feature space.

We have shown, that the poor performance of a nearest neighbor classifier, computed directly on the dissimilarity data, may be improved by generalizing classifiers built in the dissimilarity spaces. Another interesting point is designing parametrized dissimilarity measures which could be tuned to a given type of data. Design of dissimilarity measures utilizing the connectivity in the spectral domain is a subject for future research.

Acknowledgments

Authors would like to thank here Marina Skurichina for providing a challenging dataset, used in this study.

References

[1] D. C. G. de Veld, M. Skurichina, and M. J. H. e. Witjes. Autofluorescence characteristics of healthy oral mucosa at different anatomical sites. submitted to *Lasers in Surgery and Medicine*, 2003.

[2] R. P. W. Duin, E. Pekalska, and D. de Ridder. Relational Discriminant Analysis. *Pattern Recognition Letters*, 20(11-13):1175–1181, 1999.

[3] D. Landgrebe. Hyperspectral image data analysis as a high dimensional signal processing problem. *Special Issue of the IEEE Signal Processing Magazine*, 19(1):17–28, January 2002.

[4] E. Pekalska, P. Paclík, and R. P. W. Duin. A generalized kernel approach to dissimilarity based classification. *Journal of Machine Learning Research*, (2):175–211, 2001. Special Issue "New Perspectives on Kernel Based Learning Methods".

[5] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[6] M. Skurichina, A. I. Belousov, D. C. G. de Veld, R. P. W. Duin, M. J. H. Witjes, H. J. C. M. Sterenborg, and J. L. N. Roodenburg. Combining different normalizations in lesion diagnostics. Technical report, Pattern Recognition Group, TUDelft, January 2003.

[7] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 1998. ISBN 0-12-686140-4.