Available online at www.sciencedirect.com
SCIENCE DIRECT®
Real-Time Imaging 9 (2003) 237–244
www.elsevier.com/locate/rti

Dissimilarity-based classification of spectra: computational issues

Pavel Paclík*, Robert P.W. Duin

Pattern Recognition Group, TU Delft, Lorentzweg 1, Delft 2628 CJ, The Netherlands

Abstract

For the sake of classification, spectra are traditionally represented by points in a high-dimensional feature space, spanned by spectral bands. An alternative approach is to represent spectra by dissimilarities to other spectra. This relational representation enables one to treat spectra as connected entities and to emphasize characteristics such as shape, which are difficult to handle in the traditional approach. Several classification methods for relational representations were developed and found to outperform the nearest-neighbor rule. Existing studies focus only on the performance measured by the classification error. However, for real-time spectral imaging applications, classification speed is of crucial importance. Therefore, in this paper, we focus on the computational aspects of the on-line classification of spectra. We show, that classifiers built in dissimilarity spaces may also be applied significantly faster than the nearest-neighbor rule.

1. Introduction

For the sake of classification, spectra are usually represented by points in a high-dimensional space, spanned by spectral bands, see Fig. 1. The disadvantage of this approach is that the spectral connectivity is lost. It is difficult to handle characteristics, such as spectral shape or peak positions in such a feature space. In cases where such indications are important for the identification of target classes, large labeled data sets and statistical learning procedures are necessary to recover these, originally known, clues.

Alternatively, spectra may be represented by their dissimilarities to other spectra [1]. Pairs of spectra are compared by a dissimilarity measure reflecting their mutual resemblance. For a given training set, a square matrix of pair-wise dissimilarities is produced forming the *dissimilarity representation* of original spectra [2,3]. The advantage is that the dissimilarity measure may treat the spectrum as a connected entity and may emphasize its shape or other, application specific clues. In this paper, we show how to classify spectral data using this relative representation putting the emphasis on the classification speed.

Three different classification strategies may be followed, starting from dissimilarities. The first method, traditionally applied to dissimilarity data, is based on the nearest-neighbor paradigm. According to the compactness hypothesis [4,5], similar objects should also be close in their representation. Previously unseen examples are, therefore, assigned to the classes of their most similar neighbors in the training set. Although the nearest neighbor rule is usually applied in Euclidean feature spaces, it may be successfully used also for other, more specialized, types of dissimilarities computed from spectral data.

The two remaining approaches convert the set of dissimilarities into a feature space, where traditional statistical classifiers may be built. The first one is based on *isometrical embedding* of dissimilarity data into the Euclidean space preserving the distances between objects as good as possible. The second method views the distance matrix directly as a new training data set. The rows represent individual training examples and the columns form the dimensions of a new, so-called *dissimilarity space* [6]. Each dimension of this space measures the dissimilarity to a particular training prototype. In this way, the distances to prototypes may be treated as features. The set of prototypes is called the *representation set*.

An important difference exists between embeddeding and the use of a dissimilarity space concerning the implementation aspects. These methods are, in fact, equivalent to feature extraction and feature selection

*Corresponding author. Pattern Recognition Group, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, Netherlands. Tel.: +31-15-278-3191; fax: +31-15-278-6740.
E-mail address: pavel@ph.tn.tudelft.nl (P. Paclík).
URL: http://www.ph.tn.tudelft.nl/~pavel/.

1077-2014/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rti.2003.09.002

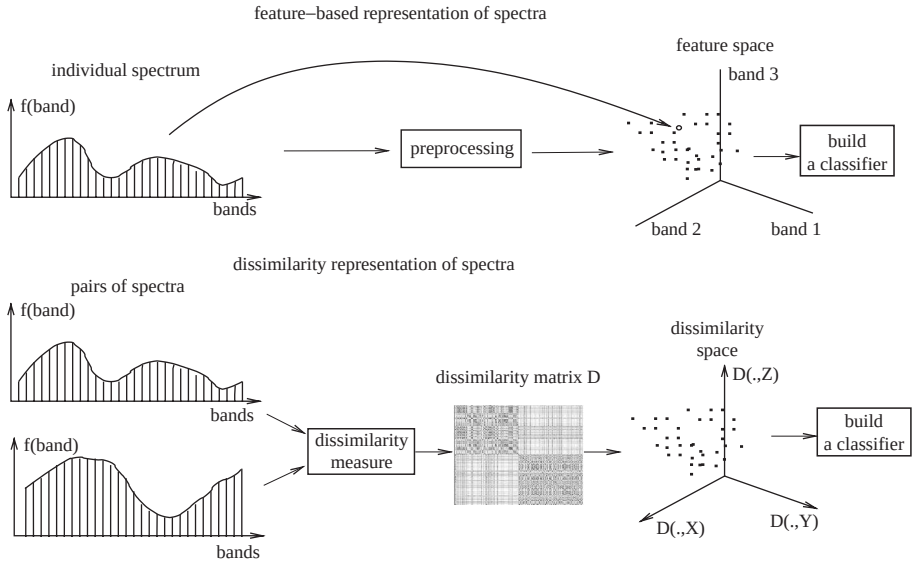

Fig. 1. Feature-based and dissimilarity representations of spectral data.

applied to a general dataset. In case of embedding, dissimilarities to all training objects must be computed to constitute a space where the classification can be done. The same transformation is also required for all new spectra to be classified which makes the embedding approach too complex for real-time applications. On the other hand, methods based on the dissimilarity space consider dissimilarities to training examples (i.e. columns of a distance matrix) as individual dimensions of a feature space. Therefore, the number of dissimilarities, computed for each new data sample, may be significantly reduced by existing feature selection techniques.

Although dissimilarity-based classification methods were examined and compared to other approaches before from the performance point of view, little attention has been paid to the classification speed of these algorithms. Due to the sheer volumes of data processed in spectral imaging, high-speed classification is a basic pre-requisite for application of dissimilarity-based methods.

In this paper, we show how to build classifiers for hyperspectral data using dissimilarity-based representation with an emphasis on fast implementation. We focus on the speed of classification (i.e. labeling new spectra by a trained classification system). We use several dissimilarity measures emphasizing different notions of spectral similarity and having different computational requirements.

In the first section, we discuss the classification system built using the dissimilarity representation. Later, we present four dissimilarity measures for spectral data and discuss their computational complexity. In Section 4, several experiments on real hyper-spectral data illustrate the use of dissimilarity-based classifiers. Finally, we give some conclusions in the last section.

2. Classification of spectra using dissimilarity representation

In this section, we explain how dissimilarity-based classification methods may be applied to spectral data and discuss the computational complexity of the classification of new data. We focus on the nearest neighbor rule (NN), directly applied to the distance matrix and on linear classifiers built in the dissimilarity space. Computational complexity is often expressed in terms of the order of an algorithm [7]. This measure is, however, not ideal for judging the actual computational requirements of classification algorithms [8]. Therefore, we also consider the number of floating point operations (FLOPS), performed for each new spectrum to be labeled [9].

In order to build a relational representation for a training set with N labeled spectra, the $N \times N$ distance matrix is formed from all pair-wise dissimilarities (see Fig. 1). Several types of dissimilarity measures, applicable to spectral data, are further discussed in Section 3.

In order to classify new spectral measurements by the nearest-neighbor rule, the dissimilarities to all the training examples must be evaluated. To label each of M objects, N dissimilarities are computed and ranked to find the closest prototype requiring:

$$t_{NN} = MNt_D + MN \tag{1}$$

operations. The symbol t_D stands for the number of operations needed to calculate a dissimilarity between two spectra. The computational complexity of the nearest-neighbor rule is therefore

$$\mathcal{O}(MN\mathcal{D}), \tag{2}$$

where $\mathscr{D}$ denotes the complexity of a single dissimilarity evaluation. The performance of the nearest-neighbor rule increases with growing number of training examples, being asymptotically bounded by twice the Bayes error for an infinite sample size [7]. To reach low error rates, the number of training examples N must be usually quite high, especially in case of overlapping classes.

Fig. 2 illustrates the different treatments of dissimilarity data by the nearest-neighbor rule and by classifiers built in the dissimilarity space. The nearest-neighbor classifier ranks the values in the row of a dissimilarity matrix searching for the closest training example. For a new object, we compute all distances to the training samples and assign the object into the class of the nearest training example. Note, that the distances between the training samples themselves are actually not used in the process. A different approach is taken in the middle image of Fig. 2 emphasising two randomly chosen columns in the distance matrix (i.e. training examples). We may observe that dissimilarities of the training objects to these two prototypes convey a typical class pattern. By using the distance to each of the two prototypes as a separate axis, we form a dissimilarity space (see the scatter plot in the rightmost image). Classifiers trained in a dissimilarity space may be based on all training objects and thereby utilise global rather than local information which often leads to a better performance [6].

In general, any classifier may be trained in a dissimilarity space. Dissimilarities, computed by summing many small differences, tend to produce normally distributed data as a consequence of the law of large numbers. Linear classifiers, applied to dissimilarity data, often lead to a very good performance [3]. In the following, we focus on linear classifiers such as the Fisher linear discriminant (FLD).

Let us assume, that a D-dimensional dissimilarity space is based on computing distances to D training prototypes, $D \leqslant N$. If a linear classifier, trained to distinguish C classes, is applied to M new samples, the data is projected from the original D-dimensional space to C dimensions, e.g. the confidences of each of the C classes. The outcome is a $M \times C$ matrix which may be ranked to find the most probable class for each data sample. To perform this classification, $C(2D+1)$ operations are needed. The total number of operations required to construct the dissimilarity representation and to classify M spectra by a linear discriminant is

$$t_{FLD} = MDt_D + 2CMD + CM. \tag{3}$$

In order to judge the classification complexity of new spectra using a linear classifier, built in a dissimilarity space, we need to know the complexity of dissimilarity computation. For the dissimilarity measures, used in this study, the number of operations needed to compare two spectra depends linearly on the number of bands K, as shown in Section 3. Because both the number of spectral bands K and the number of training prototypes D usually exceed the number of classes C, the first term in Eq. (3) will dominate and the complexity becomes:

$$\mathscr{O}(MD\mathscr{D}). \tag{4}$$

Comparing Eqs. (2) and (4), we can see that the classification complexity of both methods grows linearly with the number of evaluated dissimilarities (i.e. the number of prototypes). In case of the nearest-neighbor rule, all training examples are used as prototypes. The time, spent on classifying new spectrum, grows with each employed training example. For the classifiers built in a dissimilarity space, on the other hand, the distances to the selected prototypes define separate space dimensions. The classification time, therefore, scales with the number of these prototypes. Usually, only a subset of a training set is included in a representation set. The rest of the training examples is, however, exploited by the classifier and improves its performance.

Objects in the representation set may be selected either in a systematic or in a random way. It was previously shown that the random selection of

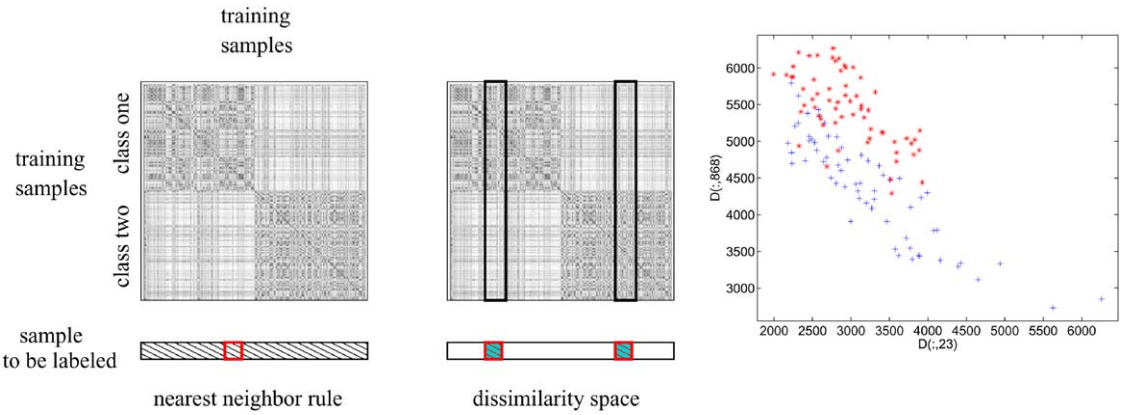

Fig. 2. Direct ranking of the dissimilarities and the formation of the dissimilarity space.

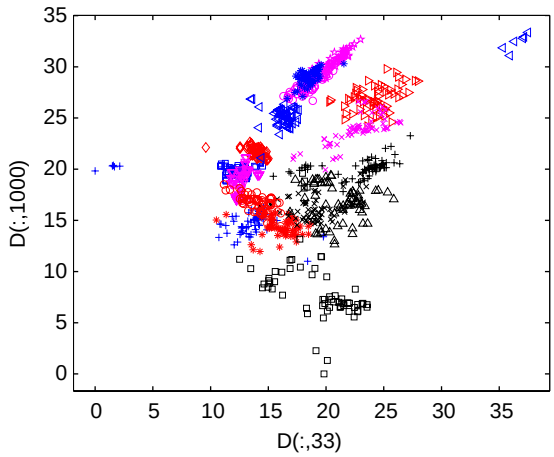

Fig. 3. An example scatter plot of a dissimilarity space with randomly selected prototypes (i.e. dimensions) and 16 classes.

prototypes yields classification results comparable to the systematic selection [10]. This, at first counter-intuitive result, becomes more clear when we investigate the example scatter plot in Fig. 3 showing a two-dimensional scatter plot of a dissimilarity space, computed from a labeled set of spectra with 16 classes. Each axis measures the derivative-based distance to a spectrum from the representation set (see Section 3). Although only two representation objects are used, we can visually separate a number of classes. In our experiments, we use the random prototype selection.

3. Dissimilarity measures for spectral data

The spectra X and Y are given by K samples x_i, and y_i, $i = 1, \ldots, K$, respectively. The Euclidean distance between spectra X and Y is defined as

$$D_E(X, Y) = \sqrt{\sum_{i=1}^{K} (x_i - y_i)^2}. \tag{5}$$

This measure views the spectra as K-dimensional feature vectors, neglects the connectivity of the spectra and merely sums the band-to-band differences. Euclidean dissimilarity between two spectra yields the same value if all the bands are randomly permuted.

Another dissimilarity measure, developed specifically for spectral data, is the Spectral Angle Mapper (SAM) [11].

$$D_{SAM}(X, Y) = \operatorname{acos}\left(\frac{\sum_{i=1}^{K} x_i y_i}{\sum_{i=1}^{K} x_i^2 \sum_{i=1}^{K} y_i^2} \right). \tag{6}$$

The SAM dissimilarity measures the angle between two vectors and is, therefore, insensitive to scaling. In order to speed-up computation of the SAM measure for new data, the norm of prototypes $1/\sum_{i=1}^{K} y_i^2$ may be pre-computed and stored.

Spectra, normalized to a unit area, may be compared using dissimilarity measures, developed for probability distributions. An example is the Kolmogorov distance which measures maximal difference between cumulative histograms:

$$D_K(X, Y) = \max_i \left(|\hat{x}_i - \hat{y}_i| \right), \tag{7}$$

where $\hat{x}$ and $\hat{y}$ are cumulative distribution functions, $\hat{x}_i = \sum_{j \leqslant i} x_j$, similarly for $\hat{y}$. This measure compares areas under original distributions (i.e. spectra) and therefore reflects the spectral shape. Instead of taking the maximum, the match dissimilarity D_M sums all the differences under the cumulative histograms [12]:

$$D_M(X, Y) = \sum_{i}^{K} |\hat{x}_i - \hat{y}_i|. \tag{8}$$

The match dissimilarity is a special case of the Earth Mover's Distance [13] which measures the effort needed to transform one distribution into another. In case of the Kolmogorov and the match distances, the speed of computation may increased by pre-computing and storing cumulative spectra for the training examples $\hat{y}$.

In order to reflect the shapes of spectra, the differences between first Gaussian derivatives of spectra may be integrated:

$$X^{\sigma} = \frac{\mathrm{d}}{\mathrm{d}i} G(i, \sigma) * X, \tag{9}$$

where $*$ denotes convolution and σ stands for a smoothing parameter. Please note, that although i is treated as a continuous variable in the last equation it is actually a discrete, densely sampled, quantity. The shape dissimilarity D_s is then a sum of absolute differences between derivatives X^{σ} and Y^{σ} [2].

$$D_s(X, Y) = \sum_{i}^{K} \left(|x_i^{\sigma} - y_i^{\sigma}| \right). \tag{10}$$

Note, that also the derivatives of the training spectra Y^{σ} may be stored to speed up the classification.

It may be seen, that the computational complexity of all presented dissimilarity measures is $\mathcal{O}(K)$. Actual computational requirements of these dissimilarities are, however, quite different. Table 1 summarizes the number of operations required for computation of a dissimilarity between two spectra with K bands. Quantities that can be pre-computed for the training data such as the derivatives or cumulative histograms are not included. Note, that the number of operations of the derivative-based dissimilarity depends on the smoothing parameter σ. For the smoothing σ, a window with $2\sigma + 1$ bands is used. The normalization of spectra to a unit area is included in case of the matching distance.

Table 1
Number of operations required for evaluation of a dissimilarity between two spectra with K bands

Dissimilarity measures	Number of				
	Add./sub.	Mult.	Div.	Sqrt	acos
Euclidean distance	$2K$	K		1	
Spectral angle mapper	$2K$	$2K$	1		1
Matching distance	$4K$	K			
Derivative-based dist.	$K(2\sigma+1)+3K$	$K(2\sigma+1)$			

4. Experiments

In this section, we present experiments with classifiers based on dissimilarities, applied to hyperspectral data. Our aim is to compare the performance and complexity of the two investigated classification methods. We have used NIR spectra of plastics acquired with InGaAs-camera using spectral range between 1 and 1.6 μm sampled into 120 bands. The dataset contains 16 classes of plastics.

We have prepared independent training and test sets with 1000 and 2000 samples, respectively. Four distance measures, presented in Table 1, were applied to the training data, producing the training dissimilarity matrices. In case of the derivative-based distance, we used two different settings of the smoothing parameter: $\sigma = 1.0$ and 2.0.

We have investigated the performance of the Fisher linear discriminant [7], applied to dissimilarity spaces of different dimensionalities. The random selection method was used to choose the representation set, i.e. the set of training examples used as the dimensions of the dissimilarity space. For each representation set, the Fisher discriminant was trained. Its performance was then estimated on the test set. This procedure was repeated 30 times for dimensionalities ranging from 5 to 500. Averaged results are presented in Fig. 4 as a function of the classification cost. The cost is estimated as the total number of operations required to compute the dissimilarity representation of a single spectrum and to label it (see Table 1 and Eqs. (2) and (4)). The lines with thick markers indicate the best overall performance of the nearest-neighbor classifier built using the complete training set with 1000 spectra. The endpoints for the derivative-based distances with $\sigma = 1.0$ and 2.0 are out of the plot at 10.8×10^5 and 15.6×10^5 operations, respectively.

It follows from the results, that the Fisher discriminant outperforms the nearest-neighbor rule in all cases except for the spectral angle mapper dissimilarity. The improvement is especially significant for matching and derivative-based distances and somewhat less in case of the Euclidean dissimilarity. Both methods reach comparable results for the spectral angle mapper distance.

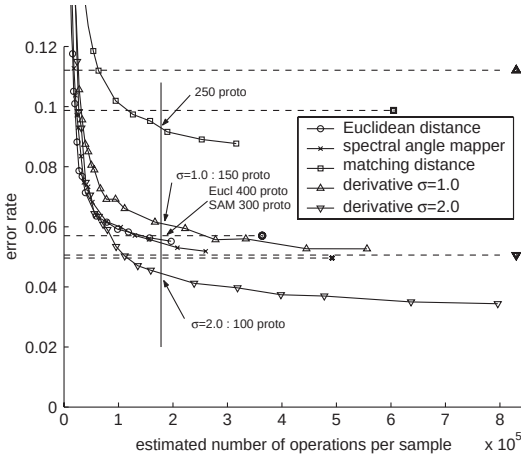

Fig. 4. Classification error versus the number of operations for dissimilarity-based classifiers. Curves represent the results of a Fisher linear discriminant trained in dissimilarity spaces based on randomly selected representation sets of sizes ranging from 5 to 500 dimensions. The dashed lines show the performance of the nearest-neighbor classifier using all 1000 training examples. For a fixed number of operations (vertical line), prototype counts are also given for each measure.

The best overall results were attained by the linear discriminant built on derivative-based dissimilarities. This suggests that it is, in this case, beneficial to treat spectra as connected entities and emphasise the differences in their shapes. Which distance measure (i.e. data representation) is the best in a particular situation remains, of course, application dependent.

If the number of operations for classification of new spectra is taken into account, all dissimilarity-based classifiers show a significant improvement over the nearest-neighbor rule. The Fisher classifier, trained on Euclidean distances consumes only 40% of operations to reach the performance of the nearest-neighbor rule. For the matching distance this is 20% and for the derivative-based dissimilarity with $\sigma = 2.0$ only 7% of operations.

It is interesting to note, that for each measure, a different number of prototypes is used for a fixed amount of operations (see the vertical line in the graph). A computationally complex approach may, therefore, reach lower errors employing less prototypes than simpler methods while spending the same time on the labeling of a new spectrum. An example is the derivative-based distance with $\sigma = 2.0$, outperforming with 100 prototypes both the SAM and the Euclidean dissimilarities using 300 and 400 prototypes, respectively.

Low error rates, obtained in the first experiment, suggest that classes are well separable. In the following experiment, we study the performance of dissimilarity-based classifiers in case of class overlap. We have used two classes, almost completely overlapping in the

spectral domain. The training set contains 120 and the independent test set 250 examples. Fig. 5 shows the error rate as a function of the number of operations spent on the labeling of a single new spectrum. Each sub-figure depicts the results for a different dissimilarity measure: the spectral angle mapper is in the left and the derivative-based distance with $\sigma = 2.0$ in the right pane. The results were averaged over 30 experiments.

Dashed lines correspond to the results of the nearest-neighbor rule, trained on a growing number of examples. We can observe, that including more training examples gradually improves the classification performance. In case of the Fisher linear discriminant, four curves are given, representing four training set sizes: 60, 80, 100, and 120 samples, respectively. Varying dimensionalities of the dissimilarity space are denoted by different markers. Please note the performance deterioration in the right part of the curves. In this area, the dimensionality of the dissimilarity space gets close to the number of training examples. Inverting the ill-conditioned covariance matrices becomes inaccurate and the classification error increases.

The vertical lines illustrate a situation when dissimilarities to 40 training prototypes are computed for each new spectrum to be labeled. For the nearest-neighbor classifier, this representation set coincides with the training set. In the case of the Fisher linear discriminant, the representation set merely defines the problem dimensionality. We can observe, how the classifier benefits from larger training sets without altering the classification time. This explains why the classifiers, built in a dissimilarity space, are more attractive from the implementation point of view, than the nearest neighbor rule.

For the sake of comparison, Table 2 presents the results of a minimum distance classifier, traditionally used to classify hyperspectral data [1]. We have applied the minimum distance classifier to each of the four dissimilarity measures. A mean spectrum was computed for each class using the whole training dataset. For each spectrum from the test set, a dissimilarity was computed to each class spectrum and a label was assigned based on the minimum distance. Results in the table are the means and standard deviations computed from 30

Table 2
Results of minimum distance classifiers in both discussed experiments: the 16-class problem and the 2-class problem with class overlap. The numbers are means and standard deviations computed from 30 repetitions

Dissimilarity measure	Exp 1:16 classes		Exp 2:2 classes with overlap	
	Mean error	Std. dev.	Mean error	Std. dev.
Euclidean distance	0.26	0.01	0.25	0.03
Spectral angle mapper	0.23	0.02	0.20	0.03
Matching distance	0.46	0.02	0.34	0.03
Derivative dist. $\sigma = 1.0$	0.10	0.02	0.22	0.03
Derivative dist. $\sigma = 2.0$	0.06	0.01	0.20	0.02

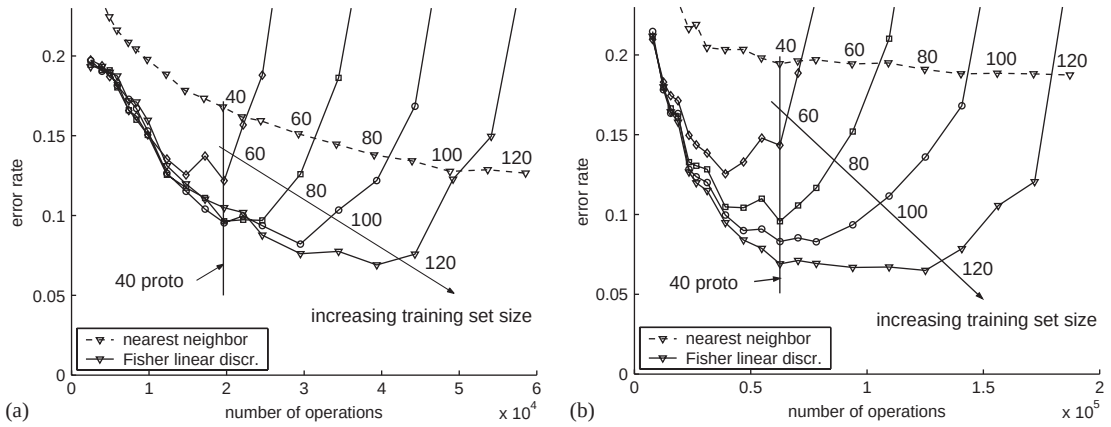

Fig. 5. Performance as a function of the number of operations needed for labeling a new spectrum in case of overlapping classes. The dashed lines correspond to the nearest-neighbor classifier and the solid curves to the Fisher linear discriminant, built on the training sets of increasing size. The numbers along the curves represent the training set size. The vertical lines show the performance using a fixed number of 40 training prototypes.

repetitions using independent training and test sets. The sizes of the training and the test sets are identical with the full sets used in the experiments above.

By comparing the results in the Table 2 with the Figs. 4 and 5, we can observe that, as expected, the minimum distance classifier gives significantly worse results compared to the classifiers built using larger numbers of prototypes. The best result of the minimum distance classifier is 6% of error using the derivative-based distance with $\sigma = 2.0$ on the first data set. By building classifiers in a dissimilarity space the result may be improved to about 4% of error using more than 100 prototypes. In case of overlapping classes, the minimum distance classifier fails to find any solution comparable to that of Fisher linear discriminant, built in the dissimilarity space.

When spectral data are represented by dissimilarities, the minimum distance classifier offers a fast solution if classes are not overlapping. However, more training prototypes need to be employed to reach low error rates in case of class overlap. We show, that classifiers built in a dissimilarity space provide better and more time effective solutions than the traditionally applied nearest-neighbor rule.

5. Conclusions

Spectra may be, for the sake of classification, represented by dissimilarities to other spectra. The advantage is that spectra may be treated as connected entities and expert knowledge may be used to emphasize application specific clues. The traditional way of classifying dissimilarity data is based on the nearest-neighbor classifier. It has been shown before, that the nearest-neighbor rule is often outperformed by more global classifiers built directly on dissimilarities. We have found out that classifiers, built in dissimilarity spaces, also provide a significantly faster way to label new objects. The reason is that they exploit the dissimilarity representation more effectively than the nearest-neighbor rule.

Each prototype is viewed as a separate problem dimension. The classifier may benefit from large training sets without impact on classification speed. The nearest-neighbor rule considers, on the other hand, each distance to a prototype as the individual training example. Using large training data sets slows down the classification process but this is necessary to reach low error rates.

In a dissimilarity approach, the spectrum is treated in its entirety which may be a viable alternative in cases when class characteristics cannot be easily handled in the feature space (i.e. slight shape variations). Generalizing classifiers may be built directly on dissimilarities leading to a low error classification. The computational requirements of this approach are, however, considerably higher than for the feature-based representation operating only on a subset of bands in a single spectrum.

Please note, that the dissimilarity measures, used in this study, considerably differ in the amount of exploited apriori information. For example, the derivative-based dissimilarity is based on our knowledge that the spectral shape is of importance and should be emphasized when building the data representation. It may be argued that differentiating spectra is, in fact, a pre-processing step making the comparison of presented dissimilarity measures unfair. We intentionally included the derivative-based dissimilarities to illustrate our point that the more sophisticated and computationally demanding measures may not only outperform the simpler ones but, eventually, may also be more time efficient. We believe that the very possibility to incorporate the pre-processing stage in the dissimilarity measure and hence into the representation is the most interesting aspect of the dissimilarity-based classification methods.

Acknowledgements

We would like to thank anonymous reviewers for careful reading and valuable comments on the manuscript and Wouter Caarls for fruitful discussions on the computational requirements. Further, we would like to express our gratitude to CTR Carinthian Tech Research AG for the experimental dataset. This work was supported by the Dutch Foundation for Applied Sciences (STW).

References

[1] Landgrebe D. Signal theory methods in multispectral remote sensing. New York: Wiley; 2003.
[2] Paclík P, Duin RPW. Classifying spectral data using relational representation. In: Leitner R, editor. Spectral imaging. Proceedings of the International Workshop of the Carinthian Tech Research AG. Vienna: Austrian Computer Society; 2003. p. 31–4.
[3] Pekalska E, Paclík P, Duin RPW. A generalized kernel approach to dissimilarity based classification. Journal of Machine Learning Research (special issue New Perspectives on Kernel Based Learning Methods) 2002;2(2):175–211.
[4] Arkedev AG, Braverman EM. Computers and pattern recognition. Washington, DC: Thompson; 1966.
[5] Duin RPW. Compactness and complexity of pattern recognition problems. In: Perneel C, editor. Proceedings of the International Symposium on Pattern Recognition In Memoriam Pierre Devijver, Brussels, B, February 12. Brussels: Royal Military Academy; 1999. p. 124–8.
[6] Pekalska E, Duin RPW. Dissimilarity representations allow for building good classifiers. Pattern Recognition Letters 2002; 23(8):943–56.
[7] Duda RO, Hart PE, Stork DG. Pattern classification, 2nd ed. New York: Wiley; 2001.
[8] Hennessy J, Patterson D. Computer architecture: a quantitative approach. San Mateo, CA: Morgan Kaufmann; 1990.

[9] de Ridder D, Pekalska E, Duin RPW. The economics of classification: error vs. complexity. In: Proceedings of the 16th IAPR International Conference on Pattern Recognition, vol. II, IAPR, ICPR 2002, Que., Canada, August 11–15. Los Alamitos, CA: IEEE Computer Society Press; 2002. p. 244–7.

[10] Duin RPW, Pekalska E, de Ridder D. Relational discriminant analysis. Pattern Recognition Letters 1999;20(11–13):1175–81.

[11] Yuhas RH, Goetz AFH, Boardman, JW. Discrimination among semiarid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Summaries of the Third Annual JPL Airborne Geoscience Workshop, Jet Propulsion Laboratory, Pasadena, CA, 1992. p. 147–9.

[12] Werman M, Peleg S, Rosenfeld S. A distance metric for multidimensional histograms. Computer, Vision, Graphics, and Image Processing 1985;32:328–36.

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