The use of dissimilarities for object recognition

Elżbieta Pękalska[1,2] **and Robert P.W. Duin**[1]

[1] Delft Univ. of Technology, Electr. Eng., Math. and Comp. Science, The Netherlands
[2] University of Manchester, School of Computer Science, Manchester, United Kingdom
email: e.pekalska@ewi.tudelft.nl

Summary

Pairs of objects, represented by histograms, contours, shapes or images, can be compared by the use of a dissimilarity measure. Pattern classes may be learned from dissimilarity matrices enabling more efficient object recognition than by using the traditional nearest neighbor rule.

Introduction

The ability to see differences supports the human recognition of patterns. When knowledge on differences between objects, phenomena or events is gathered, one starts to observe their similarities. This makes the process of finding good discriminating features easier.

Groups of objects can be studied, starting from the perspective that some are dissimilar, while others are similar. This is the basis on which methods for learning and recognizing objects by automatic means should rely on. This mimics the human ability of recognition by computers and sensors. In the area of statistical pattern recognition objects are described by a set of numerical features and represented as points in a feature vector space. Usually, it is a Euclidean space, equipped with the algebraic structure of inner product, hence the associated norm and metric. The difference between objects is interpreted as the Euclidean distance between their point representations in this space.

The concept of dissimilarity leads to dissimilarity representations. These are numerical representations, in which each value captures the degree of commonality between pairs of objects. Since a dissimilarity measure can be defined on arbitrary data such as sensor measurements, shapes, strings, graphs, or even vectors in a feature space, this representation becomes very general. Statistical approaches offer the well-developed theory of vector spaces and ways of dealing with noise. Syntactical approaches posses the structure-handling capability lacked by the first. As they both make use of distances, their strengths can be combined on the level of dissimilarity representation; see also [5]. Below, we will briefly review some developed dissimilarity measures and introduce the methodology of learning from dissimilarity representations.

Dissimilarity representation and learning methods

There exists a plethora of dissimilarity measures used for object matching and comparison. We will mention some in relation to histograms, spectra and shapes. When normalized histograms and spectra are considered as probability distributions, probabilistic measures like the divergence can be then applied. The order of bins or wavelengths is important as neighboring elements are expected to be correlated. This may be incorporated in the measure e.g. by computing the ℓ_p-distance between cumulative histograms or between approximated histograms or spectra derivatives [9, 10].

Shape dissimilarity measures are usually developed to solve pattern matching problems. A typical example is to find geometric transformations (from a specified class) of one pattern (e.g. shape or contour) into another one such that a predefined cost is minimized [16, 1]. For the matching of binary images (or contours), variants of Hausdorff distances are used [4, 8]. Ideas for curve matching are presented in [6], the definition of an elastic distance is considered in [18] and some framework for deformable matching is developed in [2].

Shape contours can be compared by the use of an edit distance applied to the chain codes [17]. Since starting points may be arbitrary, matching is performed between all cyclic permutations. Alternatively, contours may be represented as sequences of points in 2D. Then, string descriptions consist of the direction vectors between consecutive point pairs. Different substitution costs used in the edit distance lead to different distance measures [3]. Dissimilarities can also be evaluated as edit distances between shock graphs representing shapes [15].

In pattern matching one usually looks for the object which yields the smallest mismatch. In classification, this translates into the 1-Nearest Neighbor (1-NN) rule, assigning objects to the class of the nearest neighbor, defined by the smallest dissimilarity. Alternative learning approaches based on dissimilarities, that can be applied more efficiently, however, exist.

Consider a set of objects $R = \{p_1, p_2, \ldots, p_n\}$, called a representation set, and a dissimilarity measure d, derived from the objects directly, their sensor representations, or some other intermediate representations. It is nonnegative and obeys the reflexivity condition, $d(x, x) = 0$, but it may be non-metric. A dissimilarity representation (DR) of an example x is a vector $D(x, R) = [d(x, p_1),\, d(x, p_2), \ldots, d(x, p_n)]$ of the dissimilarities between x and the objects of R. Consequently, for a set T of N objects, it extends to an $N \times n$ dissimilarity matrix $D(T, R)$. R is assumed to be a relatively small set of representative objects. R may be a subset of T or they might be completely distinct. In the former case, R should be selected out of T for the specific learning task to characterize variability in the data. Various random and systematic approaches can be used; see [9, 13, 10].

Assume a dissimilarity representation $D(T, R)$, where T is a training set. Three frameworks for automatic learning are considered. In the first one, the given dissimilarities are directly used, as carrying information on local neighbors. This leads to variants of the k-Nearest Neighbor (k-NN) method. The second approach addresses a dissimilarity representation as a data-dependent mapping specified by the set R. A mapping $D(x, R) : \mathcal{X} \to \mathbb{R}^n$ is defined as $D(x, R)$. Hence, learning takes place in a dissimilarity space, in which each dimension corresponds to a dissimilarity with a representation object, $d(\cdot, p_i)$. This vector space is assumed to be equipped with the traditional inner product and the associated norm, and Euclidean metric. Additionally, if beneficial, other distance measures may be considered. The third approach relies on an approximate embedding of dissimilarities into a pseudo-Euclidean space, an extension of a Euclidean space [7]. This is a real vector space $\mathcal{E} = \mathbb{R}^{(p,q)}$ equipped with a non-degenerate, indefinite inner product. $\mathcal{E}$ admits a direct orthogonal decomposition $\mathcal{E} = \mathbb{R}^p \oplus \mathbb{R}^q$ such that the inner product is positive definite on $\mathbb{R}^p$ and negative definite on $\mathbb{R}^q$. Many traditional classifiers, especially those relying on (generalized) inner product operations, can be applied or adapted for dissimilarity spaces and pseudo-Euclidean spaces. These three frameworks are used for clustering and data characterization, novelty detection and classification. The details can be found in [11, 14, 10].

Clustering example

Two fish contours are compared such that one is fixed and the other is transformed by a scaling, shift and rotation to yield the smallest sum of square Euclidean distances. A dissimilarity matrix D is derived for a collection of $n = 60$ examples. As the clustering results based on D are imperfect (as judged by us visually), a new matrix D_n is computed such that $d_n(i, j) = \frac{d(i,j)}{\sum_k d(i,k) + \sum_l d(l,j)}$. Every dissimilarity value between the i-th and j-th examples is hereby scaled with respect to their overall dissimilarities. In this way, clusters are emphasized. The results of three clustering techniques, EM-clustering in a PCA-dissimilarity space, k-centers and hierarchical clustering with complete linkage, are shown in Fig. 1. All three give good results. As the first two methods depend on initialization, they are run 50 times and

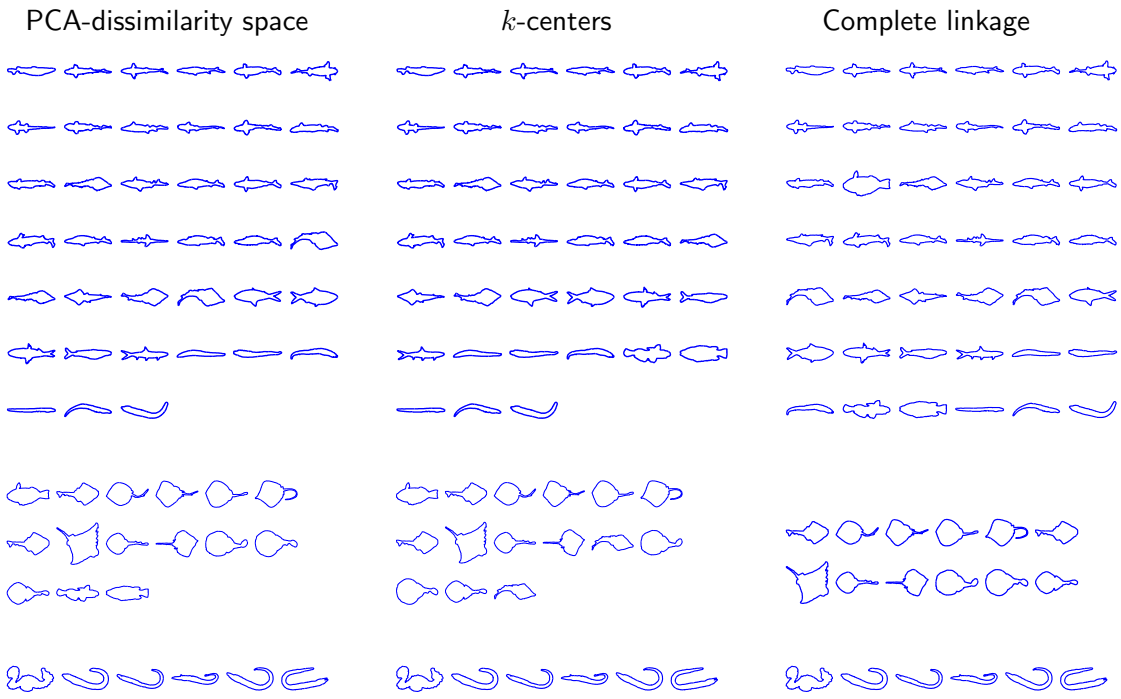

Figure 1: Clustering of fish shapes into three clusters.

the best results, as judged by some appropriate criteria, are presented. For the EM-clustering, the criterion is $J = \frac{\sum_i ni/(n-n_i)\sum_j n_j A_{ij}}{2\sum_i n_i A_{ii}}$, where n_i is the number of objects per class and A_{ij} is the average between-class dissimilarity and A_{ii} is the average within-class dissimilarity. For the k-centers, the best result is found by minimizing the maximum of the dissimilarities to the selected centers.

Classification example

An edit non-metric distance between string contours of pen-written digits is considered; see [12] for details. The insertion and delation costs are fixed, while the substitution cost is the distance between neighboring direction vectors. There are 10 classes, represented by approximately 350 examples each. The training set T consists of $10{\cdot}100$ examples and the test set S has $10{\cdot}200$ examples. R is a random subset of T. A pseudo-Euclidean space and a PCA-dissimilarity space are determined with the dimensionality of $0.3|R|$. A regularized quadratic classifier is trained in both spaces. The experiments are repeated 50 times for representations sets of various sizes and the results are averaged. Fig. 2 presents the resulting error curves as a function of $|R|$. The legend refers to the element-wise transformations applied to dissimilarity representations. As a comparison, the 1-NN rule (the best k-NN) is directly applied to $D(\cdot, R)$. The averaged error of the 1-NN on $D(\cdot, T)$ is approximately 1% and it relies on $|T| = 10 \cdot 100$ examples to which dissimilarities need to be computed. The same result is reached for a quadratic classifier in pseudo-Euclidean and PCA-dissimilarity spaces and $|R| = 10{\cdot}15$ examples. When the computation of dissimilarities is very costly, the latter offer an alternative to the k-NN methods.

Conclusion

Dissimilarity representations can be considered as a way of integrating statistical and structural approaches to pattern recognition. Some learning methods are developed, much is still open for the future research.

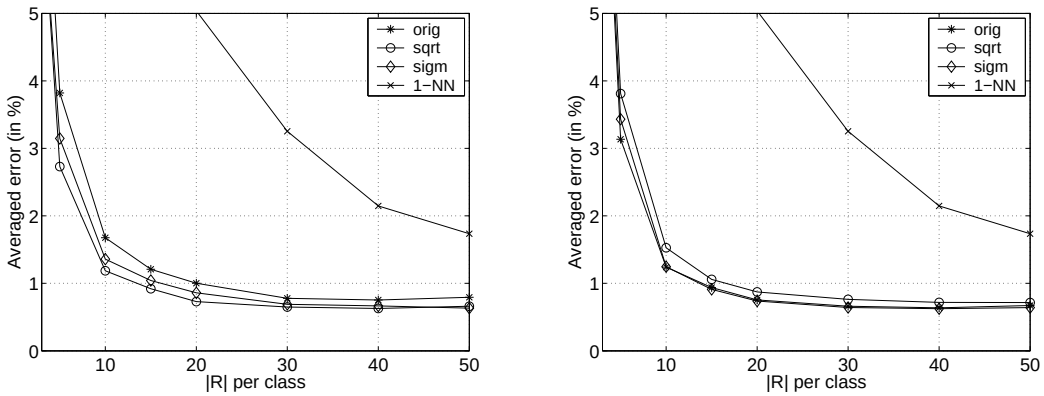

Figure 2: The averaged classification error for quadratic classifiers trained in pseudo-Euclidean (left) and dissimilarity (right) spaces. As a comparison, the best k-NN rule, the 1-NN, is directly applied to $D(\cdot, R)$. The 1-NN rule on $D(\cdot, T)$ yields an average error of 1%. The standard deviations reach on average 0.3% and maximally 1% for very small R.

References

[1] R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38:2365–2385, 1998.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on PAMI*, 24(24):509–522, 2002.

[3] H. Bunke, X. Jiang, K. Abegglen, and A. Kandel. On the weighted mean of a pair of strings. *Pattern Analysis and Applications*, 5(1):23–30, 2002.

[4] M.P. Dubuisson and A.K. Jain. Modified Hausdorff distance for object matching. In *International Conference on Pattern Recognition*, volume 1, pages 566–568, 1994.

[5] R.P.W. Duin, E. Pękalska, P. Paclík, and D.M.J. Tax. The dissimilarity representation, a basis for domain based pattern recognition? In *Pattern representation and the future of pattern recognition, ICPR 2004 Workshop Proc.*, pages 43–56, United Kingdom, 2004.

[6] Y. Gdalyahu and D. Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. *IEEE Trans. on PAMI*, 21(12), 1999.

[7] L. Goldfarb. A unified approach to pattern recognition. *Pattern Recognition*, 17:575–582, 1984.

[8] D.P. Huttenlocher, G.A. Klanderman, and J.R. William. Comparing images using the Hausdorff distances. *IEEE Trans. on PAMI*, 15:850–863, 1993.

[9] P. Paclík and R.P.W. R.P.W. Duin. Dissimilarity-based classification of spectra: computational issues. *Real Time Imaging*, 9(4):237–244, 2003.

[10] E. Pękalska. *Dissimilarity representations in pattern recognition. Concepts, theory and applications.* PhD thesis, ASCI dissertation series no. 109, Delft University of Technology, January 2005.

[11] E. Pękalska and R.P.W. Duin. Dissimilarity representations allow for building good classifiers. *Pattern Recognition Letters*, 23(8):943–956, 2002.

[12] E. Pękalska, R.P.W. Duin, S. Günter, and H. Bunke. On not making dissimilarities euclidean. In *Joint IAPR Int. Workshops on SSPR and SPR, LNCS*, pages 1145–1154. Springer-Verlag, 2004.

[13] E. Pękalska, R.P.W. Duin, and P. Paclík. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, to appear, 2005.

[14] E. Pękalska, P. Paclík, and R.P.W. Duin. A Generalized Kernel Approach to Dissimilarity Based Classification. *Journal of Machine Learning Research*, 2(2):175–211, 2002.

[15] T.B. Sebastian, P.N. Klein, and B.B. Kimia. Recognition of shapes by editing shock graphs. In *International Conference on Computer Vision*, pages 755–762, 2001.

[16] R. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands, 1999.

[17] R.A. Wagner and M.J. Fisher. The string-to-string correction problem. *Journal of the Association for Computing Machinery*, 21(1):168–173, 1974.

[18] L. Younes. Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, 1998.