Non-Euclidean Dissimilarities: Causes and Informativeness

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

[1] Faculty of Electrical Engineering, Mathematics and Computer Sciences, Delft University of Technology, The Netherlands
r.duin@ieee.org
[2] School of Computer Science, University of Manchester, United Kingdom
pekalska@cs.man.ac.uk

Abstract. In the process of designing pattern recognition systems one may choose a representation based on pairwise dissimilarities between objects. This is especially appealing when a set of discriminative features is difficult to find. Various classification systems have been studied for such a dissimilarity representation: the direct use of the nearest neighbor rule, the postulation of a dissimilarity space and an embedding to a virtual, underlying feature vector space.

It appears in several applications that the dissimilarity measures constructed by experts tend to have a *non*-Euclidean behavior. In this paper we first analyze the causes of such choices and then experimentally verify that the non-Euclidean property of the measure can be informative[1].

1 Introduction

Dissimilarities are a natural way to represent objects. Some consider them as more fundamental than features [1]. This paper studies particular aspects of dissimilarities. First, we analyze why *non*-Euclidean dissimilarities arise in recognition. Then, we discuss how non-Euclidean relations can become informative.

Dissimilarities have been studied in [2] for both supervised and unsupervised learning as an alternative to the use of features in building representations. They are especially useful in two contexts. First, when no clear properties are available to become features and, secondly, when objects can be compared globally such as shapes in images, time signals or spectra. Classifiers relying on dissimilarity relations can outperform nearest neighbor approaches or template matching.

There are two main approaches for building vector spaces from dissimilarities. One postulates a Euclidean space, the so-called dissimilarity space, in which features are defined by dissimilarities to a representation set of objects. The other relies on a linear embedding of the given dissimilarity matrix. The first is very general and can always be used. It demands a proper choice of the representation

[1] We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250) as well as the Engineering and Physical Sciences Research Council in the UK.

E.R. Hancock et al. (Eds.): SSPR & SPR 2010, LNCS 6218, pp. 324–333, 2010.

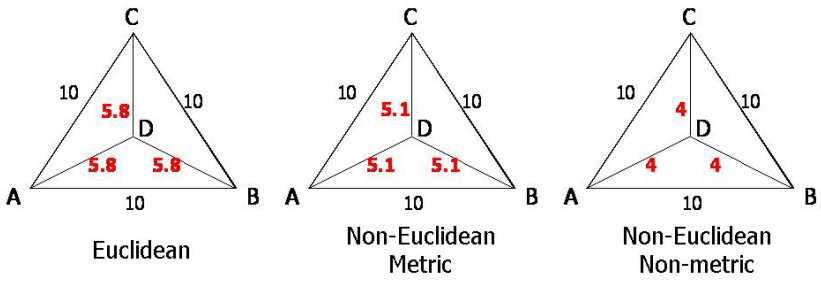

Fig. 1. Illustration of the difference between Euclidean, metric, but non-Euclidean and non-metric dissimilarities. If the distances between the four points A, B, C and D are given as in the left plot, then an exact 2-dimensional Euclidean embedding is possible. If the distances are given as given as in the middle plot, the triangle inequality is obeyed. So the given distances are metric, but no isometric Euclidean embedding exists. The distances in the right plot are non-Euclidean as well as non-metric.

set, a problem similar to feature selection [3]. In the selection of the representation set the intrinsic nature of the dissimilarities may be used, e.g. such that the closer or the more likely objects belong to the same class. In the construction of classifiers in this space, dissimilarities may be used in the same way as features. This, however, neglects their original character of pairwise dissimilarities.

On the contrary, in the embedding of a dissimilarity matrix to a space with a given metric, the nature of dissimilarities is preserved. It is natural to search for an embedding to a Euclidean space as the Euclidean metric is assumed either implicitly or explicitly in many classification systems. It appears however that in many applications *non*-Euclidean and even non-metric dissimilarities are used due to their good performance in template matching. (See Fig. 1 to understand the difference between non-Euclidean and non-metric dissimilarities.) An early example is given by Dubuisson and Jain [4] who showed that in a set of image object matching examples the non-metric modified Hausdorff distance outperforms the original metric Hausdorff distance. Non-Euclidean distances can only be approximately embedded in a Euclidean space. Goldfarb showed how a so-called Pseudo-Euclidean (PE) embedding can be found [5] for any symmetric dissimilarity matrix. It is error free, but requires a different distance measure. Some classifiers can be defined in this space, such as the nearest mean, nearest neighbor, Parzen classifier, LDA and QDA. The relation of the latter three with densities is not clear yet, as the concept of probability density distributions has not been well defined for the PE space.

The question on usefulness or non-importance of non-Euclidean distances has been around for some time. Goldfarb did not find good applications for the PE space and abandoned all vector space approaches [5]. Instead, he focussed on the Evolving Transformation System (ETS) which by a structural representation aimed to model relations between objects [6], but for which it was difficult to find classifiers [7]. This is common for structural approaches, but is solved in a heuristic way by the use of a dissimilarity space, in which the non-Euclidean nature of the object relations in neglected. In [8] some studies are presented

indicating that non-Euclideaness of the data (i.e. the deviation from Euclidean distances) might contribute to the classification performance. In [9] Euclidean corrections of non-Euclidean data are discussed.

Finding classifiers for non-Euclidean dissimilarities is directly related to study of indefinite kernels, important for the optimization of Support Vector Machines (SVMs). The quadratic programming solution used for SVMs is not guaranteed to be optimal for kernels that violate the Mercer conditions. As the inner product definition for the pseudo-Euclidean space leads to indefinite kernels, the construction of SVMs in such a space is thereby hampered. For that reason there is a strong tendency in the machine learning community to design positive semidefinite (psd), i.e. Mercer, kernels. On the other hand, since non-Euclidean dissimilarities are frequently used in pattern recognition applications, it is relevant to know how to deal with them. Should we avoid them, correct them into Euclidean distances to make them suitable for the full set of traditional classification tools, or keep them as they are and design special classifiers for non-Euclidean data?

In this paper we want to contribute to this discussion in two ways. In Section 3 we will analyze the causes behind non-Euclidean dissimilarities. In Section 4 we will argue why non-Euclidean dissimilarities can be informative and we will present some examples. First, however, the dissimilarity space and PE embedded space will be briefly introduced in Section 2.

2 Dissimilarity Representations

The dissimilarity representation has extensively been discussed, e.g. in [2] or [10], so we will only focus here on aspects that are essential for this paper.

Traditionally, dissimilarity measures were often optimized for the nearest neighbor classification performance. In addition, they were also widely used in hierarchical cluster analysis. Later, the resulting dissimilarity matrices served for the construction of vector spaces and the computation of classifiers. Only more recently proximity measures have been designed for classifiers that are more general than the nearest neighbor rule. These are usually similarities and kernels (but not dissimilarities), used in combination with SVMs. So, research on the design of dissimilarity measures such that they fit to a wider range of classifiers is still in an early stage. Consequently, we will restrict ourselves in this paper to the common practice of measures optimized for nearest neighbor classifiers. New objects are thereby classified just on the basis of pairwise comparisons. They are not represented in a vector space. An additional step is necessary to create such a space, and as a result, this will allow the use of other classifiers. The two ways investigated so far are the dissimilarity space and PE embedded space.

2.1 Dissimilarity Space

Let $\mathcal{X} = \{o_1, \ldots, o_n\}$ be a training set of objects o_i. These are not necessarily vectors but can be real world objects, or e.g. images or time signals. Given a dissimilarity function and/or dissimilarity data, we define a data-dependent mapping $D(\cdot, R) : \mathcal{X} \to \mathbb{R}^k$ from $\mathcal{X}$ to the so-called *dissimilarity space* [11,12,13]. The

k-element set R consists of objects that are representative for the problem. This set, the representation or prototype set, may be a subset of $\mathcal{X}$. In the dissimilarity space each dimension $D(\cdot, p_i)$ describes a dissimilarity to a prototype p_i from R. Here we will choose $R := \mathcal{X}$. As a result, every object is described by an n-dimensional vector $D(o, \mathcal{X}) = [d(o, o_1) \ \ldots \ d(o, o_n)]^T$, which are just the rows of the given dissimilarity matrix D. The resulting vector space is endowed with the traditional inner product and the Euclidean metric. As we have n training objects in an n-dimensional space, a classifier such as SVM is needed to handle this situation.

2.2 Pseudo-Euclidean Embedded Space

A Pseudo-Euclidean (PE) space $\mathcal{E} = \mathbb{R}^{(p,q)} = \mathbb{R}^p \oplus \mathbb{R}^q$ is a vector space with a non-degenerate indefinite inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$ such that $\langle \cdot, \cdot \rangle_{\mathcal{E}}$ is positive definite on $\mathbb{R}^p$ and negative definite on $\mathbb{R}^q$ [5,2]. The inner product in $\mathbb{R}^{(p,q)}$ is defined (using an orthonormal basis) as $\langle x, y \rangle_{\mathcal{E}} = x^T \mathcal{J}_{pq} y$, where $\mathcal{J}_{pq} = [I_{p \times p} \ 0; \ 0 \ -I_{q \times q}]$ and I is the identity matrix. As a result, $d^2_{\mathcal{E}}(x, y) = (x - y)^T \mathcal{J}_{pq} (x - y)$.

Any symmetric $n \times n$ dissimilarity matrix D can be embedded into a $(n-1)$-dimensional PE space [5,2]. The eigenvalue decomposition needed for the embedding results in p positive and q negative eigenvalues λ_j, $p+q = n-1$, and the corresponding eigenvectors. To inspect the amount of non-Euclidean influence in the derived PE space, we use the negative eigenfraction (*NEF*)

$$NEF = \sum_{j=p+1}^{p+q} |\lambda_j| / \sum_{i=1}^{p+q} |\lambda_i| \in [0, 1] \tag{1}$$

as a measure for the non-Euclidean behavior of the dissimilarity matrix.

If the negative eigenvalues are considered as the result of noise or errors, they may be neglected. As a result, a 'corrected' dissimilarity matrix D_p may be computed by using a positive subspace $\mathbb{R}^p$ of the embedded space $\mathbb{R}^{(p,q)}$:

$$d^2_{\mathcal{E}p}(x, y) = (x_p - y_p)^T (x_p - y_p), \tag{2}$$

where x_p, y_p are projections of the vectors x, y from $\mathbb{R}^{(p,q)}$ onto the subspace $\mathbb{R}^p$ and all diagonal values of $\mathcal{J}_{pq}$ become $+1$. In order to investigate a possible contribution of the negative eigenvalues, the residue can be computed by:

$$d^2_{\mathcal{E}q}(x, y) = -(x_q - y_q)^T (x_q - y_q) \tag{3}$$

where x_q, y_q are projections of the vectors x, y from $\mathbb{R}^{(p,q)}$ onto the negative subspace $\mathbb{R}^q$ and all diagonal values of $\mathcal{J}_{pq}$ become -1. The complete dissimilarity matrix D can thereby be decomposed as

$$D^{*2} = D^{*2}_p - D^{*2}_q \tag{4}$$

in which the values of D^{*2}_q are positive and ^{*2} denotes an element-wise squaring. n-dimensional dissimilarity spaces may also be defined for D_p and D_q.

3 Causes of Non-Euclidean Dissimilarity Measures

In the previous section two procedures for deriving vector spaces are presented. One is general, but neglects the pairwise dissimilarity characteristics. The other is specific but suffers from the possible non-Euclidean relations. If we want to make use of the specific dissimilarity character, but suffer from the non-Euclidean behavior, it is important to analyze why this happens. Should we avoid it, should we correct it, or should we design special classifiers that deal with it?

First, it should be emphasized how common non-Euclidean measures are. In [2] an extensive overview of such measures has been given, but in many occasions we have encountered that this fact is not fully recognized. Almost all probabilistic distance measures are non-Euclidean. This implies that by dealing with object invariants, the dissimilarity matrix resulting from the overlap between the object pdfs is non-Euclidean. Also the Mahalanobis class distance as well as the related Fisher criterion are non-Euclidean. Consequently many non-Euclidean distance measures are used in cluster analysis and in the analysis of spectra in chemometrics and hyperspectral image analysis.

In shape recognition, various dissimilarity measures are used based on the weighted edit distance, on variants of the Hausdorff distance or on non-linear morphing. Usual parameters are optimized within an application w.r.t. the performance based on template matching and other nearest neighbor classifiers [14]. Almost all have non-Euclidean behavior and some are even non-metric [4].

In the design and optimization of the dissimilarity measures for template matching, their Euclidean behavior is not an issue. With the popularity of support vector machines (SVMs), it has become important to design kernels (similarities) which fulfill the Mercer conditions. This is equivalent to a possibility of an isometric Euclidean embedding of such a kernel (or dissimilarities). Next subsections discuss reasons that give rise to violations of these conditions leading to non-Euclidean dissimilarities or indefinite kernels.

3.1 Non-intrinsic Non-Euclidean Dissimilarities

Below we identify some non-intrinsic causes for non-Euclidean dissimilarities.

Numeric inaccuracies. Non-Euclidean dissimilarities arise due the numeric inaccuracies caused by the use of a finite word length. If the intrinsic dimensionality of the data is lower than the sample size, eigenvalues that should be zero during embedding, may become negative due to numeric inaccuracies. It is thereby advisable to neglect dimensions (features) that correspond to very small positive and negative eigenvalues.

Overestimation of large distances. Complex measures are used when dissimilarities are derived from raw data such as (objects in) images. They may define the distance between two objects as the length of the path that transforms one object into the other. Examples are the weighted edit distance [15] and deformable templates [16]. In the optimization procedure that minimizes the path length, the procedure may approximate the transformation costs from

above. As a consequence, too large distances are found. If the distance measure is Euclidean, such errors make it non-Euclidean or even non-metric.

Underestimation of small distances. The underestimation of small distances has the same result as the overestimation of large distances. It may happen when the pairwise comparison of objects is based on different properties for every pair, like in studies on consumer preference data. Another example is the comparison of partially occluded objects in computer vision.

3.2 Intrinsic Non-Euclidean Dissimilarities

The causes discussed in the above may be judged as accidental. They result either from computational or observational problems. If better computers and observations were available, they would disappear. Now we will focuss on dissimilarity measures for which this will not happen. We will present three possibilities.

Non-Euclidean Dissimilarities. As already indicated at the start of this section, there can be arguments from the application side to use another metric than the Euclidean one. An example is the l_1-distance between energy spectra as it is related to energy differences. Although the l_2-norm is very convenient for computational reasons and it is rotation invariant in a Euclidean space, the l_1-norm may naturally arise from the demands in applications.

Invariants. A very fundamental reason is related to the occurrence of invariants. Frequently, one is not interested in the dissimilarity between objects A and B, but between their equivalence classes i.e. sets of objects $A(\theta)$ and $B(\theta)$ in which θ controls an invariant. One may define the dissimilarity between the A and B as the minimum difference between the sets defined by all their invariants.

$$d^*(A,B) = \min_{\theta_A} \min_{\theta_B} (d(A(\theta_A), B(\theta_B))) \tag{5}$$

This measure is non-metric: the triangle inequality may be violated as for different pairs of objects different values of θ are found minimizing (5).

Sets of vectors. Complicated objects like multi-region images may be represented by sets of vectors. Distance measures between such sets have already been studied for a long time in cluster analysis. Many are non-Euclidean or even non-metric, e.g. the single linkage procedures. It is defined as the distance between the two most neighboring points of the two clusters being compared, is non-metric. It even holds that if $d(A,B) = 0$, then it does not follow that $A \equiv B$.

For the single linkage dissimilarity measure it can be understood why the dissimilarity space may be useful. Given a set of such dissimilarities between clouds of vectors, it can be concluded that two clouds are similar if the entire sets of dissimilarities with all other clouds are about equal. If just their mutual dissimilarity is (close to) zero, they may still be very different.

The problem with the single linkage dissimilarity measure between two sets of vectors points to a more general problem in relating sets and even objects. In [17] an attempt has been made to define a proper Mercer kernel between two

sets of vectors. Such sets are in that paper compared by the Hellinger distance derived from the Bhattacharyya's affinity between two pdfs $p_A(x)$ and $p_B(x)$ found for the two vector sets A and B:

$$d(A,B) = \left[\int \left(\sqrt{p_A(x)} - \sqrt{p_B(x)} \right)^2 \right]^{1/2}. \tag{6}$$

The authors state that by expressing $p(x)$ in any orthogonal basis of functions, the resulting kernel K is automatically positive semidefinite (psd). This is only correct, if all vector sets $A, B, ...$ to which the kernel is applied have the same basis. If different bases are derived in a pairwise comparison of sets, the kernel will become indefinite.

This makes clear that indefinite relations may arise in any pairwise comparison of real world objects if they are first represented in some joint space for the two objects, followed by a dissimilarity measure. These joint spaces may be different for different pairs! Consequently, the total set of dissimilarities will likely have a non-Euclidean behaior, even if a single comparison is defined as Euclidean, as in (6).

4 Informativeness

Are non-Euclidean dissimilarity measures informative? How should this question be answered? It is different than the question whether non-Euclidean measures are better than Euclidean ones. This second question can certainly not be answered in general. After we study a set of individual problems and compare a large set of dissimilarity measures we may find that for some problems of interest the best measure is non-Euclidean. Such a result is always temporal. A new Euclidean measure may later be found that outperforms the earlier ones.

The question of informativeness however may be answered in an absolute sense. Even if a particular measure is not the best one, its non-Euclidean characteristic can be informative as by removing it, performance deteriorates. Should this result also be found by a classifier in the non-Euclidean space? If an Euclidean correction can be found for an initial non-Euclidean representation that enables the construction of a good classifier, is the non-Euclidean dissimilarity measure then informative? We answer this question positively as any transformation can be included in the classifier and thereby effectively a classifier for the non-Euclidean representation has been found.

We will therefore state that the non-Euclidean character of a dissimilarity measure is non-informative if the classification result improves by removing its non-Euclidean characteristic. The answer may be different for different classifiers. The traditional way of removing the non-Euclidean characteristic is by neglecting the negative eigenvectors in the pseudo-Euclidean embedding. This is represented by the recomputed dissimilarities in the positive part of the pseudo-Euclidean space, D_p in (4). The dissimilarity representation based on D_q isolates the non-Euclidean characteristic of the given dissimilarity matrix D and can be used as a check to see whether there is any class separability visibility in the removed part of the embedding.

Table 1. Classification errors of the linear SVM for several representations using leave-one-out crossvalidation

	size	classes	Non-Metric	NEF	Rand Err	Original, D	Positive, D_p	Negative, D_q
Chickenpieces45	446	5	0	0.156	0.791	**0.022**	0.132	0.175
Chickenpieces60	446	5	0	0.162	0.791	**0.020**	0.067	0.173
Chickenpieces90	446	5	0	0.152	0.791	**0.022**	0.052	0.148
Chickenpieces120	446	5	0	0.130	0.791	**0.034**	0.108	0.148
FlowCyto	612	3	1e-5	0.244	0.598	0.103	**0.100**	0.327
WoodyPlants50	791	14	5e-4	0.229	0.928	**0.075**	0.076	0.442
CatCortex	65	4	2e-3	0.208	0.738	**0.046**	0.077	0.662
Protein	213	4	0	0.001	0.718	0.005	**0.000**	0.634
Balls3D	200	2	3e-4	0.001	0.500	**0.470**	0.495	0.000
GaussM1	500	2	0	0.262	0.500	**0.202**	**0.202**	0.228
GaussM02	500	2	5e-4	0.393	0.500	0.204	**0.174**	0.252
CoilYork	288	4	8e-8	0.258	0.750	**0.267**	0.313	0.618
CoilDelftSame	288	4	0	0.027	0.750	**0.413**	0.417	0.597
CoilDelftDiff	288	4	8e-8	0.128	0.750	**0.347**	0.358	0.691
NewsGroups	600	4	4e-5	0.202	0.733	**0.198**	0.213	0.435
BrainMRI	124	2	5e-5	0.112	0.499	0.226	**0.218**	0.556
Pedestrians	689	3	4e-8	0.111	0.348	**0.010**	0.015	0.030

We analyze a set of public domain dissimilarity matrices used in various applications, as well as a few artificially generated ones. The details of the sets are available from the D3.3 deliverable of the EU SIMBAD project[2]. See Table 1 for some properties: *size* (number of objects), (number of) *classes*, *non-metric* (fraction of triangle violations), *NEF* (negative eigenfraction) and *Rand Err* (classification error by random assignment). Every dissimilarity matrix is made symmetric by averaging with its transpose and normalized by the average off-diagonal dissimilarity. We compute the linear SVM in the dissimilarity spaces based on the original, 'positive' and 'negative' dissimilarities D, D_p and D_q. Error estimates are based on leave-one-out crossvalidation. These experiments are done in a transductive way: test objects are included in the derivation of the embedding as well as the dissimilarity representations.

The four Chickenpieces datasets are the averages of 11 dissimilarity matrices derived from a weighted edit distance between blobs [15]. FlowCyto is the average of four specific histogram dissimilarities including an automatic calibration correction. WoodyPlants is a subset of the shape dissimilarities between leaves of woody plants [10]. We used classes with more than 50 objects. Catcortex is based on the connection strength between 65 cortical areas of a cat, [12]. Protein measures protein sequence differences using an evolutionary distance measure [18]. Balls3D is an artificial dataset based on the surface distances of randomly

[2] `http://simbad-fp7.eu/`

positioned balls of two classes having a slightly different radius. GaussM1 and GaussM02 are based on two 20-dimensionally normally distributed sets of objects for which dissimilarities are computed using the Minkowsky distances 1 respectively 0.2. The three Coil datasets are based on the same sets of SIFT points in COIL images compared by different graph distances. BrainMRI is the average of 182 dissimilarity measures obtained from MRI brain images. Pedestrians is a set of dissimilarities between detected objects (possibly pedestrians) in street images of the classes 'pedestrian', 'car', 'other'. They are based on cloud distances between sets of feature points derived from single images.

5 Discussion and Conclusions

In this paper we identify a number of causes that give rise to non-Euclidean and non-metric dissimilarities and we wonder whether they can play an informative role for classification purposes. In the above table some phenomena can be observed that illustrate these issues and answer some questions.

- From the negative eigenfraction column (NEF) it can be understood that all datasets are non-Euclidean. Protein set has a nearly Euclidean measure as it has just a very small contribution from the negative eigenvalues.
- A number of datasets is metric. Chickenpieces, as we used the dataset here, based on averages of weighted edit-distances between contours, should be metric as the edit-distance searches for the smallest edit path. In the individual dissimilarities matrices some violations can be observed due to approximations in the path optimization procedure. After averaging this is solved. Interesting is that this procedure improves the results significantly. The performances found in the dissimilarity space are to our knowledge the best ever published for this dataset .
- The original, uncorrected, pseudo-Euclidean dissimilarities are the best (in bold) in many cases. For these the deletion of the negative eigenvectors works counter-productive.
- For the other datasets the Euclidean correction works out well.
- However, in almost all cases the negative part of the space alone shows some separability of the classes (compare with the random assignment error), proving that it contains some information.
- In a few cases the negative space shows very good results, e.g. Pedestrians.
- In the Balls3D example all information is concentrated in the negative space.

In conclusion it is stated that the non-Euclidean characteristic of dissimilarity data, resulting from the search of the best representation for nearest neighbor assignment should not be directly removed from the representation as by using the positive space only. This space performs often similar or worse compared to the original dissimilarities. The negative space itself, concentrating all non-Euclidean characteristics, yields usually a better than random performance and surprisingly leads to a very good result in some problems. From these two observations, removing the negative space often deteriorates results and the negative

space alone shows a better than random performance, it is concluded that negative space and thereby the non-Euclideaness of the data is informative. It should be realized that these conclusions are classifier dependent.

References

1. Edelman, S.: Representation and Recognition in Vision. MIT Press, Cambridge (1999)
2. Pękalska, E., Duin, R.: The Dissimilarity Representation for Pattern Recognition. In: Foundations and Applications. World Scientific, Singapore (2005)
3. Pękalska, E., Duin, R., Paclík, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39(2), 189–208 (2006)
4. Dubuisson, M., Jain, A.: Modified Hausdorff distance for object matching. In: Int. Conference on Pattern Recognition, vol. 1, pp. 566–568 (1994)
5. Goldfarb, L.: A new approach to pattern recognition. In: Kanal, L., Rosenfeld, A. (eds.) Progress in Pattern Recognition, vol. 2, pp. 241–402. Elsevier, Amsterdam (1985)
6. Goldfarb, L., Gay, D., Golubitsky, O., Korkin, D.: What is a structural representation? second version. Technical Report TR04-165, University of New Brunswick, Fredericton, Canada (2004)
7. Duin, R.P.W.: Structural class representation and pattern recognition by ets; a commentary. Technical report, Delft University of Technology, Pattern Recognition Laboratory (2006)
8. Pękalska, E., Harol, A., Duin, R., Spillmann, B., Bunke, H.: Non-Euclidean or non-metric measures can be informative. In: S+SSPR, pp. 871–880 (2006)
9. Duin, R.P.W., Pekalska, E., Harol, A., Lee, W.J., Bunke, H.: On euclidean corrections for non-euclidean dissimilarities. In: Structural, Syntactic, and Statistical Pattern Recognition, pp. 551–561 (2008)
10. Jacobs, D., Weinshall, D., Gdalyahu, Y.: Classification with Non-Metric Distances: Image Retrieval and Class Representation. IEEE TPAMI 22(6), 583–600 (2000)
11. Duin, R., de Ridder, D., Tax, D.: Experiments with object based discriminant functions; a featureless approach to pattern recognition. Pattern Recognition Letters 18(11-13), 1159–1166 (1997)
12. Graepel, T., Herbrich, R., Bollmann-Sdorra, P., Obermayer, K.: Classification on pairwise proximity data. In: Advances in Neural Information System Processing, vol. 11, pp. 438–444 (1999)
13. Pękalska, E., Paclík, P., Duin, R.: A Generalized Kernel Approach to Dissimilarity Based Classification. J. of Machine Learning Research 2(2), 175–211 (2002)
14. Bunke, H., Shearer, K.: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters 19(3-4), 255–259 (1998)
15. Bunke, H., Bühler, U.: Applications of approximate string matching to 2D shape recognition. Pattern recognition 26(12), 1797–1812 (1993)
16. Jain, A.K., Zongker, D.E.: Representation and recognition of handwritten digits using deformable templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(12), 1386–1391 (1997)
17. Kondor, R.I., Jebara, T.: A kernel between sets of vectors. In: Fawcett, T., Mishra, N. (eds.) ICML, pp. 361–368. AAAI Press, Menlo Park (2003)
18. Graepel, T., Herbrich, R., Schölkopf, B., Smola, A., Bartlett, P., Müller, K.R., Obermayer, K., Williamson, R.: Classification on proximity data with LP-machines. In: ICANN, pp. 304–309 (1999)