Dissimilarity-Based Multiple Instance Learning

Lauge Sørensen[1], Marco Loog[1,2], David M.J. Tax[2], Wan-Jui Lee[2], Marleen de Bruijne[1,3], and Robert P.W. Duin[2]

[1] The Image Group, Department of Computer Science, University of Copenhagen, Denmark
lauges@diku.dk
[2] Pattern Recognition Laboratory, Delft University of Technology, The Netherlands
[3] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands

Abstract. In this paper, we propose to solve multiple instance learning problems using a dissimilarity representation of the objects. Once the dissimilarity space has been constructed, the problem is turned into a standard supervised learning problem that can be solved with a general purpose supervised classifier. This approach is less restrictive than kernel-based approaches and therefore allows for the usage of a wider range of proximity measures. Two conceptually different types of dissimilarity measures are considered: one based on point set distance measures and one based on the earth movers distance between distributions of within- and between set point distances, thereby taking relations within and between sets into account. Experiments on five publicly available data sets show competitive performance in terms of classification accuracy compared to previously published results.

Key words: dissimilarity representation, multiple instance learning, bag dissimilarity measure

1 Introduction

In multiple instance learning (MIL), complex objects are represented by sets of "sub-objects" where only the sets have an associated label, not the sub-objects. Following MIL terminology, the sets are termed bags and the sub-objects are termed instances. This kind of problem might, e.g., arise in medical image classification where a subject is known to suffer from a certain disease, but it is not clear exactly which regions in the associated medical image that contain the corresponding pathology. In this case, local image patches are the instances, the whole image is the bag, and the label of the bag is either ill or healthy.

The traditional approach to solving MIL problems involves explicit learning of a decision boundary in instance space that separates the instances capturing the concept from the remaining instances [1, 2]. A bag is then classified based on whether it contains an instance falling in this area. An alternative instance space approach involves labeling all instances with the same label as the bag they belong to. The problem is then treated as a standard supervised learning problem where all instances are classified in instance space, ultimately disregarding the multiple instance aspect of the original problem, and a bag is classified by combining the individual instance classifications in that bag [3].

The above mentioned approaches treat instances in the same bag independently in the learning step thereby disregarding potentially useful information. In some MIL problems, instances from the same bag collectively constitute that bag and should as such all contribute to the classification of that bag. Several authors have looked into using this information by applying learning at bag level with kernel-based methods. To name a few: Andrews *et al.* reformulated a support vector machine (SVM) optimization problem to operate directly on MIL problems at bag level [4]. Gärtner *et al.*, Tao *et al.*, and Zhou *et al.* designed specialized kernels for MIL problems and used standard SVMs with these kernels [5–7]. Chen *et al.* represented bags in an n-dimensional space where each dimension was the similarity between one of the n instances in the training set and the closest instance in a bag. Then a 1-norm SVM was used to simultaneously select the relevant features, or instances, and train a bag classifier [8].

In this paper, we propose to use the dissimilarity representation approach to learning [9] for solving MIL problems at the bag level. Once the bag dissimilarity space has been constructed, the problem is turned into a standard supervised learning problem that can be solved with a general purpose supervised classifier. This is a proximity-based approach as are kernel-based methods, however, the dissimilarity representation approach does not require Mercer kernels as do kernel-based methods. A broader range of proximity measures, such as well known measures in pattern recognition like the Hausdorff distance and the single linkage distance, can therefore be used for solving MIL problems with this approach. We further propose, not only to consider all instances collectively in bag classification, but also to consider the relations among the instances within and between bags. This is similar in spirit to [7] where graphs capturing instance relations were constructed and used in a SVM with a graph kernel [5]. A novel non-Mercer bag dissimilarity measure that is based on the earth movers distance (EMD) between instance distance distributions is proposed for this purpose. Compared to the graph kernel approach used in [7], the proposed bag dissimilarity measure is less rigid since distributions of instance distances are considered instead.

Dissimilarity-based learning has previously been applied in MIL. Wang and Zucker applied the k nearest neighbor (kNN) classifier to MIL problems by using the Hausdorff distance between the instances in two bags as the distance between these bags [10]. They showed that this was not sufficient to get good performance on the classical MIL data sets MUSK1 and MUSK2 [1], due to noise in the presence of negative instances in the positive bags, and suggested two adaptations of kNN instead. A key observation is that kNN using Hausdorff distance between instances is working on dissimilarities between bags, and one way of arriving at a more global and robust decision rule when dissimilarities between objects are available is via a dissimilarity representation [9]. Building a global classifier like the Fisher linear discriminant classifier (Fisher) on such a representation leads to a global decision rule that uses a weighted combination of the dissimilarities to all training set objects in classification. This means better utilization of the available training data, with possibly increased accuracy and less sensitivity to noise.

The rest of the paper is organized as follows: Sections 2 and 3 briefly describe the MIL problem and the dissimilarity representation approach to learning. Section 4 presents two conceptually different types of dissimilarity measures between bags of instances. The first type is points set distance measures and the second type is based on EMD between distributions of instance distances within- and between bags. The proposed approach is evaluated by training and testing traditional supervised classifiers on dissimilarity representations of five publicly available MIL data sets. This is reported in Section 5. Finally, Section 6 provides a discussion and conclusions.

2 Multiple instance learning in short

In MIL [1], an object x_i is represented by a set, or bag, $B_i = \{\mathbf{x}_{ij}\}_{n_i}$ of n_i instances $\mathbf{x}_{ij}$, and a label $Y_i = \{+1, -1\}$ is associated with the entire bag. There are no labels y_{ij} associated directly with the instances, only indirectly via the label of the bag. This is different from standard supervised learning where objects are represented by a single instance, i.e., $B_i = \mathbf{x}_i$ and all instances therefore are directly labeled. The bag labels are interpreted in the following way in the original MIL formulation [1]: if $Y_i = -1$, then $\forall \mathbf{x}_{ij} \in B_i : y_{ij} = -1$. If $Y_i = +1$, then $\exists \mathbf{x}_{ij} \in B_i : y_{ij} = +1$. In other words, if a bag is labeled as positive, then at least one instance in that bag is a positive example of the underlying concept. This formulation can be relaxed to cope with a large and noisy set of instances by requiring that a positive bag contains a number or fraction of positive instances instead. In this work, we only consider two-class problems, but MIL can also be generalized to multi-class problems.

3 Dissimilarity representations in short

Objects x are traditionally represented by feature vectors in a feature vector space, and classifiers are built in this space. Alternatively, one can represent the objects by their pair-wise dissimilarities $d(x_i, x_j)$ and build classifiers on the obtained dissimilarity representation [9]. From the matrix of pair-wise object dissimilarities $D = [d(x_i, x_j)]_{n \times n}$ computed from a set of objects $\{x_1, \ldots, x_n\}$, there are different ways of arriving at a feature vector space where traditional vector space methods can be applied. In this work, we consider the dissimilarity space approach [9].

Given a training set $T = \{x_1, \ldots, x_n\}$, a subset $R = \{p_1, \ldots, p_k\} \subseteq T$ called the representation set containing prototype objects p_i is selected. An object x is represented with respect to R by the vector $D(x, R) = [d(x, p_1), \ldots, d(x, p_k)]$ of dissimilarities computed between x and the prototypes in R. This k-dimensional vector space based on R is called a dissimilarity space, and it is in this space that we propose to solve MIL problems at the bag level. In this work, we apply learning in the full dissimilarity space, i.e., $R = T$.

4 Bag dissimilarity space

The idea we propose is to map the bags into a dissimilarity space $D(\cdot, R = \{B_i\}_k)$. Here the bags are represented as single objects, positioned with respect to their dissimilarities to the prototype bags in R. In this space, the MIL problem can be considered as a standard supervised classification problem where each object $x_i = B_i$ has label Y_i and general purpose supervised classifiers can be directly applied. The separation of the bags in the obtained dissimilarity space is very much dependent on the choice of bag dissimilarity measure $d(B_i, B_j)$. In the following, we present two conceptually different types of dissimilarity measures for bags of instances.

4.1 Point set distance measures

The instances $\mathbf{x}$ reside in a common space and bags B can therefore be thought of as sets of objects in this space. In the case of vectorial instances, these objects are points in a vector space. This leads to the idea of computing dissimilarities between bags using point set distance measures. In this work, we experiment with the minimum distance

$$d_{min}(B_i, B_j) = \min_{p,q} ||\mathbf{x}_{ip} - \mathbf{x}_{jq}||_2 \tag{1}$$

and the Hausdorff distance

$$d_H(B_i, B_j) = \max\{d_{dir}(B_i, B_j), d_{dir}(B_j, B_i)\} \tag{2}$$

which is based on the directed distance $d_{dir}(B_i, B_j) = \max_p \min_q ||\mathbf{x}_{ip} - \mathbf{x}_{jq}||_2$. These point set distance measures were also used in a modified kNN classifier in [10].

Both point set distance measures (1) and (2) use the distance between two single instances in the end. These measures may therefore be sensitive to noisy instances, and they are in general insensitive to the number of positive instances in a positive bag. This may not be desirable when constructing a bag dissimilarity representation, and taking more information about the instances in a bag into account in the bag dissimilarity measure may lead to a better representation of the bags.

4.2 Measures based on between- and within bag instance distances

Zhou *et al.* conjectured that instances in a bag are rarely independently and identically distributed and that relations among the instances may convey important information when applying learning at bag level [7]. In a similar spirit, we propose two bag dissimilarity measures that take relations among instances into account, or more precisely, the distribution of instance distances within a bag and between bags. It is assumed that the instances in the two bag classes follow distributions in the common instance space that are very similar, with the slight difference that one distribution contains additional modes capturing the concept(s). This situation is illustrated, for one additional mode, to the left in Figure 1. This could, e.g., be the situation in a MIL problem in medical image

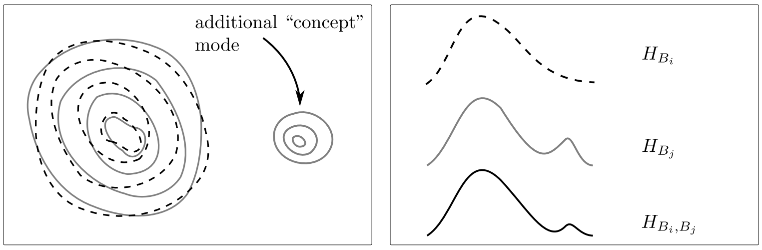

Fig. 1. Left: Illustration of two similar bag class distributions where one of the distributions, typically the positive bag distribution, has an extra mode corresponding to the positive instances. **Right**: Distributions of instance distances, from top to bottom: within bag instance distances in a bag from the class with no additional mode, typically the negative class; within bag instance distances in a bag from the class with an additional mode, typically the positive class. Notice the extra "bump" in the distribution; instance distances between two bags, one from each class.

classification where the positive medical images contain lesions surrounded by healthy tissue whereas the negative images only contain healthy tissue. The additional mode in one of the bag class distributions gives rise to an extra "bump" in the distribution of instance distances within bags from that class, compared to bags from the other class, as illustrated to the right in Figure 1. Further, the bump can also be seen in the histogram of instance distances computed between bags from the two classes.

We propose to use the within bag instance distance histograms H_{B_i} and H_{B_j}, computed from bag B_i and B_j, respectively, and the between bag instance distance histogram H_{B_i,B_j}, computed between bag B_i and B_j. The bag dissimilarity measure is then computed as the pair-wise histogram dissimilarity $d_{i,ij} = d(H_{B_i}, H_{B_i,B_j})$. $d_{i,ij}$ can be seen as the directed dissimilarity from B_i to B_j. The maximum and the mean of the directed dissimilarities from each of the two bags are proposed as two symmetric dissimilarity measures for bags

$$d_{BWmax}(B_i, B_j) = \max\{d_{i,ij}, d_{j,ij}\} \tag{3}$$

and

$$d_{BWmean}(B_i, B_j) = \frac{1}{2}(d_{i,ij} + d_{j,ij}). \tag{4}$$

The histogram dissimilarities are computed using EMD [11] between the normalized empirical distributions. For one-dimensional histograms $H = [h_1, \ldots, h_n]^T$ and $K = [k_1, \ldots, k_n]^T$ of equal number of bins n and equal mass, EMD can be computed as the L1-norm between the cumulative histograms of H and K: $d_{EMD}(H, K) = \sum_{i=1}^{n} | \sum_{j \leq i} h_j - \sum_{j \leq i} k_j |$.

4.3 A second dissimilarity space

Initial experiments showed that linear classifiers performed poorly when built on the obtained bag dissimilarity representations whereas the nearest neighbor classifier (1NN) performed quite well. This indicates that the bags are separated in the obtained dissimilarity representations, but that the decision boundaries between the positive bags and the negative bags are complicated and non-linear, and/or that the class distributions are multi-modal in these new representations. An extra preprocessing step is therefore done before applying linear classifiers. From $D(\cdot, X)$ computed on the full data set X, a new dissimilarity representation $D2$ is constructed such that $D2(x_i, x_j) = ||D(x_i, X) - D(x_j, X)||_2, \forall_{x_i, x_j} \in X$. The linear classifiers are built on this representation. This is a transductive learning approach since all objects are used to construct the representation D2. It is, however, important to note that the labels of the objects are not considered in this construction. Tao *et al.* also used transductive learning to solve MIL problems [6].

5 Experiments and results

The proposed approach is evaluated on the two standard data sets in MIL, namely MUSK1 and MUSK2 originally used in [1], and on three recently published image retrieval data sets [4].

5.1 MUSK1 and MUSK2

These are the standard MIL data sets, and they consist of descriptions of aromatic molecules that have been labeled according to whether they smell "musky" or not. A bag represents a molecule, and the instances in a bag are low energy shapes of the molecule described by 166-dimensional feature vectors. The MUSK1 data set comprises 47 positive bags and 45 negative bags, and each bag is represented by 2 to 40 instances. The MUSK2 data set comprises 39 positive bags and 63 negative bags, and each bag is represented by 1 to 1044 instances. The data was obtained from the UCI Machine Learning Repository [12], and we refer to this source as well as to [1] for further information about the data.

5.2 Image retrieval

This data comprises three data sets that are subsets of the Corel data set. Each data set consists of 100 positive bags, or example images; elephant, fox, or tiger, and 100 negative bags, or background images, which are randomly drawn from a pool of photos of other animals. Each image is represented by 2-13 instances (apart from a single image in the tiger data set that is represented by a single instance), which are 230-dimensional feature vectors describing the color, texture and shape in subsegments of the image. The data was obtained from the homepage[4] associated with [4] and we refer to these sources for further information about the data.

5.3 Evaluation

Classifier	Bag dissimilarity measure	MUSK1	MUSK2
1NN (on D)	d_{min} (1)	90.2 / **91.3**	86.9 / 84.6
	d_H (2)	88.0 / 87.9	86.1 / 82.5
	d_{BWmax} (3)	85.8 / 86.9	82.8 / 77.7
	d_{BWmean} (4)	89.1 / 91.2	85.3 / 80.7
SVM (on $D2$)	d_{min} (1)	90.0 / 90.1	92.2 / 87.5
	d_H (2)	88.0 / 88.0	91.2 / 85.5
	d_{BWmax} (3)	89.1 / 89.0	82.2 / 88.3
	d_{BWmean} (4)	**91.2** / 89.0	85.3 / 85.0
Fisher (on $D2$)	d_{min} (1)	90.1 / 90.1	**93.5** / **92.7**
	d_H (2)	88.0 / 86.9	90.3 / 88.2
	d_{BWmax} (3)	90.1 / 87.9	87.7 / 87.4
	d_{BWmean} (4)	**91.2** / 91.2	89.8 / 90.3
citation-kNN [10]		**92.4** / -	86.3 / -
iterated discrim APR [1]		- / **92.4**	- / 89.2
diverse density [2]		- / 88.9	- / 82.5
mi-SVM [4]		- / 87.4	- / 83.6
MI-SVM [4]		- / 77.9	- / 84.3
SVM polynomial minimax kernel [5]		**92.4** / -	86.3 / -
SVM MI kernel [5]		87.0 / -	92.2 / -
MILES [8]		86.3 / 87.0	87.7 / **93.1**
$k_{\wedge}$ *emph* transduction [6]		- / 91.2	- / 90.3
$k_{\wedge/\vee}$ *emph* transduction [6]		- / 90.2	- / 92.2
MIGraph [7]		- / 90.0	- / 90.0
miGraph [7]		- / 88.9	- / 90.3

Table 1. Classification accuracy on the MUSK1 and MUSK2 data set, reported as leave-one-out / ten-fold cross-validation. Accuracies reported in the literature are shown in the bottom part of the table. Cases in the literature where the classification accuracy is not reported using leave-one-out or ten-fold cross-validation are marked with "-". The highest accuracy among the dissimilarity representation-based classifiers as well as the highest accuracy in general is marked in boldface in each column.

[4] http://www.cs.columbia.edu/~andrews/mil/datasets.html

The proposed dissimilarity representations are evaluated by training and testing three supervised classifiers on the bags in the given dissimilarity space. These classifiers are: 1NN; SVM with a linear kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ where $\mathbf{x}_i = D2(B_i, X)$ and trade-off parameter $C = 1$; and Fisher. For 1NN and Fisher we use the pattern recognition toolbox PRTools [13], and for SVM we use LIBSVM [14].

Classification accuracies are estimated using leave-one-out and 10-fold cross-validation, since these are commonly used performance measures in the MIL literature [1, 2, 10, 4, 5, 3, 7]. 10-fold cross-validation is sometimes performed once and sometimes the average of a repeated number of 10-fold cross-validation procedures is reported. Here we perform one 10-fold cross-validation. The results are presented in Table 1 and Table 2 where also previously published results are reported.

Classifier	Bag dissimilarity measure	elephant	fox	tiger
1NN (on D)	d_{min} (1)	78.0 / 78.0	60.0 / 59.5	77.0 / 74.0
	d_H (2)	70.0 / 69.5	52.0 / 50.0	67.0 / 64.5
	d_{BWmax} (3)	75.0 / 77.5	57.5 / 57.0	68.0 / 66.0
	d_{BWmean} (4)	80.0 / 79.0	59.5 / 59.0	70.5 / 71.5
SVM (on $D2$)	d_{min} (1)	85.5 / 83.5	**67.5** / 65.0	77.5 / 78.0
	d_H (2)	84.0 / 84.5	37.5 / 49.0	73.5 / 73.5
	d_{BWmax} (3)	**89.0** / **89.0**	64.5 / 56.0	69.5 / 62.0
	d_{BWmean} (4)	87.0 / 87.0	62.5 / 58.5	78.0 / 76.5
Fisher (on $D2$)	d_{min} (1)	86.0 / 84.5	66.0 / **66.0**	78.5 / 78.0
	d_H (2)	84.5 / 85.0	59.0 / 59.0	73.5 / 72.0
	d_{BWmax} (3)	88.5 / 88.5	66.5 / 63.0	81.0 / 78.5
	d_{BWmean} (4)	**89.0** / 88.5	64.5 / 64.0	**81.5** / **79.5**
mi-SVM [4]		- / 82.2	- / 58.2	- / 78.9
MI-SVM [4]		- / 81.4	- / 59.4	- / 84.0
MIGraph [7]		- / 85.1	- / 61.2	- / 81.9
miGraph [7]		- / 86.8	- / 61.6	- / **86.0**

Table 2. Classification accuracy on the image retrieval data. See the caption of Table 1 for further details.

The classification accuracies of 1NN are quite close to the ones previously reported in the literature. The high 1NN classification accuracies on the MUSK1 and MUSK2 data set indicate that the bags are well separated in the obtained bag dissimilarity space defined by D. Fisher performs poorly when built on D with an average classification accuracy of 62.1% whereas SVM performs decent when built on D with an average classification accuracy of 78.4%. However, building them on a second dissimilarity representation $D2$ constructed from D, as described in Section 4.3, improves performance considerably for Fisher with an average absolute increase of 19.3% and slightly for SVM with an average absolute increase of 4%. 1NN performs slightly worse when applied to $D2$ compared to D, and the numbers reported in Table 1 and Table 2 for 1NN are therefore based on D. SVM and Fisher generally perform better than 1NN. We also tried kNN with k optimized using cross-validation on the training set in each fold which achieved similar performance to 1NN.

Across all five data sets, SVM and Fisher built on dissimilarity representations show excellent performance. On the MUSK1 and MUSK2 data set, the classifiers achieve accuracies close to the best reported accuracies in the literature. On the image retrieval data sets, SVM with a linear kernel, as well as Fisher, perform better than the SVM's adapted to MIL problems [4] in two out of three data sets. This indicates that taking instance relations into account is beneficial in this kind of problems, as is also seen in [7].

6 Discussions and conclusions

The linear classifiers built on the proposed dissimilarity representations performed better than the best results in the MIL literature in some cases, and in the remaining cases close to the best published results [1, 2, 10, 4, 5, 8, 6, 7]. It should be noted that the classifiers were applied "off the shelf" and that, e.g., the trade-off parameter C in SVM was not tuned by cross-validation but fixed to 1. Also, the classifiers were trained and tested in dissimilarity spaces of dimension equal to the number of training samples. This is no problem for SVM. For Fisher, the pseudo-inverse was used. It may be possible to obtain even better results than the ones reported in Table 1 and Table 2 by proper regularization or by reducing the dimensionality of the dissimilarity space, e.g., by prototype selection [15].

SVM shows worse than random performance on some of the image retrieval data sets, in particular when built on the dissimilarity representation obtained using the Hausdorff distance, d_H, on the fox data set. This could be caused by a strong class overlap in the dissimilarity space. This is also indicated by the fact that both 1NN and Fisher perform worse on this representation compared to the other representations.

The minimum point set distance, d_{min}, works well as bag dissimilarity measure. Similar results were reported in [10]. This is somewhat surprising since classes are expected to be overlapping in MIL due to positive bags also containing negative instances. The explanation is that the distribution of the positive instances is more dense compared to the negative instances in the used data sets, and therefore a bag containing at least one positive instance is more likely to be close to another bag containing at least one positive instance than to a bag containing only negative instances.

To conclude, we have shown that the dissimilarity representation approach can be used to solve MIL problems. Global decision rules in the form of general purpose supervised linear classifiers built in a bag dissimilarity space achieves excellent classification accuracies on publicly available MIL data sets. The approach is general, and we see this as a promising direction that allows for using a wider range of proximity measures between bags in solving MIL problems compared to the popular kernel-based approaches. Further, there are indications that taking relations among instances into account improves the performance on certain MIL problems, such as the image retrieval problems.

Acknowledgements. This work is partly funded by the Danish Council for Strategic Research (NABIIT); the Netherlands Organisation for Scientific Research (NWO); AstraZeneca, Lund, Sweden; the FET programme within the EU FP7, under the SIMBAD project (contract 213250); and the ICT-2007.2.2 programme within the EU FP7, under the SSPNet project (contract 231287).

References

1. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. **89**(1-2) (1997) 31–71
2. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In Jordan, M.I., Kearns, M.J., Solla, S.A., eds.: NIPS, The MIT Press (1997)
3. Cannon, A., Hush, D.: Multiple instance learning using simple classifiers. In: ICMLA. (2004) 123–128
4. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In Becker, S., Thrun, S., Obermayer, K., eds.: NIPS, MIT Press (2002) 561–568
5. Gärtner, T., Flach, P.A., Kowalczyk, A., Smola, A.J.: Multi-instance kernels. In Sammut, C., Hoffmann, A.G., eds.: ICML, Morgan Kaufmann (2002) 179–186
6. Tao, Q., Scott, S.D., Vinodchandran, N.V., Osugi, T.T., Mueller, B.: Kernels for generalized multiple-instance learning. IEEE Trans. Pattern Anal. Machine Intell. **30**(12) (2008) 2084–2098
7. Zhou, Z.H., Sun, Y.Y., Li, Y.F.: Multi-instance learning by treating instances as non-i.i.d. samples. In Danyluk, A.P., Bottou, L., Littman, M.L., eds.: ICML. Volume 382 of ACM International Conference Proceeding Series., ACM (2009) 1249–1256
8. Chen, Y., Bi, J., Wang, J.Z.: MILES: Multiple-instance learning via embedded instance selection. IEEE Trans. Pattern Anal. Machine Intell. **28**(12) (2006) 1931–1947
9. Pekalska, E., Duin, R.P.W.: Dissimilarity representations allow for building good classifiers. Pattern Recog. Lett. **23**(8) (2002) 943–956
10. Wang, J., Zucker, J.D.: Solving the multiple-instance problem: A lazy learning approach. In Langley, P., ed.: ICML, Morgan Kaufmann (2000) 1119–1126
11. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40**(2) (2000) 99–121
12. Asuncion, A., Newman, D.: UCI machine learning repository (2007)
13. Duin, R.P.W., Juszczak, P., Paclík, P., Pekalska, E., de Ridder, D., Tax, D.M.J.: PRTools, a Matlab toolbox for pattern recognition, version 4.0 (January 2004)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.
15. Pekalska, E., Duin, R.P.W., Paclík, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition **39**(2) (2006) 189–208