Pattern Recognition 48 (2015) 264–275

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Multiple instance learning with bag dissimilarities

Veronika Cheplygina [a,*], David M.J. Tax [a], Marco Loog [a,b]

[a] Pattern Recognition Laboratory, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands
[b] Image Group, University of Copenhagen, Universitetsparken 5, DK-2100, Denmark

ARTICLE INFO

Article history:
Received 26 June 2013
Received in revised form
26 June 2014
Accepted 18 July 2014
Available online 5 August 2014

Keywords:
Multiple instance learning
Dissimilarity representation
Point set distance
Image classification
Drug activity prediction
Text categorization

ABSTRACT

Multiple instance learning (MIL) is concerned with learning from sets (bags) of objects (instances), where the individual instance labels are ambiguous. In this setting, supervised learning cannot be applied directly. Often, specialized MIL methods learn by making additional assumptions about the relationship of the bag labels and instance labels. Such assumptions may fit a particular dataset, but do not generalize to the whole range of MIL problems. Other MIL methods shift the focus of assumptions from the labels to the overall (dis)similarity of bags, and therefore learn from bags directly. We propose to represent each bag by a vector of its dissimilarities to other bags in the training set, and treat these dissimilarities as a feature representation. We show several alternatives to define a dissimilarity between bags and discuss which definitions are more suitable for particular MIL problems. The experimental results show that the proposed approach is computationally inexpensive, yet very competitive with state-of-the-art algorithms on a wide range of MIL datasets.

1. Introduction

Many pattern recognition problems deal with complex objects that consist of parts: images displaying several objects, documents with different paragraphs, proteins with various amino acid subsequences. The success of supervised learning techniques forces such complex objects to be represented as a single feature vector. However, this reduction may cause important differences between objects to be lost, degrading classification performance. Rather than representing such a complex object by a single feature vector, we can represent it by a set of feature vectors, as in multiple instance, or multi-instance learning (MIL) [1]. For example, an image can be represented as a bag of segments, where each segment is represented by its own feature vector. This is a more flexible representation that potentially can preserve more information than a single feature vector representation.

In MIL terminology, an object is called a *bag* and its feature vectors are called *instances*. MIL problems are often considered to be two-class problems, i.e., a bag can belong either to the positive or the negative class. During training, the bag labels are available, but the labels of the instances are unknown. Often assumptions are made about the instance labels and their relationship with the bag labels. The standard assumption is that positive bags contain at least one positive or *concept* instance, whereas negative bags contain only negative, *background* instances [1,2]. An image labeled as "tiger" would therefore contain a tiger in at least one of its segments, whereas images with other labels would not depict any tigers. Many MIL approaches therefore focus on using the labeled bags to model the concept region in the instance space. To classify a previously unseen bag, the instances are labeled according to the best candidate model for the concept, and the bag label is then obtained from these instance labels.

It has been pointed out [3] that although for many problems the bag representation is useful, the assumptions on the bag and instance labels typically do not fit the application. For instance, for an image of the "desert" category, it would be wrong to say that "sand" is the concept, if images of the "beach" category are also present. Therefore, methods in which the standard assumption is relaxed have emerged. In [4] an adaptive parameter is used to determine the fraction of concept instances in positive bags. Generalized MIL [5,6] examines the idea that there could be an arbitrary number of concepts, where each concept has a rule of how many (just one, several or a fraction) positive instances are needed to satisfy each concept. A review of MIL assumptions can be found in [7].

This line of thought can be extended further to cases where it is difficult to define a concept or concepts, and where most, if not all, instances, contribute to the bag label. The implicit assumption is that bag labels can be related to distances between bags, or to distances between bags and instances. Such approaches have used bag distances [8], bag kernels [9], instance kernels [3] or dissimilarities between bags [10–12].

* Corresponding author. Tel.: +31 152787243.
E-mail addresses: v.cheplygina@tudelft.nl (V. Cheplygina), d.m.j.tax@tudelft.nl (D.M.J. Tax), m.loog@tudelft.nl (M. Loog).

http://dx.doi.org/10.1016/j.patcog.2014.07.022
0031-3203/

Bag-based approaches are attractive because they transform the MIL dataset back to a standard feature vector representation such that regular supervised classifiers can be used. Unfortunately, some of the representational power of MIL can be lost when converting a bag to a single feature vector of (dis)similarities. It has indeed been pointed out that the definition of distance or similarity can influence how well the representation is suited for one or more concepts [7]. The question is how to do this in a way that still preserves information about the class differences. Furthermore, competing approaches offer a variety of definitions of (dis)similarity, and it is not always clear which definition should be preferred when a new type of MIL problem presents itself.

In this paper we propose a general MIL dissimilarity approach called MInD (*M*ultiple *In*stance *D*issimilarity). We discuss several ways in which dissimilarities between bags can be defined, show which assumptions these definitions are implicitly making, and hence which definitions are suitable for different types of MIL problems. We have collected several examples of such problems in a single repository online.[1] Furthermore, we discuss why the dissimilarity space is an attractive approach for MIL in general. An important advantage is that there are no restrictions on the dissimilarity measure (such as metricity or positive-definiteness). This allows the use of expert-defined dissimilarities which often violate these mathematical restrictions. Similarly, there is no restriction on the classifier used in the dissimilarity space, which is attractive for potential end-users faced with MIL problems, and who already have experience with a certain supervised classifier. Lastly, with a suitable choice of dissimilarity and classifier, the approach is very computationally efficient, yet still provides interpretable state-of-the-art results on many datasets. For example, the average minimum distance between bags with a logistic classifier achieves very good performances, is easy to implement, and the classifier decisions can be explained in terms of dissimilarities to the prototypes.

After a review of MIL approaches in Section 2, we propose MInD in Section 3. In Section 4, we show some examples of MIL problems and demonstrate which dissimilarities are most suitable in each case. We then compare results to a range of MIL methods in Section 5, and discuss practical issues of dissimilarities and other bag-level methods in Section 6. A conclusion is given in Section 7.

2. Review of MIL approaches

In multiple instance learning (MIL), an object is represented by a bag $B_i = \{\mathbf{x}_{ik} | k = 1, \ldots, n_i\} \subset \mathbb{R}^d$ of n_i feature vectors or instances. The training set $\mathcal{T} = \{(B_i, y_i) | i = 1, \ldots N\}$ consists of positive ($y_i = +1$) and negative ($y_i = -1$) bags. We will also denote such bags by B_i^+ and B_i^-. The standard assumption for MIL is that there are instance labels which relate to the bag labels as follows: a bag is positive if and only if it contains at least one positive, or *concept* instance [1].

Under this standard assumption, the strategy has been to model the concept: a region in the feature space which contains at least one instance from each positive bag, but no instances from negative bags. The original class of MIL methods used an axis-parallel hyper-rectangle (APR) [1] as a model for the concept, and several search strategies involving such APRs have been proposed.

Diverse Density [2] is another approach for finding the concept in instance space. For a given point t in this space, a measure $DD(t)$ is defined as the ratio between the number of positive bags which have instances near t, and the distance of the negative instances to t. The point of maximum Diverse Density should therefore correspond to the target concept. The maximization problem does not have a closed form solution and gradient ascent is used to find the maximum. The search may therefore converge to a local optimum, and several restarts are needed to find the best solution.

EM-DD [13] is an expectation-maximization algorithm that refines Diverse Density. The instance labels are modeled by hidden variables. After an initial guess for the concept t, the expectation step selects the most positive instance from each bag according to t. The maximization step then finds a new concept t' by maximizing DD on the selected, most positive instances. The steps are repeated until the algorithm converges.

Furthermore, several regular supervised classifiers have been extended to work in the MIL setting. One example is mi-SVM [14], an extension of support vector machines which attempts to find hidden labels of the instances under constraints posed by the bag labels. Another example is MILBoost [15], where the instances are reweighted in each of the boosting rounds. The bag labels are decided by applying a noisy OR [2] rule to the instance labels, which reflects the standard assumption.

It has been recognized that the standard assumption might be too strict for certain types of MIL problems. Therefore, relaxed assumptions have emerged [5,6], where a fraction or a particular number of positive instances are needed to satisfy a concept, and where multiple concept regions are considered.

A similar notion is used in MILES [3], where multiple concepts, as well as the so-called negative concepts (concepts that only negative bags have) are allowed. All of the instances in the training set are used as candidate concept targets, and each bag is represented by its similarities to these instances. A sparse 1-norm SVM is then used to simultaneously maximize the bag margin, and select the most discriminative similarities, i.e., instances that are identified as positive or negative concepts.

A step further are methods that do not make explicit assumptions about the instances or the concepts, but only assume that bags of the same class are somehow similar to each other, and then learn from distances or similarities between bags. Such methods include Citation-*k* NN [8], which is based on the Hausdorff distance between bags, bag kernels [9] and bag dissimilarities [10,16]. In [9], a bag kernel is defined either as a sum of the instance kernels, or as a standard (linear or RBF kernel) on a transformed, single instance representation of the bag. One example is the Minimax representation, where each bag is represented by the minimum and maximum feature values of its instances.

Last but not least, a way to learn in MIL problems is to propagate the bag labels to the instances, and use supervised learners on these propagated labels. We call this approach SimpleMIL. To obtain a bag label from predicted instance labels, the instance labels have to be combined. Here, the noisy OR rule or other combining methods can be used [17,18]. It has been demonstrated that supervised methods can be quite effective in dealing with MIL problems [19].

All MIL methods can be more globally summarized by the representation that they use: the standard instance-vector-based representation, a bag dissimilairity representation and a bag-instance representation (see Fig. 1). The first representation is the standard representation of MIL, where each bag consists of several instances, and the dimensionality is equal to the dimensionality of the instance space. In this example, there are two bags which are represented in a 2D feature space. This is the representation used by EM-DD, mi-SVM, MILBoost and SimpleMIL. The representation on the top right is the bag representation, used by Citation-*k* NN, bag kernels and our bag dissimilarity approach. The representation in the bottom is the instance representation, used by MILES. In the latter two representations, regular supervised

[1] http://www.miproblems.org