Pattern Recognition 58 (2016) 83–99

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust multiple-instance learning ensembles using random subspace instance selection

Marc-André Carbonneau [a,b,*], Eric Granger [b], Alexandre J. Raymond [a], Ghyslain Gagnon [a]

[a] Laboratoire de communications et d'intégration de la microélectronique (LACIME), École de technologie supérieure, Université du Québec, 1100, rue Notre-Dame Ouest, Montréal (Qc), Canada H3C 1K3

[b] Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA), École de technologie supérieure, Université du Québec, 1100, rue Notre-Dame Ouest, Montréal (Qc), Canada H3C 1K3

ARTICLE INFO

Article history:
Received 10 August 2015
Received in revised form
3 February 2016
Accepted 11 March 2016
Available online 14 April 2016

Keywords:
Multiple-instance learning
Random subspace methods
Classifier ensembles
Instance selection
Weakly supervised learning
Classification
MIL

ABSTRACT

Many real-world pattern recognition problems can be modeled using multiple-instance learning (MIL), where instances are grouped into bags, and each bag is assigned a label. State-of-the-art MIL methods provide a high level of performance when strong assumptions are made regarding the underlying data distributions, and the proportion of positive to negative instances in positive bags. In this paper, a new method called Random Subspace Instance Selection (RSIS) is proposed for the robust design of MIL ensembles without any prior assumptions on the data structure and the proportion of instances in bags. First, instance selection probabilities are computed based on training data clustered in random subspaces. A pool of classifiers is then generated using the training subsets created with these selection probabilities. By using RSIS, MIL ensembles are more robust to many data distributions and noise, and are not adversely affected by the proportion of positive instances in positive bags because training instances are repeatedly selected in a probabilistic manner. Moreover, RSIS also allows the identification of positive instances on an individual basis, as required in many practical applications. Results obtained with several real-world and synthetic databases show the robustness of MIL ensembles designed with the proposed RSIS method over a range of witness rates, noisy features and data distributions compared to reference methods in the literature.

1. Introduction

Multiple-instance learning (MIL) is a form of weakly-supervised learning [1], where data *instances* are grouped into *bags*. A label is not provided for each instance, but for a whole bag. Typically, a negative bag contains only negative instances, while positive bags contain instances from both classes [2].

Since the first formulations of the MIL problem [2,3] many solutions have been proposed. In many cases, MIL algorithms were developed with a specific application in mind. For instance, Dietrich [2] proposed Axis Parallel Rectangle (APR) to solve a molecule classification problem. Later, many methods were proposed to solve image categorization [4–8], web mining [9,10], object and face detection [11–15] and tracking [16] problems. While they can achieve a high level of performance in their respective application domains, many of these methods are less efficient over a wide variety of data distributions and pattern classification problems.

For instance, many methods rely on the assumption that the proportion of positive instances in positive bags, hereafter called *witness rate*, is high. Sometimes, these methods implicitly assume that all instances in a positive bag are positive. This is the case for methods such as APR [2], Citation-kNN [17] and diverse density-based (DD) methods [5,6,18,19]. This assumption is also made in the initialization of the optimization process in mi-SVM and MI-SVM [4]. Other methods assume a high witness rate by representing bags as the average of the instances it contains, as in MI-Kernel [20] and MIBoosting [21]. The performance of all these methods decreases when the high witness rate assumption is not verified, which limits the applicability of MIL methods to many problems. For instance, until recently, object identification systems were limited to problems where instances represent slight translational and scale uncertainties around localization bounding boxes [15].

* Corresponding author.
E-mail addresses: marcandre.carbonneau@gmail.com (M.-A. Carbonneau), eric.granger@etsmtl.ca (E. Granger), alexandre.raymond@lacime.etsmtl.ca (A.J. Raymond), ghyslain.gagnon@etsmtl.ca (G. Gagnon).

http://dx.doi.org/10.1016/j.patcog.2016.03.035
0031-3203/

To deal with lower witness rates, Gehler and Chapelle [22] applied deterministic annealing to an SVM-based MIL algorithm. Bunescu and Mooney [23] enforced the constraint that positive bags contain at least one positive instance in their SVM formulation. Both obtained good results with lower witness rates, but observed performance degradation with higher witness rates. SVR-SVM [24] and the γ-rule [25] have been proposed to overcome these problems by estimating the witness rate and then using it as a system parameter. These techniques provide a high level of performance over a range of high and low witness rates, yet, the witness rate is assumed to be constant across all bags. This assumption proves to be problematic in some applications, such as image categorization [26], where images are segmented and features are extracted from the different segments [4,5]. The image corresponds to a bag, while each segment is an instance. Depending on the visual complexity of the image, a different proportion of target and non-target segments will be obtained. Therefore, the witness rate of a bag depends on the image content, and is likely to vary from one bag to another.

Another challenge of MIL problems is the fact that the shape of positive and negative distributions affect the performance of some algorithms. For instance, some methods such as APR [2] are not designed to deal with multi-modal distributions where instances are grouped in distinct clusters. Methods based on DD [5,6,18,19] assume that positive instances form a compact cluster [7]. In MILIS [7], the negative distribution is modeled with Gaussian kernels, which can be difficult when the quantity of data available is limited. On the other hand, in Citation-kNN [17] the presence of compact data cluster in the negative distribution increases the probability of misclassification.

Finally, some methods classify bags as a whole instead of trying to label each instance individually. Some of these methods [17,20,27,28] use different types of bag distance measure, while others embed bags using distance to a set of prototypes [6,7,5], vocabulary [29] and sparse coding [30]. Bag-level classification approaches cannot identify instances individually, which is necessary in certain applications such as object detection and tracking in images or videos. Moreover, by considering bags as a whole, the performance of these methods often decreases in problems where the witness rate is low.

To address these limitations, this paper proposes a new ensemble-based method for MIL called Random Subspace Instance Selection (RSIS). Classifier ensembles are generally known to provide accurate and robust classification systems when data is limited [31]. The key feature of RSIS is that it constructs classifier ensembles based on a probabilistic identification of positive instances. The proposed method allows to classify instances individually and does not rely on a specific witness rate or specific type of data distribution. It can therefore be applied in a wide variety of context.

In the proposed method, the training data is projected onto several random subspaces before being clustered. The proportion of instances from positive and negative bags is computed for every cluster. Based on these bag proportions, a *positivity score* is computed for every instance in the data set. These scores are later converted into selection probabilities, and used to select diverse training sets to generate base classifiers in the ensemble. The general intuition for RSIS is that it is easier to identify positive instance clusters while only considering a discriminant subset of features. The optimal feature subset to represent a given concept is unknown, and may vary from one concept to another. However, if a data set is projected into all possible subspaces, instances from the same concept are more likely to be grouped together than with the other instances.

The RSIS method allows to design MIL ensembles that are robust to various witness rate, because each time one of the classifiers in the ensemble is trained, only one instance is used from each bag. The instances are drawn based on their probability of being positive. If the witness rate is low and only one instance is likely to be positive, this instance will be the only one selected. In contrast, if many instances appear to be positive, each instance will have a similar probability of being selected, and thus being used as a training instance in one or another classifier. Since selection probabilities are computed for each bag separately, the witness rate does not have to be constant across all bags. Moreover, by clustering the data in many different subspaces, RSIS can inherently uncovers multiple underlying concepts in the data distributions. This makes the algorithm resistant to multi-modal distributions of various shapes, and robust to noisy or irrelevant features.

In this paper, the performance of MIL ensembles designed using RSIS is compared to several methods in the literature using benchmark data sets. Further experiments are performed on synthetic data sets to study the algorithm's tolerance to various multi-modal distributions, witness rate and irrelevant features. Five well-known baseline methods, APR [2], Citation-kNN [17], mi-SVM [4], AL-SVM [22] and CCE [32] are also used for comparison. Finally, the sensitivity of the proposed approach to internal parameters is also characterized experimentally, and some general guidelines for parameter selection are provided.

The remainder of this paper is organized as follows. The MIL problem is formalized and state-of-the-art techniques are reviewed in Section 2. Then, in Section 3, the proposed RSIS algorithm is described. Section 4 presents the experimental methodology. Section 5 presents robustness experiments on synthetic data, while Sections 6 and 7 present experimental results on benchmark data sets, and experiments on parameter sensitivity respectively. Time complexity is discussed in Section 8.

2. Multiple instance learning

Let $\mathcal{B} = \left\{B^1, B^2, \ldots, B^Z\right\}$ be a set composed of Z bags.[1] Each bag B^i corresponds to a positive or negative label $L^i \in \{-1, +1\}$ in the set $\mathcal{L} = \left\{L^1, L^2, \ldots, L^Z\right\}$, and contains N^i feature vectors: $B^i = \left\{\mathbf{x}^i_1, \mathbf{x}^i_2, \ldots, \mathbf{x}^i_{N^i}\right\}$ where $\mathbf{x}^i_j = (x^i_{j1}, x^i_{j2}, \ldots, x^i_{jd}) \in \mathbb{R}^d$. Each of these feature vector instances corresponds to a positive or negative label in the set $Y^i = \left\{y^i_1, y^i_2, \ldots, y^i_{N^i}\right\}$, where $y^i_j \in \{-1, +1\}$. Instance labels are unknown in positive bags, but are assumed negative in negative bags. A bag is labeled positive if at least one instance contained in the bag is labeled positive [2]:

$$L^i = \begin{cases} +1 & \text{if } \exists y \in Y^i : y^i_j = +1; \\ -1 & \text{if } \forall y \in Y^i : y^i_j = -1. \end{cases} \quad (1)$$

Many methods have been proposed over the years to address MIL problems in a variety of domains. An overview of these methods and a review of the MIL assumptions can be found in recent surveys by Amores [33] and Foulds and Frank [34]. In the taxonomy proposed by Amores, [33] MIL methods are divided in three categories, based on how bags are represented. A first corpus of methods operates at the instance level. Each instance is classified individually, and scores are aggregated to label bags. The two other types of method operate on the bag level. In one case, bags are mapped to a vector representation, which reformulate the MIL problem as a standard supervised classification problem, while in

[1] Throughout this paper, upper indexes are used to denote bags, while lower indexes designate instances. For the sake of clarity, when unnecessary, these indexes are omitted.