Information Sciences 367–368 (2016) 783–808

Contents lists available at ScienceDirect

Information Sciences

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

Designing bag-level multiple-instance feature-weighting algorithms based on the large margin principle

Jing Chai [a,*], Zehua Chen [a], Hongtao Chen [a], Xinghao Ding [b]

[a] College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, PR China
[b] School of Information Science and Engineering, Xiamen University, Xiamen 361005, PR China

ARTICLE INFO

Article history:
Received 29 December 2014
Revised 7 July 2016
Accepted 10 July 2016
Available online 12 July 2016

Keywords:
Large margin
Feature weighting
Multiple-instance learning
Bag level

ABSTRACT

In multiple-instance learning (MIL), class labels are attached to bags instead of instances, and the goal is to predict the class labels of unseen bags. Existing MIL algorithms generally fall into two types: those designed at the bag level and those designed at the instance level. In this paper, we aim to employ bags directly as learning objects and design multiple-instance feature-weighting algorithms at the bag level. In particular, we initially provide a brief introduction of the bag-level large margin feature-weighting framework and then adopt the three bag-level distances minimal Hausdorff (minH), class-to-bag (C2B) and bag-to-bag (B2B) as examples to design the corresponding bag-level feature-weighting algorithms. Experiments conducted on synthetic and real-world datasets empirically demonstrate the effectiveness of our work in improving MIL performances.

1. Introduction

The main difference between traditional supervised learning (SL) and multiple-instance learning (MIL) [56] is that SL takes instances as learning objects and aims to predict the class labels of unseen instances, whereas MIL takes bags (a set of instances is termed a bag) as learning objects with the goal of predicting the class labels of unseen bags. In MIL, only the class labels of bags are known in advance. Usually, a bag is classified as positive if it contains at least one positive instance. Otherwise, this bag is classified as negative, which means that all instances in it are negative. Hence, one obvious property of MIL is that there are label ambiguities for instances in positive bags because a positive bag can contain positive and negative instances simultaneously.

The terminology "multiple-instance learning" was originally proposed by Dietterich et al. [18] when they were investigating the drug-activity prediction problem. In their seminal paper, Dietterich et al. considered the problem of predicting whether a candidate drug molecule binds to the target protein. In particular, a molecule can take on many different shapes, and if any of these shapes conforms closely to the structure of the binding site, the candidate molecule binds to the target protein. By treating each molecule as a bag and each shape of a molecule as an instance, drug activity prediction can be considered a typical MIL problem. In addition to drug activity prediction, MIL can be applied in many other domains such as image categorization [13,29,30,31,35,44,57], image retrieval [12,33,54], text classification [4,40], stock selection [36], protein sequence classification [33,46], computer-aided diagnosis [8,22], and security applications [42]. Moreover, with the rapid development of MIL, many representative algorithms such as ID-APR [18], Diverse Density (DD) [35] and its improvement

* Corresponding author. Fax: +86 351 6010029.
E-mail address: jingchai@aliyun.com, jing_chai@hotmail.com (J. Chai).

http://dx.doi.org/10.1016/j.ins.2016.07.029
0020-0255/

EM-DD [55], Bayesian-KNN and Citation-KNN [49], MI-SVM and mi-SVM [4], MI-Kernel [25], MI-Graph and mi-Graph [58], Simple MI [19], and MILES [13] have been proposed to cope with various MIL tasks.

Based on the difference of levels at which an algorithm exploits the required discriminative information from MIL data, existing MIL algorithms can be roughly divided into two types [3]: instance-level algorithms and bag-level algorithms. Instance-level algorithms such as ID-APR, DD, and EM-DD, exploit the discriminative information from the instance level; they initially attempt to obtain instance labels (or information related to instance labels) and then derive the bag labels from instance labels (or related information). In contrast, bag-level algorithms, such as Bayesian-KNN, Citation-KNN, MI-Kernel, MI-Graph, mi-Graph, Simple MI, and MILES, exploit the discriminative information from the bag level; they treat each bag as an inseparable entity and aim to obtain the bag labels directly. Based on the difference in how the discriminative information is exploited, i.e., explicitly or implicitly, bag-level algorithms can be further divided into two subtypes [3]: bag-space algorithms and embedded-space algorithms. Because each bag is a set of vectors with each vector describing an instance, and because usually the numbers of instances in different bags are different, the bag space is a non-vector space. Therefore, bag-space algorithms such as Bayesian-KNN, Citation-KNN, MI-Kernel, MI-Graph, and mi-Graph use a similarity function (e.g., kernel function) or a distance function to evaluate explicitly the similarity or dissimilarity of any two bags. Different from bag-space algorithms, embedded-space algorithms such as Simple MI and MILES initially map each bag to a feature vector that implicitly summarizes the relevant information of the bag and then use supervised classifiers to conduct classifications on the embedded vectors.

Note that although some MIL algorithms obtained very promising classification results, most did not consider the feature-weighting problem; i.e., they treated different features equivalently and adopted all original features in classifications. Usually, the contributions of different features to multiple-instance classifications are different. Some features can contain valuable discriminative information and are helpful to improve MIL performances. These features should be highlighted. In contrast, some features can contain redundant and noisy information only, which makes them useless and even harmful to discriminations. These features should be suppressed.

Feature weighting, which highlights some features and suppresses others, is an important research direction in machine learning. By endowing each feature with a weighting coefficient, we can quantitatively describe the relevance of different features to the learning task (e.g., classification) by comparing the relative magnitudes of different weighting coefficients. Some representative feature-weighting algorithms, e.g., RELIEF [27], I-RELIEF [45], LESS [48], and LMFW [14], had obtained very promising learning performances in different SL domains. Considering that the task of MIL is to separate heterogeneous bags, in this paper, we term features that can improve the discrimination of heterogeneous bags relevant features, and we term features that are useless or harmful to discriminations irrelevant ones, respectively. We hope to improve MIL performances by feature weighting, i.e., by highlighting relevant features and suppressing irrelevant ones. We focus our study on the bag level (more accurately, on the bag space); i.e., we take bags directly as learning objects and attempt to obtain large margins among heterogeneous bags in the weighted feature space via different bag-level distances.

Because we would like to propose three different bag-level multiple-instance feature-weighting algorithms based on different bag-level distances, for convenience of description, we term our work the Bag-level Large Margin Multiple-instance Feature-Weighting (B-LM2FW) framework, and we term different algorithms different realizations of this framework. The three adopted bag-level distances are the minimal Hausdorff (minH) distance [49], the class-to-bag (C2B) distance [50], and the bag-to-bag (B2B) distance, respectively. Thus, the resulting feature-weighting algorithms based on the above three distances can, respectively, be termed LM2FW-minH, LM2FW-C2B, and LM2FW-B2B. Note that of the above three distances, minH and C2B are off-the-shelf, whereas B2B is newly proposed in this paper. Moreover, because the original C2B distance is class specific [50], i.e., different classes can have different metrics, we also treat the minH and B2B distances as class specific; thus, all of the three proposed bag-level feature-weighing algorithms are locally adaptive.

After the introduction of the above three feature-weighting algorithms, we also provide a discussion about how to optimize them. In particular, LM2FW-minH can be optimized by solving a linear programming (LP) problem, whereas both LM2FW-C2B and LM2FW-B2B are non-convex and optimized by the block coordinate descent algorithm; i.e., different types of unknown variables in LM2FW-C2B and LM2FW-B2B are updated alternatively and iteratively. Moreover, according to the properties of the three proposed feature-weighting algorithms, we also provide the corresponding multiple-instance classifiers after their feature-weighting preprocesses.

In summary, the main contributions of this paper are listed below.

(1) We propose a bag-level multiple-instance feature-weighting framework based on the large margin principle, namely, B-LM2FW, which is applicable to different bag-level distances and can provide guidance for future algorithm design.
(2) Enlightened by the off-the-shelf C2B distance, which is able to measure the similarity between any bag and the positive/negative super-bag, we propose a new bag-level distance, B2B, which enables us to measure the similarity between any two bags and is more flexible than C2B.
(3) We adopt three locally adaptive bag-level distances (minH, C2B, and B2B) as examples to demonstrate how to construct the corresponding bag-level feature-weighting algorithms under the B-LM2FW framework.

The remainder of this paper is organized as follows. We introduce several related works and discuss their relationship to our work in Section 2. In Section 3, we introduce useful notation and provide the formulation of the B-LM2FW framework. From Section 4 to Section 6, we substitute three different bag-level distances into the B-LM2FW framework to derive the corresponding bag-level feature-weighting algorithms and discuss how to optimize these algorithms. In Section 7, we analyze