Information Sciences 385–386 (2017) 353–377

Contents lists available at ScienceDirect

Information Sciences

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

Multiple-Instance feature extraction at the bag and instance levels using the maximum trace-difference criterion

Jing Chai[a,*], Bo Chen[b], Fan Liu[a], Zehua Chen[a], Xinghao Ding[c]

[a] College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, PR China
[b] National Lab of Radar Signal Processing, Xidian University, Xi'an 710071, PR China
[c] School of Information Science and Engineering, Xiamen University, Xiamen 361005, PR China

ARTICLE INFO

Article history:
Received 23 September 2015
Revised 30 October 2016
Accepted 31 December 2016
Available online 2 January 2017

Keywords:
Multiple-instance learning
Feature extraction
Dimensionality reduction
Maximum trace-difference

ABSTRACT

Multiple-Instance Learning (MIL) refers to the problem wherein each object is a bag consisting of multiple instances and only the bags' labels are provided. MIL data can contain irrelevant, redundant, and noisy components, which makes feature-extraction preprocessing essential for performance improvement. In this paper, we propose a Multiple-Instance Feature Extraction (MIFE) framework to design algorithms at both the bag and instance levels based on the Maximum Trace-Difference criterion, which simultaneously maximizes between-class scattering and minimizes within-class scattering. MIFE not only treats the existing Multiple-Instance Discriminant Analysis algorithm as an instance-level realization but also enables us to adopt different bag-level distances to design corresponding bag-level algorithms. In particular, we introduce the Class-to-Bag (C2B) and Bag-to-Bag (B2B) distances into the MIFE framework and obtain the MIFE-C2B and MIFE-B2B algorithms, respectively. The experimental results show that both MIFE-C2B and MIFE-B2B obtain competitive classification performance, and MIFE-B2B obtains the best performance on most tested datasets. The dimensionality reduction results show that both MIFE-C2B and MIFE-B2B obtain their best performance with no more than approximately 30% of the original dimensions on most tested datasets.

1. Introduction

Feature extraction copes with the problem of extracting valuable components for a particular learning task via linear or nonlinear transformations [20]. In particular, for Supervised Learning (SL), we hope to extract discriminative components to improve separation between heterogeneous samples (samples from different classes) [4]. In addition to the improvement in discrimination, by feature extraction, we can also reduce the data's dimensionality, save memory space, reduce time complexity in the testing phase, and weaken the disadvantage caused by the curse-of-dimensionality problem.

Multiple-Instance Learning (MIL) is a machine learning branch which has attracted many researchers' interest in the last two decades [8,10,11,14,24,26,43,46]. Different from SL, in MIL, the class label is not attached to a single instance but to a bag (a set of multiple instances), and the goal is to predict the class labels of unseen bags. A bag is labeled positive iff it contains at least one positive instance; otherwise, it is labeled negative. According to this rule, we can deduce that all

* Corresponding author.
E-mail addresses: jingchai@aliyun.com, jing_chai@hotmail.com (J. Chai).

http://dx.doi.org/10.1016/j.ins.2016.12.042
0020-0255/

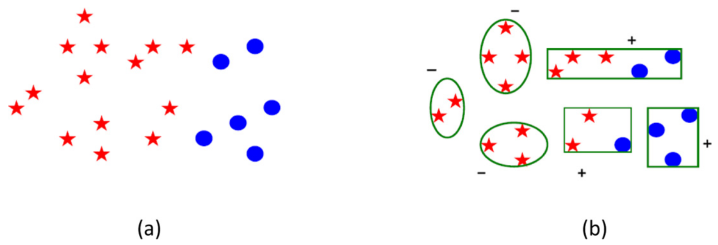

Fig. 1. Illustration of supervised learning and multiple-instance learning. (a) is for supervised learning, with each instance being a basic learning object; (b) is for multiple-instance learning, with each bag consisting of multiple instances being a basic learning object. The notation "+" or "−" around each bag denotes the positive or negative class label of the corresponding bag.

instances in a negative bag are negative. Unfortunately, there are label ambiguities for instances in positive bags, i.e., an instance in a positive bag can be either positive or negative.

We illustrate the difference between SL and MIL in Fig. 1, in which blue circles and red stars denote positive and negative instances, respectively. A rectangular contour containing multiple instances denotes a positive bag, and the notation "+" around the contour denotes the positive bag label. Similarly, an ellipsoidal contour denotes a negative bag, and the notation "−" around it denotes the negative bag label. Fig. 1 shows that the basic learning object of SL is an instance, that of MIL is a bag, and that a bag's label depends upon whether it contains at least one positive instance.

MIL can be applied in, for example, drug activity prediction [14], image annotation [9,27,28,34], text classification [3,33], stock selection [29], protein sequence classification [40], computer aided diagnosis [5,17], and security applications [36]. The MIL data applied in these domains might contain irrelevant, redundant, and noisy components whose existence is useless or even harmful to discriminating heterogeneous bags. Hence, feature extraction preprocessing is essential for MIL. In this paper, we focus on linear feature extraction for MIL, in which "linear" refers to learning a linear transformation and projecting original data onto a linear subspace. The subspace dimensionality is usually lower than that of the original space. Thus, conducting linear feature extraction can lead to dimensionality reduction.

There exist several multiple-instance linear feature extraction algorithms. For example, Sun et al. [39] proposed the Multi-Instance Dimensionality Reduction (MIDR) algorithm based on the criterion of maximizing the posterior probability and optimized MIDR by gradient descent along the tangent space of the orthonormal projection matrix. Ping et al. [32] considered the structural information conveyed by instances within each bag and proposed the algorithm termed Multi-Instance Dimensionality reduction by Learning a mAximum Bag margin Subspace (MidLABS). Kim and Choi [22] proposed the Citation Local Fisher Discriminant Analysis (CLFDA) algorithm to adopt citation and reference information to extract locally discriminative components. Chai et al. [8] coped with the ambiguity problem by seeking positive instances in positive bags and proposed the Multiple-Instance Discriminant Analysis (MIDA) algorithm to use unambiguous instances to extract discriminative components. Among the four above-listed algorithms, MIDR, CLFDA and MIDA were instance-level learning algorithms because they tried to select positive or eliminate negative instances in positive bags to cope with the ambiguity problem and then converted MIL into SL with the help of unambiguous instances. In contrast, MidLABS was a bag-level learning algorithm because it directly measured the similarities among bags and used these similarities to conduct feature extraction.

Different from these algorithms, our work focuses on constructing a Multiple-Instance Feature Extraction (MIFE) framework based on the Maximum Trace-Difference (MTD) criterion [25] to unify the algorithm design at both the bag and instance levels. In particular, MIFE not only treats MIDA as a specific instance-level realization but also enables us to design MIL feature extraction algorithms at the bag level. MTD refers to maximizing the difference between the trace of the between-class scattering matrix and that of the within-class one; MTD also is a variation of the well-known Maximum Trace-Ratio (MTR) criterion [20] used in Linear Discriminant Analysis (LDA) [15]. To show how to design the bag-level algorithms under the MIFE framework, we use the Class-to-Bag (C2B) [42] and Bag-to-Bag (B2B) distances as two examples and substitute them into MIFE to derive the corresponding bag-level algorithms: MIFE-C2B and MIFE-B2B. Different from C2B, which is an off-the-shelf distance, B2B is newly proposed in this paper.

Both MIFE-C2B and MIFE-B2B contain two types of unknown variables: the weighting coefficients for instances in each bag or super-bag and the transformation matrix. The formulation of MIFE-C2B/MIFE-B2B is non-convex w.r.t. all unknown variables. However, by fixing any type of unknown variables, we find that either the formulation is convex w.r.t. the other type of unknown variables or the analytical solutions are accessible. Therefore, it is difficult to optimize all unknown variables simultaneously; instead, we use the block coordinate ascent approach [31] to update them alternately and iteratively.

The contribution of this paper lies in three aspects:

(1) A framework termed MIFE which follows the MTD criterion is proposed to guide the design of MIL feature extraction algorithms at both the bag and instance levels; i.e., we can incorporate different bag level and instance-level distances into this framework to derive the corresponding MIL feature extraction algorithms.
(2) Enlightened by the existing C2B distance, a new bag-level distance, B2B distance, is proposed and adopted in the design of MIL feature extraction algorithms. Compared with C2B, B2B is more flexible, and the resulting MIFE-B2B algorithm outperforms MIFE-C2B in most tested datasets.