

Multiple-Instance Learning: Multiple Feature Selection on Instance Representation

I-Hong Jhuo^{1,2} and D. T. Lee^{1,2}

¹Dept. of CSIE, National Taiwan University, Taipei, Taiwan

²Institute of Information Science, Academia Sinica, Taipei, Taiwan
ihjhuo@gmail.com; dtlee@iee.org

Abstract

In multiple-Instance Learning (MIL), training class labels are attached to sets of *bags* composed of unlabeled instances, and the goal is to deal with classification of bags. Most previous MIL algorithms, which tackle classification problems, consider each instance as a represented feature. Although the algorithms work well in some prediction problems, considering diverse features to represent an instance may provide more significant information for learning task. Moreover, since each instance may be mapped into diverse feature spaces, encountering a large number of irrelevant or redundant features is inevitable. In this paper, we propose a method to select relevant instances and concurrently consider multiple features for each instance, which is termed as *MIL-MFS*. MIL-MFS is based on multiple kernel learning (MKL), and it iteratively selects the fusing multiple features for classifier training. Experimental results show that the MIL-MFS combined with multiple kernel learning can significantly improve the classification performance.

Introduction

Multiple-Instance Learning (MIL) was first introduced by Dietterich97 et al. (1997) when they were investigating the problem of binding ability of a drug activity prediction. In MIL framework, a positive training class (bag) contains at least one positive instance, whereas a negative class (bag) contains only negative instances. Although the labels of training bags are given, the labels of the instances in the bags are unknown. The aim of MIL is to construct a learned classifier to predict the labels of unseen bags. An advantage of multiple-instance learning is that it may provide more information than learning that uses only single-instance representation. For example, we could capture more image information from an image partitioned into several regions, each of which is regarded as an instance, than the whole image represented by a single-instance. When an instance that corresponds to a region contains meaningful information, such as salient cues, the region information may be useful for learning task. However, the useful information may be lost if each region is represented by only one type of feature. Considering the above situation, we concurrently represent each image region with different features, such as shape, texture,

and then propose a multiple kernel learning to select multiple features for multiple-instance learning problem.

Multiple-Instance Learning with Multiple Feature Selection

We propose an approach to MIL, by generating various instances and selecting multiple features, combined with MKL to solve the classification problem. The approach not only captures diverse relevant information but also selects representative instances for each bag. In this paper, each image is regarded as a bag and the regions into which the image is segmented, are regarded as instances, so the image classification problem can be viewed as an MIL problem. We denote B_i^+ as the i^{th} positive bag and the j^{th} instance in that bag as x_{ij}^+ . Each bag B_i^+ consists of l_i^+ instances, x_{ij}^+ , $j = 1, \dots, l_i^+$. Similarly, a negative bag, the j^{th} instance in the bag, the number of instances in the bag are denoted as B_i^- , x_{ij}^- , l_i^- , respectively. The number of positive bags is denoted as m^+ , and the number of negative bags is denoted as m^- . Let $n = m^+ + m^-$ denote the total number of bags.

Bag to Instance Representation

We describe how to represent a bag (image) by using representative instances for MIL on the binary problem; otherwise, multi-class problem can be solved to several binary cases by using the one-against-all strategy. In MIL setting, we first process each positive image into several instances(regions) and extract different features from instances. Since the instances consist of relevant and irrelevant information, we adopt k -means clustering algorithm method to pick up the first k co-occurrences of instances and then construct a bag-level feature map using bag-to-instance similarities :

$$\psi_i(v) = [sim(x_{1,v_1}, B_i), \dots, sim(x_{k^+,v_{k^+}}, B_i)] \quad (1)$$

Here $v = \{v_j \in \{1, \dots, l_i\} | j = 1, \dots, k^+\}$ is the set of evaluated co-occurrence, called representative instances. We use a similarity based feature representation (Chen et al. 2006) that represents each $\psi_i(v)$ as follows

$$sim(x_{j,v_j}, B_i) = \max_{l=1, \dots, l_i} \exp\left(-\frac{d(x_{i,l}, x_{j,v_j})}{\sigma^2}\right) \quad (2)$$

Here σ is a constant and $d(x,y)$ is the distance between two instances. The similarity between a representative instance x_{j,v_j} and a bag B_i is evaluated by the instance and the closest instance in the bag.

Multiple Kernel Learning for Feature Selection

While multiple kernel learning, e.g., Xu et al. (2010), Jhuo et al. (2010), achieves good performance on the classification tasks, we generate diverse features for each instance and develop an MKL algorithm to more effectively fusing multiple features. To extract different instance information for MKL algorithm, we adopt two instance descriptors and their associated distance functions for kernel construction. **Gist**: To obtain texture information, we apply the gist descriptor Oliva et al. (2001). The Euclidean distance is used to construct the RBF kernel. **Pyramid HOG (PHOG)**: The PHOG descriptor Bosch et al. (2007) is a summarized distribution of oriented gradients. The kernel is established by χ^2 distance. The forms of the different visual features may be manifold, so we derive a kernel matrix to describe bags under each visual feature. Suppose we have n training bags $B = \{b_i, y_i\}_{i=1}^n$, where $y_i \in \{+, -\}$. For each bag $b_i \in B$, we have F kinds of diverse image features. For each feature f , we convert pairwise distances between training bags to the f th kernel matrix based on radial basis function (RBF): $K_f(\psi_i, \psi_j) = \exp(-\gamma_f \|\psi_{i,f} - \psi_{j,f}\|^2)$, for $f = 1, 2, \dots, F$. Here γ_f is a positive constant and $\psi_{i,f}$ is evaluated from Eq. (1). Finally, a unified data representation of multiple features, i.e., a collection of the kernel matrices $\{K_f\}_{f=1}^F$, is established. We exploit the *dyadic hypercuts* to generate weak learners from the established kernels $\{K_f\}_{f=1}^F$. For each kernel and each pair of opposite labeled training data, a dyadic hypercut $h_{i,j}^f$ is constructed by a positive bag b_i , a negative bag b_j and a kernel matrix K_f . Specifically, the constructed dyadic hypercut weak learner has the following expression: $h_{i,j}^f(b) = \text{sign}(K_f(b, b_i) - K_f(b, b_j) + \delta_{i,j}^f)$. Here $\delta_{i,j}^f$ is a threshold. With diverse visual features transferring information embedded in each kernel f as a set of weak learners, a strong classifier is obtained via the boosting algorithm. The goal of the learning algorithm is to iteratively select hypercuts $\{h_t\}$ that minimizes weighted error, ϵ_t followed by updating the weight w_{t+1} in the next iteration (Jhuo et al. 2010). The resulting classifier, in the form of a linear combination of weak classifiers, is

$$H(x) = \text{sign}\left(\sum_{i=1}^T \alpha_i h_i(b)\right). \quad (3)$$

Experiments

One of the most successful applications on multiple-instance learning is image classification. The dataset *1000-Image* is a collection of natural scene images from COREL, which contains 10 categories according to object presence. Each image is regarded as a bag, and the ROIS (Region of Interests) in the image are regarded as instances represented by *Gist* and *PHOG* features. We randomly divide the images within each category in half such that one for training and the others for testing. We compare ours with MILES (Chen

